

University of Sheffield

Exploring the Landscape of σ -hole Interactions



Harry W. Nash

Supervisors: Dr Grant Hill & Prof. Lee Brammer

A thesis submitted as part of the fulfilment of the requirements for a PhD in Chemistry

July 29, 2025

Acknowledgements

"When the fun stops, double down." - J. Preston

What an ordeal this has been. No one could have prepared me for this undertaking, yet now it draws to a close. Thank **** for that. I would like to take the time to acknowledge the wonderful people who have supported me and helped make this possible.

To Elna and Jo, and to everyone at CCDC, thank you so much for all the tech support, encouragement, and community over these past few years. Oh and for funding this project!

Lee, you have been a great help throughout, reassuring me and giving confidence that I am doing the right thing, along with providing a different and fresh perspective. Tom, Josh and the rest of the Brammer group, thank you being so welcoming and for making me less scared of crystallography.

Grant, well, where would I be without you? The reason I am here in the first place, and a key factor in my success. You have been a truly fantastic supervisor during both masters and PhD, and I count you as a true friend. Amongst the many members of the Hill Group, past and present, I must thank Zuzanna, Adam, Rob, and Shaun for their help and support along the way. I have thoroughly enjoyed being a bad influence on you all.

I couldn't have made it this far with my sanity intact if not for my pals, especially Henry, Craigh, Michael, Ed, James, and Travis, now they no longer have to hear me moaning about my PhD. I could not possibly forget the absolute darling that is Ciaran, who is a brother to me in the most uniquely weird way.

Of course I have to thank my family for their unending support throughout my life and my time at university. I cannot believe how lucky I am to have the opportunities to get this far, and I truly could not have done it without them.

And last, but by no means least, I want to thank the love of my life, Jaz. Going through this PhD experience with you has made it all worthwhile, and something I can never regret. You have shown me what it means to be a team, and helped make me the man I am today. I love you to the moon! Thank you for everything that you have done, and I look forward to the future with you.

Abstract

σ -hole interactions are a class of noncovalent interactions involving p -block elements. Comparable in strength and directionality to hydrogen bonds, σ -hole interactions present a useful tool for a wide range of applications.

The σ -hole arises when a polarisable p -block element is covalently bonded to an electron withdrawing group (EWG). The σ -bond to the EWG creates a charge anisotropy, resulting in a depletion of electron density along the extension of the σ -bond, the electropositive σ -hole.

Detailed searches of the CSD were constructed to investigate the geometric characteristics of σ -hole interactions to various nitrogen and oxygen lone-pair-possessing Lewis bases. Trends in the identity of the σ -hole donor atom and EWG were explored. The σ -hole acceptors also play a key role in the interactions and so were investigated on a functional group basis. Nitrogen Lewis bases are generally better σ -hole acceptors than oxygen Lewis bases, but oxygen Lewis bases show greater variety.

To predict the interaction energies for σ -hole interactions, the σ -Hole Approximation Tool (σ HAT) was constructed. Based on a XGBoost regression machine learning model, and trained on geometric measurements and adapted molecular fingerprints of σ -hole complexes from a data set of high accuracy calculated σ -hole interactions. This model performs comparably to density functional theory (DFT) on the test set, with an RMSD of 1.03 kcal/mol, which is very close to the 1 kcal/mol "chemical accuracy" threshold.

This model was validated using cutouts derived from crystal structures in the CSD, which possess halogen bonds (a subset of σ -hole interactions). DFT calculated interaction energies for these cutouts were used to compare against the σ HAT predicted interaction energies. The performance on this validation set showed an RMSD of 2.02 kcal/mol, but differed depending on the halogen bond donor atom, suggesting different character of the interactions.

Contents

1	Introduction	5
1.1	Overview	5
1.2	Types of σ -hole Interaction	7
1.2.1	Halogen Bonding	7
1.2.2	Chalcogen Bonding	9
1.2.3	Pnictogen Bonding	10
1.2.4	Tetrel Bonding	11
1.2.5	Triel & Aerogen Bonding	13
1.3	The Nature of σ -hole Interactions	13
1.3.1	Composition of the Interaction Energy	14
1.3.2	Nature of the Acceptor	15
1.4	Searching the Cambridge Structural Database	16
1.5	Predictive Models for Noncovalent Interaction Energies	17
1.6	Aims and Objectives	18
2	Theory & Methodology	19
2.1	CSD Search Methodology	19
2.1.1	Query & Search Construction	20
2.1.2	On van der Waals Radii	22
2.2	Calculation of Non-covalent Interaction Energies	23
2.2.1	Density Functional Theory	26
2.2.2	Basis Set Superposition Error	29
3	Nitrogen as an Acceptor	30
3.1	Introduction to Nitrogen Lewis Bases	30
3.2	CSD Query Details	31
3.3	sp^3 Hybridised Nitrogen	32

3.3.1	$X \cdots sp^3$ -Nitrogen Contacts	32
3.3.2	$Ch \cdots sp^3$ -Nitrogen Contacts	34
3.3.3	$Pn \cdots sp^3$ -Nitrogen Contacts	36
3.3.4	$Tt \cdots sp^3$ -Nitrogen Contacts	38
3.3.5	Prevalence of $\sigma_{hole} \cdots sp^3$ -Nitrogen Contacts	40
3.3.6	Notable sp^3 -Nitrogen Moieties	42
3.4	sp^2 Hybridised Nitrogen	44
3.4.1	$X \cdots sp^2$ -Nitrogen Contacts	44
3.4.2	$Ch \cdots sp^2$ -Nitrogen Contacts	46
3.4.3	$Pn \cdots sp^2$ -Nitrogen Contacts	48
3.4.4	$Tt \cdots sp^2$ -Nitrogen Contacts	50
3.4.5	Prevalence of $\sigma_{hole} \cdots sp^2$ -Nitrogen Contacts	52
3.4.6	Notable sp^2 -Nitrogen Moieties	54
3.5	sp -Hybridised Nitrogen	57
3.5.1	$X \cdots sp$ -Nitrogen Contacts	57
3.5.2	$Ch \cdots sp$ -Nitrogen Contacts	59
3.5.3	$Pn \cdots sp$ -Nitrogen Contacts	61
3.5.4	$Tt \cdots sp$ -Nitrogen Contacts	63
3.5.5	Prevalence of $\sigma_{hole} \cdots sp$ -Nitrogen Contacts	65
3.5.6	Notable sp -Nitrogen Moieties	67
3.6	Trends in the σ -hole Donor Atom	68
3.6.1	Effect of Donor Electron-Withdrawing Group	70
3.7	Summary	76
4	Oxygen as an Acceptor	77
4.1	Introduction to Oxygen Lewis Bases	77
4.2	CSD Query Details	78
4.3	Organic Carbonyls	80
4.3.1	$X \cdots$ Organic Carbonyl Contacts	81
4.3.2	$Ch \cdots$ Organic Carbonyl Contacts	86
4.3.3	$Pn \cdots$ Organic Carbonyl Contacts	91
4.3.4	$Tt \cdots$ Organic Carbonyl Contacts	95
4.3.5	Prevalence of $\sigma_{hole} \cdots$ Organic Carbonyl Oxygen Contacts	99
4.3.6	Notable Organic Carbonyl Moieties	103
4.4	Divalent Oxygen	106

4.4.1	<i>X</i> ··· Divalent Oxygen Contacts	107
4.4.2	<i>Ch</i> ··· Divalent Oxygen Contacts	112
4.4.3	<i>Pn</i> ··· Divalent Oxygen Contacts	117
4.4.4	<i>Tt</i> ··· Divalent Oxygen Contacts	122
4.4.5	Prevalence of σ_{hole} ··· Divalent Oxygen Contacts	127
4.4.6	Notable Divalent Oxygen Functional Groups	131
4.5	Oxides	134
4.5.1	<i>X</i> ··· Oxide Oxygen Contacts	135
4.5.2	<i>Ch</i> ··· Oxide Oxygen Contacts	141
4.5.3	<i>Pn</i> ··· Oxide Oxygen Contacts	146
4.5.4	<i>Tt</i> ··· Oxide Oxygen Contacts	151
4.5.5	Prevalence of σ_{hole} ··· Oxide Oxygen Contacts	157
4.5.6	Notable Oxide Functional Groups	161
4.6	Metal Carbonyls	165
4.6.1	<i>X</i> ··· Metal Carbonyl Oxygen Contacts	166
4.6.2	<i>Ch</i> ··· Metal Carbonyl Oxygen Contacts	169
4.6.3	<i>Pn</i> ··· Metal Carbonyl Oxygen Contacts	171
4.6.4	<i>Tt</i> ··· Metal Carbonyl Oxygen Contacts	173
4.6.5	Prevalence of σ_{hole} ··· Metal Carbonyl Oxygen Contacts	175
4.6.6	Notable Metal Carbonyl Functional Groups	179
4.7	Summary	183
5	Machine Learning Prediction of σ-hole Interaction Energies	187
5.1	Overview	187
5.1.1	Introduction to Machine Learning in Chemistry	187
5.1.2	General Considerations for ML Applications	188
5.1.3	ML Approaches to Predicting Noncovalent Interactions	189
5.2	The SHEF1552 σ -hole dataset	190
5.2.1	Train-Test Split	193
5.3	Descriptors	199
5.3.1	Molecular Fingerprints	200
5.3.2	Dimensionality Reduction	202
5.3.3	Descriptor Selection for σ -hole Interaction Energy Prediction	203
5.4	Models	204
5.4.1	Model Selection for σ -hole Interaction Energy Prediction	206

5.5	The σ -Hole Approximation Tool (σ HAT)	207
5.5.1	Fitting & Hyperparameter Optimisation	207
5.5.2	Validation Performance	211
5.6	Summary	220
6	Conclusions & Future Work	222
6.1	Conclusions	222
6.2	Future Work	225
	Appendix A CSD Geometric Data	227
A.1	Nitrogen Heterocycle Queries	227
A.2	Effect of Donor Electron-Withdrawing Group Continued	228
	Appendix B Crystal Structures	236
	Appendix C ML Work Appendices	239
C.1	Fingerprint Performance	239
C.2	Random Forest & Linear Regression Performance	242
C.3	σ HAT Model Details	243
	SHEF1552 σ -hole data set	244

Chapter 1

Introduction

1.1 Overview

Non-covalent interactions can be intra- or intermolecular and, although generally weaker than covalent bonds that dominate much of chemistry, they play a crucial role in supramolecular assembly and in molecular conformations. Non-covalent interactions have been shown to play key roles in protein-drug interactions, catalyst-substrate interactions, nanomaterial self assembly, crystal engineering, and a range of different chemical reactions¹⁻⁵.

This thesis focuses on a family of non-covalent interactions known as σ -hole interactions. These σ -holes, as first coined by Clark, are regions of positive electrostatic potential on an atom's surface, opposite to and along the extension of a covalent bond⁶. This positive potential is results from the anisotropic charge distribution around an atom caused by a σ -bond to an electron-withdrawing substituent. An example of a σ -hole is shown in Figure 1.1 on the molecular electrostatic potential of FCl, where the covalent bond to fluorine creates a σ -hole on the chlorine. The σ -hole phenomenon has recently been imaged at subatomic resolution via Kelvin probe force microscopy, a form of non-contact atomic force microscopy. This demonstrates what had only been previously determined indirectly via diffraction or theoretical means⁷.

These electropositive σ -holes can form attractive interactions with Lewis bases, for example; anions, lone-pair-containing groups, and π -systems⁸. The groups of elements that can form these σ -holes are groups 13-18 of the Periodic Table, and the subfamilies of this class of interaction are known as triel bonding (group 13), tetrel bonding (group 14), pnictogen bonding (group 15), chalcogen bonding (group 16), halogen bonding (group 17), and aerogen bonding (group 18). σ -hole interactions are analogous to hydrogen bonding, both arising from charge anisotropy and forming in a linear manner, with comparable interaction strength⁹.

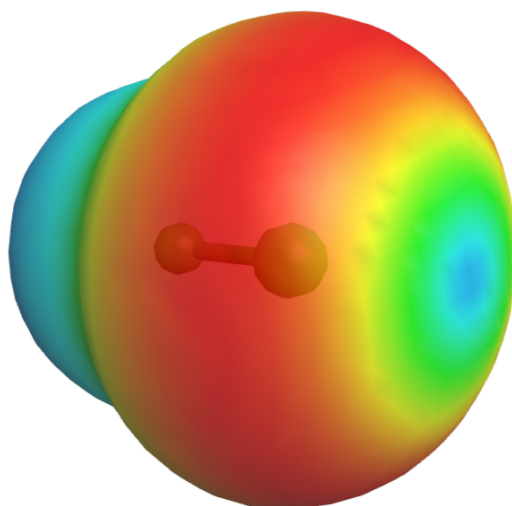


Figure 1.1: Molecular electrostatic surface potential showing the σ -hole of FCl. Red denotes negative electrostatic potential, and blue denotes positive electrostatic potential. Calculated at B3-LYP/6-31G level of theory on a 0.001 a.u. electron density isosurface.

Non-covalent interactions involving these groups have been known about for some time, especially in the case of halogen bonds, but their nature and similarity were not well understood. In recent years, the wide-reaching significance of these interactions has been a major area of study in many disciplines of chemistry due to the abundance of the involved elements in chemical and biological systems. The quantitative measurement of the strength of noncovalent interactions have proven to cause some difficulty experimentally, with it being hard to resolve the effects of multiple competing influences and solvent effects that may dictate the experimental outcome¹⁰. This means that theoretical methods must be employed to realise the characteristics of these interactions and identify their importance in experimental systems.

The Cambridge Structural Database (CSD) is a widely used database of small-molecule organic and metal-organic crystal structures, with over one million accurate 3D structures deposited and curated by the Cambridge Crystallographic Data Centre (CCDC)¹¹. Founded in 1965, it has been a vital tool for research concerning drug discovery, novel materials, chemical and crystallographic phenomena for decades. Database inquiry tools created by the CCDC have been used to scour this database and find common trends and relationships within these crystal structures^{12 13}. This has led to subsets of the CSD being created such as the CSD MOF collection, which contains over 100,000 metal-organic frameworks (MOFs), which has led to a greater understanding of the conserved motifs in these structures and serves as a platform for further research¹⁴. Information obtained from searches of the CSD, such as geometric measurements, and any trends in such data can be interpreted to gain insight into various chemical properties in the solid state.

1.2 Types of σ -hole Interaction

1.2.1 Halogen Bonding

The archetypal example of σ -hole interactions is that of halogen bonding. Long before the σ -hole model was first posited, several attractive complexes of dihalogens with Lewis bases were discovered. The first known halogen bonding complexes were accidentally produced by Colin in 1813, resulting in blue-black coloured product when combining iodine with both amylose and gaseous ammonia¹⁵. The composition of the iodine-ammonia product was not proposed 1863 by Guthrie after mixing powdered iodine and aqueous ammonia¹⁶. However, it was not until much later that the true nature of the interaction was determined, which will be covered below. Bromine and chlorine acting as halogen bond donors was first observed in the late 19th century as Remsen and Norris each described the formation of 1:1 dimers of Cl_2 and Br_2 with several amines¹⁷. A halogen bonding adduct involving fluorine was not seen until 1979 with Miller & Andrews isolating the F_3^- anion¹⁸.

