NEW IMPLEMENTATION OF THE EXCHANGE-HOLE DIPOLE MOMENT DISPERSION METHOD FOR LARGE-SCALE MATERIALS SIMULATIONS

by

Alastair J. A. Price

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

at

Dalhousie University
Halifax, Nova Scotia
April 2023

Do not go gentle into that good night,
Old age should burn and rave at close of day;
Rage, rage against the dying of the light.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS AND SYMBOLS USED

Abbreviation	Description
3B-69	Set of molecular Trimers
a.u.	Atomic Unit
ADF	Amsterdam Density Functional Code
AE	All-Electron
APW	Augmented Planewave Method
ATM	Axilrod-Teller-Muto Three-Body Dispersion-Energy Term
aug-cc-pVTZ	Augmented Correlation-Consistent Polarized Triple-Zeta Dunning Type Basis Set
B86bPBE	Becke's B86b Exchange and PBE GGA Density Functional
B86bPBE-nX	Global Hybrid form of B86bPBE (e.g., n=25%, 50% exact exchange)
B88	Becke's B88 Exchange Functional
BE	Binding Energy
BJ	Becke-Johnson
BLYP	Becke's B88 Exchange and Lee, Yang, and Parr's Correlation Functional
BR	Becke-Roussel
BSIE	Basis-Set Incompleteness Error
BSSE	Basis-Set Superposition Error
BTn	Blind Test (e.g., n= 1–7)
CCDC	Cambridge Crystallographic Data Centre
CCSD(T)	Coupled Cluster Singles Doubles Triples
CP	Counterpoise
CPU	Computer Processing Unit
CSP	Crystal Structure Prediction
dDsC	density-dependent correction
DFA	Density Functional Approximation
DFT(-D)	(Dispersion-Corrected) Density-Functional Theory
Dn	Grimme's Series of Dispersion Corrections (e.g., n= 2, 3, 4)
DMACRYS	Distributed Multipole Analysis for Crystals
DMC	Diffusion Monte Carlo

Abbreviation	Description
DMoL^3	Density Functional Calculations on Molecules
DNA	Deoxyribonucleic Acid
EDA	Energy Decomposition Analysis
Expt	Experimental
FI	Fractionally Ionic
G	Gigabit
gCP	Geometrical Counterpoise Method
GGA	Generalized-Gradient Approximation
GHz	Gigahertz
GTO	Gaussian-Type Orbitals
FHI-aims	Fritz Haber Institute *ab initio* molecular simulations
HF	Hartree-Fock Theory
HSE06	Heyd-Scuseria–Ernzerhof Variable Exchange Hybrid Functional
ICE10	Benchmark of 10 Ice Polymorphs
ICE13	Benchmark of 13 Ice Polymorphs
KB(49)	Kannemann-Becke Benchmark Set of Molecular Dimers
KS	Kohn and Sham
LAPW	Linearised Augmented Planewave Methods
LDA	Local Density Approximation
LDOS	Local Density of States
LJ	Lennard-Jones
LSDA	Local Spin Density Approximation
M06-L	Local Variant of the Minnesota 2006 Functional
MAE	Mean Absolute Error
ME	Mean Error
MAPE	Mean Absolute Percent Error
MBD	Many-Body Dispersion Model
MP	Monkhorst-Pack k-point Grid Sampling Scheme
MP2	Møller Plesset Second-Order Perturbation Theory
NAO	Numerical Atomic Orbitals
NUMOL	NUmerical MOLecules
PAW	Projector-Augmented Wave
PBE	Perdew-Burke-Ernzerhof GGA Density Functional
PBE0	Perdew-Burke-Ernzerhof Hybrid Density Functional

Abbreviation	Description
PBE50	Perdew-Burke-Ernzerhof Hybrid Density Functional with 50% Exact Exchange
PBEsol	Revised PBE for Solids
PLATO	Package for Linear-combination of ATomic Orbitals
PP	Pseudopotential
PW86	Perdew Wang's 1986 Exchange Functional
PW91	Perdew-Wang 91 Functional
QE	Quantum ESPRESSO
QMC	Quantum Monte Carlo
revPBE	Revised PBE Functional
RMSP	Root-Mean-Square Percent
RPA	Random Phase Approximation
r^2SCAN	Revision of the SCAN Functional
rsSCS	Range-Separated Self-Consistent Screening
rVV10	Revised Vydrov and van Voorhis Non-Local Density Functional
Ry	Rydberg
S12L	Grimme's Set of Large Complexes
S22x5	Set of Non-Covalent Complexes
S66x8	Set of Non-Covalent Complexes
SCAN	Strongly Constrained and Appropriately Normed Functional
SCF	Self-Consistent Field
SCS	Self-Consistent Screening
SIESTA	Spanish Initiative for Electronic Simulations with Thousands of Atoms
STM	Scanning Tunnelling Microscopy
STS	Scanning Tunnelling Spectroscopy
STO	Slater-Type Orbitals
TMDC	Transition Metal Dichalcogenides
TPPT	Terpyridine-Phenyl-Phenyl-Terpyridine
TPSS	Tao, Perdew, Staroverov, and Scuseria's Exchange-Correlation Functional
TS	Tkatchenko-Scheffler Dispersion Model
TSsurf	Variant of TS Dispersion Model for Surfaces
UEG	Uniform Electron Gas
vdW	van der Waals
VSXC	Van Voorhis and Scuseria's Exchange-Correlation Functional
VV	Vydrov and van Voorhis

Abbreviation	Description
X23	Benchmark Set of Molecular Organic Solids
XDM	Exchange-Hole Dipole Moment Dispersion Model

Symbol	Description
a, b, c	Unit-Cell Parameters
a_1, a_2	Becke-Johnson Damping Parameters
a_X	Fraction of Exact Exchange
$\mathbf{a}$	Unit-Cell Vector
A, a, b	Becke-Roussel Exchange-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., $n = 6, 8, 10$)
CN	Coordination Number in D3 Dispersion Model
c_X	Constant Defined in LSDA 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
E_c	Correlation Energy
E^{cut}	Planewave Kinetic-Energy Cutoff
E_{disp}	Dispersion Energy
E_{DFT}	Total DFT Energy
E_{ex}	Exfoliation Energy
E_x^{GGA}	GGA Exchange Energy
E_x^{LSDA}	LSDA Exchange Energy
E_{latt}	Lattice Energy
E^{nl}	Non-Local Energy
E_X	Exchange Energy
E_X^{HF}	Hartree-Fock Exact Exchange Energy
E_X^{hybrid}	Hybrid Exchange Energy
E_{XC}	Exchange-Correlation Energy
E_μ	Energy of State without Tunnelling
$f(E)$	Fermi Function
$f(R_{ab\mathbf{L}})$	Damping Function

Symbol	Description
$\mathbf{G}$	Reciprocal Lattice Vector
$\hat{H}$	Electronic Hamiltonian
$\hbar$	Reduced Planck's Constant
$h_{X\sigma}$	Spin-Dependant Exchange Hole
i, j	Electronic Index
$\mathrm{I_{A,B}}$	Atomic Ionization Energies
J	Classical Coulomb Self Energy
$\mathbf{k}$	Reciprocal Space Vector
∇^2	Laplacian Operator
m	Atomic Mass
M_l	l-th Order Multipole Moment
$M_{\mu\nu}$	Tunnelling Matrix Elements
N	Normalization Factor or Number
$\tilde{p}_i$	Projector Functions
e	Electron Charge
$\mathbf{q}$	Reciprocal Space Vector
Q_σ	Exchange Hole Curvature
r	Radius
r_{12}	Interelectronic Separation
r_c	Radial Cutoff
r_{cl}	Radial Cutoff for Angular Momentum l
$\mathbf{R}$	Bravais Lattice Vector
R_{ab}	Interatomic Distance
$R_{ab\mathbf{L}}$	Interatomic Distance in a Periodic System
R_{vdW}	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
$S_{n,n'}$	Overlap Matrix of Basis Functions
$\hat{T}$	Transform Operator for PAW method
T_{KS}	Kinetic-Energy of Non-Interacting Electrons
$u(r)$	Radial Function of NAOs

Symbol	Description
$\hat{U}$	Electron-Electron Interaction Energy
$\mathcal{V}$	Periodic Potential
V	Volume
V_{ext}	External Potential
Y_{lm}	Spherical Harmonics for Basis Sets
α_a	Atom-in-Solid Polarizability
α_a^{at}	Free-Atom Polarizability
β, γ	Constants Defined in B86bPBE Enhancement Factor Expression
Γ	Brillouin-Zone Center
δ	Convergence Criteria for Dispersion Energy in Periodic Systems
δ_{ij}	Kronecker Delta
ϵ	Potential Well Depth
ε_i	Orbital Energy
ε_{X}	Exchange Energy Density
ζ	Gaussian Exponential Coefficient
η, ξ	Cartesian Directions
κ	Constant Defined in PBE Enhancement Factor Expression
ρ	Electron Density
μ	Constant Defined in PBE Energy and Enhancement Factor Expressions
μ, ν	Atomic-Orbital Indices
$\mu_{A,B}$	Dipole Moment Integral
σ	Electron Spin Index
σ_{disp}	van der Waals Contribution to Cell Stress
τ_σ	Positive-Definite Spin-Kinetic-Energy Density
ϕ	Atomic Orbital
$\tilde{\phi}$	Pseudo Partial Wave
Φ	Non-Local Correlation Kernel
φ	Basis Functions
ψ	Electronic Wavefunction
Ψ	Total Electronic Wavefunction
$\tilde{\Psi}$	Pseudised Wavefunction
ψ_{k}	Bloch Wavefunction
ω_{Hirsh}	Hirshfeld Weight

ABSTRACT

Intermolecular interactions are the forces that exist between molecules and are fundamental to many aspects of chemistry and the natural sciences, including the determination of the phases of matter. A correct theoretical description of intermolecular interactions at the quantum mechanical level is required for quantitative understanding of chemistry. While highly accurate wavefunction-based methods exist, they are impractical or impossible to apply to the large molecular systems relevant to most areas of chemistry, making dispersion-corrected density-functional theory (DFT) the standard class of methods for such applications. This thesis first characterised the shortcomings and laid out the requirements for high accuracy within dispersion-corrected DFT approaches, and highlighted some advantages of the exchange-hole dipole moment (XDM) dispersion method. This thesis then focused on the implementation of XDM within the Fritz Haber Institute *ab initio* materials simulations package (FHI-aims), which allowed for the first time the routine use of XDM-corrected hybrid functionals for the study of molecular solids. Using a selection of common molecular and solid-state benchmarks as test systems allowed for the validation of the XDM model within FHI-aims. Specifically, XDM-corrected hybrid functionals were shown to yield unprecedented accuracy for predicting energies of forming crystals from their constituent molecules, as well as remarkable predictive capacity for stable crystal packing motifs, or polymorphs, of organic molecules in the solid state. Finally, these methods were then applied to the compounds forming the 7th blind test of first-principles molecular crystal structure prediction. The implementation of XDM within the FHI-aims package presented within this work now allows for the accurate and efficient computational study of a new and interesting range of important chemical compounds and materials.

ACKNOWLEDGEMENTS

I would like to first start with Erin, the impact that you have made on my life is more than I can put to words. I have grown as a person and a scientist thanks to your mentorship and friendship over the last number of years, without I wouldn't be nearly half the person I am today. Thank you. I have in you a lifelong friend.

To Brienne, Mariah, Bizou, Pip, and Dylan, you are some of my closest friends, although we might be separated by vast distances, I always know that you are there to support me. I treasure the time we all spent together during the pandemic, those times were special and important. The last few months have been some of the most difficult of my entire life, thank you for keeping me going and believing in me even when I didn't believe in myself. I have been a hard friend and I am lucky that you have stuck by my side, and owe the completion of my PhD to you all. I love you all dearly.

To the PD&RR you all have been constant pains in my side and examples to strive towards. The friendship, and scientific advice, have made me into the person and scientist I am. You have opened up opportunities in my life I would have never thought possible. Again, words cannot express how lucky I am to have had you all blaze the trail before me and help me navigate the hardships of a PhD.

To. You were there at the start and there for so much of this. I owe so much of who I am as a person to you. You were my best friend, and partner throughout the years at Dalhousie. When I didn't think I could do it you were my cheerleader, and confidant. You always had the right words to steady me, and would pick me up when I was down, and for that I will always be grateful. I am lucky to have had you make such a large mark on my life, and you will always be part of me. Thank you.

CHAPTER 1

INTRODUCTION

1.1 LONDON DISPERSION IN SOLIDS

Modern density-functional theory (DFT)[1,2] has its origins in 1964–1965 works by Hohenberg, Kohn, and Sham.[3,4] They showed that the total electron density, ρ, completely describes all ground-state properties of an N-electron system, which simplifies the complexity down to a function of three coordinates, independent of the number of electrons.[1] The variational principle must also be applied, where the density that minimises the total energy is the exact ground-state density.[1,3] The breakthroughs in the works by Hohenberg, Kohn, and Sham ultimately allow one to write the total energy as

$$E = T_{\mathrm{KS}} + \int V_{\mathrm{ext}} d\boldsymbol{r} + J(\rho) + E_{\mathrm{xc}}(\rho), \tag{1.1}$$

where T_{KS} is the kinetic energy of the non-interacting electrons, V_{ext} is the external potential from the nuclear charges, $J(\rho)$ is the classical Coulomb self-energy, and the only unknown term is the exchange-correlation functional, E_{xc}.[1] Much of the DFT work since 1965 has focused on the development of newer and more accurate functionals for E_{xc}.[1,2]

Modern DFT allows for the modelling of electronic structure of both molecules and solid materials. DFT is widely used across computational chemistry, physics, and materials science in order to investigate the fundamental nature of interactions in matter, which aid in the interpretation and prediction of experimental results. The power of DFT has helped in the design of new drugs, reactions, catalysts, and materials, amongst many other things.[1,2,5–7] While DFT has been shown to be highly accurate for most examples of intramolecular chemistry, its applications to problems in intermolecular chemistry are still not as developed. A key issue with DFT is that many of the original exchange-correlation functionals are based on the local (local density approximation) or the semi-local density (generalized-gradient approximations) and, by definition, neglect all non-local interactions. This leads to inaccuracies by many density functionals for the modelling of systems
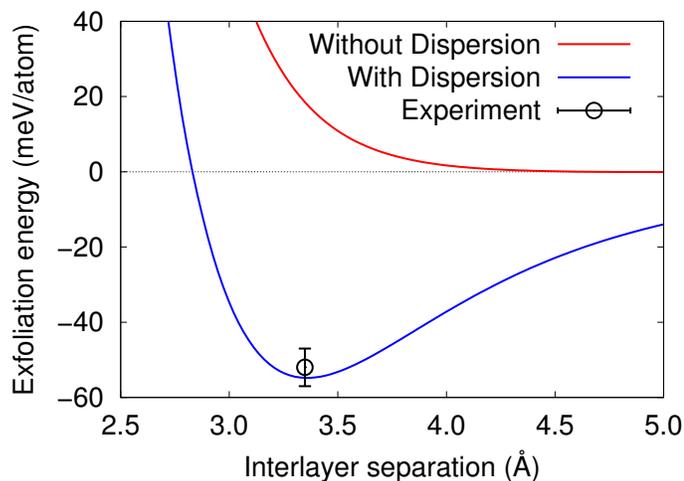
Figure 1.1: Potential energy curve for graphite exfoliation computed using B86bPBE[8,9] (without dispersion) and B86bPBE-XDM[8–10] (with dispersion), as described in Ref. 10. The experimental reference value is taken from Ref. 11.

involving London dispersion, which is a wholly non-local type of electron correlation.

London dispersion[12,13] interactions are long-range interactions that arise between instantaneous dipole (and higher-order multipole) moments in the electron distribution of atoms or molecules. These instantaneous dipoles originate from time-dependent variation of the electron distribution around the nuclei, and these changes in density can create instantaneous electric fields. These electric fields are in turn felt by neighbouring atoms or molecules around this instantaneous dipole and induce further density polarization, resulting in an attractive force. London dispersion is the driving force behind the condensation of non-polar molecules and, despite being the weakest intermolecular interaction, is ubiquitous in chemistry. London dispersion forces are key to determining the 3D structure of biomolecules such as proteins and DNA,[14] molecular self-assembly,[15] interactions within layered materials such as the carbon sheets of graphene to make graphite,[16] surface adsorption,[17] phase transitions,[18] and molecular crystal packing.[12,19,20] The importance of dispersion for layered materials can be seen in the potential energy curves for the exfoliation of graphite, as seen in Fig. 1.1. Without the inclusion of dispersion, graphene layers are not predicted to bind together. However, with the inclusion of a dispersion correction, DFT instead predicts an interlayer separation and exfoliation energy that matches experiment.[11] There are currently a number of distinct methods to address the lack of dispersion forces within the base DFT functional.[10,21]

Consider two non-polar atoms A and B. The dispersion energy between these two atoms can be determined from second-order perturbation theory, where the non-interacting atomic wave functions are used as the unperturbed reference state and the dispersion forces are treated as a small perturbation.[10,22–24] As the atoms are non-polar, they have no permanent dipoles and the first-order

energy correction can be shown to be zero.[22–24] Therefore, one must look at the second-order correction to describe the dispersion forces. Using the closure approximation,[24,25] which avoids the explicit sum over excited atomic states, the second-order dispersion energy (in atomic units) is

$$E^{(2)} \approx -\frac{2}{3} \left(\frac{1}{\Delta E_A + \Delta E_B} \right) \frac{\langle \mu_A^2 \rangle \langle \mu_B^2 \rangle}{R_{ab}^6}, \tag{1.2}$$

where R is the internuclear separation, $\langle \mu^2 \rangle$ is the dipole moment integral, and ΔE is the average excitation energy. In the absence of a permanent dipole moment, $\langle \mu_A^2 \rangle \approx \frac{3}{2} \alpha_A \Delta E_A$ due to the relationship between the mean-square dipole moment and the polarizability. Also, $\Delta E_{A,B}$ can be replaced by taking the average excitation energies as roughly equal to the atomic ionisation potentials, $I_{A,B}$. This substitution gives rise to the London formula,[24]

$$E^{(2)} \approx -\frac{3}{2} \left(\frac{I_A I_B}{I_A + I_B} \right) \frac{\alpha_A \alpha_B}{R^6}, \tag{1.3}$$

which simplifies the dispersion energy down to an expansion in terms of the atomic ionisation energies, $I_{A,B}$, and polarizabilities, $\alpha_{A,B}$. Although approximate, it begins to allow an understanding of the true physics of dispersion interactions. The London formula, for example, correctly predicts that atoms with large polarizabilites have the greatest interatomic dispersion interactions.

An early way of modelling dispersion interactions was through the use of a Lennard-Jones potential,[26] which consists of two fundamental parts: a steep repulsive R^{-12} term and a smooth attractive R^{-6} term. It is represented as potential:

$$V(R) = 4\epsilon \left[\left(\frac{\sigma}{R} \right)^{12} - \left(\frac{\sigma}{R} \right)^6 \right], \tag{1.4}$$

where ϵ is the potential well depth or strength of the interaction between two particles, and σ is the sum of van der Waals radii of the particles. Although simplistic in its formulation, the Lennard-Jones potential is still in use today in force field calculations, with empirical atom typing for the dispersion coefficients.[27–29]

Building on the insight gained from second-order perturbation theory, a number of methods have been formulated to address the missing component of dispersion within DFT. These can be broken down into two key categories: dispersion corrections and dispersion-including functionals. These two types further can be subdivided into asymptotic dispersion corrections,[10,30–43] dispersion-correcting potentials,[44–50] non-local functionals,[42,43,51–56] parameterised exchange-correlation functionals,[57–61] and others.[36,37,62]

The focus of this work is on the subclass of asymptotic dispersion corrections. This subclass is

attractive as asymptotic dispersion corrections are simple in their formulation and capable of high accuracy, whilst also being low in computational cost when compared to other types of dispersion methods. In a general asymptotic dispersion correction, the leading-order dispersion energy term is represented as

$$E_{\mathrm{disp}}^{(6)} = -\sum_{i<j} \frac{C_{6,ij} f_{6,ij}(R_{ij})}{R_{ij}^6}, \tag{1.5}$$

where the sum runs over all atomic pairs and the $C_{6,ij}$ are the interatomic dispersion coefficients which can be approximated, for example, by the London formula above. In order to ensure physicality, the dispersion energy is damped by some function, $f_{6,ij}(R_{ij})$, at short internuclear distances to avoid divergence.

Commonly employed asymptotic dispersion corrections include Grimme's D2,[30] D3,[31,32] and D4 models,[33,34] the Tkatchenko-Scheffler (TS)[35] method, and the exchange-hole dipole moment correction (XDM)[10,63] of Becke and Johnson. The difference in each correction arises from the method by which the C_6 term is calculated and whether higher-order dispersion terms, such as C_8 or C_{10}, are included. For example, the early D2 implementation of Grimme's series of corrections was limited to the C_6 term and used fixed parameterised values for the dispersion coefficients. The later D3 correction computes the C_6 terms using coordination numbers and also includes C_8. XDM improves on the accuracy of the dispersion energy through the inclusion of both C_8 and C_{10} terms, and the dispersion coefficients are calculated from the electron density, which allows the correction to respond robustly to a wide range of chemical environments.

1.2 THESIS GOALS

Currently, XDM is only implemented in a limited selection of electronic-structure codes, and its implementation within the Fritz Haber Institute *ab initio* materials simulations package (FHI-aims)[64–70] is a key objective of this thesis. FHI-aims is an all-electron electronic-structure code package, designed for both finite molecular and periodic solid-state applications. It uses numerical atomic orbitals (NAOs) as basis functions, which allows computational flexibility and speed not seen with planewave codes for periodic solids. The flexibility and speed of NAOs arises due to the fact that they are localised around atoms, allowing them to be truncated in space beyond a given distance, and the number of basis functions needed to converge the energy is usually smaller than with other competing methods.[71] The computational efficiency of FHI-aims paired with XDM will allow for the study of very large (> 200 atoms/unit cell) solid-state systems that would previously have been impractical with conventional planewave codes. Furthermore, FHI-aims is able to extend the range of applicability of hybrid-functionals to solids, which have been too computationally

expensive or impractical for planewave-based codes (except for small unit cells, of ~ 30 atoms or less), and with greater accuracy than can be attained with minimal Gaussian basis sets in the CRYSTAL code.[72,73]

The purpose of the current work is the implementation and testing of the XDM model within the FHI-aims code. The new implementation will be rigorously benchmarked against conventional systems from the literature, including the KB49,[74,75] S22×5,[76] S66×8,[77] and 3B-69[78] sets of intermolecular interactions. We will also perform benchmarking on the more challenging test of lattice energies, considering the X23 set of molecular crystals[10,79,80] and the ICE13 set of 13 ice polymorphs,[81] with the intent to compare the XDM module's performance head-to-head with other commonly employed dispersion corrections. Once the performance of the XDM module has been demonstrated, it will be tested for application to molecular crystal structure prediction (CSP) by considering the energy ranking of candidate crystal structures.

This thesis will focus on the FHI-aims implementation of XDM and its application to solid-state systems involving molecular crystals. Chapter 2 will start with a detailed description of DFT and the forms of the exchange-correlation functionals to be used in this work. The considerations of how periodic boundary conditions affect solid-state DFT calculations will also be discussed, along with an overview of both pseudopotential methods using planewave basis sets and all-electron methods using atom-centered basis sets. The last section within Chapter 2 will be a review of dispersion methods, with a detailed description of the XDM method, as well as outlines of Grimme's methods and the Tkatchenko-Scheffler and many-body dispersion corrections. Chapter 3 will discuss the requirements for the selection of an appropriate density functional and dispersion correction for the accurate description of non-bonded repulsion and dispersion attraction and will inform the proceeding chapters within this work. Chapter 4 will describe a completed computational research project that investigates surface adsorption through XDM-corrected DFT to enable the theoretical assignment of structures characterised experimentally by scanning tunneling microscopy (STM) and will directly apply the theory discussed in Chapters 2 and 3. Chapter 5 will then outline the implementation of the XDM method within the FHI-aims electronic structure package. Chapter 6 will compare the performance of the FHI-aims XDM implementation with that of other dispersion corrections for a range of both molecular and solid-state benchmarks found within the literature. Chapter 7 will then discuss an application of the FHI-aims XDM tool to crystal structure prediction for all molecular crystals found within the Cambridge Crystallographic Data Centre's (CCDC) first six CSP blind tests.[82–87] Chapter 8 will present the results obtained using the FHI-aims implementation of XDM for the recently concluded 7th blind test. Finally, Chapter 9 will discuss the conclusions formed within this thesis and aim to outline the future work left to complete.

CHAPTER 2

THEORY

2.1 DENSITY-FUNCTIONAL THEORY

Density-functional theory (DFT)[1,2] is fundamentally a quantum-mechanical modeling method underpinned by the time-independent Schrödinger equation,[88]

$$\hat{H}\psi = E\psi. \tag{2.1}$$

Here, E is the energy, $\psi(\mathbf{r}_1, \ldots, \mathbf{r}_N)$ is the N-electron wavefunction, and equation 2.2, $\hat{H}$ is the Hamiltonian,

$$\hat{H} = \hat{T} + \hat{V} + \hat{U}, \tag{2.2}$$

shown here broken into its kinetic energy term ($\hat{T}$), the external potential energy term ($\hat{V}$), and the electron-electron interaction energy ($\hat{U}$).

DFT is used in many branches of chemistry and physics to answer fundamental questions concerning the electronic structure of many-body systems in the forms of atoms, molecules, or condensed phases.[1,2,5,89,90] DFT does not directly solve the time-independent Schrödinger equation, but instead systematically maps the many-body problem onto a single-particle problem with the key variable being the electron density, $\rho(\mathbf{r})$, given by

$$\rho(\mathbf{r}) = \int \cdots \int \psi^*(\mathbf{r}_1, \ldots, \mathbf{r}_N)\psi(\mathbf{r}_1, \ldots, \mathbf{r}_N)d\mathbf{r}_2, \ldots d\mathbf{r}_N. \tag{2.3}$$

DFT was made practical by the work by Kohn and Sham in 1965.[4] Their fundamental assumption is that the exact ground-state density can be obtained from the numerical solution of a set of single-particle Schodinger-like equations for a system of non-interacting electrons.[4] The total

energy functional for a Kohn-Sham system can be written as

$$E[\rho] = T_{\mathrm{KS}} + \int v_{\mathrm{ext}}(\mathbf{r})\rho(\mathbf{r})d\mathbf{r} + J[\rho] + E_{\mathrm{XC}}[\rho]. \tag{2.4}$$

Here, T_{KS} is the kinetic energy (in atomic units) which can be written in terms of the occupied Kohn-Sham orbitals, ϕ_i, as

$$T_{\mathrm{KS}} = \sum_{i=1}^{N} \int \phi_i^*(\mathbf{r}) \left(-\frac{1}{2}\nabla^2\right) \phi_i(\mathbf{r})d\mathbf{r}, \tag{2.5}$$

where ∇^2 is the Laplacian operator. For the subsequent terms, v_{ext} is the external potential that arises from the electron-nuclei interactions and

$$J[\rho] = \frac{1}{2} \int \int \frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|}d\mathbf{r}d\mathbf{r}' \tag{2.6}$$

is the Hartree or classical Coulomb energy between electrons. Lastly, E_{XC} is the exchange-correlation energy, which contains all of the energetic contributions neglected by the other terms.

In practice, Equation 2.4 must be solved in an iterative or self-consistent fashion. This is done through the self-consistent field (SCF) procedure, as in Hartree-Fock (HF) theory, which obtains the orbitals that minimise the energy of the system. This occurs through variationally lowering the electronic energy by optimising the electron density. The density is constructed from the occupied Kohn-Sham orbitals as

$$\rho(\mathbf{r}) = \sum_{i}^{N} |\phi_i(\mathbf{r})|^2. \tag{2.7}$$

The Kohn-Sham equations are then solved for the energy and an updated set of Kohn-Sham orbitals. This iterative process is continued until the energy is converged to within a certain tolerance and the procedure is finished.

As mentioned earlier, it is the E_{XC} term in Equation 2.4 where all previously neglected energetic components are packaged. Since the exact form of E_{XC} is not known, there have been many different density-functional approximations proposed over the years of DFT.[1,2] The exchange-correlation energy can be separated into its component parts, exchange energy (E_{X}) and correlation energy (E_{C}), as $E_{\mathrm{XC}} = E_{\mathrm{X}} + E_{\mathrm{C}}$. For the remainder of this section, only the form of exchange functional will be discussed, as its contribution to E_{XC} energy is far greater than that of E_{C}.[1]

The first and simplest exchange functional is the local spin density approximation (LSDA), which has the form

$$E_{\mathrm{X}}^{\mathrm{LDSA}} = -\frac{3}{2}\left(\frac{3}{3\pi}\right)^{1/3} \sum_{\sigma} \int \rho_{\sigma}^{4/3}(\mathbf{r})d\mathbf{r}, \tag{2.8}$$

where the sum is over electron spins and ρ_σ is the spin-dependent density. The correlation energy for the LSDA is obtained from quantum Monte Carlo calculations on the uniform electron gas, as there is no closed-form expression.[91] The LSDA typically leads to an underestimation of the exchange energies and an overestimation of correlation and bond energies.[1] This led to the first improvement on the LSDA and spawned a class of functionals called generalized gradient approximations (GGA).

Real chemical systems do not contain a uniform density. For GGA-type functionals, the exchange energy depends on both the density and the density gradient, $\nabla \rho$, and has the general form

$$E_X^{GGA} = \sum_\sigma \int \varepsilon_X(\rho_\sigma, \nabla \rho_\sigma) d\mathbf{r}. \tag{2.9}$$

where ε_X is the exchange-energy density. Two common GGAs used in this work are the Becke86b (B86b)[8] and Perdew-Burke-Ernzerhof (PBE)[9] exchange functionals. These GGAs have the form

$$E_X^{GGA} = -\sum_\sigma \int c_X \rho_\sigma^{4/3} F(\chi_\sigma) d\mathbf{r}, \tag{2.10}$$

where $F(\chi_\sigma)$ is termed the enhancement factor, c_X is a constant term, and χ_σ is the dimensionless density gradient,

$$\chi_\sigma = \frac{|\nabla \rho_\sigma|}{\rho_\sigma^{4/3}}. \tag{2.11}$$

The differences between PBE and B86b arise in the enhancement factor. In PBE, it has the form[92]

$$F(\chi_\sigma) = 1 + \frac{\beta}{c_X} \frac{\chi_\sigma^2}{1 + \gamma \chi_\sigma^2}, \tag{2.12}$$

where the values of the two parameters β and γ are determined by imposing known constraints.[9] This functional corrects for the uniform-electron gas in the low-gradient limit, but misses the correct behaviour in the large-gradient limit.[8] The B86b functional was developed to address the large gradient limit. It has the same form for the exchange energy as Equation 2.10, but the enhancement factor is different:

$$F(\chi_\sigma) = 1 + \frac{\beta}{c_X} \frac{\chi_\sigma^2}{(1 + \gamma \chi_\sigma^2)^{4/5}}. \tag{2.13}$$

In the case of B86b, the parameters β and γ were fit to atomic exchange energies.[8] Both PBE and B86b exchange are typically paired with the PBE correlation functional.[9] PBE is generally the most widely used density functional for solid-state applications.[2,9] However, B86bPBE[8,9] is our functional of choice in this work due to its behaviour in the large-gradient limit, which ensures better performance for non-bonded repulsion.[93,94]

GGAs represent major improvements over the LSDA for many systems of chemical interest, but

are still limited in their performance as they are semi-local. More complicated classes of functionals include meta-GGAs[95] and hybrid functionals,[96] which can achieve greater chemical accuracy.[1] For meta-GGAs, this is achieved through the additional inclusion of the kinetic energy density, τ_σ, and the Laplacian of the density, $\nabla^2 \rho_\sigma$,[95] with more information given in Section 2.5. Moving beyond GGAs and meta-GGAs, exact (HF) exchange may be included to give a hybrid functional,

$$E_{X\sigma}^{\text{hybrid}} = a_X E_{X\sigma}^{\text{HF}} + (1 - a_X) E_{X\sigma}^{\text{GGA}}, \tag{2.14}$$

with the mixing parameter, a_X, typically varying between 20 - 50%, depending on the functional used. E_X^{HF} has the form

$$E_X^{\text{HF}} = -\frac{1}{2} \sum_\sigma \iint \phi_i^*(\mathbf{r})\phi_j^*(\mathbf{r}')\frac{1}{r_{12}}\phi_i(\mathbf{r})\phi_j(\mathbf{r}')d\mathbf{r}d\mathbf{r}' \tag{2.15}$$

and uses the occupied Kohn-Sham orbitals. One of the most common hybrid functionals is PBE0,[97] which has $a_X = 25\%$. Hybrid functionals address some of the shortcomings found within both pure HF and GGAs. They outperform both GGAs and HF by mitigating issues common in GGAs, such as delocalization error, and yet still include electron correlation absent in pure HF. However, as exact exchange is non-local in nature, hybrid functionals can become increasingly costly for large molecular systems and prohibitively expensive with increasing unit-cell size for solid-state systems (becoming impractical for roughly > 30 atoms/unit cell, within planewave-based codes).

2.2 PERIODIC BOUNDARY CONDITIONS IN THE SOLID STATE

The treatment of solids has a difficulty not seen within gas-phase calculations on finite molecules, which is the requirement to treat an infinitely extending system with an infinitely large number of electrons. This is only surmountable due to the ability to break down solid-state systems into periodic components, or unit cells. Unit cells are the most basic repeating structure of a given material from which the whole can be constructed. A unit cell provides periodicity and can be completely specified by three Bravais lattice vectors,

$$\mathbf{R} = n_1 \mathbf{a_1} + n_2 \mathbf{a_2} + n_3 \mathbf{a_3}, \tag{2.16}$$

where $\mathbf{a_1}$, $\mathbf{a_2}$, and $\mathbf{a_3}$ represent the primitive vectors which span the unit cell and n_1, n_2, and n_3 range through all integer values. As the entire repeating system can be represented by translations

of the unit cell along the lattice vectors, the problem can be reduced from considering the whole infinitely repeating solid and its orbitals down to considering just the unit cell.

Given that the potential $U(\mathbf{r})$ can be represented using the periodicity of the Bravais lattice,

$$U(\mathbf{r}) = U(\mathbf{r} + \mathbf{R}), \tag{2.17}$$

Bloch's theorem[98] allows the eigenstates ψ of the one-electron Hamiltonian,

$$\hat{H} = -\frac{1}{2}\nabla^2 + U(\mathbf{r}), \tag{2.18}$$

to be chosen such that they have the periodicity of the Bravais lattice through the use of plane waves,

$$\psi_{n\mathbf{k}}(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}}u_{n\mathbf{k}}(\mathbf{r}), \tag{2.19}$$

with

$$u_{n\mathbf{k}}(\mathbf{r}) = u_{n\mathbf{k}}(\mathbf{r} + \mathbf{R}), \tag{2.20}$$

which can be written in the terms of its Fourier expansion as

$$u_{\mathbf{k}}(\mathbf{r}) = \sum_{\mathbf{G}} u_{\mathbf{k},\mathbf{G}} e^{i\mathbf{G}\cdot\mathbf{r}}. \tag{2.21}$$

Here, $\mathbf{k}$ is a reciprocal wave vector of the form,

$$\mathbf{k} = m_1\mathbf{b}_1 + m_2\mathbf{b}_2 + m_3\mathbf{b}_3 , \tag{2.22}$$

where the reciprocal lattice vectors are,

$$\mathbf{b}_1 = \frac{2\pi}{\Omega}\mathbf{a}_2 \times \mathbf{a}_3 , \qquad \mathbf{b}_2 = \frac{2\pi}{\Omega}\mathbf{a}_3 \times \mathbf{a}_1 , \qquad \mathbf{b}_3 = \frac{2\pi}{\Omega}\mathbf{a}_1 \times \mathbf{a}_2 . \tag{2.23}$$

Where the volume of the unit cell is represented by $\Omega = \mathbf{a}_1 \cdot (\mathbf{a}_2 \times \mathbf{a}_3)$, and the real-space primitive lattice vectors satisfy $\mathbf{a}_i \cdot \mathbf{b}_j = 2\pi\delta_{ij}$, where δ_{ij} is the Kronecker delta. $\mathbf{G}$ can be any of the periodic wave vectors such that $\mathbf{G} \cdot \mathbf{R} = 2\pi m$ for some integer m, where $\mathbf{R}$ is a lattice vector of the crystal. The electronic wavefunction can then be written as

$$\psi_i(\mathbf{r}) = \frac{1}{\sqrt{\Omega}} \sum_{\mathbf{G}} c_{i,\mathbf{k}+\mathbf{G}} e^{i(\mathbf{k}+\mathbf{G})\cdot\mathbf{r}}. \tag{2.24}$$

This use of Bloch's theorem maps the problem of writing wavefunctions for an infinite number

of electrons to one of expressing the wavefunctions in terms of a finite number of electrons, but using an infinite number of reciprocal-space vectors in the first Brillouin zone. This is then solved through the sampling of $\mathbf{k}$-space within the Brilloun zone.

In principle, $\mathbf{k}$-space must be sampled completely for an exact solution. In practice, this can be avoided to give an accurate and tractable solution through an efficient sampling algorithm. Similar to the unit cell, as noted before, there exists a way to uniquely define a primitive cell in reciprocal or $\mathbf{k}$-space. The first Brillouin zone can be mapped by a set of $\mathbf{k}$-points, denoted as $x \times y \times z$, where every possible $\mathbf{k}$ in the infinite solid is equivalent to exactly one point in this region. There is an infinite density of $\mathbf{k}$-points within the first Brillouin zone and, therefore, an efficient yet accurate sampling must be used. One such scheme is that developed by Monkhorst and Pack,[99] which distributes the sampled $\mathbf{k}$-points homogeneously in columns and rows parallel to the reciprocal lattice vectors and is based on point-group symmetries. Schemes like the Monkhorst-Pack scheme are viable even though they do not span all possible $\mathbf{k}$-points because they offer an efficient route to a sufficiently dense integration grid in reciprocal space, and a well-converged set of finite $\mathbf{k}$-points can still determine the total energy within an acceptable margin of error.

2.3 PLANEWAVES AND PSEUDOPOTENTIALS

A finite set of planewaves can be used as a basis because there are only discrete $\mathbf{G}$ vectors that fit within the lattice periodicity. The kinetic-energy cutoff that limits the energy/frequency of the planewaves is represented by

$$E^{\mathrm{cut}} = \frac{1}{2}|\mathbf{k} + \mathbf{G}_{\mathrm{max}}|^2, \tag{2.25}$$

which may be used in practice to limit the basis as the contributions of the higher Fourier components are ever diminishing. Separate kinetic-energy cutoffs are employed for the charge density and potential and for wavefunctions, respectively. As many planewaves are used to build up the true wavefunction, the calculation must be converged with respect to these cutoffs to ensure an accurate total energy.

In practice, even with the use of planewaves and efficient integration grids, there are caveats for calculations on solid-state systems. One such drawback is the inability of planewaves to easily describe wavefunctions containing sharp features, such as those at electron-nuclear cusps. To address this drawback and reduce the number of planewaves required to describe the wavefunctions and electron density, all atomic core regions are replaced with fixed pseudopotentials, so that the planewaves are only used to describe the valence charge density. These pseudopotentials have a simplified wavefunction with no nodes or cusps and, in the case of norm-conserving

pseudopotentials, a correct total charge. Outside some defined radial cutoff (r_c), the all-electron and pseudo wavefunctions are the same. Additionally, pseudopotentials can be classed as hard or soft, where a pseudopotential is considered soft when it requires a small number of planewaves for an accurate representation, and hard otherwise. The pseudopotentials also should be constructed from the specific type of exchange-correlation functional to be used in calculations.

Norm-conserving pseudopotentials are a specific type of pseudopotential that allow for a basis set with a significantly lower cutoff for the highest Fourier mode of the planewaves. Norm-conserving pseudopotentials, first proposed in 1979 by Hamann, Schlüter, and Chiang[100] and later modified by Troullier and Martins,[101–103] follow four general conditions. The first is that the valence pseudo-wavefunction generated from the pseudopotential contains no nodes. Second, the normalised atomic radial pseudo-wavefunction with angular momentum l and the normalised radial all-electron wavefunction are equal for a chosen cutoff radius, r_{cl}. Third, the charge within that cutoff region r_{cl} is equal for the pseudopotential (PP) and all-electron (AE) radial wavefunctions, R_l:

$$\int_0^{r_{cl}} \left| R_l^{\mathrm{PP}}(r) \right|^2 r^2 dr = \int_0^{r_{cl}} \left| R_l^{\mathrm{AE}}(r) \right|^2 r^2 dr. \tag{2.26}$$

Lastly, the valence pseudopotential and all-electron eigenvalues, ϵ_l, are also equal:

$$\varepsilon_l^{\mathrm{PP}} = \varepsilon_l^{\mathrm{AE}}. \tag{2.27}$$

Although, norm-conserving pseudopotentials have wide applications in solid-state calculations, they are not as accurate as the projector augmented-wave (PAW) method developed by Blöchl in 1994[104] and generalized by Kresse and Joubert.[105]

The PAW method is a type of augmented-wave method. Augmented-wave methods have various forms which began with the development of the augmented planewave method (APW). The APW method aims to solve the Kohn-Sham equations, shown in matrix form as

$$\sum_{n'} (H_{n,n'} - \varepsilon_n S_{n,n'}) c_{n'} = 0, \tag{2.28}$$

where ε_n is the eigenvalue of the energy, and $c_{n'}$ are expansion coefficients. $H_{n,n'}$ is the Hamiltonian matrix,

$$H_{n,n'} = \left\langle \phi_n \middle| -\nabla^2 + V_{\mathrm{eff}} \middle| \phi_{n'} \right\rangle, \tag{2.29}$$

V_{eff} the effective potential, and $S_{n,n'}$ is the overlap matrix,

$$S_{n,n'} = \left\langle \phi_n \middle| \phi_{n'} \right\rangle. \tag{2.30}$$

APW methods solve the Kohn-Sham equations 2.28 through linearising the eigenvalue problem and forms of this method are called linearised augmented planewave (LAPW) methods.[106,107] The LAPW method uses a dual basis type, where regions within the atomic spheres use atomic-like basis functions and the interstitial regions between atoms are treated with a planewave basis. The atomic regions are represented as muffin-tin spheres, where the number of planewaves required to represent the system is dependent on the muffin-tin cutoff radius, with a large region requiring fewer planewaves and a smaller one requiring more planewaves. The LAPW method has the benefit of being a true all-electron method and, therefore, can give an accurate description of chemistry involving the atomic cores. Unfortunately, this accuracy comes at the cost of being computationally costly compared to other techniques such as using pseudopotentials.

PAW is another common method of core-region reconstruction within solid-state calculations, which aims to have the accuracy seen with LAPW methods, and the efficiency seen with pseudopotential methods. This is achieved through transforming the true all-electron wavefunction, $\Psi(r)$, into a pseudised wavefunction, $\tilde{\Psi}(r)$, through the use of a transform operator:

$$\Psi(r) = \hat{T}\tilde{\Psi}(r). \tag{2.31}$$

This transform operator, $\hat{T}$, is written as $\hat{T} = 1 + \sum_R S_R$ and has some constraints so that the pseudo-wavefunctions remain smooth, and can be expanded in terms of a minimal number of planewaves. S_R adds the difference between the all-electron and pseudo-wavefunctions, which are expanded in partial waves for free isolated atoms:

$$S_R|\tilde{\phi}_i\rangle = |\phi_i\rangle - |\tilde{\phi}_i\rangle, \tag{2.32}$$

where $|\tilde{\phi}_i\rangle$ is the pseudo partial wave and $|\phi_i\rangle$ is the all-electron valence partial wave. The augmentation region $\tilde{\Psi}_n(\mathbf{r})$ is then expanded in partial waves of $|\tilde{\phi}_i\rangle$ and is written as

$$\tilde{\Psi}_n(\mathbf{r}) = \sum_{i \varepsilon R} \tilde{\phi}_i(r)c_{in}, \tag{2.33}$$

where the c_i are determined by the projector functions $\langle \tilde{p}_i |$, which provide the contribution of each wave and is the projector part of PAW. This leaves the transform operator to be

$$\hat{T} = 1 + \sum_i \left(|\phi_i\rangle - |\tilde{\phi}_i\rangle \right) \langle \tilde{p}_i |, \tag{2.34}$$

and therefore the true wavefunction is recovered from $\Psi(\mathbf{r}) = \hat{T}\tilde{\Psi}(\mathbf{r})$. The PAW method is

Figure 2.1: Schematic representation of PAW method. Figure adapted from Ref. [108].

$$\rho_v \qquad = \qquad \tilde{\rho} \qquad + \qquad \rho^1 \qquad - \qquad \tilde{\rho}^1$$
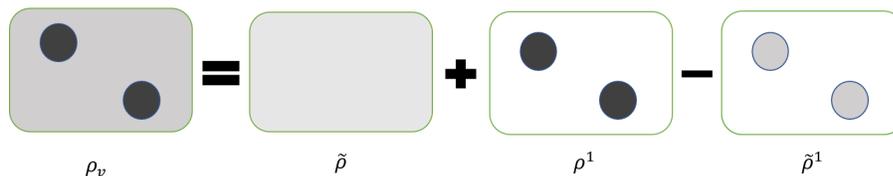
represented schematically in Fig. 2.1, where ρ_v is the total valence charge density, $\tilde{\rho}$ is the density from the pseudopotential calculation over the entire cell, ρ^1 represents the all-electron charge density within the PAW spheres, and $\tilde{\rho}^1$ the pseudo charge densities within the PAW spheres.[108] The PAW method achieves efficient and accurate densities by calculating the pseudopotential density and correcting the core region of each atom through the subtraction of the in-sphere pseudo part of the total charge density, replacing it with the all-electron in-sphere component.[108] This method has the benefit of retaining the core wavefunctions, with a reduction in the computational expense that arises from the planewave basis-set treatment of the core region.

All PAW datasets in this work were generated with B86bPBE using the "atomic" code from Quantum ESPRESSO,[109] version 5.1.1. The cutoff radii used for the pseudisation of the all-electron wavefunctions were set to those specified in Table 2.1. These components are important as PAW sphere overlap can cause spurious coulombic interactions between spheres[108] due to the unwanted augmentation of the charge densities. For the PAW calculations performed in Chapter 4 of this work, the possible minimum overlap distances are listed in Table 2.2. $C-H$ bonds have the shortest bond lengths considered and also have one of the smallest differences between the optimised bond lengths and the sum of the core radii. The minimum distance between the PAW spheres is 0.23 bohr for $C=O$ double bonds and this is the closest the spheres come to overlapping within the systems studied in this work. This is still beyond the range where the compensation densities would be expected to be overlapping and, therefore, PAW overlap would not be expected to be a major source of error within this work.

2.4 ALL-ELECTRON METHODS

2.4.1 ATOM-CENTERED BASIS SETS

There exist two main approaches to treating the core region wavefunctions, one being pseudopotentials and the other being all-electron-based methods. Although a pseudopotential-based method has the benefits of obtaining arbitrarily high accuracy with systematic increases in the kinetic-energy cutoffs for planewave basis sets and straightforward inclusion of scalar relativistic effects, they

Table 2.1: Valence configuration and core-region cutoff radii, r_c, in atomic units for the B86bPBE PAW datasets used in this work.

Element	Valence Configuration		
	shell	occupation	r_{c}
H	1S	1.00	0.80
C	2S	2.00	1.00
	2P	2.00	0.90
N	2S	2.00	1.00
	2P	3.00	0.90
O	2S	2.00	1.00
	2P	4.00	0.90
Fe	4S	1.70	1.30
	2P	0.00	1.30
	3D	6.30	1.30
Ag	5S	1.50	2.20
	5P	0.00	2.20
	4D	9.50	1.70

Table 2.2: Minimum bond distances from optimised geometries in Chapter 4 of this work, compared to the sum of maximum core cutoff radii for each element. All quantities are in atomic units.

