UC Riverside

UC Riverside Electronic Theses and Dissertations

Title

Accurately Calculating the Stability of Molecular Crystal Polymorphs With Improved Intra- and Intermolecular Energies

Permalink

https://escholarship.org/uc/item/77k79893

Author

Greenwell, Chandler

Publication Date

2020

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA
RIVERSIDE

Accurately Calculating the Stability of Molecular Crystal Polymorphs With Improved Intra- and Intermolecular Energies

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

Doctor of Philosophy

in

Chemistry

by

Chandler Scott Greenwell

December 2020

Dissertation Committee:

Dr. Gregory J. O. Beran, Chairperson
Dr. Chia-en Chang
Dr. Chirstopher Switzer

Copyright by
Chandler Scott Greenwell
2020

The Dissertation of Chandler Scott Greenwell is approved:

Committee Chairperson

University of California, Riverside

Acknowledgments

I am grateful to my advisor, Dr. Gregory Beran, who is not only an excellent scientist, but also a superb mentor. He exercised infinite patience, encouragement, and genuine excitement while I made the transition from novice to computational chemist. To my wife, Kaytee Flint, who encourages me every step of the way, and who made this experience a real life. To my parents, Scott and Deborah Greenwell, who created a stimulating environment where I could read and learn about the world, and dinosaurs. To my sister, Andréa Prevo, and her husband, Aaron Prevo, for your love and support. To my nieces and nephew; I look to be an example for all of you. Thank you to my labmates, Dr. Josh Hartman, Dr. Watit Sontising, Dr. Dominique Nocito, Pablo Unzueta, Cameron Cook, and a special thanks to Dr. Jessica McKinley. I would also like to thank Dr. Timothy Herzog, Dr. John Armstrong, Dr. Dale Arney, and Romeo Galang for constant encouragement and way more letters of recommendation than they initially bargained for.

To Monique Greenwell Penman. Everything we do is in loving memory of you.

ABSTRACT OF THE DISSERTATION

Accurately Calculating the Stability of Molecular Crystal Polymorphs With Improved Intra- and Intermolecular Energies

by

Chandler Scott Greenwell

Doctor of Philosophy, Graduate Program in Chemistry
University of California, Riverside, December 2020
Dr. Gregory J. O. Beran, Chairperson

Since the inception of computational chemistry, its practitioners have imagined the ability to predict the three-dimensional form of matter starting from only a two-dimensional representation of a molecule. The science of crystal structure prediction (CSP) starts by predicting rational 3-D arrangements of atoms or molecules. Then, the energy of those arrangements is calculated to determine thermodynamic stability. Complicating the energy determination step is the phenomena of polymorphism whereby a single molecule can adopt multiple solid-state arrangements. The differing physical and chemical properties of polymorphs present an opportunity and a great challenge to chemists and material scientists, and calculating the energy between polymorphs demands modeling intra- and intermolecular interactions with high accuracy.

Two dispersion-corrected variants of Second-Order Møller-Plesset Perturbation Theory (MP2) will be introduced. Compared to high-level benchmark calculations, both methods accurately model both intra- and intermolecular interactions of organic molecules at reasonable computational cost. The methods presented here offer a highly accurate

wavefunction alternative to density functional theory (DFT) for modeling chemical reactions, interaction energies, conformational energies, charge transfer reactions, and nuanced potential energy surfaces.

Plane-wave DFT with a dispersion correction is the current state-of-the-art method for ranking molecular conformational polymorphs; however, there are many systems for which this method does not agree with experimentally determined results. Combining dispersion-corrected MP2 with periodic Hartree-Fock provides high-accuracy polymorph rankings for several systems for which DFT is found to diverge from experiment. Furthermore, the exceptional conformational energies provided by dispersion-corrected MP2 are shown to improve DFT energy rankings simply by replacing the DFT conformational energy. This monomer correction method is applicable to the conformational polymorphs of large, flexible pharmaceuticals like axitinib and galunisertib as well as the organic semiconductors rubrene and perfluororubrene.

Contents

List of Figures **xii**

List of Tables **xix**

1 Introduction **1**
1.1 Polymorphism in Molecular Crystals 2
1.1.1 Crystal structure prediction: searching for structures 4
1.1.2 Crystal structure prediction: ranking structure energies 6
1.2 Electronic Structure Methods 11
1.2.1 Hartree-Fock Theory 11
1.2.2 Correlated Wavefunction Methods 13
1.2.3 Basis Sets 16
1.2.4 Dispersion Corrected Density Functional Theory 17
1.3 Hybrid Many-Body Interaction 21
1.4 Outline of the Dissertation 24

2 Accurate noncovalent interactions via dispersion-corrected second-order Møller-Plesset perturbation theory **28**
2.1 Introduction 28
2.2 Theory 31
2.2.1 MP2D overview 31
2.2.2 Review of the D3 approach 32
2.2.3 UCHF dispersion coefficients 34
2.2.4 Short-range damping 37
2.2.5 Modified C_6 interpolation 40
2.2.6 MP2D Parameterization 41
2.3 Computational Methods 46
2.4 Results and Discussion 47
2.4.1 Energetics 47
2.4.2 Geometries 54
2.5 Conclusions 56

3 Spin-component-scaled and dispersion-corrected second-order Møller-Plesset perturbation theory: A path toward chemical accuracy 59
3.1 Theory . . . 66
3.1.1 SCS-MP2D Energy . . . 66
3.1.2 Empirical Parameter Fitting Procedure . . . 71
3.2 Computational Methods . . . 74
3.3 Results and Discussion . . . 75
3.3.1 Parameter optimization . . . 75
3.3.2 Performance on Benchmark Data Sets . . . 80
3.3.3 Two Challenging Examples . . . 94
3.4 Conclusions . . . 102

4 Overcoming the difficulties of predicting conformational polymorph energetics in molecular crystals via correlated wavefunction methods 106
4.1 Introduction . . . 106
4.2 Theory and Methods . . . 111
4.3 Results and Discussion . . . 119
4.3.1 *o*-Acetamidobenzamide . . . 119
4.3.2 ROY . . . 124
4.3.3 Oxalyl Dihydrazide . . . 131
4.4 Conclusions . . . 139

5 Innacurate conformational energies sill hinder crystal structure prediction in flexible organic molecules 142
5.1 Introduction . . . 142
5.2 Theoretical Approach . . . 145
5.3 Results and Discussion . . . 146
5.4 Conclusions . . . 158

6 Rubrene untwisted: common density functional theory calculations overestimate its deviant tendencies 160
6.1 Introduction . . . 160
6.2 Computational Methods . . . 165
6.3 Results and Discussion . . . 166
6.3.1 Gas-phase twisting energies . . . 166
6.3.2 Theoretical origins of the DFT twisting energy errors . . . 170
6.3.3 Impact on solid state polymorph stabilities . . . 172
6.4 Conclusions . . . 175

7 Conclusions 177

A Supporting Information for "Accurate noncovalent interactions via dispersion-corrected second-order Møller-Plesset perturbation theory" 204
A.1 Additional Computational Details . . . 204
A.1.1 Separate vs Shared Tang-Toennies Damping . . . 204

A.1.2 Benchmark data sets . . . 207
A.1.3 Anthracene PES . . . 207
A.1.4 Helicene PES . . . 209

B Supporting Information for "Spin-component-scaled and dispersion-corrected second-order Møller-Plesset perturbation theory: A path toward chemical accuracy" 210
B.1 Bayesian Parameter Search Algorithm . . . 210
B.2 Comparing the Tang-Toennies damping function from MP2D and SCS-MP2D 213
B.3 Percent Weighted Root Mean Square Errors for Benchmark Data Sets . . . 214
B.4 Complete Anthracene Photodimerization Potential Energy Curve . . . 215

C Supplementary Information for "Overcoming the difficulties of predicting conformational polymorph energetics in molecular crystals via correlated wavefunction methods" 216
C.1 Additional Computational Details . . . 216
C.1.1 DFT Monkhorst-Pack grids and quality of the optimized structures . 216
C.1.2 Basis set dependence for the MP2 models . . . 219
C.1.3 Selection of the Hartree-Fock many-body treatment basis set . . . 221
C.2 *o*-Acetamidobenzamide polymorphs . . . 223
C.2.1 Enthalpies and Gibbs Free Energies . . . 223
C.3 ROY polymorphs . . . 225
C.3.1 Relative polymorph energies . . . 225
C.3.2 Impact of the MP2D dispersion correction . . . 226
C.3.3 Comparison of 0 K and room-temperature structures . . . 227
C.3.4 Omission of the R05 polymorph . . . 228
C.4 Oxalyl Dihydrazide polymorphs . . . 229
C.4.1 Energy decomposition of the polymorph contributions . . . 231

D Supporting Information for "Inaccurate conformational energies still hinder crystal structure prediction in flexible organic molecules." 233
D.1 Computational Methods . . . 233
D.1.1 General methods . . . 233
D.1.2 Periodic DFT Monomer Cell Convergence . . . 235
D.2 ROY . . . 238
D.2.1 Crystal structures . . . 238
D.2.2 Monomer conformational energies . . . 238
D.2.3 Polymorph energies . . . 239
D.3 *o*-Acetamidobenzamide . . . 242
D.3.1 Crystal structures . . . 242
D.3.2 Monomer conformational energies . . . 242
D.3.3 Polymorph energies . . . 242
D.4 Molecule X . . . 245
D.4.1 Crystal structures . . . 245
D.4.2 Monomer conformational energies . . . 245

D.4.3 Polymorph energies 246
D.5 Axitinib 248
D.5.1 Crystal structures 248
D.5.2 Monomer conformational energies 248
D.5.3 Polymorph energies 250
D.6 Galunisertib 251
D.6.1 Crystal structures 251
D.6.2 Monomer conformational energies 251
D.6.3 Polymorph energies 256

E Supporting Information for "Rubrene untwisted: common density functional theory calculations overestimate its deviant tendencies." 257
E.1 Gas-phase twisting energies 257
E.2 Delocalization error 259
E.3 Solid state polymorph energies 260

List of Figures

1.1 (a) Artificial CSP landscape and (b) energy ranking. (a) Theoretically predicted structures are shown in purple, and experimentally verified structures are in teal. Some theoretical structures are found to be lower in energy than form III, which is a commonly observed phenomenon in CSP. This has many causes including neglect of thermodynamic effects and kinetics, the accuracy of the energy ranking method, or they could be undiscovered polymorphs. (b) Here, the experimentally most stable structure, I, is predicted to be less stable than form II. This common error can be an expression of the limitations of the energy ranking method, or due to temperature dependent effects that are neglected when only energy is considered. 10

1.2 (a) A one-dimensional scan about the distance between two benzene molecules as they are separated. CCSD(T) is the reference and taken to be the "true" interaction energies. HF fails to model dispersion, and does not predict an attractive interaction. MP2 overbinds the dimer compared to CCSD(T). A dispersion corrected MP2 method, SCS-MP2D, that will be introduced in chapter 3 is in excellent agreement with CCSD(T) at greatly reduced computational cost. (b) HF is omitted to emphasize the overbinding tendency of uncorrected MP2. 16

2.1 Example of how the secondary short-range damping modifies the effective C-C interatomic separation based on a cutoff radius of $R_{0,ab} = 2.9103$ Å, $r_{cut} = 0.72$, and $w = 0.2$. 40

2.2 Behavior of the MP2D dispersion correction (C_6 only here) without damping, with Tang-Toennies damping, and the double Tang-Toennies/shorter-range damping for two sp^2-hybridized carbon atoms. 45

2.3 Performance of MP2 and other methods on the S66x10 benchmark test set used in fitting the empirical parameters. 48

2.4 Performance of MP2D and other methods for benchmark sets focusing on (a) intermolecular interaction energies, (b) conformational energies, and (c) thermochemistry. MP2C is only defined for the intermolecular interactions. Boxes and whiskers contain 50% and 95% (or 99.9% for part (a)) of the data, respectively. The table lists root-mean-square errors for all methods and sets. 51

2.5 Potential energy basins for covalent anthracene photodimer (near 1.6 Å) and non-covalent π-dimer (near 3.6 Å). B3LYP-D3 employs def2-QZVP basis; all other methods at extrapolated CBS limit. 52
2.6 Overlay of [7]helicene crystal structure[1] with that predicted by MP2 (green), MP2D (red), and B3LYP-D3(BJ) (purple) in the def2-TZVP basis. Single-point energies relative to linear heptacene as a function of the distance R. 54
2.7 Overlay of benzene dimer S4, showing how MP2D (red) reproduces the correct angle between the molecules, unlike MP2 (green) and B3LYP-D3(BJ) (purple). 56

3.1 Left: Performance of one run of the evolutionary SCS-MP2D parameter optimization, plotting the relative root-mean-square error (the objective function) versus the generation number. The figure focuses only on the low-error region, and the horizontal lines indicate the best-performing models in the initial (Gen 0) and final (Gen 14) generations. Right: Radar plot comparing the final SCS-MP2D parameter values (red) to those of the five lowest-error parameter sets (blue), the next five lowest-error models (green), and all other parameter sets (gray) from this particular optimization. 77
3.2 Comparison of the relative RMSEs for several models on benchmark data sets focusing on (a) intermolecular interactions, (b) conformational energies, (c) reaction energies, and (d) the union of all training and testing sets. The asterix indicates that S66x8, SCONF, and DARC were involved in fitting the SCS-MP2D parameters. 83
3.3 Root mean square error of MP2C, MP2D, SCS-MP2D, and ωB97M-V for the different interaction categories in SSI relative to the DW-CCSD(T)-F12 reference values of ref 2. 92
3.4 Relaxed potential energy scan curves for the θ_{thio} dihedral angle of ROY, as computed with several different density functionals and correlated wave function models. The potential energy curves are separated into two panels for easier viewing. 96
3.5 Comparison of relative lattice energies for seven ROY polymorphs against experimentally measured enthalpies. The methods listed with a Δ were used to compute the intramolecular conformational energy correction; the intermolecular part was calculated with periodic B86bPBE-XDM throughout. . 97
3.6 Potential curves along the anthracene photodimerization curve for (a) the photodimer and (b) the π-stacked non-covalent dimer. 99

4.1 The species whose conformational polymorphs are studied here. 109
4.2 Local hydrogen bonding environments for crystalline *o*-acetamidobenzamide. (a) In the α polymorph, the molecule is nearly planar and adopts an intramolecular hydrogen bond, while (b) in the β polymorph the amide and acetamide side chains rotate out of the plane to achieve better intermolecular hydrogen bonds. 118

4.3 Predicted enthalpy difference between the α and β polymorphs of *o*-acetamidobenzamide at 0 K, room temperature, and linearly extrapolated to 423 K. Experimental values were taken from from 121
4.4 Schematic relative enthalpy H and free energy G curves for the α and β polymorphs of *o*-acetamidobenzamide computed using (a) B86bPBE-XDM or (b) fragment-based MP2D/CBS + pHF. The MP2D model reverses the stability ordering, giving results that are consistent with experiment. 122
4.5 Comparison between predicted lattice energies and experimentally measured enthalpies[3] for 8 polymorphs of ROY, all relative to form Y. The PBE-D3, optPBE-vdW, PBE+MBD, and PBE0+MBD results were taken from They omit PO13 and use fully relaxed 0 K unit cells. The other results employ fixed-cell room-temperature structures. 125
4.6 Comparison of the intramolecular conformational energy scan for the key intramolecular dihedral angle in ROY at several levels of theory. Vertical dotted lines indicate the experimental dihedral angles for each polymorph. Energies are relative to the 120° conformation, which corresponds to the CCSD(T) minimum. 127
4.7 Energy decomposition of the room-temperature structure ROY polymorph lattice energies into (a) intramolecular, (b) intermolecular, and (c) total energy contributions. The energies are plotted relative to the most stable Y polymorph in each case. 128
4.8 Crystalline oxalyl dihydrazide exhibits (a) purely intermolecular hydrogen bonding in the α polymorph and (b) a mixture of intra- and intermolecular hydrogen bonding in the other four polymorphs (ϵ form shown here). 131
4.9 Relative stabilities of the oxalyl dihydrazide polymorphs at several levels of theory using both the fully relaxed 0 K structures and the fixed-cell room-temperature structures. For the MP2 methods, the relative stabilities of the β and γ forms differ depending on which structure optimization is used. . . 132
4.10 Relative enthalpy H and free energy G curves for the five polymorphs of oxalyl dihydrazide computed using (a) B86bPBE-XDM or (b) fragment-based MP2/CBS + pHF. The most notable difference between the two curves is that the β and γ forms are monotropically related in the DFT calculations, but enantiotropically related with MP2D (the G curves intersect near 175 K). The data points indicate computed results, while the curves connecting them are schematic. 135
4.11 Energy decomposition for the oxalyl dihydrazide polymorph stabilities into (a) intramolecular (1-body), (b) pairwise intermolecular (2-body), and (c) many-body intermolecular contributions. Note, in the HMBI fragment approach, both MP2D and CCSD(T) employ the same HF many-body treatment. The same 14 kJ/mol energy range is used in all three figures to facilitate comparisons. 138

5.1 The five species whose crystal polymorphs are considered here. Arrows highlight the most important flexible torsion angles. 144

5.2 Overlay of the intramolecular conformations from the ROY polymorphs highlighting how the angle between the two rings generally decreases from yellow to orange to red polymorphs. Relative 0 K lattice energies of the nine known polymorphs of ROY calculated using dispersion corrected DFT and the monomer-corrected energies, compared to experimentally measured enthalpy differences. Many DFT functionals predict the ROY polymorph energies incorrectly, but correcting the B86bPBE-XDM lattice energies with MP2D conformational energies ("+ΔMP2D") performs far better. 146
5.3 Comparison of the intramolecular conformations adopted by the α and β polymorphs of *o*-acetamidobenzamide. The α form is more planar and exhibits an intramolecular hydrogen bond. 148
5.4 (a) Relative 0 K lattice energies versus conformational energies for the low-lying candidate structures of Molecule X before and after monomer correction (see Table D.5 for data on higher-energy structures). (b) Another representation of the relative lattice energies that highlights the impact of the monomer corrections on the stability ordering. MP2D revises the intramolecular conformational energies and stability ordering of the crystal energy landscape considerably. (c) Comparison of the intramolecular conformation of the experimental and vanEijck-3 crystal structures. In contrast to the experimental structure, the amide group in vanEijck-3 is essentially planar relative to the aromatic ring. 149
5.5 Overlay of the monomer conformations and relative 0 K lattice energies for the five axitinib polymorphs compared to the experimentally inferred stability ordering. The thermodynamically stable XLI polymorph of axitinib (red) adopts a folded conformation, in contrast to the extended conformers found in the other polymorphs. B86bPBE-XDM ranks the form XLI incorrectly relative to the other forms, but monomer-correction fixes the qualitative polymorph ordering. 152
5.6 Overlay of the galunisertib monomer conformations. The viewer is looking down the central torsion angle in the middle of the image. The GM conformer is in red. Monomers III, IVb, and X have similar conformations around the central torsion angle, but they differ in the degree of planarity in the pyradine and amide groups relative to the adjacent parts of the molecule. 155
5.7 Relative 0 K lattice energies of nine experimentally known galunisertib polymorphs (higher-energy form I is omitted here) and the putative GM structure compared to the experimental enthalpies that have been measured for several polymorphs. Monomer-correcting galunisertib with MP2D destabilizes the GM structure by ∼1.7 kJ/mol and generally shifts the lattice energies closer to the experimentally measured enthalpies. 156

6.1 Rubrene in its planar (blue) and 60° twisted (gray) conformations. The twisting is defined by the dihedral angle formed from atoms 1–4. 161
6.2 Potential energy scans from 0° to 60° for (a) rubrene, (b) half-fluorinated F_{14}-rubrene, and (c) perfluororubrene. 168

6.3 The twisting energy of the tetracene backbone varies nearly linearly with the fraction of exact exchange included in the functional. Evaluating the energy of the same functionals with HF orbitals instead of the DFT ones (HF-DFT) dramatically reduces the dependence on the fraction of exact exchange and leads to results in far better agreement with SCS-MP2D. 172
6.4 Crystal structures of the (a) perfluororubrene and (b) rubrene polymorphs studied here. 173
6.5 Relative stabilities for the polymorphs of (a) perfluororubrene and (b) rubrene as computed with periodic B86bPBE-XDM and after correcting the monomer conformational energy with the method indicated ("Δ methods"). 174

7.1 (a) ROY CSP landscape generated by a structure search using the CrystalOptimizer code. (b) Refined landscape obtained by optimizing the 50 lowest energy and all known ROY forms with the B86bPBE-XDM functional. (c) SCS-MP2D CSP landscape obtained by monomer correcting the B86bPBE-XDM energies. 181

A.1 Examination of the (a) UCHF dispersion, (b) CKS dispersion, (c) net MP2 dispersion correction, and (d) Errors in the predicted interaction energies for the benzene-uracil π-π dimer from the S66x8 benchmark set. 206

B.1 Example of a Bayesian parameters search: (1) the optimizer is seeded with good parameter sets from the genetic/gradient optimizer, (2) 50 random points are sampled within a predefined sample space, and (3) new points are sampled based on a function constructed from the seeded and random guess points. 211
B.2 Zooming in on the search space shows how the Bayesian optimizer finds new parameter sets that perform well, but do not match the performance of the best parameter sets from the genetic/gradient optimizer. 212
B.3 Examination of the Tang-Toennies damping function f_6 as a function of the interatomic distance, R, for three atom-type interactions. 213
B.4 1-D potential energy scan following the dissociation of an anthracene photodimer to a separated π-stacked dimer 215

C.1 Dependence of the polymorph stabilities for *o*-acetamidobenzamide on the basis set used for the MP2D 1 & 2-body contributions. In all cases, periodic HF/pob-TZVP-rev2 was used for the many-body contribution. 219
C.2 Dependence of the ROY polymorph stabilities on the basis set used for the MP2D 1 & 2-body contributions. In all cases, periodic HF/pob-TZVP-rev2 was used for the many-body contribution. 220
C.3 Three sample oxalyl dihydrazide pentamer clusters. From left to right: α cluster #2, β cluster #3, and δ cluster #2. 222

C.4 Comparison of the predicted enthalpies and Gibbs free energies for the two polymorphs of *o*-acetamidobenzamide as computed with several different models. Data points were computed, but the curves connecting them are merely schematic. 223
C.5 The impact of the MP2D dispersion correction on the relative polymorph energies is fairly modest in ROY, as compared to uncorrected MP2. 226
C.6 The impact of using fully-relaxed (0 K) structures versus fixed-cell relaxations (298 K) on the relative lattice energies. 227
C.7 Temperature dependence of the relative lattice energies (dotted lines) and enthalpies (solid lines) for the oxalyl dihydrazide polymorphs. The qualitative stability ordering is unchanged for both lattice energies and enthalpies. . . 230

D.1 (a) The conformational energies of the five conformers with dipole moments less than 10 D converge faster with the amount of vacuum spacing than do the experimental structure (red) and other structures (gray) with dipole moments above 10 D. (b) The change in the Molecule X monomer conformational energies relative to that of the experimental conformer correlates with the molecular dipole moment. The relative conformational energies converge fastest for conformations whose dipole moment is similar to the experimental one. Each color represents the relative conformational energy change upon increasing the box size by 2.5 Å along each axis. As the box vacuum spacing increases from 12.5 Å to 25 Å, the energy differences become less sensitive to dipole moment. 237
D.2 Comparison of the ROY gas-phase intramolecular conformational energies computed at several levels of theory. The labels R05a/b correspond to the two symmetrically unique monomers in the R05 unit cell. 238
D.3 Relaxed 1-D potential energy scan along the key ROY dihedral angle θ that governs the conformational polymorphism. Dotted vertical lines indicate the value of this dihedral angle adopted in various ROY polymorphs. 239
D.4 0 K electronic energies of the ROY polymorphs relative to Form Y, computed using crystal structures optimized with the fixed experimental lattice parameters, in kJ/mol. These are compared against experimental enthalpies measured at finite temperatures. The B86bPBE-XDM monomer calculations here used a 15 Å vacuum spacing instead of the 20 Å spacing used elsewhere in this work. 241
D.5 Comparison of the Molecule X gas-phase intramolecular conformational energies computed at several levels of theory. The overlay compares the experimental (red) and vanEijck-3 (blue) conformations. 246
D.6 Comparison of the axitinib gas-phase intramolecular conformational energies computed at several levels of theory. The labels IVa and IVb correspond to the two symmetrically unique monomers in the Form IV unit cell. 249

D.7 (top) Two different overlays of the galunisertib monomers. The left one aligns them based on the quinoline ring, while the right one aligns along the pyrazole ring. Molecules in blue have similar values of dihedral angles 1 (left overlay) or 3 (right overlay). (bottom) Summary of the key dihedral angles found for the different galunisertib crystal structures (using DFT-optimized geometries). Color coding from green (most similar) to red (least similar) is used to cluster angles by how much they differ from the GM angles in terms of planarity. 253
D.8 Comparison of the galunisertib gas-phase intramolecular conformational energies computed at several levels of theory. The plot is split into two portions for viewing convenience. The labels IVa, IVb, VIIa, and VIIb correspond to the symmetrically unique monomers in the Form IV and Form VII unit cells. 254
D.9 Relaxed 1-D potential energy curve scans for the three key dihedral angle scans for galunisertib. Vertical lines indicate the values of that particular dihedral angle that are adopted in the crystal structures. Note that the curves exhibit discontinuities due to sudden changes in the other flexible degrees of freedom that occur as the primary coordinate is scanned. 255

E.1 Impact of increasing the fraction of exact exchange in the functional on the (a) Mulliken charges for the carbon atoms and (b) frontier orbital energies for the tetracene backbone. 259

List of Tables

2.1 Sample UCHF and CKS C_6 coefficients for carbon-carbon interactions between two identical carbons for different species/coordination environments as computed according to the D3 scheme. 37
2.2 Optimized values of the five global parameters used in the MP2D method. . 44
2.3 Root-mean-square deviations (rmsd) in pm for the optimized geometries for 10 benzene dimer stationary points relative to reference structures obtained with a mixed DFT/CCSD(T) approach.[4] Structure S4 causes significant problems for MP2 and B3LYP-D3(BJ), so the rmsd is reported with and without S4 included. Counterpoise corrections were not employed. 58

3.1 Comparison of the MP2D parameters from ref 5 and the SCS-MP2D ones determined here. All parameters are dimensionless except for a_2. 76
3.2 Root mean square errors calculated relative to the benchmark reference values (kcal/mol). The asterix (*) indicates data sets that were used to fit SCS-MP2D. Cell color indicates the relative RMSE: dark blue $\leq$ 5%, light blue = 5–10%, orange = 10–25%, light red = 25–50%, and dark red $\geq$ 50%. 81
3.3 Root-mean-square errors for various models in the anion-anion interaction subset of SSI compared to those for the full SSI data set, using the benchmark DW-CCSD(T**)-F12 reference values of ref 2. 93
3.4 Reaction energy ΔE_{rxn} for anthracene photodimerization, 2 $C_{14}H_{10} \longrightarrow (C_{14}H_{10})_2$,[a] in kcal/mol. Root-mean-square errors relative to the CCSD(T) benchmarks are also presented for the photodimer (1.4–2.0 Å) and π dimer (3.2–6.0 Å) basins from Figure 3.6. 103

4.1 Stability of the β o-acetamidobenzamide polymorph relative to the α one, in kJ/mol. Positive values indicate α is more stable than β. See Appendix C Table C.4 for additional details of the lattice energy and phonon contributions. Wavefunction method results are extrapolated to the CBS limit. 119

6.1 Energy differences, $\Delta E = E_{twist} - E_{planar}$, between the planar and twisted forms of rubrene and four derivatives (kJ/mol), and the root-mean-square error relative to the benchmark DLPNO-CCSD(T) results. 169

A.1 Comparison of optimal MP2D parameters when one uses shared or separate Tang-Toennies damping parameters for the UCHF and CKS dispersion energies. The shared damping set here is identical to the one used throughout the main paper. 205
A.2 1-D potential energy scan for the anthracene dimer relative to two non-interacting anthracene monomers (in kcal/mol). The CCSD(T) results are used as the reference value. 208
A.3 1-D potential energy scan for [7]helicene relative to linear heptacene (in kcal/mol). The MP2.5 results are used as the reference value. 209

B.1 Relative root mean square error (%) calculated by dividing the RMSE by the mean absolute value of the reference energies and multiplying by 100. The asterix (*) indicates data sets that were used to fit SCS-MP2D. 214

C.1 Summary of **k**-point grids and quality of the agreement between the predicted and experimental crystal structures. 217
C.2 Comparison of the experimental and predicted S-C-N-C dihedral angle θ for the different ROY polymorphs. Absolute errors between theory and experiment are indicated in parentheses. The key ON polymorph results are highlighted in bold. 218
C.3 Predicted HF many-body energies (in kJ/mol per monomer) and the root-mean-square (RMS) error versus the def2-QZVP basis set results for 15 different clusters cut from the five polymorphs of oxalyl dihydrazide. 222
C.4 Lattice energy differences, vibrational enthalpy contributions, and vibrational Gibbs free energy contributions to the α–β polymorph energy differences in *o*-acetamidobenzamide, in kJ/mol. Positive values indicate α is more stable than β. Summing the lattice energy and vibrational enthalpy contributions gives the results plotted in Figure 3 of the main paper. 224
C.5 Summary of relative lattice energies (in kJ/mol) computed here using the fixed-cell room temperature crystal structures. 225
C.6 Relative lattice energies (in kJ/mol) for the polymorphs of oxalyl dihydrazide using the fully relaxed 0 K structures or fixed-cell room-temperature structures. 229
C.7 Change in the relative lattice energies, vibrational enthalpy contributions, and vibrational Gibbs free energy contributions of oxalyl dihydrazide polymorph stabilities between the 0 K and room-temperature structures, in kJ/mol. Values are listed relative to the α polymorph. 230
C.8 Decomposition of the relative polymorph electronic energies (kJ/mol) for oxalyl dihydrazide 0 K structures into intramolecular (1-body), pairwise intermolecular, and many-body intermolecular contributions. RMS errors are computed against the CCSD(T) benchmarks. All MP2 and CCSD(T) models employ the same periodic HF many-body treatment. 231

D.1 Mean absolute change and maximum change (in parentheses) in the relative B86bPBE-XDM monomer conformational energies (kJ/mol) upon increasing the minimum intermolecular vacuum spacing between periodic images. $\Delta\mu$ indicates the dipole moment range of the monomer conformations, in Debeye from B3LYP/def2-QZVP. 235
D.2 Final ROY polymorph 0 K electronic energies relative to Form Y (kJ/mol). The "MP2D correction" is the correction applied to the B86bPBE-XDM polymorph energies, while "B86bPBE-XDM + ΔMP2D" indicates the final corrected polymorph energies. Finite-temperature experimental enthalpies were taken from Refs 6–9. 240
D.3 Electronic energy difference between the α and β polymorph conformations of *o*-acetamidobenzamide (kJ/mol). 242
D.4 Summary of the electronic energy and enthalpy differences between the α and β polymorphs of acetamidobenzamide (kJ/mol). The 0 K values use fully relaxed crystal structures, while the 298 K values used fixed-cell optimizations of the room-temperature crystal structures. The enthalpy differences are linearly extrapolated to 423 K for comparison with experiment. A positive value here indicates the α polymorph is more stable, while a negative value means the β polymorph is more stable. 244
D.5 Final 0 K electronic energies for molecule X with respect to the experimental structure (kJ/mol). 247
D.6 Details of the axitinib crystal structures used and rmsd15 values comparing the 0 K DFT-optimized and finite-temperature experimental structures. . . 248
D.7 Final 0 K electronic energies for the axitinib polymorphs relative to form XLI (kJ/mol). 250
D.8 Details of the galunisertib crystal structures used and rmsd15 values comparing the 0 K DFT-optimized and finite-temperature experimental structures. 251
D.9 Final 0 K electronic energies for galunisertib polymorphs relative to form IV (kJ/mol). 256

E.1 Root-mean-square errors for the rubrene derivative twisting potential energy scans relative to the benchmark DLPNO-CCSD(T) energies, in kJ/mol. . . 257
E.2 Predicted gas-phase twisting energies (kJ/mol) for rubrene and several derivatives as computed with three additional density functionals and the RMSE relative to the DLPNO-CCSD(T) benchmarks. 258
E.3 Relative polymorph energies for Perfluororubrene and Rubrene, in kJ/mol. The B86bPBE-XDM results were computed with periodic DFT, while the monomer-corrected values represent the B86bPBE-XDM results after correcting the monomer conformational energies with the indicated level of theory (as denoted with a Δ symbol). 260
E.4 Monkhorst-Pack k-point grids used for the periodic DFT calculations. . . . 260

Chapter 1

Introduction

Molecular crystal polymorphism refers to the tendency of some molecules to exhibit multiple crystal forms.[10,11] The chemical space that encompasses the packing variability of molecular crystals is vast and expressed in two different ways: (1) differing conformations of the molecules and (2) different three-dimensional packing arrangements of the molecules. Both are determined by the balance of intramolecular interactions in single molecules and intermolecular interactions between neighboring molecules. Often, highly polymorphic systems will exhibit instances of both types of polymorphism. Interest in polymorphism has proliferated because polymorphs can and often do exhibit different physical properties, and they are ideal test systems for determining structure-property relationships.[12]

Key to accurate modeling of molecular crystals is capturing intramolecular interactions of individual molecules and the intermolecular interactions of dimers and larger clusters of molecules. The large size of molecular crystals prohibits the direct use of the most accurate electronic structure methods, and thus a worthy goal of computational chemistry

is to develop new methods that mimic the accuracy of higher level methods at a fraction of the cost.

This dissertation will describe new techniques for improving the calculation of intra- and intermolecular interactions, and how these techniques can be applied to greatly improve the ranking of molecular crystal polymorphs. The first chapter will define the problem of crystal structure prediction, and summarize some state-of-the-art techniques currently available to computational chemists for studying these systems. Chapter 2 and 3 describe the development and application of two dispersion corrected variants of Second-Order Møller-Plesset Perturbation Theory (MP2). In chapter 4, dispersion corrected MP2 within the Hybrid Many-Body Interaction (HMBI) method is used to rectify with experimental results three examples of molecular crystal polymorphism that have eluded accurate description until now. Chapter 5 details a simple intramolecular correction that can rival HMBI in accuracy at significantly reduced computational cost when inadequately described intramolecular energies are to blame for faulty energy ranking. This advancement allows for highly accurate calculations on systems of pharmaceutical relevance. Finally, the monomer correction is applied to crystals of the organic semiconductor rubrene and rubrene derivatives.

1.1 Polymorphism in Molecular Crystals

The issue of polymorphism is relevant to the pharmaceutical,[13–15] organic semiconductor,[16,17] energetic materials,[18] food science,[19] and pesticide[20] industries, as well as the field of high pressure chemistry.[21] The most notorious instance of polymorphism was

the sudden appearance of a second, unanticipated and more thermodynamically stable form of the HIV drug Ritonavir. Form II of Ritonavir has the unique properties of being more thermodynamically stable, but it is also much harder to crystallize than form I.[14] Form II is much less soluble than form I, which forced a reformulation effort that temporarily removed this lifesaving drug from the market[22] and cost its creator, Abbot (now AbbVie) upwards of 250 million dollars.[23]

The misfortune surrounding Ritonavir greatly increased the effort pharmaceutical companies expend to thoroughly characterize polymorph landscapes during the development phase. However, despite increased scrutiny, a similar setback befell Rotigotine, a drug used for the treatment of Parkinson's and restless leg syndrome. Once again, a more thermodynamically stable form of the drug appeared after formulation during the manufacturing scale-up. The new form was more stable and less soluble than the original, which forced the FDA to order removal of Rotigotine patches from the market from 2008-2012.[15]

Furthermore, the intricacies of patent law as pertains to pharmaceutical polymorphs have caused no shortage of legal battles.[24] Polymorphs play a significant role in patents that allow pharmaceutical companies to extend their market exclusivity rights against generic competitors an average of 6.3 years.[25,26] Indeed, Celgene (now Bristol Myers Squibb) settled for an undisclosed amount with a number of generic drug manufacturers when a joint lawsuit involving polymorphs threatened to prematurely end Celgene's exclusivity of the $9-Billion+/year blockbuster drug Revlimid.[27]

Molecules that exhibit polymorphism are not limited to two forms. In fact, the polymorphic landscapes of some molecules can be exceedingly complicated. Some examples

of prolific polymorph formers include flufenamic acid (9 neat forms),[28] aripiprazole (9 neat forms),[29] nicotinamide (9 neat forms),[30] galunisertib (10 neat forms),[13] and 5-methyl-2-[(2-nitrophenyl)- amino]thiophene-3-carbonitrile, mercifully deemed ROY after its red, orange, and yellow crystals(12 neat forms).[31] The term neat form refers to polymorphs of the base molecule and does not include the often extensive list of hydrates, solvates, and salts that many molecules can form. However, no matter how many polymorphs a molecule exhibits, one of them must be the most stable form thermodynamically at a given temperature, and a hierarchy of stability will proceed from there. Therefore, it is desirable to understand the polymorphic landscape of a molecule with great detail long before decisions are made about formulation or manufacturing. One way to accomplish this is extensive screening, crystallization, and structure determination efforts. Experimental polymorph screening is hindered by the cost of laboratory experiments, by complicated thermodynamic relationships between polymorphs, differing crystallization conditions, and limited quantities of the material available during development. An alternative to experiment, and the focus of this dissertation, is the study of predicting the polymorph landscape of a given molecule using computational chemistry. The key challenge addressed will be accurately predicting the energy differences between polymorphs; a task of immense challenge in computational chemistry due to the large system sizes of molecular crystals, and the high accuracy requirements necessary for accurate ranking.

1.1.1 Crystal structure prediction: searching for structures

The first component of crystal structure prediction (CSP) is a search problem that seeks the most stable arrangement(s) of atoms within a crystal starting from the

2-dimensional representation of a molecule.[21] Although the particulars of each algorithm differ considerably, many determine the energetically favorable conformations of the target molecule, and consider how those conformations can be accommodated by a variety of crystallographic unit cells.

In terms of computational complexity, crystal searching is classified as an NP-hard problem, or non-deterministic polynomial-time hardness problem. According to a mathematical analysis by Oganov, et al., the intrinsic dimensionality for a given molecule is $d^* = 3N + 3 - \kappa$ where κ is the number of correlated dimensions.[21] The intrinsic dimensionality is less than the standard dimensionality for a crystal unit cell ($d = 3N + 3$), and is obtained by enforcing constraints regarding bond distances, and performing local optimizations of target molecules before the full crystal structure search. The estimated number of local minima on a potential energy landscape for the target molecule is $C^* \sim exp(\beta d^*)$. Here, β is a constant that is specific to the target system. In other words, even after intelligently reducing the dimensionality, the number of local minima on the potential energy surface that represents possible crystal structures for a target molecule increases exponentially with respect to the number of atoms and torsional angles. Simply put, for flexible molecules of even modest size, the number of plausible crystal packing structures can easily exceed a million.

Despite the daunting computational challenge posed by structure searching, the last decade has seen tremendous progress generating 3-D packing arrangements of molecules starting from only 2-D representations.[32–34] Searches are capable of identifying experimentally discoverable structures of large and flexible pharmaceutical molecules.[13,35–37] Search

algorithms employ several different strategies such as global searches,[38,39] evolutionary algorithms,[21,40] ab-initio random structure search,[41] and data-driven search algorithms.[42,43]

1.1.2 Crystal structure prediction: ranking structure energies

Once the possible structures have been predicted by the search algorithm, the thermodynamic stability of the structures must be determined. The second component of CSP seeks to rank the energies of candidate structures. This phase typically neglects the effects of temperature and nucleation kinetics.[44] Within this approximation, the structure or structures with the lowest energies are considered to be good candidates for the actual solid-state forms of the target molecule at 0 K. Accurate ranking provides the structures most likely to be crystallized, and the thermodynamic relationships that exist between forms at 0 K. Inaccurate ranking fails to decrease uncertainty regarding what forms should be selected for development and how they can be expected to behave, and in the worse case scenario would actually mislead decision makers.

A structure search can generate hundreds of thousands of potential crystal structures. The energy of each structure must be calculated accurately in order to determine which structures are most likely to exist. Ideally, one would use quantum mechanics, but the computational demands of quantum mechanical calculations make it infeasible to perform these calculations on more than $\sim$100 prime targets. For this reason, the ranking of structures is often divided into two steps:[45] (1) classical force fields are employed to remove energetically uncompetitive structures, then (2) the most promising structures can be refined and reranked using a higher level quantum mechanical calculation. Often the first step is tied directly to the search algorithm. Calculations for the second step must be sig-

nificantly more accurate in order to discern between the small kJ/mol differences between polymorphs or between real and unobserved CSP structures. By far the most common and successful method is plane-wave dispersion corrected density functional theory,[34] but other useful models include those that are built from the theory of intermolecular forces,[46] force fields fitted to relevant computational or experimental data then modified with *ab initio* conformational energies,[47] and potentials fitted from symmetry adapted perturbation theory (SAPT).[48]

More recently, machine learning algorithms have been introduced to the ranking of crystals, and promise to help improve accuracy and reduce computational cost.[44,49,50] There also exists the longstanding observation that CSP predicts many more energetically reasonable structures than are ever realized experimentally. This is so common it has prompted the question "why don't we find more polymorphs?"[51] Even with a perfect energy ranking method, it is likely that this problem will persist due to effects of nucleation and kinetics in crystallization. To this end, recent works have sought to establish a connection between structures and ease of crystallization.[52]

A survey of polymorphic systems from the Cambridge Structural Database (CSD) revealed that more than 50% of polymorphs differ in energy by less than 2 kJ/mol.[53] Traditionally, an electronic structure calculation is considered to achieve chemical accuracy if the result is within 1 kcal/mol of reference calculations or experimental data. The kJ/mol accuracy necessary for accurately calculating the energy differences between polymorphs is very stringent. Furthermore, crystals are often large periodic systems, and molecules of interest to the pharmaceutical industry are often large flexible molecules. Many electronic

structure methods are capable of achieving chemical accuracy; however, they are often very expensive in terms of computation time, and are not applicable to large systems. Often times a computational chemist can choose to have high accuracy results or study large systems, but the energy ranking problem requires both.

The aforementioned dispersion-corrected plane-wave DFT methods have proven to be both computationally tractable and capable of kJ/mol accuracy in many instances of polymorphism. However, the generalized gradient approximation (GGA) density functionals often employed in these methods are susceptible to the delocalization error, and yield errors well in excess of kcal/mol accuracy in some cases. Delocalization error, or self-interaction error, refers to the spurious tendency of electrons to interact with themselves within the framework of DFT.[35,37,54–56] Delocalization error can even result in incorrect structures during geometry optimizations due to spurious proton transfer[57] and predicting monoatomic forms of dihalogens like Br_2 and Cl_2.[58] One potential solution is to use hybrid functionals with a plane-wave basis. However, these models require considerably more computational effort, and are still susceptible to delocalization error.[59] Wavefunction methods provide an alternative to DFT without susceptibility to delocalization error. However, correlated wavefunction methods are often too expensive to apply directly to molecular crystals. There are periodic implementations of MP2, but their accuracy is limited due to the computational necessity of using small basis sets.[38] The energies from wavefunction methods are well known to converge slowly with respect to basis set size. An alternative to periodic methods is the use of fragment-based energy methods. These techniques partition large systems, like molecular crystals, into different regions or by interaction. Two

fragment-based techniques for studying molecular crystal polymoprhs with wavefunction methods will be described in chapter 4 and 5 of this dissertation. Both methods are highly accurate with respect to experimental results, and the latter is highly transferable to systems of pharmaceutical relevance. Critically, the proposed fragment-based methods succeed at reproducing experimentally determined stability orderings for several polymorphic systems for which plane-wave DFT diverges from experiment.

Figure 1.1 demonstrates a typical potential energy landscape for a theoretical crystal structure prediction search, and highlights commonly encountered problems associated with these landscapes. Two common challenges are highlighted. In panel (a) some predicted but unobserved structures are found to be more stable than one of the experimentally verified structures. Causes include neglect of temperature and kinetics, errors inherent to the energy ranking method, or the structures could be real polymorphs that have yet to be observed. In panel (b) the energy ranking method predicts form II to be more stable than the experimentally determined most stable structure, form I. This error could indicate that the relationship between I and II is temperature dependent, or highlight limitations in the energy ranking method itself.

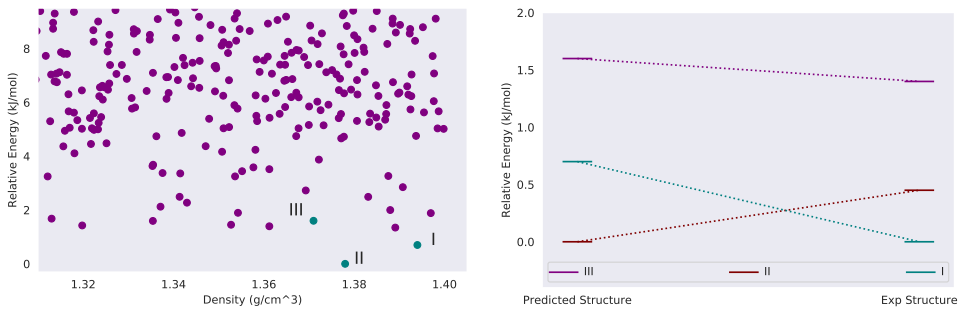


Figure 1.1: (a) Artificial CSP landscape and (b) energy ranking. (a) Theoretically predicted structures are shown in purple, and experimentally verified structures are in teal. Some theoretical structures are found to be lower in energy than form III, which is a commonly observed phenomenon in CSP. This has many causes including neglect of thermodynamic effects and kinetics, the accuracy of the energy ranking method, or they could be undiscovered polymorphs. (b) Here, the experimentally most stable structure, I, is predicted to be less stable than form II. This common error can be an expression of the limitations of the energy ranking method, or due to temperature dependent effects that are neglected when only energy is considered.

1.2 Electronic Structure Methods

Ranking polymorphic energies accurately requires: (1) calculations capable of kJ/mol accuracy, and (2) affordable scaling due to the large system size of many molecular crystals. The list of methods capable of achieving kJ/mol accuracy has grown considerably in recent years; however, the intricacies of these calculations often require computational efforts that grow from $O(N^3) - O(N^7)$ with system size. The following is a brief summary of the methods available to computational chemists for the study of intra- and intermoleculer interactions.

The fundamental equation of quantum mechanics is the Schrödinger equation.

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

Here, $\hat{H}$ is the Hamiltonian operator for a system of N electrons and the corresponding atomic nucei, Ψ is the wavefunction, and E is the Energy corresponding to a given Ψ. In computational chemistry we are often most concerned with the behavior of electrons. This allows us to impose the Born-Oppenheimer approximation, which assumes that nuclei are massive and slow relative to electrons; therefore, they may be treated as stationary potentials within the Hamiltonian operator.

1.2.1 Hartree-Fock Theory

The simplest wavefunction method is the Hartree-Fock (HF) or self-consistent-field (SCF) method. When considering an electron, the explicit positions of the other N-1

electrons are not considered. Rather their charges are averaged over all positions creating a "mean field." The Hamiltonian operator for an N electron system is:

$$\hat{H}_{elec} = \sum_{i}^{N} -\frac{1}{2}\Delta_i^2 - \sum_{i}^{N}\sum_{A}^{M}\frac{Z_A}{r_{iA}} + \sum_{i}^{N}\sum_{j=i+1}^{N}\frac{1}{r_{ij}} \tag{1.2}$$

The first two terms in the electronic Hamiltonian represent the kinetic energy of the electrons, and the potential energy of electron i with respect to nucleus A respectively. The final term, the Coulomb term, considers the potential energy between each electron i and another electron j. Every electron in the system is subject to the movements of every other electron. This creates a "many-body" problem, and prevents us from obtaining exact solutions to the Schrödinger equation for systems with more than 1 electron. It is this final term that necessitates the "mean field" approximation used in Hartree-Fock. In the HF approximation the third term in the Hamiltonian is replaced with a mean field operator for the cumulative effect of the electrons. The Hamiltonian can then be written as:

$$\hat{H}_{elec} = \sum_{i}[h(i) + v^{HF}(i)] \tag{1.3}$$

This approximation means that Hartree-Fock theory will neglect what is known as Fermi-correlation. For this reason, Hartree-Fock is inadequate for the description of noncovalent interactions dominated by long-range electron-electron correlation commonly referred to as dispersion. Correlation refers to the correlated movements of electrons about an atom or molecule caused as electrons simultaneously avoid other electrons and are attracted to atomic nuclei. In the case of dispersion, the electronic fluctuations are from instantaneous excitations, and create a net attraction. Without dispersion, HF cannot predict the

stabilizing $\pi - \pi$ stacking attractions in DNA doubles helices, or explain why gecko's can climb vertical surfaces.[60]

1.2.2 Correlated Wavefunction Methods

The simplest correlated wavefunction method is Second-Order Møller-Plesset Perturbation Theory (MP2). MP2 improves upon HF by describing the effect of two-electron excitations. These two-electron excitations represent the simplest description of the electron correlation responsible for van der Waals dispersion. Amongst correlated wavefunction methods, MP2 distinguishes itself as the least computationally expensive wavefunction method that includes dispersion. Although dispersion extends beyond two-electron excitations, to three-electron, four-electron, and higher, there is a steep increase in computational cost for each type of electronic excitation that is considered.

Perturbation theory partitions a Hamiltonian into a Hamiltonian for which a solution already exits plus a perturbation to that Hamiltonian for which we need to adapt a solution. $\hat{H} = \hat{H}_0 + v$. In Second-Order Møller-Plesset Perturbation Theory, the Hamiltonian we already know how to solve is the HF method, and the perturbation involves the $1/r_{ij}$ term that was approximated in HF. The perturbation can be written as the difference between the mean-field electron-electron repulsion and the true 1/r repulsion:

$$v = \sum_{i<j} \frac{1}{r_{ij}} - \sum_{i} v^{HF}(i) \tag{1.4}$$

According to the perturbation theory expansion, the second-order energy, or MP2 correlation energy is given as[61]:

$$E_0^2 = \sum_{\substack{a<b \\ r<s}} \frac{|\langle\psi_0|\sum_{i<j}\frac{1}{r_{ij}}|\psi_{ab}^{rs}\rangle|^2}{\epsilon_a + \epsilon_b - \epsilon_r - \epsilon_s} \tag{1.5}$$

The summation is over all ground state orbitals a and b, and all virtual orbitals r and s. The ϵ terms in the denominator are the Hartree-Fock energies for each orbital. ψ_0 is the ground state wavefunction, and ψ_{ab}^{rs} is the doubly excited wavefunction. In this way, the MP2 energy is simply the Hartree-Fock energy plus the MP2 correlation energy $E_{MP2} = E_{HF} + E_{correlation}$.

Furthermore, since the MP2 correlation energy is attributed to two-electron excitations, we can divide the correlation energy into contributions from same-spin and opposite-spin electronic excitations: $E_{MP2} = E_{HF} + E_{ss,correlation} + E_{os,correlation}$. Grimme realized that the same-spin and opposite-spin energies can be empirically scaled.[62] This spin-component-scaled (SCS-MP2) method was found to provide significantly improved accuracy for thermochemical reactions at no additional computational cost. However, Grimme's spin-component-coefficients transfer poorly to interaction energies. By optimizing the coefficients against interaction energy data, DiStasio and Head-Gordon obtained coefficients that lead to excellent performance on interaction energies (SCS-MI-MP2).[63] Spin-component scaling provides an easy way to substantially improve MP2 calculations without increasing the computational cost; however, coefficients that are ideal for one family of chemical problems rarely transfer well to different problems. This issue of spin-component transferability will be covered in detail in chapter 3.

Theoretically it is possible to obtain exact solutions to the Schrödinger equation with the full configuration interaction (CI) or full coupled cluster (CC) method. However, these calculations are only tractable for very small systems. The gold standard of quantum chemistry, in terms of accuracy and it's applicability to small organic molecules, is the Coupled Cluster Singles, Doubles, and Perturbative Triples method (CCSD(T)).[64,65] CCSD(T) has earned this moniker due to it's high accuracy when compared to experimental data, or higher level benchmark energies.[66,67] The parent method of CCSD(T), CCSDT, where the triples are treated exactly scales as $O(N^8)$, CCSD(T) scales as $O(N^7)$, and CCSD scales as $O(N^6)$. Simply doubling the size of a molecule treated with CCSD, CCSD(T), or CCSDT would increase the computation time by 64x, 128x, and 256x respectively. The unfortunate scaling of CCSD(T) cannot be overcome simply by increasing the available disk and memory resources either. The computational demands exceed the physical limitations of modern computer hardware, and even hardware that could exist 12 years from now assuming that Moore's Law continues to hold true.[68] However, a linear-scaling implementation of CC theory that uses a series of mathematical approximations called domain-based local pair-natural orbital theory (DLPNO) agrees with canonical CCSD and CCSD(T) to within 0.5 kcal/mol accuracy for a test set of medium-sized organic molecules.[69] The DLPNO-CC methods scale linearly with system size, thereby allowing approximate CC calculations on large chemical systems. However, DLPNO calculations require an SCF calculation and MP2 correlation calculation, so they are still much more computationally intensive than HF or MP2. Figure 1.2 demonstrates how different wavefunction methods calculate the interaction energy for a $\pi - \pi$ stacked benzene dimer as the distance between the molecules

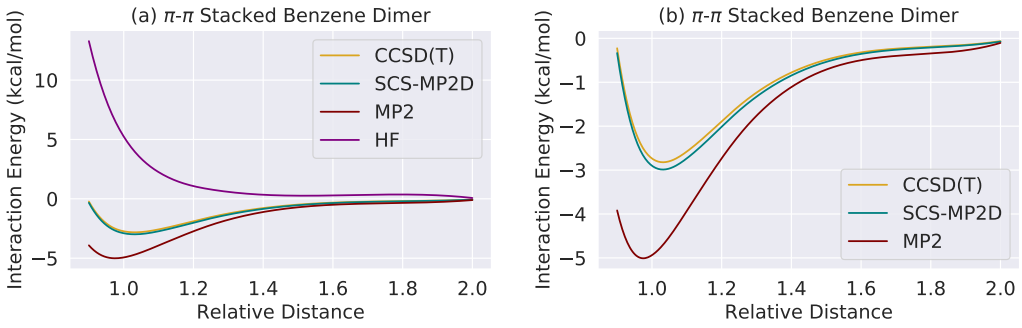


Figure 1.2: (a) A one-dimensional scan about the distance between two benzene molecules as they are separated. CCSD(T) is the reference and taken to be the "true" interaction energies. HF fails to model dispersion, and does not predict an attractive interaction. MP2 overbinds the dimer compared to CCSD(T). A dispersion corrected MP2 method, SCS-MP2D, that will be introduced in chapter 3 is in excellent agreement with CCSD(T) at greatly reduced computational cost. (b) HF is omitted to emphasize the overbinding tendency of uncorrected MP2.

is incrementally increased. Dispersion due to electron-electron correlation is known to be the dominant attractive interaction for this system.

1.2.3 Basis Sets

Basis sets are the set of mathematical functions used to construct the molecular orbitals that are optimized self-consistently during an electronic structure calculation. The basis set size used for the calculation is of the utmost importance for benchmark level calculations, and a complete basis set is required. Basis set size refers to the number of atomic orbitals that are used to construct the molecular wavefunctions during an electronic struc-

ture calculation. Technically, a complete basis set would be infinitely large, but in practice, for an organic molecule with < 50 atoms, a basis set with several thousand functions is often sufficient for convergence. Unfortunately, as basis set size increases, so does the computation time, which for CCSD(T) is $O(N^7)$. However, it is possible to reach near CCSD(T) accuracy in a complete basis set by correcting MP2 energies with CCSD(T) correlation energies computed in a smaller basis set.[64] Furthermore, it is possible to extrapolate smaller basis sets to the basis set limit.[70]

$$E^{CCSD(T)} = E^{HF}_{large\ basis} + E^{MP2,correlation}_{large\ basis} + \delta E^{CCSD(T),correlation}_{small\ basis} \tag{1.6}$$

The $\delta E^{CCSD(T),correlation}_{small\ basis}$ term is obtained by performing a CCSD(T) and MP2 calculation in a smaller basis set. Although neither method is converged with respect to the basis set, the difference in energy between the two methods converges much faster than the base methods themselves; therefore, the energy difference is nearly converged even when computed in the smaller basis set.

$$\delta E^{CCSD(T),correlation}_{small\ basis} = E^{CCSD(T)}_{small\ basis} - E^{MP2}_{small\ basis} \tag{1.7}$$

1.2.4 Dispersion Corrected Density Functional Theory

Rather than solve the Schrödinger equation by considering multi-dimensional wavefunctions, it is also possible to consider the electron density, ρ, which is a 3-dimensional property. Here, the ground state energy of the system is a functional of the electronic density, ρ.[71]

$$E_{DFT}[\rho] = T_S[\rho] + E_{ne}[\rho] + J[\rho] + E_{xc}[\rho] \tag{1.8}$$

Here, T_S is the kinetic energy for non-interacting electrons, E_{ne} is the energy for attractions between electrons and nuclei, J is a Coulomb repulsion term, and E_{xc} is the exchange-correlation functional which contains the remaining kinetic energy terms and the exchange energy terms. The exchange-correlation term is the focus of much of the research into DFT functionals, and the different approximations for treating it lead to many different types of density functionals. In practice, Kohn-Sham density functional theory is often used, which requires calculating the electron density from orbitals.

$$\rho_{approx} = \sum_{i=1}^{N_{elec}} |\phi_i|^2 \tag{1.9}$$

This approach makes Kohn-Sham density functional theory similar to Hartree-Fock in terms of the kinetic and potential energy terms that must be considered, but density dependent functionals can allow significantly improved accuracy at nearly identical computational cost. For this reason, DFT is probably the most used quantum mechanical method in computational chemistry. However, like HF, typical Kohn-Sham density functionals neglect long-range electron correlation, and in addition, they neglect the exact treatment of electron exchange that is obtained even with the simple Hartree-Fock method. One also sacrifices the systematic improvement of accuracy that is a natural feature of wavefunction methods.

To rectify these shortcomings, much effort has been expended to improve the DFT description of dispersion and exact exchange. Including dispersion corrections in

DFT takes 3 main forms: (1) semi-empirical post-hoc dispersion corrections (D corrections and TS),[72–75] (2) post-hoc dispersion corrections that are computed ab-initio (XDM and MBD),[76,77] and density functionals that include dispersion (vdW-DF and VV10).[78,79] Although DFT is not systematically improvable like wavefunction methods, there does exist a hierarchy or "Jacob's Ladder" that trends toward increased accuracy. The final term of the DFT energy functional contains a term for exact exchange and correlation of electrons. The exchange-correlation term contains all the complicated many-body terms for the system. Each rung of Jacob's Ladder includes additional terms to better describe exact exchange. The rungs of the ladder are: (1) local density approximations, (2) generalized gradient approximations, (3) meta-generalized gradient approximations, (4) hybrid density functionals, and (5) double hybrid functionals. Roughly, each rung of the ladder can be distinguished from previous rungs as follows.

- local density approximation (LDA): treats the electron density as homogenous
- generalized gradient approximation (GGA): introduces the gradient, or first derivative of the density, which accounts for heterogeneity of the electron density.
- meta-generalized gradient approximation (meta-GGA): includes the second derivative of the electron density or Laplacian.
- hybrid functionals: incorporate Hartree-Fock terms for describing exact exchange. These methods are hybrids of DFT and HF.
- double hybrid density functionals: Incorporate Hartree-Fock terms, and MP2 correlation terms to correct the description of dispersion.

Each rung of the ladder can provide increased computational accuracy, but also increases the computational cost. Rung 4 functionals are as expensive as HF, and rung 5 functionals are similar in computation time to MP2. The details of the D3 dispersion correction will be discussed in Chapter 2. D3 is a versatile dispersion correction scheme because it can be adapted to functionals on any rung of Jacob's Ladder. Basic details and popular top performing rung 4 and rung 5 functionals are discussed in the next two sections.