X-ray crystallographic studies on complexes of dihalogens with Lewis bases, such as Hassel's work on $\text{Br}_2 \cdots \text{O}(\text{CH}_2\text{CH}_2)\text{O}$ adducts, helped to realise the configuration of halogen bonds in the solid state¹⁹. Hassel described the complexes as a chain of "halogen molecule bridges", finding the $\text{Br} \cdots \text{O}$ intermolecular distance to be less than the sum of their van der Waals radii and having a $\text{Br}-\text{Br} \cdots \text{O}$ angle close to 180° . Hassel also published research on crystal structures of $\text{Br}_2 \cdots$ benzene and $\text{Cl}_2 \cdots$ benzene complexes, demonstrating the ability of π systems to act as halogen bond acceptors in the solid state²⁰. An interesting set of interactions was found in 1963 by Sakurai *et al.* of $\text{R}-\text{X} \cdots \text{X}-\text{R}$ halogen-halogen complexes²¹. This raised some questions as halogens are generally considered to be negative centres and thus the like-like interaction should be expected to be repulsive. However, these contacts were found to favour two modes of binding, later classified by Desiraju and Parthasarathy into type I and type II contacts, see Figure 1.2²². Type I contacts ($\theta_1 \approx \theta_2$) are consequences of close packing requirements and are not true halogen bonds. However, type II contacts ($\theta_1 \approx 180^\circ$ and $\theta_2 \approx 90^\circ$) are true halogen bonds according to the IUPAC (The International Union of Pure and Applied Chemistry) definition (see below) and are the result of the σ -hole of one halogen interacting with the electron-rich equatorial region of the other halogen.

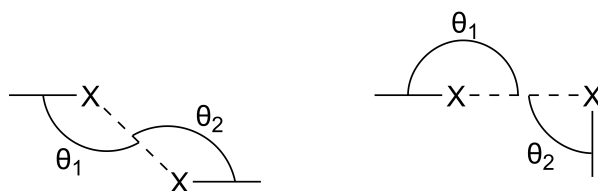


Figure 1.2: An illustration of Type I halogen $\cdots$ halogen contacts, where $\theta_1 \approx \theta_2$ (left) and Type II (right) halogen $\cdots$ halogen contacts, where $\theta_1 \approx 180^\circ$, $\theta_2 \approx 90^\circ$. Recreated from Metrangolo, P. & Resnati, G. (2014). *IUCrJ*, 1, 5-7.²³

The strength of the halogen bond increases with the size of the halogen bond donor - $\text{F} < \text{Cl} < \text{Br} < \text{I}$, see Table 1.1. This is due to the larger atomic size making it more polarisable, and thus needing a less electron-withdrawing group to

cause the charge anisotropy required to form a σ -hole. Fluorine can only be considered a halogen bond donor when it is covalently bonded to very strongly electron-withdrawing groups, due to its small size and high charge density²⁴. Astatine, the largest naturally occurring halogen, is an unstable radioactive element and its most stable isotope only has a half life of around eight hours, so very little is known about its chemistry. Astatine is calculated to have a higher polarisability than iodine, so theoretically could function as an even stronger halogen bond donor than iodine²⁵. However, there have been recent theoretical studies into astatine halogen bonding, which is calculated to be weaker than for iodine due to spin-orbit coupling effects^{26 27}.

Table 1.1: Table of static scalar dipole polarisabilities(atomic units) for neutral halogens recreated from *Molecular Physics*, 117(9–12), 1200–1225²⁸. These static scalar dipole polarisabilities are a measure of how easily the electron cloud of the neutral atom can be distorted by a static external electric field, resulting in a dipole.

Atom	Dipole Polarisability
F	3.7
Cl	14.6
Br	21.0
I	32.9

Halogen bonding can be of comparable strength to hydrogen bonding, sometimes even exceeding it, although the two non-covalent interactions are not fully alike^{3 29 30}. Halogen atoms are considerably larger than hydrogen, with the van der Waals radii of F, Cl, Br, I being 1.47, 1.75, 1.85, and 1.98 Å respectively (compared to 1.2 Å of hydrogen)¹⁶. Due to their larger size, this makes halogens more susceptible to steric hindrance of their noncovalent interactions than hydrogens, thus their ability to form halogen bonds is more likely to be impeded by the steric bulk of the acceptor, which can be seen in some examples where halogen bonds with bromine as the donor are stronger than with iodine as the donor, even though iodine is more polarisable³¹.

The exact point when the term halogen bond was coined is not clear, however it became commonly used after a review by Legon in 1999³². The term has now been accepted and defined by IUPAC which recommended the definition of the halogen bond from 2013³³:

"A halogen bond occurs when there is evidence of a net attractive interaction between an electrophilic region associated with a halogen atom in a molecular entity and a nucleophilic region in another, or the same molecular entity. A typical halogen bond is denoted by the three dots in $R-X \cdots Y$. $R-X$ is the halogen bond donor, X is any halogen atom with an electrophilic region, and R is a group covalently bound to X . In some causes, X may be covalently bound to more than one group. It may also form more than one halogen bond. Y is the halogen bond acceptor and is typically a molecular entity possessing at least one nucleophilic region."

In this IUPAC recommendation further characteristics are defined for a typical halogen bond, namely the $R-X \cdots Y$ angle tending to be close to 180° and the $X \cdots Y$ contact distance being shorter than the sum of their van der Waals radii. The halogen bond strength increases with the electron-withdrawing ability of R , and decreases with increasing electronegativity of X . The nature of the acceptor Y also plays a key role in halogen bond strength, with a stronger Lewis base producing

a stronger interaction with the halogen bond donor.

Although most typically halogen bonding involves monovalent halogen species, hypervalent halogen structures have also been reported as halogen bond donors. Hypervalent iodine(III) salts have been determined to be biaxial halogen bond donors, having two electrophilic regions compared to the one in iodine(I) complexes³⁴. Halonium salts containing iodine and bromine have also been shown to be halogen bond donors in X-ray crystallographic and theoretical studies¹⁶. Scheiner *et al.* have made theoretical investigations into pentavalent halogen species of halopentafluorides, which shows that greater valency of the halogen bond donor increases the sensitivity to steric effects of the surroundings³⁵.

The sensitivity of halogen bonding to both the size of halogen atom and the electron-withdrawing ability of its covalently bonded substituents brings about the desirable property of a tuneable stabilising interaction. This, and the strong directionality of the bond, meaning that the interaction forms specifically near 180° along the extension of the covalent bond, make it of great interest for the formation of supramolecular structures, drug development, and many other applications.

1.2.2 Chalcogen Bonding

As the understanding of the characteristics of halogen bonding began to develop, curiosity regarding whether other groups of elements could partake in a similar type of non-covalent interaction increased. It has been known for a long time that compounds containing elements in group 16, the chalcogen group, display amphiphilic character caused by anisotropic charge distributions. As early as the 1930s selenium and tellurium species were found to show what were then known as secondary bonding interactions with linear contacts to oxygen, which we now understand to be chalcogen bonds^{36–38}. Surveys of the Protein Data Bank (PDB) have revealed the regular occurrence of S ... O interactions that play an important role in stabilising three-dimensional structure and enabling overall function³⁹.

The main difference compared to halogen bonding is that chalcogens typically have more than one covalently bonded substituent, thus have the potential to form more than one σ -hole. Chalcogen bonding primarily involves divalent, neutral chalcogenic species, however higher valency charge-assisted chalcogen bonds have been observed⁴⁰. Hypervalent chalcogen compounds have also been reported to act as chalcogen bond donors, with tetravalent sulfur compounds found in CSD searches and shown using theoretical methods to possess a σ -hole and undergo chalcogen bonding⁴¹. Tetravalent chalcogens as chalcogen bond donors are also reported in the theoretical studies of Scheiner and co-workers⁴².

The naming of these non-covalent interactions as chalcogen bonds was first posited by Minyaev and Minkin in 1998, but the name has now become widely used in the literature⁴³. In 2019 IUPAC published in their recommendations a definition for the chalcogen bond⁴⁴:

"a net attractive interaction between an electrophilic region associated with a chalcogen atom in a molecular entity and a nucleophilic region in another, or the same, molecular entity."

IUPAC goes on to define a list of features of a typical chalcogen bond, including the interatomic distance between donor and acceptor being less than the sum of their van der Waals radii. However chalcogen bonds often form with some deviation from linearity with respect to the extension of the covalently bonded electron-withdrawing group, this is likely due to the

presence of other substituents covalently bonded to the chalcogen, and/or the presence of a lone-pair on the chalcogen. Similar trends apply to chalcogen bonding as to halogen bonding, with the strength of the interaction increasing with the polarisability of the atom: $O < S < Se < Te$, see Table 1.2. Much like the ability of fluorine to partake in halogen bonding, oxygen requires strongly electron-withdrawing groups covalently bonded to it in order to bring about significant charge anisotropy due to oxygen’s small size and charge density. Although polonium has a much greater stability than astatine and is more polarisable than tellurium, no research has been done concerning its ability to form a σ -hole or chalcogen bond.

Table 1.2: Table of static scalar dipole polarisabilities in atomic units for chalcogens recreated from *Molecular Physics*, 117(9–12), 1200–1225²⁸. These static scalar dipole polarisabilities are a measure of how easily the electron cloud of the neutral atom can be distorted by a static external electric field, resulting in a dipole.

Atom	Dipole Polarisability
O	5.3
S	19.4
Se	28.9
Te	38.0

Due to the typical chalcogen bond donor having two covalently bonded constituents, of the form R_2Ch , where $Ch = S, Se, Te$, this offers a greater ability to tune the character of its σ -hole bonding behaviour. However, this additional coordination to the Lewis acidic atom means that chalcogen bonding is more significantly affected by steric crowding of the interacting pair’s moieties. The fact that chalcogens can have more than one σ -hole opens up more options for stabilising interactions and the prevalence of sulfur in many disciplines of chemistry, especially in biological systems, makes chalcogen bonding and the nature of its bonding of significant interest^{45–48}.

1.2.3 Pnictogen Bonding

Analogous to the above behaviour of chalcogens and halogens, group 15 the pnictogen group also shows the amphiphilic character indicative of anisotropic charge distribution. Despite this, compared to halogen and chalcogen bonding, research into pnictogen bonding is far less extensive to date. Theoretical modelling of the pnictogen bond and the effects of substituents have been investigated but far greater focus is given to σ -hole and π -hole interactions of groups 14, 16 and 17, along with hydrogen bonding⁴⁹.

A study from 1995 by Matile and co-workers show intramolecular $P \cdots N$ contacts in hypervalent phosphorous². Further phosphorous donor $\cdots$ acceptor interactions have also been observed in NMR and X-ray crystallographic structural analysis⁵⁰. Arsenic has been found to form pnictogen $\cdots \pi$ interactions from searches of the CSD combined with NMR and X-ray crystallographic study⁵¹. Pnictogen bonds involving antimony have been observed in multiple systems, including alkoxide cages as investigated by theoretical and X-ray crystallographic means⁵². Weak pnictogen bonding has been realised in bismuth structures, with the change in $^{209}Bi - ^{31}P$ coupling observed in its ^{31}P NMR spectra giving insight into the specific interaction energy⁵³.

The term pnictogen bond was first employed to describe this type of interaction in publications of the Hey-Hawkins group in 2011, but now the name has become accepted by the wider scientific community⁵⁴. IUPAC has recently recommended a definition for pnictogen bonding as part of the same initiative that led to the above definitions for halogen and chalcogen bonds⁵⁵.

"A weak attractive interaction between an electrophilic region on a pnictogen atom in a molecular entity (wherein the pnictogen is involved in other stronger bonds) and a nucleophilic region in another, or the same, molecular entity"

Pnictogen bonding follows very similar trends as chalcogen and halogen bonding, with the more polarisable species showing greater tendency to form these non-covalent interactions, following the trend $N < P < As < Sb < Bi$, see Table 1.3. Much like chalcogen bonds, pnictogen bonds form with a slight deviation from linearity with respect to the covalently bonded electron-withdrawing group, likely due to the presence of other substituents on the pnictogen and its lone-pair.

Table 1.3: Table of static scalar dipole polarisabilities in atomic units for pnictogens recreated from *Molecular Physics*, 117(9–12), 1200–1225²⁸. These static scalar dipole polarisabilities are a measure of how easily the electron cloud of the neutral atom can be distorted by a static external electric field, resulting in a dipole.

Atom	Dipole Polarisability
N	7.4
P	25.0
As	30.0
Sb	43.0
Bi	48.0

A typical pnictogen bond donor would be trivalent, following the general form R_3Pn , where $Pn = P, As, Sb, Bi$, allowing the tuning of the interaction by varying the nature of the covalently bonded substituents R. Hypercoordinated species are also capable of pnictogen bonding, with valencies up to pentavalent pnictogens being investigated by theoretical studies⁵⁶. Charge-assisted pnictogen bonds have also been observed in phosphonium salts in CSD searches and theoretical calculations by Gomila and Frontera⁵⁷. The tuneability of pnictogen bonding makes the interaction of interest mainly for the control of molecular self assembly and catalysis, with less interest from a medicinal standpoint, given that pnictogens have a less abundant role in biological systems.

1.2.4 Tetrel Bonding

Group 14, known as the tetrel (Tt) group, show the same charge anisotropy as described above, allowing them to act as Lewis acids and potentially as σ -hole donors. An early example of tetrel bonding from 1997 by Mitzel *et al.* shows short contacts between Si and N atoms in crystal structures of $Si(ONMe_2)_4$ ⁵⁸. Germanium crystal structures have been observed in CSD studies to act as a tetrel bond donor to Lewis bases, for example chloro-tris((2-methoxymethyl)phenyl)germanium showing $Cl-Ge \cdots O$ tetrel bonds to oxygen⁵⁹. Tetrel bonding interactions have been observed in tin complexes such as R_2Sn complexes of salicylaldehyde acyldihydrazone as reported by Baker and co-workers⁶⁰. Organo-lead compounds

have been found to display tetrel bonding, such as diphenyl-(2,20-sulfanediyldibenzenethiolato-S,S0)-lead undergoing self assembly to form dimers in the solid state via tetrel bonding between the Pb and S centres⁶¹. Theoretical studies by Grabowski indicate tetrel bonding to tetravalent tin and lead centres can cause them to be considered penta- or hexavalent when non-covalently bonded to certain acceptors, showing the continuum between non-covalent interaction and hypervalency⁶².

The term tetrel bonding was first coined in a paper by Bauza, Mooibroek, and Frontera in 2013 as a continuation of the σ -hole interaction family⁶³. Scheiner posits a definition for the tetrel bond as⁶⁴:

"An attractive interaction in which a tetrel atom serves as electron acceptor in a Lewis acid, to an electron-donating Lewis base."

Much like the above non-covalent interactions, one might anticipate that the strength of the bond increases with the trend of the size of the tetrel atom $C < Si < Ge < Sn < Pb$, which normally relates to the polarisability of the tetrel atom. However, shown in Table 1.4, the polarisability of lead is notably lower than the polarisability of tin. This is reflected in the strength of tetrel bonds for tetrel bond donors up to tin, yet the strength of lead tetrel bonding interactions is very similar to those of tin^{65,66}. Tetrrels, as their nomenclature indicates, have a strong preference for tetravalency with the form R_4Tt , where $Tt = Si, Ge, Sn, Pb$. This makes the approach of a nucleophile along the extension of one of the covalent bonds much more susceptible to steric hindrance. These steric effects give an extra characteristic of tetrel bonding in that the formation of the non-covalent interaction causes a distortion of the bond angles of the covalent bonds away from the tetrel bond, as shown in Figure 1.3⁶⁴. Analogous to the rest of the first row of halogens, chalcogens, and pnictogens, carbon requires very strongly withdrawing electron-withdrawing constituents in order to form a significant electropositive potential to form a tetrel bond⁶⁷.