Bond	Optimised Distance	$\sum r_c$	Difference
C-H	2.06	1.80	0.26
C-C	2.65	2.00	0.65
C-N	2.64	2.00	0.64
C-O	2.23	2.00	0.23
N-Fe	3.63	2.30	1.33
Fe-Ag	5.50	3.30	2.20
Fe-C	3.31	2.30	1.01
Ag-Ag	5.39	4.40	0.99

still have the drawbacks of transferability and the introduction of an approximation error. The alternative to the use of pseudopotentials is to treat all of the electrons in the core and valence on an equal footing by employing atom-centred basis sets. This allows for increased accuracy in calculations on excited states, electric and magnetic response properties, and high-pressure applications where a pseudopotential may either overlap with the valence wavefunction or not allow for a density change to occur.[64,65,110] Many types of atom-centred basis sets can be used both with pseudopotentials and in all-electron calculations. The specific focus in this section will be on the all-electron implementation of numerical atomic orbitals (NAOs), with a survey of other common atom-centered basis sets.

The electronic wavefunction is expanded as a linear combination of the functions that comprise the basis set. The use of basis sets converts the three-dimensional partial differential equations from the base mathematical formalism of KS-DFT into algebraic equations, whose solution can be implemented on a computer. These basis functions can be either constructed from the atomic orbitals or planewaves, each of which has its own benefits and drawbacks. The KS orbitals can be written as linear combinations of basis functions, $\varphi_p(\mathbf{r})$, with optimised coefficients c_{ip}:

$$\phi_i(\mathbf{r}) = \sum_p c_{ip}\varphi_p(\mathbf{r}). \tag{2.35}$$

As mentioned previously, one major source of error in DFT calculations is the selection of the exchange-correlation energy functional. One of the next major sources of error can be found in the basis-set selection and its completeness. Basis set superposition error (BSSE) is a manifestation of the issues of basis set incompleteness. In practical quantum-chemical calculations, a finite basis must be used, as a complete basis would not be computationally tractable. The use of a finite basis means that, as subunits of the calculation (for example atoms, in atom-centred basis sets) are brought together, their basis functions begin to overlap. This allows for the subunits to borrow functions from the other nearby interacting components, which effectively increases its basis set. Thus, the energies of molecular complexes are artificially lowered when compared to the separated molecules. While not used in this work, methods to address or minimise BSSE beyond the use of an ever-increasing basis set exist, such as the chemical Hamiltonian approach,[111,112] the traditional Boys-Bernardi counterpoise (CP) correction (which approximates the energy contribution arising due to this overlap and subtracts it from the energy),[113] and Grimme's geometric CP correction,[114–116] all of which aim to address BSSE in a systematic way but can themselves introduce error.

There are many different types of basis sets in common employ within computational chemistry codes. What follows is a list of the most common basis sets, along with their form and uses. As mentioned earlier, planewave basis sets are common within the solid state community. These basis

sets have the form

$$\varphi_k(\mathbf{r}) = \frac{1}{\sqrt{V}} e^{i\mathbf{k}\cdot\mathbf{r}}, \tag{2.36}$$

where the prefactor ensures normalization for integration over a volume, V. Planewave basis sets are desirable for periodic systems, although they do have the caveat of requiring pseudopotentials, as mentioned in the section on planewave DFT above.

Slater-type orbitals (STOs) are an obvious first choice for atom-centered basis sets for electronic structure theory due to their exponential decay at long range and their nuclear cusp. These have the form

$$\varphi_{nlm}(\mathbf{r}) = N r^{n-1} e^{-\alpha r} Y_{lm}(\theta, \phi), \tag{2.37}$$

where n, l, and m are the quantum numbers, N is a normalisation factor, α an adjustable exponent, and $Y_{lm}(\theta, \phi)$ are the spherical harmonics. STOs correctly capture the nature of the wavefunction within the nuclear cusp region,[117] although they are difficult to use in a computationally expedient manner due to the derivative discontinuity at the nuclei and integration of the radial part.[118,119]

In order to address these mathematical shortcomings, basis sets composed of Gaussian-type orbitals (GTOs) were developed. While missing the nuclear cusp, these functions allow for easier treatment of the nuclear core regions and can be constructed of many primitive Gaussian functions. By Gaussian product theorem, we can create a function that is analytically solvable/integrable but closely matches the form of the exact hydrogen-like atomic orbitals.[118–121] Gaussian-type orbitals have the form

$$\varphi_{nlm}(\mathbf{r}) = N r^{n-1} e^{-\alpha r^2} Y_{lm}(\theta, \phi), \tag{2.38}$$

which is comparable to the STO form, except for the power of r in the exponential. GTOs and STOs are both atom-centered and commonly used in all-electron calculations, although they can also be used in conjunction with pseudopotentials, as is common for heavy elements. GTOs have issues with their radial tails, which decay very slowly, and require many primitive gaussians in order to correctly capture the true electronic wavefunction. This increases the computational expense and can lead to poor SCF convergence, which can make the choice of a Gaussian basis set unappealing for periodic systems, although codes such as CRYSTAL[73] do employ GTOs for solid-state calculations.

NAOs are another type of atom-centred, all-electron basis set, which allow for the treatment of both molecular and periodic systems, but without the need for pseudopotentials to be computationally tractable.[64] NAOs have the general form

$$\varphi_{nlm}(\mathbf{r}) = \frac{u(r)}{r} Y_{lm}(\theta, \phi), \tag{2.39}$$

where the angular component $Y_{lm}(\theta, \phi)$ is the same as in GTOs and STOs.[64,110] The flexibility of NAOs comes in the form of the radial component, $u(r)$, which is numerically constructed and fully flexible. The Fritz Haber *ab initio* materials simulations program (FHI-aims) is one program that utilises basis sets of this form. Although FHI-aims will be the main discussion of this thesis, other codes do exist that utilise NAOs, such as the SIESTA code,[122] DMOL^3,[123] PLATO,[124] and more. FHI-aims is mainly a DFT-based code that is able to compute ground-state total energies for both molecular and solid-state matter, but also has a wide range of beyond-DFT methods built up within the code (GW, MP2, RPA, wavefunction-based methods, etc.).[64] The focus of this work will be on the pure and hybrid DFT components of the code.

Within FHI-aims, $u(r)$ is constructed in such a way that it satisfies the single-particle, Schrödinger-like equation

$$\left[-\frac{1}{2} \frac{d^2}{dr^2} + \frac{l(l+1)}{r^2} + v_i(r) + v_{\text{cut}}(r) \right] u_i(r) = \varepsilon_i u_i(r). \tag{2.40}$$

The $v_i(r)$ term can correspond to an atomic, ionic, or hydrogen-like potential of the form $v_i(r) = \frac{Z}{R}$,[125–127] as well as other forms.[128] The construction of the FHI-aims basis sets was done through an optimisation algorithm that minimised the total energy error for a selection of spin-unpolarised, symmetric dimers at 4-5 differing bond distances at the DFT-LDA level of theory,

$$\Delta_{\text{basis}} = \frac{1}{N_d} \sum_{i=1}^{N_d} \left[\varepsilon_{\text{basis}}(d_i) - \varepsilon_{\text{cb}}(d_i) \right]. \tag{2.41}$$

Here, $\varepsilon_{\text{basis}(d_i)}$ represents the non-self-consistent total energy for each atom at the specified bond length, d_i, $\varepsilon_{\text{cb}}(d_i)$ represents the converged basis set limit for the non-self-consistent total energy, and N_d non-spinpolarized symmetric dimers at approximately 4-5 different bond lengths.[64] The various tiers of basis sets, as shown for the examples of C and H in Table 2.3, are converged to varying levels. With tiers 1 and 2 for the light elements, which represent the "light" and "tight" levels referred to in this work, the energies are converged to $\sim$0.1 eV and $\sim$0.01 eV, respectively.

One key point concerning the construction of NAOs within FHI-aims is that these functions are strictly localised by the confining potential, $v_{\text{cut}}(r)$, which cuts off the slowly decaying tails:[64]

$$v_{\text{cut}}(r) = \begin{cases} 0 & r \leq r_{\text{onset}}, \\ s \cdot \exp\left(\frac{w}{r - r_{\text{onset}}}\right) \cdot \frac{1}{(r - r_{\text{cut}})^2} & r_{\text{onset}} < r < r_{\text{cut}}, \\ \infty & r \geq r_{\text{cut}}. \end{cases} \tag{2.42}$$

At r_{onset}, the confining potential begins and smoothly proceeding towards infinity at r_{cut}. Additionally, $w = (r_{\text{cut}} - r_{\text{onset}})$ and s is used as a global scaling parameter. The values of r_{onset}, r_{cut},

Table 2.3: Representation of radial functions, $u_i(r)$, that construct the various tiers of basis functions within FHI-aims. The minimal basis set is constructed from free-atom-like defining potentials, $v_i(r)$ from Eq. 2.42, while the $H(n\ell, Z)$ notation represents hydrogen-like basis functions with quantum numbers n, ℓ using the bare Coulomb potential, Z/R. Table recreated from Ref. 64.

	H	C
minimal	1s	[He] + 2s2p
Tier 1 (light)	H(2s, 2.1)	H(2p, 1.7)
	H(2p, 3.5)	H(3d, 6.0)
		H(2s, 4.9)
Tier 2 (tight)	H(1s, 0.85)	H(4f, 9.8)
	H(2p, 3.7)	H(3p, 5.2)
	H(2s, 1.2)	H(3s, 4.3)
	H(3d, 7.0)	H(5g, 14.4)
		H(3d, 6.2)

and s are preselected within the construction of the basis sets, and these defaults have been shown to provide excellent performance in other works.[64,65] The use of the confining potential drives the radial component to zero, and provides functions that are strictly confined to a certain domain in space.[64,65] This allows for the expensive steps in the SCF procedure, such as integrations or density updates, to scale linearly with system size, which is a primary source of the NAO method's efficiency.[64,65,110]

Another key aspect of the construction of NAOs in this manner is that it allows for optimised element-dependent basis sets that have both systematically achievable levels of accuracy, through the inclusion of ever more basis functions, and high transferability for a wide range of possible chemical species.[64] Previous works have illustrated that, beyond the minimal basis set, BSSE is not a major source of error within DFT calculations using the precompiled FHI-aims basis sets.[64] As an example, in the case of a water dimer for the tier 2 basis set, the energy (with no counterpoise correction) is already within 5 meV of the accepted full basis set converged values.[64]

2.4.2 REAL-SPACE INTEGRATION GRIDS

As discussed earlier, integrals that involve the exchange-correlation functionals or potentials cannot trivially be solved and require numerical methods to compute. Evaluation of the exchange-correlation energy requires numerical integration over a grid of points, with many pseudopotential codes achieving this through the use of even-spaced grids within the Brillouin zone. For chemical

systems, the density changes rapidly near the atomic nuclei. Thus, an equally distributed grid of points in Cartesian space is highly inefficient for all-electron calculations, and would either require a massively large number of total points or miss fully characterising the density changes. An alternative method must then be sought and arises in the form of the grids proposed by Becke in 1988.[129]

In numerical integration, we must evaluate 3D molecular integrals of the form

$$I = \int F(\mathbf{r})d\mathbf{r} \approx \sum_i A_i F(\mathbf{r}_i), \tag{2.43}$$

where $F(\mathbf{r})$ is an arbitrary integrand. These integrals can be approximated by the summation of the integrand over discrete integration points, $\mathbf{r}_i$, multiplied by their integration weights, A_i. Rather then spreading the points equally across the system of interest, it can be partitioned into atomic regions, where the numerical integration is carried out via spherical coordinates. This splitting of the function $F(\mathbf{r})$ is done through employing relative weight functions, $w_n(\mathbf{r}_i)$, which are assigned to each nucleus n in the system, sum to one within the vicinity of each atom, and decay to zero at all other nuclei:

$$\sum_n w_n(\mathbf{r}_i) = 1. \tag{2.44}$$

This then allows us to write

$$F(\mathbf{r}) = \sum_n F_n(\mathbf{r}) = \sum_n w_n(\mathbf{r})F(\mathbf{r}), \tag{2.45}$$

where the arbitrary integrand is partitioned into atomic regions, allowing the initial integral I to be written as

$$I = \sum_n I_n = \sum_n \left(\int w_n(\mathbf{r})F(\mathbf{r})d\mathbf{r} \right). \tag{2.46}$$

The exact selection of the integration weights can vary, but most DFT codes use the "Becke weights", which arise from smoothed Voronoi polyhedra.[129] FHI-aims uses the modified weight functions described in Appendix C of Ref. 70 to limit the extent of each integrand to the atom-centered grid that is associated with each basis function.

The selection of the integration weights, assuming that they are numerically well-behaved, lets us write each atomic subintegration, I_n, as a single-centre integral in spherical coordinates as

$$I_n = \iiint_V w_n(\mathbf{r})F(\mathbf{r})r^2 \sin\theta dr d\theta d\phi. \tag{2.47}$$

Lebedev quadratures are utilised to carry out the angular integration and the radial integrations are

20

mapped from $0 < r < \infty$ to the interval $-1 < x < 1$ used in standard Gaussian quadratures, via the transform

$$r = r_m \frac{1+x}{1-x},$$ (2.48)

where r_m is taken as half the empirical Bragg-Slater radii [130] and corresponds to the midpoint of the integration interval. [129] Overall, this definition allows for more integration points to be concentrated at the nucleus, with ever-decreasing amounts of points contained within the tail region, providing an accurate and systematic prescription for increasing numerical precision with atom-centred basis sets. [129]

2.5 THE EXCHANGE-HOLE DIPOLE MOMENT DISPERSION MODEL

The exchange-hole dipole moment (XDM) dispersion model, developed by Becke and Johnson, [10,63,131] overcomes the lack of dispersion within conventional DFT exchange-correlation functionals. This is achieved by a post-SCF correction to the base DFT energy of the form

$$E = E_{\text{base}} + E_{\text{disp}},$$ (2.49)

where E_{base} contains only the semi-local and local effects. The dispersion energy within a periodic lattice is given by

$$E_{\text{disp}} = -\frac{1}{2} \sum_n \sum_{\mathbf{L}} \sum_{ij}' \frac{C_{n,ij} f_n(R_{ij,\mathbf{L}})}{R_{ij,\mathbf{L}}},$$ (2.50)

with $\mathbf{L}$ indicating the lattice vector, $C_{n,ij}$ the n-th order interatomic dispersion coefficients, and $f_n(R_{ij,\mathbf{L}})$ a damping function for the deactivation of the dispersion interaction at short ranges. [10] Here, the primed notation on the summation indicates $i \neq j$ for $\mathbf{L} = 0$. In practice, XDM truncates equation 2.50 at the $n = 10$ term. [10,63]

For the XDM model, the dispersion coefficients are determined from second-order perturbation theory. [10,21,22,63] The first three pairwise $C_{n,ij}$ dispersion coefficients between atoms i and j are

$$C_{6,ij} = \alpha_i \alpha_j \frac{\left\langle M_1^2 \right\rangle_i \left\langle M_1^2 \right\rangle_j}{\alpha_j \left\langle M_1^2 \right\rangle_i + \alpha_i \left\langle M_1^2 \right\rangle_j},$$ (2.51)

$$C_{8,ij} = \frac{3}{2} \alpha_i \alpha_j \frac{\left\langle M_1^2 \right\rangle_i \left\langle M_2^2 \right\rangle_j + \left\langle M_2^2 \right\rangle_i \left\langle M_1^2 \right\rangle_j}{\alpha_j \left\langle M_1^2 \right\rangle_i + \alpha_i \left\langle M_1^2 \right\rangle_j},$$ (2.52)

21

$$C_{10,ij} = \alpha_i \alpha_j \frac{2 \langle M_1^2 \rangle_i \langle M_3^2 \rangle_j + 2 \langle M_3^2 \rangle_i \langle M_1^2 \rangle_j + \frac{21}{5} \langle M_2^2 \rangle_i \langle M_2^2 \rangle_j}{\alpha_j \langle M_1^2 \rangle_i + \alpha_i \langle M_1^2 \rangle_j}, \qquad (2.53)$$

where $C_{6,ij}$ captures dipole-dipole, $C_{8,ij}$ captures dipole-quadrupole, and $C_{10,ij}$ captures dipole-octupole and quadrupole-quadrupole interactions. α_i is the atom-in-solid polarizability,

$$\alpha_i = \frac{V_i}{V_{i,\text{free}}} \alpha_{i,\text{free}}, \qquad (2.54)$$

which is proportional to the ratio between the in-solid (V_i) and free ($V_{i,\text{free}}$) atomic volumes. $\langle M_l^2 \rangle$ ($l = 1, 2, 3, \dots$) are the expectation values for the exchange-hole multipole moments. Although equation 2.50 only includes pairwise contributions to the dispersion energy, some electronic many-body effects are accounted for in the construction of the dispersion coefficients from the exchange hole, where the fully-interacting electron density is used.[10,21]

As can be seen in equations 2.51–2.53, the exchange-hole multipole moments are required for the formulation of the XDM dispersion coefficients. The expectation values for these multipole moments are

$$\langle M_l^2 \rangle_i = \int w_i(\mathbf{r}) \rho_\sigma(\mathbf{r}) [r^l - (r - d_{X\sigma})^l]^2 d\mathbf{r}, \qquad (2.55)$$

where w_i are the weights used in the partitioning scheme for the density. For XDM this is achieved through Hirshfeld partitioning:[132,133]

$$w_i(\mathbf{r}) = \frac{\rho_{i,\text{free}}(\mathbf{r})}{\sum_j \rho_{j,\text{free}}(\mathbf{r})}, \qquad (2.56)$$

where $\rho_{i,\text{free}}$ is the sphericalised free atomic density of atom i. The partitioning weights have a value of one near atom i and close to zero elsewhere.

In equation 2.55, the exchange-hole dipole,

$$d_{X\sigma}(\mathbf{r}) = \int \mathbf{r}' h_{X\sigma}(\mathbf{r}, \mathbf{r}') d\mathbf{r}' - \mathbf{r}, \qquad (2.57)$$

can be calculated either exactly, which uses the occupied orbitals, or using the Becke-Roussel (BR) exchange-hole model.[10,95] In the solid-state, the BR model is employed to approximate the exchange-hole, $h_{X\sigma}$, due to the prohibitive computational cost associated with exact exchange.[131] The BR approach also has been shown to be in better agreement with reference data for molecular C_6 dispersion coefficients and binding energies, due to the fact that it models a localised exchange-hole. This is a better approximation to the full exchange-correlation hole than can be described by Hartree-Fock alone, as it captures some aspects of electron correlation by the assumption that the exchange-hole is localised to a region of atomic size.[10,134,135] The BR model hole relies on

the (spin-dependent) electron density (ρ_σ), the density gradient ($\nabla \rho_\sigma$), Laplacian ($\nabla^2 \rho_\sigma$), and the kinetic-energy density (τ_σ), [10,95] and hence is classified as a meta-GGA functional.

The BR exchange hole is modeled as an off-centre exponential function with the form Ae^{-ar}, which is centred at a distance b from the position of the reference electron. [10] There are three constraints for the BR model that uniquely determine the three parameters: the exponent (a), normalisation (A), and displacement (b). These constraints are that the hole must be normalised to -1 electron, deplete to the spin density at the reference point, and have the same curvature as the exact exchange hole at the reference point. [10] These constraints cause the normalisation to be

$$A = -\frac{a^3}{8\pi},$$
(2.58)

and the density to be

$$\rho_\sigma = \frac{a^3}{8\pi}e^{-ab}.$$
(2.59)

The exact exchange hole curvature is

$$Q_\sigma = \frac{1}{6}\left[\nabla^2 \rho_\sigma - 2\tau_\sigma + \frac{1}{2}\frac{(\nabla \rho_\sigma)^2}{\rho_\sigma}\right],$$
(2.60)

which gives the curvature constraint the form of

$$Q_\sigma = \frac{\rho_\sigma}{6b}(a^2 b - 2a),$$
(2.61)

where the positive-definite kinetic-energy density τ_σ is

$$\tau_\sigma = \sum_i |\nabla \psi_{i,\sigma}|^2,$$
(2.62)

where $\psi_{i,\sigma}$ are the occupied Kohn-Sham orbitals. The values of a and b are fixed by solving the non-linear equations that arise from combining equations 2.59 and 2.60 for $x = ab$:

$$\frac{xe^{-2x/3}}{(x-2)} = \frac{2}{3}\pi^{2/3}\frac{\rho_\sigma^{5/3}}{Q_\sigma},$$
(2.63)

which is canonically solved via the numerical Newton-Raphson method. [95,136] The exchange-hole dipole is $d_{X_\sigma} = b$, where b is calculated from the rearrangement of equation 2.59 and the valence spin density to give

$$b^3 = \frac{x^3 e^{-x}}{8\pi \rho_\sigma},$$
(2.64)

with the added constraint that, in certain cases, non-physical values of the exchange-hole dipole

23

moment can be obtained, which manifests as $b > r$. In these cases, $b > r$ is replaced with $b = r$ as a limiting value.[10,131]

The last key component for discussion of equation 2.50 is the inclusion of a damping function (f_n). The perturbation-theory expansion of the dispersion energy is valid in the limit of large atomic separation, but a damping function is used to prevent a divergence of the dispersion energy at small internuclear distances and correct errors in the multipolar expansion.[10] XDM takes a novel approach to this issue: rather than requiring the dispersion energy to approach zero at small separations, it instead requires the dispersion force to approach zero and the dispersion energy approach a constant, non-zero value.[10] This is consistent with the perturbation theory result for two hydrogen atoms at short interatomic separations.[137] The Becke-Johnson damping function used in XDM[10,138] is given by

$$f_n(R_{ij}) = \frac{R_{ij}^n}{R_{ij}^n + (a_1 R_{c,ij} + a_2)^n}. \qquad (2.65)$$

The van der Waals radius is given by $a_1 R_{c,ij} + a_2 = R_{\text{vdw}}$, where the critical damping radius, $R_{c,ij}$, denotes the internuclear distance at which the multipolar expansion breaks down and the successive dispersion-energy terms become equal:

$$R_{c,ij} = \frac{1}{3} \left[\left(\frac{C_{8,ij}}{C_{6,ij}} \right)^{\frac{1}{2}} + \left(\frac{C_{10,ij}}{C_{6,ij}} \right)^{\frac{1}{4}} + \left(\frac{C_{10,ij}}{C_{8,ij}} \right)^{\frac{1}{2}} \right]. \qquad (2.66)$$

The parameters a_1 and a_2 are fit through the minimisation of residual errors between the computed and reference binding energies for non-covalently bound dimers.[10] It is important to note that these parameters depend upon the density functional employed by the calculation, but are transferable to all elements.[10]

The contribution to the atomic forces and the stress tensor can be approximated from differentiating equation 2.50, under the assumption that the dispersion coefficients are approximately constant. For atom i, the force would be[131]

$$\mathbf{F}_{\text{disp},i} = \sum_{\mathbf{L}} {\sum_{j}}' \sum_{n=6,8,10} \frac{n C_{n,ij} R_{ij\mathbf{L}}^{n-2}}{(R_{\text{vdw},ij}^n + R_{ij\mathbf{L}}^n)^2} \mathbf{R_{ij\mathbf{L}}}, \qquad (2.67)$$

and components of the stress tensor are represented by

$$\sigma_{\text{disp},\xi\eta} = -\frac{1}{2V} \sum_{\mathbf{L}} {\sum_{ij}}' \sum_{n=6,8,10} \frac{n C_{n,ij} R_{ij\mathbf{L}}^{n-2} (R_{ij\mathbf{L}})_\xi (R_{ij\mathbf{L}})_\eta}{(R_{\text{vdw},ij}^n + R_{ij\mathbf{L}}^n)^2}, \qquad (2.68)$$

where the V is the cell volume and ξ, and η are two Cartesian coordinates of x, y, z. Again, the primed notation here denotes $i \neq j$ for $\mathbf{L} = 0$. The assumption of constant dispersion coefficients,

although not exact, does not drastically affect most geometry optimisations.[131] In practice, several iterations of geometry relaxations, each with fixed dispersion coefficients, are done in order to converge the dispersion coefficients and the structural geometry. This approach has been shown to result in the same geometry for molecular solids as if the dispersion coefficients were recalculated at each step, while being more computationally efficient.[131]

2.6 ALTERNATIVE POST-SCF DISPERSION METHODS

There are a number of alternative dispersion-correction methods for DFT. These can be broken down into two categories, the first of which builds dispersion into the base exchange-correlation functional. The second, which will be the main focus of this work, is an *a posteriori* correction added to the self-consistent DFT energy, of which the exchange-hole dipole moment (XDM) model is one example. Of this second type, there are two other classes of dispersion corrections in common use, one being the Grimme D series (D2, D3, D4) and the other being the Tkatchenko-Scheffler (TS) and subsequent many-body dispersion (MBD) correction models. These two series of models differ from XDM in the way they arrive at the components of equation 2.50. These methods will now be discussed and compared with the XDM model.

2.6.1 THE GRIMME-D SERIES OF DISPERSION MODELS

Grimme and colleagues have introduced a series of dispersion corrections, which are add-ons to the standard Kohn-Sham density functionals.[30–32,139] The Grimme-D2 model[30] contains only the leading-order dispersion-energy term and the $C_{6,ij}$ heteroatomic dispersion coefficients are derived from the geometric mean of the homoatomic values:

$$C_6^{ij} = \sqrt{C_6^i C_6^j}.$$ (2.69)

The homoatomic dispersion coefficients are based on DFT/PBE0 calculations of atomic ionisation potentials, I_p, and static dipole polarizabilities, α:

$$C_6^a = 0.05 N I_p^a \alpha^a,$$ (2.70)

where N is the maximum atomic number for the elements in each given row of the periodic table. The damping function used for Grimme-D2 is given by

$$f^{\text{WY}}(R_{ij}) = \frac{1}{1 + \exp\left[-d\left(\frac{R_{ij}}{s_6 R_r} - 1\right)\right]},$$ (2.71)

where R_{ij} is the interatomic distance and R_r is the sum of atomic van der Waals radii, treated as parameters for each element.[140–142] The D2 model's damping function has two empirical parameters, $d = 20$ is a fixed parameter, and s_6, a global scaling factor, which depends on the functional used and is determined by a least-squares minimisation of interaction-energy errors for 40 noncovalently bound complexes.[30] D2 has since been improved through the reformulation of the dispersion coefficients, the inclusion of higher-order dispersion terms, the optional inclusion of the three-body Axilrod-Teller-Muto (ATM) dispersion term, and the use of the Becke-Johnson damping function, to give the D3 model.[31,32]

Instead of using an empirical basis for the formulation of the dispersion coefficients in D2, the D3 dispersion coefficients are derived from the Casimir-Polder formula:[31]

$$C_6^{AB} = \frac{3}{\pi} \int_0^\infty \alpha^A(i\omega)\alpha^B(i\omega)d\omega, \tag{2.72}$$

where $\alpha(i\omega)$ is the averaged dipole polarizability at an imaginary frequency $(i\omega)$, which may be computed by time-dependent DFT. In practice, the dispersion contributions are calculated not directly from equation 2.72, and therefore are not computed from the $\alpha(i\omega)$ values for free atoms. Instead, they are computed for reference stable hydrides, which exist for every element in the periodic table except for the rare gases.[31] The C_6^{AB} coefficient can then be calculated using the stable hydrides A_mH_n and B_kH_l as reference molecules, with no contribution from the hydrogen atoms:

$$C_6^{AB} = \frac{3}{\pi} \int_0^\infty \frac{1}{m} \left[\alpha^{A_mH_n}(i\omega) - \frac{n}{2}\alpha^{H_2}(i\omega) \right] \cdot \frac{1}{k} \left[\alpha^{B_kH_l}(i\omega) - \frac{l}{2}\alpha^{H_2}(i\omega) \right]. \tag{2.73}$$

Here, $\alpha^{H_2}(i\omega)$ is the polarizability of the dihydrogen molecule, m, k, n, l are stoichiometric factors, and $\alpha^{A_mH_n}(i\omega)$ and $\alpha^{B_kH_l}(i\omega)$ are the reference polarizabilities of the stable hydrides. This allows for the D3 C_6 coefficients to be dependent upon the atomic coordination number (CN), which is defined as

$$CN^A = \sum_{A \neq B}^{N} \left[1 + e^{-k_1(k_2(R_{A,cov}+R_{B,cov})/r_{AB}-1)} \right]^{-1}, \tag{2.74}$$

where $R_{A,cov}$ and $R_{B,cov}$ are the scaled covalent single-bond radii for atoms A and B, and the k_1 and k_2 terms are scaling parameters fit for carbon atoms in common structures. The C_6 for any fractional coordination in a specific chemical environment may then be obtained by interpolation between a set of reference values, $C_{6,ref}^{AB}(CN^A, CN^B)$, computed according to equation 2.73.

Early implementations of Grimme-D3 utilise the damping function of Wu and Yang, as in D2.[10] Later implementations of D3 utilise the BJ damping function, detailed above in equation 2.65.[31,32]

This has been shown to improve the results from the D3 model and is now the default in most applications.[31,32]

The higher-order C_8 coefficients are also included and computed in a recursive fashion via

$$C_8^{\mathrm{AB}} = 3C_6^{\mathrm{AB}}\sqrt{Q^{\mathrm{A}}Q^{\mathrm{B}}}, \tag{2.75}$$

where

$$Q^{\mathrm{A}} = s_{42}\sqrt{Z^A}\frac{\langle r^4\rangle^{\mathrm{A}}}{\langle r^2\rangle^{\mathrm{A}}} \tag{2.76}$$

is the effective charge for atom A. $\langle r^4 \rangle$ and $\langle r^2 \rangle$ are radial expectation values, which are computed from geometrically averaged atomic densities, Z^A is the nuclear charge factor, and s_{42} is a scaling factor fit for noble-gas atoms.[31]

Grimme's D3 model also has the ability to include a three-body ATM dispersion term. This leading non-additive term is derived from third-order perturbation theory and, for three atoms ABC, is written as

$$E^{\mathrm{ABC}} = \frac{C_9^{\mathrm{ABC}}(3\cos\theta_a\cos\theta_b\cos\theta_c + 1)}{(R_{\mathrm{AB}}R_{\mathrm{BC}}R_{\mathrm{CA}})^3}, \tag{2.77}$$

with $\theta_{a,b,c}$ representing the internal angles of the triangle formed between atoms ABC, and C_9^{ABC} the triple-dipole dispersion coefficient,

$$C_9^{\mathrm{ABC}} = \frac{3}{\pi}\int_0^\infty \alpha^{\mathrm{A}}(i\omega)\alpha^{\mathrm{B}}(i\omega)\alpha^{\mathrm{C}}(i\omega)d\omega, \tag{2.78}$$

which is approximated as

$$C_9^{\mathrm{ABC}} \approx -\sqrt{C_6^{\mathrm{AB}}C_6^{\mathrm{AC}}C_6^{\mathrm{BC}}}. \tag{2.79}$$

The ATM term is similarly damped to remove non-physical effects that arise from the dispersion contributions at close range.[32,33]

Although the Grimme-D3 model addresses many of the issues found in D2, it is still fundamentally an advanced form of atom typing.[21] For many systems, D3 can be as accurate as XDM, but D3 falls short in some cases due to the fact that it is not dependent on the density and cannot respond to changing chemical environments. Some headway has been made in the formulation of D4, which is the newest iteration of the series, and includes an oxidation-state dependence but, at this time, has not been extensively tested for solids.[21,143]

2.6.2 THE TS AND MBD METHODS

The final post-SCF dispersion methods to be discussed are the works pioneered by Tkatchenko and colleagues, in the form of TS[35] and MBD.[36,37,144,145] The TS scheme, much like the Grimme-D2

model, truncates the dispersion-energy expansion at the C_6 term and does not include any higher-order terms, such as C_8. The TS method begins with the Casimir-Polder formula for the free-atom C_6's, as in equation 2.72. In order to arrive at the C_6 terms for an atom inside a molecule or solid, the TS scheme utilises a simple volume scaling:

$$C_{6,ii}^{\text{eff}} = \left(\frac{V_i^{\text{eff}}}{V_i^{\text{free}}} \right)^2 C_{6,ii}^{\text{free}}. \tag{2.80}$$

This makes use of the direct relation between polarizability and volume,

$$\frac{\alpha_i^{\text{eff}}}{\alpha_i^{\text{free}}} \propto \frac{V_i^{\text{eff}}}{V_i^{\text{free}}} = \left(\frac{\int r^3 w_i(\mathbf{r}) \rho(\mathbf{r}) d\mathbf{r}}{\int r^3 \rho_i^{\text{free}}(\mathbf{r}) d\mathbf{r}} \right), \tag{2.81}$$

where the $w_i(\mathbf{r})$ are the Hirshfeld partitioning weights, as detailed above in XDM.[63,132] From this, the effective coefficient $C_{6,ij}$ can be determined as

$$C_{6,ij} = \frac{2\alpha_i \alpha_j C_{6,ii} C_{6,jj}}{\alpha_j^2 C_{6,ii} + \alpha_i^2 C_{6,jj}}. \tag{2.82}$$

TS also uses a Fermi-type damping function of the form[35]

$$f^{\text{WY}}(R_{ij}) = \frac{1}{1 + \exp\left[-d \left(\frac{R_{ij}}{\beta R_{ij}^0} - 1 \right) \right]}, \tag{2.83}$$

where R_{ij}^0 is the sum of the van der Waal radii, and d and β are empirically fit parameters as we saw in D2. For TS, d is fixed to 20, and β is fit for a given functional.

The TS method was improved upon through the inclusion of long-range many-body terms which are neglected in pairwise dispersion models.[37,145] This method is called the many-body dispersion (MBD) model and shows significant improvements over TS.[37,145] The main idea behind MBD is to replace all of the atoms within the system with quantum harmonic oscillators and then to calculate the coupled and uncoupled energy states of that system, this difference in energy is considered to be the MBD dispersion energy.[145] This was achieved by implementing the coupled fluctuating dipole model (CFDM) Hamiltonian,[146–150]

$$\hat{H} = -\frac{1}{2} \sum_i^N \nabla_{\xi_i}^2 + \frac{1}{2} \sum_i^N \left(\omega_i^{\text{SCS}} \xi_i \right)^2 + \sum_i^N \sum_{j<i}^N \omega_i^{\text{SCS}} \omega_j^{\text{SCS}} \sqrt{\alpha_i^{\text{SCS}} \alpha_j^{\text{SCS}}} \xi_i \mathbf{T}_{ij}' \xi_j, \tag{2.84}$$

where $\xi_i = \sqrt{m_i} \Delta r_i$ is the ith atom's mass-weighted displacement from equilibrium, and $\mathbf{T}_{ij}$ is the dipole-dipole interaction tensor.[37,145] The SCS labels on the variables indicate that they were

refined through a self-consistent screening process.[37] The CFDM Hamiltonian is normally solved via matrix diagonalization to yield $3N$ eigenvalues, λ_i, which correspond to the square of the interacting quantum harmonic oscillator frequencies. The MBD interaction energy is then given by the difference between the interacting and non-interacting frequencies, ω_i^{SCS}, as

$$E_{\text{MBD}} = \frac{1}{2} \sum_{i=1}^{3N} \sqrt{\lambda_i} - \frac{3}{2} \sum_{i=1}^{N} \omega_i^{\text{SCS}}. \tag{2.85}$$

MBD attains comparable accuracy to that found with XDM for molecular solids, although convergence issues are common for metals and ionic solids.[21]

2.6.3 NON-LOCAL VAN DER WAALS FUNCTIONALS

As mentioned previously, there is an entirely different approach that can be taken within DFT to address the lack of non-local dispersion terms within the base functional. The van der Waals functionals insert these terms directly into the exchange-correlation functional[51] through a non-local kernel, which acts as a response function packaged into the correlation energy of the E_{XC} term.[51] These functionals all have the general form

$$E_{\text{C}}^{\text{nl}} = \frac{1}{2} \iint \rho(\mathbf{r}) \Phi(\mathbf{r}, \mathbf{r}') \rho(\mathbf{r}') d\mathbf{r} d\mathbf{r}', \tag{2.86}$$

where ρ is the total electron density and Φ is the non local correlation kernel. There are many different approaches to how this kernel is formed and with what base functional it is paired. Two of the main categories are the VV functionals of Vydrov and Voorhis[53,54] and the vdW-DF functionals of Langreth and Lundquvist.[42,43]

The vdW functionals have various caveats, such as the level of empiricism, the computational expense of having the dispersion energy recalculated at each SCF step,[53] and the sensitivity to the choice of base functional.[51] Whilst some work well for intermolecular interactions, they are much more computationally expensive than the post-SCF corrections for large systems due to their poor scaling with system size.

CHAPTER 3

REQUIREMENTS FOR AN ACCURATE DISPERSION-CORRECTED DENSITY FUNCTIONAL

This chapter is adapted from: A. J. A. Price, K. R. Bryenton, and E. R. Johnson, "Requirements for an accurate dispersion-corrected density functional" *J. Chem. Phys.* **154**, 230902 (2021), with the work performed by K. R. Bryenton omitted.

3.1 INTRODUCTION

The dispersion interaction, first described by London,[12] arises from the formation of instantaneous dipole moments in the electron distributions of atoms or molecules. For a pair of atoms separated by a distance R, long-range attraction between these instantaneous dipoles leads to the following expression for the dispersion energy:

$$E_{\text{disp}} = -\frac{C_6}{R^6},\tag{3.1}$$

where C_6 is referred to as the dispersion coefficient. More sophisticated treatments of dispersion use second-order perturbation theory to obtain the dispersion coefficients.[22,151] The perturbation-theory treatment allows formulation of the dispersion energy as a series expansion that includes terms arising from higher-order instantaneous multipole moments in the atomic electron-density distributions:[13,23]

$$E_{\text{disp}} = -\frac{C_6}{R^6} - \frac{C_8}{R^8} - \frac{C_{10}}{R^{10}} - \cdots\tag{3.2}$$

This series expansion is commonly truncated at either the C_8 or C_{10} term,[21,31,138] as higher-order terms will have negligible contributions to the overall dispersion energy.[152] Finally, third-order perturbation theory can be used to evaluate non-additive dispersion terms involving interaction of three instantaneous atomic dipole moments.[153,154] The leading-order term is the C_9, or Axilrod-Teller-Muto, dispersion energy. Its contribution is typically smaller in magnitude than the pairwise C_{10} dispersion energy.[21,152]

Despite the importance of London dispersion in chemistry, biochemistry, and materials science, it has proved difficult to include in quantum chemical simulations since dispersion physics are missing from mean-field electronic structure methods, such as Hartree-Fock (HF) theory and Kohn-Sham density-functional theory (DFT).[155] Initial HF+D[156] and DFT+D[135,139,157–159] dispersion corrections included only the leading-order C_6 term and used empirical dispersion corrections for a limited set of elements. Subsequently, the D2[30] and TS[35] dispersion methods provided simple general prescriptions to obtain C_6 coefficients for any element of the periodic table. The drawback of these models is that they have either no dependence, or only weak dependence, on an atom's charge state, coordination, and other factors that determine its local chemical environment. More complex methods that allow increased variation of the atom-in-molecule or atom-in-solid dispersion coefficients have been proposed and are still the focus of ongoing development. Such methods include the D3,[31,32] D4,[33,34] density-dependent correction (dDsC),[41] and exchange-hole dipole moment (XDM) models,[10,63] as well as the many-body dispersion (MBD) method[36] and its fractionally ionic (FI)[160] and uMBD[161] variants.

For each of the models listed above, the dispersion energy is added as a post self-consistent-field (SCF) correction to a base density functional:

$$E_{\mathrm{DFT}} = E_{\mathrm{base}} + E_{\mathrm{disp}}, \tag{3.3}$$

where E_{DFT} is the total DFT energy, E_{base} is the base-functional energy, and E_{disp} is the dispersion correction. As an alternative to these post-SCF dispersion models, there is also the class of explicitly non-local dispersion methods, based on the van der Waals (vdW) density functional,[162] of which there are now many variations.[42,43,53,54,163–166] However, while capable of high accuracy, these methods are inherently more computationally expensive than post-SCF models due to their non-locality and will not be discussed further here.

In this chapter, we discuss the most desirable requirements for a reliable and physically reasonable post-SCF dispersion-corrected density functional. Development of such a method necessitates high accuracy for both components of the total DFT energy in Eqn. 3.3: the base density functional and the dispersion correction. Users of dispersion-corrected DFT should be mindful of these

requirements when deciding on their computational methodology. Method developers should also consider them when designing new base functionals and dispersion models.

3.2 REQUIREMENTS FOR THE BASE DENSITY FUNCTIONAL

3.2.1 DISPERSIONLESS

The key requirement for a base density functional to pair with a dispersion correction is that it is itself dispersionless.[167] While dispersion is properly a highly non-local correlation effect, the local density approximation (LDA) and many generalized gradient approximation (GGA) functionals capture some 'dispersion-like' binding through their exchange functionals.[93,94,168–172] This arises because the large-gradient regions in the free atoms (or isolated molecules) do not contribute sufficiently to the exchange energy, causing the isolated monomers to be artificially destabilised relative to the vdW complex. However, this 'dispersion-like' binding is inherently short range and does not recover the proper C_6/R^6 asymptotic behaviour that is characteristic of London dispersion. This is shown by examination of the potential-energy curves for graphite exfoliation, in Figure 3.1, for the LDA and selected GGA functionals without any dispersion correction.

Adding a dispersion correction to an exchange functional that already mimics dispersion-like binding means that there will be some double counting of dispersion at short range,[173,174] resulting in significant overstabilisation of vdW complexes. Empirical parameterisation to equilibrium binding energies will therefore result in excessive damping of the dispersion correction at short range to reduce this double counting, resulting in under-estimation of dispersion at intermediate ranges. This leads to poor performance for bulk materials, where there are many intermediate-range atomic contacts, such as in molecular crystals.[79]

Double counting of dispersion from the base functional also creates a bias where hydrogen-bonding interactions are favoured relative to dispersion-driven interactions, such as π-stacking.[75] Hydrogen-bond strengths (as in the water dimer) are typically over-estimated with non-dispersionless functionals.[183] In such cases, the addition of a dispersion correction, unfortunately, introduces greater errors for hydrogen bonds, despite being necessary to stabilise dispersion-bound dimers (such as the methane or benzene dimers). Hence, parameterizing the damping function for a non-dispersionless base functional typically leads to over-stabilization of hydrogen-bonded complexes, and under-stablisation of dispersion-bound complexes, to minimise the overall error.[75]

To design a dispersionless GGA functional, we must ensure accurate large-gradient behaviour, which is determined by the functional's enhancement factor. GGA exchange functionals have the
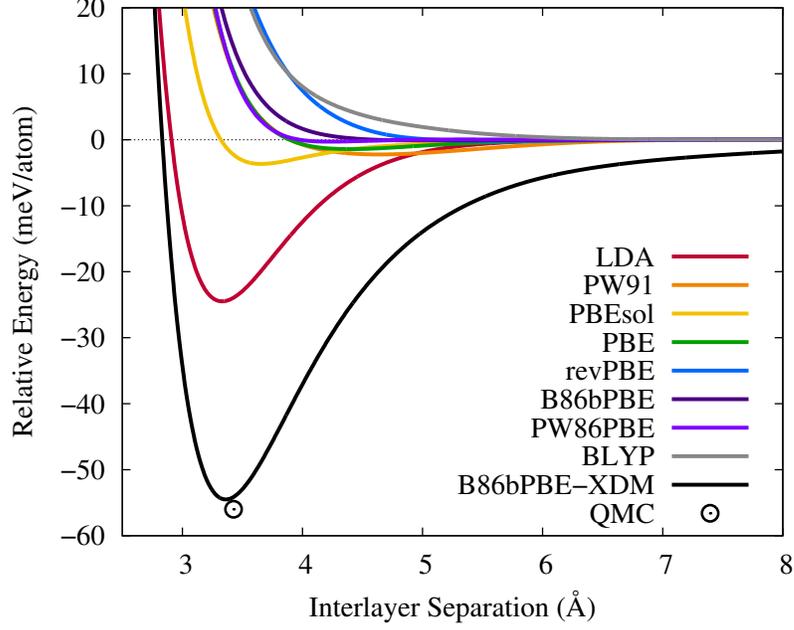
Figure 3.1: Graphite exfoliation energies computed using selected base density functionals.[8,9,175–181] Calculations were performed with Quantum ESPRESSO,[109] using the projector augmented-wave (PAW) approach[104,105] with a $12 \times 12 \times 4$ **k**-point mesh. Planewave cut-offs of 100 and 1000 Ry were used for the kinetic energy and electron density, respectively. The B86bPBE-XDM[8–10] curve and Quantum Monte Carlo equilibrium result[182] are included for reference.

general form:

$$E_X^{\mathrm{GGA}} = \sum_\sigma \int \varepsilon_{X,\sigma}^{\mathrm{LDA}} F(\chi_\sigma) d\mathbf{r}, \qquad (3.4)$$

where $\varepsilon_{X,\sigma}^{\mathrm{LDA}}$ is the spin-dependent LDA exchange-energy density

$$\varepsilon_{X,\sigma}^{\mathrm{LDA}} = -\frac{3}{2} \left(\frac{3}{4\pi} \right)^{1/3} \rho_\sigma^{4/3}, \qquad (3.5)$$

ρ_σ is the σ-spin density,

$$\chi_\sigma = \frac{|\nabla \rho_\sigma|}{\rho^{4/3}} \qquad (3.6)$$

is the reduced density gradient, and $F(\chi_\sigma)$ is the enhancement factor. Enhancement factors for several popular GGA exchange functionals[8,9,176–180] are shown in Figure 3.2.

For a proper, dispersionless exchange functional, the large-gradient limit of the enhancement factor should be[93,94,168]

$$\lim_{\chi_\sigma \to \infty} \propto \chi_\sigma^{2/5}. \qquad (3.7)$$

This limit is only obeyed by the PW86[176] and B86b[8] GGAs. As seen in Figure 3.2, the LDA,
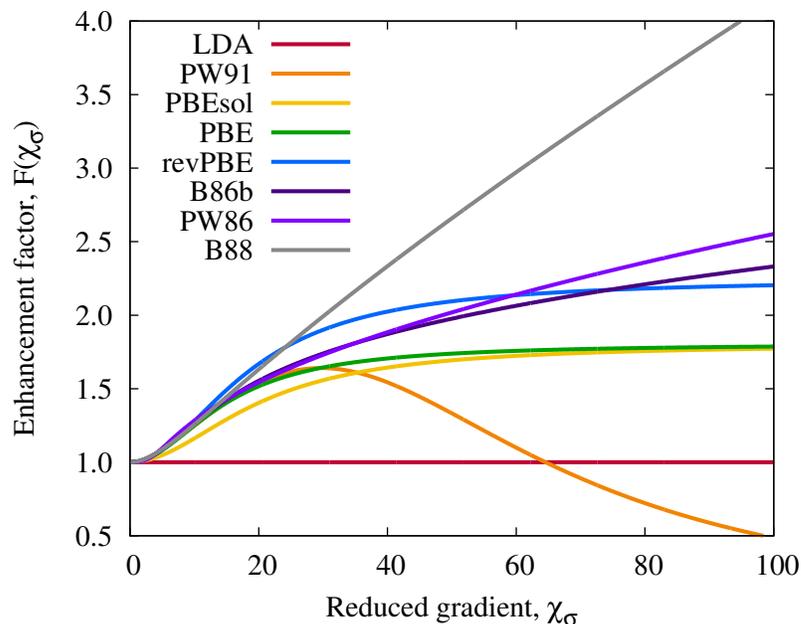
Figure 3.2: Plots of the exchange enhancement factor for the LDA and selected GGA[8,9,176–180] functionals.

PBEsol,[179] PW91,[178] and PBE[9] functionals all have enhancement factors that converge to a constant value in the large-gradient limit. As a result, these functionals mimic dispersion-like binding at short range to varying extents, making them unsuited for use with asymptotic dispersion corrections. Conversely, the B88[177] GGA has an enhancement factor that diverges too quickly, proportional to $\chi_\sigma / \ln(\chi_\sigma)$, in the large-gradient limit. While B88 exchange is dispersionless, it notoriously over-estimates non-bonded repulsion. One strategy to obtain a dispersionless GGA functional has been to take linear combinations of PBE and B88,[184] as in the APF[185] functional.

