DEVELOPMENT AND APPLICATIONS OF COMPOSITE AND LOW-COST APPROACHES IN MOLECULAR CRYSTAL STRUCTURE PREDICTION

by

Luc M. LeBlanc

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy

at

Dalhousie University
Halifax, Nova Scotia
June 2019

Спектакль окончен.

CONTENTS

List of Tables . . . . . **vii**

List of Figures . . . . . **xi**

Abstract . . . . . **xiii**

List of Abbreviations and Symbols Used . . . . . **xiv**

Acknowledgements . . . . . **xxi**

Chapter 1 Introduction **1**

1.1 Molecular Crystal Structure Prediction 1

1.2 Crystal Structure Prediction Methodologies 6

1.2.1 Overview of Structure-Generation Methods 6

1.2.2 Overview of Energy-Ranking Methods 8

1.3 Thesis Goals 11

Chapter 2 Theoretical Background and Methodological Details **14**

2.1 Dispersion-Corrected Periodic-Boundary Density-Functional Theory . . . 14

2.1.1 Overview of Density-Functional Theory 14

2.1.2 Treatment of Periodic Solid-State Systems 18

2.2 Accounting for London Dispersion Within the Density-Functional Theory Framework 21

2.2.1 Post-SCF Dispersion Correction Models 22

2.2.2 Non-Local van-der-Waals Density Functional Models 29

2.3 Low-Cost Methods as Alternatives to Plane-Wave Density-Functional Theory Methods 30

2.3.1 Density-Functional Tight-Binding Methods 31

2.3.2 Minimal-Basis-Set Approaches 33

Chapter 3 Evaluation of Shear-Slip Transitions in Crystalline Aspirin by Density-Functional Theory 35

3.1 Motivation 35

3.2 Computational Methods 38

3.3 Results and Discussion 40

3.3.1 Relative Energetics 40

3.3.2 Slip Mechanism 43

3.3.3 Applied Pressure 46

3.4 Conclusions 48

Chapter 4 Pervasive Delocalisation Error Causes Spurious Proton Transfer in Organic Acid-Base Co-Crystals 49

4.1 Motivation 49

4.2 Computational Methods 54

4.3 Results and Discussion 55

4.4 Conclusions 59

Chapter 5 Composite and Low-Cost Approaches for Molecular Crystal Structure Prediction 62

5.1 Motivation 62

5.2 Computational Methods 65

5.3 Results and Discussion 70

5.3.1 Absolute Lattice Energies 70

5.3.2 Relative Lattice Energies 73

5.3.3 Crystal Structure Prediction 79

5.4 Conclusions 83

Chapter 6 Non-Covalent Interactions in Molecular Crystals: Exploring the Accuracy of the Exchange-Hole Dipole Moment Model with Local Orbitals 85

6.1 Motivation 85

6.2 Computational Methods 87

6.3 Results and Discussion 88

6.3.1 The XDM Dispersion Model: Implementation, Parametrization, and Testing 88

6.3.2 Graphite and Phosphorene Exfoliation 93

6.3.3 Lattice Energies of Molecular Crystals 97

6.4 Conclusions 100

Chapter 7 Crystal-Energy Landscapes of Active Pharmaceutical Ingredients Using Composite Approaches 102

7.1 Motivation 102

7.2 Computational Methods 105

7.3 Results and Discussion 107

7.3.1 5-Fluorouracil 107

7.3.2 Naproxen 112

7.3.3 Carbamazepine 115

7.3.4 Olanzapine 116

7.4 Conclusions 121

Chapter 8 Conclusion 124

Appendix A Supporting Information for Chapter 5 128

Appendix B Supporting Information for Chapter 6 141

Appendix C Supporting Information for Chapter 7 147

Bibliography **156**

LIST OF TABLES

2.1 Cutoff radii used in the generation of the projector-augmented wave datasets for the elements present in the molecular crystals studied in this thesis. 21

3.1 Lattice energies of form-I aspirin calculated using various theoretical methods. 41

3.2 Unit-cell parameters, including cell lengths, angles, volumes and temperatures for the two crystal forms of aspirin. 41

3.3 Relative energies of the two forms of aspirin using selected levels of theory. 42

3.4 Convergence of the vibrational free-energy difference between aspirin-I and aspirin-II at 298 K, with respect to phonon sampling. 43

4.1 Structures predicted by selected plane-wave dispersion-corrected density-functional approximations. 56

4.2 Structures predicted by geometry optimization of the co-crystal structures with plane-wave calculations. 56

4.3 Experimental and optimized C–O bond lengths for the co-crystals studied. 58

4.4 Relative PBEh-3c and HSE-3c energies for the co-crystal and salt forms, together with literature pKa values. 58

4.5 Structures predicted from geometry optimization of the salt structures with selected density-functional approximations. 59

5.1 Statistics for the X23 set of lattice energies using various computational methods. 71

5.2 The mean absolute error for the X23 set using bare low-cost, and composite, methods. 73

5.3 Statistics for the EE14 set of relative lattice energies using various low-cost methods. 78

5.4 Relative energy differences, unit-cell parameters, and volume of the lowest-energy racemate structure from the crystal structure prediction study of 1-aza-helicene. 81

6.1 XDM damping parameters for the PBE and B86bPBE functionals, and selected basis sets, along with the resulting error statistics for the fit set. 91

6.2 Results of geometry optimizations on graphite performed with DFT-D/DZP using various input values of the c lattice parameter, as a function of the stress convergence threshold. 97

6.3 Statistics for the X23 set of lattice energies using DFT-XDM and DFT-D2 methods relative to back-corrected experimental sublimation enthalpies. 98

6.4 Statistics for the X23 set of lattice energies using composite methods, relative to plane-wave calculations with the same density-functional approximation. 100

7.1 Relative energies for the experimental polymorphs of carbamazepine computed with low-level and composite approaches. 116

7.2 Relative energies for the experimental polymorphs of olanzapine computed with low-level and composite approaches. 122

A.1 Lattice energies of the X23 benchmark calculated with the PBE-D2 method, as implemented in the SIESTA code. 129

A.2 Lattice energies of the X23 benchmark calculated with low-cost methods, as implemented in the SIESTA, DFTB+ and CRYSTAL17 codes. 130

A.3 Powder diffraction structure measures of X23 crystal geometries generated by low-cost methods, as implemented in the SIESTA and DFTB+ codes. 131

A.4 Powder diffraction structure measures of X23 crystal geometries generated by low-cost methods, as implemented in the CRYSTAL17 code. 132

A.5 Lattice energies of the X23 benchmark calculated with plane-wave DFT methods implemented in the Quantum ESPRESSO code. . . . 133

A.6 Powder diffraction structure measures of X23 crystal geometries generated by plane-wave DFT methods implemented in the Quantum ESPRESSO code. 134

A.7 Predicted energy differences between the enantiopure and racemate crystal structures and enantiomeric excesses of the EE14 benchmark set calculated with low-cost methods using the DZP basis set and the SIESTA code. 135

A.8 Predicted energy differences between the enantiopure and racemate crystal structures and enantiomeric excesses of the EE14 benchmark set calculated with the DFTB+ code. 136

A.9 Predicted energy differences between the enantiopure and racemate crystal structures and enantiomeric excesses of the EE14 benchmark set calculated with the CRYSTAL17 code. 137

A.10 Powder diffraction structure measures of EE14 crystal geometries generated by low-cost methods implemented in the SIESTA and CRYSTAL17 codes. 138

A.11 Relative energy differences and unit-cell densities of the lowest-energy enantiopure and racemate crystal structures of aza-helicene calculated using composite methods with the SIESTA and Quantum ESPRESSO codes. 139

A.12 Relative energy differences and unit-cell densities of the lowest-energy enantiopure and racemate crystal structures of aza-helicene calculated using composite methods with the CRYSTAL17 and Quantum ESPRESSO codes. 140

B.1 Binding energies of the KB49 benchmark set of van-der-Waals bound molecular dimers calculated with the SIESTA and Quantum ESPRESSO codes, using the PBE functional. 142

B.2 Binding energies of the KB49 benchmark set of van-der-Waals bound molecular dimers calculated with the SIESTA and Quantum ESPRESSO codes, using the B86bPBE functional. 143

B.3 Lattice energies of the X23 benchmark set of small molecular organic solids calculated with the SIESTA and Quantum ESPRESSO codes, using the PBE functional. 144

B.4 Lattice energies of the X23 benchmark set of small molecular organic solids calculated with the SIESTA and Quantum ESPRESSO codes, using the B86bPBE functional. 145

B.5 Powder diffraction structure measures of X23 crystal geometries generated by low-cost methods referred to plane-wave DFT-XDM or DFT-D2 equilibrium structures. 146

C.1 Relative energies and unit-cell densities of the lowest-energy crystal structures of 5-fluorouracil calculated using composite methods with the SIESTA, CRYSTAL17, and Quantum ESPRESSO codes. 148

C.2 Relative energies and unit-cell densities of the lowest-energy enantiopure and racemate crystal structures of naproxen calculated using composite methods with the SIESTA, CRYSTAL17, and Quantum ESPRESSO codes. 150

C.3 Relative energies and unit-cell densities of the lowest-energy crystal structures of carbamazepine calculated using composite methods with the SIESTA, CRYSTAL17, and Quantum ESPRESSO codes. 152

C.4 Relative energies and unit-cell densities of the lowest-energy crystal structures of olanzapine calculated using composite methods with the SIESTA, CRYSTAL17, and Quantum ESPRESSO codes. 154

LIST OF FIGURES

1.1 Chemical Structure of "ROY": A precursor to the antipsychotic drug olanzapine. 2

1.2 Structures of the compounds forming the first four crystal structure prediction blind tests organized by the Cambridge Crystallographic Data Center. 4

1.3 Structures of the compounds forming the fifth and sixth crystal structure prediction blind tests organized by the Cambridge Crystallographic Data Center. 5

3.1 Key structural differences between the two forms of aspirin. 37

3.2 $\{100\}<001>$ Slip mechanism in aspirin. 45

3.3 Variation in intermolecular interactions between acetyl groups during the phase transition of aspirin-I to aspirin-II via the $\{100\}<001>$ slip mechanism. 45

3.4 $\{100\}<001>$ Slip mechanism in aspirin forms I and II under varying pressures. 47

4.1 Molecular diagrams for the organic acid-base co-crystals considered, along with their Cambridge Structural Database codes. . . . 52

4.2 The unit cells of the organic acid-base co-crystals, together with their Cambridge Structural Database codes. 53

4.3 Molecular diagrams for the organic acid-base salts considered, along with their Cambridge Structural Database codes. 59

4.4 Predicted ionisation states of organic acid-base solids when using various amounts of exact-exchange mixing in density-functional theory and Hartree-Fock methods. 60

5.1 Molecular diagrams for the organic solids of the X23 benchmark set of lattice energies. 64

5.2 Real CPU time per SCF iteration, per optimization step, and total CPU time for selected crystal structures, of the various low-cost methods tested. 69

5.3 Molecular structures for “aldol”, “oxazoline”, “tetrazole”, and “N-helicene”, added to the set of ten amino acids to form the EE14 set. 75

5.4 Experimental enantiomeric excess as a function of calculated energy differences between the enantiopure and racemate crystal structures. 77

5.5 Lowest-energy enantiopure and racemate crystal structures of aza-helicene reranked with the B86bPBE-XDM//PBE-D2/DZP composite method. 80

5.6 Lowest-energy enantiopure and racemate crystal structures of aza-helicene reranked with the B86bPBE-XDM//sHF-3c composite method. 82

6.1 Quantities used to calculate the XDM dispersion coefficients along the internuclear coordinate in solid argon. 90

6.2 Graphite exfoliation curves calculated with selected methods, compared to experimental data. 94

6.3 Counterpoise correction of single-layer graphene in the presence of ghost orbitals centered of the atomic positions of a neighboring layer of graphene. 95

6.4 Phosphorene exfoliation curves calculated with selected methods, compared to Diffusion Monte Carlo data. 96

7.1 5-Fluorouracil and its two known polymorphs. 108

7.2 Crystal-energy landscapes of 5-fluorouracil. 109

7.3 Tetramers of 5-fluorouracil and uracil. 110

7.4 Naproxen and its known enantiopure and racemate forms. 113

7.5 Crystal-energy landscapes of naproxen. 114

7.6 Carbamazepine and its five known polymorphs. 117

7.7 Crystal-energy landscapes of carbamazepine. 118

7.8 Olanzapine and its three characterized polymorphs, along with the hypothesized structure of Form-III. 119

7.9 Crystal-energy landscapes of olanzapine. 121

ABSTRACT

Despite significant progress made in the last twenty years, the crystal structure prediction (CSP) of organic molecular solids remains challenging, as the demand to predict more complex crystal structures increases. On the one hand, relative energies between candidate crystal structures generated during a CSP protocol must be calculated accurately; on the other, the complexity of the crystal-energy landscape imposes stringent limitations on the method's computational cost. While plane-wave density-functional theory (DFT) methods have become the workhorse for the final stages of CSP protocols, due to their balance between high accuracy and efficiency, they remain prohibitively expensive during the early and intermediate stages.

The primary aim of this thesis is the development of composite approaches for CSP, which comprise a geometry optimization using a low-cost method followed by a single-point energy calculation using plane-wave DFT with the exchange-hole dipole moment (XDM) dispersion model. The composite approaches were first tested on small molecular solids; assessment based on their abilities to produce absolute lattice energies was found to be misleading, and relative lattice energies provided a much better indicator of performance in a CSP context. To allow use of the XDM dispersion model with low-level methods, it was implemented in the SIESTA code, which uses numerical finite-support local orbitals to reduce the computational cost of the calculation. Composite approaches making use of the same DFT-D method both for low- and high-level DFT frameworks yielded the best accuracy, while remaining significantly cheaper than performing full geometry optimizations with plane-wave DFT. The composite approaches were then successfully employed for CSP of organic molecules with applications ranging from chiral organic semiconductors to pharmaceutical solids.

Secondary objectives of this thesis sought to offer insight as to whether certain classes of solid-state materials are not appropriate benchmarks for method validation, and whether DFT-D methods are always suitable to describe all molecular crystals of interest. In particular, using compounds that form polytypes, e.g., crystalline aspirin, to validate computational methods was found to be inadvisable due to their high geometric and energetic similarity. Also, delocalization error, an often-overlooked limitation of most DFT methods, affected the correct identification of the protonation site in multicomponent acid-base crystals. This error greatly affects the reliability of these methods for validation of experimental (or the prediction of new) crystal structures.

Overall, the work presented in this dissertation provides appropriate methodological and benchmarking tools to accelerate the intermediate stages of CSP protocols, while retaining high levels of accuracy and reliability in the crystal-energy landscapes generated, ultimately enabling the study of increasingly complex molecular crystals.

List of Abbreviations and Symbols Used

Abbreviation	Description
3ob	Parameter Set For Density-Functional Tight Binding Method
a.u.	Atomic Unit
aDZ	Dunning's aug-cc-pVDZ Basis Set
API	Active Pharmaceutical Ingredient
aTZ	Dunning's aug-cc-pVTZ basis set
ATM	Axilrod–Teller–Muto Three-Body Dispersion-Energy Term
BLYP	Becke Exchange and Lee-Yang-Parr Correlation
	GGA Density Functional
B3LYP	Becke-3-Parameter Exchange and Lee-Yang-Parr
	Correlation Hybrid Functional
B86bPBE	Becke's B86b Exchange and PBE GGA Density Functional
BJ	Becke-Johnson
BR	Becke-Roussel
BSIE	Basis-Set Incompleteness Error
BSSE	Basis-Set Superposition Error
C21	Benchmark Set of Molecular Organic Solids
CCDC	Cambridge Crystallographic Data Center
CCSD(T)	Coupled Cluster Singles & Doubles with Perturbative Triples Method
CG	Conjugate Gradient
CI-NEB	Climbing-Image Nudged-Elastic-Band Method
COD	Crystallography Open Database
CP	Counterpoise
CPOSS	Control and Prediction of the Organic Solid State
CPU	Computer
CSD	Cambridge Structure Database
CSP	Crystal Structure Prediction
DFA	Density-Functional Approximation
DFTB	Density-Functional Tight Binding
DFT(-D)	(Dispersion-Corrected) Density-Functional Theory

Abbreviation	Description
DMA	Distributed Multipole Analysis
DMC	Diffusion (Quantum) Monte Carlo Method
DZP	Double-ζ Plus Polarization Basis Set
ee	Enantiomeric Excess
EE14	Benchmark Set of Amino Acids and Chiral Compounds
Expt	Experimental
FF	Force Field
gCP	Geometrical Counterpoise Method
GGA	Generalized-Gradient Approximation
Gn	Gaussian Gn (n = 1–5) Series of Methods
HF	Hartree-Fock Theory
(s)HF-3c	Grimme's (Scaled) Minimial Basis-Set Hartree-Fock Method with Semiempirical Corrections
HSE06	Heyd-Scuseria-Ernzerhof Variable Exchange Hybrid Functional
HSE-3c	Heyd-Scuseria-Ernzerhof Variable Exchange Hybrid Functional with Semiempirical Corrections
KB(49)	Kannemann-Becke Benchmark Set of Molecular Dimers
LDA	Local-Density Approximation
LR	Long Range
MAE	Mean Absolute Error
MAPE	Mean Absolute Percent Error
MAX	Maximum Absolute Error
MBD	Many-Body Dispersion Model
MC MP2C	Monomer-Centered Basis MP2-Coupled Method
ME	Mean Error
MEP	Minimum-Energy Path
mio	Parameter Set For Density-Functional Tight Binding Method
MP	Monkhorst-Pack k-point Grid Sampling Scheme
MP2	Møller Plesset Second-Order Perturbation Theory
PAW	Projector-Augmented Wave
PBE	Perdew-Burke-Ernzerhof Density GGA Density Functional
PBE0	Perdew-Burke-Ernzerhof Hybrid Density Functional
PBEh-3c	Grimme's Small Basis-Set PBE Hybrid Functional with Semiempirical Corrections

Abbreviation	Description
PES	Potential Energy Surface
POWDIFF	Powder Diffraction Structure Similarity Measure
pLMP2	Periodic Local MP2 Method
POB-DZPV	Pople-Type Valence Double-ζ Plus Polarization Basis Set
POLY59	Benchmark Set for Organic Crystal Polymorphism
PW86PBE	Perdew-Wang Exchange and PBE Correlation Hybrid Functional
QE	Quantum ESPRESSO
QM/MM	Quantum-Mechanical/Molecular Mechanics Embedding Scheme
ROY	5-Methyl-2-[(2-nitrophenyl)amino]-3-thiophenecarbonitrile
revPBE	Revised Perdew-Burke-Ernzerhof Exchange Functional
revPW86	Revised Perdew-Wang Exchange Functional
rVV10	Revised Vydrov and van Voorhis Non-Local Density Functional
RI-MP2	Resolution-of-the-Identity MP2 Method
SCS(MI) RI-MP2	Optimized Spin-Component-Scaled RI-MP2 Method
SCC-DFTB(n)	Self-Consistent Charge (n-th) Order Density-Functional Tight-Binding Method
SCF	Self-Consistent Field
SCS	Self-Consistent Screening
SR	Short Range
SRB	Short-Range Basis
SZP	Single-ζ Plus Polarization Basis Set
TPSS	Tao-Perdew-Straroverov-Scuseria Meta-GGA Density Functional
TS	Tkatchenko-Scheffler Dispersion Model
TZP	Triple-ζ Plus Polarization Basis Set
vdW	van-der-Waals
vdW-DF1	Dion-Rydberg-Schröder-Langreth-Lundqvist Non-Local Density Functional
vdW-DF2	Lee-Murray-Kong-Lundqvist-Langreth Non-Local Density Functional
VV	Vydrov and van Voorhis
W99	Williams-99 Force Field
xTB	Grimme's Density-Functional Tight-Binding Method
X23	Extended C21 Benchmark Set of Molecular Organic Solids
XDM	Exchange-Hole Dipole Moment Dispersion Model

Symbol	Description
a, b	Atomic Index
a, b, c	Unit-Cell Parameters
$\tilde{a}$, $\tilde{b}$, $\tilde{c}$	Super-Cell Dimensions
a_1, a_2	XDM Becke-Johnson Damping Parameters
a_{x}	Fraction of Exact Exchange
$\mathbf{a}$	Unit-Cell Vector
A	Hemholtz Free Energy
A, a, b	Becke-Roussel Hole Parameters
$\mathbf{b}$	Reciprocal Cell Vector
$c_{\mathbf{k+R}}$	Fourier Coefficient
$C_{n,a}$	Homoatomic Dispersion Coefficient (e.g., n = 6, 8, 10)
$C_{n,ab}$	Two-Body Dispersion Coefficient (e.g., n = 6, 8, 10)
$C_{9,abc}$	Three-Body Dispersion Coefficient
CN	Coordination Number in D2 Dispersion Model
c_s	Constant Defined in Dimensionless Reduced-Density-Gradient Expression
$c_{\mu,\nu}$	Density Matrix Element Between Atomic Orbitals
c_{x}	Constant Defined in LDA and GGA Exchange Energy Expressions
d, d_1, d_2	Damping Parameters Used in D2 and D3 Dispersion Models
$d_{X\sigma}$	Exchange-Hole Dipole Moment
D_σ	Difference Between Spin-Kinetic-Energy Density and Weizsaecker Term
E	Energy
E_{abc}	Three-Body Energy
E_{ab}^{rep}	DFTB Pairwise Repulsion Energy Between Atoms
E_{bsse}	Counterpoise Correction to the Energy
E_{c}	Correlation Energy
$E_{\mathrm{c}}^{\mathrm{nl}}$	Non-Local Correlation Energy
E_{disp}	Dispersion Energy
E_{DFT}	Total DFT Energy
E_{DFTBn}	n-th Order DFTB Energy
E_{el}	Electronic Energy
E_{ex}	Exfoliation Energy
$E_{\mathrm{x}}^{\mathrm{GGA}}$	GGA Exchange Energy
$E_{\mathrm{x}}^{\mathrm{LDA}}$	LDA Exchange Energy

Symbol	Description
$E_{\mathrm{L-DL}}$, $E_{\mathrm{S-RS}}$	(Electronic) Energy Difference Between Enantiopure and Racemate Crystals
E_{latt}	Lattice Energy
E_{vib}	Vibrational Contributions to the Energy
E_{x}	Exchange Energy
$E_{\mathrm{x}}^{\mathrm{HF}}$	Hartree-Fock Exact Exchange Energy
E^{high}	Energy Calculated by a High-Level Method
$E_{\mathrm{x}}^{\mathrm{hybrid}}$	Hybrid Exchange Energy
E^{low}	Energy Calculated by a Low-Level Method
E_{xc}	Exchange-Correlation Energy
$f(R_{ab\mathbf{L}})$	Damping Function
F_{disp}	van-der-Waals Contribution to Atomic Force
F_s	Enhancement Factor
F_{vib}	Vibrational Contributions to the Hemholtz Free Energy
G	Gibbs Free Energy
G_{solv}	Gibbs Free Energy of Solvation
G_{sub}	Gibbs Free Energy of Sublimation
$\mathbf{G}$	Reciprocal Lattice Vector
H	Enthalpy
H_{sub}	Enthalpy of Sublimation
h	Planck's Constant
$\hbar$	Reduced Planck's Constant
$h_{X\sigma}$	Spin-Dependant Exchange Hole
i, j	Electronic Index
IP	Ionization Potential
$\mathbf{k}$	Reciprocal Space Vector
k_{B}	Boltzmann Constant
K_{sp}	Solubility Product
K	Non-Local van-der-Waals Kernel
K_a	Acidity Constant ($\mathrm{p}K_a$ = -$\log K_a$)
$\mathbf{L}$	Arbitrary Lattice Vector
m	Atomic Mass
M_l	l-th Order Multipole Moment
N	Number of Data Points (e.g., on a Crystal-Energy Landscape)
N_{opt}	Number of Geometry Optimization Steps

Symbol	Description
p_{th}	Thermal Pressure
q	Electronic Charge
$\mathbf{q}$	Reciprocal Space Vector
Q	Effective Charge Used in D3 Dispersion Model
Q_σ	Exchange Hole Curvature
r_{12}	Interelectronic Separation, Coulomb, Operator
R	Gas Constant
$\mathbf{R}$	Bravais Lattice Vector
R_{ab}	Interatomic Distance
$R_{ab\mathbf{L}}$	Interatomic Distance in a Periodic System
$R_{\mathrm{c},ab}$	Critical (van-der-Waals) Radius
$R_{\mathrm{cov},a}$	Scaled Covalent, Single-Bond, Atomic Radius
$R_{\mathrm{vdW},ab}$	van-der-Waals Radius
s	Dimensionless Reduced Density Gradient
s_{42}	Scaling Factor Used in D3 Dispersion Model
s_6	Damping Parameter Used in D2 and D3 Dispersion Models
s_8	Damping Parameter Used in D3 Dispersion Model
s_r	Damping Parameter Used in TS Dispersion Model
t	Time
S	Similarity Index
T	Temperature
U	Hubbard Parameter
$\mathcal{V}$	Periodic Potential
V	Volume
V_{ext}	External Potential
Z	Formula Unit
Z_a	Nuclear Charge Factor
α, β, γ	Cell Angles
α_a	Atom-in-Solid Polarizability
α_a^{at}	Free-Atom Polarizability
β	Constant Defined in B86bPBE Enhancement Factor Expression
γ	Constant Defined in B86bPBE Enhancement Factor Expression
γ_{ab}	DFTB Function Describing On-Site Self Repulsion

Symbol	Description
Γ	Brillouin-Zone Center
Γ_{ab}	Charge Derivative of γ_{ab} Function
δ_i	Increment Along i-th Cell Axis
δ_{ij}	Kronecker Delta
$\delta\rho$	Electron Density Fluctuation
ϵ	Boltzmann Factor Used to Evaluate Enantiomeric Excess
ζ	Gaussian Exponential Coefficient
η,ξ	Cartesian Directions
θ	Angle
κ	Constant Defined in PBE Enhancement Factor Expression
$\varkappa$	Transmission Coefficient
ρ	Electron Density
ρ_0	Harris Density
ρ_a^{at}	Free-Atom Electron Density
$\rho_\sigma^{\mathrm{ae}}$	All-Electron Spin Density
$\rho_\sigma^{\mathrm{at}}$	Free-Atom Spin Density
$\rho_\sigma^{\mathrm{at}}$	Valence-Electron Spin Density
μ	Constant Defined in PBE Energy and Enchancement Factor Expressions
$\mu\ \nu$	Atom-Ortibal Indices
σ	Electron Spin Index
σ_{disp}	van-der-Waals Contribution to Cell Stress
τ_σ	Positive-Definite Spin-Kinetic-Energy Density
ϕ	Atomic Orbital
$\Phi_{a,b}$	Hessian Matrix
ψ	Electronic Wave Function
ψ_{k}	Bloch Wave Function
ω	Error-Function Width Parameter Used in Range-Separated Functionals
ω_{Hirsh}	Hirshfeld Weight
ω	Vibrational Frequency

ACKNOWLEDGEMENTS

I am not one to write all of this mushy stuff. But, here goes... I would be lying if I said that the last four years have been anything else than a roller-coaster ride. If it were not for all the support that I have received from family, friends, and colleagues over the years, I don't where I would be. Some of you have believed more in me than I ever have been able to, so thanks! It will be hard to forget many of the shenanigans that I have experienced with all of you... Whether that's a good a bad thing, I am still unsure.

To my supervisors, Erin and Rob: I have learned so much from you both. Your different styles of mentoring have definitely made this journey even more worthwhile. Also, to Alberto: thanks for all the advice and mentoring over the years, especially during the early stages. I have learned a lot of the technical stuff from you, and have enjoyed your sense of humour and sarcasm. Hopefully, I didn't annoy you too much.

To those with whom I have had the opportunity of collaborating and/or publishing with: it has been wonderful to work with a group of such talented and passionate people.

To my committee members: I thank you for taking the time to follow my ongoing progress and am appreciative of the feedback and advice that you have given me over the last few years.

To the organizations that have helped make the work in this thesis possible; the Natural Sciences and Engineering Research Council (NSERC) of Canada, the Walter C. Sumner Foundation, Dalhousie University, along with Compute Canada and its regional consortia: I am appreciative of the financial support I have received over the years, as well as having been able to access computational resources.

Finally, to the people that have showed me that board games are not boring: it's been a blast!

CHAPTER 1

INTRODUCTION

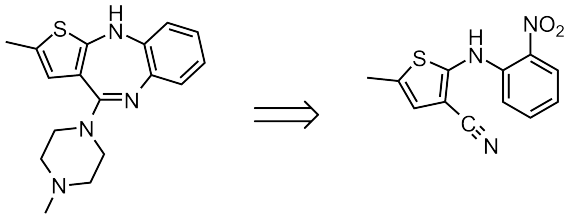
1.1 Molecular Crystal Structure Prediction

The phenomenon of polymorphism is ubiquitous in nature and is exhibited by many organic solids, which can be of importance to the development of, e.g., microporous materials,[1] semiconductors,[2,3] explosives,[4–6] pigments and dyes.[7,8] By using the term *polymorphism*, although slight variations of it exist in the literature,[9–12] it is meant that a substance consisting of identical units of an element, molecule, ionic formula, or combination thereof, can exhibit different periodic arrangements in the solid state.[13]

The ability to predict which polymorph will be formed under a given set of experimental conditions, as well as defining its stability with respect to other crystalline forms, has been of growing interest and concern in many scientific communities,[9,14–24] most notably in the pharmaceutical industries.[20,25–27] This should be unsurprising if one imagines the numerous ways molecules can pack in the solid state, which in turn will affect the physical, thermodynamic, spectroscopic, kinetic, surface, and chemical properties of a crystal structure.[13] For example, one of the most famous cases of polymorphism, 5-methyl-2-[(2-nitrophenyl)amino]-3-thiophenecarbonitrile (more commonly referred to as "ROY", Figure 1.1),[28] a precursor to the antipsychotic drug olanzapine,[29] shows how varying interactions in the crystal structure can yield different physical properties. Each of its known polymorphs (now 10)[30] exhibit different colours (red, orange, or yellow), crystal habits, and various melting points, due to the conformational flexibility and degree of co-planarity about the two aromatic rings of the parent molecule within the crystal structure.[28,30]

From the pharmaceutical industry perspective, the importance of knowing which form of a substance will crystallize, and whether or not it will remain stable, has repeatedly demonstrated the need to better understand this phenomenon.[31] In the late 90s, ritonavir,[32,33] marketed as the antiretroviral

Figure 1.1: "ROY": A precursor to the antipsychotic drug olanzapine.



drug Norvir® for the treatment of the human immunodeficiency virus by Abbott Laboratories, was removed from market just a few years after its approval. The presence of seeds of a more stable crystalline form of the drug caused the conversion over time of the manufactured metastable form to a less soluble form, making the formulation of the drug less biologically active.[33] Similarly, and more recently, rotigotine,[34] which is marketed as Neupro® transdermal patches for the treatment of Parkinson's disease, was withdrawn from market[20] due to interference from a second late-appearing form. Another notorious case of polymorphism is that of ratinidine hydrochloride (i.e., Zantac®), where discovery of a new polymorphic form of the active pharmaceutical ingredient provided fuel for a series of patent litigations between two major pharmaceutical companies (Glaxo, Inc. vs. Novopharm Ltd).[20,21] These three examples serve only to highlight a few of the landmark cases in this industry and illustrate that, even to this day, the tools and technologies (e.g., crystallization and solid-form screening methods[26]) used in industry can struggle to identify all accessible polymorphs of a compound to avoid the late appearances of more stable crystalline forms.[20,35] Furthermore, these cases are not exceptions to a rule; rather, it has been estimated that between 15-45% of the pharmaceutical drugs currently marketed are not in their thermodynamically stable forms.[36] To be able to predict, based solely from theoretical and computational grounds, which polymorphs are likely to be formed under an experimental setting and if they are stable is thus of significant interest.[26]

However, predicting the molecular crystal structure of a given compound solely from computational grounds, or 'crystal structure prediction' (CSP), is currently a great challenge in computational chemistry.[24,37] A number of CSP strategies have been used over the past few decades, and progress in the field has been tracked by a series of blind test competitions held by the Cambridge Crystallographic Data Centre (CCDC).[24,38–42] In these blind test contests, participants are given the molecular diagrams for a few organic compounds and are asked to predict the experimental crystal structure, which is unknown to them. The sixth and most recent blind test was completed in 2016.[24] Many excellent reviews have highlighted the progress made in the development of computational

protocols for CSP, while outlining the difficulties that remain.[18,21,23,24,27,43]

The chemical diagrams of the target structures given in these blind tests, labelled with roman numerals, are depicted in Figures 1.2 and 1.3 for reference. The structures have ranged from small (<25 non-H atoms) and rigid compounds made up of common elements (such as C, H, N and O) to larger (<50-60 non-H atoms) and more flexible candidate structures (i.e., with several internal degrees of freedom) consisting of more uncommon elements and/or functional groups. Co-crystals, solvates, as well as some targets approaching pharmaceutically-relevant compounds (e.g., **XX**)[27] or some possessing multiple known polymorphs (e.g., **XXI**[44] and **XXIII**[24]), have also found their way into the most recent blind tests.[24,42,45] Some of the structures thought to be monomorphic at the time of the earlier blind tests (e.g., **IV** and **VI**) have since had other crystalline forms discovered.[46,47] This has provoked potential procedural changes for future blind tests to make them more useful in assessing CSP methodologies for generating and/or ranking crystal structure candidates and polymorphs.[37]

Although the success of current CSP methods has been largely variable based on whether or not the experimentally isolable polymorphs are ranked within the lowest-energy candidate structures generated in a search,[24] several key results during the first few blind tests have directed the developments of more modern methods. The use of dispersion-corrected density-functional theory (DFT-D) to describe intermolecular interactions in solids demonstrated the importance of accurately treating these interactions.[46,48,49] For instance, the Neumann-Leusen-Kendrick group cleverly used DFT-D to generate reference data and derive force-fields (i.e., tailor-made force fields),[50,51] which were then used to assist in the generation and initial ranking of structures. A final ranking of the lowest-energy structures was then done by a more computationally expensive treatment with DFT-D. This approach led to the successful prediction of all target structures in the fourth blind test.[41] During the fifth blind test, a successful prediction for compound **XX** by two groups[52] was equally encouraging, given that its scale and complexity approaches that of pharmaceutically-relevant compounds, and led to confirmation of the capabilities of using CSP as a complement to polymorph-screening methods in the pharmaceutical industry.[20,26,27,35]

Another important realization from the CSP blind tests (although pointed out from the onset of this field[15]) has been that kinetics plays an important role in deciding which polymorph will crystallize under a given set of experimental conditions.[21,27,53] This translates into structures that are crystallized from experiment as not necessarily being the most thermodynamically stable

Figure 1.2: Structures of the compounds forming the first four CCDC CSP blind tests. CSP1999[38] **I**: 3-oxabicyclo[3.2.0]hepta-1,4-diene, **II**: 4-hydroxy-2-thiophenecarbonitrile, **III**: 2-(2-phenylethenyl)-1,3,2-benzodioxaborole. CSP2001[39] **IV**: 3-azabicyclo[3.3.1]nonane-2,4-dione, **V**: 7-*endo*-(bromocamphorylsulfonyl)imine, **VI**: 6-amino-2-phenylsulfonylimino-1,2-dihydropyridine, **VII**: propane. CSP2004[40] **VIII**: hydantoin, **IX**: 2,9-diiodoanthanthrone, **X**: 2-acetamido-4,5,-dinitrotoluene, **XI**: azetidine. CSP2007[41] **XII**: 2-propenal, **XIII**: 1,3-dibromo-2-chloro-5-fluorobenzene, **XIV**: *N*-(dimethylthiocarbamoyl)benzothiazole-2-thione, **XV**: 2-amino-4-methylpyrimidine:2-methylbenzoic acid.

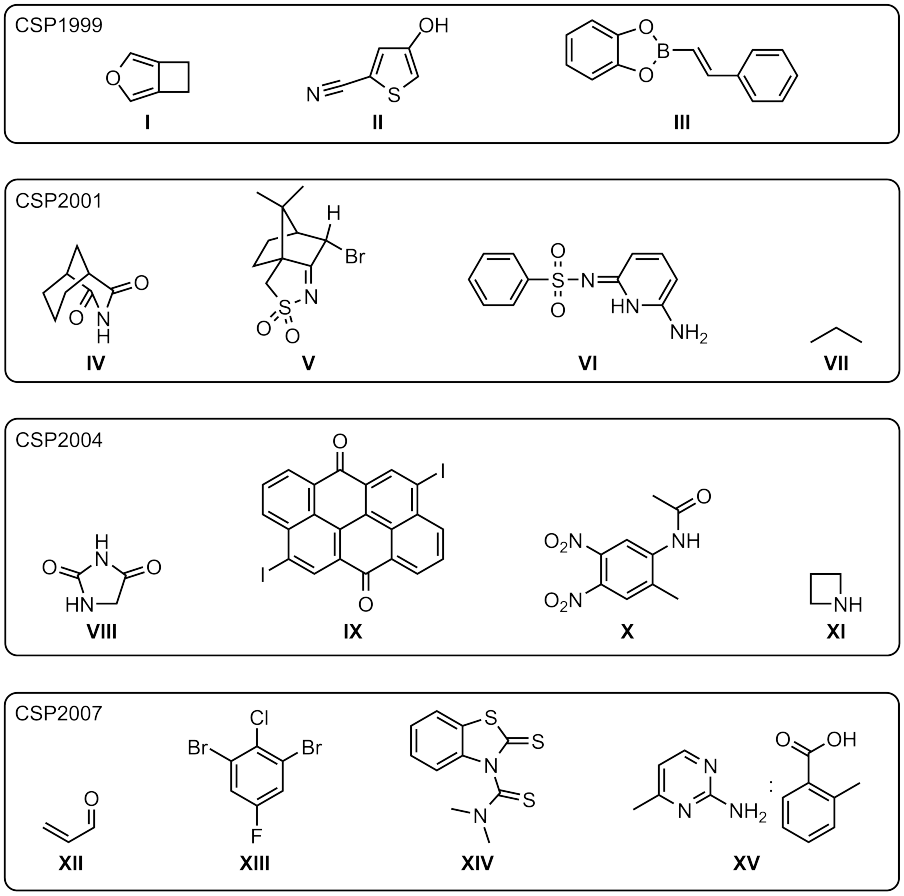
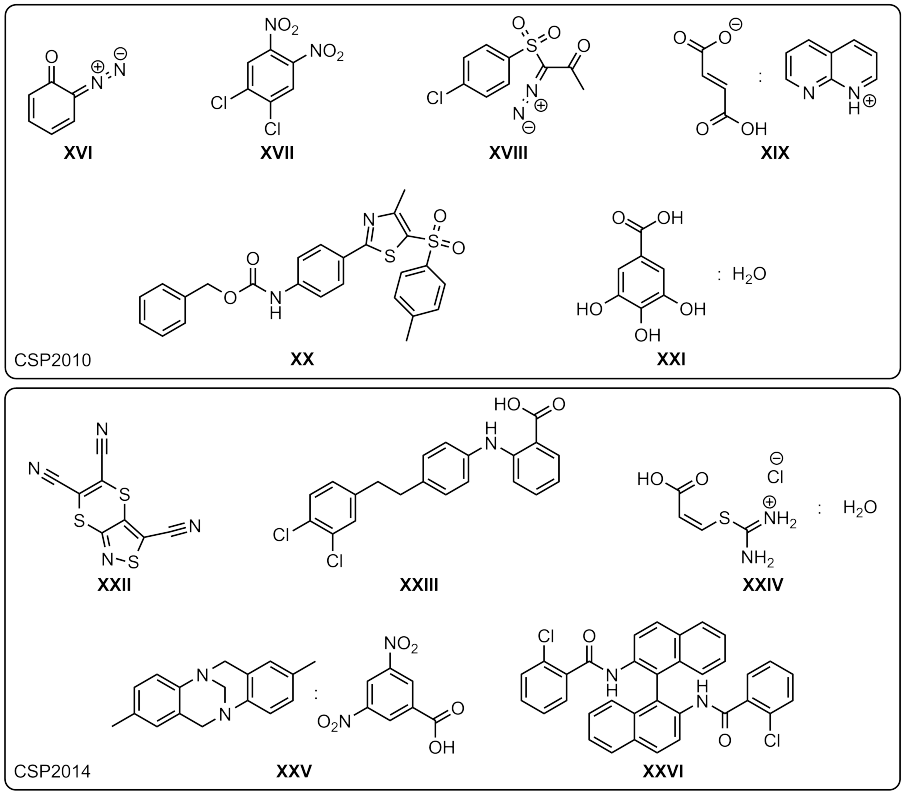


Figure 1.3: Structures of the compounds forming the fifth and sixth CCDC CSP blind tests. CSP2010[42] **XVI**: 2-diazo-3,5-cyclohexadiene-2-one, **XVII**: 1,2-dichloro-4,5-dinitrobenzene, **XVIII**: (1-((4-chlorophenyl)sulfonyl)-2-oxopropylidene)diazenium, **XIX**: 1,8-naphthyridinium fumarate, **XX**: benzyl-(4-(4-methyl-5-(*p*-tolyl-sulfonyl)-1,3-thiazol-2-yl)phenyl)carbamate, **XXI**: gallic acid monohydrate. CSP 2014[24] **XXII**: tricyano-1,4-dithiino[*c*]-isothiazole, **XXIII**: 2-((4-(3,4-dicholorophenethyl)phenyl)amino)benzoic acid, **XXIV**: chloride salt hydrate of *(*Z)-3-((diaminomethyl)thio)acrylic acid, **XXV**: 2,8-dimethyl-6*H*,12*H*-5,11-methanodibenzo[*b*,*f*][1,5]diazocine:3,5-dinitrobenzoic acid, **XXVI**: *N*,*N'*-([1,1'-Binaphthalene]-2,2'-diyl)bis(2-cholorobenzamide).



(lowest-energy) form. Therefore, the initial aim of CSP methods to match the lowest-energy predicted candidate structure to the experimental crystal structure has recently evolved into locating a set of plausible candidate structures, which could be potentially isolable under different sets of conditions.[21,24,27,54] The study of nucleation and related kinetics of the crystallization, however, remains to this day one of the greatest challenges in the field of crystal structure prediction and polymorphism.[53,55]

1.2 CSP Methodologies

How complex is it to devise CSP methods? The crystal structure prediction problem is formulated as follows: from the knowledge of only the chemical formula, or the two-dimensional molecular diagram, find the lowest-energy crystal structure, or set of minimal-energy structures, under a given set of conditions (e.g., temperature and pressure). This implies that *(a)* the crystal-energy landscape (a high-dimensional hypersurface of the atomic positions and crystal lattice parameters) be explored exhaustively in an efficient manner, and that *(b)* the energies of the crystal structures be determined to establish the likelihood of formation. Of course, both of these points are not inherently separable problems. Searching the potential-energy surface requires determining the energies of the points sampled, and finding all of the energy minima requires efficiently sampling the potential-energy surface. Nevertheless, depending on the CSP approach taken, the two can be considered as separate tasks or problems that must be solved.

1.2.1 Overview of Structure-Generation Methods

Searching for energy minima on the crystal-energy landscape requires exploring a high-dimensional surface, which is a function of the number of atoms or molecules in the unit cell and the lattice parameters of the cell itself. Global-optimization methods[56,57] have been developed over the years to deal with this type of problem and are now briefly discussed.

The most rudimentary of all of these methods is that of random sampling. As the name implies, the idea is to generate a random structure, relax it to a local minimum, and repeat *ad infinitum*, or until no lower-energy minima are found. The obvious drawback to this method is that the probability of finding the global minimum decreases rapidly as the system increases in scale and complexity, and that the method has no memory of where it has already sampled the potential-energy surface. Random sampling methods can be improved by enforcing chemical or symmetry constraints on the

search,[58] or by making use of better sampling algorithms, such as that of low-discrepancy Sobol' sequences,[59,60] which offer a more efficient sampling of the potential-energy surface. Codes such as GLEE[61] and CystalPredictor[59,60] are two examples that perform the quasi-random sampling just described.

Other classes of methods are based on Monte-Carlo algorithms or molecular dynamics. These include simulated annealing,[62] minima hopping,[63] and basin hopping.[64] The premise of these methods relies on perturbing the initial structure in a random, but controlled, way to overcome energy barriers between points on the potential-energy surface, thereby accessing nearby minima. The overall efficiency of these methods depends on the magnitudes of the perturbations applied and the choice of temperatures (or temperature gradients) used throughout the simulation, the latter dictating whether or not new structures are accepted or rejected throughout the search from a Boltzmann-distribution-like metric.

Metadynamics[65,66] is another method aimed at finding the global minimum along the potential-energy surface. Here, however, it is the free-energy surface that is explored. During the simulation, a set of collective variables are first defined and the energy of the system is evaluated. During the simulation, bias potentials are constructed from these collective variables, and then are gradually added to the potential energy surface wherever the candidate structure visits, forcing it to explore other nearby minima. The advantage of this method is that phase transitions can be explored efficiently, as the candidate structure will likely step from one minimum to another via the lowest-energy path between them. A major disadvantage to the metadynamics approach, however, lies in the choice of collective variables to be used during the simulation, which are often kept to a small number and/or chosen based on the structure of known polymorphic phases. Recently however, the use of collective variable surrogates to the enthalpy and entropy[67] have been shown to resolve some of these issues, thus allowing for the metadynamics methods to be appropriate in discovering new polymorphs of small molecular solids such as urea and naphthalene.[68]

Evolutionary algorithms have also found a place as a solution to the optimization problem,[69] adopting a "survival of the fittest" approach. Here, a population of structures is first produced and then subjected to "variation" operators. These variation operators combine features of the lowest-energy candidate structures from the initial population or affect one (low-energy) candidate structure (by, e.g., deforming the lattice, exchanging atoms with a different species, etc.) in systematic ways to generate a new population of structures. It is thus likely that the newer structures

sampled will be of similar or lower energy as the previous lowest-energy candidates. This has the advantage of exploring the low-energy regions of the potential energy surface more efficiently, while neglecting the higher-energy regions. Furthermore, to avoid a biased population and to preserve diversity, duplicate (or too similar) structures are removed by the algorithm and random structures are periodically inserted into the population. Amongst all evolutionary algorithms, the USPEX algorithm[69,70] is the most popular, most likely due to the many developments made to the program over the years. Variants of metadynamics and the variable-cell nudged elastic band method have also been implemented in a recent version of the core code to evaluate phase transitions.[71,72] Other codes such as XOpt[73,74] and GAtor[75,76] equally make use of evolutionary algorithms similar to what was just described. Particle-swarm optimization algorithms[77] have many similarities to evolutionary algorithms, although no variation operators are used. Instead, each candidate structure in an initial population is relaxed to a local minimum. The structures are further perturbed in such a way that the direction of perturbation is dependant on the overall lowest-energy structures found and, during subsequent iterations of the algorithm, on the lowest-energy structure visited in the past. These history-dependant terms in the perturbation are weighted, and multiplied by a random number between 0 and 1, to avoid having structures trapped in a local minimum. Such algorithms can be found in codes like CALYPSO.[77,78]

All of these methods are continuously improved upon in crystal structure prediction, and their progress discussed in the results of the most recent of blind tests.[24] None of these, however, has stood out as the best method. Whether or not an experimental crystal structure ends up as one of the lowest-energy candidate structures depends not only on its generation by the global optimization method,[52] but also on its retention during the many stages of energy re-ranking and refinement.[24]

1.2.2 Overview of Energy-Ranking Methods

Thousands (or even millions) of structures can be generated from the searching algorithms discussed above. The subsequent (and often concurrent) ranking of polymorphs must therefore be done in such a way that a balance between computational cost and accuracy is attained. For example, with limited resources, it would be impractical to evaluate all of the points sampled on the crystal-energy landscapes with high-level computations. Thus, ranking of structures on the crystal-energy landscape is often performed in stages, using empirical methods (e.g., force fields[79–82]) to reduce the pool of candidate structures to a more manageable number. The gradual use of more elaborate and refined methods for the treatment of electronic structure (e.g., quantum-mechanical methods) and determination of the energy is then executed until the desired accuracy is obtained.[24] Thus,

the aim of empirical methods is not to obtain an exact energy ranking between the polymorph structures, but rather to separate out the relatively more stable structures (which ideally contain the experimentally observed crystal structure(s)) from higher-energy structures. With higher-level methods, the relative energies can then be refined in the hopes of recovering the exact energy differences between polymorph pairs.

The challenge in efficiently ranking isolable polymorphs stems from the small energy differences between them. Multiple surveys over the years of isolable polymorphs, either computational[20,83–85] or experimental,[86,87] have estimated the relative energies to be on the order of 4 kJ/mol for 90% of polymorphic pairs, with 50% being separated by less than 2 kJ/mol. Quantum-mechanical calculations must therefore be capable of describing the different types of intermolecular interactions in organic solids, which are often weak and anisotropic, with extreme accuracy. This remains challenging to this day for many systems.[23] Only recently has the lattice energy of a benzene crystal been determined at a sub-kJ/mol accuracy.[17,19,88] The method used to obtain such a level of accuracy (a fragment-based approach using high-level wavefunction theory methods,[89]) is, however, still limited to molecular crystals with relatively small unit cells.[21,23]

The electronic-structure methods used throughout the ranking and refinement process serve as mediators during refinement of crystal structures and as the final arbiters when establishing their relative stabilities. It is thus important to correctly describe the numerous types of interactions between molecules in the solid state. There are continuing developments in methods to efficiently rank polymorphs,[23] either by sacrificing accuracy to speed up the computations,[90] or by working to establish sub-kJ/mol accuracy for lattice energies of molecular crystals.[88] Some of these techniques treat the systems completely with quantum mechanics, while others partition long- and short-range inter- and/or intramolecular interactions and use a combination of quantum-mechanical and classical-mechanical treatments to establish the relative energies of the solid-state systems. Amongst these methods are those based on specialized force-fields.[24,52] Of note are the intermolecular potentials developed by Price and co-workers over the last few decades, which incorporate a distributed multipole analysis scheme.[91] In this approach (implemented in the DMACRYS software[91]), the focus is placed on evaluating an anisotropic intermolecular potential between molecules, which are considered as rigid units. In short, the total energy is calculated by summing terms parametrized for inductive, repulsive, and attractive interactions, and an electrostatic term is derived from atomic multipoles, for a specific conformation of the molecule.[92]

The successful application of DFT-D-based methods used to construct efficient "tailor-made" force fields[48,49] in the fourth and fifth blind tests[41,42] has since sanctioned the use of stand-alone DFT-D methods within CSP protocols.[24] Because of their lower scaling with system size compared to wavefunction theory, DFT-D methods have become the computational workhorse of CSP during the final stages of a search. These methods allow for treatment of organic molecular solids whilst balancing computational cost and accuracy, the latter neighbouring that of the coveted "chemical accuracy" of 4.2 kJ/mol.[93]

Alternatively, post-Hartree-Fock (HF) wavefunction theories, such as Møller-Plesset perturbation theory (MP2) and coupled-cluster (CCSD(T)), can be used in CSP. Although formally computationally expensive, post-HF methods have been adapted for use in solid-state calculations.[23] An example of a class of methods making use of the post-HF formalism are those dubbed "fragment-based approaches". These techniques (such as the hybrid many-body interaction scheme developed by Beran and co-workers[89,94]) partition the short- and long-range inter- and/or intramolecular interactions and then treat these using a combination of low to high levels of theory. They have shown promise in providing an accurate description of intermolecular interactions in organic solids, as demonstrated by the sub-kJ per mole accuracy obtained for the computation of the lattice energy of crystalline benzene.[88] Overall, the advantage of these wavefunction theory-based methods is that they can achieve very accurate results and be improved upon systematically (e.g., by using more flexible basis sets, or by extrapolating to the complete basis-set limit). In the end, however, these methods are far from being routinely used in CSP studies because of their high computational costs relative to other electronic structure methods, such as DFT.[23,90]

The use of DFT-D (and post-HF wavefunction theory) methods, however, are not amenable to the early/intermediate stages of CSP, when thousands/hundreds of structures need to be sorted and ranked. To circumvent this issue, some have suggested the use of semiempirical methods,[95] such as minimal-basis Hartree-Fock (HF-3c[96]) or tight-binding DFT (DFTB[97]), as pre-screening methods during a CSP search. These would be used to refine the candidate pool further after the structure generation, but before the use of more computationally expensive DFT and/or wavefunction techniques. These low-cost methods provide modest accuracies in determining lattice energies of crystal structures, but offer a substantial speed-up for the calculations.[97]

Up to the very final stages of most CSP protocols, candidate structures are exclusively ranked in terms of the lattice energy.[23,24,27] However, such a lattice-energy ranking may not necessarily

reflect which polymorphs will be isolable experimentally at finite temperatures. This is because it is the free energy of a crystal structure that dictates its thermodynamic stability, and thus the likelihood of observing it through experiment (neglecting kinetic effects that can contribute to observing higher-energy metastable forms[53]). It is often only possible to account for thermal and vibrational effects for a handful of structures due to the high computational cost associated with computing the free energy.

A recent survey conducted by Nyman and Day[85] on 508 polymorphic organic molecules (totalling 1061 crystal structures) found that the entropic contributions affect the relative energies of isolable polymorphs only in a minor fashion, causing the structures to be reordered by less than 1 kJ/mol for 70% of the polymorph pairs, and by greater than 2 kJ/mol for less than 6% of cases. However, as the lattice energy differences between polymorphs are similarly small, including these finite-temperature effects can alter the relative stability of the polymorph pairs. Specifically, 9% of the cases examined in the Nyman and Day survey displayed a reversal of polymorph-pair ordering upon inclusion of vibrational and thermal effects.[85] Neglecting these vibrational corrections to the energy has been justified in the past if pairs of polymorphs have similar vibrational modes;[83] unfortunately, this is not always the case.[85,98] Because of the computational expense needed to compute these free-energy corrections, they are usually used only when low-energy candidate structures are separated by a few kJ/mol, so that thermal and vibrational effects are taken into account if there may be an inversion in the ordering of the candidate structures.[24]

1.3 Thesis Goals

All current CSP protocols work under the assumption that the experimentally observed crystal structure coincides with the thermodynamically stable phase (i.e. kinetic effects are neglected).[53] A CSP protocol must sample a complex high-dimensional energy landscape efficiently and must also be able to rank reliably candidate crystal structures by their free energy.[57] Usually, the electronic lattice energies are used instead of free energies due to the cost and the difficulty of evaluating the vibrational free-energy contributions accurately. These vibrational terms are typically added only in the final candidate ranking step, if at all.[85,99] Even if one focuses on calculating electronic lattice energies alone, it is estimated that over 80% of all isolable polymorphs are separated by less than ca. 4.2 kJ/mol.[85,100] This is a very strict requirement that poses a challenge to current computational methods regarding their ability to produce accurate relative lattice energy differences in molecular crystals.

While dispersion-corrected DFT methods provide lattice energies accurate enough for reliable CSP (except in specific cases[101,102]), their computational cost is quite high.[24] This is especially problematic in the early stages of the search, when it is necessary to discard the less stable candidate structures.[96] To address this problem, one can resort to using a multi-step approach, in which computationally inexpensive (in the following "low-cost" or "cheap") methods are used to prune the list of candidate structures.[90] These low-cost methods are often empirical or semiempirical in nature (e.g., force fields) or resort to drastic approximations, such as the use of empirically corrected Gaussian minimal-basis-set calculations.[95] Low-cost methods speed up the energy calculation by several orders of magnitude and allow for more points on the crystal energy landscape to be surveyed. However, it is critical that they are able to identify reliably whether a candidate is within a certain energy range above the global minimum. Otherwise, low-energy polymorphs will be lost in the pruning stage.[90]

As the work described in this thesis aims to show, the multilevel refinement strategy for CSP protocols described above can be improved substantially by applying techniques used in the past several decades to speed up gas-phase molecular calculations, while retaining high levels of accuracy (e.g. the Gn series of methods[103–107]). The technique in question consists of using composite approaches, which make use of low-cost methods to cheaply optimize a given molecular (or in this work, crystal) geometry, followed by a single-point energy calculation with a higher-level method. The main assumption in this approach is that the structures obtained with the low-cost methods are transferable to the high-level single-point calculations, such that the results obtained have a comparable accuracy to full calculations with the high-level method. For this reason, the successful application of composite methods relies on performing appropriate benchmark calculations in order to establish the compatibility between the low- and high-level techniques that are being combined.

In this thesis, the proposed composite methods were first tested on small molecular organic solids (Chapter 5), where several low-cost methods were investigated for use in concert with high-level DFT-D methods. The ability of these composite methods to capture absolute energies and to properly describe relative energies between crystal structures was examined, and a benchmark set of chiral-molecules crystal structures was introduced.

It was postulated that the best levels of accuracy for composite methods can be obtained if both the low- and high-level methods used in a composite approach share the same theoretical

framework. For this purpose, the B86bPBE-XDM method was implemented in the SIESTA code, which uses a basis set of numerical orbitals with finite support to speed up calculations. Then, the application of a composite approach making use of the SIESTA method followed by plane-wave calculations, both employing the B86bPBE density functional and XDM dispersion correction, was examined for small molecular solids (Chapter 6). Finally, these approaches were demonstrated to effectively reduce the computational cost, while retaining high-level accuracy, for CSP of pharmaceutical solids (Chapter 7). Ultimately, it was shown that the composite methods in question yield desirable crystal-energy landscapes for a variety of drug molecules, making these types of approaches attractive for performing routine CSP in the pharmaceutical industry.

However, before moving on to the main topic of this thesis, the theoretical foundations of the computational methods used in this work, in particular of periodic-boundary DFT-D and low-cost semiempirical methods, are detailed in the following chapter (Chapter 2). The two subsequent chapters provide illustrations of the limitations and potential failures of DFT-D methods when certain classes of systems are studied, i.e., polytypes (Chapter 3) and organic acid-base co-crystals and salts (Chapter 4).

CHAPTER 2

THEORETICAL BACKGROUND AND METHODOLOGICAL DETAILS

The crystalline structures adopted by organic molecules often involve a compromise between many competing weak interactions. Thus, the area of crystal structure prediction presents a stringent challenge for computational methods, where the relative energies between different packing arrangements must be assessed to within accuracies of a few kJ mol^{-1} or less.[24] The impressive success of dispersion-corrected DFT methods in CSP blind tests[41] has led to increased use of these methods for modelling the organic molecular solid state.[24] A review of its theoretical foundations is now undertaken.

2.1 Dispersion-Corrected Periodic-Boundary Density-Functional Theory

2.1.1 Overview of Density-Functional Theory

From a DFT point of view, evaluating the electronic energy is not done by solving the Schrödinger equation for the *N*-electron wavefunction, $\psi(\mathbf{r}_1, \mathbf{r}_2, ..., \mathbf{r}_N)$, but rather is obtained from a lower-dimensional quantity, the electron density, $\rho(\mathbf{r})$. Ultimately, the DFT electronic energy, E_{DFT}, is formulated as a functional of the density (i.e., a function of the electron density function):

$$E_{\mathrm{DFT}} \equiv E[\rho] = -\sum_i \int \psi_i(\mathbf{r}) \nabla^2 \psi_i(\mathbf{r}) \mathrm{d}\mathbf{r} + \int \rho V_{\mathrm{ext}} \mathrm{d}\mathbf{r} + \frac{1}{2} \int \int \frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{r_{12}} \mathrm{d}\mathbf{r}\mathrm{d}\mathbf{r}' + E_{\mathrm{xc}}[\rho], \tag{2.1}$$

where ∇^2 is the Laplacian operator. The first term represents the kinetic energy of non-interacting

electrons occupying Kohn-Sham orbitals, ψ_i, and the sum runs over all occupied orbitals for an N-electron system. The second term is the electron-nuclei interaction energy (where the external potential, V_{ext}, is determined by the nuclear positions), and the third term is the classical Coulomb repulsion between each electron, $J[\rho]$. All other energy contributions are swept into the exchange-correlation functional term, E_{xc}. This term effectively contains all other electron-electron interactions (both kinetic and potential), which are not taken into account from the reference system of non-interacting electrons by the first three terms in the DFT formalism.

To be able to compute the DFT energy in Eq. 2.1 above, one must do so iteratively, i.e., through a self-consistent field (SCF) calculation. This is because one must know the electron density in order to compute the energy, but the density itself is not known. One thus starts from an initial guess for the electron density, which is constructed from the Kohn-Sham orbitals for an initial configuration of the system,

$$\rho(\mathbf{r}) = \sum_{i}^{N} |\psi_i(\mathbf{r})|^2 \,. \tag{2.2}$$

The effective potential (consisting of the sum of the external, Coulomb, and exchange-correlation potentials) is constructed next, and the Kohn-Sham equations solved to find the minimal-energy orbitals, $\psi_i(\mathbf{r})$. An improved electron density is then constructed by placing the (N) electrons into these non-interacting orbitals and the entire process (using this new density to determine the potential) is repeated until convergence of the total energy is achieved.

One of the main drawbacks and sources of error in DFT calculations is that the exact form of the $E_{\mathrm{xc}}[\rho]$ term in Eq. 2.1 is unknown and must be approximated. If this last term was known, the electronic energy obtained would be exact. Typical density-functional approximations (DFAs) used to express E_{xc} include the local-density approximation (LDA), generalized-gradient approximations (GGAs), and hybrid GGAs, although more exist.[108] Furthermore, E_{xc} can be split in two contributions: exchange, E_{x}, and correlation, E_{c}. However, the exchange portion of E_{xc} dominates over correlation.[108] Correlation functionals will therefore not be discussed any further. Throughout the work presented, PBE correlation is used.[109]

The forms of exchange functionals are now addressed for the various DFAs used in this thesis work. The LDA assumes that the system of interest will behave as a uniform electron gas and, as

such, the E_{x} term can be written as

$$E_{\mathrm{x}}^{\mathrm{LDA}} = c_{\mathrm{x}} \int \rho^{4/3} \mathrm{d}\mathbf{r}, \tag{2.3}$$

where c_{x} is a constant. Because $E_{\mathrm{x}}^{\mathrm{LDA}}$ involves only the electron density, ρ, this type of density functional is referred to as being "local" in the density. However, rather than being uniform, the electron density is piecewise exponential for atomic systems. Thus, GGAs (such as PBE[109] and B86bPBE[109,110]) were developed to include the gradient of the density, $\nabla\rho$, within the E_{x} term:

$$E_{\mathrm{x}}^{\mathrm{GGA}} = c_{\mathrm{x}} \int F_s \rho^{4/3} \mathrm{d}\mathbf{r}. \tag{2.4}$$

The enhancement factor in, F_s, is a functional of the dimensionless reduced density gradient, $s = c_s|\nabla\rho|/\rho^{4/3}$, where c_s is a constant. The enhancement factors for the PBE and B86b exchange functionals are, respectively,

$$F_s^{\mathrm{PBE}} = 1 + \kappa - \frac{\kappa}{1 + \mu s^2/\kappa} \tag{2.5}$$

and

$$F_s^{\mathrm{B86b}} = 1 + \frac{\beta}{c_{\mathrm{x}}} \frac{s^2}{(1+\gamma s^2)^{4/5}}, \tag{2.6}$$

where, β, γ, κ, and μ are constants.

The use of GGAs over the LDA affords a drastic improvement in describing thermochemistry and reaction barriers, but using these DFAs tends to yield results that remain slightly overbinding.[108] To resolve this, hybrid density functionals were developed.[111] These DFAs are based on GGAs, but a fraction of the DFT exchange is replaced by Hartree-Fock exact exchange:

$$E_{\mathrm{x}}^{\mathrm{hybrid}} = a_{\mathrm{x}} E_{\mathrm{x}}^{\mathrm{HF}} + (1 - a_{\mathrm{x}}) E_{\mathrm{x}}^{\mathrm{GGA}}, \tag{2.7}$$

where a_{x} is a fraction between 0 and 1 and E_x^{HF} is the exact exchange energy, which has the form

$$E_x^{\mathrm{HF}} = -\frac{1}{2} \sum_{\sigma} \int \int \psi_i^*(\mathbf{r}) \psi_j^*(\mathbf{r}') \frac{1}{r_{12}} \psi_j(\mathbf{r}) \psi_i(\mathbf{r}') \mathrm{d}\mathbf{r} \mathrm{d}\mathbf{r}'. \tag{2.8}$$

An example of a commonly used hybrid functional is PBE0,[112] which replaces 25% of the PBE exchange by exact exchange ($a_{\mathrm{x}} = 0.25$ in Eq. 2.7). On the other hand, when a_{x} is set to 1 and no correlation is used in the formulation of E_{xc}, the resultant DFT hybrid is equivalent to a

Hartree-Fock description of electronic structure. The last two terms of Eq. 2.1 would consist of a classical "mean-field" description of the Coulomb interaction between electrons along with the exact exchange term defined in Eq. 2.8, which corresponds to the Hartree-Fock formalism.

The overbinding tendency of GGAs is a symptom of a problematic error known as self-interaction, delocalisation, or charge-transfer error,[113,114] which causes excessive delocalisation of charge. Because of approximations made to establish the functional form of E_{xc}, the electron self-interaction present within $J[\rho]$ is never fully cancelled out by E_{xc} in GGAs. However, if exact exchange (as in Eq. 2.8) is used in E_{xc}, as is the case in Hartree-Fock theory, this error can be fully eliminated. The non-locality of exact exchange, i.e., the dependence on not only $\rho(\mathbf{r})$, but also $\rho(\mathbf{r}')$ within Eq. 2.8, allows for the proper physical description of electronic structure for large interelectronic separations. The end result is that the tendency of GGAs to delocalise charge is to a certain degree removed when pure GGA exchange is replaced by a fraction of exact exchange.[108] The unfortunate disadvantage of computing Hartree-Fock exact exchange is its non-local nature, which makes calculations significantly more expensive for large molecular systems, and often prohibitively expensive for solid-state analogues.

To make computation of non-local exact exchange more tractable, modifications have been made to the density-functional hybrids just discussed (often termed as "global hybrids" because of a constant application of exact exchange in both the short- and long-range). These new range-separated hybrid functionals differ in how exact exchange replaces GGA exchange in the short and long ranges. That is, the Coulomb operator, $\frac{1}{r_{12}}$, is partitioned using an error function with parameter ω,

$$\frac{1}{r_{12}} = \underbrace{\frac{\mathrm{erfc}(\omega r_{12})}{r_{12}}}_{\mathrm{SR}} + \underbrace{\frac{\mathrm{erf}(\omega r_{12})}{r_{12}}}_{\mathrm{LR}}, \tag{2.9}$$

such that either exact or GGA exchange is applied in the short (SR) or long range (LR), respectively. Delocalisation error can be reduced by considering full or partial GGA exchange in the short range and transitioning to full HF exchange in the long-range. Conversely, in the HSE06[115] range-separated hybrid functional, exact exchange is applied in the short range only, as this functional was developed with the intent to study solid-state systems so as to avoid the expense of computing full non-local exact exchange in the long range.[116] Ultimately, HSE06 is based on the PBE0 hybrid, and so incorporates 25% of exact exchange in the short-range, while none is applied the long-range, and the ω value is set to 0.11 Bohr^{-1}.[115,117]

2.1.2 Treatment of Periodic Solid-State Systems

When treating solids, as opposed to individual molecules in the gas phase, one is faced with the problem of expressing an effectively infinite number of orbitals distributed over the entirety of the solid for an equally large number of electrons. Fortunately, a crystalline solid can be broken down into smaller identical pieces, i.e., unit cells, as by definition a crystalline structure displays periodicity in three spatial dimensions. When these unit cells are translated by (Bravais) lattice vectors, $\mathbf{R} = \sum_{i=1}^{3} n_i \mathbf{a}_i$, along these directions, one obtains the bulk crystal structure. The problem is then reduced to considering only the wavefunction, potentials, and electron density within one unit-cell. Furthermore, because of this periodicity, the potential, $\mathcal{V}(\mathbf{r})$, felt by an electron at any point, $\mathbf{r}$, in a unit cell will be of equal value when one considers any displacements from this point by a Bravais vector $\mathbf{R}$, i.e.,

$$\mathcal{V}(\mathbf{r}) = \mathcal{V}(\mathbf{r} + \mathbf{R}). \tag{2.10}$$

Bloch's theorem states that the eigenstates, ψ_i, satisfying the Schrödinger equation with an effective periodic potential, such as $\mathcal{V}(\mathbf{r})$ above, can be written as the product of a plane wave with reciprocal wave vector $\mathbf{k}$, times a potential function expanded in terms of a finite number of plane waves with (reciprocal) wave vector $\mathbf{G}$:

$$\psi_i(\mathbf{r}) = \sum_{\mathbf{G}} c_{i,\mathbf{k+G}} e^{i(\mathbf{k+G})\cdot\mathbf{r}}, \tag{2.11}$$

where $c_{i,\mathbf{k+G}}$ are Fourier coefficients. The reciprocal wave vectors satisfy $\mathbf{k} \cdot \mathbf{R} = 2\pi\delta_{ij}$ and $\mathbf{G} \cdot \mathbf{R} = 2\pi m$, and arise from the requirement that the wavefunction and periodic potential have the periodicity of the Bravais lattice. δ_{ij} is the Kronecker delta, and m is an integer. The set of the allowed $\mathbf{k}$-vectors then define the energy levels of the electrons.