Table 1.4: Table of static scalar dipole polarisabilities in atomic units for tetrrels recreated from *Molecular Physics*, 117(9–12), 1200–1225²⁸. These static scalar dipole polarisabilities are a measure of how easily the electron cloud of the neutral atom can be distorted by a static external electric field, resulting in a dipole.

Atom	Dipole Polarizability
C	11.3
Si	37.3
Ge	40.0
Sn	53.0
Pb	47.0

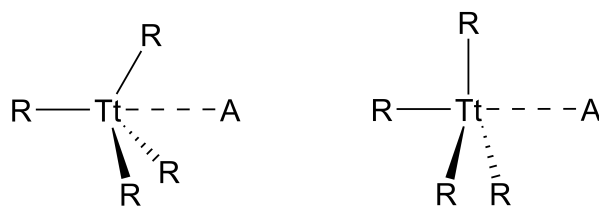


Figure 1.3: An illustration the distortion of a tetrel bond donor from a tetrahedral-like geometry to a trigonal-bipyramidal-like geometry on formation of a tetrel bond with acceptor A.

1.2.5 Triel & Aerogen Bonding

Group 13, the triel group, participate in non-covalent interactions, but the majority of these interactions, known as triel bonds, are a type of π -hole interaction. π -hole interactions are similar to σ -hole interactions, but the positive region of electrostatic potential is found perpendicular to the covalent bond rather than along the extension of it⁶⁸. Scheiner and co-workers have undertaken two theoretical studies on the triel groups ability to form σ -hole interactions in highly fluorinated species, however even with these highly electron-withdrawing constituents these interactions prove to be very weak in comparison to the π -hole interactions that triels are known for^{69 70}.

Group 18, the noble gases or aerogen group, although studied extensively in the past, have been found to participate in σ -hole interactions, with Frontera and co-workers publishing theoretical studies of XeO₃ and XeFO₂ that indeed reveal the presence of a σ -hole. This group's other work on aerogenic species includes theoretical work on the prevalence of the π -hole interactions which they support with structures found in searches of the CSD^{71 72}.

1.3 The Nature of σ -hole Interactions

Before going further, it is important to emphasise that the σ -hole is just a model for this class of non-covalent interaction, and that several other models have been proposed over the years that still have merit. Clark and co-workers proposed the σ -hole model in 2007, which has gone on to become widely accepted by the scientific community, but non-covalent interactions that fit the same criteria have been noted in literature for decades⁷³.

Valence Shell Electron Pair Repulsion (VSEPR) theory can be used to evaluate the geometric preferences of these non-covalently bound complexes, which can be considered to be "incipient hypervalency"⁷⁴. This approach emphasises efficient space filling and the avoidance of lone-pair directions the optimal arrangement of chemical bonds and lone-pairs. As an example system, take the halogen bonding complex R-I $\cdots$ NR'₃. If considering the iodine to have become hypervalent it adopts an octahedral geometric arrangement of electron pairs around the central iodine atom, with R and NR'₃ in the axial positions, and the three lone pairs occupying the equatorial positions. This model predicts the linearity of these interactions but gives little insight as to the nature of the interaction.

In the past, σ -hole interactions were mostly referred to as secondary bonding interactions and were most commonly realised through a charge-transfer lens. A common approach to these interactions was justified by molecular orbital (MO) theory, in that the σ^* orbital, which is located opposite and along the extension of the σ -bond, would receive lone pair donation from the Lewis base, forming an $n \rightarrow \sigma^*$ interaction⁷⁵. This model can explain the linearity of interaction, with the σ^* orbital lying along the extension of the σ -bond, and the lengthening of the σ -bond associated with the non-covalent interaction due to the donation into an antibonding orbital.

Another model for these interactions is that of Mulliken complexes, which gives a discrete difference between a mostly electrostatic interaction, and a charge-transfer system - being denoted respectively as outer complexes and inner complexes. Using notation proposed with respect to halogen bonding, outer complexes are weakly bound and have the form B $\cdots$ X-A,

whereas inner complexes are more tightly bound with a defined charge-transfer of the form $[R-D^+]\cdots[A^-]$. In this example B represents a Lewis base and XA represents a dihalogen or other halogen bond donor⁷⁶. However, this model does not have clear criteria for differentiating between the two complexes, beyond the emphasis on charge-transfer⁷⁷.

When the σ -hole model was first posited, the interaction was considered to be purely electrostatic in nature. The main argument for this is the excellent correlation shown between the strength of the interaction and the magnitude of the electropositive maxima of the electrostatic surface potential⁷⁸. The size and magnitude of the σ -hole increases with the polarisability of the atom, and with the electron-withdrawing ability of the covalently bound substituent. This provides an explanation for the tunability and directionality of these interactions. This purely electrostatic σ -hole model does not adequately explain the lengthening of the covalent bond that is associated experimentally with these non-covalent interactions. The σ -hole model has since been revisited to emphasise the importance of polarisation in causing the necessary charge anisotropy, which can also be used to justify this lengthening of the covalent bond⁷⁹.

There is a striking amount of similarity between the charge-transfer model and the σ -hole model, which is understandable in that they are both trying to explain the same phenomenon via different semantics. The σ^* orbital as used in the charge-transfer model is co-localised with that of the σ -hole as the majority of the electron density is localised in the σ bond. Whether there is a distinction between polarisation and charge transfer is a widely argued topic, with some considering charge-transfer to be merely an extreme form of polarisation⁸⁰. Although electrostatics do explain a large component of the interaction energy, it clearly is not the sole component in interactions of these species. There are notable edge cases that are poorly explained by a σ -hole model and more adequately realised by that of charge-transfer, such as the work of Huber and co-workers demonstrating a family of halogen bonding adducts that follow trends in interaction energy that are contrary to what is expected from a σ -hole model but do follow the trends of charge donation⁸¹. As such, the different components of these non-covalent interactions must be taken into account to get a full understanding of their nature, and taking any of these bonding models at face value is an oversimplification.

1.3.1 Composition of the Interaction Energy

The partitioning of interaction energies into discrete components can be viewed as controversial. Taking a physical observable interaction energy and trying to extract different contributions that have different distance dependencies has no unique rigorous basis, this has led to multiple different approaches that each have their own strengths and weaknesses^{82 83}. The typical components that make up non-covalent interactions are electrostatics, polarisation (often referred to as induction), exchange, and dispersion. Charge-transfer is often included as part of the polarisation term in standard energy decomposition analysis (EDA) schemes and thus quantifying the charge-transfer requires a separate calculation. Electrostatics refers to the Coulombic interaction between the two unperturbed interacting fragments in their frozen geometry, and is usually attractive. Polarisation describes the mixing of virtual and occupied orbitals in each fragment and associated charge rearrangements, which results in a stabilising effect. Exchange refers to that of Pauli exchange, which is a repulsive component caused by indistinguishable electrons being subject to exchange symmetry and the repulsion is

due to the Pauli exclusion principle. Dispersion is a weak attractive force associated with induced dipole-induced dipole interactions that can stabilise and fine tune these sorts of non-covalent interactions.

Although electrostatics are usually the primary component in these non-covalent interactions; dispersion, polarisation, exchange, and charge-transfer play important roles in certain cases, and it is important to treat these components with their due consideration, along with other factors to realise energies that are true to their nature. For example certain EDA schemes, such as Natural Bond Orbital (NBO) analysis are known to largely overestimate charge-transfer in its calculation, which can lead to false conclusions⁸⁴. Likewise the often used Møller-Plesset Perturbation Theory (MP2) *ab initio* method, when applied to non-covalently interacting systems, can considerably overestimate the dispersion, which can lead to overestimation of noncovalent interaction energies⁸⁵.

Hence, it is important to stress that there is ambiguity in energy decomposition and there is no "correct" way of separating these components of the interaction, and that much like familiar concepts such as orbitals, hybridisation, and resonance; these are just useful models to help rationalise what is observed.

1.3.2 Nature of the Acceptor

σ -hole interactions involve both a σ -hole donor and an acceptor, with the nature of these dictating the character of the interaction. The three families of acceptors are n lone-pair-possessing atoms, π systems, and anions.

Lone pair based acceptors of σ -hole interactions have the ability to interact with as many σ -holes as the acceptor has available lone pairs. The geometric preference of these interactions generally aligns with the location of the lone pair, which changes based on the hybridisation of the atom. For example, sp -hybridised nitrogen has a lone pair located along the extension of its triple bond, whereas carbonyl oxygens have two lone pairs each at 120° to carbonyl bond.

For π -systems as acceptors the σ -hole generally approaches perpendicularly to the bonds of the π -system, but the preferred interaction site depends on the size and structure of the π -system. For aromatic systems such as benzene, the σ -hole adduct forms an arrangement where the donor approaches perpendicular to the bonding plane but closer to the carbons than to the centroid of the π -system⁸⁶. If both lone pairs and π -systems both feature on a potential acceptor, the σ -hole generally interacts preferentially with the lone pair¹⁵.

Anions pose an issue in that their nature more dramatically effects the charge distribution of the σ -hole possessing atom, inducing a high degree of polarisation that causes the standard σ -hole model to become less watertight. Huber and co-workers have reported multiple studies on so-called "anti-electrostatic" halogen bonds in crystals with halogenic anion-anion interactions that show the importance of polarisation and charge-transfer in halogen bonding⁸⁷. However, some could argue that these "anti-electrostatic" halogen bonds are not a halogen bond but instead a non-covalent interaction that so happens to involve halogens.

1.4 Searching the Cambridge Structural Database

In order to find crystal structures deposited in the CSD that contain possible σ -hole interactions, one first must define a query that contains the structural motif of the desired interaction. A query begins with some substituent constraints, for example constraining the search to just find moieties that contain a terminally-bound halogen. However, for a σ -hole interaction there needs to be a suitable number of covalently bonded substituents that are sufficiently electron-withdrawing and satisfy the valency of the atom.

Next, close contacts can be defined, which is a signifier of a noncovalent interaction, usually an interatomic distance that is shorter than the sum of the interacting atom's van der Waals (vdW) radii. However, not all close contacts involving groups 14-18 are σ -hole interactions and thus additional constraints are required, but not so strict as to miss potential interactions. Since σ -hole interactions show a tendency towards linearity along the extension of a covalent bond, an angular constraint can be imposed on the R-D...A angle that can be tailored to the specific σ -hole interaction type in question. Geometrical measurements and observables such as atomic numbers and atomic valencies of the query structures can also be defined in much the same way, which can then be tabulated and used to analyse the results so as to learn more about the interacting moieties.

Once a query has been described, further restrictions on the structures within the CSD that are searched can then be put into place. These include whether the full 3D atomic coordinates have been determined for the crystal structure, the option to exclude disordered structures or structures containing errors, the option to exclude polymeric structures, the option to also include powder X-ray determined structures along with single crystal structures, and the option to limit the maximum crystallographic R factor, which determines the quality of the structure model.

Once the search has been run, the output is a number of "hits" that correspond to crystal structures that contain the structural motif provided in the query. These hits are identified by their CSD Refcode which is usually a 6 letter name providing a unique identifier for the crystal structure. The crystal structures associated with these hits can be visualised and other information gathered such as associated literature, crystallographic information file (cif), and any geometric measurements that have been defined in the query. These measurements can be used to analyse the hits to determine trends in the data so as to further constrain the searches with the eventual aim of obtaining structures that primarily show σ -hole interactions. The specific details of the CSD searches constructed as part of this thesis are included in Section 2.1, Section 3.2 and Section 4.2.

Analogous searches of the CSD have been made for hydrogen bonding, such as those by Brammer and co-workers, in which halogens were investigated for their ability to act as hydrogen bond acceptors⁸⁸. In this survey the hydrogen bond donors were represented as C-H, O-H, and N-H, and the acceptors were X-C, X-M, and X⁻, where X is a halogen and M is a transition metal. As hydrogen bonds tend towards linearity, an angular constraint was placed on the D-H...X angle of 110° and the H...X contacts were normalised as a fraction of the sum of van der Waals radii. This study led to the conclusion that due to the correlation in hydrogen bond distance and D-H...X angle that all halogens are capable of accepting hydrogen bonds.

Searches for halogen, chalcogen, and pnictogen bonds in the CSD have been undertaken by Frontera and co-workers, focusing on interactions between nitrogen atom or arene ring-based acceptors and monovalent halogens, divalent chalcogens and trivalent pnictogens as donors⁶³. This is a promising starting point that is supported by theoretical calculations, but could be broadened in terms of σ -hole interaction variety and more in-depth investigation into functional groups.

1.5 Predictive Models for Noncovalent Interaction Energies

Given the weakness of σ -hole interactions, and that multiple energetic components are involved in their formation, the only way to reliably compute their interaction energies is to utilise very high accuracy electronic structure methods. However, these are only viable for use in small molecular systems as the computational cost becomes prohibitive for larger systems. As such, there is considerable interest in the development of sound, simple models that can accurately predict the strength of σ -hole interactions. From these quantitative models, it can be deduced whether molecules will behave differently to similar examples, which can lead to insight into the nature of these interactions.

Such models have been produced for hydrogen bonding, and for several σ -hole interaction classes. Legon and co-workers developed a scheme for the prediction of hydrogen bond interaction energies (see Equation (1.1)) based on experimental spectroscopic force constants k_σ , which are involved in the interaction energy and then parameterised to the molecules involved in the hydrogen bond⁸⁹.

$$k_\sigma = cNE \quad (1.1)$$

In this scheme N and E are dubbed the nucleophilicity and electrophilicity of the hydrogen bond donor/acceptor, respectively, even though there is no physical basis for these names, and c is a proportionality constant. This model shows good agreement with experiment, though the errors, expectedly, are larger new, unseen systems. Another approach by Legon and co-workers is a generalised model that includes hydrogen, halogen, chalcogen, pnictogen, and tetrel bonding⁹⁰. It is important to note that not all of the interactions are σ -hole based, with some such as the carbon dioxide molecule possessing a π -hole instead. This scheme has much the same form as the above for hydrogen bonding, but is based on high accuracy CCSD(T)/CBS calculations rather than experimental data, see Equation (1.2).

$$D_e = c'N_BE_A \quad (1.2)$$

Where D_e is the dissociation energy, N_A and E_A are the numerical nucleophilicity/electrophilicity of Lewis base B and Lewis acid A respectively, and c' is another proportionality constant.

A rigorous yet simple model for halogen bonding has been reported by Shaw and Hill where, instead of fitting by arbitrarily choosing unit parameters for a single halogen bond donor/acceptor like Legon, an unbiased fitting regime was employed, fitting purely statistical values over the whole dataset⁹¹. A distance dependence was also implemented in Equation (1.3) by the inclusion of the equilibrium separation between donor and acceptor $R_{e,ij}$. The different models produced in this

work are dubbed the Pn models, in which the $P0$ model would have no distance term, and the best performing model was found to be the $P4$ model.

$$E_{ij}^{Pn} = \frac{cX_iB_j}{R_{e,ij}^n} \quad (1.3)$$

In Equation (1.3) X_i and B_j are statistical parameters for the donor and acceptor, respectively, and c is a real constant that sets the energy scale. This model was fitted to high accuracy CCSD(T)-F12b calculations on a large molecular dataset, and shows much more robust fitting than the prior schemes.

A key issue with each of these predictive models is that they still require computational calculations, albeit at lower levels of theory, as a starting point. It is hoped that in this thesis, machine learning can be leveraged to produce a predictive model for σ -hole interaction energies that, after training, will bypass the need for electronic structure calculations.

For large supramolecular complexes such as those seen in the CSD, a predictive model to estimate the stabilising influence of σ -hole interaction energies would be highly valuable. This could then be implemented into the CSD to accurately predict the stabilising effect of a σ -hole interaction motif within newly deposited structures into the database. This would be of interest to those studying crystal structures in the CSD as this possible influence could have been missed or unquantified.