Rung 4 hybrid functionals

One of the first useful hybrid functionals, and certainly the most pervasive is B3LYP.[80] In fact, the original paper to describe this functional is one of the most cited of all time. That it was introduced in 1992 and still finds widespread application attests to its usefulness. The modern implementation is often paired with a D3-like dispersion correction for geometry optimizations and energy calculations. However, B3LYP can exhibit significant errors for systems with complicated electron correlation, charge transfer reactions, and reaction thermochemistry. Modern hybrid functionals seek to address these weaknesses by incorporating additional physical insight into the functionals. One of the most successful, ωB97X-V, improves upon the treatment of exact exchange via range separation of the XC terms, and incorporates the VV10 method for treatment of nonlocal correlation.[81] This prescription leads to excellent performance on the GMTKN55 dataset that includes non-covalent interactions, thermochemistry reactions, and molecular conformations.[67] Further improvements in chemical accuracy are obtained by incorporating nonlocal correlation and range-separation of the exchange-correlation terms within a meta-GGA functional, ωB97M-V.[82]

Rung 5 double hybrid functionals

Double hybrid density functionals incorporate Hartree-Fock for the treatment of exact exchange, and MP2 correlation energy for the treatment of electron-electron correlation. By including correlation at the MP2 level of theory, these functionals surrender the computational advantage offered by DFT; however, they do so to obtain a significant increase in accuracy. These methods have proven capable of achieving near CCSD(T) accuracy at significantly decreased $O(N^5)$ computational cost. The hybridization of DFT and MP2 creates an opportunity to take advantage of techniques that have been used to improve the accuracy of both methods. The highly accurate DSD functionals incorporate D3 and D4 dispersion corrections for improved treatment of electron correlation, and spin-component scaling of the MP2 correlation energy.[62,63] The resulting revDSD-PBEP86-D4 and revDSD-BLYP-D4 functionals represent two of the most accurate functionals ever developed based upon their performance on the GMTKN55 benchmark dataset.[83,84] In terms of chemical accuracy, ωB97M(2)[85] is currently the state-of-the-art density functional. It is the double hybrid analogue of the ωB97X-V and ωB97M-V functionals.

1.3 Hybrid Many-Body Interaction

The Hybrid Many-Body Interaction (HMBI) is a mathematical framework that decomposes the interactions of molecular crystals into their constituent interactions.[86,87] HMBI works by partitioning the energy of a cluster or periodic system based on a many-body expansion of the interacting terms. In this way HMBI distinguishes itself from

previous fragment QM/MM models that partitioned systems based on regions rather than by interaction type.

$$E_{total} = \sum_{i} E_i + \sum_{ij} \Delta^2 E_i + \sum_{ijk} \Delta^3 E_{ijk} + ... \tag{1.10}$$

Here, E_i is the energy of monomer i, $\Delta^2 E_{ij}$ is the interaction energy between monomers i and j, and $\Delta^3 E_{ijk}$ is the 3-body interaction energy between monomers i, j, and k. In this way, the energy can be expanded to explicitly include any interactions deemed necessary. However, in practice, only the monomers and dimers are treated explicitly, and the higher order terms are calculated with a less expensive periodic method.

$$E_{total}^{HMBI} = E_{1-body}^{QM} + E_{2-body,SR}^{QM} + E_{2-body,LR}^{MM} + E_{many-body}^{MM} \tag{1.11}$$

Here, QM stands for quantum mechanics. Any sufficiently accurate correlated wavefunction method, or density functional can be used for the high level quantum mechanical calculations. MM stands for molecular mechanics, and in the past a classical or polarizable force-field was used for the MM terms, but in this work the MM term is replaced with periodic Hartree-Fock. Formally, this substitution makes HMBI equivalent to the method of increments.[88–90] The formulation of HMBI found to be sufficient for modeling molecular crystal polymorphs uses dispersion corrected MP2 (MP2D or SCS-MP2D) for the quantum mechanical calculations of monomers and short-range dimers. These methods will be described in chapter 2 and 3. Periodic Hartree-Fock (pHF) is used for the long-range dimer and many-body interactions.

$$E^{HMBI} = E^{pHF}_{many-body} + \sum_i (E_i^{MP2D} - E_i^{HF}) + \sum_{ij} f_{ij}^{damp} (\Delta^2 E_{ij}^{MP2D} - \Delta^2 E_{ij}^{HF}) + \frac{1}{2} \sum_i \sum_k^{images} f_{ik}^{damp} (\Delta^2 E_{ik}^{MP2D} - \Delta^2 E_{ik}^{pHF}) \tag{1.12}$$

In the expression, i and j are indices of molecules in the central unit cell. The index k runs over the periodic image molecules according to a cutoff distance defined with respect to monomer i. E_i is the energy of monomer i, $\Delta^2 E_{ij}$ is the interaction energy between monomer i and monomer j, and $\Delta^2 E_{ik}$ is the interaction energy between monomer i and monomer k. $E_{many-body}$ is the total periodic energy for the entire system. Finally, thef_{ij}^{damp} term is a damping function that ensures smooth transitions between the high level quantum mechanical region and lower level QM or MM region.[91]

$$f_{ij}^{damp}(R) = \frac{1}{1 + e^{2\frac{|r_1 - r_0|}{(r_1 - R)} - \frac{|r_1 - r_0|}{(R - r_0)}}} \tag{1.13}$$

R is the shortest intermolecular distance between monomer i and monomer j. A dimer is treated with dispersion corrected MP2 if the shortest intermolecular distance is less than r_1. A dimer is treated with Hartree-Fock if the shortest intermolecular distance is greater than r_0. The dimer interaction energy is calculated as a linear combination of the pHF and QM energies if the shortest intermolecular distance falls within r_1 and r_0.

The ability to control the level of theory with which certain interactions are modeled allows for highly accurate calculations on large, periodic systems. Furthermore, as electronic structure methods and force-fields improve, state of the art methods can easily be incorporated within the HMBI framework. HMBI has been used to predict the thermodynamic stability of the 5 polymorphs of Oxalyl dihydrazide.[56,92] In this case, the

energy decomposition proved to be highly accurate and amended itself to revealing the extent to which DFT methods may benefit from fortuitous error cancellation. HMBI provides the requisite accuracy to distinguish between the two nearly degenerate polymorphs of aspirin.[93] Finally, HMBI provides accurate polymorph energy rankings for systems like ortho-acetamidobenzamide and ROY that prove difficult for DFT methods due to the delocalization error.[56]

Beyond energies, the partitioning of interactions in HMBI finds practical applications in many aspects of theoretical solid-state chemistry. It is possible to obtain optimized crystal geometries.[94] Spectroscopic properties of crystals, like Raman and nuclear magnetic resonance can be simulated by taking derivatives of the HMBI energy terms.[95–99] Thermodynamic (enthalpy and entropy), and physical properties (bulk modulus) can be predicted both at finite temperature and by simulating the effects of temperature.[100–102] By including the effects of temperature and pressue, it is even possible to predict phase diagram behavior of small molecules like methane, carbon dioxide, and molecular nitrogen.[99,103–105]

1.4 Outline of the Dissertation

MP2 theory is the least expensive wavefunction method that treats electron-electron correlation. However, MP2 is well known to overstabilize systems where dispersion is the key noncovalent interaction. This error is exemplified by a dimer interaction energy that is nearly double that of the CCSD(T) benchmark for a $\pi - \pi$ stacked benzene dimer. This error manifests itself in many systems of biological interest and condensed phase systems. In chapter 2, a dispersion-corrected variant of MP2 is proposed. The MP2D method

is an amalgam of a previously proposed but computationally more expensive dispersion corrected MP2 method (MP2C), and Grimme's D3 dispersion correction. Significantly, the MP2D method achieves higher accuracy than MP2 with no additional computational cost.

Historically, spin-component scaling has provided a computationally "free" method for improving the accuracy of MP2. However, the empirical nature of this method means that spin-component scaling constants only work well for specific chemical problems. In an ideal world, a single set of spin coefficients would be widely transferable, but until now this has not been the case. In Chapter 3 a spin-component scaled version of MP2D (SCS-MP2D) is presented that offers improved accuracy over the original MP2D method, and is competitive with some of the very best hybrid and double hybrid functionals in terms of accuracy and computational cost. Most importantly, SCS-MP2D does not increase the computational cost of MP2, and the spin coefficients prove highly transferable to thermochemistry reactions, interaction energies, and conformational changes. SCS-MP2D provides excellent agreement with CCSD(T) for difficult potential energy surfaces like those for the dissociation of an anthracene photodimer to a $\pi - \pi$-stack dimer, and a 1-D scan about the key dihedral angle in ROY.

The state-of-the-art method for ranking polymorphs is plane-wave DFT with a dispersion correction. These methods are applicable to large periodic systems, and have demonstrated themselves to be highly accurate in many important instances. However, there exist many systems for which DFT does not provide polymorph energy rankings that are in good agreement with established experimental results. Wavefunction methods provide many benefits that compensate for the deficiencies seen in DFT methods, but are

far too expensive to be applied directly to periodic systems of interest to pharmaceutical companies or other industries. The hybrid many-body interaction method provides an alternative to plane-wave DFT for calculations on large periodic systems. In chapter 4, dispersion corrected MP2 is used within the HMBI model in order to calculate the energy of several polymorphic systems for which DFT does not agree with experimental results. The MP2D/HMBI method is found to yield excellent agreement compared to experiment for all three systems. Furthermore, each monomer and dimer can be run individually and in parallel.

Although MP2D/HMBI has proved to be a highly accurate energy ranking method for deviant polymorph energy ranking problems, the high level quantum mechanical calculations that must be performed on 50-100 dimer interactions are cost prohibitive for organic molecules of even modest size. For example, 21,000 CPU hours/polymorph were consumed for the 54-atom dimers of ROY. Modest pharmaceutical molecules can easily exceed 50 atoms. Not only would the 100-atom dimers require an enormous computational effort, recent work has shown that the perturbation expressions in MP2 may diverge for systems with $>$ 100 atoms.[106] Aside from providing high accuracy energies for periodic systems, HMBI is also useful for analyzing the contributions of each energy term within the fragment energy decomposition. Solving polymorph ranking problems with MP2D/HMBI revealed that many of the inaccurate rankings at the DFT level can be attributed to poor monomer conformational energies. This insight revealed that an HMBI-like approach that only replaces DFT monomer energies, rather than dimer and monomer energies, can provide highly accurate polymorph energy rankings at drastically reduced computational cost.

Chapter 5 describes this monomer correction technique for molecular polymorph energy ranking. The monomer correction is used to provide high accuracy energy rankings for 5 difficult examples of molecular polymorphism, and proves to be computationally tractable for systems of pharmaceutical relevance.

The pharmaceutical industry is not the only industry that can benefit or suffer from polymorphism. Despite 66 years of intense study, five new polymorphs of the B-vitamin nicotinamide were reported in November of 2020.[30] Polymorphism is also prevalent in the semiconductor industry; Rubrene has three known polymorphs,[107] pentacene has four,[17] and perfluorinated rubrene has three polymorphs.[108,109] The potential exists for CSP to assist in the search for new organic semiconductors or reveal undiscovered polymorphic forms of known organic semiconducting molecules. CSP provides the three-D structural information that is useful for determining the type of physical properties that a potential material might exhibit. Chapter 6 reveals how polymorph energy rankings at the DFT level are also susceptible to delocalization error for the perfluorinated derivatives of Rubrene. Monomer correction flips the DFT stability ordering. Furthermore, delocalization error consistently overstabilizes the twisted conformations of rubrene derivatives by about 3 kJ/mol with respect to DLPNO-CCSD(T) benchmarks. MP2D and SCS-MP2D are found to be in excellent agreement with the benchmark calculations.

Chapter 2

Accurate noncovalent interactions via dispersion-corrected second-order Møller-Plesset perturbation theory

2.1 Introduction

Non-covalent interactions govern protein folding, chemistry in solution, molecular crystal polymorphism, and many other important phenomena. Simulating such systems requires theoretical models capable of accurately reproducing the often delicate balances among the different types of non-covalent interactions both within molecules and between them. Large-basis coupled cluster methods can achieve this accuracy for small

systems,[64] but they are computationally prohibitive for larger ones. Dispersion-corrected density functional theory (DFT) models provide a much more affordable option,[110] though DFT cannot always provide the requisite accuracy due to self-interaction error and other inherent limitations in the functionals.[111] Here, we report a new, computationally practical dispersion-corrected second-order Møller-Plesset perturbation theory (MP2) model which provides high-quality energetics and structures in systems where non-covalent interactions are important, filling an important gap between DFT and higher-level techniques.

With formal computational cost scaling with the fifth power with system size (though this scaling can be reduced via Laplace transform, local correlation models, etc[112]), MP2 provides a valuable and computationally affordable alternative to DFT for organic systems, but it has well-known problems describing van der Waals dispersion interactions. It overestimates the interaction energy in the π-stacked benzene dimer by a factor of two, for example.[113] From the perspective of intermolecular perturbation theory, this deficiency in MP2 stems from its uncoupled Hartree-Fock (UCHF) treatment of intermolecular dispersion[114,115] which approximates the excited states and excitation energies that contribute to the dispersion energy using unrelaxed ground-state Hartree-Fock orbitals.

Various models empirically scale the same-spin and opposite-spin correlation components in MP2 to improve its performance,[63,116–120] though the optimal parameters often vary with the nature of the chemistry being modeled. The very successful non-empirical MP2C method[115,121] replaces the problematic UCHF dispersion with an improved coupled Kohn-Sham (CKS) treatment of dispersion,

$$E^{MP2C} = E^{MP2} - E^{UCHF}_{disp} + E^{CKS}_{disp} \tag{2.1}$$

effectively using time-dependent density functional theory to obtain an improved description of the excited states. The excellent performance of MP2C for intermolecular interactions earned it "the bronze-standard of quantum chemistry" moniker.[122] Unfortunately, MP2C has two major limitations. First, the dispersion correction is derived from intermolecular perturbation theory and is not defined for intramolecular interactions. Intramolecular dispersion can be crucial in larger molecules. Second, MP2C is not currently used for structure optimization due to the complexity of its analytical nuclear gradients.

Here, we combine the ideas of MP2C with Grimme's DFT-D3 dispersion correction[123] to develop a new dispersion-corrected MP2D model. Recasting MP2C in terms of atom-centered two-body dispersion coefficients offers clear advantages. Atomic dispersion coefficients can be applied to both intra- and intermolecular atom-atom interactions. Furthermore, both the energy and analytical gradients of the dispersion correction can be computed with trivial computational cost. On the other hand, it introduces some empiricism to the model in the form of five global parameters. MP2D is also similar to the MP2+ΔvdW model,[124] but it improves upon that model in several important ways. It includes both the C_6 and C_8 terms, instead of only C_6 like MP2+ΔvdW. More significantly, MP2D solves the problem of how to determine the atomic C_6 dispersion coefficients for different chemical environments by adopting the D3 dispersion correction approach.[123]

The following sections present the MP2D model, including how the dispersion coefficients were obtained, modifications to the short-range damping necessary to treat both covalent- and non-covalent chemistry, several minor changes to the D3 procedure, and the strategy used to ensure physically appropriate parameters were obtained. We then demon-

strate that MP2D performs very well across thousands of benchmark energies, including intermolecular interactions, conformation energies, and thermochemistry. We examine in detail the performance of MP2D on the challenging anthracene photodimerization, in which inter- and intramolecular interactions compete strongly. Finally, we study several examples of geometry optimization where dispersion effects play a major role. Throughout these tests, MP2D significantly improves MP2 in cases where van der Waals dispersion is important, and it does so with negligible additional computational cost. At the same time, the MP2D dispersion correction has little impact on MP2 in cases where dispersion does not contribute significantly.

2.2 Theory

2.2.1 MP2D overview

MP2D corrects MP2 by subtracting out the pairwise interatomic UCHF dispersion energy and replacing it with the equivalent contribution calculated at the CKS level of theory,

$$E^{MP2D} = E^{MP2} - \tilde{E}_{disp}^{UCHF} + \tilde{E}_{disp}^{CKS} \tag{2.2}$$

where

$$\tilde{E}_{disp} = s_6 \sum_{a,b} f_6(R_{AB}) \frac{C_{6,ab}}{R_{AB}^6} + s_8 \sum_{a,b} f_8(R_{AB}) \frac{C_{8,ab}}{R_{AB}^8} \tag{2.3}$$

In these expressions, C_6 and C_8 are the interatomic two-body dispersion coefficients calculated at either the UCHF or CKS levels of theory, R_{AB} is the distance between atoms A and B, f_n are short-range damping functions, and s_n are empirical scaling factors. One might further augment MP2D with a 3-body dispersion term,[125] since those contributions

are missing in MP2[126] and can become significant in large systems,[127,128] but that is not done here.

MP2D adopts Grimme's D3 model[123] to compute the UCHF and CKS dispersion contributions. In fact, MP2D uses Grimme's existing D3 C_6 coefficients for the CKS dispersion energy. New UCHF dispersion coefficients are computed here. Several other minor modifications are made to the D3 approach with regard to the damping at short non-covalent and covalent distances and the evaluation of the continuous coordination approach, as described below.

2.2.2 Review of the D3 approach

It is worthwhile to review Grimme's D3 approach briefly before discussing the MP2D-specific changes. D3 computes frequency-dependent dipole-dipole polarizabilities $\alpha(i\omega)$ for a series of different hydrides with different coordination numbers (CNs). For carbon, for instance, it computes them for C ($CN = 0$), CH ($CN \approx 1$), C_2H_2 ($CN \approx 2$), C_2H_4 ($CN \approx 3$), and C_2H_6 ($CN \approx 4$). The C_6 dispersion coefficients for all possible pairwise combinations of atoms and coordination numbers are calculated via Casimir-Polder integration with these polarizabilities after subtracting out the approximate hydrogen contribution,

$$C_{6,ab}(CN_i^a, CN_j^b) = \frac{3}{\pi}\int_0^\infty d\omega \frac{1}{m}\left[\alpha^{A_mH_n}(i\omega) - \frac{n}{2}\alpha^{H_2}(i\omega)\right]\frac{1}{k}\left[\alpha^{B_kH_l}(i\omega) - \frac{l}{2}\alpha^{H_2}(i\omega)\right] \tag{2.4}$$

In this expression, $\alpha^{A_mH_n}(i\omega)$ and $\alpha^{B_kH_l}(i\omega)$ are the frequency dependent polarizabilities for the reference hydrides, and $\alpha^{H_2}(i\omega)$ is the corresponding value for H_2.

Key to the success of the D3 model is how it interpolates the pre-tabulated C_6 coefficients to adapt them to the current chemical environment via these coordination numbers. It computes continuous coordination numbers for each atom a in a given system as,

$$CN^a = \sum_{b \neq a}^{N} \frac{1}{1 + e^{-16\left(\frac{4}{3}\frac{R_{ab}^{cov}}{R_{ab}} - 1\right)}} \tag{2.5}$$

where R_{ab}^{cov} are sums of pre-tabulated, modified covalent radii for each element pair. Refer to the original D3 paper for details on the set of the radii used.[123] The C_6 dispersion coefficients for atom a interacting with atom b in their current coordination environments is then computed via a weighted average of the C_6 values from the tabulated coordination environments,

$$C_{6,ab} = \frac{\sum_i \sum_j C_{6,ab}^{ref}(CN_i^a, CN_j^b) L_{ij}}{\sum_i \sum_j L_{ij}} \quad \text{where} \quad L_{ij} = e^{-4[(CN^a - CN_i^a)^2 + (CN^b - CN_i^b)^2]} \tag{2.6}$$

Indices i and j sum over the all reference hydrides for the given element. Once the C_6 coefficients are obtained for an atom-pair, the C_8 coefficients are estimated according to,

$$C_{8,ab} = 3 C_{6,ab} \sqrt{Q_a Q_b} \tag{2.7}$$

and

$$Q_a = \frac{1}{2} \sqrt{Z_a} \frac{\langle r^4 \rangle_a}{\langle r^2 \rangle_a} \tag{2.8}$$

where Z_a is the nuclear charge, and $\langle r^4 \rangle_a$ and $\langle r^2 \rangle_a$ are pretabulated multipole expectation values for the element. See Grimme's original work for more details.[123]

In MP2D, Grimme's reference hydride C_6 coefficients (as implemented in Cuby4[129]) are used for the CKS portion of the model. Other D3 parameters, such as the multipole expectation values $\langle r^n \rangle_a$, covalent radii R_{ab}^{cov}, and cutoff radii $R_{0,ab}$ are also employed in MP2D

without modification. However, several modifications and new ingredients were added in the development of MP2D, as described in the following sections.

2.2.3 UCHF dispersion coefficients

Before adding the CKS dispersion to MP2, one must subtract out the UCHF dispersion that is already present. This requires computing UCHF dispersion coefficients that are analogous to the existing D3 CKS ones. The general expression for the frequency dependent dipole-dipole polarizability tensor $\alpha_{\lambda\sigma}(i\omega)$ from intermolecular perturbation theory is,

$$\alpha_{\lambda\sigma}(i\omega) = \sum_{M\neq 0} \frac{\omega_M \left[\langle 0|\hat{\mu}_\lambda|M\rangle\langle M|\hat{\mu}_\sigma|0\rangle + \langle 0|\hat{\mu}_\sigma|M\rangle\langle M|\hat{\mu}_\lambda|0\rangle\right]}{\hbar(\omega_M^2 + \omega^2)} \tag{2.9}$$

where 0 and M refer to ground and excited states, ω_M is the excitation energy, and λ and σ refer to different Cartesian components of the dipole operator $\hat{\mu}$. The resulting polarizability tensor $\alpha_{\lambda\sigma}(i\omega)$ is a symmetric 3×3 matrix with unique xx, xy, xz, yy, yz, and zz elements.

At the UCHF level of theory, the excited state wavefunctions involve vertical excitation of an electron from occupied orbital i to virtual orbital a with no orbital relaxation. In that case, the matrix elements simplify to matrix elements of the dipole operator involving occupied orbital i and virtual orbital a, and the excitation energy ω_M reduces to the energy difference between orbitals i and a, $\omega_M = \epsilon_i - \epsilon_a = \epsilon_{ia}$. Employing these simplifications and recognizing that the molecular orbitals are real, Eq 2.9 becomes,

$$\alpha_{\lambda\sigma}(i\omega) = 2\sum_{ia} \frac{\epsilon_{ia}\langle i|\hat{\mu}_\lambda|a\rangle\langle a|\hat{\mu}_\sigma|i\rangle}{\hbar(\epsilon_{ia}^2 + \omega^2)} \tag{2.10}$$

where the sums run over all spin orbitals i and a. Spin integration yields the following spin-unrestricted expression,

$$\alpha_{\lambda\sigma}(i\omega) = 2 \sum_{ia}^{\alpha\ spin} \frac{\epsilon_{ia}\langle i|\hat{\mu}_\lambda|a\rangle\langle a|\hat{\mu}_\sigma|i\rangle}{\hbar(\epsilon_{ia}^2+\omega^2)} + 2 \sum_{\bar{i}\bar{a}}^{\beta\ spin} \frac{\epsilon_{\bar{i}\bar{a}}\langle \bar{i}|\hat{\mu}_\lambda|\bar{a}\rangle\langle \bar{a}|\hat{\mu}_\sigma|\bar{i}\rangle}{\hbar(\epsilon_{\bar{i}\bar{a}}^2+\omega^2)} \tag{2.11}$$

where i and a refer to α spin orbitals and $\bar{\imath}$ and $\bar{a}$ refer to β spin orbitals. In the spin restricted case, the expression for the frequency-dependent polarizability further simplifies to,

$$\alpha_{\lambda\sigma}(i\omega) = 4 \sum_{ia} \frac{\epsilon_{ia}\langle i|\hat{\mu}_\lambda|a\rangle\langle a|\hat{\mu}_\sigma|i\rangle}{\hbar(\epsilon_{ia}^2+\omega^2)} \tag{2.12}$$

Finally, the isotropic frequency-dependent polarizabilities used as inputs for the D3 model are computed as the trace of the frequency dependent polarizability tensors $\alpha_{\lambda\sigma}(i\omega)$.

From these isotropic UCHF polarizabilities, the C_6 coefficients were computed by subtracting out the approximate hydrogen contribution and performing Casimir-Polder integration over imaginary frequency according to Eq 2.4. The integration was performed via quadrature at ten frequencies given by,

$$i\omega_j = \frac{i}{\tan\left[\frac{\pi}{4N}(2j-1)\right]} \tag{2.13}$$

for $j = 1, 2, \cdots, 10$ and with integration weights g_j:

$$g_j = \frac{\pi}{2N\sin^2\left[\frac{\pi}{4N}(2j-1)\right]} \tag{2.14}$$

Empirical testing indicates that ten quadrature points is sufficient to obtain well-converged dispersion coefficients.

The frequency-dependent polarizabilities $\alpha(i\omega)$ were evaluated using a modified version of Molpro 2012.[130] While most of the hydrides used in the D3 model involve closed-shell species (spin restricted wavefunctions), there are some open-shell species for which the

unrestricted spin formalism is necessary. Open-shell species include many bare elements (e.g. H, C, N, O) and low-coordination number hydrides (e.g. CH, OH). The spin-restricted expressions were already available in Molpro as part of the MP2C implementation, and the spin-unrestricted variant was implemented in a local version of Molpro.

At present, UCHF frequency-dependent polarizabilities have been computed for all necessary hydrides of H, B, C, N, O, F, Ne, P, S, Cl, Ar, and Br. These represent some of the most common elements occurring in organic chemistry. Extending the list of elements further would be straightforward, though not all elements would be well-described with an MP2-based model (e.g. transition metals). The hydride geometries and basis sets used to obtain these frequency-dependent polarizabilities are identical to those used in the original D3 work.[123]

Table 2.1 presents several sample C_6 coefficients for C-C interactions. Broadly speaking, as the level of hydrogen saturation decreases, the atomic polarizability and therefore magnitude of the contribution to the dispersion energy should increase. Accordingly, the dispersion coefficients increase in magnitude from ethane to ethene and ethyne. The dispersion energy for such systems is typically overestimated at the UCHF level. This manifests in the UCHF coefficients listed in Table 2.1, which are 30–40% larger than the CKS ones.

It is notable that the local coordination number scheme used to interpolate the C_6 coefficients for the given chemical environment barely differentiates between the aromatic bonds in a species like benzene and the double-bond environment of ethene. The resulting dispersion coefficients are nearly identical for both cases. In reality, the disper-

Table 2.1: Sample UCHF and CKS C_6 coefficients for carbon-carbon interactions between two identical carbons for different species/coordination environments as computed according to the D3 scheme.

	UCHF C_6	CKS C_6
Ethane	24.1	18.3
Ethene	34.9	25.7
Ethyne	41.2	29.5
Benzene	34.8	25.6

sion coefficients should be somewhat larger for the aromatic species. This translates to MP2D underestimating the magnitude of the dispersion correction in the benzene π dimer at the S66x8 equilibrium separation, for example. MP2D reduces the MP2 binding energy from 4.9 kcal/mol to 3.3 kcal/mol, versus 2.8 kcal/mol for MP2C and 2.7 kcal/mol for the CCSD(T) benchmark. The dispersion correction here also would not capture the sorts of system-size-dependent changes in the C_6 coefficients observed for large carbon nanotubes or graphene,[131] for example. Nevertheless, the results in Section 3.3 will demonstrate that MP2D performs well overall across a broad range of chemical systems.

2.2.4 Short-range damping

The MP2D dispersion correction must be damped at short-ranges to avoid unphysical behavior. Here, Tang-Toennies damping,[132] is used to attenuate the dispersion correction at short interatomic separations instead of the Becke-Johnson or zero-damping

used in D3[133] The physically-motivated Tang-Toennies damping function is well-suited for reproducing the dispersion energy in correlated methods.[134] The Tang-Toennies damping expression is given by,

$$f_N(R_{ab}) = 1 - \exp(s_R R_{ab}) \sum_{k=0}^{N} \frac{(s_R R_{ab})^k}{k!} \tag{2.15}$$

where N is the order of the dispersion term, i.e. 6 and 8, and s_R is a distance scaling factor calculated from the cutoff radius $R_{0,ab}$ (taken from the D3 dispersion correction) using two empirical parameters:

$$s_R = a_1 R_{0,ab} + a_2. \tag{2.16}$$

Fitting the damping functions to UCHF and CKS dispersion energies separately, we found that optimal damping parameters a_1 and a_2 for the UCHF dispersion energies differ from those for the CKS ones. This would give rise to four parameters, a_1^{UCHF}, a_2^{UCHF}, a_1^{CKS} and a_2^{CKS}. However, the number of parameters can be reduced as described in Section A.1.1 below.

Although the Tang-Toennies damping provides a physically sound treatment in the non-covalent regime, it damps insufficiently at covalent distances. This leads to deteriorated MP2D description of reaction energies, for example. We thus introduce a secondary short-ranged damping that ensures that dispersion correction becomes constant at covalent

distances. This damping is achieved by modifying the value of the interatomic distance R_{ab} that enters Eq 2.3 as

$$R'_{ab} = \begin{cases} r_{cut} R_{0,ab} & \text{if } R_{ab} \leq R_{0,ab}(r_{cut} - w/2) \\ R_{ab} & \text{if } R_{ab} \geq R_{0,ab}(r_{cut} + w/2) \\ r_{cut} R_{0,ab} + f(R_{ab}, R_{0,ab} r_{cut}, R_{0,ab} w) & \text{otherwise;} \end{cases} \tag{2.17}$$

$$f(R_{ab}, r'_{cut}, w') = (-2.5x^8 + 10x^7 - 14x^6 + 7x^5) * w'; \tag{2.18}$$

$$x = \frac{R_{ab} - (r'_{cut} - w'/2)}{w'} \tag{2.19}$$

where two more parameters are introduced, r_{cut} and w, which are defined as dimensionless factors scaling the radius $R_{0,ab}$ taken from the D3 correction. Eq 3.12 leaves R_{ab} untouched for distances greater than $R_{0,ab}(r_{cut} + w/2)$, and it fixes R_{ab} at a constant fraction of $R_{0,ab}$ for distances that are shorter than $R_{0,ab}(r_{cut} - w/2)$. The third portion of the function in Eq 3.12 smooths the transition between these two regimes. The high-order polynomial ensures smooth first, second and third derivatives at the end points of the switching interval.

This damping is applied at very short distances so that it practically does not affect intermolecular non-covalent interactions. For example, using the final optimized parameters described in Section A.1.1, this damping smoothly alters the effective interatomic separation for two carbon atoms from the actual separation to a fixed value near 2 Å and below (Figure 2.1). The fractional nature of r_{cut} and w means that these distances adapt depending on the threshold radius $R_{0,ab}$ for the given atom pair. The overall dispersion energy is therefore doubly damped: first by Tang-Toennies in the non-covalent regime, and second by this short-range damping in the covalent regime.

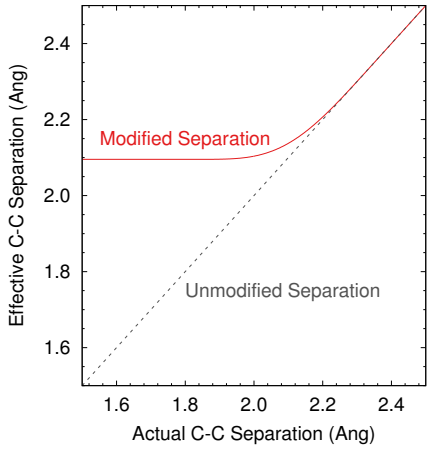


Figure 2.1: Example of how the secondary short-range damping modifies the effective C-C interatomic separation based on a cutoff radius of $R_{0,ab} = 2.9103$ Å, $r_{cut} = 0.72$, and $w = 0.2$.

2.2.5 Modified C_6 interpolation

In D3, the C_6 coefficients are interpolated using a continuous coordination number CN calculated from distances to all other atoms using a switching function (Eq 2.5). Although this switching function decays quickly, it yields small but nonzero contributions even at non-covalent distances. This makes the C_6 coefficients in a dimer slightly different to these in isolated monomers in the same geometry, and this propagates also to the C_8 coefficients. This proves problematic at short distances (where the C_8 term becomes important) when a weaker damping function is used.[134] In MP2D, we eliminated this issue by replacing the switching function with one that drops exactly to zero at larger separations. Again, the polynomial interpolating between the short- and long-range regimes was designed to have smooth first and second derivatives at the end points of the switching interval. The scaling

factors in this function were fitted to closely reproduce the original D3 one. The MP2D coordination number is now calculated as,

$$CN = \sum_{B \neq A}^{N} f(R_{ab}^{cov}, R_{ab}); \tag{2.20}$$

$$f(R_{ab}^{cov}, R_{ab}) = \begin{cases} 1.0 & \text{if } R_{ab} \leq 0.95 R_{ab}^{cov} \\ 0.0 & \text{if } R_{ab} \geq 1.75 R_{ab}^{cov} \\ f'(x) & \text{otherwise;} \end{cases} \tag{2.21}$$

$$f'(x) = 1.0 - (-20x^7 + 70x^6 - 84x^5 + 35x^4); \tag{2.22}$$

$$x = \frac{R_{ab} - 0.95 R_{ab}^{cov}}{1.75 R_{ab}^{cov} - 0.95 R_{ab}^{cov}}. \tag{2.23}$$

In the geometry optimizations reported here, we used integer coordination numbers to simplify the calculation of the gradient. All the studied systems have well-defined geometries where the continuous valence numbers differ only negligibly from integer ones, so this assumption does not introduce any appreciable error. One could implement gradients for the continuous valence coordination numbers if desired.

2.2.6 MP2D Parameterization

The MP2D model described thus far could conceivably employ up to ten potential global parameters: separate s_6 and s_8 scaling terms for each of the UCHF and CKS C_6 and C_8 dispersion energies (four parameters), two parameters for the UCHF Tang-Toennies damping function (a_1^{UCHF} and a_2^{UCHF}), two parameters for the CKS Tang-Toennies damping function (a_1^{CKS} and a_2^{CKS}), and two parameters for the secondary short-range damping (r_{cut} and w). To obtain correct dispersion energies at long distances, the s_6 parameter is

set to unity for both UCHF and CKS dispersion. In exploring the parameterization, we found that the value of the s_8 parameter is similar in both the UCHF and CKS cases, so we use a single global s_8 parameter for both. This reduces the number of possible global fitting parameters from ten to seven. At this point, we tested multiple variants of the parameterization protocol, and analyzed the results obtained with different parameter sets, exploring the possibilities to simplify the method further.

First, we tested fitting the CKS and UCHF Tang-Toennies terms separately to the corresponding CKS and UCHF energies from MP2C calculations performed on the S66x8 benchmark data set.[135] The resulting dispersion coefficient model reproduced the original CKS and UCHF energies rather well. The optimal parameter values in the CKS and UCHF Tang-Toennies damping functions differed significantly. Next, to capture some higher-order contributions not covered by MP2C, the Tang-Toennies fit parameters were refined against CCSD(T)/CBS interaction energies on the same S66x8 data (i.e. fitting to the energy difference between MP2 and CCSD(T) instead of the raw MP2C energy components). When the parameterization was started from the UCHF and CKS parameter values fitted to the CKS and UCHF dispersion energy components separately, the resulting Tang-Toennies damping parameters changed only very slightly.

However, further testing found that an equally good fit to the post-MP2 correlation energy can be obtained when the same values of the parameters are used in both the CKS and UCHF damping functions. Using common parameters reduces the fidelity with which MP2D reduces the individual UCHF and CKS dispersion energy components, but it has no appreciable negative impact on the quality of the net dispersion correction. We decided

that reduction in the number of adjustable parameters was more useful than reproducing the individual dispersion energy components. Setting

$$a_1^{UCHF} = a_1^{CKS} = a_1, \text{ and} \tag{2.24}$$

$$a_2^{UCHF} = a_2^{CKS} = a_2. \tag{2.25}$$

reduces the number of global parameters down to five (s_8, a_1, a_2, r_{cut}, and w). Figure A.1 in Appendix A provides a sample comparison for MP2D with and without constraining the Tang-Toennies parameters to be identical.

During the parameterization procedure thus far, the S66x8 data set proved sufficient for a robust and transferable parameterization of the correction at non-covalent distances. The three parameters affecting the dispersion energy in this range of distances (a_1, a_2 and $s8$) were thus optimized first on the S66x8 data set with the short-ranged damping disabled. Subsequently, the initial values of the remaining two parameters in the short-range damping ($r_{cut} = 0.7$ and $w = 0.2$) were manually chosen to yield the best compromise between short intermolecular interactions (in the S66x10 data set) and conformation energies (using all the conformer data sets considered later in the paper). Overall, a fairly broad range of values for r_{cut} and w provide comparably good performance. Larger values of r_{cut} would be obtained if the method was optimized on thermochemistry data, but the description of non-covalent interactions at short distances would be compromised while reaction energies would improve only by about 0.5 kcal/mol.

Finally, all the five parameters were fitted again to the S66x8 data set, resulting in only small change to the values of r_{cut}. The resulting parameters thus represent a minimum with respect to non-covalent interactions around and above equilibrium distance

Table 2.2: Optimized values of the five global parameters used in the MP2D method.

s_8	1.187	dimensionless
a_1	0.944	dimensionless
a_2	0.480	Ångstrom
r_{cut}	0.72	dimensionless
w	0.20	dimensionless

(represented by the S66x8 data set) which can be expected to be the main application targets for the method. At the same time, the formulation of the model and the choice of the initial values of the parameters used additional information from conformational energies needed to provide a seamless connection between the covalent and non-covalent regime.

Table 3.1 lists the final values of the parameters. Figure 2.2 plots the MP2D dispersion correction energy (C_6 contributions only for simplicity) for two carbon atoms as a function of distance with no damping, just Tang-Toennies damping, and finally the actual doubly-damped model that also includes the covalent regime short-range damping.

The MP2D dispersion correction has been fitted to reproduce the counterpoise-corrected ΔCCSD(T) energy correction. The remaining question is how to treat the basis set superposition error (BSSE) in the MP2 part of the calculation. When the MP2 energy is calculated in a large basis set or extrapolated to the CBS limit, the MP2D method should

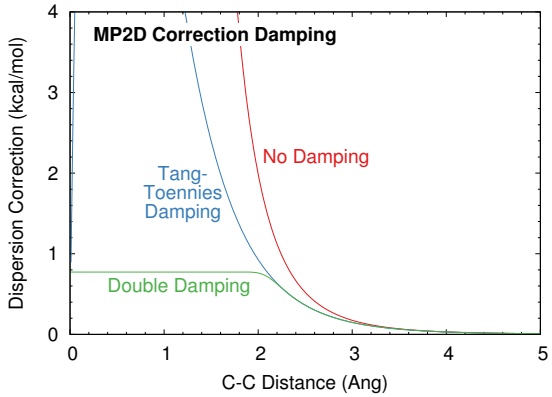


Figure 2.2: Behavior of the MP2D dispersion correction (C_6 only here) without damping, with Tang-Toennies damping, and the double Tang-Toennies/shorter-range damping for two sp^2-hybridized carbon atoms.

be universally applicable to both inter- and intramolecular energies because the BSSE would be smaller than the error of the dispersion correction.

This can be demonstrated on the calculations of the interaction energies in the S66 data set performed with and without counterpoise (CP) correction. At the CBS limit, using MP2D without CP correction yields smaller RMSE (0.19 kcal/mol) than when CP correction is applied (0.26 kcal/mol). The good performance without CP correction results from error cancellation, but it shows that the energy changes associated with the CP correction are several times smaller than the overall error. In the aug-cc-pVQZ basis, the CP-corrected and uncorrected results are very similar (RMSE 0.35 and 0.37 kcal/mol). In smaller basis sets, the CP uncorrected interaction energies become significantly worse.

2.3 Computational Methods

The dispersion correction was implemented in the freely available Cuby4 framework,[129] which interfaces multiple computational chemistry packages that could provide the MP2 calculation. The MP2D dispersion correction implementation, including the dispersion coefficients for the 13 common first- and second-row elements noted above, documentation, and input examples, are provided at the Cuby website.[136]

Electronic structure calculations were carried out using a mixture of PSI4[137], Molpro 2012.1,[130] and TURBOMOLE.[138,139] All single-point[139] MP2, MP2D, MP2C, MP2.5, and CCSD(T) results reported here were extrapolated to the complete basis set (CBS) limit,[70] typically from the Dunning aug-cc-pVXZ basis sets.[140] For MP2.5 and CCSD(T), the CBS limit was estimated using the standard focal point technique which combines

MP2/CBS with post-MP2 correlation estimated in a smaller basis set.[112,141] B3LYP-D3(BJ) results employ the nearly-complete def2-QZVP basis. Geometry optimizations were performed in the def2-TZVP basis with no counterpoise correction. Integer coordination numbers of the atoms were employed for the dispersion coefficients in MP2D geometry optimizations. All calculations here employed density fitting with standard auxiliary basis sets throughout. Counterpoise corrections for basis set superposition error were employed in the MP2-based methods (excluding the geometry optimizations). Because the DFT-D3(BJ) damping parameters were fitted without counterpoise correction,[142] no counterpoise corrections were applied to the DFT results here.

2.4 Results and Discussion

2.4.1 Energetics

The results here will demonstrate that MP2D performs very competitively with other techniques across a wide variety of systems, including intermolecular interactions, conformational energies, and thermochemistry. For comparison purposes, B3LYP-D3(BJ) was chosen as a representative, widely used density functional that generally performs well for non-covalent interactions.[2,143–145] Comparison against other functionals (particularly double-hybrid density functionals[145]) would make for an interesting subject of future work.

Non-covalent interactions at short range. Consider first the sixty-six dimers at ten intermolecular separations comprising the S66x10 benchmark test set.[146] The eight largest intermolecular separations are identical to the S66x8 set used in the parameter fit-

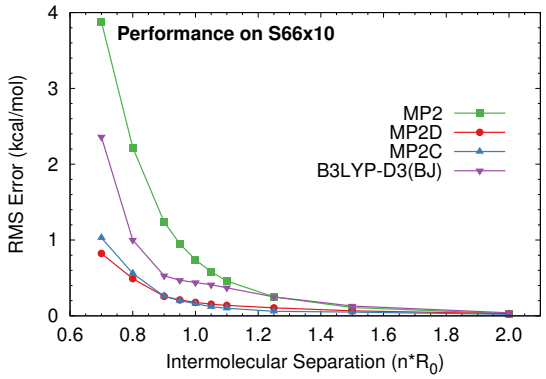


Figure 2.3: Performance of MP2 and other methods on the S66x10 benchmark test set used in fitting the empirical parameters.

ting, while the shortest two distances ($0.7R_0$ and $0.8R_0$) were not part of the final fit. Figure 2.3 plots the root-mean-square (rms) errors versus intermolecular distance. At equilibrium distances $1.0R_0$, the 0.18 kcal/mol rms error for MP2D is much better than MP2 (1.24 kcal/mol), appreciably smaller than B3LYP-D3(BJ) (0.44 kcal/mol), and almost as good as MP2C (0.16 kcal/mol).

The MP2D performance improves further relative to the other methods at shorter separations. As the dimer separation decreases, the fraction of the MP2C UCHF and CKS dispersion energies captured by MP2D decreases, with the short-range damping effectively mimicking some of the repulsive exchange-dispersion terms which are not corrected in MP2C. At $0.7R_0$ for instance, which was not included in the training data, the MP2D error is only 0.82 kcal/mol, versus 1.03 kcal/mol for MP2C and 2.36 kcal/mol for B3LYP-D3(BJ). Even the recently proposed B3LYP-D3M(BJ), which seeks to improve the short-range

behavior of D3, exhibits a significantly larger error of 1.57 kcal/mol at this separation (those reported results were counterpoise corrected).[146]

Of course, MP2D is only useful if it also performs well broadly, beyond the systems the empirical parameters were fitted for. Figure 2.4 reports box plot error distributions for MP2, MP2D, MP2C, and B3LYP-D3(BJ) across 10 different benchmark sets consisting of several thousand diverse examples and lists the rms errors.

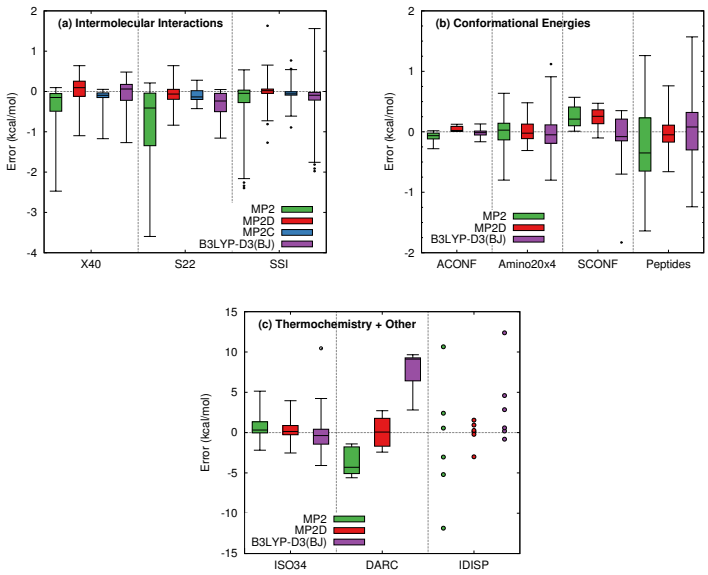
Interaction energies in more diverse systems. For the 3380 protein side chain-side chain interactions in SSI,[2] the MP2D errors are two-thirds smaller than MP2, comparable to MP2C, and half those for B3LYP-D3(BJ) (Figure 2.4a). For the halogen-containing dimers in X40[147] (excluding iodine-containing species for which the Molpro MP2C implementation fails) and the S22 set,[148,149] MP2D reduces the MP2 errors 2–4-fold, though they are larger than the MP2C ones and only moderately better than B3LYP-D3(BJ). Some of the largest MP2D errors occur for π-stacked cases, for which it corrects much of the MP2 error, but not as effectively as MP2C. Such π-stacked cases make up a disproportionately large fraction of the S22 set compared to S66x8 or SSI, which helps explain the larger difference between MP2C and MP2D for that set.

Conformation energies. Figure 2.4b examines conformational energy benchmarks for alkanes (ACONF[150]), amino acids (Amino20x4[145]), sugars (SCONF[145]), and short peptides (from MPCONF196[151]). MP2D improves upon MP2 in all cases, reducing the rms errors and generally narrowing the width of the error distributions. The largest improvements occur for the amino acids and peptides where van der Waals interactions are relatively large. Dispersion plays the smallest role the sugar conformers (SCONF), and the

MP2D improvement there is correspondingly small. MP2D outperforms B3LYP-D3(BJ) appreciably for Amino20x4, SCONF, and the Peptides.

Thermochemistry. While intermolecular interactions and conformational energies represent traditional applications for methods like MP2D, it is equally important that MP2D should not interfere with thermochemistry. Figure 2.4c plots errors for small-molecule isomerizations (ISO34[152]) and Diels-Alder reactions (DARC[153]). Dispersion plays a minimal role in the small-molecule isomerization reactions found in ISO34, and the MP2D correction only reduces the MP2 error by 8%. Dispersion is much more important in the Diels-Alder reactions (DARC), and MP2D cuts the MP2 error in half and substantially outperforms B3LYP-D3(BJ). Also impressive, however, is the MP2D performance for IDISP,[145] which consists of four chemical reactions and two conformational changes involving intramolecular dispersion. MP2D reduces the 7.0 kcal/mol rms error for MP2 to only 1.4 kcal/mol, compared to 5.5 kcal/mol for B3LYP-D3(BJ).

While a thorough comparison of density functionals is beyond the scope of this letter, many of the test sets considered here are part of the GMTKN55 suite,[145] for which results from many density functionals have been reported. Analysis of those results suggests that MP2D is competitive with or better than the best dispersion-corrected hybrid functionals. The best double hybrid functionals sometimes perform moderately better (with similar computational cost to MP2D), but those functionals frequently employ empirical spin-component scaling of the MP2 correlation (e.g. DSD-BLYP-D3(BJ)[154]), which is not exploited here.



Root-Mean-Square Error (kcal/mol)

Set	MP2	MP2C	MP2D	B3LYP-D3(BJ)
S66x10	1.54	0.39	0.33	0.88
X40	0.72	0.27	0.34	0.35
S22	1.38	0.19	0.35	0.43
SSI	0.36	0.14	0.16	0.32
ACONF	0.11	–	0.06	0.07
Amino20x4	0.26	–	0.17	0.29
SCONF	0.31	–	0.29	0.51
Peptides	0.71	–	0.27	0.56
ISO34	1.68	–	1.55	2.64
DARC	3.97	–	1.90	8.26
IDISP	7.03	–	1.44	5.53

Figure 2.4: Performance of MP2D and other methods for benchmark sets focusing on (a) intermolecular interaction energies, (b) conformational energies, and (c) thermochemistry. MP2C is only defined for the intermolecular interactions. Boxes and whiskers contain 50% and 95% (or 99.9% for part (a)) of the data, respectively. The table lists root-mean-square errors for all methods and sets.

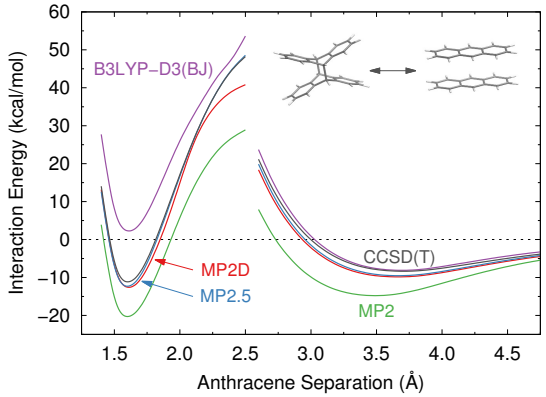


Figure 2.5: Potential energy basins for covalent anthracene photodimer (near 1.6 Å) and non-covalent π-dimer (near 3.6 Å). B3LYP-D3 employs def2-QZVP basis; all other methods at extrapolated CBS limit.

Anthracene dimerization. For further insight on how MP2D performs, consider the anthracene photodimerization reaction, which is the most challenging case in the IDISP set. This reaction includes both a non-covalent π-stacked dimer with intermolecular separation $\sim$3.6 Å and a covalent photoreacted dimer with intramolecular separation $\sim$1.6 Å. It proves problematic for many electronic structure methods.[155] Whereas the MP2C dispersion correction is defined only for the intermolecular dimer, MP2D can describe both regimes. A benchmark CCSD(T)/CBS potential energy scan was created for this reaction, as shown in Figure 2.5 and described in the Appendix A.

Figure 2.5 shows that various methods perform well for the intermolecular dimer energy basin, with MP2 performing the worst due to its overestimation of the π–π interactions. MP2D and MP2.5[156,157] perform very similarly, overbinding the π-dimer by 1.2–1.5 kcal/mol relative to CCSD(T). B3LYP-D3(BJ) does even better, underbinding it by 0.2 kcal/mol. The real challenge, however, occurs in modeling the covalent basin and the competition among the long covalent bonds between the two anthracenes, the anthracene ring distortion, and the very short-range dispersion interactions between the anthracene rings. Unsurprisingly, MP2 overbinds by 9.1 kcal/mol. At the other extreme, B3LYP-D3(BJ) underestimates the stability of the covalent dimer by 13.4 kcal/mol, predicting the photodimerization reaction to be significantly endothermic instead of slightly exothermic. In contrast, MP2.5 and MP2D reproduce CCSD(T) nicely across most of the potential energy surface, with 1.2–1.5 kcal/mol errors at the minima. Both methods reproduce the energy difference between the two minima to within less than 0.1 kcal/mol. Only near $\sim$2.5 Å does MP2D perform appreciably worse than CCSD(T) or MP2.5, where the increasingly

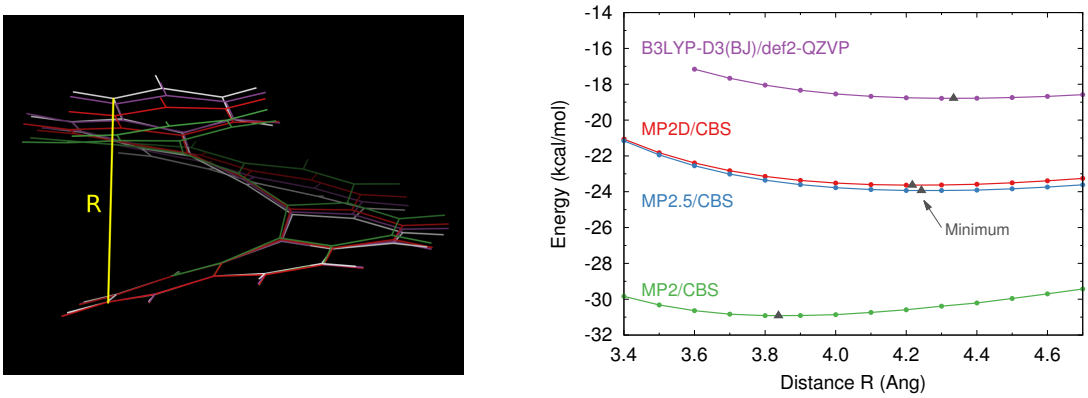


Figure 2.6: Overlay of [7]helicene crystal structure[1] with that predicted by MP2 (green), MP2D (red), and B3LYP-D3(BJ) (purple) in the def2-TZVP basis. Single-point energies relative to linear heptacene as a function of the distance R.

significant static correlation arising from the stretching of the two inter-anthracene covalent bonds is ill-described by spin-restricted MP2. The ability to accurately describe both intra- and intermolecular interactions simultaneously is a key feature of MP2D.

2.4.2 Geometries

Facile geometry optimization represents another advantage of MP2D over MP2C for systems with significant non-covalent interactions. Consider the challenging example of [7]helicene—seven fused benzene rings arranged in a helix. Figure 2.6 overlays the structures optimized with MP2, MP2D, and B3LYP-D3(BJ) in the def2-TZVP basis with the X-ray crystal structure. MP2 overestimates the van der Waals interactions and artificially compresses the helical spacing R at 3.72 Å, versus ∼4.4–4.6 Å from the experimental crystal structures.[1,158,159] MP2D significantly corrects this to 4.08 Å, and B3LYP-D3(BJ) predicts a seemingly even better 4.32 Å.

However, the potential energy surface for compressing [7]helicene is very flat, and the experimental crystal structure may differ from the gas-phase electronic energy one due to solid-state packing forces. To investigate, a one-dimensional B3LYP-D3(BJ)/def2-TZVP relaxed scan over R was performed (see Appendix A Information for details). Large-basis single-point energies were computed with B3LYP-D3(BJ), MP2, MP2D, and MP2.5 and plotted relative to the energy of the isomeric linear heptacene in Figure 2.6. Linear heptacene provides a useful reference structure, since the planar molecule has no significant non-covalent interactions. Taking MP2.5 as the reference energy, MP2 significantly overestimates the interactions and underestimates the optimal R. B3LYP-D3(BJ) underestimates the interaction energy by a quarter, and it overestimates the distance R, while MP2D reproduces MP2.5 quite well and drastically lower computational cost. These results imply that the MP2D optimized structure is actually closer than the B3LYP-D3(BJ) one to the true gas-phase structure.

The ten stationary points on the benzene dimer potential energy surface[4] provide another interesting case. For nine of the ten structures, B3LYP-D3(BJ) and MP2D reproduce the reference DFT + CCSD(T) structures well, with root-mean-square RMSD of 2.8 and 3.2 pm respectively (Table 2.3). However, the geometry of the S4 structure varies strongly with the method, as shown in Figure 2.7. In the reference structure, the benzene molecules interact at a 53.8° angle. MP2 fails completely for this structure, optimizing to a parallel π-stacked structure, while B3LYP-D3(BJ) underestimates the angle at 45.2° (RMSD 12.2 pm). In contrast, MP2D predicts a 51.8° angle of the reference structure and gives an RMSD of only 3.5 pm.

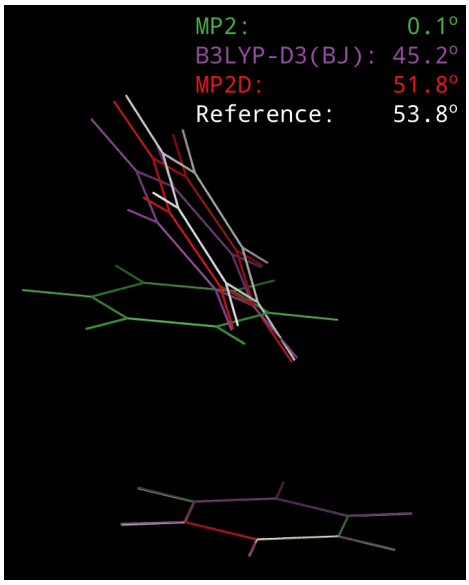


Figure 2.7: Overlay of benzene dimer S4, showing how MP2D (red) reproduces the correct angle between the molecules, unlike MP2 (green) and B3LYP-D3(BJ) (purple).

2.5 Conclusions

In summary, MP2D largely corrects the key dispersion-related flaws of MP2 with trivial computational cost using a dispersion correction based on the Grimme D3 scheme. It relies on pre-tabulated *ab initio* dispersion coefficients and five universal empirical parameters designed to attenuate the correction at short range and compensate for higher-order dispersion contributions. Unlike MP2C, MP2D improves the description of intramolecular dispersion and can be used for geometry optimizations. The results here indicate that MP2D provides a valuable alternative to DFT in systems where van der Waals interactions are important, ranging from organics to biomolecules. A software implementation of the MP2D dispersion correction that can be easily coupled with MP2 calculations in many computational chemistry programs is freely available.[136]

Future work should compare MP2D against a broader suite of density functionals. It would also be interesting to pursue a spin-component-scaled version of MP2D. One of the key problems in spin-component-scaled methods has been the difficulty of finding parameters that simultaneously improve the treatment of thermochemistry and non-covalent interactions. The MP2D dispersion correction addresses the non-covalent interactions, which would allow the spin-component scaling to correct the thermochemistry errors. Such an approach could provide an interesting alternative to some of the best-performing double-hybrid density functionals which also employ spin-component-scaling and a similar number of global empirical parameters.

Table 2.3: Root-mean-square deviations (rmsd) in pm for the optimized geometries for 10 benzene dimer stationary points relative to reference structures obtained with a mixed DFT/CCSD(T) approach.[4] Structure S4 causes significant problems for MP2 and B3LYP-D3(BJ), so the rmsd is reported with and without S4 included. Counterpoise corrections were not employed.

Structure	MP2	MP2D	B3LYP-D3(BJ)
	def2-TZVP	def2-TZVP	def2-TZVP
M1	11.4	3.1	2.7
M2	6.2	2.3	0.7
S1	7.2	4.9	1.2
S2	11.7	3.4	2.9
S3	8.3	3.9	2.7
S4	90.5	3.5	12.2
S5	3.8	1.0	1.0
S6	2.5	1.3	0.9
S7	11.1	3.6	4.7
S8	10.4	2.9	4.6
RMSD (all)	29.8	3.2	4.7
RMSD (excluding S4)	8.7	3.2	2.8

Chapter 3

Spin-component-scaled and dispersion-corrected second-order Møller-Plesset perturbation theory: A path toward chemical accuracy

Accurately modeling many chemically-interesting systems with electronic structure theory requires models capable of describing diverse mixtures of covalent and non-covalent interactions. Chemical reactions occurring in enzyme active sites demand models that can treat the thermochemistry associated with changes in the substrate chemical bonding together with the hydrogen bonding, electrostatic, and dispersion interactions that govern the

substrate-protein interaction.[160] Furthermore, the stabilities of molecular crystal conformational polymorphs are governed by the competition between intramolecular conformation and intermolecular packing.[12] In principle, high-accuracy methods like coupled cluster singles, doubles, and perturbative triples (CCSD(T)) can provide the requisite accuracy for modeling systems like these, but the steep $O(N^7)$ computational cost with system size N frequently makes it cost-prohibitive in practice.

Instead, Kohn-Sham density functional theory (DFT) has become the standard tool of choice for modeling such systems. Many successful density functionals have been developed over the years, some of which can approach CCSD(T) accuracy. Large benchmark studies[145,161] have identified some of the best current functionals. These include, for example, the hierarchy of functionals developed by Mardirossian and Head-Gordon: the range-separated hybrid functional ωB97X-V,[162] its hybrid meta-GGA variant ωB97M-V,[163] and the double-hybrid meta-GGA functional ωB97M(2).[164] The family of dispersion-corrected, spin-component-scaled double hybrid (DSD) density functionals developed in the Martin group are also highly competitive, both in their original[165] and recently-revised forms.[83]

At the same time, there have been many efforts to achieve near-coupled cluster accuracy using wave function methods that are less computationally demanding than CCSD(T). The domain-based local pair natural orbital variant (DLPNO-CCSD(T))[166] achieves most of the accuracy of CCSD(T) at far lower computational cost, for example. At the other extreme, inexpensive machine learning models that target CCSD(T) accuracy continue to improve.[120,167] Models based on second-order Møller-Plesset perturbation theory (MP2), the least-expensive correlated wavfunction method, have also garnered con-

siderable attention over the years. Although MP2 scales $O(N^5)$ with system size, efficient density-fitting algorithms mean that the computational cost of evaluating the non-iterative MP2 correlation energy is small compared to that associated with the underlying iterative Hartree-Fock (HF) self-consistent field energy for many systems. This means that MP2 can be computationally competitive with modern state-of-the-art density functionals for systems with up to ∼100 atoms. Furthermore, MP2 inherently includes exact exchange and does not suffer from the issues of self-interaction error/delocalization error that plague most existing density functionals.[111,168]

Despite its advantages, MP2 has its own significant deficiencies which manifest in systems with strong static correlation (e.g. stretched covalent bonds) or in systems where van der Waals dispersion interactions are important (e.g. benzene dimer). The second-order correlation energy can be partitioned into contributions arising from same-spin ($\alpha\alpha$ and $\beta\beta$) and opposite-spin ($\alpha\beta$ and $\beta\alpha$) electron pairs. The same-spin contributions are more connected with long-range static correlation, while the opposite-spin ones are more important for the dynamic correlation that is associated with dispersion. The MP2 perturbation series is biased toward the same-spin correlation and frequently overestimates its contribution.[169] In 2003, Grimme's spin-component-scaled MP2 (SCS-MP2) model[170] demonstrated how scaling the same-spin and opposite-spin energy components of the MP2 correlation energy with constant coefficients improves the accuracy of MP2 on systems that would otherwise be poorly described, without any increase in the computational effort required. The SCS-MP2 prescription greatly improves upon canonical MP2 for predicting reaction thermochemistry.