An important consideration in these types of calculations is that properties of the system, such as the energy, span the entire set of allowed wave vectors (an infinity of them), i.e., $\mathbf{k}$-space. Thus, in order to compute these properties, one must sample $\mathbf{k}$-space efficiently in order to make the computations feasible. A common sampling scheme used to select these $\mathbf{k}$-vectors is that developed by Monkhorst and Pack,[118] which will be referred to as an "MP"-scheme. These MP-schemes establish a set of $\mathbf{k}$-vectors by producing an equally-spaced mesh of points in each of the three dimensions of the cell in $\mathbf{k}$-space. The notation used to specify the number of points along each

cell direction is $i \times j \times k$.

Solid-state calculations use plane-wave basis sets in order to represent the overall wavefunction or electron density at each lattice point of a crystal structure. Expressing these functions in terms of such a basis set requires the superposition of many plane waves. In practice, the size of basis set is controlled by setting cutoffs for the kinetic energy and electron density, which control how many plane waves are used for the expansion of the wavefunction and the electron density throughout the calculations. Increasing the cutoffs for the kinetic energy leads to the inclusion of plane waves with higher oscillating frequencies in the basis-set expansion. When performing solid-state calculations with plane-wave basis sets, it is important to check that both the cutoffs and the **k**-point sampling used lead to converged results, such that there is no significant change in the properties of the system examined upon using higher cutoffs or denser sampling grids.

Using a plane-wave basis set can be problematic for solid-state calculations, however. This is because of the rapidly oscillating behaviour of the wavefunction near the nuclei, in contrast to the slowly varying behaviour between them. As a consequence, calculations can require a very large number of plane waves to achieve converged results and correctly capture the fact that valence orbitals must be orthogonal to core orbitals. In order to remedy this problem, pseudopotential methods have been developed.[119] The idea behind these methods is that, since core electrons tend to be inert, it becomes advantageous to ignore the latter by adding a term in Eq. 2.1 to account for them: the so-called pseudopotential. Many types of pseudopotentials exist, with two of the most common types being norm-conserving[120–122] and ultrasoft.[123] Pseudopotentials also depend on the type of exchange-correlation functional used. However, the errors arising from the functionals to treat a particular system are typically much larger than the approximation of replacing the core electrons by a pseudopotential.[119]

Another approach that is widely used in solid-state calculations is that of projector-augmented waves (PAW) developed by Blöchl,[124] and later generalized by Kresse and Joubert.[125] Here, a combination of linear-augmented plane waves[126] and pseudopotentials is used to treat the valence electrons of a solid-state system. By augmented plane waves, it is meant that the region about each atom in the solid is treated using atomic-like functions (i.e., with radial functions and spherical harmonics), whereas the interstitial region (defined to be outside a given radius from the centre of an atom) is treated using plane waves only. The unique feature of the PAW method is that the rapid oscillations of valence orbital functions near atomic nuclei are removed through the use of

a one-to-one linear-mapping of the all-electron wavefunctions to pseudo-valence wavefunctions. This ultimately makes manipulation of valence-electron wavefunctions more convenient and efficient. The advantage of the PAW method, as opposed to other pseudopotentials, is that the core wavefunctions are retained, which can be of importance when one wants to describe chemical properties that depend on core electrons in addition to valence electrons.[119]

For the calculations described in this thesis, B86bPBE PAW datasets for the H, C, N, O, and F atoms were generated using the "atomic" code bundled within the Quantum ESPRESSO code,[127] version 5.1. These have been previously tested and validated by others for applications to molecular crystals, such as the racemate and enantiopure crystals of chiral amino acids (which are part of the EE14 set studied in Chapter 5),[128] as well as the previous CCDC blind test candidate structures.[101,102]

The cutoff radii used for the pseudisation of the all-electron wavefunctions were set to the values specified in Table 2.1. A concern during the generation of PAW datasets is to verify that, for the systems to be studied, no overlap between PAW spheres occurs.[129] This ensures that no spurious Coulomb interactions between spheres results from the augmentation of the charge densities by a "compensation" charge density term, this latter term being used to remove the multipole moments resulting from replacing the pseudised core density with its all-electron counterpart. Substantial PAW sphere overlap can lead to significant, and unpredictable, errors in the energies computed for crystalline systems, although it has a lesser effect on interatomic forces.[129]

For the molecular crystals studied in this thesis, the shortest conceivable inter-atomic distances would be for elements bonding with hydrogen. Of these, the shortest bond would be that of a typical single-bond between O and H, which is ca. 0.98 Å. The sum of PAW sphere radii for these two elements is 0.95 Å, and thus there should be little to no PAW overlap. Any other interatom distance, either of intra- or intermolecular nature, would be much greater than the sum of any two radii listed in Table 2.1.

The norm-conserving pseudopotentials used in conjunction with the SIESTA code,[130,131] were generated with the pseudopotential generator "ATOM".[132] A set of pseudopotentials for the H, C, N, O, F, Si, P, S and Cl atoms with the PBE functional were obtained from the SIESTA distribution, originally part of the Abinit pseudopotential database.[133] Additional pseudopotentials for the other functionals (e.g., B86bPBE, vdW-DF1, ...) were generated using similar pseudopotential core-radii

Table 2.1: Cutoff radii, in Å, used in the generation of the PAW datasets for the elements present in the molecular crystals studied in this thesis.

Element	ψ_s	ψ_p
H	0.42	n/a
C, N, O	0.53	0.48
F	0.64	0.64
S, P	0.79	0.79

cutoffs to the ones for used with PBE.

The pseudised and all-electron wavefunctions were verified to match beyond the core radii specified for each valence orbital and, in addition, to yield reasonably similar energy eigenvalues for the free-atom calculations. Additional calculations testing transferability were performed by ensuring that the pseudopotentials were able to reproduce all-electron energy levels and wavefunctions in arbitrary environments, i.e., by examining various electronic configurations of each element and comparing the excitation energies produced from pseudopotential and all-electron calculations. The differences in excitation energies between the pseudopotential and all-electron calculations were generally found to be close to, or less than, 1 mRy, which is typically considered to be acceptable.[132]

2.2 Accounting for Non-Local Interactions Within the DFT Framework

All density-functional approximations of E_{xc} based on the local-density or generalized-gradient approximation are, by definition, devoid of describing nonlocal phenomena. As such, London dispersion, which arises from the attraction of instantaneous dipoles formed between two fluctuating electron densities, is not correctly described by these DFAs.[134–136] Being of paramount importance for the correct description of molecular solids,[137] methods which account for dispersion interactions are necessary.

In this regard, there have been long-standing efforts to develop methods capable of modelling dispersion interactions.[138–141] Two types of approaches are commonly taken to include dispersion in DFT: (i) the addition of a dispersion contribution to the DFT energy calculated *a priori*,[138] and (ii) the development of exchange-correlation functionals, E_{xc}, which contain an explicit contribution for dispersion in their formulation.[141]

2.2.1 Post-SCF Dispersion Correction Models

Post-SCF dispersion-correction models supplement the total DFT electronic energy, E_{DFT}, with a dispersion energy term, E_{disp}, of the form

$$E = E_{\mathrm{DFT}} + E_{\mathrm{disp}} = E_{\mathrm{DFT}} - \frac{1}{2}\sum_{\mathbf{L}}\sum_{ab}{}'\sum_{n}\frac{C_{n,ab}f(R_{ab\mathbf{L}})}{R^{n}_{ab\mathbf{L}}}, \tag{2.12}$$

where the $C_{n,ab}$ are the n-th order interatomic dispersion coefficients, $f(R_{ab\mathbf{L}})$ is a damping function that deactivates the dispersion interaction at short range, and

$$R_{ab\mathbf{L}} = |\mathbf{R_b} - \mathbf{R_a} + \mathbf{L}| \tag{2.13}$$

is the distance between atoms a and b in cells separated by lattice vector $\mathbf{L}$. The prime indicates that, for $\mathbf{L} = \mathbf{0}$, a cannot equal b. In practice, the sum over lattice vectors is truncated at a point such that all remaining interatomic contributions to the total dispersion energy fall below a specified energy threshold.

Three commonly used post-SCF dispersion correction models are: (i) the exchange-hole dipole moment (XDM) model,[139] (ii) the Grimme Dn (n = 2–4) models,[138] and (iii) the Tkatchenko-Scheffler (TS) and many-body dispersion (MBD) correction models.[140] The aforementioned dispersion corrections have all shown successes in providing reasonable accuracies (on the order of 4 kJ/mol or less) for computed lattice energies of molecular solids,[96,139,142–146] which neighbours the experimental uncertainty (i.e., 4.9 kJ/mol).[93,147]

These dispersion correction models differ in how the dispersion coefficients, $C_{n,ab}$, are computed, to which order they are summed (either pairwise, or many-body), what damping function is used, and whether or not additional (semiempirical) damping parameters are included in front of the summation in Eq. 2.12. These methods are now surveyed, with an emphasis placed on the XDM dispersion model.

2.2.1.1 Exchange-Hole Dipole Moment (XDM) Dispersion Correction Model

The XDM model has previously been implemented in the Quantum ESPRESSO solid-state code, which makes use of plane-wave basis sets.[142,148] It has demonstrated excellent accuracy in the treatment of non-covalent interactions in the gas phase,[149,150] as well as for surfaces,[151] layered materials,[152,153] and molecular crystals.[142]

The XDM dispersion coefficients are determined using second-order perturbation theory.[139,154,155] Dispersion forces are derived from interactions between instantaneous atomic multipole moments, which originate from the distribution of electrons plus exchange-hole dipoles.[155] The first three leading-order pairwise dispersion coefficients are

$$C_{6,ab} = \frac{\alpha_a \alpha_b \langle M_1^2 \rangle_a \langle M_1^2 \rangle_b}{\langle M_1^2 \rangle_a \alpha_b + \langle M_1^2 \rangle_b \alpha_a}, \tag{2.14}$$

$$C_{8,ab} = \frac{3}{2} \frac{\alpha_a \alpha_b \left(\langle M_1^2 \rangle_a \langle M_2^2 \rangle_b + \langle M_2^2 \rangle_a \langle M_1^2 \rangle_b \right)}{\langle M_1^2 \rangle_a \alpha_b + \langle M_1^2 \rangle_b \alpha_a}, \tag{2.15}$$

$$C_{10,ab} = 2 \frac{\alpha_a \alpha_b \left(\langle M_1^2 \rangle_a \langle M_3^2 \rangle_b + \langle M_3^2 \rangle_a \langle M_1^2 \rangle_b \right)}{\langle M_1^2 \rangle_a \alpha_b + \langle M_1^2 \rangle_b \alpha_a} + \frac{21}{5} \frac{\alpha_a \alpha_b \langle M_2^2 \rangle_a \langle M_2^2 \rangle_b}{\langle M_1^2 \rangle_a \alpha_b + \langle M_1^2 \rangle_b \alpha_a}, \tag{2.16}$$

where $\langle M_l^2 \rangle$ ($l = 1, 2, 3, ...$) are the expectation values of the l-th order exchange-hole multipole moments, and α_a is the atom-in-solid polarisability of atom a.

Combined, the three dispersion energy terms corresponding to the these coefficients have been shown to describe long-range interactions in solids accurately.[139,142] Thus, in the canonical XDM implementation, the summation over n in Eq. 2.12 is truncated at the $n = 10$ term, and only takes into consideration atomic-pairwise contributions. Nevertheless, it is important to note that the XDM dispersion model does take into account electronic many-body effects to all orders by way of the construction of the dispersion coefficients from the exchange hole, which is evaluated using the fully-interacting electron density.

In order to compute the XDM dispersion coefficients (Eqs. 2.14-2.16), the exchange-hole multipole moments and the atom-in-solid polarisabilities are needed:

$$\langle M_l^2 \rangle_a = \sum_\sigma \int \omega_{\mathrm{Hirsh},a}(\mathbf{r}) \rho_\sigma^{\mathrm{ae}}(\mathbf{r}) \left[r_a^l - (r_a - d_{X\sigma}(\mathbf{r}))^l \right]^2 \mathrm{d}\mathbf{r}, \tag{2.17}$$

$$\alpha_a = \frac{\int r^3 \omega_{\mathrm{Hirsh},a}(\mathbf{r}) \rho_\sigma^{\mathrm{ae}}(\mathbf{r}) \mathrm{d}\mathbf{r}}{\int r^3 \rho_a^{\mathrm{at}}(\mathbf{r}) \mathrm{d}\mathbf{r}} \alpha_a^{\mathrm{at}}, \tag{2.18}$$

where $d_{X\sigma}$ is the exchange-hole dipole moment, σ is a spin index, $\rho_\sigma^{\mathrm{ae}}$ is the all-electron spin-density, and r_a is the distance to atom a. ρ_a^{at} and α_a^{at} are the reference free-atom densities, and polarisabilities, respectively, and $\omega_{\mathrm{Hirsh},a}$ is the weight of that atom's contribution to the spin-density. The weights can, in principle, be constructed using any partitioning method. In XDM, the

Hirshfeld partitioning scheme is used:[156,157]

$$\omega_{\mathrm{Hirsh},a}(\mathbf{r}) = \frac{\rho_a^{\mathrm{at}}(\mathbf{r})}{\sum_b \rho_b^{\mathrm{at}}(\mathbf{r})}, \tag{2.19}$$

which is relatively simple to implement. The atomic polarisabilities (Eq. 2.18) are calculated from their *in vacuo* (free-atom) counterparts (α_a^{at}) by exploiting their proportionality with the atomic volumes.[158] The in-solid and free-atomic volumes are the numerator and denominator in Eq. 2.18, respectively. The all-electron spin-density, $\rho_\sigma^{\mathrm{ae}}$, is approximated from the valence spin-density by adding the core electron density.

In Eq. 2.17, $d_{X\sigma}$ is the dipole moment between the electron at the reference point $\mathbf{r}$ and its associated exchange-hole ($h_{X\sigma}$) given by

$$d_{X\sigma}(\mathbf{r}) = \int \mathbf{r}' h_{X\sigma}(\mathbf{r}, \mathbf{r}') \mathrm{d}\mathbf{r}' - \mathbf{r}. \tag{2.20}$$

The exact expression for $h_{X\sigma}$ is computationally prohibitive in solids because it involves a double sum over occupied states. Instead, the Becke-Roussel (BR) semi-local model for the spherically averaged exchange hole is used.[159] The BR hole provides a better approximation to the full exchange-correlation hole than $h_{X\sigma}$ and, consequently, it results in improved accuracy of the resulting molecular dispersion coefficients.[139,160]

The BR model hole has the form of an off-centred exponential function ($-Ae^{-ar}$) displaced from the electron's reference point by a distance b.[142] The three parameters (A, a, b) are determined by enforcing three exact constraints related to the hole normalisation and its value and curvature at the reference point. This leads to $b = d_{X\sigma}$, where b is calculated as

$$b^3 = \frac{x^3 e^{-x}}{8\pi\rho_\sigma^{\mathrm{val}}}, \tag{2.21}$$

with $\rho_\sigma^{\mathrm{val}}$ being the valence spin-density. The new variable $x = ab$ is the solution to the non-linear equation

$$\frac{xe^{-2x/3}}{x-2} = \frac{2}{3}\pi^{2/3}\frac{{\rho_\sigma^{\mathrm{val}}}^{5/3}}{Q_\sigma}, \tag{2.22}$$

which is solved numerically using Newton's method. The hole curvature (Q_σ) is

$$Q_\sigma = \frac{1}{6}(\nabla^2\rho_\sigma^{\mathrm{val}} - 2D_\sigma), \tag{2.23}$$

with

$$D_\sigma = \tau_\sigma - \frac{1}{4}\frac{|\nabla\rho_\sigma^{\mathrm{val}}|^2}{\rho_\sigma^{\mathrm{val}}}, \tag{2.24}$$

where τ_σ is the positive-definite valence spin kinetic-energy density,

$$\tau_\sigma = \sum_i |\nabla\psi_{i\sigma}|^2 . \tag{2.25}$$

It is important to ensure that the calculated value of b does not unphysically overshoot the distance to the closest nucleus.[142,155] For this reason, $d_{X\sigma}$ is set to $\min(b, r_a)$, instead of simply b, in Eq. 2.17.

Finally, returning to Eq. 2.1, the Becke-Johnson damping function[139,161] has the form

$$f(R_{ab\mathbf{L}}) = \frac{1}{R_{\mathrm{vdW},ab}^n + R_{ab\mathbf{L}}^n}, \tag{2.26}$$

where

$$R_{\mathrm{vdW},ab} = a_1 R_{c,ab} + a_2, \tag{2.27}$$

and

$$R_{c,ab} = \frac{1}{3}\left[\left(\frac{C_{8,ab}}{C_{6,ab}}\right)^{1/2} + \left(\frac{C_{10,ab}}{C_{6,ab}}\right)^{1/4} + \left(\frac{C_{10,ab}}{C_{8,ab}}\right)^{1/2}\right]. \tag{2.28}$$

The sum of van-der-Waals radii ($R_{\mathrm{vdW},ab}$) is constructed from a critical radius, $R_{c,ab}$, corresponding to the point where dispersion contributions from the first three leading-order pairwise dispersion coefficients, $C_{6,ab}$, $C_{8,ab}$, and $C_{10,ab}$, are equal. The a_1 and a_2 parameters are found by minimising the residual errors between computed and reference binding energies for a benchmark set of non-covalently bound dimers (the Kannemann-Becke set[162]). These parameters depend on the functional and they serve to match the long-range dispersion and short-range exchange-correlation contributions.

Lastly, from the dispersion energy expression (Eq. 2.12), it is possible to determine the dispersion contribution to the atomic forces and the stress tensor. For atom i, the force is

$$\mathbf{F}_{\mathrm{disp},a} = \sum_{\mathbf{L}}\sum_b\sum_n \frac{nC_{n,ab}R_{ab\mathbf{L}}^{n-2}}{(R_{\mathrm{vdW},ab}^n + R_{ab\mathbf{L}}^n)^2}\mathbf{R}_{\mathbf{abL}}, \tag{2.29}$$

and the components of the stress tensor are

$$\sigma_{\mathrm{disp},\eta\xi} = -\frac{1}{2V}\sum_{\mathbf{L}}\sum_{ab}{}'\sum_{n}\frac{nC_{n,ab}R_{ab\mathbf{L}}^{n-2}(R_{ab\mathbf{L}})_{\eta}(R_{ab\mathbf{L}})_{\xi}}{(R_{\mathrm{vdW},ab}^{n}+R_{ab\mathbf{L}}^{n})^{2}}, \tag{2.30}$$

where η, ξ are the Cartesian coordinates x, y, z and V is the unit-cell volume. Note that, in these expressions, the dispersion coefficients are assumed to be constant with respect to changes in the crystal geometry, which is not strictly correct. However, practise has shown that this approximation does not noticeably affect the geometry minimisation in most cases. In fact, it is more computationally efficient to calculate the dispersion coefficients at the first ionic step and then keep them constant throughout the geometry minimisation. "Relaxed" geometries are then subject to additional geometry optimisation calculations, until the newly computed dispersion coefficients, and consequently the total energy, cease to to change. The effect of fixing the dispersion coefficients at the first ionic step, as opposed to recalculating them at every step, has been tested and shown to yield equivalent geometries in both molecules and solids.

2.2.1.2 Grimme-Dn ($n = 2 - 4$) Dispersion Correction Models

Over the years, Grimme and co-workers have introduced the "D"-series of dispersion correction models[163] to obtain accurate approximations to molecular and atomic dispersion coefficients.[138,164]

In the Grimme-D2 model,[165] the dispersion coefficients, $C_{6,ab}$, are determined from the geometric mean of homoatomic dispersion coefficients $C_{6,a}$:

$$C_{6,ab} = \sqrt{C_{6,a}C_{6,b}}. \tag{2.31}$$

The homoatomic dispersion coefficients themselves are determined from DFT calculation of the ionization potentials, $\mathrm{IP_a}$, and the static dipole polarisabilities, α_a:

$$C_{6,a} = 0.05N\mathrm{IP}_a\alpha_a, \tag{2.32}$$

where N is the maximal number of electrons for the row in which atom a sits in the periodic table. Only the leading-order dispersion coefficients ($C_{6,ab}$) in the energy expression (see Eq. 2.1), are used. The damping function is

$$f(R_{ab\mathbf{L}}) = 1 + \exp^{-d(R_{ab\mathbf{L}}/R_{\mathrm{vdW},ab}-1)}, \tag{2.33}$$

where a value of $d = 20$ is suitable to correct for dispersion energies in the intermediate ranges,

while not affecting the short-range, or covalent bonding between atoms. $R_{\mathrm{vdW},ab}$ is the sum of atomic van-der-Waals radii. In addition, a DFA-specific global scaling factor, s_6, is used in front of the summation in Eq. 2.12 to optimize the performance of the dispersion-corrected functionals on atomization energies, ionization potentials, proton affinities, chemical reactions, and non-covalently-bound complexes.[165]

The D2 model has benefited from its computational expediency as, unlike the XDM dispersion model, the dispersion coefficients are not density-dependent or modified according to the electronic structure of the atoms in various chemical environments. This compromise, however, can lead to poor performance (or sometimes fortuitous error cancellation) when the atoms are in different environments than what was used to generate the atomic dispersion coefficients (a free-atom in its ground state). To partially address this issue, the Grimme-D3 model was developed.[166] This newer iteration of the Grimme dispersion model re-formulates the form of the molecular dispersion coefficients as a function of a simulated "coordination number" of the atom in its environment. While this improvement remains electronic-structure independent, its usage introduces some flexibility in describing an atom in various environments, all while retaining computational simplicity.

In the D3 model, the dispersion coefficients are computed from the Casimir-Polder relationship via time-dependent DFT calculations performed on reference stable hydrides $\mathrm{A}_m\mathrm{H}_n$ and $\mathrm{B}_k\mathrm{H}_l$:

$$C_{6,ab} = \frac{3}{\pi}\int_0^\infty \frac{1}{m}\left[\alpha_{\mathrm{A}_m\mathrm{H}_n}(a\omega) - \frac{n}{2}\alpha_{\mathrm{H}_2}(a\omega)\right] \times \left[\alpha_{\mathrm{B}_k\mathrm{H}_l}(a\omega) - \frac{l}{2}\alpha_{\mathrm{H}_2}(a\omega)\right] \mathrm{d}\omega, \tag{2.34}$$

with α being the averaged dipole polarisability at imaginary frequency ω. The higher-order C_8 dispersion coefficients are then determined recursively:

$$C_{8,ab} = 3C_{6,ab}\sqrt{Q_a Q_b}, \tag{2.35}$$

where $Q_a \equiv s_{42}\sqrt{Z_a}\frac{\langle M_4\rangle_a}{\langle M_2\rangle_a}$ is an effective charge for atom a. This definition of Q_a involves the expectation values of the lth-order multipole moments, $\langle M_l\rangle$, and a nuclear charge factor, Z_a, along with a scaling factor, s_{42}, chosen to give reasonable C_8's for noble-gas atoms. The reason for evaluating these coefficients for molecules with respect to their hydridic state (which exists for every element in the periodic table, except rare gases) is to be able to afford a proper description of the polarisabilities within any given system, as opposed to using free-atom polarisabilities. These coefficients are then adapted further to the chemical environment of each atom in the system by

considering its "coordination number", CN, which is defined as

$$\mathrm{CN}_a = \sum_{a \neq b}^{N} (1 + \exp^{-d_1(d_2(R_{\mathrm{cov},a}+R_{\mathrm{cov},b})/R_{ab}\mathbf{L}-1)})^{-1}. \tag{2.36}$$

$R_{\mathrm{cov},a}$ and $R_{\mathrm{cov},b}$ are scaled covalent, single-bond, radii of atoms a and b, respectively. The d_1 and d_2 parameters are scaled to yield appropriately meaningful results for the coordination numbers (e.g., between 2 and 4 for carbon in various molecules). These "coordination numbers" are first calculated for a particular system, and the values of the pairwise dispersion coefficients (Eq. 2.34) are then determined by interpolating between the coordination values and dispersion coefficients computed from reference data.

Additionally, in order to account for interatomic interactions that are beyond pairwise and non-additive, Grimme-D3 models can be supplemented with a three-body (Axilrod–Teller–Muto, ATM) term

$$E_{abc} = \frac{C_{9,abc}(3\cos\theta_a \cos\theta_b \cos\theta_c + 1)}{(R_{ab}R_{bc}R_{ac})^3}, \tag{2.37}$$

where the three-body dispersion coefficients, $C_{9,abc}$, are approximated to be the geometrical mean of the homoatomic C_6 coefficients. θ_a, θ_b, θ_c are the internal angles of the triangle formed by the atoms separated by the three distances, R_{ab}, R_{bc} and R_{ac}.

Contrary to the D2 dispersion model, the newer iteration includes dispersion coefficients up to second-order (i.e., truncation of the terms in the summation for $C_{n,ab}$ in Eq. 2.12 includes $n = 6, 8$). For the D3 model, the choice of damping function was re-examined and, ultimately,[167] choosing a Becke-Johnson damping function (Eq. 2.26) proved satisfactory. Global scaling factors remain in use, with s_6 being set to 1, and s_8 (placed in front of the second-order term in Eq. 2.12) being DFA-dependent and fitted via a least-squares minimization procedure to yield the best results on a training set.[166]

Although initially deemed to be of lesser relevance as opposed to geometrical considerations (e.g., coordination number),[166] the oxidation/charge state of elements has been found to affect the magnitude of dispersion coefficients, and neglect of this is one of the main limitations of the D3 model.[164] The newest iteration of this series of dispersion corrections (the D4 model), which includes an oxidation-state dependence, has been recently developed for molecules, but remains to be fully applicable to solids.[164] Systems treated in this thesis would likely not benefit from inclusion

of oxidation-state-dependent dispersion coefficients, as would, e.g., systems with transition metal elements with highly ionized character.[168] As such, no more information is given on this revision of the Grimme-type model; rather, see Ref. 164 for more details.

2.2.1.3 Tkatchenko-Scheffler (TS-vdW) and Many-Body Dispersion (MBD) Correction Model

The last two dispersion correction schemes to be discussed are the TS[169–171] and MBD[172,173] schemes.

In the TS scheme, dispersion coefficients are computed from the mean-field electron density, instead of being computed from ground-state orbitals (as in XDM) or from time-dependent DFT (as in D3). The coefficients themselves are derived from the Casimir-Polder relation, and can ultimately be written in terms of the free-atom $C_{6,a}$'s and the static free-atom polarisabilities, α_a:

$$C_{6,ab} = \frac{2C_{6,a}C_{6,b}}{\left[\frac{\alpha_b}{\alpha_a}C_{6,a} + \frac{\alpha_a}{\alpha_b}C_{6,b}\right]}. \tag{2.38}$$

In order to establish the dispersion coefficients of atoms within molecules, the same Hirshfeld partitioning scheme (refer back to Eq. 2.19) as is utilized in XDM[155] is used. This approach derives effective (atom-in-molecule) polarisabilites and volumes from free-atom quantities,[158] which are then used to scale free-atom dispersion coefficients to effective dispersion coefficients.

This dispersion scheme was further improved by taking into account the effects of distant interacting fluctuating dipoles, which cannot be computed with dispersion coefficients based on exponentially decaying electronic densities.[170] These revisions of the TS scheme, known as TS-vdW + SCS (or TS with self-consistent screening)[170,171] and MBD (many-body dispersion),[172,173] describe many-body atomic contributions at long-range and are found to generally improve the results for dispersion coefficients of atoms within solids over the initially proposed TS model.[171–173]

2.2.2 Non-Local van-der-Waals Density Functional Models

An alternative approach to account for non-local interactions within the DFT framework is to include a non-local term in the formulation of the exchange-correlation functional, E_{xc} (Eq. 2.1).[141] Generally, the form of the non-local correlation term, $E_{\mathrm{c}}^{\mathrm{nl}}$, is given by

$$E_{\mathrm{c}}^{\mathrm{nl}}[\rho] = \int \mathrm{d}^3\mathbf{r} \int \rho(\mathbf{r}) K(\mathbf{r},\mathbf{r}') \rho(\mathbf{r}') \mathrm{d}^3\mathbf{r}, \tag{2.39}$$

where $K(\mathbf{r}, \mathbf{r}')$ is a non-local kernel serving as a response function, which accounts for the interactions of electron densities at two separate points ($\mathbf{r}$ and $\mathbf{r}'$). As the non-local term is added to E_{xc}, van-der-Waals interactions are incorporated into the self-consistent field equations that are solved to establish the electron density of a system. This brings about a significant computational overhead to the use of these methods, as opposed to the post-SCF methods discussed in previous sections, as the non-local term must be evaluated at every SCF iteration throughout the calculation. On the other hand, total DFT densities tend to be affected only slightly upon inclusion of non-local contributions during the SCF calculation,[174,175] such that the post-SCF summation remains a reasonable approach to account for dispersion forces in molecular and solid-state systems.

Many formulations of the non-local kernel have been developed over the years,[141] including the vdW-DF1 and vdW-DF2 non-local functionals of Langreth, Lundquvist, and co-workers,[176,177] and the rVV10 functional of Vydrov and Van Voorhis.[178] All of them differ not only in the formulation of the non-local kernel, but also in which exchange and correlation functionals are used to construct the kernel along with the base (semi-local) functional.[141] Improved description of geometries can be obtained if the base exchange functional has a better treatment of non-bonded repulsion. For instance, vdW-DF2, which uses revPW86 as its base functional, yields better results in this regard than does vdW-DF1, which uses revPBE.[141]

The VV class of vdW-DF functionals, such as rVV10,[178] make several approximations and/or use relaxed constraints in formulating the non-local kernal, and introduce empirical parameters fitted to training sets of small van-der-Waals molecular dimers. This has been noted to result in the functional being less transferable to bulk systems or layered materials, although the parameters can be tuned to yield better results for specific classes of systems.[141]

2.3 Low-Cost Methods as Alternatives to Plane-Wave DFT-D Methods

Over the years, several approximate approaches have been developed as alternatives DFT-D and post-HF theory, with the aim of making calculations cheaper and/or applicable to larger systems. Two such classes of approximations, density-functional tight-binding and minimal-basis set approaches, are used in parts of the thesis work, and will now be described in detail.

2.3.1 Density-Functional Tight-Binding Methods

Within density-functional tight-binding (DFTB) methods,[95,179] the electron density of an N-electron system, ρ, is first approximated to be equivalent to the sum of the free-atom densities (a.k.a the Harris density) within the system, i.e., $\rho_0 \equiv \sum_a \rho_0^a$. In order to account for interactions between atoms, ρ_0 is subsequently allowed to fluctuate by some small amount $\delta\rho$, such that the actual density of the system, ρ, is recovered, i.e., $\rho = \rho_0 + \delta\rho$. Then, the DFT expression for the total energy, Eq. 2.1, is rewritten in terms of these free-atom densities and density fluctuations, and Taylor-expanded to arbitrary order in terms of $\delta\rho$:

$$E_{\mathrm{DFTB}} \equiv E[\rho] = E^{(0)}[\rho_0] + E^{(1)}[\rho_0, \delta\rho] + E^{(2)}[\rho_0, (\delta\rho)^2] + E^{(3)}[\rho_0, (\delta\rho)^3] + ... \tag{2.40}$$

Eq. 2.40 is considered to be exact and equivalent to the DFT formalism, as effectively no approximation has yet been made. However, in practice, truncation of Eq. 2.40 is usually resolved to second[180] or third order.[181,182]

Substituting for each term in Eq. 2.1, following the first-order Taylor expansion of Eq. 2.40, yields

$$E_{\mathrm{DFTB}} = E^{(0)}_{\mathrm{DFT}} + \frac{1}{2}\sum_{ab} E^{\mathrm{rep}}_{ab}, \tag{2.41}$$

where E^0_{DFT} is the energy contribution due to the reference density of the free atoms, and E^{rep}_{ab} is a term introduced to reflect repulsion between nearest-neighbour pairs of atoms a and b. Taking only these two terms yields the standard tight-binding approach.[180] Here, requiring only free-atom densities to perform the calculations means that parameter files can be tabulated for some reference systems and used subsequently to study other systems of interest. The '3ob' parameter set of "Slater-Koster" files, which consist of parameters for atoms useful to study of organic and biomolecules,[183–185] consists of pair-pair potentials for C, H, N, O, halogens, and some alkali- and rare-earth s-block metals. Having pre-computed the required integrals and pair-pair potentials in Eq. 2.41 in an *a priori* fashion affords DFTB a substantial speed-up of 2-3 orders of magnitude over performing DFT calculations. However, its successful use relies on the quality of the input guess for the reference density (in this case, the Harris density).

To obtain higher accuracy, higher-order terms in $\delta\rho$ must be included in the expansion of Eq. 2.40. Truncation of Eq. 2.40 to second order in $\delta\rho$ yields E_{DFTB2}:[180]

$$
\begin{aligned}
E_{\mathrm{DFTB2}} &= E_{\mathrm{DFTB}} + \frac{1}{2}\int\int\left(\frac{1}{r_{12}} + \frac{\delta^2 E_{\mathrm{xc}}[\rho]}{\delta\rho(\mathbf{r})\delta\rho(\mathbf{r}')}\right)\delta\rho(\mathbf{r})\delta\rho(\mathbf{r}')\mathrm{d}\mathbf{r}\mathrm{d}\mathbf{r}' \qquad (2.42)\\
&\approx E_{\mathrm{DFTB}} + \frac{1}{2}\sum_{ab}\Delta q_a \Delta q_b \gamma_{ab}.
\end{aligned}
$$

The last term in the equation above is the result of making a series of approximations to the Coulomb and exchange-correlation terms when they are expanded to second order in $\delta\rho$. Essentially, the fluctuations in density, $\delta\rho$, are approximated to be fluctuations in point (monopole) charges, Δq_a and Δq_b. The γ_{ab} function reduces to a pure Coulomb interaction between the two monopole charges at long-range ($a \neq b$) whereas in the short-range ($a = b$), the function is approximated to equal twice the chemical hardness of an element, or its Hubbard parameter, U_a. This latter approximation is made to avoid computing how E_{xc} varies with regard to density fluctuations. The Hubbard parameters are also pre-computed from DFT calculations and stored in the Slater-Koster files for each atom type. Inclusion of this charge dependence within the second-order tight-binding scheme improves the description of systems having non-Harris-like charge densities, such as polar molecules.[180]

The use of a single Hubbard parameter within the γ_{ab} function, however, limits the applicability of second-order DFTB to neutral molecules. Extension of DFTB to third order in $\delta\rho$ yields

$$
\begin{aligned}
E_{\mathrm{DFTB3}} &= E_{\mathrm{DFTB2}} + \frac{1}{6}\int\int\int\frac{\delta}{\delta\rho(\mathbf{r}'')}\frac{\delta^2 E_{\mathrm{xc}}[\rho]}{\delta\rho(\mathbf{r})\delta\rho(\mathbf{r}')}\delta\rho(\mathbf{r})\delta\rho(\mathbf{r}')\delta\rho(\mathbf{r}'')\mathrm{d}\mathbf{r}\mathrm{d}\mathbf{r}'\mathrm{d}\mathbf{r}'' \qquad (2.43)\\
&\approx E_{\mathrm{DFTB2}} + \frac{1}{3}\sum(\Delta q_a)^2 \Delta q_b \Gamma_{ab}.
\end{aligned}
$$

The density fluctuations are again expressed as monopole charge fluctuations, as was done for Eq. 2.42. Considering only one-centre terms (a=b=c), all other features of these charge fluctuations (e.g., exponential decay as a function of interatomic distance) yield a function Γ_{ab}, which incorporates a charge-dependent Hubbard parameter, i.e., $\partial U_a/\partial q_a$. Therefore, an improved description of systems which have atoms in various charged states, such as, such as zwitterions, can be made by using third-order DFTB methods.[181,182]

Finally, while DFTB methods are several orders of magnitude faster than DFT methods, they suffer from limitations which are hard to improve upon systematically.[95,179] On the one hand, as pre-computed terms (E^0_{DFT}, E^{rep}_{ab}, U_a, and $\partial U_a/\partial q_a$) are determined from PBE calculations, DFTB has the same limitations as the parent functional: (i) Delocalisation error will be an issue

for DFTB as it is for typical GGA functionals, and (ii) dispersion interactions are not described within the DFTB formalism and it must therefore be supplemented with a dispersion correction (e.g, D3). On the other hand, the approximations made while establishing Eq. 2.43 can introduce random error within the results obtained from DFTB: (i) Multicentre interactions being neglected in the evaluation of the zeroth-/first-order terms, (ii) approximating charge fluctuations as being monopole/point-charges, and (iii) using minimal-basis sets to perform the calculations cheaply and efficiently.

The approach just described in this section (ultimately, self-consistent-charge third-order DFTB, or SCC-DFTB3) and used in part of this thesis work was developed by Eltsner and co-workers.[179] Other tight-binding approaches are also under development, such as Grimme's xTB method.[186,187]

2.3.2 Minimal-Basis-Set Approaches

As an alternate to approximations made to the theoretical foundations of computational methods, acceleration of calculations can also be achieved through the use of minimal or small basis sets when expanding the wavefunction or the electron density.[90] However, speed-up comes at the cost of introducing basis-set superposition error (BSSE) and basis-set incompleteness error (BSIE)[188] to the calculation of the energy and properties of a system.

BSSE leads to artificial overstabilisation (or overbinding) of vdW-bound complexes when basis functions centred on atoms of neighbouring molecule(s) are used to artificially stabilize the parent molecule, which leads to a far more stable complex than the sum of its isolated constituents. Conversely, the use of an incomplete basis set effectively introduces a BSIE because short- to intermediate-range interactions are not described properly, as they would be by a complete basis set. This latter error is always present for any calculations using a finite basis set, but is often reduced significantly by using appropriately sized and flexible basis sets. Thus, BSIE tends to be largest when the most compact basis sets are used.

In contrast to the errors that arise from DFTB methods,[95] errors that stem from using minimal basis sets can be corrected much more systematically. The standard approach to address BSSE is to apply a counterpoise (CP) correction.[189] This correction consists of performing additional calculations for each constituent molecule in the presence of "ghost" basis functions from the other molecules of the complex (i.e., basis functions centred in place of the atoms, but without actual

nuclei present). If a dimeric system, AB, is considered the counterpoise correction is

$$E_{\mathrm{bsse}} = (E_{\mathrm{A}}^{\mathrm{A}} - E_{\mathrm{A}}^{\mathrm{AB}^*}) + (E_{\mathrm{B}}^{\mathrm{B}} - E_{\mathrm{B}}^{\mathrm{A}^*\mathrm{B}}). \tag{2.44}$$

$E_{\mathrm{A}}^{\mathrm{A}}$ and $E_{\mathrm{B}}^{\mathrm{B}}$ are the energies of the monomers obtained using the basis sets associated with each monomer only, whereas $E_{\mathrm{A}}^{\mathrm{AB}^*}$ and $E_{\mathrm{B}}^{\mathrm{A}^*\mathrm{B}}$ are the energies of the monomers obtained using basis sets supplemented with the basis functions of the other monomers (denoted by the $*$ superscript).

A more recent method to account for BSSE is the geometrical counterpoise (gCP) correction suggested by Grimme and co-workers,[190,191] Here, the correction is applied based on information provided by the geometry of the system only. The sum of atomic contributions to the gCP correction is obtained from calculating *a priori* the energetic bias that the use of a smaller versus larger basis set will have on a system. This error is then weighed by using a decaying exponential function based on the inter-atom distance and normalized by the extent that basis functions on pairs of atoms overlap. The method has four semiempirical parameters fitted for various combinations of basis sets and quantum-mechanical methods (HF or DFT). Approaches to correct for BSIE, on the other hand, involve the addition of either a single or linear combination of Gaussian functions (with fit parameter(s)) to the Hamiltonian of a system, such that the error due to the short-range nature of the basis (SRB) is corrected.[188] The fit parameters of these Gaussian-type functions are obtained from minimizing the errors between using small and large basis sets on various training data.

The "3c" class of methods developed by Grimme and co-workers over recent years are minimal- (or small-) basis-set methods that can be used to cheaply study solids. For instance, HF-3c[192,193] performs Hartree-Fock calculations with a MINI(x) minimal basis set, applying several corrections to counteract the errors introduced: (i) a gCP correction is used to reduce the BSSE,[193] (ii) a SRB correction is added to reduce the BSIE, and (iii) a D3 correction is applied to treat dispersion interactions. In a similar vein, PBEh-3c[193] is a minimal-basis DFT method employing the PBE GGA functional in conjunction with a tailored double-ζ basis set (def2mSVP), with additional gCP and D3 corrections. PBEh-3c also mixes in 42% HF exact exchange, and thus functions as a global hybrid functional, to reduce delocalisation error. The D3 correction uses Becke-Johnson damping, and includes the three-body ATM correction. Finally, HSE-3c was developed as a more efficient method than PBEh-3c, which includes variable exact exchange, gradually switching from 42% exact exchange in the short range to 0% in the long range via an error function. That said, all other features of HSE-3c remain consistent to those of PBEh-3c.

CHAPTER 3

EVALUATION OF SHEAR-SLIP TRANSITIONS IN CRYSTALLINE ASPIRIN BY DENSITY-FUNCTIONAL THEORY

Reprinted with permission from **L. M. LeBlanc***, A. Otero-de-la-Roza*, E. R. Johnson*. Evaluation of Shear-Slip Transitions in Crystalline Aspirin by Density-Functional Theory. *Cryst. Growth Des.* **2016**, *16*, 6867-6873. DOI: https://pubs.acs.org/doi/full/10.1021/acs.cgd.6b01038.

Contributions to the manuscript LML performed all the calculations, and wrote the first draft of the manuscript. AOR and ERJ contributed to the final version of the manuscript. ERJ supervised the project.

3.1 Motivation

Aspirin (*o*-acetylsalicylic acid)[194] has received considerable attention over the last decade, largely due to the prediction of a second nearly degenerate polymorphic crystalline form[195] and the subsequent work leading to its isolation in "pure" form.[196–200] This second form, further referred to as aspirin-II or form-II, is nearly energetically degenerate to the first solved crystal structure[194] (aspirin-I or form-I), with the lattice-energy difference between the two forms being less than half a kJ/mol.[201,202]

The structural differences between the two crystal forms are subtle. In both forms, aspirin molecules are arranged in centrosymmetric dimers, interacting via strong carboxyl O–H$\cdots$O hydrogen bonds, which are then arranged in bilayers propagating along the bc-plane. The bilayers are periodically stacked along the a-axis. Along the c-axis, neighbouring dimers have alternating orientations and are inclined with respect to the ac-plane.[203,204]

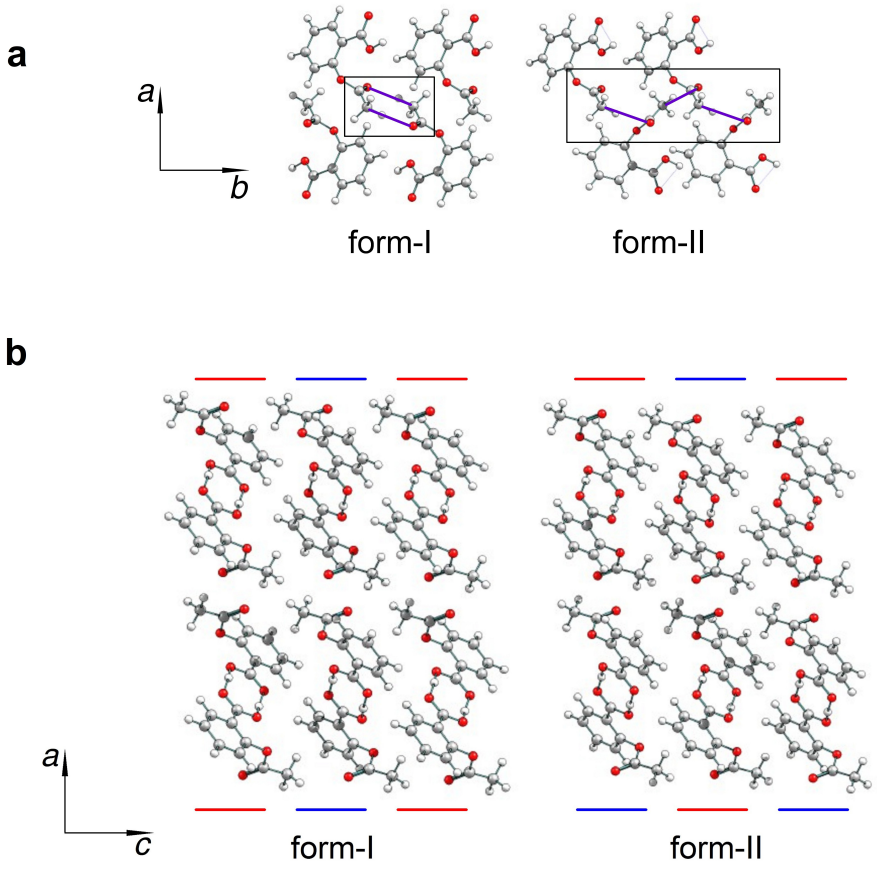
One of the key differences between the two forms, shown in Figure 3.1, lies in whether the pseudo-hydrogen-bonding C–H$\cdots$O motif between the interacting acetyl and carboxyl groups of two neighbouring bilayers along the b-axis is dimeric (form-I, inversion centres) or catemeric (form-II, 2_1 screw axes).[196,205] Indeed, the similarity of these interactions has been proposed as the reason for the near-degeneracy of the two forms of aspirin. Form-I possesses the more-favourable intermolecular dimeric configuration, while intermolecular C–H$\cdots$O bonding cooperativity favours form-II.[201] The other difference lies in the observation that two neighbouring layers of aspirin dimers along the $\{100\}$ plane are related by translation of one layer relative to the other by $c/2$ along the crystallographic c-axis, as shown in Figure 3.1.[197–200,204]

The relationship between the two forms makes choosing the proper unit cell crucial during refinement of X-ray diffraction data, as was illustrated by an unsatisfactory analysis of the data and the premature claim that aspirin-II could be prepared as a stable crystalline solid.[196,197] This was closely followed by proposal of an intergrowth structure for aspirin, in which domains of form-I and form-II are present within the same crystal in variable ratios.[198] Crystal composition and disorder in these intergrowth structures were later detected by the extent of diffuse scattering and/or streaks in the diffraction patterns.[198,203,206]

The mechanical stability (or instability) of the two forms has also been debated in multiple experimental[196,199,200,204,207] and computational[195,207,208] studies. The initial prediction of the form-II structure was accompanied by the calculation of a low shear coefficient (using an analytic second-derivatives method for rigid molecules and anisotropic atom-atom potentials as implemented in the DMAREL crystal structure modelling program).[209] This supported its mechanical instability and potential conversion to form-I, if the latter turned out to be thermodynamically favoured.[195] However, other studies of the elastic and thermal properties,[207] including density-functional computation of the bulk and shear modulus values,[208] have since showed no evidence of shear instability.

In 2011, a method for the preparation of "pure" aspirin-II (to the detection limit of the X-ray

Figure 3.1: Key structural differences between the two forms of aspirin. (a) Dimer/catemer motifs viewed along the b-axis. (b) 2D-bilayers of aspirin dimers viewed along the c-axis. The red and blue lines highlight the different molecular orientations with respect to the ac-plane.



diffraction method used) was reported by crystallizing aspirin in the presence of aspirin anhydride.[199] Studies of these single crystals at pressures of up to 2.2 GPa indicated no phase transition from form-II to form-I, thus providing another piece of evidence for the mechanical stability of the new form.[199] A recent experimental study of the behaviour of both forms of aspirin under higher pressures was conducted and coupled with micro-Raman spectroscopy.[204] Echoing previously published results,[199] no phase transition was observed[204] even upon application of pressures of up to 10 GPa to aspirin-II crystals, which is consistent with the established behaviour of the polymorphs of chlorpropamide under similar conditions.[210] However, the use of a nanoindentation technique on these "pure" form-II crystals led to the detection of form-I domains in some of the form-II crystals produced.[200] It was further noted that these crystals would convert to aspirin-I over *a period of several months* through the observation that characteristic peaks for form-II would

disappear over time in X-ray diffraction spectra of bulk powders stored under ambient conditions. When subjecting these same crystals to mechanical grinding (i.e., under the influence of shear stress), quick transformation from form-II to form-I took place.[200] This implies that slip along the $\{100\}<001>$ system in aspirin is the most likely phase-transition mechanism between the two forms. Additionally, calculations of attachment energies between adjacent dimeric layers along the a-axis (using the Dreiding force field model) also revealed the $\{100\}$ planes to be the most weakly bound and thus more susceptible to slip,[200] in agreement with previous models for calculating attachment energies in organic solids.[211]

To summarize, the near-degeneracy of both forms of aspirin and the evaluation of their thermodynamic stabilities[201,202,208] supports the formation of intergrowth structures.[198,212] It does not, however, give sufficient foundation to explain the metastability of form-II and its eventual conversion to form-I under ambient conditions in the presence or absence of shear stress.[200] Investigation of the potential ease with which interconversion between polymorphic forms through slip systems along one dimension (i.e., polytypes) can occur is necessary to better understand this class of compounds.

The aim of this chapter is to resolve contrasting experimental and computational results of whether or not form-II is mechanically stable under compression or shear stress. The viability of the $\{100\}<001>$ slip mechanism is considered using periodic-boundary dispersion-corrected density-functional theory. The barriers for interconversion between the two forms, at equilibrium conditions and subject to applied pressure, are calculated and related to the observed conversion time.[200] The limited applicability of near-degenerate, layered (i.e., polytypical) systems, such as aspirin, as benchmarks for polymorphism is also discussed.

3.2 Computational Methods

The crystal structures of aspirin-I and aspirin-II were retrieved from the Cambridge Structural Database (CSD):[22] form-I (CSD refcode ACSALA14), and form-II (CSD refcode ACSALA22). These were fully optimized using Quantum ESPRESSO,[127] version 5.1, with the PAW method[124] and the plane-waves / pseudopotentials approach.[213,214] The cutoff energy and density were 60 and 600 Ry, respectively. A $2 \times 2 \times 2$ MP $\mathbf{k}$-point grid was used in all cases, for both the single cells and doubled supercells (see below). The use of a denser $4 \times 4 \times 4$ $\mathbf{k}$-point grid had no significant effect on the relative energies of the two forms of aspirin. Specifically, for the single cells, an

energy difference of $\Delta E_{\mathrm{I}\rightarrow\mathrm{II}}$ = -0.29 kJ/mol per molecule with a 2×2×2 **k**-point grid vs. -0.31 kJ/mol per molecule with a 4×4×4 **k**-point grid was found.

Calculations employed the B86bPBE functional[109,110] with the XDM dispersion correction,[139,142,148,155] in which the dispersion energy is evaluated as a sum over all atomic pairs and includes C_6, C_8, and C_{10} terms. The values of the two parameters in the XDM damping function were set to a_1 = 0.6512 and a_2 = 1.4633 Å.

To account for thermal effects in both forms of aspirin, the phonon frequencies at Γ and the corresponding phonon density of states for the relaxed structures were calculated using the phonopy program.[215] The phonopy program uses a finite-displacement approach to compute and diagonalize the dynamical matrix, $\Lambda_{a,b}$ (a mass-weighted Hessian, $\Phi_{a,b}$, times a phase factor), which yields phonon frequencies, ω, for the crystal system:

$$\Lambda_{a,b} = \frac{\Phi_{a,b}}{\sqrt{m_a m_b}} e^{i\mathbf{q}\cdot|\mathbf{r_b}-\mathbf{r_a}|}, \tag{3.1}$$

where m_a is the mass of atom a, $\mathbf{q}$ is a point in reciprocal space, and $\mathbf{r}$ is the equilibrium position of atom a. In order to determine the vibrational frequencies of the crystal system, the Hessian must first be computed: A series of calculations is performed in which each atom, a, is displaced by some small value around its equilibrium position and the effect of moving all other atoms, b, by some small value is then evaluated. This leads to (at most) $3N \times 3N$ calculations for an N-atom system, although the phonopy program takes crystal symmetry into account when determining the minimal set of force calculations to be performed by an external program (in this case, Quantum ESPRESSO). Supercells of the reference unit cell are typically constructed in these finite-displacement methods, to avoid image interactions when atoms are displaced from their equilibrium positions, and the forces calculated.

The resulting phonon frequencies of the system, which were calculated using a harmonic approximation, can then be used to compute vibrational contributions to the Helmholtz free energies, $F_{\mathrm{vib}}(T)$, at a given temperature, T. $F_{\mathrm{vib}}(T)$, under the harmonic approximation,[216,217] is defined as

$$F_{\mathrm{vib}}(T) = \frac{1}{2}\sum_{i,\mathbf{q}} \hbar\omega_{i,\mathbf{q}} + k_B T \sum_{i,\mathbf{q}} \ln\left[1 - \exp\left(\frac{-\hbar\omega_{i,\mathbf{q}}}{k_B T}\right)\right], \tag{3.2}$$

where the sum spans all phonon modes and **q**-points. k_B is the Boltzmann constant and $\hbar$ is the

reduced Planck constant. The first term in Eq. 3.2 is the zero-point energy and the second term represents the thermal contributions to the free energy. The Hemholtz free energy, $A(T)$, of a crystal structure, is then given

$$A(T) = E_{latt} + F_{\mathrm{vib}}(T), \tag{3.3}$$

where E_{latt} is the electronic lattice energy of the crystal.

The unit cell (Z = 4) of aspirin-I was doubled along the <100> direction in order to investigate the {100}<001> slip mechanism. In this supercell, one layer of dimers was translated relative to the other in increments, δ_c, of $c/20$ (c being the lattice parameter of the unit cell) along the <001> direction up to $c/2$, i.e., resulting in a supercell of form-II. Each intermediate structure along the reaction coordinate was then optimized, allowing for relaxation of the cell. The incremented scan coordinate (the crystallographic <001> direction or the equivalent Cartesian z-coordinate) was frozen so as to avoid reversion to either form-I or form-II upon optimization. Because of this, the described procedure was performed starting from an aspirin-II supercell as well, in order to ensure that there was no bias due to constraints imposed by the choice of initial geometry during optimization.

3.3 Results and Discussion

3.3.1 Relative Energetics

The experimental sublimation enthalpy is only available for aspirin-I,[202,218] so there is no reference data for the lattice-energy difference between the two forms. The lattice energy of aspirin-I obtained with the XDM-corrected periodic-boundary DFT approach, shown in Table 3.1, is in fairly good agreement with the experimental result[219] (back-corrected for thermal effects) and the best reference RI-MP2 results,[201] further supporting the good performance of DFT-XDM for molecular crystals seen in previous works.[128,142] Additionally, the computed structural parameters (shown in Table 3.2) are well within typical deviations from experimental data given that thermal and vibrational contributions have been neglected.[201,208,220]

As shown in Table 3.3, aspirin-I is predicted to be 0.3 kJ/mol per molecule less stable than aspirin-II with B86bPBE-XDM. Including zero-point vibrational energies decreases this difference slightly by 0.1 kJ/mol. These results compare well with previous predictions using RI-MP2,[201]

Table 3.1: Lattice energies, E_{latt} in kJ/mol per molecule, of form-I aspirin calculated using various theoretical methods. The experimental value at 0 K is also shown.

Method	E_{latt}
RI-MP2/aDZ [a, 201]	-113.7
RI-MP2/aTZ [a, 201]	-132.1
SCS(MI) RI-MP2/aDZ [a, 201]	-132.5
SCS(MI) RI-MP2/aTZ [a, 201]	-135.6
MC MP2C [202]	-116.1
B86bPBE-XDM	-124.2
Experiment [219]	-115.0 [b]

[a] The RI-MP2 calculations used the many-body expansion. aDZ and aTZ indicate the aug-cc-pVDZ and aug-cc-pVTZ basis sets, respectively. [b] This value was obtained based on the heat of sublimation of form-I of -109.7 $\pm$ 0.5 kJ/mol at 298 K. [218]

Table 3.2: Unit-cell parameters, including cell lengths (a, b, and c in Å), angles (β in degrees), volumes (V in Å^3) and temperatures (T in K) for the two crystal forms of aspirin.

Form	a	b	c	β	V	T	CSD refcode
B86bPBE-XDM							
I	11.132	6.544	11.200	96.90	810.0	0	
II	11.891	6.448	11.270	110.77	808.6	0	
Experiment							
I	11.278	6.552	11.274	95.84	828.7	123	ACSALA14 [197]
I	11.446	6.596	11.388	95.55	855.7	298	ACSALA [194]
I	11.430	6.591	11.395	95.68	854.2	298	ACSALA01 [221]
II	12.102	6.472	11.334	111.59	825.4	123	ACSALA19 [200]
II	12.152	6.506	11.368	111.57	835.8	180	ACSALA22 [198]
II	12.358	6.532	11.496	112.44	857.7	298	ACSALA20 [200]

MP2C, [202] and PBE with both the TS and many-body dispersion corrections. [208] Like PBE-TS, B86bPBE-XDM predicts form-II to be marginally more stable, while the RI-MP2 calculations slightly favour form-I. However, it is not within the scope of this paper to establish the relative stability of these nearly degenerate forms to a sub-kJ/mol accuracy. Small deviations of the computational results from the nominal ordering of the two forms ($\Delta E_{\mathrm{I}\longrightarrow\mathrm{II}} > 0$, i.e., form-I energetically favoured) are certainly within the expected errors of the methods used.

Additionally, the thermodynamic preference for form-I has recently been attributed to thermal effects from the lattice vibrations, due to a change in the C–H$\cdots$O pseudo-hydrogen-bonding motifs between dimeric layers along the $\{100\}$ plane. [208] Accounting for thermal effects with

Table 3.3: Relative energies, in kJ/mol per molecule, of the two forms of aspirin using selected levels of theory. ΔE corresponds to the difference in electronic energies between forms II and I, i.e., $\Delta E_{(\mathrm{I}\longrightarrow\mathrm{II})} = E_{\mathrm{II}} - E_{\mathrm{I}}$, whereas $\Delta G(0\ K)$ and $\Delta G(298\ K)$ correspond to the sum of the electronic energy and the zero-point vibrational energy at 0 K, and to the sum of electronic energy with an additional thermal free-energy correction at 298 K, respectively.

Method	ΔE	ΔG(0 K)	ΔG(298 K)
QM/MM[195]	-0.2		
COMPASS MM[207]	-1.2		
B3LYP-D/6-31G(d,p)[222]	-2.5		
B3LYP-D/6-311G(d,p)[222]	-1.9		
B3LYP-D*/6-31G(d)[201]	-2.5		
B3LYP-D*/TZP[201]	-1.4	-0.4	
RI-MP2/aDZ [a, 201]	0.2	-0.2	
RI-MP2/aTZ [a, 201]	0.1	-0.3	
SCS(MI) RI-MP2/aDZ [a, 201]	0.1	-0.3	
SCS(MI) RI-MP2/aTZ [a, 201]	0.0	-0.4	
MC MP2C[202]	-0.1		
PBE-TS[208]	-0.2	-0.4	-0.7
PBE-MBD[208]	0.0	0.4	2.6
B86bPBE	0.2		
B86bPBE-XDM	-0.3	-0.2	0.3

[a] The RI-MP2 calculations used the many-body expansion. aDZ and aTZ indicate the aug-cc-pVDZ and aug-cc-pVTZ basis sets, respectively.

B86bPBE-XDM and phonons calculated at the Γ-point resulted in a predicted free-energy difference of 0.3 kJ/mol per molecule at 298 K, with aspirin-I now more stable than aspirin-II, following the experimental preference. Results given in Table 3.4 show that the computed thermal correction is well converged with respect to phonon sampling in the a- and c-directions. However, convergence with respect to sampling in the b-direction is much slower and the value of the thermal correction is not converged even with a $1\times4\times1$ $\mathbf{q}$-point grid. It should also be noted that the b-direction is the one in which the unit cell is narrowest and is also the direction in which the C–H$\cdots$O pseudo-hydrogen bonds are oriented. In light of these results, there remains a fairly large uncertainty of as much as ±1 kJ/mol per molecule associated with the calculation of thermal correction. Hence, the error associated with the thermal corrections is too large to yield a reliable predicition of which phase is most thermodynamically stable, given that the two forms are found to be essentially degenerate from the electronic energy calculations presented in Table 3.3.

Given the high similarity of the packing arrangements and intermolecular interactions in the

Table 3.4: Convergence of the vibrational free-energy difference between aspirin-I and aspirin-II at 298 K, with respect to phonon sampling, in kJ/mol per molecule. All results use the full phonon density of states in the computation of the free energy, except where indicated by Γ, where only the phonons at the Γ-point are used.

Phonon sampling	$\Delta G_{\mathrm{I}\rightarrow\mathrm{II}}$, 298 K
1×1×1, Γ	0.3
1×1×1	0.2
1×2×1	0.7
1×3×1	-0.4
1×4×1	-0.2
1×2×1	0.7
2×2×1	0.8
1×2×2	0.6

two forms, it is of no surprise that most computational methods correctly predict them to be nearly degenerate. Indeed, in many of the cases shown in Table 3.3, this is likely fortuitous rather than a result of an accurate treatment of all intermolecular interactions in the lattice. In particular, the result from the B86bPBE base functional alone is already in good agreement with RI-MP2, despite the necessity of including the dispersion correction to obtain a reasonable lattice energy (the base-functional contribution to the lattice energy is only -26.1 kJ/mol for aspirin-I). High geometrical similarity will necessarily result in near energetic degeneracy, similar to what is seen for ABCABC versus ABAB stacking in graphite.[223,224] Such polytypes will be predicted to be nearly degenerate with virtually all theoretical treatments, regardless of their treatment of intermolecular interactions, and are therefore not recommended benchmarks for computational approaches to polymorph ranking.

3.3.2 Slip Mechanism

While apparently stable under application of high pressures,[199,204] aspirin-II is reported to be unstable with respect to shear stress,[200] and converts to aspirin-I after several months under ambient conditions.[200] This implies a relatively low barrier for interconversion via a shear-slip mechanism.

The results of the potential energy surface (PES) scan for the {100}<001> slip mechanism are presented in Figure 3.2. The barriers going from form-I to form-II, and vice versa, are predicted to be 10.7 and 10.0 kJ/mol per molecule, respectively. The difference between the two barriers is due to the slightly different geometries arising from the constraining procedure employed to calculate the PES.

The transition state along the PES scan is located at approximately $\delta_c = 3/10$, slightly skewed towards aspirin-II. This can be explained by the changing intermolecular interactions while sliding one layer along the <001> direction, as shown in Figures 3.2(d) and 3.3. First, consider aspirin-I as the starting point. In this form, the acetyl groups of aspirin interact via stabilizing dimeric C–H$\cdots$O interactions. As the top layer of aspirin-I slides along the <001> direction, these interactions are weakened (depicted by the O^1–$C^{1'}$ distance in Figure 3.2). This destabilization is eventually offset when the acetyl groups are brought closer to those of the neighbouring molecules (O^1–C^2 in Figure 3.2), forming the catemeric C–H$\cdots$O interactions seen in the aspirin-II crystal structure. The magnitude of the barrier is reasonable in light of the fact that two C–H$\cdots$O pseudo-hydrogen bonds are broken en route to the transition state.[225] For comparison, the binding energy of the formaldehyde dimer, which is bound by two analogous C–H$\cdots$O interactions (albeit with sp^2 rather than sp^3 carbons) is 14.3 kJ/mol.[149] Thus the predicted barrier of ca. 10 kJ/mol per molecule is well within the realm of the expected interaction strength.

In order to further validate the results obtained with the constrained PES scan, the Nudged Elastic Band method with the Climbing-Image option enabled[226,227] (CI-NEB), as implemented in Quantum ESPRESSO,[127] version 5.1, was used. The nudged-elastic band method aims to find the minimum-energy path (MEP) between two input structures in order to establish the likely path for a given phase transition mechanism. Essentially, intermediate structures (referred to as "images") are generated between the two input structures, initially set up as equidistant from one another. Then, the total force acting on an image is deconstructed into components parallel and perpendicular to the direction of the initial path. The parallel component of the force, known as the spring force, ensures that the images are kept appropriately spaced from one another, whereas the perpendicular component, known as the true force, is used to relax each image to a minimum-energy configuration along the MEP. The CI-NEB variant of the NEB method additionally seeks to converge the highest-energy image computed between the initial and final input structures to the saddle point of the MEP. An advantage to the CI-NEB method is that the points are optimized along the minimum-energy path (without reversion to the local minima), while the PES scan constrains optimization along the reaction coordinate. However, the NEB method implemented in Quantum ESPRESSO does not allow for variable-cell relaxation. The CI-NEB results shown in Figure 3.2 clearly reflect the same features of the reaction energy profile, and essentially predict the same barrier as the constrained PES scan. The resulting barriers are 10.6 kJ/mol (I$\longrightarrow$II) and 10.3 kJ/mol per molecule (I$\longleftarrow$II) using 11 points along the minimum-energy path. The difference between the two barriers is likely

Figure 3.2: $\{100\}<001>$ Slip mechanism in aspirin. δ_c is the slip increment along the $<001>$ direction and is the fraction of the lattice parameter c by which both dimer layers $\{100\}$ are offset. (a) PES scan for the interconversion between the aspirin-I ($\delta_c = 0$) and aspirin-II ($\delta_c = 1/2$) crystal structures. XDM-corrected energies are given relative to aspirin-I. (b) Variation of the unit-cell volume, relative to aspirin-I, during the phase transition. (c) Minimum-energy path located for the slip mechanism using the CI-NEB method. (d) Distances between key interacting functional groups during the phase transition from aspirin-II to aspirin-I. Atom labels are shown in Figure 3.3.