1.6 Aims and Objectives

1. Identify relevant donor and acceptor moieties associated with σ -hole interactions. Define geometric and other constraints to enable identification of σ -hole interactions in crystal structures present in the CSD.
2. Use the CSD to determine geometric characteristics and trends for σ -hole interactions as a function of donor and acceptor characteristics to provide insight into the nature of the interactions.
3. Fit a predictive model for σ -hole interaction energies based on quantum-chemical calculations to enable rapid estimation of these energies with good accuracy. Apply this energy estimation to σ -hole interaction motifs found in the CSD.

Chapter 2

Theory & Methodology

2.1 CSD Search Methodology

When searching the Cambridge Structural Database for σ -hole interactions, one must first identify the indicators of such an interaction within a crystal structure, and then define a set of criteria to ensure that only the desired interactions are retrieved.

In order to identify σ -hole interactions, substituent and geometric constraints are imposed. Firstly, the σ -hole donor must be defined, this being the atom possessing the σ -hole, along with the suitably electron-withdrawing substituents needed to cause the charge anisotropy. Likewise, the acceptor must be defined, this being the lone-pair possessing Lewis base, along with varieties of substituents and valencies. This enables the subsequent classification and subdivision of the interactions by the acceptor functional group, from which trends can be discerned.

A common indicator of a noncovalent interaction within a crystal structure is a so-called close contact, two atoms that are relatively close together but not so close as to be considered covalently bonded. A common cut-off for defining a close contact is the sum of the van der Waals radii of the two atoms involved in the contact⁹². It is important to note that noncovalent interactions can still act at longer ranges, so it is also common to add tolerance to this cutoff so as not to exclude structures^{93 94}. To compare these distances between the various possible donor and acceptor atoms, this contact distance $d_{contact}$ is normalised, resulting in d_{norm} , by the sum their van der Waals radii $r_{vdW}^{D/A}$, as shown in Equation (2.1)⁹⁵.

$$d_{norm} = \frac{d_{contact}}{r_{vdW}^D + r_{vdW}^A} \quad (2.1)$$

Given what is understood about σ -hole interactions and their directionality with respect to the location of the σ -hole, it is also appropriate to apply angular constraints to the searches. The angle that defines this directionality, referred to herein as the σ -hole angle and shown in Figure 2.1, is the angle between the σ -hole interaction and the electron-withdrawing group that causes the charge anisotropy. In the interest of studying not just the directionality with respect to the σ -hole, but

the directionality with respect to the lone-pair located on the acceptor atom, a second angle must be defined. This angle, hereby referred to as the lone-pair angle, is related to the orientation of the Lewis basic lone-pair on the acceptor. This angle varies depending on the Lewis base, and is defined in the acceptor-specific sections in Section 3.2 and Section 4.2. This angle is not intended to constrain the search criteria, merely to explore the directionality at the acceptor side of the interaction.

The CSD suite of software has long included a package to search the database, ConQuest, which is incredibly useful for its graphical user interface and query constructor. However, for post-processing needs, and integration with other software packages, the CSD Python API is utilised with the query files made in ConQuest^{96 97}.

2.1.1 Query & Search Construction

The queries were constructed in ConQuest as *.qry* files, and then exported as *.con* files, a text based format that can be utilised in the *SubstructureSearch* program of the CSD Python API. The subgroups of these queries are two-fold, that of the donor and that of the acceptor. The donors studied are split into monovalent halogens, divalent chalcogens, trivalent pnictogen and tetravalent tetrels. These are each covalently bonded to a number, respective of their valency, of suitably electron-withdrawing elements, these being; B, C, N, O, F, Si, P, S, Cl, Se, Br, Te, and I. A close contact is defined between the donor and acceptor atoms, with the contact distance needing to be smaller than the sum of vdW radii with an added 0.2 Å to the maximum possible contact distance.

On the donor side of the complex, all angles between the electron-withdrawing groups and the contact are measured. In the case of halogen bonding, this is constrained between 110 and 180° within the query as there is only one electron-withdrawing group to define this angle. For the higher valency chalcogens, pnictogens and tetrels, all angles between potential electron-withdrawing groups and the contact are recorded, and then the largest angle is selected as the σ -hole angle, provided it is greater than 110°, and that substituent recorded as the electron-withdrawing group. All atomic numbers are recorded for later analysis, as are the number of connected atoms to the potential electron-withdrawing groups.

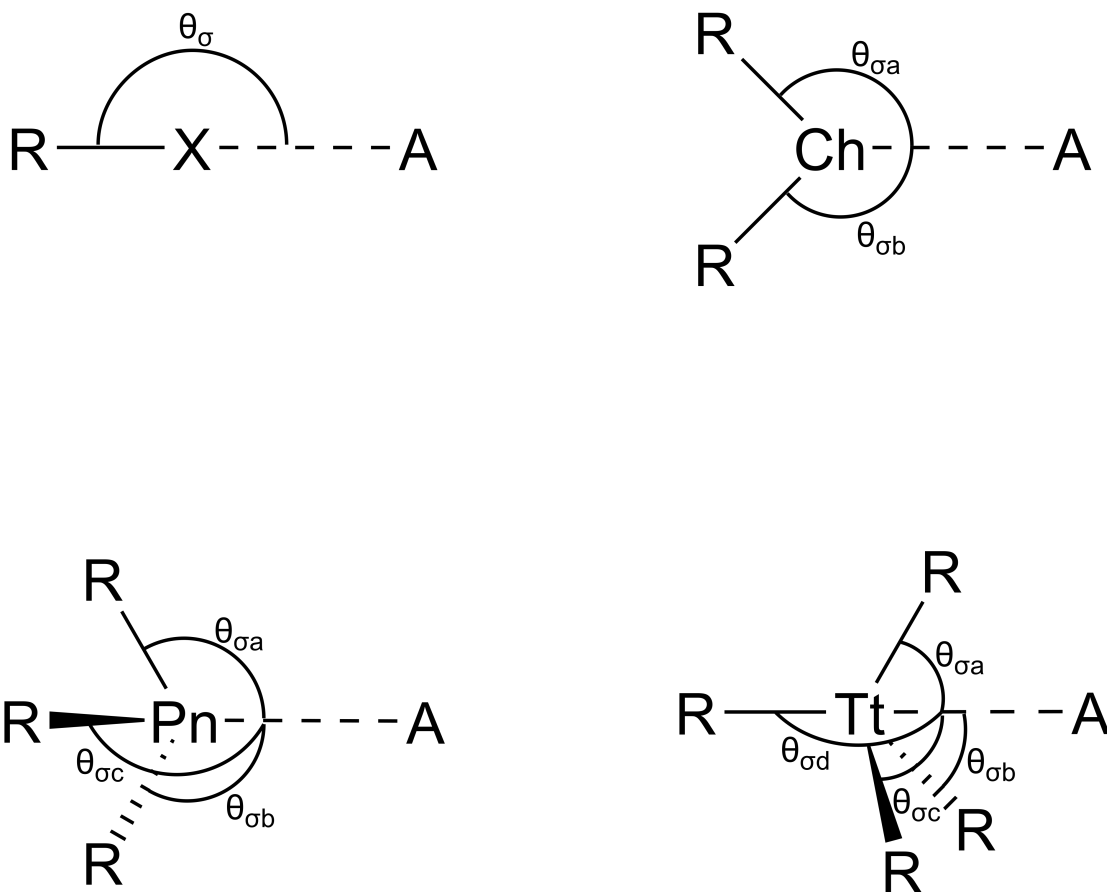


Figure 2.1: A representation of the donor components for the halogen (X), chalcogen (Ch), pnictogen (Pn), and tetrel (Tt) bonding queries. $\theta_{\sigma x}$ refers to the potential σ -hole angle(s). $R = B, C, N, O, F, Si, P, S, Cl, Se, Br, Te,$ and I .

The acceptors are split into Lewis bases with one or more lone pairs located on nitrogen or oxygen. The nitrogen Lewis bases are further subdivided by hybridisation; N_{sp} , N_{sp^2} , N_{sp^3} . Oxygen Lewis bases are subdivided by functional group, these being organic carbonyls, divalent oxygen, oxides, terminal metal carbonyls, di-bridging metal carbonyls, and tri-bridging metal carbonyls, with further detail included in Chapter 4. The atomic numbers of the varied substituents making up the acceptor moiety are recorded for later analysis, along with the number of atoms connected to these substituents. A lone-pair angle is also defined, which varies in definition for different acceptor moieties, the specifics of which can be found in Section 3.2 & Section 4.2.

To investigate the likelihood of a given σ -hole interaction motif forming a contact, the methodology of Allen *et al.* is used⁹⁸. This involves the comparison of the number of unique crystal structures that contain the required donor and acceptor moieties, and the number of unique crystal structures that contain those moieties in a close contact. The queries that are constructed from Figure 2.1 and the various acceptor moieties (Figure 3.1 & Figure 4.1) are duplicated but with the close contact and any angular constraints removed, while maintaining any post-processing detailed in Chapter 3 and Chapter 4. To obtain the number of unique crystal structures, the CSD refcodes are stripped of any suffix numbers before the unique refcodes are counted, to eliminate duplicate structures under refcodes of the form XXXXXX01 that are commonly used. The comparison of the numbers of motif-containing (N_{motif}^{unique}) and contact containing structures

($N_{contact}^{unique}$) are given as a percentage, herein referred to as the prevalence.

$$Prevalence = \frac{N_{contact}^{unique}}{N_{motif}^{unique}} \times 100 \quad (2.2)$$

Searches were performed using the 2024.3 release of the CSD software suite, with database version 5.46 (November 2024) using a combination of ConQuest query files and version 3.3.0 of the CSD Python API^{96 97 11}. General restrictions of the search criteria were applied as follows: the crystallographic R-factor of the crystal structures must be less than 0.1, the entry contains no errors, 3D coordinates have been determined, there is no disorder and the structure is not polymeric. Any structures that have identical geometric measurements yet different CSD refcodes were also removed to prevent duplicate data. Structures retrieved from these searches have any unknown bond types assigned and any missing hydrogens added to allow for proper definition of functional groups and hybridisation. The bond types available between atoms in the CSD include single, double, triple, quadruple, aromatic, delocalised, and π .

2.1.2 On van der Waals Radii

The almost-ubiquitous approach to recognising noncovalent interactions within a crystal structure is to identify atoms that are separated by a distance smaller than the sum of the atom's van der Waals radii^{99–101}. These radii have no thorough physical or theoretical basis, yet see widespread use in crystallography and structural chemistry, and there have been multiple different attempts at determining their values. Looking back, the earliest recognised set of van der Waals radii came from Pauling in 1931, estimating the vdW radius of an element by using its negative ion radius¹⁰². For many years, the established dataset of van der Waals radii was that determined by Bondi for 38 elements in 1964¹⁰³. These determinations were impressive for the time, and some still hold up surprisingly well today, being determined via a combination of solid, liquid, and gaseous properties. It is also worth noting that Bondi himself wrote in said work "It cannot be overemphasized that the van der Waals radii of this paper have been selected for the calculation of volumes. They may not always be suitable for the calculation of contact distances in crystals", yet the latter has been their primary use¹⁰⁴. These are the van der Waals radii that have been utilised within the CSD for the majority of its existence, while attributing a radius of 2.00 Å to any elements missing from the dataset compiled by Bondi.

Roland and Taylor made the first use of the huge amount of structural information within the CSD to deduce van der Waals radii from statistical analysis of atomic contacts in order to provide more comprehensive coverage of the Periodic Table¹⁰⁵. These radii were found to be reasonably consistent with those determined by Bondi, apart from that for hydrogen ($r_{vdW}^{Bondi} = 1.2$ Å, $r_{vdW}^{R\&T} = 1.1$ Å), but did not extend to the whole Periodic Table. A more extensive undertaking is that of Batsanov in 2001, which covered 68 elements, with the inclusion of transition metals but still lacking the f-block elements¹⁰⁶. This approach involved the evaluation of X-ray diffraction data, physical properties and molar volumes. Computational approaches have also been utilised to determine van der Waals radii, such as that employed by Mantina *et al.*¹⁰⁷. This is a set of van der Waals radii that is compatible with that of Bondi, adding the remaining 16 main group elements. This approach involved high accuracy relativistic coupled cluster calculations using various probe molecules to

determine the repulsive wall distances.

The most complete is the work of Alvarez in 2013, which extends the methodology of Roland and Taylor¹⁰⁸. The resulting radii were determined from a large survey of distance distributions found within the CSD, where the van der Waals peak (second maximum of interatomic distances, with the first maximum corresponding to covalent bonds) is superimposed on a random distribution function with a cubic dependence on the interatomic distance. This set of van der Waals radii covers 99 elements, the majority of the Periodic Table, with different levels of uncertainty owing to the approximations made and sparsity of data in some cases. As of CSD version 2024.3, the CCDC has transitioned to using the van der Waals radii derived by Alvarez in their software suite, which in some cases drastically changes the distance ranges for contacts pertaining to σ -hole interactions, as shown in Table 2.1. This particularly affects pnictogen and tetrel bonding contacts where, for some of the heavier elements, the van der Waals radii have increased by over half an Ångström changing from Bondi to Alvarez values. This change will likely have many implications for the works pertaining to geometry based searches of the database. This change has also been reflected in the predictive work done in Chapter 5, leading to improved training and predictions.

Table 2.1: Comparison of van der Waals radii in Ångström determined by Bondi and Alvarez used in the CSD to determine close contacts relevant to σ -hole interactions^{103 108}. *Bondi vdW radii utilised wherever possible in CSD 2024.2 and older, with a value of 2.0 Å used if a radius was not tabulated in Bondi's work.

Element	Bondi*	Alvarez	Change
I	1.98	2.04	+0.06
Br	1.85	1.86	+0.01
Cl	1.75	1.82	+0.07
F	1.47	1.46	−0.01
Te	2.06	1.99	−0.07
Se	1.90	1.82	−0.08
S	1.80	1.89	+0.09
Bi	2.0	2.54	+0.54
Sb	2.0	2.47	+0.47
As	1.85	1.88	+0.03
P	1.80	1.90	+0.10
Pb	2.02	2.60	+0.58
Sn	2.17	2.42	+0.25
Ge	2.0	2.29	+0.29
Si	2.10	2.19	+0.09
N	1.55	1.66	+0.11
O	1.52	1.50	−0.02

2.2 Calculation of Non-covalent Interaction Energies

In contrast to the study of covalent bonds, non-covalent interactions are much weaker and governed by multiple factors that, as previously mentioned, prove difficult to separate with any validity. There are also other potential sources of error that need to be accounted for when calculating these interaction energies so as to give a meaningful result.