In light of their correct behaviour in the large-gradient limit, PW86 and B86b are the GGA functionals of choice to pair with dispersion corrections to avoid any double counting of dispersion. Due to their accurate description of non-bonded repulsion, these functionals have demonstrated improved performance for solid-state benchmarks when paired with unscaled dispersion corrections,[10,79] as well as with non-local vdW functionals.[43] This is illustrated in Table 3.1 for several XDM-corrected GGA functionals, which shows mean absolute errors (MAEs) for the KB49[74,75] benchmark of molecular-dimer binding energies and the X23[79,80] benchmark of molecular-crystal lattice energies. Similar results are obtained with the two dispersionless base functionals: B86bPBE and PW86PBE. However, we tend not to favour PW86PBE for the practical reason that it shows poorer SCF convergence. Larger errors are obtained for the non-dispersionless base functionals: PBE, PBEsol, and PW91. While non-dispersionless functionals tend to perform better when paired with a zero-damped dispersion correction due to error cancellation (see Sec 3.3.1), we note that the

Table 3.1: Mean absolute errors, in kcal/mol, obtained with selected XDM-corrected GGA functionals for the KB49[74,75] and X23[80] benchmarks. All calculations were performed with Quantum ESPRESSO[109] using the projector augmented-wave (PAW) approach[104,105] and planewave cut-offs of 80 and 800 Ry. A $4 \times 4 \times 4$ **k**-point mesh was used for all molecular crystals, while only the Γ point was used to treat the isolated molecules. The B86bPBE, PW86PBE, and PBE results were taken from Ref. 10, while the PW91 and PBEsol results were determined for the present work. The a_1 and a_2 parameters used in the XDM damping function are also shown.

Functional	a_1	a_2 (Å)	KB49	X23
B86bPBE	0.6512	1.4633	0.41	0.85
PW86PBE	0.6836	1.5045	0.41	0.88
PBE	0.3275	2.7673	0.50	1.11
PW91	0.0000	4.0228	0.63	1.89
PBEsol	0.5432	2.3686	0.78	2.11

largest such error reduction for the KB49 set is only 0.06 kcal/mol for PBEsol. In our opinion, the large-gradient limit of Eqn. 3.7 is a drastically underutilised constraint in functional development.

3.2.2 NUMERICALLY STABLE

A further desirable requirement for a base density functional is that it should be numerically stable and give smooth potential energy curves for vdW complexes. While this criterion is met by the LDA and GGA functionals, some meta-GGAs display substantial numerical sensitivities that lead to oscillations in potential energy curves of vdW dimers.[170,186–190] These oscillations can result in errors in vibrational frequencies[191] and geometry optimisation to spurious high-energy conformations,[192] unless extremely fine integration grids are used. Previous works observing such oscillations have focused on meta-GGA calculations for gas-phase complexes, such as the π-stacked benzene dimer.[193] However, similar behaviour can also be seen for solid-state systems, as shown in Figure 3.3 using the SCAN functional for the example of graphite exfoliation. From the figure, we additionally note that the SCAN functional is not dispersionless and it provides significant, spurious dispersion-like binding for graphite.

In the Quantum ESPRESSO program,[109] the real-space integration grid is controlled by the choice of the planewave cut-off for the density expansion (the `ecutrho` parameter). It is common to take this value as only four times the wavefunction planewave kinetic-energy cutoff, `ecutwfc`, when using norm-conserving pseudopotentials. However, as seen in Figure 3.3, low values of `ecutrho` are clearly insufficient and result in massive oscillations of the potential energy curve. Indeed, a smooth potential can only be obtained with SCAN if `ecutrho` is increased to near 2500 Ry or higher. The need for very large real-space integration grids contributes to the increased
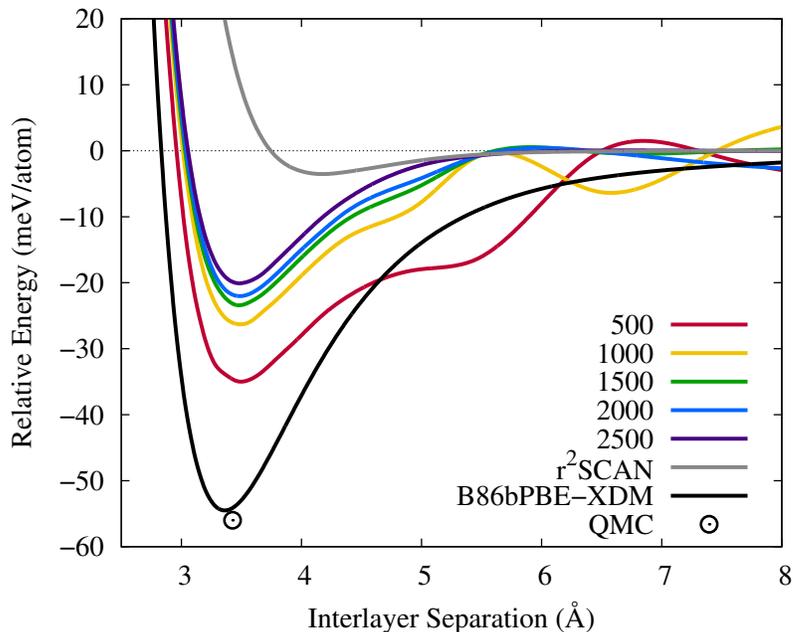
Figure 3.3: Graphite exfoliation energies computed using the SCAN[194] meta-GGA functional with varying planewave cut-offs (in Ry) for the electron density expansion. Calculations were performed using Quantum ESPRESSO,[109] with norm-conserving pseudopotentials[102,103] and a $12 \times 12 \times 4$ **k**-point mesh. A planewave cut-off of 80 Ry was used for the kinetic energy. Results from the r²SCAN[195] functional using the same options and a 1000 Ry density cut-off are also shown. The B86bPBE-XDM[8–10] curve, using the same calculation options described in Figure 3.1, and the Quantum Monte Carlo equilibrium result,[182] are included for reference.

computational cost of meta-GGA functionals relative to GGAs, making them unappealing for large systems even if they were dispersionless.

In previous work,[186] we showed that the numerical sensitivity of meta-GGAs arises from the behaviour of the τ-dependent ratios employed in these functionals for the low-density and low-gradient regions near the bond critical points of vdW dimers. Here, we focus on the SCAN meta-GGA functional,[194] which is finding increasingly widespread use in solid-state applications. Comparisons will be made to a new, more numerically stable modification of SCAN, termed r²SCAN.[195] As seen in Figure 3.3, the r²SCAN functional remedies the numerical sensitivities seen with SCAN, giving a well-behaved potential energy curve for graphite exfoliation with a modest value of `ecutrho`.

The τ-dependent term used in SCAN is[194]

$$a = \frac{\tau_\sigma - \tau_\sigma^W}{\tau_\sigma^{\text{UEG}}},$$
(3.8)

where we have used a, as opposed to the symbol α used in Ref. 194, to avoid confusion with the atomic polarizability. Here,

$$\tau_\sigma^W = \frac{1}{4} \frac{(\nabla \rho_\sigma)^2}{\rho_\sigma} \tag{3.9}$$

is the von Weizsäcker kinetic-energy density and

$$\tau_\sigma^{\text{UEG}} = \frac{3}{5} \left(6\pi^2\right)^{2/3} \rho_\sigma^{5/3} \tag{3.10}$$

is the kinetic-energy density of the uniform electron gas (UEG). A similar τ-dependent ratio is used in the r^{2}SCAN functional:

$$\bar{a} = \frac{\tau_\sigma - \tau_\sigma^W}{\tau_\sigma^{\text{UEG}} + \eta \tau_\sigma^W}, \tag{3.11}$$

where $\eta = 10^{-3}$. As seen previously[186] for ingredients of other meta-GGAs, such as VSXC,[196] M06-L,[58] and TPSS,[197] there is sharp variation in both a and $\bar{a}$ near the bond critical point of van der Waals dimers.

The key difference between the SCAN and r^{2}SCAN functionals that controls their numerical stability lies in the form of the "x" functional, which is one component of the exchange term. In SCAN, this functional is

$$
\begin{aligned}
x(\chi_\sigma, a) \;=\; & \mu \chi_\sigma^2 \left[1 + \left(\frac{b_4}{\mu} \chi_\sigma^2 \right) \exp\left(-\frac{b_4}{\mu} \chi_\sigma^2 \right) \right] \\
& + \left[b_1 \chi_\sigma^2 + b_2 \left(1 - a\right) \exp\left\{ -b_3(1-a)^2 \right\} \right]^2,
\end{aligned}
\tag{3.12}
$$

which involves τ_σ dependence. However, in r^{2}SCAN, this functional is replaced by

$$x(\chi_\sigma) = \left[C \exp\left(-d\chi_\sigma^4\right) + \mu \right] \chi_\sigma^2, \tag{3.13}$$

which is independent of τ_σ. Here, C, d, μ, and all the b_n's are constants.

Figure 3.4 shows the values of the x functionals used in the SCAN and r^{2}SCAN meta-GGAs along the internuclear coordinate of the argon dimer. While both functionals approach zero at the bond critical point, they give very different behaviour on either side of it. The SCAN results show a divergence that increases with internuclear distance, which is not present in the r^{2}SCAN results. The region surrounding the critical point is only sparsely sampled by atom-centred integration grids,[129] and the exchange-energy contributions can vary substantially depending on the precise location of the sampled grid points relative to the peaks of the diverging function. Overall, r^{2}SCAN is well behaved for molecular dimers,[200] while SCAN and TPSS show reduced numerical sensitivity[186,193] compared to VSXC[196] and several of the related Minnesota functionals.[58,59]
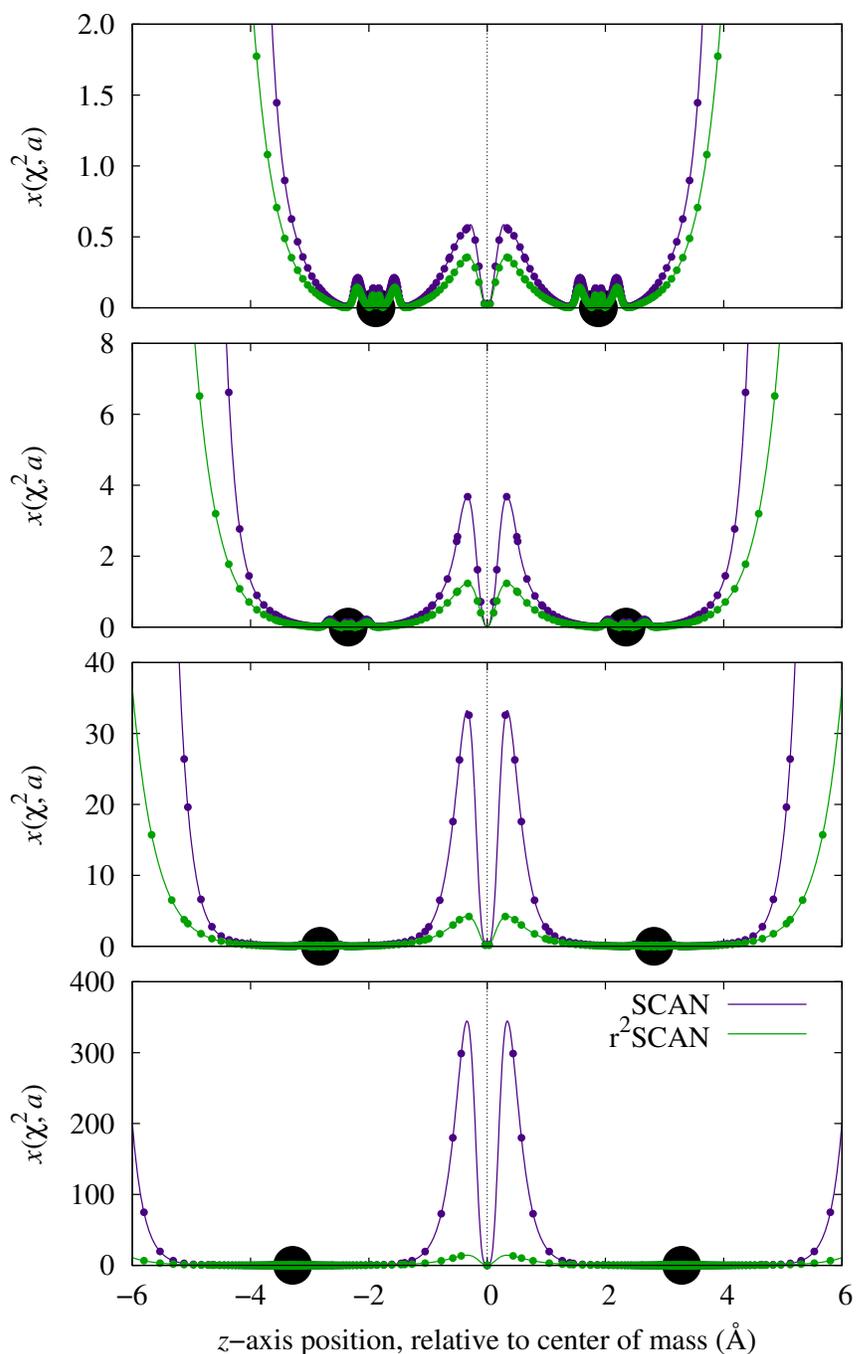
Figure 3.4: Values of the x function (Eqns. 3.12 and 3.13) used in the SCAN[194] and r²SCAN[195] functionals along the internuclear axis of Ar_2. Results are shown for internuclear distances of 1.00, 1.25, 1.50, and 1.75 times its equilibrium separation of 3.76 Å.[198] Calculations were performed using the NUMOL program.[199] The black circles indicate the positions of the Ar nuclei and the points correspond to the integration grid for a mesh containing 120 radial points per atom.

Care should be taken by developers to select well-behaved τ-dependent terms for use in the design of meta-GGA functionals. Meta-GGAs can offer significant improvements over GGAs for electronic properties, such as the band gap, and tend to give lower errors for molecular thermochemistry. However, they are much more computationally costly than GGAs for solid-state applications, frequently suffer from poor SCF convergence, and cannot be used for variable-cell relaxations in most planewave codes. Users are strongly cautioned to avoid most meta-GGA functionals for applications where dispersion plays a role since they are frequently numerically unstable. Even meta-GGA functionals that are well-behaved numerically are still not dispersionless, as is shown for r^{2}SCAN in Figure 3.3.

3.2.3 MINIMAL DELOCALIZATION ERROR

Delocalization error is almost certainly the key outstanding challenge in DFT today. This error is a result of overstabilisation of systems with highly delocalised electrons, especially in situations with fractionally charged atoms.[201–205] It affects a diverse range of chemical problems, including charge-transfer complexes[206–210] and excitations,[211–218] transition states of radical reactions,[219–224] and molecules with extended π-conjugation,[225–230] to list only a few examples. Manifestations of delocalization error can be broken down[231] into energy-driven examples, as in the case of stretched H_2^+,[232–234] and density-driven examples, such as solution-phase ions.[204,235,236] Delocalization error is most prevalent with GGA functionals and can be reduced through use of hybrid functionals that include some fraction of exact (HF) exchange mixing. Moreover, performing DFT energy evaluations on HF densities, termed density-corrected DFT,[237,238] is particularly effective in mitigating density-driven errors.

While delocalization error is well studied in finite molecules, much less is known about how it affects solid-state calculations, other than causing the notorious band-gap problem.[239–244] This is largely because hybrid calculations can be prohibitively expensive in planewave codes for all but the smallest unit-cell sizes. Despite this restriction, it has been recently shown[245] that dispersion-corrected hybrid functionals yield more accurate lattice energies than dispersion-corrected GGAs for the X23 benchmark and, particularly, for halogen-bonded molecular crystals. Dispersion-corrected hybrid functionals also significantly improve upon GGAs for predictions of alkali-halide lattice constants[246] and the relative stabilities of diamond and graphite.[230] A dramatic example of delocalization error in the solid state occurs for organic acid-base co-crystals, where geometry optimisation with GGA functionals can result in a spurious proton transfer, incorrectly yielding an organic salt structure.[247]

In some cases, particularly involving strong halogen bonding, delocalization error causes over-binding of a vdW complex with the base functional alone.[248–251] This means that the addition of

Table 3.2: Lattice energies (in kcal/mol per molecule) of the 7 ice polymorphs for which reference experimental data, back corrected for thermal and vibrational effects, is available from the ICE10 benchmark.[253] Also shown are lattice energies computed using a series of hybrid functionals[245] based on B86bPBE-XDM,[8–10] with specified percentages of exact (HF) exchange mixing. Calculations used the Quantum ESPRESSO program,[109] norm-conserving pseudopotentials,[102,103] and a $4 \times 4 \times 4$ **k**-point mesh. Plane-wave cutoffs of 100 and 1000 Ry were used for the kinetic energy and electron density, respectively. MAE: mean absolute error, relative to the reference data.

| Phase | % HF exchange mixing | | | | | | Ref. |
	0%	10%	20%	30%	40%	50%	
Ih	16.77	16.28	15.87	15.52	15.22	14.97	14.07
II	15.97	15.57	15.24	14.96	14.73	14.55	14.05
III	16.15	15.70	15.32	15.00	14.72	14.50	13.85
VI	15.55	15.20	14.90	14.67	14.47	14.33	13.68
VII	14.46	14.22	14.04	13.92	13.85	13.81	13.07
VIII	14.46	14.23	14.05	13.93	13.85	13.81	13.31
IX	16.28	15.85	15.48	15.17	14.92	14.71	13.97
MAE	1.95	1.58	1.27	1.02	0.82	0.67	

a dispersion correction results in larger binding-energy errors, despite adding necessary, missing physics. GGA functionals also tend to overstabilize co-operative hydrogen bonding due to delocalization error.[252] To highlight an example of delocalization error in molecular crystals, we focus on the lattice energies of seven ice polymorphs.[253] As shown in Table 3.2, the B86bPBE-XDM functional over-estimates the ice lattice energies. This error can be reduced, although not eliminated, by using hybrid functionals with increased fractions of exact-exchange mixing. Perhaps further increases in accuracy could be obtained with dispersion-corrected range-separated hybrids with long-range exact exchange,[254–257] although such calculations are not yet feasible in plane-wave codes.

An ideal base functional to pair with a dispersion correction would be free of delocalization error and dispersionless. In practice, this cannot yet be achieved as hybrids do not completely resolve the delocalization error problem and the optimal amount of exact exchange required to minimize it can be highly system dependent.[258–260] Moreover, hybrid functionals still are not practical for most solid-state calculations with planewave basis sets. The use of finite, numerical[64,122] or Gaussian[261] basis sets for molecular crystals leads to significant errors from basis-set incompleteness. This highlights the need for ongoing density-functional development to reduce and ideally resolve delocalization error, with a particular eye to the solid state.

3.3 REQUIREMENTS FOR THE DISPERSION CORRECTION

3.3.1 FINITE DAMPING

We now turn to the requirements for the dispersion correction itself. The use of a perturbation-theory expansion for the dispersion energy, as in Eqn. 3.2, is correct at infinite separation, but will break down for short interatomic distances. As a result, the dispersion energy for each atom pair is typically multiplied by an empirical damping function, $f(R)$, to avoid divergence as $R \to 0$. Considering only the leading-order C_6 dispersion term for simplicity, the damped dispersion energy between a single pair of atoms is

$$E_{\text{disp}} = -\frac{C_6 f(R)}{R^6}. \tag{3.14}$$

Many possibilities have been proposed for the damping function. The D2[30] and TS[35] dispersion corrections use a Fermi-type damping function of the form

$$f^{\text{WY}}(R) = \frac{1}{1 + \exp\left[-d\left(\frac{R}{sR_{\text{vdW}}} - 1\right)\right]} \tag{3.15}$$

proposed by Wu and Yang,[158] where d and s are empirical parameters and R_{vdW} is the sum of the atomic van der Waals radii. In the D3[31] dispersion correction, an alternative damping function

$$f^{\text{CHG}}(R) = \frac{1}{1 + 6\left(\frac{R}{sR_{\text{vdW}}}\right)^{-\gamma}} \tag{3.16}$$

proposed by Chai and Head-Gordon[257] was used, where γ and s are again empirical parameters. Both the Wu-Yang and Chai–Head-Gordon functions ensure that the damped dispersion energy reaches zero, as shown in Figure 3.5. However, the Chai–Head-Gordon form prevents the divergence seen with the Wu-Yang damping function at very small internuclear separations. Damping the dispersion energy to zero means that its magnitude will reach a maximum value at some intermediate interatomic distance. At shorter distances, a non-physical, repulsive dispersion force will be introduced.

Using a convergent multipole expansion, it can be shown that the dispersion energy should approach a small, but finite, value for two hydrogen atoms in the united-atom limit.[137] As a result, we and others[32] favour damping the dispersion energy to a constant, finite value as $R \to 0$. One way to achieve this is through the Becke-Johnson damping function,[138] which has the form

$$f_n^{\text{BJ}}(R) = \frac{R^n}{R^n + R_{\text{vdW}}^n}, \tag{3.17}$$
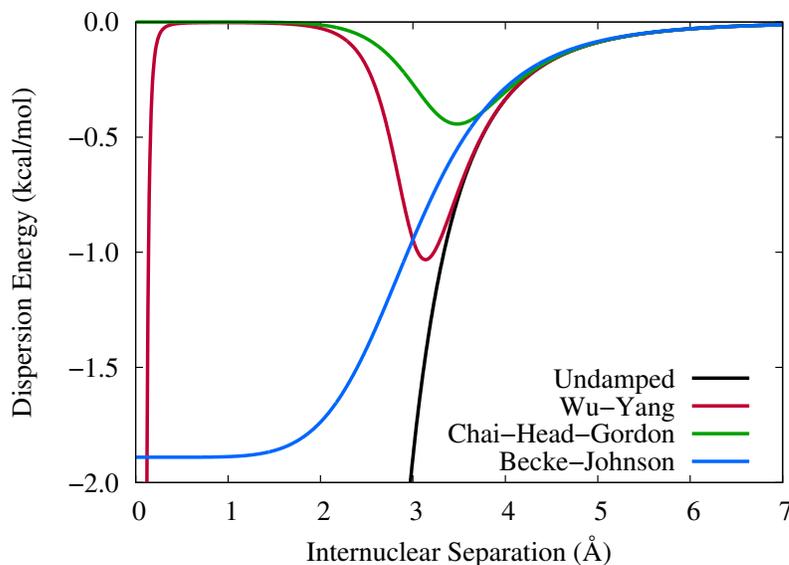
Figure 3.5: Comparison of the effect of various damping functions[138,158,257] on the leading-order dispersion energy between two atoms with $C_6 = 100$ a.u. and $R_{\mathrm{vdW}} = 3$ Å. Empirical parameters in the damping functions are set to $s = 1$, $d = 20$, $\gamma = 14$.

where $n = 6$ for the leading-order C_6 dispersion term. It should also be noted that the definitions of the vdW radii differ between various damping functions, but will not be discussed further here. The BJ form ensures that the dispersion energy approaches a non-zero constant at short interatomic separations, as shown in Figure 3.5, and the dispersion force is never repulsive.

The choice of damping function can significantly affect the performance of a dispersion correction, as demonstrated for the D3 dispersion model.[32] When D3 was paired with dispersionless or near-dispersionless base functionals, finite damping produced substantially more accurate results than zero damping for several molecular benchmarks. However, zero-damping does provide improved performance for exchange functionals that mimic short-range dispersion binding,[173,174] such as SCAN[194] and the Minnesota functionals.[58,59] In these cases, zero damping ensures that the dispersion energy will reach a maximum at intermediate atomic distances, while being significantly damped near equilibrium geometries of vdW complexes to offset the dispersion-like binding from the base functional. This results in accurate potential energy curves through a cancellation of errors. An example of such a cancellation of errors is shown in Figure 3.6. Here, the SCAN meta-GGA is paired with the zero-damped D3 dispersion correction, and the results are compared with B86bPBE-XDM for graphite exfoliation. While the component base functionals and dispersion corrections show very different behaviours, the overall potential energy curves from SCAN-D3 and B86bPBE-XDM are similar.
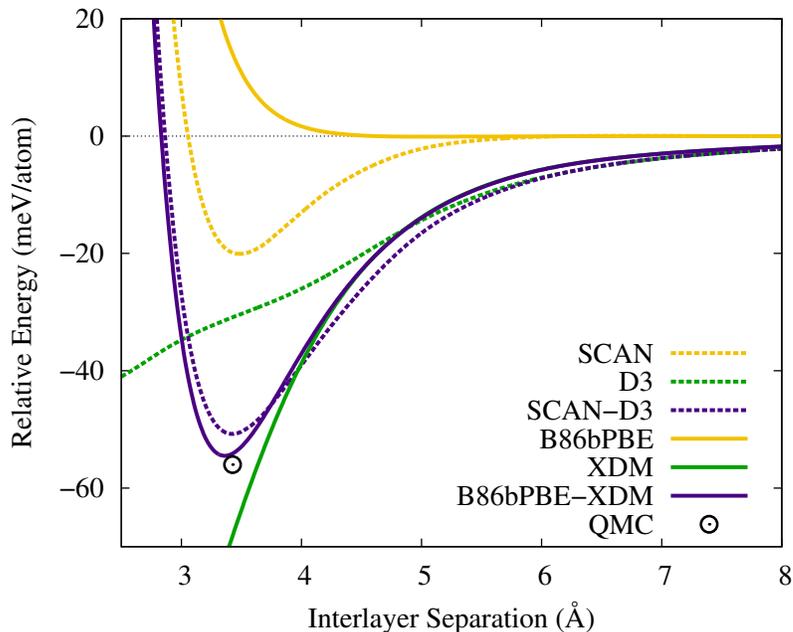
Figure 3.6: Comparison of SCAN-D3 (dashed lines) and B86bPBE-XDM (solid lines) results for graphite exfoliation. Results are shown for the separate base functionals (yellow) and dispersion corrections (green), as well as for the overall potentials (purple).

3.3.2 HIGHER-ORDER DISPERSION TERMS

Force-field implementations[27–29] of the LJ potential,[262] as well as early DFT+D dispersion corrections[135,139,157–159] including the D2[30] and TS[35] models, limited themselves to inclusion of only the leading-order C_6 dispersion term. If empirical scaling of the dispersion coefficients[30] is introduced, such methods can be capable of very high accuracy for small molecular dimers.[21,30] However, without empirical scaling, C_6-only models severely under-bind π-stacked complexes.[135] It has since been shown that the inclusion of at least the C_8 dipole-quadrupole term in the dispersion-energy expansion, as in the D3[31] and XDM[10] models, is required for an accurate treatment of π-stacking,[138] graphite exfoliation,[21] the S12L benchmark[263] of supermolecular complexes,[264] and lattice energies of molecular crystals.[21,79]

We recently presented a detailed discussion of the importance of higher-order terms in a dispersion model.[21] Table 3.3 shows a summary of our previous results, where we employed a damped dispersion-energy, summed over all atom pairs, of the form

$$E_{\mathrm{disp}} = - \sum_{i<j} \sum_{n=6,8,10} \frac{C_{6,ij} f_n^{\mathrm{BJ}}(R_{ij})}{R_{ij}^n}, \tag{3.18}$$

truncating it at either the C_6, C_8, or C_{10} terms. The BJ damping function involves two parameters (a_1 and a_2) in the definition of the vdW radii (Eqn. 3.17) that were optimized in each case to

Table 3.3: Mean absolute errors (MAE), and mean errors (ME), in kcal/mol, for the KB49[74,75] and X23[80] benchmarks using the B86bPBE[8,9] base functional and selected XDM dispersion terms. Multiplication of a term by s indicates empirical scaling. Table adapted from Ref. 21.

Dispersion Energy	KB49		X23	
Expression	MAE	ME	MAE	ME
$E^{(6)}$	0.83	-0.23	1.97	-1.59
$E^{(6)} + E^{(8)}$	0.48	0.02	0.94	-0.38
$E^{(6)} + E^{(8)} + E^{(10)}$	0.41	0.03	0.86	-0.26
$sE^{(6)}$	0.38	0.01	1.78	1.61
$E^{(6)} + sE^{(8)}$	0.40	0.01	0.84	0.08

minimise the error for the KB49[74,75] molecular-dimer benchmark. In the last two rows of the table, the value of an empirical scaling coefficient multiplying the C_6 or C_8 dispersion-energy terms was also optimised. Errors for the X23[80] lattice-energy benchmark were then evaluated using these optimised parameters.

The results in Table 3.3 show that the unscaled C_6 dispersion-energy term alone performs poorly for the KB49 molecular benchmark, since neglect of higher-order dispersion terms results in under-estimation of the dispersion energy. This model similarly results in systematic under-binding of the X23 molecular-crystal benchmark, leading to a mean error of -1.59 kcal/mol. Conversely, the use of a scaled C_6 term gives excellent performance for molecules, but this requires doubling of the dispersion coefficients, with $s = 2.09$.[21] However, the scaled C_6 dispersion correction now results in systematic over-estimation of the molecular-crystal lattice energies, with a mean error of 1.61 kcal/mol. This over-binding occurs because the scaled C_6 term's asymptotic decay is too gradual to properly account for higher-order C_8 dispersion. Consequently, the dispersion stabilisation is overestimated for the many atomic contacts in a molecular crystal with large internuclear separations. This error is not seen for small molecular dimers since, unlike for periodic solids, there will be few distant atomic contacts. Overall, inclusion of a C_8 term, either scaled or unscaled, is necessary for simultanous good performance on both benchmarks.

Interestingly, contrasting the X23 results using scaled and unscaled C_6 dispersion terms provides insight into systematic trends in atomic dispersion coefficients employed in common molecular-mechanics force fields.[27–29] We recently found that force-field dispersion coefficients are typically 1.5 times greater than the corresponding XDM values for small organic molecules,[265] as well as for biomolecules.[266] It was argued that the 1.5 factor allowed the force fields to approximate contributions from the higher-order C_8 dispersion term using only the C_6 dispersion term available in the LJ potential. From the XDM results above, the mean errors for the X23 set obtained using

scaled ($s = 2.09$) and unscaled ($s = 1$) C_6 dispersion terms are equal in magnitude, but opposite in sign. Thus, the scaling of $s = 1.5$ found for the force fields should evenly balance the over- and under-binding tendencies and would be expected to give near zero mean error for the molecular crystals and other condensed-phase systems. However, using a scaled C_6 dispersion term to account for C_8 dispersion contributions neglects important physics and can still result in substantial error. Explicit inclusion of the C_8 term in both DFT dispersion corrections and molecular-mechanics force field is favoured and, for example, can considerably increase the accuracy of a four-point-charge water model.[267]

3.3.3 RESPONSE TO ATOMIC ENVIRONMENT, A.K.A. ELECTRONIC MANY-BODY EFFECTS

There has been substantial confusion in the literature regarding the meaning of many-body dispersion. This pertains to deconvoluting non-additivities in the dispersion coefficients themselves, termed electronic many-body effects in our recent work,[21] from the triple-dipole Axilrod-Teller-Muto (ATM)[153,154] and higher-order non-pairwise terms in the perturbation-theory expansion of the dispersion energy. The ATM terms can be modeled even using fixed dispersion coefficients, and can optionally be included in calculations using the D3 dispersion correction.[31] However, such triple-dipole terms are not required for an accurate dispersion correction as they are negligible for most chemical systems,[21,152,268] although they can be significant in π-stacked dimers depending on the choice of damping function.[269] Conversely, electronic many-body effects are not included in empirical dispersion methods that use fixed dispersion coefficients, but can have massive effects on the dispersion energies.[16,21,270,271] Manifestations of electronic many-body effects include changes in dispersion coefficients due to varying atomic charge state, coordination, and even the presence of neighbouring molecules, all of which collectively determine an atom's chemical environment.

A particularly striking example of electronic many-body effects is the change in leading-order C_6 dispersion coefficients for common transition-metal elements when going from a free atom to the bulk metal.[271] Table 3.4 shows computed C_6 dispersion coefficients for five bulk transition metals,[273] obtained with various dispersion corrections, compared to the free-atomic values.[272] Unfortunately, there is no accurate reference data for atomic C_6 coefficients in bulk metals. However, XDM[10] and TSsurf[272] (which is a pre-cursor to the MBD[36] method) aim to account for variations in dispersion coefficients with chemical environment. As both of these methods provide highly accurate binding energies for the adsorption of benzene on copper, silver, and gold surfaces,[271,272] they should provide the most reasonable C_6's for these metals.

Free transition-metal atoms are highly reactive, possessing loosely bound s-shell electrons. This leads to a much more diffuse and polarizable electron density distribution for a free atom compared

Table 3.4: Homoatomic dispersion coefficients for selected metals, in atomic units; free-atom values[272] are compared with results for the bulk metals[273] using various dispersion corrections. The D3[31] and D4[33] dispersion coefficients are determined entirely by the atom positions, while the TSsurf data was taken from Ref. 272. The XDM results were obtained from B86bPBE calculations using the Quantum ESPRESSO program,[109] the PAW[104,105] approach, a $12 \times 12 \times 12$ **k**-point mesh, and a cold smearing[274] parameter of 0.01 Ry. Planewave cutoffs of 120 and 1200 Ry were used for the kinetic energy and electron density, respectively. The TS[35] dispersion coefficients were evaluated by scaling the free-atom values using the ratio of Hirshfeld volumes[132] obtained from the XDM calculations.

Element	Free Atom	Bulk Metal				
		TS	D3	D4	TSsurf	XDM
Cu	253	230	175	62	59	106
Pd	158	172	266	192	102	92
Ag	339	349	269	134	122	180
Pt	347	383	337	331	120	124
Au	298	342	317	41	134	130

to the bulk metal. As a result, one would expect a marked decrease in homoatomic C_6 coefficients for bulk metals; this is exactly what is seen in Table 3.4 with the TSsurf and XDM methods, albeit to varying extents. Conversely, the TS[35] and D3[31] dispersion models, in which the C_6 coefficients depend only on atomic volumes or local coordination numbers, are too simplistic to correctly account for the expected physics in bulk metals. In several cases, TS and D3 predict higher C_6 coefficients for the bulk metal than for the free atoms, explaining why they overestimate adsorption energies of benzene on these metal surfaces.[271] The very recent D4[33,34] model, which has not yet been comprehensively benchmarked for metal-containing systems, predicts dispersion coefficients that are similar to the free-atom values for Pd and Pt, while showing good agreement with TSsurf for Cu, Ag, and Au.

In prior works, we have highlighted how the description of molecular surface adsorption,[271] and of alkali-halide lattice constants,[246] is improved due to the leading-order C_6 dispersion coefficients' response to variations in chemical environment. More dramatically, inclusion of electronic many-body effects is essential for accurate prediction of the interlayer distances and exfoliation energies of the transition-metal dichalcogenides[16] using a post-SCF dispersion model. Accounting for changes in dispersion coefficients with atomic environment is also essential for modeling exfoliation of other layered materials, such as inorganic minerals.[275] While simple dispersion models have proved highly successful for organic chemistry, incorporating the physics of electronic many-body effects constitutes an additional requirement for a density-functional dispersion model to attain high accuracy for inorganic chemistry.

3.4 SUMMARY

In this chapter, we have presented what we believe are the key requirements for a general and accurate dispersion-corrected density functional. Accuracy of the base functional and dispersion correction should be enforced separately to ensure that their combination properly describes the physics of both non-bonded repulsion and dispersion attraction, rather than relying on error cancellation.

A reliable dispersion correction should use finite-damping to prevent introduction of an artificial, repulsive dispersion force at short internuclear separations. It should involve minimal empiricism and evaluate the dispersion coefficients in such as way that they are dependent on the electron density and, consequently, can respond to changes in chemical environment. Variations in the dispersion coefficients with chemical environment, including changes in charge state, coordination, and the presence of neighbouring molecules, can be captured through density dependence and are collectively termed electronic many-body effects.

A reliable base density functional must be dispersionless, which can be achieved by imposing a constraint on the large-gradient limit of the exchange enhancement factor for GGAs, as well as being numerically stable. These requirements are not met by the majority of meta-GGAs, where the semi-local exchange functionals tend to mimic short-range 'dispersion-like' binding, which is properly a non-local correlation effect. Moreover, many popular functional forms that build in kinetic-energy dependence result in significant numerical sensitivities to the spacing of the integration mesh relative to low-density bond critical points, as occur in vdW complexes. Finally, an accurate base functional should be free of delocalization error. This error can be reduced through use of hybrid and range-separated hybrid functionals, but such approaches are not practical for planewave calculations on most periodic solids. Development of a general method to eliminate delocalization error in both molecular and solid-state systems constitutes what is, in our opinion, the greatest outstanding challenge in density-functional theory.

CHAPTER 4

INTERPRETING SURFACE-ADSORPTION CHEMISTRY FROM STM IMAGES GENERATED VIA DISPERSION-CORRECTED DFT

This chapter is adapted from: M. DeJong, A. J. A. Price, E. Mårsell, T. Gary, G. Nguyen, E. R. Johnson, and S. Burke "Small molecule binding to surface-supported single-site transition-metal reaction centres" *Nat. Comm.* **13**, 7407 (2022). Note: All experimental data included in this chapter were provided by Dr. Miriam DeJong and Prof. Sarah Burke, at the University of British Columbia. All DFT calculations were performed by the author.

4.1   INTRODUCTION

Heterogeneous catalysis employs a catalyst that is a different phase from the reactants and/or products. It is an important tool for the chemical industry, as it allows for the efficient large-scale generation of easily separable chemical products without the need for a further step of specialized separation. Heterogeneous catalysis is estimated to be a part of almost 90% of industrial chemical processes world wide,[276,277] such as in the food, petrochemical, pharmaceutical, and automotive industries.[278–281]

A key step to further develop heterogeneous catalysts is understanding of the fundamental processes at work. Solid-state heterogeneous catalysis follows a process consisting of adsorption of reactant molecules to a surface, reaction, and then dissociation or desorption from the surface. Previous work by Burke and colleagues used supramolecular chemistry and a metal-organic process for self-assembly to make a novel active catalyst material.[282,283] The experimental focus

was generation of atomically precise, self-assembled metal-organic Fe-terpyridine-phenyl-phenyl-terpyridine [4′,4′′′′-(1,4-11-Phenylene)bis(2,2′:6′,2′′-terpyridine), TPPT] complexes on an Ag(111) surface, which were characterized using low-temperature scanning tunnelling microscopy (STM).

In order to investigate the initial steps of heterogeneous catalysis, the adsorbed Fe-TPPT complexes were exposed to either of the gaseous reagents CO or C_2H_4. These reactants have applications in many industrial catalytic processes, such as those in the refining industry, or combustion within car exhausts, where the metal catalyst can activate C=C, C–H, and C–O bonds.[284] Experimental characterization was performed by STM, which allowed for the imaging of Fe-TPPT complexes on the surface, before and after exposure to the gaseous CO and C_2H_4 reagents. The imaging before and after exposure aided in understanding of the aforementioned adsorption process, which is a key step in heterogeneous catalysis. However, STM images alone can be challenging to interpret and assign to specific chemical species on the surface.[285,286] Here, we will use theory in order to provide a more rigorous structural assignment for the experimental results.

Computational modelling of surface adsorption can be done with density-functional theory (DFT). Previous work studying adsorption of benzene on noble-metal surfaces has demonstrated the importance of dispersion and sensitivity to the choice of dispersion correction.[16,271] The process of physisorption is dominated by weak, long-range noncovalent interactions, but has little impact on the electronic levels. A full treatment of these long-range effects is essential to accurately describe the physisorption geometries and energies. A key missing long-range effect within conventional DFT is an accurate treatment of London dispersion interactions, where non-dispersion-corrected functionals yield drastically underestimated adsorption energies relative to those seen in experiment.[271] Empirical dispersion models like D2, D3, and TS[30,31,35] can model these missing forces, but tend to overestimate dispersion binding to metal surfaces.[16,271] This overestimation occurs due to their dispersion coefficients, which are based on free-atom values, being too large for bulk metals. Conversely, the XDM and MBD models have been found to treat the process of surface adsorption accurately, since their dispersion coefficients have been formulated to vary with chemical environment and both show similar changes in C_6 going from a free metal atom to a metal surface or bulk.[271] Such a flexible treatment of dispersion is central to the accurate generation of computational STM results.

In this chapter, we modelled surface adsorption of two common gaseous reactants, CO and C_2H_4, with a model of the Fe-TPPT heterogeneous catalyst on the Ag(111) surface. This was achieved through the use of dispersion-corrected DFT using the XDM dispersion model. For the DFT simulations, a truncated Fe-terpyridine (Fe-tpy) complex was selected in order to accurately recreate the active site of the complex, while reducing the molecular size to make the calculations

49

feasible. From the DFT results, STM images were generated computationally for both prebonded and bonded complexes of CO/C_2H_4 with Fe-tpy on the surface. Excellent agreement has been found between the experimental and theoretical results, allowing us to unambiguously assign structures.

4.2 SCANNING TUNNELLING MICROSCOPY THEORY

STM was developed in 1981 by Binnig and Rohrer and has become a key tool in the imaging of nanostructures at the atomic scale.[287,288] STM offers the possibility of direct real-space determination of 3D surface structures using a sharp single-atom tip to apply a voltage to a sample surface. The STM tip is brought towards the surface at a constant resistance and moved across it via a piezodrive, which consists of 3 independent piezoelectric transducers for each of the cartesian axis (x, y and z).[289] The mapping of the surface then comes via the weakly flowing current, I, which occurs due to the tunnelling of electrons through a vacuum between the tip and surface. The vacuum tunnelling resistance is measured as the tip is scanned, and the height adjusted to maintain a constant tunnelling resistance, which yields a contour map of the surface.[289] The experimental formalism comes from the Bardeen approach,[290] which uses time-dependent perturbation theory. The tunnelling current is given to first order by

$$I = \frac{2\pi e}{\hbar} \sum_{\mu,\nu} f(E_\mu) \left[1 - f(E_v + eV)\right] |M_{\mu\nu}|^2 \delta(E_\mu - E_\nu), \tag{4.1}$$

where $\hbar$ is the reduced Planck's constant, e is the electron charge, $f(E)$ is a Fermi function, V is the applied voltage, E_μ is the energy of the state ψ_μ without tunnelling, and $M_{\mu\nu}$ is the tunnelling matrix elements between the states ψ_μ of the probe and ψ_ν of the surface.[289]

This work utilizes DFT in order to generate STM plots, which are compared to those found via experiment. The DFT calculations rely on the Tersoff-Hamann approximation, where the current is taken to be proportional to the calculated density of states at the Fermi level,

$$I \approx V \rho_{loc}(\mathbf{r}, V), \tag{4.2}$$

where I is the current, and V is the bias voltage. ρ_{loc} is the local density of states at the Fermi-level E_F,

$$\rho_{loc}(\mathbf{r}, V) = \sum_{\mathbf{k},n}^{E_F - eV \to E_F} |\psi_{\mathbf{k},n}(\mathbf{r})|^2, \tag{4.3}$$

where the sum runs over the one-electron states that have energies between $E_F - eV$ and E_F.[291]

50

4.3 COMPUTATIONAL METHODS

This work utilized DFT in order to generate both theoretical STM plots and relative energy rankings for the possible catalytic species. Pseudopotential/plane-wave calculations were conducted using the projector-augmented-wave (PAW) method[104] with Quantum ESPRESSO,[109] version 5.1.1. All calculations employed the B86bPBE functional[8,9] with the XDM dispersion correction.[10,63,131] As discussed previously, the XDM model evaluates the dispersion energy as a sum over all atomic pairs, and is damped to prevent divergence at short internuclear separations. This damping function is controlled through two adjustable parameters based on the functional used. The value of the two adjustable parameters in the XDM damping function were set to their standard values for use with B86bPBE of $a_1 = 0.6512$ and $a_2 = 1.4633$ Å.[10]

The Ag (111) surface was modelled as a 4-layer, 4×4 supercell with lattice constants fixed to their XDM-optimized values throughout this work. This model surface was selected as it has been previously shown to correctly capture the physisorption interaction between organic molecules and noble-metal surfaces when combined with XDM.[271] Next a truncated form of the Fe-TPPT complex was generated, consisting of Fe coordinated to a single terpyridine molecule (Fe-tpy), as shown in Fig. 4.1(a). Initial geometry relaxation was performed with the Fe-tpy placed on the Ag (111) surface and all surface atoms frozen. The planewave cutoffs for the kinetic energy and charge density were 50 and 400 Ry, respectively. These first calculations were done at the gamma point for both the singlet and triplet states of the Fe-tpy. For the triplet state, the initial spin bias was set to 2 and an unconstrained starting magnetization of 0.1 was placed on the Fe atom. Following this a second relaxation was performed, where the atomic positions of the Fe-tpy molecule and top two layers of the Ag (111) surface were allowed to relax. These calculations used an increased **k**-point grid of $2 \times 2 \times 1$, with the same planewave cutoffs for the kinetic energy and charge density. Finally, single-point energy calculations were conducted for evaluation of the adsorption energy. For the single-point energy calculations, a **k**-point mesh of $2 \times 2 \times 1$ and planewave cutoffs for the kinetic energy and charge density of 60 and 600 Ry were used, following previous work on molecular surface adsorption.[271]

Starting from the optimized Fe-tpy structure, two catalytically active reactants, carbon monoxide (CO) and ethylene (C_2H_4), were placed on the surface in a number of candidate configurations involving prebonding and bonding to the Fe-tpy complex to probe the energetic landscape and reflect the possible structures observed in the experimental STM images. Their geometries were then relaxed following the same computational protocol described above for the reactant-free Fe-tpy complex. These structures either collapsed to the prebonding and bonding configurations shown in Fig. 4.2 or were considerably higher in energy and not considered further. For all structures, except

that of CO in the bonded configuration which has a single ground state, both the triplet and singlet states were considered and the triplet found to be more stable. Supercell calculations were also undertaken on CO, C_2H_4, and triplet Fe-tpy at the gamma point, with cutoffs for the kinetic energy and charge density of 60 and 600 Ry, in order to evaluate the binding energies.

The STM images were generated using the critic2[291] code from the resultant densities from the single-point calculations with 1.2 times the number of occupied bands. critic2[291] uses the Tersoff-Hamann approximation,[289] where the modelled current, I, is proportional to the applied bias voltage and local density of states (LDOS) at the Fermi level. To generate the STM plots a sample bias of -0.02 eV was selected to match experiment.

4.4 RESULTS AND DISCUSSION

4.4.1 STM IMAGES

STM's ability to directly image reactive sites is a key tool in the understanding of experimental patterning of novel substrates on a surface. One key weakness for pure STM studies is the assignment of structures, where educated guesses can be made as to what each image reflects, but structures cannot be conclusively assigned with 100% accuracy. Here, we compare both experimental and theoretical STM results. As noted above, the experimental system consists of a Fe adatom coordinated to a single terpyridine (tpy) group of a terpyridine-phenyl-phenyl-terpyridine (TPPT) molecule adsorbed on an Ag (111) surface. The DFT calculations used a truncated Fe-tpy molecule, containing just the active site of the complex with a central Fe adatom coordinated to the terminal terpyridine (tpy) group, adsorbed on a 4-layer Ag (111) surface.

A schematic structural representation, as well as DFT-generated and experimental STM images are shown Fig. 4.1 for the Fe-tpy (or Fe-TPPT) complex, without any coordinated gaseous reactants. Fig. 4.1(a) is a schematic representation of the minimum-energy DFT structure. Both the singlet and triplet configurations of the Fe-tpy were considered, with the triplet yielding a lower energy by 5.9 kcal/mol. The DFT results place the z-distance of the Fe and the average position of the rings at 2.34 and 3.2 Å, respectively, above the surface, which is in line with the physisorption seen for organics on metal surfaces.[271]

The experimental and simulated STM images for Fe-typ are compared in Fig. 4.1(b,c). In the experimental STM in Fig. 4.1(c), the iron is coordinated to the tpy on the right side of the TPPT molecule, as indicated by the orange diamond. The increased brightness upon iron coordination is due to the increased electron density on the negatively charged tpy ring. In both DFT and experimental STM images, the tpy bonded to the iron has a fairly uniform height, with the iron appearing less bright. This can be explained by both the partial positive charge on the Fe and the

Figure 4.1: Minimum-energy geometry of the Fe-terpyridine complex on the Ag(111) surface (a), along with the simulated STM image (b). All computational simulations were truncated to just the orange region of subfigure (c), due to computational costs. The experimental image for the full Fe-TPPT complex (d) is shown for comparison, adapted from Ref. 292.