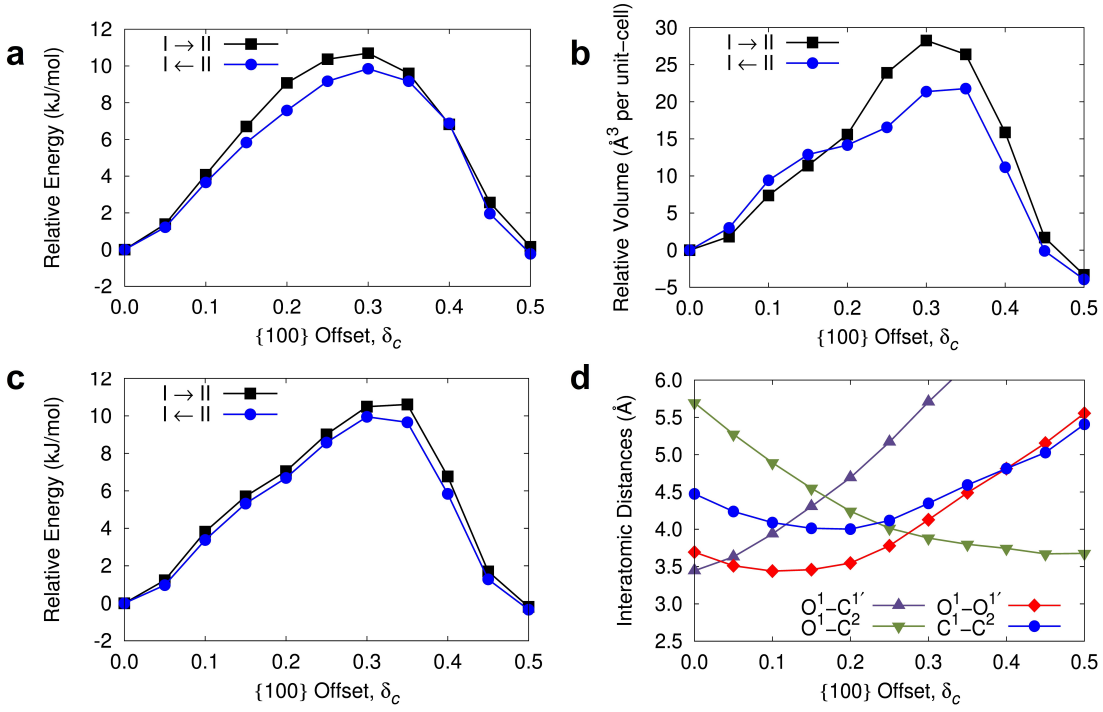
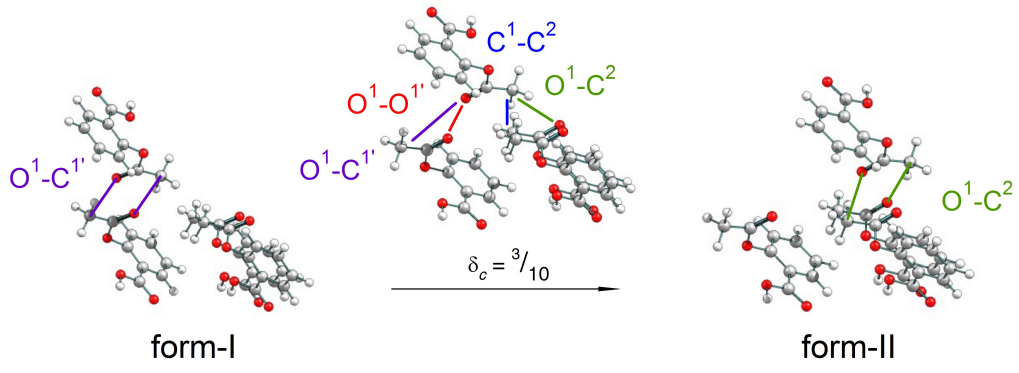


Figure 3.3: Variation in intermolecular interactions between acetyl groups during the phase transition of aspirin-I to aspirin-II via the $\{100\}<001>$ slip mechanism.



due to the fixed unit-cell geometries.

Lastly, it is considered whether the predicted barrier height for the $\{100\}<001>$ slip mechanism is consistent with the experimental observation that aspirin-II crystals, roughly 0.1 mm in size, transform to aspirin-I over a period of several months under ambient conditions.[199,200] Using a kinetic model for an upper bound to the conversion time for an ideal crystal via plastic deformation,[228] it can be shown that the time required for a complete conversion of an ideal crystal, with dimension $\tilde{c}$ along the crystallographic c-axis, is given by

$$t = \left(\frac{\tilde{c}}{c}\right)^2 \frac{h}{\varkappa k_{\mathrm{B}} T} \left|1 - \frac{\Delta G}{RT}\right|^{-1} e^{\Delta G/RT}. \tag{3.4}$$

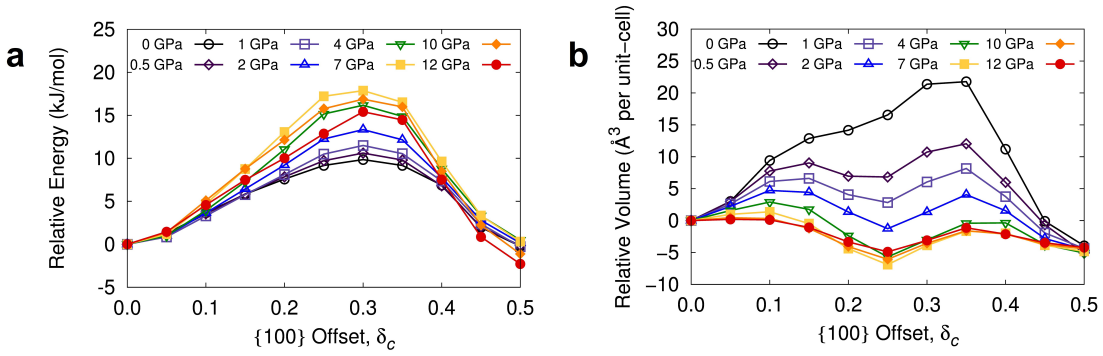
In this equation, ΔG is the free energy of activation for conversion of form-II supercells to form-I supercells. h, k_{B}, R, and T are Planck's constant, Boltzmann's constant, the gas constant, and the temperature, respectively. $\varkappa$ is the transmission coefficient, assumed to equal unity.

The calculated barrier for the slip mechanism in aspirin is ca. 80 kJ/mol for the supercell composed of 8 molecules. However, the calculated barrier is the electronic-energy barrier. Estimation of the free-energy correction at the approximate transition state ($\delta_c = 3/10$), by calculation of the phonon frequencies at Γ, reduces the barrier to ca. 76 kJ/mol. For a crystal measuring 0.1 mm along the c-axis, this would translate to a period on the order of 27 years to undergo complete conversion. However, while the crystal size of 0.1 mm was taken from nanoindentation experiments,[200] conversion was observed for a bulk powder. Considering crystal grains a quarter or a fifth of the 0.1 mm size would reduce the estimated time scale to a more realistic period of $\sim$1-2 years. Additionally, there is significant uncertainty in the predicted thermal correction (potentially as large as ± 8 kJ/mol for the supercell), due to the slow convergence of the phonon density of states with respect to sampling along the b-axis as discussed above. Given the exponential dependence on the free-energy barrier, the experimental conversion time falls well within the expected uncertainty of our calculation.

3.3.3 Applied Pressure

Finally, the high-pressure studies performed on the aspirin-I and aspirin-II crystals are examined.[199,204] These studies established that no phase transition from aspirin-II to aspirin-I occurs under high pressures (hydrostatic conditions, up to 10 GPa). These works were partly motivated by the assumption that high pressures would promote conversion of aspirin-II to aspirin-I, given

Figure 3.4: {100}<001> Slip mechanism in aspirin forms I and II under varying pressures. (a) PES scan for the I⟵II phase transition between the aspirin-I ($\delta_c = 0$) and aspirin-II ($\delta_c = 1/2$) crystal structures for various applied pressures. Energies are given relative to aspirin-I. (b) Variation of the unit-cell volume, relative to aspirin-I, during the phase transition.



that the former has a slightly larger unit-cell volume at room temperature.[200,221] As results in the previous section have shown (Figure 3.2), interconversion between the two forms is accompanied by an expansion of the unit-cell volume by ca. 25 Å^3 at the highest-energy point on the reaction path, which does not support this assumption. However, these results were obtained under normal pressure conditions.

Conducting the same PES scan under various applied pressures (ranging from 0.5 to 12 GPa) yields the results depicted in Figure 3.4. The shear-slip (electronic) energy barrier increases from ca. 10 kJ/mol per molecule, with no applied pressure, up to a maximum value of ca. 18 kJ/mol per molecule at 7 GPa (Figure 3.4). This is expected on the basis that the key interacting groups at the slip interface are brought closer together due to compression and is typical of sliding processes under applied load.[229,230]

Figure 3.4 shows that a local minimum in the relative unit-cell volume begins to develop near $\delta_c = 1/4$, as the applied pressure increases. This corresponds to the region where each aspirin dimer of one layer along the a-axis falls exactly midway between two aspirin dimers of the adjacent layers, allowing for further compression along the <100> direction. Additionally, as the applied pressure surpasses 7 GPa, the energy barriers begin to decrease, which is likely due to this "interlocking" of layers. Further increases in pressure will cause greater destabilization of the C–H$\cdots$O dimer and catemer interactions at $\delta_c = 0$ and $\delta_c = 1/2$ than the interlocked contacts in the region of the

transition state.

Despite the complex dependence of the shear-slip barrier on applied pressure, none of the barriers computed for pressures of up to 12 GPa fall below the barrier predicted under ambient conditions. This indicates that application of hydrostatic pressures will delay, rather than promote, the conversion of aspirin-II to aspirin-I crystals. Thus, it is of no surprise that the previous high-pressure experiments[199,204] did not reveal any indication of a phase transition between the two forms. To verify the validity of this claim would require keeping crystals of form-II aspirin under high pressures for periods longer than what was reported for the eventual conversion of the crystals under ambient conditions.[199,200]

3.4 Conclusions

In this work, dispersion-corrected density-functional calculations were performed to investigate the relative stabilities of two polytypes of aspirin and the shear-slip mechanism for their interconversion under equilibrium conditions and subject to applied pressure. The barrier calculated with B86bPBE-XDM for the conversion between aspirin-I and aspirin-II is approximately 10 kJ/mol per molecule. Given the uncertainty in the computation of the thermal correction and the limitations of the kinetics model, this value is consistent with the observed conversion rate between the two forms of several months to a year, which is accelerated when subject to shear stress.[200] Additionally, the computed barrier generally increases under compression and the lowest barrier is found at zero applied pressure. This result explains the observed lack of conversion between aspirin-II and aspirin-I under hydrostatic pressures up to 10 GPa within the limited timescales of the experiments.[199,204]

It is proposed here that nearly degenerate structures that are related by polytypism, such as the two forms of aspirin, are not reasonable tests of the accuracy of computational methods for polymorph ranking. This is especially true when precise experimental measurements for the relative stabilities of polytypes is non-existent. Given the very subtle structural differences between the polytypes being compared, almost all density functionals, even without dispersion corrections, predict a small energy difference between the two forms. Whether or not the nominal ordering is obtained is likely to be accidental.

CHAPTER 4

PERVASIVE DELOCALISATION ERROR CAUSES SPURIOUS PROTON TRANSFER IN ORGANIC ACID-BASE CO-CRYSTALS

L. M. LeBlanc, S. G. Dale, C. R. Taylor, A. D. Becke, G. M. Day*, E. R. Johnson*. Pervasive Delocalisation Error Causes Spurious Proton Transfer in Organic Acid-Base Co-Crystals. *Angew. Chem., Int. Ed.* **2018**, *57*, 14906-14910. DOI: `https://onlinelibrary.wiley.com/doi/full/10.1002/anie.201809381`. **Order Detail ID:** 71838285. **Angewandte Chemie by Gesellschaft Deutscher Chemiker Reproduced with permission of JOHN/WILEY & SONS, INC. in the format Educational/Instructional Program via Copyright Clearance Center.**

Contributions to the manuscript LML performed all the calculations, except for the CASTEP/VASP calculations reported in Table 4.2, which were performed by CRT. SGD, GMD and ERJ conceived the original idea. ERJ wrote the first draft of the manuscript. LML and SGD edited the manuscript with input from all authors. GMD, ADB and ERJ supervised the project.

4.1 Motivation

Dispersion interactions often dominate the lattice energy of molecular crystals; interactions that typical density-functional approximations do not include in their formulation. Consequently, the successful application of DFT to molecular crystals has relied on the development and parametrization of dispersion-correction schemes that are added to the DFA-calculated energy.[23,24,101,102,138–141]

However, a source of error that is not commonly acknowledged in this application area of DFT is delocalisation error.

It has long been established that delocalisation error in local density functionals results in overstabilisation of charge-transfer complexes and other species with separated charges.[108,231] This error is not seen in correlated wavefunction theories and can be reduced for DFAs through mixing of large amounts of exact (or Hartree-Fock, HF) exchange. Typically ca. 50% exact-exchange mixing is required to obtain accurate energetics for charge-transfer complexes,[232–234] charge-transfer excitation energies,[235–237] halogen-bonded complexes,[238] barrier heights of radical reactions,[239,240] and other cases where delocalisation error plagues local functionals.

Delocalisation error can be classified as either an energy-driven error or a density-driven error.[241] An example of the former is stretched H_2^+, where local DFAs correctly predict fractional charges of 0.5 e^- on each H atom, but give an energy that is far too low.[108,242] An example of the latter is the dissociation limit of NaCl,[243] where local DFAs incorrectly predict large, fractional charges (of ca. 0.4 e^-) on the separated atoms. Further examples of density-driven delocalisation error are charge-transfer complexes,[244] halide anions in water,[114,245] and solvated-electron model systems.[246] In density-driven cases, performing DFA calculations using a HF electron density can substantially improve the predicted thermochemistry.[240,241,247,248]

Several examples of density-driven delocalisation error have been noted, in which an improper density leads to significant errors in optimized geometries. This has been observed for the pre-reaction complex for H-atom abstraction from 1,4-diazabicyclo[2.2.2]octane by the benzyloxyl radical, where delocalisation error results in excessive stretching of one of the benzyloxyl C–H bonds.[249] Even more dramatic is the example of the carbanion intermediate for the Michael addition reaction of a thiolate to an olefin. In this case, a geometry optimization can yield either the thiolate anion or carbanion, depending on the degree of exact-exchange mixing.[250] This is an excellent challenge for testing new functionals designed to reduce delocalisation error.[247]

Unlike for molecular systems, correlated wavefunction calculations on solid-state systems are uncommon and are restricted to the smallest unit cells.[251,252] Therefore, accurate reference data for comparison is scarce. Indeed, performing efficient exact-exchange calculations using plane-wave basis sets is an ongoing challenge.[253] Consequently, dramatic examples of delocalisation error have not yet been illustrated for solid-state applications, with the exception of the consistent

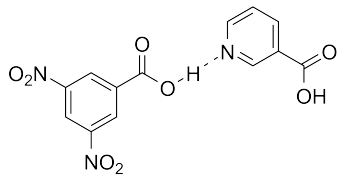
underestimation of band gaps.[231,254,255] While such energy-driven delocalisation errors exist, even the LDA and GGA functionals typically give reliable geometries for network solids. Moreover, dispersion-corrected GGAs perform extremely well in prediction of geometries of molecular solids, although organic salts have proved challenging[24,42,49] and exhibit fractionally charged ions.[101]

One area where delocalisation error may be prevalent in solids is for multicomponent crystals, where significant charge transfer between the species may exist in nature, or falsely be predicted by DFAs. Multicomponent crystals consisting of an organic acid-base pair serve as a good example. Such materials can exist as a neutral co-crystal, or proton transfer from the acid to the base can result in the formation of an organic salt. This has importance in, for example, the pharmaceutical industry, where the solubility and bioavailablity of active pharmaceutical ingredients depends strongly on their crystalline form.[256–260] The extent of the proton transfer has been shown, in many cases,[261] to be predictable by the difference in pKa's (ΔpKa) of the acid and base components.[262] Generally, if this value is less than zero (or greater than ca. 3), a co-crystal (or salt) will likely form from the acid-base pair.[262,263] However, reliable predictions cannot be made for intermediate ΔpKa's,[261–263] where the material can also possess a mixed ionisation state.[261,262,264,265] It is thus beneficial to turn to computational methods as a means to determine whether a co-crystal or salt will form from a given pair (or group) of compounds.[266,267]

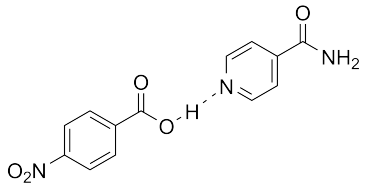
A recent study reported the energetic driving forces for co-crystal formation[265] using a dispersion-corrected GGA functional, specifically PBE-D3.[109,166] It was found that geometry optimization of 6 of the 350 co-crystals considered resulted in spurious proton transfer to give the corresponding organic salt.[265] Similar conclusions were drawn by others for nicotinamide-based crystals and their analogues.[264] B86bPBE-XDM[139] plane-wave calculations performed herein, which yielded identical results with respect to the spurious proton transfers, indicating that the choice of dispersion model is not the culprit for this erroneous stabilisation of salts over co-crystals. As a salt possesses delocalised charges, these results appear consistent with the delocalisation error inherent in GGAs, and in other local or semi-local DFAs.

In this chapter, the effect of exact-exchange mixing on the predicted geometries of six co-crystals previously identified to form salts preferentially during the course of DFA geometry optimizations is investigated.[264,265] For simplicity, crystal structures that were found to possess mixed ionisation states, or had proton positions mis-assigned in their experimental X-ray structures, have been excludedl; their molecular diagrams and unit-cells are depicted in Figures 4.1 and 4.2, respectively.

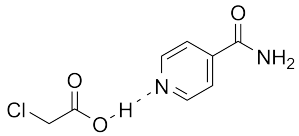
Figure 4.1: Molecular diagrams for the organic acid-base co-crystals considered, along with their CSD codes. pKa values for the acid (AH), the protonated base (BH+), and their corresponding difference, ΔpKa, are given. pKa values were taken from Ref. 268 unless otherwise indicated; [a] See Ref. 269, [b] or Ref. 270, [c] the pKa listed has been averaged from the reported values for 4-nitrophenol (7.14) and 3,5-dimethyl-4-nitrophenol (8.25) reported in Ref. 268.



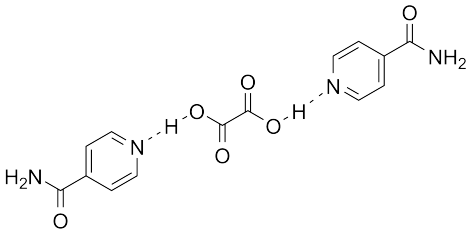
3,5-dinitrobenzoic acid nicotinic acid (**AWUDEB**)
pKa (AH) = 2.85, pKa (BH+) = 4.75, ΔpKa = 1.90



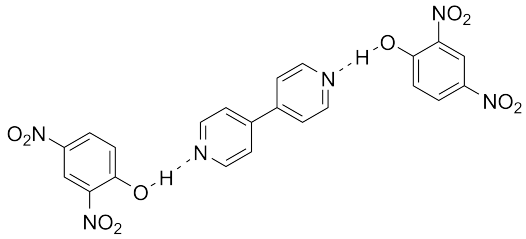
isonicotinamide 4-nitrobenzoic acid (**AJAKEB**)
pKa (AH) = 3.44, pKa (BH+) = 3.67,[b] ΔpKa = 0.23



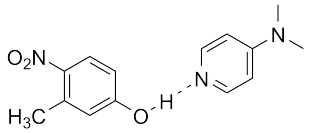
isonicotinamide monochloroacetic acid (**LUNNAJ**)
pKa (AH) = 2.87, pKa (BH+) = 3.67,[a] ΔpKa = 0.80



trans-oxalic acid bis(isonicotinamide) (**ULAWAF02**)
pKa (AH) = 1.27/4.27, pKa (BH+) = 3.67,[b] ΔpKa = 1.40/-0.60

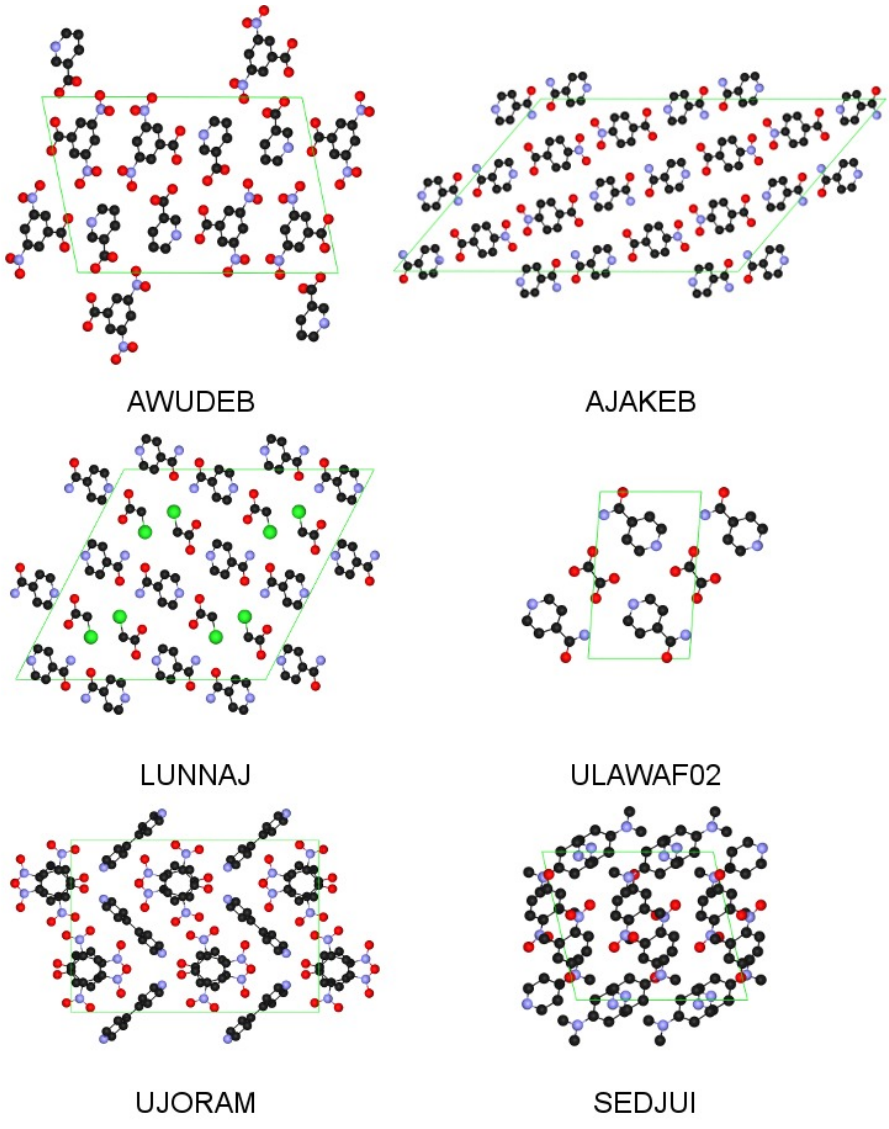


4,4'-bipyridine bis(2,4-dinitrophenol) (**UJORAM**)
pKa (AH) = 4.08, pKa (BH+) = 4.80/8.05,[b] ΔpKa = 0.72/3.97



3-methyl-4-nitrophenol 4-(dimethylamino)pyridine (**SEDJUI**)
pKa (AH) = 7.70,[c] pKa (BH+) = 6.09, ΔpKa = -1.61

Figure 4.2: The unit cells of the organic acid-base co-crystals, together with their CSD codes. Hydrogen atoms are omitted for clarity. Colours: black = carbon, blue = nitrogen, red = oxygen, green = chlorine.



4.2 Computational Methods

Initial geometries of all crystals were obtained from the Cambridge Structural Database (CSD).[22] All calculations were performed using the CRYSTAL17[271] software package, which employs atom-centred Gaussian basis sets and therefore permits efficient computation of HF exchange, even when subject to periodic-boundary conditions. Full geometry optimizations of both the atomic positions and unit-cell parameters were performed for all crystals. Convergence thresholds for the root mean square of atomic displacements and gradients between consecutive optimization steps were set to 0.0012 a.u. and 0.0003 a.u., respectively. The thresholds for the absolute value of the largest components for both quantities were set to 1.5 times these parameters.

The PBE[109] and PBE0[112] density functionals were used with double-ζ quality basis sets (i.e., POB-DZVP) and Grimme's D3 dispersion correction.[166] The HF-3c,[96,192] PBEh-3c,[193] and HSE-3c[272] methods, which use minimal basis sets together with geometrical-counterpoise and dispersion-correction terms, were also employed. In addition, both PBE0 and PBEh-3c include fractions of exact exchange, i.e., 25% and 42%, respectively. As for the screened exchange hybrid HSE-3c, it includes 42% of exact exchange in the short-range, and a standard error function with parameter ω set to 0.11 which gradually attenuates the exact exchange to yield a pure GGA in the long-range.

For all functionals, Becke-Johnson damping[161,167] was applied to the D3 dispersion correction. The default damping parameters from the CRYSTAL17 program were retained for all functionals except HF-3c, where the fit parameter was set to s_8 = 0.6143, and a non-additive Axilrod-Teller-Muto (ATM) dispersion term was included, as this is the recommended variant of this method, i.e., "sHF-3c", for optimal prediction of molecular-crystal geometries and lattice energies.[273]

Additional B86bPBE-XDM[139] calculations within the PAW formalism[213,214] were also performed, as implemented in Quantum ESPRESSO,[127] version 5.1, to ensure that the spurious proton transfer observed previously with PBE-D2/D3 optimizations (CASTEP, VASP) of co-crystals to form salts,[265] was not due to the type of dispersion correction or functional used. In these calculations, variable-cell relaxations were performed with convergence thresholds of 10^{-8} Ry, 10^{-4} Ry, and 10^{-3} Ry/bohr for electronic steps, total energies, and forces, respectively. Plane-wave kinetic-energy cutoffs for the wavefunction and density expansions were set to 60 and 600 Ry, respectively. A $2 \times 2 \times 2$ MP **k**-point grid was used. XDM a_1 and a_2 damping parameters were set to 0.6512, and 1.4633 Å, respectively. For the CASTEP/VASP PBE-D2/D3 calculations, fixed-cell, and subsequent variable-cell, optimizations using PBE-D2 (CASTEP) were performed with

tolerances of 10^{-7} eV and 0.03 eV/Å for electronic steps and forces, respectively. The plane-wave kinetic-energy cutoff was set to 500 eV. A $\mathbf{k}$-point sampling of 0.05 $Å^{-3}$ was used for reciprocal space sampling. Becke-Johnson damping was used with the dispersion correction. The resulting structures were then subjected to additional fixed-, followed by variable-cell, optimizations within VASP, using PBE-D3 and parameters equal to those just described. Additional details can be in found in the previous report.[265]

4.3 Results and Discussion

The results of the DFA geometry optimizations are summarised in Table 4.1. As with the previous plane-wave calculations using the same dispersion-corrected functional,[265] PBE-D3 optimization causes each of the co-crystals to undergo spontaneous proton transfer, from the acid to the base, forming the corresponding salts.

Similar results are also obtained with PBE0-D3 (which contains 25% exact exchange). Conversely, sHF-3c, PBEh-3c (42% exact exchange), and HSE-3c (42% exact exchange in the short range, gradually attenuated to 0% in the long range) optimizations consistently preserve the neutral co-crystal when starting from the experimental geometry. Thus, relatively high fractions of exact exchange (in excess of 25% and possibly around 40%) are needed to predict a stable co-crystal. Results for B86bPBE-XDM (Quantum ESPRESSO) and PBE-D2/D3 (CASTEP, followed by VASP) optimizations starting from the experimental co-crystal structures are also shown in Table 4.2 and are consistent with previous work[265] and the work presented herein, where the salts are preferentially formed.

To verify that the co-crystal is indeed the "correct" structure for these compounds, the C–O bond lengths are compared with the experimental X-ray crystallographic data in Table 4.3. For the four carboxylic-acid-containing crystals (AJAKEB, AWUDEB, LUNNAJ, and ULAWAF02), both PBE and PBE0 predict effectively equal C–O distances, implying the formation of a carboxylate group and a salt. Conversely, PBEh-3c, HSE-3c, and sHF-3c predict different C–O bond lengths, as are seen experimentally, implying formation of a carboxylic acid and a co-crystal. For the two alcohols, SEDJUI and UJORAM, assignment of the protonation state is slightly more difficult. However, the PBE and PBE0 C–O bond lengths are considerably shorter than experiment, implying formation of a phenoxide anion and salt, while the PBEh-3c, HSE-3c, and sHF-3c distances for the phenols are in better agreement with experiment, implying a stable alcohol and co-crystal.

Table 4.1: Structures predicted by selected plane-wave dispersion-corrected DFAs (c = co-crystal, s = salt). The initial input geometry was either the experimental co-crystal X-ray structure (Expt.) or the PBE-D3 optimized structure.

Basis Set	Plane-wave		Gaussian							
Method	B86bPBE-XDM	PBE-D3	PBE-D3	PBE0-D3	PBEh-3c		HSE-3c		sHF-3c	
Input Structure	Expt.	Expt.	Expt.	Expt.	Expt.	PBE-D3	Expt.	PBE-D3	Expt.	PBE-D3
AWUDEB	s	s	s	s	c	s	c	s	c	c
LUNNAJ	s	s	s	s	c	s	c	s	c	c
UJORAM	s	s	s	s	c	s	c	s	c	c
AJAKEB	s	s	s	s	c	c	c	c	c	c
ULAWAF02	s	s	s	s	c	c	c	c	c	c
SEDJUI	s	s	s	s	c	c	c	s	c	c

Table 4.2: Structures predicted by geometry optimization of the co-crystal structures with plane-wave calculations. Experimental and optimized bond lengths are in Å. The mean absolute error (MAE) with respect to the experimental reference data, is given for each method.

Method	Expt.	PBE-D2/D3		B86bPBE-XDM	
	C–O	C–O	N–H, O–H	C–O	N–H, O–H
AWUDEB	1.219, 1.256	1.147, 1.263	1.106, 1.485	1.261, 1.274	1.110, 1.464
LUNNAJ	1.220, 1.293			1.254, 1.286	1.117, 1.453
UJORAM	1.313	1.302	1.111, 1.462	1.283	1.110, 1.452
AJAKEB	1.222, 1.307			1.263, 1.286	1.162, 1.361
ULAWAF02	1.223, 1.288			1.252, 1.285	1.147, 1.393
SEDJUI	1.328	1.283	1.120, 1.430	1.300	1.106, 1.472
MAE	–	0.033	–	0.025	–

Additional calculations were performed with sHF-3c, PBEh-3c, and HSE-3c starting from the salt geometries obtained from PBE-D3. sHF-3c consistently transfers the proton back to its original placement, returning to the neutral co-crystal. With PBEh-3c, this was also the case for AJAKEB, ULAWAF02, and SEDJUI. However, for the other three crystals, the salt could be obtained as a stable, minimum-energy structure. With HSE-3c, only AJAKEB and ULAWAF02 returned to their co-crystal forms. The relative PBEh-3c and HSE-3c energies of the salt and co-crystal forms are given in Table 4.4. Similar to previous conclusions from the literature,[261–263] there is only a weak correlation between these relative energies and the ΔpKa, and no correlation with the individual pKa's.

While it could be argued that the preference for co-crystals over salts found with sHF-3c, PBEh-3c, and/or HSE-3c is due to the use of minimal (or small) basis sets with these methods, it is illustrated here that this is not the case. Two experimental salt structures[264] (Figure 4.3) were optimized with these small basis-set methods. Both crystals depicted in Figure 4.3 remained salts upon optimization with PBEh-3c and HSE-3c. Only EMINUJ gave a co-crystal upon optimization with sHF-3c, which is evident from the two significantly different C–O bond lengths as well as from the N–H and O–H distances in Table 4.5. Thus, the use of minimal basis sets does not unequivocally bias co-crystal over salt structures.

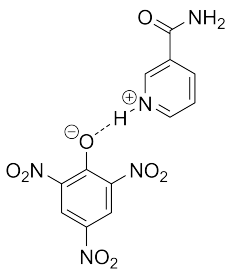
Table 4.3: Experimental and optimized C–O bond lengths (in Å) for the co-crystals studied. The initial geometry for the optimizations was either the experimental co-crystal structure, or the PBE-D3 optimized structure. The MAE with respect to the experimental reference data, is given for each method.

Method	Expt.	PBE-D3	PBE0-D3	PBEh-3c		HSE-3c		sHF-3c	
Input Structure		Expt.	Expt.	Expt.	PBE-D3	Expt.	PBE-D3	Expt.	PBE-D3
AWUDEB	1.219, 1.256	1.276, 1.282	1.262, 1.267	1.218, 1.287	1.239, 1.252	1.219, 1.286	1.240, 1.252	1.213, 1.326	1.213, 1.326
LUNNAJ	1.220, 1.293	1.273, 1.290	1.259, 1.275	1.212, 1.301	1.232, 1.267	1.212, 1.301	1.233, 1.267	1.210, 1.339	1.210, 1.339
UJORAM	1.313	1.290	1.278	1.306	1.267	1.304	1.265	1.318	1.316
AJAKEB	1.222, 1.307	1.281, 1.291	1.266, 1.275	1.218, 1.299	1.218, 1.299	1.219, 1.299	1.219, 1.299	1.217, 1.336	1.217, 1.336
ULAWAF02	1.223, 1.288	1.272, 1.289	1.257, 1.273	1.208, 1.294	1.208, 1.294	1.209, 1.293	1.209, 1.293	1.208, 1.332	1.209, 1.332
SEDJUI	1.328	1.305	1.291	1.317	1.318	1.316	1.282	1.337	1.337
MAE	–	0.031	0.031	0.010	0.015	0.009	0.019	0.024	0.024

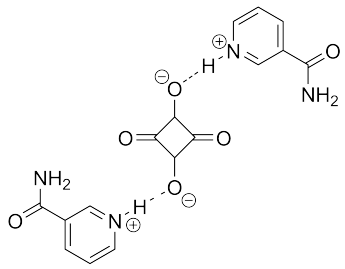
Table 4.4: Relative PBEh-3c and HSE-3c energies (in kcal/mol per proton transferred) for the co-crystal and salt forms (co-crystal energy - salt energy), together with literature pKa values (see Figure 4.1 for references).

Quantity	pKa (AH)	pKa (BH+)	ΔpKa	ΔE (PBEh-3c)	ΔE (HSE-3c)
AWUDEB	2.85	4.75	1.90	1.10	1.15
LUNNAJ	2.87	3.67	0.80	-1.42	-1.31
UJORAM	4.08	4.80	0.72	-1.69	-1.10
AJAKEB	3.44	3.67	0.23	–	–
ULAWAF02	4.27	3.67	-0.60	–	–
SEDJUI	7.70	6.09	-1.61	–	-1.69

Figure 4.3: Molecular diagrams for the organic acid-base salts considered, along with their CSD codes, shown in parentheses, bolded. pKa values for the acid (AH), the protonated base (BH+), and their corresponding difference, ΔpKa, are given. pKa values were taken from Ref. 268 unless otherwise indicated; [a] See Refs. 274,275.



nicotinamidium picrate (**IPOZAO**)
pKa (AH) = 0.42, pKa (BH+) = 3.33, ΔpKa = 2.91



dinicotinamidium squarate (**EMINUJ**)
pKa (AH) = 0.60/3.50,[a] pKa (BH+) = 3.33, ΔpKa = 2.73/-0.17

Table 4.5: Structures predicted from geometry optimization of the salt structures with selected DFAs. Experimental and optimized bond lengths are in Å. The initial geometry for the optimizations was the experimental salt crystal structure.

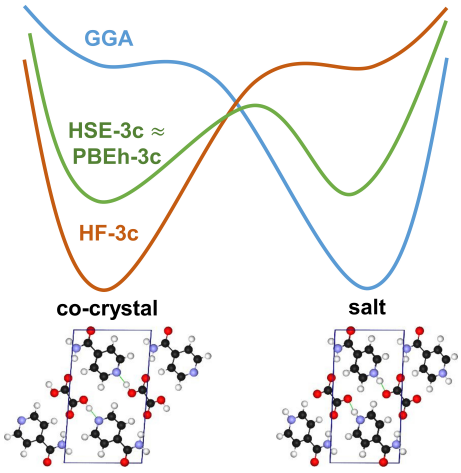
Method	Expt.	PBE-D3	PBE0-D3	PBEh-3c	HSE-3c	sHF-3c
C–O						
IPOZAO	1.263	1.279	1.273	1.249	1.249	1.240
EMINUJ	1.242, 1.257	1.270, 1.277	1.258, 1.263	1.239, 1.249	1.239, 1.250	1.218, 1.287
N–H, O–H						
IPOZAO	n/a	1.050, 1.817	1.036, 1.832	1.047, 1.630	1.046, 1.638	1.099, 1.456
EMINUJ	n/a	1.076, 1.564	1.058, 1.600	1.091, 1.463	1.092, 1.460	1.500, 1.054

4.4 Conclusions

The present results illustrate the potential of delocalisation error to cause qualitative disagreements between calculated and reference crystal structures, resulting in differing ionisation states and chemical-bonding arrangements. Depending on the extent of exact-exchange mixing, either the organic salt, neutral co-crystal, or both, can be obtained from DFA geometry optimization of the multicomponent crystals considered herein (Figure 4.4).

The pure GGA (PBE-D3) and the low-%HF hybrid (PBE0-D3) both favour proton transfer, resulting in salt structures. This error can be understood by the tendency of GGA exchange to over-stabilise delocalised charges. Using full exact exchange, as in sHF-3c, correctly preserves the neutral co-crystal in all cases, but tends to overestimate the bond lengths, which is characteristic of HF theory. Using an intermediate exact-exchange mixing of 42%, as in PBEh-3c, or a screened

Figure 4.4: Predicted ionisation states of organic acid-base solids when using various amounts of exact-exchange mixing in DFT and HF Methods



exchange hybrid, such as HSE-3c, leads to prediction of both salt and co-crystal structures as local minima in 50% or more of the cases studied. The similarity of the PBEh-3c and HSE-3c results implies that the truncated Fock-exchange distance regime employed in HSE-3c is not severe enough to encourage charge transfer. Overall, the results clearly demonstrate the sensitivity to exact-exchange mixing that is the signature of delocalisation error.

The consistent failing of the GGA functional to stabilise the six co-crystals considered in this work is a significant issue because the success and convenience of dispersion-corrected GGAs has led to their widespread use in studying molecular crystals. For example, DFT-D optimization has been suggested as a tool for validation of experimental crystal structures[276] and has been demonstrated to be among the most promising approaches to crystal structure prediction (CSP).[49,102,277] Multicomponent acid-base crystals are common and are of great importance within the pharmaceutical industry,[256,262] which is one of the main users of CSP. In this context, the formation of a multicomponent solid can modify the physical properties of an active pharmaceutical ingredient. The correct identification of protonation is important for understanding the resulting properties, as well as for regulatory and intellectual property reasons. The erroneous prediction of the ground protonation state means that GGA-based CSP approaches cannot be applied to multicomponent crystals capable of such acid-base equilibria with any expectation of reliability.

Finally, because the delocalisation error was found to affect the geometries of the multicomponent crystals, rather than simply energies or electron densities, density-corrected DFAs[241] are not a

viable solution. Moreover, full HF exchange or hybrid functionals are not practical for large-scale applications with plane-wave / pseudopotential calculations, which are the workhorse of solid-state electronic-structure theory. Thus, new approaches to reduce delocalisation error using local-density ingredients are needed. The crystalline acid-base equilibria considered herein should constitute an excellent benchmark for development of functionals with reduced delocalisation error for solid-state applications.

CHAPTER 5

COMPOSITE AND LOW-COST APPROACHES FOR MOLECULAR CRYSTAL STRUCTURE PREDICTION

Reprinted with permission from **L. M. LeBlanc***, A. Otero-de-la-Roza*, E. R. Johnson*. Composite and Low-Cost Approaches for Molecular Crystal Structure Prediction. *J. Chem. Theory Comput.* **2018**, *14*, 2265-2276. DOI: https://pubs.acs.org/doi/full/10.1021/acs.jctc.7b01179.

Contributions to the manuscript LML performed the calculations, except for the Quantum ESPRESSO calculations reported in Table 5.1, which were performed by AOR. LML wrote the first draft of the manuscript. AOR and ERJ contributed to the final version of the manuscript. ERJ supervised the project.

5.1 Motivation

The X23 benchmark set[145] is commonly used to validate computational methods on their ability to reproduce lattice energies of molecular solids. This set of small molecular organic solids extends the C21 set first proposed by Johnson and Otero-de-la-Roza[142] by adding two additional structures, hexamine and succinic acid (Figure 5.1). The X23 set provides static electronic lattice energies, ΔE_{el}, which are back-corrected from experimental sublimation enthalpies, ΔH_{sub}. For non-linear molecules, this gives

$$\Delta E_{\mathrm{el}} \approx \Delta H_{\mathrm{sub}} - \Delta F_{\mathrm{vib}} - 4RT, \tag{5.1}$$

where ΔF_{vib} is computed using Eq. 3.2 for both the solid and gas phases, and by assuming that the gas produced via the sublimation process is ideal (hence, the 4RT term). Plane-wave B86bPBE-XDM calculations are able to reproduce lattice energies of the X23 set to within 3.6 kJ/mol,[139] which is well at the limit of precision for this benchmark set, i.e., an error of ca. 4.9 kJ/mol for experimentally measured sublimation enthalpies.[147]

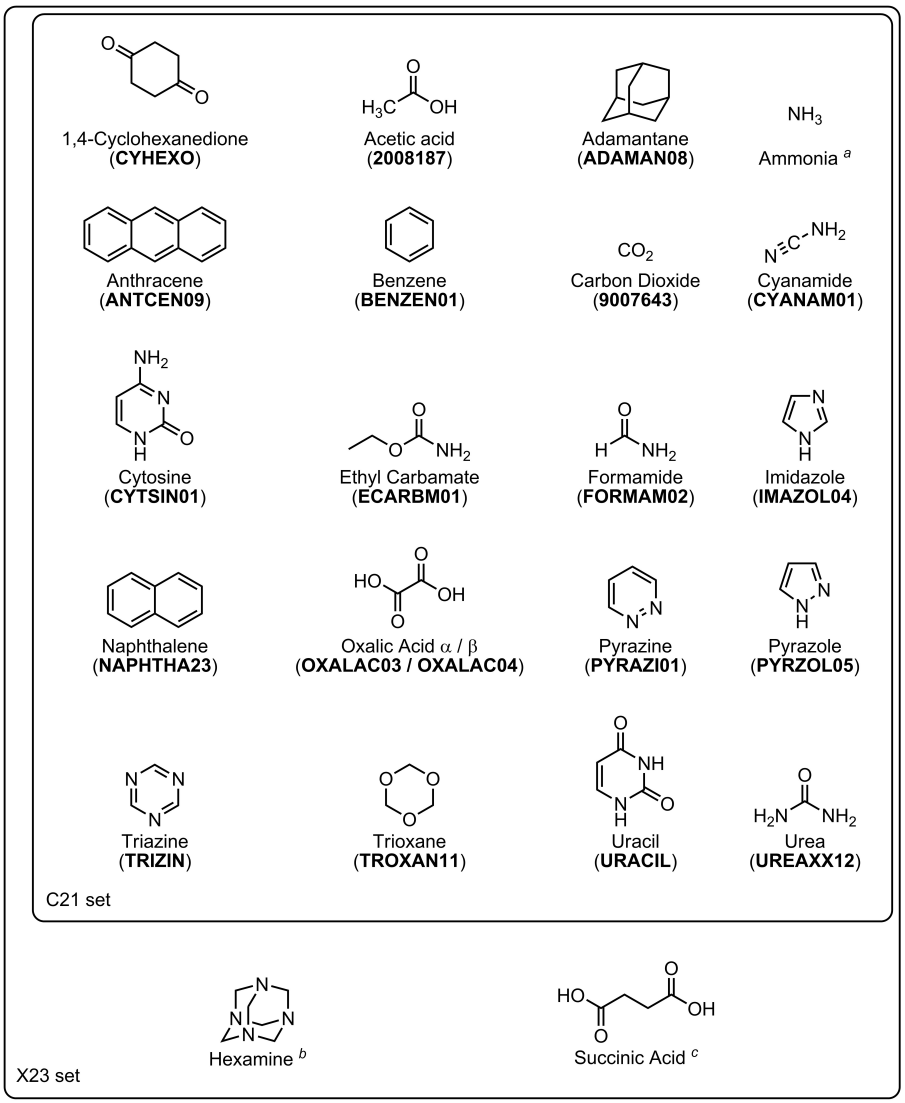
Several low-cost electronic-structure methods have been developed recently, and tested for molecular crystals. Density-functional tight-binding[179,180,182] paired with the D3 dispersion correction was shown to yield a MAE on the X23 lattice energies of 10.4 kJ/mol,[97] although it is important to note that cell parameters were kept fixed to those of benchmark geometries[142,145] during optimization. The PBEh-3c[193] and HF-3c[96,188,192] methods use small or minimal Gaussian basis sets coupled with *ad hoc* corrections, and have MAEs of 5.4 kJ/mol[193] and 8.4 kJ/mol,[273] respectively. In the case of HF-3c, the MAE can be reduced to 6.7 kJ/mol if some of the correction parameters are tuned such that the resulting geometries are closer to the experimental crystal structures.[273]

In a recent study,[282] Carter and Rohl evaluated the performance of the vdW-DF1[176] and vdW-DF2[177] non-local functionals with locally-supported numerical orbitals (SIESTA[130,131]) on the C21 set,[142] and compared their results to the same calculation using plane-waves / pseudopotentials.[127] The authors found that, by using vdW-DF2 in SIESTA with a double-ζ plus polarization (DZP) basis set and a counterpoise correction, the resulting MAE (3.8 kJ/mol) is similar to B86bPBE-XDM (3.6 kJ/mol)[139] and much smaller than either using vdW-DF2 in Quantum ESPRESSO (8.2 kJ/mol) or the same method without the counterpoise correction (22.6 kJ/mol). Other options, such as using a triple-ζ plus polarization (TZP) basis set at the DZP geometries (MAE = 10.2 kJ/mol) were also explored.[282,283]

While low-cost methods are typically assessed based on their ability to produce accurate lattice energies, this is a flawed measure of the accuracy and reliability of a CSP protocol; instead, relative lattice energies would be a more suitable measure. For instance, the enantiomeric excesses of a scalemic solution in contact with the racemate and enantiopure phases of various organic compounds can be used. This measure is accessible experimentally, and provides an excellent benchmark tool for energy differences between crystal structures, as shown in a recent study of a set of amino-acid crystals.[128]

It is also interesting to find which methods yield accurate crystal geometries in addition to

Figure 5.1: Molecular diagrams for the organic solids of the X23[142,145] benchmark set of lattice energies. Their CSD[22] and/or Crystallography Open Database (COD)[278] codes are shown in parentheses, bolded. See [a] Ref. 279, [b] Ref. 280, or [c] Ref. 281 for information on these crystal structures.



absolute or relative lattice energies.[186] A low-cost method able to generate crystal geometries comparable to B86bPBE-XDM would avoid the geometry optimizations with the latter, resulting in a large saving in computational cost. This composite approach, is herein called a "multilevel method" in the rest of this chapter. A multilevel approach (TPSS-D3 at the HF-3c geometries) has been previously tested for CSP by Brandenburg and Grimme[284] using the POLY59 benchmark set. This benchmark set consists of 9 experimental crystal structures from the sixth blind test, 5 of which are polymorphic, supplemented with 10 low-energy candidate structures generated for the blind test. While Brandenburg and Grimme did demonstrate an improvement in the polymorph energy ranking using their multilevel TPSS-D3//HF-3c approach (as opposed to HF-3c alone), the results were not accurate enough to predict the experimental crystal structure as being the lowest energy structure for each system. Note, however, that this result does not necessarily reflect the quality of their approach, given that more than just thermodynamics, i.e., kinetics,[53,285] can dictate what phases are observed experimentally. Even so, testing low-cost methods for their ability to reproduce equilibrium crystal geometries calculated at a higher level of theory is as important as the relative lattice energies themselves when considering a multilevel method.

In this chapter, the ability of various low-cost methods (local Gaussian and numerical basis sets, HF-3c, DFTB) to produce accurate absolute and relative lattice energies and geometries is assessed, with the idea of building a multilevel approach using one of these methods with a final B86bPBE-XDM single-point energy calculation. The X23 benchmark set[142,145] and the set of ten relative solubilities of chiral amino acids reported previously[128] supplemented with four additional chiral compounds (dubbed herein the EE14 set) are used as test sets for this analysis. As a practical CSP example, the resulting methods are applied to the crystal structure prediction of 1-aza[6]helicene, an organic semiconductor on which a B86bPBE-XDM-based CSP study was recently performed.[2]

5.2 Computational Methods

The high- and low-cost methods used in this work are now described.

SIESTA calculations: SIESTA (Spanish Initiative for Electronic Simulations with Thousands of Atoms) is a DFT method that uses finite-support numerical orbitals to achieve linear asymptotic scaling with system size.[131,286–288] The homonymous SIESTA program, version 4.0b, was used in this work. The PBE[109] semi-local functional and the vdW-DF1 and vdW-DF2 non-local functionals[176,177] are considered here. PBE was coupled with Grimme's D2 dispersion correction,[165]

with the functional- and basis-set-specific fitting parameters equal to those reported by Peverati and co-workers[289] (e.g. for PBE with a DZP-quality basis set, s_6 = 0.5 and s_r = 1.1). Troullier-Martins norm-conserving pseudopotentials[120,121,290] were generated using the ATOM code.[132]

DZP basis sets were used for the SIESTA calculations, except for occasional TZP basis-set calculations. DZP is the standard basis set implemented in SIESTA, while TZP was constructed and optimized by Louwerse and Rothenberg[283] for H, N, and O atoms, and further extended to the C atom by Carter and Rohl.[282] The effect of changing the confinement radius of the finite-support orbitals using SIESTA's "energy shift" parameter—the energy increase experienced by the orbital when it is confined—was also explored. By lowering the energy shift (0.02 Ry is SIESTA's default), the orbital confinement radii are extended, increasing the computational cost of the calculation, but reducing errors due to basis-set superposition error (BSSE).[291] In previous studies, an energy shift of 0.001 Ry proved to give converged results with respect to relative lattice or binding energies for various systems.[282,291] Unless otherwise indicated, this value was used in this work. The fineness of the real-space integration grid for charge densities and potentials was set by having plane waves of kinetic energy of 200 Ry or less be represented without aliasing.

CRYSTAL calculations: CRYSTAL[271] is a code for electronic-structure calculations in periodic solids using local Gaussian orbitals. In this work, CRYSTAL17[271] was used to run Hartree-Fock (HF) and DFT calculations. Specifically, the HF-3c[96,192] and PBEh-3c[193] methods were utilized. Both of these methods make use small or minimal basis sets (MINI(x)[96,192] for HF-3c and def2-mSVP[193] for PBEh-3c) and are supplemented with Grimme's D3 dispersion[166,167] and a geometrical counterpoise correction (gCP).[190,191] HF-3c has an additional correction for short-range basis-set (SRB) incompleteness errors.[96,192,273] The HF-3c calculations used the Becke-Johnson damping function in D3,[161,167] three-body interactions, and a fit parameter of s_8 = 0.6143, as described elsewhere.[273] This dispersion-scaled version of HF-3c will be referred to as "sHF-3c" throughout the rest of this chapter. For comparison, results for HF combined with the MINI(s) basis set plus Grimme's D2 dispersion correction[165] and no further corrections are presented to evaluate the importance of additional BSSE corrections. In this case, the s_6 fit parameter for the dispersion correction was set to 1. Similarly, PBE combined with the POB-DZVP basis set and Grimme's D2 or D3 dispersion corrections were also utilized for means of comparison. In the case of the D2 dispersion correction, the s_6 fit parameter was set to 0.75.

DFTB+ calculations: DFTB+,[292] version 1.3, and its implementation of the third-order self-consistent-charge density-functional tight-binding method (SCC-DFTB3-D3(BJ)) was used as another low-cost method. The DFTB semiempirical method is based on the n^{th}-order expansion of local density fluctuations with respect to a reference superposition of atomic electron densities.[179,182] DFTB uses precomputed two-centre electron integrals; the '3ob' parametrization set was used in this work.[183,185] Damping of the hydrogen pair potentials was set to ζ = 4.2 in order to obtain an improved description of hydrogen-bonding.[97,183] The method was also supplemented with Grimme's D3 dispersion correction,[166] with Becke-Johnson damping,[161,167] and fit parameters set to s_8 = 0.5883, a_1 = 0.5719, and a_2 = 3.6017, as described elsewhere.[191]

Plane-wave calculations: Quantum ESPRESSO,[127] version 5.1, a plane-waves / pseudopotentials code, was used for the high-level reference calculations in this work. Periodic-boundary DFT calculations employed the B86bPBE exchange-correlation functional[109,110] supplemented with the XDM dispersion correction.[148,155] The damping function parameters in XDM, a_1 and a_2, were set to 0.6512, and 1.4633Å, respectively. The PAW method was used.[124] Wavefunction and density cutoffs were set to 80 Ry and 800 Ry, respectively. Structure relaxations were performed with tighter thresholds for convergence of the energies and forces, i.e., 10^{-5} Ry and 10^{-4} Ry/bohr, respectively.

k-point grid sampling and structure relaxation: Unless otherwise stated, a $4 \times 4 \times 4$ MP **k**-point mesh sampling of the Brillouin zone was used throughout for crystal structures, and sampling at the Γ-point was used for molecules. Atomic coordinates and cell parameters for all crystals were fully relaxed to obtain the minimum-energy structures. For isolated molecules, a sufficiently large simulation cell was used to avoid image interactions, and only the atomic coordinates were relaxed (the exception being with CRYSTAL17, where the "MOLECULE" keyword was used to treat molecules as non-periodic systems).

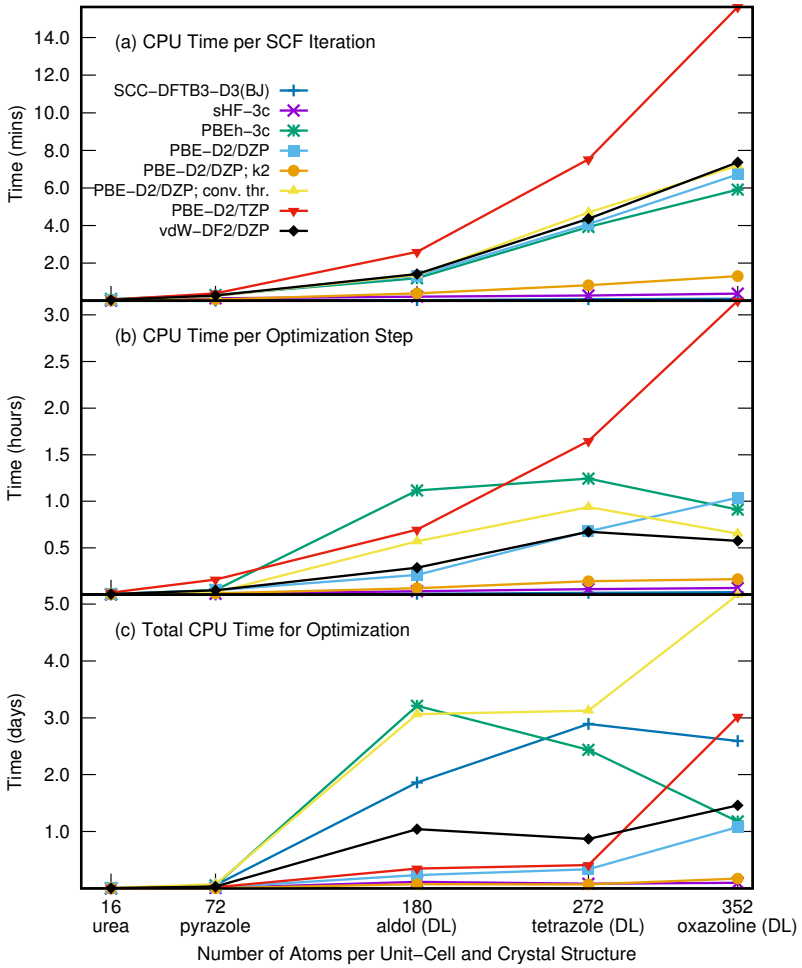
Similarity index: The POWDIFF utility in CRITIC2[293] was used to measure the degree of similarity between low-cost and B86bPBE-XDM crystal structures. POWDIFF is based on the comparison of powder diffraction patterns using a cross-correlation function.[294] A POWDIFF result of zero is an exact match, and one represents maximum dissimilarity.

The computational cost of each method above depends on the type of crystal and, in the case of geometry optimizations, also on the starting geometry. According to some tests (Figure 5.2),

a rough guide is that SIESTA methods employing a DZP basis set and PBEh-3c are about 5 times faster than a plane-waves / pseudopotentials (Quantum ESPRESSO) calculation. In contrast, sHF-3c and DFTB are roughly around 50 and 500 times faster than Quantum ESPRESSO, respectively. However, crystal geometry optimizations using SCC-DFTB3-D3(BJ) as implemented in DFTB+ require approximately two orders of magnitude more steps than the other methods, which is detrimental to its performance. Qualifying a method as "low-cost" in this chapter solely reflects how quickly a crystal structure was optimized, keeping all other parameters of the calculation constant (system under study, system size, symmetry, input geometry, memory and/or disk usage, number of CPUs requested, etc.), and does not necessarily relate to the ideal scaling of the algorithms or parallelism efficiency within each of the codes in which the methods are implemented.

In the rest of the chapter, multilevel methods built from a geometry optimization using a low-cost method (A), followed by single-point energy calculation (B; almost always B = B86bPBE-XDM) have been considered. The usual notation for composite methods in chemistry (B//A) is also used.

Figure 5.2: CPU real time (a) per SCF iteration, (b) per optimization step, and (c) total CPU time for selected crystal structures, of the various low-cost methods tested. Calculations were run in parallel each using 32 Intel "Broadwell" CPUs at 2.1 GHz, model E5-2683 v4. These timings serve only as a rough estimate of the time required to perform geometry optimizations on selected systems. The resources required (memory and/or disk) did not differ significantly between codes/methods to merit discussion. "k2": a $2\times2\times2$ k-point mesh was used instead of the $4\times4\times4$ mesh used in all other computations. "conv. thr.": Tighter force and stress convergence thresholds are used (0.01 eV/Å and 0.02 GPa for maximum component values) instead of the default thresholds within the SIESTA code (see Computational Methods section).



5.3 Results and Discussion

5.3.1 Absolute Lattice Energies

First, the performance of all methods for the crystals of the X23 benchmark set is examined. The relevant statistics are summarised in Table 5.1. The best-performing expensive method (Quantum ESPRESSO block) is B86bPBE-XDM, with a MAE of 3.6 kJ/mol. Similar in performance, although with slightly higher MAEs, are other XDM-corrected methods: PW86PBE-XDM (3.7 kJ/mol), PBE-XDM (4.7 kJ/mol), and BLYP-XDM (5.5 kJ/mol). The popular PBE-D2 functional and the non-local vdW-DF1 and vdW-DF2 functionals have MAEs around 6 kJ/mol, with a computational cost similar to B86bPBE-XDM. Finally, alternative expensive methods such as the rVV10[178] (as implemented in Quantum ESPRESSO) are clearly not suitable for lattice energies in molecular crystals.

In agreement with previous studies, the use of low-cost methods to determine lattice energies of molecular crystals showed a less than desirable accuracy. The MAE ranges from 5.7 kJ/mol for the relatively expensive PBEh-3c to up to 22.3 kJ/mol for vdW-DF1/DZP. The best-performing low-cost methods are PBEh-3c and sHF-3c, although PBEh-3c failed to converge its SCF in some cases (anthracene and naphthalene) due to their small band gaps.[193] These values are still relatively far from the 3.6 kJ/mol obtained using B86bPBE-XDM, although the most accurate of them, sHF-3c and PBEh-3c, are reasonably accurate, and similar in performance to some expensive methods available in Quantum ESPRESSO.[24,284] All low-cost methods except for sHF-3c and PBEh-3c also show a very strong tendency to overestimate lattice energies. This is reasonable if the source of the error is basis set superposition, but it is also the case for SCC-DFTB3-D3(BJ), for which this error is nil.

The MAEs obtained using B86bPBE-XDM single-point energy calculations at the geometry resulting from any of the low-cost methods tested are, in general, smaller than using the energies from those same methods, except in the case of SCC-DFTB3-D3(BJ). The MAEs correlate very roughly with the methods' ability to produce crystal structures similar to the B86bPBE-XDM equilibrium geometries, as shown in the POWDIFF column of Table 5.1. Importantly, the lattice energies improve substantially for all methods where the primary source of error is BSSE, indicating that BSSE affects very strongly the calculation of absolute lattice energies, but only has a minor effect on crystal geometries. This seems to be particularly true for the finite-support numerical orbital calculations in SIESTA, which grossly overestimate the absolute lattice energies but yield surprisingly good crystal geometries. Table 5.1 shows that the B86bPBE-XDM//PBE-D2/DZP

Table 5.1: Statistics for the X23 set of lattice energies using various computational methods.[d] Mean error (ME), mean absolute error (MAE), and maximum absolute error (MAX) relative to electronic energies back-corrected from experimentally-measured sublimation enthalpies (Expt.)[142,145] or to fully-relaxed B86bPBE-XDM results (DFT). "A" and "B//A" refer, respectively, to the results obtained from the low-cost optimization directly or to a B86bPBE-XDM single-point energy calculation at the low-cost equilibrium geometry. All values are in kJ/mol per molecule, except POWDIFF similarity measures, which are dimensionless.

	Expt.				DFT				
	A			B//A	A		B//A		
Method (A)	ME	MAE	MAX [c]	MAE	MAE	POWDIFF	ME	MAE	MAX [c]
Quantum ESPRESSO									
B86bPBE-XDM	0.5	3.6	13.4 (cyt)	3.6	0.0	0.0000	0.0	0.0	0.0
PW86PBE-XDM	0.4	3.7	14.0 (cyt)	3.7	1.2	0.0191	0.5	0.6	3.1 (oxα)
vdW-DF2	4.2	6.0	14.6 (cyt)	3.5	5.4	0.2136	-1.1	1.1	3.0 (suc)
PBE-XDM	-3.2	4.7	17.9 (cyt)	3.7	2.6	0.2168	0.3	0.5	3.0 (oxβ)
rVV10	15.8	15.8	27.5 (ada)	3.7	16.4	0.2168	-0.6	0.6	1.7 (oxβ)
PBE-D2	3.7	5.8	18.4 (ada)	4.2	4.3	0.2688	-1.4	1.4	5.5 (ada)
BLYP-XDM	4.1	5.5	15.4 (ada)	4.2	4.6	0.4207	-1.5	1.8	4.8 (ant)
vdW-DF1	4.4	6.3	19.1 (ada)	5.0	6.3	0.6031	-3.5	3.5	7.1 (ant)
SIESTA									
PBE-D2/DZP	7.4	11.4	25.2 (ant)	3.6	10.8	0.0102	0.8	0.8	1.5 (ant)
PBE-D2/DZP [a]	7.3	11.3	26.0 (ant)	6.3	10.6	0.1130	-2.2	3.0	31.6 (ura)
vdW-DF2/DZP	20.1	20.1	37.0 (suc)	3.4	20.6	0.1406	-0.2	0.6	2.0 (ada)
vdW-DF1/DZP	22.3	22.3	47.2 (suc)	4.4	22.8	0.2363	1.1	1.8	5.1 (oxα,suc)
CRYSTAL									
PBEh-3c [b]	-1.3	5.7	13.8 (amm)	3.1	4.7	0.1660	-1.0	1.1	2.9 (ure)
PBE-D2/POB-DZVP	122.4	122.4	424.3 (ant)	18.0	122.9	0.6711	-17.0	17.0	86.9 (ant)
PBE-D3/POB-DZVP	114.3	114.3	401.5 (ant)	13.5	114.8	0.5487	-12.0	12.0	71.0 (ant)
sHF-3c	-0.5	6.5	17.2 (oxα)	5.8	7.0	0.4146	-3.8	3.8	11.3 (suc)
HF-D2/MINI(s)	21.3	21.9	51.8 (oxβ)	5.3	23.5	0.5717	-3.9	4.0	14.9 (suc)
DFTB+									
SCC-DFTB3-D3(BJ)	7.1	12.8	34.6 (oxβ)	14.0	12.8	0.6689	-13.4	13.4	30.1 (ura)

[a] SCC-DFTB3-D3 geometries were used as the starting point for the PBE-D2/DZP optimization.
[b] Anthracene and naphthalene, which have small band gaps, have been excluded from the statistics for PBEh-3c because the SCF cycle failed to converge. This is in agreement with previous reports.[193]
[c] Labels in parentheses identify the X23 system that yields the maximum absolute error (ada: adamantane, amm: ammonia, ant: anthracene, cyt: cytosine, oxα: α-oxalic acid, oxβ: β-oxalic acid, suc: succinic acid, ura: uracil, ure: urea).
[d] Computed lattice energies for the X23 set of molecular crystals can be found in Appendix A.

multilevel method is particularly effective in reproducing pure B86bPBE-XDM results, with both giving the same MAE. The powder diffraction similarity measure indicates that the crystal geometries of these two methods are essentially coincident (POWDIFF = 0.0102) and the maximum absolute deviation between both methods across the whole X23 set is only 1.5 kJ/mol.

Slightly lower MAEs can be obtained using vdW-DF2 or PBEh-3c for the geometry optimization, but the POWDIFF similarity measure is not as good as PBE-D2/DZP. It is also interesting to note that HF-D2/MINI(s), which is essentially sHF-3c with a different dispersion and minus the gCP and SRB corrections, achieves a smaller MAE than sHF-3c when combined with B86bPBE-XDM in a multilevel method, indicating that the gCP and SRB corrections are not necessarily beneficial when sHF-3c is supplemented with a final single-point calculation, as in the method proposed

by Brandenburg and Grimme.[284] While faster than PBEh-3c and moderately slower than PBE-D2/DZP in SIESTA, PBE-D2/POB-DZVP and PBE-D3/POB-DZVP, within CRYSTAL17, grossly overestimate lattice energies because they also lack corrections for BSSE. Additional single-point calculations do not improve on the MAEs as much as for the other methods given that the structures differ significantly from the B86bPBE-XDM reference geometries.

While the structures obtained with the PBE-D2/DZP method (SIESTA) are quite comparable to the B86bPBE-XDM geometries, this could be an artefact of the optimization procedure, since for all entries in Table 5.1, B86bPBE-XDM starting geometries were used. To test this, PBE-D2/DZP and B86bPBE-XDM//PBE-D2-DZP calculations were rerun starting from the worst geometries available (SCC-DFTB3-D3(BJ)). Although the pure PBE-D2/DZP results are relatively independent of the starting geometry, the multilevel method is penalized, with a MAE of 6.3 kJ/mol, relative to experiment, when the poor starting geometries are used. The discrepancy between the two sets of B//A results is a consequence of the fairly loose default convergence thresholds for geometry optimization within SIESTA. The MAE using the DFTB starting geometries and tighter optimization convergence criteria (0.01 eV/Å force and 0.02 GPa stress convergence thresholds) decreases to 4.8 kJ/mol. Thus the choice of convergence thresholds has a non-negligible effect on the composite lattice energies, particularly when the starting geometries are quite dissimilar to those obtained with B86bPBE-XDM, but the looser thresholds are retained in practice to lower the computational cost.

The extraordinary agreement between B86bPBE-XDM and PBE-D2/DZP geometries prompts the question of whether it is possible to further reduce the cost of these SIESTA calculations by adjusting one or more input parameters, while maintaining a similar quality in the predicted geometries. In a practical CSP context, it is undesirable to explore the convergence of these parameters for each individual candidate structure, so there is value in studying their effect beforehand. Specifically, the effect of reducing the energy-shift parameter (i.e. the "size" of the atomic orbitals), and basis-set size was explored. Table 6.3 shows the results of these calculations. There is a small impact from using a DZP, rather than a TZP, basis set; the computational savings from using the former outweigh the 0.7 kJ/mol improvement in the agreement with the reference B86bPBE-XDM energies. Likewise, a $2\times2\times2$ $\mathbf{k}$-point mesh seems to be enough for the crystals in the X23 set and, since these are sufficiently small compared to the typical molecular crystal sizes in CSP, it is reasonable to assume that this $\mathbf{k}$-point mesh can be used routinely. In contrast, the value of the "energy-shift" parameter, which is linked to the spatial extent of the numerical atomic

Table 5.2: The MAE for the X23 set using bare low-cost (PBE-D2/DZP; A), and composite (B86bPBE-XDM//PBE-D2/DZP; B//A), methods. Comparisons are made to the B86bPBE-XDM lattice energies, in kJ/mol per molecule.[f] The POWDIFF similarity measure referred to the B86bPBE-XDM equilibrium structures is also shown.