This short review on the theory of quantum theory was compiled from the reference material found in the work of

Szabo & Ostlund, Jensen, and Cramer^{109–111}. Electronic structure calculations revolve around the concept of a system's wavefunction Ψ , in which all possible information about the system is contained. The application of operators to the wavefunction of the system in question will allow calculation of specific properties, most commonly being the energy of the system E . This energy is commonly obtained through the non-relativistic time independent Schrödinger equation, shown in Equation (2.3). This is an eigenproblem, where the Hamiltonian operator $\hat{H}$ acts on an eigenfunction Ψ to return this same eigenfunction multiplied by an eigenvalue E . The Hamiltonian operator is made up of the sum of kinetic $\hat{T}$ and potential energy operators $\hat{V}$.

$$\begin{aligned}\hat{H}\Psi &= E\Psi \\ \hat{H} &= \hat{T} + \hat{V}\end{aligned}\tag{2.3}$$

For a system of i & j electrons and A & B nuclei, the Hamiltonian operator can be expressed in terms of the kinetic and potential energy operators, shown in Equation (2.4). In Equation (2.5) the kinetic energy operators $\hat{T}$ and potential energy operators $\hat{V}$ are represented, where m_A is the mass of nucleus A, ∇^2 is the Laplacian operator (see Equation (2.6)), Z is an atomic number, and r_{xy} is the distance between two particles, with atomic units used throughout.

$$\hat{H} = \hat{T}_N + \hat{T}_e + \hat{V}_{eN} + \hat{V}_{NN} + \hat{V}_{ee}\tag{2.4}$$

$$\begin{aligned}\hat{T}_N + \hat{T}_e &= -\sum_A \frac{1}{2m_A} \hat{\nabla}_A^2 - \sum_i \frac{1}{2} \hat{\nabla}_i^2 \\ \hat{V}_{eN} + \hat{V}_{NN} + \hat{V}_{ee} &= -\sum_i \sum_A \frac{Z_A}{r_{iA}} + \sum_A \sum_{A < B} \frac{Z_A Z_B}{r_{AB}} + \sum_i \sum_{i < j} \frac{1}{r_{ij}}\end{aligned}\tag{2.5}$$

$$\hat{\nabla}_i^2 = \left(\frac{\partial^2}{\partial x_i^2} + \frac{\partial^2}{\partial y_i^2} + \frac{\partial^2}{\partial z_i^2} \right)\tag{2.6}$$

It is not possible to solve the non-relativistic time-independent Schrödinger equation for systems larger than the hydrogen atom analytically. This is due to the many-body problem introduced by $\hat{V}_{ee}$ and $\hat{V}_{NN}$, which means the multiple electrons and the nuclei cannot be separated from each other. Atomic, and by extension molecular, systems are comprised of lighter electrons and heavier nuclei, allowing them to be treated differently. Nuclei are sufficiently massive to only display small quantum effects, whereas electrons are very light and cannot be described by classical mechanics. A means to circumvent the $\hat{V}_{NN}$ term is the Born-Oppenheimer approximation, which involves the decoupling of the electronic and the nuclear contributions to the wavefunction Ψ . This approximation is based on the concept that the nuclei are "infinitely" more massive than the electrons, and thus electrons will move "infinitely" faster than the nuclei. This allows the assumption of fixed or "clamped" nuclei, which allows electronic movement to be treated separately to the nuclei. Thus the wavefunction Ψ can be described by an electronic Ψ_{el} and a nuclear wavefunction Ψ_{nu} .

$$\Psi = \Psi_{el} \Psi_{nu}\tag{2.7}$$

The uncoupling the nuclear and electronic components of the wavefunction gives rise to the electronic Schrödinger equation and electronic Hamiltonian $\hat{H}_{el}$ resulting in an electronic energy E_{el} . The total energy includes a constant nuclear repulsion from the fixed nuclei, there is only a parametric dependence on nuclear coordinates.

$$\begin{aligned}
\hat{H}_{el}\Psi_{el} &= E_{el}\Psi_{el} \\
\hat{H}_{el} &= -\sum_i \frac{1}{2}\nabla_i^2 - \sum_i \sum_A \frac{Z_A}{r_{iA}} + \sum_i \sum_{i<j} \frac{1}{r_{ij}} + \sum_A \sum_{A<B} \frac{Z_A Z_B}{r_{AB}} \\
\hat{H}_{nu} &= -\sum \frac{1}{2}\hat{\nabla}_A^2 + E_{el} \\
\hat{H}_{nu}\Psi_{nu} &= E_{total}\Psi_{nu}
\end{aligned} \tag{2.8}$$

The solution to the Schrödinger equation is found variationally using a trial wavefunction Φ , meaning that for any trial wavefunction the calculated energy will always be greater than the exact ground-state energy of the system E_0 . In Equation (2.9) Φ^* is the complex conjugate of the wavefunction, and τ is the position vector in three-dimensional space.

$$\tilde{E} = \frac{\int \Phi^* \hat{H} \Phi d\tau}{\int \Phi^* \Phi d\tau} \geq E_0 \tag{2.9}$$

Equation (2.9) is applied iteratively, varying the trial wavefunction Φ each time until the expectation value $\tilde{E}$ no longer lowers. This means that any lowering of calculated energy $\tilde{E}$ results in a trial wavefunction is a better approximation to the true wavefunction.

To construct a multi-electron wavefunction, it is assumed that it can be described in terms of independent single-electron wavefunctions, dubbed spinorbitals. Electrons can exist as spin-up (α), where the spin magnetic quantum number $m_s = +\frac{1}{2}$, and spin-down (β) where $m_s = -\frac{1}{2}$. Thus to make a spinorbital one combines the spatial and spin wavefunctions of an electron.

$$\begin{aligned}
\phi(i) &= \phi(\tau_i)\alpha(s_i) \\
\bar{\phi}(i) &= \phi(\tau_i)\beta(s_i)
\end{aligned} \tag{2.10}$$

The issue with expressing a multi-electron wavefunction as a product of spinorbitals is the requirement of a wavefunction to be antisymmetric (change sign) upon the interchange of any two electrons, known as the Pauli antisymmetry principle. To keep this requirement of the wavefunction, one combines spinorbitals into a determinant, which inherently contains antisymmetry, called a Slater determinant $\tilde{\Phi}$, with a two-electron example shown in Equation (2.11).

$$\tilde{\Phi}(1,2) = \frac{1}{\sqrt{2}} \begin{vmatrix} \phi_1(1) & \phi_2(1) \\ \phi_2(2) & \phi_1(2) \end{vmatrix} \tag{2.11}$$

In order to represent the spinorbitals ϕ_i needed to construct a Slater determinant for a molecular system, they are approximated via a linear combination of atomic orbitals χ_p . In Equation (2.12) where c_{pi} is a coefficient showing the contribution of the p -th atomic orbital out of m total atomic orbitals to the overall molecular orbital. These atomic orbitals

χ_p are referred to as basis functions, with the full set of χ being known as the basis set.

$$\phi_i = \sum_p^m c_{pi} \chi_p \quad (2.12)$$

The two most common methods of presenting basis functions are Slater type and Gaussian type orbitals. Slater-type orbitals are generally more accurate but are challenging to compute, whereas Gaussian-type orbitals can be easily integrated over larger systems. The general form of Slater-type orbitals is shown in Equation (2.13), and are closer in character to actual atomic orbitals, yet lead to complex integrals resulting in high computational cost. In contrast, Gaussian-type orbitals shown in Equation (2.14) are easily integrated and thus much less computationally expensive, yet have a poorer fit to actual atomic orbitals. In these orbital equations, N is a normalisation constant, a, b, c correspond to n, l, m principal, angular, and magnetic quantum numbers that dictate the size, shape, and orientation of the orbital, and the exponent or ζ controls the width of the orbital, with larger ζ giving a tighter orbital and a smaller ζ resulting in a wider, more diffuse orbital.

$$\chi_{abc}^{Slater} = N x^a y^b z^c e^{-\zeta r} \quad (2.13)$$

$$\chi_{abc}^{Gaussian} = N x^a y^b z^c e^{-\zeta r^2} \quad (2.14)$$

Basis sets are often specified in terms of their 'zeta' value, which corresponds to the number of exponents per atomic orbital, so a 'triple-zeta' basis set would have three exponents per atomic orbital. Additional functions can be added to the the basis set, such as polarisation functions, which are functions from the next-highest orbital angular momentum for each atom. These polarisation functions describe the polarisation of the electron density of the atom when in a molecule, and thus are crucial for describing chemical bonding. Diffuse functions can also be added to a basis set, which are extra functions with small exponents that better describe the orbitals at large separations from the nuclei, which is very important for calculating noncovalent interactions.

2.2.1 Density Functional Theory

Although methods exist to calculate the wavefunction of a molecular system directly, accurate wavefunction methods of electronic structure determination can prove to be computationally expensive, especially with molecules containing a large number of electrons. An alternative to this is found in methods that are derived as a function of the electron density, which can be summarised under the umbrella of density functional theory (DFT).

The basis of DFT stems from the Hohenberg-Kohn Theorems, the first of which is that the ground state electronic energy E_0 is completely determined by the electron density ρ . Thus if electron density is a function of the positions of electrons, then the energy is a functional of the electron density, hence the name density functional theory. The second Hohenberg-Kohn theorem is that this density based approach can be applied in a variational manner, shown in Equation (2.15), that

is analogous to that shown in Equation (2.9).

$$E[\rho] = \frac{\int \Phi^* \hat{H} \Phi d\tau}{\int \Phi^* \Phi d\tau} \geq E_0 \quad (2.15)$$

Kohn and Sham proposed the theory of treating the DFT Hamiltonian operator as non-interacting electrons, which allows the separation of the energy functional $E[\rho]$ in Equation (2.16). In which T_{ni} is the kinetic energy of non-interacting electrons, V_{ne} is the nucleus-electron interaction, V_{ee} is the classical electron-electron interaction, ΔT is the kinetic energy correction for an interacting system, and ΔV_{ee} is the non-classical correction for the electron-electron repulsion.

$$E[\rho] = T_{ni}[\rho] + V_{ne}[\rho] + V_{ee}[\rho] + \Delta T[\rho] + \Delta V_{ee}[\rho] \quad (2.16)$$

The two corrections, ΔT and ΔV_{ee} , are combined into a new functional, dubbed the exchange-correlation functional E_{xc} , which describes the quantum nature of the system. The formula for the exact exchange-correlation functional is unknown, but various theoretical and empirical approaches have been developed to approximate it. This approximation means that although the variational approach is applied in Equation (2.15), it is not truly variational as energies lower than E_0 can be calculated.

Exchange-Correlation Functionals

The most basic approach to approximating E_{xc} is the local-density approximation (LDA), which is that $E_{xc}[\rho]$ is that of the density of a uniform electron gas near the nucleus. The exchange functional is shown in Equation (2.17), where C_x is the exchange energy of a uniform electron gas. There exist several expressions for the correlation functional, which are obtained from fitting to quantum Monte Carlo calculations of a uniform electron gas, for example the VWN approach from Vosko, Wilk and Nusair¹¹².

$$E_x^{LDA} = -C_x \int \rho^{\frac{4}{3}}(\mathbf{r}) d\mathbf{r} \quad (2.17)$$

$$C_x = \frac{3}{4} \left(\frac{3}{\pi} \right)^{\frac{1}{3}} = 0.7386$$

There are many issues with using the LDA method, such as underestimating exchange energies for molecular systems and largely overestimating correlation energies, resulting in errors in bond strength up to 100 kJ/mol. Attempts to correct E_x^{LDA} require the electron density to be treated as non-uniform, such as through the use of the gradient of the electron density. These approaches, dubbed generalised gradient approximation (GGA), exist in various forms, with an example being the B88 functional shown in Equation (2.18) where β is a parameter fitted to six noble gas atoms He through Rn¹¹³.

$$E_x^{B88} = E_x^{LDA} - \beta \rho^{\frac{1}{3}} \frac{x^2}{1 + 6\beta x \sinh^{-1} x} \quad (2.18)$$

$$x = \frac{|\nabla \rho|}{\rho^{\frac{4}{3}}}$$

More advanced exchange-correlation functionals implement some wavefunction-based terms to the exchange. These

"hybrid" functionals introduce some Hartree-Fock (HF) exchange, along with LDA and/or GGA components. A typical example of this is the B3LYP functional which is shown in Equation (2.19), where a , b , and c are parameters fitted to experimental data. This utilises LDA and B88 exchange functionals, along with the Lee-Yang-Parr correlation functional E_c^{LYP} ^{114–116}. Some so-called double-hybrid functionals attempt to capture more accurate correlation through the use of post-HF perturbation terms, yet this incurs greater computational cost and poorer scaling.

$$E_{xc}^{B3LYP} = (1 - a)E_x^{LDA} + aE_x^{HF} + b\Delta E_x^{B88} + (1 - c)E_c^{LDA} + cE_c^{LYP} \quad (2.19)$$

Modern adaptations of GGA exchange-correlation functionals, known as meta-GGA functionals, introduce additional variables based on the orbital kinetic energy density τ or its spin-orbital equivalent τ_σ , detailed in Equation (2.20).

$$\begin{aligned} \tau(\mathbf{r}) &= \frac{1}{2} \sum_i^{occ} |\nabla \phi_i(\mathbf{r})|^2 \\ \tau_\sigma(\mathbf{r}) &= \frac{1}{2} \sum_i^{occ} |\nabla \phi_{i\sigma}(\mathbf{r})|^2 \end{aligned} \quad (2.20)$$

For example, components of the Minnesota M06 variant meta-GGA functionals involve the above spin kinetic energy density τ_σ , along with the spin density ρ_σ and the reduced spin density gradient x_σ of the form shown in Equation (2.21)^{117–118}. This is only a small review into the breadth of exchange-correlation functionals, with more detailed insight beyond the scope of this work. Interested readers are referred to a thorough review of exchange-correlation functionals by Mardirossian and Head-Gordon¹¹⁹.

$$x_\sigma = \frac{|\nabla \rho_\sigma|}{\rho_\sigma^{\frac{4}{3}}} \quad \sigma = \alpha, \beta \quad (2.21)$$

Dispersion Correction

A development in DFT that is particularly relevant to the calculation of noncovalent interactions is the recognition that many density functionals do not properly capture dispersion interactions, especially at long range. One approach to remedy this is through a simple empirical correction known as dispersion corrected DFT, or DFT-D^{120–121}. These corrections were developed by the Grimme group, with the most commonly used being the DFT-D3 method that involves a two-body correction detailed in Equation (2.22), along with a Axilrod-Teller-Muto (ATM) triple-dipole three-body term. Thus the total corrected energy $E_{DFT-D3} = E_{DFT} + E_{D3-2body} + E_{ATM-3body}$. The DFT-D3 correction contains a damping function f_{damp} for the short range, the R_{AB}^{-6} term would cause the dispersion energy to tend towards infinity, whereas in reality the Pauli repulsion would keep the atoms separated. This damping function smoothly scales down the dispersion energy at short ranges, ensuring this non-physical divergence of energy does not occur. There are two commonly used expressions for f_{damp} , the first being the so-called zero-damping function shown in Equation (2.23), and the Becke-Johnson damping version shown in Equation (2.24)¹²². In these equations C_n^{AB} corresponds to n th-order dispersion coefficients for atom pairs AB, R_{AB} are internuclear distances, $S_{r,n}$ parameters are scaling factors, and α and β are fitted parameters.

$$E_{D3-2body} = - \sum_A^{atoms} \sum_{B < A}^{atoms} [s_6(\frac{C_{6,AB}}{R_{AB}^6})f_{damp,6}(R_{AB}) + s_8(\frac{C_{8,AB}}{R_{AB}^8})f_{damp,8}(R_{AB})] \quad (2.22)$$

$$f_{damp,n}^{D3(0)}(R_{AB}) = [1 + 6(\frac{R_{AB}}{S_{r,n}R_{0,AB}})^{-\beta_n}]^{-1} \quad (2.23)$$

$$f_{damp,n}^{D3(BJ)}(R_{AB}) = \frac{R_{AB}^n}{R_{AB}^n + (\alpha_1 R_{0,AB} + \alpha_2)^n} \quad (2.24)$$

2.2.2 Basis Set Superposition Error

The standard approach for the calculation of the strength of non-covalent interactions is given in Equation (2.25), where A and B correspond to the two interacting "monomers" and AB corresponds to the "dimer". The subscripts correspond to the energy being calculated and the superscripts are the basis that the calculation uses.

$$E_{int} = E_{AB}^{AB} - E_A^A - E_B^B \quad (2.25)$$

However, when using this calculation, the basis of the two separate monomers is not equivalent to the basis of the dimer. This means that when the dimer energy is calculated from the basis of the two monomers, it 'borrows' basis functions from each of the monomers thus improving its basis set, giving a non-physical lowering of the energy. This erroneous lowering of the energy is known as basis set superposition error (BSSE) and should always be accounted for when calculating non-covalent interaction energies. A common scheme for removing this is to implement a so called counterpoise correction, developed by Boys and Bernardi¹²³.

$$E_{int}^{CP} = E_{AB}^{AB} - E_A^{AB} - E_B^{AB} \quad (2.26)$$

This involves the use of "ghost" atoms of one monomer when calculating the energy of the other. These ghost atoms have the basis functions of the corresponding atoms but possess no electrons, allowing this "borrowing" of basis functions from the ghost atoms so as to effectively calculate the energy of each monomer in the dimer basis. This method is quite widely accepted among the computational chemistry community, but there are debates over its use and new methods have been developed to avoid BSSE arising in the first place^{124 125}.