However, it soon became apparent that the spin-component scaling coefficients appropriate for one type of chemical problem do not always transfer well to other chemical problems. For example, the original SCS-MP2 model scales the same-spin correlation energy by $c_{ss} = 1/3$, while the opposite-spin correlation energy is scaled up by $c_{ss} = 6/5$.[170] These values were initially determined from studying reaction energies, though subsequent work established a theoretical basis for these scaling parameter values.[169,171,172] Studying the S22 benchmark set a few years later, Distasio and Head-Gordon found optimal scaling coefficients $c_{ss} = 1.29$ and $c_{os} = 0.40$ for molecular interactions (MI), denoting the resulting model as SCS(MI)-MP2.[173] These SCS(MI)-MP2 c_{ss} and c_{os} scaling coefficients are nearly reversed compared to those found in the original SCS-MP2.

Given that no single set of spin-scaling coefficients can fully address the MP2 problems, a few strategies for SCS-MP2 models have emerged over the years. One approach tailors the SCS coefficients for specific chemical systems, such as for nucleic acid base pair interactions,[174] ethylene dimers,[175] or ionic liquids[176]. Such models can potentially work well, though this parameterization strategy inherently limits transferability of the model. Another approach adapts the spin-component scaling coefficients to each given system on the fly. This has been done via spin-ratio scaled spin components (SRS-MP2)[177] or by machine learning the optimal scaling parameters as in SNS-MP2.[178] Adaptive spin-scaling approaches can be more universal, though care must be taken to ensure that the coefficient adaptations retain smooth and continuous potential energy surfaces. Moreover, adaptive schemes can still have limited applicability: the design of the neural network-based SNS-MP2 model limits its application to dimer intermolecular interactions, for instance.

Other SCS models seek to exploit the greater computational efficiency associated with the opposite-spin correlation,[179] to improve the long-range behavior of SCS methods,[180] or to apply these ideas to higher-levels of theory such as MP3,[181] coupled cluster models,[182,183] and excited state approaches.[184–187]

The fundamental challenge for existing spin-component-scaled MP2 methods is that they attempt to use spin component scaling to address multiple, physically distinct weaknesses inherent in MP2 simultaneously. These limitations can generally be partitioned into the suitability of MP2 pair correlations for covalent bond chemistry (termed "thermochemistry" here for simplicity) versus the problems associated with describing van der Waals dispersion in non-covalent interactions. The former typically include more significant amounts of static correlation energy, while the dispersion interaction arises from dynamical correlation. The differences between the optimal scaling coefficients in SCS-MP2 and SCS(MI)-MP2 highlight the challenge associated with addressing both problems simultaneously with spin-component scaling.

In the language of intermolecular perturbation theory, the dispersion problem arises from the uncoupled Hartree-Fock (UCHF) description of dispersion that is inherent in MP2.[114] Hesselmann's corrected MP2 model (MP2C),[115,121] addresses this by subtracting out the UCHF dispersion energy and replacing it with a better treatment computed at the coupled Kohn-Sham (CKS) level of theory. MP2C has proved very successful,[122] though its intermolecular perturbation theory formulation limits its application to dimer intermolecular interactions. We recently developed a new version of MP2C, called MP2D,[5] which recasts the MP2C dispersion correction in terms of Grimme's D3 dispersion correc-

tion.[123] MP2D is similar to the MP2 plus van der Waals approach proposed by Tkatchenko et al;[124] however, use of the D3 model in MP2D makes the dispersion correction more straightforward to compute.

The MP2D dispersion correction computes the CKS and UCHF dispersion contributions using atom-centered dispersion coefficients which are inexpensively interpolated from a small set of *ab initio* dispersion coefficients computed for simple hydrides of the elements. Because MP2D employs atomic dispersion coefficients, the dispersion correction is applicable to both intra- and intermolecular interactions. Its performance for intermolecular interactions is similar to that of MP2C, and it has proved very useful in describing intramolecular interactions that prove difficult for many widely used density functionals, such as in conformational polymorphs of molecular crystals.[188,189]

The present study introduces spin-component-scaled MP2D (SCS-MP2D). Because the MP2D dispersion correction already addresses the MP2 problems for non-covalent interactions well, the spin-component scaling coefficients can focus solely on correcting the residual problems that impact MP2 performance for thermochemistry. The proposed SCS-MP2D model employs seven global empirical parameters, which is a modest number of parameters compared to many models derived from big data and machine learning approaches. Furthermore, the solid physical foundations of these parameters leads to a model that appears highly transferable to new systems, despite those empirical parameters being fitted to only a modest amount of benchmark data.

The addition of spin component scaling to MP2D makes it similar in many ways to the family of DSD double-hybrid density functionals.[83,165] Those density functionals

start from a hybrid functional DFT treatment with a modest fraction of exact exchange, mix in some amount of spin-component-scaled MP2-like correlation (evaluated in terms of the Kohn-Sham orbitals), and include long-range Grimme dispersion. SCS-MP2D has its foundation in HF (instead of DFT), includes spin-component-scaled MP2 correlation from which the long-range dispersion has been removed, and Grimme D3 dispersion. Starting from the exact exchange treatment in HF circumvents the problems of delocalization error that hinder the performance of many density functionals, including double hybrids, for charge transfer reactions, ionic hydrogen bonds, and certain intramolecular conformational energies. By subtracting the UCHF dispersion from the MP2 correlation energy before adding the Grimme dispersion correction, SCS-MP2D avoids any issues of double-counting dispersion energies that can hinder dispersion-corrected DFT models.

As will be demonstrated below, SCS-MP2D is competitive with some of the very best density functionals on a large set of benchmark data sets of organic species that span intermolecular interactions, conformational energies, and reaction energies. SCS-MP2D also performs well for two particularly challenging potential energy profiles. The cost of SCS-MP2D is effectively identical to that of MP2, and SCS-MP2D is applicable to systems with up to $\sim$100 atoms when density-fitting algorithms are used. Overall, the results presented below highlight how MP2-based wave function methods offer a viable route toward high-accuracy quantum chemistry in organic systems.

3.1 Theory

3.1.1 SCS-MP2D Energy

The canonical MP2 energy can be decomposed into the HF energy plus the same-spin (ss) and opposite-spin (os) correlation energies,

$$E_{MP2} = E_{HF} + c_{os}E_{corr}^{os} + c_{ss}E_{corr}^{ss} \tag{3.1}$$

In canonical MP2, the spin-scaling coefficients c_{os} and c_{ss} both equal one. Spin-component-scaled MP2 methods change those spin-scaling coefficients to improve the performance of the model. Grimme's original SCS-MP2 model employed $c_{os} = 6/5$ and $c_{ss} = 1/3$. As described earlier, however, these coefficients can vary considerably depending on the nature of the chemical system being studied.

One of the key problems in canonical MP2 lies in its treatment of van der Waals dispersion, such as its well-known over-estimation of the benzene dimer interaction energy and many other π-π interactions. The successful MP2C model addresses this for intermolecular interactions by subtracting out the UCHF dispersion that is inherent in MP2 and replacing it with a more reliable CKS description,[115,121]

$$E_{MP2C} = E_{MP2} - E_{disp}^{UCHF} + E_{disp}^{CKS} \tag{3.2}$$

However, the reliance on intermolecular perturbation theory for the dispersion correction limits MP2C to dimer intermolecular interactions, and the analytic nuclear gradients that would facilitate geometry optimizations are complicated and have not yet been implemented.

Our recently proposed dispersion-corrected MP2D model[5] addresses both limitations. MP2D adopts the same basic formalism as MP2C, but it computes the UCHF and

CKS dispersion contributions according to Grimme's D3 strategy,[123] which estimates the atom-centered dispersion coefficients via interpolation among pre-tabulated reference values for each element in different coordination environments. The D3 dispersion correction can be computed with trivial force-field like cost, is readily differentiated for analytic nuclear gradients, and is applicable to both intra- and intermolecular interactions.

SCS-MP2D combines spin-scaling of the correlation energy with a CKS dispersion correction. The SCS-MP2D energy is given by,

$$E_{SCS\text{-}MP2D} = E_{HF} + c_{os} E^{os}_{MP2\text{-}corr} + c_{ss} E^{ss}_{MP2\text{-}corr} - c_{os} E^{disp,os}_{UCHF} - c_{ss} E^{disp,ss}_{UCHF} + E^{disp,tot}_{CKS} \quad (3.3)$$

Because the MP2 correlation energies are scaled by the spin-scaling coefficients, the UCHF dispersion energy being removed from the correlation energy must also be scaled accordingly to obtain a "dispersion-free" SCS-MP2 energy. One is then free to add an appropriate dispersion treatment onto it—the D3 CKS dispersion energy in this case, without any spin-component scaling. Since the CKS dispersion contribution is unchanged compared to MP2D, the following discussion focuses on the UCHF contribution.

Spin-component scaling of the UCHF dispersion energy begins with partitioning the molecular frequency-dependent dipole-dipole polarizabilities into their α ($\uparrow$) and β ($\downarrow$) spin contributions,

$$\alpha_{\lambda\sigma}(i\omega) = \alpha^{\uparrow}_{\lambda\sigma}(i\omega) + \alpha^{\downarrow}_{\lambda\sigma}(i\omega) = \sum_{ia} \frac{2\epsilon_{ia}\langle i|\hat{\mu}_\lambda|a\rangle\langle a|\hat{\mu}_\sigma|i\rangle}{\hbar(\epsilon^2_{ia} + \omega^2)} + \sum_{ia} \frac{2\epsilon_{\bar{i}\bar{a}}\langle \bar{\imath}|\hat{\mu}_\lambda|\bar{a}\rangle\langle \bar{a}|\hat{\mu}_\sigma|\bar{\imath}\rangle}{\hbar(\epsilon^2_{\bar{i}\bar{a}} + \omega^2)} \quad (3.4)$$

where i and a are α spin occupied and virtual orbitals, $\bar{\imath}$ and $\bar{a}$ are the analogous β spin orbitals, ϵ_{ia} is the HF energy difference between orbitals i and a, $\hat{\mu}_\lambda$ is the λ-th component of the dipole moment operator, and $i\omega$ is the imaginary frequency at which the polarizability

is being evaluated. The isotropic frequency-dependent polarizability α_{iso} is computed as the trace of $\alpha_{\lambda\sigma}$ divided by 3.

In the D3 approach, the isotropic molecular frequency-dependent polarizabilities are computed and tabulated for a series of elemental hydrides A_mH_n (e.g. C_2H_6, C_2H_4, C_2H_2, CH, and C). The atomic frequency-dependent polarizability for atom A is then determined by subtracting out the hydrogen contributions and distributing the remaining polarizability evenly across the heavy atoms. For the spin up contributions, this takes the form:

$$\alpha_{iso}^{A\uparrow}(i\omega) = \frac{1}{m}[\alpha_{iso}^{A_mH_n\uparrow}(i\omega) - \frac{n}{2}\alpha_{iso}^{H_2\uparrow}(i\omega)] \tag{3.5}$$

An analogous expression can be written for the β ($\downarrow$) polarizabilities. The total UCHF C_6 dispersion coefficients for the interaction of atoms A and B can be computed from the spin-partitioned isotropic atomic frequency-dependent polarizabilities as,

$$C_6^{AB} = \frac{3}{\pi}\int_0^\infty d\omega \left(\alpha_{iso}^{A\uparrow}(i\omega) + \alpha_{iso}^{A\downarrow}(i\omega)\right)\left(\alpha_{iso}^{B\uparrow}(i\omega) + \alpha_{iso}^{B\downarrow}(i\omega)\right) \tag{3.6}$$

By multiplying out integrand and regrouping terms, one can partition the total UCHF C_6 coefficient into separate same-spin and opposite spin contributions,

$$\begin{aligned} C_6^{AB} &= \frac{3}{\pi}\int_0^\infty d\omega \left(\alpha_{iso}^{A\uparrow}(i\omega)\alpha_{iso}^{B\uparrow}(i\omega) + \alpha_{iso}^{A\downarrow}(i\omega)\alpha_{iso}^{B\downarrow}(i\omega)\right) \\ &\quad + \frac{3}{\pi}\int_0^\infty d\omega \left(\alpha_{iso}^{A\uparrow}(i\omega)\alpha_{iso}^{B\downarrow}(i\omega) + \alpha_{iso}^{A\downarrow}(i\omega)\alpha_{iso}^{B\uparrow}(i\omega)\right) \qquad (3.7) \\ &= C_6^{AB,ss} + C_6^{AB,os} \qquad (3.8) \end{aligned}$$

Once the same-spin UCHF, opposite-spin UCHF, and total CKS C_6 coefficients have been obtained for each atom type in each coordination number environment, the dispersion energies are computed according to the D3 scheme as described previously.[5,123]

Specifically, the final C_6 coefficients for a given atom in a particular chemical environment are interpolated using a slightly modified version[5] of the original D3 coordination number scheme, the C_8 coefficients are estimated as proscribed by the D3 model, and then the dispersion energy is obtained as,

$$E_{disp} = s_6 \sum_{AB} f_6(R_{AB}) \frac{C_{6,AB}}{R_{AB}^6} + s_8 \sum_{AB} f_8(R_{AB}) \frac{C_{8,AB}}{R_{AB}^8} \tag{3.9}$$

where s_6 and s_8 are scaling coefficients and R_{AB} is the distance between atoms A and B. The Tang-Toennies damping function $f_N(R_{AB})$ is given by,

$$f_N(R_{AB}) = 1 - \exp(s_R R_{AB}) \sum_{k=0}^{N} \frac{(s_R R_{AB})^k}{k!} \tag{3.10}$$

where N is the order of the dispersion term (6 or 8), and s_R is a distance scaling factor calculated from the cutoff radius $R_{0,AB}$ (taken from the D3 dispersion correction[123]) using two empirical parameters a_1 and a_2:

$$s_R = a_1 R_{0,AB} + a_2. \tag{3.11}$$

Finally, during the development of the original MP2D model, it was found that the Tang-Toennies damping function decays too slowly at covalent-bond distances.[5] To address this, the interatomic distance R_{AB} was modified via a secondary damping at very short distances according to,

$$R'_{AB} = \begin{cases} r_{cut} R_{0,AB} & \text{if } R_{AB} <= R_{0,AB}(r_{cut} - w/2) \\ R_{AB} & \text{if } R_{AB} >= R_{0,AB}(r_{cut} + w/2) \\ r_{cut} R_{0,AB} + g(R_{AB}, R_{0,AB} r_{cut}, R_{0,AB} w) & \text{otherwise;} \end{cases} \tag{3.12}$$

$$g(R_{AB}, r'_{cut}, w') = (-2.5x^8 + 10x^7 - 14x^6 + 7x^5) * w'; \tag{3.13}$$

$$x = \frac{R_{AB} - (r'_{cut} - w'/2)}{w'} \tag{3.14}$$

with empirical parameters r_{cut} and w defining the distance and width over which the damping occurs. This short-range damping leaves R_{AB} in Eq 3.9 unchanged at longer distances, while fixing it at a constant fraction of $R_{0,AB}$ value for very short distances. The polynomial $g(R_{AB}, r'_{cut}, w')$ smoothly interpolates between the two regimes. See ref 5 for more details.

For restricted wave functions, the spin up and spin down frequency-dependent polarizabilities in Eq 3.4 are identical and each equal to half the total polarizability. As a result, the same-spin and opposite-spin UCHF C_6 coefficients in Eqs 3.7 and 3.8 are each equal to one half to the total C_6 coefficient, and the spin components each contribute half of the UCHF dispersion energy E_{disp} (Eq 3.9),

$$E_{UCHF}^{disp,ss} = E_{UCHF}^{disp,os} = \frac{1}{2} E_{UCHF}^{disp,tot} \tag{3.15}$$

In this scenario, the final SCS-MP2D energy (Eq 3.3) can be expressed as,

$$E_{SCS\text{-}MP2D} = E_{HF} + c_{os} E_{MP2\text{-}corr}^{os} + c_{ss} E_{MP2\text{-}corr}^{ss} - \frac{1}{2}(c_{os} + c_{ss}) E_{UCHF}^{disp,tot} + E_{CKS}^{disp,tot} \tag{3.16}$$

Because the present study focuses only on closed-shell species with restricted wave functions for which the open-shell reference hydrides coordination environments contribute negligibly, Eq 3.16 represents the final equation implemented here.

The SCS-MP2D model contains seven empirical parameters in total: c_{os}, c_{ss}, s_8, a_1, a_2, r_{cut}, and w. The s_6 parameter scaling the C_6 dispersion energy contribution in Eq 3.9 is set to unity unless otherwise noted. The atom-pairwise dispersion coefficients C_6^{AB} coefficients from the reference hydrides are identical to those used in ref 5. More specifically, the CKS coefficients are taken directly from the original D3 model, while the UCHF ones were computed for H, B, C, N, O, F, Ne, P, S, Cl, Ar, and Br atoms according to the same Grimme D3 scheme. Accordingly, SCS-MP2D is presently applicable to typical organic and biological systems.

3.1.2 Empirical Parameter Fitting Procedure

The seven empirical parameters in MP2D were fitted to a total of 559 benchmark data points taken from the S66x8 set of dimer intermolecular interactions,[135] the Diels-Alder reaction energy (DARC) subset of the GMTKN55 data set,[145,190] and the sugar conformational energy (SCONF) subset of GMTKN55.[145] Because each of the three data sets differ in the number of data points contained and the magnitudes of the benchmark energy values, an objective function F was constructed from the weighted sum of the relative root-mean-square error (RMSE) for each data set. Relative RMSE values were obtained by dividing the RMSE of each set by the mean absolute value of the benchmark energies in the set, $\langle|E|\rangle$,

$$F = w_{S66x8}\frac{RMSE_{S66x8}}{\langle|E|\rangle_{S66x8}} + w_{DARC}\frac{RMSE_{DARC}}{\langle|E|\rangle_{DARC}} + w_{SCONF}\frac{RMSE_{SCONF}}{\langle|E|\rangle_{SCONF}} \tag{3.17}$$

Using relative RMSE values compensates for the fact that the Diels-Alder reaction energies are many-fold larger than the typical intermolecular or conformational energies. Different weights w_i for the three relative RMSEs in the objective function were tested during the fitting. By trial and error, it was determined that increasing the weight of the DARC data set in the objective function led to particularly good, transferable parameters. The final SCS-MP2D parameters were obtained with weight $w_{DARC} = 2.0$ and $w_{S66x8} = w_{SCONF} = 1.0$. The enhanced weight on the DARC set is consistent with the desire to use spin-component scaling to improve reaction energies and the importance of including short-range interactions in fitting the damping function parameters.

Initial exploratory optimizations of the empirical parameters revealed a rugged landscape containing many local minima. Therefore, a custom evolutionary algorithm was implemented to seek out (nearly) globally optimal parameters that minimize the objective function. Initial values of the parameters were generated randomly within a predefined range of plausibly physical values. Initial spin-component scaling coefficients and most other parameters were restricted to lie between 0 and 2, though a_2 was given the range -1 to 2. Each generation of the search algorithm was populated with 30–40 distinct parameter sets. Each parameter set was optimized using a quasi-Newton algorithm in Cuby4.[129]. This gradient-based optimization converged slowly, but it was observed that the parameters varied little after the first ten optimization cycles; therefore, 10 cycles were used for each optimization during the evolutionary search.

After randomly seeding the initial generation, subsequent generations were created as a mixture of, for example: the five best-performing parameter sets from the previous generation, five parameter sets obtained by randomly combining parameters from the top five performers (inheritance), five sets where new damping function parameters were generated for the top 5 performers (mutation), 5 populations where the spin-component constants were replaced by new randomly generated constants (mutation), and 10 entirely new randomly generated parameter sets to add diversity to the population. Optimization runs which varied the partitioning among inheritance, mutation, and random generation and the total population size were explored. The genetic algorithm was allowed to run for between 5 and 20 generations. To avoid biasing the search in favor of the top performing parameters against newly generated parameter sets, each generation passed initial starting parameters to the next generation rather than the optimized parameters. Once the optimal parameters were found after many searches, they were fully-optimized to ensure a minimum had been reached in the parameter landscape (though their values changed only minimally). Convergence of the evolutionary optimization algorithm was tested both by (1) performing dozens of independent runs of the evolutionary optimizer, and (2) by seeding a Bayesian search algorithm with good parameter sets from the genetic algorithm (see Appendix B Section B.1). The final parameter set discussed in Section 3.3.1 was discovered relatively early and repeatedly in the search process, and subsequent searching did not reveal any better-performing parameter sets.

3.2 Computational Methods

The MP2D and SCS-MP2D dispersion corrections were calculated using a developmental version of the freely available Cuby4 software.[129] Energies were computed at the complete basis set limit by combining HF/aug-cc-pVQZ with correlation energies extrapolated[70] from aug-cc-pVTZ and aug-cc-pVQZ results. Counterpoise corrections were employed for all benchmark sets involving purely intermolecular interactions: S66x8, 3B-69, SSI, HBC6, NBC10, Charge Transfer, HB375, and IHB100. MP2 data for S66x8, 3B-69, SSI, and IDISP was taken from the original sources; data for all other sets was computed here using PSI4 version 1.3.[137] For ISOL24, the 24th isomerization reaction was omitted because MP2D and SCS-MP2D dispersion coefficients have not been computed for silicon. For consistency, reaction 24 was excluded for all other tested methods as well.

DFT calculations were performed using the DSD-BLYP-D3(BJ), ωB97X-V, and ωB97M-V functionals in PSI4 and the revDSD-PBEP86-D3(BJ) functional in Orca version 4.2.[191] The PSI4 DSD-BLYP-D3(BJ) calculations employ the parameters reported in ref 165, rather than the earlier parameters[144,154] which were used in the GMTKN55 benchmarks.[145] Counterpoise corrections were employed for all intermolecular benchmark sets, except for the revDSD-PBEP86-D3(BJ) functional, for which the counterpoise-corrected results were substantially worse than the uncorrected ones. In general, DSD-BLYP-D3(BJ), revDSD-PBEP86-D3(BJ), and ωB97X-V calculations were performed using the def2-QZVP basis set, while the ωB97M-V ones were performed in aug-cc-pVQZ since the latter basis set is one of the recommended ones from Ref 163. There are a few exceptions, however: For the SSI data set, aug-cc-pVTZ results were taken from Burns et al[2] for ωB97X-V and

ωB97M-V. The aug-cc-pVQZ basis set was used with DSD-BLYP-D3(BJ) and revDSD-PBEP86-D3(BJ) for the SSI set due to the presence of anionic species. The aug-cc-pVQZ basis set was similarly employed for all functionals on the IHB100 set of ionic species. Finally, ωB97X-V results for SCONF, ACONF, Amino20x4, MCONF, PCONF21, DARC, ISOL24, ISO34, and IDISP were taken directly from the GMTKN55 database.[145]

The genetic optimization algorithm was implemented by the authors. The searches employing Bayesian optimization with Gaussian processes utilized the python scikit-learn library `skopt.gp_minimize`.[192]

3.3 Results and Discussion

3.3.1 Parameter optimization

The seven empirical parameters were optimized using the evolutionary algorithm discussed in Section 3.1.2, and the final SCS-MP2D model parameters are listed in Table 3.1. Figure 3.1 shows the progress over a single run of the evolutionary optimizer. As the algorithm proceeds through the generations, the population is enriched with low-error parameter sets. By the ninth generation, quite a few parameter sets have been found that perform well, and these best-performing models change little over the next five generations. As shown in the radar plot in Figure 3.1, the five best-performing members of the population in this optimization run in blue exhibit parameters that are quite similar to those in the final SCS-MP2D model in red. The only appreciable variations occur for the a_2 Tang-Tonnies damping parameter. The next five best-performing parameter sets in green

Table 3.1: Comparison of the MP2D parameters from ref 5 and the SCS-MP2D ones determined here. All parameters are dimensionless except for a_2.

	MP2D	SCS-MP2D
c_{os}	1	0.8263
c_{ss}	1	0.9004
a_1	0.9436	1.5359
a_2 (Å)	0.4802	-0.7595
s_8	1.1873	1.2092
r_{cut}	0.72	0.8254
w	0.20	0.1198

differ a little more from the SCS-MP2D ones, most notably in s_8 and a_2, but they are again fairly similar. Moving beyond the ten best parameter sets from this search, one finds greater parameter diversity, indicating that algorithm is searching widely. Overall, many independent runs of the optimizer generated parameter sets that are similar to the final SCS-MP2D ones, and they were always among the very best performing models.

Some searches revealed a few alternative parameter sets that also performed very well, though they had unphysical parameters such as a negative s_8 value. In those models, the MP2D dispersion correction had the wrong sign, with the CKS term effectively increasing the van der Waals binding energy compared to UCHF for systems such as the π-stacked benzene dimer. This behavior is contrary to the well-known behaviors from intermolecular perturbation theory where UCHF over-binds such systems.[114,121] Other tests that allowed

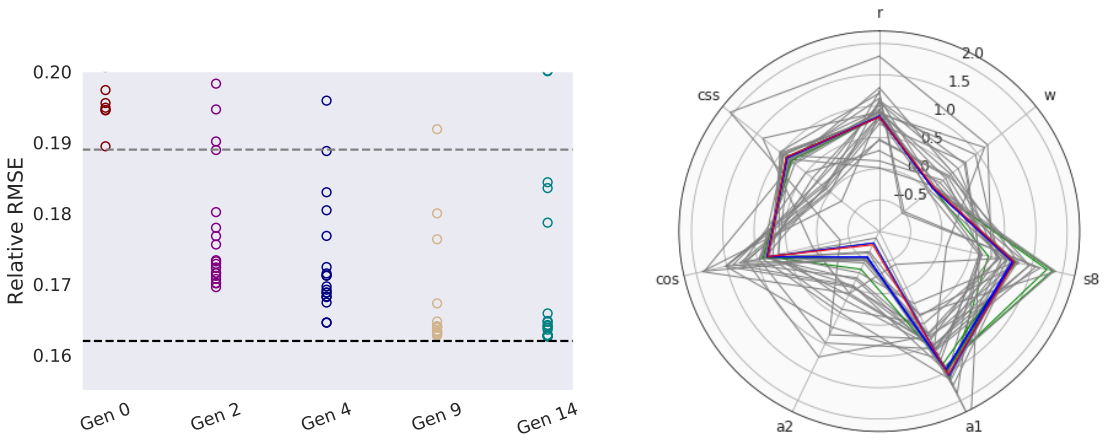


Figure 3.1: Left: Performance of one run of the evolutionary SCS-MP2D parameter optimization, plotting the relative root-mean-square error (the objective function) versus the generation number. The figure focuses only on the low-error region, and the horizontal lines indicate the best-performing models in the initial (Gen 0) and final (Gen 14) generations. Right: Radar plot comparing the final SCS-MP2D parameter values (red) to those of the five lowest-error parameter sets (blue), the next five lowest-error models (green), and all other parameter sets (gray) from this particular optimization.

s_6 to deviate from unity (increasing the number of parameters from seven to eight) produced good-performing parameter sets with $s_6 > 1$ and $s_8 \approx 2$. From intermolecular perturbation theory, s_6 and s_8 should both equal one, though the D3 model typically allows s_8 to deviate from unity to compensate for the neglect of higher-order dispersion terms. Given the rapid decay of those higher-order terms with distance, however, it seems unlikely that those neglected contributions should effectively double the s_8 contribution. When tested for transferability to other data sets not employed in the parameter fitting, the final chosen parameter set in Table 3.1 performed as well as or better than any of these alternative

parameter sets. Accordingly, the final chosen parameter set in Table 3.1 was selected on the basis of its performance on the training set, physically reasonable parameter values, and its transferability to other benchmark sets (Section 3.3.2).

Table 3.1 compares the final SCS-MP2D parameters against those published previously for MP2D, which does not scale the spin components of the correlation energy. The SCS-MP2D dispersion correction parameters are fairly similar to those in MP2D. For example, s_8 differs by only 2% between the two models. Values of s_8 near 1.2 are intermediate relative to the range of $s_8 \sim 0.8$–1.7 typically found for D3 with various density functionals.[123] The parameters r_{cut} and w that govern damping at very short (covalent) distances differ by $\sim$0.1 between MP2D and SCS-MP2D. However, as discussed in the original MP2D study,[5] a relatively broad range of parameters r_{cut} and w performs well, and the differences in these parameters between the two models has a small impact on the overall performance.

In contrast, the SCS-MP2D Tang-Toennies dispersion damping parameters differ noticeably from the earlier MP2D ones. Compared to MP2D, the new parameters enhance the SCS-MP2D dispersion correction contribution at shorter distances (Appendix B Section B.2). This increased contribution from the dispersion correction offsets the diminished contribution of the MP2 correlation energy that results from having spin-component scaling coefficients less than one. Interestingly, the spin-scaling coefficients $c_{ss} = 0.8263$ and $c_{os} = 0.9004$ are similar to each other, in contrast to many previous SCS-type MP2 models. Furthermore, the SCS-MP2D spin-component scaling enhances the same-spin contribution relative to the opposite-spin one, which is contrary to theoretical arguments that MP2 typically overestimates the same-spin correlation more than the opposite-spin contributions.[169]

On the other hand, such arguments may no longer apply when such a sizable fraction of the final SCS-MP2D correlation energy originates from the CKS dispersion correction.

The importance of the CKS dispersion can be seen from a few example systems. Along the eight points of the S66x8 π-stacked benzene dimer potential energy curve, for instance, an average 93% of the SCS-MP2D correlation energy arises from the CKS dispersion energy. Only $\sim$7% stems from what remains of the spin-component-scaled MP2 correlation energy after subtracting out the UCHF contribution. Of course, dispersion is expected to be very important for the benzene π dimer. However, even in the DARC data set chemical reaction energies for which non-dispersion components of the correlation energy are more important, the CKS dispersion still contributes an average 24 kcal/mol, compared to only 10 kcal/mol from the residual spin-component-scaled MP2 correlation energy that remains after removing the UCHF dispersion component.

That said, a few additional considerations should be noted. First, there are multiple ways to decompose and group the SCS-MP2D energy components. If one partitions it into the SCS-MP2 correlation energy plus a dispersion correction (computed as the difference between the UCHF and CKS dispersion energies), the dispersion correction amounts to only 0.5 kcal/mol for benzene dimer at its equilibrium geometry and an average of 1 kcal/mol (a few percent) for the reaction energies in the DARC set. In other words, the individual UCHF and CKS dispersion terms are very large individually, but the difference between the two contributions is far smaller. Finally, note that the parameters were fitted to the total interaction, conformational, or reaction energies, rather than to individual components of the correlation energy. As discussed for MP2D,[5] this choice leads to good overall

performance with fewer empirical parameters, but the individual UCHF and CKS components in SCS-MP2D do not quantitatively reproduce those from intermolecular perturbation theory and/or MP2C.

3.3.2 Performance on Benchmark Data Sets

As discussed above, SCS-MP2D was trained on three datasets consisting of noncovalent interactions (S66x8), reaction energies (DARC), and sugar conformational energies (SCONF). MP2D was fitted against S66x8.[5] To assess overall performance and transferability, SCS-MP2D was tested on 14 additional benchmark data sets for intermolecular interactions (3B-69 dimers,[193] NBC10,[194] HBC6,[194] HB375,[195] IHB100,[195] SSI,[2] & charge transfer reactions[196]), conformational energies (Amino20x4,[145] ACONF,[197] MCONF,[198] and PCONF21[145]), and thermochemical reaction energies (ISO34,[199] ISOL24,[145] IDISP[145]). Note that IDISP contains a mixture of interaction types that all involve substantial changes in intramolecular dispersion energy, but it is grouped with the reaction energy data sets here because four of the six examples involve chemical reactions and/or isomerizations. Table 3.2 summarizes root-mean-square errors for each data set as computed with several different quantum chemistry models. Figure 3.2 plots the relative RMSEs, where the RMSE for each model is divided by the mean absolute value of the target reference energies for that data set.

Table 3.2: Root mean square errors calculated relative to the benchmark reference values (kcal/mol). The asterix (*) indicates data sets that were used to fit SCS-MP2D. Cell color indicates the relative RMSE: dark blue $\leq$ 5%, light blue = 5–10%, orange = 10–25%, light red = 25–50%, and dark red $\geq$ 50%.

Data Set	MP2 CBS	MP2D CBS	SCS-MP2D CBS	DSD-BLYP -D3(BJ) def2-QZVP	revDSD-PBEP86 -D3(BJ) def2-QZVP	ωB97X-V def2-QZVP	ωB97M-V aQZ
Intermolecular Interactions							
S66x8	0.67	0.16	0.13*	0.18	0.16	0.21	0.11
3B-69 Dimers	0.33	0.21	0.18	0.19	0.19	0.20	0.17
SSI	0.36	0.16	0.17	0.15^a	0.12	0.16^b	0.15^b
HBC6	0.32	0.26	0.25	0.37	0.17	0.32	0.24
NBC10	1.55	0.29	0.14	0.33	0.07	0.34	0.17
Charge Transfer	2.72	0.56	0.34	0.77	0.62	0.57	0.45
HB375	0.43	0.16	0.13	0.14	0.13	0.17	0.19
IHB100	0.45	0.46	0.34	0.41^a	0.26^a	0.37^a	0.35
Conformational Energies							
SCONF	0.31	0.35	0.18*	0.26	0.13	0.21^c	0.24
ACONF	0.11	0.07	0.12	0.08	0.24	0.06^c	0.08
Amino20x4	0.26	0.17	0.18	0.16	0.17	0.24^c	0.24
MCONF	1.02	0.40	0.33	0.55	0.19	0.27^c	0.39
PCONF21	1.11	0.42	0.31	0.48	0.23	0.35^c	0.69

Data Set	MP2 CBS	MP2D CBS	SCS-MP2D CBS	DSD-BLYP -D3(BJ) def2-QZVP	revDSD-PBEP86 -D3(BJ) def2-QZVP	ωB97X-V def2-QZVP	ωB97M-V aQZ
Reaction Energies							
DARC	3.97	1.90	1.41*	1.10	0.64	4.38[c]	0.98
ISO34	1.68	1.42	0.96	1.06	0.49	1.56[c]	0.82
ISOL24	3.72	2.81	2.24	2.71	1.73	4.20[c]	2.41
IDISP	7.03	1.42	1.29	1.60	0.67	3.88[c]	2.83
Overall Relative RMSE Statistics (%)							
Mean	21.3	7.7	5.9	8.1	4.9	9.5	8.2
Median	11.5	5.8	4.4	5.5	4.1	5.3	4.5

[a] aug-cc-pVQZ basis [b] Ref 2, aug-cc-pVTZ basis. [c] Ref 145

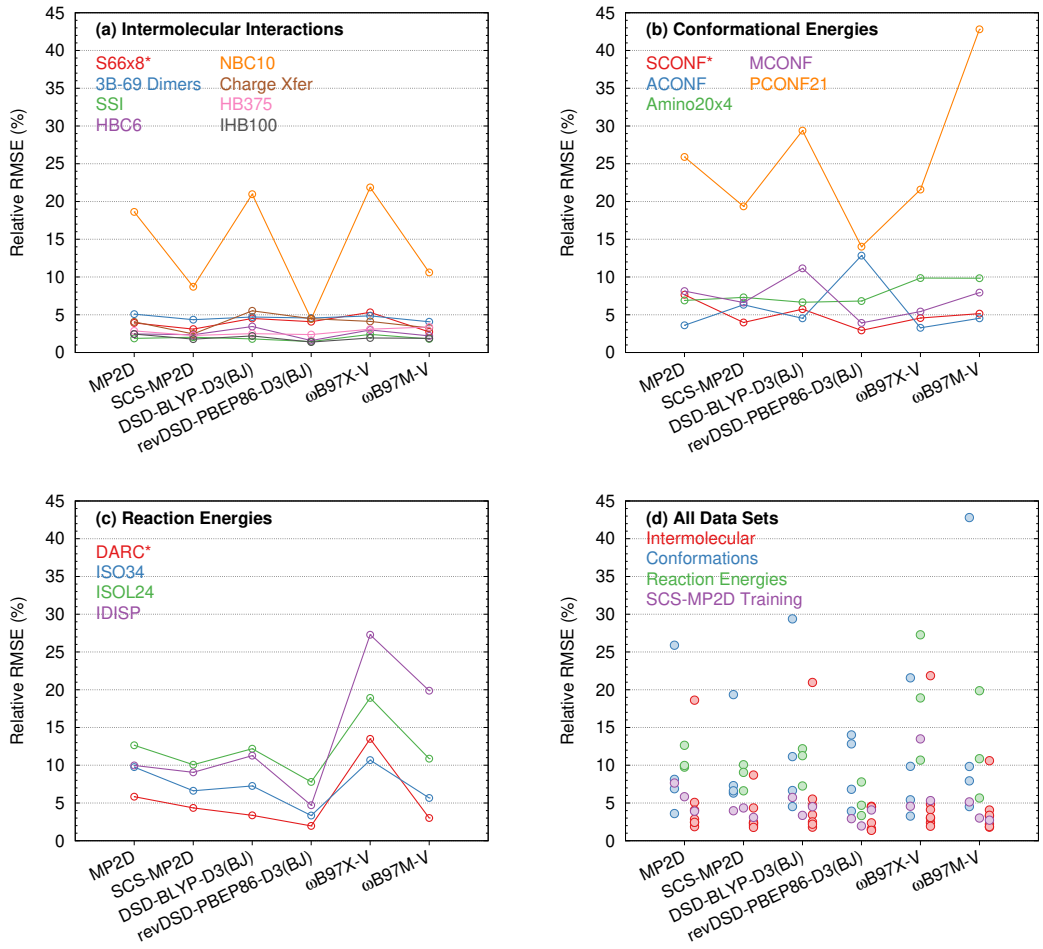


Figure 3.2: Comparison of the relative RMSEs for several models on benchmark data sets focusing on (a) intermolecular interactions, (b) conformational energies, (c) reaction energies, and (d) the union of all training and testing sets. The asterix indicates that S66x8, SCONF, and DARC were involved in fitting the SCS-MP2D parameters.

Performance of wave function methods

Consider first the performance of MP2 and MP2D for the intermolecular interaction data sets. The MP2D dispersion correction seeks to address systems like the benzene dimer where dispersion is important and which are often considerably over-bound by MP2. The dispersion correction has minimal impact on hydrogen-bonded systems like the water dimer, for which MP2 already performs fairly well. Overall, MP2D reduces the MP2 RMSE on S66x8 four-fold, from 0.67 to 0.16 kcal/mol. That improvement partly reflects that MP2D was trained against S66x8 benchmark data, but the MP2D parameters also prove highly transferable to other benchmark sets. The MP2D dispersion correction provides several-fold error reductions in the non-bonded potential energy curves of NBC10, the protein side-chain side-chain interactions of SSI, the large database of hydrogen bonds in HCNO-containing species (HB375), and a set of charge transfer reactions. Smaller MP2D improvements occur in 3B-69 and the hydrogen-bonded dimer curves of HBC6, and no appreciable improvement is found for the ionic H-bonds of IHB100. The smaller improvements seen in those latter sets largely reflects the lesser importance of dispersion interactions in those dimers rather than any weaknesses in MP2D.

Incorporating spin-component-scaling into MP2D leads to further modest improvements for the intermolecular interaction data sets. In S66x8 (a training set), the RMSE reduces from 0.16 to 0.13 kcal/mol. Improvements are observed for many of the testing sets as well. For example, spin-component-scaling reduces the NBC10 RMSE by a factor of two, from 0.29 kcal/mol for MP2D to 0.14 kcal/mol for SCS-MP2D. More typically, SCS-MP2D reduces the MP2D errors by around a third or less for many of the intermolecular

interaction sets. Spin-component scaling tends to improve the MP2D performance on ionic and hydrogen-bonded systems, even though those generally are reasonable even with MP2 and MP2D already. Surprisingly, however, SCS-MP2D also performs better than MP2D for systems like the benzene dimer, which explains much of the improvement observed for NBC10. As discussed previously,[5] the D3 correction has the weakness that its highly local interpolation scheme for the C_6 dispersion coefficients distinguishes poorly between benzene and ethene, for example. So while MP2D performs well for many systems where dispersion is important, its performance for the benzene dimer is actually somewhat worse than the fully *ab initio* treatment in MP2C (though MP2D still improves dramatically upon MP2). SCS-MP2D suffers from the same limitations of the dispersion coefficients, but apparently the spin-component-scaling compensates somewhat. For instance, the RMSE for the π-stacked benzene dimer decreases from from 0.33 kcal/mol with MP2D to 0.13 kcal/mol with SCS-MP2D. SCS-MP2D performs marginally worse than MP2D on the SSI data set; this issue will be explored in detail in Section 3.3.2.

One of the key strengths of MP2D and SCS-MP2D over MP2C is that the atom-pairwise definition of the dispersion correction allows for correcting both intra- and intermolecular dispersion. Because it is based on intermolecular perturbation theory, the MP2C dispersion correction has no effect on intramolecular conformational energies. Intramolecular dispersion corrections can be essential in systems such as the conformational polymorphs of organic crystals.[188,189]

Looking at the six conformational energy data sets (1 training and 5 testing), the MP2D and SCS-MP2D performance trends are similar to what was observed for the

intermolecular interaction data sets. The MP2D dispersion correction already improves upon MP2 nicely. Adding the dispersion correction reduces the MP2 RMS errors from 1.11 to 0.41 kcal/mol in peptide conformers (PCONF21) and from 1.02 to 0.42 kcal/mol in melatonin conformers (MCONF). Including spin-component-scaling reduces those errors ∼20–25% further. Both MP2D and SCS-MP2D perform about one third better than MP2 for the amino acid conformations in Amino20x4. The alkane conformations in (ACONF) are the only data set here where SCS-MP2D (RMSE 0.12 kcal/mol) does not improve upon MP2 (0.11 kcal/mol) and is somewhat worse than MP2D (0.07 kcal/mol). Fortunately, these errors are small in both absolute and relative terms (e.g. 6.3% for SCS-MP2D).

Finally, we examine the reaction energies associated with Diels-Alder reactions (DARC, training set), the smaller- and larger-molecule isomerizations (ISO34, ISOL24), and the IDISP set, which contains several dimerzation and isomerization reactions for which intramolecular dispersion matters. Non-dispersion contributions to the correlation energy are expected to be sizable for reaction energies, so spin-component scaling might be expected to have a significant impact in these data sets. Indeed, while MP2D did improve upon MP2 for all four data sets, SCS-MP2D performs even better. For example, MP2D (1.42 kcal/mol) only improved upon MP2 (1.68 kcal/mol) by about 15% for the ISO34 small-molecule isomerizations. The MP2D dispersion correction has a slightly larger impact on the larger-molecule isomerizations of ISOL24, reducing the MP2 error by ∼ 25% (3.72 to 2.81 kcal/mol). In both cases, however, SCS-MP2D reduces the errors by ∼40% compared to MP2, with RMSE values of 0.96 and 2.24 kcal/mol, respectively. For DARC, which was included in the SCS-MP2D fitting, SCS-MP2D reduces the MP2 error by 65%, and it

reduces the MP2D error by ∼25%. The IDISP set tends to exhibit considerable variability in the error statistics achieved by different models, due to the diverse chemistry and the disparate energy scales for the different reactions. Regardless, both MP2D and SCS-MP2D perform very well for this set, with RMS errors of 1.42 and 1.29 kcal/mol, respectively.

As a whole, these benchmark results demonstrate that the SCS-MP2D model is highly transferable to a wide variety of organic chemistry, despite being fitted to a modest amount of training data. In most of the benchmarks performed here, the MP2 dispersion contributions account for the largest share of the improvement, but the spin-component scaling almost always improves the quality of the predicted energies further. Given the error statistics presented here, SCS-MP2D is arguably one of the best-performing $O(N^5)$ correlated wave function methods available today for describing intra- and intermolecular interactions in organic chemistry. However, a better understanding of its overall performance requires comparing it to state-of-the-art density functionals.

Comparisons to selected density functional models

Four top-performing density functionals were chosen for comparison against the dispersion-corrected MP2 models: the range-separated hybrid functional ωB97X-V, the meta-GGA variant ωB97M-V, and the double-hybrid spin-component-scaled functionals DSD-BLYP-D3(BJ) and revDSD-PBEP86-D3(BJ). The ωB97X-V and DSD-BLYP-D3(BJ) functionals were selected based on their excellent performance on the GMTKN55 test suite,[145] while ωB97M-V was selected because it represents the meta-GGA rung on Jacob's ladder of density functionals and generally performs even better than ωB97X-V.[161,163]

The recently revised DSD functionals (revDSD) perform even better than the original DSD functionals.[83] The improvements stem primarily from replacing the D3 dispersion correction with the newer D4 one[200] in some of the functionals and from fitting the empirical parameters to a much larger set of training data. The revDSD-PBEP86-D3(BJ) functional was selected as a representative example of these new functionals. Because SCS-MP2D could plausibly be developed based on the D4 correction instead of the D3 one as well, we opted to compare against the D3 version of revDSD-PBEP86 for the sake of consistency. On the GMTKN55 data set, the revDSD-PBEP86-D3(BJ) functional performs about 0.1 kcal/mol worse than the D4 version in the weighted mean absolute deviation.[83]

The double-hybrid ωB97M(2) functional[164] would be another interesting potential comparison, since it performs noticeably better than the ωB97X-V and ωB97M-V functionals tested here and even slightly better than the revDSD functionals (e.g. $\sim$0.1 kcal/mol better than revDSD-PBEP86-D4 for GMTKN55).[161] However ωB97M(2) is not presently implemented in any of the software packages used here. Overall, the four functionals selected here are representative of top-performing functionals in their respective categories.

To facilitate comparisons between the MP2-based methods and the DFT functionals, Table 3.2 employs color-coding based on the relative RMS errors (i.e. RMSE divided by the average magnitude of the benchmark energy in each set). Dark blue corresponds to relative RMSEs of 5% or less, light blue to relative RMSEs in the range 5–10%, orange to those in the range 10–25%, light red for the range 25–50%, and dark red for larger relative RMSEs. Figure 3.2 plots the relative RMSEs for all models except MP2. MP2 is

omitted from Figure 3.2 because its large errors would obscure the comparison among the better-performing methods.

Considering first the intermolecular interactions, Figure 3.2a highlights how most of these dispersion-corrected MP2 and DFT models perform well, but revDSD-PBEP86-D3(BJ), SCS-MP2D and ωB97M-V are the clearly the top performers. The most noticeable differences among the models occur for NBC10 and the charge transfer set. In NBC10, revDSD-PBEP86-D3(BJ) (RMSE 0.07 kcal/mol), SCS-MP2D (RMSE 0.14 kcal/mol) and ωB97M-V (0.16 kcal/mol) exhibit errors that are a factor of 2–3 times smaller than those for MP2D and the other two functionals. For the charge transfer set, all four functionals perform noticeably worse than SCS-MP2D, especially DSD-BLYP-D3(BJ). Delocalization error in approximate functionals hinders the description of such systems. The inclusion of exact exchange and/or range-separation in these functionals reduces delocalization error considerably, but the RMSEs for this set remain appreciably larger than for any of the other intermolecular interaction data sets.

The performance of SCS-MP2D is also competitive with these density functionals for the conformational energy data sets. No single model performs uniformly well across all five test sets, but SCS-MP2D and revDSD-PBEP86-D3(BJ) exhibit the most consistent performance. The revDSD-PBEP86-D3(BJ) functional performs noticeably better for SCONF, MCONF, and PCONF21, about the same for Amino20x4, and appreciably worse for ACONF. However, SCS-MP2D either performs better than or is on par with the other three functionals. ACONF provides the most notable exception, with the other three functionals performing quite a bit better than SCS-MP2D and revDSD-PBEP86-D3(BJ). The

peptide conformations in PCONF21 are interesting for two reasons. First, due to the small average conformational energies, the relative RMSE values for most models are more than double those of the other sets, even if the absolute RMSE values are similar to those in MCONF. Second, while the ωB97M-V functional generally performs very well for conformational energies, its 0.69 kcal/mol RMSE for PCONF21 is roughly double that of SCS-MP2D (0.32 kcal/mol) and noticeably worse than the other three functionals (0.23–0.48 kcal/mol).

The behavior of the various models for the reaction energy data sets follows similar patterns. SCS-MP2D exhibits root-mean-square errors that are typically $\sim$20–70% smaller than those for DSD-BLYP-D3(BJ) and ωB97X-V for DARC, ISO34, ISOL24, and IDISP. The only exception is that SCS-MP2D and DSD-BLYP-D3(BJ) perform about the same on IDISP (1.29 and 1.31 kcal/mol RMSE, respectively). The comparison between SCS-MP2D and ωB97M-V is more mixed, with SCS-MP2D performing better on ISOL24 and IDISP, and ωB97M-V giving smaller errors for DARC and ISO34. As shown in Figure 3.2c, SCS-MP2D does notably provide much more consistent relative errors than ωB97M-V, however (Figure 3.2). Finally, revDSD-PBEP86-D3(BJ) is the clear winner for these reaction energies, with errors that are often only half those of SCS-MP2D.

Figure 3.2d aggregates the relative errors for all methods except MP2 on all the data sets. It highlights how SCS-MP2D, MP2D, and all of the density functionals examined here generally perform well. The best-performing model is revDSD-PBEP86-D3(BJ), but SCS-MP2D is only moderately worse. The general consistency of both models across the different data sets is particularly notable. This can also be seen from the mean and median statistics of the relative RMSEs for all data sets in Table 3.2: revDSD-PBEP86-D3(BJ)

exhibits the smallest mean and median errors of 4.9% and 4.1%, respectively. SCS-MP2D performs a little worse at 5.9% (mean) and 4.4% (median), and ωB87M-V is close behind with a median error of 4.5%, though its mean error of 8.2% is much larger due to its poor performance on PCONF21. The statistics also highlight that MP2D is fairly competitive with DSD-BLYP-D3(BJ) and ωB87X-V, but those three models represent a noticeable decrease in accuracy compared to the best three.

In other words, SCS-MP2D is highly competitive with some of the best density functionals on these benchmark sets. Based on earlier benchmarks, one anticipates that revDSD-PBEP86-D4 and ωB87M(2) would perform even a little better than any of the models here. On the other hand, there are some very encouraging features of SCS-MP2D here. First, while SCS-MP2D has 7 empirical parameters and the DSD functionals have six, ωB97X-V has 10 and ωB97M-V has 12. The fact that SCS-MP2D exhibits good performance and transferability with a modest number of empirical parameters speaks well to the physical foundations of the model. Second, substantial error reduction was obtained with the revDSD functionals compared to the original DSD versions by optimizing the empirical paramaters against a much larger data set.[83] This raises the prospect that a similar strategy might lead to further improvements for SCS-MP2D as well.

Anion-Anion interactions in the SSI data set

Despite overall good performance on the SSI data set[2] (RMSE of 0.17 kcal/mol), SCS-MP2D actually performs slightly worse than the original MP2D method (0.16 kcal/mol). The subset breakdown in Figure 3.3 makes clear that the interactions involving anions, es-

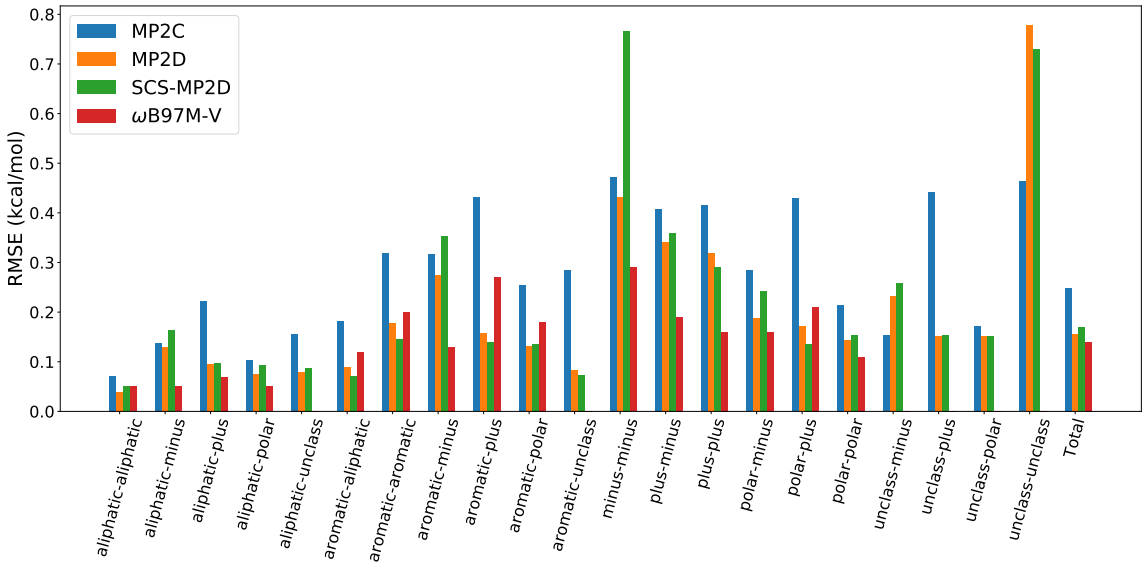


Figure 3.3: Root mean square error of MP2C, MP2D, SCS-MP2D, and ωB97M-V for the different interaction categories in SSI relative to the DW-CCSD(T)-F12 reference values of ref 2.

pecially the anion-anion subset, are the primary driver of this larger RMSE. In fact, MP2 actually performs better than SCS-MP2, MP2D, and MP2C on this anion-anion subset, which indicates there may be a general error in the dispersion correction scheme used in these methods for ionic species (Table 3.3). This behavior contrasts the results of the IHB100 data set of 100 hydrogen-bonded ion pairs, where SCS-MP2D performs somewhat better than either MP2 or MP2D (Table 3.2).

The SSI reference data uses Sherrill's silver standard DW-CCSD(T**)-F12 approach. For comparison, we also computed the energies using conventional CCSD(T)/CBS as computed from MP2/aug-cc-pV[TQ]Z and CCSD(T)/aug-cc-pVDZ. Changing the reference data reduces the SCS-MP2D RMSE modestly from 0.77 to 0.65 kcal/mol, but it does

not alter the fundamental story that SCS-MP2D is performing worse than MP2D for the anion-anion interactions.

For comparison, Table 3.3 also compares the performance of several density functionals on the anion-anion subset. With an RMSE of 0.73 kcal/mol, ωB97X-V performs comparably to SCS-MP2D. revDSD-PBEP86-D3(BJ) and DSD-BLYP-D3(BJ) perform only moderately better at 0.62 and 0.50 kcal/mol, respectively. In contrast, ωB97M-V performs outstandingly with an RMSE of only 0.29 kcal/mol. These anion-anion interactions represent only a small fraction of the full set, however, and the different models exhibit much smaller variations in RMSE across the full SSI data set.

Table 3.3: Root-mean-square errors for various models in the anion-anion interaction subset of SSI compared to those for the full SSI data set, using the benchmark DW-CCSD(T**)-F12 reference values of ref 2.

Method	Anion-Anion Subset	Full SSI Set
MP2/CBS (Ref 5)	0.36	0.36
MP2C/CBS (Ref 5)	0.43	0.12
MP2D/CBS (Ref 5)	0.43	0.16
SCS-MP2D/CBS (this work)	0.77	0.17
DSD-BLYP-D3(BJ)/aug-cc-pVQZ (this work)	0.50	0.15
revDSD-PBEP86-D3(BJ)/aug-cc-pVQZ (this work)	0.62	0.12
ωB97X-V/aug-cc-pVTZ (Ref 2)	0.73	0.16
ωB97M-V/aug-cc-pVTZ (Ref 2)	0.29	0.15

One possible source of the SCS-MP2D behavior (and the two DSD functionals, to a lesser extent) could lie in the D3 dispersion correction. The D3 dispersion coefficients are interpolated from neutral hydrides based on the geometry-dependent coordination number. They do not, however, directly differentiate between neutral and ionic environments. Anions tend to be more polarizable and likely exhibit stronger dispersion interactions that are perhaps not handled ideally in the MP2D and SCS-MP2D models. The newer D4 dispersion correction[200] accounts for the effect of atomic charge in the coordination numbers used to calculate the dispersion coefficients. Indeed, the performance of the DSD-family of double-hybrid functionals improves noticeably over a variety of benchmark sets when the D4 correction is used in place of D3.[83] Perhaps a version of the MP2D-type methods based on D4 would perform better for these anion-anion interactions. On the other hand, it is also worth noting that MP2C, which computes the UCHF and CKS dispersion contributions from first principles, still performs worse than canonical MP2 and ωB97M-V. In other words, the worse performance of the dispersion-corrected MP2 models is probably not entirely due to the D3 dispersion correction. Exchange-dispersion and induction-dispersion couplings are also important in ionic interactions,[201] and it is possible that the CKS dispersion correction and/or spin-component scaling in the MP2D-type methods disrupt some favorable error cancellation between the UCHF dispersion and those other terms.

3.3.3 Two Challenging Examples

The data sets examined above provide a broad perspective for the performance of SCS-MP2D relative to other models, but it can also be instructive to look at specific, challenging systems. Here we focus on two: the torsional scan about the key dihedral angle

in the ROY molecule[188,202,203] and the dissociation of the anthracene photodimer.[5,155,204] Both have proved challenging for common GGA and hybrid density functionals, and comparing how several top-performing models behave on these potential energy curves provides further insights into their capabilities.

ROY dihedral angle scan

The ROY molecule (Figure 3.4a) holds the current record for the largest number of fully characterized crystal polymorphs.[3,205–209] These polymorphs exhibit vibrant red, orange, or yellow crystals, depending on the degree of conjugation between the two aromatic rings as governed by the dihedral angle θ_{thio} (Figure 3.4). The 12 characterized polymorphs lie within a narrow ∼1 kcal/mol energy window, and even modest failures to predict these conformational energies correctly inhibits accurate energy ranking of the different crystal polymorphs.[188,189,202,203,205] GGA density functionals such as B86bPBE-XDM typically predict the thermodynamically preferred Y polymorph to be one of the least stable crystal forms (Figure 3.5). Compared to CCSD(T) benchmarks, MP2D predicts the conformational energies much more reliably than conventional GGA and hybrid functionals, and it is one of the relatively few methods that has been shown to predict the polymorph stabilities largely correctly.[188,189]

Figure 3.4 plots the one-dimensional conformational energy scan for θ_{thio}. The geometries were taken from ref 188, where they were obtained by constraining the dihedral angle at different angles 0–150° (in 10° steps) and relaxing all other degrees of freedom using B3LYP-D3(BJ)/def2-TZVP. Single-point energies were then computed on these geometries

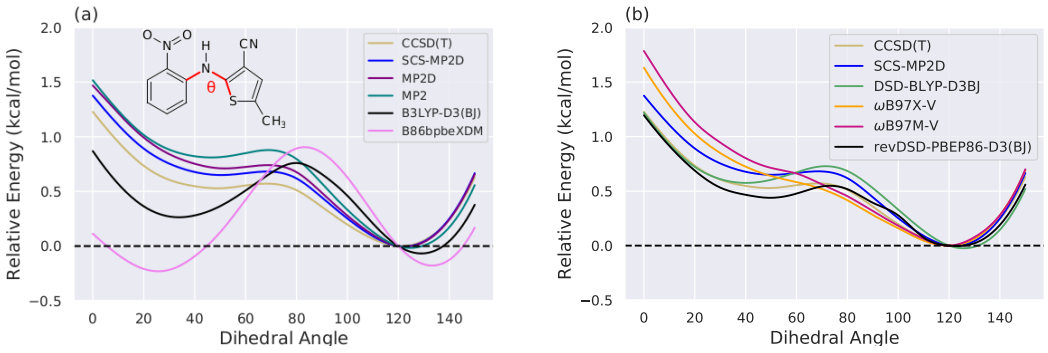


Figure 3.4: Relaxed potential energy scan curves for the θ_{thio} dihedral angle of ROY, as computed with several different density functionals and correlated wave function models. The potential energy curves are separated into two panels for easier viewing.

with the various methods considered in Figure 3.4. CCSD(T) benchmarks[188] predict a global minimum around 120°. A secondary, more shallow minimum occurs around 50°, and it is separated from the global minimum by a small barrier near 70°. The biggest challenge along this potential energy coordinate occurs in the ∼0–80° region.