(a) Fe-typ on Ag(111)

(b) Simulated STM

● tpy
✦ tpy-Fe

(c) Fe-tpy monomer
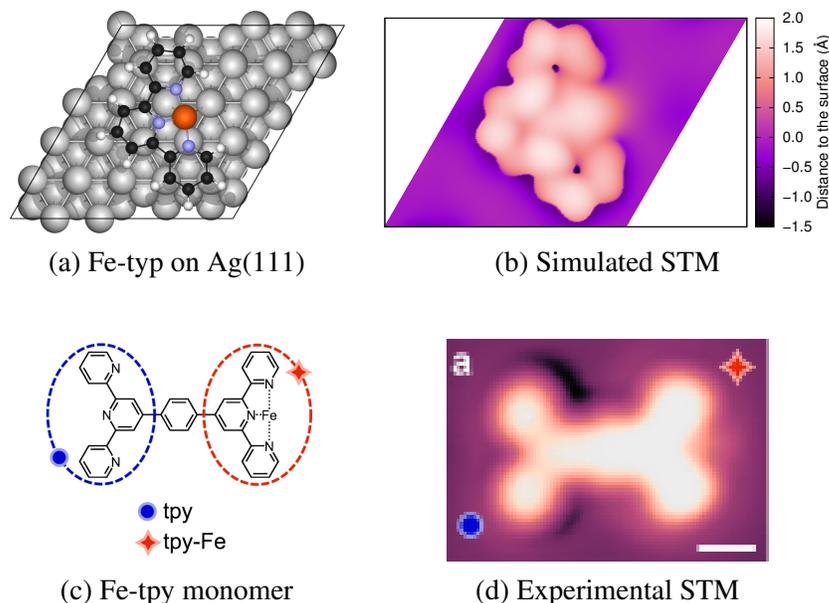
(d) Experimental STM

fact that, in the DFT structure, the Fe has a vertical height above the surface of 0.9 Å less than the average ring height.

The experimental focus of the work by Burke and colleagues was to use STM imaging to explore the reactivity of the active Fe sites when exposed to low concentrations of CO and C_2H_4. In the experimental results after CO exposure, two key motifs were observed. The first involves CO in a prebonded configuration, where it physisorbs to the Ag(111) surface, and the second motif corresponds to a bonded configuration, where the CO is fully coordinated to the metal centre through a metal-carbon bond. As the proportion of the bonded species increased compared to the prebonded when annealed, the prebonded structure was determined to be a metastable surface-bound intermediate.[292]

The DFT results for CO yielded only two stable motifs, seen in Fig. 4.2. All placements of the CO near the Fe would collapse to either the bonded or prebonded configuration (or some much higher energy state) depending on the choice of starting geometry. For the prebonded motif in Fig. 4.2(a), the CO is perpendicular to the surface and forms a C-Ag σ-bond, as well as η^2 coordination to the Fe atom. For the bonding motif in Fig. 4.2(d), the CO is σ-bonded to the Fe through the carbon atom and lies parallel to the surface. The DFT calculations found that, in the prebonded configuration, the CO-Fe-tpy complex was most stable in the triplet state, consistent with

53

Figure 4.2: Low-energy structures of CO and C_2H_4 coordinated with the adsorbed Fe-tpy complex, along with the simulated STM image. The experimental images for the full Fe-TPPT complexes with the adsorbed species are shown for comparison, adapted from Ref. 292.

(a) CO prebonded   (b) Simulated STM   (c) Experimental STM

(d) CO bonded   (e) Simulated STM   (f) Experimental STM

(g) C_2H_4 prebonded   (h) Simulated STM   (i) Experimental STM

(j) C_2H_4 bonded   (k) Simulated STM   (l) Experimental STM
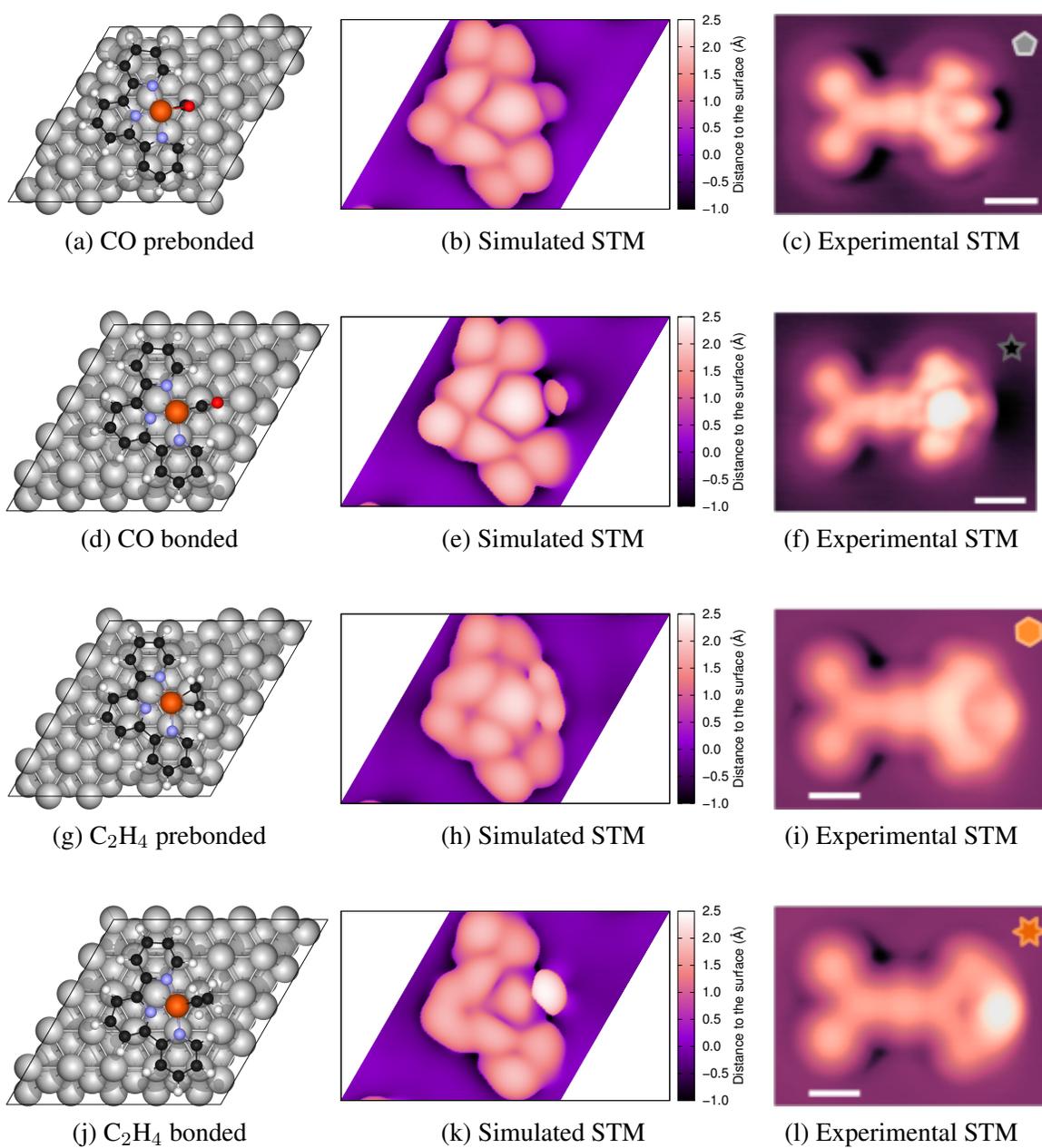
the free Fe-tpy complex, whereas the CO-Fe-tpy complex is a singlet in the bonded configuration. This is likely due to the strong Fe–CO bond, which is stabilized by the back-bonding from the metal d-orbital to the π^* orbital of the CO that occurs in the bonded form of the complex. Back donation of electron density to the π system of the CO extends the C–O bond and shortens the metal-carbon bond. This phenomenon changes the electronic state of the Fe centre to the singlet, instead of the triplet found for the CO prebonded or free Fe-tpy complexes.

The DFT-optimized geometries were then used to generate simulated STM images, as seen in the middle column of Fig. 4.2. For the simulated STM of the bonded state in Fig. 4.2(e), the two parts of the CO molecule can be seen, with the carbon corresponding to the bright half sphere to the right of the Fe centre and a darker shadow behind it, which corresponds to the oxygen atom. When comparing the CO simulated and experimental STM images (Fig. 4.2(e-f)), both have the bright spot to the right of the Fe from the electron-rich carbon. Additionally, the central iron is much brighter and more prominent than in any other of the simulated or experimental STM images.

The simulated STM of the prebonded CO complex in Fig. 4.2(b) shows a single, slightly less-intense sphere to the right of the iron centre, which is the top of the CO oxygen. The simulated image also shows a region that appears to be below the surface as a black halo around the CO molecule. Although comparing the simulated and experimental STM images for the CO prebonded configuration, as in Figs. 4.2(b,c), provides a less convincing assignment, both do contain rounded features with a darkened halo to the right of the Fe centre. These consistent features and the prediction of only two low-energy motifs allowed for conclusive assignments of the experimental structures.

For C_2H_4, two key motifs were observed experimentally for direct interaction with the Fe-tpy complexes, which were again labelled as prebonded and bonded. The experimental work found that switching between the prebonded and bonded forms could be induced by a voltage pulse of -0.7 V for 200 ms and was entirely reversible.[292] The assignment of which is the prebonded versus the bonded configuration was based on the measurement of differential conductance by scanning tunnelling spectroscopy (STS) localized around the Fe site, which would indicate a new Fe-centred state.[282,283] The bonded structure showed a new tunnelling resonance at -0.30 V, that was absent in the prebond species.[292] This indicated a downward shift of the Fe-centred state and a stabilization of the Fe-centred HOMO from the interaction with C_2H_4. Again via differential conductance measurements, the C_2H_4 prebonded configuration was found to be more stable than the CO prebonded configuration.[292]

The experimental STM images were compared to DFT results to further assign specific structures. As for CO, the C_2H_4 calculations either collapsed into one of two motifs, or were found to be much

higher in energy. When investigating the two low-energy DFT motifs, the less-stable prebonded configuration in Fig. 4.2(g) was found to have the two carbons parallel to the Ag surface, whereas the carbons are perpendicular to the surface for the more-stable bonded configuration in Fig. 4.2(j). In the case of the prebonded motif, the iron centre lies 2.76 Å from the silver surface, which is significantly higher than in the bonded form, where the distance is 2.47 Å. Additionally, as for the bare Fe-tpy complex, the complex is most stable in the triplet state for both C_2H_4 motifs.

For the prebonded C_2H_4 configuration in Fig. 4.2(h), the simulated STM shows a rounded termination with a uniform intensity to the right of the Fe centre, which corresponds to the C_2H_4 molecule as it associates with the iron. For the bonded configuration in Fig. 4.2(k), the simulated STM image has two distinct features. The first is an altered iron centre, which is triangular in shape with a reduced intensity similar to the rest of the Fe-tpy complex. The second is a much brighter lobe to the right of the Fe centre, which corresponds to the C_2H_4 molecule oriented perpendicular to the surface.

Unlike those for CO, the experimental and simulated STM images for C_2H_4 show clear similarities that allowed for straightforward assignments. The STM images for the prebonded motif in Fig. 4.2(h,i) share the key features of the low-intensity hemispherical lobes to the right of the Fe centre, arising from the C_2H_4 molecule parallel to the Ag surface. For the bonded configuration of C_2H_4 in Fig. 4.2(k,l), both STM images have a localized bright spot to the right of the Fe due to one ethylene CH_2 group that is pointed up, away from the Ag surface.

4.4.2 BINDING ENERGIES

The Fe-tpy complex adsorbs to the Ag(111) surface via a physisorption process. Due to this, accurate prediction of geometries and energies requires the inclusion of London dispersion interactions. This is illustrated in Table 4.1, which shows the binding energy of Fe-tpy on the surface, along with the decomposition into the base-functional and XDM contributions. In the case of Fe-tpy, dispersion accounts for 77% of the binding energy, illustrating the importance of its inclusion for these systems. Previous calculations for benzene adsorbed on an Ag(111) surface,[271] which employed the same methodology used here, gave a binding energy of 19.6 kcal/mol, coincidentally in exact agreement with the reference experimental value.[293] That work also gave a binding energy of 18.7 kcal/mol for pyridine on Ag(111).[271] Our binding-energy result for Fe-tpy, which contains 3 pyridine rings, is slightly over 3 times higher than that for the single pyridine on Ag(111). The additional contribution to the binding energy primarily comes from the dispersion interactions between the Fe atom and the surface.

Table 4.1 also shows the binding energies for both CO and C_2H_4 with the adsorbed Fe-tpy complex, for both bonded and prebonded arrangements. Although reduced here, the dispersion

56

Table 4.1: Computed binding energies (BE) for the Fe-tpy species, along with the base-functional (BE_{base}) and XDM dispersion (BE_{disp}) contributions. Values for Fe-tpy are relative to the clean Ag(111) surface and isolated molecule. Values for CO and C_2H_4 are relative to the isolated molecules and the adsorbed Fe-tpy complex on the surface. All quantities are in kcal/mol.

Species	Configuration	BE	BE_{base}	BE_{disp}
Fe-tpy	–	78.3	21.3	57.0
CO	prebond	17.6	10.5	7.1
C_2H_4	prebond	19.9	8.1	11.8
C_2H_4	bond	26.4	14.7	11.7
CO	bond	48.6	42.0	6.6

contributions range from 13.6% to 60% of the total binding energies and are still important to the overall picture of reactant binding. The binding energies of the prebonded configurations for both CO and C_2H_4 are more sensitive to dispersion than their bonded counterparts are. The smallest dispersion contribution, of only 13.6%, is seen for the CO bonded species, which contains a direct Fe–C bond and has the highest overall binding energy of all four CO or C_2H_4 species discussed.

The DFT calculations for both CO and C_2H_4 found that the prebonded motifs were less stable than the bonded configurations, which confirmed our assignments of the STM results. The small binding-energy difference between the C_2H_4 prebonded and bonded configurations listed in Table 4.1 is consistent with the reversible conversion found in the STS experiments. For the cases of the CO motifs, the large binding-energy difference between the prebonded and bonded species in Table 4.1 is also consistent with the non-reversible conversion seen experimentally upon annealing. Another feature of the STS results that was recovered by the DFT binding energies is the fact that the CO prebonded complex was found to be less stable than the C_2H_4 prebonded complex. The inclusion of dispersion is key for this specific result, as the base-functional contribution alone (Table 4.1) would invert this ordering.

4.5 CONCLUSIONS

The ability of STM to image at the atomic level makes it a powerful tool for the fundamental understanding of surface chemistry. However, a major weakness of STM is that explicit assignment of the images to exact chemical species and subtle bonding arrangements can be extremely difficult. This work aimed to show DFT as a promising tool to aid in the assignment of STM images to specific species and bonding motifs, through the generation of high-quality physisorption geometries.

The calculations employed the B86bPBE functional[8,9] with the XDM dispersion correction[10,63,131]

to study the physisorption of Fe-tpy, and its complexes with CO and C_2H_4, on an Ag(111) surface. The resulting binding energies and simulated STM images were then compared with the experimental work done by Burke and colleagues.[292] Previous work has shown that dispersion is important for predicting accurate geometries and energetics for the physisorption of organics on metal surfaces.[271] As theoretical STM plots are sensitive to geometry changes, we therefore expected the inclusion of dispersion to be key for the structural assignment between theory and experiment. The importance of including dispersion was also apparent in the energetic ordering as, without it, the relative stabilities of the CO and C_2H_4 prebonded complexes would be switched and no longer reflect the trend seen in experiment.

Further experimental trends seen for the adsorption energies were also reflected within theory. The ability of CO to form a strong $Fe-C$ bond made the structural assignment between prebonded and bonded complexes relatively simple. The large energetic difference agrees with the irreversible conversion to the bonded form seen experimentally after annealing. For the C_2H_4 complexes, it was not initially clear from experiment which structures would correspond to the bonded and prebonded motifs. Inspection of the DFT binding-energy results allowed assignment of the lower-energy structure as the bonded motif and the higher-energy as prebonded, and the calculated STM images had straightforward analogues with experiment. Confidence in the C_2H_4 result arose from the small difference in DFT adsorption energies, consistent with the observed reversibility of interconversion between the two motifs. Ultimately, the calculated binding energies, in addition to the generated STM images, made correct structural assignment of experimental results possible. The power of high-quality dispersion-corrected DFT calculations to aid interpretation of STM studies of surface chemistry can be seen from the results of this work.

CHAPTER 5

XDM IMPLEMENTATION WITHIN FHI-AIMS

The implementation of XDM within the FHI-aims code base, for both finite molecules and periodic solids, is the main goal of this thesis and will be the focus of this chapter. This combination of methods was selected due to the high accuracy of XDM to treat dispersion physics, and the near linear scaling and the routine application of hybrid density functionals to solids enabled with FHI-aims. The XDM implementation relies on the theory described within Chapter 2, particularly the equations found within Section 2.5 of this thesis. FHI-aims is written in the Fortran programming language, with the majority of the code using the Fortran 90 standard. This chapter will detail the modules and subroutines that were directly written for this work, the choices that were made for the implementation, and any modules that were already present within FHI-aims that were modified in the implementation of XDM. The added section present within the FHI-aims manual, describing how users can run an XDM calculation using FHI-aims, is also recreated here. Finally, this chapter will detail the preliminary tests that were undertaken during the creation of the module to ensure accuracy and consistency with other implementations of XDM.

5.1 IMPLEMENTATION

5.1.1 THE XDM MODULE

A new module within the FHI-aims software package[64] was written to contain the calculation of the XDM dispersion energy and the resulting contributions to the atomic forces and, for periodic solids, the stress tensor. The module contains a collection of subroutines:

- `xdm_collect` uses a flag to determine if the XDM dispersion coefficients (Eqns. 2.51-2.53) should be computed or reused from the last time it was called. If the coefficients are to

be computed, it obtains ρ_σ and Q_σ (Eqn. 2.60) at every grid point, and calls a number of routines listed below to obtain the moment integrals and polarizabilities, and then computes the dispersion coefficients. Once the coefficients are in hand, it calls `energy_xdm`.

- `bhole` computes the exchange-hole dipole moment at an arbitrary grid point using the Becke-Roussel[95] (BR) model (Eqn. 2.64), (where `xdm_collect` applies it on a each grid point).

- `xfuncs` computes the difference between both sides of Eqn. 2.63 during solution for the x parameter used in the BR model using the Newton-Raphson algorithm[95,136].

- `xdm_hirshfeld_free` computes the free atomic volumes and polarizabilities.

- `run_xdm_hirshfeld` computes the atomic volumes and exchange-hole multipole moment integrals.

- `energy_xdm` computes the XDM dispersion energy, as well as the contributions to the atomic forces and stresses.

There exist other subroutines that were already present within FHI-aims that are called within the XDM module to aid in the evaluation of the density-based properties, the MPI parallelism, and writing output that are not discussed in this work for brevity. Additional modifications were also made to existing routines within the code, which will be described in Section 5.1.3.

If the XDM dispersion method is selected, the program will enter the `xdm.f90` module once the self-consistent field (SCF) calculation is converged. Within the xdm module, the parent routine is referred to as `xdm_collect`, which returns the xdm energy, atomic forces, and unit-cell stresses. Once the dispersion coefficients are first computed within `xdm_collect`, they are held fixed for all subsequent energy evaluations within a geometry optimisation. However, after the optimisation has converged, they are recalculated to assess whether this results in a significant energy change that would require resubmission of the optimisation. Therefore the XDM dispersion coefficients are only calculated at the first and last instances of calling `xdm_collect`. To accomplish this, the routine is passed a boolean variable, `recalc_c6`, which indicates whether the XDM dispersion coefficients should be recalculated or if values from a previous energy evaluation should be used.

If the a_1 and a_2 damping parameters are not provided by the user in the FHI-aims input (see Section 5.1.4), then given default values will be used depending on the particular XC functional. Within the `xdm_collect` routine, the default damping parameters of a_1 and a_2 are specified for two GGAs, B86bPBE and PBE, as well as five hybrid functionals, HSE06, PBE0, PBE50 (50% exact exchange), B86bPBE-25X (25% exact exchange), and B86bPBE-50X (50% exact exchange).

Discussion of these functionals and the parameters themselves are included in Chapter 6. The program will terminate with an error message if no damping parameters are specified and the functional selected is not one of those listed here.

At the first instance of running xdm_collect, the arrays for the C_6, C_8, and C_{10} dispersion coefficients and the sums of van der Waals radii (R_{vdW}) are allocated, and consist of matrices of values with entries for all atom pairs. Arrays are also allocated for the free atomic volumes and polarizabilities, which are populated through calling the xdm_hirshfeld_free subroutine. Within this subroutine, the free atomic polarizabilities from the CRC handbook of chemistry and physics[294] are tabulated and the free atomic volumes are computed by integration of the reference free atomic densities contained within the free_atoms module. Similarly, the atomic volume and moment integral arrays are allocated within the xdm_collect routine and populated by the run_xdm_hirshfeld routine, which requires the Becke-Roussel[95] exchange-hole dipole moment, b, at each integration point as input. Next, the xdm_collect routine calculates the atomic polarizabilities by scaling the free-atom polarizabilties using a ratio of environment-dependent and free-atom volumes (Eq 2.54). The dispersion coefficients are then computed from the polarizabilities and moment integrals using Eq 2.51-2.53, via a nested loop over all atom pairs. Finally, the XDM energy, forces, and stresses (if necessary) are computed by calling the energy_xdm subroutine.

The run_xdm_hirshfeld subroutine requires the values of the BR exchange-hole dipole moment, b, on a grid to calculate the moment integrals of Eq. 2.55, which are passed back to xdm_collect and used to obtain the dispersion coefficients. In order to calculate all of the quantities needed to obtain the b values, a number of other subroutines are required, which are also called by xdm_collect. First, the existing evaluate_density_direct_from_densmat routine is called by xdm_collect to obtain the electron density (ρ), gradient ($\nabla\rho$), Laplacian ($\nabla^2\rho$), and kinetic-energy density (τ) on the grid. These quantities are then combined to obtain the curvature of Eqn. 2.60. The density and curvature are then passed to the bhole subroutine to obtain the BR b values. For the calculation of the exchange-hole dipole within bhole, the method is copied directly from the postg code,[75,295] which was one of the first implementations of XDM within a molecular electronic-structure code. The subroutine uses the Newton-Raphson method[95,136] to find the root, x, of a real-valued function given by Eqn. 2.63, from which the b value can be obtained by Eqn. 2.64. The final b array is passed back to xdm_collect before being used within the run_xdm_hirshfeld subroutine for the ultimate calculation of the XDM moment integrals.

Finally, in the energy_xdm subroutine, the dispersion energy, forces, and stresses are computed from the dispersion coefficients. As FHI-aims can perform both molecular and solid-state

61
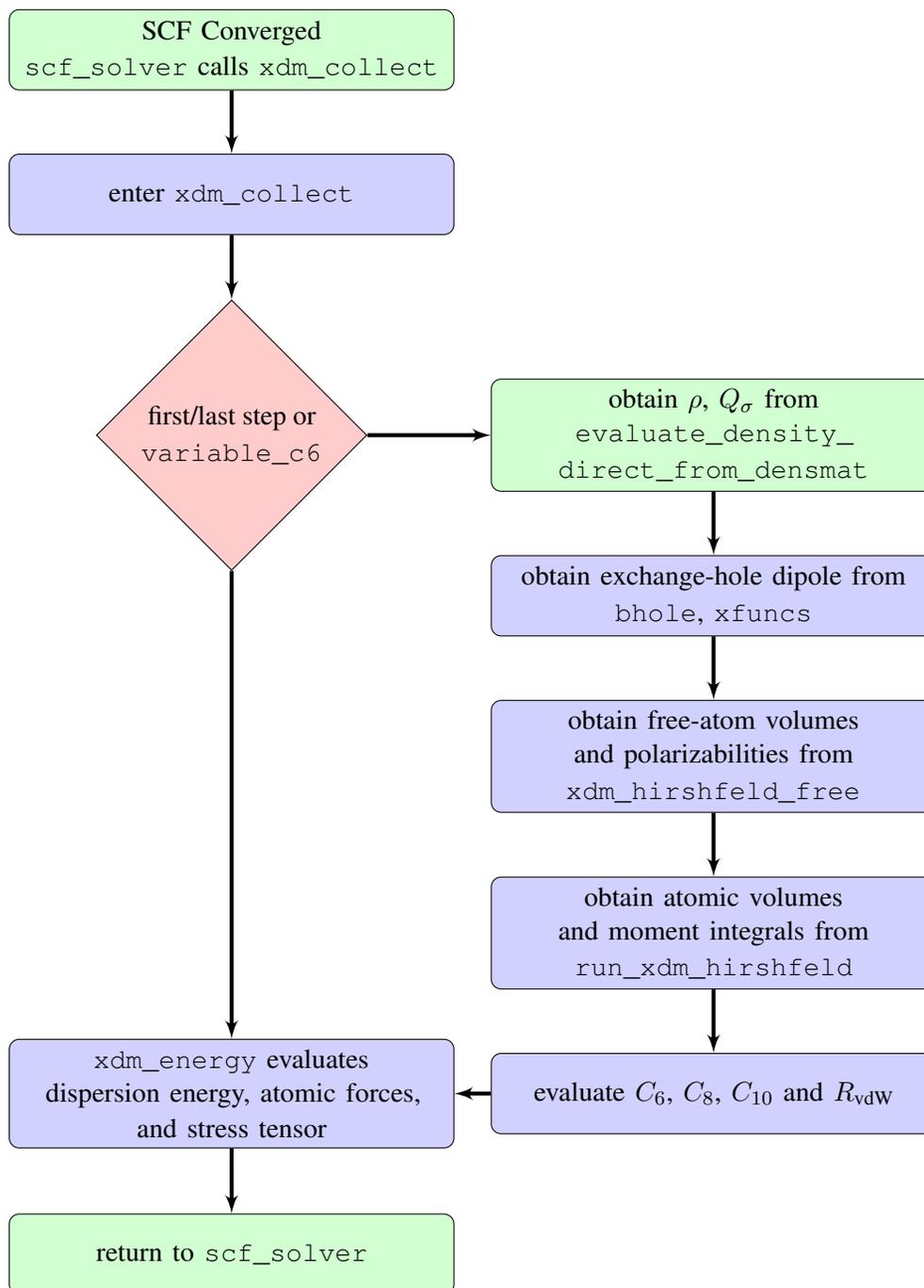
Figure 5.1: Flow chart to illustrate the computational procedure for XDM within FHI-aims. Green boxes indicate existing routines within the code, while blue boxes indicate new routines added by the author.

calculations, there is a fork within this routine. For finite-molecular calculations the dispersion energy and atomic forces are simply computed via a nested do loop over all atom pairs. The

calculation of the XDM energy for a solid is detailed by Eq 2.50, with the energy calculation also requiring summation over the lattice vectors to include the dispersion energy between atoms in the reference unit cell and its periodic replicates. The atomic forces and the stress tensor are calculated within the same nested loop as for the energy, using Eq. 2.67 and Eq. 2.68, respectively, which assume that the dispersion coefficients do not depend on the atomic positions.[131] For solid-state calculations, much of the same machinery is utilised as for the finite-molecule calculations, except another layer of convergence is required – that of over periodic cells. The sum over atomic pairs is therefore nested in a further loop over supercell size, where the change in dispersion energy (resulting from including successively many unit-cell replicates at increasing distances from the reference cell) is checked against a predetermined energy convergence threshold and, once it falls below this value, no further image cells are considered in the summation. This procedure and the choice of convergence threshold was present already within FHI-aims for use with the TS[35] dispersion method. Similarly, if a geometry optimisation is undertaken for a periodic solid, the convergence of the forces will be checked using another predetermined threshold value within the code. We note, however, that the convergence of the stress is not tested with either of the TS or XDM methods. Once the summation over supercells has converged, the final dispersion energy, forces, and stresses are passed back to the `xdm_collect` subroutine.

5.1.2 CHOICES MADE WITHIN THE IMPLEMENTATION

A number of choices were made during construction of the XDM module in order to achieve good performance, accuracy, and ease of use for those who wish to employ the XDM dispersion correction within FHI-aims. The first was that, after the first calculation of the dispersion coefficients at the outset of the calculation, the coefficients are held fixed until the last step of a geometry optimisation, when they are recalculated. The default behaviour of the code when the `xdm` keyword is used is to do exactly this and hold the dispersion coefficients fixed. This is done because continually recalculating the XDM dispersion coefficients increases the computational cost and, more importantly, causes poor convergence of geometry optimisations due to the minor force-energy inconsistency that results from the assumption that the dispersion coefficients are independent of atomic position when evaluating the atomic forces and stress tensor (Eqns 2.67 and 2.68).[131] All calculations in the proceeding chapters were achieved through the use of constant dispersion coefficients during optimisation, followed by resubmission from the converged geometry.

Another choice was to hard-code the damping parameters for a selection of common functionals employed both in the literature and in our own work. As detailed in Subsection 5.1.4, if only the `xdm` keyword is used, the code will read the selected exchange-correlation functional from `control.in`, and use the corresponding a_1 and a_2 parameters for the light-dense basis sets for

that functional. Having predetermined a_1 and a_2 parameters within the XDM module makes it easier for others to use XDM within FHI-aims.

In periodic-boundary implementations, a choice must be made to define the spatial extent for summation over the periodic unit cells surrounding the initial reference cell when evaluating the dispersion energy, as well as the forces and stress tensor. In FHI-aims, a cutoff was determined by evaluating the change in the total dispersion energy with respect to adding each successive shell of image cells surrounding the reference cell. When this energy change falls below a predetermined threshold, the evaluation of the dispersion energy is considered converged. This choice was made for consistency with the TS dispersion method already implemented within FHI-aims, where a dispersion energy convergence threshold, in eV, was set to

$$\delta = \max(n_{\text{atoms}} \times 10^{-8}, 10^{-6}), \tag{5.1}$$

where n_{atoms} is the number of atoms within the unit cell. This differs from the previous XDM implementation in Quantum ESPRESSO,[109] where the number of cells that contribute to the dispersion energy is summed up to a set radius (determined based on the C_6 value) from a given atom within the initial reference cell. It will be seen later in this work that there are minimal differences in accuracy achieved between both implementations of XDM at the GGA level for molecular crystals.

5.1.3 DESCRIPTION OF OTHER CODE MODIFICATIONS

FHI-aims contains two mandatory input file types for a general calculation: `control.in` and `geometry.in`. The `geometry.in` file contains the atomic positions and, additionally, the lattice vectors if a periodic calculation is performed. The XDM implementation did not modify any parts of the code involving the reading or writing of `geometry.in`, which will not be discussed further. The other mandatory input file is the `control.in` and this was where modifications were made within the code to allow the running of XDM. The `control.in` file contains the bulk of the runtime-specific information, such as type of calculation, choice of exchange-correlation functional (for DFT calculations), total-spin and charge, the type of dispersion correction and its flags, as well as the species subtags which detail the NAO basis set for each of the elements found in `geometry.in`. For the selection of the dispersion correction within `control.in`, the details specific to XDM will be described in subsection 5.1.4.

In order to run an XDM calculation, a new `xdm` keyword must be defined that can be invoked within the `control.in` file. To define this keyword, several key files were modified within the code: `read_control.f90`, `runtime_choices.f90`, and `dimensions.f90`. The

`control.in` file is read within `read_control.f90`, which sets the flag of `xdm` to true within the code if the keyword is present. This routine will also read any additional input options, or set these options to their default values, depending on what, if anything, is present in the input line after invoking the `xdm` keyword. Variables for these additional input options were also defined in `runtime_choices.f90`. Here two real variables were defined for the a_1 and a_2 damping parameters, as well as two boolean variables to specify whether the default damping parameters (preprogrammed and specific to each functional) will be used and whether the XDM calculation will use variable or fixed dispersion coefficients. Any keyword defined within `read_control.f90` must also be defined within `dimension.f90`, which is the module where all directly accessible array dimensions are allocated through parsing the input files. Here, if the `xdm` keyword has been invoked within `control.in`, a boolean varible flag to specify whether an XDM calculation should be performed is set to true, starting the cascade of choices within the rest of the code to run the `xdm.f90` module and calculate the dispersion energy.

The last parts of the code that were modified for the inclusion of XDM were `scf_solver.f90`, `get_total_energy.f90`, `get_total_forces.f90`, `analytical_stress.f90`, and `main.f90`. Each of these modifications can be broken up into calling the XDM module, adding the dispersion contributions to the total energy, forces, and stresses, writing the results to the output file, and determining the stage of the overall calculation. It is within the routine `scf_solver.f90` that the XDM module is called with `call xdm_collect`. Upon its first invocation, the dispersion coefficients, energy, atomic forces, stress tensor, are calculated. At the end of each subsequent SCF, the energy, forces, and stresses are recalculated but, the dispersion coefficients are held fixed, unless `xdm_variable_c6` is set to true. The `scf_solver` routine also calls `get_total_energy.f90`, `get_total_forces.f90`, and `analytical_stress.f90`, which sum up the SCF and dispersion contributions to the total energy, atomic forces, and stress tensor, respectively. These routines were modified to include the dispersion contribution if the `xdm` keyword is invoked within the `control.in` file. Finally, `main.f90` is a high-level routine within FHI-aims from which all main tasks branch. This routine calls the SCF solver and it is also in `main.f90` where the convergence of a geometry optimisation is assessed. Here, if the optimisation is considered converged, a boolean variable `xdm_last_pass` is set to true and there is a final call to the XDM module such that the dispersion coefficients are recalculated to update the dispersion energy.

5.1.4 INPUT OPTIONS

A new section was added to the FHI-aims manual to provide instructions for using the `xdm` keyword, as well as examples. The implementation of XDM within FHI-aims is able to calculate

the dispersion energy for both molecules and solids and is achieved as a post-SCF correction to the base DFT energy as in Eq. 2.43. With the dispersion energy calculated as a damped asymptotic expression as in Eq. 2.65. XDM can be run within the code by using the tag of `xdm` within the `control.in` file.

Usage: `xdm a1 a2 variable_c6`

Purpose: activates the calculation of the XDM dispersion energy and its derivatives.

Options:

- `a1 a2`: XDM dispersion coefficients (a_2 in units of Å).
  Default: the default damping coefficients (a_1 and a_2) are chosen based on the functional used (see the `xc` keyword), and correspond to a "light" basis set with a dense grid. The dispersion coefficients for the following functionals are available: B86bPBE, B86bPBE-25, B86bPBE-50, PBE, PBE0, PBE50, HSE06. If your functional is not on this list, or if you want to use parameters optimised for the "tight" basis set, you need to supply the XDM parameters.

- `variable_c6`: in a geometry relaxation, recompute the dispersion coefficients at every step.
  Default: the dispersion coefficients are calculated in the first and last step of a relaxation.

Examples:

- `xdm` (calculate XDM with default settings)

- `xdm 0.4572 0.5921` (use $a_1 = 0.4572$ and $a_2 = 0.5921$Å)

- `xdm variable_c6` (recalculate XDM coefficients at every relaxation step)

5.1.5 B86B-BASED EXCHANGE FUNCTIONALS

The functional that has been shown to pair best with the XDM dispersion correction is B86bPBE.[8,9] The PBE correlation[9] functional was already present within the FHI-aims code base and is used as is. The B86b-based functionals (including B86bPBE and its hybrid counterparts) were newly implemented within FHI-aims for this thesis. PBE exchange and B86b exchange differ only within the description of their enhancement factors (Fig 3.2), which can lead to subtle differences in their treatment of non-bonded repulsion; see Fig 3.1 for graphite exfoliation energies, for example. The enhancement factor for PBE is

$$F(\chi_\sigma) = 1 + \frac{\beta}{c_x} \frac{\chi_\sigma^2}{1 + \gamma\chi_\sigma}, \tag{5.2}$$

where $\chi_\sigma = \frac{|\nabla\rho|}{\rho^{4/3}}$ is the reduced density gradient, and β and γ are parameters. As stated within Chapter 3, the B86b enhancement factor follows the correct behaviour in the large-gradient limit of $\chi_\sigma^{4/3}$, and is written as

$$F(\chi_\sigma) = 1 + \frac{\beta}{c_x}\frac{\chi_\sigma^2}{(1 + \gamma\chi_\sigma)^{4/5}} \tag{5.3}$$

for two different β and γ parameters.

For the implementation of B86b exchange within FHI-aims, the PBE exchange subroutine, `pbe_partial_derivs.f90`, was utilised as a template. Only the enhancement factor within the subroutine was modified to create the new `b86b_partial_derivs.f90` exchange subroutine. In order for the B86bPBE exchange to be recognised as a possible call within `control.in`, two other components of the code were modified: `xc.f90`, and `xc_library.f90`. The `xc.f90` module provides the functions that calculate the exchange-correlation energy, where again the subroutine for PBE was used as a template. Both subroutines are passed the electron density (ρ) and the gradient ($\nabla\rho$), and pass back the exchange-correlation energy. Finally, within `xc_library.f90`, additions were made for the handling of input cases within `control.in` for B86bPBE, as well as the B86bPBE-25X and B86bPBE-50X hybrid functionals. It is within `xc_library.f90` that the string input within `control.in` is parsed and translated to an index or flag used by FHI-aims. It is also within this module that the percentage of exact exchange mixing is set for the B86bPBE-25X (25%) and B86bPBE-50X (50%) hybrid density functionals, depending on the flag set by the parsed string.

5.2 TESTS

A selection of tests was undertaken in order to verify the XDM implementation within FHI-aims. The central quantity of the exchange-hole dipole, b, must be calculated using Eq. 2.64, which relies on the spin density ρ_σ, the Weizsäcker term $\frac{|\nabla\rho_\sigma|^2}{\rho_\sigma}$, the Laplacian of the spin density, $\nabla^2\rho_\sigma$, and the spin kinetic density, τ_σ. The results from the calculation of these quantities at every point on the integration grid can be seen in Fig. 5.2 for an argon dimer. Here, Quantum ESPRESSO[109] uses the PAW method and an effectively complete planewave basis set, Numol[199] is all-electron fully numerical code, and FHI-aims uses a "tight" NAO basis set. The FHI-aims result for the exchange-hole dipole closely follows that of the Numol code within the core regions due to their all-electron nature, and differs from that of the Quantum ESPRESSO implementation due to the pseudized nature of the core region. For each of the quantities listed above it can be seen that FHI-aims accurately reproduces the Numol implementation. The only notable difference is that FHI-aims slightly underestimates the exchange-hole dipole at grid points far from the atomic nuclei.
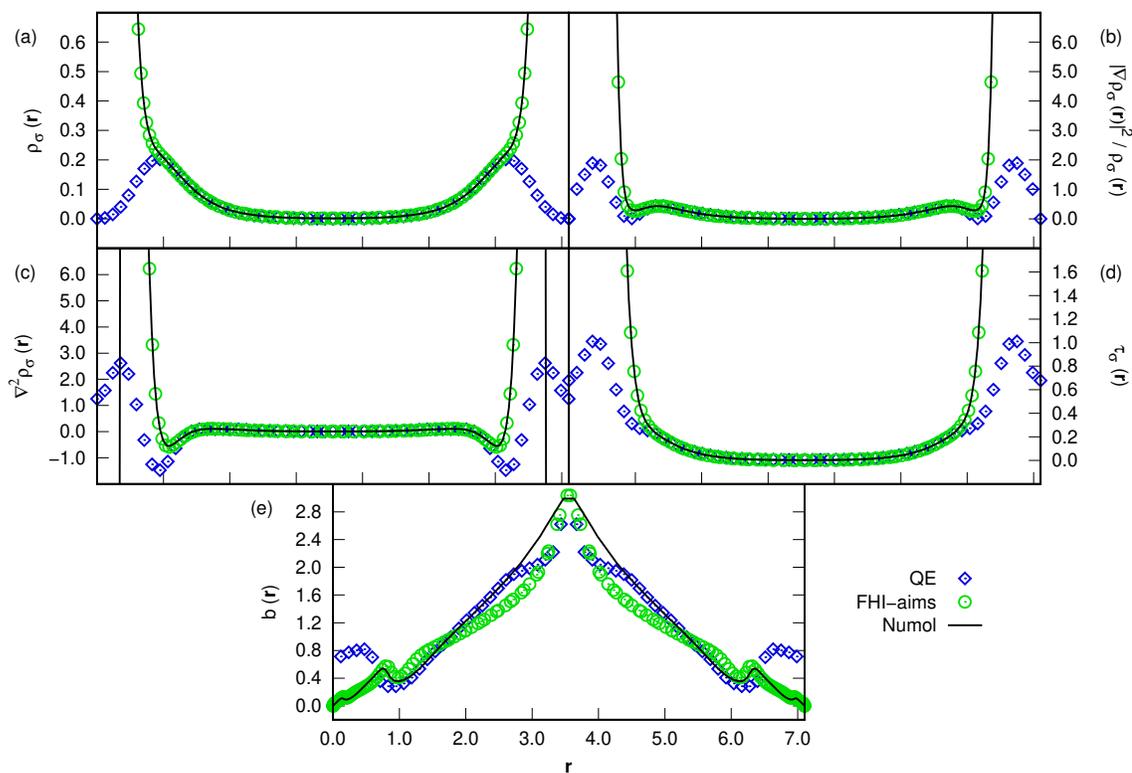
Figure 5.2: Comparison of the exchange-hole dipole and its component quantities, obtained with selected codes, for the argon dimer along the internuclear coordinate. Shown are: (a) ρ_σ, the spin density, (b) $\frac{|\nabla\rho_\sigma|^2}{\rho_\sigma}$, the Weizsäcker term, (c) $\nabla^2\rho_\sigma$, the Laplacian of the spin density, (d) τ_σ the spin-dependent kinetic density, and (e) $b = d_{\chi\sigma}$, the exchange-hole dipole moment. The Numol[199] calculations used the LDA[91] and an isolated argon dimer with a 3.70 Å internuclear distance, the Quantum ESPRESSO[109] calculations used the PBE functional[9] and an argon atom in the centre of a cubic box with lattice vectors of 3.70 Å, and the FHI-aims calculations used the PBE functional,[9] the "tight" basis set, and an isolated argon dimer. All values are presented in atomic units.

This is likely due to not fully capturing the long-rage exponential decay of the electron density, although it may also be partially due to differences in the long-range behaviour of PBE and the LDA.

Further tests were run in order to compare the C_6 dispersion coefficients calculated using FHI-aims with those from Gaussian[296] (and the postg implementation[75] of XDM) for the isolated molecules, and with Quantum ESPRESSO[109] for the molecular crystals, present within the X23 set.[80] The comparison for the homoatomic C_6 dispersion coefficients can be seen in Fig. 5.3, with each point representing an atom found within each of the respective molecules or unit cells. For the FHI-aims vs. Gaussian comparison, the plot shows a slight skew from linearity, likely due to the truncation of the NAO basis set tails when compared to the aug-cc-pVTZ Gaussian
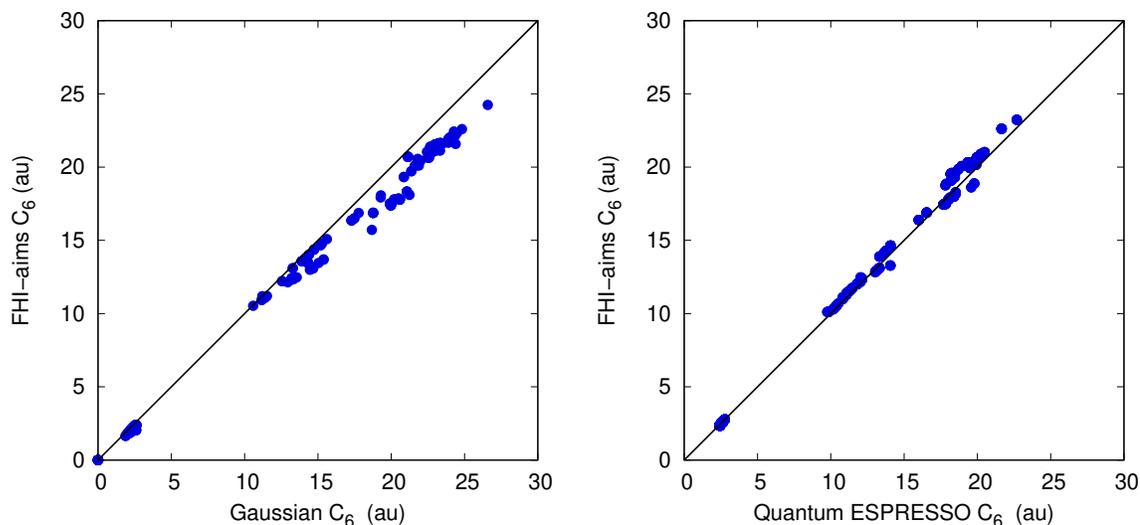
Figure 5.3: Comparison of homoatomic C_6 dispersion coefficients, in atomic units, computed using FHI-aims and either Gaussian[296] (isolated molecules, left) or Quantum ESPRESSO[109] (molecular crystals, right) for the components of the X23 lattice-energy benchmark.[80] All calculations were undertaken with the PBE functional.[9] The FHI-aims calculations used the light dense basis set described later in this work, the Gaussian calculations used the aug-cc-pVTZ basis set and the ultrafine integration grids, and the Quantum ESPRESSO calculations used kinetic energy cutoffs for the wavefunction and charge density of 80 and 800 Ry, respectively, and employed PAW data sets. The solid black line ($y = x$) represents a slope of one, indicating identical values.

basis set. Although the density is low far away from an atom in an isolated molecule, there are large contributions to the exchange-hole dipole moments. This becomes most apparent for points corresponding to carbon atoms, seen at the top right of the plot. Carbon atoms have the largest dispersion coefficients of the elements appearing within the X23 set (H, C, N, O), while hydrogen atoms have the smallest. For the FHI-aims vs. Quantum ESPRESSO comparison using molecular crystals, there are negligible differences between the codes due to two factors. First, while the core region is not fully described within QE, it does contribute greatly to the exchange-hole dipole for the light elements considered here due to its small spatial extent; this can be contrasted with transition-metal atoms, where the core contributions are larger.[297] Second, although there is a truncation of the NAOs in FHI-aims, periodic solids lie within a regime where the overlap of the valence densities dominates, which is where the largest contribution to the dispersion coefficients exists. Overall, the FHI-aims results were found to be close to those from previous XDM implementations when comparing C_6 coefficients for both molecular and solid-state systems. This provided us with confidence to apply our new implementation to energy-based benchmarks in the next chapter.

CHAPTER 6

XDM-CORRECTED HYBRID DFT WITH NUMERICAL ATOMIC ORBITALS FOR MOLECULAR CRYSTAL ENERGETICS

This chapter is adapted from: A. J. A. Price, A. Otero de la Roza, and E. R. Johnson, "XDM-corrected hybrid DFT with numerical atomic orbitals predicts molecular crystal energetics with unprecedented accuracy" *Chem. Sci.* **14**, 1252-1262 (2023).