Chapter 3

Nitrogen as an Acceptor

3.1 Introduction to Nitrogen Lewis Bases

Nitrogen Lewis bases have a single lone-pair located on a nitrogen atom which can span a range of hybridisations, and can act as an acceptor for Lewis acidic donors, forming noncovalent interactions. These Lewis bases are widely used within synthesis, including use in forming hydrogen bonds, halogen bonds, and other noncovalent interactions^{126–129}. This makes nitrogen Lewis bases an ideal starting point in exploring σ -hole interactions within the CSD

To hypothesise on the effect that hybridisation of the nitrogen has on its ability to act as an acceptor for σ -hole interactions, there are some factors one must consider. As the lone-pair containing orbital on the nitrogen progresses from sp to sp^2 to sp^3 , the p character of the orbital increases, meaning the lone-pair becomes more diffuse. This hybridisation change is associated with differing numbers of substituents, resulting in changes in steric bulk, which can perturb the geometry of the σ -hole interactions. Another factor to consider is that in sp - and sp^2 -hybridised nitrogen moieties, a π -system will also be present in the molecule, which can also act as an acceptor for σ -hole interactions. It is known that lone pairs are more Lewis basic than π -systems, and are the preferred interaction site for σ -hole interactions, but it is an important factor to consider when analysing these contacts as they may also perturb the geometries of the σ -hole interactions¹⁵.

Looking to the literature, a computational study on triel bonding interactions by Yang *et al.* uses HCN, CH₂NH, and NH₃ as models to compare the calculated interaction energies with a triel bond donor species¹³⁰. This is an interesting comparison to this work on σ -hole interactions, as these triel bonding systems have a similar nature being π -hole interactions. This study suggests that the trend in interaction energies follows the trend in strength of $N_{sp} < N_{sp^3} < N_{sp^2}$, but this is likely due to the N_{sp^2} systems containing secondary interactions between a hydrogen on the CH₂NH molecule and the substituents on the triel atom, which would enhance the interaction energy. Another computational study, this time involving hydrogen bonding to N $\equiv$ C-R, NH=CHR, and NH₂CH₂R shows an increase in interaction strength going from sp to sp^2 to sp^3 hybridised nitrogen¹³¹. This trend is also seen in calculated interaction energies of tetrel bonding complexes, done by Liu *et al.* with various nitrogen Lewis bases¹³². Experimental work carried out by Kjaergaard and co-workers on gas-phase

spectroscopic measurements of hydrogen bonding complexes with acetonitrile, pyridine, and trimethylamine shows that the hydrogen bond acceptor strength of the nitrogen decreases with increased *s*-character of the lone-pair¹³³.

In this chapter geometric data from searches of the CSD pertaining to halogen, chalcogen, pnictogen, and tetrel bonding contacts to *sp*³-, *sp*²- and *sp*-hybridised nitrogen moieties will be compared. This geometric data includes contact distances normalised by the van der Waals radii of the contacting atoms, the σ -hole angle, and the lone-pair angle. Details of these measurements are found in Section 2.1.1 and Section 3.2 and are analysed via scatter plots of either the σ -hole or the lone-pair angle against the normalised contact distance. This allows the visualisation and tabulation of the trends with respect to the interaction types, σ -hole donors, and acceptors to be interpreted from these geometric measurements.

Before proceeding it is important to discuss the concept of selection bias among both crystal structures containing contacts indicative of a σ -hole interaction, and crystals which contain the required moieties. Some classes of σ -hole interactions involve elements that are utilised at different frequencies in synthesis, thus meaning there are wildly differing numbers of crystals in the CSD that, for example, contain chlorine and those that contain tellurium. This can mean that the crystal structures, and the contacts found within, are sometimes not representative of the full scope of the class of σ -hole interaction. This is especially seen in the less frequently occurring σ -hole interactions, where the research into the field is less developed, so any possible trends seen in this data must be interpreted with selection bias in mind.

3.2 CSD Query Details

The donor moieties detailed in Figure 2.1 from Section 2.1.1 were combined with the various acceptor moieties shown in Figure 3.1 to form the query files used to complete these searches, with contact distance $d_{contact}$ defined between the D_{σ} donor atom (*X/Ch/Pn/Tt*) and the nitrogen atom. The 12 queries resulting from this are used to search for halogen, chalcogen, pnictogen, and tetrel bonding contacts to *sp*³-, *sp*²- and *sp*-hybridised nitrogen moieties. The bond typing in the *sp*-hybridised nitrogen moiety can have N–C bonds that are doubly or triply bonded, and the C–R bond can be of any type. In the *sp*²-hybridised moiety the N–R bond types can be single, double, aromatic, or delocalised, and in the *sp*³-hybridised moiety all bond types are single.

In order to investigate the directionality of σ -hole interactions with respect to the lone-pair, a so-called lone-pair angle is defined relative to the location of the lone-pair based on orbital hybridisation theory. Detailed in Figure 3.1, the lone pair angles of these acceptors differs based on the number of substituents on the nitrogen atom. The *sp*-hybridised acceptor is the simplest with the lone-pair angle being the D_{σ} –N–C angle, with the *sp*²- and *sp*³-hybridised acceptors the lone pair angle is the D_{σ} –N–centroid angle, with the centroid being between the two or three substituents, respectively, on the nitrogen. These lone-pair angles are constrained between 110–180° within the queries, and mean that a contact aligning with the lone-pair of the nitrogen atom should approach with a lone-pair angle of 180°. Post-processing of the search results is required to ensure that the σ -hole acceptors contain nitrogen with the correct valency, as the number of nitrogen substituents alone does not determine hybridization.

For *sp*²-hybridised nitrogens the R–N–R angle is measured and constrained to be less than 165° to remove any potential

sp -hybridised nitrogen moieties. For sp^3 - hybridised acceptors the three NM–N–NM angles are summed and the structure excluded from the results if the sum is greater than 345° , to ensure that the nitrogen is not planar and thus part of an sp^2 -hybridised system. To investigate the different ring sizes of sp^2 -hybridised nitrogen, altered sp^2 -hybridised nitrogen queries were constructed of ring sizes from 3-12. These queries are monocyclic and contain only the same substituent elements as in the default sp^2 -hybridised nitrogen queries. The full details of the construction of these queries is shown in Appendix A.1.

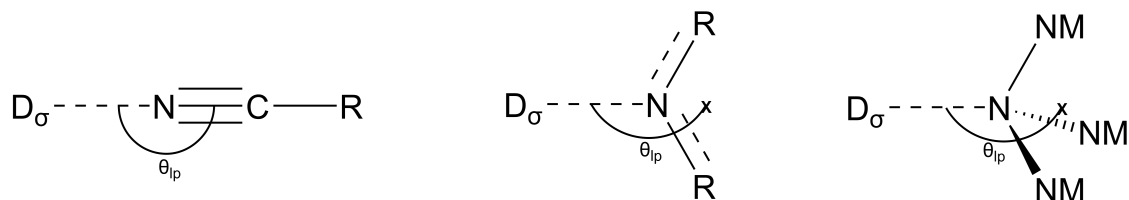


Figure 3.1: A representation of the various lone-pair angle measurements for the sp -, sp^2 - and sp^3 -hybridised nitrogen components of the CSD search queries. R = H, C, N, O, Si, P, S, F, Cl, Br, I, Se, Te, B. NM = Non-metal

3.3 sp^3 Hybridised Nitrogen

Starting with the trivalent sp^3 -hybridised nitrogen σ -hole acceptors, one might expect that sp^3 -hybridised nitrogen σ -hole acceptors have a notable effect on the geometric properties of the contact, due to the steric influence on these contacts. In particular, higher valency σ -hole donors such as pnictogens and tetrels may possess substituents that make the approach of a trivalent nitrogen less favourable when forming a σ -hole interaction. These Lewis bases possess a lone-pair in an orbital with the greatest amount of p -character, which are the least contracted and thus most available lone-pair for participating in σ -hole interactions. This presents the likelihood that these two characteristics will act in a competitive manner when considering σ -hole interactions to sp^3 -hybridised nitrogen moieties. An overall summary of the geometric observables of contacts indicative of σ -hole interactions to sp^3 -hybridised nitrogen is tabulated in Table 3.1.

Table 3.1: Overall geometric summary of the contacts indicating σ -hole interactions to sp^3 -hybridised nitrogen

σ -hole Interaction Type	No. Contacts	Mean Normalised Contact Distance	Mean σ -hole Angle/ $^\circ$	Mean Lone-Pair Angle/ $^\circ$
Halogen-bonding	707	0.893	158	155
Chalcogen-bonding	153	0.984	148	142
Pnictogen-bonding	25	0.947	156	145
Tetrel-bonding	42	0.977	146	153

3.3.1 $X \cdots sp^3$ -Nitrogen Contacts

Halogen bonding contacts involving sp^3 -hybridised nitrogen Lewis bases are the most numerous out of the σ -hole interactions to sp^3 -hybridised nitrogen. With 707 contacts, they are an excellent example to highlight the directionality of σ -hole

interactions.

The distribution of normalised contact distances, as defined in Equation (2.1), in relation to the σ -hole angle is illustrated in the scatterplot of Figure 3.2. Contacts shorter than the sum of van der Waals radii have a distinct "wedge" shaped distribution, with the very close (0.6-0.7 normalised contact distance) contacts having a σ -hole angle ranging between 170-180°, with the range broadening to 150-180° at 0.9 normalised contact distance. At contact distances beyond the sum of van der Waals radii (normalised contact distance > 1.0) there is a much wider, approximately random distribution of σ -hole angles, suggesting that the interactions are likely weak in nature and are easily perturbed by other crystal packing effects. Figure 3.3 shows that although the directionality with respect to the lone-pair angle is not as stark as the σ -hole angle it is still notable, especially at contacts with short normalised contact distances. The directionality with respect to these angles is further shown as mean averages in Table 3.2, with the trends following the polarisability of the donor; average normalised contact distances follow the trend of $I < Br < Cl$, and the trends in both σ -hole angle and lone-pair angle follows the trend $I > Br > Cl$. This suggests that the larger more polarisable halogens are forming stronger σ -hole interactions with the nitrogen lone-pair, indicated by better angular alignment of the σ -hole and lone-pair in the closest contacts.

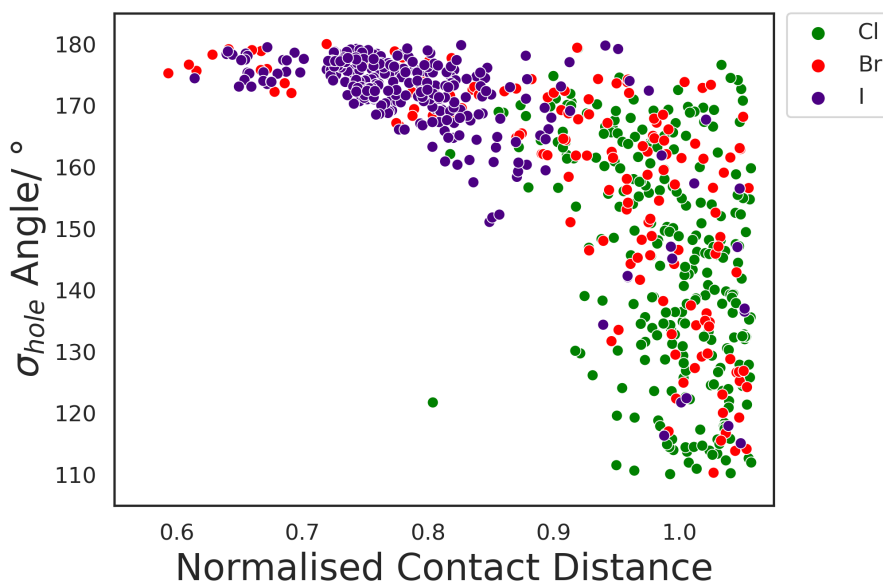


Figure 3.2: Scatter plots showing the distribution of normalised contact distance with respect to the σ -hole angle of halogen bonding contacts to sp^3 -hybridised nitrogen found in crystal structures in the CSD. Points colour coded by donor atom type.

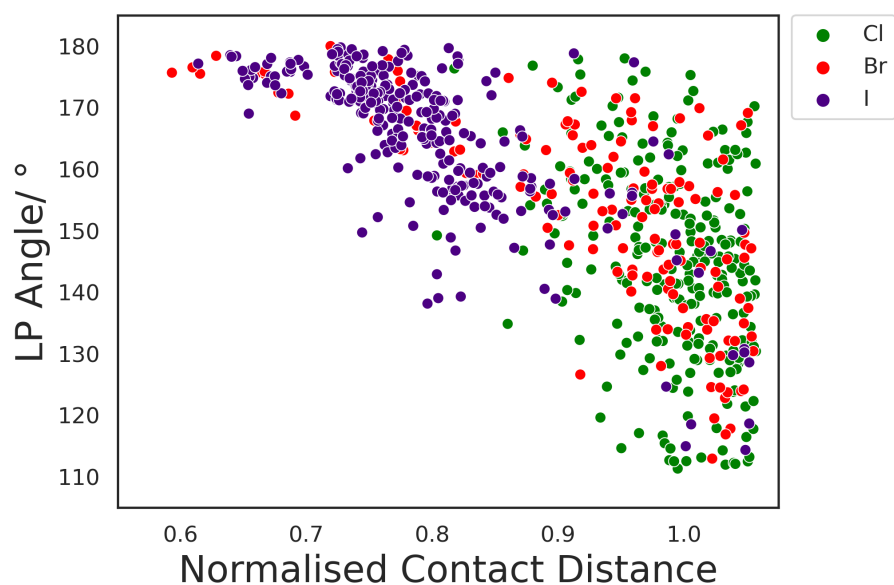


Figure 3.3: Scatter plots showing the distribution of normalised contact distance with respect to the lone-pair angle of halogen bonding contacts to sp^3 -hybridised nitrogen found in crystal structures in the CSD. Points colour coded by donor atom type.

Table 3.2: Summary of geometric characteristics of halogen (X) bonding contacts to sp^3 -hybridised nitrogen

Donor	Contacts	Mean Contact Distance/ Å	Mean Normalised Contact Distance	Mean σ -hole Angle/ °	Mean Lone-Pair Angle/ °
X	707	3.187	0.893	158	155
I	291	2.927	0.791	170	165
Br	153	3.232	0.918	157	154
Cl	263	3.449	0.991	144	145

3.3.2 Ch ... sp^3 -Nitrogen Contacts

In comparison to halogen bonding contacts to sp^3 -hybridised nitrogen, there are an order of magnitude fewer contacts found involving chalcogen bonding interactions, see Table 3.1, to sp^3 -hybridised nitrogen, but still some analogous trends emerge.