Typical GGA functionals like B86bPBE-XDM dramatically over-stabilize the lower-angle minimum (Figure 3.4a), and this leads to over-stabilization of the polymorphs with red and orange colors (R, OP, ON, & ORP; $\theta_{thio} \sim$ 20–60°) relative to the yellow ones (Y, YN, & YT04; $\theta_{thio} \sim$ 100–120°).[188,202,203] This behavior is attributed to delocalization error in the functionals artificially stabilizing conformations that allow greater conjugation between the two rings.[189] B86bPBE-XDM also shifts the global minimum of the scan closer to 130° than the true 120° value, and it incorrectly predicts both the position and magnitude of the barrier between the two wells. The hybrid B3LYP-D3(BJ) partially corrects the energies and positions of the two minima and the barrier, but not enough.

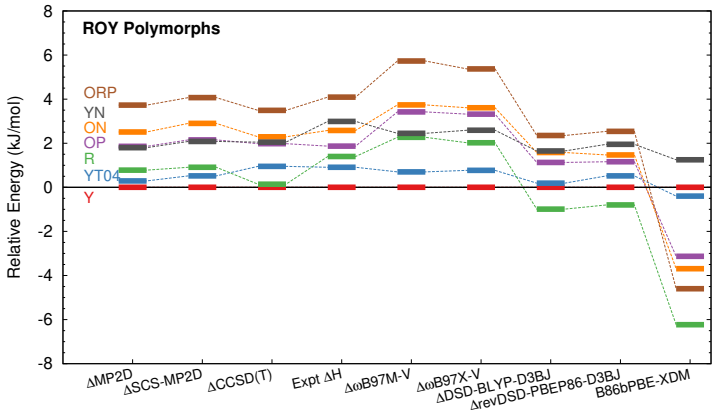


Figure 3.5: Comparison of relative lattice energies for seven ROY polymorphs against experimentally measured enthalpies. The methods listed with a Δ were used to compute the intramolecular conformational energy correction; the intermolecular part was calculated with periodic B86bPBE-XDM throughout.

MP2D performs considerably better than these traditional GGA and hybrid functionals. It predicts the correct position of the global minimum, and overestimates the barrier height and secondary minimum stability by < 0.25 kcal/mol. While the position of the secondary minimum is $\sim 10^\circ$ degrees too high with MP2D, it does position the barrier maximum correctly. SCS-MP2D improves upon MP2D modestly throughout the low-angle range, such that the minima and barrier maximum are all positioned correctly. The overall RMSE relative to CCSD(T) is about a third smaller than that of MP2D (0.12 kcal/mol vs 0.17 kcal/mol).

Consider next the double hybrid DSD-BLYP-D3(BJ) functional (Figure 3.4b). It performs fairly well near 120° and below 30°, but it overestimates the barrier near 80° and shifts the angle at which the minima occur by about 10° in opposite directions. This leads

to a distorted potential energy curve. In contrast, the range-separated hybrid ωB97X-V and hybrid meta-GGA ωB97M-V functionals under-stabilize the low-angle conformations, such that no minimum occurs near 50° at all. So while the magnitude of the errors at any individual point along the energy surface is not especially large for those three functionals, the shapes of the potential energy curves are qualitatively incorrect, particularly for the ωB97-based functionals. The revDSD-PBEP86-D3(BJ) functional performs the best among the DFT models for this curve, though whether it or SCS-MP2D performs better is debatable and depends on which regions of the curve one focuses on.

Figure 3.5 examines the impact of the conformational energy differences on the crystal polymorph stabilities by comparing the relative lattice energies for the seven polymorphs with experimentally reported enthalpies.[6–9] The lattice energies were computed via the monomer-correction approach,[189] which models the crystal energy as a combination of periodic DFT for the intermolecular interactions and a higher-level of theory for the intramolecular conformational energy. Here, the intermolecular part is computed with the B86bPBE-XDM functional (results and fixed-cell optimized geometries taken from ref 188), while the intramolecular conformational energy correction is computed with SCS-MP2D and the other methods listed in Figure 3.5.

Figure 3.5 highlights how the GGA B86bPBE-XDM overstabilizes the red and orange polymorphs (R, ON, OP, ORP) relative to the yellow ones (YN, YT04, Y). Correcting the conformational energies with MP2D or SCS-MP2D gives results in much better agreement with experiment, and only the position of the YN polymorph differs appreciably from experiment. In contrast, correcting the conformational energies with DSD-BLYP-D3(BJ)

or revDSD-PBEP86-D3(BJ) only partially resolves the B86bPBE-XDM problems, with the R polymorph still predicted to be more stable than form Y. ωB97M-V performs somewhat better, though it seemingly over-estimates the destabilization of R, OP, ON, and ORP, which is consistent with the errors seen in the low-angle region of conformational energy scan (Figure 3.5).

Some caution is warranted in interpreting these results, since the comparison against experiment assumes that B86bPBE-XDM is adequate for the intermolecular component and neglects phonon contributions.[188] Performing the conformational energy correction with CCSD(T) leads to polymorph stabilities that are seemingly slightly worse than the MP2D or SCS-MP2D ones, which is presumably an artifact of the other approximations and differing degrees of fortuitous error cancellation. Nevertheless, Figure 3.5 highlights how the conformational energy errors seen in Figure 3.4 can impact polymorph energy differences.

Anthracene photodimerization

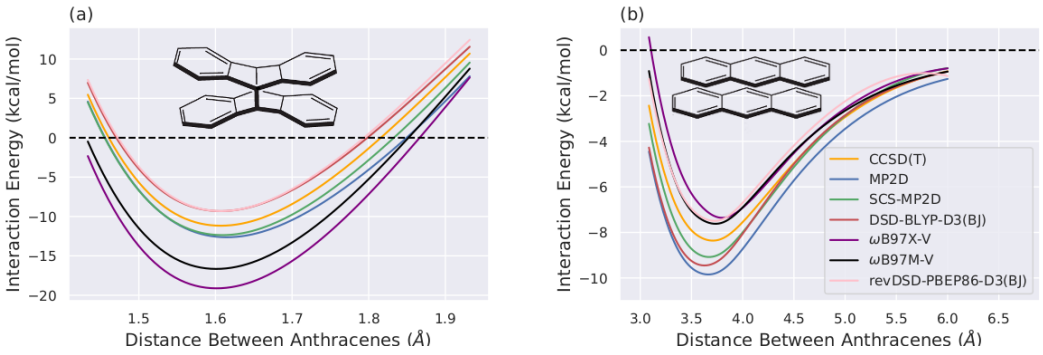


Figure 3.6: Potential curves along the anthracene photodimerization curve for (a) the photodimer and (b) the π-stacked non-covalent dimer.

The anthracene photodimerization potential energy surface is another system that taxes wave function and DFT methods alike. This photochemical reaction converts two non-covalent π-stacked anthracene molecules to form the covalently linked "butterfly" photodimer product (Figure 3.6). The difficulty of modeling the energy difference between reactants and products with electronic structure methods was first highlighted by Grimme.[155] The largest difficulty lies in the photodimer product, which exhibits atypically long C-C single bonds between the rings, highly distorted anthracene rings, and strong close-range van der Waals dispersion interactions between rings. This reaction is also of practical importance, as there has been considerable interest in anthracene-based photomechanical materials,[210–213] Problematic energy predictions for this reaction directly impact the ability for modeling to help understand those systems[204] and design anthracene derivatives with improved photomechanical switching properties.

To study this system, a one-dimensional reaction coordinate as a function of the separation between the two anthracenes was constructed previously.[5] At each constrained separation, all other degrees of freedom were relaxed. Two caveats regarding this energy curve here should be noted: (1) the constraints applied to generate this curve enforce a symmetrically stacked sandwich π dimer, while the true system would offset the anthracene molecules laterally at longer interdimer separations. (2) no effort was made to model the excited-state chemistry or the multi-reference character associated with the forming/breaking of two covalent C-C bonds simultaneously in the intermediate region between $\sim$2–3 Å. The spin-restricted wave functions used here are surely incorrect in this region;

our focus therefore lies on the reactant and product basins. A full potential energy curve is provided in Appendix B Section B.4 for completeness.

Figure 3.6 shows how several different methods perform in the non-covalent π stacked region (near 3.6–3.8 Å) and the covalent photodimer region (near 1.6 Å). Though not shown here, traditional GGA and hybrids like B86bPBE-XDM and B3LYP-D3(BJ) perform poorly for this system, exhibiting errors of up to tens of kcal/mol and predicting the photodimer product to be less stable than two non-interacting anthracene molecules.[5,155,204] In contrast, MP2D performs quite well relative to CCSD(T).[5,204] It binds both the π-stacked dimer and the photodimer a few kcal/mol too tightly, but the systematic nature of the error between the two energy wells leads to a photodimerization reaction energy of -2.8 kcal/mol that agrees almost perfectly with the CCSD(T) value of -2.9 kcal/mol (Table A.2). It also improves considerably over MP2 (not shown here).[5] SCS-MP2D performs a little better than MP2D—it slightly reduces the errors relative to CCSD(T) in the two minima, and it actually mirrors CCSD(T) almost perfectly in the bond-breaking region near 2.5 Å, unlike MP2D. The SCS-MP2D reaction energy of -3.3 kcal/mol is also in excellent agreement with CCSD(T).

The density functionals examined here perform better than traditional GGAs and hybrids for this reaction, but flaws remain. In particular, ωB97X-V and ωB97M-V both overbind the photodimer and underbind the π dimer. This means that the errors compound when computing the reaction energy, and the resulting ΔE is much too exothermic (-9.2 and -12.3 kcal/mol, respectively, Table A.2). DSD-BLYP-D3(BJ) reverses the binding trends compared to those two functionals, and it incorrectly predicts almost zero energy difference

between the two species. revDSD-PBEP86-D3(BJ) gives the best DFT reaction energy of -2.0 kcal/mol, though it still performs a little worse that SCS-MP2D.

Table A.2 also lists RMSE values relative to the CCSD(T) benchmarks, as computed across all data points used in each of the two basins. In the non-covalent π dimer basin (3.2–6 Å), MP2D, SCS-MP2D, and all three functionals perform very well, with errors of 0.5 kcal/mol for SCS-MP2D and 0.7–1.2 kcal/mol for the four functionals. In contrast, the errors in the photodimer basin (1.4–2.0 Å) are somewhat larger for ωB97X-V (6.4 kcal/mol) and ωB97M-V (4.6 kcal/mol), compared to only 1.2 kcal/mol for SCS-MP2D and 1.8 kcal/mol for revDSD-PBEP86-D3(BJ). Overall, spin-component scaling reduces the MP2D errors by about a factor of two across these two basins, and SCS-MP2D reproduces this challenging CCSD(T) potential energy curve more faithfully than any of the four density functionals.

3.4 Conclusions

The last decade has witnessed substantial performance improvements in lower-cost models based on DFT and MP2. This study presented a new spin-component-scaled, dispersion-corrected MP2 model that provides accuracy that is competitive with some of the best density functional models for intermolecular interactions, conformational energies, and thermochemistry in organic systems. The largest SCS-MP2D performance improvements over MP2D occur for the non-bonded complexes in NBC10, charge transfer reactions, the ionic hydrogen bonds of IHB100, and many of the conformational and reaction energy data sets. It also behaves well for challenging ROY and anthracene photodimerization potential

Table 3.4: Reaction energy ΔE_{rxn} for anthracene photodimerization, 2 $C_{14}H_{10} \longrightarrow (C_{14}H_{10})_2$,[a] in kcal/mol. Root-mean-square errors relative to the CCSD(T) benchmarks are also presented for the photodimer (1.4–2.0 Å) and π dimer (3.2–6.0 Å) basins from Figure 3.6.

		Root-Mean-Square Error	
Method	ΔE_{rxn}[a]	Photodimer	π Dimer
CCSD(T)	-2.9		
MP2	-5.7	10.0	5.1
MP2D	-2.8	2.0	1.3
SCS-MP2D	-3.3	1.2	0.5
DSD-BLYP-D3(BJ)	0.1	1.5	0.9
revDSD-PBEP86-D3(BJ)	-2.0	1.8	0.8
ωB97X-V	-12.3	6.4	1.2
ωB97M-V	-9.2	4.6	0.7

[a] $\Delta E_{rxn} = E(1.6\text{ Å}) - E(3.6\text{ Å})$

energy curves. The fact that the seven empirical parameters in SCS-MP2D could be trained using a relatively small amount of training data while maintaining excellent transferability to new systems suggest that SCS-MP2D is properly capturing the important physics. The accuracy and computational cost of SCS-MP2D is highly competitive with the high-quality range-separated and double-hybrid functionals considered here for systems with many tens of atoms.

Looking forward, a few potential paths for improving SCS-MP2D models are apparent: First, replacing the D3 correction with D4 is likely to improve performance on systems involving ions, since it would allow the dispersion coefficients to adapt better to the different charge states. This might address the most notable weakness of SCS-MP2D discovered thus far—its performance for anion-anion interactions. The switch might offer more general performance improvements as well, based on the gains observed upon switching from D3 to D4 in the DSD-family of double-hybrid density functionals.[83]

Second, the spin-component scaling performed here does not address the limitations of the highly local D3 dispersion coefficient implementation. As discussed in the original MP2D paper,[5] the D3 coordination number scheme used to estimate the dispersion coefficients does not distinguish clearly between localized and extended/aromatic sp^2 environments. One could likely adapt dispersion corrections such as XDM[214] or the many-body dispersion (MBD) model[215,216] to compute higher-quality *ab initio* dispersion contributions on the fly, albeit with a somewhat increased computational cost. Alternatively, it might be possible to machine-learn the dispersion coefficients; however, care would need to be taken to learn and predict atomic dispersion coefficients rather than molecular ones. Obtaining reliable atomic coefficients would also require careful thought—no single, universally agreed-upon partitioning scheme exists for the frequency-dependent polarizabilities and/or dispersion coefficients.

Finally, the revised versions of the DSD double-hybrid functionalss improved dramatically upon refitting the empirical parameters to the much larger GMTKN55 data set. On the one hand, the excellent SCS-MP2D performance obtained after fitting to a modest

amount of data speaks well to its physical foundations. On the other hand, it is possible that fitting to a much larger data set might lead to even more transferable parameters.

Chapter 4

Overcoming the difficulties of predicting conformational polymorph energetics in molecular crystals via correlated wavefunction methods

4.1 Introduction

Crystal packing influences the physical properties of organic crystals. The occurrence of multiple crystalline packing motifs, or polymorphs, of a pharmaceutical can impact its solubility, bioavailability, shelf-life/stability, and tabletting properties, for example. The

importance of polymorphism to the pharmaceutical industry is highlighted by examples such as ritonavir[22,217] and rotigotine,[15] where the late-stage appearance of more stable, less soluble crystal forms forced product recalls and reformulations. It was recently suggested that the thermodynamically stable crystal form has not been realized experimentally for ∼15–45% of pharmaceutical molecules,[218] raising speculation that more such examples may occur in the future. Moreover, solid form patents play an important role in the commercial life cycle of a drug, as evidenced by the recent legal wrangling over a new polymorph of Celgene's blockbuster drug revlimid that was discovered by generic drug manufacturer Natco.[219]

The ability to predict the molecular crystal energy landscape, which is the set of possible low-energy crystal structures for a given compound, would be a tremendous boon to the pharmaceutical industry and others. Crystal structure prediction has long been challenging[220] due to the complexity of the search space, the small energy differences that separate polymorphs, and the complexities of crystallization kinetics. The accuracy requirements for predicting the crystal energy landscape are severe: Surveys suggest that about half of all polymorph pairs are separated by less than 2 kJ/mol in lattice energy, and around 95% are separated by less than 8 kJ/mol.[221–223]

The advent of high-quality, dispersion-corrected density functional theory (DFT) models[110,224–226] has enabled tremendous progress in the energy ranking aspects of crystal structure prediction, as evidenced by results from the recent blind tests[227–230] and other studies.[231–243] Increasingly, DFT is being called on to explore pharmaceutical crystal energy landscapes as a complement to experimental solid form screening.[244–252] Computational

prediction of a highly stable, unrealized polymorph of galunisertib played a key role in the extensive characterization of its solid form landscape, for example.[251]

Despite many successes of DFT-driven crystal structure prediction, close inspection of the literature also finds polymorphic crystals for which widely-used DFT models fail dramatically. Many of these difficult cases involve conformational polymorphs, in which different intramolecular conformations enable different intermolecular crystal packing motifs. For example, DFT methods invert the polymorph stability ordering of α and β *o*-acetamidobenzamide, with errors of 5–10 kJ/mol.[253] The prolific polymorph-former 5-methyl-2-[(2-nitrophenyl)amino]-3-thiophenecarbonitrile, nicknamed "ROY" after its colorful red-orange-yellow crystals, is another example. State-of-the-art density functionals predict the Y polymorph to be one of the least stable forms,[203,205] when it is actually the most stable one. These backwards stability rankings reflect errors approaching 10 kJ/mol. In another case, crystal structure prediction failed for two of six conformationally flexible species resulting from mechanochemical aromatic disulfide metathesis reactions, with errors exceeding 6 kJ/mol due in large part to poor intramolecular DFT conformational energies.[254] Erroneous intramolecular conformational energies caused a similar failure for a recent DFT study of Molecule X from an earlier blind test of crystal structure prediction.[237]

The large errors in the relative polymorph stabilities found for many of these examples greatly exceed the few kJ/mol errors or less one typically finds for DFT in successful crystal structure prediction cases. Furthermore, these errors are catastrophically large compared to the small energy differences that are characteristic of polymorphism. The pharmaceutical industry trend toward developing larger, more flexible drug molecules[255] makes

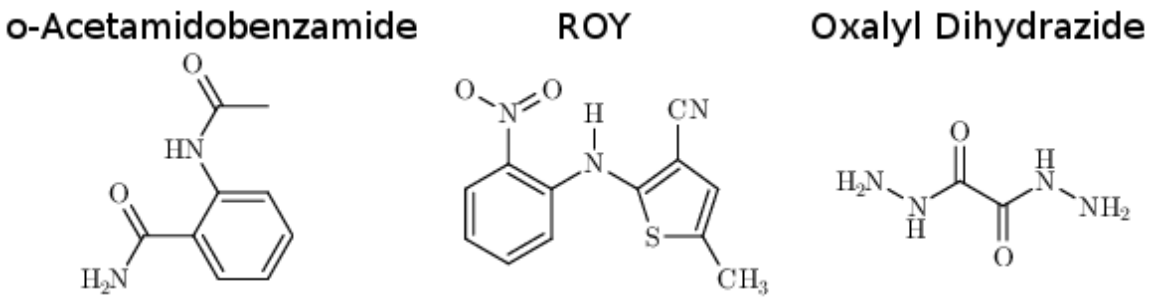


Figure 4.1: The species whose conformational polymorphs are studied here.

problems with ranking conformational polymorphs particularly concerning, since it raises the possibility that such ranking problems will become more prevalent as crystal structure prediction is applied to increasingly complicated species.

The problems with popular DFT functionals are not limited to conformational polymorphism either. Delocalization error in commonly used DFT generalized gradient approximation (GGA) functionals can cause spurious salt formation in co-crystals[256] and the substantial overbinding of crystals containing halogen bonds.[257] Many functionals erroneously predict the exothermic anthracene photodimerization reaction to be strongly endothermic,[5,155] which is problematic[204] when studying a class of interesting anthracene-based photomechanical materials.[210,211]

Switching to a hybrid functional can address some of the limitations of GGA-type functionals that cause incorrect polymorph rankings and other problems,[233,241,256–259] but it does not rectify the incorrect ROY polymorph rankings,[205] for example. Approaches based on periodic second-order Møller-Plesset perturbation theory (MP2),[260–274] the random phase approximation (RPA),[274–277] and quantum Monte Carlo[278–281] are also being developed that can improve the reliability of polymorph stability rankings. Computational

cost is the fundamental challenge that inhibits applying these more accurate electronic structure methods to molecular crystals. As evident from the studies cited above, higher-level electronic structure methods have generally been applied only to small-molecule crystals at present.

Fragment-based methods provide one means of lowering the computational cost of correlated electronic structure methods by decomposing the molecular crystal into monomers, dimers, and many-body contributions.[224,282–286] These methods typically compute only the key monomer and dimer contributions at the most accurate level of theory, while the many-body contributions are approximated in some fashion. Employing coupled cluster singles, doubles, and perturbative triples (CCSD(T)) in the context of a fragment method can give very reliable results, as demonstrated by quantitative prediction of the benzene lattice energy[287] or the prediction of the methanol polymorph phase diagram with ∼0.5 kJ/mol accuracy.[288]

Unfortunately, even with fragment methods, CCSD(T) calculations are cost prohibitive for crystals involving pharmaceutical-sized species. MP2 is more feasible computationally, but it suffers from well-known problems in the description of van der Waals interactions[112] that cause it to substantially overestimate the interaction energy in π-stacking complexes[289] and to over-bind the benzene crystal by ∼10–20%,[224] for example. This difficulty is overcome here by using the recently developed MP2D model, which employs a Grimme D3-like[123] dispersion correction that removes the problematic dispersion treatment inherent to MP2 and replaces it with a more reliable treatment. MP2D performs well across extensive benchmark calculations of dimer interactions, molecular conformations, and re-

action energies,[5] and it correctly describes the energetics of the aforementioned anthracene photodimer.[204]

The present study examines three challenging cases of conformational polymorphism in detail: *ortho*-acetamidobenzamide,[253] ROY,[3,203] and oxalyl dihydrazide (Figure 5.1).[253,290] It demonstrates how problems in the intramolecular conformational energies and, to a lesser extent, the intermolecular interactions with well-regarded dispersion-corrected DFT functionals lead to incorrect polymorph stabilities. However, modeling these systems with fragment-based correlated wavefunction methods overcomes these difficulties, restoring the crucial balance between intra- and intermolecular interactions[291–293] that is required to predict the correct stabilities in conformational polymorphs. The results here highlight how despite considerable progress with DFT, polymorph stability ranking remains challenging, and models that can achieve higher accuracy than that of commonly used DFT approximations are needed before polymorph ranking can be considered a "solved" problem.

4.2 Theory and Methods

When ranking or predicting polymorph stabilities, it is important to recognize that those stabilities depend on temperature. Sometimes the free energy variations are large enough to produce an enantiotropic relationship where the thermodynamically preferred polymorph changes depending on the temperature. However, even in monotropic cases where one polymorph is always preferred thermodynamically, the magnitude of the enthalpy and free energy differences between two polymorphs will depend on temperature. The temperature dependence of the relative stabilities arises from both phonon contri-

butions to the vibrational partition function and from phonon-driven thermal expansion. The molar volume of typical organic crystals expands several percent upon heating from 0 K to room temperature.[294] This expansion alters the lattice energy and introduces anharmonicity into the phonons. Accounting for it is important for quantitatively comparing thermochemistry,[53,288,295,296] mechanical properties,[295] and spectroscopic observables[294,297] between theory and experiment.

It is therefore important to consider the thermodynamic conditions under which experimental measurements were made when making theoretical predictions. Ideally, one would capture temperature effects via molecular dynamics (including nuclear quantum effects, since zero-point contributions can be significant[296,298]). However, molecular dynamics simulations based on high-level electronic structure methods are very computationally expensive.[298] The quasi-harmonic approximation is often used to approximate the volume-dependent contributions to the phonons and lattice energies.[53,226,236,241,288,294–296,299–302] Nevertheless, quasi-harmonic calculations remain considerably more expensive than purely harmonic calculations that neglect thermal expansion.

Here, a simple approximation is employed to estimate the temperature dependence of the thermochemical stabilities and facilitate comparison with experiment. Two types of crystal structure optimizations are performed. Fully relaxed crystal structures that optimize both the atomic positions and the unit cell vectors approximate the structure at 0 K (albeit without zero-point vibrational expansion[296]). Room-temperature structures are mimicked via fixed-cell optimizations that relax the atomic positions subject to the constraint of the room-temperature experimental lattice parameters. Harmonic phonons are computed

separately on each set of structures, thereby approximately capturing the anharmonicity that results from the change in unit cell dimensions.

Similar fixed-cell optimizations have been used by many other authors previously to examine room-temperature crystal properties, including in earlier studies on the same polymorphic systems studied here.[203,253] Constraining the lattice parameters effectively captures the thermal expansion effects and its associated phonon anharmonicity, while relaxing the atomic positions addresses any issues in the experimental molecular geometries (hydrogen atom placement, for example) and ensures the structure is at a minimum for harmonic vibrational frequency calculations. Additional support for this approach comes from the fact that nuclear magnetic resonance chemical shift predictions performed on structures relaxed with fixed lattice parameters reproduce experimental chemical shifts better than those obtained from fully relaxed structures[294] or even neutron diffraction structures.[303]

Note that the approximations used here neglect thermal/large-amplitude dynamical motions that can occur in molecular crystals. Fortunately, the structures of the systems considered here do not exhibit significant disorder and are likely amenable to static modeling treatments. The differences between quasi-harmonic and molecular dynamics models are frequently (but not always) small.[304–306] In the end, combining information from the fully-relaxed 0 K structures and fixed-cell room-temperature structures provides information regarding the topology of crystal energy landscapes that facilitates comparison with experiment.

Experimental crystal structures were obtained from the Cambridge Structure Database for *o*-acetamidobenzamide[307] (reference codes ACBNZA and ACBZNA01), ROY[6,7,206]

(refernce codes QAXMEH–QAXMEH05, QAXMEH12, and QAXMEH52), and oxalyl dihydrazide[290] (reference codes VIPKIO01–VIPKIO05). Crystal structures were optimized using periodic DFT with the B86bPBE density functional[308,309] and exchange-hole dipole moment (XDM) dispersion correction.[310] This particular combination performs well in many molecular crystal applications.[235–237,310]

Single-point refinement of the electronic energies was carried out using correlated wavefunction methods via the fragment-based hybrid many-body interaction (HMBI) model.[284,311–313] HMBI partitions the total energy of the crystal into intramolecular contributions (1-body interactions), pairwise intermolecular interactions (2-body interactions), and the remaining many-body intermolecular lattice contributions. The important 1-body and short-range (SR) 2-body terms are modeled with high-level electronic structure methods (e.g. CCSD(T) or MP2-based methods here), while the longer-range (LR) 2-body and many-body contributions are modeled with periodic Hartree-Fock (HF) theory (which makes it comparable to Stoll's method of increments[282]).

$$\begin{aligned} U_{el}^{HMBI} &= E_{1-body}^{High} + E_{SR\ 2-body}^{High} \\ &\quad + E_{LR\ 2-body}^{HF} + E_{many-body}^{HF} \end{aligned} \tag{4.1}$$

As noted above, MP2 suffers from problematic description of van der Waals interactions. The related and highly successful[122,314] MP2C model addresses this problem by adding a non-empirical intermolecular dispersion correction to MP2.[115,121] However, the MP2C correction is derived from intermolecular perturbation theory and does not address problems with intramolecular dispersion. MP2D[5] expresses the dispersion correction in terms of atom-centered C_6 and C_8 dispersion coefficients[124] computed using the scheme be-

hind Grimme's D3 dispersion correction.[123] The five global empirical parameters in MP2D were determined previously[5] on small-molecule systems that do not include the species studied here.

In the cases of oxalyl dihydrazide and *o*-acetamidobenzamide, enthalpies and free energies are computed for comparison with experiment. This requires evaluating harmonic phonon contributions to the enthalpy and Helmholtz vibrational free energy F_{vib} via the standard statistical mechanical expressions.[296] The phonons and their thermodynamic contributions are calculated at the B86bPBE-XDM level, using either the 0 K or room-temperature crystal structures, and these are used to augment the electronic energy computed with either DFT or HMBI. For example, the Gibbs free energy at the MP2D level is estimated as,

$$G(T,P) = U_{el}^{HMBI} + F_{vib}^{DFT} + PV \tag{4.2}$$

For a crystal at ambient conditions, the PV term contributes negligibly and can be ignored. This combination of DFT geometries and phonons with higher-level single-point electronic energies has been validated previously.[314]

The DFT calculations were performed using Quantum Espresso v6.3[315] using a 50 Ry planewave cutoff and well-converged Monkhorst-Pack **k**-point sampling grids (Appendix C Section C1.1[†]). Core electrons were treated according to the projector augmented wave (PAW) approach using PAW potentials for H, C, N, O, and S produced with A. Dal Corso's Atomic code v6.1.[316] Gas-phase monomer and dimer DFT calculations used in the energy decompositions were performed in large unit cells with a minimum of 15 Å spacing between the central monomer/dimer atoms and all periodic image atoms. Using an even larger 18 Å

spacing altered the gas-phase energies by ∼0.15 kJ/mol or less, indicating that this vacuum spacing is appropriately large to mimic the gas phase.

Harmonic DFT phonon frequencies for oxalyl dihydrazide and *o*-acetamidobenzamide were computed at the Γ-point using Phonopy v1.12.6-r66[317] with the same B86bPBE-XDM functional and basis set used for the energies and geometry optimizations. To ensure equal numbers of molecules ($Z = 4$) in the cell for each of the five oxalyl dihydrazide polymorphs, supercells were constructed for the α, β, δ, and ϵ forms by doubling the cell along the shortest crystallographic axis. Using larger supercells and/or capturing phonon dispersion away from the Γ point would certainly improve the quality of the predicted thermochemistry.[302] Still, earlier quasi-harmonic sublimation enthalpy calculations for several small-molecule crystals agreed with experiment to within a couple kJ/mol despite neglecting phonon dispersion.[314] Percentage errors in the entropic contributions were considerably larger for the same species, however. Phonons were not computed for the larger ROY system for reasons of computational expense.

For the HMBI fragment calculations employing correlated wave function methods, large basis sets must be used to ensure convergence of the polymorph energetics, as demonstrated in Appendix C Section C1.2[†] and many previous studies.[224,287,288,295,296,314,318,319] Here, MP2 and MP2D monomer and dimer energies at the complete-basis-set (CBS) limit were obtained using a development version of PSI4.[137] The correlation energy was extrapolated[70] to the complete-basis-set (CBS) limit using data from the aug-cc-pVTZ and aug-cc-pVQZ basis sets[140] and combined with HF/aug-cc-pVQZ. The MP2C dispersion corrections were obtained with Molpro 2012.1[130] in the aug-cc-pVTZ basis set. The MP2C

dispersion correction typically converges faster with basis set than the raw correlation energy.[121] CCSD(T) results at the CBS limit were obtained by correcting MP2/CBS energies with the difference between CCSD(T) and MP2 in the aug-cc-pVDZ (ROY, oxalyl dihydrazide) or cc-pVTZ (acetamidobenzamide intramolecular contributions) basis sets. The periodic HF many-body contributions were evaluated using Crystal 17[320] and the pob-TZVP-rev2 basis set.[321] This basis set was chosen based on cluster benchmarks described in Appendix C Section C1.3.† Note that due to computational expense, large-basis set calculations were performed only on the room-temperature structures of ROY, since those are more directly comparable with experiment. Smaller-basis results on the fully relaxed structures are provided in Appendix C Section C3.3.† As expected, large basis sets are required to converge the relative polymorph stabilities.

As part of the analysis of the ROY system, a one-dimensional, gas-phase conformational energy scan over the key S-C-N-C dihedral angle was performed. At each of 16 fixed dihedral angle values ranging 0–150° in 10° intervals, all other degrees of freedom were fully relaxed at the B3LYP-D3(BJ)/def2-TZVP level of theory. Single-point energies were then computed on these geometries using B86bPBE-XDM, MP2, MP2D, and CCSD(T).

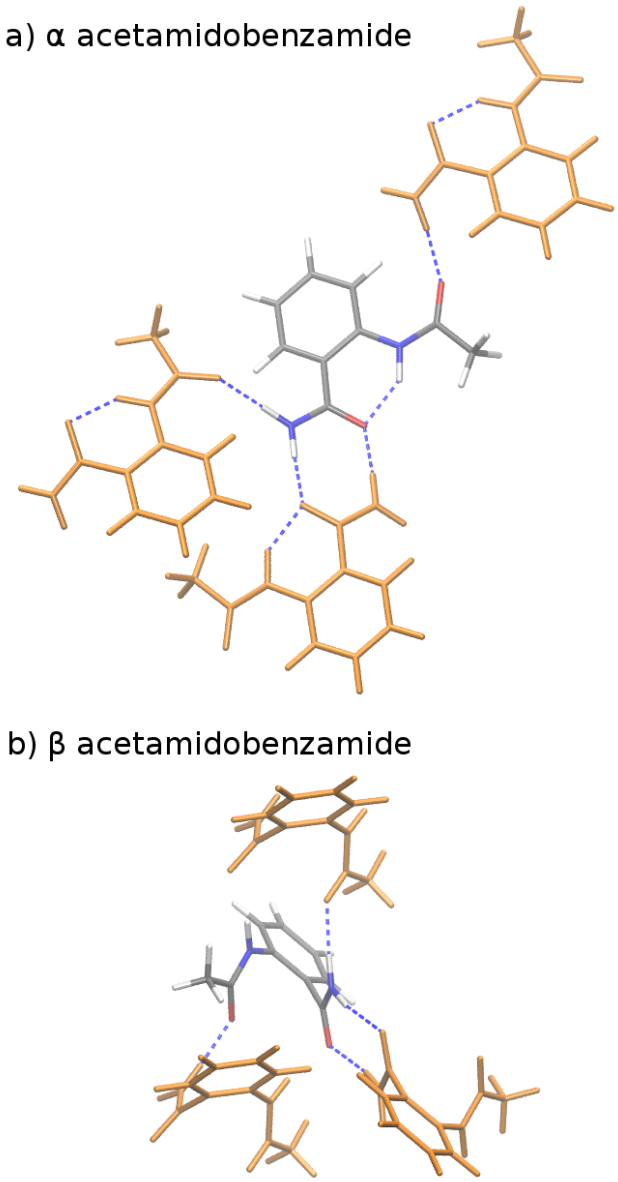


Figure 4.2: Local hydrogen bonding environments for crystalline *o*-acetamidobenzamide. (a) In the α polymorph, the molecule is nearly planar and adopts an intramolecular hydrogen bond, while (b) in the β polymorph the amide and acetamide side chains rotate out of the plane to achieve better intermolecular hydrogen bonds.

Table 4.1: Stability of the β o-acetamidobenzamide polymorph relative to the α one, in kJ/mol. Positive values indicate α is more stable than β. See Appendix C Table C.4 for additional details of the lattice energy and phonon contributions. Wavefunction method results are extrapolated to the CBS limit.

Method	ΔE_{intra} (0 K)	ΔE_{inter} (0 K)	ΔE_{latt} (0 K)	ΔH (0 K)	ΔH (298 K)	ΔH^a (423 K)	ΔG (298 K)
B86bPBE-XDM	58.0	-52.2	5.8	4.6	5.5	5.9	4.3
MP2+pHF	50.1	-46.8	3.2	1.9	-1.6	-3.1	-2.8
MP2C+pHF	50.1	-52.2	-2.1	-3.4	-4.9	-5.5	-6.1
MP2D+pHF	52.6	-51.2	1.4	0.2	-1.4	-2.0	-2.6
CCSD(T)	52.3						
Experiment						-1.9[b], -2.9[c]	

[a] Linearly extrapolated to 423 K from ΔH(0 K) and ΔH(298 K) values.

4.3 Results and Discussion

4.3.1 *o*-Acetamidobenzamide

o-Acetamidobenzamide has two conformational polymorphs: The α form adopts the more stable intramolecular conformation containing an intramolecular hydrogen bond between the acetamide hydrogen and the amide oxygen (Figure 4.2).[307] The β polymorph sacrifices the intramolecular hydrogen bond to adopt a conformation that allows better intermolecular hydrogen bonding. Experimentally, the α form converts exothermically and irreversibly to β upon heating to 150°C (423 K), with $\Delta H_{\alpha\rightarrow\beta}$ values of -1.9 kJ/mol[253]

or -2.9 kJ/mol[307]. In other words, the β form is clearly preferred over the α one at high temperatures, though the stability ordering at lower temperatures is unclear.

Earlier calculations using force fields and several different GGA and hybrid DFT functionals all predict the α form lattice energy to be $\sim$5–10 kJ/mol more stable than β.[253] New B86bPBE-XDM DFT lattice energy calculations performed here similarly favor the α form by 5.8 kJ/mol (Table D.4). The B86bPBE-XDM harmonic zero-point vibrational energy contribution for the fully-relaxed 0 K structures stabilizes the β form by 1.2 kJ/mol relative to α, which is reasonably similar to the 2 kJ/mol value estimated previously.[253] In other words, the difference in zero-point energy contributions between polymorphs are much too small to alter the B86bPBE-XDM polymorph stability ordering. Moreover, the DFT enthalpic preference for the α form increases from 4.6 kJ/mol at 0 K to 5.5 kJ/mol at room temperature (Figure 4.3). As shown in Appendix C Table C.4,[†] this temperature-dependence between the 0 K and room-temperature structures arises from a 0.6 kJ/mol relative destabilization of the α form caused by the lattice energy changes that is canceled by a larger 1.6 kJ/mol stabilization due to the vibrational enthalpy contribution.

To compare more directly against experiment, the transition enthalpy at the phase transition temperature is estimated here via linear extrapolation of the 0 K and room-temperature results. The resulting endothermic $\Delta H_{\alpha\rightarrow\beta}$(423 K) value of 5.9 kJ/mol contradicts the exothermic phase transition observed experimentally. The B86bPBE-XDM Gibbs free energies exhibit a clear preference for the α polymorph of 4.6 kJ/mol at 0 K that decreases only slightly to 4.3 kJ/mol at room temperature (Table D.4). Linear extrapolation of $\Delta G_{\alpha\rightarrow\beta}$ to 423 K gives 4.2 kJ/mol. This strong DFT free energy preference for

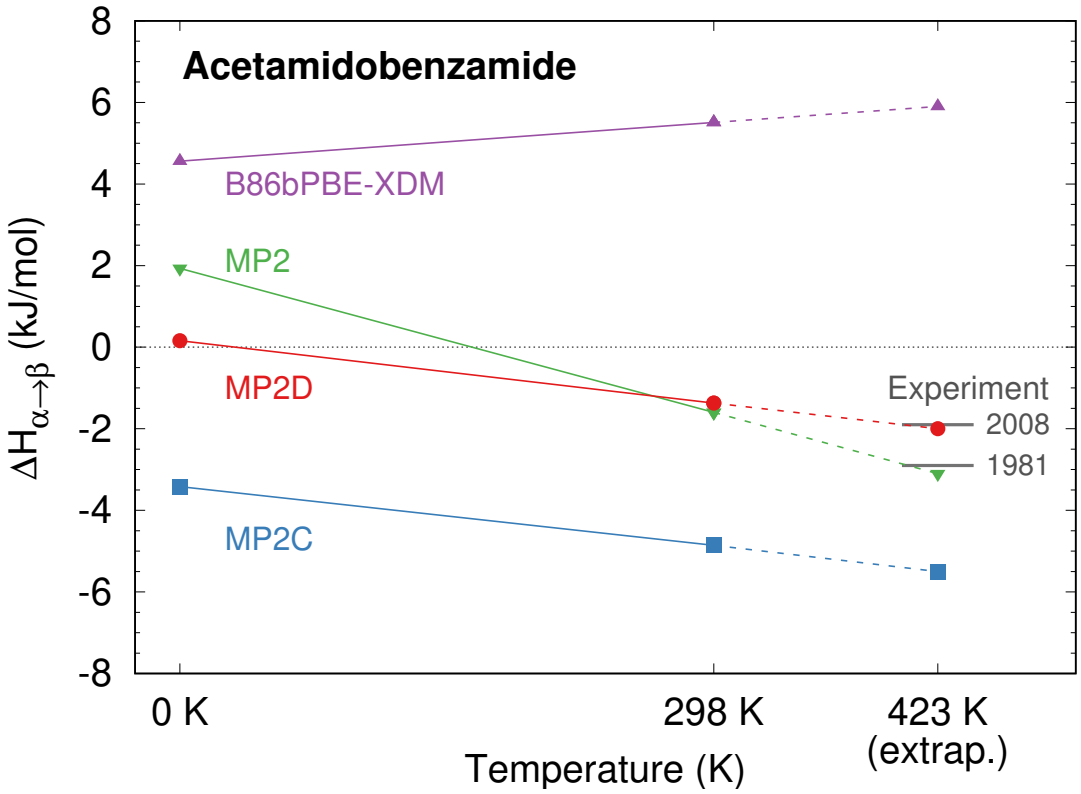


Figure 4.3: Predicted enthalpy difference between the α and β polymorphs of *o*-acetamidobenzamide at 0 K, room temperature, and linearly extrapolated to 423 K. Experimental values were taken from from .

the α form is inconsistent with the irreversible $\alpha \rightarrow \beta$ phase transition seen experimentally. Earlier attempts to rationalize the DFT lattice energy preference for the α polymorph suggested that the two polymorphs might be enantiotropically related.[253] However, the DFT harmonic free energy calculations here predict the α polymorph to be considerably more stable throughout the temperature range (i.e. monotropically related, see Figure 4.4a). Taken together, both the B86bPBE-XDM enthalpies and free energies computed here and those reported previously[253] are inconsistent with experiment.

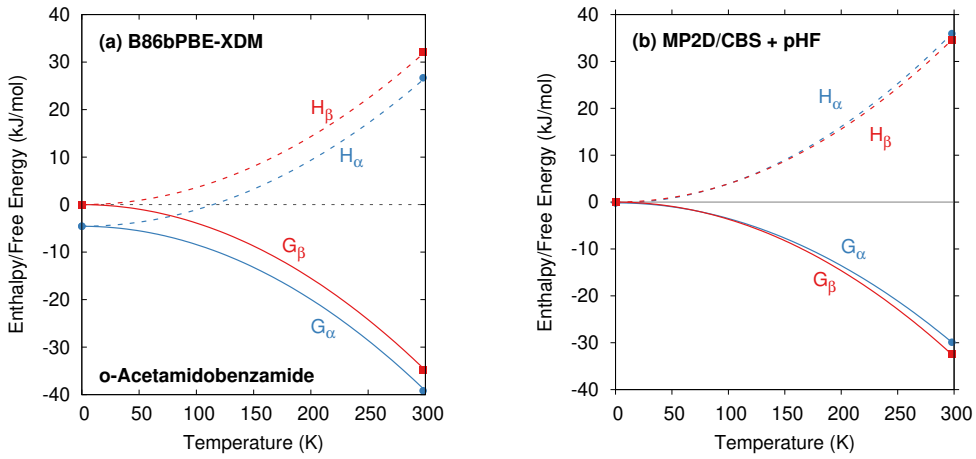


Figure 4.4: Schematic relative enthalpy H and free energy G curves for the α and β polymorphs of *o*-acetamidobenzamide computed using (a) B86bPBE-XDM or (b) fragment-based MP2D/CBS + pHF. The MP2D model reverses the stability ordering, giving results that are consistent with experiment.

In contrast, fragment-based MP2D predicts polymorph stabilities that agree very well with experiment (Table D.4). Calculations on the fully-relaxed 0 K structures suggest that the α polymorph is still more stable in lattice energy, but by a much smaller 1.4 kJ/mol. Including the B86bPBE-XDM zero-point vibrational contribution preferentially stabilizes the β form, such that the two forms become nearly degenerate (α is 0.2 kJ/mol more stable than β). Heating further stabilizes the β form, with $\Delta H_{\alpha\rightarrow\beta}$ = -1.4 kJ/mol at 298 K. This temperature dependence is largely driven by a 3.1 kJ/mol stabilization of the β form arising from the lattice energies, which is partially canceled by the DFT phonon contribution (Appendix C Table C.4†). The exothermic phase transition at higher temperatures predicted by MP2D is consistent with experiment. Linearly extrapolating

$\Delta H_{\alpha\to\beta}$ to 423 K gives -2.0 kJ/mol, in excellent agreement with both experimental values of -1.9 and -2.9 kJ/mol (Figure 4.3).

Furthermore, MP2D Gibbs free energies indicate that the β form is thermodynamically preferred at elevated temperatures (Figure 4.4b), which is consistent with the irreversible $\alpha \to \beta$ transition seen experimentally at 423 K. Nominally, the MP2D calculations predict an enantiotropic relationship between the two forms, though the 0.2 kJ/mol free energy difference between the two forms at 0 K is likely smaller than the inherent uncertainties in the models. The MP2D free energies suggest that the experimentally observed $\alpha \to \beta$ phase transition at 423 K corresponds to a kinetically activated transformation from the metastable α form to the stable β one, rather than a true thermodynamic phase boundary.

To understand why MP2D performs well in this system while B86bPBE-XDM does not, Table D.4 decomposes the lattice energy differences between the two polymorphs into their intra- and intermolecular contributions. MP2D and B86bPBE-XDM actually predict similar intermolecular energies that differ by only 0.3 kJ/mol for the fully relaxed 0 K structure, and by 1.3 kJ/mol for the room-temperature structures. Rather, the erroneous DFT predictions arise almost entirely from the intramolecular conformational energies: B86bPBE-XDM over-stablizes the intramolecular hydrogen bond conformation by $\sim$6 kJ/mol (11% error) compared to gas-phase CCSD(T) benchmarks. Delocalization error is known to cause GGA and hybrid functionals to overstabilize aromatic systems.[237,322,323] The intramolecular conformation in the β polymorph disrupts not only the intramolecular hydrogen bond, but also π conjugation between the aromatic ring and the amide/acetamide

side chains. In contrast to B86bPBE-XDM, MP2D reproduces the CCSD(T) conformational energy difference to within a few tenths of a kJ/mol (<1% error).

Finally, the *o*-acetamidobenzamide polymorphs demonstrate the importance of correcting both the intra- and intermolecular description of dispersion in MP2, as is done in MP2D. MP2 underestimates the intermolecular preference for the β phase by 4–5 kJ/mol, and it underestimates the penalty for disrupting the intramolecular hydrogen bond found in the α form by 2 kJ/mol (Table D.4). These errors cancel somewhat, but the resulting ΔH and ΔG values appear to change too rapidly with temperature (due to how the lattice energy varies with the temperature-dependent changes in crystal structure). MP2C corrects the description of the intermolecular interactions, but it does not alter the intramolecular description. This disrupts the fortuitous error cancellation found in MP2, and MP2C overestimates the stability of the β form substantially.

4.3.2 ROY

ROY is among the most prolific conformational polymorph formers known. Seven polymorphs have been well-characterized for years.[3,6,7] Since 2018, the structures of two more polymorphs, R05[205] and PO13,[206] have been solved, and a structure for the RPL polymorph was proposed.[203] However, as multiple recent studies have noted, predicting the energetics of these polymorphs has proved challenging.[202,203,205] Many well-regarded van der Waals-inclusive GGA and hybrid density functionals predict highly incorrect polymorph orderings, including PBE-D3, PBE-NP, optPBE-vdW, PBE+MBD, and PBE0+MBD (Figure 4.5).[203,205] Most strikingly, the DFT calculations frequently suggest that the Y poly-

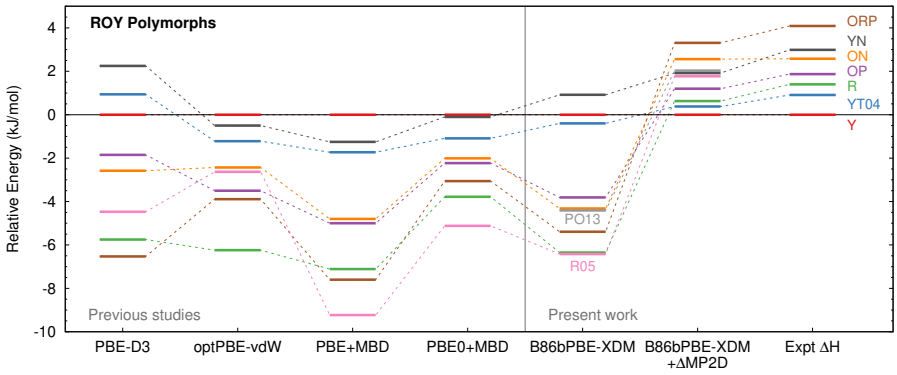


Figure 4.5: Comparison between predicted lattice energies and experimentally measured enthalpies[3] for 8 polymorphs of ROY, all relative to form Y. The PBE-D3, optPBE-vdW, PBE+MBD, and PBE0+MBD results were taken from They omit PO13 and use fully relaxed 0 K unit cells. The other results employ fixed-cell room-temperature structures.

morph is one of the least stable forms, when it is actually the most stable polymorph experimentally. Inclusion of zero-point energies or thermal contributions does not correct the rankings, either.[203]

Experimentally, the relative free energies of the ROY polymorphs were measured in the ∼40–120°C range by eutectic melting experiments.[6–9] Relative enthalpies were then obtained from the slopes of $\Delta G/T$ vs. $1/T$ plots, where T is temperature. Because the fitted enthalpies lack explicit temperature dependence, they are most valid in the elevated temperature regime under which they were measured. To facilitate comparison against the experimental data, fixed-cell room-temperature crystal structures are used for the polymorph stability calculations here, since they will mimic the structures under the experimental conditions better than fully relaxed 0 K ones.

B86bPBE-XDM calculations on the room-temperature structures predict stabilities that are similar to earlier DFT studies (Figure 4.5), with the Y form being the second least-stable polymorph in terms of lattice energy (below only the YN form). Differences in the polymorph stabilities between the fixed-cell room-temperature and fully-optimized 0 K structures are modest. Refining the lattice energies of the room-temperature structures with single-point energy calculations at the fragment-based MP2D level completely transforms the crystal energy landscape. MP2D correctly predicts the Y form to be the most stable in terms of lattice energy. Furthermore, with the exception of the ON polymorph, the MP2D stability ordering qualitatively matches the experimental enthalpy data perfectly. The predicted lattice energy differences are somewhat larger than the experimental enthalpies. That discrepancy may in part be due to the omission of phonon contributions in the predictions here. The reason for the incorrect ordering of the ON polymorph is unclear. While an earlier study using a different density functional had difficulty reproducing the experimental ON crystal structure, the structure obtained here with B86bPBE-XDM agrees well with experiment (Appendix C Section C1.1†).

The MP2D calculations also predict the recently discovered PO13 polymorph to be among the less stable polymorphs, below only ON and ORP. Though experimental thermochemical data is not available for the PO13 form, this prediction is consistent with experimental data that indicates PO13 is less stable than the Y polymorph and that its heat of fusion is in the mid-range compared to the other forms. Predictions for the R05 form are omitted here because, unlike the other polymorphs, its cell has a net dipole which creates difficulties for the fragment-based approach (Appendix C Section C3.4†).

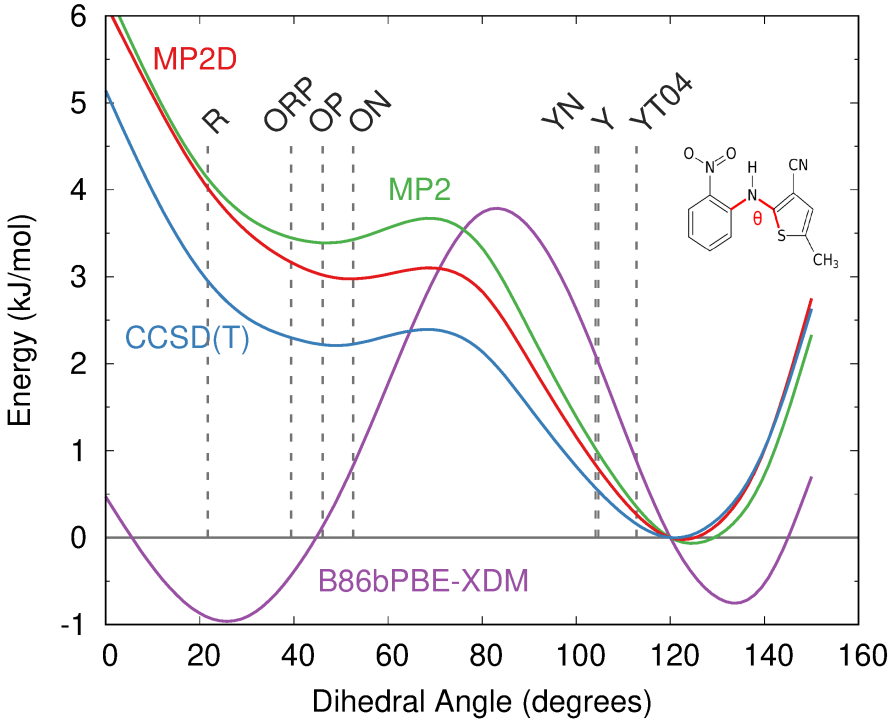


Figure 4.6: Comparison of the intramolecular conformational energy scan for the key intramolecular dihedral angle in ROY at several levels of theory. Vertical dotted lines indicate the experimental dihedral angles for each polymorph. Energies are relative to the 120° conformation, which corresponds to the CCSD(T) minimum.

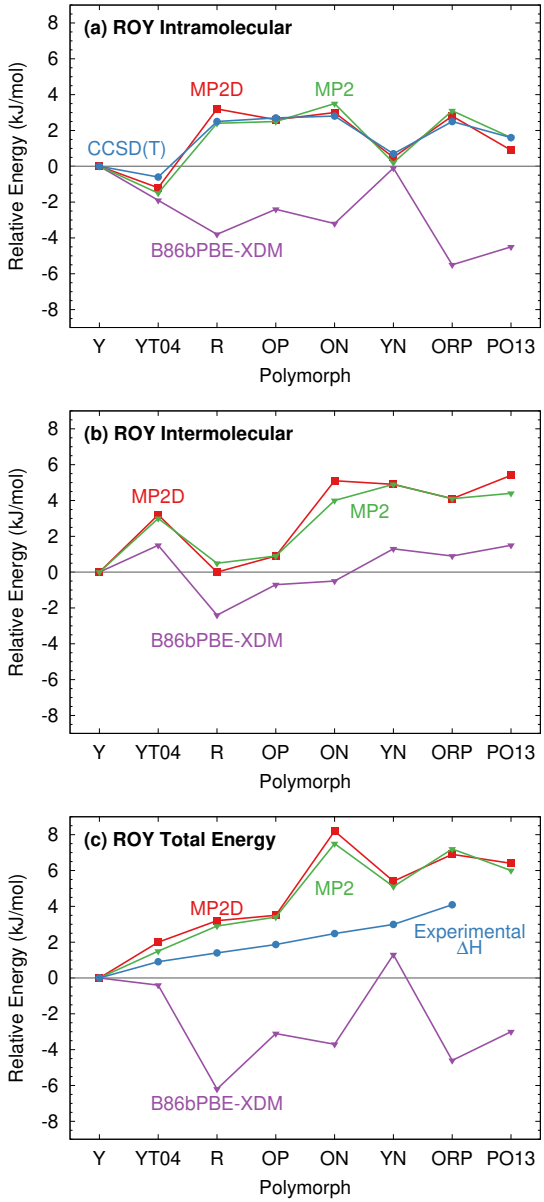


Figure 4.7: Energy decomposition of the room-temperature structure ROY polymorph lattice energies into (a) intramolecular, (b) intermolecular, and (c) total energy contributions. The energies are plotted relative to the most stable Y polymorph in each case.

Like for *o*-acetamidobenzamide, the differences between MP2D and B86bPBE-XDM for the ROY polymorph energies stem from the problematic B86bPBE-XDM intramolecular treatment of the conformational energies. Figure 4.6 plots the one-dimensional conformational energy scan along a key dihedral angle associated with the different conformations found in the ROY polymorphs. For convenience, vertical lines in the figure highlight the corresponding values of this dihedral angle found in the different polymorphs (though the other degrees of freedom along this scan may differ from those found in the actual crystals). Similar conformational energy scans can be found in earlier studies,[202,203] albeit without the CCSD(T) benchmarks provided here. Compared to CCSD(T), B86bPBE-XDM dramatically overstabilizies the conformations found in the red (R) and orange (O) polymorphs relative to those occuring in the yellow (Y) forms. This explains why the earlier DFT calculations predict the yellow forms to be so much less stable than the others. In contrast, MP2D mimics the CCSD(T) conformational energy profile much more faithfully. MP2D does underestimate the stability of the conformations with the dihedral angles adopted by the red and orange forms by up to ∼1 kJ/mol relative to CCSD(T). Nevertheless, MP2D represents a substantial improvement over B86bPBE-XDM. The intramolecular MP2D dispersion correction improves the MP2 conformational energies modestly, by up to ∼1 kJ/mol (Figure 4.6). At the same time, the MP2D dispersion correction does not alter the qualitative MP2 polymorph stability ordering in this system (Appendix C Section C3.2†).

For further insight, Figure 4.7 decomposes the relative polymorph energies into their intra- and intermolecular contributions. As expected from the one-dimensional con-

formational energy scan, B86bPBE-XDM overstabilizes the intramolecular conformations of the orange and red polymorphs, while MP2 and MP2D give conformational energies which are quite faithful to CCSD(T). The intermolecular trends plotted in Figure 4.7b are qualitatively similar between the three models. Indeed, the intermolecular energies are roughly parallel between methods for most of the polymorphs. The most notable difference is that B86bPBE-XDM appears to stabilize the intermolecular interactions of the other forms relative to Y more so than do MP2 and MP2D, which shifts all points other than Y in the B86bPBE-XDM curve down relative to the MP2-based ones in Figure 4.7b. Unfortunately, coupled cluster benchmarks that could assess the quality of the two models for the intermolecular interactions are computationally infeasible.

Combining the intra- and intermolecular contributions (Figure 4.7c), one sees once again that the MP2 and MP2D curves are in much better agreement with experiment than the B86bPBE-XDM. The MP2D energy of the ON polymorph is the most notable outlier relative to experiment. The fact that MP2D predicts the intramolecular conformational energy of the ON polymorph to within 0.2 kJ/mol of CCSD(T) suggests that any problem in the predicted energy ranking arises from the intermolecular contributions.

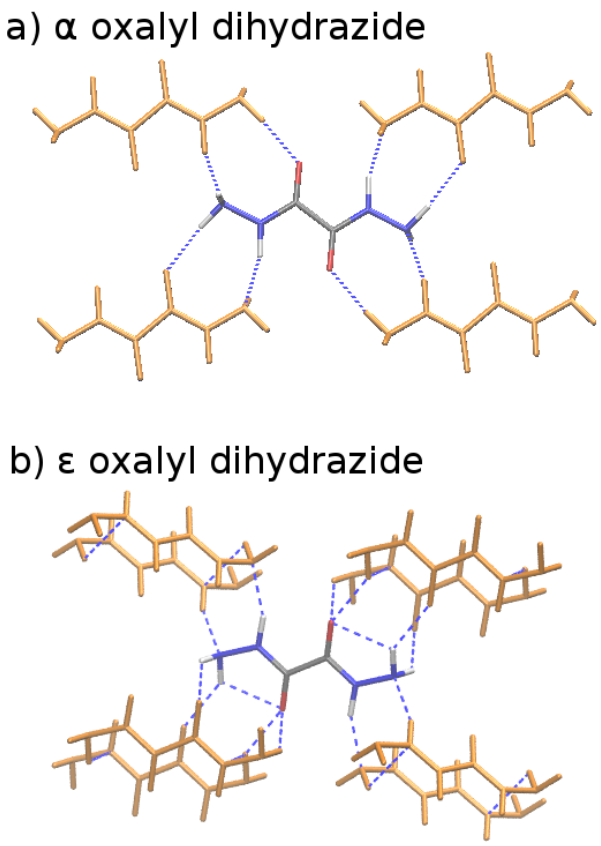


Figure 4.8: Crystalline oxalyl dihydrazide exhibits (a) purely intermolecular hydrogen bonding in the α polymorph and (b) a mixture of intra- and intermolecular hydrogen bonding in the other four polymorphs (ϵ form shown here).