6.1 INTRODUCTION

The accurate description of molecular crystals is a challenge for current computational methods. Molecular crystal structures typically have unit cells containing hundreds of atoms, meaning a high computational expense, and feature a delicate balance between weak non-covalent (intermolecular) and strong covalent (intramolecular) interactions, both of which have to be described accurately by the chosen method. The computational description of these systems is important in the study of polymorphism, which is particularly prevalent in molecular crystals,[298,299] pressure-temperature phase diagrams,[299] and for any discipline in which the solid form of a molecular material controls a property of interest: pharmaceuticals (solubility/bioavailability and patentability[300–303]), foodstuffs (organoleptic properties[304]), energetic materials (sensitivity to detonation[305–307]), organic semiconductors (charge carrier mobility[308–312]), and others.[309]

Having a method that is able to rank molecular crystal structures accurately is essential for crystal structure prediction (CSP) – the prediction of the crystal structure of a compound from its molecular diagram only.[303,313,314] A reliable CSP protocol would be extremely useful in the disciplines listed above, as it would allow circumventing experimental solid-form screening processes, which are expensive and time-consuming,[315–319] and would provide a detailed energy-structure-function map for the chosen molecule and property of interest.[308,309] To gauge progress in the field, the Cambridge Crystallographic Data Centre (CCDC) periodically runs CSP blind test competitions in which participant groups try to predict the observed crystal structures of a few molecular compounds.[82–87] The 5th blind test, held in 2011, showed that final ranking of the candidate structures using dispersion-corrected DFT is quite effective, and far superior to force fields in most cases.[86,320–323] Although other techniques such as fragment-based methods,[299,324–328] wavefunction theory,[329,330] and machine-learning methods[328,331] have been used, DFT is arguably the current workhorse for modeling molecular materials.[314,332–345]

Dispersion-corrected functionals based on the exchange-hole dipole moment (XDM) model,[10,63,131,346] in particular the semilocal functional B86bPBE-XDM,[8,9] have shown excellent performance for description of molecular crystals[131,335,336,347] and non-covalent interactions in general.[16,348] In its current plane-wave/pseudopotentials implementation, while still effective for CSP, B86bPBE-XDM is affected by outstanding drawbacks shared by all semilocal functionals. First, the use of a plane-wave basis set makes the computational requirements scale significantly with system size, such that calculations involving unit cells with hundreds to thousands of atoms are on the verge of being infeasible. Second, GGA functionals spuriously over-stabilize systems affected by delocalization error,[204,231,349] which negatively impacts the modeling of molecular salts, acid-base co-crystals, hydrogen bonding, and halogen bonding, to list only a few examples.[245,247,250,335] Lastly, GGA functionals give a poor description of conformational energies, which are important when comparing crystal polymorphs composed of flexible molecules.[336,350–352]

In this work, we address these shortcomings by combining XDM functionals with the numerical atomic orbital (NAO) basis sets in the Fritz Haber Institute *ab initio* materials simulations (FHI-aims) package.[64,65,68,353] Being finite-support functions, NAOs have the dual advantages of providing linear scaling with system size and enabling relatively inexpensive use of hybrid functionals, compared to plane-wave approaches. This is important because hybrid functionals can be used to mitigate delocalization error[75,245,349,354–357]

and are generally more accurate than GGAs for conformational energies.[352] One draw-back of NAOs is the possible appearance of basis-set incompleteness error (BSIE), which is known to have a deleterious effect on the description of non-covalent interactions,[358–360] although we show that BSIE can be effectively mitigated by parametrization of the dispersion damping function.

Dispersion-corrected DFT methods with NAOs have been applied to molecular crystals in combination with the Tkachenko-Scheffler (TS)[35] and many-body dispersion (MBD)[36,37] family of corrections.[340–344] To assess the new XDM-corrected hybrid functionals, we focus on molecular crystal lattice energies as they are the key property for CSP ranking[361] and one of the most demanding tests for computational methods regarding non-covalent inter-actions.[299] In particular, we consider the lattice energies of the X23 set[80,131,326,329,330,362] and of 13 ice phases, for which diffusion Monte Carlo (DMC) data has been generated.[81] The latter is a particularly stringent test because determining accurate lattice energies for ice relies on a fine balance of dispersion, electrostatic, and many-body induction effects. At present, there is no functional that gives a good description of the absolute and relative energies of all ice phases,[81] and therefore the reliable treatment of water and ice with DFT methods remains an unsolved problem.[81,363,364]

Herein, we show that the NAO implementation of XDM-corrected functionals provides excellent performance for the description of molecular dimers, ice, and molecular crys-tal lattice energies in general, with high computational efficiency. In particular, a com-posite method combining B86bPBE-XDM and its sequent 25% hybrid functional achieves mean absolute errors (MAEs) for the X23 and ice lattice energies of 0.48 kcal/mol and 0.19 kcal/mol, respectively. For the X23, the reported MAE is roughly half the previous best value, making the new XDM methods the most accurate DFT approaches for modeling of molecular materials currently available.

6.2 METHODS

6.2.1 THEORY

A summary of the XDM dispersion model and its implementation in the FHIaims package is presented in this section. More details about the XDM method can be found in previous works (see Ref. 10 and references therein). In XDM, the dispersion energy is calculated

using a damped asymptotic pairwise dispersion expression,

$$E_{\text{XDM}} = -\sum_{n=6,8,10} \sum_{i>j} \frac{C_{n,ij}}{R_{ij}^n + R_{\text{vdW},ij}^n}, \tag{6.1}$$

which is then added to the energy from the base density functional,

$$E = E_{\text{base}} + E_{\text{XDM}}. \tag{6.2}$$

In Eq. 6.1, i and j run over atoms, R_{ij} are the interatomic distances, $C_{n,ij}$ are the dispersion coefficients, and the $R_{\text{vdW},ij}$ are damping lengths calculated as

$$R_{\text{vdW},ij} = a_1 R_{c,ij} + a_2, \tag{6.3}$$

with

$$R_{c,ij} = \frac{1}{3}\left[\left(\frac{C_{8,ij}}{C_{6,ij}}\right)^{\frac{1}{2}} + \left(\frac{C_{10,ij}}{C_{8,ij}}\right)^{\frac{1}{2}} + \left(\frac{C_{10,ij}}{C_{6,ij}}\right)^{\frac{1}{4}}\right]. \tag{6.4}$$

The a_1 and a_2 parameters are the damping function coefficients, which are determined for every functional and basis set combination by minimizing the root-mean-square percent error in binding energies for 49 small molecular dimers, relative to high-level reference data (the Kannemann-Becke set,[74,75,171] KB49). The damping function is therefore used to match the XDM dispersion contribution to the particular exchange-repulsion behavior of the chosen functional, as well as to mitigate any (moderate) BSIE from a less-than-complete basis set. Importantly, once the a_1 and a_2 parameters are determined, they remain the same for every system to which the functional and basis set are applied, molecular or periodic, and are never re-parametrized for specific cases.

The dispersion coefficients in Eq. 6.1 ($C_{n,ij}$) are calculated non-empirically from the self-consistent electron density, its derivatives, and the kinetic energy density. It has been shown that the dependence of these coefficients on the chemical environment (the electronic many-body dispersion effects) is essential to the accuracy of the XDM method.[21] Calculation of three-body and higher-order dispersion coefficients, of which the Axilrod-Teller-Muto (C_9) is the leading term, is possible in XDM,[152] but we have found that including this term has either little impact or degrades the accuracy of XDM-corrected functionals.[21]

The performance of a XDM-corrected method depends critically on the base functional with which it is paired. In this article, we consider two generalized-gradient-approximation

(GGA) functionals: PBE,[9] due to its popularity in the solid-state community, and B86bPBE,[8,9] which is our GGA of choice when non-covalent interactions are dominant, thanks to its ability to accurately describe non-bonded repulsion.[93,94,131,171,348] In addition, we consider multiple hybrid density functionals with exchange-correlation (XC) energies of the form

$$E_{\text{XC}} = (1 - a_{\text{X}})E_{\text{X}}^{\text{PBE/B86b}} + a_{\text{X}}E_{\text{X}}^{\text{HF}} + E_{\text{C}}^{\text{PBE}}. \tag{6.5}$$

The exchange GGA is either PBE or B86b, a_X controls the fraction of exact (Hartree-Fock, HF) exchange used in the functional, and the correlation contribution comes from PBE. The PBE0 functional[97] corresponds to the choice of PBE exchange and $a_{\text{X}} = 0.25$. Functionals with 50% exact exchange ("half-and-half") have been shown to minimize delocalization error,[245,250,365] so we also considered "PBE-50" with PBE exchange and $a_{\text{X}} = 0.5$. Given the good behavior of B86bPBE for intermolecular closed-shell repulsion, we define 25% and 50% hybrids built on B86b exchange as well, termed B86bPBE-25 and B86bPBE-50, respectively.[245] Finally, we included the range-separated HSE06 hybrid functional[366] as its use is fairly common in solid-state applications.

6.2.2 Computational Details

All calculations in this work were carried out with the FHI-aims program (version 210513). The XDM method, B86b exchange, and the ensuing hybrid functionals, were all implemented in a copy of the code. The basis sets used for the calculations were either the "light" or the "tight" settings, which correspond to double-ζ and triple-ζ basis sets, respectively. Based on our initial exploration, the choice of integration mesh can substantially affect the stability of the geometry relaxation procedure for molecular crystals. We therefore chose to always use the integration meshes from the tight settings, with up to 434 angular grid points.

The memory requirements of hybrid functional calculations with the tight basis set exceeded our current computational resources, so we approximated the hybrid/tight result using a correction calculated by evaluating the energy difference between tight and light bases at the GGA level:

$$E_{\text{hybrid}}^{\text{tight}} \approx E_{\text{hybrid}}^{\text{light}} + \left(E_{\text{GGA}}^{\text{tight}} - E_{\text{GGA}}^{\text{light}} \right). \tag{6.6}$$

This type of basis-set correction is analogous to using the difference between large- and small-basis MP2 energies to correct small-basis CCSD(T) energies, as in common practice

in wavefunction theory calculations.[367] In addition to the XDM-corrected functionals mentioned above, we also considered the Tkachenko-Scheffler (TS)[35] and many-body dispersion (MBD)[36,37] methods for comparison, since they are already implemented in FHIaims and are routinely used for molecular crystals and CSP.[340–344] In the case of MBD, we used MBD@rsSCS[37] (where rsSCS indicates range-separated self-consistent screening of the polarizabilities used as input to MBD) as recommended by the FHI-aims documentation. In the rest of the article, MBD@rsSCS is referred to simply as MBD.

All molecular calculations were carried out as single-point energy evaluations at the literature geometries. The molecular datasets considered are the KB49,[74,75] S22×5,[76] and S66×8[77,368] sets of gas-phase dimer binding energies, and the 3B-69[78] set of three-body interaction energies in molecular trimers. All solid-state calculations were carried out by performing a full geometry relaxation with each functional. Reciprocal-space $\mathbf{k}$-point grids were selected with the number of points, $n_1 \times n_2 \times n_3$, given by

$$n_i = \mathrm{int}\left[\max\left(1, R_k|b_i| + 0.5\right)\right], \tag{6.7}$$

where $|b_i|$ is the length of the ith reciprocal lattice vector and $R_k = 50$ bohr. The solid-state benchmarks are the lattice energies of the X23[79,80,362] set of molecular crystals (with the most recent reference values of Dolgonos et al.[362]) and the ICE13 set of ice polymorphs.[81]

6.3 XDM PARAMETRIZATION

Before using the new FHIaims XDM implementation, we first need to parametrize the XDM damping function (Eq. 6.3) and find the optimal a_1 and a_2 for all chosen functional and basis set combinations. This is done in the same way as in previous studies, by minimizing the root-mean-square percent (RMSP) error in the binding energies of the 49 molecular dimers comprising the Kannemann-Becke set.[74,75] The optimal parameter values, along with the resulting KB49 error statistics, are collected in Table 6.1. It is important to note that these a_1 and a_2 values are fixed for each particular functional and basis set combination, and do not change with the system to which XDM is applied.

The errors shown in Table 6.1 are comparable to those obtained with our previous plane-wave (Quantum ESPRESSO[131,369]) and Gaussian basis-set (using Gaussian[296] or psi4[370] with the postg program[295]) results contained in the current XDM parametrization database.[371] For example, the MAPE for B86bPBE/tight in Table 6.1 (11.0%) is very close to the MAPE obtained for the same functional using the projector augmented wave

Table 6.1: Optimal XDM parameters (a_1 and a_2) for selected functionals, with exact-exchange mixing fractions (a_X) indicated. The mean absolute errors (MAE, in kcal/mol) and mean absolute percent errors (MAPE) for the KB49 fit set are also shown. The best overall results for each basis set are indicated in bold.

Functional	a_X	a_1	a_2 (Å)	MAE	MAPE
Light basis set					
PBE	0.00	0.5312	2.3270	0.67	19.0
B86bPBE	0.00	0.8219	1.2069	0.54	14.9
HSE06	0.11[a]	0.3268	3.0431	0.52	13.6
PBE0	0.25	0.3302	3.0042	0.46	12.5
PBE-50[b]	0.50	0.0000	4.1971	0.38	9.8
B86bPBE-25	0.25	0.5235	2.1995	0.35	9.7
B86bPBE-50	0.50	0.0831	3.7362	**0.30**	**8.5**
Tight basis set					
PBE	0.00	0.6438	1.8533	0.50	14.1
B86bPBE	0.00	0.8976	0.8518	0.38	11.0
HSE06	0.11[a]	0.5020	2.3000	0.46	11.1
PBE0	0.25	0.5053	2.2527	0.41	10.2
PBE-50	0.50	0.3983	2.5986	0.42	9.6
B86bPBE-25	0.25	0.6546	1.6097	**0.32**	**8.4**
B86bPBE-50	0.50	0.4887	2.1855	0.36	8.5

[a]This value is the range-separation parameter (ω) instead of the exact-exchange fraction.
[b]The optimal a_1 value was negative, so it was set to zero during the parametrization.

(PAW) method[104] (11.8%), plane waves plus norm-conserving pseudopotentials (12.4%), and the reasonably complete aug-cc-pVTZ Gaussian basis set (11.4%). The MAPEs obtained with other functionals, such as PBE, PBE0, or HSE06, also deviate from those in the parametrization database by around 1% at most. This is a strong indication that our FHIaims XDM implementation is working correctly.

Focusing on the results for the tight basis set, Table 6.1 shows that hybrid functionals outperform GGAs, and that B86b-based functionals consistently give lower errors than the analogous PBE-based functionals. This is also in agreement with our previous works.[75,131] The lowest errors among the functionals studied are obtained for the B86bPBE-25 hybrid, with a MAE of 0.32 kcal/mol and a MAPE of 8.4%.

Because the tight basis set is too expensive for routine geometry optimizations, we resort to using a less-than-complete basis set and relying on the XDM damping function

to partially alleviate any BSIE.[75,358] Table 6.1 shows the average errors for the light (double-ζ) basis set. While lower errors are obtained with the tight (triple-ζ) basis set, the good performance of the light basis set indicates a reasonably low impact on the accuracy caused by BSIE. This is in stark contrast to our previous results using the double-ζ basis set in the SIESTA NAO code,[122,372] where the MAPE was in the 20% to 30% range and could not mitigated by using counterpoise corrections. The small magnitude of the BSIE can be confirmed by comparing the dispersion-uncorrected binding energies calculated with the light and tight basis sets using the same functional. For example, the mean absolute difference between the light and tight binding energies obtained using B86bPBE (without XDM) is 0.32 kcal/mol, with individual errors not exceeding 0.75 kcal/mol.

6.4 MOLECULAR BENCHMARKS

In order to build a method that works reliably for molecular crystals, it is imperative to avoid error cancellation as much as possible. Therefore, it is interesting to examine whether individual interactions between monomer pairs are accurately represented. For this reason, we first evaluate the performance of the new implementation of XDM for selected molecular benchmarks comprising gas-phase dimers, and compare it to the TS[35] and MBD[36,37] dispersion corrections also implemented in FHI-aims. We consider the S22×5[76] and S66×8[77,368] benchmarks, which comprise non-covalent interaction energies of small molecular dimers at and around their equilibrium geometries. It is worth noting that the single damping parameter employed in MBD was fit to minimize the mean absolute relative error in the S66×8 binding energies,[37] although this parameter was only fit for use with the tight setting (termed the "tier 2" basis set in Ref. 37) and, unlike the XDM damping parameters, is not basis-set dependent. TS and MBD are paired only with the PBE, HSE06, and PBE0 functionals for which damping parameters are available.

The S22×5 and S66×8 error statistics for the various combinations of functional, basis set, and dispersion correction are shown in Table 6.2. As for the KB49 set, the average errors are lower for the tight basis set and hybrid functionals slightly outperform GGAs, regardless of the dispersion correction employed. The XDM values in the table are also similar to those reported for the same benchmarks using a nearly complete Gaussian basis set.[10] While all basis set, functional, and dispersion method combinations perform generally well, B86bPBE-25-XDM consistently gives the lowest errors by a small margin, with MAEs in the range of 0.2-0.4 kcal/mol.

Beyond-pairwise intermolecular interactions are also important in molecular crystals,

Table 6.2: Mean absolute errors (in kcal/mol) for the S22×5,[76] S66×8,[77] and 3B-69[78] molecular benchmarks using selected functionals and dispersion corrections. The best overall results in each column are indicated in bold.

Functional	Dispersion Correction	S22×5		S66×8		3B-69	
		Light	Tight	Light	Tight	Light	Tight
PBE	TS	0.57	0.39	0.60	0.38	0.078	0.080
HSE06	TS	0.63	0.45	0.64	0.38	0.046	0.042
PBE0	TS	0.58	0.42	0.59	0.33	0.044	0.039
PBE	MBD	0.55	0.44	0.44	0.28	0.113	0.113
HSE06	MBD	0.53	0.48	0.45	0.29	0.069	0.066
PBE0	MBD	0.50	0.46	0.40	0.26	0.060	0.055
PBE	XDM	0.58	0.44	0.45	0.29	0.101	0.099
B86bPBE	XDM	0.46	0.34	0.35	0.20	0.050	0.052
HSE06	XDM	0.52	0.45	0.41	0.28	0.054	0.055
PBE0	XDM	0.49	0.43	0.38	0.25	0.044	0.045
PBE-50	XDM	0.47	0.47	0.37	0.28	0.047	**0.030**
B86bPBE-25	XDM	**0.39**	**0.35**	**0.30**	**0.19**	**0.037**	0.040
B86bPBE-50	XDM	0.40	0.41	0.32	0.24	0.055	0.051

since they represent a small but significant fraction of the total lattice energy.[299,373] For this reason, we consider as an additional benchmark the 3B-69 set of molecular trimers.[78] In this case, the reference data corresponds to the difference between the trimer binding energy and the pairwise sum of the constituent dimer binding energies. This is a good measure of whether the considered methods can describe non-additive many-body inter-molecular interactions[21] and, as such, highlight their performance in the treatment of beyond-pairwise effects. Table 6.2 shows that BSIE has less impact on the three-body energies than it does for the pairwise binding energies, with the light and tight MAEs being approximately the same.

MBD might be expected to be the most accurate dispersion correction for the 3B-69 benchmark due to the many-body nature of the interactions. However, we observe that all three dispersion methods provide roughly comparable performance, with XDM being slightly superior to MBD for the same functional and basis set combination. Instead, it is the choice of base functional that is the determining factor, with PBE consistently giving the largest errors, while use of either B86b or HF exchange improves performance in the treatment of three-body interactions. This confirms our previous observation that the choice of base functional is critical for accurate treatment of beyond-pairwise non-covalent interactions.[374] Overall, XDM paired with either the B86bPBE-25 or PBE-50 hybrid functionals (depending on basis set) gives the lowest MAE. The fact that XDM (which does not

incorporate a three-body dispersion contribution) outperforms MBD (which does) for the description of three-body intermolecular interactions suggests that electronic many-body effects are much more important than the atomic many-body dispersion effects encapsulated by the Axilrod-Teller-Muto term.[21]

6.5 X23 LATTICE ENERGIES

The lattice energy of a molecular crystal is the energy required to separate the crystal at its equilibrium geometry into its component molecules. Lattice energies are key quantities in CSP,[299,361] and they are an essential parameter when assessing the accuracy of computational methods for modeling molecular crystals because their value is determined by a delicate balance between intermolecular and intramolecular interactions.[299] The accurate calculation of lattice energies is also a stricter performance test for computational methods than energy differences between molecular crystal pairs because the benefits from error cancellation are minimized, while longer-range interactions and many-body effects become far more important.

Reference lattice energies for molecular crystals are typically derived from experimental sublimation enthalpies[375] using a back-correction for vibrational effects.[80,131,326,329,330,362] The X23,[79,80] which comprises 23 reference lattice energies, has become the standard benchmark and DFT methods have been extensively tested using this set.[80,131,329,376,377] Here, we use the most recent re-determination of the X23 reference data[362] to assess the performance of the various functionals and dispersion corrections examined in this work. The error statistics are shown in Table 6.3.

The table is separated into two sections, with the upper part showing results obtained with full geometry optimization of the molecular crystals at each listed level of theory. As noted in the computational methods section, we were only able to perform calculations using the tight basis set for GGA functionals due to the high memory requirements for hybrids. However, literature results[37] for PBE0-TS and PBE0-MBD with tight settings (which used the earlier X23 reference data[80]) are provided as these combinations give the lowest MAEs obtained with each of these dispersion corrections. We note that updating the reference data causes the MAEs to change by at most 0.25 kcal/mol, although often the deviation is lower.

As observed previously,[37,79] TS massively overbinds these molecular crystals. With TS and MBD, there is a significant difference between the light and tight results, which occurs because the damping parameters within these dispersion corrections are not optimized

Table 6.3: Mean absolute errors (in kcal/mol) for the X23 solid-state benchmark with selected functionals, dispersion corrections, and basis sets.

Functional	Dispersion	Light	Tight
Full Geometry Relaxation			
PBE	TS	4.17	3.14
HSE06	TS	4.57	—
PBE0	TS	4.44	2.39[a]
PBE	MBD	1.61	0.94
HSE06	MBD	2.12	—
PBE0	MBD	1.98	0.84[a]
PBE	XDM	1.14	1.04
HSE06	XDM	1.20	—
PBE0	XDM	1.14	—
PBE50	XDM	1.25	—
B86bPBE	XDM	0.83	**0.72**
B86bPBE-25	XDM	**0.81**	—
B86bPBE-50	XDM	1.06	—
Single Points at GGA/Light Geometries			
PBE0//PBE	MBD	1.97	1.07[b]
PBE0//PBE	XDM	1.01	0.96[b]
PBE-50//PBE	XDM	1.00	0.87[b]
B86bPBE-25//B86bPBE	XDM	**0.66**	**0.48**[b]
B86bPBE-50//B86bPBE	XDM	0.70	0.53[b]

[a]Literature value obtained from Ref. 37. [b]The hybrid energies with the light settings are corrected using the difference between light and tight results at the GGA level (via Eq. 6.6).

for each basis set independently. As a result, XDM significantly outperforms MBD with the light basis set, although the two methods give comparable MAEs with the tight basis. Also, the PBE-XDM and B86bPBE-XDM MAEs with light are in excellent agreement with previous results obtained using the Quantum ESPRESSO plane-wave code.[348]

B86bPBE-XDM with the tight basis set yields the lowest MAE (0.72 kcal/mol) yet obtained for the X23 set with any dispersion-corrected GGA, although this is largely due to the improvement in the reference data (the MAE compared to the values in Ref. 80 is 0.90 kcal/mol). For comparison, the MAE for PBE0-MBD in Table 6.3 is 0.84 kcal/mol and the lowest MAE reported by Thomas et al.[378] for the X23 is 0.81 kcal/mol, obtained with the TPSS-D3 dispersion-corrected meta-GGA. (The D3 dispersion correction by Grimme et al. is not available in FHIaims, so a direct comparison was not possible.) Regarding the dispersion-corrected hybrid functionals, the best results are obtained with B86bPBE-25/light (0.81 kcal/mol with either the Dolgonos et al.[362] or the older Reilly et al.[80]

reference data) followed by PBE0-MBD/tight (0.84 kcal/mol), with the former being considerably more efficient. For comparison, Thomas et al.[378] report MAEs of 0.93 kcal/mol for PBE0-MBD at the PBE-TS optimized geometries and 1.03 kcal/mol for PBE0-D3.

Computational efficiency is an important consideration in CSP, where hundreds to thousands of candidate crystal structures must be ranked with DFT for a given compound. Composite approaches, in which a relatively low level of theory is used for geometry optimization, followed by single-point energy evaluation at a higher level of theory, are an excellent strategy to reduce the computational cost without losing accuracy.[337,379] In this work, we consider composite approaches that use dispersion-corrected GGA functionals (PBE-MBD, PBE-XDM, or B86bPBE-XDM) and the light basis set for geometry optimization. Single-point energies are then evaluated with the corresponding 25% or 50% hybrid functionals and the light basis set and, in some cases, also with the same GGA and the tight basis set. This allows us to obtain energies (via Eq. 6.6) with an accuracy comparable to what would be expected from full hybrid/tight calculations, but with a drastically reduced computational cost. MAEs in the X23 lattice energies obtained using this type of composite approach are shown in the lower portion of Table 6.3. The notation in the table is high-level (hybrid)//low-level (GGA).

The MAEs obtained with the composite approach using B86bPBE-25-XDM and B86bPBE-50-XDM are the lowest errors yet obtained for the X23 set with any DFT method. The composite B86bPBE-25-XDM//B86bPBE-XDM approach with basis-set correction gives an MAE of only 0.48 kcal/mol, well below the usual target of 1 kcal/mol deemed to be chemical accuracy and almost exactly on the 2 kJ/mol mark commonly cited as the average energy difference between polymorphs.[380] It is also reassuring that the average error for molecular crystals is similar to that for dimers formed from molecules with similar sizes (Table 6.2), for which the MAEs were in the 0.2 kcal/mol to 0.4 kcal/mol range.

While good performance for absolute lattice energies is highly desirable, it does not necessarily ensure reliable polymorph ranking, which is dependent on accurate lattice-energy differences (as well as thermal and kinetic factors). The performance of the proposed methods for relative lattice energies will be examined in detail elsewhere. Nonetheless, improvements in absolute lattice energies do tend to result in more accurate relative lattice energies, as seen for the two oxalic acid polymorphs (α and β forms) appearing in the X23 set. The choice of dispersion correction has only a minor effect on this energy difference, although the polymorph ordering is highly dependent on exact-exchange mixing

in the base functional. As shown in Figure 6.1, the GGAs predict the incorrect energy ordering and relatively large fractions of exact exchange (near 50%) are needed to recover the reference lattice-energy difference. This suggests that delocalization error is a factor in determining the most stable oxalic acid polymorph, with the β form likely favored by GGAs due to its dimeric hydrogen-bonding.
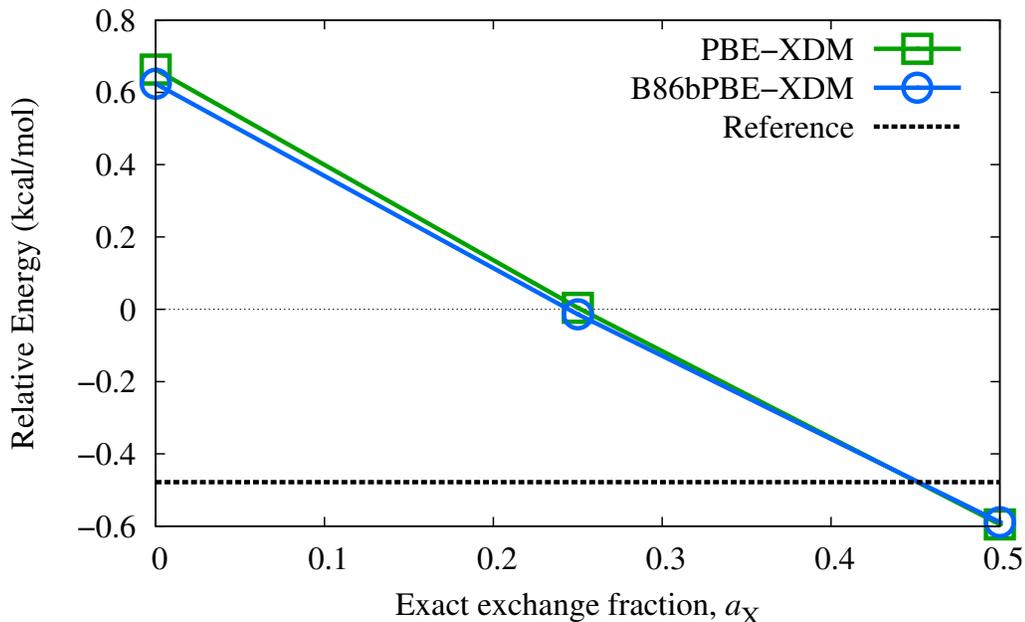
Figure 6.1: Relative energies of the α and β polymorphs of oxalic acid computed with various XDM-corrected GGA and hybrid functionals with the light basis set, evaluated at the corresponding GGA geometries. The α form is the most stable experimentally.

6.6 ICE LATTICE ENERGIES

Lastly, we examine the calculation of the lattice energies for the various phases of ice. The study of intermolecular interactions in water is both very important, because of its central role in many disciplines, and very challenging computationally, as electrostatics, induction, and dispersion all play a role. In general, it is agreed that the dispersion contribution, albeit smaller than in other non-covalently bound systems, is still necessary to describe water-water interactions accurately.[81,363] There are also significant many-body effects in water arising from intermolecular electron delocalization[364] that lead to delocalization error. As a result, a functional that describes the properties of water and ice accurately is still missing.[81,364]

Table 6.4: Mean absolute errors (in kcal/mol) for the ICE13 ice phases benchmark with selected functionals, dispersion corrections, and basis sets.

Functional	Dispersion Correction	Absolute		Relative	
		Light	Tight	Light	Tight
Full Geometry Relaxation					
PBE	TS	3.69	2.18	0.56	0.51
HSE06	TS	2.68	—	0.40	—
PBE0	TS	2.38	—	**0.37**	—
PBE	MBD	3.70	2.19	0.66	0.60
HSE06	MBD	2.68	—	0.40	—
PBE0	MBD	2.38	—	**0.37**	—
PBE	XDM	2.79	**1.71**	0.91	0.73
HSE06	XDM	1.60	—	0.72	—
PBE0	XDM	1.35	—	0.65	—
PBE50	XDM	0.79	—	0.46	—
B86bPBE	XDM	2.69	1.78	0.67	**0.45**
B86bPBE-25	XDM	1.16	—	0.53	—
B86bPBE-50	XDM	**0.55**	—	0.42	—
Single Points at GGA/Light Geometries					
PBE0//PBE	MBD	2.13	0.61[a]	0.34	0.29[a]
PBE0//PBE	XDM	1.11	0.30[a]	0.61	0.43[a]
PBE-50//PBE	XDM	**0.25**	1.16[a]	0.35	0.21[a]
B86bPBE-25//B86bPBE	XDM	0.93	**0.19**[a]	0.49	0.28[a]
B86bPBE-50//B86bPBE	XDM	0.32	1.20[a]	**0.31**	**0.19**[a]

[a]The hybrid energies with the light settings are corrected using the difference between light and tight results at the GGA level (via Eq. 6.6).

A strict test of density functionals and dispersion corrections for water is calculation of the absolute lattice energies of the various (ordered) ice phases. Different ice phases vary in molecular arrangements and in the extent of electron delocalization, which has been shown to correlate with the absolute lattice energy.[381] In a recent work, Della Pia et al.[81] reported absolute lattice energies of 13 ordered ice phases calculated using Diffusion Monte Carlo (DMC), and subsequently benchmarked a number of functionals using a plane-wave approach. Relative to the X23, this set has the advantage that no vibrational or nuclear quantum effects need to be removed before comparing to DFT results. We now use this ICE13 set, which is a superset of the previous ICE10 set proposed by Brandenburg et al.,[363] to evaluate the performance of our XDM-corrected methods.

Table 6.4 shows the MAEs calculated for the absolute and relative lattice energies of the ICE13 set with respect to the DMC data. The MAE of the relative energies is calculated by considering all 78 pairs of crystals in the ICE13 set, to avoid singling out any particular

ice phase. As in the case of the X23, the MAEs with the tight basis set are lower (in most cases) than with light, both for the absolute and for the relative lattice energies. XDM outperforms TS and MBD for absolute lattice energies by around 0.4–0.5 kcal/mol, but gives higher errors by a few tenths of a kcal/mol for the relative lattice energies. The average errors from the GGA functionals are quite high, in the vicinity of 2 kcal/mol for the absolute lattice energies and ca. 0.5 kcal/mol or more for the relative lattice energies. Hybrids give improved results, providing another indication that the cooperative hydrogen bonding networks in ice exhibit considerable delocalization error.

While we could not run the hybrid calculations with the tight basis set, the light results indicate that 25% hybrid functionals reduce the MAE, and 50% hybrids reduce it even further. However, the statistics for the composite methods in Table 6.4 allow us to understand the effects of BSIE and the incorporation of exact exchange separately, and reveal that this may be ascribed to error cancellation. While the results with the 25% hybrids improve when the basis-set correction of Eq. 6.6 is added, the 50% hybrid functionals perform better with the light basis set and no BSIE correction. This suggests that, when half-and-half functionals are used for water, there is error cancellation between delocalization error and BSIE.

As for the X23, the best-performing method for ice is found to be the composite approach using B86bPBE-25-XDM with the additive BSIE correction, which yields MAEs of 0.19 kcal/mol and 0.28 kcal/mol for the absolute and relative lattice energies, respectively. This absolute lattice-energy error is lower than the MAEs of all functionals studied by Della Pia et al.[81], and the relative-energy error is also among the best. For comparison, the best-performing functional reported[81] for absolute lattice energies is the revPBE-D3 GGA, with a MAE of 0.22 kcal/mol, and the best functionals in each of the other classes are rSCAN (meta-GGA, 0.23 kcal/mol), vdw-DF2 (non-local, 0.32 kcal/mol), and revPBE0-D3 (hybrid, 0.39 kcal/mol). Naturally, all these functionals are well within the "good functional" category established by the authors (MAEs < 0.96 kcal/mol and < 0.48 kcal/mol for absolute and relative lattice energies, respectively).

6.7 CONCLUSIONS

The calculation of lattice energies, the energy required to separate a molecular crystal into its component molecules, is of fundamental importance and a particularly stringent test for computational methods. The plane-wave implementation of exchange-hole dipole

moment (XDM) dispersion model, in particular in combination with the B86bPBE functional, has been shown to give excellent results for the calculation of absolute and relative lattice energies. This makes it a good choice for the final energy ranking in molecular crystal structure prediction (CSP). However, the reliance on plane waves imposes a poor computational scaling with system size and limits the applicability of XDM to semilocal functionals, which results in poor performance for systems with high conformational flexibility or significant delocalization error. In this work, we presented the implementation of XDM with numerical atomic orbitals (NAO) in the FHIaims package. This makes our methods effectively linear scaling and enables efficient combination of XDM with hybrid functionals without significant basis-set incompleteness errors, thus mitigating the aforementioned problems.

To test the accuracy of the new XDM-corrected hybrid functionals, we assessed their performance for binding energies of molecular gas-phase dimers and trimers, as well as lattice energies of small molecular crystals (the X23 set) and 13 phases of ice (the ICE13 set). The results were compared to the Tkachenko-Scheffler (TS) and state-of-the-art many-body dispersion (MBD@rsSCS) methods. For molecular dimers, XDM-corrected functionals achieve a mean average error (MAE) of between 0.2 and 0.4 kcal/mol, slightly outperforming TS and MBD. More importantly, XDM-corrected functionals also show excellent performance for three-body interaction energies (the 3B-69 set), suggesting that electronic many-body effects are much more important than atomic many-body dispersion effects, which are not included in the canonical XDM methods.

The XDM-corrected methods also yield very low average errors for the X23 set of lattice energies, particularly if hybrid methods or relatively large ("tight") basis sets are used. The most intriguing result is the spectacular performance of composite methods, in which a GGA geometry optimization (e.g. B86bPBE-XDM) is followed by a single-point energy calculation to incorporate the benefits of using a hybrid functional (e.g. B86bPBE-25-XDM), and perhaps an additional single-point correction to treat basis-set incompleteness error (the difference between tight and light energies at the GGA level). The best-performing composite method (B86bPBE-25-XDM with basis-set correction, at the B86bPBE-XDM equilibrium geometries) achieves a MAE of only 0.48 kcal/mol for the X23 set, roughly half the error of other similar DFT methods.

The excellent performance of the basis-corrected B86bPBE-25-XDM//B86bPBE-XDM composite method extends to the calculation of the absolute lattice energies of 13 ice

phases, for which it achieves an MAE of only 0.19 kcal/mol, outperforming all DFT functionals reported to date. The calculation of absolute lattice energies of ice is particularly difficult due to the presence of delocalization error and the delicate balance between electrostatics, dispersion, and induction. It is a key point that one single methodology works well for molecular dimers and trimers, and achieves the lowest MAE for both the X23 and ICE13 lattice energies. Obtaining good across-the-board performance in all tests examined is of paramount importance when modeling complex materials that feature several disparate types of non-covalent interactions. This makes us confident that the proposed XDM-corrected methods will serve nicely for accurate energy ranking in crystal structure prediction.

CHAPTER 7

ACCURATE AND EFFICIENT POLYMORPH ENERGY RANKING WITH XDM-CORRECTED HYBRID DFT

This chapter is adapted from: A. J. A. Price, R. A. Mayo, A. Otero de la Roza, and E. R. Johnson, "Accurate and efficient polymorph energy ranking with XDM-corrected hybrid DFT" *CrystEngComm*, **25**, 953-960 (2023).

7.1 INTRODUCTION

Molecular crystals are of central importance as pharmaceuticals,[300–302] energetic materials,[305,306] and in the emerging field of organic electronics.[309,310] Due to the sensitivity of solid-state properties such as solubility and charge transport on crystal packing, it is important to identify all likely polymorphs when developing compounds for these applications.[311,315–319] The problem of theoretical identification of isolable polymorphs is termed first-principles crystal structure prediction (CSP).

Periodically, the Cambridge Crystallographic Data Centre (CCDC) organises blind tests of CSP methods in which crystal structures of small sets of compounds are determined by X-ray diffraction, but are not released to the community until researchers have attempted to predict the structure(s) of the isolated polymorph(s).[82–87] There are two challenges at the core of CSP – that of exhaustive structure generation and that of accurate energy ranking of the resulting structure candidates. Ideally, the experimentally observed polymorph(s) should be among the most energetically stable of the putative crystal structures.
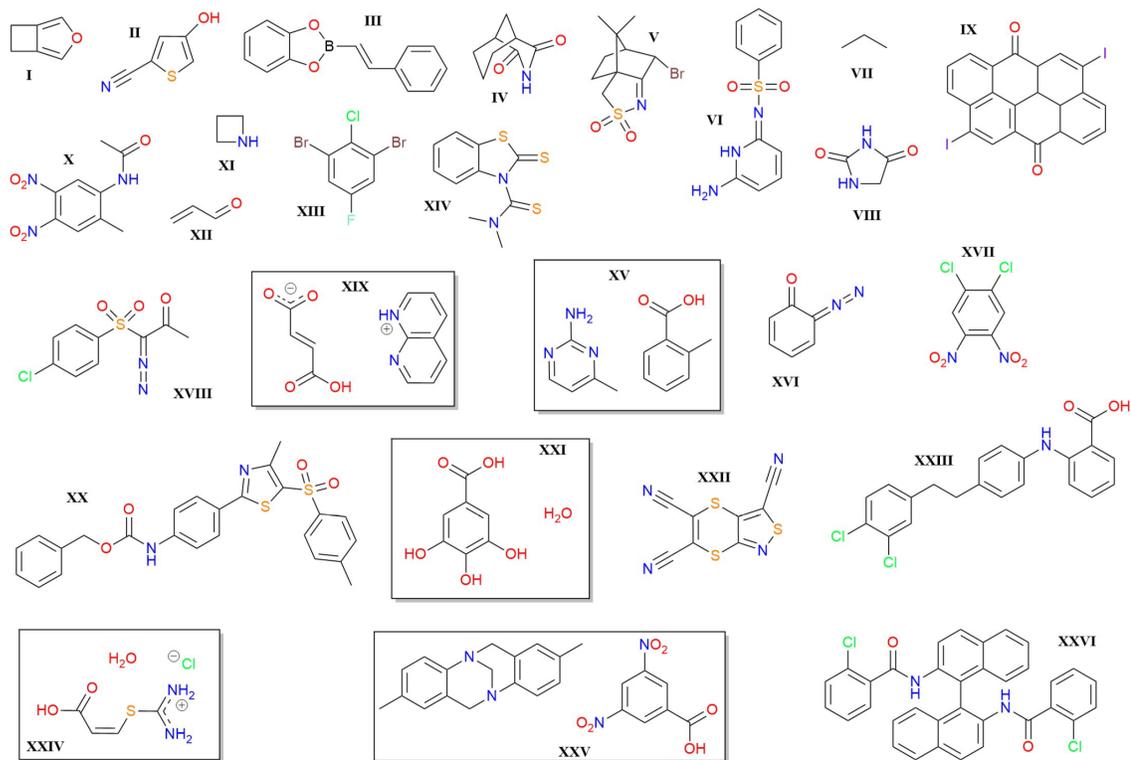
Figure 7.1: Structures of the compounds provided by the CCDC for the first 6 CSP blind tests. Rectangles are used for molecule grouping in co-crystals. I: 3-oxabicyclo[3.2.0]hepta-1,4-diene, II: 4-hydroxy-2-thiophenecarbonitrile, III: 2-(2-phenylethenyl)-1,3,2-benzodioxaborole, IV: 3-azabicyclo[3.3.1]nonane-2,4-dione, V: 7-endo-(bromocamphorylsulfonyl)imine, VI: 6-amino-2-phenylsulfonylimino-1,2-dihydropyridine, VII: propane, VIII: hydantoin, IX: 2,9-diiodoanthanthrone, X: 2-acetamido-4,5,-dinitrotoluene, XI: azetidine, 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, XVI: 2-diazo-3,5-cyclohexadiene-2-one, XVII: 1,2-dichloro4,5-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, XXII: tricyano1,4-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-6H,12H-5,11-methanodibenzo[b,f][1,5]diazocine:3,5-dinitrobenzoic acid, XXVI: N,N'-([1,1'-Binaphthalene]-2,2'-diyl)bis(2-cholorobenzamide).

The focus of the present work is on the energy ranking step of CSP, where computational methods that provide well-balanced descriptions of electrostatics, charge transfer, polarization, non-bonded repulsion, and London dispersion are required. The advent of dispersion corrections has opened the door for the use of periodic-boundary density-functional theory (DFT) for CSP, as illustrated by Neumann and coworkers[320–323] in the 5th CSP blind test[86] and further demonstrated in many subsequent studies.[311,314,332–344] As dispersion-corrected DFT outperforms alternative energy-ranking methods for CSP, it is essential to have a DFT-based dispersion method that is as accurate and efficient as possible for molecular crystals.

Many successful applications of DFT to molecular crystals have used planewave basis sets and the projector augmented-wave method.[311,320–323,332,334–336,347,382] However, such an approach has two key limitations. The first is that planewave methods have unfavourable computational scaling for large unit cells. This drastically limits their applicability to CSP since many polymorphs of fairly complex active pharmaceutical compounds, and other materials of interest, can contain many molecules per unit cell. The second is that planewave calculations are limited to generalized gradient approximation (GGA) methods and do not allow routine use of hybrid density functionals. This is important due to delocalization error,[204,231,349] which is prevalent in GGAs. Delocalization error can affect polymorph ranking for flexible molecules where there is a competition between intramolecular conjugation and intermolecular interactions,[336,350,351] as well as for organic salts[335] and acid-base co-crystal systems.[247] Hybrid functionals typically provide better performance than GGAs for non-covalent interactions in finite molecular systems,[75,349,355–357] particularly in cases with significant delocalization error, and we see similar improvement for solids based on lattice energies of small-molecule crystals.[245,383]

To avoid the limitations of planewave basis sets, we turn to numerical atom-centered orbitals (NAOs). NAOs of finite extent are a highly promising alternative since they allow elimination of integrals involving distant atomic centers in DFT calculations, resulting in formal linear scaling for large systems.[65] In particular, the Fritz Haber Institute *ab initio* materials simulation (FHI-aims) package[64,68,353] is a very robust NAO code that, thanks to its design, has minimal basis-set incompleteness. Implementation in FHI-aims has allowed application of the many-body dispersion (MBD) method,[36,37] paired with hybrid DFT, to molecular crystal benchmarks and CSP studies.[338,340–344] Moreover, we recently implemented our exchange-hole dipole moment (XDM) dispersion model in FHI-aims and showed that, when paired with selected hybrid functionals, it provides unprecedented

accuracy for evaluation of molecular-crystal lattice energies.[383]

In this article we assess the performance of XDM paired with hybrid functionals and the light NAO basis set for the energy ranking step of CSP. Specifically, we consider the 26 compounds that formed the first 6 CSP blind tests, shown in Figure 7.1.[82–87] Of the submitted putative crystal structures, the experimental polymorph is consistently ranked among the 10 most-stable unique candidates. Use of a hybrid functional with 50% exact-exchange mixing is shown to significantly improve on GGA results for challenging crystal energy landscapes where delocalization error in GGA functionals is known to adversely impact the ranking.

7.2 THEORY AND COMPUTATIONAL METHODS

7.2.1 THE XDM DISPERSION CORRECTION

XDM[10,63] is a post-self-consistent correction to the energy computed with some base density functional approximation (DFA):

$$E = E_{\text{DFA}} + E_{\text{XDM}}. \tag{7.1}$$

The XDM dispersion energy itself is expressed as a sum over all pairs of atoms, i and j, in the system

$$E_{\text{XDM}} = -\frac{1}{2} \sum_{n=6,8,10} \sum_{ij} \frac{C_{n,ij}}{R_{ij}^n + R_{\text{vdW},ij}^n}, \tag{7.2}$$

where R_{ij} is the internuclear distance. In a solid, the sum runs over all surrounding unit cells. Here, C_6, C_8, and C_{10} are termed the atomic dispersion coefficients, which are functions of the self-consistent electron density, its gradient and Laplacian, the kinetic-energy density, and Hirshfeld partitioning weights. As a result of their density and derivative dependence, the dispersion coefficients are highly responsive to changes in the chemical environment of an atom due to charge transfer, coordination, hydrogen bonding, and even weaker van der Waals interactions with distant atoms.[21]

The $R_{\text{vdW},ij}$ in Eqn. 7.2 is the sum of effective van der Waals radii of atoms i and j, which is defined using an average of three possible ratios of the dispersion coefficients and involves two empirical fit parameters, a_1 and a_2:

$$R_{\text{vdW},ij} = \frac{a_1}{3} \left[\left(\frac{C_{8,ij}}{C_{6,ij}} \right)^{\frac{1}{2}} + \left(\frac{C_{10,ij}}{C_{8,ij}} \right)^{\frac{1}{2}} + \left(\frac{C_{10,ij}}{C_{6,ij}} \right)^{\frac{1}{4}} \right] + a_2. \tag{7.3}$$

The a_1 and a_2 parameters are fit once for each choice of density functional and basis set by minimising the root-mean-square percent error in computed binding energies of 49 molecular dimers.[75,131] After this, the parameters are kept fixed and are transferable to all elements of the periodic table, as well as between finite (molecular) and periodic-boundary (solid-state) calculations.[383]

7.2.2 EXCHANGE-CORRELATION FUNCTIONALS

All density-functional dispersion corrections must be paired with a base DFA. In this work, we consider three DFA exchange-correlation functionals of the general form

$$E_{\mathrm{XC}} = (1 - a_{\mathrm{X}})E_{\mathrm{X}}^{\mathrm{B86b}} + a_{\mathrm{X}}E_{\mathrm{X}}^{\mathrm{HF}} + E_{\mathrm{C}}^{\mathrm{PBE}}, \tag{7.4}$$

which combine the B86b[8] exchange functional with PBE[9] correlation. B86b is our exchange functional of choice for molecular crystals due to its high accuracy for describing non-bonded repulsion.[93,94,131,171,348] The parameter a_{X} controls the extent of exact-exchange mixing, and we will consider values of 0, 0.25, and 0.50, which correspond to the B86bPBE GGA, and the B86bPBE-25X and B86bPBE-50X hybrid functionals, respectively.[245,383]

7.2.3 COMPUTATIONAL METHODS

All calculations in this work were carried out with the FHI-aims program (version 210513). The XDM method and the B86bPBE, B86bPBE-25X, and B86bPBE-50X functionals were all implemented in a copy of the code,[383] and these methods are now available in the distribution version of the software. For compounds from the first to fifth blind tests (I to XXI), the starting structures were the equilibrium B86bPBE-XDM structures obtained in our previous studies,[335,336] which in turn were obtained from the Supplementary Information of the corresponding blind test articles, including the reference experimental structures.[82–86] For the sixth blind test compounds (XXII to XXVI), in addition to the top three structures submitted by the various participant groups,[87] we considered the 100 structures submitted by Neumann's group[87] as well as the all structures in the POLY59 set compiled by Brandenburg and Grimme.[334] The numbers of candidate structures considered for each compound is given in the ESI of Ref. 384. All atomic positions and lattice vectors in all molecular crystals were fully optimised with B86bPBE-XDM. Subsequent single-point energy calculations were performed with the B86bPBE-25X and B86bPBE-50X hybrids, again paired with XDM. While geometry optimizations are possible with the hybrid functionals,

91

we do not expect this to lead to improved accuracy based on our previous work,[383] so we consider only GGA geometries to reduce the computational expense.