Test	A	B//A	POWDIFF
Reference [a]	10.8	0.8	0.0102
Tighter opt. convergence [b]	10.8	0.7	0.0382
Reference, DFTB geom. [a]	10.6	3.0	0.1130
Tighter opt. convergence, DFTB geom. [b]	10.8	1.8	0.0811
k-point mesh 2×2×2 [c]	10.5	0.8	0.0358
Energy shift (0.02 Ry) [d]	124.2	6.9	0.4876
TZP basis set [e]	4.7	0.1	0.0253

[a] The PBE-D2/DZP method used in Table 5.1 and the rest of this chapter, with maximum force and stress components convergence criteria of 0.04 eV/Å and 1.0 GPa, a 0.001 Ry energy shift parameter, and a 4×4×4 k-point mesh starting from B86bPBE-XDM geometries, or where indicated, from SCC SCC-DFTB3-D3 geometries.
[b] Tighter force and stress convergence thresholds (0.01 eV/Å and 0.02 GPa for maximum component values),
[c] a 2×2×2 k-point mesh,
[d] the default SIESTA energy shift parameter (0.02 Ry), and
[e] a TZP basis set were used.
[f] Computed lattice energies for the X23 set of molecular crystals can be found in Appendix A.

orbitals in SIESTA calculations,[282,291] is crucial in obtaining good-quality geometries and accurate single-point energies. A 0.001 Ry cutoff was found to be sufficient, but the default 0.02 Ry value used in SIESTA results in unacceptable errors.

5.3.2 Relative Lattice Energies

Next, focus is shifted to whether low-cost and multilevel methods can describe lattice energy differences using the enantiomeric excess (ee) data for chiral compounds. For the subset of these compounds in which the racemate is more stable than the enantiopure crystal, the relative free energy between the two phases can be directly calculated from the experimental ee of a solution saturated with both enantiomers, in contact with its enantiopure crystal and the racemic crystal forms. At equilibrium, i.e., at the eutectic point, the following equilibria can be written down:[128,295]

$$\mathrm{L}_{(\mathrm{s})} \rightleftharpoons^{\Delta G_{\mathrm{sub}}} \mathrm{L}_{(\mathrm{g})} \rightleftharpoons^{\Delta G_{\mathrm{solv}}} \mathrm{L}_{(\mathrm{solv})} \tag{5.2}$$

$$\mathrm{DL_{(s)}} \overset{\Delta G_{\mathrm{sub}}}{\rightleftharpoons} \mathrm{D_{(g)}} + \mathrm{L_{(g)}} \overset{G_{\mathrm{solv}}}{\rightleftharpoons} \mathrm{D_{(solv)}} + \mathrm{L_{(solv)}} \tag{5.3}$$

where ΔG_{sub} and ΔG_{solv} are Gibbs free energies associated to the sublimation of the crystalline forms, and to the solvation of the gas-phase molecules, of L and DL, respectively.

Then, the solubility products, K_{sp}, of both the enantiopure, L, and racemate, DL, crystals in solution can be written as

$$K_{\mathrm{sp}}(\mathrm{L}) = [\mathrm{L}] = \exp\left(-\frac{\Delta G_{\mathrm{sub}}(\mathrm{L}) + \Delta G_{\mathrm{solv}}}{RT}\right), \tag{5.4}$$

and

$$K_{\mathrm{sp}}(\mathrm{DL}) = [\mathrm{D}][\mathrm{L}] = \exp\left(-2\frac{\Delta G_{\mathrm{sub}}(\mathrm{DL}) + \Delta G_{\mathrm{solv}}}{RT}\right). \tag{5.5}$$

From the above expressions, the ee can be calculated using:

$$\mathrm{ee} = \frac{[\mathrm{L}] - [\mathrm{D}]}{[\mathrm{L}] + [\mathrm{D}]} \times 100 = \frac{(1/4) - \epsilon^2}{\epsilon^2 + (1/4)} \times 100, \tag{5.6}$$

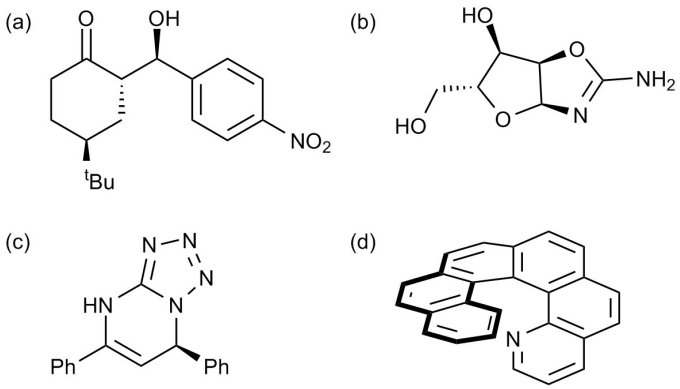
where [L] and [D] are concentrations in solution and:

$$\epsilon = \exp\left(-\frac{\Delta G_{\mathrm{L-DL}}}{RT}\right) \approx \exp\left(-\frac{\Delta E_{\mathrm{L-DL}}}{RT}\right) \tag{5.7}$$

with $\Delta G_{\mathrm{L-DL}}$ the Gibbs free energy difference between the enantiopure (L) and the racemate (DL) crystals at the experimental temperature and $\Delta E_{\mathrm{L-DL}}$ the corresponding lattice energy difference. In the last term, vibrational free energy contributions are neglected. The $(1/4)$ factors in equation 5.6 are different from previous work[128] in that an $RT \ln 2$ contribution proposed by Price *et al.*[295] to account for the entropy of mixing in solution was added to ΔE, in favor of dissolution of the racemate form in an enantiopure solution. This formula is valid only in cases where the racemate (DL) is more stable than the enantiopure (L) phase by at least the entropy of mixing term ($\Delta E_{\mathrm{L-DL}} > RT \ln 2$). Otherwise, a conglomerate of L and D crystals is formed instead, and the ee in solution is zero.

The ee information for the 10 pairs of amino acid crystals previously compiled[128] is supplemented with four additional compounds for which experimentally measured ee's are available,[2,296–298] shown in Figure 5.3. Given the exponential dependence of the ee on the computed electronic-energy differences, and the precision with which it can be experimentally determined,

Figure 5.3: Molecular structures for (a) aldol[296] (2S,4S)-4-(*tert*-butyl)-2-[(R)-hydroxy(4-nitrophenyl)methyl]cyclohexanone, (b) oxazoline[297] (ribo-amino-oxazoline), (c) tetrazole[298] (5,(7R)-diphenyl-4,7-dihydrotetrazolo[1,5-a]pyrimidine), and (d) N-helicene[2] (P-1-aza[6]helicene), added to the set of ten amino acids to form the EE14 set.



these measures provide an excellent benchmark for the relative energy differences between crystal forms.

Figure 5.4 shows the performance of different low-cost and multilevel methods in the calculation of relative lattice energies, while Table 5.3 gives the relevant statistics relative to computed B86bPBE-XDM data. Overall, the agreement with the theoretical model (black curve) is improved when final B86bPBE-XDM single-point calculations are performed, which is consistent with the discussion in Section 5.3.1. In this case, however, the performance of the methods based on a DZP basis set of numerical orbitals (PBE-D2/DZP and vdW-DF1,2/DZP) is surprisingly good, comparable to that of B86bPBE-XDM. This is in striking contrast with the poor results for absolute lattice energies shown in Table 5.1. This is an indication that, while BSSE originating from the finite nature of the orbitals in SIESTA methods affects absolute lattice energies very strongly, its impact on the relative lattice energies, which are the quantity of interest in CSP, is comparatively small. The use of tighter convergence criteria again provided similar results to those discussed for the X23 set in the previous section.

Pure sHF-3c has a MAE similar to the DZP methods and, surprisingly, so does SCC-DFTB3-D3(BJ). However, the relatively poor sHF-3c and PBEh-3c geometries make the corresponding multilevel method less accurate than with the DZP methods, which may explain the relatively poor performance of the TPSS-D3//HF-3c multilevel method in the sixth blind test.[284] It is important

to note that PBEh-3c and the DZP methods are one order of magnitude slower than sHF-3c. Furthermore, as in the case of the X23 crystals, PBEh-3c could not be applied in some cases due to SCF convergence failures (N-helicene and tetrazole), for similar reasons as in anthracene and naphthalene (band gap closing).

The performance of the multilevel method based on HF-D2/MINI(s) is only slightly worse than sHF-3c, which is another indication that the two corrections other than dispersion in 3c are not very effective at improving the accuracy of the corresponding multilevel method. In contrast, the MAE of SCC-DFTB3-D3(BJ) is on par with other methods, which is notable considering the enormous average error in the calculation of absolute lattice energies (Table 5.1). Contrarily, single-point calculations at the sHF-3c, HF-D2/MINI(s), and SCC-DFTB3-D3(BJ) result in higher average errors than using the low-cost methods alone. This observation highlights the importance of assessing low-cost methods regarding their ability to produce accurate equilibrium geometries, and not just lattice energies.

Figure 5.4: Experimental enantiomeric excess as a function of calculated energy differences between the enantiopure and racemate crystal structures, $\Delta E_{\mathrm{L-DL}}$. The solid black line represents the ee as a function of calculated $\Delta E_{\mathrm{L-DL}}$ (Eq. 5.6). Accurate relative energies yield good agreement between the experimental ee and the black curve (vibrational contributions are neglected). Data points for energies resulting from the geometry relaxation using a low-cost method and those obtained from additional B86bPBE-XDM single-point energy calculations are represented by open circles and closed diamonds, respectively.

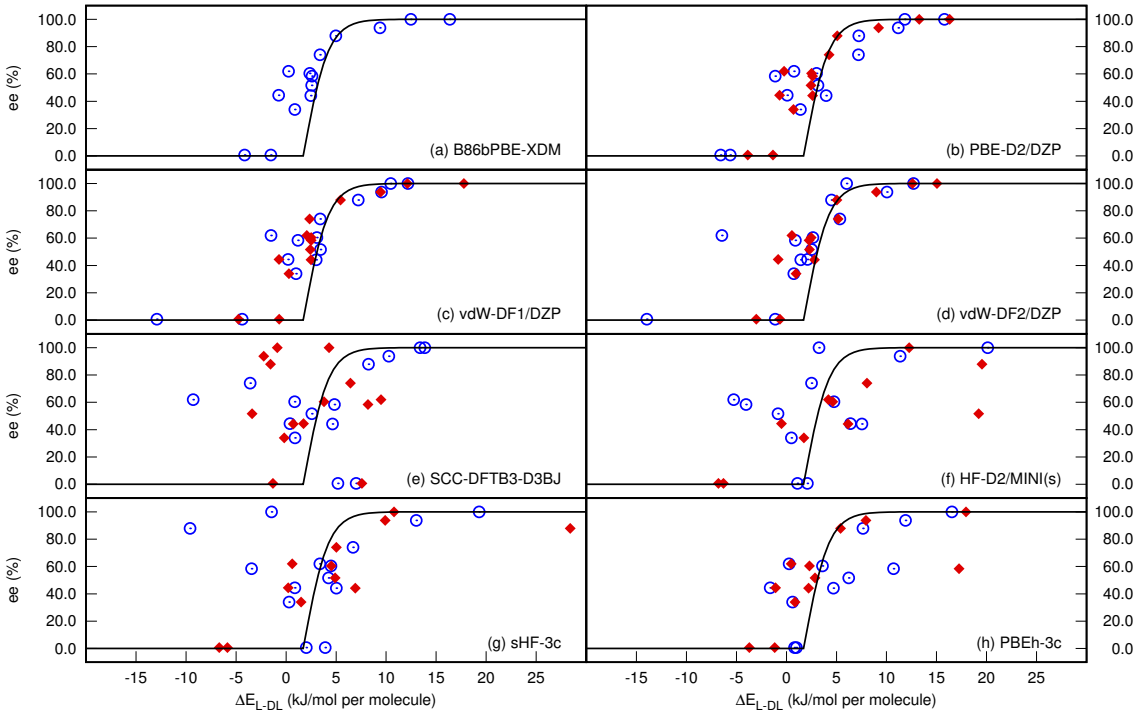


Table 5.3: Statistics for the EE14 set of relative lattice energies using various low-cost methods.[c] Mean error (ME), mean absolute error (MAE), and maximum absolute error (MAX) relative to experiment (Expt.) or fully-relaxed B86bPBE-XDM results (DFT). "A" and "B//A" refer, respectively, to the results obtained from the low-cost optimization directly or to a B86bPBE-XDM single-point energy calculation at the low-cost equilibrium geometry. All values are in kJ/mol per molecule, except POWDIFF similarity measures, which are dimensionless.

	Expt.				DFT				
	A			B//A	A		B//A		
Method	ME	MAE	MAX [b]	MAE	MAE	POWDIFF	ME	MAE	MAX [b]
Quantum ESPRESSO									
B86bPBE-XDM	-1.5	2.1	5.9 (glu)	2.1	0.0	0.0000	0.0	0.0	0.0
SIESTA									
PBE-D2/DZP	-1.3	3.2	8.3 (glu)	2.0	1.7	0.0169	0.1	0.3	0.9 (nhe)
vdW-DF2/DZP	-2.7	3.5	15.7 (asp)	1.9	3.0	0.1112	0.2	0.5	1.8 (nhe)
vdW-DF1/DZP	-2.3	3.3	14.7 (asp)	2.1	1.9	0.1930	0.1	0.5	1.8 (oxa)
CRYSTAL									
PBEh-3c [a]	4.2	4.5	11.9 (his)	4.1	2.4	0.1411	1.3	1.7	14.7 (cys)
sHF-3c	-0.8	3.5	14.7 (leu)	8.3	5.4	0.2011	6.7	7.9	42.0 (cys)
HF-D2/MINI(s)	3.1	6.1	28.6 (leu)	6.6	9.4	0.3007	8.3	9.4	27.8 (his)
DFTB+									
SCC-DFTB3-D3(BJ)	-0.7	3.9	12.8 (oxa)	4.2	3.6	0.3611	-1.3	6.2	17.3 (ser)

[a] N-helicene and tetrazole encountered similar SCF convergence problems as anthracene and naphthalene, and have thus been excluded from the statistics for PBEh-3c.[193]

[b] Labels in parentheses identify the system that yields the maximum absolute error (asp: aspartic acid, cys: cysteine, glu: glutamic acid, his: histidine, leu: leucine, nhe: N-helicene, oxa: oxazoline, ser: serine).

[c] Computed energy differences between the enantiopure and racemate forms of EE14 set chiral crystal structures can be found in Appendix A

5.3.3 Crystal Structure Prediction

Finally, a practical CSP application using the best-performing multilevel approach in the previous sections, B86bPBE-XDM//PBE-D2/DZP, is considered. In a recent article, it has been shown how the electronic properties of an organic semiconductor based on the chiral 1-aza[6]helicene molecule[2] (see Figure 5.3(d)) can be dramatically affected by the racemic or enantiopure nature of the material through differences in the racemate and enantiopure crystal structures. In the CSP part of the study, initial candidate structures were ranked using the W99 repulsion-dispersion force field,[79] with electrostatics described by a distributed multipole analysis.[299] Fifty candidate crystal structures (47 racemate and 3 enantiopure crystals) were re-optimized and re-ranked with B86bPBE-XDM. The set of candidates includes the experimentally-observed racemic and enantiopure crystal structures. The initial ranking by W99 is shown in Figure 5.5(a), and the re-ranking with B86bPBE-XDM is shown in panel (b). Detailed data can be found in Appendix A.

The initial ranking from the W99 force field proved to be inadequate as the experimentally isolated form of the racemate crystal was ranked highest in energy and the wrong relative ordering of the enantiopure and racemate crystal phases was predicted. According to the experimental measurements, the ee in solution for this crystal is 74%, which corresponds to a racemate more stable than the enantiopure crystal by an energy difference of 4.1 kJ/mol at room temperature.[2] Figure 5.5 shows that B86bPBE-XDM not only recovers the correct relative ordering of both phases, but also gives the experimentally-observed racemate as the phase with the lowest electronic energy among all the candidates. The energy difference between racemate and enantiopure crystals is 3.8 kJ/mol, in excellent agreement with the EE14 value.

Figure 5.5(c) and (d) show how the energy profile of this crystal-energy landscape is affected by re-optimization using PBE-D2/DZP. The equilibrium crystal structures are different depending on whether the W99 or B86bPBE-XDM geometries are used as starting points of the optimization. As in previous sections, this disagreement is caused by the loose geometry optimization convergence criteria (0.04 eV/Å for the force and 1.0 GPa for the stress components). Subsequent tests on the minimum-energy racemate structure (circled full point in Figure 5.5) showed that the true PBE-D2/DZP minimum-energy cell dimensions fall between the two sets of results (see Table 5.4). Although this disagreement could be resolved using tighter convergence thresholds for forces and stresses, this would increase the computational cost by approximately a factor of three.

Figure 5.5: Lowest-energy enantiopure and racemate crystal structures of 1-aza[6]helicene reranked with the B86bPBE-XDM//PBE-D2/DZP composite method. Crystal structures were obtained from a previous CSP study on chiral helicenes,[2] which had been initially ranked with the W99 force field (a) and re-ranked using B86bPBE-XDM (b). Panels (c) and (d) depict the energy profiles obtained when performing full relaxations with PBE-D2/DZP starting from either W99 or Quantum ESPRESSO geometries, respectively. Panels (e) and (f) show results from further re-ranking the PBE-D2/DZP structures using B86bPBE-XDM single-point energy calculations. The colour scheme follows the relative energetic ordering obtained with full B86bPBE-XDM relaxations (b) and is kept constant for all other panels in order to compare how the energies shift when using the force fields and low-cost methods. Experimentally observed structures are encircled. The experimentally isolated enantiopure crystal is set as the zero of energy.

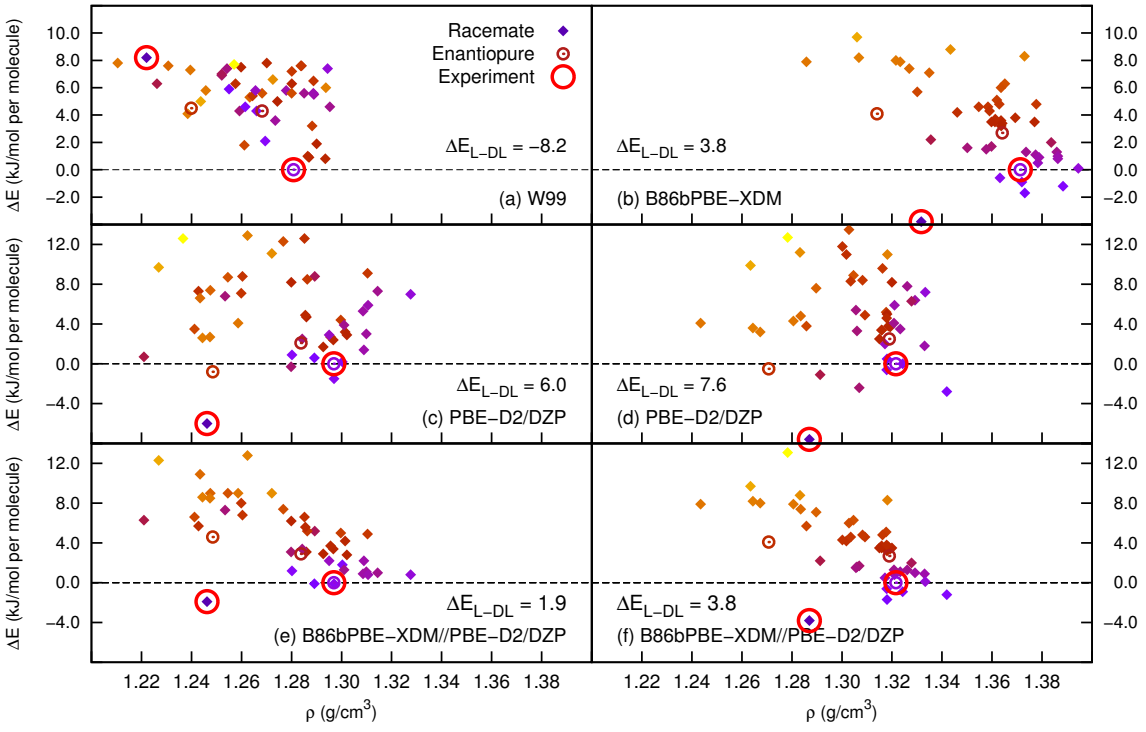


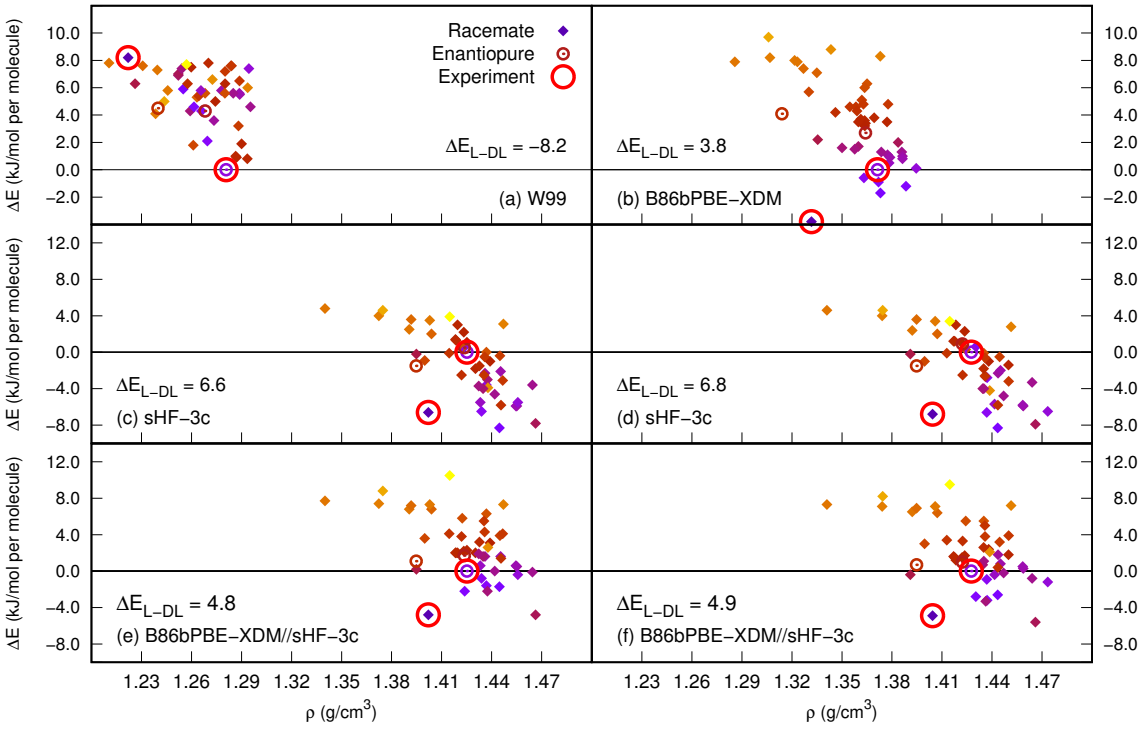
Table 5.4: Relative energy differences, in kJ/mol per molecule, unit-cell parameters, in Å, and volume, in Å^3, of the lowest-energy racemate structure from the CSP study of 1-aza[6]helicene computed by fully relaxing the structure using PBE-D2/DZP (SIESTA) and by starting from either the FF (W99) or DFT (B86bPBE-XDM) geometry. Using tighter convergence criteria for the maximum force (0.01 eV/Å) and stress components (0.02 GPa), instead of the program defaults (0.04 eV/Å and 1.0 GPa, respectively), yield effectively the same minimum-energy structure, but at the expense of requiring a significantly greater number of optimization steps and computational cost.

Test	ΔE	a	b	c	Unit-Cell Volume
FF geom.	0.1	9.223	10.273	17.712	1681.4
FF geom., Tighter opt. convergence	0.4	9.204	10.390	17.615	1684.5
DFT geom.	2.6	9.178	10.416	17.358	1659.5
DFT geom., Tighter opt. convergence	0.0	9.217	10.412	17.552	1684.4

As expected, the relative energies from using the low-cost PBE-D2/DZP method do not reproduce the B86bPBE-XDM results in Figure 5.5(b), but simply performing an additional single-point energy calculation on each crystal structure, as shown in Figure 5.5(e) and (f), recovers crystal-energy landscapes in good agreement with the reference high-level method. These plots show that the B86bPBE-XDM//PBE-D2/DZP method predicts the correct energy ordering, even if poor W99 starting geometries are used. The MAE between B86bPBE-XDM//PBE-D2/DZP and the B86bPBE-XDM results for all 50 relative energies is 0.3 kJ/mol when the B86bPBE-XDM geometries are used as the starting point of the PBE-D2/DZP optimizations and 1.6 kJ/mol when the W99 geometries are used instead. This slight discrepancy between the MAEs, or alternatively the energy landscapes (compare either panels (c) and (d) or (e) and (f)), is considered to be the result of using loose convergence thresholds for forces and stresses during geometry optimization. Nonetheless, the accuracy provided by the use of loose convergence criteria is sufficient to correctly identify the most-promising candidate structures for subsequent high-level structure refinement. Thus, given that both energy landscapes are similar and the proper relative ordering between enantiopure and racemate phases is recovered, this multilevel approach is suitable for crystal structure prediction purposes.

One can also use sHF-3c as an alternative method to generate geometries cheaply, as it is 1-2 orders of magnitude faster than PBE-D2/DZP (SIESTA). The results from B86bPBE-XDM//sHF-3c are roughly equivalent to the B86bPBE-XDM//PBE-D2/DZP energy landscapes (see Figure 5.6). In addition, starting from either the W99 or B86bPBE-XDM geometries, similar energy landscapes were obtained upon structural optimization within CRYSTAL17, something that cannot be said from using the default convergence criteria in SIESTA.

Figure 5.6: Lowest-energy enantiopure and racemate crystal structures of 1-aza[6]helicene reranked with the B86bPBE-XDM//sHF-3c composite method. Crystal structures were obtained from a previous CSP study on chiral helicenes,[2] which had been initially with the W99 force field (a) and re-ranked using B86bPBE-XDM (b). Panels (c) and (d) depict the energy profiles obtained when performing full relaxations with sHF-3c starting from either W99 or plane-wave DFT geometries, respectively. Panels (e) and (f) show results from further re-ranking the sHF-3c structures using B86bPBE-XDM single-point energy calculations. The colour scheme follows the relative energetic ordering obtained with full B86bPBE-XDM relaxations (b) and is kept constant for all other panels in order to compare how the energies shift when using the force fields and low-cost methods. Experimentally observed structures are encircled. The experimentally isolated enantiopure crystal is set as the zero of energy.



5.4 Conclusions

In this chapter, the applicability and performance of computationally inexpensive (cheap) and multilevel approaches to CSP were studied. Multilevel methods are a composite of two techniques in which a cheap method is used for geometry optimization and a more expensive and accurate method is used in a final single-point calculation. Multilevel methods are very popular in molecular quantum chemistry, as they furnish an accuracy similar to the expensive method with a much reduced computational cost. In particular, several cheap methods were examined: PBE-D2, vdW-DF1, and vdW-DF2 with a double-zeta basis set of numerical orbitals (SIESTA); minimal or small Gaussian-basis-set calculations (PBEh-3c, sHF-3c, HF-D2/MINI(s)); and self-consistent-charge dispersion-corrected density-functional tight binding (SCC-DFTB3-D3). These cheap methods have been evaluated both alone and in a multilevel method where the final single-point calculation is run with B86bPBE-XDM, which has the best performance of all current density-functional methods in the calculation of absolute lattice energies (measured for the X23 set).

The performance of various cheap and multilevel methods was evaluated using three tests: absolute lattice energies (the X23 set), relative lattice energies (the EE14 set), and a practical CSP application (enantiopure and racemic forms of 1-aza[6]helicene, an organic semiconductor). The EE14 set is a new benchmark set for relative lattice energies derived from experimental measurements of enantiomeric excess of a solution in contact with the racemic and enantiopure crystals of the same chiral compound, comprising 14 relative lattice energies.

The results show that absolute lattice energies are much more difficult to model than relative lattice energies, the latter being the important quantity in CSP. This is especially true for cheap methods. For instance, SCC-DFTB3-D3(BJ) gives a MAE of 12.8 kJ/mol on the X23, but only of 3.6 kJ/mol on the EE14. In comparison, PBEh-3c, whose corrections have been developed to improve binding- and lattice-energy accuracy relative to minimal-basis-set HF, has a relatively low MAE (5.7 kJ/mol) for the X23, but the MAE on the EE14 is higher than SCC-DFTB3-D3(BJ) (4.5 kJ/mol). This emphasizes the importance of targeting relative lattice energies in the development of cheap methods for CSP.

Regarding the performance of multilevel methods (cheap methods with a final B86bPBE-XDM step), in general the performance of all multilevel methods for the X23 and EE14 improves relative to the corresponding cheap methods alone, except for those cheap methods that give equilibrium

crystal geometries very different from B86bPBE-XDM. In particular, these results show that PBE-D2 with a double-ζ basis of numerical orbitals (PBE-D2/DZP) is particularly efficient at recovering B86bPBE-XDM geometries, despite its poor performance for absolute lattice energies. This leads to an optimal multilevel method, with MAEs for the X23 and EE14 of 3.6 and 2.0 kJ/mol. In comparison, the MAEs of pure B86bPBE-XDM on the X23 and EE14 are 3.6 and 2.1 kJ/mol, respectively.

It should be noted, however, that the performance and computational cost of B86bPBE-XDM//PBE-D2/DZP depends critically on the choice of calculation parameters for the SIESTA method. In particular, the energy shift parameter, which controls the spatial extent of the orbitals, needs to be considerably lower than the default (0.001 Ry, as suggested previously in the literature[282,291] is a good option). Likewise, choosing convergence criteria for the geometry optimization that strike a balance between performance and cost is essential. It has been determined here that the default convergence thresholds in SIESTA (0.04 eV/Å and 1.0 GPa maximum values for the components of force and stress, respectively) give an adequate compromise between the two, although they introduce a small error if the initial geometries are poor.

Because of its good performance, the B86bPBE-XDM//PBE-D2/DZP composite method was tested on the set of helicene candidate structures from a recent CSP study.[2] It was found that, starting from the B86bPBE-XDM equilibrium geometries, the relative energies from B86bPBE-XDM//PBE-D2/DZP are essentially indistinguishable from pure B86bPBE-XDM. When relatively poor starting geometries are used (the equilibrium geometries using the W99 force field), then the performance is slightly worse, but the multilevel method is still able to recover a reliable energy landscape. Thus, it can be concluded that multilevel approaches, in particular B86bPBE-XDM//PBE-D2/DZP, are excellent candidates for energy-ranking functions in molecular CSP.

CHAPTER 6

NON-COVALENT INTERACTIONS IN MOLECULAR CRYSTALS: EXPLORING THE ACCURACY OF THE EXCHANGE-HOLE DIPOLE MOMENT MODEL WITH LOCAL ORBITALS

Reprinted with permission from **L. M. LeBlanc***, J. A. Weatherby, A. Otero-de-la-Roza*, E. R. Johnson*. Non-Covalent Interactions in Molecular Crystals: Exploring the Accuracy of the Exchange-Hole Dipole Moment Model with Local Orbitals. *J. Chem. Theory Comput.* **2018**, *14*, 5715-5724. DOI: https://pubs.acs.org/doi/full/10.1021/acs.jctc.8b00797/

Contributions to the manuscript LML implemented the B86bPBE-XDM model within SIESTA and performed the calculations, except for the graphite exfoliation curves reported in Figure 6.2, and the convergence testing of force and stress thresholds, along with the use of Broyden and conjugate-gradient methods, which were performed by JAW, and the Numol calculations, which were performed by ERJ. LML wrote the first draft of the manuscript. JAW, AOR and ERJ contributed to the final version of the manuscript. ERJ supervised the project.

6.1 Motivation

As was dicussed in previous chapters of this thesis and, in particular in Chapter 2, there have been long-standing efforts to develop methods capable of modelling dispersion interactions accurately in

both molecular and solid-state systems, especially in the field of density-functional theory. While satisfactory accuracy can often be obtained with several dispersion-corrected DFT methods, it is interesting to develop computationally inexpensive variants of these methods for purposes such as *ab initio* molecular dynamics simulations or crystal structure prediction.[164,300–302] In the previous chapter, the use of composite approaches was examined for such purposes.

One of the ways that was been examined to reduce the computational cost of dispersion-corrected DFT methods is to represent the Kohn-Sham orbitals in a basis set that ensures asymptotic linear scaling with system size, known as the SIESTA method.[130,131] In SIESTA, linear scaling is achieved through a combination of $\mathcal{O}(N)$ algorithms and basis sets composed of finitely supported atom-centred numerical orbitals.[286–288] The SIESTA method is implemented in the program of the same name,[130,131] whose fourth major version release allows the treatment of London dispersion interactions. Two dispersion methods are currently implemented in SIESTA: Grimme's semiempirical dispersion model (DFT-D2)[165,289] and the non-local van-der-Waals density functionals developed by Langreth and co-workers (vdW-DF1[176] and vdW-DF2[177]). DFT-D2 uses an asymptotic energy expression with fixed empirical dispersion coefficients; it is cheap and relatively accurate, but the coefficients are empirical and do not change with the chemical environment, which is essential in certain systems.[151] The vdW-DF methods incorporate dispersion effects using a non-local correlation energy functional. Non-local vdW-DF functionals are non-empirical and "seamless", but their use increases the computational cost significantly.[138]

An alternative approach that includes dispersion effects is the exchange-hole dipole moment (XDM) dispersion model.[139,155] In XDM, the interatomic dispersion coefficients are calculated from first principles using the self-consistent density and kinetic-energy density. This makes the dispersion coefficients sensitive to the chemical environment and non-empirical, while retaining the computational and conceptual simplicity of an asymptotic, pairwise dispersion expansion.

In this chapter, the first implementation of the XDM dispersion model in combination with a finite-support local-orbital method for periodic solids in SIESTA is presented. The resulting XDM implementation is then parametrized using the Kannemann-Becke (KB49) set of molecular dimers,[162] and tested on solid-state systems such as graphite and phosphorene exfoliation, and the X23 set of molecular solids.[142,145] The implications of using local basis sets in SIESTA, as opposed to delocalised plane-waves, for the treatment of non-covalent interactions are discussed.

6.2 Computational Methods

SIESTA calculations: The B86bPBE[109,110] functional was implemented in an in-house version of the 4.0b-485 SIESTA code, as it typically yields the best results when paired with the XDM dispersion model.[139,142,148] Additional calculations were performed using the PBE[109] functional, with either XDM or Grimme's D2 dispersion correction.[165] Both DZP and TZP basis sets were considered. The D2 damping parameters were set to $s_r = 1.1$ and $s_6 = 0.50$ or 0.64, for DZP and TZP, respectively.[289] DZP is the standard basis set implemented in SIESTA, whereas TZP was constructed and optimized for H, N, and O atoms by Louwerse and Rothenberg,[283] and further extended to C atoms by Carter and Rohl.[282] The confinement radius of the finite numerical orbitals was set by an "energy shift" parameter of 0.001 Ry, found to be sufficient in reducing basis-set superposition error and yielding converged energies.[282,291] The real-space integration grid cutoff value for charge densities and potentials was set to 200 Ry, consistent with the work presented in Chapter 5. Troullier-Martins-type[121,122] norm-conserving pseudopotentials[120,290] were generated and tested for both density functionals with the ATOM code.[132] These pseudopotentials included nonlinear core corrections.[303]

Plane-wave calculations: Reference benchmark calculations were also performed with B86bPBE-XDM, PBE-XDM, and PBE-D2 as implemented in Quantum ESPRESSO[127] version 5.1, using plane-waves / pseudopotentials[213,214] within the PAW formalism.[124] The damping function parameters, a_1 and a_2, were set to 0.6512 and 1.4633 Å for B86bPBE-XDM, and to 0.3275 and 2.7673 Å for PBE-XDM.[139] The s_6 damping parameter for PBE-D2 was set to a value of 0.75. Wave-function and density cutoffs were set to 80 Ry and 800 Ry, respectively. Structure relaxations were performed with tighter thresholds for convergence of the energies and forces, i.e., 10^{-5} Ry and 10^{-4} Ry/bohr, respectively.

k-point grid sampling and structure relaxation: For all calculations, a Brillouin-zone sampling with a $4 \times 4 \times 4$ **k**-point Monkhorst-Pack (MP) scheme was used to treat crystal structures, whereas isolated molecules were studied at the Γ-point only. During structure optimization, unit-cell parameters and atomic positions for crystalline systems were allowed to fully relax, while for molecules in a large-vacuum simulation box, only the atomic coordinates were allowed to vary. Geometry relaxations were carried out using the conjugate-gradient (CG) algorithm. However, in the last stages of this study, the modified Broyden algorithm[304] implemented in SIESTA was found to be more efficient than CG, and yielded the same results with tight convergence criteria (see below). Therefore, its use for geometry relaxations in molecular crystals is recommended.

Similarity index: As a tool to measure similarity between crystal structures, the POWDIFF utility in CRITIC2[293] was used. This tool is based on the comparison of powder diffraction patterns using a cross-correlation function,[294] and ranges in value between zero and one. A result of zero indicates an exact match, while a result of one indicates maximum dissimilarity between two crystal structures.

6.3 Results and Discussion

6.3.1 The XDM Dispersion Model: Implementation, Parametrization, and Testing

6.3.1.1 Theory and Implementation Details

As the XDM dispersion model was detailed in Chapter 2 (Eqs. 2.12 to 2.25), the following discussion serves only to highlight how the density and kinetic-energy density are obtained within the SIESTA code, which utilizes numerical orbitals. In particular, in SIESTA the electron density and kinetic energy density are written as a sum over local orbitals:

$$\rho_\sigma^{\mathrm{val}} = \sum_i |\psi_{i,\sigma}|^2 = \sum_i \left| \sum_{\mu,\nu} c_{\mu,\nu} \phi_\mu \phi_\nu \right| \tag{6.1}$$

and

$$\tau_\sigma = \sum_i |\nabla \psi_{i,\sigma}|^2 = \sum_i \left| \sum_{\mu,\nu} c_{\mu,\nu} \left(\nabla \phi_\mu \phi_\nu + \phi_\mu \nabla \phi_\nu \right) \right|^2 , \tag{6.2}$$

where the $c_{\mu,\nu}$ are the density matrix elements. The sum extends only over the atomic orbitals, ϕ_μ and ϕ_ν, that give a non-zero contribution at the reference point.

The value of the exchange-hole dipole moment, b, (Eq. 2.21) is calculated at every point on the integration grid from the corresponding values for the valence spin density ($\rho_\sigma^{\mathrm{val}}$), its gradient norm ($|\nabla \rho_\sigma^{\mathrm{val}}|^2$), its Laplacian ($\nabla^2 \rho_\sigma^{\mathrm{val}}$), and the spin kinetic-energy density (τ_σ). To validate the implementation, these quantities and the resulting b parameter for argon are compared to the values obtained using the XDM implementations in Quantum ESPRESSO (plane waves),[148] and the numerical-orbital Numol program,[305] as shown in Figure 6.1. The SIESTA density, its derivatives, and the kinetic-energy density are in good agreement with Numol and Quantum ESPRESSO, except close to the core regions, where the three methods differ in their treatment (Numol is an all-electron

code, while SIESTA and Quantum ESPRESSO use different pseudopotentials). Figure 6.1 confirms that the use of finite-support numerical orbitals in SIESTA does not adversely affect the computed exchange-hole dipole moment values, when compared to either plane-wave or numerical-orbital implementations of XDM.

The XDM molecular dispersion coefficients calculated using SIESTA were also compared to those calculated with Quantum ESPRESSO. The C_6 dispersion coefficients were calculated for the isolated molecules of the X23 lattice-energy benchmark set.[142,145] The mean percent errors obtained with the DZP and TZP finite-support basis sets were 5.6 and -4.9%, respectively, relative to the B86bPBE plane-wave calculations. Again, this indicates that the use of relatively compact, atom-centred basis functions will not strongly affect the dispersion energies calculated using the XDM method.

6.3.1.2 Parametrization of the XDM dispersion model

The XDM dispersion model was parametrized for both the (newly implemented) B86bPBE and PBE functionals by minimising residual errors with respect to high-level benchmark data from the KB49 set of 49 weakly bound molecular dimers.[162] Hybrid functionals, involving a fraction of exact exchange, are not currently implemented in the SIESTA code and were not considered here. In all cases, single-point energy calculations were performed at the fixed benchmark geometries. The optimal parameters obtained by least-squares fit, and the performance of the XDM-corrected methods, are shown in Table 6.1 for the standard DZP basis set and the extended TZP basis set.[282,283] Because the TZP basis is not available for all elements contained in the KB49 data (namely, Si, S, and F), the XDM model was parametrized for a subset of the KB set containing 34 dimers (of a total of 49, not counting the noble gas dimers). For consistency, the parameters obtained by fitting to the 34-dimer set will be used for both the DZP and TZP basis sets in the rest of the chapter.

Figure 6.1: Quantities used to calculate the XDM dispersion coefficients along the internuclear coordinate in solid argon. (a) the valence spin density, $\rho_\sigma^{\mathrm{val}}$; (b) the Weizsaecker term, $|\nabla\rho_\sigma^{\mathrm{val}}|^2/\rho_\sigma^{\mathrm{val}}$; (c) the Laplacian of the spin density, $\nabla^2\rho_\sigma^{\mathrm{val}}$; (d) the spin kinetic-energy density, τ_σ; and (e) the exchange-hole dipole moment, $b = d_{X\sigma}$. All values are in atomic units. The SIESTA (red circles) and Quantum ESPRESSO[127] (blue boxes) calculations used an argon atom centred at the origin of a 3.760 Å cubic box, the PBE functional[109], and only one **k**-point at Γ. Troullier-Martins[121,122,132] and Goedecker/Hartwigsen/Hutter/Teter[306,307] norm-conserving pseudopotentials were used, respectively. The Numol (black lines) calculations used the LDA and an isolated argon dimer. The plots represent the values along the shortest argon-argon contact ($d_{\mathrm{Ar-Ar}} = 3.76$ Å).

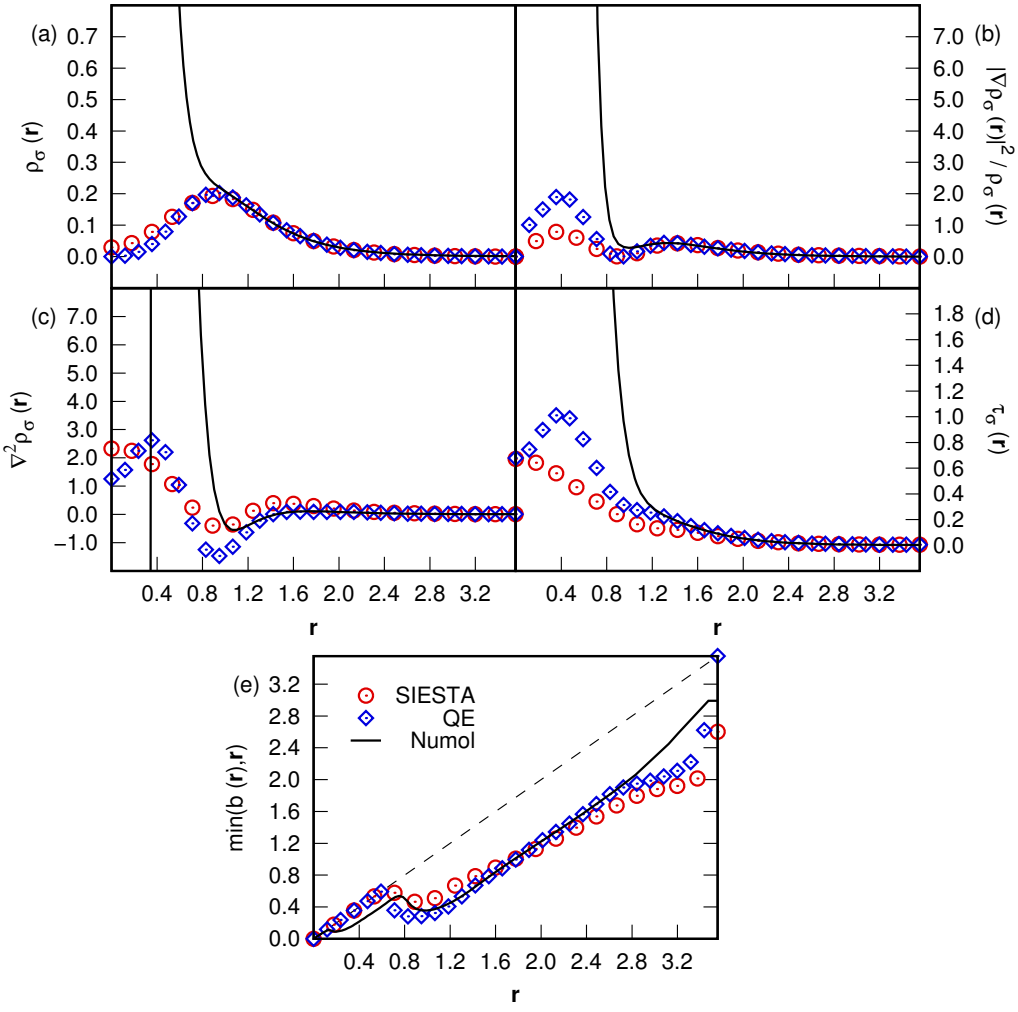


Table 6.1: XDM damping parameters for the PBE and B86bPBE functionals, and selected basis sets, along with the resulting error statistics for the fit set.[g]

Method	N^a	a_1	a_2 (Å)	MPE [b]	MAPE [c]	MAX [d]
SIESTA						
PBE/DZP	49	–	–	-17.0	39.5	232.2 (c6h6-c6h6-stack)
PBE-D2/DZP	49	–	–	20.8	33.5	130.6 (cf4-cf4)
PBE-D2/TZP	34	–	–	25.9	26.5	62.0 (ch4-nh3)
PBE-XDM/DZP	49	1.4588	0.0000 [e]	14.7	30.3	132.0 (cf4-cf4)
PBE-XDM/DZP	34	1.4025	0.0000 [e]	11.2	22.5	94.3 (ch4-nh3)
PBE-XDM/DZP+CP [f]	34	1.2901	0.0000 [e]	-3.6	19.2	73.9 (ch4-c2h4)
PBE-XDM/TZP	34	0.7086	2.3542	1.1	**12.3**	56.0 (ch4-c2h4)
PBE-XDM/TZP+CP [f]	34	1.2480	0.0000 [e]	-3.1	18.6	74.0 (ch4-c2h4)
B86bPBE/DZP	49	–	–	-32.7	43.3	158.6 (c6h6-c6h6-stack)
B86bPBE-XDM/DZP	49	0.2307	3.4210	8.7	21.7	62.4 (ch4-nh3)
B86bPBE-XDM/DZP	34	0.5000	2.5556	4.9	18.2	65.2 (ch4-nh3)
B86bPBE-XDM/DZP+CP [f]	34	1.2343	0.0000 [e]	-9.7	20.4	80.4 (c6h6-ch4)
B86bPBE-XDM/TZP	34	1.3543	0.0000 [e]	-1.2	**11.2**	40.4 (ch4-c2h4)
B86bPBE-XDM/TZP+CP [f]	34	1.1874	0.0000 [e]	-8.4	19.2	82.7 (c6h6-ch4)
Quantum ESPRESSO						
PBE-D2	49	–	–	13.2	18.4	69.7 (ch4-hf)
PBE-XDM	49	0.3275	2.7673	3.9	**13.7**	37.8 (h2s-h2s)
B86bPBE-XDM	49	0.6512	1.4633	2.6	**11.4**	23.1 (ch4-nh3)

[a] Number of molecular dimers contained in the parametrization set. [b] Mean percent error; a negative (positive) sign indicates underbinding (overbinding) with respect to the benchmark data. [c] Mean absolute percent error. [d] Maximum absolute percent error. Labels in parentheses identify the dimer that gives the maximum error. [e] In order to avoid unphysical (negative) values, a_2 was set to zero in the fit. [f] Counterpoise corrections were applied to the computed dimer binding energies. [g] Computed binding energies for the KB49 set of molecular dimers can be found in Appendix B.

The B86bPBE absolute errors in Table 6.1 are consistently lower than those from PBE, when paired with XDM. This has been observed in previous studies,[148,149] and is expected from the large-gradient-limit behaviour of the exchange enhancement factor in these functionals.[139,308,309] Using the largest basis set in this study (TZP), the MAPE of both functionals (12.3% for PBE and 11.2% for B86bPBE) are similar to those obtained using Quantum ESPRESSO (13.7% and 11.4%) and also to the near-complete-basis-set values using Gaussian basis sets reported in a previous study[149] (14.3% and 13.1% for the full KB49 set, with the latter value obtained using the psi4 program[310]). For comparison, PBE-D2 gives MAPEs of 33.5% (DZP), 26.5% (TZP), and 18.4% using the Quantum ESPRESSO implementation. However, it must be noted that PBE-D2 was not specifically fitted to the KB set.

The performance of B86bPBE-XDM and PBE-XDM suffers considerably from basis-set incompleteness when the smaller DZP basis set provided in the SIESTA package is used, with MAPEs of 21.7% (B86bPBE) and 30.0% (PBE), if the full KB49 set is considered. These values are consistent with the results for double-ζ Gaussian basis sets lacking diffuse functions, previously shown to be inadequate for non-covalent interactions.[149,311] For example, the MAPE of the PBE-XDM parametrization to the 49-dimer KB set with non-diffuse double-ζ Gaussian basis sets is 37.2% (6-31G$*$) and 45.9% (cc-pVDZ). In contrast, the MAPE of 6-31+G$*$, which contains one set of diffuse functions, is 17.8%. The DZP MAPE of 30.3% is intermediate between these two results, while the TZP basis set benefits from the increased cutoff radii compared to DZP. Other functionals, such as PBE-D2, are similarly affected by basis-set incompleteness, with MAPEs of 26.5% (TZP) and 33.5% (DZP). Therefore, it is clear that reliable calculation of non-covalent binding energies in SIESTA necessitates the use of larger basis sets than the standard DZP, or the design of new basis sets with increased cutoff radii.

The use of counterpoise (CP) corrections[189] to account for basis set superposition error (BSSE), in conjunction with the DZP or TZP basis sets, does not improve the performance of any method other than PBE-XDM/DZP. The CP correction tends to over-compensate, causing the dimers to be underbound on average, in agreement with previous reports.[312] The size of the CP correction indicates that there is still significant basis-set incompleteness error in the TZP results, probably stemming from the finite-support nature of the basis set.

6.3.2 Graphite and Phosphorene Exfoliation

The exfoliation[313] of graphite and phosphorene using the new XDM-corrected functionals in SIESTA is now considered. Graphite exfoliation is a simple test of the accuracy in the treatment of non-covalent interactions for which high-level experimental reference data exists.[314] For each method in Table 6.1, a scan was performed by systematically varying the interlayer distance between graphene sheets in graphite, while the intralayer hexagonal lattice parameter was kept fixed at 2.456 Å. The resulting potential energy curves are shown in Figure 6.2.

In all cases, the energy curves for the uncorrected functionals (PBE and B86bPBE) are very slightly binding, or non-binding, and a dispersion correction is necessary to obtain physically meaningful results. The equilibrium interlayer separations using DZP and TZP are slightly higher ($\approx$0.1–0.2 Å) than the plane-wave results and the experimental reference data. The exfoliation energies computed with SIESTA are in close agreement with both experiment and the Quantum ESPRESSO calculations, although slightly underestimated with the DZP basis set. Overall, the DFT-XDM methods perform slightly better at reproducing the experimental graphite interlayer distance and exfoliation energy than PBE-D2, regardless of the choice of local or plane-wave basis sets.

Counter-intuitively, while it is the larger basis set, TZP was found to to have increased BSSE for this system, as quantified by the greater counterpoise correction (see Figure 6.3). This is due to the more diffuse nature of the TZP basis for carbon, leading to greater orbital overlap. The BSSE leads to increased interlayer binding with TZP for all functionals, relative to the plane-wave results. In particular, uncorrected B86bPBE and PBE show between 10–20 meV/atom of spurious binding, while they give potential energy surfaces that are entirely repulsive with the plane-wave basis.

Phosphorene,[315] the single- (or few-) layer black phosphorus analogue of graphene, has attracted great interest recently due to its properties for device applications, which could be superior to those of graphene.[316–319] It has also been reported to be an "extremely challenging system from a computational point of view, given that its properties are regulated by a delicate equilibrium between dispersion forces and covalent interactions".[320]

Figure 6.2: Graphite exfoliation curves calculated with plane-waves / pseudopotentials in (a) Quantum ESPRESSO (QE), and in SIESTA using (b) DZP and (c) TZP basis sets, compared to experimental data.[314]

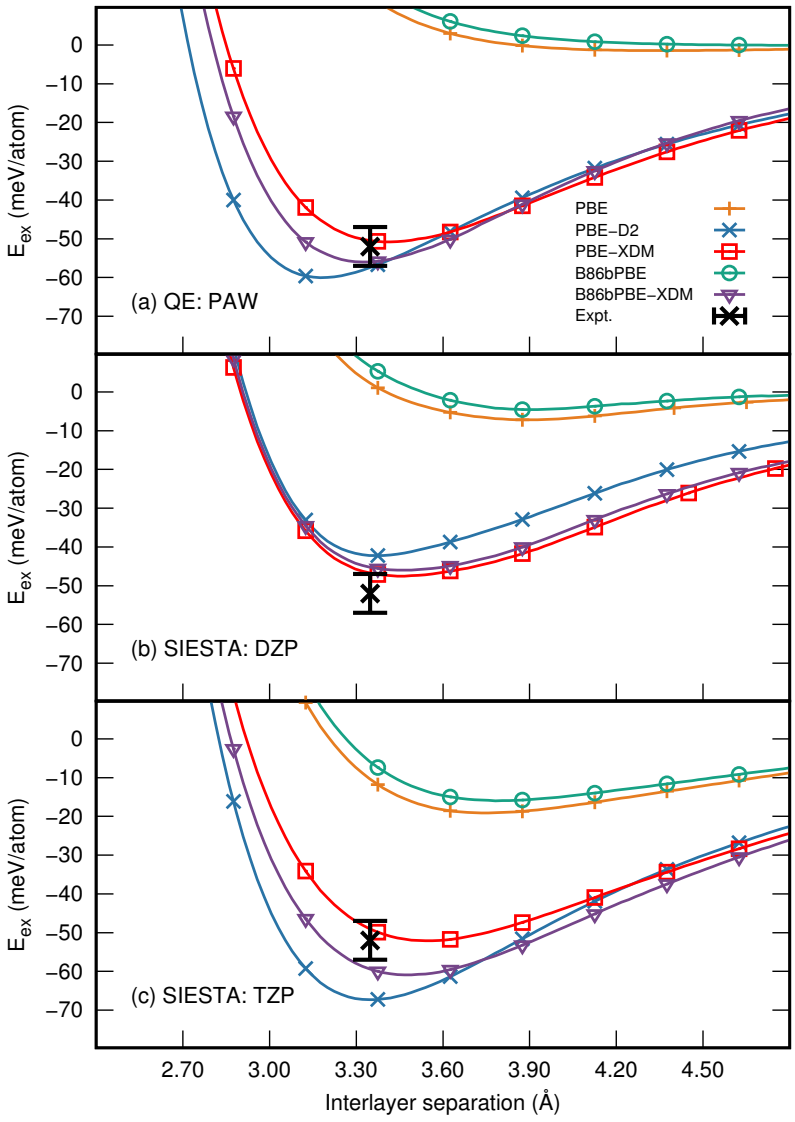
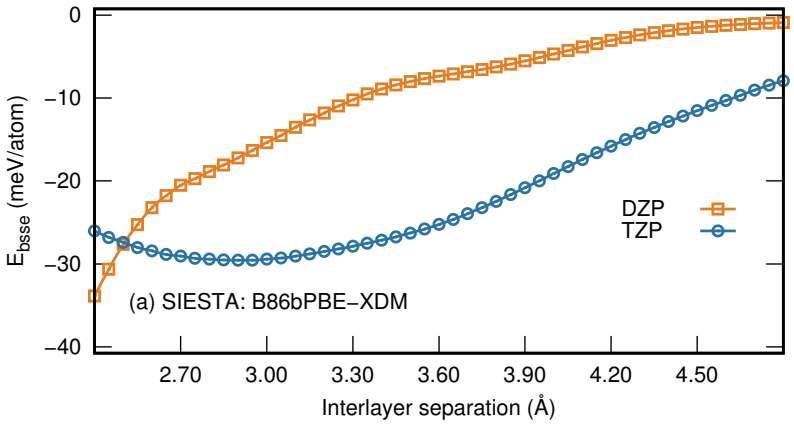


Figure 6.3: Counterpoise correction of single-layer graphene, E_{bsse}, in the presence of ghost orbitals centred around the atomic positions of a neighbouring layer of graphene. Here, E_{bsse} is defined as the difference in energy between the graphene sheet in the presence of neighbouring ghost orbitals with that of the graphene sheet in the absence of ghost orbitals. The TZP orbitals have greater overlap between two layers than do the DZP orbitals, and thus produce greater artificial stabilisation than the DZP orbitals for graphite.



While no experimental measurement for the mechanical exfoliation of phosphorene has been reported, many theoretical predictions have been made: quantum diffusion Monte Carlo (DMC, -81±6 meV/atom),[321] post-HF periodic correlated-wavefunction theory (p-LMP2 and pHF with CCSD(T) corrections, -151 meV/atom and -92 meV/atom, respectively),[322,323] along with several dispersion-corrected GGA DFT computations (falling in the ca. -80 to -150 meV/atom range).[320,322]

For each method in Table 6.1, a similar scan to what was done for graphite exfoliation was performed for the phosphorene analogue. The interlayer distance between sheets of black phosphorus (3.108Å at equilibrium) was systematically varied, while the intralayer orthorhombic lattice parameters were kept fixed at the experimental values of 3.314Å and 4.376Å.[324] The resulting potential energy curves are shown in Figure 6.4.

As Quantum Monte Carlo methods are often used as benchmarking tools,[325,326] the DMC result (-81±6 meV/atom) for the phosphorene exfoliation energy is used as a reference.[321] The curves depicted in Figure 6.4 clearly show that the equilibrium geometry and exfoliation energy are well captured by both the plane-wave and numerical-basis-set calculations, although, as seen for graphite, the exfoliation energy is somewhat underestimated with DZP. In this instance, a TZP basis set was not available for the phosphorus atom and was not considered.

Figure 6.4: Phosphorene exfoliation curves calculated with plane-waves / pseudopotentials in (a) Quantum ESPRESSO (QE), and (b) SIESTA using a DZP basis set, compared to DMC data.

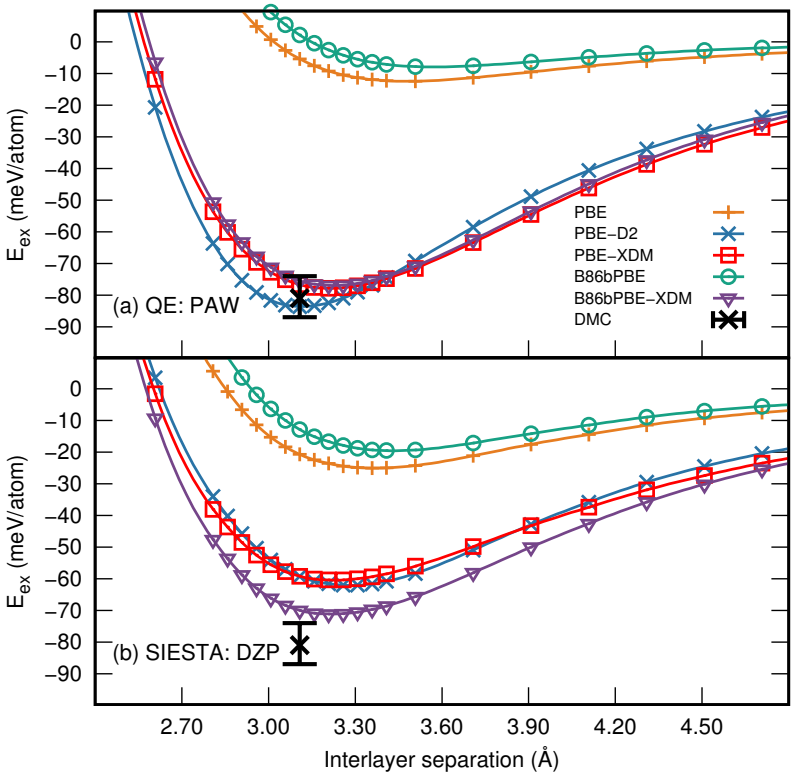


Table 6.2: Results of geometry optimizations on graphite performed with DFT-D/DZP using various input values of the c lattice parameter (specified in parentheses), as a function of the stress convergence threshold.[a]

Stress thr.	N_{opt}	c (6.0)	N_{opt}	c (7.0)	N_{opt}	c (8.0)
B86bPBE-XDM						
1.0	21	6.939	6	6.998	31	7.315
0.02	78	6.936	58	6.937	146	6.940
0.002	126	6.936	179	6.937	248	6.937
PBE-XDM						
1.0	21	6.988	6	6.999	30	7.069
0.02	47	6.996	23	6.998	163	6.995
0.002	309	6.976	291	6.977	314	6.991
PBE-D2						
1.0	13	6.871	2	6.994	19	6.820
0.02	19	6.796	46	6.797	62	6.798
0.002	66	6.795	105	6.795	137	6.795

[a] The units are GPa for the cell stresses, and Å for the c lattice parameter. N_{opt} is the number of optimization steps.

Before proceeding to the following section, there is an important point to note about geometry optimizations in weakly bound crystals using SIESTA. The default convergence thresholds for maximum atomic force and maximum stress components in SIESTA are 0.04 eV/Å and 1.0 GPa, respectively. While these are suitable for hard solids, they lead to unfinished geometry optimizations in molecular crystals, which makes the energy landscapes dependent on the choice of input geometries (*cf.* Chapter 5). This is illustrated by varying maximum stress component threshold for calculations on graphite (Table 6.2). The default convergence threshold (1.0 GPa) leads to incomplete geometry optimizations and erroneous equilibrium geometries. From these tests and the previous work in Chapter 5, maximum force and stress thresholds of 0.01 eV/Å and 0.02 GPa, which are tighter than the defaults, seem to be sufficient. These convergence thresholds will be used in the rest of the chapter. It is important to note that excessively tight stress thresholds are also problematic because of numerical instabilities that arise near the cutoff radii of the finite-support atomic orbitals when performing numerical integration.[130,131] This can lead to "endless" geometry optimizations, as pointed out by the SIESTA developers.

6.3.3 Lattice Energies of Molecular Crystals

The SIESTA implementation of XDM for molecular crystals is now assessed. The statistics for the X23 benchmark set are shown in Table 6.3. With the DZP basis set, B86bPBE-XDM performs best overall, yielding an MAE of 8.2 kJ/mol, followed by PBE-D2 and PBE-XDM. However, the use

Table 6.3: Statistics for the X23 set of lattice energies using DFT-XDM and DFT-D2 methods,[f] in kJ/mol per molecule, relative to back-corrected, experimental sublimation enthalpies.[142,145]

Method	ME[a]	MAE[b]	MAX[c]
SIESTA			
PBE-D2/DZP	7.5	11.4	25.1 (ant)
PBE-D2/TZP	3.9	5.7	14.5 (ada)
PBE-XDM/DZP	11.8	14.1	29.9 (suc)
PBE-XDM/TZP	-9.4	10.0	23.0 (cyt)
B86bPBE-XDM/DZP	3.1	8.2	24.3 (ant)
B86bPBE-XDM/TZP	-8.4	8.9	21.3 (cyt)
PBE-XDM/DZP+CP [d]	-0.8	4.4	14.4 (ant)
PBE-XDM/TZP+CP [d]	-2.1	4.4	18.2 (ant)
B86bPBE-XDM/DZP+CP [d]	-3.2	4.7	15.8 (ant)
B86bPBE-XDM/TZP+CP [d]	-3.6	4.7	17.5 (ant)
Quantum ESPRESSO			
PBE-D2	3.7	5.8	18.4 (ada)
PBE-XDM	-3.2	4.7	17.9 (cyt)
B86bPBE-XDM	0.5	3.6	13.4 (cyt)
Quantum ESPRESSO//SIESTA [e]			
PBE-D2/DZP	2.9	5.5	13.7 (cya)
PBE-XDM/DZP	-2.5	4.3	17.2 (cyt)
B86bPBE-XDM/DZP	-0.1	3.7	12.7 (cyt)

[a] Mean error; a negative (positive) ME indicates underbinding (overbinding) with respect to the benchmark data. [b] Mean absolute error. [c] Maximum absolute error; labels in parentheses identify the crystal responsible (ada: adamantane, ant: anthracene, cya: cyanamide, cyt: cytosine, suc: succinic acid). [d] Counterpoise corrections were applied to the crystal lattice energies, following the approach of Carter and Rohl.[282] [e] Composite methods using plane-wave single-point energies evaluated using the geometries obtained from SIESTA with the same functional and the DZP basis. [f] Computed lattice energies for the X23 set of molecular crystals can be found in Appendix B.

of the larger TZP basis set does not necessarily improve the quality of the results. Similar to the work of Carter and Rohl,[282] it is observed that the lattice-energy error statistics with finite-support basis sets are somewhat poorer than using plane waves. As an example, Carter and Rohl obtained a MAE of ca. 23 kJ/mol on the C21 subset of X23 (with 21 lattice energies[142]) using the non-local vdW-DF methods and a DZP basis set. For comparison, plane-wave calculations with the same functionals yield MAEs of ca. 10 kJ/mol.[142] Thus, for molecular crystals, B86bPBE-XDM/DZP outperforms the vdW-DF methods,[139] even when employed with plane waves, at a considerably reduced computational cost.

Carter and Rohl also demonstrated that the application of counterpoise corrections greatly reduced the errors in the computed lattice energies with SIESTA, restoring the performance of the vdW-DF functionals to the same quality as their plane-wave implementation.[282] Upon applying CP corrections, all of the DFT-XDM methods yield MAEs of 4.4–4.7 kJ/mol, approaching those of the plane-wave reference calculations. However, as argued previously,[282] although CP corrections significantly improve lattice energies and return values similar to plane-wave calculations, they are not straightforward to apply.

The impact of basis set incompleteness error on the lattice energies in Table 6.3 raises the question of whether the equilibrium geometries are similarly affected. This question is also important in the context of composite methods, in which high-level single-point calculations (using plane waves) are used at geometries obtained using a low-level calculation (SIESTA). These composite methods (*cf.* Chapter 5) were proposed as a computationally efficient alternative to plane waves for the calculation of lattice energies in molecular crystals,[282] but their performance relies on whether the low-level method offers equilibrium geometries close to those of the high-level method.

Table 6.4 evaluates the ability of DZP calculations with SIESTA to reproduce plane-wave equilibrium geometries, using the same density functional and dispersion correction. The plane-wave geometries are close to the complete-basis-set limit and, in this author's experience, the choice of PAW dataset / pseudopotential has very little effect on non-covalent equilibrium geometries, provided enough plane waves are used in the calculation. The powder similarity measure (POWDIFF) indicates that there are significant differences between the SIESTA and plane-wave geometries, and that this deviation is somewhat smaller for XDM than for D2. However, the lattice energies calculated using the corresponding composite methods are excellent, with MAEs being within at most 1.2 kJ/mol from the pure plane-wave calculations. Therefore, even though the lattice energies calculated using the DZP basis set are significantly affected by basis-set incompleteness error (Table 6.3), the corresponding equilibrium geometries seem to be reasonably close to the plane-wave reference and, consequently, the composite methods built with a DFT-XDM/DZP low-level approach are quite accurate and significantly cheaper than a pure plane-wave optimization.