4.3.3 Oxalyl Dihydrazide

The five polymorphs of oxalyl dihydrazide differ in whether the crystal packing contains purely intermolecular hydrogen bonding (α form) or exhibits a mixture of both intra- and intermolecular hydrogen bonds (β, γ, δ, and ϵ forms).[290] Five additional high-pressure polymorphs have been reported,[324] but their structures are unknown and they are not considered here. Experimentally, the α, ϵ, and δ forms are the most stable poly-

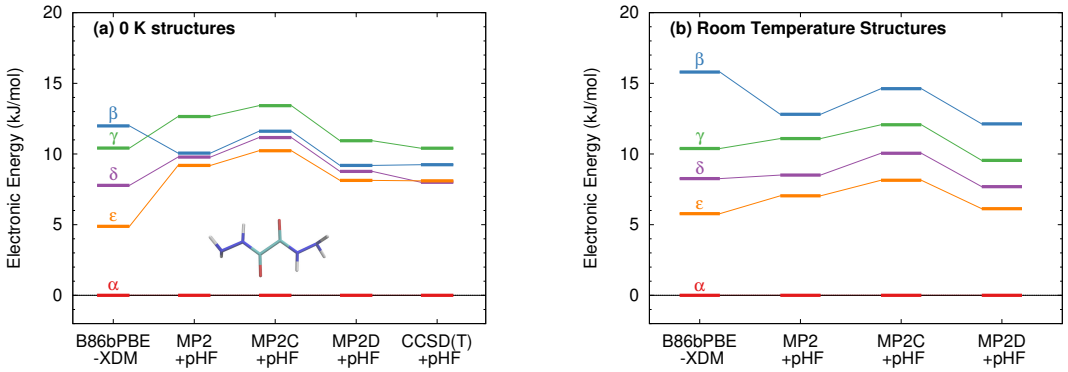


Figure 4.9: Relative stabilities of the oxalyl dihydrazide polymorphs at several levels of theory using both the fully relaxed 0 K structures and the fixed-cell room-temperature structures. For the MP2 methods, the relative stabilities of the β and γ forms differ depending on which structure optimization is used.

morphs, though the ranking among those three is uncertain. The α form should arguably be the most stable polymorph based on its high density,[290] but exceptions to density-based stability arguments can occur in hydrogen bonded crystals. All three forms convert endothermically to γ near 200–210°C. This suggests that the γ form has a lower free energy at these temperatures, but that the α, ϵ, and δ forms forms have lower enthalpies.[290] This indicates an enantiotropic relationship between γ and the other three forms according to the heat-of-transition rule.[325] Finally, the β form has proved difficult to produce and characterize, and it converts readily to the α form.[290,324] Therefore, it is assumed to be the least stable form. Overall, the inferred lattice energy ranking is (from most to least stable): α, δ, $\epsilon < \gamma < \beta$.

These polymorphs have been studied theoretically by several groups,[224,253,264,318] and the results have been summarized by a couple authors.[222,224] Initial DFT calcula-

tions involving empirical dispersion corrections predicted lattice energies consistent with the aforementioned stability ordering, but the lattice energies spanned a surprisingly large range of ∼15 kJ/mol.[253] Subsequent DFT calculations employing more modern dispersion corrections narrowed the polymorph energy range modestly to ∼10–12 kJ/mol. Prior HMBI fragment-based MP2 and MP2C calculations also achieved the same stability ordering, albeit with much smaller ∼3–4 kJ/mol energy window.[318] That study found that the competition between intra- and intermolecular basis set superposition error plays a substantial role in the energetics and that large basis sets are needed when atom-centered basis functions are used. Fully periodic local MP2 calculations performed two years later found a ∼10 kJ/mol energy range for the polymorphs,[264] which are consistent with the DFT calculations, though the double-zeta basis set is probably too small to draw firm conclusions.

Figure 4.9 presents new B86bPBE-XDM DFT results and HMBI-based MP2, MP2D, and MP2C single-point energy refinements of those DFT structures. For the fully relaxed 0 K structures, all four methods exhibit energy gaps in the ∼10–12 kJ/mol energy range, consistent with the earlier DFT and periodic local MP2 calculations. The HMBI results here should be more reliable than the previously published ones,[318] since the geometries were optimized with a more robust B86bPBE-XDM dispersion-corrected DFT functional and because the many-body terms are evaluated with periodic HF instead of a polarizable force field. The strong, favorable polarization that is found only in the α polymorph makes the energy gap between the α and other four forms sensitive to the many-body description. The small energy range for the polymorphs predicted in appears to be an erroneous artifact of the polarizable force field contributions used there.

Like earlier calculations, the relative B86bPBE-XDM lattice energies for the fully relaxed 0 K structures are consistent with the inferred experimental lattice energy stability ordering: $\alpha < \delta < \epsilon < \gamma < \beta$. Using the same structures, the MP2-based methods agree that $\alpha < \delta < \epsilon < \gamma$, though the ϵ form is somewhat less stable than it is with DFT. However, MP2-based methods predict that the β form is more stable than the γ one, contrary to what has been inferred experimentally. Fragment CCSD(T) calculations similarly predict the β form to be more stable. They also predict the δ form to be marginally (0.1 kJ/mol) more stable than ϵ. As noted earlier, the experimental stability ranking among α, δ, and ϵ is unclear.

Because the limited experimental knowledge of the β form is based on its instability at ambient conditions, the calculations were repeated using the fixed-cell room-temperature crystal structures. Using these structures destabilizes the lattice energy of the β polymorph relative to the α one by 4 kJ/mol with B86bPBE-XDM, and by 2–2.5 kJ/mol for the MP2-based methods (Figure 4.9). Moreover, the MP2 models stabilize the room-temperature γ, δ, and ϵ structures by 1–2 kJ/mol relative to the α one. The end result is that all methods predict the $\alpha < \delta < \epsilon < \gamma < \beta$ stability ordering when room-temperature structures are used.

The predicted temperature dependence translates directly to the enthalpies and free energies shown in Figure 4.10. Augmenting the electronic energies with phonon contributions stabilizes the β, γ, δ, and ϵ polymorphs relative to α, but it does not alter the predicted stability orderings for the DFT or MP2 models (Appendix C Table C.7 and Figure S7).[†]

The difference between the lattice energies and free energies is interesting. First, the free energies span a much narrower range than the lattice energies—e.g. 6 kJ/mol versus 15 kJ/mol for MP2D at 298 K—indicating the importance of entropic contributions in this system. Second, the free energy of the γ form stabilizes more rapidly with increasing temperature than do those of the α, δ, and ϵ forms. That is consistent with the experimental evidence for the γ form being enantiotropically related to the other three polymorphs.

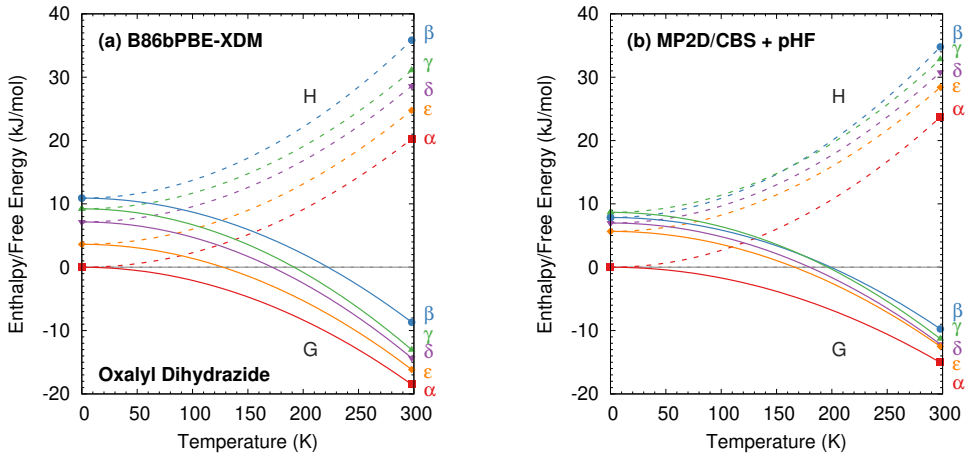


Figure 4.10: Relative enthalpy H and free energy G curves for the five polymorphs of oxalyl dihydrazide computed using (a) B86bPBE-XDM or (b) fragment-based MP2/CBS + pHF. The most notable difference between the two curves is that the β and γ forms are monotropically related in the DFT calculations, but enantiotropically related with MP2D (the G curves intersect near 175 K). The data points indicate computed results, while the curves connecting them are schematic.

Overall, the DFT and MP2-based results here are all consistent with available experimental observations. More experimental data would be needed to discriminate between the distinct DFT and MP2-based predictions for the β and γ polymorph stabilities.

However, two points can be made based on currently available information. First, the experimental crystal structure of β has greater uncertainty than the other forms.[290] That could affect the β fixed-cell room-temperature structure moreso than the fully relaxed one. Notably, the β form contracts $\sim$8% upon full geometry optimization, compared to only $\sim$4% for the other four polymorphs. This bigger structural change between 0 K and room-temperature manifests in the correspondingly large lattice energy change seen for the β form. Of course, this larger structural change in the β form might also also simply reflect differences in the crystal packing that increase its thermal expansivity.

Second, energy decomposition and CCSD(T) benchmarks on the fully-relaxed structures in Figure 4.11 suggest that the MP2D energetics are more accurate than those from B86bPBE-XDM. MP2D and CCSD(T) both predict the intramolecular conformations found in the β and γ polymorphs to be considerably more stable than those adopted in the δ and ϵ polymorphs. In contrast, B86bPBE-XDM predicts erroneously small energy differences between the conformations. For the total pairwise intermolecular interactions, B86bPBE-XDM and MP2D predict fairly similar energetics, with the DFT results actually agreeing better with CCSD(T). This is due to fortuitous error cancellation: examining all the individual dimer interactions, MP2D exhibits a root-mean-square (rms) error of 0.4 kJ/mol versus CCSD(T), compared to 0.8 kJ/mol for B86bPBE-XDM. Similarly, looking at total two-body contributions instead of relative lattice energies, the B86bPBE-XDM 2-body contributions are systematically under-bound by rms error 7.3 kJ/mol compared to CCSD(T). In contrast, MP2D overbinds by a much smaller rms 1.6 kJ/mol. In other words,

the B86bPBE-XDM 2-body terms have much larger errors, but they exhibit better systematic error cancellation here to produce the agreement seen in Figure 4.11b.

The other major difference between models occurs for the many-body contributions. The many-body contributions in all three systems explored in this paper generally amount to only ∼5% of the total lattice energy (and always less than 10%), which is typical for organic molecular crystals.[313] However, because the many-body contributions in a system like oxalyl dihydrazide vary considerably between polymorphs, they play an out-sized role in determining the relative lattice energies. As shown in Figure 4.11c, the relative many-body contributions in oxalyl dihydrazide have the same magnitude as the overall lattice energy differences. Compared to B86bPBE-XDM, periodic HF predicts a considerably stronger polarization effect for the α form that effectively shifts the relative energies of the other polymorphs up. Furthermore, HF predicts the β and γ forms to have more repulsive many-body contributions than does B86bPBE-XDM. While experimental or higher-level theoretical benchmarks are not available to determine which many-body treatment is more accurate, it is clear that B86bPBE-XDM obtains the correct experimental stability ordering only by error cancellation between the intramolecular and many-body intermolecular contributions. For example, if one corrected the erroneous intramolecular B86bPBE-XDM conformational energies with CCSD(T) ones, the result would incorrectly predict that the δ form is the least stable polymorph by 2 kJ/mol. Obtaining the proper stability ordering would also require replacing the B86bPBE-XDM many-body energies with values similar to those obtained from periodic HF. This suggests that the HF many-body contributions are likely closer to the true values.

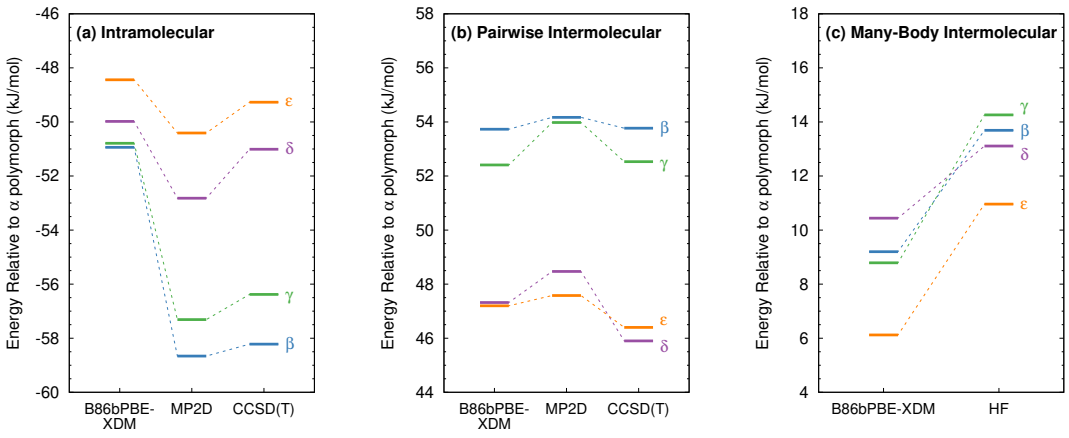


Figure 4.11: Energy decomposition for the oxalyl dihydrazide polymorph stabilities into (a) intramolecular (1-body), (b) pairwise intermolecular (2-body), and (c) many-body intermolecular contributions. Note, in the HMBI fragment approach, both MP2D and CCSD(T) employ the same HF many-body treatment. The same 14 kJ/mol energy range is used in all three figures to facilitate comparisons.

Finally, as shown in Appendix C Table C.8,[†] the MP2-based methods all predict the 1-body energies with similar accuracy. However, for the 2-body interaction energies, MP2 and MP2C systematically overbind the dimers with rms errors of 8.1 kJ/mol and 4.2 kJ/mol, respectively, both of which are several-fold larger than the 1.6 kJ/mol error for MP2D (also mildly over-bound). In other words, the similar relative polymorph stabilities seen for the different MP2-based methods in Figure 4.9 arise from systematic cancellation of the overbinding errors that occur in MP2 and MP2C.

4.4 Conclusions

For many years, it has been widely recognized that balancing intra- and intermolecular conformational energies is one of the primary obstacles to crystal structure prediction in conformational polymorphs. Since the widespread adoption of DFT in crystal structure prediction, however, this issue of balancing the intra- and intermolecular interactions has been given much less attention. While periodic DFT models may have reduced the prevalence of such balance issues compared to earlier force field studies, the results here clearly demonstrate that widely used density functionals have not yet solved the problem of ranking conformational polymorphs in crystal structure prediction.

This study examined three well-known and challenging examples of conformational polymorphism: *o*-acetamidobenzamide, ROY, and oxalyl dihydrazide. In the first two systems, a variety of dispersion-corrected DFT models predict catastrophically wrong relative polymorph stabilities. The ∼8 kJ/mol DFT errors in relative polymorph stabilities found in these two systems greatly exceed the few kJ/mol error often associated with DFT polymorph rankings. More importantly, given that over 95% of polymorph pairs exhibit energy differences less than 8 kJ/mol, and the majority have energy differences less than 2 kJ/mol, such errors are unacceptable in crystal structure prediction. In both systems, the problem for DFT arises largely from a poor description of the intramolecular conformational energy.

The third system, oxalyl dihydrazide, is more nuanced, in part due to greater ambiguity in the experimental data. High-quality experimental thermochemical measurements of this (and other) polymorphic systems would be valuable for assessing the performance of different models more clearly. Based on presently available data, the overall DFT energy

rankings for the oxalyl dihydrazide polymorphs do appear generally consistent with experiment. However, energy decomposition and coupled cluster theory benchmarks suggest that such nominal agreement arises from substantial cancellation of errors between the intra- and intermolecular interactions.

While the present study focused on the B86bPBE-XDM functional, it is important to recognize that these issues transcend any individual density functional. They occur for both GGA and hybrid density functionals with a variety of dispersion treatments, as evidenced by Figure 4.5. Futhermore, the conformational polymorph examples here raise the question: How common are such failures of DFT for crystal structure prediction? Are these three systems outliers? Will more examples be uncovered as DFT-based crystal structure prediction techniques are increasingly applied to larger, more flexible pharmaceutical molecules?

This work also demonstrates that fragment-based MP2D calculations provide a promising path forward. For ROY and *o*-acetamidobenzamide, the higher-level calculations not only correct the qualitative polymorph stability ordering, but they also predict relative stabilities that are in generally good agreement with experiment. For oxalyl dihydrazide, MP2D also appears to perform better than DFT based on the energy decomposition results, though more experimental data to confirm the predictions would be helpful.

The study also examined a few variants of MP2. The balanced description of both intra- and intermolecular dispersion makes MP2D superior to MP2C, which corrects intermolecular dispersion only, and to MP2, which performs poorly for both intra- and intermolecular dispersion. The intramolecular dispersion correction contributes in all three

systems, but it proves critical in predicting the stability for the *o*-acetamidobenzamide polymorphs. Fortunately, the MP2D dispersion correction can be computed with trivial effort once MP2 results are available.

The overall computational cost of these calculations is considerably higher than DFT, with complete-basis-set MP2D single-point energies requiring approximately 900, 7,000, and 21,000 central processing unit (CPU) hours per polymorph of oxalyl dihydrazide ($C_2H_6N_4O_2$), acetamidobenzamide ($C_9H_{10}N_2O_2$), and ROY ($C_{12}H_9N_3O_2S$). This computational cost is dominated by the evaluation of ∼40–50 dimer interactions at the large-basis MP2D limit and the N^5 MP2 scaling with monomer/dimer size.[224] Fortunately, the number of monomer and dimer fragments scales linearly with increasing number of molecules in the asymmetric unit.[285,312] Furthermore, those fragment calculations can be run independently and in parallel, making it feasible to perform crystalline calculations in much shorter amounts of wall time in modern high performance computing environments. The present study demonstrates that such correlated wavefunction calculations are feasible in species that are comparable in size to small-molecule pharamceuticals. Efforts are currently underway to improve the accuracy and reduce the computational costs of these correlated wavefunction models further.

Chapter 5

Innacurate conformational energies sill hinder crystal structure prediction in flexible organic molecules

5.1 Introduction

Molecular crystal packing plays a key role in determining properties of the solid state, and the ability to predict crystal structures *a priori* would have major benefits for the pharmaceutical industry and other areas of chemistry. Crystal structure prediction (CSP) is increasingly used to help solve challenging crystal structures[224] and to aid solid

form screening.[326] CSP driven by dispersion-corrected density functional theory (DFT) models has a long record of successes in blind test species,[227–230,236,237,241,327] pharmaceuticals,[240,244,245,247,251,252] and other species.[231,235,238,239,328–330]

However, polymorph stability depends on both intra- and intermolecular contributions to the crystal energy, and delocalization error in widely-used generalized gradient approximation (GGA) and hybrid density functionals can significantly impact the accuracy with which these interactions are predicted. Delocalization error spuriously stabilizes the salt forms of some neutral co-crystals, for example.[256] In conformational polymorphs,[12] where different intramolecular conformations enable alternate intermolecular packings, the artificially strong preference for extended π conjugation leads to highly incorrect polymorph rankings in systems such as ROY and *o*-acetamidobenzamide (Figure 5.1).[188] We recently showed how computing polymorph stabilities with higher-level correlated wave function methods via the fragment-based hybrid many-body interaction (HMBI) model[312] dramatically improves polymorph rankings in these systems.[188] Unfortunately, such calculations are computationally infeasible for many pharmaceuticals.

The present study demonstrates how conformational polymorph energy rankings can often be corrected by a simple and computationally affordable "monomer-corrected" DFT strategy that refines the conformational energy contribution to the lattice energy using a higher level of theory. The need for accurate conformational energies in flexible molecules has been recognized since the early years of CSP. Polymorph stability rankings computed from classical force fields improved considerably when they were augmented with quantum chemical conformational energies,[291,293,331] but this idea has largely been aban-

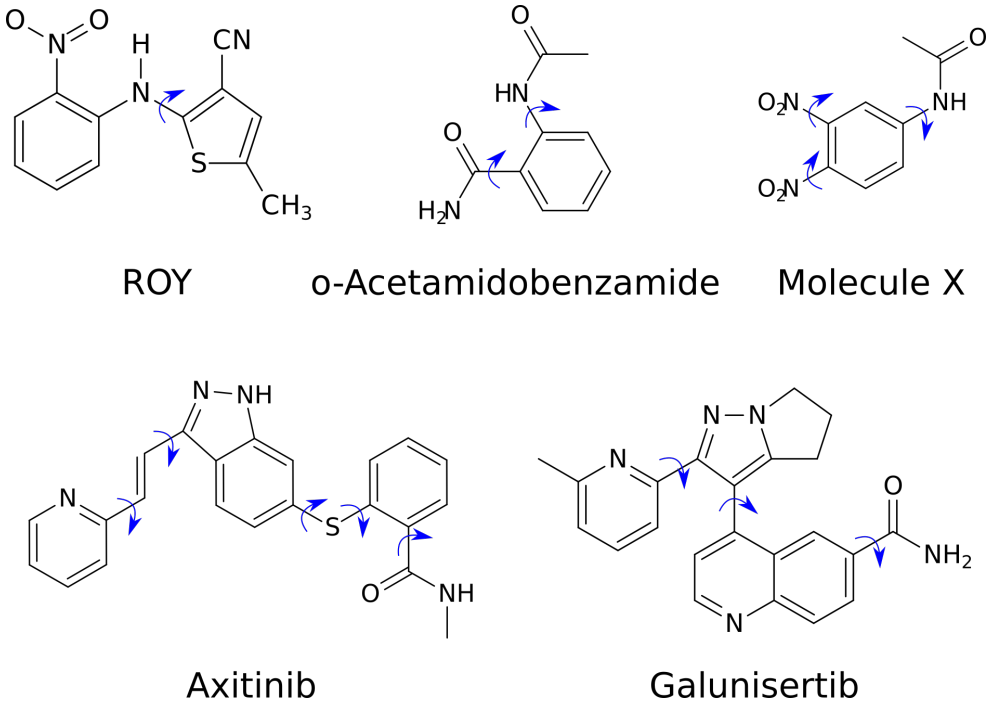


Figure 5.1: The five species whose crystal polymorphs are considered here. Arrows highlight the most important flexible torsion angles.

doned with the shift toward computing lattice energies entirely with DFT. The three small molecules and two pharmaceuticals considered here (Figure 5.1) emphasize how important intramolecular conformational energy corrections to DFT lattice energies can be for obtaining polymorph stabilities that agree with experiment and/or higher-level calculations.

Intramolecular corrections to DFT crystal energies and properties have occasionally been used previously.[204,237,332] Monomer-correction can also be viewed as the simplest version of incremental,[333] fragment,[224] and multi-layer methods.[259] Despite these precedents, the systematic failures of commonly-used density functionals for conformational polymorphs and the route for overcoming them have not been widely appreciated.

5.2 Theoretical Approach

The monomer correction is performed here with our recently developed second-order Møller-Plesset perturbation (MP2D) theory model,[5] which describes conformational energies well and is affordable for molecules with up to ∼100 atoms. The corrected crystal energies are computed as:

$$E_{crystal} = E_{crystal}^{DFT} + \sum_{i} \left(E_{mon,i}^{MP2D} - E_{mon,i}^{DFT} \right) \tag{5.1}$$

where $E_{crystal}^{DFT}$ is the energy of the crystal computed with periodic DFT, while the terms in parentheses correct the crystal energy based on the difference between MP2D and DFT for each isolated monomer in the unit cell. The gas-phase monomer correction only needs to be computed for the symmetrically unique monomers, and for systems of the size considered here, it can be evaluated with considerably less computational effort than the preceding DFT crystal geometry optimization. This intramolecular correction approach assumes that the chosen DFT functional performs well for the intermolecular interactions. While this assumption will often be true,[110,224,225] there are notable exceptions for cases such as ions[257] and halogen bonds[195,256,257] for which hybrid (or better) functionals may be needed for the intermolecular interactions.

All MP2D conformational energies are extrapolated to the complete basis set limit using standard Gaussian basis sets. The planewave DFT calculations here are performed using the dispersion-corrected B86bPBE-XDM GGA functional that has performed well in many previous studies.[235–237,310] Importantly, the conformational energy problems demonstrated here are not unique to B86bPBE-XDM—they frequently occur for other common dispersion-corrected GGA and hybrid functionals like PBE-D3(BJ), PBE0-D3(BJ), and

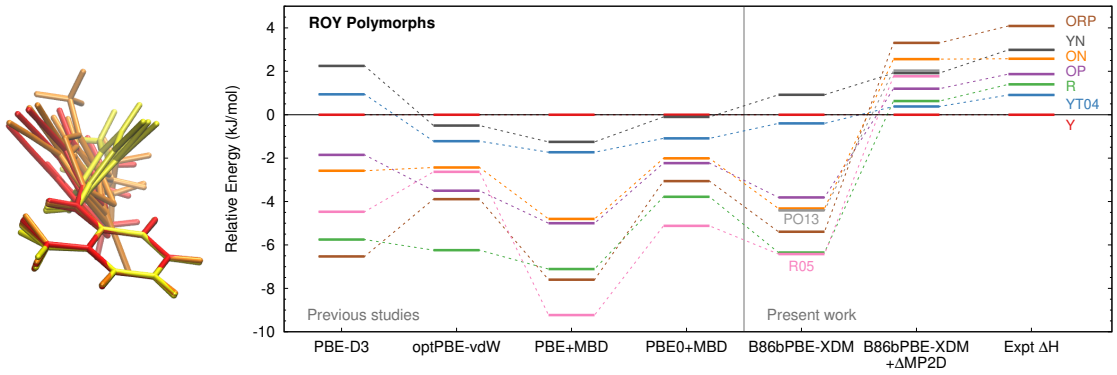


Figure 5.2: Overlay of the intramolecular conformations from the ROY polymorphs highlighting how the angle between the two rings generally decreases from yellow to orange to red polymorphs. Relative 0 K lattice energies of the nine known polymorphs of ROY calculated using dispersion corrected DFT and the monomer-corrected energies, compared to experimentally measured enthalpy differences. Many DFT functionals predict the ROY polymorph energies incorrectly, but correcting the B86bPBE-XDM lattice energies with MP2D conformational energies ("+ΔMP2D") performs far better.

B3LYP-D3(BJ) as well. Further details of the computational methods, crystal structures, and detailed analysis of conformational and polymorph energies for each system are found in Appendix D.

5.3 Results and Discussion

First, we consider two systems which were previously studied[188] using the more accurate and computationally demanding HMBI fragment model, which treats both intramolecular and intermolecular interactions with wave function-based methods. ROY has

nine well-characterized polymorphs[3,205,206] whose red, orange, and yellow colors are closely correlated with the torsion angle coupling the two aromatic rings.[202,203] Accurate experimental enthalpies have been measured for seven of the forms.[3] Both GGA and hybrid density functionals with high-quality dispersion corrections predict ROY polymorph stabilities incorrectly (Figure 5.2).[188,203,205] They over-stabilize the intramolecular conformations found in the red and orange polymorphs that exhibit stronger coupling between the aromatic rings.[188,202,203] In contrast, MP2D predicts ROY conformational energies in good agreement with CCSD(T) benchmarks (Appendix D Figures D.2–D.3).[188] Simply monomer-correcting the B86bPBE-XDM crystal energies with MP2D dramatically improves the agreement between the predicted polymorph lattice energies and the experimental enthalpies. Only the YN and ON polymorph orderings are reversed compared to experiment, with YN over-stabilized by ∼2 kJ/mol.

The two polymorphs of *o*-acetamidobenzamide represent another difficult case for common DFT functionals.[188,253] In the α form, the side chains form an intramolecular hydrogen bond and are planar to the benzene ring, extending the π conjugation. The β form disrupts this planarity/conjugation in pursuit of better intermolecular hydrogen bonding. Experimentally, the β form is more stable by 1.9–2.9 kJ/mol,[253,307] but a variety of GGA and hybrid functionals incorrectly predict the α form to be 5–10 kJ/mol more stable.[188,253] As we identified previously,[188] the primary error in the DFT calculations arises from over-stabilization of the highly-conjugated planar α form conformation (Table D.3).

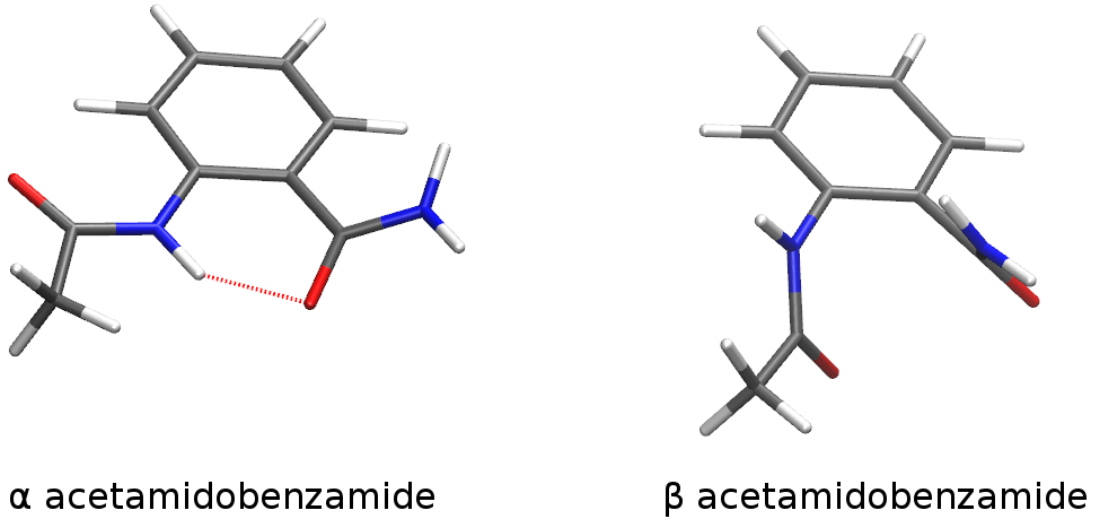


Figure 5.3: Comparison of the intramolecular conformations adopted by the α and β polymorphs of *o*-acetamidobenzamide. The α form is more planar and exhibits an intramolecular hydrogen bond.

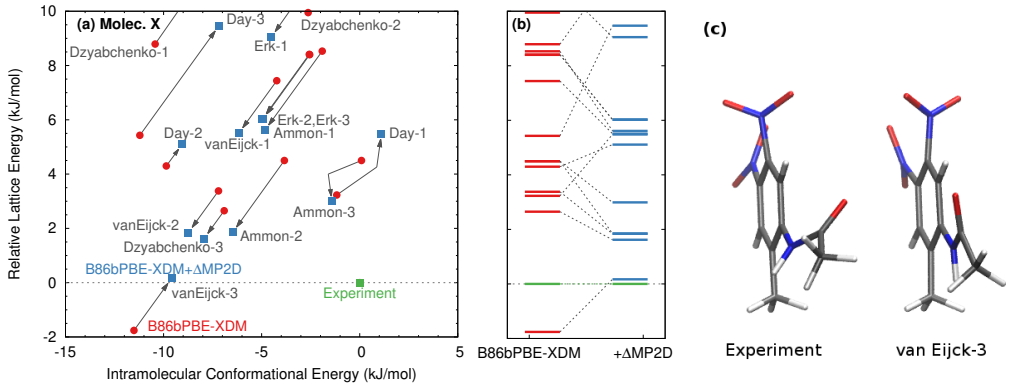


Figure 5.4: (a) Relative 0 K lattice energies versus conformational energies for the low-lying candidate structures of Molecule X before and after monomer correction (see Table D.5 for data on higher-energy structures). (b) Another representation of the relative lattice energies that highlights the impact of the monomer corrections on the stability ordering. MP2D revises the intramolecular conformational energies and stability ordering of the crystal energy landscape considerably. (c) Comparison of the intramolecular conformation of the experimental and vanEijck-3 crystal structures. In contrast to the experimental structure, the amide group in vanEijck-3 is essentially planar relative to the aromatic ring.

Applying the simple MP2D monomer correction to the B86bPBE-XDM lattice energies dramatically stabilizes the β form relative to α and corrects the stability ordering. Estimating the enthalpy difference between the two forms at the experimental 423 K transition temperature via the procedures used in (see Appendix D Section D.3 for details), the monomer-corrected model predicts $\Delta H_{\alpha\rightarrow\beta}(423\text{K}) = -0.8$ kJ/mol, versus -1.9 or -2.9 kJ/mol experimentally. This represents a major improvement over the qualitatively incorrect B86bPBE-XDM value of +5.9 kJ/mol and is close to the far more expensive HMBI MP2D value of -2.0 kJ/mol (Table D.4).

Next, we examine Molecule X from the Third Blind Test of CSP,[334] in which the amide and two nitro groups can potentially conjugate with the benzene ring. Whittleton et al[237] re-ranked twenty-four candidate structure submissions with B86bPBE-XDM. Unlike earlier PW91 calculations with empirical dispersion,[327] they found the vanEijck-3 structure to be 1.8 kJ/mol more stable than the experimentally known structure (Figure 5.4), raising the question of whether the experimental structure is actually the most stable polymorph. The vanEijck-3 structure exhibits a nearly planar amide conformation that is artificially stabilized by the GGA functional. Whittleton et al showed how correcting the intramolecular conformational energy with MP2/aug-cc-pVTZ raises the energy of the vanEijck-3 structure to 1 kJ/mol higher than the experimental one, restoring the experimental structure as the most stable one.

However, this story proves incomplete due to the finite basis set used in that work and the limitations of MP2. Performing the monomer correction with complete-basis-set MP2D instead leads to the vanEijck-3 structure being a mere 0.2 kJ/mol above the experi-

mental one (Figure 5.4). Higher-level benchmarks confirm these MP2D-corrected stabilities: performing the monomer correction with CCSD(T) instead of MP2D places vanEijck-3 0.4 kJ/mol above the experimental structure, while full, fragment-based HMBI MP2D calculations that also refine the intermolecular description place vanEijck-3 at 0.1 kJ/mol above the experimental structure. In other words, vanEijck-3 is energetically competitive with the experimental structure. If these electronic energies are combined with earlier phonon calculations that suggest the vibrational free energy contribution preferentially stabilizes vanEijck-3 by ∼1 kJ/mol relative to the experimental one,[237] vanEijck-3 may indeed be more stable than the experimental crystal structure.

More broadly, the low computational cost of the monomer correction enables re-ranking of the entire crystal energy landscape. As shown in Figure 5.4, doing so transforms the Molecule X landscape, with a mean absolute energy change of 1.9 ± 1.0 kJ/mol and substantial re-ordering of the candidate structures. Such changes are significant when half of all experimentally known polymorphs are separated by 2 kJ/mol or less in lattice energy.[221,222] Even if the monomer-corrected rankings turn out to be imperfect due to the limitations of the GGA intermolecular interaction treatment, the high sensitivity of the crystal energy landscape to the conformational energy indicate that the DFT rankings may warrant skepticism.

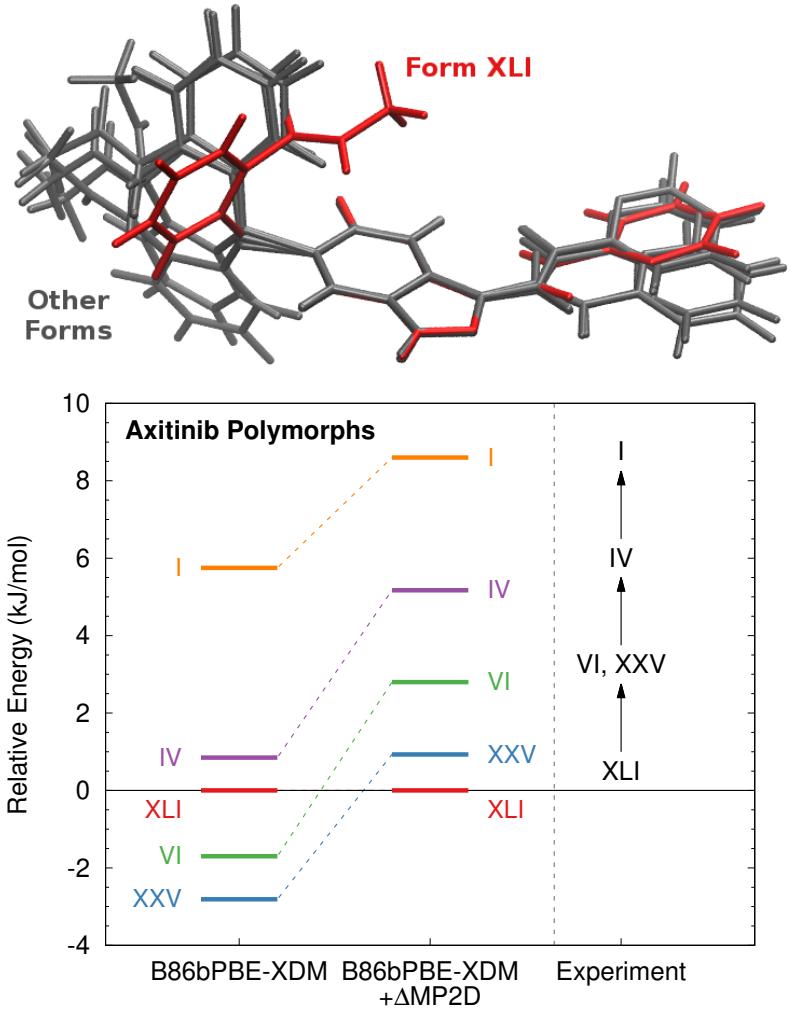


Figure 5.5: Overlay of the monomer conformations and relative 0 K lattice energies for the five axitinib polymorphs compared to the experimentally inferred stability ordering. The thermodynamically stable XLI polymorph of axitinib (red) adopts a folded conformation, in contrast to the extended conformers found in the other polymorphs. B86bPBE-XDM ranks the form XLI incorrectly relative to the other forms, but monomer-correction fixes the qualitative polymorph ordering.

Finally, DFT conformational energy problems can have significant impacts when studying the solid forms of pharmaceuticals, as demonstrated by two challenging examples here. Pfizer's anti-cancer drug axitinib has five known neat polymorphs and 66 solvates to date.[335,336] Forms I, IV, VI, and many solvates were discovered via standard experimental solid form screening procedures. Form IV was believed to be the thermodynamically stable form and was initially targeted for development.[335] However, more stable forms XXV and XLI were fortuitously discovered later on during the manufacturing campaign, and thermodynamically stable form XLI became the commercial form.[335,336] The difficulty in crystallizing Form XLI has been attributed to its distinct intramolecular conformation and unique 1-D chains of hydrogen bonds instead of the hydrogen bonded dimers found in the other polymorphs. Differential scanning calorimetry and solubility experiments indicate the stability ranking:

$$\text{XLI (most stable)} > \text{XXV, VI} > \text{IV} > \text{I}$$

where the energy ordering of XXV and VI is uncertain.[335,336]

The difficulties in discovering the Axitinib polymorphs experimentally make it an excellent example where CSP could help. Previous CSP studies based largely on force fields generated the experimental crystal structures but ranked them poorly.[35,337] In , for example, form XLI lies 10 kJ/mol above form VI. Here, B86bPBE-XDM predicts most of the polymorph stability orderings reasonably, but it incorrectly suggests that form XLI is 2–3 kJ/mol less stable than forms XXV and VI (Figure 5.5).

Once again, the key problem with the GGA ranking of the axitinib polymorphs lies in the intramolecular conformation energies. Whereas axitinib adopts an extended con-

formation in most polymorphs, form XLI folds the molecule to interact the amide π system with the indazole ring, disrupting the π conjugation between the amide and the benzene ring (Figure 5.5). GGA functionals like B86bPBE-XDM and PBE-D3(BJ) predict the XLI conformation to be one of the least stable intramolecular conformations found in the polymorphs. In contrast, higher-level models like MP2D, MP2, or B3LYP-D3(BJ) reveal that the XLI conformation is actually one of the most stable conformations (Figure D.6). The favorable intramolecular conformation and strong intermolecular hydrogen bond network actually make form XLI the most stable polymorph. Indeed, applying the MP2D monomer correction stabilizes form XLI by several kJ/mol relative to the other forms and reproduces the experimental stability ordering (Figure 5.5). Advance prediction of the highly stable form XLI enabled by the monomer-correction approach might have accelerated the experimental discovery of this thermodynamically stable form.

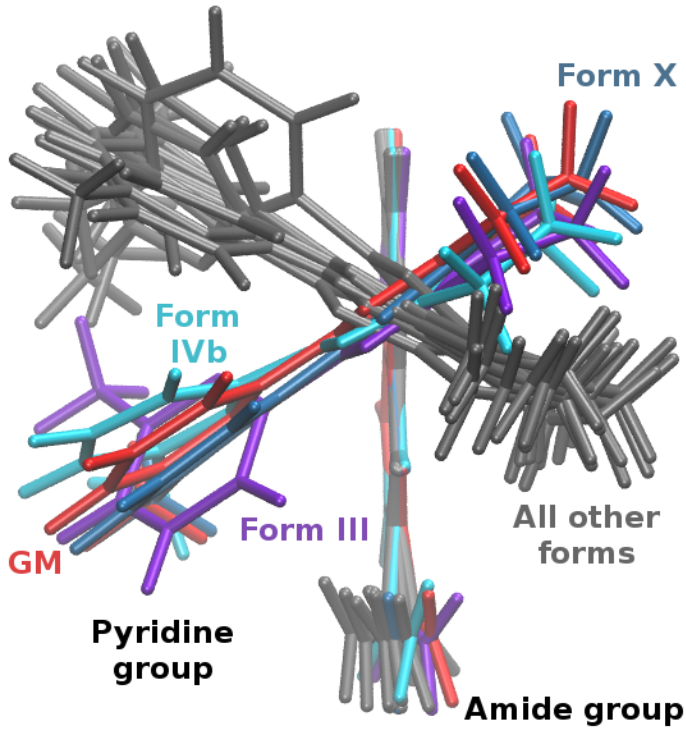


Figure 5.6: Overlay of the galunisertib monomer conformations. The viewer is looking down the central torsion angle in the middle of the image. The GM conformer is in red. Monomers III, IVb, and X have similar conformations around the central torsion angle, but they differ in the degree of planarity in the pyradine and amide groups relative to the adjacent parts of the molecule.

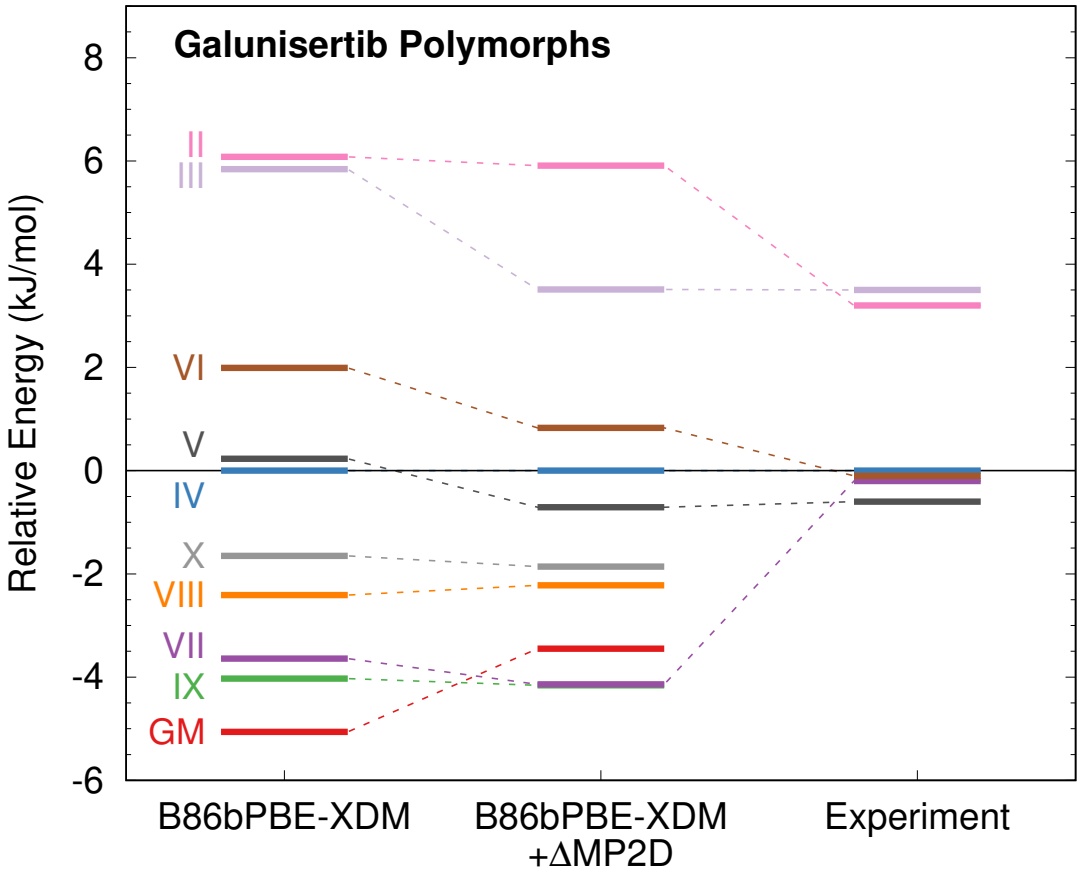


Figure 5.7: Relative 0 K lattice energies of nine experimentally known galunisertib polymorphs (higher-energy form I is omitted here) and the putative GM structure compared to the experimental enthalpies that have been measured for several polymorphs. Monomer-correcting galunisertib with MP2D destabilizes the GM structure by ∼1.7 kJ/mol and generally shifts the lattice energies closer to the experimentally measured enthalpies.

Lilly's anti-cancer drug candidate galunisertib provides another example where monomer correction could have been useful. The solid form landscape for this drug has been studied extensively, with ten well-characterized polymorphs and over 50 solvates.[251] Two different CSP studies were commissioned during the solid form screening, and both predicted a global minimum (GM) structure that was more stable than any of the experimental forms. All attempts to crystallize the GM structure have failed, despite years of effort. It was argued that the intramolecular conformation of the GM structure is too strained to crystallize effectively.[251]

However, we propose an additional factor that inhibited its crystallization: the GM structure is actually somewhat less stable than the earlier DFT studies suggested. Galunisertib contains three key flexible torsion angles that impact the degree of π conjugation throughout the molecule. While none of the conformations considered exhibits appreciable conjugation between the pyrazole and quinoline rings (i.e. across the central torsion angle), the GM structure has the most planar conformations at the other two torsion angles, thereby extending conjugation from pyrazole to pyridine and from quinoline to amide (Figures 5.6 and D.7). Accordingly, GGA and hybrid functionals tend to over-stabilize the GM relative to the other forms (Figures D.8 and D.9).

Applying the MP2D monomer correction destabilizes the GM structure by about 1.7 kJ/mol relative to form IV, such that it becomes less stable than two of the experimentally known polymorphs (Figure 5.7). The monomer correction shifts several other polymorphs closer to the experimental enthalpies, though considerable differences remain. Further work is needed to understand the role of intermolecular interactions (including dis-

persion treatments[251]) and phonon contributions to the polymorph stabilities. Errors from those sources could each be similar to or larger than the ∼2 kJ/mol monomer corrections here. Regardless, the destabilizing impact of the monomer correction on the GM structure calls its exceptional stability into question. Such results might have reduced the effort spent trying to crystallize the GM form unsuccessfully.

5.4 Conclusions

In conclusion, commonly used GGA and hybrid density functionals struggle to rank crystal structures in which the degree of intramolecular π conjugation varies significantly. A simple, intramolecular monomer correction computed at a higher level of theory improves the conformational energies and therefore the polymorph rankings significantly. The specific electronic structure model used to correct the conformational energies is at the discretion of the user: dispersion-corrected MP2 was used here because it provides high-quality conformational energies at modest computational cost, but one could also use double-hybrid density functionals or any other model that predicts conformational energies reliably.

Monomer correction has limits. It cannot solve problems in the lattice energies that stem from the intermolecular DFT interactions. In cases like oxalyl dihydrazide,[188] the monomer correction would disrupt the fortuitous error cancellation observed between the inter- and intramolecular interactions that enables reasonable GGA polymorph stability orderings. Common hybrid functionals do not fix the conformational energies in several of the examples discussed here, but perhaps they would be sufficient to improve the in-

termolecular interactions when GGAs prove inadequate. Nevertheless, hybrid functionals still exhibit mean absolute lattice energy errors of ∼4 kJ/mol in even relatively simple crystals.[110,224,258,338] Even if some of that intermolecular interaction error cancels when computing energy differences between polymorphs, the residual errors in the intermolecular description will often probably exceed the errors in the monomer-corrected conformational energies.

In addition, most of the work presented here examined 0 K lattice energies and omitted thermal vibrational contributions to the free energy. Vibrational contributions cause thermal expansion of the unit cell and can impact polymorph stability appreciably.[221] Enantiotropism, or temperature-dependent changes in polymorph stability orderings, is relatively common.[53] Several of the axitinib polymorphs are enantiotropically related, for example, and it would be interesting to examine to what extent the combination of monomer-correction and thermal contributions can correctly reproduce those temperature-dependent stability changes.

Even if the monomer-corrected polymorph rankings are imperfect, a large monomer correction serves as a warning that the DFT polymorph rankings may not be reliable. Use of such a diagnostic might have increased skepticism regarding the putative GM form of galunisertib, for example. Given the low computational cost, monomer conformational energy corrections can and should be applied routinely in CSP when the conformations vary widely across crystal forms, especially if they alter the degree of π conjugation or the nature of the intramolecular non-covalent interactions changes appreciably.

Chapter 6

Rubrene untwisted: common density functional theory calculations overestimate its deviant tendencies

6.1 Introduction

High carrier mobility of 20–40 cm^2/Vs[339] makes rubrene exceptional among organic semiconductors.[340] However, the rubrene carrier mobility drops precipitously in solution or thin film devices due to a change in the intramolecular conformation from a planar tetracene backbone in the solid-state to a twisted backbone structure in the other phases (Figure 6.1).[341] The unique semiconducting properties in crystalline rubrene result from

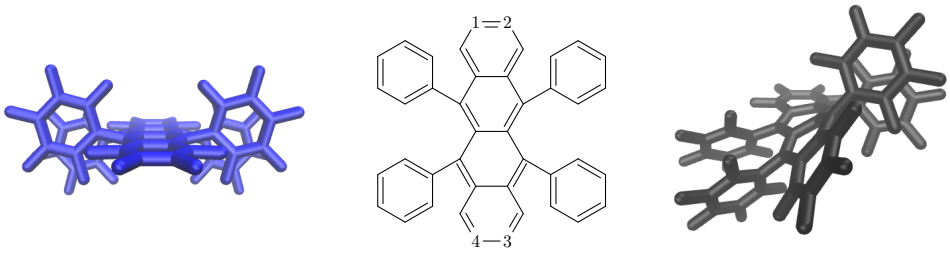


Figure 6.1: Rubrene in its planar (blue) and 60° twisted (gray) conformations. The twisting is defined by the dihedral angle formed from atoms 1–4.

favorable overlap between rubrene molecules with planar tetracene backbones and favorable intermolecular interactions with neighboring molecules.[342,343] Twisting of the tetracene backbone and the concomitant changes in crystal packing disrupt both of these favorable interactions.

For many years, researchers have probed the chemical space of rubrene derivatives and other oligoacenes in pursuit of a superior organic semiconductor. Although many interesting materials have emerged from this research,[340,344] none has yet matched rubrene's superior charge mobility. A successful rubrene derivative will likely need to adopt a planar tetracene core and favorable intermolecular packing. While all three known crystal polymorphs of rubrene exhibit planar tetracene backbones,[345,346] other rubrene derivatives can adopt either planar or twisted conformations in the solid-state.[344,347–356] The solid-state conformation adopted by a given derivative is difficult to predict *a priori* from chemical intuition alone, although the electron donating and withdrawing properties of functional groups are loosely correlated with twisted or planar structures respectively[353]. In some cases, a given derivative can form both planar and twisted crystal polymorphs.[348,357,358]

Extensive analysis of the intramolecular conformational energies and intermolecular packing interactions found that symmetric intermolecular interactions are needed in the crystal to stabilize the planar tetracene backbone.[354] This same study highlights how when chemically modifying rubrene, one must consider both the opportunity to form new interactions and the potential for disrupting existing favorable interactions.

The complex interplay of intra- and intermolecular interactions makes crystal engineering particularly challenging for conformationally flexible molecules like rubrene. Crystal structure prediction seeks to predict the entire crystal energy landscape—the set of all energetically favorable crystal structures.[326,359] From this landscape, it is ideally possible to identify both the most stable crystal packing motifs and perhaps other low-energy crystal forms with desired properties. CSP has grown increasingly reliable in recent years, as evidenced by successes in recent blind tests[227,229,230] and other applications.[209,239,241,247,251,252,360] This raises the prospects of using crystal structure prediction to search for crystal structures of rubrene derivatives which exhibit good intramolecular structural and electronic properties as well as suitable intermolecular packing. Crystal structure prediction has already been used to design new highly porous organic materials[360] and to screen organic semi-conductor materials.[361–363]

However, *in silico* design strategies are effective only if the models used correctly predict the stabilities of the crystal structures. Density functional theory (DFT) is widely used for modeling the energetic, structural, and electronic properties of organic materials[364–366] because it generally balances between accuracy and computational efficiency well. The omission of van der Waals dispersion interactions from widely used generalized gradient

approximation (GGA) and hybrid density functionals was previously a problem, but this limitation has been overcome through the development of good-quality dispersion corrections.[110,225] On the other hand, delocalization error remains pervasive in these functionals[111,168] and can lead to surprisingly large problems in organic materials. For example, delocalization error can cause spurious proton transfer which artificially converts a neutral multi-component crystal into its salt form.[256] It leads to underestimation of barrier heights[367–370] and dissociation energies.[371–373] Delocalization error can artificially stabilize molecules with more extended π conjugation relative to those with less.[190,322,323] A number of organic and pharmaceutical conformational polymorphs have been found where delocalization error and the competition between highly conjugated planar conformations and less-conjugated non-planar conformations leads to incorrect polymorph stabilities.[188,189]

Delocalization error can often be attributed to errors in the electron density.[374] In such cases, improving the electron density can improve the energetics for a given density functional considerably.[375–378] However, obtaining quantitatively accurate results may require using more advanced density functionals (e.g. double-hybrid functionals) or wave function-based models. Unfortunately, such models are often too computationally demanding for direct application to solid-state organic materials. We have recently demonstrated that dispersion-corrected second-order Møller-Plesset perturbation theory (MP2D)[5] and its spin-component-scaled analog SCS-MP2D[379] can predict conformational energies far more accurately than typical GGA and hybrid density functionals. Combining intramolecular MP2D (to describe the conformational energy) with intermolecular DFT (to describe the lattice contributions) leads to greatly improved polymorph stabilities in a number of con-

formational polymorphs ranging from small molecules like ortho-acetamidobenzamide[189] to pharmaceuticals like galunisertib and axitinib[189] and photomechanical materials like 9-tert-butyl anthracene ester.[380]

The present study revisits the energetics associated with planar and twisted rubrene conformations with high-quality electronic structure methods. We demonstrate how commonly used density functionals artificially stabilize the twisted form of rubrene relative to the planar one by several kJ/mol, which represents a chemically important fraction of the twisting energy. More significantly, we show how these poor conformational energy differences can artificially bias predictions of crystal stability toward polymorphs containing twisted rubrene molecules instead of planar ones. This means that naive DFT modeling of these crystals can produce misleading results regarding the thermodynamic stability of polymorphs for rubrene-like molecules that differ in the conformations of their tetracene cores. Moreover, in contrast to many earlier DFT estimates,[353,354,358,381] the smaller conformational energy differences between twisted and planar rubrenes found here lie well within the typical energy window associated with conformational polymorphism.[12,222,382] This suggests that conformational polymorphs are more likely to occur than previous results would have indicated. Therefore, it could be worthwhile screening more extensively for planar-backbone polymorphs of rubrene derivatives which are known to crystallize in a twisted conformation.

6.2 Computational Methods

To investigate the twisting energies of rubrene and its derivatives, gas-phase structures were optimized at the B3LYP/cc-pVDZ level of theory using Gaussian 09.[383] Although this functional omits description of van der Waals dispersion, it was chosen to match earlier work.[353] These optimizations include a series of constrained optimizations with tetracene backbone twists ranging from 0–60° to map out the twisting energy curves for rubrene, half-fluorinated rubrene (F_{14}-rubrene), and perfluororubrene. For the rubrene derivatives taken from ref 353, twisted structures were obtained via complete relaxation of the structures, while planar ones were obtained via constrained optimization.

Single-point energies were then computed for these structures using a variety of dispersion-inclusive density functionals and correlated wavefunction methods, primarily using PSI4.[137] The DFT functionals employ the def2-QZVP basis,[384] while the MP2 ones were extrapolated to the complete-basis-set (CBS) limit[70] from the aug-cc-pVTZ and aug-cc-pVQZ basis sets.[140] SCS-MP2D results were obtained from the PSI4 MP2 values and an in-house code which performs the correction. Benchmark domain-based local pair natural orbital coupled cluster singles, doubles, and perturbative triples (DLPNO-CCSD(T))[166] calculations were performed in the cc-pVDZ basis using ORCA[385] and extrapolated to the CBS limit by combining them with the MP2/CBS results using the standard focal point technique.[141] The DLPNO-CCSD(T) calculations employed both tight self-consistent field and pair-natural orbital settings to achieve better fidelity to canonical CCSD(T) results.

Finally crystalline calculations were performed for three polymorphs of rubrene (CSD reference codes QQQCIG07,[345] QQQCIG13,[346] QQQCIG14[346]) and two polymorphs

of perfluororubrene (INELUK02 and INELUK03[358]). The structures were fully relaxed with the dispersion-corrected B86bPBE-XDM functional using QuantumEspresso.[315]. A kinetic energy cutoff of 50 Ry was used for the wavefunctions, and a charge density and potential cutoff of 500 Ry was used. Core electrons were treated via the projector augmented wave (PAW) approach using PAW potentials for H, C, and F generated with A. Dal Corso's Atomic code v6.1. Monkhorst-Pack k-point sampling grids were used as specified in Appendix E Table E.4. The crystal energies were then refined via the monomer-correction approach,[189] in which the crystal energy is corrected based on the gas-phase molecular conformational energy differences between B86bPBE-XDM and a higher level of theory.

6.3 Results and Discussion

6.3.1 Gas-phase twisting energies

We begin by benchmarking the potential energy curves generated by twisting the tetracene backbone dihedral angle from 0-60° for rubrene, its half-fluorinated F_{14}-rubrene derivative, and the completely fluorinated perfluororubrene. Figure 6.2 compares the potential energy curves generated by PBE-D3(BJ), B86bPBE-XDM, B3LYP-D3(BJ), SCS-MP2D, and DLPNO-CCSD(T). PBE and B3LYP were chosen as representative of widely-used GGA and hybrid density functionals; similar results are also obtained with other common GGA and hybrid functionals as well. The B86bPBE-XDM GGA functional is included in the gas-phase calculations here because it is used in the solid-state calculations in the next section.

Backbone twisting stabilizes all three rubrene systems, though fluorination increases the optimal extent of twisting. More significantly, the energy stabilization achieved by twisting the tetracene backbone varies considerably depending on the computational model used. For example, all five methods shown in Figure 6.2a predict a minimum near 40° for rubrene. However, whereas the three density functionals predict that the twisted form is about 12 kJ/mol more stable than the planar one, SCS-MP2D and DLPNO-CCSD(T) find the minimum to be closer to 9 kJ/mol.

For perfluororubrene, the SCS-MP2D, and DLPNO-CCSD(T) methods agree that the minimum energy occurs at 50°, whereas the curve continues to decrease with PBE-D3(BJ) and B86bPBE-XDM, and it becomes flat for B3LYP-D3(BJ). Furthermore, SCS-MP2D and DLPNO-CCSD(T) predict the minimum energy to be near 8 kJ/mol. On the other hand, the three density functionals predict the energy to be 4–6 kJ/mol lower. Similar behavior is seen for the half-fluorinated rubrene derivative in Figure 6.2b, for which the density functionals twist the backbone too much and over-stabilize the twisted form by about 5 kJ/mol.

Taking the DLPNO-CCSD(T) results as benchmark values, the density functionals exhibit root-mean-square errors of 3–4 kJ/mol across these three potential curves, compared to only 0.6 kJ/mol for SCS-MP2D. So while the SCS-MP2D calculations are moderately more computationally demanding than the DFT ones, they cost only a fraction of what the DLPNO-CCSD(T) ones cost.