Calculations used the "light" (double-ζ) NAO basis sets for computational efficiency, along with the dense integration grids (including Lebedev meshes up to 434 angular grid points) that are used by default in conjunction with the tight settings. The **k**-point grids were chosen to have dimensions $n_1 \times n_2 \times n_3$ such that

$$n_i = \text{int}\left[\max\left(1, R_k|b_i| + 0.5\right)\right], \tag{7.5}$$

where $|b_i|$ is the magnitude of the corresponding reciprocal lattice vector and the length parameter, R_k, was set to 50 Bohr. The XDM damping parameters were taken from Ref. 383 and are $a_1 = 0.8219$, $a_2 = 1.2069$ Å for B86bPBE, $a_1 = 0.5235$, $a_2 = 2.1995$ Å for B86bPBE-25X, and $a_1 = 0.0831$, $a_2 = 3.7362$ Å for B86bPBE-50X.

7.3   RESULTS AND DISCUSSION

7.3.1   SUMMARY OF POLYMORPH RANKING

Results of the energy ranking of putative crystal structures for the 26 compounds constituting the first 6 CSP blind tests are summarized in Table 7.1. The B86bPBE-XDM results are similar to those from previous planewave calculations performed with the same methodology,[335,336] confirming the minimal effects of basis-set incompleteness in these NAO calculations. Overall, the GGA and B86bPBE-25X-XDM hybrid identify an experimental polymorph as the minimum-energy structure in 17/26 cases, and this fraction increases to 18/26 cases with the B86bPBE-50X-XDM hybrid functional.

Notably, with the 50% hybrid, an experimental polymorph is ranked second in another 6/26 cases. In 3 of these, the experimental form is nearly degenerate with the minimum-energy candidate, while it lies 1.2 kJ/mol or less above the minimum otherwise. For compound V, the experimental structure is ranked 4th in energy, but still lies within 1.5 kJ/mol of the minimum. These are all sufficiently small energy differences that thermal free-energy contributions (which are generally <2 kJ/mol)[385,386] may be sufficient to reverse the ordering. However, evaluating the thermal corrections requires very computationally expensive phonon calculations, which is beyond the scope of the present work.

7.3.2   COMPOUND XX

All three functionals considered perform consistently poorly for the large, flexible compound XX (benzyl-(4-(4-methyl-5-(*p*-tolyl-sulfonyl)-1,3-thiazol-2-yl)phenyl)carbamate) from

92

Table 7.1: Ranking of the most stable experimental polymorph on crystal energy landscapes computed with the B86bPBE-XDM (GGA), B86bPBE-25X-XDM (25X), and B86bPBE-50X-XDM (50X) functionals. Also shown is the energy (ΔE, in kJ/mol) of that experimental polymorph relative to the most stable candidate structure on each landscape.

Compound	GGA		25X		50X	
	Rank	ΔE	Rank	ΔE	Rank	ΔE
I	1[a]	0.0	1[b]	0.0	1[b]	0.0
II	4	1.1	2	0.7	2	0.5
III	1	0.0	1	0.0	1	0.0
IV	1	0.0	1	0.0	1	0.0
V	4	1.9	4	1.9	4	1.5
VI	1	0.0	1	0.0	1	0.0
VII	1	0.0	1	0.0	1	0.0
VIII	1	0.0	1	0.0	1	0.0
IX	1	0.0	1	0.0	1	0.0
X	2	0.9	2	0.4	1	0.0
XI	1	0.0	1	0.0	1	0.0
XII	1	0.0	1	0.0	1	0.0
XIII	1	0.0	1	0.0	1	0.0
XIV	1	0.0	1	0.0	1	0.0
XV	2	0.3	2	0.8	2	1.2
XVI	1	0.0	1	0.0	1	0.0
XVII	1	0.0	1	0.0	1	0.0
XVIII	1	0.0	1	0.0	1	0.0
XIX	6	4.3	4	2.2	2	0.1
XX	4	6.7	4	6.5	6	6.7
XXI	1	0.0	1	0.0	1	0.0
XXII	9	2.7	3	0.4	2	0.1
XXIII	3[c]	0.8	2[c]	0.0	2[c]	0.0
XXIV	8	2.0	3	0.7	2	0.4
XXV	1	0.0	1	0.0	1	0.0
XXVI	1	0.0	1	0.0	1	0.0

[a]polymorph 1, [b]polymorph 2, [c]polymorph c

the 5th blind test, the structure of which is shown in Figure 7.1. Here, the minimum-energy structure is predicted to be more stable than the isolated experimental form by 6.5-6.7 kJ/mol.

One potential source of error for application of DFT to flexible molecules is the conformational energy[350,351] and examination of the structures shows that the DFA-predicted minimum has a different molecular conformation than the experimental form. However,

we expect conformational changes to be much better described by hybrid functionals than with GGAs. Additionally, since the conjugated central portion of the molecule must remain planar, the conformational differences primarily involve a 180° degree twist of the amide group. This conformational change should be well described by DFT and not responsible for a large energy reordering of the putative crystal structures.

An alternative explanation for the relatively high energy of the experimental structure may be that the crystallisation is controlled by the rugosity.[387] This refers to the roughness of a cleaved crystal surface – crystals with a smooth cleavage plane should be more likely to form than those with a rough surface. We examined the rugosity of the putative crystal structures for compound XX using the CSD-Particle suite of tools within the Mercury program[388] (2022.2.0) and the resulting crystal rugosity-energy landscape is shown in Figure 7.2. While the minimum-energy crystal structure is found to have a higher rugosity than the experimental form, several other low-energy candidates also have lower rugosities, so this does not reconcile theory and experiment.
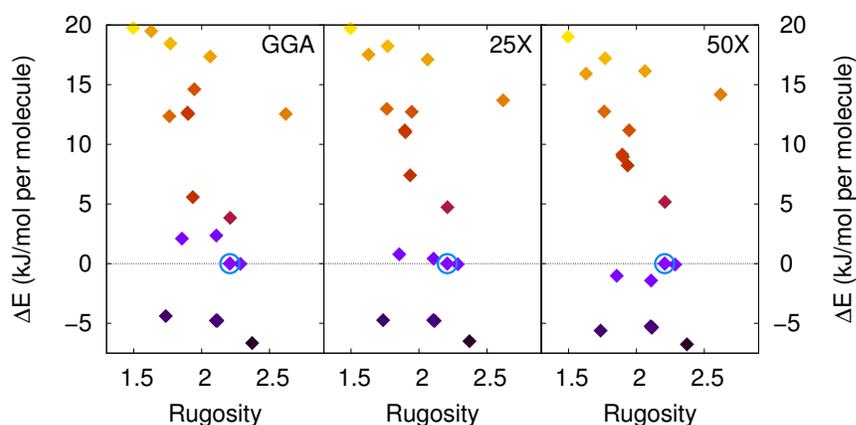
Figure 7.2: Computed crystal rugosity-energy landscapes for compound XX, using the rugosity as the ordinate. Results are shown for full geometry relaxations using B86bPBE-XDM (left), as well as single-point energy calculations with B86bPBE-25X-XDM (middle) and B86bPBE-50X-XDM (right). The energy range is truncated to focus on the low-energy candidates and the experimental form is circled.

A final potential source of error in the DFT ranking is the neglect of thermal free-energy corrections. It is notable that the experimental crystal structure of compound XX was successfully predicted by two participating groups[389] in the 5th CSP blind test,[86] both of whom used the DMACRYS force field[390] with an empirical exp-6 repulsion-dispersion term. It is possible that the force field partly captures effects of thermal expansion on the crystal structure through its empirical parameterization. The 4.4-6.7 kJ/mol energy differences between the low-energy and experimental forms obtained with DFT are much larger than

the contributions from thermal effects seen for small, rigid molecules ($<$2 kJ/mol).[385,386] However, it would be expected that the magnitude of the thermal free-energy corrections would increase with the number of rotatable bonds, leading to much larger contributions for compound XX (and XXIII), relative to the other members of the blind test set. Indeed, for compound XXIII, Hoja and Tkatchenko[342] noted that inclusion of the vibrational free energy correction could alter the relative stabilities by as much as 8.4 kJ/mol. Thus, one potential reason that compound XX is an outlier for DFT is a much larger contribution from thermal effects than what is seen for rigid molecules.

7.3.3 EFFECTS OF DELOCALIZATION ERROR ON LANDSCAPES

To highlight the improved performance of hybrid functionals based on B86bPBE-XDM over the parent GGA for CSP, we focus on four challenging crystal energy landscapes. The first two are for compounds X (2-acetamido-4,5-dinitrotoluene)[84] and XIX (1,8-naphthyridinium fumarate),[86] where the delocalization error inherent in GGA functionals has been previously shown to adversely affect the predicted landscapes.[335,336,350] This results in one or more candidate structures being predicted to lie lower in energy than the known polymorph, as shown in Figure 7.3.

The crystal energy landscape of compound X has been extensively investigated by both ourselves[336] and by Beran and coworkers.[350] Here, there are two competing low-energy structures. The experimental form has the acetamide group twisted somewhat out of plane to maximise intermolecular hydrogen bonding. Conversely, GGA functionals favour a competing structure in which the acetamide group lies in plane to maximise conjugation. The energy ordering can be corrected by evaluating the conformational energy change of the isolated molecule using a correlated-wavefunction theory, such as MP2D.[350] Conversely, compound XIX is an organic salt with rigid molecular components. Salts are problematic for GGA functionals, which are known to give large errors for polarization and charge transfer.[206,208,236,247,252,349] Here, GGA functionals favour the structure with the most cooperative H-bonding network.[335]

The other two landscapes highlighted in Fig. 7.3 are for compounds XXII (tricyano-1,4-dithiino[c]-isothiazole) and XXIV (chloride salt hydrate of (Z)-3-((diaminomethyl)thio)acrylic acid) from the 6th CSP blind test.[87] Like compound XIX, compound XXIV is also an organic salt, so we again expect significant delocalization error. Examination of the minimum-energy GGA structure shows an unusual arrangement in which the H atom of the carboxlic acid moiety points away from the sp^2 oxygen to form a hydrogen bond to the chloride (see Fig. 7.4). It is likely that the strength of this ionic H-bond is overestimated by the GGA
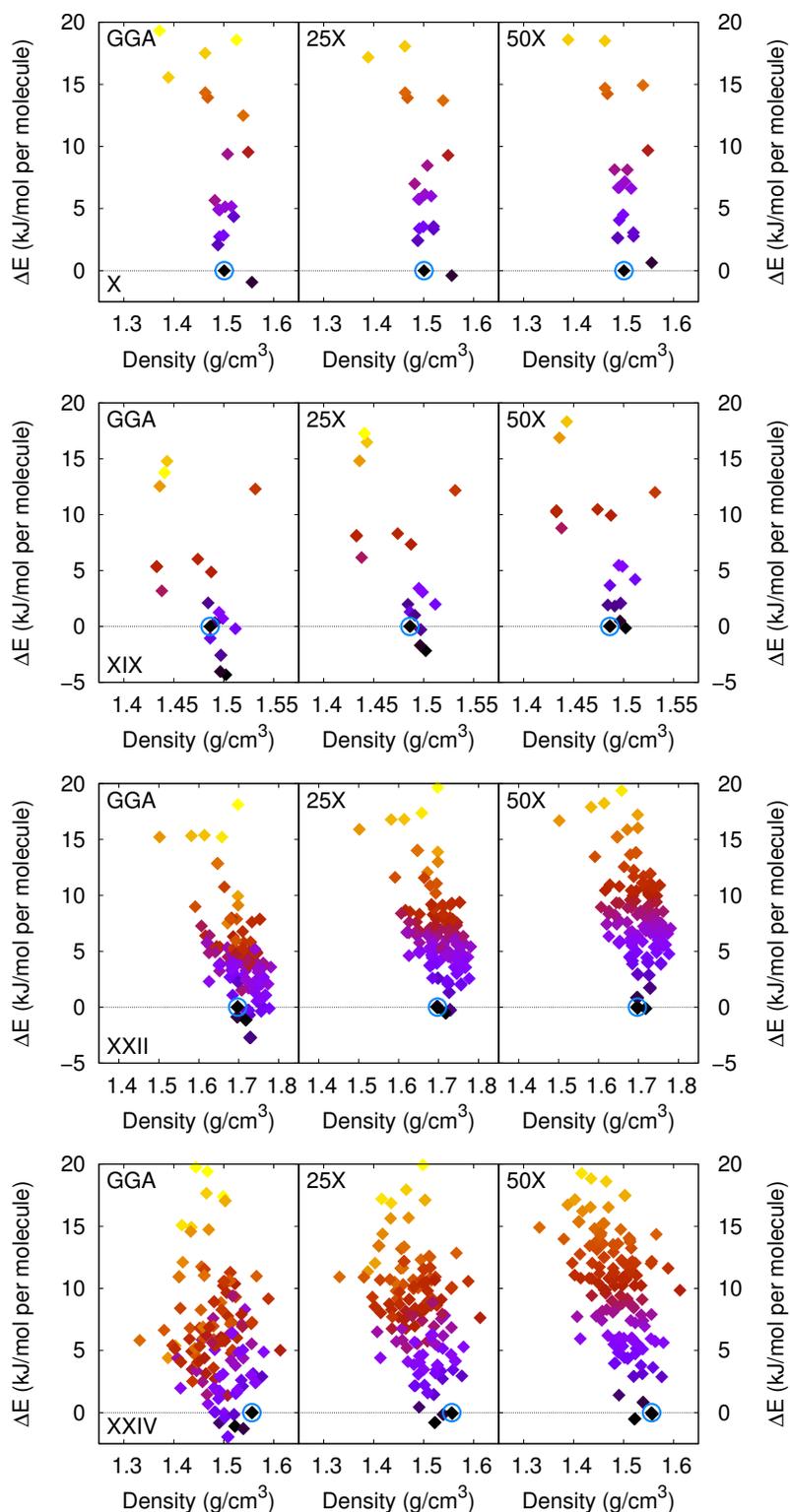
Figure 7.3: Computed crystal energy landscapes for compounds X, XIX, XXII, and XXIV (top to bottom). Results are shown for full geometry relaxations using B86bPBE-XDM (left), as well as single-point energy calculations with B86bPBE-25X-XDM (middle) and B86bPBE-50X-XDM (right). The energy ranges are truncated to focus on the low-energy candidates and the experimental forms are circled.
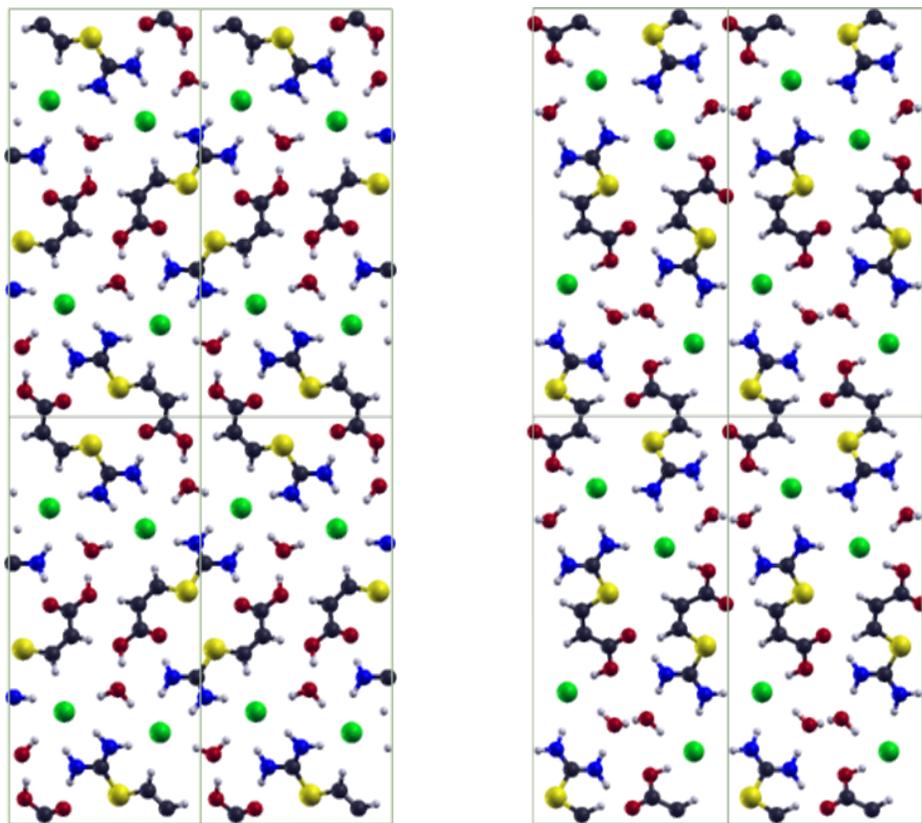
Figure 7.4: The experimental (left) and the DFT-predicted minimum-energy (right) crystal structures for compound XXIV, chloride salt hydrate of (Z)-3-((diaminomethyl)thio)acrylic acid, viewed in the yz plane. C: grey, H: white, N: blue, O: red, S: yellow, Cl: green.

functional. Finally, it is not immediately obvious how delocalization error would affect the crystal energy landscape of compound XXII. Examination of the minimum-energy GGA and experimental crystal structures reveals that, while the experimental structure has the molecules stacked in dimers, the GGA minimum has all the molecules aligned along the c axis (see Fig. 7.5). It is possible that this results in excessive polarization at the GGA level, overestimating the electrostatic stabilization.

Hybrid functionals, particular those with near 50% exact-exchange mixing, are known to reduce delocalization error. As such we expect B86bPBE-50-XDM to significantly improve the crystal energy landscapes in cases where this error plays a role. The hybrid results shown in Figure 7.3 confirm this to be the case. B86bPBE-50-XDM predicts the experimental form of compound X to be lowest in energy. Moreover, it almost entirely corrects the energy ranking for compound XXII and both organic salts, with the most-stable candidate structures now nearly degenerate with the experimental forms.
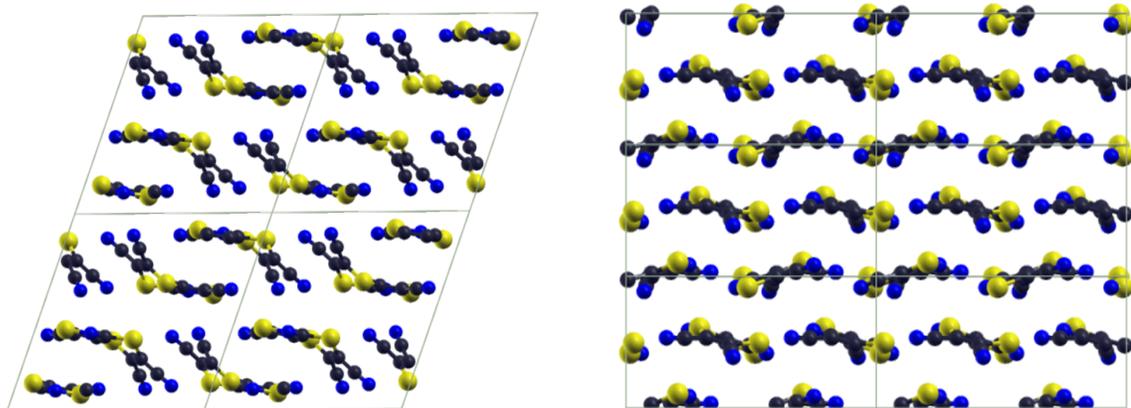
Figure 7.5: The experimental (left) and the DFT-predicted minimum-energy (right) crystal structures for compound XXII, tricyano1,4-dithiino[c]-isothiazole, viewed in the ac plane. C: grey, H: white, N: blue, S: yellow.

7.4 SUMMARY

In this work, we applied a recent NAO implementation of the XDM dispersion model to rank the submitted structures for the first 6 CSP blind tests. Unlike planewave basis sets, NAOs allow efficient use of hybrid DFT, so the B86bPBE GGA and two related hybrid functionals with 25% and 50% exact exchange mixing were considered. Pairing XDM with the B86bPBE-50X hybrid was found to rank an isolated experimental polymorph as the most stable structure for 18/26 compounds considered, and as the second most-stable structure for another 6/26. In 7 cases, the DFT minimum and the experimental form were separated by 1.5 kJ/mol or less, meaning that thermal free-energy corrections from the phonons may be sufficient to reverse the ranking. The only clear failure of hybrid DFT occurred for compound XX, where the experimental structure lies 6.7 kJ/mol above the DFT minimum, where the larger number of rotatable bonds may magnify the importance of thermal effects. Finally, B86bPBE-50X-XDM was found to provide excellent performance for several challenging crystal energy landscapes where GGA calculations suffer from delocalization error. This includes both cases where the error affects intramolecular conformational energies, and intermolecular charge transfer, making it a more general approach than monomer energy corrections. Overall, the FHI-aims implementation of XDM-corrected hybrid functionals provides an unprecedented combination of accuracy and efficiency for DFT-based modelling of molecular crystals that should facilitate high-throughput use of first-principles CSP.

CHAPTER 8

ASSESSMENT OF XDM-CORRECTED DENSITY FUNCTIONALS FOR THE ENERGY RANKING STAGE OF THE 7TH CSP BLIND TEST

8.1 INTRODUCTION

First-principles molecular crystal structure prediction (CSP) is a grand challenge in the field of computational chemistry and is fruitful ground for ongoing method development. To assess the performance of various methods for both the structure generation and energy ranking stages of CSP, the Cambridge Crystallographic Data Centre (CCDC) regularly holds blind test competitions. Solved but unpublished crystals structures for a small set of compounds are held in reserve by the CCDC and participants have a set period of time to submit predicted crystal structures given only the compounds' chemical diagrams. There have been 6 previous blind tests to date,[82-87] with the 7th held from 2020-2022 and the results as yet unpublished. This most recent blind test was separated into two phases, with the first focusing on structure generation and the second on energy ranking. In the second phase, lists of either 100 or 500 putative crystal structures were provided by the CCDC for each of 5 compounds, the chemical diagrams of which are shown in Figure 8.1.

In 5th[86] and 6th[87] blind tests, dispersion-corrected density-functional theory (DFT) was shown to provide large improvements in energy ranking of candidate crystal structures compared to classical force-field methods. While the Neumann group was the first to use DFT in the 5th blind test with their own empirical dispersion correction, the more
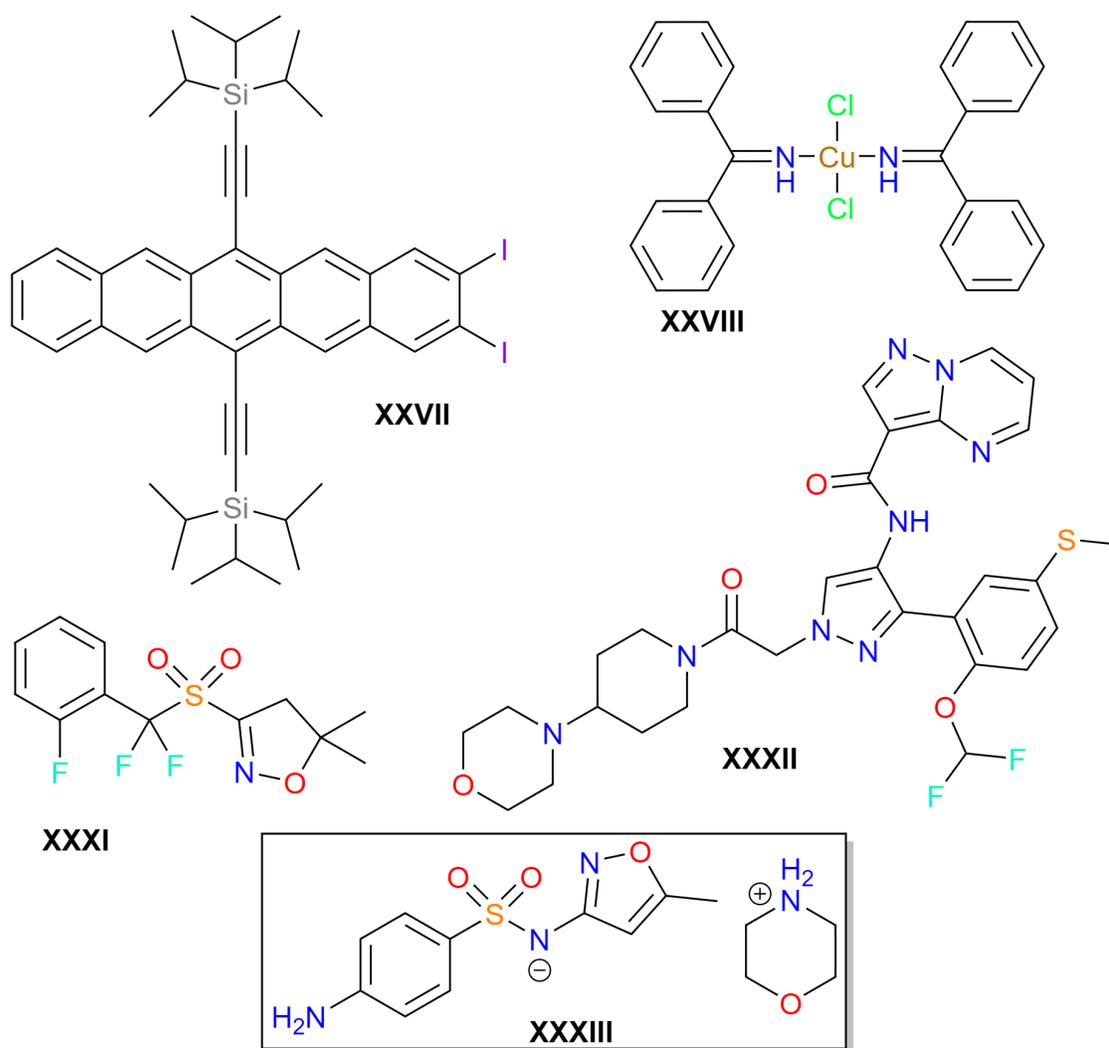
Figure 8.1: The 5 compounds considered for the energy-ranking phase of the 7th CSP blind test.

popular D3[31,32] and MBD[36,37] dispersion corrections were used in submissions to the following competition. Quite recently, we additionally assessed the performance of our XDM dispersion correction[10,383] for the compounds appearing in all six previous blind tests.[384] While delocalization error has been shown to adversely affect the energy ranking provided with generalized gradient approximation (GGA) functionals for some compounds, XDM-corrected hybrid functionals yielded improved performance.[384]

The present work reports on the performance of several XDM-corrected density functionals for the energy ranking phase of the 7th CSP blind test. Overall, our methodology was able to rank the most stable experimentally isolated polymorph as the lowest in energy for two of the five compounds considered (XXVIII and XXXIII), and second lowest for

one additional compound (XXVII). The additional two cases (XXXI and XXXII) are large organic molecules involving a number of rotatable bonds, for which the isolated polymorphs were ranked surprisingly high energetically.

8.2 COMPUTATIONAL METHODS

The CCDC provided lists of candidate structures to each of the participating teams,[391] including the atomic positions and unit-cell dimensions, in .cif format. The numbers of candidates provided were 100 each for compounds XXVII and XXXI and 500 each for compounds XXVIII, XXXII, and XXXIII. These lists were selected to include at least one structure representative of each polymorph identified from single-crystal x-ray diffraction (XRD).

All calculations used an in-house modified version of FHI-aims[64] version 210513. Geometry optimisations of all provided candidate crystal structures were performed using the B86bPBE functional[8,9] with the XDM dispersion correction,[383] the "light" basis set, and a "dense" integration grid, with a relaxation convergence threshold of 0.025 eV/Å. For all structures within 1.5 kcal/mol per molecule of the minimum, further geometry optimisation was performed with a tighter relaxation threshold of 0.005 eV/Å. Subsequent single-point energy calculations were then performed on all optimised structures using either the hybrid B86bPBE-25X or B86bPBE-50X functionals with XDM dispersion, with the damping function parameters matching those reported in Ref. 383. For compound XXVIII, all calculations were run for a ferromagnetic configuration with one unpaired electron per copper atom.

For the blind test submission, we selected results from the B86bPBE-25X functional for compounds XXVII and XXVIII, and results from B86bPBE-50X for the remaining three compounds, as we expected them to be potentially more susceptible to delocalization error. In particular, we thought this error might affect the relative conformational energies for the flexible molecules, and delocalization error was previously found to lead to poor energy ranking for an organic salt (compound XIX in the 5th blind test).[384]

8.3 RESULTS AND DISCUSSION

At the conclusion of the blind test, the CCDC shared the experimental crystal structures and the identities of the matching structures within the lists of candidates provided.[391]

This allowed us to assess the quality of our predicted crystal energy landscapes. A summary of our computational results is presented in Table 8.1, which shows the ranking of the experimental polymorphs obtained with each functional considered, as well as their relative energies above the minimum on each landscape.

Table 8.1: Summary of computational results for the 7th CSP blind test. For each compound, we show the energetic rank of all isolated polymorphs and their relative energies (ΔE in kJ/mol per formula unit) above the minimum on the corresponding crystal energy landscape for each of the three density functionals considered. LT: low temperature; Maj and Min indicate the major and minor components of a disordered structure.

Compound	Form	GGA		25X		50X	
		Rank	ΔE	Rank	ΔE	Rank	ΔE
XXVII	LT	2	1.9	2	1.5	2	1.3
XXVIII	–	1	0.0	1	0.0	1	0.0
XXXI	A_{Maj}	4	1.5	7	2.8	8	4.3
	A_{Min}	10	3.0	10	4.3	12	5.6
	B	14	4.6	16	5.6	18	6.2
	C	76	10.7	69	11.6	61	12.2
XXXII	A_{Maj}	25	7.2	26	9.2	32	11.6
	B	29	7.7	34	9.7	37	12.2
XXXIII	A	6	4.5	5	4.5	5	5.3
	B	1	0.0	1	0.0	1	0.0

We begin with discussion of compounds XXVII, XXVIII, and XXXIII, for which the predicted landscapes are shown in Figure 8.2. For compound XXVII, there was some ambiguity in determining the conformations of the isopropyl groups in the experimental structure due to their ability to rotate freely at room temperature. As a result, two candidates within the provided list of 100 structures corresponded to the same polymorph as the low-temperature experimental form. Upon geometry optimisation, both these candidates converged to an identical structure, yielding equal energies within the convergence criteria of the optimisation. This candidate is the 2nd-ranked structure on our landscape, independent of functional. It also lies < 2 kJ/mol above the minimum and it is likely that thermal-free energy corrections could alter the ordering, particularly since there could be considerable entropy differences between polymorphs depending on the mobilities of the isopropyl groups. For compound XXVIII, a single polymorph was identified experimentally. Here, all there functionals performed extremely well and each predicted the match to the experimental structure to be the minimum energy form. For compound XXXIII, two polymorphs were identified experimentally, with form B being the more stable.[391] While
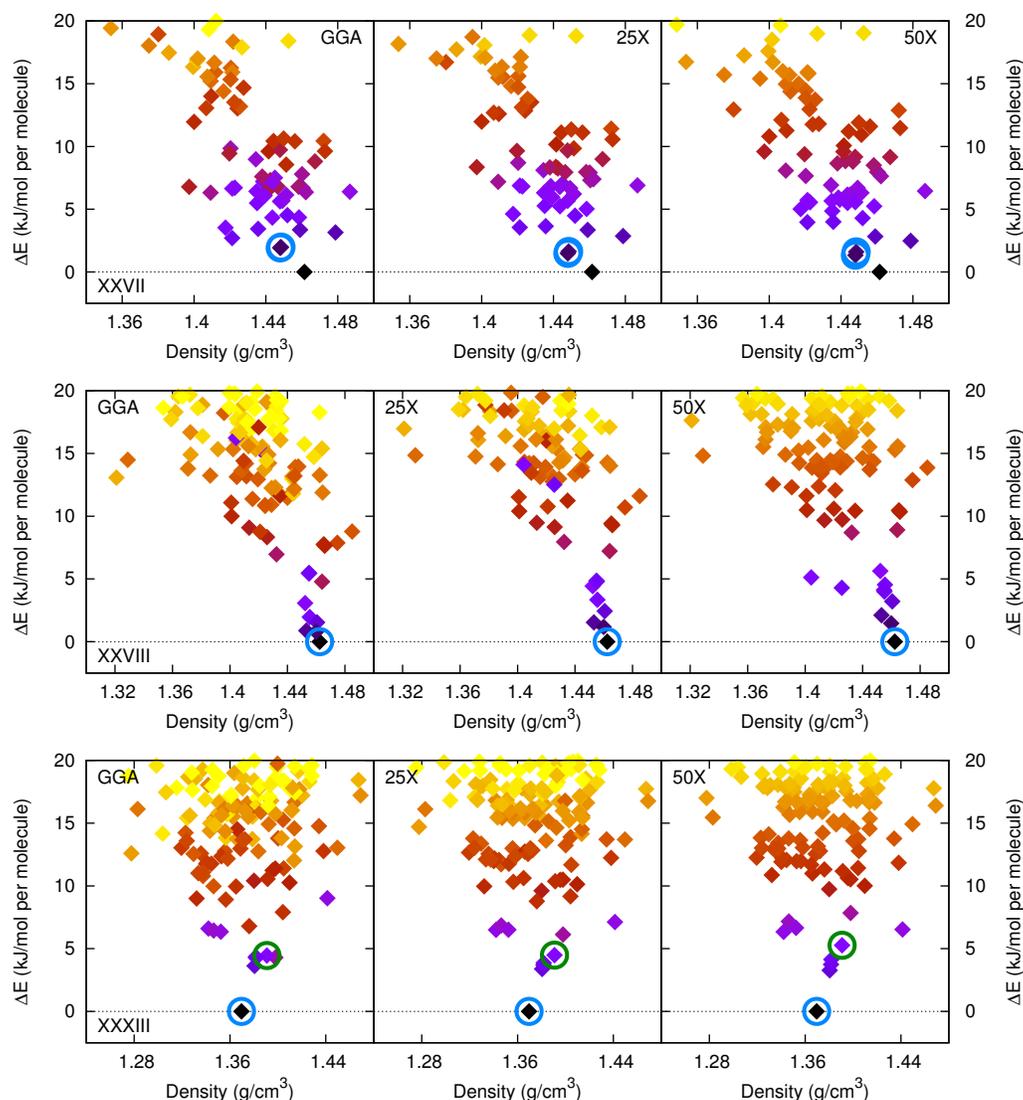
Figure 8.2: Computed crystal energy landscapes for compounds XXVII, XXVIII, and XXXIII from the 7th CSP blind test, where DFT performs well. The blue circles indicate the structure that is the best match to the experimental polymorph. For compound XXXIII, a second, less-stable polymorph (form A) was isolated and that structure is indicated by the green circles.

we had expected that delocalization error may significantly affect the crystal energy landscapes for this compound, similar to compound XIX from the 5th blind test,[384] we instead found that all three functionals considered again consistently predicted form B to be the lowest-energy structure on each landscape. Form A was either the 5th or 6th ranked structure, depending on functional, lying ca. 5 kJ/mol in energy above form B. Thus, our XDM-corrected functionals all performed well, consistently generating reasonable crystal energy landscapes for these three compounds.

The good performance seen so far can be contrasted with the results for the flexible molecules, compounds XXXI and XXXII. The computed crystal energy landscapes for these two compounds are shown in Figure 8.3. As both problematic compounds contain fluorine, it is possible that the light basis set used here is not able to provide a sufficiently accurate description of fluorine and the effect of using the tight basis set for this element should be investigated in future work.

For compound XXXI, three polymorphs were found experimentally. Of these, form B was found to be the most stable, while form A was found to be disordered with major and minor components and was the next most stable. Finally, a less stable and very low-density form C was identified, possessing large crystal voids and was likely formed from solvent evaporation from a solvate. As expected, form C was found to be a very high energy structure, ranked 61-76 on the computed crystal energy landscapes and similar results have been seen for analogous low-density structures obtained from solvates.[392,393] It is notable that our DFT calculations reversed the energy ordering of the other two experimental forms, predicting both major and minor components of form A to be more stable than form B. Additionally, form B and both the major and minor components of form A are all surprisingly high energy structures for isolable polymorphs, lying 1-6 kJ/mol above the DFT-predicted minimum, depending on functional. The specific energy rankings of each experimental form, with each functional, are summarised in Table 8.1. The GGA was found to give the best energy ordering for compound XXXI, providing lower rankings for forms A and B than seen with the hybrid functionals.

The high relative energies obtained for the experimental polymorphs for compound XXXI are reminiscent of what was seen previously for compound XX in the 6th blind test.[384] Since compound XXXI is a flexible molecule with three rotatable bonds, it is possible that the thermal free energy corrections are significantly larger[342] than the 0-2 kJ/mol typically seen for small, rigid molecules,[380,385,386] and are sufficient to alter the ranking. However, it should be noted that the individual molecules in Form B, the major component of Form A, and the predicted minimum-energy structures all have the same conformation, making this possibility less likely. For form A, the additional entropy contribution from the disorder arising from the major and minor components may also provide significant stability, although this does not explain the greater stability of form B seen experimentally. It is also possible that rugosity[387] or other kinetic effects may explain why the DFT predicted minimum-energy structure was not observed experimentally and perhaps it could be isolated under different crystallisation conditions.
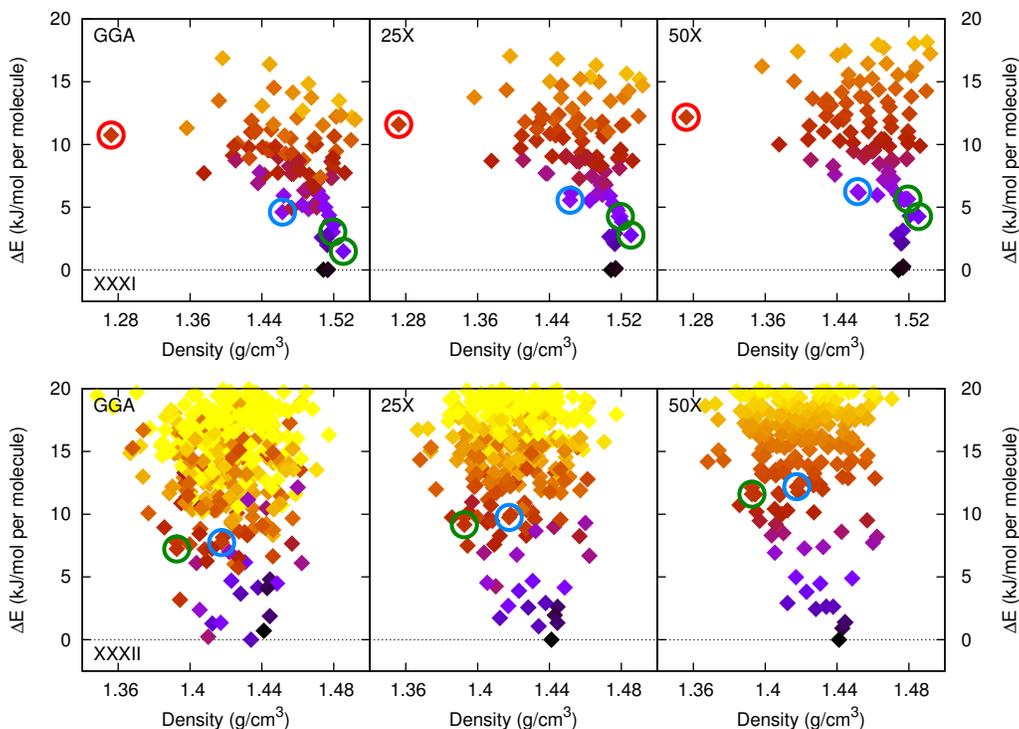
Figure 8.3: Computed crystal energy landscapes for the two flexible molecules, XXXI and XXXII, from the 7th CSP blind test, where DFT performs poorly. The blue circles indicate the structure that is the best match to the most stable experimental polymorph, form B, while the green circles indicate the structure of a second, less-stable polymorph, form A (major and minor components are shown for compound XXXI). For compound XXXI, an additional low-density form with large crystal voids, indicated by the red circles, was identified after removal of solvent from a solvate structure.

Finally, for compound XXXII, the computed crystal energy landscapes are also shown in Figure 8.3. This molecule is extremely flexible, with four rotatable bonds. Multiple polymorphs were observed experimentally, with single-crystal x-ray structures being obtained for two forms, and several other forms only able to be characterised via powder x-ray diffraction.[391] Of the two characterised polymorphs, form B was found to be the more stable. Form A was again disordered, and composed of both major and minor components, although only the major component was represented in the set of 500 supplied candidate structures. Our XDM-corrected methods were found to perform quite poorly for this compound, predicting the experimental polymorphs to lie quite high in energy, from 7-12 kJ/mol above the minimum depending on the functional. Again, somewhat surprisingly, the best performance was obtained with the XDM-corrected GGA functional, which predicts the two experimental forms to lie 7-8 kJ/mol higher in energy than the

minimum. This energy range is fairly comparable to the result obtain previously for the similarly large and flexible compound XX.[384] It is likely that neglect of vibrational effects is at least partly responsible for the poor energy ranking, since both experimental polymorphs have different conformations from the minimum-energy structures identified with DFT. However, it is possible that both of these forms are meta-stable and one or more of the experimental structures that could only be characterised by PXRD is actually more stable. It would be very interesting to determine if any of the low-energy DFT structures match the experimental powder patterns and this will be the subject of future work once these powder diffractograms are released by the CCDC.

8.4 SUMMARY

This chapter presented the results from the energy ranking phase of the 7th CSP blind test. This work relied on the application of the XDM dispersion model using FHI-aims to rank the putative crystal structures provided by the CCDC to the investigators, as we could not have completed all the calculations within the allotted time with our available resources using planewave DFT. The functionals considered were the B86bPBE based GGA and its hybrid counterparts with 25% and 50% exact exchange mixing. For compounds XXVII, and XXVIII, all functionals considered performed equally well, ranking the experimental structure as the lowest in energy in the case of XXVIII, and the second lowest (and within 2 kJ/mol of the minimum) for XXVII. For compound XXXIII, there were two experimentally isolated polymorphs, where form B was found to be the more stable. The crystal energy landscape for this compound was expected to exhibit significant delocalization error with the GGA; however, as all functionals ranked form B as the minimum-energy structure, this was not observed to be the case. Form A of compound XXXIII was found be less stable than form B by ca. 5 kJ/mol.

The high accuracy seen for compounds XXVII, XXVIII, and XXXIII was unfortunately not observed for compounds XXXI and XXXII. Compound XXXI was reported to have 3 experimentally isolable polymorphs. DFT predicted surprisingly high energy structures for form B and both major and minor components of the disordered form A and, when compared with the experimental stabilities, the ranking between the forms B and A was found to be reversed. Additionally, the GGA was found to provide lower relative energies for the known polymorphs compared to the hybrid functionals. For compound XXXII, a much larger molecule with 4 rotatable bonds, multiple polymorphs were observed experimentally but only two single-crystal XRD structures could be obtained. The XDM-corrected

methods performed quite poorly for this compound, with the two characterised experimental polymorphs found to lie quite high in energy (7-12 kJ/mol from the DFT predicted minima), with the GGA functional again outperforming the hybrids. A possible source of error is the neglect of vibrational and anharmonic effects as the higher flexibility of these two molecules, and their abilities to adopt different conformations, could potentially lead to large thermal free-energy corrections. Another source of error may be the use of an insufficiently large basis set for fluorine. It is also possible that one or more of the low-energy structures predicted from DFT results would match other polymorphs that were unable to be characterised via single-crystal methods. Overall, the XDM-based/DFT methods performed better on the smaller more rigid molecules and failed drastically for compounds that showed a high number of rotatable bonds, in line with the results from the previous Chapter 7 in this thesis.

CHAPTER 9

CONCLUSIONS AND FUTURE WORK

9.1 CONCLUSIONS

The work presented in this thesis aimed to showcase a new implementation of the XDM dispersion model within the FHI-aims code. We began with a discussion of the theory that is used within this document: the underpinnings of DFT, periodic boundary conditions, various types of basis sets, the XDM model itself, and alternative and competing post-SCF dispersion models. We then discussed what goes into the accurate selection and pairing of a density functional with a dispersion correction, as not all functionals are created equal. Next, we applied XDM and planewave/PAW DFT to the modelling of a surface-based experimental catalytic system through the use of the Quantum ESPRESSO code. Highly accurate calculations can be achieved through this approach, but we are strongly limited in terms of system size by the high computational cost. The shortcomings of the planewave/PAW-based approach led to an investigation of alternative methods for solid-state DFT and lead to the decision to use the FHI-aims code, which employs NAO basis sets of finite extent. This was proposed to be a good platform for implementation of XDM as the use of NAOs would allow for the application of XDM to larger and more complex systems, particularly with hybrid functionals. Once XDM was implemented within FHI-aims, it was benchmarked using a suite of common data sets for post-SCF dispersion corrections and was determined to be one of the most accurate methods available for predicting lattice energies of molecular crystals. The FHI-aims implementation of XDM was then applied to a series of compounds, including pharmaceutical analogues, that have made up the previous 6 blind tests of crystal structure prediction (CSP) organised by the Cambridge Crystallographic Data Centre (CCDC). This work illustrated the importance of using a functional/dispersion correction pair that minimises delocalization error, and is as

efficient as possible.

Despite the importance of dispersion to the correct and full understanding of chemical processes, it has not been trivial to include dispersion physics within quantum chemical simulations. This is due to the fact that dispersion physics is missing from mean-field electronic structure methods, such as Hartree-Fock and most density functionals. Much work has been done over the last 20 years in an effort to remedy this through the development of dispersion models, such as the D2 and TS methods, that provide simplified and general prescriptions to obtain the leading order C_6 term. A drawback of these methods is that they have, at best, only minimal dependence of the dispersion coefficients on the chemical environment. More complex methods have since been developed, such as the D3 and D4 methods, XDM, and MBD, each of which is of the post-SCF variety of dispersion corrections, where the dispersion coefficients have more sophisticated environment dependence. However, as discussed in Chapter 3, the dispersion correction alone does not give the complete picture when attempting to obtain higher accuracy from DFT-based calculations. When using modern dispersion corrections, particularly XDM, the choice of base density functional should be held to three key points: the functional should be dispersionless, numerically stable, and have minimal delocalization error. For pairing with XDM, the choice of a dispersionless functional is made to prevent a kind of double counting of dispersion at short range. This means that the treatment of non-bonded repulsion is left entirely to the base functional and dispersion effects to the correction. It can be seen within Chapter 3, and in the rest of this thesis, that this division of labour is key to obtaining the highest accuracy when pairing XDM with particular density functionals. Next, there are both practical and scientific implications for the numerical sensitivity constraint. Functionals that do not adhere to the constraint of numerical sensitivity are meta-GGAs where one or more of the τ-dependent terms show divergence at the bond critical points of van der Waals dimers. Here, using the very large integration grids required to obtain a smooth potential results in a significantly increased computational cost relative to GGAs, with no gains in overall accuracy. The final constraint, which is also a key motivation for the work described in the later chapters within this thesis, is the selection of a functional that minimises delocalization error. Until recently, a user would be restricted to GGA functionals (which have significant inherent delocalization error) for most solid-state systems, due to the prohibitive cost of evaluating the exact exchange energy with planewave-based codes. This prevented the application of XDM-corrected DFT to study certain classes of chemical problems with any expectation of reliability. The conclusions within Chapter 3

were fundamental to all of the method choices that were made for the work that followed in this thesis.