Table 6.4: Statistics for the X23 set of lattice energies using composite methods,[e] in kJ/mol per molecule, relative to plane-wave calculations with the same DFA.

	SIESTA	Quantum ESPRESSO//SIESTA			
Method	MAE [a]	ME [b]	MAE [a]	MAX [c]	POWDIFF [d]
PBE-D2/DZP	10.8	-0.8	1.2	5.4 (ada)	0.2935
PBE-XDM/DZP	15.0	0.8	0.8	2.0 (pyr)	0.1578
B86bPBE-XDM/DZP	7.0	0.5	0.6	1.2 (eth)	0.1975

[a] Mean absolute error. [b] Mean error; a negative (positive) ME indicates underbinding (overbinding) with respect to the benchmark data. [c] Maximum absolute error; labels in parentheses identify the crystal responsible (ada: adamantane, eth: ethylcarbamate, pyr: pyrazole). [d] The deviation between the SIESTA and Quantum ESPRESSO equilibrium geometries is quantified by the powder diffraction similarity measure. [e] Computed binding energies for the X23 set of molecular crystals can be found in Appendix B.

6.4 Conclusions

In this chapter, the first implementation of the XDM dispersion model with the numerical finite-supported orbital method in the SIESTA software package was presented. The new XDM-enabled SIESTA code was verified by comparing the calculated dispersion coefficients to other XDM implementations using plane waves (Quantum ESPRESSO) and numerical orbitals in the gas phase (Numol). The XDM method in SIESTA was then parametrized against the KB49 set of gas-phase binding energies of molecular dimers for the PBE and B86bPBE functionals, with double-ζ (DZP) and triple-ζ (TZP) basis sets. DZP is the default basis set in SIESTA, while TZP has been recently formulated for C, H, O, and N by Louwerse and Rothenberg[283] and Carter and Rohl.[282] The performance of the new XDM-corrected methods for the molecular dimers is of similar quality to the results obtained using plane waves, provided that the TZP basis set is employed. DZP, on the other hand, suffers from significant basis-set incompleteness, and its performance is similar to double-ζ Gaussian basis sets without diffuse functions.

The new XDM implementation was tested for three cases: graphite and phosphorene exfoliation, and the calculation of lattice energies of molecular crystals. The XDM dispersion correction brings the potential energy curves for graphite and phosphorene exfoliation into close agreement with experimental data (or theoretical benchmark data, in the case of phosphorene). For the molecular-crystal lattice energies, the XDM-corrected SIESTA methods also give good results, although care needs to be taken in using/setting the proper stress convergence thresholds for geometry optimizations.

Both the DZP and TZP basis sets display considerable basis set incompleteness effects, causing the predicted lattice energies to be inferior to the plane-wave results unless counterpoise corrections are applied, which is undesirable in practice. This reveals the necessity of designing specialized SIESTA basis sets for non-covalent interactions. However, the equilibrium geometries obtained from all XDM-corrected methods in SIESTA are quite close to the plane-wave results. It was shown that composite approaches, in which single-point plane-wave calculations are performed at SIESTA's DZP equilibrium geometries, are quite accurate and computationally efficient.

B86bPBE-XDM/DZP showed particular promise as an excellent balance between accuracy and efficiency when determining lattice energies. The performance of this method is a significant improvement over the currently implemented post-SCF dispersion corrections (DFT-D2) and non-local functionals (vdW-DF1 and vdW-DF2) in SIESTA. While the applications examined in this work were focused on layered materials and organic molecular crystals, there are no reasons that would impede the use of the local-orbital implementation of XDM for computations on ionic or covalent bulk solids.

CHAPTER 7

CRYSTAL-ENERGY LANDSCAPES OF ACTIVE PHARMACEUTICAL INGREDIENTS USING COMPOSITE APPROACHES

Contributions to the manuscript LML performed the calculations, and wrote the first draft of the manuscript. ERJ contributed to the final version of the manuscript, and supervised the project.

7.1 Motivation

Stringent limitations on the choice of computational methods[24] for CSP are imposed by the increasing demand to study larger and more complex systems, with applications to predict the stability and properties of active pharmaceutical ingredients[27,327] and to the design of new materials.[328,329] If executed properly, one can benefit from using multilevel methodologies to perform CSP studies more efficiently. In this fashion, one can gradually increase the accuracy of the computational method used, while concomitantly reducing the initial pool of candidate structures to more manageable numbers, to hone in on the low-energy parts of the landscape. While density-functional tight-binding[97] or small-basis-set (e.g., HF-3c)[90,273] approaches are popular low-cost methods for CSP purposes, their energetic accuracy for the intermediate stages of a search is often suboptimal.[330] This requires a larger subset of structures to be brought forward in the CSP protocol[331] to avoid experimentally isolable structures being thrown out during the refinement steps.[24] However, low-cost methods can yield reasonably accurate geometries,[330,332,333] particularly if the dispersion correction parameters are fitted not only to energies, but also to reproduce benchmark geometries

of molecular dimers[331,332] and/or geometries of molecular crystals.[273,333]

In Chapter 5, the use of composite approaches with small-basis-set methods (HF-3c and PBE-D2/DZP), followed by high-level plane-wave DFT single-point energy calculations (e.g., B86bPBE-XDM) to examine and assess their ability to provide absolute and relative lattice energies of small molecular crystals. These composite approaches were then applied to the previously reported CSP for a chiral organic semiconductor, 1-aza-[6]-helicene, and were found to reduce the computational cost of generating accurate electronic-energy landscapes, relative to full plane-wave DFT calculations. These composite methods worked best when geometries were transferable between the low- and high-level methods. This prompted the implementation of B86bPBE-XDM in the numerical-basis SIESTA code,[130,131] to allow use of the same functional for both low- and high-level calculations (Chapter 6).

Others have also recently applied similar ideologies to perform CSP studies. In the work by Iuzzolino and co-workers,[331] a re-parametrized version of the D3-dispersion-corrected DFTB method allowed for low-cost optimization of several thousands of candidate structures generated during CSP studies of drug-like molecules (compounds **XX**, **XXIII**, and **XXVI** of the most recent CCDC blind tests *viz.* Figure 1.3,[24,42] two tautomers of mebendezole, and a lead compound in a pharmaceutical study). While DFTB did not improve on the initial ranking from the CrystalPredictor-generated structures, the geometries were amenable to higher-level single-point calculations using a multipole-based atom-atom force field. These single-point energy calculations aided in placing the experimentally isolated polymorphs in the low-energy regions of the CSP landscapes. The success of DFTB in this instance was a consequence of refitting the dispersion damping parameters to minimize errors in geometries of small molecular dimers or molecular crystals within the X23 set,[142,145] rather than minimizing energies.[331–333] It would seem that utilising DFTB with newly fitted dispersion parameters would be the best choice for a low-cost method in a composite approach, given the speed-up it can offer in pre-optimizing geometries. However, several challenges to its use in CSP remain. For instance, in the study by Iuzzolino and co-workers,[331] DFTB was unable to provide an improved energy landscape relative to the force fields used in crystal structure generation. While DFTB improved the agreement of computed crystal geometries with experimental X-ray data, it generally caused the experimentally isolated polymorphs to be destabilized significantly with respect to other candidate structures on the energy landscape, meaning that a larger number of structures would need to be considered using the composite approach. It should be noted that the DFTB method also altered the covalent bonding in

some situations.

In another study,[334] the use of the sHF-3c method as a cheap alternative to full DFT geometry optimizations allowed the unit-cell volumes of zeolites to be reproduced within 2% of experimental values. Additional high-level dispersion-corrected DFT energy calculations were then used to improve upon the sHF-3c results, reducing the errors in formation enthalpies from ca. 5.9 kJ/mol to 0.8 kJ/mol per silica unit for the quartz-based zeolites. Other fragment-based approaches using a combination of low- and high-cost methods to efficiently compute lattice energies of the small molecular crystals of the X23 set,[330] or applying machine-learning methods to correct energies of crystal structures on landscapes generated via anisotropic atom-atom force-field methods,[335] have been developed. In the latter study, two-body interaction energies within a fragment-based method initially computed by force fields were corrected with high-level methods (such as DFT or MP2) for a series of polymorphic structures of small organic molecules (oxalic acid, maleic hydrazide, and tetrolic acid, to name a few). However, the large number of dimer calculations and corrections applied with high-level methods prompted the authors to develop a machine-learned variant, capable of reproducing the high-level corrections much more cheaply, and improving the accuracy of the relative energies of these polymorphs on generated crystal-energy landscapes.

In this chapter, the robustness of composite approaches utilizing small-basis set methods (in particular, B86bPBE-XDM/DZP and sHF-3c) in combination with plane-wave DFT calculations (B86bPBE-XDM/PAW) to efficiently produce accurate energy landscapes is validated for a selection of active pharmaceutical ingredients (APIs): 5-fluorouracil, naproxen, carbamazepine, and olanzapine. Each of these have several experimentally characterized polymorphs and have been the subject of previous CSP studies.[336–339] The work presented herein focuses solely on energy-ranking methods, and not on the generation of candidate structures. The initial sets of structures were obtained from the Control and Prediction of the Organic Solid State (CPOSS) database,[339] courtesy of Prof. Sally L. Price and Dr. Louise S. Price, University College London. The results in this chapter show that the use of composite approaches can lead to the generation of more realistic energy landscapes for pharmaceutical-like molecules, in contrast to previous force-field calculations. The experimentally isolated polymorphs of the drugs are generally placed as the lowest in energy and the proper experimental stability ordering between polymorphs is well described (within the tolerance for thermal effects). Ultimately, these composite approaches could also serve as a means to identify the limitations of (semi-)empirical methods, by further examining the candidate structures which are most affected by re-ranking.

7.2 Computational Methods

CPOSS database structures: Three of the four APIs selected, 5-fluorouracil,[336] naproxen,[337] and olanzapine,[338] have had their crystal-energy landscapes published, while the data for carbamazepine is currently unpublished. All of the force-field energy rankings were obtained from empirically fitted dispersion-repulsion potentials, supplemented with distributed multipole analysis (DMA) for the description of electrostatics, derived from post-Hartree-Fock (e.g., MP2) wavefunction or DFT methods.[91] Details can be found in each of the references pertaining to the initial CSP studies of the APIs. Briefly, (i) 5-fluorouracil:[336] crystal structures were generated with MOLPAK, using MP2/6-31G(d,p) optimized molecular structures; lattice energy calculations were performed with DMAREL keeping the molecules completely rigid throughout. (ii) Naproxen:[337] a rigid molecule search in MOLPAK was conducted, and further refinements were performed with DMAflex-2, which allows minimization with some molecular flexibility, i.e., alternating between geometry optimizations and charge density calculations with GAUSSIAN and lattice-energy minimizations with DMACRYS. (iii) Olanzapine:[338] a flexible molecule search was performed with CrystalPredictor-I, then lattice energies were computed and refined with CrystalOptimizer using a DMA-refined force-field. (iv) Carbamazepine:[339] a similar procedure to what was used for olanzapine was performed, and is detailed in Ref. 340.

The low-cost and high-level methods used for the composite approaches taken in this work are now detailed.

SIESTA calculations: B86bPBE[109,110]-XDM,[139] with the DZP basis set, as implemented in an in-house version of the 4.0b-485 SIESTA code, was used for the low-cost calculations. Similar parameters to what were used in Chapters 5 and 6 are also used here. The "energy shift" parameter controlling the the confinement radius of the orbitals was set to 0.001 Ry, the real-space integration-grid cutoff value was set to 400 Ry, and Trouillier-Martins-type[121,122] norm-conserving pseudopotentials[120,290] generated with the ATOM code[132] were utilized. The XDM damping parameters, a_1 and a_2, were set to 0.5000 and 2.5556 Å, respectively. The Broyden optimizer was used for geometry relaxations, and convergence thresholds for forces and stresses were set to 0.04 eV/Å and 0.02 GPa, respectively.

CRYSTAL calculations: sHF-3c,[96,192,273] as implemented in the CRYSTAL17[271] code, was also

used for the low-cost calculations. sHF-3c makes use of a minimal basis set (MINI(x)[96,192]) and is supplemented with Grimme-D3 dispersion,[166,167] a geometrical counterpoise correction,[190,191] and a correction for short-range basis-set incompleteness errors.[96,192,273] Becke-Johnson damping was used for the D3 dispersion model,[161,167] three-body interactions were included, and a fit parameter of s_8 = 0.6143 was applied.[273]

Plane-wave calculations: High-level plane-wave B86bPBE-XDM calculations using PAW data sets[124] were performed with the Quantum ESPRESSO[127] code, version 5.1. For single-point energy calculations, energy cutoffs for the wavefunction and density were set to 60 Ry and 600 Ry, respectively. For the full geometry relaxations of 5-fluorouracil, these two parameters were set to 80 Ry and 800 Ry, respectively. The XDM damping parameters, a_1 and a_2, were set to 0.6512, and 1.4633 Å, respectively. When structures were fully relaxed (e.g., in the case of 5-fluorouracil), convergence thresholds for energies and forces of 10^{-5} Ry and 10^{-4} Ry/bohr, respectively, were used.

k-point grid sampling: A $4 \times 4 \times 4$ MP k-point mesh sampling of the Brillouin zone was used for all calculations.

Similarity metric between crystal-energy landscapes: In order to compare the degree of similarity between crystal-energy landscapes a similarity index, S, is indicated on each of the panels of the figures presented in this chapter. The value of this index is given by

$$S = \frac{1}{N} \sum_{i=1}^{N} ||\Delta E_{\text{low}} - \Delta E_{\text{high}}||_2, \tag{7.1}$$

where the sum runs over the N data points (ca. 50-60) present on a crystal energy landscape, and $||x||_2$ is the Euclidean l^2-norm. Overall, the index represents a mean deviation between ΔE's on two different energy landscapes, from low and high levels of theory, with the high level serving as the reference landscape. ΔE_{low} and ΔE_{high} are the energy differences between a given data point on the landscape and the centre-of-mass energy of all the points on this same landscape. The centre-of-mass energy is chosen to eliminate dependence on an arbitrary choice of reference point. The larger this index, the more dissimilar the points are with respect to their corresponding points on the reference landscape. A value of zero indicates an exact match.

Relative energies and densities of crystal structures from the generated energy landscapes for 5-fluorouracil, naproxen, carbamazepine, and olanzapine can be found in Appendix C.

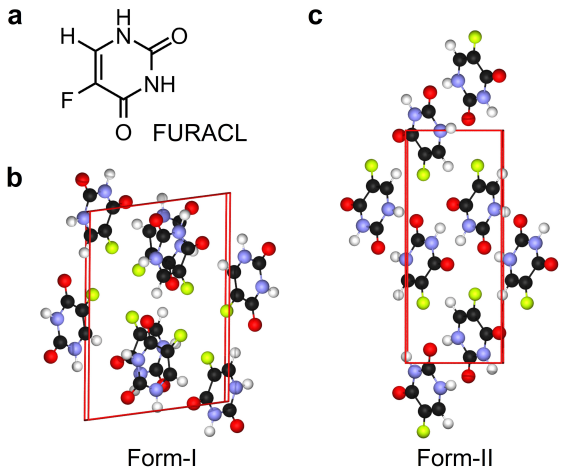
7.3 Results and Discussion

7.3.1 5-Fluorouracil

5-Fluorouracil, first synthesized in 1957 and used as an anti-cancer API for over forty years,[341,342] forms two known polymorphs (Figure 7.1) to date. Its first crystal structure (Form-I) was solved in 1973,[343] but a second form (Form-II) was predicted by computational means and subsequently isolated thirty years later.[336] From this initial CSP study,[336] Form-II was predicted via force-field computations to be the thermodynamic minimum, being ca. 6 kJ/mol more stable than Form-I at 0 K. However, experimental evidence from thermal analysis measurements, reported in that same study, pointed to the opposite stability ordering and supported a monotropic relationship between the two forms. That is, Form-I was found to be more stable than Form-II, and no phase transitions were observed upon differential scanning calorimetry measurements between 298 K and the melting point of each form. The fitted force-field potential was able to reproduce the geometries of crystal structures at room temperature within reasonable limits, given that thermal expansion was neglected. However, the neglect of thermal expansion was not deemed to be the main reason for the disagreement between the experimental and computational results, as the quality of the computed lattice energies did not improve upon comparing to low-temperature (150 K) experimental crystal structures. Instead, it was suggested that the force field would not necessarily describe the relative lattice energies of these two forms of 5-fluorouracil properly, because it was fitted for geometries using a set of molecules containing only a small number of perfluorohydrocarbons.

The prediction of a stable Form-II of 5-fluorouracil prompted several investigations into why only Form-I had been isolable for several decades past. In particular, as this second form could only be prepared in dry nitromethane, it was postulated that the solvation of 5-fluorouracil could affect the mechanisms of nucleation and crystal growth of both forms.[336] Subsequent molecular dynamics simulations showed that the presence of water in organic solvents, or of water as a solvent itself, could affect which polymorph of 5-fluorouracil would be kinetically favoured.[344,345] Ultimately, the hydrogen bonding of water molecules to the C=O and N–H sites leaves the 5-fluorouracil molecules open to self-associate via F–F interactions only (as in Form-I, Figure 7.1(b)). Conversely, in the absence of water, more weakly-interacting solvents (such as nitromethane) do not disallow formation of hydrogen-bonded dimers between the parent 5-fluorouracil molecules, which translates to yield Form-II (Figure 7.1(c)) in the solid state.

Figure 7.1: 5-Fluorouracil (CSD[22] refcode FURACL) and its two known polymorphs. Form-I crystallises in the $P-1$ space group and contains 4 molecules in the asymmetric unit, which adopt a hydrogen-bonded sheet structure with four fluorine atoms in close proximity. Form-II, on the other hand, crystallizes in the $P2_1/c$ space group, displaying a ribbon motif of molecules, each forming two hydrogen-bonds with its neighbours.[336] Form-I is the more stable experimentally at 298 K.

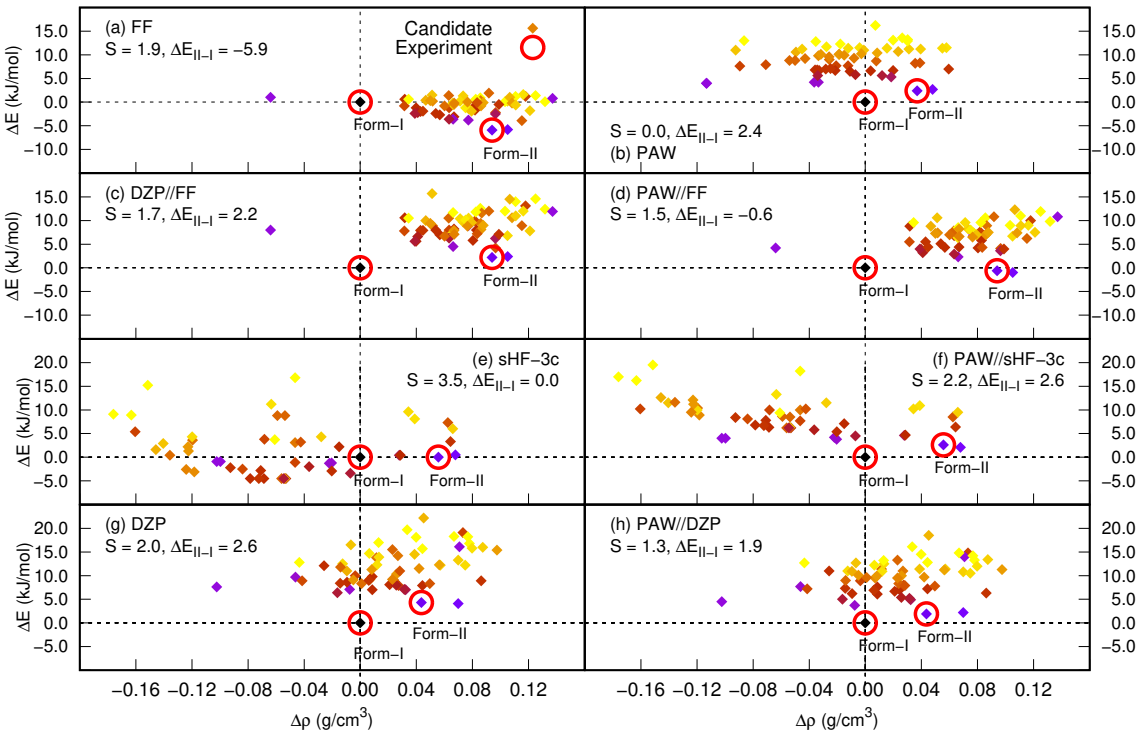


The computational evidence for kinetically-controlled pre-nucleation of 5-fluorouracil has offered a more complete understanding of crystallisation of 5-substituted uracils.[346] Yet, despite additional efforts in establishing free-energy differences between the two polymorphs, and accounting for thermal expansion through the use of classical molecular-dynamics computations, Form-II has continually been predicted to be more stable than Form-I.[345,347] This remains in disagreement with the experimental evidence reported when Form-II was first isolated/solved.[336]

The force-field ranking results reported in Ref. 336 for the crystal-energy landscape are presented in Figure 7.2(a). Figure 7.2(b), on the other hand, shows the re-ranking of crystal structures with fully-relaxed geometries obtained from plane-wave DFT calculations (B86bPBE-XDM). In contrast to the force-field results, Form-I is now predicted to be more stable than Form-II by 2.4 kJ/mol, which (while these results neglect thermal/entropic effects) recovers the same relative stability of the crystals forms as experiment.[336] A possible reason that the force fields utilized in earlier studies predict the incorrect relative energy ordering of Forms I and II could be an inability to describe halogen–halogen interactions,[348–350] which impact the stability of Form-I of 5-fluorouracil.[351]

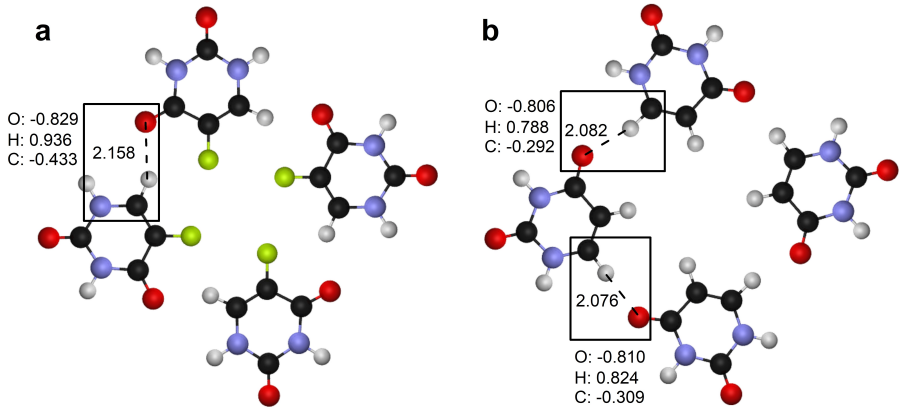
Gas-phase calculations on the tetrameric arrangement of molecules present in Form-I of 5-fluorouracil, and the equivalent arrangement for uracil molecules (Figure 7.3), indicate a binding

Figure 7.2: Crystal-energy landscapes of 5-fluorouracil. Energy landscapes were generated via (a) a force-field (FF, data obtained from the CPOSS database[336,339]); plane-wave DFT calculations, (b) PAW: B86bPBE-XDM, and two small- and/or minimal-basis low-cost methods, (e) sHF-3c/MINI(x), and (g) DZP: B86bPBE-XDM/DZP. Crystal structures within panels (a,e,g) were then re-ranked with single-point energy calculations using plane-wave B86bPBE-XDM to give panels (d,f,h), respectively. The force-field structures within panel (a) have also been re-ranked with single-point energy calculations using the small-basis B86bPBE-XDM/DZP method to give panel (c). The computed energies, ΔE, and densities, $\Delta\rho$, are expressed relative to Form-I for all methods. The similarity index, S, is relative to the crystal-energy landscape in panel (b). $\Delta E_{\mathrm{II-I}}$ is the energy difference between the two isolable polymorphs, relative to Form-I.



energy per molecule of -35.4 and -29.1 kJ/mol for these two systems, respectively. The majority of this interaction is due to the base functional used, with the XDM dispersion correction representing only -7.2 and -5.2 kJ/mol per molecule, respectively. It is rather unlikely that the ca. 6 kJ/mol difference in binding energy between the 5-fluorouracil and uracil systems is due to halogen bonding between the fluorine atoms, given that these interactions are weakest for fluorine,[352] and that the interatomic distance between fluorines (ca. 2.94 Å) is at the edge of what is typically observed for fluorine–fluorine halogen bonding (ca. 2.95 Å).[351] Instead, the larger binding energy observed in 5-fluorouracil is likely due to the C-H bonds being more acidic, due to the fluorine atom withdrawing electron density, which results in stronger C–H$\cdots$O interactions between 5-fluorouracil molecules.

Figure 7.3: Tetramers of (a) 5-fluorouracil and (b) uracil. BHandHLYP[111]-XDM[155]/aug-cc-pVTZ//BHandHLYP-XDM/aug-cc-pVDZ gas-phase geometry optimizations of the tetramers of 5-fluorouracil (found in the Form-I polymorph) and uracil were performed with GAUSSIAN 09[353] software package and the postg code.[354] Bond lengths (in Å), indicated by dashed lines, and Mulliken charges (in atomic units), indicated for the boxed atoms, are shown.



Having shown that plane-wave DFT calculations can afford a crystal-energy landscape for 5-fluorouracil that agrees with available experimental findings, focus is now shifted to whether composite approaches can reproduce similar-quality landscapes. As has been shown in earlier chapters, it is important to choose low-cost methods that will produce geometries of similar quality to the higher-level methods (here, plane-wave DFT), so that the application of a single-point energy calculation on the "cheaply" optimized crystal structures yields reliable relative energies. In Chapter 5, the best low-cost method for use in composite approaches with plane-wave B86bPBE-XDM methods were sHF-3c and PBE-D2/DZP. This prompted the implementation of B86bPBE-XDM in SIESTA (Chapter 6). It is expected that these geometries will generally be more compatible with plane-wave B86bPBE-XDM, given the improved energetics obtained over PBE-D2 with DZP and because there will be consistency in the choice of functional and dispersion correction between low- and high-level calculations. The small-basis B86bPBE-XDM method alone (Figure 7.2(g)) affords a more realistic energy landscape than that offered by the force-field calculations (Figure 7.2(a)), given that both experimentally observed crystal forms fall within the lower-energy regions. However, this is not the case for sHF-3c (Figure 7.2(e)), which gives many lower-energy structures than the experimentally isolated forms at larger cell volumes. An improved description of the crystal-energy landscapes is obtained by performing single-point energy calculations with B86bPBE-XDM using a plane-wave basis set, as shown in Figure 7.2(d,f,h). In all three cases, performing these single-point energy calculations restores energy differences between the two

isolable crystal forms in very close agreement to the full plane-wave DFT results.

The overall agreement between energy landscapes can be observed qualitatively by taking into account the heat map, where each data point in Figure 7.2(a)-(h) has its colour fixed to that of the corresponding point for the full plane-wave results in Figure 7.2(b). Alternatively, a distance metric is also presented for each panel to give a more quantitative descriptor of similarity between the low-level, or composite, and full plane-wave crystal-energy landscapes. Interestingly, it is clear from this distance metric that the force-field energy ranking alone yields a similar-quality landscape to that of plane-wave DFT calculations, given that the relative energies are only 1.9 kJ/mol away from the reference landscape on average. This is on par with the average deviation computed for the landscape generated with the small-basis B86bPBE-XDM method (2.0 kJ/mol). The major issue encountered with the force-field ranking is clearly its inability to describe the proper energies for systems where close halogen–halogen contacts are important (such as in Form-I).[336] The sHF-3c landscape is comparatively of poorer quality than the other two methods, with a distance metric from the reference landscape of 3.5 kJ/mol. However, application of DFT single-point energy calculations, in general, does reduce deviations from the reference plane-wave DFT landscape.

Overall, although more computationally expensive than the force field and sHF-3c methods, B86bPBE-XDM/DZP gives geometries that seem to be most compatible with plane-wave B86bPBE-XDM, making it more amenable to be combined with this particular high-level method in a composite approach. That being said, in order to remain competitive with other approaches making use of the faster DFTB or sHF-3c methods, one could instead resort to pre-screening, by performing single-point energy calculations with B86bPBE-XDM/DZP on the force-field geometries (Figure 7.2(c)). The force-field geometries in this case are amenable to a composite approach with the small-basis DZP method, and reproduce an energy landscape of similar quality to the one where full DZP optimizations were performed (Figure 7.2(g)), with Form-I again predicted to be more stable than Form-II. Then, a subset of low-energy structures from the composite DZP//FF energy landscape (Figure 7.2(c)) could be subjected to full optimization with DZP (Figure 7.2(g)) and subsequent single-point energy calculations with plane-wave basis sets (Figure 7.2(h)).

Because the B86bPBE-XDM composite approach, making use of DZP geometries followed by plane-wave single-point energies, has afforded the best results so far (including the study of small molecular solids, Chapter 6), this method will be used as a reference for establishing energy landscapes for the following APIs: naproxen, olanzapine and carbamazepine. Performing full

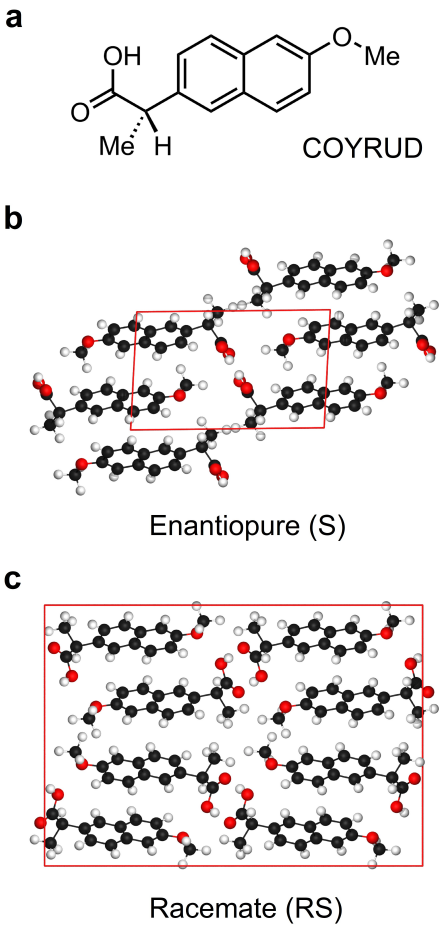
plane-wave DFT calculations for these molecules' crystal-energy landscapes would be extremely costly, as their unit cells contain several hundred atoms. Thus, it is now shown that physically reasonable energy landscapes, agreeing with available experimental evidence, can be obtained via composite approaches, without having to rely on high-level geometry optimizations.

7.3.2 Naproxen

The (S)-enantiomer of naproxen is marketed as a non-steroidal anti-inflammatory drug. Its enantiopure crystal structure was first reported in 1985,[355] and refined two years later (Figure 7.4).[356] In 2011, the existence of a racemate crystal structure for naproxen was predicted by CSP and experimentally isolated.[337] The racemate and enantiopure crystal structures were ultimately found to have similar stabilities: (i) differential scanning calorimetry measurements determined the racemate form to have a similar melting point (within $< 1^{\circ}$C), while having a higher heat of fusion than the enantiopure form (by ca. 1.5 kJ/mol), and (ii) solubility measurements determined the relative heats of solution to be in favour of the racemate by ca. 2.4 kJ/mol. Force-field computations of the relative lattice energies estimated the racemate form as being 6.1 to 9.2 kJ/mol more stable than the enantiopure crystal form, depending on the quality of the method used to describe intra- and intermolecular forces. This overestimation of the relative energy difference has been attributed to neglect of thermal expansion and entropic contributions.[337]

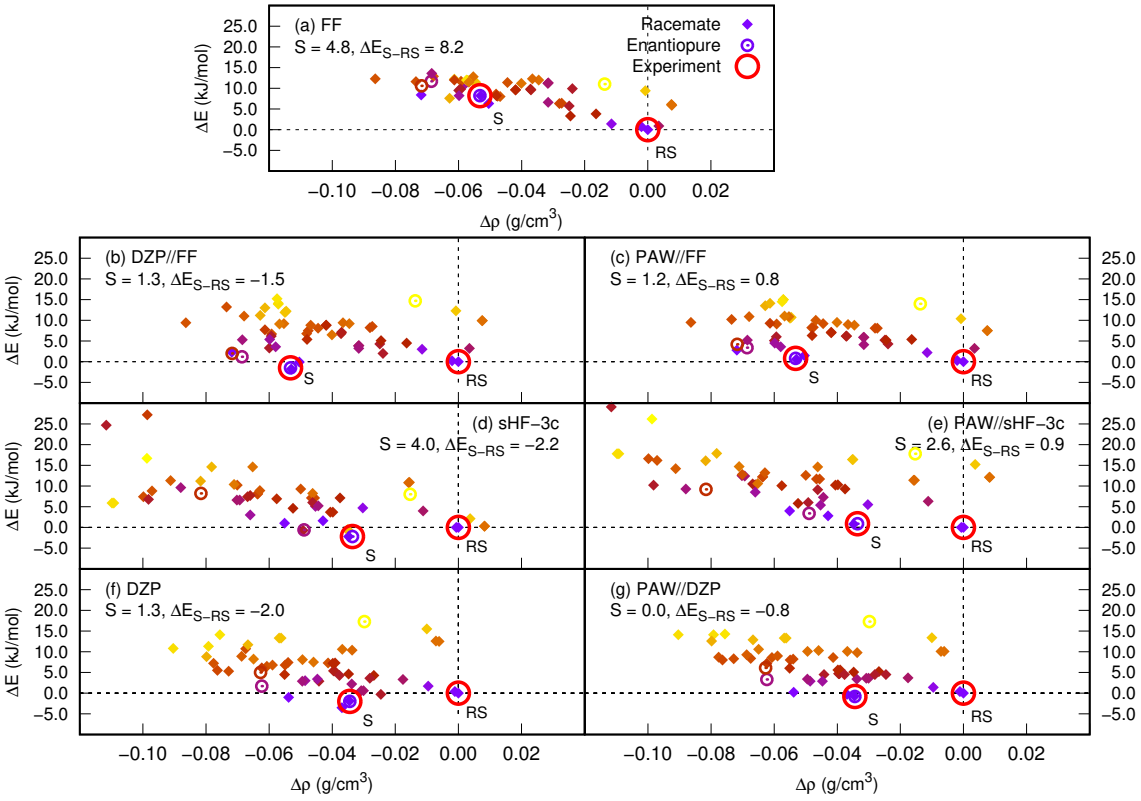
The crystal-energy landscape obtained from the previous force-field computations, reported in Ref. 337 and reproduced Figure 7.5(a), predicts both the experimental racemate and enantiopure crystal forms as the overall minima of all candidate racemic and enantiopure crystal structures, respectively. An even more dramatic improvement in the quality of the energy landscape to what was observed for 5-fluorouracil is observed for naproxen when only B86bPBE-XDM/DZP single-point energy calculations, instead of full geometry optimizations, are performed on force-field geometries (Figure 7.5(b)). This indicates that an initial composite DZP//FF approach could be taken on a larger number of structures prior to moving forward only a subset of low-energy structures to the PAW//DZP composite approach. Making use of a composite approach with plane-wave DFT using the force-field geometries (Figure 7.5(c)) also gives the enantiopure and racemate crystal structures as the lowest-energy structures and maintains the correct ordering. While the energy difference between these two minima is overestimated with the force field,[337] the composite approach brings both forms into closer proximity (ca. 0.8 kJ/mol apart), in good agreement with the experimental measurements.

Figure 7.4: Naproxen (CSD[22] refcode COYRUD) and its known enantiopure and racemate forms. The two forms crystallise in the $P2_1$ and $Pbca$ space groups, respectively. Hydrogen-bonded interactions occur between the carboxylic acid moeities, related by 2_1 symmetry in the enantiopure form, and by inversion between the two enantiomers in the racemate form. The racemate form is the more stable experimentally.[337]



The energy landscape generated solely from the low-cost B86bPBE-XDM/DZP method (Figure 7.5(f)) produces a similar-quality energy landscape to the reference (Figure 7.5(g)), showing that relative energies need not be corrected for basis-set superposition and or incompleteness errors (e.g., with counterpoise corrections), which are typically needed to compute accurate absolute lattice energies (see Chapters 5 and 6). The sHF-3c calculations (Figure 7.5(d)), again do not yield as good results as the B86bPBE-XDM/DZP method (Figure 7.5(f)), when compared to the reference landscape, as quantified by the larger S value (*viz.* S = 4.0 vs. 1.3 kJ/mol). Applying the composite approach on force-field or sHF-3c geometries leads to the racemate structure being slightly more favoured than the enantiopure (Figures 7.5(c.e), respectively), in agreement with experiment,

Figure 7.5: Crystal-energy landscapes of naproxen. Energy landscapes were generated via (a) a force-field (FF, data obtained from the CPOSS database[337,339]); and two small- and/or minimal-basis low-cost methods, (d) sHF-3c/MINI(x), and (f) DZP: B86bPBE-XDM/DZP. Crystal structures within panels (a,d,f) were then re-ranked with single-point energy calculations using plane-wave B86bPBE-XDM to give panels (c,e,g), respectively. The force-field structures within panel (a) have also been re-ranked with single-point energy calculations using the small-basis B86bPBE-XDM/DZP method to give panel (b). The computed energies, ΔE, and densities, $\Delta\rho$, are expressed relative to the racemate form, "RS", for all methods. "S" is the enantiopure form. The similarity index, S, is relative to the crystal-energy landscape in panel (g). $\Delta E_{\mathrm{S-RS}}$ is the energy difference between the two isolable crystalline forms, relative to the racemate form.



whereas with the B86bPBE-XDM/DZP geometries, the reverse is true, and thus is no longer in agreement with experiment. However, the enantiopure structure is predicted to be more stable by only ca. 1 kJ/mol, so thermal effects could potentially reverse the stability ordering of the two forms.

7.3.3 Carbamazepine

Carbamazepine, a drug used in the treatment of epilepsy and trigeminal neuralgia, has five known polymorphs,[357–359] as depicted in Figure 7.6. The most stable form at room temperature is Form-III,[358] and both Forms I and II are related enantiotropically to it at higher temperatures, meaning that they become more stable than Form-III at a temperature higher than 298 K.[357] The stability of Form-IV has been reported to lie between that of Forms I and II.[360] Forms II-IV have been shown to have an enantiotropic relationship, and display conversion to Form-I at high temperatures, prior to the melting point of Form-I.[361] Nevertheless, this means that, at 0 K, one would expect Form-III to be the most stable polymorph. Several computational studies[362–364] are generally in qualitative agreement with experimental findings,[357,358,360] giving the following stability ordering of the first four known polymorphs: III $<$ I,IV $<$ II, with the latter three forms being separated by no more than ca. 3-4 kJ/mol. Force-field lattice-energy calculations have stipulated Form-V to be of similar stability to the other carbamazepine polymorphs, within ca. 2 kJ/mol from Forms I and IV,[359,365] but no experimental measurement of its stability relative to the other forms has been reported to date. Presumably, this is because of its later prediction, and eventual isolation, than the first four forms.[359,365]

The relative lattice-energy rankings derived via force-field computations reported in the CPOSS database[339] are presented in Figure 7.7(a), together with the data obtained herein using the sHF-3c (Figure 7.7(d)) and B86bPBE-XDM/DZP (Figure 7.7(f)) low-cost methods. Applying plane-wave DFT single-point energy calculations to all of these geometries yields the energy landscapes depicted in Figure 7.6(c,e,g), respectively, while the DZP//FF results are depicted in Figure 7.6(b). The force-field computations again place many of the isolable polymorphs as high-energy structures, whereas re-ranking from the plane-wave DFT calculations place them all, except for Form-II, in the lower-energy regions of the crystal-energy landscape. The small-basis B86bPBE-XDM/DZP single-point energy calculations perform less well than the plane-wave calculations, but are significantly cheaper and, in addition, generate an energy landscape on par with the fully optimized DZP case, Figure 7.7(f). Similar results as seen for 5-fluorouracil and naproxen are obtained for sHF-3c, showing poorer performance than B86bPBE-XDM/DZP. Re-ranking via the composite approach from the fully-optimized B86bPBE-XDM/DZP geometries, however, affords energy landscapes giving the same stability ordering as previously described by other force-field calculations[362–364] and, more importantly, through experiment.[357,358,360] While the low-cost methods disagree on the relative stability of Form-V with respect to Forms IV and II, results from the composite approaches all suggest that it should be less stable than Form-II only. The overall advantage of using composite approaches remains that most isolable polymorphs observed through experiments are predicted to

Table 7.1: Relative energies, in kJ/mol, for the experimental polymorphs of carbamazepine computed with low-level and composite approaches. Energies are expressed relative to the most stable form of carbamazepine at 0 K, Form-III. B//A: corresponds to the use of the low-level and composite approach, where A = the lower-level method indicated in the table header, B = the higher-level B86bPBE-XDM/DZP and B86bPBE-XDM/PAW methods (DZP and PAW, respectively). The established experimental ordering is III $<$ I,IV $<$ II.

	FF			sHF-3c		DZP	
	A	DZP//A	PAW//A	A	PAW//A	A	PAW//A
Form-III	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Form-I	5.0	-0.2	3.9	6.2	5.8	1.6	5.1
Form-IV	3.8	4.5	3.5	8.6	6.2	5.8	4.2
Form-V	3.0	-2.0	4.6	12.9	8.3	2.2	8.1
Form-II	8.6	4.0	7.6	9.4	9.4	6.0	9.3

be thermodynamic minima, instead of being of similar relative energy as close to 50 other candidate structures.

7.3.4 Olanzapine

Olanzapine, an anti-psychotic drug used in the treatment of schizophrenia,[366,367] currently has three characterized anhydrous polymorphic forms, one other uncharacterised form, and many isolated hydrates and solvates.[338,368] The anhydrous solid-state forms have been inconsistently reported as Forms I-IV in the literature. The two first forms that were isolated have since been found to be the same and are now both labelled as Form-I.[369] The third form (Form-III) remains uncharacterised and not isolable as a pure crystal.[370] The fourth form has been isolated as a second distinct polymorph (Form-II),[371] although it has been questioned whether it too can be isolated without contamination from Forms I and/or III.[370] That being said, Forms II and III of olanzapine are known to be metastable with regards to Form I and solid-solid phase transitions can be observed through hot-stage microscopy experiments.[370] Very recently, a fourth distinct crystalline form of olanzapine (Form-IV) has been successfully isolated and characterized.[368] This form differs from other known crystal structures of olanzapine (including the known hydrates and solvates), in that it does not contain the same centrosymmetric dimer stacking motif of olanzapine molecules within its unit cell. Rather, olanzapine molecules in Form-IV interact via hydrogen bonds between neighbouring molecules.[368] The structures of the characterized olanzapine polymorphs are depicted in Figure 7.8.

Figure 7.6: Carbamazepine (CSD[22] refcode CBMZPN) and its five known polymorphs. Forms I-IV crystallise in the $P-1, R-3, P2_1/c$, and $C2/c$ space groups, respectively, and all display hydrogen-bonded dimers of carbamazepine through the carboxamide moiety.[364] In contrast to this, Form-V, which crystallises in the $Pbca$ space group, is the only form which displays catameric hydrogen-bonding interactions between its molecular constituents.[359,365]

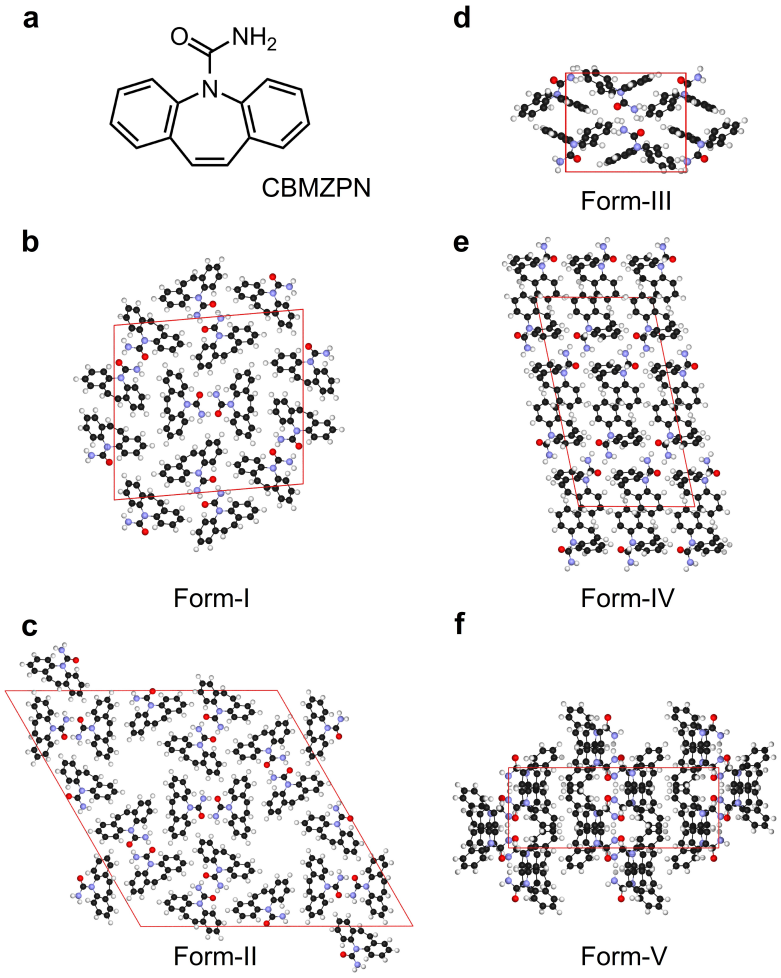


Figure 7.7: Crystal-energy landscapes of carbamazepine. Energy landscapes were generated via (a) a force-field (FF, data obtained from the CPOSS database[339]); and two small- and/or minimal-basis low-cost methods, (d) sHF-3c/MINI(x), and (f) DZP: B86bPBE-XDM/DZP. Crystal structures within panels (a,d,f) were then re-ranked with single-point energy calculations using plane-wave B86bPBE-XDM to give panels (c,e,g), respectively. The force-field structures within panel (a) have also been re-ranked with single-point energy calculations using the small-basis B86bPBE-XDM/DZP method to give panel (b). The computed energies, ΔE, and densities, $\Delta\rho$, are expressed relative to Form-I for all methods. The similarity index, S, is relative to the crystal-energy landscape in panel (g). Relative energies between the isolable polymorphs are tabulated in Table 7.1.

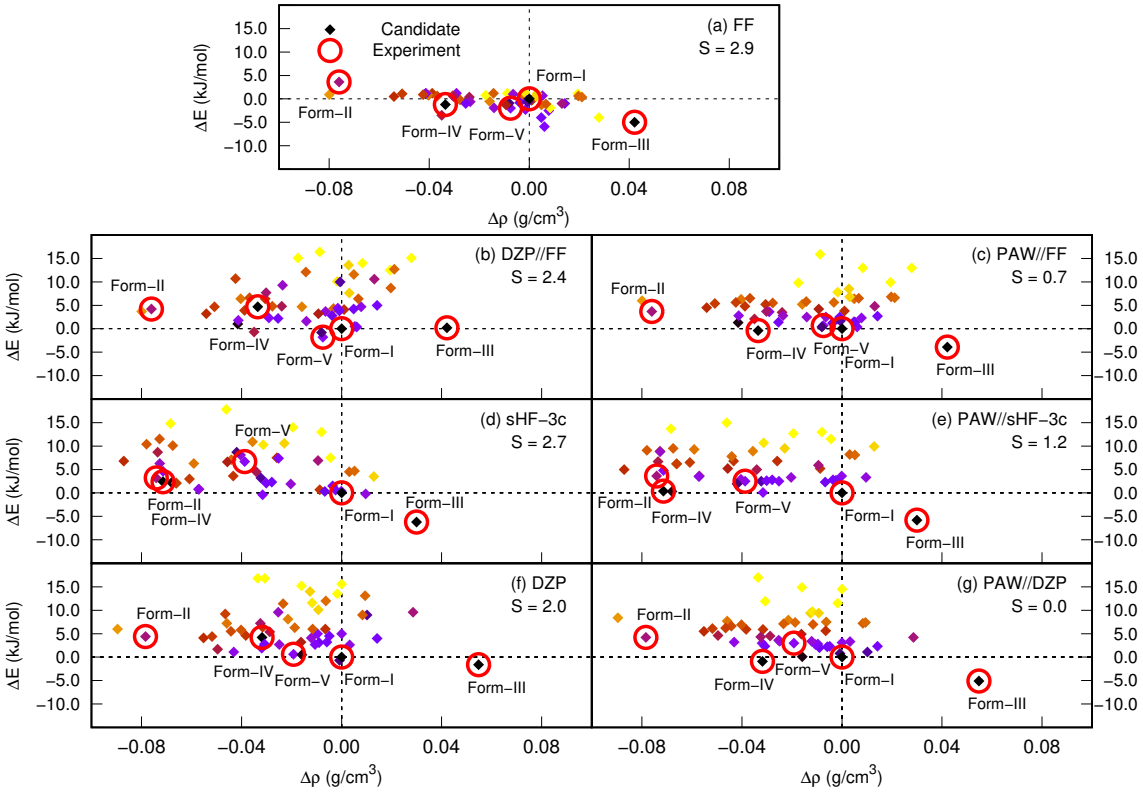
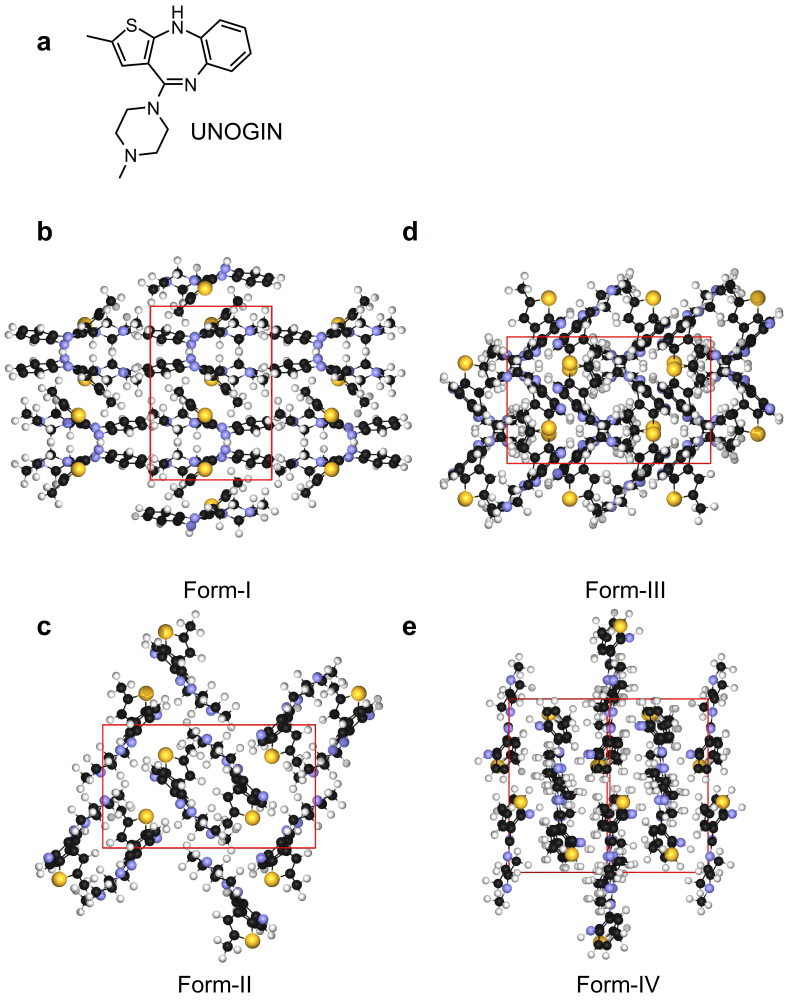


Figure 7.8: Olanzapine (CSD[22] refcode UNOGIN) and its three characterized polymorphs, along with the hypothesized structure of Form-III.[338] Forms I-III all display centrosymmetric dimers of the olanzapine molecules, the first two crystallising in the $P2_1/c$ space group, and the third in the $Pbca$ space group.[338] Analogous to the case of carbamazepine Form-V, Form-IV of olanzapine (space group: $P2_1/c$) does not display centrosymmetric dimers of olanzpine molecules, but rather forms "ladder-like" hydrogen-bonding interactions between each molecule and two of its neighbours.[368] Form-I is the most stable experimentally, but Form-IV was predicted to be thermodynamically competitive from initial CSP studies.[338]



Previous force-field calculations had predicted the new form (Form-IV) as a potentially isolable structure, given its near degeneracy with Form-I,[338] but conventional methods of crystallization had not been able to obtain polymorphs that incorporated anything other than centrosymmetric dimers.[368] The recent report on olanzapine's Form-IV[368] shows that alternative crystallization conditions to solution-based methods (e.g., using the drug-polymer dispersion system) can yield

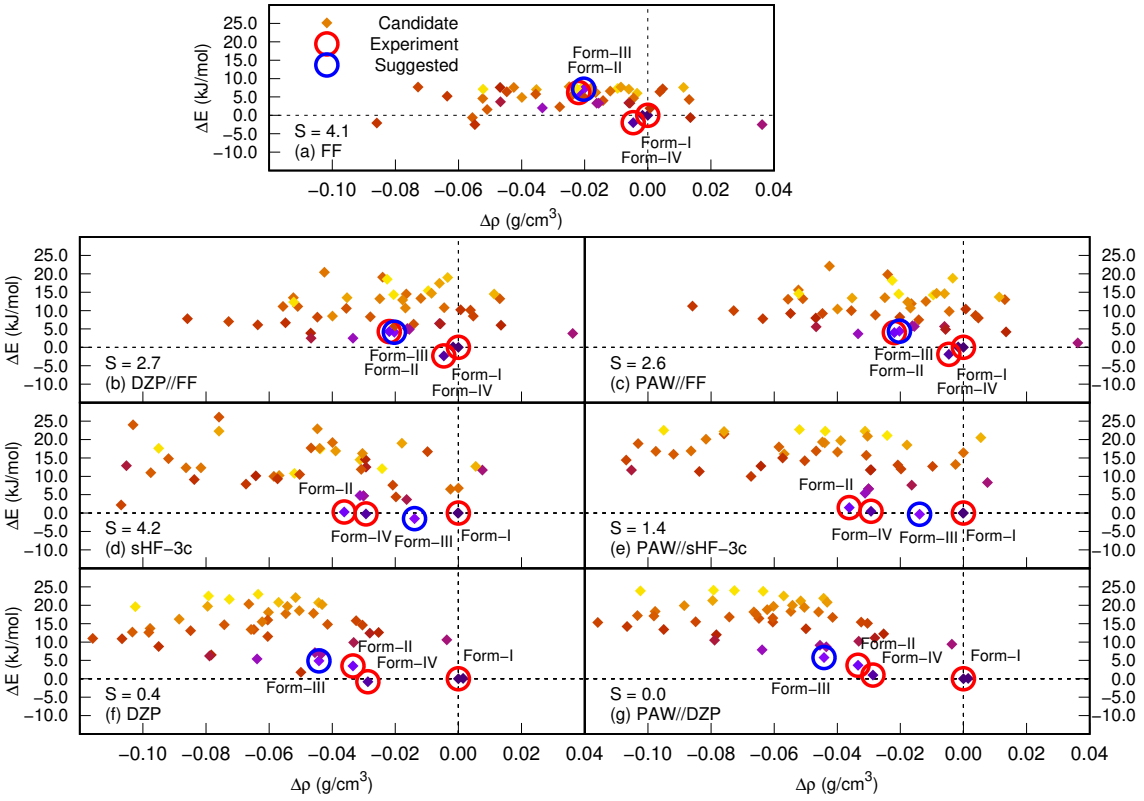
low-energy structures predicted by CSP methods,[338] but never previously isolated. Through differential scanning calorimetry measurements combined with powder X-ray diffraction data, Form-IV was found to be less stable than Form-I, given its lower melting point (ca. 189 vs. 194°C) and lower heat of fusion (ca. 36.7 vs. 40.5 kJ/mol), and no conversion between the forms was observed.[368]

Equally recently, the Gibbs free-energy landscape of olanzapine has been determined through embedded-fragment quantum mechanical methods and Form I has been confirmed to be more stable than Form-II by ca. 2.8 kJ/mol and 4.8 kJ/mol at 5 K and 350 K, respectively.[372] These results are in qualitative agreement with the earlier reported force-field computations of Bhardwaj *et al.*[338] (ca. 6.7 kJ/mol in favour of Form-I at 0 K).

The force-field energy ranking for the 50 lowest-energy structures reported in the previous CSP study[338] is depicted in Figure 7.9(a). The force-field computations predict both Forms I and IV to be in the lower region of the energy landscape, whereas Form-II is predicted to be significantly less stable. Form-II has nearly 40 or so forms more stable than it, making it quite unlikely that this form and not others would be isolated experimentally. Re-ranking of the force-field crystal structures using composite approaches with B86bPBE-XDM/DZP and B86bPBE-XDM/PAW (Figures 7.9(b,c), respectively) now predicts Form II to be significantly lower in energy than most other candidate structures, making these landscapes more realistic. However, Form-IV is still predicted to be lower in energy than Form-I, contrary to what experimental data suggests.[368] The sHF-3c and the B86bPBE-XDM/DZP methods (Figure 7.9(d,f)), along with the corresponding composite approaches (Figure 7.9(e,g)), all generate more plausible energy landscapes than did the force field computations. The reference energy landscape (Figure 7.9(g)) reproduces an energy difference between Forms I and II in good agreement with high-level theory results[372] and, in addition, now predicts Form-I to be slightly more stable than Form-IV by 1.0 kJ/mol, which is in qualitative agreement with experimental data, barring any inclusion of thermal effects.

As for Form-III, its structure has been hypothesized based on structural and energetic similarities present between it and Form-II.[338] Preparing Forms II and III in pure form can be problematic, with mixtures of these forms often resulting instead.[370] The suggested structure by Bhardwaj and co-workers[338] is a likely possibility for this form. The data presented herein with the use of composite approaches support this, given that the proposed Form-III is now brought into the lower-energy regions of the landscape, along with Form-II. It is satisfying to see that all four known polymorph structures are the lowest-energy candidate structures on the crystal-energy landscape

Figure 7.9: Crystal-energy landscapes of olanzapine. Energy landscapes were generated via (a) a force-field (FF, data obtained from the CPOSS database[338,339]); and two small- and/or minimal-basis low-cost methods, (d) sHF-3c/MINI(x), and (f) DZP: B86bPBE-XDM/DZP. Crystal structures within panels (a,d,f) were then re-ranked with single-point energy calculations using plane-wave B86bPBE-XDM to give panels (c,e,g), respectively. The force-field structures within panel (a) have also been re-ranked with single-point energy calculations using the small-basis B86bPBE-XDM/DZP method to give panel (b). The computed energies, ΔE, and densities, $\Delta\rho$, are expressed relative to Form-I for all methods. The similarity index, S, is relative to the crystal-energy landscape in panel (g). Relative energies between the isolable polymorphs are tabulated in Table 7.2



generated with the most accurate composite approach presented in this work (Figure 7.9(g)).

7.4 Conclusions

The work presented in this chapter aimed at applying small-basis set methods, and composite approaches, to the CSP of four APIs. The results obtained were compared to anisotropic force-field data from the CPOSS database. In general, the force fields typically did not yield accurate energy landscapes, as many of the isolable polymorphs were higher in energy than other candidate structures. In addition, the relative stability ordering of the polymorphs did not always agree with the

Table 7.2: Relative energies, in kJ/mol, for the experimental polymorphs of olanzapine computed with low-level and composite approaches. Energies are expressed relative to the most stable form of olanzapine, Form-I. B//A: corresponds to the use of the low-level and composite approach, where A = the lower-level method indicated in the table header, B = the higher-level B86bPBE-XDM/DZP and B86bPBE-XDM/PAW methods (DZP and PAW, respectively).

	FF			sHF-3c		DZP	
	A	DZP//A	PAW//A	A	PAW//A	A	PAW//A
Form-I	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Form-II	6.2	4.3	4.0	0.3	1.5	3.5	3.7
Form-III	7.1	4.1	4.4	-1.5	-0.3	4.9	5.8
Form-IV	-2.0	-2.3	-1.9	-0.2	-0.5	-0.8	1.0

experimental data. This was also the case for the sHF-3c results, most notably for 5-fluorouracil and carbamazepine. In contrast to this, the small-basis B86bPBE-XDM/DZP method implemented in the SIESTA code provided more realistic energy landscapes, consistently placing isolable polymorphs in the lowest-energy regions of the crystal-energy landscapes.

Composite methods, in which high-level single-point energies are calculated at the low-level geometries, yielded improved results in agreement with full plane-wave DFT calculations (for 5-fluorouracil) and, most importantly, in agreement with available experimental evidence. Specifically, for 5-fluorouracil, composite methods recovered the correct experimental ordering, confirming that the previous error is not due to thermal effects, but is a limitation of the force field. For naproxen, it is conjectured that thermal effects may be important to recover the correct ordering, since the known racemic and enantiopure forms are nearly degenerate and the order flips depending on the choice of geometries employed in the composite methods. However, in a CSP protocol, high-level plane-wave DFT optimizations could be performed, together with phonon calculations, to resolve the thermodynamic minimum. For carbamazepine, the composite approaches place most isolable forms within the low-energy region of the crystal-energy landscape. This work adds computational evidence that Form-V is slightly less stable than Forms I and IV, but remains close in energy to the other experimentally isolable polymorphs overall. Inclusion of thermal effects, which remains challenging to do accurately and efficiently,[373] would again be beneficial when comparing to experimental data, or in determining whether other competitive forms still remain to be discovered. For olanzapine, the data obtained via composite methods confirm that there should be four isolable polymorphs, given that four low-energy structures are predicted from the given candidates. Form-I is predicted to be the most stable polymorph, in agreement with experiment, while Form-IV is

slightly less stable than the thermodynamic minimum, and Forms II and III are nearly degenerate in energy. This work also confirms the proposed structure for Form-III from the CSP study of Bhardwaj and co-workers, given that it is one of the four lowest-energy structures identified on the crystal-energy landscape.

Overall, the small-basis B86bPBE-XDM/DZP method yielded the geometries most amenable to further single-point energy calculations with B86bPBE-XDM/PAW, followed by the force-field and sHF-3c geometries, respectively. The poorer performance of the force-field and sHF-3c methods is most likely due to the high levels of (semi-)empiricism involved in these methods' constructions. However, while force-field calculations do not always yield the proper energetic ordering, the use of a composite approach could be useful in identifying crystal-packing motifs where force fields are biased, and subject to systematic errors. The results confirm that, for composite approaches to be most successful and reliable, the low- and high-level methods should be compatible, i.e., be based on similar levels of theory. Finally, while the cost of performing small-basis set calculations with B86bPBE-XDM/DZP is slightly more expensive than sHF-3c, and both of these are relatively more expensive than density-functional tight-binding methods (*viz.* Chapter 5), it should be noted that no empiricism or parameter fitting is involved in B86bPBE-XDM/DZP. This is in stark contrast to the other low-cost methods (force fields, DFTB, and sHF-3c) considered for use in CSP, which makes it a more reliable and generally applicable method for systems where DFT can provide an appropriate description of the electronic structure. Finally, to make the B86bPBE-XDM composite approach competitive with lower-cost semiempirical methods, one could first utilize B86bPBE-XDM/DZP single-point calculations to refine the preliminary energy landscape produced by a force-field. Then, taking only the lower-energy candidate structures, one could further refine the energy ranking and geometries of candidate structures with the composite B86bPBE-XDM/PAW//B86bPBE-XDM/DZP approach.

CHAPTER 8

CONCLUSION

The work presented in this thesis aimed primarily to show how the use of composite approaches can be beneficial in producing crystal-energy landscapes of similar quality and accuracy to high-level quantum-mechanical methods (such as plane-wave DFT) in a more cost-effective manner. That is, geometry optimizations performed by well-chosen low-cost methods can be used to generate crystal geometries amenable to high-level single-point energy calculations, thus foregoing expensive high-level geometry optimizations. The use of composite approaches has several benefits, as it allows not only for more complex structures to be examined by CSP protocols, such as for new materials and active pharmaceutical ingredients, but also allows reliable sampling of a larger number of structures on the energy landscape in the intermediate stages of a CSP protocol.

In design of a composite approach, it is crucial that the low-level geometries are compatible with the high-level energy methods. This was observed in Chapter 5, where the dispersion-corrected DFTB methods gave low-quality geometries for small molecular crystals, which led to lower-quality relative energies obtained from the resulting composite approach. Therefore, while it is less computationally expensive than the other small-basis set methods considered in this thesis, its reliability and/or transferability in describing the many candidate structures on an energy landscape with equal footing remains questionable. Small-basis DFT-D methods implemented in the local-orbital SIESTA code, and the HF-3c method present in the CRYSTAL code, were considered in Chapter 5. Their performance was examined for a proposed benchmark set of chiral amino acids (the EE14 set), and further assessed for a simulated CSP study of 1-aza[6]helicene. These small-basis set methods afforded sub-par relative energies between crystal structures, which is more important for CSP purposes than requiring a method to compute accurate lattice energies. The accuracy of low-cost methods like PBE-D2/DZP and HF-3c was substantially improved upon the use of additional high-level single-point energy calculations with the dispersion-corrected B86bPBE-XDM

functional and plane-wave basis sets.

Given that the choice of low-cost methods impacted whether or not the resulting composite method yielded improved results, it was postulated that composite approaches would be most successful when the low- and high-level methods employ similar levels of theory, such as using the same functional and dispersion correction, and only altering the quality or type of the basis set (e.g., using finite-support numerical orbitals instead of delocalised plane-wave basis sets). This led to the implementation of the B86bPBE-XDM method in the SIESTA code, and testing of its compatibility with numerical basis sets consisting of atomic orbitals of finite support in Chapter 6. Encouraging results were obtained when computing absolute lattice energies of the small molecular solids of the X23 set and, additionally, when refining these energies within a composite approach. The XDM dispersion model has several advantages over the existing dispersion models currently implemented within the code: (i) providing a post-SCF dispersion correction which offers environment-sensitive dispersion coefficients, in contrast to the D2 dispersion model, and (ii) being less computationally expensive than non-local van-der-Waals functionals implemented therein, such as the vdW-DF functionals. Its implementation in the SIESTA code should thus provide a more suitable description of van-der-Waals interactions, applicable to systems of interest in various fields of chemistry and physics.

By the estimates provided in Chapter 5, there is roughly an order of magnitude separating the computational cost of the B86bPBE-XDM/DZP, HF-3c, and DFTB methods, with the first being most expensive. This presents a disadvantage for using B86bPBE-XDM/DZP for CSP purposes. However, the higher computational cost of this method can be circumvented if one reduces the number of candidate structures to fully optimize by first performing DZP single-point energy calculations on the initial set of generated force-field crystal structures. The performance of the B86bPBE-XDM composite method, combining DZP optimizations, followed by single-point energy calculations using plane-wave methods, was ultimately validated by generating accurate crystal-energy landscapes for drug-like molecules in Chapter 7. This composite approach excelled in generating more realistic crystal-energy landscapes for the drug-like molecules studied, in contrast to the cheaper, but more empirical force-field and HF-3c methods. It was found that methods which are heavily parametrized can lead to unforeseeable errors or present bias towards the description of certain crystal structures on an energy landscape. However, the use of composite approaches can help to identify problematic cases where force-fields or other semiempirical approaches may fail to properly rank crystal structures, by identifying data points on the landscape that

are substantially shifted by the use of high-level single-point energy calculations. The combined B86bPBE-XDM composite approach generally predicted the experimentally isolated polymorphs as the lowest-energy structures on each landscape and provided the same stability ordering of the polymorphs as seen experimentally.