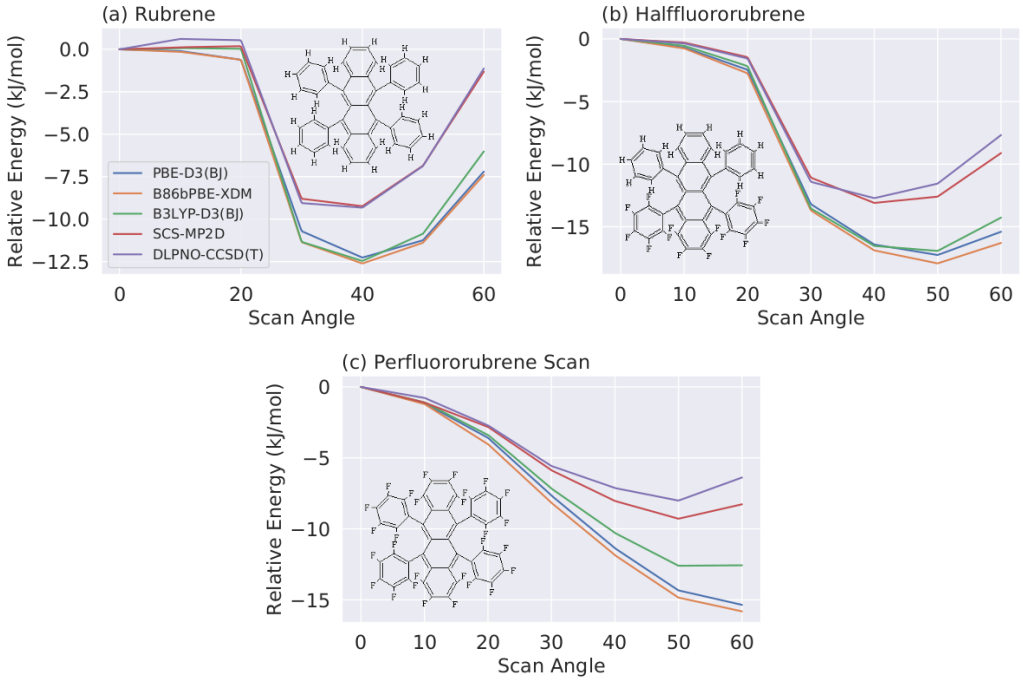
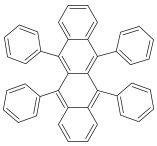
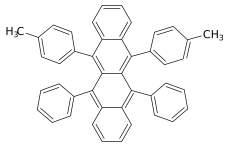


Figure 6.2: Potential energy scans from 0° to 60° for (a) rubrene, (b) half-fluorinated F_{14}-rubrene, and (c) perfluororubrene.

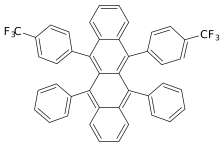
Table 6.1: Energy differences, $\Delta E = E_{twist} - E_{planar}$, between the planar and twisted forms of rubrene and four derivatives (kJ/mol), and the root-mean-square error relative to the benchmark DLPNO-CCSD(T) results.



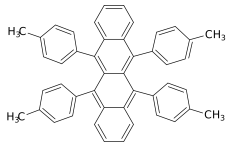
1



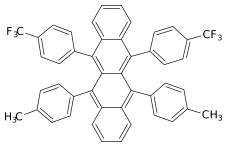
2



3



4



5

Species	PBE-D3(BJ)	B86bPBE-XDM	B3LYP-D3(BJ)	SCS-MP2D	DLPNO-CCSD(T)
1	-12.0	-12.2	-12.1	-9.4	-9.4
2	-11.9	-12.1	-11.8	-9.1	-8.9
3	-11.2	-11.0	-11.6	-8.6	-8.4
4	-11.5	-11.7	-11.2	-8.4	-8.3
5	-10.6	-10.6	-10.4	-7.5	-7.5
RMSE	3.0	3.0	2.9	0.2	

The same tendency for density functional theory models to over-stabilize the twisted tetracene backbone manifests in other rubrene derivatives. In 2015, Sutton and co-workers synthesized a series of rubrene derivatives with fluoro, methyl, or trifluromethyl groups substituted on the phenyl rings.[353] All of these species adopt a gas-phase geometry in which the tetracene backbone twists by ∼30–40°. The density functionals used here once again over-stabilize the twisted form relative to the planar one for all of these derivatives, with root-mean-square errors (RMSE) of 2.6–3.0 kJ/mol relative to DLPNO-CCSD(T) benchmarks (Table 6.1). Previous studies have also employed functionals such

as ωB97, ωB97X-D, and M06-L for these systems.[353,354] As summarized in Appendix E Table E.2, only the long-range-corrected hybrid functional ωB97X-D performs moderately better, with an RMSE of 1.5 kJ/mol. However, all of these functionals exhibit considerably larger errors than SCS-MP2D, which reproduces the DLPNO-CCSD(T) twisting energies to within a RMSE of 0.2 kJ/mol (and at fraction of the computational cost of coupled cluster theory). The next section investigates the origins of these erroneous DFT twisting energies in terms of delocalization error, while Section 6.3.3 demonstrates how these errors can impact the modeling of rubrene derivatives in the solid state.

6.3.2 Theoretical origins of the DFT twisting energy errors

The problematic twisting energies observed in Section 6.3.1 arise from delocalization error in the density functionals. The literature contains numerous examples where delocalization error over-stabilizes planar structures with extended π-conjugation relative to non-planar ones with reduced conjugation,[188,237,322,323] and one might therefore expect semi-local density functionals like those used here to artificially stabilize the planar form relative to the twisted one. However, the rubrene data above contradicts this assumption—the twisted form is artificially stabilized, not the planar one. Moreover, twisting the tetracene backbone does not significantly disrupt the π conjugation; electronic properties such as frontier orbital energies and Mulliken atomic charges change minimally upon twisting (Appendix E Section E.2).

Rubrene twisting is governed by two competing factors: Twisting the tetracene backbone is energetically unfavorable, but it allows the phenyl rings to shift apart and reduce the repulsive exchange interactions between them.[353] The tetracene backbone twisting is

the largest single source of error, accounting for about 60% of discrepancy between PBE-D3(BJ) and SCS-MP2D, and it bears the hallmarks of electron density-driven delocalization error.[374] As shown in Figure 6.3, the tetracene twisting energy varies nearly linearly with the fraction of exact exchange included in the functional. Similar linear variation is observed for the Mulliken atomic charges and frontier orbital energies (Appendix E Figure E.1). This behavior also helps explain the somewhat improved performance of the long-range corrected hybrid functional ωB97X-D noted above.

In systems plagued by density-driven delocalization error, correcting the electron density greatly improves the energies obtained from the density functionals. For example, replacing DFT orbitals with Hartree-Fock ones substantially improves the barrier heights, dissociation energies, and potential energy surfaces obtained across a variety of systems.[374–378] Here, the same orbital substitution considerably improves the rubrene twisting energetics for all three functionals shown in Figure 6.3. Taken together, this data demonstrates that density-driven delocalization error plays a significant role in the DFT twisting energy errors.

The next section shows how these errors can have a chemically significant impact on the predicted stabilities of rubrene derivatives in the solid state. Unlike these density functionals, correlated wavefunction methods like SCS-MP2D and DLPNO-CCSD(T) are not plagued by these delocalization error effects and can be used to improve the description of the solid state stabilities.

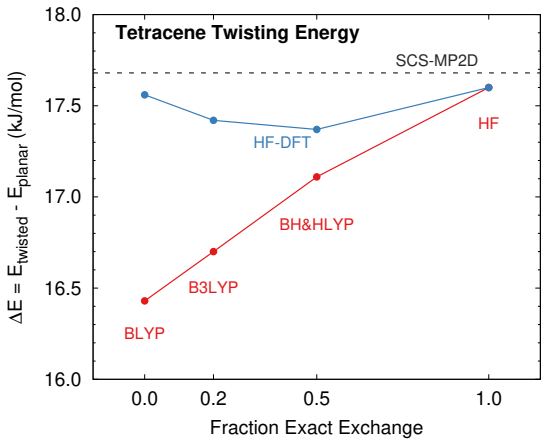


Figure 6.3: The twisting energy of the tetracene backbone varies nearly linearly with the fraction of exact exchange included in the functional. Evaluating the energy of the same functionals with HF orbitals instead of the DFT ones (HF-DFT) dramatically reduces the dependence on the fraction of exact exchange and leads to results in far better agreement with SCS-MP2D.

6.3.3 Impact on solid state polymorph stabilities

Having seen how standard density functionals artificially stabilize the twisted conformation of rubrene and its derivatives, we now examine how this behavior can lead to incorrect predictions of rubrene behavior in the solid state. Perfluororubrene can adopt multiple crystal forms, including a monoclinic polymorph with a planar rubrene backbone conformation and another with a twisted conformation (Figure 6.4a).[357,358] Although the relative thermodynamic stabilities of these two polymorphs has not been reported experimentally, we can investigate them computationally. After relaxing their experimental crystal structures with the dispersion-corrected GGA functional B86bPBE-XDM, the twisted polymorph is predicted to be 3.5 kJ/mol more stable (Figure 6.5a). However, this functional

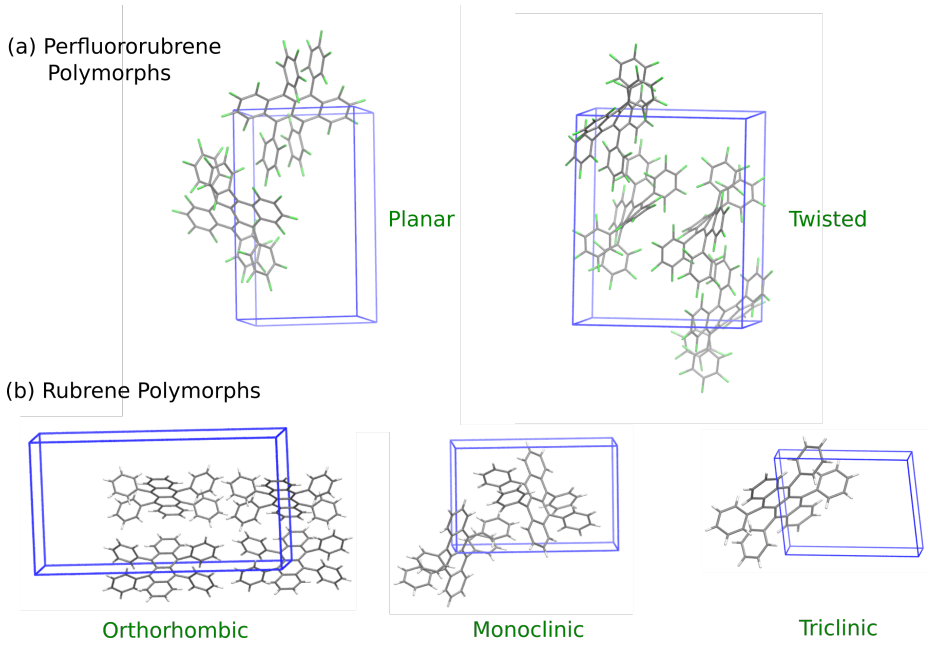


Figure 6.4: Crystal structures of the (a) perfluororubrene and (b) rubrene polymorphs studied here.

is expected to artificially stabilize the twisted polymorph due to the conformational energy errors discussed above.

Calculating the periodic perfluororubrene crystal energies with state-of-the-art density functionals or correlated wave function methods would be very computationally expensive. However, in cases where the primary deficiency of the DFT functional lies in the intramolecular conformational energy, a monomer-correction approach that replaces the DFT conformational energy with a higher-level one can dramatically improve the polymorph energies.[189] Figure 6.5a plots the relative polymorph energy differences for the two perfluororubrene polymorphs after correcting the intramolecular conformational energies. Refining the conformational energies with PBE-D3(BJ) or B3LYP-D3(BJ) has little impact

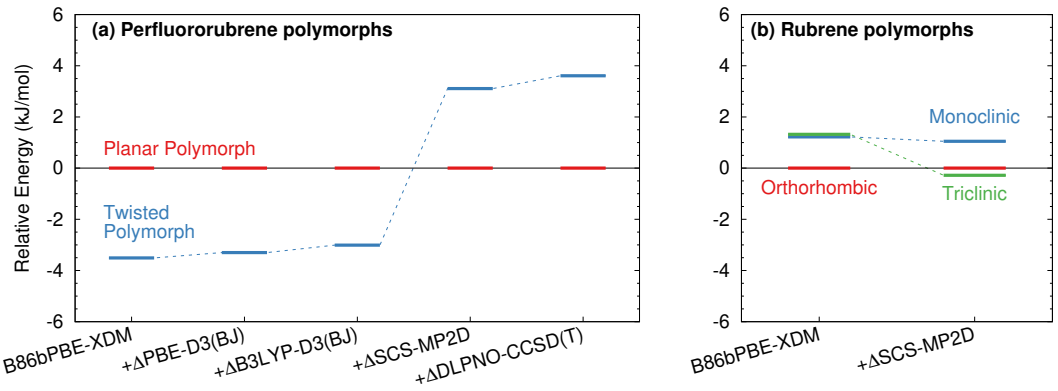


Figure 6.5: Relative stabilities for the polymorphs of (a) perfluororubrene and (b) rubrene as computed with periodic B86bPBE-XDM and after correcting the monomer conformational energy with the method indicated ("Δ methods").

on the polymorph energies, since they also over-stabilize the twisted conformation. On the other hand, refining the conformational energies with SCS-MP2D or DLPNO-CCSD(T) completely reverses the polymorph stabilities, shifting the the planar form from being 3.5 kJ/mol less stable than the twisted one to being 3.1–3.6 kJ/mol more stable. This example highlights the chemically important impact of these DFT errors—incorrectly predicting the thermodynamically stable crystal form by a substantial margin.

In contrast to the perfluororubrene polymorphs, one expects much better performance for common density functionals when ranking among a set of purely planar polymorphs or within a set of purely twisted ones. For example, the orthorhombic, monoclinic, and triclinic polymorphs of rubrene all contain planar tetracene backbones (Figure 6.4b).[345,346] As shown in Figure 6.5b, the monomer correction shifts the B86bPBE-XDM relative energies of the triclinic and monoclinic forms by only 1.6 kJ/mol and 0.2 kJ/mol, respectively, compared to the ∼7 kJ/mol shift seen for the perfluororubrene polymorphs.

The monomer-corrected results do suggest that the triclinic form is 0.3 kJ/mol more stable than the commonly crystallized orthorhombic form. However, previous modeling has found that many-body dispersion and zero-point vibrational energy contributions destabilize both the monoclinic and triclinic forms relative to the orthorhombic one.[366] Including such corrections would likely eliminate the slight nominal preference for the triclinic form seen here and restore the orthorhombic form as the thermodynamically preferred one. In other words, the relative DFT stabilities are probably far more reasonable for these three planar polymorphs of rubrene than they are for comparing the planar and twisted forms of perfluororubrene.

6.4 Conclusions

Overall, these results highlight how standard GGA and hybrid density functionals over-stabilize twisted forms of rubrene and its derivatives, and this error can lead to serious problems when trying to understand rubrene behavior in the solid state. In cases like perfluororubrene, the functionals incorrectly suggest that the twisted structure is more stable, when it appears that the planar rubrene polymorph is actually preferred. This error is a manifestation of density-driven delocalization error which needs to be addressed in order to model these species correctly. These limitations can be overcome by using higher-level electronic structure methods like SCS-MP2D or DLPNO-CCSD(T). State-of-the-art double-hybrid density functionals[83,164] might also perform reasonably as well.

More generally, a number of previous studies have synthesized derivatives and tested to see whether they crystallize in planar or twisted forms. In many of those cases,

however, little effort was spent screening for alternative polymorphs. Conformational polymorphs occur when adopting a strained intramolecular conformation is compensated for by the new favorable intermolecular interactions it enables. This conformational energy difference is typically ∼10 kJ/mol or less unless: one survey of 55 organic conformational polymorph pairs found that almost a quarter of them exhibited intramolecular conformational energy differences between 6–10 kJ/mol.[222] Another survey found that around 20% exhibited conformations in the solid state that lie 4.5–10.5 kJ/mol above the optimal gas-phase conformation.[12] Conformational energy differences exceeding ∼10 kJ/mol typically occur under specific circumstances, such as when switching from intra- to intermolecular hydrogen bonding[12] or when dramatically increasing the effective molecular surface area (e.g. changing from a folded to unfolded conformation).[382] Neither of those scenarios is applicable to rubrene.

If the conformational energy differences between twisted and planar were in the ∼10–14 kJ/mol (or more) range as previously believed,[353,354,358,381] conformational polymorphism would be somewhat less likely. However, the fact that the energy difference between the twisted and planar conformations of rubrene species is actually only ∼8–10 kJ/mol considerably increases the likelihood that new favorable intermolecular packing interactions could overcome the gas-phase preference for twisted rubrene-like molecules. In other words, it may be worthwhile to search for planar crystal polymorphs of rubrene derivatives that have previously only been observed to crystallize in twisted conformations, in the hopes of finding improved crystal packing motifs for species that exhibit promising intrinsic electronic properties.

Chapter 7

Conclusions

In conclusion, dispersion-corrected MP2 has proven to be a computationally tractable and highly accurate strategy for obtaining energies of small organic molecules. MP2D shows marked improvements over many commonly used density functionals on benchmarks of noncovalent interactions. Unlike the similar MP2C method, MP2D corrects both the intra- and intermolecular dispersion and proved to be differentiable allowing us to obtain computationally efficient energy gradients. MP2D provides highly accurate structures for 10 conformations of benzene dimers, and [7]helicene.

MP2D can be improved upon by including spin-component scaling of the same-spin and opposite-spin correlation energies. Incorporating spin-component scaling yields high accuracy for noncovalent interactions, conformational energies, thermochemistry reactions, charge transfer reactions, and energetically nuanced potential energy surfaces. SCS-MP2D rivals the best hybrid and double hybrid density functionals with similar cost, providing a

much needed wavefunction method alternative to DFT. Along with MP2D, SCS-MP2D can be used to accurately predict the energy differences between molecular crystal polymorphs.

Possible strategies for improving dispersion-corrected MP2 include incorporating Grimme's D4 dispersion correction to better describe ionic systems, and expanding the dataset used to train the model parameters, strategies that have been used to improve the performance and transferability of double hybrid functionals. Another possibility would be to replace the semi-empirical dispersion correction with an *ab initio* correction like MBD or XDM. This would increase the computation time, but would allow the dispersion correction to correctly adapt to different electronic environments. Dispersion-corrected MP2 has the potential to improve the training of tailor-made force-fields[386] and for geometry optimizations of small organic molecules. Both MP2D and SCS-MP2D were fit to reference data with basis sets extrapolated to the CBS limit. For this reason, care must be taken when using them in conjunction with smaller basis sets. For force-field parameterization and geometry optimization a smaller triple-zeta basis set is often used for the sake of computational efficiency. Implementations of MP2D or SCS-MP2D specifically fit to smaller triple-zeta basis set data could offer excellent performance for force-field parameterization and geometry optimizations. However, these methods would be more empirical in nature because the model parameters would be fit to compensate for slow convergence of MP2 with respect to the basis set size.

Critical to the success of the HMBI model is the ability to control the level of theory with which the different interactions are treated. As electronic structure methods, force-fields, and periodic codes improve, they can be incorporated within HMBI to advance

the theoretical study of molecular crystals. By combining MP2D for monomer and close-range dimer energies with periodic Hartree-Fock for long-range dimers and the many-body contributions, it was possible to accurately calculate the energy differences between polymorphs of three challenging systems. In the case of ortho-acetamidobenzamide and ROY, energy rankings that are in good agreement with experiment were obtained for the first time. For oxalyl dihydrazide MP2D/HMBI served a dual purpose. First, MP2D/HMBI provided a satisfactory ranking of the polymorphs with respect to the experimentally determined order. Second, HMBI revealed the extent to which periodic DFT methods benefit from error cancellation. This insight reveals that caution should be used when considering polymorphic energy rankings obtained with DFT.

HMBI offers the ability to apply wavefunction methods to the study of molecular crystals. However, high-level QM methods can quickly become infeasible for the dimer calculations needed in a full HMBI treatment. We discovered that often times delocalization error manifests itself by overstabilizing and thereby erroneously giving preference to certain monomer conformations. This error leads to overstabilization of planar molecular conformations in pharmaceuticals, and preference for the twisted conformations of rubrene and rubrene derivatives. Replacing the DFT monomer energies with MP2D and SCS-MP2D yielded polymorph energy rankings in excellent agreement with experiment for ortho-acetamidobenzamide, ROY, molecule X, galunisertib, axitinib, and perfluororubrene.

Although monomer correction works well for systems where monomer conformations are ill-described by DFT, two important issues must still be addressed. First, for systems like oxalyl dihydrazide the monomer correction method does not work. DFT works

by fortuitous error cancellation in this system, and applying the monomer correction disrupts this. Plane-wave hybrid functionals may be better at describing intermolecular interactions, in which case a monomer correction could then be applied successfully. Second, the monomer correction method has not been shown to work with finite temperature effects. Most polymorph energy rankings are obtained at 0 K, but phonon vibrational frequencies are sometimes necessary in order to obtain the correct relationship between polymorphs. It remains to be seen whether including these effects will further improve monomer corrected energy rankings, or if their inclusion will disrupt fortuitous error cancellation, thereby yielding inaccurate energy rankings.

Currently, the monomer correction method is being applied to full CSP landscapes. Past CSP studies on ROY have succeeded at predicting many of the experimentally verified forms, but the landscapes are plagued by unobserved structures, and inconsistent energy ordering of the known forms. Figure 7.1 shows preliminary results obtained by applying SCS-MP2D monomer corrections to the 50 lowest energy structures and known experimental structures from a ROY CSP landscape obtained with the largely force-field based CrystalOptimizer code.[387] Although the CrystalOptimizer landscape in panel (a) identifies nearly all of the experimentally verified structures, SCS-MP2D improves the ordering of the known forms while dramatically reducing the number of unobserved structures in the low energy regime. There are at least 12 known polymorphs of ROY, but the experimentally determined stability ordering is known for 8 forms: Y > YT04 > R > OP > ON > YN > ORP. Notable problems with the CrystalOptimizer landscape include overstabilization of the R form with respect to Y by 2 kJ/mol, the known third most stable polymorph, YT04,

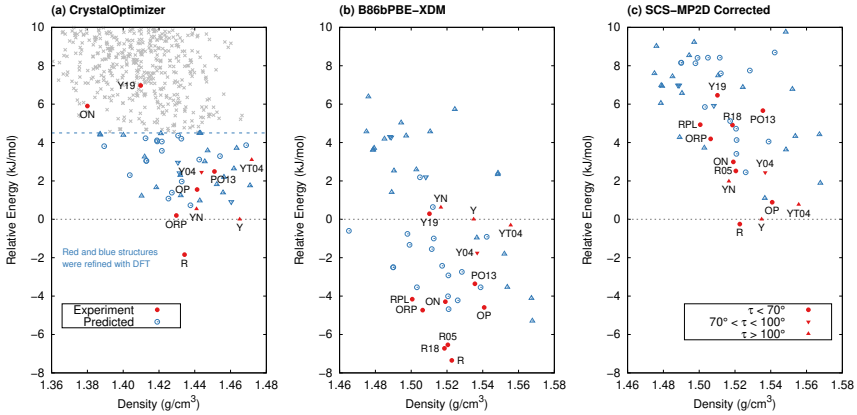


Figure 7.1: (a) ROY CSP landscape generated by a structure search using the CrystalOptimizer code. (b) Refined landscape obtained by optimizing the 50 lowest energy and all known ROY forms with the B86bPBE-XDM functional. (c) SCS-MP2D CSP landscape obtained by monomer correcting the B86bPBE-XDM energies.

is ranked as the 29th most stable structure, and within the top 10 most stable structures, 6 of them have never been verified experimentally. SCS-MP2D improves the CrystalOptimizer landscape in several ways: (1) the 4 lowest energy structures correspond to the 4 lowest energy experimental structures, (2) the energy difference between R and Y is reduced to 0.2 kJ/mol, and (3) 8 of the 11 lowest energy structures correspond to experimentally determined structures.

The SCS-MP2D landscape is not perfect. First, the R form should be less stable than Y and YT04, but is found to be slightly more stable. However, finite temperature effects have not been included, so it is possible that these effects are necessary. Second, SCS-MP2D stabilizes the 49th structure from CrystalOptimizer and pushes it into the top 10 most stable strucutres. The CrystalOptimizer landscape includes over 1000 structures,

so it is possible that more structures could be artificially stabilized by SCS-MP2D. Also, the monomer correction is considerably more computationally efficient than HMBI, but it would still be a considerable effort to perform a monomer correction on 1000 structures. The most computationally efficient approach is directly correcting the CrystalOptimzer landscape with SCS-MP2D. However, preliminary results suggest this approach does not meaninfully change the CrystalOptimizer landscape the way monomer correction changes refined DFT landscapes. It would also be considerably less expensive to perform monomer corrections on DFT single point energies obtained directly from the CrystalOptimizer structures. This would still require 1000 DFT calculations, but single point calcualtions are considerably faster than full geometry optimizations. Two possible solutions exist. First, we are currently training a tailor-made force-field, trained on the structures refined from the SCS-MP2D landscape. This force-field will be used to produce a new ROY CSP landscape. We hypothesize that the tailor-made forcefield will yield a new ROY CSP landscape that can be monomer corrected directly with SCS-MP2D. Second, if the tailor-made landscape is not ammenable to direct correction, we are hopeful that the single point energy correction will be successful. Alternatively, single point energy monomer corrections do seem capable of identifying the structures that are most sensitive to the monomer correction. Pre-screening with single-point corrections may substantially reduce the number of structures that must be refined by optimization. Regardless, even in its current implementation the monomer correction greatly improves upon pure DFT for polymorph rankings, and the ROY CSP presented in this conclusion is closer to the experimental reality than any obtained before. Monomer correction combined with high-level QM models like SCS-MP2D is very promising

for producing highly accurate crystal structure prediction landscapes, thereby improving the ability of chemists and material scientists to predict the three-dimensional form of matter from only the two-dimensional molecular representation.

Bibliography

[1] M. J. Fuchter, M. Weimar, X. Yang, D. K. Judge, and A. J. White, Tetrahedron Lett. **53**, 1108 (2012).

[2] L. A. Burns, J. C. Faver, Z. Zheng, M. S. Marshall, D. G. Smith, K. Vanommeslaeghe, A. D. MacKerell, K. M. Merz, and C. D. Sherrill, J. Chem. Phys. **147**, (2017).

[3] L. Yu, Acc. Chem. Res. **43**, 1257 (2010).

[4] O. Bludský, M. Rubeš, P. Soldán, and P. Nachtigall, J. Chem. Phys. **128**, 114102 (2008).

[5] J. Řezáč, C. Greenwell, and G. J. O. Beran, J. Chem. Theory Comput. **14**, 4711 (2018).

[6] L. Yu, G. A. Stephenson, C. A. Mitchell, C. A. Bunnell, S. V. Snorek, J. J. Bowyer, T. B. Borchardt, J. G. Stowell, and S. R. Byrn, J. Am. Chem. Soc. **122**, 585 (2000).

[7] S. Chen, I. A. Guzei, and L. Yu, J. Am. Chem. Soc. **127**, 9881 (2005).

[8] S. Chen, H. Xi, and L. Yu, J. Am. Chem. Soc. **127**, 17439 (2005).

[9] L. Yu, J. Huang, and K. J. Jones, J. Phys. Chem. B **109**, 19915 (2005).

[10] J. Bernstein, **11**, 632 (2011).

[11] A. J. Cruz-Cabeza, S. M. Reutzel-Edens, and J. Bernstein, Chem. Soc. Rev. **44**, 8619 (2015).

[12] A. J. Cruz-Cabeza and J. Bernstein, Chem. Rev. **114**, 2170 (2014).

[13] R. M. Bhardwaj, J. A. McMahon, J. Nyman, L. S. Price, S. Konar, I. D. Oswald, C. R. Pulham, S. L. Price, and S. M. Reutzel-Edens, J. Am. Chem. Soc. **141**, 13887 (2019).

[14] J. Bauer, S. Spanton, R. Henry, J. Quick, W. Dziki, W. Porter, and J. Morris, Technical Report No. 6 (unpublished).

[15] I. B. Rietveld and R. Céolin, J. Pharm. Sci. **104**, 4117 (2015).

[16] S. Bergantin and M. Moret, (2012).

[17] C. C. Mattheus, A. B. Dros, J. Baas, G. T. Oostergetel, A. Meetsma, J. L. De Boer, and T. T. Palstra, Synth. Met. **138**, 475 (2003).

[18] G. Liu, R. Gou, H. Li, and C. Zhang, (2018).

[19] P. Sanphui, N. R. Goud, U. B. Khandavilli, S. Bhanoth, and A. Nangia, Chem. Commun. **47**, 5013 (2011).

[20] J. Yang, X. Zhu, C. T. Hu, M. Qiu, Q. Zhu, M. D. Ward, and B. Kahr, (2019).

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

[22] S. R. Chemburkar, J. Bauer, K. Deming, H. Spiwek, K. Patel, J. Morris, R. Henry, S. Spanton, W. Dziki, W. Porter, J. Quick, P. Bauer, J. Donaubauer, B. A. Narayanan, M. Soldani, D. Riley, and K. Mcfarland, Org. Proc. Res. Dev. **4**, 413 (2000).

[23] D.-K. Bučar, R. W. Lancaster, and J. Bernstein, Angew. Chem. Int. Ed. **54**, 6972 (2015).

[24] I. Partes, R. Process, and M. Overtake, Technical Report No. September 2012 (unpublished).

[25] A. Kapczynski, C. Park, and B. Sampat, .

[26] R. Tandon, N. Tandon, and R. Kumar Thapar, .

[27] A. Konakanchi and Us008877932b2, Technical report (unpublished).

[28] V. López-Mejías, J. W. Kampf, and A. J. Matzger, J. Am. Chem. Soc. **134**, 9872 (2012).

[29] V. N. Serezhkin and A. V. Savchenkov, Cite This Cryst. Growth Des **20**, (2020).

[30] X. Li, X. Ou, B. Wang, H. Rong, B. Wang, and M. Lu, Commun. Chem. 1 .

[31] A. Lévesque, T. Maris, and J. D. Wuest, J. Am. Chem. Soc. **142**, 11873 (2020).

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

[33] D. A. Bardwell, C. S. Adjiman, Y. A. Arnautova, E. Bartashevich, S. X. M. Boerrigter, D. E. Braun, A. J. Cruz-Cabeza, G. M. Day, R. G. Della Valle, G. R. Desiraju, B. P. van Eijck, J. C. Facelli, M. B. Ferraro, D. Grillo, M. Habgood, D. W. M. Hofmann, F. Hofmann, K. V. J. Jose, P. G. Karamertzanis, A. V. Kazantsev, J. Kendrick, L. N. Kuleshova, F. J. J. Leusen, A. V. Maleev, A. J. Misquitta, S. Mohamed, R. J. Needs,

M. A. Neumann, D. Nikylov, A. M. Orendt, R. Pal, C. C. Pantelides, C. J. Pickard, L. S. Price, S. L. Price, H. A. Scheraga, J. van de Streek, T. S. Thakur, S. Tiwari, E. Venuti, I. K. Zhitkov, and IUCr, Acta Crystallogr. Sect. B Struct. Sci. **67**, 535 (2011).

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

[35] M. Vasileiadis, C. C. Pantelides, and C. S. Adjiman, Chem. Eng. Sci. **121**, 60 (2015).

[36] R. M. Bhardwaj, L. S. Price, S. L. Price, S. M. Reutzel-Edens, G. J. Miller, I. D. H. Oswald, B. F. Johnston, and A. J. Florence, (2013).

[37] M. Vasileiadis, A. V. Kazantsev, P. G. Karamertzanis, C. S. Adjiman, C. C. Pantelides, and IUCr, Acta Crystallogr. Sect. B Struct. Sci. **68**, 677 (2012).

[38] P. G. Karamertzanis and C. C. Pantelides, Mol. Phys. **105**, 273 (2007).

[39] M. A. Neumann, F. J. J. Leusen, and J. Kendrick, Angew. Chemie **120**, 2461 (2008).

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

[41] C. J. Pickard and R. J. Needs, J. Phys. Condens. Matter **23**, (2011).

[42] V. L. Deringer, D. M. Proserpio, G. . Abor Csányi, C. Csányi, and C. J. Pickard, (2018).

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

[44] N. F. Francia, L. S. Price, J. Nyman, S. L. Price, and M. Salvalaglio, Cryst. Growth Des. (2020).

[45] M. Habgood, I. J. Sugden, A. V. Kazantsev, C. S. Adjiman, and C. C. Pantelides, (2015).

[46] M. L. Sarah L. Price, L. S. P. Gareth W. A. Welch, Matthew Habgood, P. G. K. Day, and G. M., Phys. Chem. Chem. Phys (2010).

[47] B. P. V. A. N. Eijck, W. T. M. Mooij, and J. A. N. Kroon, **22**, 805 (2001).

[48] A. J. Misquitta, R. Podeszwa, B. Jeziorski, and K. Szalewicz, J. Chem. Phys. **123**, (2005).

[49] O. Egorova, (2020).

[50] D. McDonagh, C. K. Skylaris, and G. M. Day, J. Chem. Theory Comput. **15**, 2743 (2019).

[51] S. L. Price, Acta Crystallogr. Sect. B Struct. Sci. Cryst. Eng. Mater. **69**, 313 (2013).

[52] R. Montis, R. J. Davey, S. E. Wright, G. R. Woollam, and A. J. Cruz-Cabeza, .

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

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

[55] C. Greenwell, J. L. McKinley, P. Zhang, Q. Zeng, G. Sun, B. Li, S. Wen, and G. J. O. Beran, Chemical Science **1**, 1 (2020).

[56] C. Greenwell, J. L. Mckinley, P. Zhang, Q. Zeng, G. Sun, B. Li, S. Wen, and G. J. O. Beran, (2020).

[57] L. M. LeBlanc, S. G. Dale, C. R. Taylor, A. D. Becke, G. M. Day, and E. R. Johnson, Angew. Chemie **130**, 15122 (2018).

[58] A. Otero-De-La-Roza, L. M. Leblanc, and E. R. Johnson, (2019).

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

[60] K. Autumn, M. Sitti, Y. A. Liang, A. M. Peattie, W. R. Hansen, S. Sponberg, T. W. Kenny, R. Fearing, J. N. Israelachvili, and R. J. Full, Technical Report No. 19 (unpublished).

[61] N. O. A. Szabo, Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory - Attila Szabo, Neil S. Ostlund - Google Books.

[62] S. Grimme, J. Chem. PhysicsPu J. Chem. Phys. J. Chem. Phys. J. Chem. Phys. J. Chem. Phys. J. Chem. Phys. J. Chem. Phys. **1181**, 154104 (2003).

[63] R. A. Distasio and M. Head-Gordon, Mol. Phys. **105**, 1073 (2007).

[64] J. Řezáč and P. Hobza, Chem. Rev. **116**, 5038 (2016).

[65] J. F. Stanton, Chem. Phys. Lett. **281**, 130 (1997).

[66] S. Jan and M. L. Martin, Phys. Chem. Chem. Phys **18**, 20905 (2016).

[67] L. Goerigk, A. Hansen, C. Bauer, S. Ehrlich, A. Najibi, and S. Grimme, Phys. Chem. Chem. Phys. **19**, 32184 (2017).

[68] C. Riplinger and F. Neese, J. Chem. Phys. **138**, (2013).

[69] C. Riplinger, B. Sandhoefer, A. Hansen, and F. Neese, J. Chem. Phys. **139**, (2013).

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

[71] F. Jensen, Introduction to Computational Chemistry - Frank Jensen - Google Books.

[72] S. Grimme, J. Comput. Chem. **25**, 1463 (2004).

[73] S. Grimme, J. Comput. Chem. **27**, 1787 (2006).

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

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

[76] A. Otero-De-La-Roza and E. R. Johnson, J. Chem. Phys. **136**, (2012).

[77] A. Tkatchenko, R. A. Distasio, R. Car, and M. Scheffler, Phys. Rev. Lett. **108**, (2012).

[78] O. A. Vydrov and T. Van Voorhis, J. Chem. Phys **133**, 244103 (2010).

[79] K. Berland, V. R. Cooper, K. Lee, E. Schröder, T. Thonhauser, P. Hyldgaard, and B. I. Lundqvist, Reports Prog. Phys. **78**, (2015).

[80] A. D. Becke, J. Chem. Phys. **98**, 1372 (1993).

[81] N. Mardirossian and M. Head-Gordon, Phys. Chem. Chem. Phys **16**, 9904 (2014).

[82] N. Mardirossian and M. Head-Gordon, J. Chem. Phys. **144**, 214110 (2016).

[83] G. Santra, N. Sylvetsky, and J. M. L. Martin, J. Phys. Chem. A **123**, 5129 (2019).

[84] J. M. L. Martin and G. Santra, (2020).

[85] N. Mardirossian and M. Head-Gordon, J. Chem. Phys **148**, 241736 (2018).

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

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

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

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

[90] G. Beran, Chem. Rev. **116**, (2016).

[91] J. E. Subotnik, A. Sodt, and M. Head-Gordon, J. Chem. Phys. **128**, 034103 (2008).

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

[93] S. Wen and G. J. O. Beran, Cryst. Growth Des. **12**, 2169 (2012).

[94] K. D. Nanda and G. J. O. Beran, **174106**, 1 (2012).

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

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

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

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

[99] W. Sontising, Y. N. Heit, J. McKinley, and G. J. O. Beran, Chem. Sci. 1 (2017).

[100] Y. Heit, K. Nanda, and G. Beran, Chem. Sci. **7**, 246 (2015).

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

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

[103] C. Červinka and G. J. O. Beran, Chem. Sci. 4622 (2018).

[104] W. Sontising and G. J. O. Beran, Phys. Rev. Mater. **3**, 095002 (2019).

[105] W. Sontising and G. J. O. Beran, Phys. Rev. Mater **4**, 063601 (2020).

[106] B. D. Nguyen, G. P. Chen, M. M. Agee, M. Burow, M. P. Tang, and F. Furche, Cite This J. Chem. Theory Comput **16**, (2020).

[107] X. Wang, T. Garcia, S. Monaco, B. Schatschneider, and N. Marom, CrystEngComm **18**, 7353 (2016).

[108] Y. Sakamoto and T. Suzuki, (2017).

[109] Z. Zhang, W. A. Ogden, V. G. Young, and C. J. Douglas, Chem. Commun. **52**, 8127 (2016).

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

[111] A. J. Cohen, P. Mori-Sanchez, and W. Yang, Chem. Rev. **112**, 289 (2012).

[112] K. E. Riley, M. Pitonak, P. Jurecka, and P. Hobza, Chem. Rev. **110**, 5023 (2010).

[113] M. O. Sinnokrot and C. D. Sherrill, J. Phys. Chem. A **108**, 10200 (2004).

[114] S. M. Cybulski and M. L. Lytle, J. Chem. Phys. **127**, 141102 (2007).

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

[116] M. Gerenkamp and S. Grimme, Chem. Phys. Lett. **392**, 229 (2004).

[117] J. G. Hill and J. A. Platts, J. Chem. Theory Comput. **3**, 80 (2007).

[118] S. Y. Tan, L. Wylie, I. Begic, D. Tran, and E. I. Izgorodina, Phys. Chem. Chem. Phys. **19**, 28936 (2017).

[119] O. Marchetti and H. J. Werner, J. Phys. Chem. A **113**, 11580 (2009).

[120] R. T. McGibbon, A. G. Taube, A. G. Donchev, K. Siva, F. Hernández, C. Hargus, K. H. Law, J. L. Klepeis, and D. E. Shaw, J. Chem. Phys. **147**, (2017).

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

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

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

[124] A. Tkatchenko, R. A. DiStasio, M. Head-Gordon, and M. Scheffler, J. Chem. Phys. **131**, 094106 (2009).

[125] Y. Huang and G. J. O. Beran, J. Chem. Phys. **143**, 044113 (2015).

[126] G. Chalasinski, M. M. Szczesniak, and S. M. Cybulski, J. Chem. Phys. **92**, 2481 (1990).

[127] O. A. von Lilienfeld and A. Tkatchenko, J. Chem. Phys. **132**, 234109 (2010).

[128] T. Risthaus and S. Grimme, J. Chem. Theory Comput. **9**, 1580 (2013).

[129] J. Řezáč, J. Comp. Chem. **37**, 1230 (2016).

[130] MOLPRO, version 2012.1, a package of ab initio programs, H.-J. Werner, P. J. Knowles, G. Knizia, F. R. Manby, M. Schütz, P. Celani, T. Korona, R. Lindh, A. Mitrushenkov, G. Rauhut, K. R. Shamasundar, T. B. Adler, R. D. Amos, A. Bernhardsson, A. Berning, D. L. Cooper, M. J. O. Deegan, A. J. Dobbyn, F. Eckert, E. Goll, C. Hampel, A. Hesselmann, G. Hetzer, T. Hrenar, G. Jansen, C. Köppl, Y. Liu, A. W. Lloyd, R. A. Mata, A. J. May, S. J. McNicholas, W. Meyer, M. E. Mura, A. Nicklass, D. P. O'Neill, P. Palmieri, D. Peng, K. Pflüger, R. Pitzer, M. Reiher, T. Shiozaki, H. Stoll, A. J. Stone, R. Tarroni, T. Thorsteinsson, and M. Wang, see http://www.molpro.net.

[131] V. V. Gobre and A. Tkatchenko, Nature Commun. **4**, 2341 (2013).

[132] K. T. Tang and J. P. Toennies, J. Chem. Phys. **80**, 3726 (1984).

[133] L. Goerigk, H. Kruse, and S. Grimme, ChemPhysChem **12**, 3421 (2011).

[134] R. Sedlak and J. Řezáč, J. Chem. Theory Comput. **13**, 1638 (2017).

[135] J. Řezáč, K. E. Riley, and P. Hobza, J. Chem. Theory Comput. **7**, 2427 (2011).

[136] J. Řezáč, Cuby 4, software framework for computational chemistry: MP2D method, 2015.

[137] R. M. Parrish, L. A. Burns, D. G. A. Smith, A. C. Simmonett, A. E. DePrince, E. G. Hohenstein, U. Bozkaya, A. Y. Sokolov, R. Di Remigio, R. M. Richard, J. F. Gonthier, A. M. James, H. R. McAlexander, A. Kumar, M. Saitow, X. Wang, B. P. Pritchard, P. Verma, H. F. Schaefer, K. Patkowski, R. A. King, E. F. Valeev, F. A. Evangelista, J. M. Turney, T. D. Crawford, and C. D. Sherrill, J. Chem. Theory Comput. **13**, 3185 (2017).

[138] TURBOMOLE V7.2 2017, a development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, 1989-2007, TURBOMOLE GmbH, since 2007; available from
`http://www.turbomole.com`.

[139] F. Furche, R. Ahlrichs, C. Hättig, W. Klopper, M. Sierka, and F. Weigend, WIREs Comput. Mol. Sci. **4**, 91 (2014).

[140] T. H. Dunning, J. Chem. Phys. **90**, 1007 (1989).

[141] A. G. Császár, W. D. Allen, and H. F. Schaefer, J. Chem. Phys. **108**, 9751 (1998).

[142] S. Grimme, S. Ehrlich, and L. Goerigk, J. Comp. Chem. **32**, 1456 (2011).

[143] L. A. Burns, A. V. Mayagoitia, B. G. Sumpter, and C. D. Sherrill, J. Chem. Phys. **134**, 084107 (2011).

[144] L. Goerigk and S. Grimme, Phys. Chem. Chem. Phys. **13**, 6670 (2011).

[145] L. Goerigk, A. Hansen, C. Bauer, S. Ehrlich, A. Najibi, and S. Grimme, Phys. Chem. Chem. Phys. **19**, 32184 (2017).

[146] D. G. A. Smith, L. A. Burns, K. Patkowski, and C. D. Sherrill, J. Phys. Chem. Lett. **7**, 2197 (2016).

[147] J. Řezáč, K. E. Riley, and P. Hobza, J. Chem. Theory Comput. **8**, 4285 (2012).

[148] P. Jurečka, J. Šponer, J. Černý, and P. Hobza, Phys. Chem. Chem. Phys. **8**, 1985 (2006).

[149] M. S. Marshall, L. A. Burns, and C. D. Sherrill, J. Chem. Phys. **135**, 194102 (2011).

[150] D. Gruzman, A. Karton, and J. M. L. Martin, J. Phys. Chem. A **113**, 11974 (2009).

[151] J. Řezáč, D. Bím, O. Gutten, and L. Rulíšek, J. Chem. Theory Comput. **14**, 1254 (2018).

[152] S. Grimme, M. Steinmetz, and M. Korth, J. Org. Chem. **72**, 2118 (2007).

[153] E. R. Johnson, P. Mori-Sánchez, A. J. Cohen, and W. Yang, The Journal of Chemical Physics **129**, 204112 (2008).

[154] S. Kozuch, D. Gruzman, and J. M. L. Martin, J. Phys. Chem. C **114**, 20801 (2010).

[155] S. Grimme, C. Diedrich, and M. Korth, Angew. Chem. Int. Ed. **45**, 625 (2006).

[156] M. Pitonak, P. Neogrady, J. Cerny, S. Grimme, and P. Hobza, ChemPhysChem **10**, 282 (2009).

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

[158] S. D. C. Organique, U. L. D. Bruxelles, A. F. D. Roosevelt, G. Germain, U. C. D. Louvain, and P. L. Pasteur, **60**, (1977).

[159] S. Fujino, M. Yamaji, H. Okamoto, T. Mutai, I. Yoshikawa, H. Houjou, and F. Tani, Photochem. Photobiol. Sci. **16**, 925 (2017).

[160] *Multi-scale Quantum Models for Biocatalysis*, Vol. 7 of *Challenges and Advances in Computational Chemistry and Physics*, edited by D. M. York and T.-S. Lee (Springer Netherlands, Dordrecht, 2009).

[161] G. Santra and J. M. L. Martin, AIP Conf. Proc. **2186**, 030004 (2019).

[162] N. Mardirossian and M. Head-Gordon, Phys. Chem. Chem. Phys. **16**, 9904 (2014).

[163] N. Mardirossian and M. Head-Gordon, J. Chem. Phys. **144**, 214110 (2016).

[164] N. Mardirossian and M. Head-Gordon, J. Chem. Phys. **148**, 241736 (2018).

[165] S. Kozuch and J. M. L. Martin, J. Comput. Chem. **34**, 2327 (2013).

[166] C. Riplinger, B. Sandhoefer, A. Hansen, and F. Neese, J. Chem. Phys. **139**, 134101 (2013).

[167] J. S. Smith, B. T. Nebgen, R. Zubatyuk, N. Lubbers, C. Devereux, K. Barros, S. Tretiak, O. Isayev, and A. E. Roitberg, .

[168] A. J. Cohen, P. Mori-Sanchez, and W. Yang, Science **321**, 792 (2008).

[169] S. Grimme, L. Goerigk, and R. F. Fink, WIRES: Comput. Mol. Sci. **2**, 886 (2012).

[170] S. Grimme, J. Chem. Phys. **118**, 9095 (2003).

[171] Á. Szabados, J. Chem. Phys. **125**, (2006).

[172] R. F. Fink, J. Chem. Phys. **133**, 174113 (2010).

[173] R. A. DiStasio and M. Head-Gordon, Mol. Phys. **105**, 1073 (2007).

[174] J. G. Hill and J. A. Platts, J. Chem. Theory Comput. **3**, 80 (2007).

[175] R. A. King, Molecular Physics **107**, 789 (2009).

[176] J. Rigby and E. I. Izgorodina, J. Chem. Theory Comput. **10**, 3111 (2014).

[177] S. Tan, S. Barrera Acevedo, and E. I. Izgorodina, J. Chem. Phys. **146**, 064108 (2017).

[178] R. T. McGibbon, A. G. Taube, A. G. Donchev, K. Siva, F. Hernández, C. Hargus, K.-H. Law, J. L. Klepeis, and D. E. Shaw, J. Chem. Phys. **147**, 161725 (2017).

[179] Y. Jung, R. C. Lochan, A. D. Dutoi, and M. Head-Gordon, J. Chem. Phys. **121**, 9793 (2004).

[180] R. C. Lochan, Y. Jung, and M. Head-Gordon, J. Phys. Chem. A **109**, 7598 (2005).

[181] S. Grimme, J. Comp. Chem. **24**, 1529 (2003).

[182] T. Takatani, E. G. Hohenstein, and C. D. Sherrill, J. Chem. Phys. **128**, 124111 (2008).

[183] M. Pitonak, J. Rezac, and P. Hobza, Phys. Chem. Chem. Phys. **12**, 9611 (2010).

[184] S. Grimme and E. I. Izgorodina, Chem. Phys. **305**, 223 (2004).

[185] Y. M. Rhee and M. Head-Gordon, J. Phys. Chem. A **111**, 5314 (2007).

[186] D. Casanova, Y. M. Rhee, and M. Head-Gordon, J. Chem. Phys. **128**, 164106 (2008).

[187] A. Hellweg, S. A. Grün, and C. Hättig, Phys. Chem. Chem. Phys. **10**, 4119 (2008).

[188] C. Greenwell, J. L. McKinley, P. Zhang, Q. Zeng, G. Sun, B. Li, S. Wen, and G. J. O. Beran, Chem. Sci. **11**, 2200 (2020).

[189] C. Greenwell and G. J. O. Beran, Cryst. Growth Des. **20**, 4875 (2020).

[190] E. R. Johnson, P. Mori-Sánchez, A. J. Cohen, and W. Yang, J. Chem. Phys. **129**, 204112 (2008).

[191] F. Neese, WIREs Comput. Molec. Sci. **2**, 73 (2012).

[192] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, Journal of Machine Learning Research **12**, 2825 (2011).

[193] J. Řezáč, Y. Huang, P. Hobza, and G. J. O. Beran, J. Chem. Theory Comput. **11**, 2065 (2015).

[194] M. S. Marshall, L. A. Burns, and C. D. Sherrill, J. Chem. Phys. **135**, 194102 (2011).

[195] J. Řezáč, J. Chem. Theory Comput. **16**, 2355 (2020).

[196] J. Řezáč and A. de la Lande, J. Chem. Theory Comput. **11**, 528 (2015).

[197] D. Gruzman, A. Karton, and J. M. L. Martin, J. Phys. Chem. A **113**, 11974 (2009).

[198] U. R. Fogueri, S. Kozuch, A. Karton, and J. M. Martin, J. Phys. Chem. A **117**, 2269 (2013).

[199] S. Grimme, M. Steinmetz, and M. Korth, J. Org. Chem. **72**, 2118 (2007).

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

[201] K. U. Lao, R. Schaeffer, G. Jansen, and J. M. Herbert, J. Chem. Theory Comput. 150417132228001 (2015).

[202] S. P. Thomas and M. A. Spackman, Austral. J. Chem. **71**, 279 (2018).

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

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

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

[206] K. S. Gushurst, J. Nyman, and S. X. M. Boerrigter, CrystEngComm **21**, 1363 (2019).

[207] A. R. Tyler, R. Ragbirsingh, C. J. McMonagle, P. G. Waddell, S. E. Heaps, J. W. Steed, P. Thaw, M. J. Hall, and M. R. Probert, Chem **6**, 1755 (2020).

[208] A. Lévesque, T. Maris, and J. D. Wuest, J. Am. Chem. Soc. **142**, 11873 (2020).

[209] X. Li, X. Ou, H. Rong, S. Huang, J. Nyman, L. Yu, and M. Lu, Cryst. Growth Des. acs.cgd.0c01017 (2020).

[210] R. O. Al-Kaysi, A. M. Müller, and C. J. Bardeen, J. Am. Chem. Soc. **128**, 15938 (2006).

[211] L. Zhu, R. O. Al-Kaysi, and C. J. Bardeen, J. Am. Chem. Soc. **133**, 12569 (2011).

[212] S. R. Jezowski, L. Zhu, Y. Wang, A. P. Rice, G. W. Scott, C. J. Bardeen, and E. L. Chronister, J. Am. Chem. Soc. **134**, 7459 (2012).

[213] L. Zhu, F. Tong, C. Salinas, M. K. Al-Muhanna, F. S. Tham, D. Kisailus, R. O. Al-Kaysi, and C. J. Bardeen, Chem. Mater. **26**, 6007 (2014).

[214] A. D. Becke and E. R. Johnson, J. Chem. Phys. **127**, 154108 (2007).

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

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

[217] J. Bauer, S. Spanton, R. Henry, J. Quick, W. Dziki, W. Porter, and J. Morris, Pharm. Res. **18**, 859 (2001).

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

[219] D. P. Konakanchi, B. Gongalla, K. B. Sikha, C. Kandaswamy, K. Sataya, B. R. Adibhatala, and V. C. Nannapaneni., US Patent 8,877,932 B2 (2014).

[220] J. Maddox, Nature **335**, 201 (1988).

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

[222] A. J. Cruz-Cabeza, S. M. Reutzel-Edens, and J. Bernstein, Chem. Soc. Rev. **44**, 8619 (2015).

[223] A. Burger and R. Ramberger, Mikrochim. Acta **72**, 273 (1979).

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

[225] J. Hermann, R. A. DiStasio, and A. Tkatchenko, Chem. Rev. **117**, 4714 (2017).

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

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

[228] M. A. Neumann, F. J. J. Leusen, and J. Kendrick, Angew. Chem. Int. Ed. **47**, 2427 (2008).

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

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

[231] J. Kendrick, M. D. Gourlay, M. A. Neumann, and F. J. J. Leusen, CrystEngComm **11**, 2391 (2009).

[232] H. C. S. Chan, J. Kendrick, M. A. Neumann, and F. J. J. Leusen, CrystEngComm **15**, 3799 (2013).

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

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

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

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

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

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

[239] J. Yang, C. T. Hu, X. Zhu, Q. Zhu, M. D. Ward, and B. Kahr, Angew. Chem. Int. Ed. **56**, 10165 (2017).

[240] P. Zhang, G. P. F. Wood, J. Ma, M. Yang, Y. Liu, G. Sun, Y. A. Jiang, B. C. Hancock, and S. Wen, Cryst. Growth Des. **18**, 6891 (2018).

[241] J. Hoja, H.-Y. Ko, M. A. Neumann, R. Car, R. A. DiStasio, and A. Tkatchenko, Science Adv. **5**, eaau3338 (2019).

[242] E. Schur, J. Bernstein, L. S. Price, R. Guo, S. L. Price, S. H. Lapidus, and P. W. Stephens, Cryst. Growth Des. **19**, 4884 (2019).

[243] L. M. LeBlanc and E. R. Johnson, CrystEngComm **21**, 5995 (2019).

[244] M.-A. Perrin, M. A. Neumann, H. Elmaleh, and L. Zaske, Chem. Commun. 3181 (2009).

[245] J. Kendrick, G. A. Stephenson, M. A. Neumann, and F. J. J. Leusen, Cryst. Growth Des. **13**, 581 (2013).

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

[247] M. A. Neumann, J. van de Streek, F. P. A. Fabbiani, P. Hidber, and O. Grassmann, Nature Commun. **6**, 7793 (2015).

[248] D. E. Braun, S. R. Lingireddy, M. D. Beidelschies, R. Guo, P. Müller, S. L. Price, and S. M. Reutzel-Edens, Cryst. Growth Des. **17**, 5349 (2017).

[249] G. R. Woollam, M. A. Neumann, T. Wagner, and R. J. Davey, Faraday Disc. **211**, 209 (2018).

[250] D. E. Braun, J. A. McMahon, R. M. Bhardwaj, J. Nyman, M. A. Neumann, J. Van De Streek, and S. M. Reutzel-Edens, Cryst. Growth Des. **19**, 2947 (2019).

[251] R. M. Bhardwaj, J. A. McMahon, J. Nyman, L. S. Price, S. Konar, I. D. H. Oswald, C. R. Pulham, S. L. Price, and S. M. Reutzel-Edens, J. Am. Chem. Soc. **141**, 13887 (2019).

[252] M. Mortazavi, J. Hoja, L. Aerts, L. Quéré, J. van de Streek, M. A. Neumann, and A. Tkatchenko, Commun. Chem. **2**, 80 (2019).

[253] P. G. Karamertzanis, G. M. Day, G. W. A. Welch, J. Kendrick, F. J. J. Leusen, M. A. Neumann, and S. L. Price, J. Chem. Phys. **128**, 244708 (2008).

[254] P. J. Bygrave, D. H. Case, and G. M. Day, Faraday Disc. **170**, 41 (2014).

[255] M. J. Bryant, S. N. Black, H. Blade, R. Docherty, A. G. Maloney, and S. C. Taylor, J. Pharm. Sci. **108**, 1655 (2019).

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

[257] A. Otero-de-la Roza, L. M. LeBlanc, and E. R. Johnson, J. Chem. Theory Comput. **15**, 4933 (2019).

[258] A. M. Reilly and A. Tkatchenko, J. Chem. Phys. **139**, 024705 (2013).

[259] O. A. Loboda, G. A. Dolgonos, and A. D. Boese, J. Chem. Phys. **149**, 124104 (2018).

[260] L. Maschio, D. Usvyat, F. R. Manby, S. Casassa, C. Pisani, and M. Schütz, Phys. Rev. B **76**, 075101 (2007).

[261] D. Usvyat, L. Maschio, F. R. Manby, S. Casassa, C. Pisani, and M. Schütz, Phys. Rev. B **76**, 075102 (2007).

[262] L. Maschio, D. Usvyat, and B. Civalleri, CrystEngComm **12**, 2429 (2010).

[263] L. Maschio, J. Chem. Theory Comput. **7**, 2818 (2011).

[264] D. Presti, A. Pedone, M. C. Menziani, B. Civalleri, and L. Maschio, CrystEngComm **16**, 102 (2014).

[265] M. Marsman, A. Grüneis, J. Paier, and G. Kresse, J. Chem. Phys. **130**, 184103 (2009).

[266] A. Grüneis, M. Marsman, and G. Kresse, J. Chem. Phys. **133**, 074107 (2010).

[267] S. Hirata and T. Shimazaki, Phys. Rev. B **80**, 1 (2009).

[268] T. Shiozaki and S. Hirata, J. Chem. Phys. **132**, 151101 (2010).

[269] M. Katouda and S. Nagase, J. Chem. Phys. **133**, 184103 (2010).

[270] Y.-Y. Ohnishi and S. Hirata, J. Chem. Phys. **133**, 034106 (2010).

[271] D. Usvyat, J. Chem. Phys. **139**, 194101 (2013).

[272] M. Del Ben, J. Hutter, and J. Vandevondele, J. Chem. Theory Comput. **8**, 4177 (2012).

[273] M. Del Ben, J. Hutter, and J. VandeVondele, J. Chem. Phys. **143**, 102803 (2015).

[274] M. Del Ben, J. VandeVondele, and B. Slater, J. Phys. Chem. Lett. **5**, 4122 (2014).

[275] D. Lu, Y. Li, D. Rocca, and G. Galli, Phys. Rev. Lett. **102**, 1 (2009).

[276] M. Macher, J. Klimeš, C. Franchini, and G. Kresse, J. Chem. Phys. **140**, 084502 (2014).

[277] J. Klimeš, J. Chem. Phys. **145**, 094506 (2016).

[278] K. Hongo, M. A. Watson, R. S. Sanchez-Carrera, T. Iitaka, and A. Aspuru-Guzik, J. Phys. Chem. Lett. **1**, 1789 (2010).

[279] K. Hongo, M. a. Watson, T. Iitaka, A. Aspuru-Guzik, and R. Maezono, J. Chem. Theory Comput. **11**, 907 (2015).

[280] G. H. Booth, A. Grüneis, G. Kresse, and A. Alavi, Nature **493**, 365 (2013).

[281] A. Zen, J. G. Brandenburg, J. Klimeš, A. Tkatchenko, D. Alfè, and A. Michaelides, Proc. Nat. Acad. **0**, 201715434 (2018).

[282] B. Paulus, Phys. Rep. **428**, 1 (2006).

[283] S. Wen, K. Nanda, Y. Huang, and G. J. O. Beran, Phys. Chem. Chem. Phys. **14**, 7578 (2012).

[284] G. J. O. Beran, S. Wen, K. Nanda, Y. Huang, and Y. Heit, Top. Curr. Chem. **345**, 59 (2014).

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

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

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

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

[289] M. O. Sinnokrot and C. D. Sherrill, J. Phys. Chem. A **110**, 10656 (2006).

[290] S. Ahn, F. Guo, B. M. Kariuki, and K. D. M. Harris, J. Am. Chem. Soc. **128**, 8441 (2006).

[291] W. T. M. Mooij, B. P. van Eijck, and J. Kroon, J. Am. Chem. Soc. **122**, 3500 (2000).

[292] B. P. van Eijck, W. T. M. Mooij, and J. Kroon, J. Comp. Chem. **22**, 805 (2001).

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

[294] J. L. McKinley and G. J. O. Beran, J. Chem. Theory Comput. **15**, 5259 (2019).

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

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

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

[298] H. Y. Ko, R. A. Distasio, B. Santra, and R. Car, Phys. Rev. Mater. **2**, 1 (2018).

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

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

[301] C. Červinka and M. Fulem, Phys. Chem. Chem. Phys. **21**, 18501 (2019).

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

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

[304] E. C. Dybeck, N. S. Abraham, N. P. Schieber, and M. R. Shirts, Cryst. Growth Des. **17**, 1775 (2017).

[305] E. C. Dybeck, D. P. McMahon, G. M. Day, and M. R. Shirts, Cryst. Growth Des. **19**, 5568 (2019).

[306] N. S. Abraham and M. R. Shirts, Cryst. Growth Des. **19**, 6911 (2019).

[307] L. A. Errede, M. C. Etter, R. C. Williams, and S. M. Darnauer, J. Chem. Soc. Perkin Trans. **2**, 233 (1981).

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

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

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

[311] G. J. O. Beran, J. Chem. Phys. **130**, 164115 (2009).