The computational modeling of STM images is a powerful tool for the understanding of atomic-level features for surface chemistry, as the explicit assignment of the images to the exact chemical species is difficult experimentally. The work present within Chapter 4 employed the B86bPBE functional with the XDM dispersion correction using a planewave/projector augmented-wave (PAW) method that, up until the work presented within this thesis, was the standard platform for applying XDM to solid-state systems. The resultant computational STM images were then compared to the experimental results obtained by Burke and colleagues. Dispersion largely determines the optimal geometry, as well as the relative stabilities, for adsorbates on a surface. The geometry, in turn, affects the appearance of the theoretical STM plots. Although conclusions could be drawn from this work to unambiguously assign chemical structures to the observed STM data, a key caveat of the method used was the requirement to truncate the modeled system to just the active site, to make the calculations tractable. Even with this truncation, the size of the unit cell used within Chapter 4 approached the limit of what was possible with the compute resources available at the time of publication. This led to the exploration of alternative methods which might improve the efficiency of solid-state calculations with XDM-corrected DFT without a marked reduction in accuracy.

The issues of delocalization error and limitations on unit-cell size lead to the search for alternative programs in which XDM could be implemented to allow for the investigation of larger and more complex systems than possible with planewave DFT. The selection of the FHI-aims code was made due to its use of NAOs of finite extent, which led to the ability to perform routine calculations on significantly larger unit cells, and to pivot to the use of hybrid density functionals. The implementation of XDM within FHI-aims was detailed in Chapter 5, along with the results of preliminary tests, comparing both the exchange-hole dipole moments and the homoatomic C_6 dispersion coefficients with results from previous implementations in NUMOL, Quantum ESPRESSO, and Gaussian/postg. As XDM contains two parameters that go into the damping function that are functional and basis set dependent, Chapter 6 presents optimal damping parameters for a set of common functionals, including hybrids. Following the determination of the damping parameters, a common set of benchmark calculations were undertaken and XDM was compared against two competing post-SCF dispersion methods, TS and XDM. The XDM-corrected hybrid functionals presented within Chapter 6 provided the lowest mean absolute errors for molecular crystal

lattice energies across all combinations of functionals and dispersion methods considered both by ourselves and within the broader literature.

Accurate and efficient computation of relative energies of molecular crystal polymorphs is a problem of central importance for solid-state pharmaceuticals. DFT has emerged as the preeminent tool within crystal structure prediction, but previous implementations of these methods have been hindered by poor scaling and delocalization error, as detailed in this work. Within Chapter 7, XDM-corrected functionals implemented within FHI-aims were shown to provide an accurate and efficient route of candidate crystal structures for CSP. However, we saw that GGA calculations alone are not enough for the full description of all molecular polymorphs, and the use of hybrid functionals was shown to be key for some classes of compounds, such as organic salts, that suffer from delocalization error. Finally, the implementation of XDM within FHI-aims was critical to the timely completion of the calculations required for participation in the CCDC 7th Blind Test, as detailed in Chapter 8. For compounds XXVII, XXVIII, and XXXIII, our XDM-corrected DFT methods performed extremely well, predicting the most stable experimental polymorph to be the lowest or second-lowest energy structure of the provided candidates. For compounds XXXI and XXXII, the XDM-corrected functionals did not fair as well, perhaps due to thermal free-energy corrections or kinetic effects being more significant for highly flexible molecules. Notably, delocalization error did not appear to be a major contributor to any of the crystal energy landscapes.

In summary, the implementation of XDM within FHI-aims has been shown to set new standards for both accuracy and efficiency for DFT calculations on molecular crystals and has unlocked the ability to address some of the shortcomings of GGA functionals for use in CSP. Going forward, it will facilitate high-throughput computational studies of a wide range of new and exciting chemical systems, some of which will be detailed in Section 9.2.

9.2 FUTURE WORK

There are a wide array of projects still to complete within the scope of this work since the implementation of FHI-aims enables routine XDM-corrected DFT calculations to be performed on larger and more complex systems. One such area where the efficiency and accuracy of FHI-aims will be useful is for the investigation of the errors seen within Chapter 7 and 8 for compounds XX, XXXI, and XXXII in the 6th[87] and 7th blind tests of molecular crystal structure predictions. While we conjectured that these errors are likely due to neglect of thermal effects, this can be directly probed by FHI-aims through the

calculation of the phonons and resulting thermal free-energy corrections for the respective compounds, which would be prohibitively costly with plane-wave DFT.

Another potential application of the FHI-aims XDM implementation is the study of transition-metal dichalcogenide (TMDC) heterojunctions. TMDCs are 2D layered materials with the general chemical formula MX_2, where M is a transition metal, such as Mo or W, and X is a chalocogen (group 16).[394] An extremely common example of a TMDC is MoS_2, which has a trigonal prismatic 2H hexagonal structure. TMDCs are isostructural analogs of graphene, where the 2D layers are bound by weak van der Waals interactions, while the metal atoms are sandwiched between two chalcogenide layers through strong covalent bonding.[394] As a result, charge transport in TMDCs occurs efficiently in-plane through the single layers, while there exists a tunneling barrier for charge transport between layers due to the weak van der Waals interactions and large interlayer spacing.[395] TMDCs are semiconductors and have been proposed to have applications in electronics, batteries, and as catalysts. Focusing on electronics applications, TMDCs are proposed to be the future of classical computing, with the aim of replacing Si within computing devices and extending Moore's law.[396–399] However, there exist barriers to their wide adoption in the form of high contact resistance between the TMDC material and required metal contact, common within transistors.[400–403] Computational work in this space has primarily used density-functional methods such as PBE-D2 and the LSDA.[403–405] DFT studies of TMDC-metal interfaces can be computationally expensive due to the need for large unit cells in order to ensure commensurate lattices. Previous studies of metal-graphite interfaces have already employed unit cells that are on the order of 100s of atoms.[271,406,407] As we have shown in Fig. **??**, this is already in the regime where significant time savings can be achieved with NAO basis sets as opposed to plane waves. Additionally, the XDM correction has been shown to provide improved accuracy relative to other post-SCF dispersion corrections both for properties of bulk TMDCs[16] and for graphene-metal interfaces.[406]

Another potential application of the work within this thesis is the computational modeling of crystal growth, important to understanding the origins of the chiral selectivity found within all living things.[408–410] Previous work considered hydroxylated α-quartz as a seed crystal, present within the early earth, that may have caused an enrichment of D or L amino acids within the local environment.[408–414] This work was limited to the placement of single amino acid residues on the hydroxylated α-quartz surface and, although an important first step, is far too simplistic to draw conclusions regarding crystal growth.[413–415] Quantum mechanical study of these systems is key due to the minute adsorption-energy

differences between enantiomers, which might not be resolved through alternative methods. Much like the TMDC heterojunctions mentioned above, explicit modeling of the interface between an enantiopure amino acid crystal and a hydroxylated α-quartz surface will require large unit cells to ensure commensurate lattices. The FHI-aims XDM implementation will provide the accuracy and speed needed to model such interfaces and will allow more in-depth study of crystal growth of amino acids and other biological precursor molecules on chiral substrates than might otherwise be intractable. The computational work could then be combined with experiments to draw conclusions about the emergence of preferential chirality in early life on earth and possible life on other planets.

Another piece of future work is the implementation of XDM into other electronic structure packages, allowing for a wider array of users to access the method. One example is the Amsterdam Density Functional (ADF) code that, like many other DFT codes, has a wide user base with interests in a range of chemistry and material science applications. One of ADF's headline features is energy decomposition analysis (EDA),[416,417] which is a quantitative analysis method used to partition the energy for a chemical bonding interaction obtained from electronic structure calculations into chemically intuitive components. EDA is a widely used tool for the investigation of bonding in inorganic and organometallic compounds and provides a quantitative picture for the traditional understanding of chemical bonding, providing a bridge between synthetic chemistry and the quantum mechanical world. Within EDA, the bond energy can be represented as

$$\Delta E_{\text{bond}} = \Delta E_{\text{strain}} + \Delta E_{\text{int}}, \tag{9.1}$$

where the ΔE_{strain} is the strain energy involved in the deformation of the geometry into the pro-molecule and ΔE_{int} is the bonding interaction energy between fragments represented by

$$\Delta E_{\text{int}} = V_{\text{elestat}} + \Delta E_{\text{Pauli}} + \Delta E_{\text{orb}} + \Delta E_{\text{disp}}. \tag{9.2}$$

V_{elestat} represents the electrostatic attraction, ΔE_{Pauli} is the Pauli repulsion energy, ΔE_{orb} represents the stabilising orbital interactions or mixing and charge transfer between fragments, and ΔE_{disp} is the dispersion energy.[220,417] EDA relies on the accurate treatment of the geometry of compounds, which in turn relies on the accurate treatment of dispersion interactions, and these are also a key component of the interaction energy. The inclusion of XDM within ADF will allow for a new tool for users to apply XDM within EDA for molecular systems and will also allow direct comparison with the new Grimme-D4 correction

within the code base.

Although DFT is a widely used and a broadly successful tool for the modeling of matter, it still has a major outstanding challenge facing it today in the form of delocalization error,[349] as discussed within this work. Delocalization error has been categorised broadly within the literature and given many different names, but ultimately all are different manifestations of the same issue.[175,202,204,349,418,419] It can be seen in even some of the simplest chemical systems, such as the H_2^+ molecular ion near its dissociation limit.[232,233,420–423] Here, the molecule is greatly over-stabilized due to the highly delocalized nature of the single electron, shared equally between the two hydrogen atoms. Other examples include systems with extended conjugation[225–227,229,350,424–428] and charge-transfer complexes.[206,208–210] Delocalization error comes about due to the inherently local nature of the construction of practical DFT methods, which are unable to "see" beyond the local environment. Delocalization error can go from being a small nuisance to giving entirely incorrect results that make drawing reasonable conclusions from DFT impossible. There are prescriptions for fixing this issue through the inclusion of a portion of the exact exchange energy calculated from Hartree-Fock theory.[1] However, depending on the system, it may require anywhere from 0-100% exact exchange for high accuracy and, in some cases, any answer you wish for a particular property may be obtained based on the amount of exact exchange included in the calculation.[247] This has led to the proliferation of large numbers of DFT methods targeted for specific applications and having no single "one size fits all" choice that is consistently reliable for all classes of chemical systems.[1,2,349,429–434] In order for DFT to continue to make materials predictions and be used beyond well-understood chemistries, a more general method that does not rely on empirical fitting is needed. One promising alternative is the B05 method, developed by Axel Becke.[233,435] It is non-empirical in design and is exact for any one-electron system, such as the aforementioned H_2^+ example. However, due to the complexity of B05, this method has not been widely adopted and is not generally available in computational codes beyond the specialised software used for its development. As a result, the B05 functional has never been applied to large molecules or to solid-state systems. The efficiency seen within FHI-aims for HF-based exact exchange provides an ideal platform for the implementation of B05 and this would be a major boost to the computational tools available for solid-state applications.

Finally, within the scope of improving XDM, work can be done to address shortcomings in the way the atomic forces and stress tensor are calculated. XDM currently relies on

an approximation made in the analytical formula used to calculate the dispersion contribution to the forces and stresses, where the dispersion coefficients are taken to be constants, independent of the atomic positions. For traditional geometry optimisation of finite molecules and molecular crystals, this has been shown not to have detrimental effects on the overall accuracy of the XDM model, although errors have been observed for inorganic solids.[348] However, for *ab initio* molecular dynamics, this force-energy inconsistency is posited to be the source of a potential energy leak within the simulations, rendering XDM useless for this application in its current state. A full rigorous analytical definition of the XDM force and stress equations, allowing for the dispersion coefficients to evolve with the atom positions, is therefore necessary before any dynamics work can be undertaken with confidence. Addressing this shortcoming within XDM would push it to the forefront of dispersion corrections, allowing XDM to cover the greatest chemical space with the highest accuracy compared to competing methods.

BIBLIOGRAPHY

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

[2] Burke, K. Perspective on Density Functional Theory. *J. Chem. Phys.* **2012**, *136*, 150901.

[3] Hohenberg, P.; Kohn, W. Inhomogeneous Electron Gas. *Phys. Rev.* **1964**, *136*, B864–B871.

[4] Kohn, W.; Sham, L. J. Self-Consistent Equations Including Exchange and Correlation Effects. *Phys. Rev.* **1965**, *140*, A1133–A1138.

[5] Sherrill, C. D. Frontiers in Electronic Structure Theory. *J. Chem. Phys.* **2010**, *132*, 110902.

[6] Casely, I. J.; Ziller, J. W.; Fang, M.; Furche, F.; Evans, W. J. Facile Bismuth-Oxygen Bond Cleavage, C-H Activation, and Formation of a Monodentate Carbon-Bound Oxyaryl Dianion,$(C_6H_2tBu_2 − 3, 5−O−4)^{2−}$. *J. Am. Chem. Soc.* **2011**, *133*, 5244–5247.

[7] Mo, Y.; Ong, S. P.; Ceder, G. First-Principles Study of the Oxygen Evolution Reaction of Lithium Peroxide in the Lithium-air Battery. *Phys. Rev. B* **2011**, *84*, 205446.

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

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

[10] Johnson, E. R. In *Non-covalent Interactions in Quantum Chemistry and Physics*; Otero-de-la-Roza, A., DiLabio, G. A., Eds.; Elsevier, 2017; Chapter 5, pp 169–194.

[11] Zacharia, R.; Ulbricht, H.; Hertel, T. Interlayer Cohesive Energy of Graphite from Thermal Desorption of Polyaromatic Hydrocarbons. *Phys. Rev. B* **2004**, *69*, 155406.

[12] London, F. On the Theory and Systematic of Molecular Forces. *Z. Physik* **1930**, *63*, 245–279.

[13] Stone, A. *The Theory of Intermolecular Forces*; Oxford University Press, 2013.

[14] Kolar, M.; Kubar, T.; Hobza, P. On the Role of london Dispersion Forces in Biomolecular Structure Determination. *J. Phys. Chem. B* **2011**, *115*, 8038–8046.

[15] 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.

[16] Otero-de-la-Roza, A.; LeBlanc, L. M.; Johnson, E. R. Pairwise Dispersion Corrections Can Describe Layered Materials Accurately. *J. Phys. Chem. Lett.* **2020**, *11*, 2298–2302.

[17] Liu, W.; Tkatchenko, A.; Scheffler, M. Modeling Adsorption and Reactions of Organic Molecules at Metal Surfaces. *Acc. Chem. Res.* **2014**, *47*, 3369–3377.

[18] Micoulaut, M.; Piarristeguy, A.; Flores-Ruiz, H.; Pradel, A. Towards Accurate Models for Amorphous GeTe: Crucial Effect of Dispersive van der Waals Corrections on the Structural Properties Involved in the Phase-change Mechanism. *Phys. Rev. B* **2017**, *96*, 184204.

[19] Moellmann, J.; Grimme, S. Importance of London Dispersion Effects for the Packing of Molecular Crystals: A Case Study for Intramolecular Stacking in A Bis-thiophene Derivative. *Phys. Chem. Chem. Phys.* **2010**, *12*, 8500–8504.

[20] Otero-de-la-Roza, A.; DiLabio, G. *Non-Covalent Interactions in Quantum Chemistry and Physics*; Elsevier: Amsterdam, Netherlands, 2017.

[21] Otero-de-la-Roza, A.; LeBlanc, L. M.; Johnson, E. R. What Is "Many-Body" Dispersion and Should I Worry About It? *Phys. Chem. Chem. Phys.* **2020**, *22*, 8266–8276.

[22] Salem, L. The Calculation of Dispersion Forces. *Mol. Phys.* **1960**, *3*, 441–452.

[23] Dalgarno, A.; Davison, W. D. The Calculation of van der Waals Interactions. *Adv. At. Mol. Phys.* **1966**, *2*, 1–32.

[24] Atkins, P. W.; Friedman, R. S. *Molecular Quantum Mechanics*; Oxford University Press, 2011; Chapter 4, pp 414–417.

[25] Unsöld, A. Quantentheorie des Wasserstoffmolekülions und der Born-Landéschen Abstoßungskräfte. *Z. Phys.* **1927**, *43*, 563–574.

[26] Jones, J. E. On the Determination of Molecular Fields.—II. From the Equation of State of a Gas. *Proc. R. Soc. Lond.* **1924**, *106*, 463–477.

[27] Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; MacKerell Jr., A. D. CHARMM General Force Field: A Force Field for Drug-like Molecules Compatible with the CHARMM All-Atom Additive Biological Force Fields. *J. Comput. Chem.* **2010**, *31*, 671–690.

[28] Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and Testing of a General Amber Force Field. *J. Comput. Chem.* **2004**, *25*, 1157–1174.

[29] Jorgensen, W. L.; Tirado-Rives, J. Potential Energy Functions for Atomic-Level Simulations of Water and Organic and Biomolecular Systems. *Proc. Natl. Acad. Sci.* **2005**, *102*, 6665–6670.

[30] Grimme, S. Semiempirical GGA-type Density Functional Constructed with a Long-range Dispersion Correction. *J. Comput. Chem.* **2006**, *27*, 1787–1799.

[31] 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.

[32] Grimme, S.; Ehrlich, S.; Goerigk, L. Effect of the Damping Function in Dispersion Corrected Density Functional Theory. *J. Comput. Chem.* **2011**, *32*, 1456–1465.

[33] Caldeweyher, E.; Ehlert, S.; Hansen, A.; Neugebauer, H.; Spicher, S.; Bannwarth, C.; Grimme, S. A Generally Applicable Atomic-charge Dependent London Dispersion Correction. *J. Chem. Phys.* **2019**, *150*, 154122.

[34] Caldeweyher, E.; Mewes, J.-M.; Ehlerta, S.; Grimme, S. Extension and Evaluation of the D4 London-dispersion Model for Periodic Systems. *Phys. Chem. Chem. Phys.* **2020**, *22*, 8499–8512.

[35] 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.

[36] Tkatchenko, A.; DiStasio, R. A.; Car, R.; Scheffler, M. Accurate and Efficient Method for Many-Body van der Waals Interactions. *Phys. Rev. Lett.* **2012**, *108*, 236402.

[37] 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.

[38] Steinmann, S. N.; Csonka, G.; Corminboeuf, C. Unified Inter-and Intramolecular Dispersion Correction Formula for Generalized Gradient Approximation Density Functional Theory. *J. Chem. Theory Comput.* **2009**, *5*, 2950–2958.

[39] Steinmann, S. N.; Corminboeuf, C. A System-dependent Density-based Dispersion Correction. *J. Chem. Theory Comput.* **2010**, *6*, 1990–2001.

[40] Steinmann, S. N.; Corminboeuf, C. A Generalized-gradient Approximation Exchange Hole Model for Dispersion Coefficients. *J. Chem. Phys.* **2011**, *134*, 044117.

[41] Steinmann, S. N.; Corminboeuf, C. Comprehensive Benchmarking of a Density-dependent Dispersion Correction. *J. Chem. Theory Comput.* **2011**, *7*, 3567–3577.

[42] 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.

[43] Lee, K.; Murray, É. D.; Kong, L.; Lundqvist, B. I.; Langreth, D. C. Higher-accuracy van der Waals Density Functional. *Phys. Rev. B* **2010**, *82*, 081101.

[44] DiLabio, G. A. Accurate Treatment of van der Waals Interactions Using Standard Density Functional Theory Methods with Effective Core-type Potentials: Application to Carbon-containing Dimers. *Chem. Phys. Lett.* **2008**, *455*, 348–353.

[45] Mackie, I. D.; DiLabio, G. A. Interactions in Large, Polyaromatic Hydrocarbon Dimers: Application of Density Functional Theory with Dispersion Corrections. *J. Phys. Chem. A* **2008**, *112*, 10968–10976.

[46] Torres, E.; DiLabio, G. A. A (Nearly) Universally Applicable Method for Modeling Noncovalent Interactions Using B3LYP. *J. Phys. Chem. Lett.* **2012**, *3*, 1738–1744.

[47] DiLabio, G. A.; Koleini, M.; Torres, E. Extension of the B3LYP–dispersion-correcting Potential Approach to the Accurate Treatment of Both Inter-and Intra-molecular Interactions. *Theor. Chem. Acc.* **2013**, *132*, 1389.

[48] Von Lilienfeld, O. A.; Tavernelli, I.; Rothlisberger, U.; Sebastiani, D. Optimization of Effective Atom Centered Potentials for London Dispersion Forces in Density Functional Theory. *Phys. Rev. Lett.* **2004**, *93*, 153004.

[49] Lin, I.-C.; Coutinho-Neto, M. D.; Felsenheimer, C.; von Lilienfeld, O. A.; Tavernelli, I.; Rothlisberger, U. Library of Dispersion-Corrected Atom-centered Potentials for Generalized Gradient Approximation Functionals: Elements H, C, N, O, He, Ne, Ar, and Kr. *Phys. Rev. B* **2007**, *75*, 205131.

[50] DiLabio, G. A.; Otero-de-la Roza, A. Noncovalent Interactions in Density Functional Theory. *Rev. Comput. Chem.* **2016**, *29*, 1–97.

[51] 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.

[52] Vydrov, O. A.; Van Voorhis, T. Nonlocal van der Waals Density Functional Made Simple. *Phys. Rev. Lett.* **2009**, *103*, 063004.

[53] Vydrov, O. A.; Van Voorhis, T. Nonlocal van der Waals Density Functional: The Simpler the Better. *J. Chem. Phys.* **2010**, *133*, 244103.

[54] Sabatini, R.; Gorni, T.; de Gironcoli, S. Nonlocal van der Waals Density Functional Made Simple and Efficient. *Phys. Rev. B* **2013**, *87*, 041108.

[55] Vydrov, O. A.; Van Voorhis, T. Benchmark Assessment of the Accuracy of Several van der Waals Density Functionals. *J. Chem. Theory Comput.* **2012**, *8*, 1929–1934.

[56] Langreth, D.; Lundqvist, B. I.; Chakarova-Käck, S. D.; Cooper, V.; Dion, M.; Hyldgaard, P.; Kelkkanen, A.; Kleis, J.; Kong, L.; Li, S.; Moses, P. G.; Murray, E.; Puzder, A.; Rydberg, H.; Schröder, E.; Thonhauser, T. A Density Functional for Sparse Matter. *J. Phys. Condens. Matter* **2009**, *21*, 084203.

[57] Zhao, Y.; Schultz, N. E.; Truhlar, D. G. Exchange-correlation Functional with Broad Accuracy for Metallic and Nonmetallic Compounds, Kinetics, and Noncovalent Interactions. *J. Chem. Phys.* **2005**, *123*, 161103.

[58] Zhao, Y.; Truhlar, D. G. A New Local Density Functional for Main-Group Thermochemistry, Transition Metal Bonding, Thermochemical Kinetics, and Noncovalent Interactions. *J. Chem. Phys.* **2006**, *125*, 194101.

[59] Zhao, Y.; Truhlar, D. G. The M06 Suite of Density Functionals for Main Group Thermochemistry, Thermochemical Kinetics, Noncovalent Interactions, Excited States, and Transition Elements: Two New Functionals and Systematic Testing of Four M06-Class Functionals and 12 Other Functionals. *Theor. Chem. Acc.* **2008**, *120*, 215–241.

[60] Peverati, R.; Zhao, Y.; Truhlar, D. G. Generalized Gradient Approximation That Recovers the Second-order Density-gradient Expansion with Optimized Across-the-board Performance. *J. Phys. Chem. Lett.* **2011**, *2*, 1991–1997.

[61] Peverati, R.; Truhlar, D. G. Exchange–correlation Functional With Good Accuracy for Both Structural and Energetic Properties While Depending Only on the Density and Its Gradient. *J. Chem. Theory Comput.* **2012**, *8*, 2310–2319.

[62] Bučko, T.; Lebègue, S.; Gould, T.; Ángyán, J. G. Many-body Dispersion Corrections for Periodic Systems: An Efficient Reciprocal Space Implementation. *J. Phys. Condens. Matter* **2016**, *28*, 045201.

[63] Becke, A. D.; Johnson, E. R. Exchange-hole Dipole Moment and the Dispersion Interaction Revisited. *J. Chem. Phys.* **2007**, *127*, 154108.

[64] Blum, V.; Gehrke, R.; Hanke, F.; Havu, P.; Havu, V.; Ren, X.; Reuter, K.; Scheffler, M. Ab Initio Molecular Simulations with Numeric Atom-Centered Orbitals. *Comp. Phys. Comm.* **2009**, *180*, 2175–2196.

[65] Havu, V.; Blum, V.; Havu, P.; Scheffler, M. Efficient O(N) Integration for All-Electron Electronic Structure Calculation Using Numeric Basis Functions. *J. Comput. Phys.* **2009**, *228*, 8367–8379.

[66] Ren, X.; Rinke, P.; Blum, V.; Wieferink, J.; Tkatchenko, A.; Sanfilippo, A.; Reuter, K.; Scheffler, M. Resolution-of-identity Approach to Hartree–Fock, Hybrid Density Functionals, RPA, MP2 and GW with Numeric Atom-centered Orbital Basis Functions. *New J. Phys.* **2012**, *14*, 053020.

[67] Marek, A.; Blum, V.; Johanni, R.; Havu, V.; Lang, B.; Auckenthaler, T.; Heinecke, A.; Bungartz, H.-J.; Lederer, H. The ELPA Library: Scalable Parallel Eigenvalue Solutions for Electronic Structure Theory and Computational Science. *J. Phys. Condens. Matter* **2014**, *26*, 213201.

[68] Levchenko, S.; Ren, X.; Wieferink, J.; Johanni, R.; Rinke, P.; Blum, V.; Scheffler, M. Hybrid Functionals for Large Periodic Systems in an All-Electron, Numeric Atom-Centered Basis Framework. *Comput. Phys. Comm.* **2015**, *192*, 60–69.

[69] Huhn, W. P.; Blum, V. One-Hundred-Three Compound Band-Structure Benchmark of Post-Self-Consistent Spin-Orbit Coupling Treatments in Density Functional Theory. *Phys. Rev. Mater.* **2017**, *1*, 033803.

[70] Knuth, F.; Carbogno, C.; Atalla, V.; Blum, V.; Scheffler, M. All-electron Formalism for Total Energy Strain Derivatives and Stress Tensor Components for Numeric Atom-Centered Orbitals. *Comput. Phys. Commun.* **2015**, *190*, 33–50.

[71] Junquera, J.; Paz, O.; Sánchez-Portal, D.; Artacho, E. Numerical Atomic Orbitals for Linear-Scaling Calculations. *Phys. Rev. B* **2001**, *64*, 235111.

[72] Dovesi, R.; Orlando, R.; Erba, A.; Zicovich-Wilson, C. M.; Civalleri, B.; Casassa, S.; Maschio, L.; Ferrabone, M.; De La Pierre, M.; D'Arco, P.; Noël, Y.; Causà, M.; Rérat, M.; Kirtman, B. CRYSTAL14: A Program for the *Ab Initio* Investigation of Crystalline Solids. *Int. J. Quantum Chem.* **2014**, *114*, 1287–1317.

[73] 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.

[74] Kannemann, F. O.; Becke, A. D. van der Waals Interactions in Density-Functional Theory: Intermolecular Complexes. *J. Chem. Theory Comput.* **2010**, *6*, 1081–1088.

[75] 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.

[76] Gráfová, L.; Pitoňák, M.; Řezáč, J.; Hobza, P. Comparative Study of Selected Wave Function and Density Functional Methods for Noncovalent Interaction Energy Calculations Using the Extended S22 Data Set. *J. Chem. Theory Comput.* **2010**, *6*, 2365–2376.

[77] Řezáč, J.; Riley, K. E.; Hobza, P. Extensions of the S66 Data Set: More Accurate Interaction Energies and Angular-Displaced Nonequilibrium Geometries. *J. Chem. Theory Comput.* **2011**, *7*, 3466–3470.

[78] Řezáč, J.; Huang, Y.; Hobza, P.; Beran, G. J. O. Benchmark Calculations of Three-Body Intermolecular Interactions and the Performance of Low-Cost Electronic Structure Methods. *J. Chem. Theory Comput.* **2015**, *11*, 3065–3079.

[79] Otero-de-la-Roza, A.; Johnson, E. R. A Benchmark for Non-covalent Interactions in Solids. *J. Chem. Phys.* **2012**, *137*, 054103.

[80] 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.

[81] Della Pia, F.; Zen, A.; Alfè, D.; Michaelides, A. DMC-ICE13: Ambient and High Pressure Polymorphs of Ice from Diffusion Monte Carlo and Density Functional Theory. *J. Chem. Phys.* **2022**,

[82] Lommerse, J. P. M.; Motherwell, W. D. S.; Ammon, H. L.; Dunitz, J. D.; Gavezzotti, A.; et al., A Test of Crystal Structure Prediction of Small Organic Molecules. *Acta Cryst. B* **2000**, *58*, 647–661.

[83] Motherwell, W. D. S.; Ammon, H. L.; Dunitz, J. D.; Dzyabchenko, A.; Erk, P.; et al., Crystal Structure Prediction of Small Organic Molecules: A Second Blind Test. *Acta Cryst. B* **2002**, *58*, 647–661.

[84] Day, G. M.; Motherwell, W. D. S.; Ammon, H. L.; Boerrigter, S. X. M.; Della Valle, R. G.; et al., A Third Blind Test of Crystal Structure Prediction. *Acta Cryst. B* **2005**, *61*, 511–527.

[85] Day, G. M.; Cooper, T. G.; Cruz-Cabeza, A. J.; Hejczyk, K. E.; Ammon, H. L.; et al., 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.

[86] Bardwell, D. A.; Adjiman, C. S.; Arnautova, Y. A.; Bartashevich, E.; Boerrigter, S. X. M.; et al., Towards Crystal Structure Prediction of Complex Organic Compounds – a Report on the Fifth Blind Test. *Acta Cryst. B* **2011**, *67*, 535–551.

[87] 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., et al. Report on the Sixth Blind Test of Organic Crystal Structure Prediction Methods. *Acta Cryst. B* **2016**, *72*, 439–459.

[88] Szabo, A.; Ostlund, N. S. *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*, 1st ed.; Dover Publications, Inc., 1996.

[89] Parr, R. G. Density Functional Theory. *Annu. Rev. Phys. Chem.* **1983**, *34*, 631–656.

[90] Jain, A.; Shin, Y.; Persson, K. A. Computational Predictions of Energy Materials Using Density Functional Theory. *Nat. Rev. Mater.* **2016**, *1*, 1–13.

[91] Perdew, J. P.; Wang, Y. Accurate and Simple Analytic Representation of the Electron-gas Correlation Energy. *Phys. Rev. B* **1992**, *45*, 13244–13249.

[92] Becke, A. D. Density Functional Calculations of Molecular Bond Energies. *J. Chem. Phys.* **1986**, *85*, 4524–4529.

[93] 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.

[94] 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.

[95] Becke, A. D.; Roussel, M. R. Exchange Holes in Inhomogeneous Systems: A Coordinate Space Model. *Phys. Rev. A* **1989**, *39*, 3761–3767.

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

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

[98] Bloch, F. Quantum Mechanics of Electrons in Crystal Lattices. *Z. Phys.* **1928**, *52*, 555–600.

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

[100] Hamann, D.; Schlüter, M.; Chiang, C. Norm-conserving Pseudopotentials. *Phys. Rev. Lett.* **1979**, *43*, 1494.

[101] Troullier, N.; Martins, J. A Straightforward Method for Generating Soft Transferable Pseudopotentials. *Solid State Commun.* **1990**, *74*, 613–616.

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

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

[104] Blöchl, P. E. Projector Augmented-wave Method. *Phys. Rev. B* **1994**, *50*, 17953.

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

[106] Slater, J. C. Wave Functions in a Periodic Potential. *Phys. Rev.* **1937**, *51*, 846–851.

[107] Slater, J. C. In *Energy Band Calculations by the Augmented Plane Wave Method*; Löwdin, P.-O., Ed.; Advances in Quantum Chemistry; Academic Press, 1964; Vol. 1; pp 35–58.

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

[109] Giannozzi, P.; Andreussi, O.; Brumme, T.; Bunau, O.; Nardelli, M. B.; Calandra, M.; Car, R.; Cavazzoni, C.; Ceresoli, D.; Cococcioni, M. Advanced Capabilities for Materials Modelling with Quantum ESPRESSO. *J. Condens. Matter Phys.* **2017**, *29*, 465901.

[110] Jensen, S. R.; Saha, S.; Flores-Livas, J. A.; Huhn, W.; Blum, V.; Goedecker, S.; Frediani, L. The Elephant in the Room of Density Functional Theory Calculations. *J. Phys. Chem. Lett.* **2017**, *8*, 1449–1457.

[111] Mayer, I. Towards a "Chemical" Hamiltonian. *Int. J. Quantum Chem.* **1983**, *23*, 341–363.

[112] Mayer, I.; Valiron, P. Second Order Møller–Plesset Perturbation Theory Without Basis Set Superposition Error. *J. Chem. Phys.* **1998**, *109*, 3360–3373.

[113] 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.

[114] 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.

[115] 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.

[116] 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.

[117] Van Lenthe, E.; Baerends, E. J. Optimized Slater-type Basis Sets for the Elements 1–118. *J. Comput. Chem.* **2003**, *24*, 1142–1156.

[118] Hehre, W. J.; Stewart, R. F.; Pople, J. A. Self-consistent Molecular-orbital Methods. I. Use of Gaussian Expansions of Slater-type Atomic Orbitals. *J. Chem. Phys.* **1969**, *51*, 2657–2664.

[119] Hehre, W. J.; Ditchfield, R.; Stewart, R. F.; Pople, J. A. Self-Consistent Molecular Orbital Methods. IV. Use of Gaussian Expansions of Slater-Type Orbitals. Extension to Second-Row Molecules. *J. Chem. Phys.* **1970**, *52*, 2769–2773.

[120] 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.

[121] 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.

[122] 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.

[123] Delley, B. From Molecules to Solids With the DMol3 Approach. *J. Chem. Phys.* **2000**, *113*, 7756–7764.

[124] Kenny, S. D.; Horsfield, A. P. Plato: A Localised Orbital Based Density Functional Theory Code. *Comp. Phys. Comm.* **2009**, *180*, 2616–2621.

[125] Zunger, A.; Freeman, A. J. Self-consistent Numerical-basis-set Linear-combination-of-atomic-orbitals Model for the Study of Solids in the Local Density Formalism. *Phys. Rev. B* **1977**, *15*, 4716–4737.

[126] Delley, B.; Ellis, D. E. Efficient and Accurate Expansion Methods for Molecules in Local Density Models. *J. Chem. Phys.* **1982**, *76*, 1949–1960.

[127] Delley, B. An All-electron Numerical Method for Solving the Local Density Functional for Polyatomic Molecules. *J. Chem. Phys.* **1990**, *92*, 508–517.

[128] Wilson, A. K.; van Mourik, T.; Dunning Jr, T. H. Gaussian Basis Sets for Use in Correlated Molecular Calculations. VI. Sextuple Zeta Correlation Consistent Basis Sets for Boron Through Neon. *J. Mol. Struc-Theochem* **1996**, *388*, 339–349.

[129] Becke, A. D. A Multicenter Numerical Integration Scheme for Polyatomic Molecules. *J. Chem. Phys.* **1988**, *88*, 2547–2553.

[130] Slater, J. C. Atomic Radii in Crystals. *J. Chem. Phys.* **1964**, *41*, 3199–3204.

[131] Otero-de-la-Roza, A.; Johnson, E. R. Van der Waals Interactions in Solids Using the Exchange-hole Dipole Moment. *J. Chem. Phys.* **2012**, *136*, 174109.

[132] Hirshfeld, F. L. Bonded-atom Fragments for Describing Molecular Charge Densities. *Theor. Chim. Acta* **1977**, *44*, 129–138.

[133] 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.

[134] Becke, A. D. A Real-space Model of Nondynamical Correlation. *J. Chem. Phys.* **2003**, *119*, 2972–2977.

[135] Becke, A. D.; Johnson, E. R. A Density-Functional Model of the Dispersion Interaction. *J. Chem. Phys.* **2005**, *123*, 154101.

[136] Gil, A.; Segura, J.; Temme, N. M. *Numerical Methods for Special Functions*; SIAM, 2007.

[137] Koide, A. A New Expansion for Dispersion Forces and its Application. *J. Phys. B* **1976**, *9*, 3173–3183.

[138] 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.

[139] Grimme, S. Accurate Description of van der Waals Complexes by Density Functional Theory Including Empirical Corrections. *J. Comput. Chem.* **2004**, *25*, 1463–1473.

[140] Wu, Q.; Yang, W. Empirical Correction to Density Functional Theory for van der Waals Interactions. *J. Chem. Phys.* **2002**, *116*, 515–524.

[141] Dirac, P. A. M. On the Theory of Quantum Mechanics. *Proc. R. Soc. London, Ser. A* **1926**, *112*, 661–677.

[142] Fermi, E. Sulla Quantizzazione Del Gas Perfetto Monoatomico. *Rend. Lincei* **1926**, *145*.

[143] Caldeweyher, E.; Bannwarth, C.; Grimme, S. Extension of the D3 Dispersion Coefficient Model. *J. Chem. Phys.* **2017**, *147*, 034112.

[144] 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.

[145] 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.

[146] Lucas, A. Collective Contributions to the Long-Range Dipolar Interaction in Rare-Gas Crystals. *Physica* **1967**, *35*, 353–368.

[147] Donchev, A. G. Many-body Effects of Dispersion Interaction. *J. Chem. Phys.* **2006**, *125*, 074713.

[148] Cao, J.; Berne, B. J. Many-body dispersion forces of polarizable clusters and liquids. *J. Chem. Phys.* **1992**, *97*, 8628–8636.

[149] Cole, M. W.; Velegol, D.; Kim, H.-Y.; Lucas, A. A. Nanoscale van der Waals Interactions. *Mol. Simul.* **2009**, *35*, 849–866.

[150] Berthoumieux, H.; Maggs, A. C. Collective Dispersion Forces in the Fluid State. *Europhys. Lett.* **2010**, *91*, 56006.

[151] Dalgarno, A.; Lewis, J. The Representation of Long Range Forces by Series Expansions I: The Divergence of the Series II: The Complete Perturbation Calculation of Long Range Forces. *Proc. Phys. Soc. A* **1955**, *69*, 57.

[152] Otero-de-la-Roza, A.; Johnson, E. R. Many-body Dispersion Interactions from the Exchange-hole Dipole Moment Model. *J. Chem. Phys.* **2013**, *138*, 054103.

[153] Axilrod, B. M.; Teller, E. Interaction of the van der Waals Type Between Three Atoms. *J. Chem. Phys.* **1943**, *11*, 299–300.

[154] Muto, Y. *J. Phys. Math. Soc. Japan* **1943**, *17*, 629.

[155] Grimme, S.; Hansen, A.; Brandenburg, J. G.; Bannwarth, C. Dispersion-Corrected Mean-Field Electronic Structure Methods. *Chem. Rev.* **2016**, *9*, 5105–5154.

[156] Ahlrichs, R.; Penco, R.; Scoles, G. Intermolecular Forces in Simple Systems. *Chem. Phys.* **1977**, *19*, 119–130.

[157] Wu, X.; Vargas, M. C.; Nayak, S.; Lotrich, V.; Scoles, G. Towards Extending the Applicability of Density Functional Theory to Weakly Bound Systems. *J. Chem. Phys.* **2001**, *115*, 8748–8757.

[158] Wu, Q.; Yang, W. Empirical Correction to Density Functional Theory for van der Waals Interactions. *J. Chem. Phys.* **2002**, *116*, 515.

[159] Sato, T.; Tsuneda, T.; Hirao, K. Van der Waals Interactions Studied by Density Functional Theory. *Mol. Phys.* **2005**, *103*, 1151–1164.

[160] Gould, T.; Lebègue, S.; Ángyán, J. G.; Bučko, T. A Fractionally Ionic Approach to Polarizability and van der Waals Many-Body Dispersion Calculations. *J. Chem. Theory Comput.* **2016**, *12*, 5920–5930.

[161] Kim, M.; Kim, W. J.; Gould, T.; Lee, E. K.; Lebègue, S.; Kim, H. uMBD: A Materials-Ready Dispersion Correction That Uniformly Treats Metallic, Ionic, and van der Waals Bonding. *J. Am. Chem. Soc.* **2020**, *142*, 2346–2354.

[162] Andersson, Y.; Langreth, D. C.; Lundqvist, B. I. van der Waals Interactions in Density-Functional Theory. *Phys. Rev. Lett.* **1996**, *76*, 102–105.

[163] Cooper, V. R. van der Waals Density Functional: An Appropriate Exchange Functional. *Phys. Rev. B* **2010**, *81*, 161104.

[164] Klimeš, J.; Bowler, D. R.; Michaelides, A. Chemical Accuracy for the van der Waals Density Functional. *Phys. Cond. Matter* **2010**, *22*, 022201.

[165] Klimeš, J.; Bowler, D. R.; Michaelides, A. Van der Waals Density Functionals Applied to Solids. *Phys. Rev. B* **2011**, *83*, 195131.

[166] Hamada, I. van der Waals Density Functional Made Accurate. *Phys. Rev. B* **2014**, *89*, 121103(R).

[167] Pernal, K.; Podeszwa, R.; Patkowski, K.; Szalewicz, K. Dispersionless Density Functional Theory. *Phys. Rev. Lett.* **2009**, *103*, 263201.

[168] Éamonn D. Murray,; Lee, K.; Langreth, D. C. Investigation of Exchange Energy Density Functional Accuracy for Interacting Molecules. *J. Chem. Theory Comput.* **2009**, *5*, 2754–2762.

[169] Perez-Jorda, J. M.; Becke, A. D. A Density-Functional Study of van-der-Waals Forces – Rare-Gas Diatomics. *Chem. Phys. Lett.* **1995**, *233*, 134–137.

[170] Johnson, E. R.; Wolkow, R. A.; DiLabio, G. A. Application of 25 Density Functionals to Dispersion-Bound Homomolecular Dimers. *Chem. Phys. Lett.* **2004**, *394*, 334–338.

[171] Kannemann, F. O.; Becke, A. D. Van der Waals Interactions in Density-Functional Theory: Rare-Gas Diatomics. *J. Chem. Theory Comput.* **2009**, *5*, 719–727.

[172] Gillan, M. J. Many-body Exchange-Overlap Interactions in Rare Gases and Water. *J. Chem. Phys.* **2014**, *141*, 224106.

[173] Goerigk, L. Treating London-Dispersion Effects with the Latest Minnesota Density Functionals: Problems and Possible Solutions. *J. Phys. Chem. Lett.* **2015**, *6*, 3891–3896.

[174] Goerigk, L.; Grimme, S. A Thorough Benchmark of Density Functional Methods for General Main Group Thermochemistry, Kinetics, and Noncovalent Interactions. *Phys. Chem. Chem. Phys.* **2011**, *13*, 6670–6688.

[175] Perdew, J. P.; Zunger, A. Self-interaction Correction to Density-functional Approximations for Many-Electron Systems. *Phys. Rev. B* **1981**, *23*, 5048–5079.

[176] Perdew, J.; Yue, W. Accurate and Simple Density Functional for the Electronic Exchange Energy: Generalized Gradient Approximation. *Phys. Rev. B* **1986**, *33*, 8800.

[177] Becke, A. D. Density-functional Exchange-energy Approximation with Correct Asymptotic Behavior. *Phys. Rev. A* **1988**, *38*, 3098.

[178] Perdew, J. P.; Chevary, J. A.; Vosko, S. H.; Jackson, K. A.; Pederson, M. R.; Singh, D. J.; Fiolh ais, C. Atoms, Molecules, Solids, and Surfaces: Applications of the Generalized Gradient Approximation for Exchange and Correlation. *Phys. Rev. B* **1992**, *46*, 6671–6687.

[179] Perdew, J. P.; Ruzsinszky, A.; Csonka, G. I.; Vydrov, O. A.; Scuseria, G. E.; Constantin, L. A.; Zhou, X.; Burke, K. Restoring the Density-gradient Expansion for Exchange in Solids and Surfaces. *Phys. Rev. Lett.* **2008**, *100*, 136406.

[180] Zhang, Y.; Yang, W. Comment on "Generalized Gradient Approximation Made Simple". *Phys. Rev. Lett.* **1998**, *80*, 890–890.

[181] Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti Correlation-energy Formula into a Functional of the Electron Density. *Phys. Rev. B* **1988**, *37*, 785.

[182] Spanu, L.; Sorella, S.; Galli, G. Nature and Strength of Interlayer Binding in Graphite. *Phys. Rev. Lett.* **2009**, *103*, 196401.

[183] Gillan, M. J.; Alfè, D.; Michaelides, A. Perspective: How good is DFT for water? *J. Chem. Phys.* **2016**, *144*, 130901.

[184] Becke, A. D.; Johnson, E. R. A Unified Density-Functional Treatment of Dynamical, Nondynamical and Dispersion Correlations. *J. Chem. Phys.* **2007**, *127*, 124108.

[185] Austin, A.; Petersson, G. A.; Frisch, M. J.; Dobek, F. J.; Scalmani, G.; Throssell, K. A Density Functional with Spherical Atom Dispersion Terms. *J. Chem. Theory Comput.* **2012**, *8*, 4989–5007.

[186] Johnson, E. R.; Becke, A. D.; Sherrill, C. D.; DiLabio, G. A. Oscillations in Meta-Generalized-Gradient Approximation Potential Energy Surfaces for Dispersion-Bound Complexes. *J. Chem. Phys.* **2009**, *131*, 034111.

[187] Wheeler, S. E.; Houk, K. N. Integration Grid Errors for Meta-GGA-Predicted Reaction Energies: Origin of Grid Errors for the M06 Suite of Functionals. *J. Chem. Theory Comput.* **2010**, *6*, 395–404.

[188] Dasgupta, S.; Herbert, J. M. Standard Grids for High-Precision Integration of Modern Density Functionals: SG-2 and SG-3. *J. Comp. Chem.* **2017**, *38*, 869–882.

[189] Morgante, P.; Peverati, R. The Devil in the Details: A Tutorial Review on Some Undervalued Aspects of Density Functional Theory Calculations. *Int. J. Quantum Chem.* **2020**, *120*, e26332.

[190] Mardirossian, N.; Head-Gordon, M. How Accurate Are the Minnesota Density Functionals for Noncovalent Interactions, Isomerization Energies, Thermochemistry, and Barrier Heights Involving Molecules Composed of Main-Group Elements? *J. Chem. Theory Comput.* **2016**, *12*, 4303–4325.

[191] Jiménez-Hoyos, C. A.; Janesko, B. G.; Scuseria, G. E. Evaluation of Range-Separated Hybrid Density Functionals for the Prediction of Vibrational Frequencies, Infrared Intensities, and Raman Activities. *Phys. Chem. Chem. Phys.* **2008**, *10*, 6621–6629.

[192] Csonka, G. I.; French, A. D.; Johnson, G. P.; Stortz, C. A. Evaluation of Density Functionals and Basis Sets for Carbohydrates. *J. Chem. Theory Comput.* **2009**, *5*, 679–692.

[193] Gould, T.; Johnson, E. R.; Tawfik, S. A. Are Dispersion Corrections Accurate Outside Equilibrium? A Case Study on Benzene. *Beilstein J. Org. Chem.* **2018**, *14*, 1181–1191.

[194] Sun, J.; Ruzsinszky, A.; Perdew, J. P. Strongly Constrained and Appropriately Normed Semilocal Density Functional. *Phys. Rev. Lett.* **2015**, *115*, 036402.