In cases where there are small energy differences between some polymorph pairs, additional calculations could be performed to account for for thermal expansion and free-energy contributions to effectively determine the relative stabilities of the polymorphs at finite temperatures.[295] In this regard, similar composite approaches can be applied using, e.g., the quasi-harmonic approximation, in which low-level methods are used to compute the phonons and obtain the thermal expansion and free-energy corrections.[331,373,374] The cost of performing these phonon calculations, however, remains significant and is therefore practically restricted to a few low-energy structures of interest,[373] i.e., candidate structures within a small energy window in the lowest-lying regions of the crystal-energy landscape.[85] Therefore, if electronic crystal-energy landscapes can first be efficiently and accurately produced, additional computational resources can be diverted to compute free-energy corrections for a small subset of structures.

The reliability of a composite approach relies not only on the ability of the low-cost method to provide geometries amenable to the higher-level energy calculations, but also on the reliability of the high-level method for properly describing the system of interest. For instance, the work detailed in Chapter 4 came to the conclusion that existing DFT methods, as they stand, are not yet capable of describing systems where the inherent delocalisation error substantially impacts the relative energies of systems in different ionization states. In the case of organic acid-base salts and co-crystals, this can lead to spurious proton transfers between the acidic and basic sites. Existing DFT methods are thus not recommended to the study of organic acid-base multicomponent systems. Furthermore, the choice of systems used to benchmark computational methods is equally of importance: polytypical systems, such as aspirin (Chapter 3), were found to not be suitable systems for these purposes. The low energetic barriers that exist for conversion between two phases can often lead to difficulties in experimentally isolating a given polytype (or polymorph) in a pure phase. Without precise experimental evidence to establish the relative stabilities of polytypes, assessing the capabilities of computational methods at producing the correct energy differences remains a futile exercise.

As the field of crystal structure prediction is moving towards describing more complex and

challenging systems, there is a need to develop protocols that strike a balance between efficiency, accuracy, and reliability, all while being automatable and transferable to many systems of interest in academic and/or industrial settings.[375] It is hoped that the work highlighted in this thesis will provide the necessary tools to do just this, all while outlining the subtleties which can arise during the refinement stages of a CSP protocol. For production CSP calculations, it would seem reasonable to use the following approach to yield reliable results in the most cost-effective manner, while minimizing empiricism: (i) structures are first generated and ranked with a force-field method, (ii-a) these are then subjected to single-point energy calculations with a low-level quantum-mechanical method such as B86bPBE-XDM/DZP (as implemented in SIESTA); (ii-b) a subset of these are then fully-optimized the same low-level method, and (iii) single-point energies are obtained from higher-level plane-wave B86bPBE-XDM calculations. Finally, if needed, (iv) thermal and entropic effects are included for a handful of structures through the use of composite approaches following the same ideology detailed elsewhere in the literature.[331,373,374]

APPENDIX A

SUPPORTING INFORMATION FOR CHAPTER 5

Contents: Computed and benchmark lattice energies for the X23 set crystal structures, predicted and experimental enantiomeric excesses for the EE14 set crystal structures, and relative energy differences and densities of crystal structures from the CSP study of 1-aza[6]helicene.

Table A.1: Lattice energies, in kJ/mol per molecule, of the X23 benchmark calculated with the PBE-D2 method, as implemented in the SIESTA code. Reference lattice energies back-corrected from experimentally-measured sublimation enthalpies [145] (Expt.), and computed with B86bPBE-XDM (DFT) are also presented. DZP or TZP: a double-zeta (by default) or triple-zeta plus polarization basis set was used, with default convergence thresholds for forces and stresses during optimization, a $4\times4\times4$ k-point mesh, and reduced SIESTA energy shift parameter (0.001 Ry). FC/ST: tighter force and stress convergence thresholds (0.01 eV/Å and 0.02 GPa for maximum component values). ES: default SIESTA energy shift parameter (0.02 Ry). k2: using a $2\times2\times2$ k-point mesh. DFTB: SCC-DFTB3-D3 geometries were used as the starting point for the PBE-D2/DZP optimization.

no	structure	Reference		PBE-D2													
		Expt.	DFT	DZP	B//A	TZP	B//A	FC/ST	B//A	ES	B//A	k2	B//A	DFTB	B//A	DFTB; FC/ST	B//A
1	14-cyclohexanedione	88.6	88.6	100.5	89.6	93.3	88.8	100.6	89.2	244.5	81.4	100.4	89.6	100.3	89.0	100.7	89.2
2	acetic acid	72.8	74.4	85.7	74.9	76.5	74.4	85.8	74.8	193.9	67.5	85.1	74.8	85.7	74.7	85.8	74.6
3	adamantane	69.4	72.0	68.4	71.6	81.5	71.0	69.2	72.7	107.6	71.8	67.8	71.6	69.2	72.4	69.3	72.7
4	ammonia	37.2	41.4	46.6	41.5	45.4	41.4	46.8	41.4	89.8	41.2	46.4	41.5	46.9	41.0	46.9	41.0
5	anthracene	112.7	101.7	87.5	103.2	115.0	101.6	87.6	102.9	240.9	92.5	86.8	103.2	86.7	101.7	87.6	102.9
6	benzene	51.7	51.8	47.7	52.9	59.9	51.8	47.7	52.4	122.1	48.1	47.7	52.9	47.5	52.4	47.8	52.5
7	carbon dioxide	28.4	24.2	29.9	24.8	22.2	24.1	29.8	24.9	102.8	12.1	29.9	24.8	29.8	24.7	29.8	24.9
8	cyanamide	79.7	90.9	100.7	91.9	92.9	91.0	100.7	91.9	192.1	86.3	100.7	91.9	100.7	91.8	100.7	91.9
9	cytosine	169.8	156.4	173.2	157.7	164.8	156.7	173.3	157.6	334.3	146.6	172.8	157.6	173.2	129.4	173.2	129.6
10	ethylcarbamate	86.3	86.7	100.7	87.6	90.7	86.8	101.0	87.4	216.6	80.2	100.1	87.7	100.6	87.6	101.0	87.5
11	formamide	79.2	81.1	93.4	81.7	83.6	81.1	93.4	81.5	188.5	76.0	92.1	81.4	93.4	81.3	93.4	81.4
12	hexamine	86.2	85.5	93.4	86.2	90.8	85.5	93.6	86.0	221.4	84.6	94.0	86.2	93.4	85.5	93.6	85.7
13	imidazole	86.8	90.4	98.4	90.9	95.2	90.4	98.6	90.8	191.2	85.6	98.4	90.8	98.5	90.8	98.6	90.7
14	naphthalene	81.7	76.4	66.2	77.6	86.2	76.3	66.3	77.4	181.8	69.0	65.9	77.6	66.1	76.8	66.3	77.1
15	oxalic acid α	96.3	94.8	108.1	95.7	92.3	94.9	107.9	95.6	257.7	80.3	108.1	95.7	107.8	95.4	107.9	95.5
16	oxalic acid β	96.1	97.1	111.5	97.8	96.1	97.2	111.6	97.8	259.3	80.6	109.7	98.1	110.3	97.2	111.6	97.8
17	pyrazine	61.3	62.2	66.2	62.9	66.0	62.2	66.2	63.0	180.8	56.7	65.8	62.9	66.2	63.0	66.2	62.9
18	pyrazole	77.1	78.8	85.9	79.4	84.6	78.8	86.2	79.3	178.4	73.0	85.9	79.4	86.0	79.0	86.1	79.2
19	succinic acid	130.3	133.0	154.4	134.3	136.4	133.9	153.9	133.5	356.0	123.7	154.4	134.3	153.8	132.7	153.8	132.9
20	triazine	61.7	57.4	66.7	58.4	59.9	57.5	66.8	58.3	192.0	49.3	66.8	58.4	66.9	58.4	67.0	58.6
21	trioxane	66.4	60.7	72.0	61.4	61.5	60.8	71.9	61.1	197.8	54.8	72.0	61.4	71.7	61.0	72.1	61.0
22	uracil	135.7	134.0	151.4	135.0	138.2	134.1	151.5	135.0	323.5	123.7	150.0	134.3	150.6	102.4	151.4	135.1
23	urea	102.5	106.4	120.7	107.0	110.1	106.4	120.7	107.2	230.2	102.1	119.0	107.1	120.9	107.1	120.6	107.1

Table A.2: Lattice energies, in kJ/mol per molecule, of the X23 benchmark calculated with low-cost methods, as implemented in the SIESTA, DFTB+ and CRYSTAL17 codes. Reference lattice energies back-corrected from experimentally-measured sublimation enthalpies[145] (Expt.), and computed with B86bPBE-XDM (DFT) are also presented.

no	structure	Reference		vdW-DF1		vdW-DF2		DFTB3-D3		HF-D2		sHF-3c		PBE-D2		PBE-D3		PBEh-3c	
		Expt.	DFT	DZP	B//A	DZP	B//A	A	B//A	A	B//A	A	B//A	A	B//A	A	B//A	A	B//A
1	14-cyclohexanedione	88.6	88.6	124.6	88.7	123.6	88.8	105.7	75.9	108.3	86.3	76.4	84.1	200.0	77.1	187.9	82.3	85.9	86.9
2	acetic acid	72.8	74.4	97.3	77.5	92.7	73.6	84.0	63.1	90.5	67.8	68.9	70.3	138.5	69.8	135.6	71.9	70.3	73.5
3	adamantane	69.4	72.0	92.1	69.9	86.5	69.9	92.8	67.7	103.8	65.5	68.2	70.1	155.5	55.8	129.4	67.2	59.3	70.8
4	ammonia	37.2	41.4	42.8	40.8	46.8	41.1	44.2	27.1	49.5	38.0	42.3	41.0	67.2	40.6	64.2	41.3	51.0	41.1
5	anthracene	112.7	101.7	124.8	99.4	112.9	101.8	125.8	87.9	126.0	95.3	111.4	94.9	537.0	14.9	514.2	30.7	– [a]	– [a]
6	benzene	51.7	51.8	65.4	51.1	60.6	51.8	65.4	44.5	63.4	47.9	57.1	48.1	252.9	18.3	240.0	24.8	52.1	51.6
7	carbon dioxide	28.4	24.2	44.9	25.0	43.1	24.4	25.3	17.7	41.1	19.6	32.5	20.3	74.1	19.4	74.0	19.5	27.3	24.5
8	cyanamide	79.7	90.9	98.8	93.1	102.1	90.5	69.8	79.7	72.9	84.8	73.0	87.0	163.5	81.0	162.7	84.5	90.2	90.0
9	cytosine	169.8	156.4	183.1	159.8	181.3	157.5	154.8	129.1	184.2	156.0	155.8	151.8	309.7	136.6	302.3	142.1	156.4	154.0
10	ethylcarbamate	86.3	86.7	112.4	89.0	114.0	86.3	100.2	76.9	107.3	80.0	79.3	82.5	157.6	79.7	151.4	84.2	80.1	84.9
11	formamide	79.2	81.1	95.1	82.2	97.2	81.1	81.4	69.7	93.1	78.7	72.6	76.7	139.5	77.6	138.1	79.3	76.5	79.8
12	hexamine	86.2	85.5	111.7	85.1	110.5	84.4	80.9	82.0	129.7	81.2	92.9	84.4	212.7	68.0	191.7	79.7	92.2	84.6
13	imidazole	86.8	90.4	106.0	91.4	101.8	90.4	74.9	77.9	104.6	87.9	94.0	88.0	204.0	73.9	196.8	78.6	92.8	90.5
14	naphthalene	81.7	76.4	93.4	74.8	85.7	76.1	95.0	65.7	96.3	70.5	85.8	70.4	393.1	16.1	376.1	27.4	– [a]	– [a]
15	oxalic acid α	96.3	94.8	131.2	99.9	128.4	95.1	125.4	65.0	144.6	93.0	113.5	90.8	205.8	91.2	204.2	91.6	99.9	95.1
16	oxalic acid β	96.1	97.1	136.1	99.2	126.4	96.5	130.7	71.2	147.8	92.5	109.1	91.5	205.3	91.7	202.2	93.4	98.1	96.2
17	pyrazine	61.3	62.2	83.8	61.8	82.0	62.2	64.1	54.8	72.8	62.4	67.1	58.7	165.7	42.9	161.7	47.7	63.5	61.7
18	pyrazole	77.1	78.8	94.8	79.6	91.2	79.0	62.1	59.5	87.2	78.4	77.9	77.5	204.4	65.1	197.6	68.3	79.6	78.9
19	succinic acid	130.3	133.0	177.5	138.1	167.3	131.2	153.4	108.6	174.2	118.1	125.3	121.7	254.2	121.8	245.5	127.3	120.9	130.9
20	triazine	61.7	57.4	83.8	58.2	83.0	58.4	59.4	49.4	67.2	58.3	57.0	56.1	146.9	42.4	141.8	46.6	60.1	57.6
21	trioxane	66.4	60.7	90.6	61.1	94.8	59.2	64.5	57.8	100.9	57.7	61.9	58.7	132.9	55.3	123.5	58.3	58.9	58.4
22	uracil	135.7	134.0	162.8	136.1	161.3	134.6	149.5	104.0	157.3	132.5	125.7	128.0	276.1	114.1	273.0	118.3	125.4	131.8
23	urea	102.5	106.4	118.5	108.6	127.0	107.3	112.7	102.7	126.5	103.4	99.6	105.5	176.7	101.7	172.8	104.7	95.6	103.6

[a] Anthracene and naphthalene, which have small band gaps, have been excluded from the statistics for PBEh-3c because the SCF cycle failed to converge. This is in agreement with previous reports.[193]

Table A.3: POWDIFF structure measures[293,294] of X23 crystal geometries generated by low-cost methods, as implemented in the SIESTA and DFTB+ codes. Structures are referenced to B86bPBE-XDM equilibrium structures.

no	structure	PBE-D2							vdW-DF1	vdW-DF2	DFTB3-D3
		DZP	TZP	FC/ST	ES	k2	DFTB	DFTB; FC/ST	DZP	DZP	DZP
1	14-cyclohexanedione	0.0193	0.0000	0.0866	0.4443	0.0194	0.0770	0.0873	0.2366	0.0578	0.6932
2	acetic acid	0.0009	0.0001	0.0096	0.5791	0.0107	0.0215	0.0161	0.0787	0.0421	0.5625
3	adamantane	0.0128	0.0002	0.0244	0.1487	0.0129	0.0362	0.0271	0.0820	0.0751	0.5295
4	ammonia	0.0010	0.0002	0.0904	0.1940	0.0011	0.4130	0.3270	0.9865	0.6132	0.9547
5	anthracene	0.0274	0.0000	0.1269	0.5297	0.0277	0.1883	0.1161	0.4400	0.1839	0.6381
6	benzene	0.0415	0.0000	0.0451	0.4643	0.0415	0.0326	0.0692	0.4836	0.0866	0.7278
7	carbon dioxide	0.0089	0.4806	0.0053	0.8246	0.0089	0.1016	0.0051	0.0070	0.4257	0.9759
8	cyanamide	0.0014	0.0000	0.0187	0.5809	0.0012	0.1067	0.0279	0.5587	0.0930	0.8179
9	cytosine	0.0073	0.0003	0.0316	0.3740	0.0971	0.0317	0.0339	0.2024	0.1482	0.6183
10	ethylcarbamate	0.0030	0.0000	0.0418	0.6663	0.0010	0.0157	0.0391	0.0966	0.3493	0.5022
11	formamide	0.0041	0.0000	0.0595	0.4956	0.1443	0.1114	0.0672	0.1651	0.0851	0.5002
12	hexamine	0.0069	0.0000	0.0007	0.1098	0.0069	0.0019	0.0012	0.1010	0.0646	0.2781
13	imidazole	0.0061	0.0000	0.0057	0.4790	0.0058	0.0072	0.0007	0.1753	0.0661	0.7128
14	naphthalene	0.0229	0.0000	0.1187	0.5512	0.0231	0.1538	0.1448	0.4271	0.1749	0.6502
15	oxalic acid α	0.0053	0.0982	0.0158	0.7388	0.0051	0.0244	0.0136	0.0907	0.1278	0.7447
16	oxalic acid β	0.0175	0.0010	0.0446	0.7647	0.0091	0.0329	0.0372	0.2464	0.1286	0.8221
17	pyrazine	0.0118	0.0000	0.0630	0.4122	0.2920	0.1626	0.0842	0.3135	0.0306	0.6238
18	pyrazole	0.0241	0.0002	0.0373	0.7793	0.0241	0.0441	0.0331	0.2889	0.1750	0.6113
19	succinic acid	0.0011	0.0001	0.0308	0.7965	0.0012	0.4850	0.3891	0.0882	0.1490	0.6925
20	triazine	0.0000	0.0000	0.0002	0.0041	0.0000	0.4682	0.2884	0.0000	0.0010	0.7322
21	trioxane	0.0000	0.0000	0.0001	0.0035	0.0000	0.0537	0.0502	0.0001	0.0012	0.8774
22	uracil	0.0049	0.0004	0.0219	0.5451	0.0849	0.0286	0.0055	0.0996	0.0383	0.4509
23	urea	0.0059	0.0000	0.0007	0.7286	0.0053	0.0006	0.0008	0.2670	0.1172	0.6371

Table A.4: POWDIFF structure measures[293,294] of X23 crystal geometries generated by low-cost methods, as implemented in the CRYSTAL17 code. Structures are referenced to B86bPBE-XDM equilibrium structures.

no	structure	HF-D2	sHF-3c	PBE-D2	PBE-D3	PBEh-3c
1	14-cyclohexanedione	0.2979	0.0355	0.6671	0.5568	0.1674
2	acetic acid	0.5163	0.1736	0.4453	0.2575	0.1621
3	adamantane	0.5383	0.2321	0.8118	0.5324	0.0072
4	ammonia	1.0000	0.8924	0.7522	0.0108	0.4737
5	anthracene	0.5024	0.5024	0.5860	0.6003	– [a]
6	benzene	0.5740	0.5009	0.9113	0.9178	0.0341
7	carbon dioxide	0.9231	0.9947	0.9941	0.9932	0.0017
8	cyanamide	0.5377	0.5757	0.7939	0.6267	0.0847
9	cytosine	0.2968	0.2482	0.4116	0.4069	0.0727
10	ethylcarbamate	0.5510	0.2357	0.5853	0.5623	0.3612
11	formamide	0.4969	0.4644	0.4538	0.4223	0.2032
12	hexamine	0.4404	0.2229	0.7466	0.3918	0.0306
13	imidazole	0.4129	0.2742	0.7731	0.5275	0.0101
14	naphthalene	0.5847	0.5736	0.5991	0.5548	– [a]
15	oxalic acid α	0.4225	0.4356	0.4660	0.4291	0.0520
16	oxalic acid β	0.6989	0.4945	0.6440	0.5992	0.1013
17	pyrazine	0.3521	0.3754	0.6234	0.5844	0.0439
18	pyrazole	0.5289	0.2125	0.8449	0.9026	0.0445
19	succinic acid	0.7435	0.2227	0.7861	0.6644	0.3135
20	triazine	0.4793	0.3311	0.7147	0.7054	0.0436
21	trioxane	0.8512	0.6871	0.6713	0.4898	0.7500
22	uracil	0.6275	0.5199	0.4363	0.5164	0.0986
23	urea	0.7722	0.3307	0.7181	0.3677	0.4275

[a] Anthracene and naphthalene, which have small band gaps, have been excluded from the statistics for PBEh-3c because the SCF cycle failed to converge. This is in agreement with previous reports.[193]

Table A.5: Lattice energies, in kJ/mol per molecule, of the X23 benchmark calculated with plane-wave DFT methods implemented in the Quantum ESPRESSO code. Reference lattice energies back-corrected from experimentally-measured sublimation enthalpies[145] (Expt.), and computed with B86bPBE-XDM (DFT) are also presented.

no	structure	Reference		BLYP-XDM		PBE-D2		PBE-XDM		PW86PBE-XDM		rVV10		vdW-DF1		vdW-DF2	
		Expt.	DFT	A	B//A	A	B//A	A	B//A	A	B//A	A	B//A	A	B//A	A	B//A
1	14-cyclohexanedione	88.6	88.6	98.0	86.5	96.1	87.0	85.5	88.7	90.5	89.0	111.2	87.7	102.4	85.0	101.1	88.2
2	acetic acid	72.8	74.4	76.2	73.6	76.1	74.0	72.9	74.7	73.7	74.7	84.3	73.7	75.3	71.6	75.4	73.1
3	adamantane	69.4	72.0	84.8	69.5	87.8	66.4	70.1	71.4	75.4	72.2	96.9	71.7	88.5	65.5	83.2	70.8
4	ammonia	37.2	41.4	42.0	41.5	45.9	41.0	40.3	41.7	42.2	41.7	46.2	41.4	37.1	39.0	40.1	40.6
5	anthracene	112.7	101.7	109.4	96.9	106.1	98.0	97.1	101.6	103.2	102.0	127.0	101.5	115.8	94.6	107.4	100.0
6	benzene	51.7	51.8	57.0	49.0	56.3	48.9	50.1	51.8	53.0	51.9	62.9	51.6	59.5	48.1	55.4	51.3
7	carbon dioxide	28.4	24.2	26.3	22.7	23.9	24.1	23.7	24.2	25.2	24.1	32.1	22.8	35.0	24.0	33.3	23.9
8	cyanamide	79.7	90.9	92.5	89.7	92.7	90.5	89.7	90.4	91.0	91.0	99.5	90.5	84.7	87.6	88.6	89.0
9	cytosine	169.8	156.4	159.8	154.2	162.6	155.4	151.9	156.5	155.8	157.0	177.6	155.8	151.9	151.9	155.2	154.3
10	ethylcarbamate	86.3	86.7	91.6	85.4	91.4	85.0	84.9	87.2	87.8	87.2	101.8	86.0	93.6	83.9	92.8	86.0
11	formamide	79.2	81.1	82.6	80.0	82.9	80.4	79.2	81.1	81.3	81.2	90.3	80.6	78.0	78.6	81.1	80.1
12	hexamine	86.2	85.5	91.0	83.9	97.4	81.9	81.6	85.9	89.1	86.4	112.6	84.8	99.5	81.7	95.5	85.0
13	imidazole	86.8	90.4	92.3	88.6	94.4	88.8	88.8	90.3	90.3	90.6	100.9	90.0	89.3	87.3	88.7	89.0
14	naphthalene	81.7	76.4	83.0	72.5	80.5	73.0	72.9	76.2	77.6	76.5	94.5	76.2	87.3	70.9	81.0	75.1
15	oxalic acid α	96.3	94.8	101.2	96.6	94.6	94.8	90.4	97.8	96.1	97.9	117.2	93.7	101.5	91.7	106.4	93.7
16	oxalic acid β	96.1	97.1	102.6	98.6	98.0	97.0	93.0	100.1	97.1	100.0	118.3	95.4	101.9	94.0	105.8	94.9
17	pyrazine	61.3	62.2	65.4	59.9	64.3	61.6	60.0	62.5	64.3	62.5	76.3	62.0	68.5	59.0	67.1	61.9
18	pyrazole	77.1	78.8	81.7	77.4	83.6	77.5	77.2	78.8	78.7	79.0	89.5	78.4	78.7	75.8	77.9	77.5
19	succinic acid	130.3	133.0	138.3	130.3	136.8	133.5	128.6	132.7	131.5	132.7	156.5	131.3	133.7	128.4	135.4	130.0
20	triazine	61.7	57.4	60.8	55.5	60.8	56.7	55.5	57.5	59.9	57.8	72.9	56.8	67.4	55.0	65.6	57.3
21	trioxane	66.4	60.7	68.0	59.6	65.0	60.0	57.9	61.6	63.2	61.8	78.8	59.6	74.2	59.2	73.8	60.5
22	uracil	135.7	134.0	137.6	132.6	136.3	133.2	130.7	134.2	133.8	134.6	155.2	133.4	134.6	130.0	136.7	132.2
23	urea	102.5	106.4	109.7	106.1	110.7	105.6	103.6	106.7	106.9	106.7	119.6	106.2	100.3	103.6	106.4	105.6

Table A.6: POWDIFF structure measures[293,294] of X23 crystal geometries generated by plane-wave DFT methods implemented in the Quantum ESPRESSO code. Structures are referenced to B86bPBE-XDM equilibrium structures.

no	structure	BLYP-XDM	PBE-D2	PBE-XDM	PW86PBE-XDM	rVV10	vdW-DF1	vdW-DF2
1	14-cyclohexanedione	0.3120	0.2213	0.1444	0.0022	0.1854	0.5268	0.0420
2	acetic acid	0.1768	0.0950	0.0691	0.0037	0.0900	0.4185	0.0737
3	adamantane	0.3927	0.5715	0.0000	0.0764	0.1925	0.3655	0.0172
4	ammonia	0.3489	0.4803	0.2019	0.0212	0.0641	0.9992	0.8337
5	anthracene	0.4167	0.3411	0.1522	0.0279	0.0087	0.6225	0.4126
6	benzene	0.4459	0.4781	0.1169	0.0028	0.0534	0.7734	0.3164
7	carbon dioxide	0.8561	0.0048	0.3948	0.1130	0.8026	0.5637	0.1899
8	cyanamide	0.6280	0.1689	0.2777	0.0144	0.3211	0.7155	0.3022
9	cytosine	0.2302	0.1566	0.2201	0.0116	0.0854	0.3127	0.1696
10	ethylcarbamate	0.5645	0.4850	0.2928	0.0201	0.3965	0.4601	0.0327
11	formamide	0.3612	0.3024	0.2339	0.0230	0.2614	0.4048	0.1297
12	hexamine	0.1522	0.3120	0.0579	0.0000	0.0538	0.5008	0.1303
13	imidazole	0.3571	0.3033	0.2399	0.0025	0.1345	0.6249	0.2171
14	naphthalene	0.4598	0.4073	0.1732	0.0285	0.0189	0.6846	0.4306
15	oxalic acid α	0.2794	0.0068	0.2005	0.0154	0.2317	0.5238	0.1391
16	oxalic acid β	0.5197	0.0248	0.3074	0.0305	0.5361	0.5597	0.1229
17	pyrazine	0.4148	0.2098	0.2296	0.0079	0.1130	0.5791	0.2625
18	pyrazole	0.3730	0.3299	0.1568	0.0017	0.1189	0.7524	0.2256
19	succinic acid	0.6022	0.4117	0.3326	0.0152	0.4475	0.5950	0.0369
20	triazine	0.6608	0.2513	0.4509	0.0083	0.1859	0.8145	0.3496
21	trioxane	0.3926	0.0447	0.1868	0.0025	0.1553	0.7223	0.0781
22	uracil	0.6091	0.3663	0.4882	0.0040	0.4622	0.6442	0.1130
23	urea	0.1221	0.2086	0.0587	0.0072	0.0677	0.7075	0.2874

Table A.7: Predicted energy differences between the enantiopure (L) and racemate (DL) crystal structures, $\Delta E_{\mathrm{L-DL}}$ in kJ/mol, and enantiomeric excesses, "ee" in %, of the EE14 benchmark set calculated with low-cost methods using the DZP basis set and the SIESTA code. Experimentally-measured enantiomeric excesses (Expt.) and computed B86bPBE-XDM relative energies/predicted ee's (DFT) are also presented. A: full optimization using the stated method. B//A: an additional B86bPBE-XDM single-point energy calculation was performed.

no	structure	Expt.	DFT		PBE-D2				vdW-DF1				vdW-DF2			
		ee	A	ee	A	ee	B//A	ee	A	ee	B//A	ee	A	ee	B//A	ee
1	ala	60.4	2.4	25.6	3.0	48.1	2.5	30.5	3.1	50.0	2.5	30.5	2.6	35.5	2.5	29.2
2	aldol	34.0	0.9	0.0	1.4	0.0	0.7	0.0	1.0	0.0	0.3	0.0	0.7	0.0	1.0	0.0
3	asp	0.7	-1.5	0.0	-5.6	0.0	-1.3	0.0	-12.9	0.0	-0.7	0.0	-14.0	0.0	-0.7	0.0
4	oxazoline	62.0	0.3	0.0	0.8	0.0	-0.2	0.0	-1.5	0.0	2.1	13.8	-6.5	0.0	0.6	0.0
5	cys	58.4	2.6	32.4	-1.1	0.0	2.6	34.2	1.2	0.0	2.5	30.5	0.9	0.0	2.3	22.0
6	glu	0.7	-4.1	0.0	-6.6	0.0	-3.9	0.0	-4.4	0.0	-4.8	0.0	-1.1	0.0	-3.0	0.0
7	his	93.7	9.4	99.6	11.2	99.9	9.2	99.5	9.6	99.6	9.5	99.7	10.0	99.8	9.0	99.5
8	leu	87.9	5.0	86.5	7.2	97.7	5.1	87.3	7.2	97.7	5.4	90.8	4.5	80.8	5.0	87.3
9	N-helicene	74.0	3.4	58.7	7.2	97.6	4.3	77.1	3.4	59.3	2.4	25.6	5.3	89.8	5.2	88.2
10	pro	44.4	-0.7	0.0	0.1	0.0	-0.7	0.0	0.2	0.0	-0.7	0.0	2.1	15.2	-0.8	0.0
11	ser	100.0	16.4	100.0	15.8	100.0	16.3	100.0	10.5	99.8	17.8	100.0	6.0	93.9	15.1	100.0
12	tetrazole	100.0	12.5	100.0	11.8	99.9	13.3	100.0	12.2	100.0	12.1	99.9	12.7	100.0	12.6	100.0
13	tyr	51.7	2.6	32.7	3.2	52.0	2.5	29.2	3.4	60.0	2.4	28.0	2.5	31.6	2.3	23.2
14	val	44.1	2.5	29.8	4.0	72.0	2.6	34.2	3.0	48.1	2.5	30.5	1.4	0.0	2.8	41.9

Table A.8: Predicted energy differences between the enantiopure (L) and racemate (DL) crystal structures, $\Delta E_{\mathrm{L-DL}}$ in kJ/mol, and enantiomeric excesses, "ee" in %, of the EE14 benchmark set calculated with the DFTB+ code. Experimentally-measured enantiomeric excesses (Expt.) and computed B86bPBE-XDM relative energies/predicted ee's (DFT) are also presented. A: full optimization using the stated method. B//A: an additional B86bPBE-XDM single-point energy calculation was performed.

no	structure	Expt.	DFT		DFTB3-D3			
		ee	A	ee	A	ee	B//A	ee
1	ala	60.4	2.4	25.6	0.9	0.0	3.8	68.8
2	aldol	34.0	0.9	0.0	0.9	0.0	-0.2	0.0
3	asp	0.7	-1.5	0.0	7.0	97.2	7.6	98.4
4	oxazoline	62.0	0.3	0.0	-9.3	0.0	9.5	99.7
5	cys	58.4	2.6	32.4	4.8	85.1	8.2	99.0
6	glu	0.7	-4.1	0.0	5.2	88.5	-1.3	0.0
7	his	93.7	9.4	99.6	10.3	99.8	-2.2	0.0
8	leu	87.9	5.0	86.5	8.2	99.0	-1.6	0.0
9	N-helicene	74.0	3.4	58.7	-3.6	0.0	6.4	95.6
10	pro	44.4	-0.7	0.0	0.4	0.0	1.7	1.0
11	ser	100.0	16.4	100.0	13.4	100.0	-0.9	0.0
12	tetrazole	100.0	12.5	100.0	13.9	100.0	4.3	78.2
13	tyr	51.7	2.6	32.7	2.6	33.4	-3.4	0.0
14	val	44.1	2.5	29.8	4.7	82.8	0.7	0.0

Table A.9: Predicted energy differences between the enantiopure (L) and racemate (DL) crystal structures, $\Delta E_{\mathrm{L-DL}}$ in kJ/mol, and enantiomeric excesses, "ee" in %, of the EE14 benchmark set calculated with the CRYSTAL17 code. Experimentally-measured enantiomeric excesses (Expt.) and computed B86bPBE-XDM relative energies/predicted ee's (DFT) are also presented. A: full optimization using the stated method. B//A: an additional B86bPBE-XDM single-point energy calculation was performed.

no	structure	Expt.	DFT		HF-D2				sHF-3c				PBEh-3c			
		ee	A	ee	A	ee	B//A	ee	A	ee	B//A	ee	A	ee	B//A	ee
1	ala	60.4	2.4	74.2	4.8	84.0	4.6	82.5	4.5	80.5	4.5	81.4	3.6	64.0	2.3	23.2
2	aldol	34.0	1.0	36.9	0.5	0.0	1.7	1.0	0.3	0.0	1.5	0.0	0.6	0.0	0.9	0.0
3	asp	0.7	-1.5	0.0	2.1	14.4	-6.3	0.0	3.9	71.0	-5.9	0.0	0.8	0.0	-1.2	0.0
4	oxazoline	62.0	1.4	51.3	-5.3	0.0	4.2	75.9	3.4	58.2	0.6	0.0	0.3	0.0	0.5	0.0
5	cys	58.4	2.6	77.3	-4.0	0.0	30.4	100.0	-3.5	0.0	44.6	100.0	10.7	99.9	17.3	100.0
6	glu	0.7	-4.1	0.0	1.1	0.0	-6.8	0.0	2.0	12.4	-6.7	0.0	1.0	0.0	-3.7	0.0
7	his	93.7	9.4	99.9	11.4	99.9	35.2	100.0	13.0	100.0	9.9	99.7	11.9	100.0	7.9	98.7
8	leu	87.9	5.0	96.4	33.7	100.0	19.5	100.0	-9.6	0.0	28.4	100.0	7.6	98.3	5.4	90.0
9	N-helicene	74.0	3.9	91.9	2.5	30.1	8.0	98.7	6.7	96.4	5.0	87.3	– [a]	– [a]	– [a]	– [a]
10	pro	44.4	-0.7	0.0	6.4	95.5	-0.5	0.0	0.9	0.0	0.2	0.0	-1.6	0.0	-1.1	0.0
11	ser	100.0	16.4	100.0	3.3	55.2	39.4	100.0	-1.5	0.0	40.0	100.0	16.5	100.0	17.9	100.0
12	tetrazole	100.0	12.2	100.0	20.1	100.0	12.3	100.0	19.3	100.0	10.8	99.8	– [a]	– [a]	– [a]	– [a]
13	tyr	51.7	2.6	77.5	-0.9	0.0	19.2	100.0	4.2	76.9	4.9	85.5	6.2	94.8	2.9	43.2
14	val	44.1	2.5	76.2	7.6	98.2	6.2	94.4	5.0	87.1	6.9	97.2	4.7	83.3	2.2	19.6

[a] N-helicene and tetrazole encountered similar SCF convergence problems as anthracene and naphthalene, and have thus been excluded from the statistics for PBEh-3c.[193]

Table A.10: POWDIFF structure measures[293,294] of EE14 crystal geometries generated by low-cost methods implemented in the SIESTA and CRYSTAL17 codes. Structures are referenced to B86bPBE-XDM equilibrium structures. L: enantiopure. DL: racemate.

no	structure	PBE-D2	vdW-DF1	vdW-DF2	DFTB3-D3	HF-D2	sHF-3c	PBEh-3c
1	ala (DL)	0.0035	0.3335	0.1897	0.5167	0.0064	0.4642	0.1123
2	ala (L)	0.0158	0.3091	0.1520	0.3861	0.2853	0.3203	0.0815
3	aldol (DL)	0.0192	0.1262	0.0371	0.4003	0.2134	0.1122	0.1053
4	aldol (L)	0.0237	0.1459	0.0402	0.3319	0.1719	0.1053	0.2057
5	asp (DL)	0.0287	0.3582	0.1377	0.6584	0.6479	0.4214	0.3457
6	asp (L)	0.0060	0.1933	0.0667	0.2851	0.6017	0.3043	0.1707
7	oxazoline (DL)	0.0035	0.0539	0.0714	0.4036	0.3180	0.0041	0.0654
8	oxazoline (L)	0.0086	0.1421	0.0905	0.4504	0.3591	0.0292	0.1695
9	cys (DL)	0.0283	0.2994	0.2537	0.2428	0.4296	0.2509	0.0539
10	cys (L)	0.0875	0.5175	0.3156	0.2720	0.4695	0.4904	0.6789
11	glu (DL)	0.0104	0.1883	0.0981	0.2415	0.1470	0.1281	0.1364
12	glu (L)	0.0099	0.2074	0.0718	0.3935	0.4154	0.2622	0.1678
13	his (DL)	0.0105	0.1629	0.1913	0.5233	0.4654	0.3839	0.5556
14	his (L)	0.0037	0.2309	0.0871	0.4519	0.3722	0.2549	0.0553
15	leu (DL)	0.0020	0.0350	0.0115	0.4650	0.1826	0.0474	0.0185
16	leu (L)	0.0011	0.0430	0.0156	0.4565	0.6150	0.0627	0.0089
17	N-helicene (DL)	0.0340	0.1264	0.1019	0.2047	0.0628	0.1244	– [a]
18	N-helicene (L)	0.0440	0.1699	0.1258	0.3514	0.2300	0.1954	– [a]
19	pro (DL)	0.0012	0.1361	0.0956	0.3834	0.2315	0.0414	0.0483
20	pro (L)	0.0094	0.1876	0.1286	0.4072	0.1750	0.1167	0.0497
21	ser (DL)	0.0022	0.2486	0.1146	0.1977	0.3648	0.3504	0.1410
22	ser (L)	0.0018	0.3283	0.1249	0.3920	0.6559	0.1123	0.1410
23	tetrazole (DL)	0.0335	0.1387	0.1101	0.3633	0.1641	0.1936	– [a]
24	tetrazole (L)	0.0559	0.1844	0.0955	0.2350	0.1723	0.1539	– [a]
25	tyr (DL)	0.0097	0.2005	0.1577	0.2008	0.3427	0.3438	0.0145
26	tyr (L)	0.0103	0.1766	0.1513	0.1636	0.1208	0.2848	0.0127
27	val (DL)	0.0064	0.0893	0.0608	0.3451	0.1004	0.0453	0.0377
28	val (L)	0.0013	0.0703	0.0172	0.3866	0.0987	0.0274	0.0098

[a] N-helicene and tetrazole encountered similar SCF convergence problems as anthracene and naphthalene, and have thus been excluded from the statistics for PBEh-3c.[193]

Table A.11: Relative energy differences, in kJ/mol per molecule, and unit-cell densities, ρ, in g/cm^3, of the lowest-energy enantiopure (L) and racemate (DL) crystal structures of 1-aza[6]helicene generated by the W99 force field from a previously conducted CSP study[2] (FF; W99), and fully relaxed/reranked using B86bPBE-XDM, as implemented in Quantum ESPRESSO (DFT; QE), or PBE-D2/DZP (SIESTA) starting from the W99 geometries (FF; A) and B86bPBE-XDM geometries (DFT; A). Single-point energy calculations performed on PBE-D2/DZP geometries using B86bPBE-XDM are also tabulated (B//A). The experimentally-observed crystal structures are indicated by "(Expt.)".

no	structure	FF					DFT				
		W99	ρ	A	B//A	ρ	QE	ρ	A	B//A	ρ
1	L (Expt.)	0.0	1.2808	0.0	0.0	1.2969	0.0	1.3713	0.0	0.0	1.3216
2	DL	0.8	1.2935	3.2	4.2	1.3014	3.6	1.3636	3.8	3.6	1.3186
3	DL	0.9	1.2870	2.4	3.4	1.2967	3.5	1.3612	3.4	3.5	1.3159
4	DL	1.0	1.2865	2.7	3.7	1.2956	3.7	1.3613	3.4	3.7	1.3159
5	DL	1.8	1.2611	3.5	6.6	1.2412	5.7	1.3301	3.8	5.7	1.2858
6	DL	1.9	1.2900	2.9	2.8	1.3021	3.4	1.3641	3.7	3.4	1.3193
7	L	4.5	1.2399	-0.8	4.6	1.2485	4.1	1.3140	-0.5	4.1	1.2707
8	DL	2.1	1.2695	0.9	1.2	1.2802	-0.9	1.3719	-0.0	-0.9	1.3243
9	DL	3.2	1.2882	4.4	5.0	1.2997	5.1	1.3619	5.2	5.1	1.3177
10	DL	3.6	1.2735	3.0	1.1	1.3099	1.1	1.3774	3.5	1.1	1.3233
11	DL	4.1	1.2385	2.6	8.6	1.2444	8.2	1.3068	3.6	8.2	1.2644
12	DL	4.3	1.2660	-1.5	-0.1	1.2970	-0.6	1.3632	-0.6	-0.6	1.3179
13	L	4.3	1.2682	2.1	2.9	1.2838	2.7	1.3641	2.5	2.7	1.3189
14	DL	4.3	1.2592	6.8	7.3	1.2534	1.6	1.3501	3.3	1.6	1.3060
15	DL	4.6	1.2615	0.2	1.8	1.3003	-1.2	1.3884	-2.8	-1.2	1.3419
16	DL	4.6	1.2953	7.3	1.0	1.3144	1.3	1.3859	7.8	1.3	1.3261
17	DL	5.0	1.2743	4.7	3.1	1.2859	3.2	1.3637	5.0	3.2	1.3180
18	DL	5.0	1.2437	9.7	12.3	1.2269	9.7	1.3060	9.9	9.7	1.2634
19	DL	5.3	1.2633	7.4	9.0	1.2475	7.1	1.3349	7.6	7.1	1.2897
20	DL	5.4	1.2644	7.1	8.0	1.2599	4.6	1.3547	4.9	4.6	1.3092
21	DL	5.5	1.2889	5.3	0.9	1.3086	1.0	1.3863	6.4	1.0	1.3292
22	DL	5.6	1.2682	8.7	9.0	1.2545	6.3	1.3651	8.9	6.3	1.3046
23	DL	5.6	1.2887	5.9	0.8	1.3106	0.8	1.3863	5.9	0.8	1.3210
24	DL	5.6	1.2850	3.9	1.3	1.3010	1.3	1.3736	4.1	1.3	1.3208
25	DL	5.8	1.2656	2.9	2.2	1.2950	0.5	1.3783	2.0	0.5	1.3172
26	DL	5.8	1.2779	1.4	2.2	1.3089	0.9	1.3788	1.8	0.9	1.3330
27	DL	5.8	1.2457	2.7	8.5	1.2474	8.0	1.3216	3.2	8.0	1.2673
28	DL	5.9	1.2550	0.6	-0.1	1.2891	-1.7	1.3731	0.5	-1.7	1.3180
29	DL	6.0	1.2937	12.9	12.8	1.2624	8.3	1.3730	11.0	8.3	1.3182
30	DL	6.3	1.2576	7.3	5.7	1.2428	4.2	1.3461	11.0	4.2	1.3018
31	DL	6.3	1.2262	0.7	6.3	1.2210	2.2	1.3355	-1.1	2.2	1.2913
32	DL	6.3	1.2801	8.2	6.2	1.2799	3.5	1.3770	8.2	3.5	1.3200
33	DL	6.5	1.2888	8.8	6.8	1.2604	4.6	1.3584	8.3	4.6	1.3035
34	DL	6.6	1.2725	11.1	9.0	1.2721	8.8	1.3434	11.2	8.8	1.2832
35	DL	6.9	1.2521	-0.3	3.1	1.2798	1.7	1.3598	-2.4	1.7	1.3069
36	DL	7.0	1.2521	8.8	5.2	1.2893	2.0	1.3837	6.3	2.0	1.3278
37	DL	7.2	1.1636	23.0	29.2	1.1738	28.8	1.2402	23.6	28.8	1.1999
38	DL	7.2	1.2801	8.5	5.2	1.2863	4.8	1.3629	8.4	4.8	1.3082
39	DL	7.3	1.2395	4.1	9.0	1.2586	7.9	1.3233	4.3	7.9	1.2806
40	DL	7.4	1.2944	7.0	0.8	1.3276	0.1	1.3945	7.2	0.1	1.3333
41	DL	7.4	1.2542	2.5	3.4	1.2843	1.5	1.3577	5.4	1.5	1.3056
42	DL	7.5	1.2598	1.7	2.9	1.2927	3.5	1.3598	2.5	3.5	1.3150
43	DL	7.6	1.2307	6.6	10.9	1.2434	7.4	1.3269	4.8	7.4	1.2835
44	DL	7.6	1.2840	12.6	6.6	1.2852	4.3	1.3590	11.8	4.3	1.3002
45	DL	5.6	1.2800	12.3	7.4	1.2767	6.0	1.3636	13.5	6.0	1.3028
46	DL	7.6	1.2837	9.1	4.9	1.3104	4.8	1.3777	9.6	4.8	1.3162
47	DL	7.7	1.2570	12.6	14.7	1.2366	13.1	1.3220	12.7	13.1	1.2783
48	DL	7.8	1.2701	4.9	5.6	1.2855	3.8	1.3693	4.6	3.8	1.3179
49	DL	7.8	1.2105	5.4	12.7	1.1804	7.9	1.2858	4.1	7.9	1.2435
50	DL (Expt.)	8.2	1.2220	-6.0	-1.9	1.2462	-3.8	1.3317	-7.6	-3.8	1.2870

Table A.12: Relative energy differences, in kJ/mol per molecule, and unit-cell densities, ρ, in g/cm^3, of the lowest-energy enantiopure (L) and racemate (DL) crystal structures of 1-aza-helicene generated by the W99 force field from a previously conducted CSP study[2] (FF; W99), and fully relaxed/reranked using B86bPBE-XDM, as implemented in Quantum ESPRESSO (DFT; QE), or sHF-3c starting from the W99 geometries (FF; A) and B86bPBE-XDM geometries (DFT; A). Single-point energy calculations performed on sHF-3c geometries using B86bPBE-XDM are also tabulated (B//A). The experimentally-observed crystal structures are indicated by "(Expt.)".

no	structure	FF					DFT				
		W99	ρ	A	B//A	ρ	QE	ρ	A	B//A	ρ
1	L (Expt.)	0.0	1.2808	0.0	0.0	1.4252	0.0	1.3713	0.0	0.0	1.4276
2	DL	0.8	1.2935	3.0	2.0	1.4196	3.6	1.3636	3.0	1.2	1.4183
3	DL	0.9	1.2870	1.4	2.0	1.4183	3.5	1.3612	1.2	1.6	1.4169
4	DL	1.0	1.2865	1.3	2.0	1.4186	3.7	1.3613	1.2	1.5	1.4176
5	DL	1.8	1.2611	-0.9	3.6	1.3999	5.7	1.3301	-1.0	3.0	1.3996
6	DL	1.9	1.2900	2.2	2.2	1.4233	3.4	1.3641	2.3	1.7	1.4236
7	L	4.5	1.2399	-1.5	1.1	1.3948	4.1	1.3140	-1.5	0.7	1.3948
8	DL	2.1	1.2695	-3.6	-1.6	1.4369	-0.9	1.3719	-2.8	-3.2	1.4371
9	DL	3.2	1.2882	-0.4	3.9	1.4450	5.1	1.3619	-0.5	3.2	1.4446
10	DL	3.6	1.2735	-2.3	1.6	1.4360	1.1	1.3774	-2.3	1.8	1.4436
11	DL	4.1	1.2385	4.0	7.4	1.3723	8.2	1.3068	4.0	7.1	1.3741
12	DL	4.3	1.2660	-6.5	-0.8	1.4339	-0.6	1.3632	-6.6	-0.9	1.4369
13	L	4.3	1.2682	0.5	1.7	1.4236	2.7	1.3641	0.9	1.0	1.4223
14	DL	4.3	1.2592	-3.0	-2.2	1.4375	1.6	1.3501	-2.7	-3.3	1.4362
15	DL	4.6	1.2615	0.6	-2.2	1.4238	-1.2	1.3884	0.6	-2.8	1.4303
16	DL	4.6	1.2953	-3.6	-0.1	1.4645	1.3	1.3859	-3.3	-0.8	1.4640
17	DL	5.0	1.2743	-1.0	3.1	1.4390	3.2	1.3637	-1.0	2.4	1.4380
18	DL	5.0	1.2437	4.6	8.8	1.3748	9.7	1.3060	4.6	8.2	1.3744
19	DL	5.3	1.2633	2.0	6.8	1.4039	7.1	1.3349	2.0	6.4	1.4072
20	DL	5.4	1.2644	-0.1	4.1	1.4146	4.6	1.3547	-0.1	3.4	1.4129
21	DL	5.5	1.2889	-5.9	0.5	1.4551	1.0	1.3863	-5.9	0.3	1.4587
22	DL	5.6	1.2682	-0.0	6.3	1.4369	6.3	1.3651	-0.1	5.5	1.4351
23	DL	5.6	1.2887	-5.9	0.6	1.4546	0.8	1.3863	-5.8	0.5	1.4584
24	DL	5.6	1.2850	-4.6	0.0	1.4419	1.3	1.3736	-4.8	-0.2	1.4470
25	DL	5.8	1.2656	-5.5	0.6	1.4332	0.5	1.3783	-5.7	-0.4	1.4415
26	DL	5.8	1.2779	-2.1	1.6	1.4454	0.9	1.3788	-2.0	0.8	1.4451
27	DL	5.8	1.2457	3.5	7.3	1.4029	8.0	1.3216	3.4	7.1	1.4059
28	DL	5.9	1.2550	-8.3	-1.7	1.4447	-1.7	1.3731	-8.3	-2.6	1.4434
29	DL	6.0	1.2937	3.1	7.3	1.4471	8.3	1.3730	2.8	7.2	1.4514
30	DL	6.3	1.2576	-2.5	3.8	1.4220	4.2	1.3461	-2.5	3.3	1.4224
31	DL	6.3	1.2262	-0.2	0.2	1.3949	2.2	1.3355	-0.2	-0.4	1.3910
32	DL	6.3	1.2801	-1.8	2.0	1.4303	3.5	1.3770	-1.4	1.8	1.4499
33	DL	6.5	1.2888	-2.5	5.5	1.4354	4.6	1.3584	-2.6	5.0	1.4357
34	DL	6.6	1.2725	-3.9	2.6	1.4378	8.8	1.3434	-4.2	2.1	1.4386
35	DL	6.9	1.2521	-4.0	1.6	1.4349	1.7	1.3598	-4.0	0.7	1.4345
36	DL	7.0	1.2521	-7.8	-4.8	1.4663	2.0	1.3837	-7.9	-5.6	1.4660
37	DL	7.2	1.1636	17.4	20.3	1.3101	28.8	1.2402	17.2	19.7	1.3105
38	DL	7.2	1.2801	-0.5	4.3	1.4358	4.8	1.3629	-0.6	3.8	1.4358
39	DL	7.3	1.2395	2.5	6.8	1.3906	7.9	1.3233	2.4	6.5	1.3921
40	DL	7.4	1.2944	-5.5	-0.4	1.4557	0.1	1.3945	-6.5	-1.2	1.4734
41	DL	7.4	1.2542	-3.7	1.9	1.4323	1.5	1.3577	-4.0	1.1	1.4351
42	DL	7.5	1.2598	1.1	2.3	1.4252	3.5	1.3598	1.0	1.5	1.4221
43	DL	7.6	1.2307	3.6	7.2	1.3917	7.4	1.3269	3.6	6.9	1.3948
44	DL	7.6	1.2840	-5.8	1.4	1.4456	4.3	1.3590	-5.8	0.4	1.4435
45	DL	5.6	1.2800	0.5	5.8	1.4224	6.0	1.3636	0.3	5.5	1.4242
46	DL	7.6	1.2837	-3.1	4.1	1.4467	4.8	1.3777	-3.2	3.9	1.4500
47	DL	7.7	1.2570	3.9	10.5	1.4149	13.1	1.3220	3.4	9.5	1.4147
48	DL	7.8	1.2701	-1.5	3.2	1.4327	3.8	1.3693	-1.8	2.6	1.4350
49	DL	7.8	1.2105	4.8	7.7	1.3401	7.9	1.2858	4.6	7.3	1.3410
50	DL (Expt.)	8.2	1.2220	-6.6	-4.8	1.4020	-3.8	1.3317	-6.8	-4.9	1.4043

APPENDIX B

SUPPORTING INFORMATION FOR CHAPTER 6

Contents: Computed and benchmark binding energies for the KB49 set of molecular dimers and lattice energies for the X23 set crystal structures.

Table B.1: Binding energies, in kJ/mol, of the KB49 benchmark set of van-der-Waals bound molecular dimers calculated with the SIESTA and Quantum ESPRESSO codes, using the PBE functional. Reference binding energies[162] (REF) are also presented. DZP or TZP: a double-zeta (by default) or triple-zeta plus polarization basis set was used (SIESTA). PAW: a plane-wave basis set was used (Quantum ESPRESSO). CP: a counterpoise correction was applied. Negative values indicate that dimers are not bound.

no	structure	REF	PBE	PBE-D2			PBE-XDM					
			DZP	DZP	TZP	PAW	DZP	DZP	DZP+CP	TZP	TZP+CP	PAW
1	ch4_c2h4	2.1	2.0	3.3	3.7	3.0	3.3	3.5	3.6	3.2	3.6	2.9
2	cf4_cf4	3.6	5.0	8.4		5.1	6.4					3.5
3	sih4_ch4	3.3	3.6	5.9		4.4	5.3					4.3
4	ocs_ocs	7.2	3.8	6.6		5.3	6.8					6.4
5	c10h8_c10h8_p	17.0	-4.0	11.3	22.9	13.4	13.1	14.9	12.7	17.6	13.6	16.6
6	c10h8_c10h8_pc	24.0	-1.5	17.0	30.4	19.5	16.4	18.8	16.4	21.8	17.4	21.9
7	c10h8_c10h8_t	21.6	5.4	18.0	25.3	21.0	19.6	19.9	18.9	20.0	18.0	18.2
8	c10h8_c10h8_tc	16.5	4.1	13.1	18.4	14.7	17.0	16.7	15.6	16.7	15.3	14.1
9	sih4_hf	3.2	5.2	6.4		4.0	5.4					3.3
10	ch4_nh3	3.1	5.1	6.1	5.0	4.4	5.7	6.0	4.9	4.3	4.5	4.1
11	ch4_hf	6.7	7.2	9.3		11.4	7.6					8.0
12	c2h4_hf	18.7	23.5	25.6		24.8	24.1					22.4
13	ch3f_ch3f	10.0	10.9	13.5		11.3	12.1					9.1
14	h2co_h2co	14.3	13.6	16.0	12.9	14.2	14.7	15.9	12.3	10.9	12.9	12.1
15	ch3cn_ch3cn	25.9	24.3	28.5	27.8	27.7	27.0	27.5	24.9	24.4	24.8	25.7
16	hcn_hf	31.0	41.9	42.6		33.7	42.3					33.0
17	co2_co2	6.0	7.2	9.1	7.8	5.8	8.4	9.4	5.3	6.4	6.0	5.1
18	nh3_nh3	13.1	13.7	15.4	14.9	16.1	14.4	15.0	13.6	13.5	14.2	13.7
19	h2o_h2o	20.9	23.5	24.8	25.2	23.3	23.9	24.7	21.9	24.1	23.6	22.2
20	h2co2_h2co2	78.5	93.2	96.8	87.7	84.8	95.2	97.8	86.0	85.0	85.9	81.8
21	formamide_formamide	67.2	74.9	79.0	71.5	70.2	77.3	79.1	69.8	68.4	69.7	67.3
22	uracil_uracil_hb	86.4	90.8	96.2	88.4	88.2	95.4	99.4	87.4	86.0	88.0	85.0
23	pyridoxine_aminopyridine	70.9	77.7	84.5	77.8	77.0	87.2	88.1	77.3	75.9	77.0	73.2
24	adenine_thymine_wcc1	69.7	76.2	83.6	74.4	73.5	84.6	88.2	71.2	71.8	71.0	69.5
25	ch4_ch4	2.2	2.2	3.9	3.6	3.0	3.3	3.7	2.9	2.3	3.0	2.6
26	c2h4_c2h4	6.2	4.1	8.5	9.9	8.1	6.4	6.7	5.9	6.1	5.6	6.1
27	c6h6_ch4	6.1	2.6	7.0	9.6	7.1	5.4	6.0	3.0	6.4	3.1	5.7
28	c6h6_c6h6_pd	11.1	-3.5	7.9	15.7	10.2	8.7	9.3	7.9	10.0	8.4	9.8
29	pyrazine_pyrazine	17.8	4.0	15.9	20.9	16.7	11.8	14.0	8.7	12.2	9.1	14.1
30	uracil_uracil_stack	41.0	24.5	41.4	42.6	39.7	34.4	41.7	36.3	30.8	36.3	33.9
31	indole_c6h6_stack	18.9	-2.3	14.5	24.3	17.7	10.4	13.2	11.6	13.5	12.2	15.5
32	adenine_thymine_stack	49.1	24.3	48.9	53.0	47.4	39.8	46.7	36.4	35.3	36.7	38.1
33	c2h4_c2h2	6.3	6.2	8.1	8.4	8.1	7.8	7.8	7.9	7.2	7.5	7.5
34	c6h6_h2o	13.7	11.2	15.3	18.2	15.8	13.1	14.9	11.5	15.0	12.0	13.3
35	c6h6_nh3	9.7	6.2	10.5	13.1	11.1	8.4	9.7	6.6	9.7	6.9	9.3
36	c6h6_hcn	19.0	14.2	19.2	21.9	21.6	17.0	18.5	16.7	18.1	15.9	18.0
37	c6h6_c6h6_t	11.4	3.4	10.2	14.4	12.0	13.8	12.7	11.4	12.5	10.9	10.1
38	indole_c6h6_t	23.5	14.6	23.6	26.2	25.3	23.4	23.5	21.0	21.1	19.9	21.0
39	phenol_phenol	29.7	24.0	31.7	32.0	29.4	32.5	33.2	29.5	28.5	29.6	26.5
40	hf_hf	19.1	24.2	24.9		21.5	24.3					20.4
41	nh3_h2o	26.8	32.0	33.5	34.1	31.5	32.6	33.3	29.5	32.7	31.8	30.3
42	h2s_h2s	6.9	9.0	10.2		9.1	11.1					9.6
43	hcl_hcl	8.4	10.3	11.3		10.4	11.6					10.5
44	h2s_hcl	14.0	19.6	20.8		18.8	21.3					19.3
45	ch3cl_hcl	14.9	17.5	19.7		18.0	19.2					17.4
46	hcn_ch3sh	15.0	18.2	20.5		18.6	20.3					17.9
47	ch3sh_hcl	20.4	27.1	29.7		27.4	29.3					27.0
48	c2h2_c2h2	5.6	5.5	7.4	7.2	7.0	7.3	7.2	7.2	6.0	6.4	6.8
49	c6h6_c6h6_stack	7.0	-9.3	-2.0	10.5	5.9	2.0	1.8	7.0	7.6	7.7	7.8

Table B.2: Binding energies, in kJ/mol, of the KB49 set of van-der-Waals bound molecular dimers benchmark calculated with the SIESTA and Quantum ESPRESSO codes, using the B86bPBE functional. Reference binding energies[162] (REF) are also presented. DZP or TZP: a double-zeta (by default) or triple-zeta plus polarization basis set was used (SIESTA). PAW: a plane-wave basis set was used (Quantum ESPRESSO). CP: a counterpoise correction was applied. Negative values indicate that dimers are not bound.

no	structure	REF	B86bPBE	B86bPBE-XDM					
			DZ	DZP	DZP	DZP+CP	TZP	TZP+CP	PAW
1	ch4_c2h4	2.1	1.2	2.8	2.9	3.3	2.9	3.3	2.6
2	cf4_cf4	3.6	3.1	5.3					3.2
3	sih4_ch4	3.3	2.2	4.2					3.8
4	ocs_ocs	7.2	2.0	5.7					6.1
5	c10h8_c10h8_p	17.0	-7.6	14.3	15.3	12.4	19.4	13.6	16.4
6	c10h8_c10h8_pc	24.0	-5.4	17.7	18.9	16.3	23.8	17.4	22.6
7	c10h8_c10h8_t	21.6	2.1	20.4	21.1	19.2	20.1	18.5	18.4
8	c10h8_c10h8_tc	16.5	1.6	18.0	18.4	15.8	16.8	15.6	13.5
9	sih4_hf	3.2	3.9	4.3					2.7
10	ch4_nh3	3.1	4.2	5.0	5.1	4.3	3.8	3.9	3.8
11	ch4_hf	6.7	5.7	6.4					7.6
12	c2h4_hf	18.7	21.9	22.9					22.3
13	ch3f_ch3f	10.0	9.3	11.0					8.8
14	h2co_h2co	14.3	11.7	13.4	13.7	11.5	10.1	12.1	12.4
15	ch3cn_ch3cn	25.9	22.0	25.9	26.2	24.3	23.1	24.1	25.7
16	hcn_hf	31.0	40.1	40.8					32.8
17	co2_co2	6.0	5.7	7.5	7.7	4.8	6.0	5.5	5.1
18	nh3_nh3	13.1	12.2	13.3	13.5	12.8	12.6	13.5	13.5
19	h2o_h2o	20.9	21.9	22.6	22.8	21.0	23.2	22.8	22.0
20	h2co2_h2co2	78.5	89.9	93.0	93.7	85.2	84.6	85.4	82.8
21	formamide_formamide	67.2	72.0	75.6	76.2	69.0	67.5	69.1	68.0
22	uracil_uracil_hb	86.4	87.4	94.0	95.0	86.0	86.4	86.8	85.6
23	pyridoxine_aminopyridine	70.9	74.3	87.6	88.3	76.8	75.6	76.5	73.9
24	adenine_thymine_wcc1	69.7	72.7	85.0	86.3	69.2	72.9	69.3	70.4
25	ch4_ch4	2.2	1.0	2.6	2.8	2.5	2.0	2.6	2.4
26	c2h4_c2h4	6.2	2.3	5.4	5.7	5.3	5.1	5.0	6.1
27	c6h6_ch4	6.1	1.0	4.7	4.9	0.8	5.9	1.1	5.8
28	c6h6_c6h6_pd	11.1	-6.3	10.5	11.3	8.0	10.9	8.6	10.9
29	pyrazine_pyrazine	17.8	0.9	11.7	12.6	7.2	12.4	7.7	15.5
30	uracil_uracil_stack	41.0	19.8	34.2	36.2	37.3	33.4	37.6	36.2
31	indole_c6h6_stack	18.9	-6.0	11.2	12.5	11.4	14.9	12.3	17.7
32	adenine_thymine_stack	49.1	18.3	40.3	42.6	37.5	37.9	38.1	42.1
33	c2h4_c2h2	6.3	5.1	7.3	7.4	7.5	6.5	7.1	7.3
34	c6h6_h2o	13.7	9.6	12.3	12.8	9.3	15.0	10.0	13.8
35	c6h6_nh3	9.7	4.6	7.6	8.0	4.2	9.5	4.8	9.4
36	c6h6_hcn	19.0	12.2	16.2	16.7	14.7	17.6	14.2	18.3
37	c6h6_c6h6_t	11.4	1.3	15.4	15.7	11.4	12.3	10.9	10.0
38	indole_c6h6_t	23.5	11.7	23.6	24.1	20.7	20.5	19.7	21.3
39	phenol_phenol	29.7	21.1	32.6	33.2	30.0	28.1	30.2	27.1
40	hf_hf	19.1	26.7	23.0					20.0
41	nh3_h2o	26.8	30.0	31.3	31.4	28.6	31.7	30.9	30.2
42	h2s_h2s	6.9	7.5	10.2					9.2
43	hcl_hcl	8.4	8.7	10.7					10.2
44	h2s_hcl	14.0	17.5	20.2					19.0
45	ch3cl_hcl	14.9	15.0	17.8					17.3
46	hcn_ch3sh	15.0	15.9	19.4					17.9
47	ch3sh_hcl	20.4	24.2	27.8					27.0
48	c2h2_c2h2	5.6	4.2	6.8	6.9	6.9	5.2	6.1	6.6
49	c6h6_c6h6_stack	7.0	-4.1	4.0	4.4	6.8	8.5	7.5	7.2

Table B.3: Lattice energies, in kJ/mol per molecule, of the X23 benchmark set of small molecular organic solids calculated with the SIESTA and Quantum ESPRESSO codes, using the PBE functional. Reference lattice energies back-corrected from experimentally-measured sublimation enthalpies[145] (Expt.) are also presented. DZP or TZP: a double-zeta (by default) or triple-zeta plus polarization basis set was used (SIESTA). PAW: a plane-wave basis set was used (Quantum ESPRESSO). CP: a counterpoise correction was applied. B//A: plane-wave single-point energy calculations with the same dispersion-corrected DFT functional were performed on the SIESTA/DZP equilibrium geometries (composite approach).

no	structure	Expt.	PBE-D2				PBE-XDM					
			DZP	TZP	PAW	B//A	DZP	TZP	PAW	B//A	DZP+CP	TZP+CP
1	14-cyclohexanedione	88.6	100.6	92.7	96.1	94.3	105.6	75.7	85.5	87.0	87.0	87.6
2	acetic acid	72.8	85.8	76.4	76.1	76.2	88.8	67.6	72.9	73.5	74.0	74.6
3	adamantane	69.4	69.2	83.9	87.8	82.4	73.4	59.1	70.1	71.0	68.6	67.6
4	ammonia	37.2	46.8	45.6	45.9	45.8	46.1	38.8	40.3	40.3	41.0	41.5
5	anthracene	112.7	87.6	116.0	106.1	101.6	97.3	89.9	97.1	98.6	98.3	94.5
6	benzene	51.7	47.7	60.0	56.3	53.7	50.2	45.3	50.1	50.7	50.0	48.1
7	carbon dioxide	28.4	29.8	21.3	23.9	24.7	32.5	17.9	23.7	24.3	20.2	23.0
8	cyanamide	79.7	100.7	93.1	92.7	93.4	102.4	85.2	89.7	90.5	88.0	88.6
9	cytosine	169.8	173.3	165.0	162.6	162.5	177.7	146.8	151.8	152.6	157.6	154.1
10	ethylcarbamate	86.3	101.0	90.6	91.4	91.2	103.9	77.7	84.9	86.2	85.7	86.4
11	formamide	79.2	93.4	83.2	82.9	83.1	95.3	75.7	79.1	79.0	82.3	80.6
12	hexamine	86.2	93.6	92.7	97.4	95.3	100.5	71.5	81.7	82.6	90.7	86.4
13	imidazole	86.8	98.6	95.6	94.4	93.8	101.4	84.5	88.8	89.3	91.5	89.4
14	naphthalene	81.7	66.3	86.5	80.5	76.8	72.5	66.4	73.0	74.0	73.0	69.8
15	oxalic acid α	96.3	107.9	92.0	94.6	95.5	115.3	81.9	90.3	90.5	90.5	94.0
16	oxalic acid β	96.1	111.6	96.9	98.0	98.7	119.4	86.8	92.9	93.4	96.9	97.3
17	pyrazine	61.3	66.2	65.9	64.3	64.0	70.1	54.0	60.0	60.6	66.3	59.6
18	pyrazole	77.7	86.2	84.8	83.6	82.5	88.7	73.7	75.7	77.7	81.1	77.7
19	succinic acid	130.3	153.9	135.1	136.8	136.5	160.2	120.1	128.6	129.2	135.0	133.4
20	triazine	61.7	66.8	60.4	60.8	61.0	70.8	48.3	55.5	56.2	57.4	54.3
21	trioxane	66.4	71.9	63.1	65.0	64.9	77.8	50.5	57.5	58.1	64.7	63.1
22	uracil	135.7	151.5	139.0	136.3	136.4	158.6	125.6	130.6	131.5	135.9	133.5
23	urea	102.5	120.7	109.1	110.7	111.2	121.9	99.2	103.5	104.2	103.6	105.7

Table B.4: Lattice energies, in kJ/mol per molecule, of the X23 benchmark set of small molecular organic solids calculated with the SIESTA and Quantum ESPRESSO codes, using the B86bPBE functional. Reference lattice energies back-corrected from experimentally-measured sublimation enthalpies[145] (Expt.) are also presented. DZP or TZP: a double-zeta (by default) or triple-zeta plus polarization basis set was used (SIESTA). PAW: a plane-wave basis set was used (Quantum ESPRESSO). CP: a counterpoise correction was applied. B//A: plane-wave single-point energy calculations with the same dispersion-corrected DFT functional were performed on the SIESTA/DZP equilibrium geometries (composite approach).

no	structure	Expt.	B86bPBE-XDM					
			DZP	TZP	PAW	B//A	DZP+CP	TZP+CP
1	14-cyclohexanedione	88.6	95.6	77.5	88.6	89.6	85.8	86.2
2	acetic acid	72.8	80.5	67.4	74.4	75.0	71.2	71.5
3	adamantane	69.4	63.8	60.7	72.0	71.4	65.8	65.4
4	ammonia	37.2	41.6	38.2	41.4	41.5	39.3	39.9
5	anthracene	112.7	88.4	94.6	101.7	101.9	96.9	95.2
6	benzene	51.7	45.1	45.9	51.8	51.4	49.0	47.5
7	carbon dioxide	28.4	25.9	17.2	24.2	25.2	18.0	20.5
8	cyanamide	79.7	95.9	84.4	90.9	91.7	85.4	86.4
9	cytosine	169.8	166.7	148.5	156.4	157.2	155.1	152.9
10	ethylcarbamate	86.3	94.2	78.0	86.7	87.9	83.2	83.3
11	formamide	79.2	88.1	75.3	81.1	81.4	79.9	78.5
12	hexamine	86.2	89.1	74.8	85.5	85.6	89.4	87.8
13	imidazole	86.8	94.6	84.5	90.4	90.5	90.2	87.4
14	naphthalene	81.7	65.6	69.0	76.4	76.2	71.5	70.2
15	oxalic acid α	96.3	104.2	84.1	94.8	95.5	89.0	92.0
16	oxalic acid β	96.1	107.2	88.2	97.1	97.7	93.5	94.9
17	pyrazine	61.3	63.6	55.1	62.2	62.5	60.4	59.1
18	pyrazole	77.7	82.2	73.8	78.8	78.9	79.1	76.8
19	succinic acid	130.3	145.2	120.7	133.0	133.9	130.9	130.4
20	triazine	61.7	63.5	49.3	57.4	58.1	55.7	53.7
21	trioxane	66.4	68.0	52.1	60.7	61.1	63.1	62.1
22	uracil	135.7	147.1	127.0	134.0	135.1	132.1	130.9
23	urea	102.5	113.9	99.5	106.4	107.0	101.4	103.6

Table B.5: POWDIFF structure measures[293,294] of X23 crystal geometries generated by low-cost methods referred to plane-wave DFT-XDM or DFT-D2 equilibrium structures. Results are shown only for the DFT-D/PAW//SIESTA/DZP composite approach.

no	structure	PBE-D2	PBE-XDM	B86bPBE-XDM
1	14-cyclohexanedione	0.3945	0.0034	0.2733
2	acetic acid	0.0803	0.0762	0.0822
3	adamantane	0.4550	0.0323	0.2210
4	ammonia	0.2266	0.1301	0.1190
5	anthracene	0.5828	0.1696	0.4217
6	benzene	0.6341	0.1793	0.4509
7	carbon dioxide	0.0003	0.6713	0.0125
8	cyanamide	0.1014	0.1728	0.1956
9	cytosine	0.2061	0.0217	0.1872
10	ethylcarbamate	0.3236	0.1138	0.1283
11	formamide	0.2333	0.2227	0.0526
12	hexamine	0.2919	0.0012	0.1306
13	imidazole	0.3416	0.0045	0.3007
14	naphthalene	0.5974	0.1753	0.4572
15	oxalic acid α	0.0332	0.2620	0.0116
16	oxalic acid β	0.1162	0.2031	0.2310
17	pyrazine	0.3365	0.0389	0.3491
18	pyrazole	0.4762	0.0972	0.3333
19	succinic acid	0.3827	0.2554	0.1243
20	triazine	0.2538	0.4560	0.0002
21	trioxane	0.0440	0.1859	0.0002
22	uracil	0.4497	0.1517	0.3648
23	urea	0.1899	0.0056	0.0948

APPENDIX C

SUPPORTING INFORMATION FOR CHAPTER 7

Contents: Relative energies and densities of crystal structures from the generated energy landscapes for 5-fluorouracil, naproxen, carbamazepine, and olanzapine.