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

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

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

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

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

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

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

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

[320] R. Dovesi, A. Erba, R. Orlando, C. M. Zicovich-Wilson, B. Civalleri, L. Maschio, M. Rérat, S. Casassa, J. Baima, S. Salustro, and B. Kirtman, WIREs Comput. Mol. Sci. **8**, e1360 (2018).

[321] D. Vilela Oliveira, J. Laun, M. F. Peintinger, and T. Bredow, J. Comp. Chem. **40**, 2364 (2019).

[322] H. L. Woodcock, H. F. Schaefer, and P. R. Schreiner, J. Phys. Chem. A **106**, 11923 (2002).

[323] T. Heaton-Burgess and W. Yang, J. Chem. Phys. **132**, 234113 (2010).

[324] X. Tan, K. Wang, T. Yan, X. Li, J. Liu, K. Yang, B. Liu, G. Zou, and B. Zou, J. Phys. Chem. C **119**, 10178 (2015).

[325] A. Burger and R. Ramberger, Mikrochim. Acta **72**, 259 (1979).

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

[327] A. Asmadi, M. A. Neumann, J. Kendrick, P. Girard, M. A. Perrin, and F. J. J. Leusen, J. Phys. Chem. B **113**, 16303 (2009).

[328] J. van de Streek, M. A. Neumann, and M.-A. Perrin, CrystEngComm **12**, 3827 (2010).

[329] D. E. Braun, R. M. Bhardwaj, A. J. Florence, D. A. Tocher, and S. L. Price, Cryst. Growth Des. **13**, 19 (2013).

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

[331] M. Vasileiadis, A. V. Kazantsev, P. G. Karamertzanis, C. S. Adjiman, and C. C. Pantelides, Acta Cryst. B **68**, 677 (2012).

[332] M. Dračínský, P. Unzueta, and G. J. O. Beran, Phys. Chem. Chem. Phys. **21**, 14992 (2019).

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

[334] G. M. Day, W. D. S. Motherwell, H. L. Ammon, S. X. M. Boerrigter, R. G. Della Valle, E. Venuti, A. Dzyabchenko, J. D. Dunitz, B. Schweizer, B. P. van Eijck, P. Erk, J. C. Facelli, V. E. Bazterra, M. B. Ferraro, D. W. M. Hofmann, F. J. J. Leusen, C. Liang, C. C. Pantelides, P. G. Karamertzanis, S. L. Price, T. C. Lewis, H. Nowell, A. Torrisi, H. A. Scheraga, Y. A. Arnautova, M. U. Schmidt, and P. Verwer, Acta Cryst. B **61**, 511 (2005).

[335] B. P. Chekal, A. M. Campeta, Y. A. Abramov, N. Feeder, P. P. Glynn, R. W. McLaughlin, P. A. Meenan, and R. A. Singer, Org. Proc. Res. Dev. **13**, 1327 (2009).

[336] A. M. Campeta, B. P. Chekal, Y. A. Abramov, P. A. Meenan, M. J. Henson, B. Shi, R. A. Singer, and K. R. Horspool, J. Pharm. Sci. **99**, 3874 (2010).

[337] D. Lupyan, Y. A. Abramov, and W. Sherman, J. Comput. Aided Mol. Des. **26**, 1195 (2012).

[338] A. Otero-de-la Roza and E. R. Johnson, J. Chem. Phys. **137**, 054103 (2012).

[339] J. Takeya, M. Yamagishi, Y. Tominari, R. Hirahara, Y. Nakazawa, T. Nishikawa, T. Kawase, T. Shimoda, and S. Ogawa, Appl. Phys. Lett. **90**, (2007).

[340] G. Schweicher, G. Garbay, R. Jouclas, F. Vibert, F. Devaux, and Y. H. Geerts, Adv. Mater. **32**, 1905909 (2020).

[341] M. Kytka, L. Gisslen, A. Gerlach, U. Heinemeyer, J. Kováč, R. Scholz, and F. Schreiber, J. Chem. Phys. **130**, (2009).

[342] J. L. Bredas, J. P. Calbert, D. A. da Silva Filho, and J. Cornil, Proc. Nat. Acad. Sci. **99**, 5804 (2002).

[343] D. A. Da Silva Filho, E. G. Kim, and J. L. Brédas, Advanced Materials **17**, 1072 (2005).

[344] M. L. Clapham, E. C. Murphy, and C. J. Douglas, Synthesis in press (2020).

[345] O. D. Jurchescu, A. Meetsma, and T. T. M. Palstra, Acta Cryst. B **62**, 330 (2006).

[346] L. Huang, Q. Liao, Q. Shi, H. Fu, J. Ma, and J. Yao, J. Mater. Chem. **20**, 159 (2010).

[347] D. Käfer, L. Ruppel, G. Witte, and C. Wöll, Phys. Rev. Lett. **95**, 166602 (2005).

[348] S. Haas, A. Stassen, G. Schuck, K. Pernstich, D. Gundlach, B. Batlogg, U. Berens, and H.-J. Kirner, Phys. Rev. B **76**, 1 (2007).

[349] A. S. Paraskar, A. R. Reddy, A. Patra, Y. H. Wijsboom, O. Gidron, L. J. W. Shimon, G. Leitus, and M. Bendikov, Chem. Eur. J. **14**, 10639 (2008).

[350] S. Bergantin and M. Moret, Cryst. Growth Des. **12**, 6035 (2012).

[351] S. Uttiya, L. Miozzo, E. M. Fumagalli, S. Bergantin, R. Ruffo, M. Parravicini, A. Papagni, M. Moret, and A. Sassella, J. Mater. Chem. C **2**, 4147 (2014).

[352] M. Mamada, H. Katagiri, T. Sakanoue, and S. Tokito, Cryst. Growth Des. **15**, 442 (2015).

[353] C. Sutton, M. S. Marshall, C. D. Sherrill, C. Risko, and J.-L. Bredas, J. Am. Chem. Soc. **137**, 8775 (2015).

[354] W. A. Ogden, S. Ghosh, M. J. Bruzek, K. A. McGarry, L. Balhorn, V. Young, L. J. Purvis, S. E. Wegwerth, Z. Zhang, N. A. Serratore, C. J. Cramer, L. Gagliardi, and C. J. Douglas, Cryst. Growth Des. **17**, 643 (2017).

[355] Y. Wu, X. Ren, K. A. McGarry, M. J. Bruzek, C. J. Douglas, and C. D. Frisbie, Adv. Elec. Mater. **3**, 1700117 (2017).

[356] G. Xie, S. Hahn, F. Rominger, J. Freudenberg, and U. H. F. Bunz, Chem. Commun. **54**, 7593 (2018).

[357] Z. Zhang, W. A. Ogden, V. G. Young, and C. J. Douglas, Chem. Commun. **52**, 8127 (2016).

[358] Y. Sakamoto and T. Suzuki, J. Org. Chem. **82**, 8111 (2017).

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

[360] A. Pulido, L. Chen, T. Kaczorowski, D. Holden, M. A. Little, S. Y. Chong, B. J. Slater, D. P. McMahon, B. Bonillo, C. J. Stackhouse, A. Stephenson, C. M. Kane, R. Clowes, T. Hasell, A. I. Cooper, and G. M. Day, Nature **543**, 657 (2017).

[361] A. N. Sokolov, S. Atahan-Evrenk, R. Mondal, H. B. Akkerman, R. S. Sánchez-Carrera, S. Granados-Focil, J. Schrier, S. C. Mannsfeld, A. P. Zoombelt, Z. Bao, and A. Aspuru-Guzik, Nature Commun. **2**, 437 (2011).

[362] J. Yang, S. De, J. E. Campbell, S. Li, M. Ceriotti, and G. M. Day, Chem. Mater. **30**, 4361 (2018).

[363] C. Y. Cheng, J. E. Campbell, and G. M. Day, Chem. Sci. **11**, 4922 (2020).

[364] V. Coropceanu, J. Cornil, D. A. da Silva Filho, Y. Olivier, R. Silbey, and J.-L. Brédas, Chem. Rev. **107**, 926 (2007).

[365] C. Sutton, C. Risko, and J.-L. Brédas, Chem. Mater. **28**, 3 (2016).

[366] X. Wang, T. Garcia, S. Monaco, B. Schatschneider, and N. Marom, CrystEngComm **18**, 7353 (2016).

[367] B. G. Johnson, C. A. Gonzales, P. M. Gill, and J. A. Pople, Chem. Phys. Lett. **221**, 100 (1994).

[368] O. V. Gritsenko, B. Ensing, P. R. T. Schipper, and E. J. Baerends, J. Phys. Chem. A **104**, 8558 (2000).

[369] S. Patchkovskii and T. Ziegler, J. Chem. Phys. **116**, 7806 (2002).

[370] S. Andersson and M. Grüning, J. Phys. Chem. A **108**, 7621 (2004).

[371] T. Bally and G. N. Sastry, J. Phys. Chem. A **101**, 7923 (1997).

[372] A. Ruzsinszky, J. P. Perdew, G. I. Csonka, O. A. Vydrov, and G. E. Scuseria, J. Chem. Phys. **125**, 194112 (2006).

[373] A. D. Dutoi and M. Head-Gordon, Chem. Phys. Lett. **422**, 230 (2006).

[374] M.-C. Kim, E. Sim, and K. Burke, Phys. Rev. Lett. **111**, 073003 (2013).

[375] B. G. Janesko and G. E. Scuseria, J. Chem. Phys. **128**, 244112 (2008).

[376] P. Verma, A. Perera, and R. J. Bartlett, Chem. Phys. Lett. **524**, 10 (2012).

[377] M.-C. Kim, E. Sim, and K. Burke, J. Chem. Phys. **140**, 18A528 (2014).

[378] M.-C. Kim, H. Park, S. Son, E. Sim, and K. Burke, J. Phys. Chem. Lett. **6**, 3802 (2015).

[379] C. Greenwell, J. Řezáč, and G. J. O. Beran, .

[380] G. J. O. Beran, CrystEngComm (2018).

[381] D. Casanova, J. Chem. Theory Comput. **10**, 324 (2014).

[382] H. P. G. Thompson and G. M. Day, Chem. Sci. **5**, 3173 (2014).

[383] M. J. Frisch *et al.*, Gaussian 09 Revision E.01, 2009, gaussian Inc. Wallingford CT.

[384] F. Weigend and R. Ahlrichs, Phys. Chem. Chem. Phys. **7**, 3297 (2005).

[385] F. Neese, Wiley Interdiscip. Rev. Comput. Mol. Sci. **2**, 73 (2012).

[386] Z. Lin, J. Zou, C. Peng, S. Liu, Z. Li, and X. Wan, 1 .

[387] M. C. Georgiadis, *Predictive Control in Process Engineering Chemical Engineering Dynamics Logistic Optimization of Chemical Production Processes Product Design and Engineering* (PUBLISHER, ADDRESS, YEAR), Vol. 6.

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

[389] C. K. Leech, S. A. Barnett, K. Shankland, M. Gutmann, and C. C. Wilson, Acta Cryst. B **62**, 926 (2006).

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

[391] D. Vilela Oliveira, J. Laun, M. F. Peintinger, and T. Bredow, J. Comput. Chem. **40**, 2364 (2019).

[392] J. G. Brandenburg, C. Bannwarth, A. Hansen, and S. Grimme, J. Chem. Phys. **148**, 064104 (2018).

Appendix A

Supporting Information for "Accurate noncovalent interactions via dispersion-corrected second-order Møller-Plesset perturbation theory"

A.1 Additional Computational Details

This Section provides additional information regarding (1) the Tang-Toennies damping parameterization and (2) details for the anthracene and helicene benchmark calculations. The latter benchmarks have not previously been published.

A.1.1 Separate vs Shared Tang-Toennies Damping

As described in the main paper, better reproduction of individual UCHF and CKS dispersion energies is obtained using separate Tang-Toennies damping parameters for each type of dispersion energy. As shown in Table A.1, the optimal a_1 and a_2 Tang-Toennies parameters can differ significantly for the UCHF and CKS dispersion energies. Nevertheless, one can obtain very good results with a common set of Tang-Toennies parameters for both the UCHF and CKS dispersion energies.

As an example, Figure A.1 examines the components of the dispersion correction for the benzene-uracil dimer (dimer #28 from S66x8). Figures A.1a and A.1b demonstrate that both the MP2D models with shared CKS and UCHF damping parameters (the recommended MP2D model) and one with separate damping parameters underestimate the magnitude of the MP2C UCHF and CKS dispersion, particularly at short range. Using separate damping parameters instead gives UCHF and CKS component dispersion energies that are closer to the MP2C results. However, the net dispersion correction, which is the difference between the CKS and UCHF dispersion energies, is quite similar in all three

Table A.1: Comparison of optimal MP2D parameters when one uses shared or separate Tang-Toennies damping parameters for the UCHF and CKS dispersion energies. The shared damping set here is identical to the one used throughout the main paper.

Parameter	Shared Damping	Separate Damping	
		UCHF	CKS
s_8	1.187	0.997	
a_1	0.944	1.031	1.207
a_2	0.480 Å	-0.477 Å	1.207 Å
r_{cut}	0.72	0.71	
w	0.20	0.20	

cases and is close to the exact correction (i.e. the difference between MP2 and CCSD(T)), as shown in Figure A.1c. In this particular dimer, MP2D with the shared damping parameters actually gives slightly smaller errors overall (Figure A.1d) than either the separate damping parameter model or MP2C. Of course, from the data presented in the paper, MP2C generally performs modestly better than MP2D over a larger suite of benchmarks.

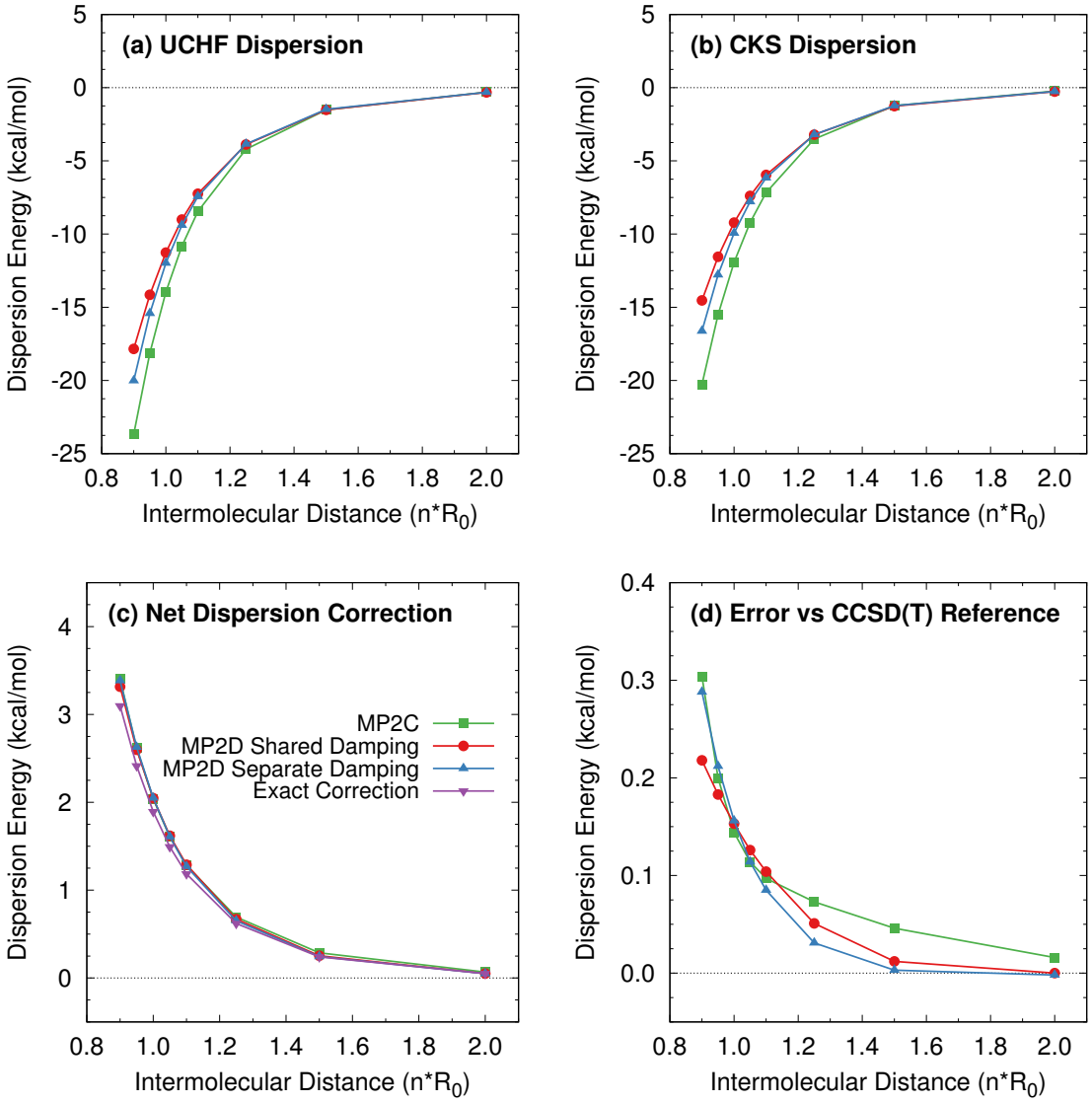


Figure A.1: Examination of the (a) UCHF dispersion, (b) CKS dispersion, (c) net MP2 dispersion correction, and (d) Errors in the predicted interaction energies for the benzene-uracil π-π dimer from the S66x8 benchmark set.

A.1.2 Benchmark data sets

The anthracene and [7]helicene potential energy curves presented here have not been published previously. Details about their generation, along with reference structures and energies are provided here.

A.1.3 Anthracene PES

The geometries for the 1-D anthracene dimerization potential energy scan were optimized at the BLYP-D3(BJ)/def2-TZVP level of theory, constraining the distance between the two anthracene molecules at C7–C21 and C8–C22 (as numbered in the provided coordinate files) and relaxing all other degrees of freedom. Complete-basis-set reference CCSD(T) energies were then approximated via the focal point approach as

$$E_{CCSD(T)/CBS} \approx E_{MP2/CBS} + \left(E_{CCSD(T)/cc-pVDZ} - E_{MP2/cc-pVDZ}\right) \tag{A.1}$$

The MP2/CBS extrapolation was performed from the cc-pVTZ and cc-pVQZ basis sets, and the post-MP2 correction evaluated in the cc-pVDZ basis set. Diffuse functions were omitted due to the large system size and the large degree of basis function overlap in the covalent regime. Ideally one would use a basis set larger than cc-pVDZ for estimating the post-MP2 correlation, but given the size of the system and the number of data points, we opted for cc-pVDZ. The identical focal-point strategy was used to compute the MP2.5/CBS energies. The CCSD(T) calculations were run using the density-fitted frozen natural orbital implementation in PSI4. However, at the default 10^{-6} natural orbital cutoff used, no virtual orbitals were actually truncated in this system.

Energies are computed relative to that of two non-interacting anthracene monomers, with the results listed in Table A.2. Note that spin-restricted orbitals were used throughout. Static correlation likely becomes significant near the transition from the covalent to non-covalent basins. Cartesian coordinates of the geometries used are provided separately.

Table A.2: 1-D potential energy scan for the anthracene dimer relative to two non-interacting anthracene monomers (in kcal/mol). The CCSD(T) results are used as the reference value.

R (Å)	CCSD(T) CBS	MP2.5 CBS	MP2 CBS	MP2D CBS	B3LYP-D3(BJ) def2-QZVP
1.40	14.05	12.60	3.92	13.17	27.71
1.50	-5.68	-6.98	-15.13	-6.67	8.13
1.60	-11.15	-12.31	-20.30	-12.56	2.30
1.70	-8.42	-9.42	-17.62	-10.45	4.33
1.80	-1.34	-2.19	-11.07	-4.05	10.46
1.90	7.69	7.04	-2.76	4.64	18.28
2.00	17.10	16.71	5.79	14.94	26.00
2.10	25.94	25.78	13.36	24.87	32.50
2.20	33.63	33.60	19.37	31.79	38.31
2.30	39.93	40.01	23.85	36.27	43.53
2.40	44.77	44.94	26.84	39.09	47.81
2.50	48.22	48.57	28.89	40.81	53.64
2.60	21.16	19.87	7.92	18.38	23.70
2.70	13.90	12.54	1.63	11.33	15.91
2.80	8.10	6.70	-3.28	5.71	9.68
2.90	3.52	2.10	-7.06	1.28	4.75
3.00	-0.05	-1.47	-9.88	-2.16	0.91

R (Å)	CCSD(T) CBS	MP2.5 CBS	MP2 CBS	MP2D CBS	B3LYP-D3(BJ) def2-QZVP
3.20	-4.81	-6.20	-13.32	-6.71	-4.24
3.40	-7.29	-8.62	-14.66	-9.01	-6.96
3.60	-8.24	-9.48	-14.60	-9.80	-8.01
3.80	-8.27	-9.33	-13.38	-9.65	-7.98
4.00	-7.55	-8.41	-11.54	-8.72	-7.11
4.20	-6.52	-7.22	-9.63	-7.51	-5.97
4.50	-4.96	-5.45	-7.11	-5.72	-4.34
5.00	-2.93	-3.21	-4.14	-3.45	-2.44
6.00	-0.93	-1.03	-1.36	-1.26	-0.83

Table A.3: 1-D potential energy scan for [7]helicene relative to linear heptacene (in kcal/mol). The MP2.5 results are used as the reference value.

R (Å)	MP2.5 CBS	MP2 CBS	MP2D CBS	B3LYP-D3(BJ) def2-QZVP
3.4	-21.158	-29.837	-21.069	-15.671
3.5	-21.945	-30.319	-21.821	-16.514
3.6	-22.551	-30.640	-22.395	-17.162
3.7	-23.014	-30.834	-22.828	-17.666
3.8	-23.357	-30.916	-23.144	-18.050
3.9	-23.603	-30.912	-23.367	-18.337
4.0	-23.775	-30.862	-23.515	-18.540
4.1	-23.878	-30.738	-23.600	-18.675
4.2	-23.930	-30.588	-23.634	-18.754
4.3	-23.930	-30.390	-23.622	-18.786
4.4	-23.908	-30.208	-23.583	-18.781
4.5	-23.837	-29.958	-23.501	-18.744
4.6	-23.738	-29.700	-23.393	-18.678
4.7	-23.612	-29.430	-23.257	-18.582

A.1.4 Helicene PES

To generate the 1-D potential energy surface scan, [7]helicene geometries were optimized at the B3LYP-D3(BJ)/def2-TZVP level with the distance between C(4)–C(42) and C(6)–C(45) were constrained (as numbered in the provided coordinate files) to identical distances ranging from 3.4–4.7 Å. All other nuclear degrees of freedom were relaxed. Complete-basis-set reference MP2.5 energies were then approximated via the focal point approach as

$$E_{MP2.5/CBS} \approx E_{MP2/CBS} + \left(E_{MP2.5/cc-pVTZ} - E_{MP2/cc-pVTZ}\right) \tag{A.2}$$

As in the anthracene case, the CBS extrapolation was performed from cc-pVTZ and cc-pVQZ calculations. Energies are computed as a function of this constrained distance relative to the energy of linear heptacene (Table A.3). Cartesian coordinates of the geometries used are provided separately.

Appendix B

Supporting Information for "Spin-component-scaled and dispersion-corrected second-order Møller-Plesset perturbation theory: A path toward chemical accuracy"

This section provides additional insight on (1) an alternative parameter search where a Bayesian search algorithm with Gaussian processes was explored, (2) how the damping function changes from MP2D to SCS-MP2D, (3) a table displaying the percent relative root mean square errors for each method on each benchmark data set, and (4) the full anthracene photodimer potential energy curve.

B.1 Bayesian Parameter Search Algorithm

Bayesian optimization is intended to find optimal parameters (x_1, ..., x_n) for an unknown function f(x_1, ... x_n). For example, Bayesian optimization is useful for finding optimal hyperparameters in machine learning models, especially when the model is expensive to train. If analytic gradients of the objective function are available, then Bayesian optimization will likely be less efficient than a gradient-based optimizer. For that reason, the evolutionary algorithm with gradient-based minimization was employed for the primary parameter search in this work. However, given the rough SCS-MP2D parameter landscape with many local minima, Bayesian optimization was employed as a secondary check for discovering any potential alternate parameter sets that might have been missed in the evolutionary search.

A Bayesian algorithm with Gaussian processes was used in this work as a probe of the likely parameter space, and as a convergence test for the genetic/gradient optimization

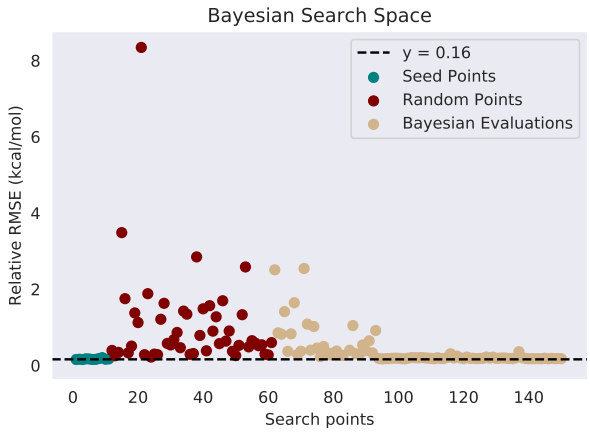


Figure B.1: Example of a Bayesian parameters search: (1) the optimizer is seeded with good parameter sets from the genetic/gradient optimizer, (2) 50 random points are sampled within a predefined sample space, and (3) new points are sampled based on a function constructed from the seeded and random guess points.

algorithm. The Bayesian optimizer was seeded with top performing parameter sets from the evolutionary algorithm/gradient-based optimizations. The progression of the optimizer is shown in Figure B.1. Initially the Bayesian algorithm probes the potential energy landscape of the optimization problem with the seed points, then a predefined number of random points are sampled and used to construct a probability model of the objective function for the parameter space. After the random points are sampled, the most promising points according to the probability model are sampled. The probability model is continuously updated as new points are sampled. As exemplified by Figure B.2, none of the Bayesian search runs identified any parameter sets that were better than what the evolutionary algorithm searches found. This result increases the confidence that an optimal or near-optimal set of parameters was discovered for the chosen training data.

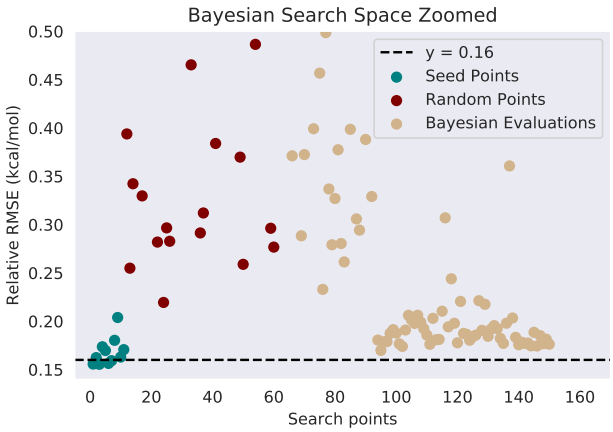


Figure B.2: Zooming in on the search space shows how the Bayesian optimizer finds new parameter sets that perform well, but do not match the performance of the best parameter sets from the genetic/gradient optimizer.

B.2 Comparing the Tang-Toennies damping function from MP2D and SCS-MP2D

Here, the Tang-Toennies damping function is plotted as a function of the interatomic distance between three different atom pairs (C-C, C-H, and H-H) for our original MP2D method and SCS-MP2D. Compared to MP2D, SCS-MP2D allows the dispersion term to contribute more at shorter ranges. It is possible that this is in response to the reduced scaling of the opposite-spin C_{os} and same-spin C_{ss} coefficients in SCS-MP2D. Their respective values of 0.8263 and 0.9004 are both less than 1. Unlike the original SCS-MP2 and SCS-MI-MP2 methods, SCS-MP2D has both spin components reduced from the canonical value of 1. It is also worth observing that the C_{os} and C_{ss} values are very similar. This suggests it could be possible to develop an SCS-MP2D type method with a single scaling coefficient for the MP2 correlation energy.

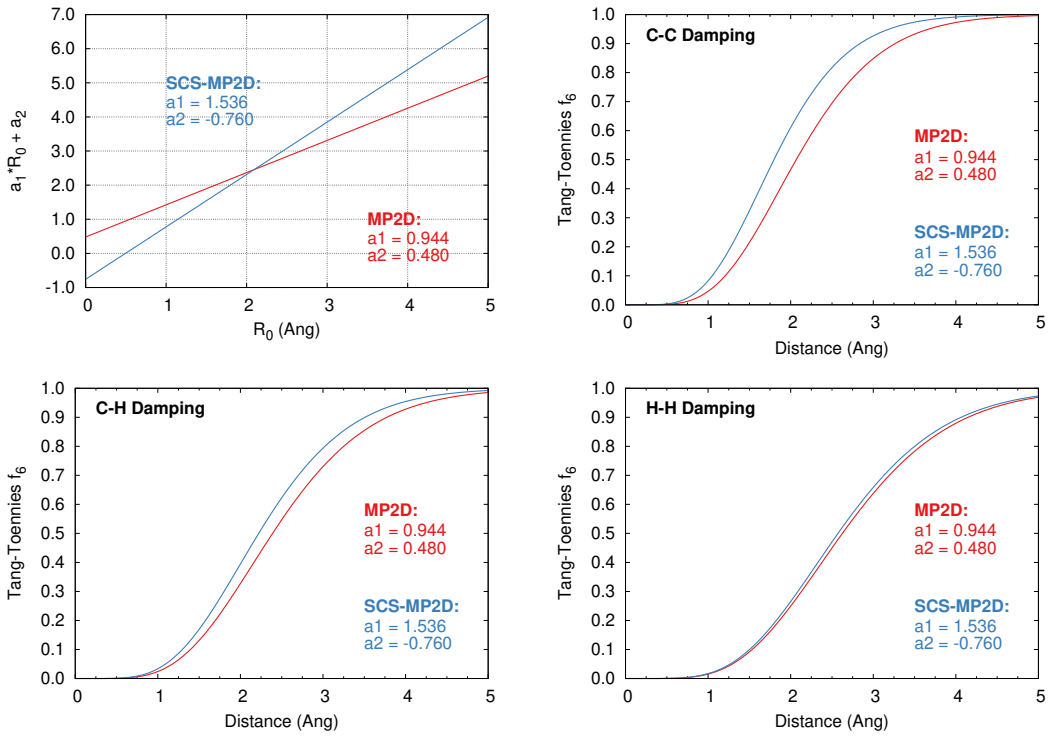


Figure B.3: Examination of the Tang-Toennies damping function f_6 as a function of the interatomic distance, R, for three atom-type interactions.

B.3 Percent Weighted Root Mean Square Errors for Benchmark Data Sets

As a companion to Table 2 in the main paper, Table B.1 summarizes the relative root-mean-square errors for the data sets

Table B.1: Relative root mean square error (%) calculated by dividing the RMSE by the mean absolute value of the reference energies and multiplying by 100. The asterix (*) indicates data sets that were used to fit SCS-MP2D.

Data Set	MP2 CBS	MP2D CBS	SCS-MP2D CBS	DSD-BLYP -D3(BJ) def2-QZVP	revDSD-PBEP86 -D3(BJ) def2-QZVP	ωB97X-V def2-QZVP	ωB97M-V aQZ
Intermolecular Interactions							
S66x8	16.71	3.92	3.10*	4.51	4.08	5.32	2.73
3B-69 Dimers	8.08	5.08	4.35	4.70	4.57	4.85	4.07
SSI	4.31	1.87	2.02	1.80	1.45	2.40	1.80
HBC6	2.96	2.46	2.36	3.44	1.58	2.98	2.20
NBC10	98.56	18.62	8.71	20.97	4.43	21.87	10.60
Charge Transfer	19.61	4.06	2.46	5.52	4.48	4.12	3.26
HB375	7.69	2.86	2.26	2.50	2.37	3.09	3.38
IHB100	2.36	2.44	1.77	2.19	1.38	1.93	1.88
Conformational Energies							
SCONF	6.72	7.65	3.98*	5.74	2.93	4.57	5.15
ACONF	6.00	3.60	6.32	4.53	12.84	3.27	4.53
Amino20x4	10.66	6.89	7.30	6.66	6.82	9.86	9.84
MCONF	20.52	8.13	6.62	11.15	3.92	5.43	7.93
PCONF21	68.45	25.90	19.36	29.39	14.02	21.58	42.81
Reaction Energies							
DARC	12.22	5.84	4.35*	3.37	1.96	13.49	3.01
ISO34	11.53	9.76	6.62	7.26	3.34	10.68	5.66
ISOL24	16.74	12.64	10.07	12.19	7.79	18.92	10.87
IDISP	49.40	9.97	9.06	11.27	4.70	27.28	19.87

B.4 Complete Anthracene Photodimerization Potential Energy Curve

This plot shows the full potential energy curve for the anthracene photodimerization. As stated in the main paper, the restricted, single-reference electronic structure models are probably not reliable in the intermediate regime between the two wells due to the substantial static correlation associated with forming/breaking two C-C single bonds.

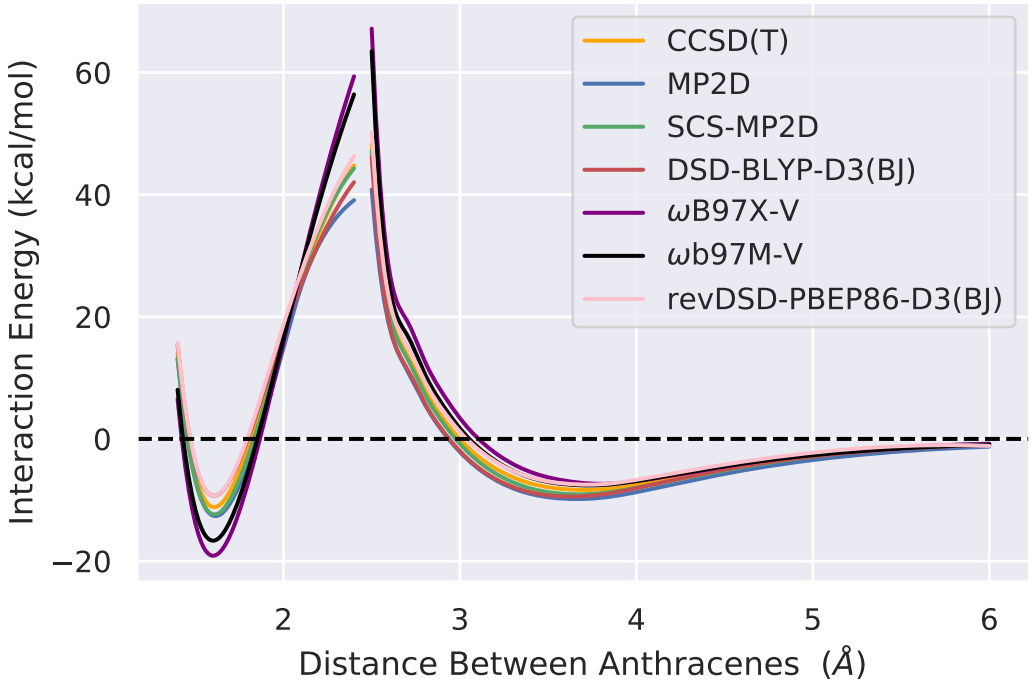


Figure B.4: 1-D potential energy scan following the dissociation of an anthracene photodimer to a separated π-stacked dimer

Appendix C

Supplementary Information for "Overcoming the difficulties of predicting conformational polymorph energetics in molecular crystals via correlated wavefunction methods"

C.1 Additional Computational Details

C.1.1 DFT Monkhorst-Pack grids and quality of the optimized structures

The planewave DFT Monkhorst-Pack grids indicated in Table E.4 were used for the B86bPBE-XDM geometry optimizations and single-point energies. These choices were made by systematically increasing the grid until the geometry no longer changed. Roughly speaking, 9 k-points were used for lattice constants below 4 Å, 7 k-points were used for lattice constants of 4–6 Å, 5 k-points for lattice constants of 6–8 Å, 3 k-points for lattice constants 8-12 Å, and 1 k-point for larger lattice constants. Table E.4 also indicates the rmsd15 value for the 0 K (fully relaxed) and 298 K (fixed experimental lattice parameters) B86bPBE-XDM geometry optimizations. The rmsd15 metric indicates the root-mean-square deviation in the non-hydrogen atomic positions for a cluster of 15 molecules in each crystal.[388]

For *o*-acetamidobenzamide, the room-temperature x-ray diffraction structures of ACBNZA and ACBNZA01 were chosen over the more recent low-temperature neutron diffraction structures (ACBNZA02 and ACBNZA03)[389] to enable examination of the temperature dependent behaviors. Full B86bPBE-XDM relaxation of the neutron diffraction structures resulted in structures that were virtually identical to those used here (rmsd15 differences less than 0.04 Å and energy differences of 0.1 kJ/mol or less).

An earlier study[203] employing the PBE-NP dispersion-corrected DFT functional found that the ON structure relaxed considerably upon optimization, with the key in-

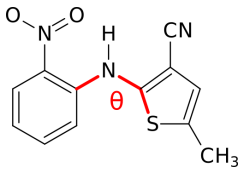
tramolecular dihedral angle θ changing by $\sim$25°. This suggested that DFT might be particularly problematic for the ON polymorph. However, the B86bPBE-XDM functional used here does not appear to suffer from the same problems in reproducing the experimental crystal structures. For the fixed-cell optimization, the dihedral angle θ changes only 2.6°, which is on par with the average 3.4 ± 2.4° change observed for the other seven polymorphs (Table C.2). Relaxing the cell fully leads to a larger 9.5° change in that dihedral angle, but this is again on par with the average change of 8.2 ± 3.6° change seen for the other forms. Structure overlays give rmsd15 values of 0.05 Å for the fixed cell and 0.22 Å for the fully relaxed cell relative to experiment, which are again compatible with values seen for the other crystals (Table E.4). In other words, there is no obvious structural discrepancy between the B86bPBE-XDM and experimentally reported structures that might account for the disagreement in relative stabilities.

Table C.1: Summary of **k**-point grids and quality of the agreement between the predicted and experimental crystal structures.

Crystal Polymorph	CSD RefCode	**k**-point grid	0 K Structure rmsd15 (Å)	298 K Structure rmsd15 (Å)
		***o*-Acetamidobenzamide**		
α form	ACBNZA	(7,3,3)	0.158	0.039
β form	ACBNZA01	(5,3,3)	0.144	0.050
		ROY		
form Y	QAXMEH01	(3,1,3)	0.220	0.048
form YT04	QAXMEH12	(5,3,3)	0.208	0.051
form R	QAXMEH02	(5,5,3)	0.163	0.052
form OP	QAXMEH03	(5,3,3)	0.199	0.075
form ON	QAXMEH	(9,1,1)	0.219	0.047
form YN	QAXMEH04	(7,3,3)	0.250	0.050
form ORP	QAXMEH05	(1,3,1)	0.454	0.053
form PO13	QAXMEH52	(7,1,3)	0.255	0.099

Crystal Polymorph	CSD RefCode	**k**-point grid	0 K Structure rmsd15 (Å)	298 K Structure rmsd15 (Å)
		Oxalyl Dihydrazide		
α form	VIPKIO01	(9,7,3)	0.071	0.040
β form	VIPKIO02	(9,3,7)	0.333	0.041
γ form	VIPKIO03	(7,1,5)	0.130	0.106
δ form	VIPKIO04	(9,1,7)	0.190	0.022
ϵ form	VIPKIO05	(7,9,3)	0.131	0.017

Table C.2: Comparison of the experimental and predicted S-C-N-C dihedral angle θ for the different ROY polymorphs. Absolute errors between theory and experiment are indicated in parentheses. The key ON polymorph results are highlighted in bold.



		B86bPBE-XDM		PBE-N2
	Experiment	298 K Structure	0 K Structure	0 K Structure
Y	104.7°	108.0° (3.2°)	112.5° (7.8°)	111.6° (6.8°)
YT04	112.8°	116.6° (3.7°)	119.8° (6.9°)	118.7° (5.8°)
R	21.7°	20.5° (1.2°)	19.5° (2.2°)	22.2° (0.4°)
OP	46.1°	43.0° (3.1°)	37.5° (8.6°)	40.1° (6.0°)
ON	**52.6°**	**50.0° (2.6°)**	**43.1° (9.5°)**	**26.2° (26.4°)**
YN	104.0°	104.4° (0.3°)	111.3° (7.3°)	112.3° (8.3°)
ORP	39.4°	34.4° (5.0°)	29.0° (10.4°)	31.0° (8.4°)
PO13	122.1°	129.5° (7.5°)	136.1° (14.0°)	135.3° (13.2°)

C.1.2 Basis set dependence for the MP2 models

The HMBI calculations here rely on monomer and dimer MP2 calculations. MP2-based models exhibit slower convergence with basis set than do HF or DFT models, and the use of large basis sets is important for obtaining well-converged polymorph stability predictions. In this work, all MP2-based methods were extrapolated to the CBS limit from aug-cc-pVTZ and aug-cc-pVQZ results. Additionally, the MP2D dispersion correction was designed for use with large basis sets, and the short-range damping functions were fitted against CBS-limit results.[5]

Figures C.1 and C.2 show the dramatic impact of basis set completeness on the predicted polymorph stabilities for acetamidobenzamide and ROY. In acetamidobenzamide, the relative stability of the two forms changes by more than 6 kJ/mol between aug-cc-pVDZ and the CBS limit. Examination of the 1-body and 2-body contributions (not explicitly shown here) indicates that the basis set dependence is approximately evenly split between the intra- and intermolecular contributions. The ROY polymorph stabilities also exhibit strong basis set dependence. For polymorph ORP, the relative stability decreases 6.9 kJ/mol between aug-cc-pVDZ and the CBS limit. The average magnitude of the basis set change between aug-cc-pVDZ and the CBS limit is 3.6 kJ/mol, which again is substantial on the energy scale of polymorphism. The basis set dependence for oxalyl dihydrazide has been examined previously,[224,318] and similar results were found to those shown here for ROY and acetamidobenzamide.

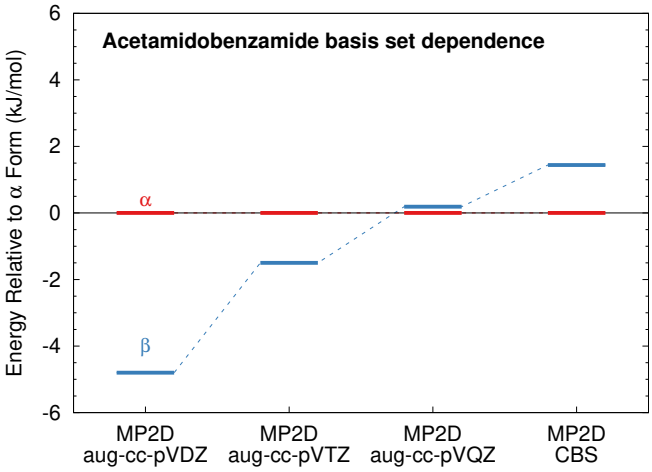


Figure C.1: Dependence of the polymorph stabilities for *o*-acetamidobenzamide on the basis set used for the MP2D 1 & 2-body contributions. In all cases, periodic HF/pob-TZVP-rev2 was used for the many-body contribution.

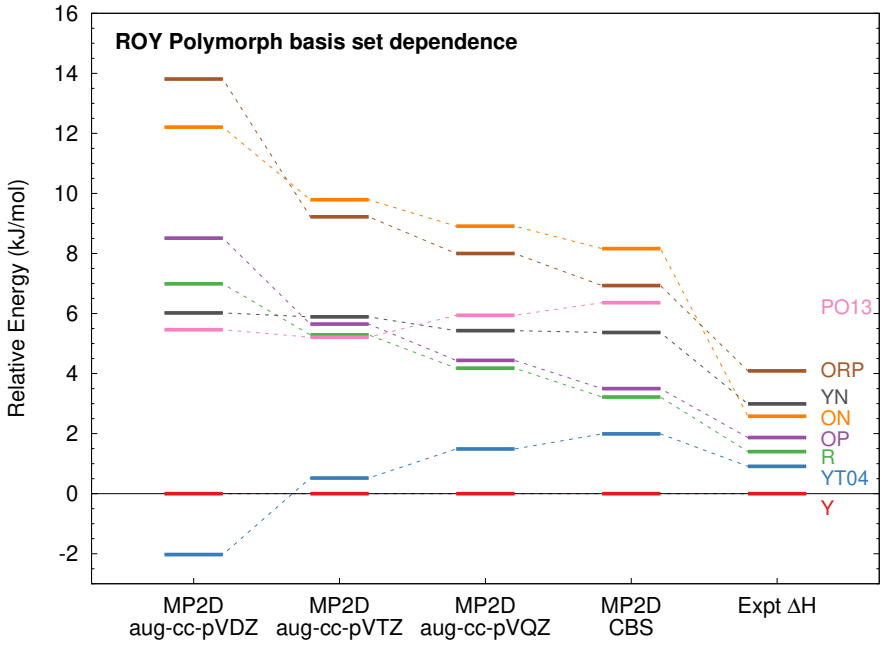


Figure C.2: Dependence of the ROY polymorph stabilities on the basis set used for the MP2D 1 & 2-body contributions. In all cases, periodic HF/pob-TZVP-rev2 was used for the many-body contribution.

C.1.3 Selection of the Hartree-Fock many-body treatment basis set

The 1-body (monomer) and 2-body (dimer) terms frequently account for 90% or more of the total lattice energy in the HMBI model, while the remaining ∼10% results from the intermolecular many-body contributions. However, in the polymorphic systems considered in this paper, the proportional impact of the many-body term on the relative polymorph energies is considerably larger. Therefore, it is important to obtain a well-converged description of the many-body term. Periodic HF is relatively inexpensive, particularly when Gaussian basis sets are employed, and can describe the frequently dominant many-body polarization effects.

Periodic HF (and DFT) calculations with large Gaussian basis sets suffer from well-known issues with linear dependencies and poor convergence of the self-consistent-field equations. Adapting a Gaussian basis sets designed for molecular systems to periodic calculations typically requires eliminating the most diffuse basis functions and subsequent re-tuning of the remaining basis functions. Several triple-ζ basis sets have been developed in recent years, including the pob-TZVP,[390] pob-TZVP-rev2,[391] and mTZVP basis set.[392] All three are modified versions of the popular def2-TZVP basis intended for periodic systems.

To assess the performance of these three basis sets, 15 five-molecule clusters were extracted from the oxalyl dihydrazide polymorphs. HF calculations were performed on the entire gas-phase cluster, and then separately subtracting out the HF monomer and dimer contributions according to a many-body expansion. Consistent with the HMBI treatment of the many-body terms, no counterpoise correction is employed in either the full cluster nor the fragment calculations. The HF/def2-QZVP results are chosen as the benchmark, and these results are compared against the traditional def2-TZVP basis set and the three aforementioned triple-ζ basis sets that are suitable for periodic calculations.

As shown in Table C.3, the def2-TZVP basis set gives the root-mean-square (RMS) error of 0.31 kJ/mol per monomer versus def2-QZVP. The pob-TZVP-rev2 basis revised the pob-TZVP basis in order to reduce the occurrence of basis set superposition error (BSSE). While BSSE is more prevalent in dimer interactions, the results here show that the pob-TZVP-rev2 basis gives considerably smaller errors (0.31 kJ/mol) than pob-TZVP (0.63 kJ/mol) relative to the benchmark def2-QZVP values. Finally, the mTZVP basis, which typically exhibits even less BSSE, predicts many-body contributions with errors intermediate between the two other basis sets. In the end, the pob-TZVP-rev2 basis was adopted in this study based on these benchmarks.

Table C.3: Predicted HF many-body energies (in kJ/mol per monomer) and the root-mean-square (RMS) error versus the def2-QZVP basis set results for 15 different clusters cut from the five polymorphs of oxalyl dihydrazide.

Cluster	def2-QZVP	def2-TZVP	pob-TZVP	pob-TZVP-rev2	mTZVP
α cluster #1	-1.87	-1.60	-1.51	-1.68	-1.20
α cluster #2	-4.05	-3.92	-4.20	-4.33	-3.82
α cluster #3	0.15	0.38	0.73	0.43	0.41
β cluster #1	-0.09	0.00	0.09	0.05	0.07
β cluster #2	0.00	0.21	0.29	0.02	0.40
β cluster #3	-0.02	0.27	0.55	0.29	0.49
γ cluster #1	0.36	0.60	0.89	0.68	0.62
γ cluster #2	0.05	0.15	0.35	0.25	0.13
γ cluster #3	0.33	0.58	0.83	0.66	0.65
δ cluster #1	0.17	0.36	0.61	0.44	0.46
δ cluster #2	-0.08	-0.11	0.45	0.23	-0.26
δ cluster #3	-0.70	-0.59	-0.50	-0.61	-0.54
ϵ cluster #1	-0.23	0.03	-0.03	-0.13	0.11
ϵ cluster #2	0.38	0.39	0.50	0.43	0.22
ϵ cluster #3	0.80	0.95	1.22	1.03	0.82
RMS Error		**0.31**	**0.63**	**0.37**	**0.51**

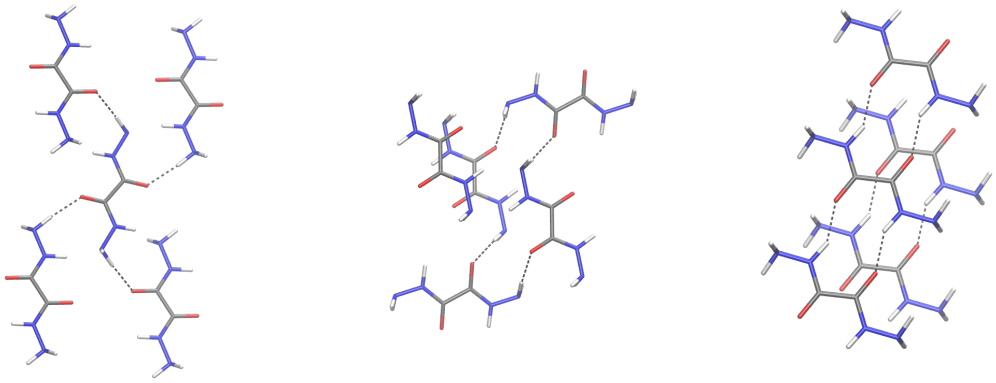


Figure C.3: Three sample oxalyl dihydrazide pentamer clusters. From left to right: α cluster #2, β cluster #3, and δ cluster #2.

C.2 *o*-Acetamidobenzamide polymorphs

C.2.1 Enthalpies and Gibbs Free Energies

To augment the enthalpy and free energy curves shown in the main paper, Figure C.4 plots the corresponding semi-schematic G and H curves for all four methods considered in the main paper. All three MP2-based models agree that the β form is thermodynamically preferred at high temperatures, though the specific free energies differ. MP2 predicts an enantiotropic relationship, while MP2C predicts a monotropic one. For MP2D, the relationship is also enantiotropic, though the two forms are nearly degenerate at 0 K. Consistent with experiment, all three MP2-based methods predict an exothermic $\Delta H_{\alpha\rightarrow\beta}$ phase transition at elevated temperatures. In contrast, B86bPBE-XDM incorrectly predicts an endothermic phase transition and that the α form is thermodynamically more stable at all temperatures. Examination of the $\Delta H_{\alpha\rightarrow\beta}$ values shown in Table 1 and Figure 3 of the main manuscript suggests that MP2C overestimates the stability of the β form due largely to the neglect of intramolecular dispersion.

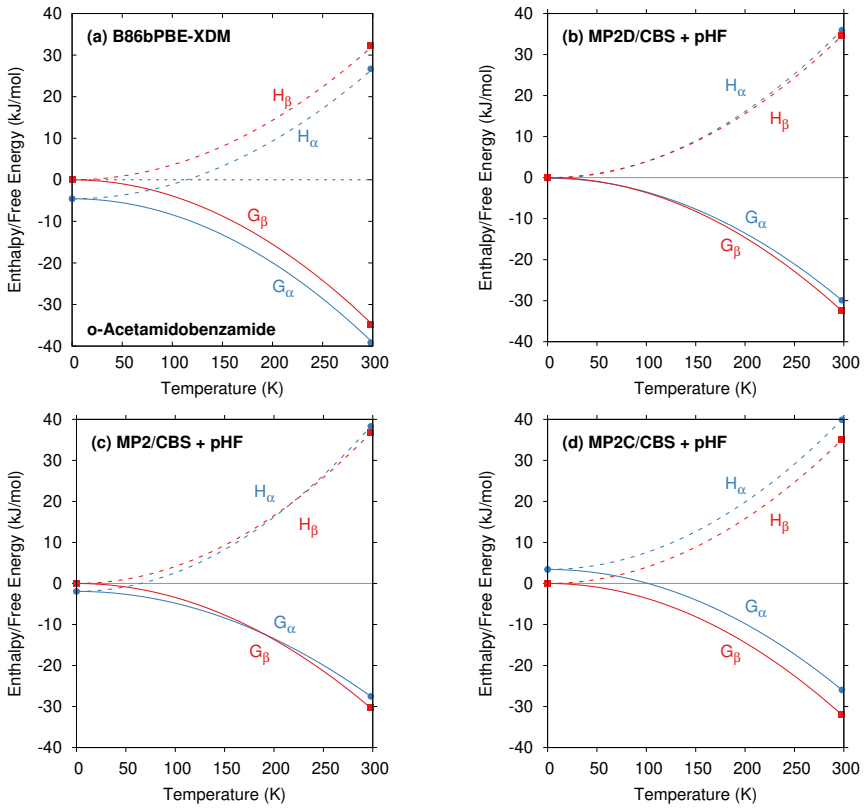


Figure C.4: Comparison of the predicted enthalpies and Gibbs free energies for the two polymorphs of *o*-acetamidobenzamide as computed with several different models. Data points were computed, but the curves connecting them are merely schematic.

Table C.4: Lattice energy differences, vibrational enthalpy contributions, and vibrational Gibbs free energy contributions to the α–β polymorph energy differences in *o*-acetamidobenzamide, in kJ/mol. Positive values indicate α is more stable than β. Summing the lattice energy and vibrational enthalpy contributions gives the results plotted in Figure 3 of the main paper.

	Lattice Energy Contributions				B86bPBE-XDM Phonon Contrib.	
T (K)	B86bPBE-XDM	MP2	MP2D	MP2C	H_{vib}	G_{vib}
0 K	5.8	3.2	1.4	-2.1	-1.3	-1.3
298 K	5.2	-1.9	-1.7	-5.2	0.3	-0.9
423 K[a]	4.9	-4.1	-3.0	-6.5	1.0	0.4
Change 0 K→298 K	-0.6	-5.1	-3.1	-3.0	1.6	0.4
Change 0 K→423 K	-0.9	-7.3	-4.4	-4.3	2.3	0.5

[a] Linearly extrapolated to 423 K from 0 K and 298 K values

To understand the temperature dependence of the *o*-acetamidobenzamide predictions better, Table C.4 decomposes the temperature dependence into lattice energy contributions (due to thermal expansion) and phonon contributions. Note that the MP2-type results use B86bPBE-XDM phonon frequencies, meaning that the vibrational contributions to the enthalpy H_{vib} and free energy G_{vib} are identical across all methods here.

For B86bPBE-XDM, the lattice energy changes stabilize the β form relative to α by less than 1 kJ/mol over the 0–423 K temperature range, while the vibrational enthalpy destabilizes the β form by 2.3 kJ/mol. The MP2-based models all predict more substantial lattice energy-driven stabilization of the β polymorph with increasing temperature, ranging from -7.3 kJ/mol for MP2 to -4.3–4.4 for the dispersion-corrected MP2D and MP2C models. These lattice energy changes favoring the β form at higher temperatures are partially canceled by the phonon contributions that increasingly favor the α form at higher temperatures, leading to the final predicted MP2D enthalpy that agrees very well with the experimentally measured transition enthalpy, as described in the main paper.

C.3 ROY polymorphs

C.3.1 Relative polymorph energies

Table C.5 summarizes the lattice energies computed here that were used to generate Figure 5 in the main paper.

Table C.5: Summary of relative lattice energies (in kJ/mol) computed here using the fixed-cell room temperature crystal structures.

Polymorph	B86bPBE-XDM	MP2	MP2D
Y	0.0	0.0	0.0
YT04	-0.4	1.5	2.0
R	-6.2	2.9	3.2
OP	-3.1	3.4	3.5
ON	-3.7	7.5	8.2
YN	1.3	5.1	5.4
ORP	-4.6	7.2	6.9
PO13	-3.0	6.0	6.4

C.3.2 Impact of the MP2D dispersion correction

Figure C.5 compares HMBI results with MP2 and MP2D. The differences in the relative polymorph energies are less than 1 kJ/mol throughout. In all cases except ORP, the dispersion correction slightly destabilizes the polymorphs relative to Y. Note that although the impact on the relative polymorph energies is small in this instance, the dispersion correction has a big impact on the overall lattice energies. For example, MP2D reduces the strength of the 2-body interaction energies by 27–31 kJ/mol, depending on the polymorph.

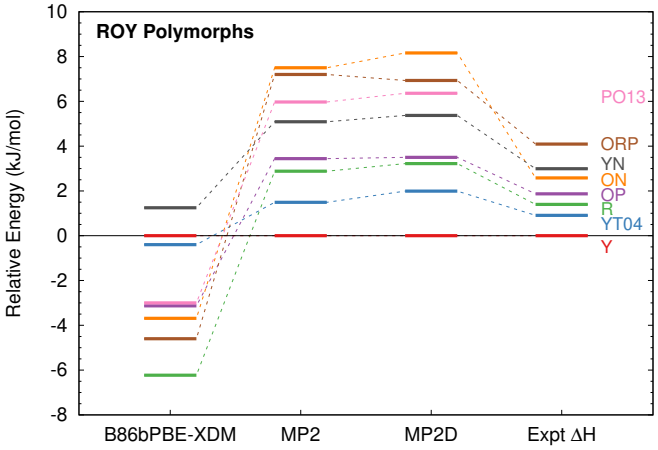


Figure C.5: The impact of the MP2D dispersion correction on the relative polymorph energies is fairly modest in ROY, as compared to uncorrected MP2.

C.3.3 Comparison of 0 K and room-temperature structures

Figure C.6 plots the relative energies with DFT and MP2D for the fully relaxed 0 K structures and the fixed-cell room-temperature structures. Note that the results here compare MP2D in the aug-cc-pVTZ basis, rather than in the CBS limit like all other results presented here. This basis set was chosen for the 0 K structures to reduce computational costs.

At the DFT level, the differences in the relative lattice energies for the two sets of structures are quite small. For MP2D, they are considerably larger. The reasons behind the larger MP2D differences are not entirely clear. One potential explanation could be that the optimal MP2D geometries might differ somewhat from the DFT ones, and therefore the DFT optimal structures correspond to steeper portions of the MP2D potential energy surface (and therefore one obtains larger changes in the energies with the geometry changes).

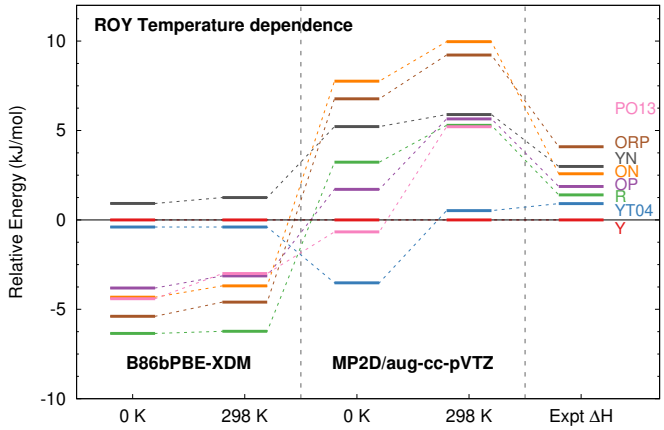


Figure C.6: The impact of using fully-relaxed (0 K) structures versus fixed-cell relaxations (298 K) on the relative lattice energies.

C.3.4 Omission of the R05 polymorph

The R05 polymorph is omitted from the present study because it was recently discovered that the HMBI model energies are poorly convergent in polar unit cells. The treatment of long-range electrostatics is generally problematic in periodic cells with a net dipole moment. In practice, this issue is frequently managed via application of tin-foil boundary conditions.

The working HMBI expression has the following form:

$$E^{HMBI} = E^{pHF}_{crystal} + \left(E^{MP2D}_{1body} - E^{HF}_{1body}\right) + \left(E^{MP2D}_{2body} - E^{HF}_{2body}\right) \tag{C.1}$$

where the 2-body terms are computed out to a cutoff radius of 9–10 Å. A smoothing function is applied across this interval to ensure smooth potential energy surfaces as the model transitions from the high-level (MP2D) to low-level (HF). The periodic HF contribution does employ tin-foil boundary conditions. In non-polar cells, one typically observes well-behaved convergence of the lattice energy as the 2-body cutoff radius is increased, as expected from the long-range decay of the interaction energies.