In terms of directionality with respect to the σ -hole angle, shown in Figure 3.4, the same "wedge" distribution is seen, albeit with a greater sparsity in data points. At contact distances beyond the sum of van der Waals radii there exists a large range of σ -hole angles, indicating their weak nature. For contacts smaller than the sum of van der Waals radii, the σ -hole angle range is 150-180°. Interestingly the contacts involving selenium are more directional with respect to the σ -hole angle than those involving tellurium. Table 3.3 shows that although the number of contacts is very low, selenium contacts have an average σ -hole angle of 170°, whereas tellurium contacts have an average σ -hole angle of only 160°. It is important to note that the selenium contacts are more varied in origin whereas the four short tellurium contacts

correspond to a single structure, CSD refcode **XUHHEQ**, which forms multiple chalcogen bonds¹³⁴.

Figure 3.5 illustrates the directionality of these contacts with respect to the lone-pair angle. Much the same trends can be seen as with halogen bonding contacts, with a very slight "wedge" distribution. A similar range in angle is seen for the lone-pair angle as compared to halogen bonds to sp^3 -hybridised nitrogen, albeit at contacts shorter than 0.9 normalised contact distance. Although the data is sparse compared to halogen bonding contacts, some observations can still be made. Notably, as highlighted in Table 3.3, the average lone-pair angles follow the trend $\text{Te} > \text{Se} > \text{S}$, which does not match the trend seen with σ -hole angle. This suggests that although the tellurium contacts deviate further from linearity with respect to the σ -hole angle, they align to a greater degree with the lone-pair of the Lewis base, indicating a stronger interaction. This can also be seen in the trend in normalised contact distance, which follows the trend $\text{Te} < \text{Se} < \text{S}$. Due to there only being 5 tellurium contacts and 7 selenium contacts, meaningful, statistically significant trends cannot be drawn and are likely due to selection bias as mentioned previously.

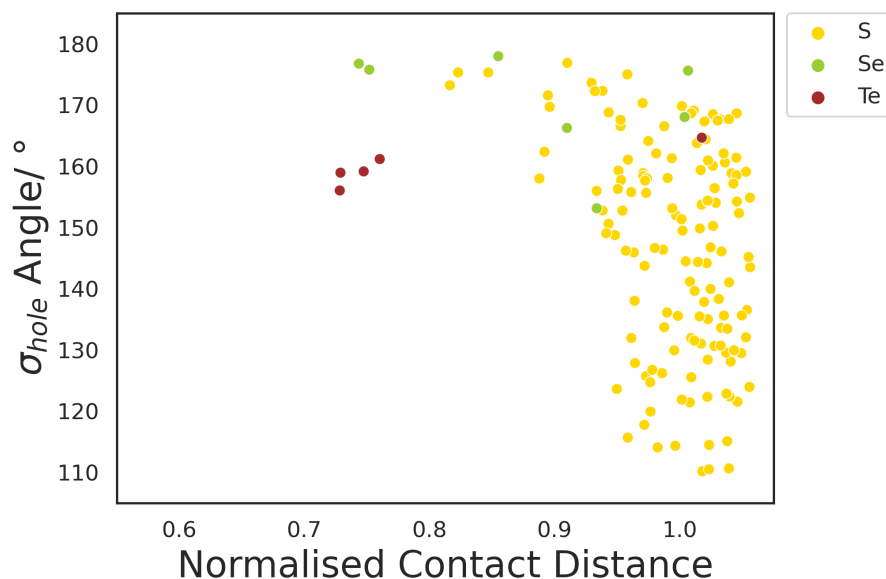


Figure 3.4: Scatter plot showing the distribution of normalised contact distance with respect to the σ -hole angle of chalcogen bonding contacts to sp^3 -hybridised nitrogen found in crystal structures in the CSD. Points colour coded by donor atom type.

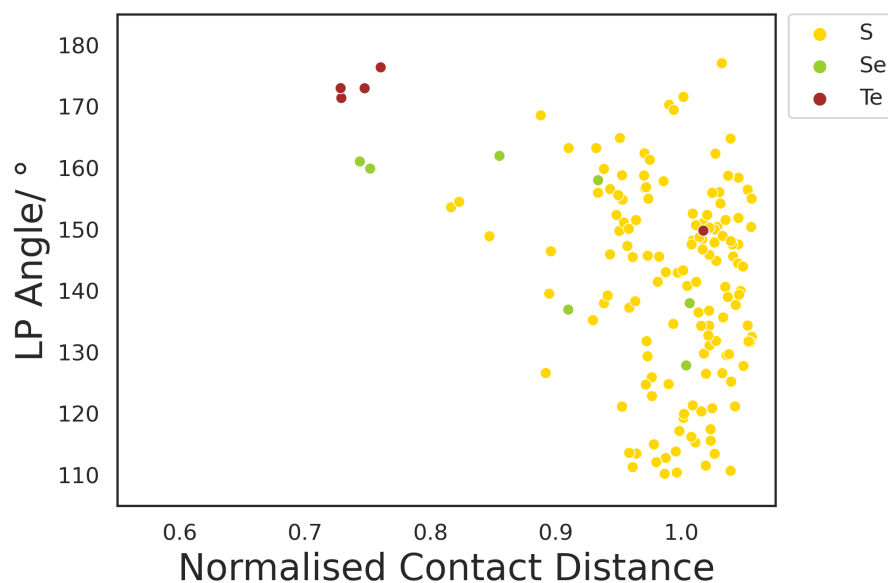


Figure 3.5: Scatter plot showing the distribution of normalised contact distance with respect to the lone-pair angle of chalcogen bonding contacts to sp^3 -hybridised nitrogen found in crystal structures in the CSD. Points colour coded by donor atom type.

Table 3.3: Summary of geometric characteristics of chalcogen (*Ch*) bonding contacts to sp^3 -hybridised nitrogen

Donor	Contacts	Mean Contact Distance/ Å	Mean Normalised Contact Distance	Mean σ -hole Angle/ °	Mean Lone-Pair Angle/ °
Ch	153	3.494	0.984	148	142
Te	5	2.907	0.796	160	168
Se	7	3.084	0.886	170	149
S	141	3.535	0.996	147	141

3.3.3 $Pn \cdots sp^3$ -Nitrogen Contacts

Pnictogen bonding contacts to sp^3 -hybridised nitrogen shows the lowest abundance out of the σ -hole interactions to sp^3 -hybridised nitrogen shown in Table 3.1. With only 25 contacts that match the search criteria, it is difficult to find any specific trends with respect to the nature of these interactions from the gathered geometric data. What can be seen from the scatter plot in Figure 3.6 is that the contacts within the sum of van der Waals radii have a σ -hole angle range of 150-180°. Likewise directionality is seen with the lone-pair angle, shown in Figure 3.7, but with the range being 150-170°. The averages in geometric measurements are tabulated and shown in Table 3.4, but only phosphorus contacts and antimony contacts have multiple contacts for which to average over, but it does show that antimony contacts have much shorter normalised contact distances on average but have comparable σ -hole angles.

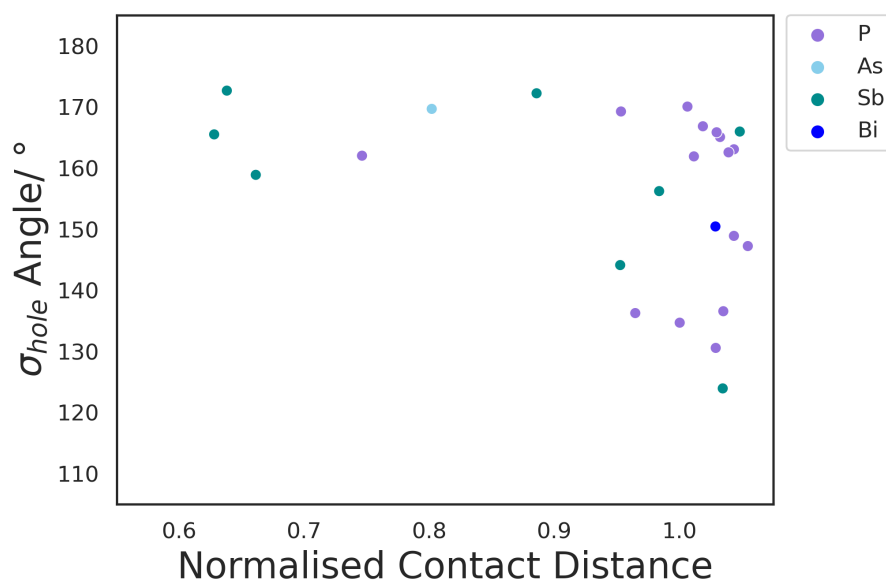


Figure 3.6: Scatter plot showing the distribution of normalised contact distance with respect to the σ -hole angle of pnictogen bonding contacts sp^3 -hybridised nitrogen found in crystal structures in the CSD. Points colour coded by donor atom type.

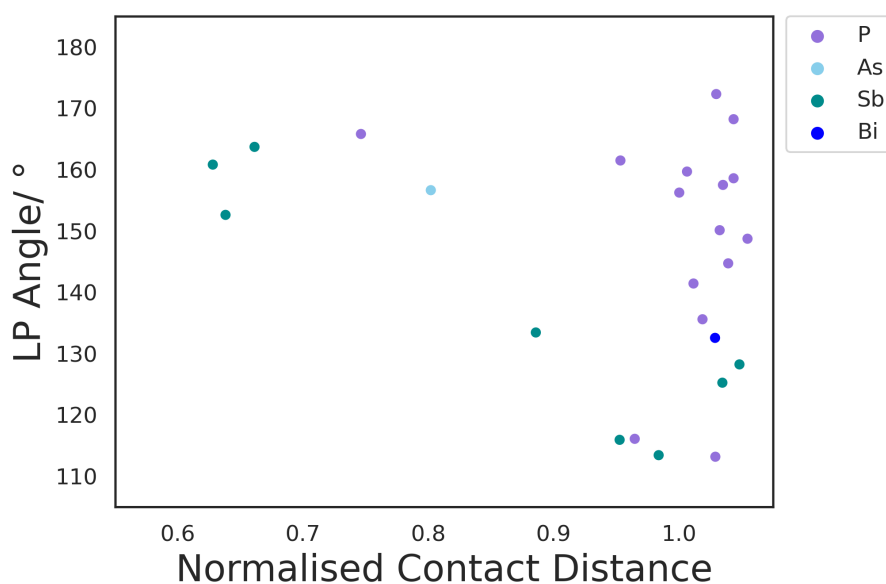


Figure 3.7: Scatter plot showing the distribution of normalised contact distance with respect to the lone-pair angle of pnictogen bonding contacts to sp^3 -hybridised nitrogen found in crystal structures in the CSD. Points colour coded by donor atom type.

Table 3.4: Summary of geometric characteristics of pnictogen (*Pn*) bonding contacts to sp^3 -hybridised nitrogen

Donor	Contacts	Mean Contact Distance/ Å	Mean Normalised Contact Distance	Mean σ -hole Angle/ °	Mean Lone-Pair Angle/ °
Pn	25	3.552	0.947	156	145
Bi	1	4.321	1.029	150	132
Sb	8	3.527	0.854	157	136
As	1	2.839	0.802	169	156
P	15	3.562	1.001	154	149

3.3.4 *Tt* ... sp^3 -Nitrogen Contacts

Despite having almost double the number of contacts as for pnictogen bonding contacts, see Table 3.1, the tetrel contacts to sp^3 -hybridised nitrogen are also very sparse. As seen in Figure 3.8, it would seem that the only contacts with a strong tetrel bond are those involving tin, and possibly a very weak near linear contact involving silicon. It is also apparent that there is an absence of contacts involving lead, the largest of the tetrel bond donors. The directionality with respect to the lone-pair, shown in Figure 3.9 would suggest that the silicon contact does not align markedly with the lone pair, suggesting that a tetrel bond is unlikely to be the dominant contribution to this contact. The contacts involving tin show more directionality with respect to the lone-pair, with the average lone-pair angle being 157° , as shown in Table 3.5.

The number of tetrel bonding contacts to sp^3 -hybridised nitrogen, shown in Table 3.5, is greater than the number of analogous pnictogen bonding contacts, and hence the average geometric measurements are potentially more useful. The trend in σ -hole angle reflects the polarisability of the tetrel atom $\text{Sn} > \text{Ge} > \text{Si}$, likely due to the electron density of these more polarisable donors is easier to perturb to form a σ -hole. Interestingly this is not seen with the trend in normalised contact distance which is $\text{Sn} < \text{Si} < \text{Ge}$, and the trend in lone-pair angle is $\text{Si} > \text{Sn} > \text{Ge}$. It is important to remember that these values are still averages over a very small number of data points, so should be cautiously compared to average geometric data from σ -hole interacting motifs with a greater abundance of contacts due to the aforementioned effects of selection bias.

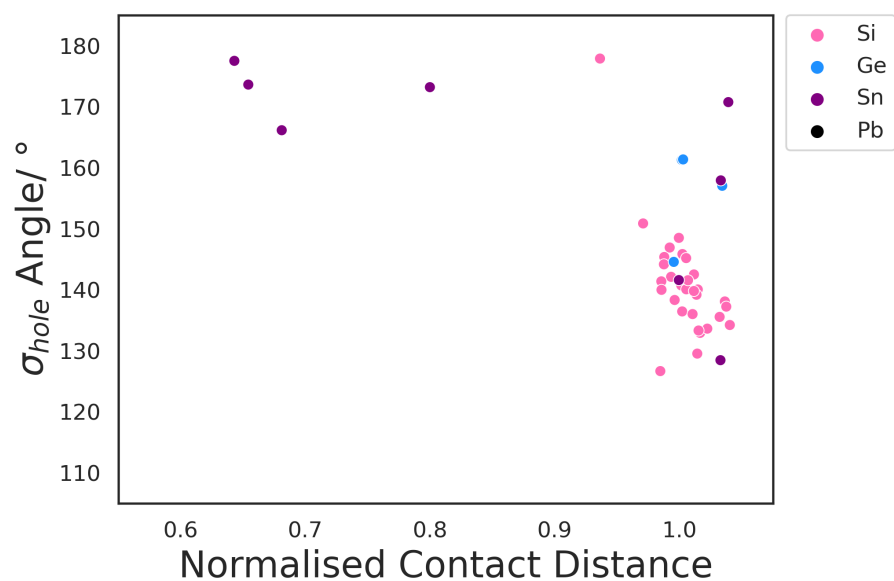


Figure 3.8: Scatter plot showing the distribution of normalised contact distance with respect to the σ -hole angle of tetrel bonding contacts to sp^3 -hybridised nitrogen found in crystal structures in the CSD. Points colour coded by donor atom type.