Table C.1: Relative energies, in kJ/mol per molecule, and unit-cell densities, ρ, in g/cm^3, of the lowest-energy crystal structures of 5-fluorouracil (FURACL) determined by a force field from a previously conducted CSP study[336] (FF; A), and fully relaxed/reranked using B86bPBE-XDM as implemented in Quantum ESPRESSO (PAW; A), B86bPBE-XDM/DZP (SIESTA; A), and sHF-3c (CRYSTAL17; A). Single-point energy calculations performed on low-cost geometries using plane-wave B86bPBE-XDM/PAW or small-basis B86bPBE-XDM/DZP are also tabulated (B//A and C//A, respectively). Energies and densities are relative to Form-I (OPT_eo1). The other experimentally-observed crystal structure is Form-II (OPT_am75).

no	structure	FF				SIESTA			CRYSTAL17			PAW	
		A	B//A	C//A	ρ	A	B//A	ρ	A	B//A	ρ	A	ρ
1	OPT_ab13	-1.3	4.2	5.8	0.0803	7.9	7.2	0.0443	-4.5	6.2	-0.0531	6.7	0.0198
2	OPT_ab94	-1.7	7.6	9.9	0.0819	10.0	8.9	-0.0036	3.8	10.0	-0.0683	7.7	-0.0208
3	OPT_ab98	-1.4	7.4	10.5	0.0772	12.5	11.0	-0.0125	-4.5	6.3	-0.0538	11.4	-0.0122
4	OPT_ad11	0.9	10.6	12.6	0.0856	19.7	16.1	0.0335	16.8	18.2	-0.0465	16.2	0.0072
5	OPT_af51	0.1	9.8	12.4	0.1317	16.8	13.3	0.0779	8.9	16.2	-0.1632	13.1	0.0198
6	OPT_af57	0.2	7.4	8.9	0.0516	9.2	9.5	0.0069	-3.1	8.9	-0.1183	9.8	-0.0334
7	OPT_ai46	1.3	8.2	9.7	0.0733	13.3	10.8	0.0698	4.3	11.5	-0.0278	11.2	0.0395
8	OPT_ai61	1.5	10.6	15.7	0.0512	16.5	12.7	-0.0067	2.9	11.5	-0.1405	11.0	-0.0925
9	OPT_ai86	-1.2	5.8	8.0	0.0828	8.7	7.0	0.0244	2.2	7.1	-0.0150	6.6	0.0038
10	OPT_aj38	0.4	8.8	10.0	0.0467	11.0	10.9	-0.0096	1.6	12.6	-0.1457	10.6	-0.0494
11	OPT_ak24	-3.1	4.4	8.2	0.0668	9.8	7.6	0.0086	-4.5	6.3	-0.0687	6.7	-0.0329
12	OPT_ak3	-0.1	8.9	11.6	0.0866	13.8	11.4	0.0116	3.6	10.4	-0.1196	9.2	-0.0430
13	OPT_ak39	-1.2	6.8	8.0	0.0633	7.9	8.0	0.0273	-1.1	7.7	-0.0465	7.7	-0.0123
14	OPT_ak90	-2.7	4.3	7.0	0.0567	8.1	6.7	0.0209	-4.5	6.8	-0.0784	6.7	-0.0215
15	OPT_am108	-3.8	6.9	8.9	0.0771	9.7	7.7	-0.0462	-1.3	4.2	-0.0221	4.2	-0.0365
16	OPT_am115	-0.2	6.4	7.4	0.0660	12.3	10.9	0.0119	8.1	10.9	0.0388	11.5	-0.0041
17	OPT_am36	-2.2	4.1	4.2	0.0965	8.3	7.8	0.0496	7.3	8.5	0.0625	8.3	0.0390
18	OPT_am51	-2.5	3.6	6.2	0.0963	7.0	4.9	0.0323	0.4	4.6	0.0278	5.3	0.0184
19	OPT_am56	-5.8	-1.0	2.4	0.1051	4.1	2.2	0.0698	0.5	2.1	0.0678	2.7	0.0480
20	OPT_am64	-1.8	7.5	7.8	0.1210	12.2	11.4	0.0571	6.0	9.5	0.0660	11.2	0.0311
21	OPT_am69	1.0	4.2	8.0	-0.064	7.6	4.5	-0.1024	-0.9	4.0	-0.1024	4.0	-0.1134
22	OPT_am75	-5.9	-0.6	2.2	0.0940	4.3	1.9	0.0436	0.0	2.6	0.0557	2.4	0.0369
23	OPT_am76	1.2	10.0	13.1	0.1179	19.1	14.8	0.0732	5.4	10.2	-0.1605	7.6	-0.0894
24	OPT_am79	0.8	10.8	11.9	0.1371	16.1	14.0	0.0707	-0.9	4.0	-0.0997	4.0	-0.1131
25	OPT_am87	-1.2	3.9	5.6	0.0398	7.0	6.1	0.0086	0.4	4.7	0.0286	6.5	-0.0109
26	OPT_am93	-3.6	2.3	4.5	0.0662	7.1	3.7	-0.0076	-1.2	3.8	-0.0205	4.2	-0.0335
27	OPT_aq103	0.6	8.8	10.6	0.0317	12.1	11.0	-0.0258	-2.9	5.4	-0.0203	6.8	-0.0343
28	OPT_aq116	0.3	7.0	9.4	0.0858	14.2	11.0	0.0268	3.1	10.0	-0.0464	9.3	-0.0080
29	OPT_aq61	-2.4	5.1	8.0	0.0539	8.6	6.2	-0.0094	-2.8	7.8	-0.0710	6.8	-0.0358
30	OPT_aq62	1.6	9.0	12.0	0.1162	15.8	12.0	0.0797	11.2	13.3	-0.0635	11.5	0.0579
31	OPT_aq90	-2.0	5.5	7.9	0.0429	8.4	7.3	-0.0143	-2.2	8.4	-0.0926	7.0	-0.0282
32	OPT_av24	1.4	7.8	10.5	0.0631	11.8	9.0	-0.0141	8.8	8.6	-0.0591	8.8	-0.0541
33	OPT_av32	0.2	6.7	6.8	0.1050	12.2	10.5	0.0748	4.3	10.0	-0.1199	11.4	0.0550
34	OPT_av64	-0.8	5.5	7.8	0.0315	8.9	7.2	-0.0414	-2.5	8.1	-0.0840	7.9	-0.0710
35	OPT_av86	1.9	10.8	12.0	0.0917	15.5	13.3	0.0229	8.8	8.5	-0.0537	8.8	-0.0501
36	OPT_ay76	1.2	11.0	14.0	0.1107	18.2	14.2	0.0767	10.1	19.9	-0.2487	13.2	0.0302

Table C.1: Relative energies, in kJ/mol per molecule, and unit-cell densities, ρ, in g/cm^3, of the lowest-energy crystal structures of 5-fluorouracil (FURACL) determined by a force field from a previously conducted CSP study[336] (FF; A), and fully relaxed/reranked using B86bPBE-XDM as implemented in Quantum ESPRESSO (PAW; A), B86bPBE-XDM/DZP (SIESTA; A), and sHF-3c (CRYSTAL17; A). Single-point energy calculations performed on low-cost geometries using plane-wave B86bPBE-XDM/PAW or small-basis B86bPBE-XDM/DZP are also tabulated (B//A and C//A, respectively). Energies and densities are relative to Form-I (OPT_eo1). The other experimentally-observed crystal structure is Form-II (OPT_am75).

no	structure	FF				SIESTA			CRYSTAL17			PAW	
		A	B//A	C//A	ρ	A	B//A	ρ	A	B//A	ρ	A	ρ
37	OPT_az50	1.4	11.9	14.6	0.1250	18.3	14.8	0.0669	9.1	17.0	-0.1760	12.9	0.0301
38	OPT_ca13	-3.9	6.2	8.9	0.1152	9.1	8.3	0.0238	-4.5	6.2	-0.0561	8.7	0.0078
39	OPT_ca23	0.1	9.0	12.0	0.1032	15.7	12.8	0.0443	3.7	9.4	-0.0610	12.3	-0.0178
40	OPT_ca33	-3.6	2.9	6.3	0.0634	7.1	5.2	0.0314	-4.5	6.2	-0.0544	5.6	0.0119
41	OPT_cd3	0.6	9.6	11.8	0.0824	18.1	14.5	0.0398	4.3	16.2	-0.2525	13.6	0.0263
42	OPT_cd39	0.2	9.6	10.4	0.0712	13.9	12.5	0.0133	9.6	10.2	0.0343	11.4	0.0155
43	OPT_ce23	0.3	7.3	9.4	0.0499	11.3	9.9	0.0082	1.9	14.9	-0.2917	10.1	-0.0317
44	OPT_dc71	-2.2	3.2	6.6	0.0409	8.0	5.4	0.0257	-2.0	5.8	-0.0365	5.8	-0.0071
45	OPT_dd13	0.6	9.6	10.5	0.0345	12.8	12.7	-0.0434	23.2	33.2	-0.2542	13.0	-0.0865
46	OPT_dd26	-0.5	8.5	11.7	0.0662	17.0	13.2	0.0131	15.2	19.5	-0.1515	12.8	-0.0353
47	OPT_dd73	-0.8	7.1	8.1	0.0861	10.3	10.0	0.0258	2.2	12.1	-0.1229	10.1	-0.0185
48	OPT_de24	-0.5	8.2	7.9	0.0668	9.1	10.1	-0.0049	1.3	11.2	-0.1225	10.1	-0.0188
49	OPT_de33	-1.2	8.0	10.1	0.0751	14.7	12.2	0.0064	3.5	15.8	-0.2637	11.7	-0.0277
50	OPT_de93	-0.8	7.0	6.7	0.0601	8.3	9.6	0.0010	0.4	11.6	-0.1357	9.0	-0.0257
51	OPT_eo1	0.0	0.0	0.0	0.0000	0.0	0.0	0.0000	0.0	0.0	0.0000	0.0	0.0000
52	OPT_fa20	-1.9	7.0	9.6	0.0793	14.5	11.4	0.0387	1.6	14.9	-0.2885	11.0	-0.0010
53	OPT_fa30	0.0	6.5	8.8	0.0765	11.5	9.7	0.0420	1.2	14.2	-0.2860	9.9	-0.0150
54	OPT_fa47	-2.4	4.0	5.6	0.0390	6.4	5.0	-0.0163	-3.4	4.5	-0.0070	5.6	-0.0339
55	OPT_fa5	-0.1	7.5	9.0	0.0877	12.3	11.2	0.0564	1.5	15.3	-0.4084	10.4	0.0085
56	OPT_fa56	1.0	12.3	14.5	0.1066	22.2	18.5	0.0452	8.2	14.7	-0.2657	10.7	0.0235
57	OPT_fc19	-1.9	4.4	7.9	0.0447	9.0	6.9	0.0061	-4.5	6.7	-0.0723	7.1	-0.0248
58	OPT_fc25	-0.5	4.0	7.1	0.0989	8.9	6.3	0.0862	3.3	6.4	0.0643	7.0	0.0595
59	OPT_fc34	-0.2	7.4	10.6	0.1082	15.4	11.3	0.0975	3.4	14.2	-0.2525	10.2	-0.0028
60	OPT_fc59	-0.1	4.3	7.2	0.0833	11.0	7.9	0.0282	3.2	10.2	-0.0424	8.2	0.0356
61	OPT_fc65	-0.6	8.9	11.0	0.1109	16.0	13.4	0.0873	3.5	15.0	-0.3693	11.2	-0.0451
62	OPT_fc86	-0.2	6.6	7.1	0.0670	9.2	9.5	0.0230	-2.6	9.5	-0.1241	9.6	-0.0258

Table C.2: Relative energy differences, in kJ/mol per molecule, and unit-cell densities, ρ, in g/cm^3, of the lowest-energy enantiopure and racemate crystal structures of naproxen (COYRUD) determined by a force field from a previously conducted CSP study[337] (FF; A), and fully relaxed/reranked using B86bPBE-XDM as implemented in Quantum ESPRESSO (PAW; A), B86bPBE-XDM/DZP (SIESTA; A), and sHF-3c (CRYSTAL17; A). Single-point energy calculations performed on low-cost geometries using plane-wave B86bPBE-XDM/PAW or small-basis B86bPBE-XDM/DZP are also tabulated (B//A and C//A, respectively). Energies and densities are relative to the experimentally isolated racemate form (dfCOS1_COS1). The experimentally isolated enantiopure form is dfaf92_af92.

no	structure	FF				SIESTA			CRYSTAL17		
		A	B//A	C//A	ρ	A	B//A	ρ	A	B//A	ρ
1	dfab52_ab52	11.6	10.2	13.2	-0.0735	10.7	8.4	-0.0674	7.4	16.6	-0.0998
2	dfab70_ab70	11.5	10.7	12.2	-0.0546	13.3	13.3	-0.0562	-0.6	16.4	-0.0350
3	dfab99_ab99	11.9	15.0	13.9	-0.0571	11.3	14.1	-0.0792	5.9	17.8	-0.1098
4	dfab9_ab9	6.4	8.1	8.5	-0.0273	7.4	8.2	-0.0540	7.1	11.7	-0.0456
5	dfaf9_af9	8.2	0.6	-1.8	-0.0530	-1.9	-0.7	-0.0345	-2.2	0.8	-0.0348
6	dfai123_ai123	11.4	9.2	8.1	-0.0445	7.3	8.6	-0.0413	9.3	12.6	-0.0500
7	dfai12_ai12	9.5	5.1	3.3	-0.0599	2.9	4.5	-0.0442	-0.7	6.0	-0.0492
8	dfai43_ai43	12.3	9.5	9.4	-0.0864	7.2	8.7	-0.0776	8.8	16.2	-0.0971
9	dfaj49_aj49	22.6	23.1	22.6	-0.1224	22.3	23.6	-0.1266	16.7	26.2	-0.0987
10	dfak24_ak24	6.3	1.5	-0.1	-0.0504	-3.5	-0.3	-0.0369	1.6	2.8	-0.0429
11	dfak35_ak35	1.4	2.2	3.0	-0.0115	1.7	1.4	-0.0096	4.7	5.5	-0.0303
12	dfak52_ak52	11.3	5.9	3.9	-0.0315	0.6	3.6	-0.0301	6.6	12.5	-0.0702
13	dfak57_ak57	0.9	3.2	3.2	0.0035	3.3	3.7	-0.0176	4.0	6.3	-0.0112
14	dfak58_ak58	11.2	5.9	3.8	-0.0317	0.6	3.6	-0.0308	6.6	12.4	-0.0693
15	dfak63_ak63	8.4	2.8	2.3	-0.0718	-1.0	0.2	-0.0538	1.0	4.0	-0.0551
16	dfak86_ak86	12.9	10.9	11.0	-0.0679	8.9	9.3	-0.0687	11.3	14.2	-0.0912
17	dfam133_am133	7.6	13.5	11.2	-0.0628	8.8	12.6	-0.0798	11.2	16.1	-0.0817
18	dfam39_am39	6.1	7.5	9.9	0.0075	12.5	10.1	-0.0061	0.3	12.1	0.0083
19	dfam57_am57	5.9	7.5	9.9	0.0076	12.6	10.1	-0.0071	0.3	12.1	0.0082
20	dfam85_am85	3.8	5.4	4.5	-0.0164	4.3	5.2	-0.0268	7.1	9.3	-0.0375
21	dfaw48_aw48	9.4	10.4	12.3	-0.0008	15.5	13.4	-0.0100	2.1	15.2	0.0037
22	dfbh18_bh18	11.0	3.6	3.6	-0.0579	3.4	2.9	-0.0447	5.1	5.4	-0.0453
23	dfca102_ca102	8.6	8.2	6.8	-0.0481	5.3	8.3	-0.0728	8.1	12.2	-0.0636
24	dfca114_ca114	9.8	6.2	6.8	-0.0372	5.4	4.7	-0.0393	7.4	10.5	-0.0668
25	dfca13_ca13	8.1	8.3	7.5	-0.0476	5.5	8.0	-0.0764	8.1	12.2	-0.0637
26	dfca39_ca39	6.3	8.1	8.2	-0.0280	6.8	8.0	-0.0549	7.1	11.7	-0.0466
27	dfca69_ca69	9.6	7.1	8.8	-0.0419	7.3	5.6	-0.0391	3.7	10.3	-0.0398
28	dfca79_ca79	9.6	6.2	7.1	-0.0370	5.4	4.7	-0.0396	7.7	10.3	-0.0660
29	dfca84_ca84	9.6	7.0	8.8	-0.0419	7.2	5.6	-0.0398	3.7	10.2	-0.0406
30	dfca87_ca87	11.4	10.5	11.9	-0.0550	13.3	13.3	-0.0568	-0.6	16.4	-0.0352
31	dfca89_ca89	11.9	15.0	14.1	-0.0571	10.8	14.1	-0.0904	5.9	17.8	-0.1092
32	dfcb96_cb96	12.0	8.8	9.2	-0.0346	10.6	10.1	-0.0368	10.9	11.4	-0.0155
33	dfcc56_cc56	11.2	9.5	6.5	-0.0401	7.5	10.3	-0.0459	14.6	10.6	-0.0652
34	dfcd137_cd137	11.7	9.1	6.7	-0.0592	6.5	8.6	-0.0607	10.2	12.8	-0.0701
35	dfCO1_CO1	0.7	0.3	0.2	-0.0020	0.4	0.4	-0.0011	-0.0	0.0	-0.0007

Table C.2: Relative energy differences, in kJ/mol per molecule, and unit-cell densities, ρ, in g/cm^3, of the lowes-energy enantiopure and racemate crystal structures of naproxen (COYRUD) determined by a force field from a previously conducted CSP study[337] (FF; A), and fully relaxed/reranked using B86bPBE-XDM as implemented in Quantum ESPRESSO (PAW; A), B86bPBE-XDM/DZP (SIESTA; A), and sHF-3c (CRYSTAL17; A). Single-point energy calculations performed on low-cost geometries using plane-wave B86bPBE-XDM/PAW or small-basis B86bPBE-XDM/DZP are also tabulated (B//A and C//A, respectively). Energies and densities are relative to the experimentally isolated racemate form (dfCOS1_COS1). The experimentally isolated enantiopure form is dfaf92_af92.

no	structure	FF				SIESTA			CRYSTAL17		
		A	B//A	C//A	ρ	A	B//A	ρ	A	B//A	ρ
36	dfCO433_CO433	12.3	9.0	9.4	-0.0366	10.4	9.8	-0.0337	10.8	11.4	-0.0158
37	dfCO70_CO70	12.8	10.9	9.2	-0.0553	8.1	10.1	-0.0494	8.2	14.6	-0.0462
38	dfdc96_dc96	12.0	14.5	15.2	-0.0575	14.1	14.3	-0.0756	17.3	25.6	-0.1995
39	dfde139_de139	12.2	14.1	13.0	-0.0613	11.7	12.9	-0.0667	14.6	17.9	-0.0782
40	dfde29_de29	9.9	4.3	2.0	-0.0239	-0.3	4.5	-0.0246	24.7	29.1	-0.1116
41	dfde83_de83	12.0	9.3	7.7	-0.0613	6.1	7.2	-0.0623	27.2	37.5	-0.0986
42	dffa104_fa104	9.8	6.0	6.2	-0.0593	4.3	4.5	-0.0377	6.8	10.2	-0.0982
43	dffa31_fa31	8.0	10.0	8.8	-0.0468	6.8	9.0	-0.0589	8.9	13.2	-0.0629
44	dffb24_fb24	5.7	5.2	4.3	-0.0249	3.6	4.5	-0.0281	6.0	9.1	-0.0462
45	dffb70_fb70	21.7	24.7	24.1	-0.2136	22.8	24.8	-0.2257	18.4	26.7	-0.2210
46	dffc100_fc100	6.6	4.1	3.2	-0.0316	2.2	3.4	-0.0338	5.2	7.3	-0.0444
47	dffc116_fc116	11.5	11.0	9.1	-0.0567	8.2	10.6	-0.0649	10.4	14.7	-0.0711
48	dffc119_fc119	8.2	6.3	5.4	-0.0480	4.5	6.0	-0.0551	6.9	10.1	-0.0576
49	dffc125_fc125	8.2	4.5	5.4	-0.0598	3.0	2.9	-0.0486	3.0	8.5	-0.0660
50	dffc15_fc15	3.3	5.2	5.1	-0.0245	4.6	5.1	-0.0348	4.6	5.8	-0.0524
51	dffc83_fc83	13.6	5.2	5.3	-0.0685	2.9	3.4	-0.0495	9.6	9.3	-0.0880
52	dffd17_fd17	90.3	45.8	48.3	-0.1178	54.7	52.5	-0.1235	46.8	53.4	-0.1206
53	dffd60_fd60	85.2	36.4	35.7	-0.1753	38.5	39.9	-0.1840	33.1	42.1	-0.1841
54	dfCOS1_COS1	0.0	0.0	0.0	0.0000	0.0	0.0	0.0000	0.0	0.0	0.0000
55	dfaq49_aq49	11.7	3.4	1.2	-0.0686	1.7	3.3	-0.0623	-0.6	3.4	-0.0489
56	dfah12_ah12	10.6	4.2	2.0	-0.0717	5.0	6.1	-0.0627	8.2	9.2	-0.0816
57	dfaf41_af41	11.0	14.0	14.7	-0.0137	17.3	17.3	-0.0298	8.0	17.8	-0.0153
58	dfaf92_af92	8.2	0.8	-1.5	-0.0532	-2.0	-0.8	-0.0344	-2.2	0.9	-0.0336

Table C.3: Relative energies, in kJ/mol per molecule, and unit-cell densities, ρ, in g/cm^3, of the lowest-energy crystal structures of carbamazepine (CBMZPN) determined ranked by a force field[339] (FF; A), and fully relaxed/reranked using B86bPBE-XDM as implemented in Quantum ESPRESSO (PAW; A), B86bPBE-XDM/DZP (SIESTA; A), and sHF-3c (CRYSTAL17; A). Single-point energy calculations performed on low-cost geometries using plane-wave B86bPBE-XDM/PAW or small-basis B86bPBE-XDM/DZP are also tabulated (B//A and C//A, respectively). Energies and densities are relative to Form-I (dfE1_E1). The other experimentally-observed crystal structures are Form-II (df193_193), Form-III (df1_1), Form-IV (df54_54), and Form-V (df145_145).

no	structure	FF				SIESTA			CRYSTAL17		
		A	B//A	C//A	ρ	A	B//A	ρ	A	B//A	ρ
1	df1045_1045	1.2	4.8	3.9	-0.0388	4.6	5.9	-0.0383	3.6	6.8	-0.0433
2	df1_1	-5.0	-3.9	0.2	0.0421	-1.6	-5.1	0.0548	-6.2	-5.8	0.0300
3	df11_11	-2.5	2.3	4.7	0.0078	5.0	3.1	0.0000	0.4	3.6	-0.0001
4	df12_12	-0.5	5.2	4.7	-0.0159	8.1	7.9	-0.0214	10.9	8.9	-0.0356
5	df1219_1219	0.4	3.5	4.8	-0.0239	5.4	4.5	-0.0287	4.5	5.0	-0.0341
6	df13_13	0.2	5.6	4.3	-0.0032	6.3	7.4	-0.0187	7.1	7.8	-0.0441
7	df141_141	0.7	1.4	0.5	0.0053	2.6	3.3	0.0032	-0.2	3.3	0.0096
8	df14_14	-1.3	6.4	11.6	0.0049	12.0	7.5	-0.0063	17.9	21.0	-0.1326
9	df144_144	1.1	9.9	12.5	0.0194	13.5	11.5	-0.0017	7.5	11.5	-0.0044
10	df145_145	-2.0	0.7	-1.8	-0.0075	0.6	3.0	-0.0192	6.7	2.5	-0.0387
11	df153_153	-0.3	3.8	4.1	0.0010	6.0	5.6	-0.0065	7.5	6.8	-0.0259
12	df1532_1532	0.8	6.3	6.4	-0.0403	7.2	7.7	-0.0459	6.3	9.3	-0.0591
13	df161_161	-1.9	13.0	14.0	0.0084	15.2	14.9	-0.0160	14.0	12.7	-0.0194
14	df17_17	-1.0	2.7	5.0	0.0142	4.0	2.3	0.0142	1.5	2.8	-0.0036
15	df1739_1739	0.5	8.5	13.6	0.0028	14.0	9.4	-0.0126	10.6	9.5	-0.0229
16	df1836_1836	-0.2	5.5	4.8	-0.0277	5.8	6.9	-0.0402	11.5	8.9	-0.0728
17	df193_193	3.6	3.7	4.2	-0.0760	4.4	4.2	-0.0784	3.2	3.6	-0.0741
18	df195_195	1.2	3.6	2.3	-0.0291	1.9	3.0	-0.0319	4.6	4.8	-0.0714
19	df211_211	-0.6	2.8	9.3	-0.0236	9.6	3.6	-0.0253	6.3	8.8	-0.0727
20	df2_2	-2.3	2.0	4.4	-0.0015	5.0	2.8	-0.0096	-0.4	0.1	-0.0316
21	df235_235	0.7	5.2	6.4	-0.0305	6.4	5.9	-0.0316	0.7	5.2	-0.0087
22	df25_25	1.3	7.8	10.1	-0.0017	11.6	9.7	-0.0118	10.3	10.7	-0.0313
23	df26_26	-0.8	1.2	2.8	-0.0038	3.1	2.0	-0.0090	0.3	2.3	-0.0067
24	df269_269	1.2	15.9	16.4	-0.0087	16.8	17.0	-0.0335	17.8	15.0	-0.0461
25	df31_31	-0.3	2.5	3.8	-0.0009	3.2	2.2	-0.0059	2.3	2.8	-0.0279
26	df3_3	-5.9	0.4	0.4	0.0061	2.7	2.7	-0.0106	7.4	2.5	-0.0252
27	df33_33	-1.9	2.5	1.6	-0.0141	2.7	3.4	-0.0249	8.1	2.9	-0.0404
28	df36_36	-1.0	4.8	10.6	0.0131	9.6	4.2	0.0286	6.9	5.9	-0.0094
29	df43_43	1.0	2.5	3.7	-0.0064	4.1	3.4	-0.0121	1.9	3.3	-0.0203
30	df44_44	-0.9	0.4	-0.8	-0.0080	-0.7	0.8	-0.0008	8.7	1.9	-0.0419
31	df444_444	-4.0	1.5	4.2	0.0048	4.5	2.3	-0.0049	0.6	2.7	-0.0024
32	df47_47	0.5	4.5	3.2	-0.0541	4.1	5.5	-0.0552	6.8	5.0	-0.0871
33	df49_49	0.9	6.0	3.7	-0.0799	6.0	8.4	-0.0896	10.8	9.1	-0.1086
34	df51_51	1.1	1.3	1.0	-0.0415	0.5	0.1	-0.0159	2.2	0.3	-0.0680
35	df53_53	-0.7	1.7	10.0	-0.0006	9.0	1.1	0.0102	3.2	2.5	-0.0323
36	df54_54	-1.2	-0.4	4.7	-0.0335	4.2	-0.9	-0.0318	2.4	0.4	-0.0714

Table C.3: Relative energies, in kJ/mol per molecule, and unit-cell densities, ρ, in g/cm^3, of the lowest-energy crystal structures of carbamazepine (CBMZPN) determined ranked by a force field[339] (FF; A), and fully relaxed/reranked using B86bPBE-XDM as implemented in Quantum ESPRESSO (PAW; A), B86bPBE-XDM/DZP (SIESTA; A), and sHF-3c (CRYSTAL17; A). Single-point energy calculations performed on low-cost geometries using plane-wave B86bPBE-XDM/PAW or small-basis B86bPBE-XDM/DZP are also tabulated (B//A and C//A, respectively). Energies and densities are relative to Form-I (dfE1_E1). The other experimentally-observed crystal structures are Form-II (df193_193), Form-III (df1_1), Form-IV (df54_54), and Form-V (df145_145).

no	structure	FF				SIESTA			CRYSTAL17		
		A	B//A	C//A	ρ	A	B//A	ρ	A	B//A	ρ
37	df60_60	0.6	6.8	8.7	0.0197	9.0	7.2	0.0084	4.5	8.2	0.0031
38	df63_63	0.4	6.6	12.7	0.0211	13.1	7.4	0.0094	4.7	8.1	0.0052
39	df654_654	1.1	5.4	4.7	-0.0508	4.4	6.1	-0.0518	3.0	6.5	-0.0608
40	df65_65	0.5	3.7	7.7	-0.0302	7.2	4.3	-0.0325	0.8	3.6	-0.0570
41	df6_6	-1.1	6.5	6.4	0.0067	6.0	7.0	-0.0116	10.4	9.1	-0.0780
42	df66_66	0.8	9.8	15.1	-0.0174	16.8	11.9	-0.0307	14.8	13.7	-0.0684
43	df68_68	1.2	5.8	12.1	-0.0143	11.4	7.1	-0.0234	10.1	9.5	-0.0675
44	df80_80	-4.0	13.0	15.1	0.0279	15.6	14.5	0.0000	13.0	13.0	-0.0080
45	df809_809	-1.3	4.2	3.2	-0.0093	3.2	4.9	-0.0162	6.6	5.2	-0.0457
46	df838_838	-1.0	1.4	2.2	-0.0254	2.8	2.3	-0.0308	2.1	2.6	-0.0301
47	df853_853	1.1	6.8	7.6	0.0031	10.1	9.4	-0.0093	3.5	9.9	0.0129
48	df863_863	1.2	2.8	1.8	-0.0413	1.1	3.2	-0.0431	0.8	3.5	-0.0572
49	df8_8	-3.5	2.1	-0.7	-0.0350	1.7	4.6	-0.0496	8.7	6.7	-0.0735
50	df90_90	0.6	6.5	6.6	-0.0368	5.5	6.6	-0.0442	2.1	6.2	-0.0662
51	df95_95	0.9	5.6	10.7	-0.0424	9.2	6.2	-0.0465	2.1	6.2	-0.0660
52	dfE1_E1	0.0	0.0	0.0	0.0000	0.0	0.0	0.0000	0.0	0.0	0.0000

Table C.4: Relative energies, in kJ/mol per molecule, and unit-cell densities, ρ, in g/cm^3, of the lowest-energy crystal structures of olanzapine (UNOGIN) determined by a force field from a previously conducted CSP study[338] (FF; A), and fully relaxed/reranked using B86bPBE-XDM as implemented in Quantum ESPRESSO (PAW; A), B86bPBE-XDM/DZP (SIESTA; A), and sHF-3c (CRYSTAL17; A). Single-point energy calculations performed on low-cost geometries using plane-wave B86bPBE-XDM/PAW or small-basis B86bPBE-XDM/DZP are also tabulated (B//A and C//A, respectively). Energies and densities are relative to Form-I (dfeq11_eq11). The other experimentally-observed crystal structures are Form-II (dfeq45_eq45) and Form-IV (dfeq1_eq1). Form-III has not yet been characterized by experiment, although previous work[338] has postulated the structure to match dfeq162_eq162.

no	structure	FF				SIESTA			CRYSTAL17		
		A	B//A	C//A	ρ	A	B//A	ρ	A	B//A	ρ
1	dfax16_ax16	7.7	10.0	7.0	-0.0728	11.0	15.3	-0.1159	2.2	14.4	-0.1069
2	dfax1_ax1	7.2	8.0	8.5	0.0047	14.6	15.1	-0.0304	4.4	12.0	-0.0198
3	dfax2074_ax2074	3.9	10.2	5.8	-0.1279	9.0	15.8	-0.1760	3.9	15.3	-0.1624
4	dfax211_ax211	3.8	9.9	5.9	-0.1210	9.1	15.5	-0.1692	3.9	15.2	-0.1611
5	dfax3_ax3	-2.1	11.2	7.8	-0.0859	8.0	14.0	-0.1387	3.9	13.6	-0.1287
6	dfax4_ax4	7.3	14.2	15.4	-0.0096	23.0	23.8	-0.0635	10.8	22.7	-0.0520
7	dfax6763_ax6763	6.2	11.9	14.5	-0.0165	20.3	18.2	-0.0664	6.8	16.4	0.0000
8	dfeq10_eq10	2.3	8.8	8.3	-0.0279	14.7	16.8	-0.0742	24.0	18.9	-0.1031
9	dfeq11_eq11	0.0	0.0	0.0	0.0000	0.0	0.0	0.0000	0.0	0.0	0.0000
10	dfeq126_eq126	7.8	14.5	14.3	-0.0205	21.6	24.0	-0.0726	17.4	22.3	-0.0438
11	dfeq12_eq12	6.6	12.5	13.3	-0.0119	17.7	18.3	-0.0547	6.5	13.2	-0.0025
12	dfeq1392_eq1392	3.3	5.9	5.1	-0.0162	7.1	9.1	-0.0455	4.7	6.6	-0.0301
13	dfeq13_eq13	3.4	5.7	6.5	-0.0060	9.9	10.2	-0.0332	3.7	7.6	-0.0164
14	dfeq157_eq157	6.4	10.7	10.7	-0.0168	18.1	19.7	-0.0602	17.6	19.1	-0.0440
15	dfeq15_eq15	4.0	7.5	6.3	-0.0142	13.4	16.4	-0.0648	26.1	21.5	-0.0759
16	dfeq162_eq162	7.1	4.4	4.1	-0.0203	4.9	5.8	-0.0442	-1.5	-0.3	-0.0139
17	dfeq17_eq17	4.9	10.4	8.5	-0.0399	16.2	19.9	-0.0885	10.2	16.1	-0.0568
18	dfeq1_eq1	-2.0	-1.9	-2.3	-0.0046	-0.8	1.0	-0.0287	-0.2	0.5	-0.0293
19	dfeq206_eq206	4.6	15.6	13.5	-0.0523	13.7	18.3	-0.0977	12.3	16.9	-0.0863
20	dfeq20_eq20	3.3	5.7	4.9	-0.0156	6.8	8.7	-0.0435	4.7	6.6	-0.0301
21	dfeq226_eq226	6.0	18.8	19.0	-0.0034	20.8	22.5	-0.0570	14.5	22.3	-0.0312
22	dfeq23_eq23	7.6	8.0	3.9	-0.0468	6.5	12.0	-0.0783	10.1	12.8	-0.0642
23	dfeq2481_eq2481	5.4	19.8	19.1	-0.0240	13.4	17.5	-0.0657	10.0	18.0	-0.0585
24	dfeq24_eq24	1.6	13.2	11.1	-0.0509	12.6	17.1	-0.0981	14.8	16.0	-0.0919
25	dfeq25_eq25	7.0	13.4	13.5	-0.0353	19.7	21.3	-0.0795	22.3	22.2	-0.0759
26	dfeq2639_eq2639	0.0	0.1	0.1	-0.0017	0.1	0.1	0.0015	0.0	0.0	-0.0002
27	dfeq27_eq27	5.2	7.8	6.1	-0.0636	10.9	14.2	-0.1066	7.9	10.0	-0.0673
28	dfeq28_eq28	4.7	9.8	10.8	-0.0045	17.8	18.2	-0.0459	19.2	16.6	-0.0399
29	dfeq2_eq2	4.9	8.2	5.8	-0.0202	11.5	15.4	-0.0604	7.6	13.1	-0.0208
30	dfeq30_eq30	-0.6	13.1	11.1	-0.0556	12.7	17.2	-0.1033	11.0	16.8	-0.0975
31	dfeq336_eq336	7.1	14.6	17.4	-0.0060	22.1	21.2	-0.0516	19.0	18.5	-0.0179
32	dfeq34_eq34	1.9	10.4	10.2	0.0007	1.8	13.6	-0.0499	9.3	15.0	-0.0573
33	dfeq36_eq36	6.4	9.2	8.2	-0.0447	13.1	15.5	-0.0849	10.5	14.2	-0.0503
34	dfeq3731_eq3731	-2.5	1.2	3.8	0.0362	10.6	9.4	-0.0037	11.7	8.3	0.0076

Table C.4: Relative energies, in kJ/mol per molecule, and unit-cell densities, ρ, in g/cm^3, of the lowest-energy crystal structures of olanzapine (UNOGIN) determined by a force field from a previously conducted CSP study[338] (FF; A), and fully relaxed/reranked using B86bPBE-XDM as implemented in Quantum ESPRESSO (PAW; A), B86bPBE-XDM/DZP (SIESTA; A), and sHF-3c (CRYSTAL17; A). Single-point energy calculations performed on low-cost geometries using plane-wave B86bPBE-XDM/PAW or small-basis B86bPBE-XDM/DZP are also tabulated (B//A and C//A, respectively). Energies and densities are relative to Form-I (dfeq11_eq11). The other experimentally-observed crystal structures are Form-II (dfeq45_eq45) and Form-IV (dfeq1_eq1). Form-III has not yet been characterized by experiment, although previous work[338] has postulated the structure to match dfeq162_eq162.

no	structure	FF				SIESTA			CRYSTAL17		
		A	B//A	C//A	ρ	A	B//A	ρ	A	B//A	ρ
35	dfeq406_eq406	3.3	4.9	6.4	-0.0057	12.4	11.1	-0.0281	12.5	11.8	-0.0293
36	dfeq42_eq42	5.8	10.0	10.6	-0.0355	16.0	16.5	-0.0604	17.7	16.9	-0.0467
37	dfeq45_eq45	6.2	4.0	4.3	-0.0219	3.5	3.7	-0.0334	0.3	1.5	-0.0362
38	dfeq49_eq49	4.3	13.0	13.2	0.0131	14.8	16.7	-0.0415	11.9	15.7	-0.0308
39	dfeq4_eq4	3.7	5.6	2.5	-0.0467	6.2	10.5	-0.0789	12.9	11.7	-0.1052
40	dfeq539_eq539	-0.6	4.2	6.0	0.0135	12.6	12.2	-0.0253	14.6	11.7	-0.0295
41	dfeq53_eq53	6.4	8.7	10.1	0.0038	15.8	15.4	-0.0324	16.7	12.7	-0.0098
42	dfeq55_eq55	7.6	13.7	14.5	0.0113	20.7	21.9	-0.0443	12.7	20.5	0.00550
43	dfeq56_eq56	7.3	18.2	18.5	-0.0226	22.5	24.1	-0.0792	12.1	21.1	-0.0242
44	dfeq63_eq63	7.1	14.6	12.3	-0.0522	19.6	23.9	-0.1024	17.6	22.5	-0.0950
45	dfeq679_eq679	7.7	12.3	12.9	-0.0177	19.7	20.0	-0.0542	16.2	20.9	-0.0304
46	dfeq6_eq6	2.0	3.7	2.5	-0.0334	5.4	7.9	-0.0638	4.8	5.4	-0.0312
47	dfeq71_eq71	7.7	14.7	14.7	-0.0085	20.2	20.8	-0.0432	16.9	19.7	-0.0389
48	dfeq75_eq75	7.6	22.1	20.4	-0.0425	15.5	18.8	-0.0622	12.3	20.1	-0.0816
49	dfeq7_eq7	-2.5	9.2	6.7	-0.0548	8.8	13.4	-0.0950	9.1	11.3	-0.0837
50	dfeq99_eq99	7.8	13.5	13.2	-0.0249	18.5	19.4	-0.0504	22.9	19.3	-0.0447

BIBLIOGRAPHY

[1] Pyzer-Knapp, E. O.; Thompson, H.; Schiffmann, F.; Jelfs, K. E.; Chong, S. Y.; Little, M. A.; Cooper, A. I.; Day, G. M. Predicted Crystal Energy Landscapes of Porous Organic Cages. *Chem. Sci.* **2014**, *5*, 2235–2245.

[2] Yang, Y.; Rice, B.; Shi, X.; Brandt, J. R.; Correa da Costa, R.; Hedley, G. J.; Smilgies, D.; Frost, J. M.; Samuel, I.; Otero-de-la-Roza, A.; Johnson, E. R.; Jelfs, K. E.; Nelson, J.; Campbell, A. J.; Fuchter, M. J. Emergent Properties of an Organic Semiconductor Driven by its Molecular Chirality. *ACS Nano* **2017**, *11*, 8329–8338.

[3] Ravva, M. K.; Risko, C.; Brédas, J. L. In *Non-covalent Interactions in Quantum Chemistry and Physics*; Otero-de-la-Roza, A., DiLabio, G., Eds.; Elsevier: Amsterdam, Netherlands, 2017; Chapter 9, pp 277–298.

[4] Gallagher, H. G.; Sherwood, J. N. Polymorphism, Twinning and Morphology of Crystals of 2,4,6-Trinitrotoluene Grown From Solution. *J. Chem. Soc., Faraday Trans.* **1996**, *92*, 2107–2116.

[5] Vrcelj, R. M.; Sherwood, J. N.; Kennedy, A. R.; Gallagher, H. G.; Gelbrich, T. Polymorphism in 2-4-6 Trinitrotoluene. *Cryst. Growth Des.* **2003**, *3*, 1027–1032.

[6] Wei, C.; Huang, H.; Duan, X.; Pei, C. Structures and Properties Prediction of HMX/TATB Co-Crystal. *Propellants Explos. Pyrotech.* **2011**, *36*, 416–423.

[7] Hao, Z.; Iqbal, A. Some Aspects of Organic Pigments. *Chem. Soc. Rev.* **1997**, *26*, 203–213.

[8] Zykova-Timan, T.; Raiteri, P.; Parrinello, M. Investigating the Polymorphism in PR179: A Combined Crystal Structure Prediction and Metadynamics Study. *J. Phys. Chem. B* **2008**, *112*, 13231–13237.

[9] McCrone, W. In *Physics and Chemistry of the Organic Solid State*; Fox, D., Labes, M. M., Weissenberg, N., Eds.; Wiley Interscience: New York, 1965; Chapter 8, pp 725–767.

[10] Sharma, B. D. Allotropes and Polymorphs. *J. Chem. Educ.* **1987**, *64*, 404–407.

[11] Gavezzotti, A. A Solid-State Chemist's View of the Crystal Polymorphism of Organic Compounds. *J. Pharm. Sci.* **2007**, *96*, 2232–2441.

[12] Purohit, R.; Venugopalan, P. Polymorphism: An Overview. *Resonance* **2009**, *14*, 882–893.

[13] Brog, J.-P.; Chanez, C.-L.; Crochet, A.; Fromm, K. M. Polymorphism, What it is and How to Identify it: A Systematic Review. *RSC Advances* **2013**, *3*, 16905–16931.

[14] Maddox, J. Crystals From First Principles. *Nature* **1988**, *335*, 201.

[15] Gavezzotti, A. Are Crystal Structures Predictable? *Acc. Chem. Res.* **1994**, *27*, 309–314.

[16] Day, G. M. Current Approaches to Predicting Molecular Organic Crystals. *Crystallogr. Rev.* **2011**, *17*, 3–52.

[17] Price, S. L. Lattice Energy, Nailed? *Science* **2014**, *345*, 619–620.

[18] Price, S. L. Predicting Crystal Structures of Organic Compounds. *Chem. Soc. Rev.* **2014**, *43*, 2098–2111.

[19] Beran, G. A New Era For *Ab Initio* Molecular Crystal Lattice Energy Prediction. *Angew. Chem. Int. Ed.* **2015**, *54*, 396–398.

[20] Cruz-Cabeza, A. J.; Reutzel-Edens, S. M.; Bernstein, J. Facts and Fiction About Polymorphism. *Chem. Soc. Rev.* **2015**, *44*, 8619–8635.

[21] Thakur, T. S.; Dubey, R.; Desiraju, G. R. Crystal Structure and Prediction. *Annu. Rev. Phys. Chem.* **2015**, *66*, 21–42.

[22] Groom, C. R.; Bruno, I. J.; Lightfoot, M. P.; Ward, S. C. The Cambridge Structural Database. *Acta Crystallographica Section B: Structural Science* **2016**, *72*, 171–179.

[23] Beran, G. Modelling Polymorphic Molecular Crystals with Electronic Structure Theory. *Chem. Rev.* **2016**, *116*, 5567–5613.

[24] Reilly, A. M.; Cooper, R. I.; Adjiman, C. S.; Bhattacharya, S.; Boese, A. D.; Brandenburg, J. G.; Bygrave, P. J.; Bylsma, R.; Campbell, J. E.; Car, R.; Case, D. H.; Chadha, R.; Cole, J. C.; Cosburn, K.; Cuppen, H. M.; Curtis, F.; Day, G. M.; DiStasio Jr, R. A.; Dzyabchenko, A.; van Eijck, B. P.; Elking, D. M.; van den Ende, J. A.; Facelli, J. C.; Ferraro, M. B.; Fusti-Molnar, L.; Gatsiou, C.-A.; Gee, T. S.; de Gelder, R.; Ghiringhelli, L. M.; Goto, H.; Grimme, S.; Guo, R.; Hofmann, D. W. M.; Hoja, J.; Hylton, R. K.; Iuzzolino, L.; Janckiewicz, W.; de Jong, D. T.; Kendrick, J.; de Klerk, N.; Ko, H.-Y.; Kuleshova, L. N.; Li, X.; Lohani, S.; Leusen, F.; Lund, A. M.; Lv, J.; Ma, Y.; Marom, N.; Masunov, A. E.; McCabe, P.; McMahon, D. P.; Meekes, H.; Metz, M. P.; Misquitta, A. J.; Mohamed, S.; Monserrat, B.; Needs, R. J.; Neumann, M. A.; Nyman, J.; Obata, S.; Oberhofer, H.; Oganov, A. R.; Orendt, A. M.; Pagola, G. I.; Pantelides, C. C.; Pickard, C. J.; Podeszwa, R.; Price, L. S.; Price, S. L.; Pulido, A.; Read, M. G.; Reuter, K.; Schneider, E.; Schober, C.; Shields, G. P.; Singh, P.; Sugden, I. J.; Szalewicz, K.; Taylor, C. R.; Tkatchenko, A.; Tuckerman, M. E.; Vacarro, F.; Vasileiadis, M.; Vazquez-Mayagoitia, A.; Vogt, L.; Wang, Y.; Watson, R. E.; de Wijs, G. A.; Yang, J. Z.; Zhu, Q.; Groom, C. R. Report on the Sixth Blind Test of Organic Crystal-Structure Prediction Methods. *Acta Crystallogr., Sect. B: Struct. Sci.* **2016**, *72*, 439–459.

[25] Haleblian, J.; McCrone, W. Pharmaceutical Applications of Polymorphism. *J. Pharm. Sci.* **1969**, *58*, 911–929.

[26] Abramov, Y. A. Current Computational Approaches to Support Pharmaceutical Solid Form Selection. *Org. Process Res. Dev.* **2013**, *17*, 472–485.

[27] Price, S. L.; Braun, D. E.; Reutzel-Edens, S. M. Can Computed Crystal Energy Landscapes Help Understand Pharmaceutical Solids? *Chem. Commun.* **2016**, *52*, 7065–7077.

[28] Yu, L. Polymorphism in Molecular Solids: An Extraordinary System of Red, Orange, and Yellow Crystals. *Acc. Chem. Res.* **2010**, *43*, 1257–1266.

[29] Calligaro, D. O.; Fairhurst, J.; Hotten, T. M.; Moore, N. A.; Tupper, D. E. The Synthesis and Biological Activity of Some Known and Putative Metabolites of the Atypical Antipsychotic Agent Olanzepine (LY170053). *Bioorg. Med. Chem. Lett.* **1997**, *7*, 25–30.

[30] Tan, M.; Shtukenberg, A. G.; Zhu, S.; Xu, W.; Dooryhee, E.; Nichols, G. M.; Ward, M. D.; Kahr, B.; Zhu, Q. ROY Revisited, Again: The Eighth Solved Structure. *Faraday Discuss.* **2018**, *211*, 477–491.

[31] Bučar, D.; Lancaster, R. W.; Bernstein, J. Disappearing Polymorphs Revisited. *Angew. Chem. Int. Ed.* **2015**, *54*, 6972–6993.

[32] Chemburkar, S. R.; Bauer, J.; Deming, K.; Spiwek, H.; Patel, K.; Morris, J.; Henry, R.; Spanton, S.; Dziki, W.; Porter, W.; Quick, J.; Bauer, P.; Donaubauer, J.; Narayanan, B. A.; Soldani, M.; Riley, D.; McFarland, K. Dealing with the Impact of Ritonavir Polymorphs on the Late Stages of Bulk Drug Process Development. *Org. Process Res. Dev.* **2000**, *4*, 413–417.

[33] Bauer, J.; Spanton, S.; Henry, R.; Quick, J.; Dziki, W.; Porter, W.; Morris, J. Ritonavir: An Extraordinary Example of Conformational Polymorphism. *Pharm. Res.* **2001**, *18*, 859–866.

[34] Perez-Lloret, S.; Rey, M. V.; Ratti, P. L.; Rascol, O. Rotigotine Transdermal Patch for the Treatment of Parkinson's Disease. *Fundam. Clin. Pharmacol.* **2013**, *27*, 81–95.

[35] Newman, A. Specialized Solid Form Screening Techniques. *Org. Process Res. Dev.* **2013**, *17*, 457–471.

[36] Neumann, M. A.; van de Streek, J. How Many Ritonavir Cases are There Still Out There. *Faraday Discuss.* **2018**, *211*, 441–458.

[37] Price, S. L. Is Zeroth Order Crystal Structure Prediction (CSP_0) Coming To Maturity? What Should We Aim For in An Ideal Structure Prediction Code. *Faraday Discuss.* **2018**, *211*, 9–30.

[38] Lommerse, J.; Motherwell, W. D. S.; Ammon, H. L.; Dunitz, J. D.; Gavezzotti, A.; Hofmann, D.; Leusen, F.; Mooij, W.; Price, S. L.; Schweizer, B.; Schmidt, M. U.; van Eijck, B. P.; Verwer, P.; Williams, D. E. A Test of Crystal Structure Prediction of Small Organic Molecules. *Acta Cryst. B* **2000**, *56*, 697–714.

[39] Motherwell, W.; Ammon, H. L.; Dunitz, J. D.; Dzyabchenko, A.; Erk, P.; Gavezzotti, A.; Hofmann, D.; Leusen, F.; Lommerse, J.; Mooij, W.; Price, S. L.; Scheraga, H.; Schweizer, B.; Schmidt, M. U.; van Eijck, B. P.; Verwer, P.; Williams, D. E. Crystal Structure Prediction of Small Organic Molecules: A Second Blind Test. *Acta Cryst. B* **2002**, *58*, 647–661.

[40] Day, G. M.; Motherwell, W.; Ammon, H. L.; Boerrigter, S.; Della Valle, R. G.; Venuti, E.; Dzyabchenko, A.; Dunitz, J. D.; Schweizer, B.; van Eijck, B. P.; Erk, P.; Facelli, J. C.; Bazterra, V. E.; Ferraro, M. B.; Hofmann, D.; Leusen, F. J. J.; Liang, C.; Pantelides, C. C.; Karamertzanis, P. G.; Price, S. L.; Lewis, T. C.; Nowell, H.; Torrisi, A.; Scheraga, H. A.; Arnautova, Y. A.; Schmidt, M. U.; Verwer, P. A Third Blind Test of Crystal Structure Prediction. *Acta Cryst. B* **2005**, *61*, 511–527.

[41] Day, G. M.; Cooper, T. G.; Cruz-Cabeza, A. J.; Hejczyk, K. E.; Ammon, H. L.; Boerrigter, S.; Tan, J. S.; Della Valle, R. G.; Venuti, E.; Jose, J.; Gadre, R. S.; Desiraju, G. R.; Thakur, T. S.; van Eijck, B. P.; Facelli, J. C.; Bazterra, V. E.; Ferraro, M. B.; Hofmann, D.; Neumann, M. A.; Leusen, F.; Kendrick, J.; Price, S. L.; Misquitta, A. J.; Karamertzanis, P. G.; Welch, G.; Scheraga, H. A.; Arnautova, Y. A.; Schmidt, M. U.; van de Streek, J.; Wolf, A. K.; Schweizer, B. Significant Progress in Predicting the Crystal Structures of Small Organic Molecules – A Report on the Fourth Blind Test. *Acta Cryst. B* **2009**, *65*, 107–125.

[42] Bardwell, D. A.; Adjiman, C. S.; Arnautova, Y. A.; Bartashevich, E.; Boerrigter, S.; Braun, D. E.; Cruz-Cabeza, A. J.; Day, G. M.; Della Valle, R. G.; Desiraju, G. R.; van Eijck, B. P.; Facelli, J. C.; Ferraro, M. B.; Grillo, D.; Habgood, M.; Hofmann, D. W. M.; Hofmann, F.; Jose, K. V. J.; Karamertzanis, P. G.; Kazantsev, A. V.; Kendrick, J.; Kuleshova, L. N.; Leusen, F.; ad A. J. Misquitta, A. V. M.; Mohamed, S.; Needs, R. J.; Neumann, M. A.; Nikylov, D.; Orendt, A. M.; Pal, R.; Pantelides, C. C.; Pickard, C. J.; Price, L. S.; Price, S. L.; Scheraga, H. A.; van de Streek, J.; Thakur, T. S.; Tiwari, S.; Venuti, E.; Zhitkov, I. K. Towards Crystal Structure Prediction of Complex Organic Compounds – A Report On the Fifth Blind Test. *Acta Crystallogr., Sect. B: Struct. Sci.* **2011**, *67*, 535–551.

[43] Cruz-Cabeza, A. J. Crystal Structure Prediction: Are We There Yet? *Acta Cryst. B* **2016**, *72*, 437–438.

[44] Braun, D. E.; Bhardwaj, R. M.; Florence, A. J.; Tocher, D. A.; Price, S. L. Complex Polymorphic System of Gallic Acid–Five Monohydrates, Three Anhydrates, and over 20 Solvates. *Cryst. Growth Des.* **2013**, *13*, 19–23.

[45] Lehmann, C. W. Crystal Structure Prediction–Dawn of a New Era. *Angew. Chem. Int. Ed.* **2011**, *50*, 5616–5617.

[46] Asmadi, A.; Neumann, M. A.; Kendrick, J.; Girard, P.; Perrin, M.-A.; Leusen, F. Revisiting the Blind Tests in Crystal Structure Prediction: Accurate Energy Ranking of Molecular Crystals. *J. Phys. Chem. B* **2009**, *113*, 16303–16313.

[47] Chan, H.; Kendrick, J.; Leusen, F. Molecule VI, a Benchmark Crystal-Structure Prediction-Prediction Sulfonamide: Are Its Polymorphs Predictable? *Angew. Chem. Int. Ed.* **2011**, *50*, 2979–2981.

[48] Neumann, M. A.; Leusen, F.; Kendrick, J. A Major Advance in Crystal Structure Prediction. *Angew. Chem. Int. Ed.* **2008**, *47*, 2427–2430.

[49] Kendrick, J.; Leusen, F.; Neumann, M. A.; van de Streek, J. Progress in Crystal Structure Prediction. *Chem. Eur. J.* **2011**, *17*, 10736–10744.

[50] Neumann, M. A. Tailor-Made Force-Fields for Crystal-Structure Predicition. *J. Phys. Chem. B* **2008**, *112*, 9810–9829.

[51] Neumann, M. GRACE has been developed by Avant-garde Materials Simulation since 2002. **2002**,

[52] Kazantsev, A. V.; Karamertzanis, P. G.; Adjiman, C. S.; Pantelides, C. C.; Price, S. L.; Galek, P.; Day, G. M.; Cruz-Cabeza, A. J. Successful Prediction of a Model Pharmaceutical in the Fifth Blind Test of Crystal Structure Prediction. *Int. J. Pharm.* **2011**, *418*, 168–178.

[53] Price, S. L. Why Don't We Find More Polymorphs? *Acta Crystallogr., Sect. B: Struct. Sci.* **2013**, *69*, 313–328.

[54] Price, S. L. Computed Crystal Energy Landscapes for Understanding and Predicting Organic Crystal Structures and Polymorphism. *Acc. Chem. Res.* **2009**, *42*, 117–126.

[55] Hylton, R. K.; Tizzard, G. J.; Threlfall, T. L.; Ellis, A. L.; Coles, S. J.; Seaton, C. C.; Schulze, E.; Lorenz, H.; Seidel-Morgenstern, A.; Stein, M.; Price, S. L. Are the Crystal Structures of Enantiopure and Racemic Mandelic Acids Determined by Kinetics or Thermodynamics? *J. Am. Chem. Soc.* **2015**, *137*, 11095–11104.

[56] Woodley, S. M.; Catlow, R. Crystal Structure Prediction From First Principles. *Nat. Mater.* **2008**, *7*, 937–946.

[57] Marqués, M.; Morales, A.; Menéndez, J. M. In *An Introduction to High-Pressure Science and Technology*; Recio, J. M., Menendez, J. M., Otero-de-la-Roza, A., Eds.; CRC Press, 2015; Chapter 4, pp 105–130.

[58] Pickard, C. J.; Needs, R. J. *Ab Initio* Random Structure Searching. *J. Phys. Condens. Matter* **2011**, *23*, 053201.

[59] Karamertzanis, P. G.; Pantelides, C. C. *Ab Initio* Crystal Structure Prediction–I. Rigid Molecules. *J. Comput. Chem.* **2005**, *26*, 304–324.

[60] Karamertzanis, P. G.; Pantelides, C. C. *Ab Initio* Crystal Structure Prediction. II. Flexible Molecules. *J. Mol. Phys.* **2007**, *105*, 273–291.

[61] Case, D. H.; Campbell, J. E.; Bygrave, P. J.; Day, G. M. Convergence Properties of Crystal Structure Prediction by Quasi-Random Sampling. *J. Chem. Theory Comput.* **2016**, *12*, 910–924.

[62] Kirkpatrick, S.; Gelat, C. D.; Vecchi, M. P. Optimization by Simulated Annealing. *Science* **1983**, *220*, 671–690.

[63] Wales, D. J.; Doyle, J. Global Optimization by Basin-Hopping and the Lowest Energy Structures of Lennard-Jones Clusters Containing up to 110 Atoms. *J. Phys. Chem. A* **1997**, *101*, 5111–5116.

[64] Goedecker, S. Minima Hopping: An Efficient Search Method for the Global Minimum of the Potential Energy Surface of Complex Molecular Systems. *J. Chem. Phys.* **2004**, *120*, 9911–9917.

[65] Laio, A.; Parrinello, M. Escaping Free-Energy Minima. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 12562–12566.

[66] Laio, A.; Gervasio, F. L. Metadynamics: A Method to Simulate Rare Events and Reconstruct the Free Energy in Biophysics, Chemistry and Material Science. *Rep. Prog. Phys.* **2008**, *71*, 126601.

[67] Piaggi, P. M.; Parrinello, M. Enhancing Entropy and Enthalpy Fluctuations to Drive Crystallization in Atomistic Simulations. *Phys. Rev. Lett.* **2017**, *119*, 015701.

[68] Piaggi, P. M.; Parrinello, M. Predicting Polymorphism in Molecular Crystals Using Orientational Entropy. *Proced. Acad. Nat. Sci.* **2018**, *115*, 10251–10256.

[69] Oganov, A. R.; Glass, C. W. Crystal Structure Prediction Using *Ab Initio* Evolutionary Techniques: Principles and Applications. *J. Chem. Phys.* **2006**, *124*, 244704.

[70] Glass, C. W.; Oganov, A. R.; Hansen, N. USPEX–Evolutionary Crystal Structure Prediction. *Comput. Phys. Commun.* **2006**, *175*, 713–720.

[71] Zhu, Q.; Oganov, A. R.; Glass, C. W.; Stokes, H. T. Constrained Evolutionary Algorithm for Structure Prediction of Molecular Crystals: Methodology and Applications. *Acta Cryst. B* **2012**, *68*, 215–226.

[72] Lyakhov, A. O.; Oganov, A. R.; Stokes, H. T.; Zhu, Q. New Developments in Evolutionary Structure Prediction Algorithm USPEX. *Comput. Phys. Commun.* **2013**, *184*, 1172–1182.

[73] Lonie, D. C.; Zurek, E. XTALOPT: An Open-Source Evolutionary Algorithm for Crystal Structure Prediction. *Comput. Phys. Commun.* **2011**, *182*, 372–387.

[74] Lonie, D. C.; Zurek, E. Identifying Duplicate Crystal Structures: XTALCOMP, and Open-Source Solution. *Comput. Phys. Commun.* **2012**, *183*, 690–697.

[75] Curtis, F.; Li, X.; Rose, T.; Vázquez-Mayagoitia, A.; Bhattacharya, S.; Ghiringhelli, L. M.; Marom, N. GAtor: A First Principles Genetic Algorithm for Molecular Crystal Structure Prediction. *J. Chem. Theory Comput.* **2018**, *14*, 2246–2264.

[76] Curtis, F.; Rose, T.; Marom, N. Evolutionary Niching in the GAtor Genetic Algorithm for Molecular Crystal Structure Prediction. *Faraday Discuss.* **2018**, *211*, 61–77.

[77] Wang, Y.; Lv, J.; Zhu, L.; Ma, Y. CALYPSO: A Method for Crystal Structure Prediction. *Comput. Phys. Commun.* **2012**, *183*, 2063–2070.

[78] Wang, H.; Wang, Y.; Lv, J.; Lu, Q.; Zhang, L.; Ma, Y. CALYPSO Structure Prediction Method and Its Wide Application. *Comput. Mater. Sci.* **2016**, *112*, 406–415.

[79] Williams, D. E. Improved Intermolecular Force Field for Molecules Containing H, C, N, and O Atoms, with Application to Nucleoside and Peptide Crystals. *J. Comput. Chem.* **2001**, *22*, 1154–1166.

[80] Gale, J. D.; Rohl, A. L. The General Lattice Utility Program (GULP). *Mol. Simul.* **2003**, *29*, 291–341.

[81] Mayo, S. L.; Olafson, B. D.; Goddard, W. A. DREIDING: A Generic Force Field for Molecular Simulations. *J. Phys. Chem.* **1990**, *94*, 8897–8909.

[82] Rappe, A. K.; Goddard III, W. A. Charge Equilibration for Molecular Dynamics Simulations. *J. Phys. Chem.* **1991**, *95*, 3358–3363.

[83] Gavezzotti, A.; Filippini, G. Polymorphic Forms of Organic Crystals at Room Conditions: Thermodynamic and Structural Implications. *J. Am. Chem. Soc.* **1995**, *117*, 12299–12305.

[84] Day, G. M.; Chisholm, J.; Shan, N.; Motherwell, W.; Jones, W. An Assessment of Lattice Energy Minimization for the Prediction of Molecular Organic Crystal Structures. *Cryst. Growth. Des.* **2004**, *4*, 1327–1340.

[85] Nyman, J.; Day, G. M. Static and Lattice Vibrational Energy Differences Between Polymorphs. *CrystEngComm* **2015**, *17*, 5154–5165.

[86] Burger, A.; Ramburger, R. On the Polymorphism of Pharmaceuticals and Other Molecular Crystals. I Theory of Thermodynamic Rules. *Mikrochim. Acta* **1979**, *72*, 259–271.

[87] Burger, A.; Ramburger, R. On the Polymorphism of Pharmaceuticals and Other Molecular Crystals. II Applicability of Thermodynamic Rules. *Mikrochim. Acta* **1979**, *72*, 273–316.

[88] Yang, J.; Hu, W.; Usvyat, D.; Matthews, D.; Schütz, M.; Chan, G. *Ab Initio* Determination of the Crystalline Benze Lattice Energy to Sub-Kilojoule/Mole Accuracy. *Science* **2014**, *345*, 640–643.

[89] Beran, G.; Nanda, K. Predicting Organic Crystal Lattice Energies with Chemical Accuracy. *J. Phys. Chem. Lett.* **2010**, *1*, 3480–3487.

[90] Brandenburg, J. G.; Hochheim, M.; Bredow, T.; Grimme, S. Low-Cost Quantum Chemical Methods for Noncovalent Interactions. *J. Phys. Chem. Lett.* **2014**, *5*, 4275–4284.