In polar unit cells, however, the contributions from the 2-body terms proves highly erratic with cutoff radius. For example, in γ-glycine, which has a polar unit cell, the energy relative to the non-polar α form can vary by many kJ/mol depending on the specific cutoff. For this reason, results on the R05 polymorph are omitted from the present study.

C.4 Oxalyl Dihydrazide polymorphs

Table C.6 presents the relative lattice energies for the oxalyl dihydrazide polymorphs using the 0 K and room-temperature structures—the same data used to generate Figure 9 in the main paper. For MP2 and CCSD(T), the many-body treatment is computed at the periodic HF/pob-TZVP-rev2 level of theory.

Table C.6: Relative lattice energies (in kJ/mol) for the polymorphs of oxalyl dihydrazide using the fully relaxed 0 K structures or fixed-cell room-temperature structures.

	0 K Structures				
Polymorph	B86bPBE-XDM	MP2	MP2C	MP2D	CCSD(T)
α form	0.0	0.0	0.0	0.0	0.0
β form	12.0	10.1	11.6	9.2	9.2
γ form	10.4	12.7	13.4	10.9	10.4
δ form	7.8	9.8	11.2	8.8	8.0
ϵ form	4.9	9.2	10.2	8.1	8.1

	298 K Structures			
Polymorph	B86bPBE-XDM	MP2	MP2C	MP2D
α form	0.0	0.0	0.0	0.0
β form	15.8	12.8	14.6	12.1
γ form	10.4	11.1	12.1	9.5
δ form	8.3	8.5	10.0	7.7
ϵ form	5.8	7.0	8.1	6.1

Table C.7 shows how the lattice energies, H_{vib} and G_{vib} of the different polymorphs change relative to the α form between the 0 K and room temperature structures. The lattice energy contributions here amount to the differences between the 0 K and 298 K results in Table C.6. At the B86bPBE-XDM, the changes in the relative lattice energies are relative modest, except for the β form, which is destabilized by 3.8 kJ/mol. For the MP2-based methods, the β form is destabilized slightly less, by $\sim$3 kJ/mol. On the other hand, increasing temperature stabilizes the γ, δ, and ϵ forms by 1–2 kJ/mol relative to α. The temperature-dependence of the vibrational contributions to the enthalpy are relatively modest as well, with the other four forms being destabilized by 1–2 kJ/mol relative to the α form. In the end, the lattice energies and enthalpies predict the same qualitative polymorph stability orderings, with only small changes in the quantitative values, as can be seen from Figure C.7. The key temperature-dependent change in the β/γ MP2D stability ordering arises from lattice energy contributions

Table C.7: Change in the relative lattice energies, vibrational enthalpy contributions, and vibrational Gibbs free energy contributions of oxalyl dihydrazide polymorph stabilities between the 0 K and room-temperature structures, in kJ/mol. Values are listed relative to the α polymorph.

	Lattice Energy Contributions Change 0 K→298 K				B86bPBE-XDM Phonon Contrib. Change 0 K→298 K	
	B86bPBE-XDM	MP2	MP2D	MP2C	H_{vib}	G_{vib}
α form	0.0	0.0	0.0	0.0	0.0	0.0
β form	3.8	2.7	2.9	3.0	0.8	-5.0
γ form	0.0	-1.6	-1.4	-1.3	1.7	-3.7
δ form	0.5	-1.3	-1.1	-1.1	1.1	-3.0
ϵ form	0.9	-2.1	-2.0	-2.1	0.9	-1.3

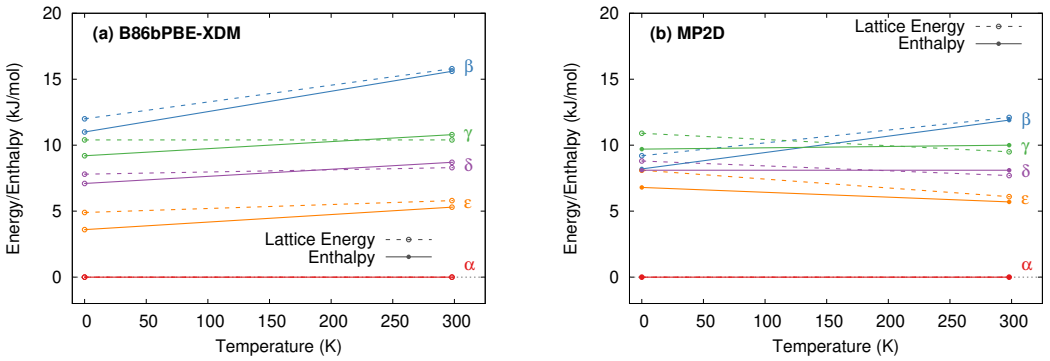


Figure C.7: Temperature dependence of the relative lattice energies (dotted lines) and enthalpies (solid lines) for the oxalyl dihydrazide polymorphs. The qualitative stability ordering is unchanged for both lattice energies and enthalpies.

C.4.1 Energy decomposition of the polymorph contributions

Table C.8 shows energy decomposition data for oxalyl dihydrazide that was used to generate Figure 11 in the main paper. MP2 and MP2D give 1-body energies in much better agreement with the CCSD(T) benchmarks than does B86bPBE-XDM. For the two-body energies, B86bPBE-XDM does extremely well in the "relative" RMS error vs. CCSD(T) (i.e. as computed from the data shown in the table, where stabilities are specified relative to the α polymorph). However, if one instead computes the "absolute" RMS error based on the total 2-body interaction energies for each form (i.e. interaction energies relative to non-interacting monomers, summed over all dimers), the DFT functional performs far worse. MP2 performs even worse than B86bPBE-XDM, while MP2C does somewhat better on the absolute RMS error. MP2D reproduces CCSD(T) extremely well, with 1.6 kJ/mol errors in both the absolute and relative RMS errors.

Table C.8: Decomposition of the relative polymorph electronic energies (kJ/mol) for oxalyl dihydrazide 0 K structures into intramolecular (1-body), pairwise intermolecular, and many-body intermolecular contributions. RMS errors are computed against the CCSD(T) benchmarks. All MP2 and CCSD(T) models employ the same periodic HF many-body treatment.

	1-Body Intramolecular Contribution			
Polymorph	B86bPBE-XDM	MP2	MP2D	CCSD(T)
α form	0.0	0.0	0.0	0.0
β form	-50.9	-58.5	-58.7	-58.2
γ form	-50.8	-57.2	-57.3	-56.4
δ form	-50.0	-52.8	-52.8	-51.0
ϵ form	-48.4	-50.4	-50.4	-49.3
RMS error	4.6	1.1	1.2	—

	2-Body Intermolecular Contribution				
Polymorph	B86bPBE-XDM	MP2	MP2C	MP2D	CCSD(T)
α form	0.0	0.0	0.0	0.0	0.0
β form	53.7	54.9	56.5	54.2	53.8
γ form	52.4	55.6	56.4	54.0	52.5
δ form	47.3	49.5	48.5	48.5	45.9
ϵ form	47.2	48.6	47.6	47.6	46.4
RMS error relative	0.8	2.7	3.8	1.6	—
RMS error absolute	7.3	8.1	4.2	1.6	—

	Many-Body Contribution	
	B86bPBE-XDM	HF
α form	0.0	0.0
β form	9.2	13.7
γ form	8.8	14.3
δ form	10.4	13.1
ϵ form	6.1	11.0

Appendix D

Supporting Information for "Inaccurate conformational energies still hinder crystal structure prediction in flexible organic molecules."

D.1 Computational Methods

D.1.1 General methods

- **DFT crystal optimizations:** Crystal structures were optimized using the B86bPBE density functional[308,309] with the exchange-hole dipole moment (XDM) dispersion correction[310] as implemented in Quantum Espresso v6.3.[316] Both the atomic coordinates and lattice parameters were relaxed during the optimization. A kinetic energy cutoff of 50 Ry was used for the wavefunctions and a cutoff of 500 Ry was used for the charge density and potential. Core electrons were treated via the projector augmented wave (PAW) approach using PAW potentials for H, C, N, O, and S generated with A. Dal Corso's Atomic code v6.1.

  The optimized crystal structures are compared against experiment using the rmsd15 metric,[388] which compares the root-mean-square deviation in non-hydrogen atom positions for a cluster of 15 molecules taken from the crystal. Fully relaxed DFT crystal structures correspond to crystals at 0 K and with no zero-point vibrational energy contributions. When comparing such DFT structures against room-temperature experimental structures, rmsd15 values of ∼0.15–0.3 are typical.

- **DFT monomer energies:** To evaluate the monomer correction, B86bPBE-XDM single-point energies were computed for each monomer in the asymmetric unit with planewave B86bPBE-XDM in a large periodic box with 20 Å of vacuum spacing in all directions. As discussed in Section D.1.2, this is typically sufficient to converge the

relative monomer conformational energies to within ∼0.1–0.2 kJ/mol. The monomer geometries were extracted directly from the crystal structures. These calculations were performed at the Γ point with the same planewave cutoff and PAW potentials as for the crystal geometry optimizations.

- **MP2D monomer energies:** Complete basis set MP2D[5] monomer energies were computed on the same monomer geometries using PSI4.[137] Density fitting was employed for both the Hartree-Fock and MP2 portions, and core electrons were frozen in the MP2 portion of the calculation. Extrapolation of the correlation energy to the complete basis set (CBS) limit was performed from the aug-cc-pVTZ and aug-cc-pVQZ basis sets using the standard two-point extrapolation approach.[70] The final MP2/CBS values combined HF/aug-cc-pVQZ with the CBS-limit correlation energies and the MP2D dispersion correction,

$$E_{MP2D}^{CBS} = E_{HF}^{aQZ} + E_{corr}^{CBS} + E_{disp}^{MP2D} \tag{D.1}$$

  Large basis sets are important for obtaining converged intramolecular conformational energies that minimize intramolecular basis set superposition error.[188,318] Note that the MP2D dispersion correction includes five global empirical parameters that were also fitted to CBS-limit benchmark data.[5]

  MP2D is freely available in PSI4. As of this writing, it can be found in the developmental versions. It should be included in future release versions. Alternatively, a free code to evaluate the stand-alone correction to MP2 can be obtained from Github:

  https://github.com/Chandemonium/MP2D

- **CCSD(T) monomer energies:** CCSD(T) benchmarks were computed for some monomer conformational energies using density-fitted frozen natural orbital CCSD(T). Complete basis set CCSD(T) monomer energies were computed via the focal point method which combines CCSD(T) in a modest basis set with MP2/CBS,

$$E_{CCSD(T)}^{CBS} = E_{MP2}^{CBS} + \left(E_{CCSD(T)}^{small} - E_{MP2}^{small}\right) \tag{D.2}$$

  The aug-cc-pVDZ basis set was used as the small basis set for ROY, while cc-pVTZ was used for *o*-acetamidobenzamide and molecule X. In MP2 basis set convergence tests on *o*-acetamidobenzamide, the cc-pVTZ basis gave results closer to the CBS limit than did aug-cc-pVDZ. For ROY, the aug-cc-pVDZ basis set (499 functions) was chosen over cc-pVTZ (670 functions) for reasons of computational cost. For the default 10^{-6} frozen natural orbital occupancy cutoff employed here, no virtual orbitals were discarded in the cc-pVTZ jobs, while about 5% of the virtual orbitals were discarded for ROY in the aug-cc-pVDZ basis.

- **Monomer-corrected crystal energies:** The monomer-corrected energies are computed as,

$$E_{crystal} = E_{crystal}^{DFT} + \sum_i \left(E_{mon,i}^{MP2D} - E_{mon,i}^{DFT}\right) \tag{D.3}$$

Table D.1: Mean absolute change and maximum change (in parentheses) in the relative B86bPBE-XDM monomer conformational energies (kJ/mol) upon increasing the minimum intermolecular vacuum spacing between periodic images. $\Delta\mu$ indicates the dipole moment range of the monomer conformations, in Debeye from B3LYP/def2-QZVP.

	$\Delta\mu$ (D)	15→20 Å	20→25 Å	25→30 Å
ROY	5.3–8.5	0.12 (0.20)	0.05 (0.08)	—
o-Acetamidobenzamide	5.2–7.4	0.17 (0.17)	0.10 (0.10)	0.04 (0.04)
Molecule X	5.2–10.8	0.14 (0.60)	0.05 (0.25)	—
Axitinib	2.5–2.9	0.07 (0.09)	—	—
Galunersitib	5.0–8.2	0.08 (0.17)	—	—

Note that while the total energy scales for the planewave DFT and MP2 differ somewhat (due to the different basis set types and pseudopotentials), this does not impact the relative crystal energies. The use of large basis sets for each component of the calculations helps minimize errors that might arise from mixing planewave and Gaussian basis set contributions with differing degrees of basis set completeness.

- **Additional monomer conformational energy calculations:** For further analysis, the performance of several other GGA and hybrid density functionals such as PBE-D3(BJ), PBE0-D3(BJ), and B3LYP-D3(BJ) are compared for various conformational energies and dihedral angle scans. These were computed with PSI4 using the def2-QZVP basis set.

D.1.2 Periodic DFT Monomer Cell Convergence

The planewave DFT monomer conformational energies depend on the size of the periodic box in which they are calculated. For each monomer, the box dimensions were determined by computing the span of the atomic coordinates along each Cartesian direction x, y, and z and then adding 20 Å of vacuum spacing to each of those values. This means that the closest two atoms between the molecule and any periodic image molecules will be at least 20 Å apart.

The box-size dependence was explored by systematically varying the amount of vacuum spacing used. The raw conformational energies vary more strongly with box size than do the relative energy differences between conformations. For example, increasing the vacuum spacing from 15 Å to 20 Å typically alters the raw energies by an average ∼0.5 kJ/mol, but the relative conformational energy changes are typically only ∼0.1–0.2 kJ/mol. Table D.1 summarizes the mean and maximum conformational energy changes that occur upon increasing the box size by 5 Å in each direction. Increasing the box size beyond 20 Å of vacuum spacing leads to only minor changes in the conformational energy that are likely smaller than other errors inherent in the calculations.

Monomer dipole moments are a key factor in determining the convergence with box size. Very polar molecules exhibit longer range interactions with their periodic images

and will therefore require larger box sizes to converge the energies to the gas-phase limit. To the extent that the dipole moments are similar across the different conformers, this box size dependence cancels somewhat. Molecule X provides an excellent example. The experimental conformation and most of the conformations in the predicted structures have molecular dipole moments $\mu \sim 10$–11 D. However, five structures have $\mu < 10$ D. The raw conformational energies of these five lower-dipole conformers converge more quickly with box size than do those for the experimental and other structures with $\mu > 10$ D (Figure D.1a). When the conformational energies are computed relative to the experimental conformation (Figure D.1b), those conformers with dipole moments similar to the experimental structure converge fastest—those structures have raw conformational energies that are similarly far from convergence, so they benefit from more error cancellation. The impact of large dipole variations on the relative conformational energies explains why Molecule X has the largest maximum errors in Table D.1. Fortunately, the energies are reasonably well converged with the 20 Å of vacuum spacing, and the five low-dipole structures which exhibit the largest box-size induced conformational energy errors have crystal energies lying ∼20–100 kJ/mol above the experimental structure (Table D.5). In other words, those five structures are experimentally irrelevant, regardless of any uncertainties in the monomer energies on the order of a few tenths of a kJ/mol.

In the end, the data presented here shows that a 20 Å vacuum spacing is typically sufficient to achieve well-converged monomer conformational energies, but that care should be taken to ensure the planewave DFT monomer calculations are converged with respect to box size when different conformations exhibit very different dipole moments, as in Molecule X. The dipole moment ranges are far smaller for the conformations that occur in the other four systems examined here (Table D.1), and none of them as polar as Molecule X.

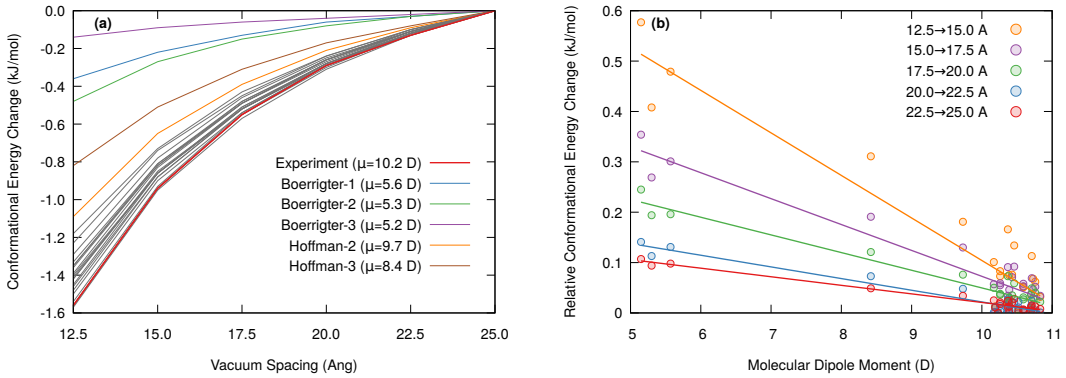


Figure D.1: (a) The conformational energies of the five conformers with dipole moments less than 10 D converge faster with the amount of vacuum spacing than do the experimental structure (red) and other structures (gray) with dipole moments above 10 D. (b) The change in the Molecule X monomer conformational energies relative to that of the experimental conformer correlates with the molecular dipole moment. The relative conformational energies converge fastest for conformations whose dipole moment is similar to the experimental one. Each color represents the relative conformational energy change upon increasing the box size by 2.5 Å along each axis. As the box vacuum spacing increases from 12.5 Å to 25 Å, the energy differences become less sensitive to dipole moment.

D.2 ROY

D.2.1 Crystal structures

Fully relaxed crystal structures for ROY were taken from Ref 188. That work optimized experimental crystal structures taken from the Cambridge Structure Database (CSD) using the same B86bPBE-XDM model as used in the present paper.

D.2.2 Monomer conformational energies

Figure D.2 compares the conformational energies for the different intramolecular conformations of the ROY molecule against CCSD(T) benchmarks. At the CCSD(T) and MP2D levels, the conformations adopted by the yellow polymorphs (Y, YT04, and YN) are considerably more stable than those in the red and orange polymorphs. The GGA functionals B86bPBE-XDM and PBE-D3(BJ) dramatically over-stabilize the red and orange ones, which is the key reason why they rank the polymorphs incorrectly. The hybrid functionals B3LYP-D3(BJ) and PBE0-D3(BJ) destabilize the red and orange conformations somewhat relative to the yellow ones, but not enough.

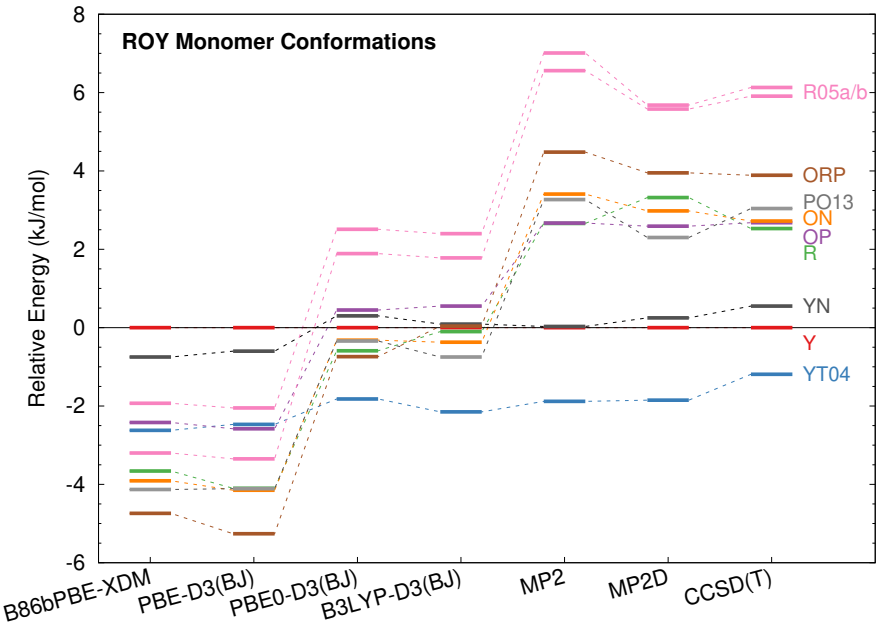


Figure D.2: Comparison of the ROY gas-phase intramolecular conformational energies computed at several levels of theory. The labels R05a/b correspond to the two symmetrically unique monomers in the R05 unit cell.

This issue can be traced directly to the potential energy curve for the dihedral angle between the two rings, as shown in Figure D.3. The GGAs substantially over-stabilize the low-angle conformations found in the red and orange polymorphs compared to CCSD(T)

benchmarks. They also place the second minimum near 120° at angles that are ∼10-15° too high. The hybrid functionals do improve upon the GGA functionals, but not enough to correct the polymorph stabilities. In contrast, MP2D gives a far more faithful representation of the potential energy curve, even if the energies are about 1 kJ/mol higher than CCSD(T) near ∼40°.

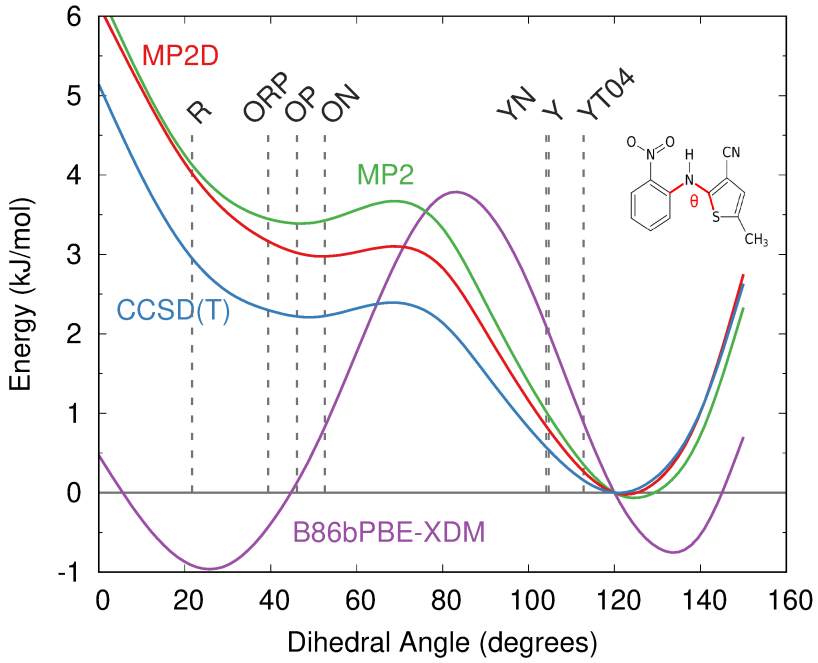


Figure D.3: Relaxed 1-D potential energy scan along the key ROY dihedral angle θ that governs the conformational polymorphism. Dotted vertical lines indicate the value of this dihedral angle adopted in various ROY polymorphs.

D.2.3 Polymorph energies

Table D.2 summarizes the relative polymorph energies used to generate Figure 2 in the main paper. It highlights how the monomer-correction is large for the orange and red polymorphs (5–9 kJ/mol), but quite small for the Y polymorphs (1 kJ/mol), as anticipated from the data presented in Section D.2.2.

Table D.2: Final ROY polymorph 0 K electronic energies relative to Form Y (kJ/mol). The "MP2D correction" is the correction applied to the B86bPBE-XDM polymorph energies, while "B86bPBE-XDM + ΔMP2D" indicates the final corrected polymorph energies. Finite-temperature experimental enthalpies were taken from Refs 6–9.

Polymorph	B86bPBE-XDM	B86bPBE-XDM + ΔMP2D	MP2D correction	Experimental ΔH
Y	0.00	0.00	0.00	0.00
YT04	-0.40	0.38	0.77	0.91
R	-6.35	0.63	6.98	1.40
OP	-3.81	1.20	5.01	1.87
ON	-4.32	2.56	6.89	2.58
YN	0.92	1.92	1.00	2.99
ORP	-5.39	3.31	8.69	4.09
R05	-6.42	1.77	8.19	
PO13	-4.41	2.03	6.44	

To connect with our earlier work,[188] Figure D.4 compares the polymorph energies computed using crystals that were optimized at the fixed experimental lattice parameters. That work computed full HMBI MP2D energies for those structures; here monomer-corrected MP2D and CCSD(T) results are added. The discrepancies between the monomer-correction models and a full HMBI treatment that also treats the pairwise intermolecular interactions with MP2D and many-body interactions with periodic HF are clear: The monomer-corrected model predicts a narrower energy range for the polymorphs and actually gives results closer to experiment. On the other hand, these electronic energies lack the vibrational contributions one would want to include when comparing to experimental enthalpies. This makes it difficult to state for certain which model is more faithful to the experimental data. Crucially, however, the energy orderings from the monomer-corrected and HMBI results are both far more consistent with experiment than are those from the DFT functionals.

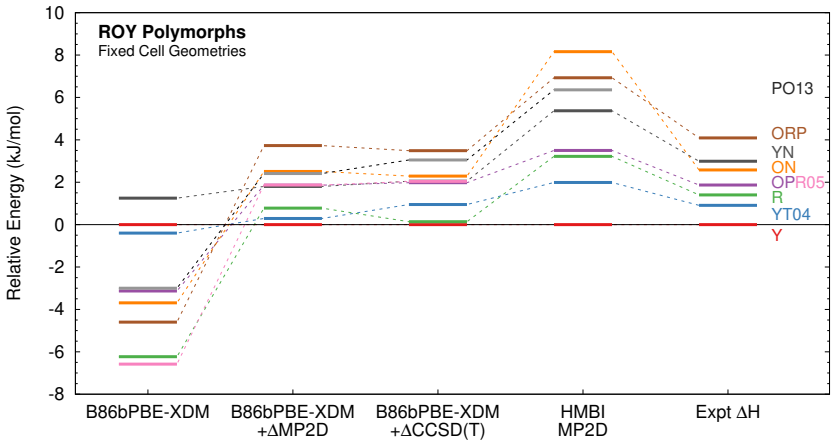


Figure D.4: 0 K electronic energies of the ROY polymorphs relative to Form Y, computed using crystal structures optimized with the fixed experimental lattice parameters, in kJ/mol. These are compared against experimental enthalpies measured at finite temperatures. The B86bPBE-XDM monomer calculations here used a 15 Å vacuum spacing instead of the 20 Å spacing used elsewhere in this work.

D.3 *o*-Acetamidobenzamide

D.3.1 Crystal structures

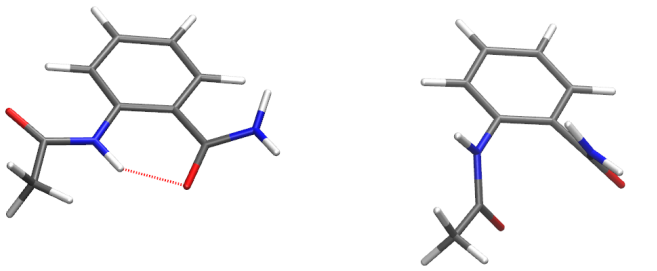
Fully relaxed crystal structures for *o*-acetamidobenzamide were taken from Ref 188. That work optimized experimental crystal structures taken from the CSD using the same B86bPBE-XDM model as used in the present paper.

D.3.2 Monomer conformational energies

Table D.3 compares the energy differences between the α and β form monomer conformations. Because the α form exhibits an intramolecular hydrogen bond while the β form does not, the conformational energy differences exceed 50 kJ/mol. All four density functionals considered over-stabilize the α form with its extended π conjugation by ∼5–8 kJ/mol relative to CCSD(T) benchmarks. In contrast, MP2D reproduces the CCSD(T) conformational energy difference to within 0.4 kJ/mol.

Table D.3: Electronic energy difference between the α and β polymorph conformations of *o*-acetamidobenzamide (kJ/mol).

Method	$\Delta E = E_\beta - E_\alpha$	Error vs CCSD(T)
B86bPBE-XDM	58.18	5.89
PBE-D3(BJ)	59.81	7.52
PBE0-D3(BJ)	60.60	8.31
B3LYP-D3(BJ)	57.24	4.95
MP2D	52.65	0.36
CCSD(T)	52.29	



α acetamidobenzamide β acetamidobenzamide

D.3.3 Polymorph energies

The polymorph energetics of *o*-acetamidobenzamide were studied in detail with HMBI-based MP2D calculations in our previous work.[188] It is known that the β form is more stable at high temperatures, with α transforming to it exothermically around 150°C (423

K) with ΔH = -1.9 or -2.9 kJ/mol. The stability ordering at 0 K is unclear experimentally, but it is believed that the two structures are close in energy at 0 K.

Accordingly, our previous work in Ref 188 approximately modeled the temperature dependence in this system by (1) fully relaxing the crystal structures to obtain 0 K structures and (2) performing fixed-cell optimizations of the room-temperature experimental structures to mimic 298 K structures. Electronic energy differences $\Delta E = E_\beta - E_\alpha$ were computed on each set of structures. Then B86bPBE-XDM Γ-point harmonic phonons were computed to evaluate the vibrational contributions for the 0 K and 298 K structures. Finally, the ΔH values at 0 K and 298 K were (crudely) linearly extrapolated to 423 K to estimate the ΔH(423K). For HMBI MP2D calculations, this process produced an enthalpy difference in quantitative agreement with the experimental values (Table D.4). In contrast, B86bPBE-XDM erroneously prefers the α form by nearly 6 kJ/mol, an error of 8–9 kJ/mol relative to experiment. See Ref 188 for details.

The present work repeats this procedure using monomer-corrected B86bPBE-XDM results (Table D.4). Performing the monomer correction with MP2D reduces ΔE(0K) from having the α form be 5.84 kJ/mol more stable than β to it being only 0.31 kJ/mol more stable than β. Zero-point vibrational energy preferentially stabilizes the β form, such that it becomes enthalpically preferred by 0.97 kJ/mol at 0 K. Repeating the process for 298 K and extrapolating to 423 K, we obtain ΔH(423K) = -0.85 kJ/mol, which is reasonably close to the experimental values.

Further validation for the monomer-corrected MP2D results comes from performing the monomer correction with CCSD(T) instead. As shown in Table D.4, the enthalpies exhibit the same qualitative trends and differ from the MP2D-corrected results by no more than a few tenths of a kJ/mol. It is worth noting, however, that whereas HMBI MP2D predicts that the enthalpy gap between the two polymorphs increases with temperature, the DFT and monomer-corrected DFT results here predict that the enthalpy gap narrows slightly with temperature. Neither behavior can be confirmed experimentally, but it seems likely that the HMBI value is closer to the truth, and that this reflects errors in the B86bPBE-XDM intermolecular interactions that are not addressed by the monomer correction.

Table D.4: Summary of the electronic energy and enthalpy differences between the α and β polymorphs of acetamidobenzamide (kJ/mol). The 0 K values use fully relaxed crystal structures, while the 298 K values used fixed-cell optimizations of the room-temperature crystal structures. The enthalpy differences are linearly extrapolated to 423 K for comparison with experiment. A positive value here indicates the α polymorph is more stable, while a negative value means the β polymorph is more stable.

Method	ΔE (0K)	ΔE (298K)	ΔH (0K)	ΔH (298K)	ΔH^a (423K)
B86bPBE-XDM (Ref 188)	5.84	5.19	4.56	5.51	5.91
B86bPBE-XDM + ΔMP2D	0.31	-1.20	-0.97	-0.89	-0.85
B86bPBE-XDM + ΔCCSD(T)	-0.05	-1.30	-1.33	-0.98	-0.83
HMBI MP2D (Ref 188)	1.44	-1.69	0.16	-1.37	-2.02
Experiment (Refs 253,307)					-1.9, -2.9

[a] Linearly extrapolated from ΔH(0K) and ΔH(298K).

D.4 Molecule X

D.4.1 Crystal structures

Candidate crystal structures for molecule X were obtained from the authors of Ref 237. They were optimized using a similar B86bPBE-XDM protocol as was used here, albeit with a larger 80 Ry planewave cutoff and a (4,4,4) **k**-point grid.

D.4.2 Monomer conformational energies

Figure D.5 shows monomer conformational energies for several of the low-energy crystal structures relative to the experimental structure. While the various DFT functionals mostly capture the correct energy ordering for the different molecular conformations (except for the Day-1 conformation, which is slightly less stable than the experimental conformation), the quantitative GGA errors relative to the CCSD(T) benchmarks are 2–3 times larger than the MP2D ones (GGAs: RMSE $\sim$ 1.6-1.7 kJ/mol, hybrid $\sim$1.0–1.4 kJ/mol, and MP2D 0.5 kJ/mol)

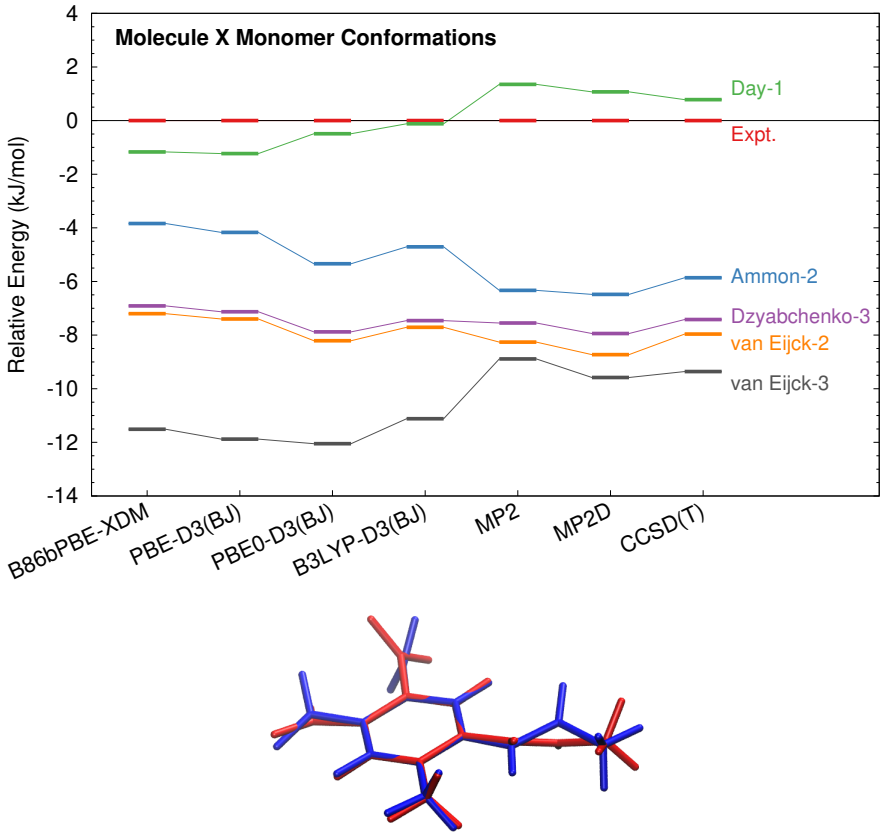


Figure D.5: Comparison of the Molecule X gas-phase intramolecular conformational energies computed at several levels of theory. The overlay compares the experimental (red) and vanEijck-3 (blue) conformations.

The overlay in Figure D.5 compares the molecular conformations in the experimental and vanEijck-3 structures. The vanEijck-3 structure (blue) is artificially stabilized by the DFT structures because the amide group and the nitro para to it are nearly planar with the ring, extending the π conjugation. In contrast, the experimental conformation (red) rotates these groups out of the plane, disrupting the conjugation.

D.4.3 Polymorph energies

Table D.5 summarizes the relative crystalline energies for several of the candidate structures of Molecule X. The B86bPBE-XDM and B86bPBE-XDM + ΔMP2D monomer correction values are the same as those used to generate Figure 3 in the main paper. Note that the smaller monomer correction found here for vanEijck-3 compared to that in Ref 237 arises in roughly equal parts from our use of the larger basis set and the dispersion correction.

As noted in the main paper, HMBI calculations were performed on the Experimental and vanEijck-3 structures to compare against the monomer-corrected values. These

found the vanEijck-3 structure to lie 0.11 kJ/mol above the experimental structure, which is very close to the monomer-corrected value of 0.17 kJ/mol in Table D.5. These calculations combined MP2D/CBS monomer and dimer contributions plus a periodic HF/pob-TZVP-rev2 treatment of the long-range and many-body contributions, the same procedures as were used in Ref[188].

Table D.5: Final 0 K electronic energies for molecule X with respect to the experimental structure (kJ/mol).

Structure	B86bPBE-XDM	B86bPBE-XDM + ΔMP2D	MP2D Correction
Expt	0.00	0.00	0.00
vanEijck-3	-1.76	0.17	1.93
Dzyabchenko-3	2.65	1.62	-1.03
vanEijck-2	3.38	1.84	-1.54
Ammon-2	4.50	1.85	-2.65
Ammon-3	4.50	2.99	-1.52
Day-2	4.30	5.11	0.82
Day-1	3.23	5.48	2.25
vanEijck-1	7.44	5.51	-1.93
Ammon-1	8.53	5.61	-2.92
Erk-3	8.41	6.02	-2.39
Erk-2	8.40	6.03	-2.37
Erk-1	10.95	9.05	-1.90
Day-3	5.43	9.47	4.04
Dzyabchenko-2	9.95	11.28	1.32
Dzyabchenko-1	8.79	11.44	2.64
Liang-1	15.23	13.13	-2.10
Liang-2	16.82	14.73	-2.09
Liang-3	18.74	17.23	-1.51
Boerrigter-1	18.15	20.65	2.49
Boerrigter-2	21.23	22.61	1.38
Boerrigter-3	31.14	30.22	-0.91
Hofmann-1	37.55	37.73	0.18
Hofmann-3	71.58	69.26	-2.32
Hofmann-2	97.86	97.13	-0.73

D.5 Axitinib

D.5.1 Crystal structures

Experimental crystal structures for axitinib were taken directly from the CSD and were fully relaxed with B86bPBE-XDM according to the procedures described in Section D.1 above. The experimental crystal structures were all determined at room temperature except for form I, which was measured at 213 K.[335,336] The rmsd15 values of 0.17–0.25 Å between the optimized and experimental structures are typical for comparison of 0 K DFT structures against experimental structures at higher temperatures. CIF files containing the optimized geometries are provided separately.

Table D.6: Details of the axitinib crystal structures used and rmsd15 values comparing the 0 K DFT-optimized and finite-temperature experimental structures.

Polymorph	RefCode	Space Group	Z / Z'	**k**-point grid	rmsd15
Form XLI	VUSDIX04	$P2_1/c$	4 / 1	1×5×1	0.170 Å
Form XXV	VUSDIX	$P2_1/c$	4 / 1	9×3×1	0.223 Å
Form VI	VUSDIX03	$P\bar{1}$	2 / 1	5×3×3	0.239 Å
Form IV	VUSDIX01	$P\bar{1}$	4 / 2	3×3×1	0.252 Å
Form I	VUSDIX06	$P\bar{1}$	2 / 1	5×3×3	0.230 Å

D.5.2 Monomer conformational energies

Figure D.6 compares the monomer energies for several different functionals against those computed with MP2 and MP2D. When compared against MP2D, the two GGAs and the hybrid PBE0 functional all under-stabilize the XLI form relative to the others to varying degrees. B3LYP-D3(BJ) actually ranks the axitinib conformational energies fairly well, unlike for ROY, *o*-acetamidobenzamide, and Molecule X. The conformational energy for monomer IVb is the only one which exhibits appreciable disagreement between B3LYP-D3(BJ) and MP2D.

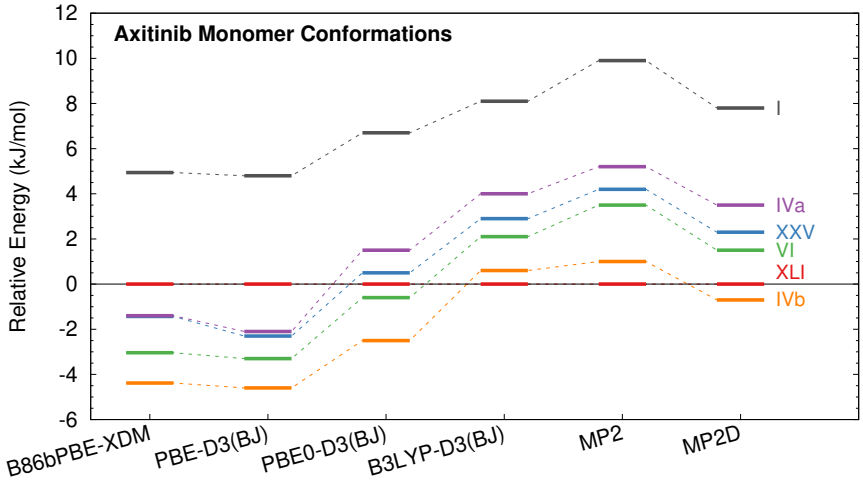


Figure D.6: Comparison of the axitinib gas-phase intramolecular conformational energies computed at several levels of theory. The labels IVa and IVb correspond to the two symmetrically unique monomers in the Form IV unit cell.

D.5.3 Polymorph energies

Table D.7 summarizes the relative B86bPBE-XDM polymorph energies, the revised energies after applying the MP2D monomer correction (B86bPBE-XDM + ΔMP2D), and the size of the MP2D correction. This data was used to generate Figure 4 in the main paper. The monomer correction destabilizes the other four polymorphs by an average of 3.8 kJ/mol relative to form XLI. The smallest correction of 2.8 kJ/mol occurs for form I. Similar to form XLI, form I also disrupts the π conjugation between the amide group and the benzene ring, so the relative DFT error between forms XLI and I would be expected to be smaller than compared to the other three forms for which DFT artificially stabilizes the extended π conjugation.

Table D.7: Final 0 K electronic energies for the axitinib polymorphs relative to form XLI (kJ/mol).

Polymorph	B86bPBE-XDM	B86bPBE-XDM + ΔMP2D	MP2D correction
Form XLI	0.00	0.00	0.00
Form XXV	-2.81	0.93	3.74
Form VI	-1.70	2.80	4.51
Form IV	0.85	5.17	4.32
Form I	5.75	8.60	2.85

D.6 Galunisertib

D.6.1 Crystal structures

The putative GM structure was taken from the CSP2 structures provided in supporting information of Ref 251. All others were taken from the CSD. The structures were fully relaxed with B86bPBE-XDM according to the procedures described in Section D.1. Form I was omitted here because it lies much higher in energy than the other forms.[251] Experimental structures for forms II, III, IV, VI, and X were determined at 100 K, while the others were solved at room temperature. Once again, agreement between the DFT and experimental structures is fairly good. As one would expect, there is closer agreement for the structures solved experimentally at 100 K (rmsd15 values ∼0.13–0.17 Å) than those solved at room temperature (rmsd values ∼0.22–0.35 Å). The agreement between the B86bPBE-XDM optimized structure of the GM form and the one reported in Ref 251 is also reasonable, with rmsd15 of 0.14 Å. CIF files containing the optimized geometries are provided separately.

Table D.8: Details of the galunisertib crystal structures used and rmsd15 values comparing the 0 K DFT-optimized and finite-temperature experimental structures.

Polymorph	RefCode	Space Group	Z / Z'	**k**-point grid	rmsd15
GM	CSP2	$P2_12_12_1$	4 / 1	3×1×1	0.141 Å[a]
Form II	DORDUM	$P\bar{1}$	2 / 1	3×1×1	0.138 Å
Form III	DORDUM01	$P\bar{1}$	2 / 1	3×3×1	0.148 Å
Form IV	DORDUM02	$P2_1/n$	8 / 2	1×3×1	0.170 Å
Form V	DORDUM03	$P2_1/c$	4 / 1	3×1×1	0.254 Å
Form VI	DORDUM04	$P2_1/n$	4 / 1	1×1×1	0.125 Å
Form VII	DORDUM10	$P\bar{1}$	4 / 2	3×1×1	0.353 Å
Form VIII	DORDUM05	$P\bar{1}$	2 / 1	3×1×1	0.246 Å
Form IX	DORDUM06	$P2_1/c$	4 / 1	1×1×3	0.222 Å
Form X	DORDUM07	$C2/c$	8 / 1	3×1×1	0.161 Å

[a] The rmsd15 value for the GM structure compares to the DFT-predicted structure from CSP2 in Ref 251, rather than against experiment.

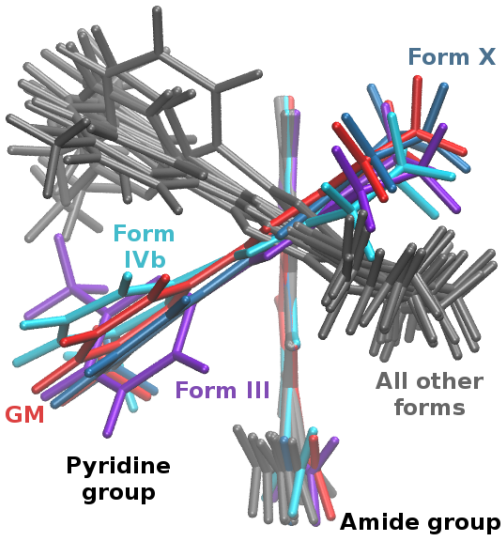
D.6.2 Monomer conformational energies

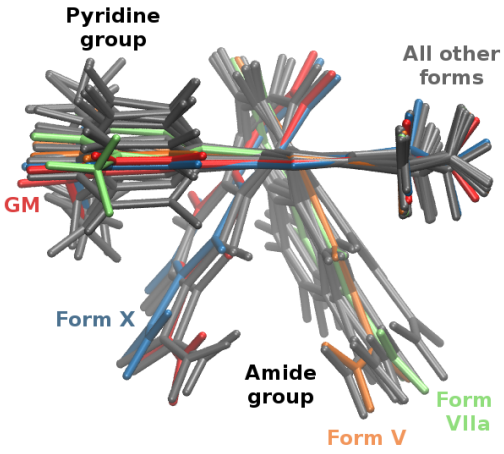
Figure D.7 overlays the galunisertib conformations from the different crystal forms in two ways and tabulates the values of the three key dihedral angles. The first overlay aligns the molecules based on the quinoline ring. It highlights how the GM form adopts a relatively rare value of dihedral angle 1, shared only by monomers III, IVb, and X. It also highlights the relatively narrow range of amide group torsion angles, though the GM structure adopts the most planar amide. The second overlay aligns the molecules based on the pyrazole ring. It is oriented such that the horizontal line through the pyridine group indicates a planar value of dihedral 3, which is also shared by a handful of the monomers (Forms V, X, and

Form VIIa with a 180-degree rotation). While the actual impact of the monomer correction depends on the interplay of these three dihedrals and other subtle geometry changes, it is clear that the GM structure has a relatively unique geometry with extended π conjugation.

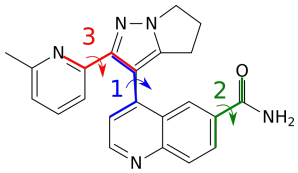
Figure D.8 plots the gas-phase intramolecular conformational energies for the different polymorphs as computed with various methods. While the monomer stabilities do not reorder as extensively between the DFT functionals and MP2D compared to the earlier systems, MP2D stabilizes all other forms relative to the GM conformation by an average 2.2 kJ/mol (max 4.0 kJ/mol for form III). This occurs because the GM form exhibits some of the most planar conformations for dihedral angles 2 and 3, which causes it to be overstabilized by delocalization error in the density functionals.

Figure D.9 plots relaxed 1-D conformational energy profiles along the three key dihedrals. The discontinuities along some of the scans result from sudden conformational changes that occur in other parts of the molecule as the coordinate is scanned. While there are no glaring failures for the DFT functionals along these three scans, it is clear that there are often differences of a few kJ/mol between MP2D and the various functionals for relevant values of the scan coordinates. Torsional angle 1 in Figure D.9 contributes the largest discrepancies. These results are consistent with the method-dependent conformational energy variations observed in Figure D.8.





	Dihedral 1	Dihedral 2	Dihedral 3
GM	120.1	-4.8	7.1
Form II	49.3	11.9	27.2
Form III	129.3	-19.2	-30.9
Form IVa	56.1	12.9	40.4
Form IVb	113.4	-12.8	20.0
Form V	56.8	26.2	-2.7
Form VI	54.3	6.9	26.2
Form VIIa	62.1	7.8	166.7
Form VIIb	66.2	26.2	-13.8
Form VIII	63.2	23.6	-15.2
Form IX	42.1	8.9	41.9
Form X	128.8	6.9	4.5



Galunisertib

Figure D.7: (top) Two different overlays of the galunisertib monomers. The left one aligns them based on the quinoline ring, while the right one aligns along the pyrazole ring. Molecules in blue have similar values of dihedral angles 1 (left overlay) or 3 (right overlay). (bottom) Summary of the key dihedral angles found for the different galunisertib crystal structures (using DFT-optimized geometries). Color coding from green (most similar) to red (least similar) is used to cluster angles by how much they differ from the GM angles in terms of planarity.

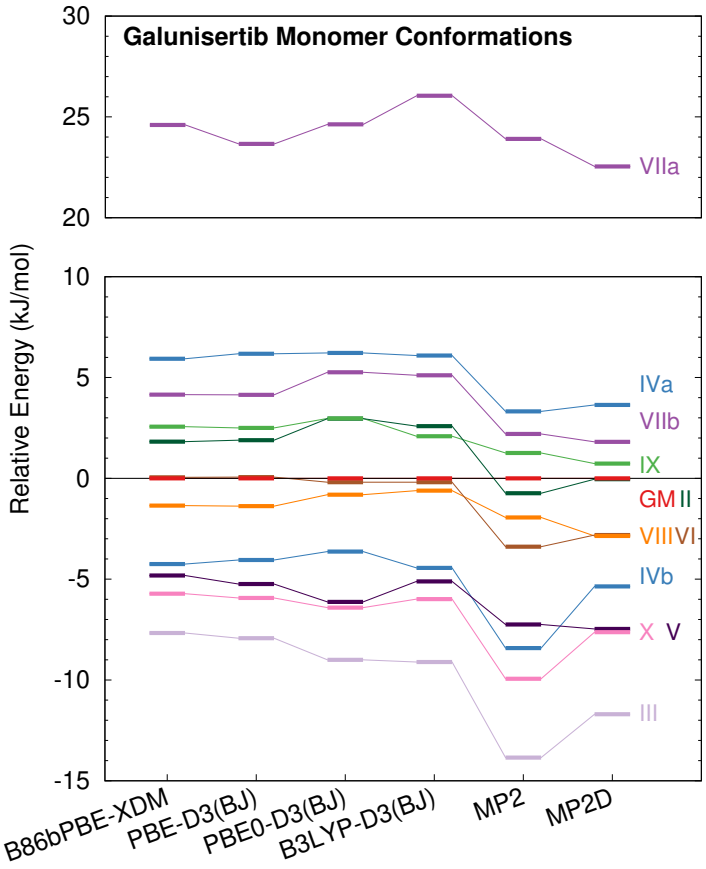


Figure D.8: Comparison of the galunisertib gas-phase intramolecular conformational energies computed at several levels of theory. The plot is split into two portions for viewing convenience. The labels IVa, IVb, VIIa, and VIIb correspond to the symmetrically unique monomers in the Form IV and Form VII unit cells.

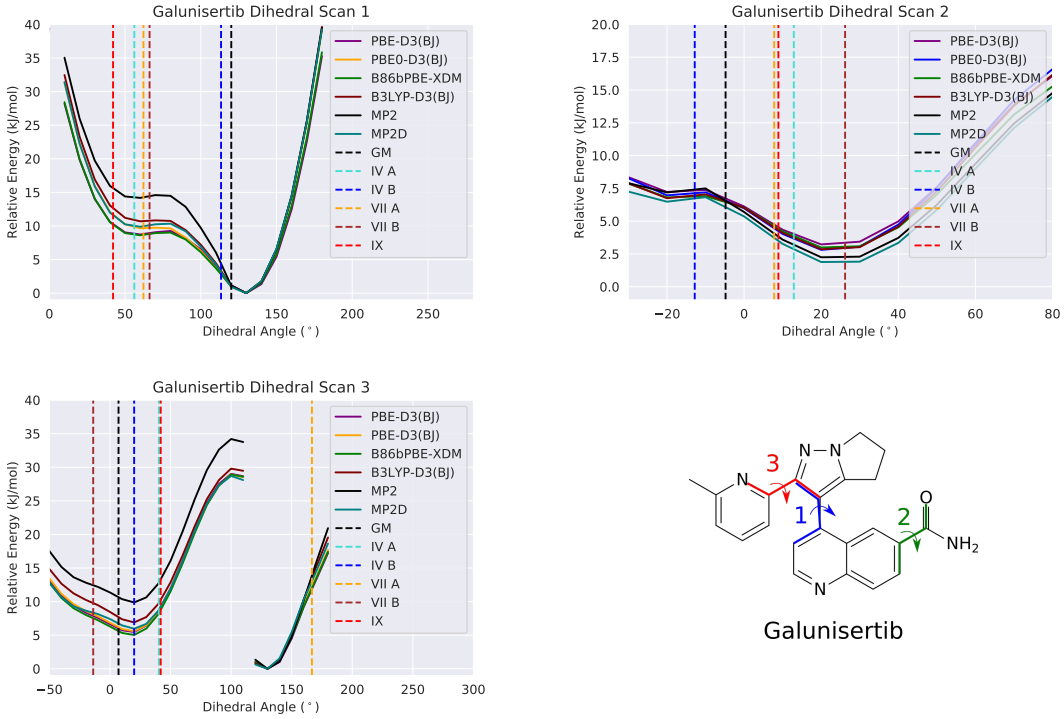


Figure D.9: Relaxed 1-D potential energy curve scans for the three key dihedral angle scans for galunisertib. Vertical lines indicate the values of that particular dihedral angle that are adopted in the crystal structures. Note that the curves exhibit discontinuities due to sudden changes in the other flexible degrees of freedom that occur as the primary coordinate is scanned.

D.6.3 Polymorph energies

Table D.9 summarizes the relative B86bPBE-XDM polymorph energies, the revised energies after applying the MP2D monomer correction (B86bPBE-XDM + ΔMP2D), and the size of the MP2D correction. This data was used to generate Figure 5 in the main paper. The MP2D monomer correction relative to form IV is less than 1 kJ/mol for many of the polymorphs, but it has particularly large ∼2 kJ/mol impacts on the GM structure and form III.

Table D.9: Final 0 K electronic energies for galunisertib polymorphs relative to form IV (kJ/mol).

Polymorph	B86bPBE-XDM	B86bPBE-XDM + ΔMP2D	MP2D correction
Form GM	-5.06	-3.36	1.70
Form II	6.08	5.91	-0.16
Form III	5.84	3.51	-2.34
Form IV	0.00	0.00	0.00
Form V	0.23	-0.71	-0.94
Form VI	1.99	0.83	-1.16
Form VII	-3.63	-4.14	-0.50
Form VIII	-2.41	-2.22	0.19
Form IX	-4.03	-4.16	0.13
Form X	-1.65	-1.86	-0.21

Appendix E

Supporting Information for "Rubrene untwisted: common density functional theory calculations overestimate its deviant tendencies."

E.1 Gas-phase twisting energies

Table E.1 lists the root-mean-square errors for each method across the twisting potential energy scans corresponding to Figure 2 in the main paper. For comparison with earlier studies,[353,354] Table E.2 supplements Table 1 from the main paper with twisting energies computed with three additional density functionals: ωB97, ωB97X-D, and M06-L. All three perform notably worse than SCS-MP2D relative to the DLPNO-CCSD(T) benchmarks.

Table E.1: Root-mean-square errors for the rubrene derivative twisting potential energy scans relative to the benchmark DLPNO-CCSD(T) energies, in kJ/mol.

System	PBE-D3(BJ)	B86bPBE-XDM	B3LYP-D3(BJ)	SCS-MP2D
Rubrene	3.14	3.14	2.81	0.26
F_{14}-Rubrene	3.96	4.46	3.62	0.70
Perfluororubrene	4.53	4.88	3.22	0.94
All three	**3.9**	**4.2**	**3.2**	**0.7**

Table E.2: Predicted gas-phase twisting energies (kJ/mol) for rubrene and several derivatives as computed with three additional density functionals and the RMSE relative to the DLPNO-CCSD(T) benchmarks.

Species	ωB97[a] cc-pVDZ	ωB97X-D def2-QZVP	M06-L def2-QZVP	SCS-MP2D CBS	DLPNO-CCSD(T) CBS
1	-11.3	-10.5	-13.7	-9.4	-9.4
2	-11.8	-10.4	-13.8	-9.1	-8.9
3	-10.7	-10.8	-12.7	-8.6	-8.4
4	-10.3	-9.4	-13.3	-8.4	-8.3
5	-11.2	-8.7	-12.3	-7.5	-7.5
RMSE	2.6	1.5	4.7	0.2	

[a] Ref 353

E.2 Delocalization error

To characterize the delocalization error associated with rubrene twisting, Figure E.1 plots two electronic properties of tetracene for the BLYP family of density functionals as a function of the exact exchange. Figure E.1a examines the Mulliken charges (computed with the 6-31G(d) basis set) for the symmetrically unique carbon atoms. Similar to what was seen for the twisting energies (Figure 3) in the main paper, the charges vary nearly linearly with the increasing amount of exact exchange included in the functional. Figure E.1b plots the change in the HOMO and LUMO orbital energies, which also exhibit the expected variation with increasing amounts of exact exchange.

Moreover, twisting tetracene has a negligible impact on these properties. The mean absolute atomic charge change is only 0.003 a.u. between the planar and twisted conformations. Similarly, twisting alters the HOMO and LUMO orbital energies by 0.03 eV or less. In other words, the twisting does not appear to impact the extent of π conjugation significantly.

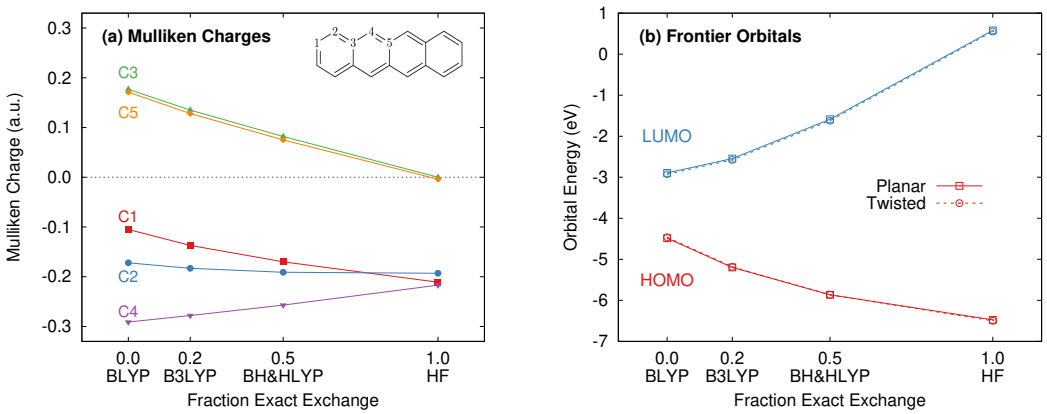


Figure E.1: Impact of increasing the fraction of exact exchange in the functional on the (a) Mulliken charges for the carbon atoms and (b) frontier orbital energies for the tetracene backbone.

E.3 Solid state polymorph energies

Table E.3 lists the relative polymorph energies for perfluororubrene and rubrene before and after monomer correction. These energies match Figure 5 in the main paper. Table E.4 lists the Monkhorst-Pack k-point grids used in the periodic DFT calculations.

Table E.3: Relative polymorph energies for Perfluororubrene and Rubrene, in kJ/mol. The B86bPBE-XDM results were computed with periodic DFT, while the monomer-corrected values represent the B86bPBE-XDM results after correcting the monomer conformational energies with the indicated level of theory (as denoted with a Δ symbol).

	Perfluororubrene		Rubrene		
Method	Planar	Twisted	Orthorhombic	Monoclinic	Triclinic
Periodic DFT					
B86bPBE-XDM	0.0	-3.51	0.0	1.22	1.32
Monomer-corrected					
+ΔPBE-D3(BJ)	0.0	-3.30	–	–	–
+ΔB3LYP-D3(BJ)	0.0	-3.01	–	–	–
+ΔSCS-MP2D	0.0	3.11	0.0	1.05	-0.28
+ΔDLPNO-CCSD(T)	0.0	3.61	–	–	–

Table E.4: Monkhorst-Pack k-point grids used for the periodic DFT calculations.

Polymorph	CSD RefCode	K-point grid
Perfluororubrene		
Planar	INELUK02	(3,3,1)
Twisted	INELUK03	(1,3,1)
Rubrene		
Orthorhombic	QQCIG07	(1,5,1)
Monoclinic	QQCIG13	(3,3,1)
Triclinic	QQCIG14	(5,3,3)