[195] Furness, J. W.; Kaplan, A. D.; Ning, J.; Perdew, J. P.; Sun, J. Accurate and Numerically Efficient r^2SCAN Meta-Generalized Gradient Approximation. *J. Phys. Chem. Lett.* **2020**, *11*, 8208–8215.

[196] Voorhis, T. V.; Scuseria, G. E. A Novel Form for the Exchange-Correlation Energy Functional. *J. Chem. Phys.* **1998**, *109*, 400–410.

[197] Tao, J.; Perdew, J. P.; Staroverov, V. N.; Scuseria, G. E. Climbing the Density Functional Ladder: Nonempirical Meta–Generalized Gradient Approximation Designed for Molecules and Solids. *Phys. Rev. Lett.* **2003**, *91*, 146401.

[198] Tang, K. T.; Toennies, J. P. The van der Waals Potentials Between All the Rare Gas Atoms from He to Rn. *J. Chem. Phys.* **2003**, *118*, 4976.

[199] Becke, A. D.; Dickson, R. M. Numerical-Solution of Schrödinger's Equation in Polyatomic-Molecules. *J. Chem. Phys.* **1990**, *92*, 3610.

[200] Ehlert, S.; Huniar, U.; Ning, J.; Furness, J. W.; Sun, J.; Kaplan, A. D.; Perdew, J. P.; Brandenburg, J. G. r^2SCAN-D4: Dispersion Corrected Meta-Generalized Gradient Approximation for General Chemical Applications. *J. Chem. Phys.* **2021**, *154*, 061101.

[201] Zhang, Y.; Yang, W. A Challenge for Density Functionals: Self-Interaction Error Increases for Systems with a Noninteger Number of Electrons. *J. Chem. Phys.* **1998**, *109*, 2604–2608.

[202] 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.

[203] Vydrov, O. A.; Scuseria, G. E.; Perdew, J. P. Tests of Functionals for Systems with Fractional Electron Number. *J. Chem. Phys.* **2007**, *126*, 154109.

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

[205] Hait, D.; Head-Gordon, M. Delocalization Errors in Density Functional Theory Are Essentially Quadratic in Fractional Occupation Number. *J. Phys. Chem. Lett.* **2018**, *9*, 6280–6288.

[206] Ruiz, E.; Salahub, D. R.; Vela, A. Charge-Transfer Complexes: Stringent Tests for Widely Used Density Functionals. *J. Chem. Phys.* **1996**, *100*, 12265–12276.

[207] Johnson, E. R.; Salamone, M.; Bietti, M.; DiLabio, G. A. Modeling Non-Covalent Radical-Molecule Interactions Using Conventional Density-Functional Theory: Beware Erroneous Charge Transfer. *J. Phys. Chem. A* **2013**, *117*, 947–952.

[208] 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.

[209] 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.

[210] 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.

[211] 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.

[212] 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.

[213] Dreuw, A.; Head-Gordon, M. Failure of Time-Dependent Density Functional Theory for Long-Range Charge-Transfer Excited States: The Zincbacteriochlorin-Bacteriochlorin and Bacteriochlorophyll-Spheroidene Complexes. *J. Am. Chem. Soc.* **2004**, *126*, 4007–4016.

[214] Gritsenko, O.; Baerends, E. J. Asymptotic Correction of the Exchange-Correlation Kernel of Time-Dependent Density Functional Theory for Long-Range Charge-Transfer Excitations. *J. Chem. Phys.* **2004**, *121*, 655.

[215] Peach, M. J. G.; Benfield, P.; Helgaker, T.; Tozer, D. J. Excitation Energies in Density Functional Theory: An Evaluation and a Diagnostic Test. *J. Chem. Phys.* **2008**, *128*, 044118.

[216] Isborn, C. M.; Mar, B. D.; Curchod, B. F. E.; Tavernelli, I.; Martínez, T. J. The Charge Transfer Problem in Density Functional Theory Calculations of Aqueously Solvated Molecules. *J. Phys. Chem. B* **2013**, *117*, 12189–12201.

[217] Moore, B.; Sun, H.; Govind, N.; Kowalski, K.; Autschbach, J. Charge-Transfer Versus Charge-Transfer-Like Excitations Revisited. *J. Chem. Theory Comput.* **2015**, *11*, 3305–3320.

[218] Kümmel, S. Charge-Transfer Excitations: A Challenge for Time-Dependent Density Functional Theory That Has Been Met. *Advan. Energy Mater.* **2017**, *7*, 1700440.

[219] 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.

[220] Patchkovskii, S.; Ziegler, T. Improving "Difficult" Reaction Barriers with Self-Interaction Corrected Density Functional Theory. *J. Chem. Phys.* **2002**, *116*, 7806–7813.

[221] Dickson, R. M.; Becke, A. D. Reaction Barrier Heights from an Exact-Exchange-Based Density-Functional Correlation Model. *J. Chem. Phys.* **2005**, *123*, 111101.

[222] Lingwood, M.; Hammond, J. R.; Hrovat, D. A.; Mayer, J. M.; Borden, W. T. MPW1K Performs Much Better than B3LYP in DFT Calculations on Reactions that Proceed by Proton-Coupled Electron Transfer (PCET). *J. Chem. Theory Comput.* **2006**, *2*, 740–745.

[223] 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.

[224] 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.

[225] Woodcock, H. L.; Schaefer, H. F.; Schreiner, P. R. Problematic Energy Differences Between Cumulenes and Poly-ynes: Does This Point to a Systematic Improvement of Density Functional Theory? *J. Phys. Chem. A* **2002**, *106*, 11923–11931.

[226] 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.

[227] Johnson, E. R.; Clarkin, O. J.; DiLabio, G. A. Density Functional Theory Based Model Calculations for Accurate Bond Dissociation Enthalpies. 3. A Single Approach for XH, XX, and XY (X, Y= C, N, O, S, Halogen) Bonds. *J. Phys. Chem. A* **2003**, *107*, 9953–9963.

[228] Jacquemin, D.; Femenias, A.; Chermette, H.; Ciofini, I.; Adamo, C.; André, J.-M.; Perpète, E. A. Assessment of Several Hybrid DFT Functionals for the Evaluation of Bond Length Alternation of Increasingly Long Oligomers. *J. Phys. Chem. A* **2006**, *110*, 5952–5959.

[229] Heaton-Burgess, T.; Yang, W. Structural Manifestation of the Delocalization Error of Density Functional Approximations: C_{4N+2} Rings and C_{20} Bowl, Cage, and Ring Isomers. *J. Chem. Phys.* **2010**, *132*, 234113.

[230] White, M. A.; Kahwaji, S.; Freitas, V. L. S.; Siewert, R.; Weatherby, J. A.; Ribeiro da Silva, M. D. M. C.; Verevkin, S. P.; Johnson, E. R.; Zwanziger, J. W. Relative Thermodynamic Stability of Diamond and Graphite. *Angew. Chem. Int. Ed.* **2021**, *60*, 1546–1549.

[231] Kim, M.-C.; Sim, E.; Burke, K. Understanding and Reducing Errors in Density Functional Calculations. *Phys. Rev. Lett.* **2013**, *111*, 073003.

[232] Chermette, H.; Ciofini, I.; Mariotti, F.; Daul, C. Correct Dissociation Behavior of Radical Ions Such as H_2^+ in Density Functional Calculations. *J. Chem. Phys.* **2001**, *114*, 1447–1453.

[233] Becke, A. D. A Real-Space Model of Nondynamical Correlation. *J. Chem. Phys.* **2003**, *119*, 2972–2977.

[234] Grafenstein, J.; Kraka, E.; Cremer, D. The Impact of the Self-Interaction Error on the Density Functional Theory Description of Dissociating Radical Cations: Ionic and Covalent Dissociation Limits. *J. Chem. Phys.* **2004**, *120*, 524–539.

[235] 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.

[236] Kim, M.-C.; Sim, E.; Burke, K. Ions in Solution: Density Corrected Density Functional Theory (DC-DFT). *J. Chem. Phys.* **2014**, *140*, 18A528.

[237] 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.

[238] Sim, E.; Song, S.; Burke, K. Quantifying Density Errors in DFT. *J. Phys. Chem. Lett.* **2018**, *9*, 6385–6392.

[239] Sham, L. J.; Schlüter, M. Density-Functional Theory of the Band Gap. *Phys. Rev. B* **1985**, *32*, 3883.

[240] Perdew, J. P. Density Functional Theory and the Band Gap Problem. *Int. J. Quantum Chem.* **1985**, *28*, 497–523.

[241] Cohen, A. J.; Mori-Sánchez, P.; Yang, W. Fractional Charge Perspective on the Band Gap in Density-Functional Theory. *Phys. Rev. B.* **2008**, *77*, 115123.

[242] 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.

[243] Crowley, J. M.; Tahir-Kheli, J.; III, W. A. G. Resolution of the Band Gap Prediction Problem for Materials Design. *J. Phys. Chem. Lett.* **2016**, *7*, 1198–1203.

[244] Perdew, J. P.; Yang, W.; Burke, K.; Yang, Z.; Gross, E. K. U.; Scheffler, M.; Scuseria, G. E.; Henderson, T. M.; Zhang, I. Y.; Ruzsinszky, A.; Peng, H.; Sun, J.; Trushin, E.; Görling, A. Understanding Band Gaps of Solids in Generalized Kohn–Sham Theory. *Proc. Nat. Acad. Sci.* **2017**, *114*, 2801–2806.

[245] Otero-de-la-Roza, A.; LeBlanc, L. M.; Johnson, E. R. Dispersion XDM with Hybrid Functionals: Delocalization Error and Halogen Bonding in Molecular Crystals. *J. Chem. Theory Comput.* **2019**, *15*, 4933–4944.

[246] Otero-de-la-Roza, A.; Johnson, E. R. Application of XDM to Ionic Solids: the Importance of Dispersion for Bulk Moduli and Crystal Geometries. *J. Chem. Phys.* **2020**, *153*, 054121.

[247] LeBlanc, L. M.; Dale, S. G.; Taylor, C. R.; Becke, A. D.; Day, G. M.; Johnson, E. R. Pervasive Delocalisation Error Causes Spurious Proton Transfer in Organic Acid-Base Co-crystals. *Angew. Chem. Int. Ed.* **2018**, *57*, 14906–14910.

[248] Kozuch, S.; Martin, J. M. L. Halogen Bonds: Benchmark and Theoretical Analysis. *J. Chem. Theory Comput.* **2013**, *9*, 1918–1931.

[249] Bauzá, A.; Alkorta, I.; Frontera, A.; Elguero, J. On the Reliability of Pure and Hybrid DFT Methods for the Evaluation of Halogen, Chalcogen, and Pnicogen Bonds Involving Anionic and Neutral Electron Donors. *J. Chem. Theory Comput.* **2013**, *9*, 5201–5210.

[250] 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.

[251] Kim, Y.; Song, S.; Sim, E.; Burke, K. Halogen and Chalcogen Binding Dominated by Density-Driven Errors. *J. Phys. Chem. Lett.* **2019**, *10*, 295–301.

[252] Otero-de-la Roza, A.; DiLabio, G. A.; Johnson, E. R. Exchange-Correlation Effects for Non-Covalent Interactions in Density-Functional Theory. *J. Chem. Theory. Comput.* **2016**, *12*, 3160–3175.

[253] Brandenburg, J. G.; Maas, T.; Grimme, S. Benchmarking DFT and Semiempirical Methods on Structures and Lattice Energies for Ten Ice Polymorphs. *J. Chem. Phys.* **2015**, *142*, 124104.

[254] Yanai, T.; Tew, D. P.; Handy, N. C. A New Hybrid Exchange-Correlation Functional Using the Coulomb-Attenuating Method (CAM-B3LYP). *Chem. Phys. Lett.* **2004**, *393*, 51–57.

[255] Vydrov, O. A.; Scuseria, G. E. Assessment of a Long-Range Corrected Hybrid Functional. *J. Chem. Phys.* **2006**, *125*, 234109.

[256] Vydrov, O. A.; Heyd, J.; Krukau, A. V.; Scuseria, G. E. Importance of Short-Range Versus Long-Range Hartree-Fock Exchange for the Performance of Hybrid Density Functionals. *J. Chem. Phys.* **2006**, *125*, 074106.

[257] Chai, J.-D.; Head-Gordon, M. Systematic Optimization of Long-range Corrected Hybrid Density Functionals. *J. Chem. Phys.* **2008**, *128*, 084106.

[258] Koerzdoerfer, T.; Sears, J. S.; Sutton, C.; Bredas, J. L. Long-Range Corrected Hybrid Functionals for pi-Conjugated Systems: Dependence of the Range-Separation Parameter on Conjugation Length. *J. Chem. Phys.* **2011**, *135*, 204107.

[259] Kronik, L.; Stein, T.; Refaely-Abramson, S.; Baer, R. Excitation Gaps of Finite-Sized Systems from Optimally Tuned Range-Separated Hybrid Functionals. *J. Chem. Theory Comput.* **2012**, *8*, 1515–1531.

[260] Garrett, K.; Sosa Vazquez, X. A.; Egri, S. B.; Wilmer, J.; Johnson, L. E.; Robinson, B. H.; Isborn, C. M. Optimum Exchange for Calculation of Excitation Energies and Hyperpolarizabilities of Organic Electro-Optic Chromophores. *J. Chem. Theory Comput* **2014**, *10*, 3821–3831.

[261] Dovesi, R.; Erba, A.; Orlando, R.; Zicovich-Wilson, C. M.; Civalleri, B.; Maschio, L.; Rérat, M.; Casassa, S.; Baima, J.; Salustro, S.; Kirtman, B. Quantum-Mechanical Condensed Matter Simulations with CRYSTAL. *WIREs Comput. Mol. Sci.* **2018**, *8*, e1360.

[262] Lennerd-Jones, J. E. Cohesion. *Proc. Phys. Sci.* **1931**, *43*, 461–482.

[263] Risthaus, T.; Grimme, S. Benchmarking of London Dispersion-Accounting Density Functional Theory Methods On Very Large Molecular Complexes. *J. Chem. Theory Comput.* **2013**, *9*, 1580–1591.

[264] 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.

[265] Mohebifar, M.; Johnson, E. R.; Rowley, C. N. Evaluating Force-field London Dispersion Coefficients Using the Exchange-hole Dipole Moment Model. *J. Chem. Theory Comput.* **2017**, *13*, 6146–6157.

[266] Walters, E.; Mohebifar, M.; Johnson, E. R.; Rowley, C. N. Evaluating the London Dispersion Coefficients of Protein Force Fields Using the Exchange-hole Dipole Moment Model. *J. Phys. Chem. B* **2018**, *122*, 6690–6701.

[267] Mohebifar, M.; Rowley, C. N. An Efficient and Accurate Model for Water with an Improved Non-bonded Potential. *J. Chem. Phys.* **2020**, *153*, 134105.

[268] Alkan, M.; Xu, P.; Gordon, M. S. Many-Body Dispersion in Molecular Clusters. *J. Phys. Chem. A* **2019**, *123*, 8406–8416.

[269] von Lilienfeld, O. A.; Tkatchenko, A. Two-and Three-body Interatomic Dispersion Energy Contributions to Binding in Molecules and Solids. *J. Chem. Phys.* **2010**, *132*, 234109.

[270] Johnson, E. R. Dependence of Dispersion Coefficients on Atomic Environment. *J. Chem. Phys.* **2011**, *135*, 234109.

[271] 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.

[272] Ruiz, V. G.; Liu, W.; Zojer, E.; Scheffler, M.; Tkatchenko, A. Density-Functional Theory with Screened van der Waals Interactions for the Modeling of Hybrid Inorganic-Organic Systems. *Phys. Rev. Lett.* **2012**, *108*, 146103.

[273] Wyckoff, R. *Crystal Structures*; Interscience publishers: New York, 1963.

[274] Marzari, N.; Vanderbilt, D.; De Vita, A.; Payne, M. C. Thermal Contraction and Disordering of the Al(110) Surface. *Phys. Rev. Lett.* **1999**, *82*, 3296–3299.

[275] Cutini, M.; Maschio, L.; Ugliengo, P. Exfoliation Energy of Layered Materials by DFT-D: Beware of Dispersion! *J. Chem. Theory Comput.* **2020**, *16*, 5244–5252.

[276] National Research Council; Division on Engineering and Physical Sciences; Commission on Physical Sciences, Mathematics, and Applications; Panel on New Directions in Catalytic Science and Technology, *Catalysis Looks to the Future*; National Academies Press, 1992.

[277] Dumesic, J. A.; Huber, G. W.; Boudart, M. Principles of Heterogeneous Catalysis. *Handbook of Heterogeneous Catalysis: Online* **2008**, 1–15.

[278] Armor, J. N. New Catalytic Technology Commercialized in the USA During the 1990s. *Appl. Catal. A: Gen.* **2001**, *222*, 407–426.

[279] Chorkendorff, I.; Niemantsverdriet, J. W. *Concepts of Modern Catalysis and Kinetics*; John Wiley & Sons, 2017.

[280] Bartholomew, C. H.; Farrauto, R. J. *Fundamentals of Industrial Catalytic Processes*; John Wiley & Sons, 2011.

[281] Averill, B. A.; Moulijn, J. A.; van Santen, R. A.; van Leeuwen, P. W. N. M. *Catalysis: An Integrated Approach*; Elsevier, 1999.

[282] Krull, C.; Castelli, M.; Hapala, P.; Kumar, D.; Tadich, A.; Capsoni, M.; Edmonds, M.; Hellerstedt, J.; Burke, S.; Jelinek, P. Iron-based Trinuclear Metal-organic Nanostructures on a Surface with Local Charge Accumulation. *Nat. Comm.* **2018**, *9*, 1–7.

[283] Schiffrin, A.; Capsoni, M.; Farahi, G.; Wang, C.; Krull, C.; Castelli, M.; Roussy, T.; Cochrane, K.; Yin, Y.; Medhekar, N.; M., F.; Shaw, A.; Ji, W.; Burke, S. Designing Optoelectronic Properties by On-Surface Synthesis: Formation and Electronic Structure of an Iron-Terpyridine Macromolecular Complex. *ACS Nano* **2018**, *12*, 6545–6553.

[284] Valero, M. C.; Raybaud, P.; Sautet, P. Interplay Between Molecular Adsorption and Metal–support Interaction For Small Supported Metal Clusters: CO and C_2H_4 Adsorption on Pd_4/γ-Al_2O_3. *J. Cata.* **2007**, *247*, 339–355.

[285] Cyr, D. M.; Venkataraman, B.; Flynn, G. W. STM Investigations of Organic Molecules Physisorbed at the Liquid-Solid Interface. *Chem. Mater.* **1996**, *8*, 1600–1615.

[286] Grinter, D.; Thornton, G. In *Encyclopedia of Interfacial Chemistry*; Wandelt, K., Ed.; Elsevier: Oxford, 2018; pp 62–85.

[287] Binnig, G.; Rohrer, H.; Gerber, C.; Weibel, E. Tunneling Through a Controllable Vacuum Gap. *Appl. Phys. Lett.* **1982**, *40*, 178–180.

[288] Binnig, G.; Rohrer, H.; Gerber, C.; Weibel, E. Surface Studies by Scanning Tunneling Microscopy. *Phys. Rev. Lett.* **1982**, *49*, 57–61.

[289] Tersoff, J.; Hamann, D. R. Theory of the Scanning Tunneling Microscope. *Phys. Rev. B* **1985**, *31*, 805–813.

[290] Bardeen, J. Tunnelling from a Many-Particle Point of View. *Phys. Rev. Lett.* **1961**, *6*, 57–59.

[291] 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.

[292] DeJong, M.; Price, A. J.; Mårsell, E.; Nguyen, G.; Johnson, E. R.; Burke, S. A. Small Molecule Binding to Surface-Supported Single-Site Reaction Centers. *Nat. Comm.* **2022**, *13*, 7407.

[293] Liu, W.; Maaß, F.; Willenbockel, M.; Bronner, C.; Schulze, M.; Soubatch, S.; Tautz, F. S.; Tegeder, P.; Tkatchenko, A. Quantitative Prediction of Molecular Adsorption: Structure and Binding of Benzene on Coinage Metals. *Phys. Rev. Lett.* **2015**, *115*, 036104.

[294] Haynes, W. M. *CRC Handbook of Chemistry and Physics*; CRC press, 2016.

[295] The postg program is freely available from https://github.com/aoterodelaroza/postg.

[296] Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A. V.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratchian, H. P.; iz, J. V. O.; Izmaylov, A. F.; Sonnenberg, J. L.; Williams-Young, D.; Ding, F.; Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranasinghe, D.; Zakrzewski, V. G.; Gao, J.; Rega, N.; Zheng, G.; Liang, W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Throssell, K.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J. J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Keith, T. A.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Millam, J. M.; Klene, M.; Adamo, C.; Cammi, R.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Farkas, O.; Foresman, J. B.; Fox, D. J. Gaussian 16 Revision A.03. 2016; Gaussian Inc. Wallingford CT.

[297] Adeleke, A. A.; Johnson, E. R. Effects of Dispersion Corrections on the Theoretical Description of Bulk Metals. *Phys. Rev. B* **2023**, *107*.

[298] Bernstein, J. *Polymorphism in Molecular Crystals*; International Union of Crystallography, 2020; Vol. 30.

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

[300] Price, S. L.; Reutzel-Edens, S. M. The Potential of Computed Crystal Energy Landscapes to Aid Solid-Form Development. *Drug Discov. Today* **2016**, *21*, 912–923.

[301] 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.

[302] Nyman, J.; Reutzel-Edens, S. M. Crystal Structure Prediction is Changing From Basic Science to Applied Technology. *Faraday Discuss.* **2018**, *211*, 459–476.

[303] Dudek, M. K.; Drużbicki, K. Along the Road to Crystal Structure Prediction (CSP) of Pharmaceutical-Like Molecules. *CrystEngComm* **2022**, *24*, 1665–1678.

[304] Peschar, R.; Pop, M. M.; De Ridder, D. J.; van Mechelen, J. B.; Driessen, R. A.; Schenk, H. Crystal Structures of 1, 3-distearoyl-2-oleoylglycerol and Cocoa Butter in the β (V) Phase Reveal the Driving Force Behind the Occurrence of Fat Bloom on Chocolate. *J. Phys. Chem. B* **2004**, *108*, 15450–15453.

[305] Bolton, O.; Matzger, A. J. Improved Stability and Smart-Material Functionality Realized in an Energetic Cocrystal. *Angew. Chem.* **2011**, *50*, 8960–8963.

[306] Bier, I.; O'Connor, D.; Hsieh, Y.-T.; Wen, W.; Hiszpanski, A. M.; Han, T. Y.-J.; Marom, N. Crystal Structure Prediction of Energetic Materials and a Twisted Arene with Genarris and GAtor. *CrystEngComm* **2021**, *23*, 6023–6038.

[307] O'Connor, D.; Bier, I.; Hsieh, Y.-T.; Marom, N. Performance of Dispersion-Inclusive Density Functional Theory Methods for Energetic Materials. *J. Chem. Theory Comput* **2022**, *18*, 4456–4471.

[308] Campbell, J. E.; Yang, J.; Day, G. M. Predicted Energy–Structure–Function Maps for the Evaluation of Small Molecule Organic Semiconductors. *J. Mater. Chem. C* **2017**, *5*, 7574–7584.

[309] 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.

[310] 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.

[311] Yang, Y.; Rice, B.; Shi, X.; Brandt, J. R.; Correa da Costa, R.; Hedley, G. J.; Smilgies, D.-M.; Frost, J. M.; Samuel, I. D. W.; 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.

[312] Rice, B.; LeBlanc, L.; Otero-de-la-Roza, A.; Fuchter, M. J.; Johnson, E. R.; Nelson, J.; Jelfs, K. E. A Computational Exploration of the Crystal Energy and Charge Carrier Mobility Landscapes of the Chiral [6]Helicene Molecule. *Nanoscale* **2018**, *10*, 1865–1876.

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

[314] 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.

[315] 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.

[316] Blagden, N.; de Matas, M.; Gavan, P. T.; York, P. Crystal Engineering of Active Pharmaceutical Ingredients to Improve Solubility and Dissolution Rates. *Adv. Drug Deliv. Rev.* **2007**, *59*, 617–630.

[317] Singhai, D.; Curatolo, W. Drug Polymorphism and Dosage Form Design: A Practical Perspective. *Adv. Drug Deliv. Rev.* **2004**, *56*, 335–347.

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

[319] Millar, D. I.; Oswald, I. D. H.; Francis, D. J.; Marshall, W. G.; Pulham, C. R.; S, C. A. The Crystal Structure of β-RDX–an Elusive Form of an Explosive Revealed. *Chem. Commun.* **2009**, 562–564.

[320] Neumann, M. A.; Perrin, M.-A. Energy Ranking of Molecular Crystals Using Density Functional Theory Calculations and an Empirical van der Waals Correction. *J. Phys. Chem. B* **2005**, *109*, 15531–15541.

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

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

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

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

[325] Wen, S.; Nanda, K.; Huang, Y.; Beran, G. J. Practical Quantum Mechanics-Based Fragment Methods for Predicting Molecular Crystal Properties. *Phys. Chem. Chem. Phys.* **2012**, *14*, 7578–7590.

[326] Červinka, C.; Fulem, M.; Ružička, K. CCSD (T)/CBS Fragment-Based Calculations of Lattice Energy of Molecular Crystals. *J. Chem. Phys.* **2016**, *144*, 064505.

[327] Beran, G. J. O.; Wen, S.; Nanda, K.; Huang, Y.; Heit, Y. In *Prediction and Calculation of Crystal Structures: Methods and Applications*; Atahan-Evrenk, S., Aspuru-Guzik, A., Eds.; Springer International Publishing: Cham, 2014; pp 59–93.

[328] McDonagh, D.; Skylaris, C.-K.; Day, G. M. Machine-Learned Fragment-Based Energies for Crystal Structure Prediction. *J. Chem. Theory Comput.* **2019**, *15*, 2743–2758.

[329] 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.

[330] Maschio, L.; Civalleri, B.; Ugliengo, P.; Gavezzotti, A. Intermolecular Interaction Energies in Molecular Crystals: Comparison and Agreement of Localized Møller–Plesset 2, Dispersion-Corrected Density Functional, and Classical Empirical Two-body Calculations. *J. Phys. Chem. A* **2011**, *115*, 11179–11186.

[331] Musil, F.; De, S.; Yang, J.; Campbell, J. E.; Day, G. M.; Ceriotti, M. Machine Learning for the Structure–Energy–Property Landscapes of Molecular Crystals. *Chem. Sci.* **2018**, *9*, 1289–1300.

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

[333] 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.

[334] Brandenburg, J. G.; Grimme, S. Organic Crystal Polymorphism: A Benchmark for Dispersion Corrected Mean Field Electronic Structure Methods. *Acta Cryst. B* **2016**, *72*, 502–513.

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

[336] Whittleton, S. R.; Otero-de-la-Roza, A.; Johnson, E. R. The 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.

[337] LeBlanc, L. M.; ; Johnson, E. R. Crystal-Energy Landscapes of Active Pharmaceutical Ingredients Using Composite Approaches. *CrystEngComm* **2019**, *21*, 5995–6009.

[338] 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.

[339] Kronik, L.; Tkatchenko, A. Understanding Molecular Crystals with Dispersion-Inclusive Density Functional Theory: Pairwise Corrections and Beyond. *Acc. Chem. Res.* **2014**, *47*, 3208–3216.

[340] 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.

[341] Shtukenberg, A. G.; Zhu, Q.; Carter, D. J.; Vogt, L.; Hoja, J.; Schneider, E.; Song, H.; Pokroy, B.; Polishchuk, I.; Tkatchenko, A.; Oganov, A. R.; Rohl, A. L.; Tuckerman, M. E.; Kahr, B. Powder Diffraction and Crystal Structure Prediction Identify Four New Coumarin Polymorphs. *Chem. Sci.* **2017**, *8*, 4926–4940.

[342] Hoja, J.; Tkatchenko, A. First-Principles Stability Ranking of Molecular Crystal Polymorphs with the DFT+MBD Approach. *Faraday Discuss.* **2018**, *211*, 253–274.

[343] Hoja, J.; Ko, H. Y.; Neumann, M. A.; Car, R.; Distasio, R. A.; Tkatchenko, A. Reliable and Practical Computational Description of Molecular Crystal Polymorphs. *Sci. Adv.* **2019**, *5*.

[344] Mortazavi, M.; Hoja, J.; Aerts, L.; Quéré, L.; van de Streek, J.; Neumann, M. A.; Tkatchenko, A. Computational Polymorph Screening Reveals Late-Appearing and Poorly-Soluble Form of Rotigotine. *Commun. Chem.* **2019**, *2*, 1–7.

[345] Otero-de-la-Roza, A.; Hein, J. E.; Johnson, E. R. Reevaluating the Stability and Prevalence of Conglomerates: Implications for Preferential Crystallization. *Cryst. Growth Des.* **2016**, *16*, 6055–6059.

[346] 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.

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

[348] Price, A. J. A.; Bryenton, K. R.; Johnson, E. R. Requirements for an Accurate Dispersion-Corrected Density Functional. *J. Chem. Phys.* **2021**, *154*, 230902, Chapter 3 in this thesis.

[349] Bryenton, K. R.; Adeleke, A. A.; Dale, S. G.; Johnson, E. R. Delocalization Error: The Greatest Outstanding Challenge in Density-Functional Theory. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2022**, e1631.

[350] Greenwell, C.; Beran, G. J. O. Inaccurate Conformational Energies Still Hinder Crystal Structure Prediction in Flexible Organic Molecules. *Cryst. Growth Des.* **2020**, *20*, 4875–2881.

[351] Greenwell, C.; McKinley, J. L.; Zhang, P.; Zeng, Q.; Sun, G.; Li, B.; Wen, S.; Beran, G. J. O. Overcoming the Difficulties of Predicting Conformational Polymorph Energetics in Molecular Crystals Via Correlated Wavefunction Methods. *Chem. Sci.* **2020**, *11*, 2200–2214.

[352] Beran, G. J.; Wright, S. E.; Greenwell, C.; Cruz-Cabeza, A. J. The Interplay of Intra-and Intermolecular Errors in Modeling Conformational Polymorphs. *J. Chem. Phys.* **2022**, *156*, 104112.

[353] Knuth, F.; Carbogno, C.; Atalla, V.; Blum, V.; Scheffler, M. All-electron Formalism for Total Energy Strain Derivatives and Stress Tensor Components for Numeric Atom-Centered Orbitals. *Comp. Phys. Comm.* **2015**, *190*, 33–50.

[354] Hait, D.; Head-Gordon, M. Delocalization Errors in Density Functional Theory are Essentially Quadratic in Fractional Occupation Number. *J. Phys. Chem. Lett.* **2018**, *9*, 6280–6288.

[355] Goerigk, L.; Kruse, H.; Grimme, S. Benchmarking Density Functional Methods against the S66 and S66x8 Datasets for Non-Covalent Interactions. *ChemPhysChem* **2011**, *12*, 3421–3433.

[356] Kříž, K.; Řezáč, J. Non-Covalent Interactions Atlas Benchmark Data Sets 4: σ-Hole Interactions. *Phys. Chem. Chem. Phys.* **2022**, *24*, 14794–14804.

[357] Řezáč, J. Non-Covalent Interactions Atlas Benchmark Data Sets 5: London Dispersion in an Extended Chemical Space. *Phys. Chem. Chem. Phys.* **2022**, *24*, 14780–14793.

[358] 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.

[359] Otero-de-la-Roza, A.; DiLabio, G. A. Transferable Atom-Centered Potentials for the Correction of Basis Set Incompleteness Errors in Density-Functional Theory. *J. Chem. Theory Comput.* **2017**, *13*, 3505–3524.

[360] Tuca, E.; DiLabio, G. A.; Otero-de-la-Roza, A. Minimal Basis Set Hartree-Fock Corrected with Atom Centered Potentials for Molecular Crystal Modeling and Crystal Structure Prediction. *J. Chem. Inf. Model.* **2022**, *62*, 4107–4121.

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

[362] Dolgonos, G. A.; Hoja, J.; Boese, A. D. Revised Values for the X23 Benchmark Set of Molecular Crystals. *Phys. Chem. Chem. Phys.* **2019**, *21*, 24333–24344.

[363] Brandenburg, J. G.; Maas, T.; Grimme, S. Benchmarking DFT and Semiempirical Methods on Structures and Lattice Energies for Ten Ice Polymorphs. *J. Chem. Phys.* **2015**, *142*, 124104.

[364] Gillan, M. J.; Alfe, D.; Michaelides, A. Perspective: How Good is DFT for Water? *J. Chem. Phys.* **2016**, *144*, 130901.

[365] Becke, A. D. A New Mixing of Hartree–Fock and Local Density-Functional Theories. *J. Chem. Phys.* **1993**, *98*, 1372.

[366] 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.

[367] Hohenstein, E. G.; Sherrill, C. D. Wavefunction Methods Noncovalent Interactions. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2012**, *2*, 304–326.

[368] Brauer, B.; Kesharwani, M. K.; Kozuch, S.; Martin, J. M. L. The S66x8 Benchmark for Noncovalent Interactions Revisited: Explicitly Correlated Ab Initio Methods and Density Functional Theory. *Phys. Chem. Chem. Phys.* **2016**, *18*, 20905–20925.

[369] Giannozzi, P.; Andreussi, O.; Brumme, T.; Bunau, O.; Nardelli, M. B.; Calandra, M.; Car, R.; Cavazzoni, C.; Ceresoli, D.; Cococcioni, M., et al. Advanced Capabilities for Materials Modelling with Quantum ESPRESSO. *J. Phys.: Condens. Matter* **2017**, *29*, 465901.

[370] Parrish, R. M.; Burns, L. A.; Smith, D. G.; Simmonett, A. C.; De-Prince, A. E.; Hohenstein, E. G.; Bozkaya, U.; Sokolov, A. Y.; Di Remigio, R.; Richard, R. M., et al. Psi4 1.1: An Open-Source Electronic Structure Program Emphasizing Automation, Advanced Libraries, and Interoperability. *J. Chem. Theory Comput.* **2017**, *13*, 3185–3197.

[371] The Canoical XDM Coefficients can be Found at https://github.com/aoterodelaroza/postg, in the xdm.param File of the Postg Distribution.

[372] LeBlanc, L. M.; Weatherby, J. A.; Otero-de-la-Roza, A.; Johnson, E. R. 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.

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

[374] Otero-de-la-Roza, A.; DiLabio, G. A.; Johnson, E. R. Exchange-Correlation Effects for Non-Covalent Interactions in Density-Functional Theory. *J. Chem. Theory Comput.* **2016**, *12*, 3160–3175.

[375] Acree Jr, W.; Chickos, J. S. Phase Transition Enthalpy Measurements of Organic and Organometallic Compounds. Sublimation, Vaporization and Fusion Enthalpies from 1880 to 2015. Part 1. C1- C10. *J. Phys. Chem. Ref. Data* **2016**, *45*, 033101.

[376] Carter, D. J.; Rohl, A. L. Benchmarking Calculated Lattice Parameters and Energies of Molecular Crystals Using van der Waals Density Functionals. *J. Chem. Theory Comput.* **2014**, *10*, 3423–3437.

[377] Moellmann, J.; Grimme, S. DFT-D3 Study of Some Molecular Crystals. *J. Phys. Chem. C* **2014**, *118*, 7615–7621.

[378] Thomas, S. P.; Spackman, P. R.; Jayatilaka, D.; Spackman, M. A. Accurate Lattice Energies for Molecular Crystals from Experimental Crystal Structures. *J. Chem. Theory Comput.* **2018**, *14*, 1614–1623.

[379] LeBlanc, L. M.; Otero-de-la-Roza, A.; Johnson, E. R. Composite and Low-Cost Approaches for Molecular Crystal Structure Prediction. *J. Chem. Theory Comput.* **2018**, *14*, 2265–2276.

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

[381] Otero-de-la-Roza, A.; Martín Pendás, A.; Johnson, E. R. Quantitative Electron Delocalization in Solids from Maximally Localized Wannier Functions. *J. Chem. Theory Comput.* **2018**, *14*, 4699–4710.

[382] Rice, B.; LeBlanc, L. M.; Otero-de-la-Roza, A.; Fuchter, M. J.; Johnson, E. R.; Nelson, J.; Jelfs, K. E. A Computational Exploration of the Crystal Energy and Charge Carrier Mobility Landscapes of the Chiral [6]Helicene Molecule. *Nanoscale* **2018**, *10*, 1865–1876.

[383] Price, A. J. A.; Otero de la Roza, A.; Johnson, E. R. XDM-Corrected Hybrid DFT with Numerical Atomic Orbitals Predicts Molecular Crystal Lattice Energies with Unprecedented Accuracy. *Chem. Sci.* **2023**, *14*, 1252–1262, Chapter 6 in this thesis.

[384] Price, A. J. A.; Mayo, R. A.; Otero de la Roza, A.; Johnson, E. R. Accurate and Efficient Polymorph Energy Ranking with XDM-Corrected Hybrid DFT. *CrystEngComm* **2023**, *25*, 953–960, Chapter 7 in this thesis.

[385] Nyman, J.; Day, G. M. Modelling Temperature-Dependent Properties of Polymorphic Organic Molecular Crystals. *Phys. Chem. Chem. Phys.* **2016**, *18*, 31132–31143.

[386] Weatherby, J. A.; Rumson, A. F.; Price, A. J. A.; Otero de la Roza, A.; Johnson, E. R. A Density-Functional Benchmark of Vibrational Free-Energy Corrections for Molecular Crystal Polymorphism. *J. Chem. Phys.* **2022**, *156*, 114108.

[387] Montis, R.; Davey, R. J.; Wright, S. E.; Woollam, G. R.; Cruz-Cabeza, A. J. Transforming Computed Energy Landscapes into Experimental Realities: The Role of Structural Rugosity. *Angew. Chem. Int. Ed.* **2020**, *59*, 20357–20360.

[388] Macrae, C. F.; Sovago, I.; Cottrell, S. J.; Galek, P. T. A.; McCabe, P.; Pidcock, E.; Platings, M.; Shields, G. P.; Stevens, J. S.; Towler, M.; Wood, P. A. *Mercury 4.0*: From Visualization to Analysis, Design and Prediction. *J. Appl. Cryst.* **2020**, *53*, 226–235.

[389] Kazantsev, A. V.; Karamertzanis, P. G.; Adjiman, C. S.; Pantelides, C. C.; Price, S. L.; Galek, P. T. A.; 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.

[390] Price, S. L.; Leslie, M.; Welch, G. W. A.; 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.

[391] Centre, C. C. D. Personal communication to the 7th CSP blind test participants.

[392] 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.

[393] Bhardwaj, R. M.; McMahon, J. A.; Nyman, J.; Price, L. S.; Konar, S.; Oswald, I. D. H.; Pulham, C. R.; Price, S. L.; Reutzel-Edens, S. M. A Prolific Solvate Former, Galunisertib, Under the Pressure of Crystal Structure Prediction, Produces Ten Diverse Polymorphs. *J. Am. Chem. Soc.* **2019**, *141*, 13887–13897.

[394] Manzeli, S.; Ovchinnikov, D.; Pasquier, D.; Yazyev, O. V.; Kis, A. 2D Transition Metal Dichalcogenides. *Nature Reviews Materials* **2017**, *2*, 1–15.

[395] Das, S.; Appenzeller, J. Where Does the Current Flow in Two-dimensional Layered Systems? *Nano Lett.* **2013**, *13*, 3396–3402.

[396] Schulman, D. S.; Arnold, A. J.; Das, S. Contact Engineering for 2D Materials and Devices. *Chem. Soc. Rev.* **2018**, *47*, 3037–3058.

[397] Kam, K. K.; Parkinson, B. A. Detailed Photocurrent Spectroscopy of the Semiconducting Group VIB Transition Metal Dichalcogenides. *J. Phys. Chem.* **1982**, *86*, 463–467.

[398] Splendiani, A.; Sun, L.; Zhang, Y.; Li, T.; Kim, J.; Chim, C.-Y.; Galli, G.; Wang, F. Emerging Photoluminescence in Monolayer MoS_2. *Nano Lett.* **2010**, *10*, 1271–1275.

[399] Ouyang, B.; Xiong, S.; Jing, Y. Tunable Phase Stability and Contact Resistance of Monolayer Transition Metal Dichalcogenides Contacts with Metal. *NPJ 2D Mater. Appl.* **2018**, *2*, 1–8.

[400] Das, S.; Chen, H.-Y.; Penumatcha, A. V.; Appenzeller, J. High Performance Multilayer MoS_2 Transistors with Scandium Contacts. *Nano Lett.* **2013**, *13*, 100–105.

[401] Liu, H.; Neal, A. T.; Ye, P. D. Channel Length Scaling of MoS_2 MOSFETs. *ACS Nano* **2012**, *6*, 8563–8569.

[402] Liu, H.; Si, M.; Najmaei, S.; Neal, A. T.; Du, Y.; Ajayan, P. M.; Lou, J.; Ye, P. D. Statistical Study of Deep Submicron Dual-gated Field-effect Transistors on Monolayer Chemical Vapor Deposition Molybdenum Disulfide Films. *Nano Lett.* **2013**, *13*, 2640–2646.

[403] Allain, A.; Kang, J.; Banerjee, K.; Kis, A. Electrical contacts to two-dimensional semiconductors. *Nature Mater.* **2015**, *14*, 1195–1205.

[404] Kang, J.; Liu, W.; Sarkar, D.; Jena, D.; Banerjee, K. Computational Study of Metal Contacts to Monolayer Transition-Metal Dichalcogenide Semiconductors. *Phys. Rev. X* **2014**, *4*, 031005.

[405] Zhang, P.; Zhang, Y.; Wei, Y.; Jiang, H.; Wang, X.; Gong, Y. Contact Engineering for Two-dimensional Semiconductors. *J. Semicond.* **2020**, *41*, 1–16.

[406] 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.

[407] 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.

[408] Blackmond, D. G. The Origin of Biological Homochirality. *Cold Spring Harb. Perspect. Biol* **2010**, *2*, a002147.

[409] Bonner, W. A.; Kavasmaneck, P. R.; Martin, F. S.; Flores, J. J. Asymmetric Adsorption by Quartz: A Model for the Prebiotic Origin of Optical Activity. *Orig. Life* **1975**, *6*, 367–376.

[410] Hazen, R. M. Mineral Surfaces and the Prebiotic Selection and Organization of Biomolecules. *Am. Min.* **2006**, *91*, 1715.

[411] Sholl, D. S.; Gellman, A. J. Developing Chiral Surfaces for Enantioselective Chemical Processing. *AIChE Journal* **2009**, *55*, 2484–2490.

[412] Hazen, R. M.; Sholl, D. S. Chiral Selection on Inorganic Crystalline Surfaces. *Nat. Mat.* **2003**, *2*, 367–374.

[413] Han, J. W.; Sholl, D. S. Enantiospecific Adsorption of Amino Acids on Hydroxylated Quartz (0001). *Langmuir* **2009**, *25*, 10737–10745.

[414] Han, J. W.; Sholl, D. S. Enantiospecific Adsorption of Amino Acids on Hydroxylated Quartz (1010). *Phys. Chem. Chem. Phys.* **2010**, *12*, 8024–8032.

[415] Price, A. J.; Johnson, E. R. Theoretical Investigation of Amino-Acid Adsorption on Hydroxylated Quartz Surfaces: Dispersion Can Determine Enantioselectivity. *Phys. Chem. Chem. Phys.* **2020**, *22*, 16571–16578.

[416] Ziegler, T.; Rauk, A. Carbon Monoxide, Carbon Monosulfide, Molecular Nitrogen, Phosphorus Trifluoride, and Methyl Isocyanide as. Sigma. Donors and. Pi. Acceptors. A Theoretical Study by the Hartree-Fock-Slater Transition-State Method. *Inorg. Chem.* **1979**, *18*, 1755–1759.

[417] Zhao, L.; von Hopffgarten, M.; Andrada, D. M.; Frenking, G. Energy Decomposition Analysis. *Wires* **2018**, *8*, e1345.

[418] Perdew, J. P.; Parr, R. G.; Levy, M.; Balduz Jr, J. L. Density-Functional Theory for Fractional Particle Number: Derivative Discontinuities of the Energy. *Phys. Rev. Lett.* **1982**, *49*, 1691.

[419] Mori-Sánchez, P.; Cohen, A. J.; Yang, W. Many-Electron Self-Interaction Error in Approximate Density Functionals. *J. Chem. Phys.* **2006**, *125*, 201102.

[420] Gräfenstein, J.; Kraka, E.; Cremer, D. The Impact of the Self-Interaction Error on the Density Functional Theory Description of Dissociating Radical Cations: Ionic and Covalent Dissociation Limits. *J. Chem. Phys.* **2004**, *120*, 524–539.

[421] 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+, Li H+, and Ne 2+. *J. Chem. Phys.* **2007**, *126*, 104102.

[422] Ruzsinszky, A.; Perdew, J. P.; Csonka, G. I. Binding Energy Curves from Nonempirical Density Functionals. I. Covalent Bonds in Closed-Shell and Radical Molecules. *J. Chem. Phys. A* **2005**, *109*, 11006–11014.

[423] Janesko, B. G.; Proynov, E.; Kong, J.; Scalmani, G.; Frisch, M. J. Practical Density Functionals Beyond the Overdelocalization–Underbinding Zero-Sum Game. *J. Chem. Phys. Lett.* **2017**, *8*, 4314–4318.

[424] Sancho-García, J.; Pérez-Jiménez, A. J.; Moscardó, F. Description of C(sp2)-C(sp2) Rotation in Butadiene by Density Functionals. *J. Phys. Chem. A* **2001**, *105*, 11541–11548.

[425] Jacquemin, D.; Femenias, A.; Chermette, H.; Ciofini, I.; Adamo, C.; André, J.-M. i.; Perpète, E. A. Assessment of Several Hybrid DFT Functionals for the Evaluation of Bond Length Alternation of Increasingly Long Oligomers. *J. Phys. Chem. A* **2006**, *110*, 5952–5959.

[426] White, M. A.; Kahwaji, S.; Freitas, V. L.; Siewert, R.; Weatherby, J. A.; Ribeiro da Silva, M. D.; Verevkin, S. P.; Johnson, E. R.; Zwanziger, J. W. The Relative Thermodynamic Stability of Diamond and Graphite. *Angew. Chem.* **2021**, *133*, 1570–1573.

[427] Nam, S.; Cho, E.; Sim, E.; Burke, K. Explaining and Fixing DFT Failures for Torsional Barriers. *J. Phys. Chem. Lett.* **2021**, *12*, 2796–2804.

[428] Beran, G. J.; Sugden, I. J.; Greenwell, C.; Bowskill, D. H.; Pantelides, C. C.; A djiman, C. S. How Many More Polymorphs of ROY Remain Undiscovered. *Chem. Sci.* **2022**, *13*, 1288–1297.

[429] Cohen, A. J.; Mori-Sánchez and W. Yang, P. Challenges for Density Functional Theory. *Chem. Rev.* **2012**, *112*, 289–320.

[430] Zhang, Y.; Yang, W. Perspective on "Density-Functional Theory for Fractional Particle Number: Derivative Discontinuities of the Energy". *Theor. Chem. Acc.* **2000**, 346–348.

[431] Henderson, T. M.; Janesko, B. G.; Scuseria, G. E. Range Separation and Local Hybridization in Density Functional Theory. *J. Phys. Chem. A* **2008**, *112*, 12530–12542.

[432] 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.

[433] Ruzsinszky, A.; Perdew, J. P. Twelve Outstanding Problems in Ground-State Density Functional Theory: A Bouquet of Puzzles. *Comput. Theor. Chem.* **2011**, *963*, 2–6.

[434] Janesko, B. G. Replacing Hybrid Density Functional Theory: Motivation and Recent Advances. *Chem. Soc. Rev* **2021**, *50*, 8470–8495.

[435] Becke, A. D. Real-Space Post-Hartree–Fock Correlation Models. *J. Chem. Phys.* **2005**, *122*, 064101.