[91] Price, S. L.; Leslie, M.; Welch, G.; Habgood, M.; Price, L. S.; Karamertzanis, P. G.; Day, G. M. Modelling Organic Crystal Structures Using Distributed Multipole and Polarizability-Based Model Intermolecular Potentials. *Phys. Chem. Chem. Phys.* **2010**, *12*, 8478–8490.

[92] Stone, A. J.; Alderton, M. Distributed Multipole Analysis Methods and Applications. *Mol. Phys.* **2002**, *100*, 221–233.

[93] Perdew, J. P. Some Fundamental Issues in Ground-State Density Functional Theory: A Guide for the Perplexed. *J. Chem. Theory Comput.* **2009**, *5*, 902–908.

[94] Wen, S.; Beran, G. Accurate Molecular Crystal Lattice Energies from a Fragment QM/MM Approach with On-the-Fly *Ab Initio* Force Field Parametrization. *J. Chem. Theory Comput.* **2011**, *7*, 3733–3742.

[95] Christensen, A. S.; Kubař, T.; Cui, Q.; Elstner, M. Semiempirical Quantum Mechanical Methods for Noncovalent Interactions for Chemical and Biochemical Applications. *Chem. Rev.* **2016**, *116*, 5301–5337.

[96] Brandenburg, J. G.; Grimme, S. Dispersion Corrected Hartree-Fock and Density Functional Theory for Organic Crystal Structure Prediction. *Top. Curr. Chem.* **2014**, *345*, 1–24.

[97] Brandenburg, J. G.; Grimme, S. Accurate Modeling of Organic Molecular Crystals by Dispersion-Corrected Density Functional Tight Binding (DFTB). *J. Phys. Chem. Lett.* **2014**, *5*, 1785–1789.

[98] Heit, Y. N.; Beran, G. How Important is Thermal Expansion for Predicting Molecular Crystal Structures and Thermochemistry at Finite Temperatures. *Acta Cryst. B* **2016**, *72*, 514–529.

[99] Hoja, J.; Reilly, A. M.; Tkatchenko, A. First-Principles Modeling of Molecular Crystals: Structures and Stabilities, Temperature and Pressure. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2017**, *7*, e1294.

[100] Cruz-Cabeza, A. J. Facts and Fictions About Polymorphism. *Chem. Soc. Rev.* **2015**, *44*, 8619–8635.

[101] Whittleton, S. R.; Otero-de-la-Roza, A.; Johnson, E. R. Exchange-Hole Dipole Dispersion Model for Accurate Energy Ranking in Molecular Crystal Structure Prediction. *J. Chem. Theory Comput.* **2017**, *13*, 441–450.

[102] Whittleton, S. R.; Otero-de-la-Roza, A.; Johnson, E. R. Exchange-Hole Dipole Dispersion Model for Accurate Energy Ranking in Molecular Crystal Structure Prediction II: Non-Planar Molecules. *J. Chem. Theory Comput.* **2017**, *13*, 5332–5342.

[103] Pople, J. A.; Head-Gordon, M.; Fox, D. J.; Raghavachari, K.; Curtiss, L. A. Gaussian-1 Theory: A General Procedure for Prediction of Molecular Energies. *J. Chem. Phys.* **1989**, *90*, 5622–5629.

[104] Curtiss, L. A.; Jones, C.; Trucks, G. W.; Raghavachari, K.; Pople, J. A. Gaussian-1 Theory of Molecular Energies for Second-Row Compounds. *J. Chem. Phys.* **1990**, *93*, 2537–2545.

[105] Curtiss, L. A.; Raghavachari, K.; Trucks, G. W.; Pople, J. A. Gaussian-2 Theory for Molecular Energies of First- and Second-Row Compounds. *J. Chem. Phys.* **1991**, *94*, 7221–7230.

[106] Curtiss, L. A.; Raghavachari, K.; Redfern, P. C.; Rassolov, V.; Pople, J. A. Gaussian-3 (G3) Theory for Molecules Containing First and Second-Row Atoms. *J. Chem. Phys.* **1998**, *109*, 7764–7776.

[107] Curtiss, L. A.; Redfern, P. C.; Raghavachari, K. Gaussian-4 Theory. *J. Chem. Phys.* **2007**, *126*, 084108.

[108] Becke, A. D. Perspective: Fifty Years of Density-Functional Theory in Chemical Physics. *J. Chem. Phys.* **2014**, *140*, 18A301.

[109] Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.

[110] Becke, A. D. On the Large-Gradient Behavior of the Density Functional Exchange Energy. *J. Chem. Phys.* **1986**, *85*, 7184–7187.

[111] Becke, A. D. Density-Functional Thermochemistry. III. The Role of Exact Exchange. *J. Chem. Phys.* **1993**, *98*, 5648–5652.

[112] Adamo, C.; Barone, V. Toward Reliable Density Functional Methods Without Adjustable Parameters: The PBE0 Model. *J. Chem. Phys.* **1999**, *110*, 6158–6170.

[113] Mori-Sánchez, P.; Cohen, A. J.; Yang, W. Localization and Delocalization Errors in Density Functional Theory and Implications for Band-Gap Prediction. *Phys. Rev. Lett.* **2008**, *100*, 146401.

[114] Cohen, A. J.; Mori-Sánchez, P.; Yang, W. Insights Into Current Limitations of Density Functional Theory. *Science* **2008**, *321*, 792–794.

[115] Henderson, T. M.; Janesko, B. G.; Scuseria, G. E. Generalized Gradient Approximation Model Exchange Holes for Range-Separated Hybrids. *J. Chem. Phys.* **2008**, *128*, 194105.

[116] Janesko, B. G.; Henderson, T. M.; Scuseria, G. E. Screened Hybrid Density Functionals for Solid-State Chemistry and Physics. *Phys. Chem. Chem. Phys.* **2009**, *21*, 443–454.

[117] Krukau, A. V.; Vydrov, O. A.; Izmaylov, A. F.; Scuseria, G. E. Influence of the Exchange Screening Parameter on the Performance of Screened Hybrid Functionals. *J. Chem. Phys.* **2006**, *125*, 224106.

[118] Monkhorst, H. J.; Pack, J. D. Special Points for Brillouin-Zone Integrations. *Phys. Rev. B* **1976**, *13*, 5188–5192.

[119] Schwerdtfeger, P. The Pseudopotential Approximation in Electronic Structure Theory. *ChemPhysChem* **2011**, *12*, 3143–3155.

[120] Hamann, D. R.; Schlüter, M.; Chiang, C. Norm-Conserving Pseudopotentials. *Phys. Rev. Lett.* **1979**, *43*, 1494–1497.

[121] Troullier, N.; Martins, J. L. Efficient Pseudopotentials for Plane-Wave Calculations. *Phys. Rev. B* **1991**, *43*, 1993–2006.

[122] Troullier, N.; Martins, J. L. Efficient Pseudopotentials for Plane-Wave Calculations. II. Operators for Fast Iterative Diagonalization. *Phys. Rev. B* **1991**, *43*, 8861–8869.

[123] Vanderbilt, D. Soft Self-Consistent Pseudopotentials in a Generalized Eigenvalue Formalism. *Phys. Rev. B* **1990**, *41*, 7892–7895.

[124] Blöchl, P. E. Projector Augmented-Wave Method. *Phys. Rev. B* **1994**, *50*, 17953–17979.

[125] Kresse, G.; Joubert, D. From Ultrasoft Pseudopotentials to the Projector Augmented-Wave Method. *Phys. Rev. B* **1999**, *59*, 1758–1775.

[126] Andersen, O. K. Linear Methods in Band Theory. *Phys. Rev. B* **1975**, *12*, 3060–3083.

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

[128] Otero-de-la-Roza, A.; Cao, B. H.; Price, I. K.; Hein, J. E.; Johnson, E. R. Predicting the Relative Solubilities of Racemic and Enantiopure Crystals by Density-Functional Theory. *Angew. Chem. Int. Ed.* **2014**, *53*, 7879–7882.

[129] Zwanziger, J. W. Computation of NMR Observables: Consequences of Projector-Augmented Wave Sphere Overlap. *Solid State Nucl. Magn. Reson.* **2016**, *80*, 14–18.

[130] Soler, J. M.; Artacho, E.; Gale, J. D.; García, A.; Junquera, J.; Ordejón, P.; Sánchez-Portal, D. The SIESTA Method for *Ab Initio* Order-*N* Materials Simulation. *J. Phys.: Condens. Matter* **2002**, *14*, 2745–2779.

[131] Artacho, E.; Gale, J. D.; García, A.; Junquera, J.; Martin, R. M.; Ordejón, P.; Pruneda, J. M.; Sánchez-Portal, D.; Soler, J. M. The SIESTA Method; Developments and Applicability. *J. Phys.: Condens. Matter* **2008**, *20*, 064208.

[132] ATOM, a program for DFT calculations in atoms and pseudopotential generation, maintained by Alberto Garcia and distributed as part of the SIESTA package. See *http://www.icmab.es/siesta/atom*.

[133] Norm-conserving pseudopotentials obtained from Abinit's pseudo database have been translated with the ATOM program for use with the SIESTA distribution. See *https://departments.icmab.es/leem/siesta/Databases/Pseudopotentials/periodictable-gga-abinit.html*.

[134] Riley, K. E.; Pitonňák, M.; Jurečka, P.; Hobza, P. Stabilization and Structure Calculations for Noncovalent Interactions in Extended Molecular Systems Based on Wave Function and Density Functional Theories. *Chem. Rev.* **2010**, *110*, 5023–5063.

[135] Klimeš, J.; Michaelides, A. Perspective: Advances and Challenges in Treating van der Waals Dispersion Forces in Density Functional Theory. *J. Chem. Phys.* **2012**, *137*, 120901.

[136] Cohen, A. J.; Mori-Sánchez and W. Yang, P. Challendges for Density Functional Theory. *Chem. Rev.* **2012**, *112*, 289–320.

[137] Otero-de-la-Roza, A.; DiLabio, G. *Non-Covalent Interactions in Quantum Chemistry and Physics*; Elsevier: Amsterdam, Netherlands, 2017.

[138] Grimme, S. Dispersion-Corrected Mean-Field Electronic Structure Methods. *Chem. Rev.* **2016**, *116*, 5105–5154.

[139] Johnson, E. R. In *Non-Covalent Interactions in Quantum Chemistry and Physics*; Otero-de-la-Roza, A., DiLabio, G., Eds.; Elsevier: Amsterdam, Netherlands, 2017; Chapter 5, pp 169–192.

[140] Hermann, J.; DiStasio Jr., R. A.; Tkatchenko, A. First-Principles Models for van der Waals Interactions in Molecules and Materials: Concepts, Theory, and Applications. *Chem. Rev.* **2017**, *117*, 4714–4758.

[141] Berland, K.; Cooper, V. R.; Lee, K.; Schröder, E.; Thonhauser, T.; Hyldgaard, P.; Lundqvist, B. I. Van der Waals Forces in Density Functional Theory: A Review of the vdW-DF Method. *Rep. Prog. Phys.* **2015**, *78*, 066501.

[142] Otero-de-la-Roza, A.; Johnson, E. R. A Benchmark for Non-Covalent Interactions in Solids. *J. Chem. Phys.* **2012**, *137*, 054103.

[143] Moellmann, J.; Grimme, S. DFT-D3 Study of Some Molecular Crystals. *J. Phys. Chem. C* **2014**, *118*, 7615–7621.

[144] Marom, N.; DiStasio, Jr., R. A.; Atalla, V.; Levchenko, S.; Reilly, A. M.; Chelikowsky, J. R.; Leiserowitz, L.; Tkatchenko, A. Many-Body Dispersion Interactions in Molecular Crystal Polymorphism. *Angew. Chem. Int. Ed.* **2013**, *52*, 6629–6632.

[145] Reilly, A. M.; Tkatchenko, A. Understanding the Role of Vibrations, Exact Exchange, and Many-Body van der Waals Interactions in the Cohesive Properties of Molecular Crystals. *J. Chem. Phys.* **2013**, *139*, 024705.

[146] Kronik, L.; Tkatchenko, A. Understanding Molecular Crystals with Dispersion-Inclusive Density Functional Theory: Pairwise Corrections and Beyond. *Acc. Chem. Res.* **2014**, *47*, 3208–3216.

[147] Chickos, J. S. Enthalpies of Sublimation after a Century of Measurement: A View as Seen Through the Eyes of a Collector. *Netsu Sokutei* **2003**, *3*, 116–124.

[148] Otero-de-la-Roza, A.; Johnson, E. R. Van der Waals Interactions in Solids Using the Exchange-Hole Dipole Moment Model. *J. Chem. Phys.* **2012**, *136*, 174109.

[149] Otero-de-la-Roza, A.; Johnson, E. R. Non-Covalent Interactions and Thermochemistry using XDM-Corrected Hybrid and Range-Separated Hybrid Density Functionals. *J. Chem. Phys.* **2013**, *138*, 204109.

[150] Otero-de-la-Roza, A.; Johnson, E. R. Predicting Energetics of Supramolecular Systems Using the XDM Dispersion Model. *J. Chem. Theory Comput.* **2015**, *11*, 4033–4040.

[151] Christian, M. S.; Otero-de-la-Roza, A.; Johnson, E. R. Surface Adsorption from the Exchange-Hole Dipole Moment Dispersion Model. *J. Chem. Theory Comput.* **2016**, *12*, 3305–3315.

[152] Christian, M. S.; Otero-de-la-Roza, A.; Johnson, E. R. Adsorption of Graphene to Nickel (111) Using the Exchange-Hole Dipole Moment Model. *Carbon* **2017**, *118*, 184–191.

[153] Christian, M. S.; Otero-de-la-Roza, A.; Johnson, E. R. Adsorption of Graphene to Metal (111) Surfaces Using the Exchange-Hole Dipole Moment Model. *Carbon* **2017**, *124*, 531–540.

[154] Salem, L. The Calculation of Dispersion Forces. *Mol. Phys.* **1960**, *3*, 441–452.

[155] Becke, A. D.; Johnson, E. R. Exchange-hole Dipole Moment and the Dispersion Interaction Revisited. *J. Chem. Phys.* **2007**, *127*, 154108.

[156] Hirshfeld, F. L. Bonded-Atom Fragments for Describing Molecular Charge Densities. *Theoret. Chim. Acta* **1997**, *44*, 129–138.

[157] Heidar-Zadeh, F.; Ayers, P. W.; Verstraelen, T.; Vinogradov, I.; Vöhringer-Martinez, E.; Bultinck, P. Information-Theoretic Approaches to Atoms-in-Molecules: Hirshfeld Family of Partioning Schemes. *J. Phys. Chem. A* **2018**, *122*, 4219–4245.

[158] Kannemann, F. O.; Becke, A. D. Atomic Volumes and Polarizabilities in Density-Functional Theory. *J. Chem. Phys.* **2012**, *136*, 034109.

[159] Becke, A. D.; Roussel, M. R. Exchange Holes in Inhomogeneous Systems: A Coordinate Space Model. *Phys. Rev. A* **1989**, *39*, 3761–3767.

[160] Becke, A. D.; Johnson, E. R. A Density-Functional Model of the Dispersion Interaction. *J. Chem. Phys.* **2005**, *123*, 154101.

[161] Johnson, E. R.; Becke, A. D. A Post-Hartree-Fock Model of Intermolecular Interactions: Inclusion of Higher-Order Corrections. *J. Chem. Phys.* **2006**, *124*, 174104.

[162] Kannemann, F. O.; Becke, A. D. Van der Waals Interactions in Density-Functional Theory: Intermolecular Complexes. *J. Chem. Theory Comput.* **2010**, *6*, 1081–1088.

[163] Grimme, S. Accurate Description of van der Waals Complexes by Density Functional Theory Including Empirical Corrections. *J. Comput. Chem.* **2004**, *25*, 1463–1473.

[164] Caldeweyher, E.; Bannwarth, C.; Grimme, S. Extension of the D3 Dispersion Coefficient Model. *J. Chem. Phys.* **2017**, *147*, 034112.

[165] Grimme, S. Semiempirical GGA-Type Density Functional Constructed with a Long-Range Dispersion Correction. *J. Comput. Chem.* **2006**, *27*, 1787–1799.

[166] Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A Consistent and Accurate *Ab Initio* Parametrization of Density Functional Dispersion Correction (DFT-D) for the 94 Elements H-Pu. *J. Chem. Phys.* **2010**, *132*, 154104.

[167] Grimme, S.; Ehrlich, S.; Goerigk, L. Effect of the Damping Function in Dispersion Corrected Density Functional Theory. *J. Comput. Chem.* **2011**, *32*, 1456–1465.

[168] Tawfik, S. A.; Gould, T.; Stampfl, C.; Ford, M. J. Evaluation of van der Waals Density Functionals for Layered Materials. *Phys. Rev. Materials* **2018**, *2*, 034005.

[169] Tkatchenko, A.; Scheffler, M. Accurate Molecular van der Waals Interactions From Ground-State Electron Density and Free-Atom Reference Data. *Phys. Rev. Lett.* **2009**, *102*, 073005.

[170] Tkatchenko, A.; DiStasio, Jr., R. A.; Car, R.; Scheffler, M. Accurate and Efficient Method for Many-Body van der Waals Interactions. *Phys. Rev. Lett.* **2012**, *108*, 236402.

[171] Bučko, T.; Lebègue, S.; Hafner, J.; Ángyán, J. G. Tkatchenko-Scheffler van der Waals Correction Method With and Without Self-Consistent Screening Applied to Solids. *Phys. Rev. B* **2013**, *87*, 064110.

[172] Ambrosetti, A.; Reilly, A. M.; DiStasio, Jr., R. A.; Tkatchenko, A. Long-Range Correlation Energy Calculated From Coupled Atomic Response Functions. *J. Chem. Phys.* **2014**, *140*, 18A508.

[173] DiStasio Jr., R. A.; Gobre, V. V.; Tkatchenko, A. Many-Body van der Waals Interactions in Molecules and Condensed Matter. *J. Phys.: Condens. Matter* **2014**, *26*, 213202.

[174] Allen, M. J.; Tozer, D. J. Helium Dimer Dispersion Forces and Correlation Potentials in Density Functional Theory. *J. Chem. Phys.* **2002**, *117*, 11113–11120.

[175] Kong, J.; Gan, Z.; Proynov, E.; Freindorf, M.; Furlani, T. R. Efficient Computation of the Dispersion Interaction with Density-Functional Theory. *Phys. Rev. A* **2009**, *79*, 042510.

[176] Dion, M.; Rydberg, H.; Schröder, E.; Langreth, D. C.; Lundqvist, B. I. Van der Waals Density Functional for General Geometries. *Phys. Rev. Lett.* **2004**, *92*, 246401.

[177] Lee, K.; Murray, E. D.; Kong, L.; Lundqvist, B. I.; Langreth, D. C. Higher-Accuracy van der Waals Density Functional. *Phys. Rev. B* **2010**, *82*, 081101.

[178] Vydrov, O. A.; Voorhis, T. V. Nonlocal van der Waals Density Functional: The Simpler the Better. *J. Chem. Phys.* **2010**, *133*, 244103.

[179] Gaus, M.; Cui, Q.; Elstner, M. Density Functional Tight Binding: Application to Organic and Biological Molecules. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2014**, *4*, 49–61.

[180] Elstner, M.; Porezag, D.; Jungnickel, G.; Elsner, J.; Haugk, M.; Frauenheim, T.; Suhai, S.; Seifert, G. Self-Consistent-Charge Density-Functional Tight-Binding Method for Simulations of Complex Materials Properties. *Phys. Rev. B* **1998**, *58*, 7260–7268.

[181] Yang, Y.; Yu, H.; York, D.; Cui, Q.; Elstner, M. Extension of the Self-Consistent-Charge Density-Functional Tight-Binding Method: Third-Order Expansion of the Density Functional Total Energy and Introduction of a Modified Effective Coulomb Interation. *J. Phys. Chem. A* **2007**, *111*, 10861–10873.

[182] Gaus, M.; Cui, Q.; Elstner, M. DFTB3: Extension of the Self-Consistent-Charge Density-Functional Tight-Binding Method (SCC-DFTB). *J. Chem. Theory Comput.* **2011**, *7*, 931–948.

[183] Gaus, M.; Goez, A.; Elstner, M. Parametrization and Benchmark of DFTB3 for Organic Molecules. *J. Chem. Theory Comput.* **2013**, *9*, 338–354.

[184] Kubař, T.; Bodrog, Z.; Gaus, M.; Aradi, B.; Frauenheim, T.; Elstner, M. Parametrization of the SCC-DFTB Method for Halogens. *J. Chem. Theory Comput.* **2013**, *9*, 2939–2949.

[185] Kubillus, M.; Kubař, T.; Gaus, M.; Řezáč, J.; Elstner, M. Parametrization of the DFTB3 Method for Br, Ca, Cl, F, I, K, and Na in Organic and Biological Systems. *J. Chem. Theory Comput.* **2015**, *11*, 332–342.

[186] Grimme, S.; Bannwarth, C.; Shushkov, P. A Robust and Accurate Tight-Binding Quantum Chemical Method for Structures, Vibrational Frequencies, and Noncovalent Interactions of Large Molecular Systems Parametrized for All spd-Block Elements (Z = 1–86). *J. Chem. Theory Comput.* **2017**, *13*, 1989–2009.

[187] Bannwarth, C.; Ehlert, S.; Grimme, S. GFN2-xTB–An Accurate and Broadly Parametrized Self-Consistent Tight-Binding Quantum Chemical Method with Multipole Electrostatics and Density-Dependent Dispersion Contributions. *J. Chem. Theory Comput.* **2019**, *15*, 1652–1671.

[188] Sure, R.; Brandenburg, J. G.; Grimme, S. Small Atomic Orbital Basis Set First-Principles Quantum Chemical Methods for Large Molecular and Periodic Systems: A Critical Analysis of Error Sources. *ChemistryOpen* **2016**, *5*, 94–109.

[189] Boys, S. F.; Bernardi, F. The Calculation of Small Molecular Interactions by the Differences of Separate Total Energies. Some Procedures with Reduced Errors. *Mol. Phys.* **1970**, *19*, 553–566.

[190] Kruse, H.; Grimme, S. A Geometrical Correction for the Inter-and Intra-Molecular Basis Set Superposition Error in Hartree-Fock and Density Functional Theory Calculations for Large Systems. *J. Chem. Phys.* **2012**, *136*, 154101.

[191] Brandenburg, J. G.; Alessio, M.; Civalleri, B.; Peintinger, M. F.; Bredow, T.; Grimme, S. Geometrical Correction for the Inter- and Intramolecular Basis Set Superposition Error in Periodic Density Functional Theory Calculations. *J. Phys. Chem. A* **2013**, *117*, 9282–9292.

[192] Sure, R.; Grimme, S. Corrected Small Basis Set Hartree-Fock Method for Large Systems. *J. Comput. Chem.* **2013**, *34*, 1672–1685.

[193] Grimme, S.; Brandenburg, J. G.; Bannwarth, C.; Hansen, A. Consistent Structures and Interactions by Density Functional Theory with Small Atomic Orbital Basis Sets. *J. Chem. Phys.* **2015**, *143*, 054107.

[194] Wheatley, P. J. The Crystal and Molecular Structure of Aspirin. *J. Chem. Soc.* **1964**, 6036–6048.

[195] Ouvrard, C.; Price, S. L. Toward Crystal Structure Prediction for Conformationally Flexible Molecules: The Headaches Illustrated by Aspirin. *Cryst. Growth Des.* **2004**, *4*, 1119–1127.

[196] Vishweshwar, P.; McMahon, J. A.; Oliveira, M.; Peterson, M. L.; Zaworotko, M. J. The Predictably Elusive Form II of Aspirin. *J. Am. Chem. Soc.* **2005**, *127*, 16802–16803.

[197] Bond, A. D.; Boese, R.; Desiraju, G. R. On the Polymorphism of Aspirin. *Angew. Chem. Int. Ed.* **2007**, *46*, 615–617.

[198] Bond, A. D.; Boese, R.; Desiraju, G. R. On the Polymorphism of Aspirin: Crystalline Aspirin as Intergrowths of Two "Polymorphic" Domains. *Angew. Chem. Int. Ed.* **2007**, *46*, 618–622.

[199] Bond, A. D.; Solanko, K. A.; Parsons, S.; Redder, S.; Boese, R. Single Crystals of Aspirin Form II: Crystallization and Stability. *CrystEngComm* **2011**, *13*, 399–401.

[200] Varughese, S.; Kiran, M. S. R. N.; Solanko, K. A.; Bond, A. D.; Ramamurty, U.; Desiraju, G. R. Interaction Anisotropy and Shear Instability of Aspirin Polymorphs Established by Nanoindentation. *Chem. Sci.* **2011**, *2*, 2236–2242.

[201] Wen, S.; Beran, G. Accidental Degeneracy in Crystalline Aspirin: New Insights from High-Level *Ab Initio* Calculations. *Cryst. Growth Des.* **2012**, *12*, 2169–2172.

[202] Huang, Y.; Shao, Y.; Beran, G. Accelerating MP2C Dispersion Corrections for Dimers and Molecular Crystals. *J. Chem. Phys.* **2013**, *138*, 224112.

[203] Chan, E. J.; Welberry, T. R.; Heerdegen, A. P.; Goosens, D. J. Diffuse Scattering Study of Aspirin Forms (I) and (II). *Acta Cryst. B* **2010**, *66*, 696–707.

[204] Crowell, E. L.; Dreger, Z. A.; Gupta, Y. M. High-Pressure Polymorphism of Acetylsalicylic Acid (Aspirin): Raman Spectroscopy. *J. Mol. Struct.* **2015**, *1082*, 29–37.

[205] Das, D.; Desiraju, G. R. Packing Modes in Some Mono- and Disubstituted Phenylpropiolic Acids: Repeated Occurence of the Rare *syn,anti* Catemer. *Chem. Asian J.* **2006**, *1–2*, 231–244.

[206] Bürgi, H.-B.; Hostettler, M.; Birkedal, H.; Schwarzenbach, D. Stacking Disorder: The Hexagonal Polymorph of Tris(Bicyclo[2.1.1]hexeno)benzene and Related Examples. *Z. Kristallogr.* **2005**, *220*, 1066–1075.

[207] Bauer, J. D.; Haussühl, E.; Winkler, B.; Arbeck, D.; Milman, V.; Robertson, S. Elastic Properties, Thermal Expansion, and Polymorphism of Acetylsalicylic Acid. *Cryst. Growth Des.* **2010**, *10*, 3132–3140.

[208] Reilly, A. M.; Tkatchenko, A. Role of Dispersion Interactions in the Polymorphism and Entropic Stabilization of the Aspirin Crystal. *Phys. Rev. Lett.* **2014**, *113*, 055701.

[209] Day, G. M.; Price, S. L. Elastic Constant Calculations for Molecular Organic Crystals. *Cryst. Growth Des.* **2001**, *1*, 13–27.

[210] Wildfong, P. L. D.; Morris, K. R.; Anderson, C. A.; Short, S. M. Demonstration of a Shear-Based Solid-State Phase Transformation in a Small Molecular Organic System: Chlorpropamide. *J. Pharm. Sci.* **2007**, *96*, 1100–1113.

[211] Roberts, R. J.; Rowe, R. C.; York, P. The Relationship Between Indentation Hardness of Organic Solids and Their Molecular Structure. *J. Mater. Sci.* **1994**, *29*, 2289–2296.

[212] Bond, A. D.; Boese, R.; Desiraju, G. R. What is a Polymorph? Aspirin as a Case Study. *Am. Pharm. Rev.* **2007**, *10*, 24–30.

[213] Pickett, W. E. Pseudopotential Methods in Condensed Matter Applications. *Comput. Phys. Rep.* **1989**, *9*, 115–197.

[214] Payne, M. C.; Teter, M. P.; Allan, D. C.; Arias, T. A.; Joannopoulos, J. D. Iterative Minimization Techniques for *Ab Initio* Total-Energy Calculations: Molecular Dynamics and Conjugate Gradients. *Rev. Mod. Phys.* **1992**, *64*, 1045–1097.

[215] Togo, A.; Tanaka, I. First Principles Phonon Calculations in Materials Science. *Scr. Mater.* **2015**, *108*, 1–5.

[216] Neto, N.; Righini, R.; Califano, S.; Walmsley, S. H. Lattice Dynamics of Molecular Crystals Using Atom–Atom and Multipole–Multipole Potentials. *Chem. Phys.* **1978**, *29*, 167–179.

[217] Fultz, B. Vibrational Thermodynamics of Materials. *Prog. Mater. Sci.* **2010**, *55*, 247–352.

[218] Perlovich, G. L.; Kurkov, S. V.; Kinchin, A. N.; Bauer-Brandl, A. Solvation and Hydration Characteristics of Ibuprofen and Acetylsalicylic Acid. *AAPS PharmSci.* **2004**, *6*, 22–30.

[219] Li, T.; Feng, S. Empirically Augmented Density Functional Theory for Predicting Lattice Energies of Aspirin, Acetaminophen Polymorphs, and Ibuprofen Homochiral and Racemic Crystals. *Pharm. Res.* **2006**, *23*, 2326–2332.

[220] Beyer, T.; Price, S. L. The Errors in Lattice Energy Minimisation Studies: Sensitivity to Experimental Variations in the Molecular Structure of Paracetemol. *CrystEngComm* **2000**, *34*, 183–190.

[221] Kim, Y.; Machida, K.; Taga, T.; Osaki, K. Structure Redetermination and Packing Analysis of Aspirin Crystal. *Chem. Pharm. Bull. (Tokyo)* **1985**, *33*, 2641–2647.

[222] Li, T. Understanding the Polymorphism of Aspirin with Electronic Calculations. *J. Pharm. Sci.* **2007**, *96*, 755–760.

[223] Charlier, J.-C.; Gonze, X.; Michenaud, J.-P. First-Principles Study of the Stacking Effect on the Electronic Properties of Graphite(s). *Carbon* **1994**, *32*, 289–299.

[224] Lipson, H.; Stokes, A. R. The Structure of Graphite. *Proc. R. Soc. A* **1942**, *181*, 101–105.

[225] Desiraju, G. R. Crystal Engineering: From Molecules to Crystal. *J. Am. Chem. Soc.* **2013**, *135*, 9952–9967.

[226] Henkelman, G.; Uberuaga, B. P.; Jónsson, H. A Climbing Image Nudged Elastic Band Method for Finding Saddle Points and Minimum Energy Paths. *J. Chem. Phys.* **2000**, *113*, 9901–9904.

[227] Henkelman, G.; Jónsson, H. Improved Tangent Estimate in the Nudged Elastic Band Method for Finding Minimum Energy Paths and Saddle Points. *J. Chem. Phys.* **2000**, *113*, 9978–9985.

[228] Krausz, A. S.; Eyring, H. Chemical Kinetics of Plastic Deformation. *J. Appl. Phys.* **1971**, *42*, 2382–2385.

[229] Riedo, E.; Gnecco, E.; Bennewitz, R.; Meyer, E.; Brune, H. Interaction Potential and Hopping Dynamics Governing Sliding Friction. *Phys. Rev. Lett.* **2003**, *91*, 084502.

[230] Szlufarska, I.; Chandross, M.; Carpick, R. W. Recent Advances in Single-Asperity Nanotribology. *J. Phys. D: Appl. Phys.* **2008**, *41*, 123001.

[231] Burke, K. Perspective on Density Functional Theory. *J. Chem. Phys.* **2012**, *136*, 150901.

[232] Ruiz, E.; Salahub, D. R.; Vela, A. Charge-Transfer Complexes: Stringent Tests for Widely Used Density Functionals. *J. Chem. Phys.* **1996**, *100*, 12265–12276.

[233] Sini, G.; Sears, J. S.; Bredas, J. L. Evaluating the Performance of DFT Functionals in Assessing the Interaction Energy and Ground-State Charge Transfer of Donor/Acceptor Complexes: Tetrathiafulvalene-Tetracyanoquinodimethane (TTF-TCNQ) as a Model Case. *J. Chem. Theory Comput.* **2011**, *7*, 602–609.

[234] Steinmann, S. N.; Piemontesi, C.; Delacht, A.; Corminboeuf, C. Why are the Interaction Energies of Charge-Transfer Complexes Challenging for DFT? *J. Chem. Theory Comput.* **2012**, *8*, 1629–1640.

[235] Cai, Z.-L.; Sendt, K.; Reimers, J. R. Failure of Density-Functional Theory and Time-Dependent Density-Functional Theory for Large Extended π Systems. *J. Chem. Phys.* **2002**, *117*, 5543–5549.

[236] Tozer, D. J. Relationship Between Long-Range Charge-Transfer Excitation Energy Error and Integer Discontinuity in Kohn–Sham Theory. *J. Chem. Phys.* **2003**, *119*, 12697–12699.

[237] Dreuw, A.; Weisman, J. L.; Head-Gordon, M. Long-Range Charge-Transfer Excited States in Time-Dependent Density Functional Theory Require Non-Local Exchange. *J. Chem. Phys.* **2003**, *119*, 2943–2946.

[238] Otero-de-la Roza, A.; Johnson, E. R.; DiLabio, G. A. Halogen Bonding from Dispersion-Corrected Density-Functional Theory: The Role of Delocalization Error. *J. Chem. Theory Comput.* **2014**, *10*, 5436–5447.

[239] Lynch, B. J.; Truhlar, D. G. How Well Can Hybrid Density Functional Methods Predict Transition State Geometries and Barrier Heights? *J. Phys. Chem. A* **2001**, *105*, 2936–2941.

[240] Janesko, B. G.; Scuseria, G. E. Hartree-Fock Orbitals Significantly Improve the Reaction Barrier Heights Predicted by Semilocal Density Functionals. *J. Chem. Phys.* **2008**, *128*, 244112.

[241] Kim, M.-C.; Park, H.; Son, S.; Sim, E.; Burke, K. Improved DFT Potential Energy Surfaces via Improved Densities. *J. Phys. Chem. Lett.* **2015**, *6*, 3802–3807.

[242] Ruzsinszky, A.; Perdew, J. P.; Csonka, G. I.; Vydrov, O. A.; Scuseria, G. E. Density Functionals That are One- and Two- are Not Always Many-Electron Self-Interaction-Free, as Shown for H_2^+, He_2^+, LiH^+, and Ne_2^+. *J. Chem. Phys.* **2007**, *126*, 104102.

[243] Ruzsinszky, A.; Perdew, J. P.; Csonka, G. I.; Vydrov, O. A.; Scuseria, G. E. Spurious Fractional Charge on Dissociated Atoms: Pervasive and Resilient Self-Interaction Error of Common Density Functionals. *J. Chem. Phys.* **2006**, *125*, 194112.

[244] Becke, A. D.; Dale, S. G.; Johnson, E. R. Correct Charge Transfer in CT Complexes From the Becke'05 Density Functional. *J. Chem. Phys.* **2018**, *148*, 211101.

[245] Kim, M.-C.; Sim, E.; Burke, K. Ions in Solution: Density Corrected Density Functional Theory (DC-DFT). *J. Chem. Phys.* **2014**, *140*, 18A528.

[246] Johnson, E. R.; Otero-de-la-Roza, A.; Dale, S. G.; DiLabio, G. A. Efficient Basis Sets for Non-Covalent Interactions in XDM-Corrected Density-Functional Theory. *J. Chem. Phys.* **2013**, *139*, 214109.

[247] Janesko, B. G. Reducing Density-Driven Error Without Exact Exchange. *Phys. Chem. Chem. Phys.* **2017**, *19*, 4793–4801.

[248] Wasserman, A.; Nafziger, J.; Jiang, K.; Kim, M.-C.; Sim, E.; Burke, K. The Importance of Being Inconsistent. *Ann. Rev. Phys. Chem.* **2017**, *68*, 555–581.

[249] Johnson, E. R.; Salamone, M.; Bietti, M.; DiLabio, G. A. Modeling Noncovalent Radical-Molecule Interactions Using Conventional Density-Functional Theory: Beware Erroneous Charge Transfer. *J. Phys. Chem. A* **2013**, *117*, 947–952.

[250] Smith, J. M.; Alahmadi, Y. J.; Rowley, C. N. Range-Separated DFT Functionals are Necessary to Model Thio-Michael Additions. *J. Chem. Theory Comput.* **2013**, *9*, 4860–4865.

[251] Maschio, L.; Usvyat, D.; Civalleri, B.
textitAb Initio Study of van der Waals and Hydrogen-Bonded Molecular Crystals With a Periodic Local-MP2 Method. *CrystEngComm* **2010**, *12*, 2429–2435.

[252] Maschio, L.; Usvyat, D.; Schütz, M.; Civalleri, B. Periodic Local Møller-Plesset Second Order Perturbation Theory Method Applied to Molecular Crystals: Study of Solid NH_3 and CO_2 Using Extended Basis Sets. *J. Chem. Phys.* **2010**, *132*, 134706.

[253] Barnes, T. A.; Kurth, T.; Carrier, P.; Wichmann, N.; Prendergast, D.; Kent, P. R. C.; Deslippe, J. Improved Treatment of Exact Exchange in Quantum ESPRESSO. *Comp. Phys. Comm.* **2017**, *214*, 52–58.

[254] Cohen, A. J.; Mori-Sanchez, P.; Yang, W. Fractional Charge Perspective on the Band Gap in Density-Functional Theory. *Phys. Rev. B.* **2008**, *77*, 115123.

[255] Tran, F.; Blaha, P. Accurate Band Gaps of Semiconductors and Insulators with a Semilocal Exchange-Correlation Potential. *Phys. Rev. Lett.* **2009**, *102*, 226401.

[256] Morissette, S. L.; Almarsson, O.; Peterson, M. L.; Remenar, J. F.; Read, M. J.; Lemmo, A. V.; Ellis, S.; Cima, M. J.; Gardner, C. R. High-Throughput Crystallization: Polymorphs, Salts, Co-Crystals and Solvates of Pharmaceutical Solids. *Adv. Drug Deliv. Rev.* **2004**, *56*, 275–300.

[257] Blagden, N.; Coles, S. J.; Berry, D. J. Pharmaceutical Co-Crystals – Are We There Yet? *CrystEngComm* **2014**, *16*, 5753–5761.

[258] Bolla, G.; Nangia, A. Pharmaceutical Cocrystals: Walking the Talk. *Chem. Commun.* **2016**, *52*, 8342–8360.

[259] Cherukuvada, S.; Kaur, R.; Guru Row, T. N. Co-Crystallization and Small Molecule Crystal Form Diversity: From Pharmaceutical to Materials Applications. *CrystEng-Comm* **2016**, *18*, 8528–8555.

[260] Duggirala, N. K.; Perry, M. L.; Almarsson, O.; Zaworotko, M. J. Pharmaceutical Cocrystals: Along the Path to Improved Medicines. *Chem. Commun.* **2016**, *52*, 640–655.

[261] Lemmerer, A.; Govindraju, S.; Johnston, M.; Motlung, X.; Savig, K. L. Co-Crystals and Molecular Salts of Carboxylic Acid/Pyridine Complexes: Can Calculated pKa's Predict Proton Transfer? A Case Study of Nine Complexes. *CrystEngComm* **2015**, *17*, 3591–3595.

[262] Childs, S. L.; Stahly, G. P.; Park, A. The Salt-Cocrystal Continuum: The Influence of Crystal Structure on Ionization State. *Mol. Pharmaceutics* **2007**, *4*, 323–338.

[263] Cruz-Cabeza, A. Acid-Base Crystalline Complexes and the pKa Rule. *CrystEngComm* **2012**, *14*, 6362–6365.

[264] Chan, H.; Kendrick, J.; Neumann, M. A.; Leusen, F. Towards *Ab Initio* Screening of Co-Crystal Formation Through Lattice Energy Calculations and Crystal Structure Prediction of Nicotinamide, Isonicotinamide, Picolinamide and Paracetamol Multi-Component Crystals. *CrystEngComm* **2013**, *15*, 3799–3807.

[265] Taylor, C. R.; Day, G. M. Evaluating the Energetic Driving Force for Cocrystal Formation. *Cryst. Growth Des.* **2018**, *18*, 892–904.

[266] Mohamed, S.; Tocher, D. A.; Vickers, M.; Karamertzanis, P. G.; Price, S. L. Salt or Cocrystal? A New Series of Crystal Structures Formed from Simple Pyridines and Carboxylic Acids. *Cryst. Growth Des.* **2009**, *9*, 2881–2889.

[267] Mohamed, S.; Tochter, D. A.; Price, S. L. Computational Prediction of Salt and Cocrystal Structures—Does a Proton Position Matter? *Int. J. Pharm.* **2011**, *418*, 187–198.

[268] Dean, J. A. *Lange's Handbook of Chemistry, Fifteenth Edition*; McGraw-Hill, Inc.: New York, NY, 1999.

[269] Martell, A. E.; Smith, R. M. *Critical Stability Constants*; Springer, 2014.

[270] Ashton, L. A.; Bullock, J. I.; Simpson, P. Effect of Temperature on the Protonation Constants of Some Aromatic, Heterocyclic Nitrogen Bases and the Anion of 8-Hydroxyquinoline. *J. Chem. Soc., Faraday Trans. 1* **1982**, *78*, 1961–1970.

[271] Dovesi, R.; Erba, A.; Orlando, R.; Zicovich-Wilson, C. M.; Civalleri, B.; Maschio, L.; Rerat, M.; Casassa, S.; Baima, J.; Salustro, S.; Kirtman, B. Quantum-Mechanical Condensed Matter Simulations with CRYSTAL. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2018**, *8*, e1360.

[272] Brandenburg, J. G.; Caldeweyher, E.; Grimme, S. Screened Exchange Hybrid Density Functional for Accurate and Efficient Structures and Interaction Energies. *Phys. Chem. Chem. Phys.* **2016**, *18*, 15519–15523.

[273] Cutini, M.; Civalleri, B.; Corno, M.; Orlando, R.; Brandenburg, J. G.; Maschio, L.; Ugliengo, P. Assessment of Different Quantum Mechanical Methods for the Prediction of Structure and Cohesive Energy of Molecular Crystals. *J. Chem. Theory Comput.* **2016**, *12*, 3340–3352.

[274] Schwartz, L. M.; Howard, L. O. Aqueous Dissociation of Squaric Acid. *J. Chem. Phys.* **1970**, *74*, 4372–4377.

[275] Wurm, F. R.; Klok, H.-A. Be Squared: Expanding the Horizon of Squaric Acid-Mediated Conjugations. *Chem. Soc. Rev.* **2013**, *42*, 8220–8236.

[276] van de Streek, J.; Neumann, M. A. Validation of Experimental Molecular Crystal Structures with Dispersion-Corrected Density Functional Theory Calculations. *Acta Cryst. B* **2010**, *66*, 544–558.

[277] Price, S. L.; Brandenburg, J. G. In *Non-Covalent Interactions in Quantum Chemistry and Physics*; Otero-de-la Roza, A., DiLabio, G. A., Eds.; Elsevier, 2017; Chapter 11, pp 333–363.

[278] Gražulis, S.; Daškevič, A.; Merkys, A.; Chateigner, D.; Lutterotti, L.; Quirós, M.; Serebryanaya, N. R.; Moeck, P.; Downs, R. T.; Le Bail, A. Crystallography Open Database (COD): An Open-Access Collection of Crystal Structures and Platform for World-Wide Collaboration. *Nucleic Acids Res.* **2012**, *40*, D420–D427.

[279] Boese, R.; Nlederprüm, N.; Bläser, D.; Maulltz, A. Single-Crystal Structure and Electron Density Distribution of Ammonia at 160 K on the Basis of X-ray Diffraction Data. *J. Phys. Chem. B* **1997**, *101*, 5794–5799.

[280] Chang, S.-S.; Westrum Jr., E. F. Heat Capacities and Thermodynamic Properties of globular Molecules. I. Adamantane and Hexamethylenetetramine. *J. Phys. Chem.* **1960**, *64*, 1547–1551.

[281] Vanderzee, C. E.; Westrum Jr., E. G. Succinic Acid. Heat Capacities and Thermodynamic Properties from 5 to 328 K. An Efficient Drying Procedure. *J. Chem. Thermodyn.* **1970**, *2*, 681–687.

[282] Carter, D. J.; Rohl, A. L. Benchmarking Calculated Lattice Parameters and Energies of Molecular Crystals Using van der Waals Density Functional. *J. Chem. Theory Comput.* **2014**, *10*, 3423–3437.

[283] Louwerse, M. J.; Rothenberg, G. Transferrable Basis Sets of Numerical Atomic Orbitals. *Phys. Rev. B* **2012**, *85*, 035108.

[284] Brandenburg, J. G.; Grimme, S. Organic Crystal Polymorphism: A Benchmark for Dispersion Corrected Mean Field Electronic Structure Methods. *Acta Crystallogr. Sect. B: Struct. Sci.* **2016**, *72*, 502–513.

[285] Hylton, R. K.; Tizzard, G. J.; Threlfall, T. L.; Ellis, A. L.; Coles, S. J.; Seaton, C. C.; Schulze, E.; Lorenz, H.; Seidel-Morgenstern, A.; Stein, M.; Price, S. L. Are the Crystal Structures of Enantiopure and Racemic Mandelic Acids Determined by Kinetics or Thermodynamics? *J. Am. Chem. Soc.* **2015**, *137*, 11095–11104.

[286] Artacho, E.; Sánchez-Portal, D.; Ordejón, P.; García, A.; Soler, J. M. Linear-Scaling Ab-Initio Calculations for Large and Complex Systems. *Phys. Stat. Sol.* **1999**, *215*, 809–817.

[287] Junquera, J.; Paz, O.; Sánchez-Portal, D.; Artacho, E. Numerical Atomic Orbitals for Linear-Scaling Calculations. *Phys. Rev. B* **2001**, *64*, 235111.

[288] Sánchez-Portal, D.; Ordejón, P.; Artacho, E.; Soler, J. M. Density-Functional Method for Very Large Systems with LCAO Basis Sets. *Int. J. Quantum Chem.* **1997**, *65*, 453–461.

[289] Peverati, R.; Baldridge, K. K. Implementation and Performance of DFT-D with Respect to Basis Set and Functional for Study of Dispersion Interactions in Nanoscale Aromatic Hydrocarbons. *J. Chem. Theory. Comput.* **2008**, *4*, 2030–2048.

[290] Bachelet, G. B.; Hamann, D. R.; Schlüter, M. Pseudopotentials That Work: From H to Pu. *Phys. Rev. B* **1982**, *26*, 4199–4228.

[291] Chapman, C.; Ting, E. C.; Kereszti, A.; Paci, I. Self-Assembly of Cysteine Dimers at the Gold Surface: A Computational Study of Competing Interactions. *J. Phys. Chem. C* **2013**, *117*, 19426–19435.

[292] Aradi, B.; Hourahine, B.; Frauenheim, T. DFTB+, a Sparse Matrix-Based Implementation of the DFTB Method. *J. Phys. Chem. A* **2007**, *111*, 5678–5684.

[293] Otero-de-la-Roza, A.; Johnson, E. R.; Luaña, V. CRITIC2: A Program for Real-Space Analysis of Quantum Chemical Interactions in Solids. *Comput. Phys. Commun.* **2014**, *185*, 1007–1018.

[294] de Gelder, R.; Wehrens, R.; Hageman, J. A. A Generalized Expression for the Similarity of Spectra: Application to Powder Diffraction Pattern Classification. *J. Comput. Chem.* **2001**, *22*, 273–289.

[295] Buchholz, H. K.; Hylton, R. K.; Brandenburg, J. G.; Seidl-Morgenstern, A.; Lorenz, H.; Stein, M.; Price, S. L. Thermochemistry of Racemic and Enantiopure Organic Crystals for Predicting Enantiomer Separation. *Cryst. Growth Des.* **2017**, *17*, 4676–4686.

[296] Rougeot, C.; Situ, H.; Cao, B. H.; Vlachos, V.; Hein, J. E. Automated Reaction Progress Monitoring of Heterogeneous Reactions: Crystallization-Induced Stereoselectivity in Amine-Catalyzed Aldol Reactions. *React. Chem. Eng.* **2017**, *2*, 226–231.

[297] Hein, J. E.; Tse, E.; Blackmond, D. G. A Route to Enantiopure RNA Precursors From Nearly Racemic Starting Materials. *Nat. Chem.* **2011**, *3*, 704–705.

[298] Price, I. K. Investigation and Application of Crystal Phase Behavior to Obtain Enantiopure Chemicals. Ph.D. thesis, University of California, Merced. Available from http://eprints.cdlib.org/uc/item/24r9v8mc, 2014.

[299] Stone, A. J. Distributed Multipole Analysis, or How to Describe a Molecular Charge Distribution. *Chem. Phys. Lett.* **1981**, *83*, 233–239.

[300] Tao, J.; Zheng, F.; Gebhardt, J.; Perdew, J. P.; Rappe, A. M. Screened van der Waals Correction to Density Functional Theory for Solids. *Phys. Rev. Materials* **2017**, *1*, 020802(R).

[301] Caldeweyher, E.; Brandenburg, J. G. Simplified DFT Methods for Consistent Structures and Energies of Large Systems. *J. Phys.: Condens. Matter* **2018**, *30*, 213001.

[302] Dolgonos, G. A.; Loboda, O. A.; Boese, A. D. Development of Embedded and Performance of Density Functional Methods for Molecular Crystals. *J. Phys. Chem. A* **2018**, *122*, 708–713.

[303] Louie, S. G.; Froyen, S.; Cohen, M. L. Nonlinear Ionic Pseudopotentials in Spin-Density-Functional Calculations. *Phys. Rev. B* **1982**, *26*, 1738–1742.

[304] Johnson, D. D. Modified Broyden's Method for Accelerating Convergence in Self-Consistent Calculations. *Phys. Rev. B* **1988**, *38*, 807–813.

[305] Becke, A. D.; Dickson, R. M. Numerical-Solution of Schrodinger-Equation in Polyatomic-Molecules. *J. Chem. Phys.* **1990**, *92*, 3610–3612.

[306] Goedecker, S.; Teter, M.; Hutter, J. Separable Dual-Space Gaussian Pseudopotentials. *Phys. Rev. B* **1996**, *54*, 1703–1710.

[307] Hartwigsen, C.; Goedecker, S.; Hutter, J. Relativistic Separable Dual-Space Gaussian Pseudopotentials from H to Rn. *Phys. Rev. B* **1998**, *58*, 3641–3662.

[308] Lacks, D. J.; Gordon, R. G. Pair Interactions of Rare-Gas Atoms as a Test of Exchange-Energy-Density Functionals in Regions of Large Density Gradients. *Phys. Rev. A* **1993**, *47*, 4681–4690.

[309] Zhang, Y.; Pan, W.; Yang, W. Describing van der Waals Interaction in Diatomic Molecules with Generalized Gradient Approximations: The Role of the Exchange Functional. *J. Chem. Phys.* **1997**, *107*, 7921–7925.

[310] Parrish, R. M.; Burns, L. A.; Smith, D.; Simmonett, A. C.; DePrince, A. E.; Hohenstein, E. G.; Bozkaya, U.; Sokolov, A. Y.; Di Remigio, R.; Richard, R. M. Psi4 1.1: An Open-Source Electronic Structure Program Emphasizing Automation, Advanced Libraries, and Interoperability. *J. Chem. Theory Comput.* **2017**, *13*, 3185–3197.

[311] Johnson, E. R.; Otero-de-la Roza, A.; Dale, S. G. Extreme Density-Driven Delocalization Error for a Model Solvated-Electron System. *J. Chem. Phys.* **2013**, *139*, 184116.

[312] Mackie, I. A.; DiLabio, G. A. Approximations to Complete Basis Set-Extrapolated, Highly Correlated Non-Covalent Interaction Energies. *J. Chem. Phys.* **2011**, *135*, 134318.

[313] Björkman, T.; Gulans, A.; Krasheninnikov, A. V.; Nieminen, R. M. Van der Waals Bonding in Layered Compounds from Advanced Density-Functional First-Principles Calculations. *Phys. Rev. Lett.* **2012**, *108*, 235502.

[314] Zacharia, R.; Ulbricht, H.; Hertel, T. Interlayer Cohesive Energy of Graphite from Thermal Desorption of Polyaromatic Hydrocarbons. *Phys. Rev. B* **2004**, *69*, 155406.

[315] Liu, H.; Neal, A. T.; Zhu, Z.; Luo, Z.; Xu, X.; Tománek, D.; Ye, P. D. Phosphorene: An Unexplored 2D Seminconductor with a High Hole Mobility. *ACS Nano* **2014**, *8*, 4033–4041.

[316] Kou, L.; Chen, C.; Smith, S. C. Phosphorene: Fabrication, Properties, and Applications. *J. Phys. Chem. Lett.* **2015**, *6*, 2794–2805.

[317] Carvalho, A.; Wang, M.; Zhu, X.; Rodin, A. S.; Su, H.; Castro Neto, A. H. Phosphorene: From Theory to Applications. *Nat. Rev. Mater.* **2016**, *1*, 16061.

[318] Jing, Y.; Zhang, X.; Zhou, Z. Phosphorene: What Can We Know from Computations. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2016**, *6*, 5–19.

[319] Batmunkh, M.; Bat-Erdene, M.; Shapter, J. G. Phosphorene and Phosphorene-Based Materials – Prospects for Future Applications. *Adv. Mater.* **2016**, *28*, 8586–8617.

[320] Sansone, G.; Karttunen, A. J.; Usvyat, D.; Schutz, M.; Brandenburg, J. G.; Maschio, L. On the Exfoliation and Anisotropic Thermal Expansion of Black Phosphorus. *Chem. Commun.* **2018**, *54*, 9793–9796.

[321] Schulenburger, L.; Baczewski, A. D.; Zhu, Z.; Guan, J.; Tománek, D. The Nature of the Interlayer Interaction in Bulk and Few-Layer Phosphorus. *Nano Lett.* **2017**, *15*, 8170–8175.

[322] Sanone, G.; Maschio, L.; Usvyat, D.; Schutz, M.; Karttunen, A. Toward an Accurate Estimate of the Exfoliation Energy of Black Phosphorus: A Periodic Quantum Chemical Approach. *J. Phys. Chem. Lett.* **2016**, *7*, 131–136.

[323] Schutz, M.; Maschio, L.; Karttunen, A. J.; Usvyat, D. Exfoliation Energy of Black Phosphorus Revisited: A Coupled Cluster Benchmark. *J. Phys. Chem. Lett.* **2017**, *8*, 1290–1294.

[324] Brown, A.; Rundqvist, S. Refinement of the Crystal Strcuture of Black Phosphorus. *Acta. Cryst.* **1965**, *19*, 684–685.

[325] Schulenburger, L.; Mattson, T. R. Quantum Monte Carlo Applied to Solids. *Phys. Rev. B* **2013**, *88*, 245117.

[326] Mostaani, E.; Drummond, N. D. Quantum Monte Carlo Calculation of the Binding Energy of Bilayer Graphene. *Phys. Rev. Lett.* **2015**, *115*, 115501.

[327] Price, S. L.; Reutzel-Edens, S. M. The Potential of Computed Crystal Energy Landscapes to Aid Solid-Form Development. *Drug Discovery Today* **2016**, *21*, 912–923.

[328] Pulido, A.; Chen, L.; Kaczorowski, T.; Holden, D.; Little, M. A.; Chong, S. Y.; Slater, B. J.; McMahon, D. P.; Bonillo, B.; Stackhouse, C. J.; Stephenson, A.; Kane, C. M.; Clowes, R.; Hasell, T.; Cooper, A. I.; Day, G. M. Functional Materials Discovery Using Energy–Structure–Function Maps. *Nature* **2017**, *543*, 657–664.

[329] Yang, J.; De, S.; Campbell, J. E.; Li, S.; Ceriotti, M.; Day, G. M. Large-Scale Computational Screening of Molecular Organic Semiconductors Using Crystal Structure Prediction. *Chem. Mater.* **2018**, *30*, 4361–4371.

[330] Loboda, O. A.; Dolgonos, G. A.; Boese, A. D. Towards Hybrid Density Functional Calculations of Molecular Crystals via Fragment-Based Methods. *J. Chem. Phys.* **2018**, *149*, 124104.

[331] Iuzzolino, L.; McCabe, P.; Price, S. L.; Brandenburg, J. G. Crystal Structure Prediction of Flexible Pharmaceutical-Like Molecules: Density Functional Tight-Binding as an Intermediate Optimisation Method and For Free Energy Estimation. *Faraday Discuss.* **2018**, *211*, 275–296.

[332] Mortazavi, M.; Brandenburg, J. G.; Maurer, R. J.; Tkatchenko, A. Structure and Stability of Molecular Crystals with Many Body Dispersion Inclusive Density Functional Tight Binding. *J. Phys. Chem. Lett.* **2018**, *9*, 399–405.

[333] Dolgonos, G. A.; Boese, A. D. Adjusting Dispersion Parameters for the Density-Functional Tight-Binding Description of Molecular Crystals. *Chem. Phys. Lett.* **2019**, *718*, 7–11.

[334] Cutini, M.; Civalleri, B.; Ugliengo, P. Cost-Effective Quantum Mechanical Approach for Predicting Thermodynamic and Mechanical Stability of Pure-Silica Zeolites. *ACS Omega* **2019**, *4*, 1838–1846.

[335] McDonagh, D.; Skylaris, C.; Day, G. M. Machine-Learned Fragment-Based Energies for Crystal Structure Prediction. *J. Chem. Theory Comput.* **2019**, *15*, 2743–2758.

[336] Hulme, A. T.; Price, S. L.; Tocher, D. A. A New Polymorph of 5-Fluorouracil Found Following Computational Crystal Structure Predictions. *J. Am. Chem. Soc.* **2005**, *127*, 1116–1117.

[337] Braun, D. E.; Ardid-Candel, M.; D'Oria, E.; anf J.-B. Arlin, P. G. K.; Florence, A. J.; Jones, A. G.; Price, S. L. Racemic Naproxen: A Multidisciplinary Structural and Thermodynamic Comparison with the Enantiopure Form. *Cryst. Growth Des.* **2011**, *11*, 5659–5669.

[338] Bhardwaj, R. M.; Price, L. S.; Price, S. L.; Reutzel-Edens, S. M.; Miller, G. J.; Oswald, I.; Florence, B. F.; Florence, A. J. Exploring the Experimental and Computed Energy Landscape of Olanzapine. *Cryst. Growth Des.* **2013**, *13*, 1602–1617.

[339] The CPOSS (Control and Prediction of the Organic Solid State) database holds details of the hypothetical crystal structures generated in the computational searches carried out at University College London, and is maintained by Dr. Louise S. Price, and led by Prof. Sally L. Price. See *http://www.chem.ucl.ac.uk/cposs/index.htm*.

[340] Srirambhatla, V. K.; Guo, R.; Price, S. L.; Florence, A. J. Isomorphous Template Induced Crystallisation: A Robust Method of the Targeted Crystallisation of Computationally Predicted Metastable Polymorphs. *Chem. Commun.* **2016**, *52*, 7384–7386.

[341] Heidelberger, C.; Chaudhuri, N. R.; Danneberg, P.; Mooren, D.; Griesbach, L. Fluorinated Pyrimidines, A New Class of Tumour-Inhibitory Compounds. *Nature* **1957**, *179*, 663–666.

[342] Grem, J. L. 5-Fluorouracil: Forty-Plus and Still Ticking. A Review of its Preclinical and Clinical Development. *Invest. New Drugs* **2000**, *18*, 299–313.

[343] Fallon III, L. The Crystal and Molecular Structure of 5-Fluorouracil. *Acta. Cryst. B* **1973**, *29*, 2549–2556.

[344] Hamad, S.; Moon, C.; Catlow, C.; Hulme, A. T.; Price, S. L. Kinetic Insights into the Role of the Solvent in the Polymorphism of 5-Fluorouracil from Molecular Dynamics Simulations. *J. Phys. Chem. B* **2006**, *110*, 3323–3329.

[345] Price, S. L.; Hamad, S.; Torrisi, A.; Karamertzanis, P. G.; Leslie, M.; Catlow, C. Applications of Dl_poly And Dl_multi to Organic Molecular Crystals. *Mol. Simul.* **2006**, *32*, 985–997.

[346] Barnett, S. A.; Hulme, A. T.; Issa, N.; Lewis, T. C.; Price, L. S.; Tochter, D. A.; Price, S. L. The Observed and Energetically Feasible Crystal Structures of 5-Substituted Uracils. *New J. Chem.* **2008**, *32*, 1761–1775.

[347] Karamertzanis, P. G.; Raiteri, P.; Parrinello, M.; Leslie, M.; Price, S. L. The Thermal Stability of Lattice-Energy Minima of 5-Fluorouracil: Metadynamics as an Aid to Polymorph Prediction. *J. Phys. Chem. B* **2008**, *112*, 4298–4308.

[348] Bui, T.; Dehaoui, S.; Lecomte, C.; Desiraju, G. R.; Espinosa, E. The Nature of Halogen–Halogen Interactions: A Model Derived from Experimental Charge-Density Analysis. *Angew. Chem. Int. Ed.* **2009**, *48*, 3838–3840.

[349] Baker, R. J.; Colavita, P. E.; Murphy, D. M.; Wallis, J. D. Fluorine–Fluorine Interactions in the Solid State: An Experimental and Theoretical Study. *J. Phys. Chem. A* **2012**, *116*, 1435–1444.

[350] Politzer, P.; Murray, J. S.; Clark, T. Halogen Bonding and Other σ-Hole Interactions: a Perspective. *Phys. Chem. Chem. Phys.* **2013**, *15*, 11178–11189.

[351] Jarzembska, K. N.; Kubsik, M.; Kamiński, R.; Woźniak, K.; Dominiak, P. M. From a Single Molecule to Molecular Crystal Architectures: Structural and Energetic Studies of Selected Uracil Derivatives. *Cryst. Growth Des.* **2012**, *12*, 2508–2524.

[352] Metrangolo, P.; Murray, J. S.; Pilati, T.; Politzer, P.; Resnati, G.; Terraneo, G. The Fluorine Atom as a Halogen Bond Donor, *viz.* a Positive Site. *CrystEngComm* **2011**, *13*, 6593–6596.

[353] Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, O.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. Gaussian 09 Revision B.01. Gaussian Inc. Wallingford CT 2010.

[354] Otero-de-la-Roza, A.; Johnson, E. R. Non-Covalent Interactions and Thermochemistry Using XDM-Corrected Hybrid and Range-Separated Hybrid Density Functionals. *J. Chem. Phys.* **2013**, *138*, 204109.

[355] Ravikumar, K.; Rajan, S. S.; Pattabhi, V. Structure of Naproxen, $C_{14}H_{14}O_3$. *Acta Cryst. C* **1985**, *41*, 280–282.

[356] Kim, Y. B.; Song, H. J.; Park, I. Y. Refined of the Structure of Naproxen, (+)-6-Methoxy-α-methyl-2-naphthaleneacetic Acid. *Arch. Pharm. Res.* **1987**, *10*, 232–238.

[357] Rustichelli, C.; Gamberini, G.; Ferioli, V.; Gamberini, M. C.; Ficarra, R.; Tommasini, S. Solid-State Study of Polymorphic Drugs: Carbamazepine. *J. Pharm. Biomed. Anal.* **2000**, *23*, 41–54.

[358] Lang, M.; Kampf, J. W.; Matzger, A. J. Form IV of Carbamazepine. *J. Pharm. Sci.* **2002**, *91*, 1186–1190.

[359] Arlin, J.; Price, L. S.; Price, S. L.; Florence, A. J. A Strategy for Producing Predicted Polymorphs: Catemeric Carbamazepine Form V. *Chem. Commun.* **2011**, *47*, 7074–7076.

[360] Grezsiak, A. L.; Lang, M. D.; Kim, K.; Matzger, A. J. Comparison of the Four Anhydrous Polymorphs of Carbamazepine and the Crystal Structure of Form I. *J. Pharm. Sci.* **2003**, *92*, 2260–2271.

[361] Clout, A. E.; Buanz, A.; Gaisford, S.; Williams, G. R. Polymorphic Phase Transitions in Carbamazepine and 10,11-Dihydrocarbamazepine. *Chem. Eur. J.* **2018**, *24*, 13573–13581.

[362] Karamertzanis, P. G.; Price, S. L. Energy Minimization of Crystal Structures Containing Flexible Molecules. *J. Chem. Theory Comput.* **2006**, *2*, 1184–1190.

[363] Florence, A. J.; Johnston, A.; Price, S. L.; Nowell, H.; Kennedy, A. R.; Shankland, N. An Automated Parallel Crystallisation Search for Predicted Crystal Structures and Packing Motifs of Carbamazepine. *J. Pharm. Sci.* **2006**, *95*, 1918–1930.

[364] Cruz-Cabeza, A. J.; Day, G. M.; Motherwell, W. D. S.; Jones, W. Amide Pyramidalization in Carbamazepine: A Flexibility Problem in Crystal Structure Prediction? *Cryst. Growth Des.* **2006**, *6*, 1858–1866.

[365] Florence, A. J.; Leech, C. K.; Shankland, N.; Johnston, A. Control and Prediction of Packing Motifs: A Rare Occurence of Carbamazepine in a Catameric Configuration. *CrystEngComm* **2006**, *8*, 746–747.

[366] Fulton, B.; Goa, K. L. Olanzapine A Review of its Pharmacological Properties and Therapeutic Efficacy in the Management of Schizophrenia and Related Psychoses. *Drugs* **1997**, *53*, 281–298.

[367] Tollefson, G. D.; Beasley Jr., C. M.; Tran, P. V.; Street, J. S.; Krueger, J. A.; Tamura, R. N.; Graffeo, K. A.; Thieme, M. E. Olanzapine Versus Haloperidol in the Treatment of Schizophrenia and Schizoaffective and Schizophreniform Disorders: Results of an International Collaborative Trial. *Am. J. Psychiatry* **1997**, *154*, 457–465.

[368] Askin, S.; Cockcroft, J. K.; Price, L. S.; Gonçalves, A. D.; Zhao, M.; Tocher, D. A.; Williams, G. R.; Gaisford, S.; Craig, D. Olanzapine Form IV: Discovery of a New Polymorphic Form Enabled by Computed Crystal Energy Landscapes. *Cryst. Growth Des.* **2019**, DOI:10.1021/acs.cgd.8b01881.

[369] Reutzel-Edens, S. M.; Bush, J. K.; Magee, P. A.; Stephenson, G. A.; Byrn, S. R. Anhydrates and Hydrates of Olanzapine: Crystallization, Solid-State Characterization, and Structural Relationships. *Cryst. Growth Des.* **2003**, *3*, 897–907.

[370] Testa, C. G.; Prado, L. D.; Costa, R. N.; Costa, M. L.; Linck, Y. G.; Monti, G. A.; Cuffini, S. L.; Rocha, H. Challenging Indentification of Polymorphic Mixture: Polymorps I, II and III in Olanzapine Raw Materials. *Int. J. Pharm.* **2019**, *556*, 125–135.

[371] Thakuria, R.; Nangia, A. Polymorphic Form IV of Olanzapine. *Acta Cryst. C* **2011**, *67*, o461–o463.

[372] Luo, H.; Hao, X.; Gong, Y.; Zhou, J.; He, X.; Li, J. Rational Crystal Polymorphism Design of Olanzapine. *Cryst. Growth Des.* **2019**, DOI:10.1021/acs.cgd.9b00068.

[373] Brandenburg, J. G.; Potticary, J.; Sparkes, H. A.; Price, S. L.; Hall, S. R. Thermal Expansion of Carbamazepine: Systematic Crystallographic Measurements Challenge Quantum Chemical Calculations. *J. Phys. Chem. Lett.* **2017**, *8*, 4319–4324.

[374] McKinley, J. L.; Beran, G. Identifying Pragmatic Quasi-Harmonic Electronic Structure Approaches for Modeling Molecular Crystal Thermal Expansion. *Faraday Discuss.* **2018**, *211*, 181–207.

[375] Nyman, J.; Reutzel-Edens, S. M. Crystal Structure Prediction is Changing From Basic Science to Applied Technology. *Faraday Discuss.* **2018**, *211*, 459–476.