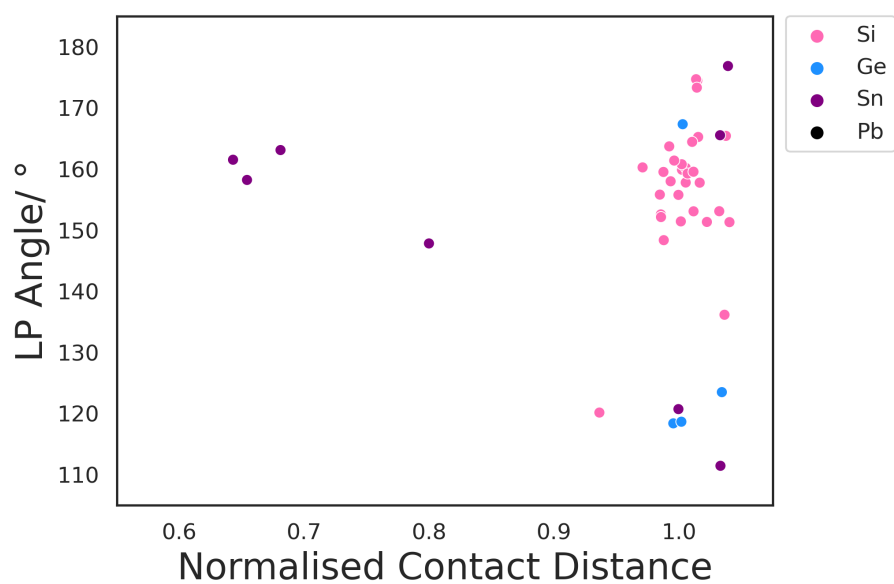


Figure 3.9: Scatter plot showing the distribution of normalised contact distance with respect to the lone-pair angle of normalised tetrel bonding contacts to sp^3 -hybridised nitrogen found in crystal structures in the CSD. Points colour coded by donor atom type.

Table 3.5: Summary of geometric characteristics of tetrel (*Tt*) bonding contacts to sp^3 -hybridised nitrogen

Donor	Contacts	Mean Contact Distance/ Å	Mean Normalised Contact Distance	Mean σ -hole Angle/ °	Mean Lone-Pair Angle/ °
Tt	42	3.809	0.977	146	153
Pb	0	0.000	0.000	0	0
Sn	8	3.510	0.860	161	150
Ge	4	3.984	1.009	156	131
Si	30	3.865	1.004	140	157

3.3.5 Prevalence of $\sigma_{hole} \cdots sp^3$ -Nitrogen Contacts

To explore the likelihood of σ -hole interactions to sp^3 -hybridised nitrogen forming within a crystal, the number of unique crystal structures that contain a contact has been compared to the number of unique crystal structures containing the necessary moieties in the same structure. This ratio is given as a percentage referred to as the prevalence, see Equation (2.2). Over all of the motifs, shown in Table 3.6, there is a large disparity between the number of crystal structures containing the required moieties and the number of crystal structures containing a contact.

Table 3.6: Overall summary of the prevalence of contacts indicating σ -hole interactions to sp^3 -hybridised nitrogen compared to the number of unique crystal structures found in the CSD that contain the necessary moieties.

σ -hole Interaction Type	Structures with Moieties	Structures with Contact(s)	Prevalence
Halogen-bonding	6340	499	7.87%
Chalcogen-bonding	2277	135	5.93%
Pnictogen-bonding	406	15	3.69%
Tetrel-bonding	1056	30	2.84%

For halogen bonding motifs, which is the most prevalent motif, Table 3.7 shows that only 7.87% of structures containing these moieties exhibit halogen bonding behaviour to sp^3 -hybridised nitrogen. This is very low in prevalence, but it is still much greater than for chalcogen, pnictogen, and tetrel bonding motifs involving sp^3 -hybridised nitrogen. Out of the three halogen atoms, motifs involving iodine show that nearly a third of the crystals containing this motif participate in halogen bonding. Motifs involving bromine and chlorine show fairly similar prevalence to each other, 6.30% and 5.34% respectively, but it is important to note that there are many more structures that contain bromine or chlorine than iodine.

Table 3.7: Summary of the prevalence of halogen (X) bonding contacts to sp^3 -hybridised nitrogen compared to the number of unique crystal structures found in the CSD that contain the necessary moieties

Donor	Structures with Motif	Structures with Contact(s)	Prevalence
X	6340	499	7.87 %
I	480	159	33.12 %
Br	1984	125	6.30 %
Cl	4060	217	5.34 %

Table 3.8 shows that the prevalence of chalcogen bonding to sp^3 -hybridised nitrogen is less than that of the analogous halogen bonding motifs, at 5.93%. Out of these chalcogens, tellurium motifs show the highest prevalence with 10.53%, however this is a percentage of a very small number of structures that contain this motif (19 structures). Selenium motifs have around half the prevalence of tellurium motifs at 4.20%, but this is still a percentage of a small number of structures (143 structures). Sulfur contacts have an increased prevalence of 5.97%, and make up the majority of the structures showing contacts indicative of chalcogen bonding, and constitute an even greater majority of the crystal structures that contain the required moieties (2128 structures). This is likely a reflection on synthetic accessibility and the widespread use of sulfur in organic chemistry, whereas selenium and tellurium chemistry is less extensively studied¹³⁵.

Table 3.8: Summary of the prevalence of chalcogen (Ch) bonding contacts to sp^3 -hybridised nitrogen compared to the number of unique crystal structures found in the CSD that contain the necessary moieties

Donor	Structures with Motif	Structures with Contact(s)	Prevalence
Ch	2277	135	5.93 %
Te	19	2	10.53 %
Se	143	6	4.20 %
S	2128	127	5.97 %

When considering the prevalence of pnictogen and tetrel contacts involving sp^3 -hybridised nitrogen, the numbers of structures compared to the total number of crystal structures containing the required moieties is very small in comparison to halogen and chalcogen contacts. This is a reflection of the aforementioned selection bias within these crystal structures, as there is comparatively fewer crystals synthesised using these elements. This combined with the fact that even fewer of those containing σ -hole contacts, thus the whole chemical space of the interactions are unlikely to be completely covered in this data.

It is interesting to note that, in Table 3.9 and Table 3.10, there are 406 structures that contain the required pnictogen and sp^3 -hybridised nitrogen moieties, and 1056 structures that contain the required tetrel and sp^3 -hybridised nitrogen moieties, yet only 15 and 30 structures containing contacts, respectively. Given that these are the two most sterically bulky classes of σ -hole donors, interacting with the most sterically bulky hybridisation of nitrogen, this suggests a notable steric hindrance

to the formation of pnictogen and tetrel bonds within structures containing the required moieties.

Table 3.9: Summary of the prevalence of pnictogen (*Pn*) bonding contacts to sp^3 -hybridised nitrogen compared to the number of unique crystal structures found in the CSD that contain the necessary moieties

Donor	Structures with Motif	Structures with Contact(s)	Prevalence
Pn	406	15	3.69 %
Bi	24	1	4.17 %
Sb	46	1	2.17 %
As	18	1	5.56 %
P	321	12	3.74 %

Table 3.10: Summary of the prevalence of tetrel (*Tt*) bonding contacts to sp^3 -hybridised nitrogen compared to the number of unique crystal structures found in the CSD that contain the necessary moieties

Donor	Structures with Motif	Structures with Contact(s)	Prevalence
Tt	1056	30	2.84 %
Pb	8	0	0.00 %
Sn	117	6	5.13 %
Ge	58	3	5.17 %
Si	884	21	2.38 %

3.3.6 Notable sp^3 -Nitrogen Moieties

It would seem, from the findings presented in Section 3.3, that although sp^3 -hybridised nitrogen Lewis bases are capable of being very good σ -hole acceptors, resulting in short, highly directional contacts, there is a notable steric impact on the interactions. This means that to suitably exploit the highly nucleophilic lone pair located on the nitrogen, either the acceptor moiety needs to be engineered to minimise steric bulk around the nitrogen, or the surroundings in the crystal environment need to be engineered to direct the donor and acceptor moieties together. A noteworthy example of a suitable acceptor for this purpose is shown in Figure 3.10, an image of the CSD refcode **SUMZEI**, which utilises a Lewis base called 1,4-diazabicyclo[2,2,2]octane, better known as DABCO¹³⁶. This acceptor minimises the steric impact of the sp^3 -hybridised nitrogen environment by exploiting the ring strain of this cage-like molecule. This highly accessible lone-pair results in a very strong Lewis base that, when paired with halogen bond donors such as this 1,2,3,4-tetrafluoro-5,6-diiodobenzene, results in a very strong σ -hole interaction. An analogous acceptor structure is shown in Figure 3.11, showing CSD refcode **GOBYOO**, featuring a 1-azabicyclo[2,2,2]octane, also known as quinuclidine, this time interacting with a strong selenazole-based chalcogen bond donor¹³⁷.

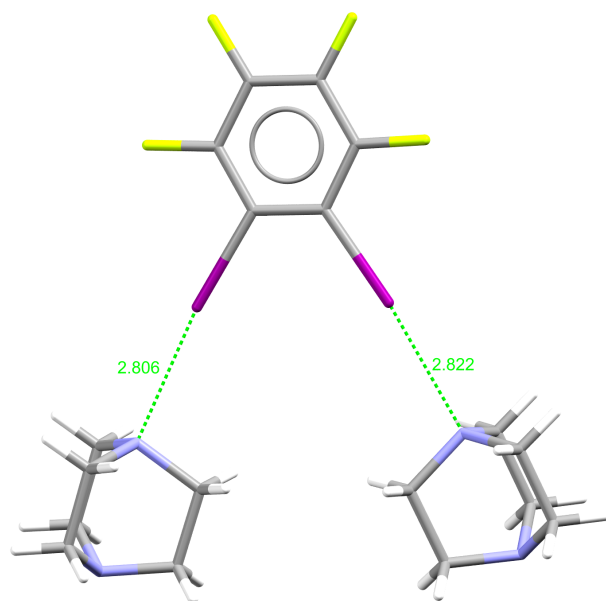


Figure 3.10: Two contacts between tetrafluoro-diiodobenzene and two DABCO molecules. This is an archetypal example of a crystal structure with a halogen bonding contact to an sp^3 -hybridised nitrogen Lewis base. I $\cdots$ N contact distances given in Ångström. CSD refcode: **SUMZEI**

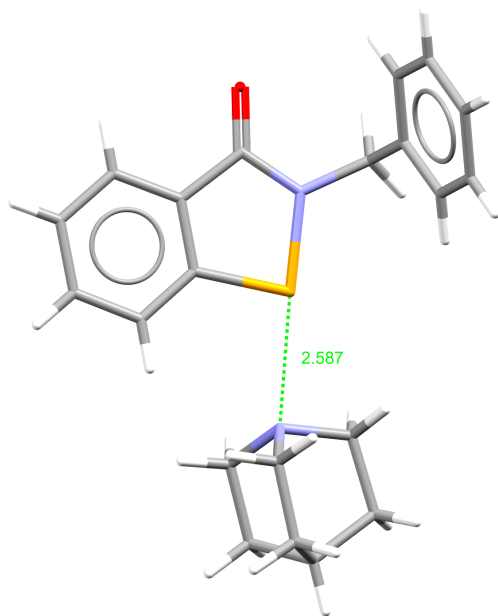


Figure 3.11: A contact between benzoselenazolinone derivative and a quinuclidine molecule. This is an archetypal example of a crystal structure with a chalcogen bonding contact to an sp^3 -hybridised nitrogen Lewis base. Se $\cdots$ N contact distances given in Ångström. CSD refcode: **GOBYOO**

3.4 sp^2 Hybridised Nitrogen

Divalent sp^2 -hybridised nitrogen σ -hole acceptors include amides, imines, nitrogen heterocycles and various other nitrogen functional groups. These moieties are, by nature of their hybridisation, likely to be planar, which would dictate that a notable portion of the molecular environment would exist in this same plane, including the lone-pair on the nitrogen. This would predict that σ -hole donors would approach these lone-pairs along this same plane, but rotation of the donor and acceptor moieties could minimise steric interactions. The lone-pairs on these nitrogen are more contracted due to greater s -character than the aforementioned sp^3 -hybridised nitrogen σ -hole acceptors, suggesting their lower effectiveness in participating in σ -hole interactions. An overall summary of the geometric observables of contacts indicative of σ -hole interactions to sp^2 -hybridised nitrogen is tabulated in Table 3.11. Overall there is a much greater number of contacts involving sp^2 -hybridised nitrogen than those involving sp^3 -hybridised nitrogen, this is likely due to the plethora of aromatic nitrogen heterocycles that are used in synthesis, compared to sp^3 -hybridised amines.

Table 3.11: Overall geometric summary of the contacts indicative of σ -hole interactions to sp^2 -hybridised nitrogen

σ -hole Interaction Type	No. Contacts	Mean Normalised Contact Distance	Mean σ -hole Angle/ $^\circ$	Mean Lone-Pair Angle/ $^\circ$
Halogen-bonding	6002	0.913	158	148
Chalcogen-bonding	5302	0.944	153	140
Pnictogen-bonding	52	0.890	160	145
Tetrel-bonding	200	0.859	153	147

3.4.1 $X \cdots sp^2$ -Nitrogen Contacts

Starting with halogen bonding contacts to sp^2 -hybridised nitrogen, Table 3.11 shows a massive increase in the number of data points compared to halogen bonding contacts to sp^3 -hybridised nitrogen (see Table 3.1). The distribution of normalised contact distances with respect to the σ -hole angle and the lone-pair angle in Figure 3.12 and Figure 3.13 respectively, shows a broadening of the "wedge" shaped distribution approaching the sum of van der Waals radii. An interesting trend in these scatter plots is seen in contacts with normalised contact distances shorter than 0.75. In these very short contacts, the identity of halogen donor atoms is more mixed, with the two closest contacts involving bromine and chlorine as halogen bond donors. For normalised contact distances greater than 0.75, the contact distance, and linearity of the σ -hole and lone-pair angles is linked to the polarisability of the halogen. The overall trends with respect to the σ -hole donor polarisability are reflected in the mean averages listed in Table 3.12. The iodine contacts are the shortest and have σ -hole and lone-pair angles closest to linearity, then bromine contacts, and then the chlorine contacts, which are the longest and have the σ -hole and lone-pair angle furthest from linearity. This further supports that the larger more polarisable halogens are forming stronger σ -hole interactions with the nitrogen lone-pair, indicated by better angular alignment of the σ -hole and lone-pair in the closest contacts.

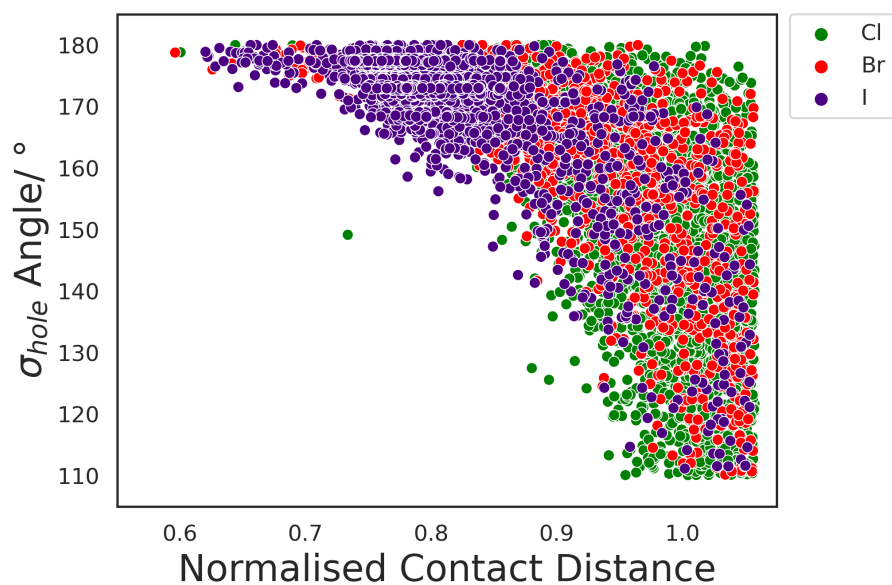


Figure 3.12: Scatter plot showing the distribution of normalised contact distance with respect to the σ -hole angle of halogen bonding contacts to sp^2 -hybridised nitrogen found in crystal structures in the CSD. Points colour coded by donor atom type.

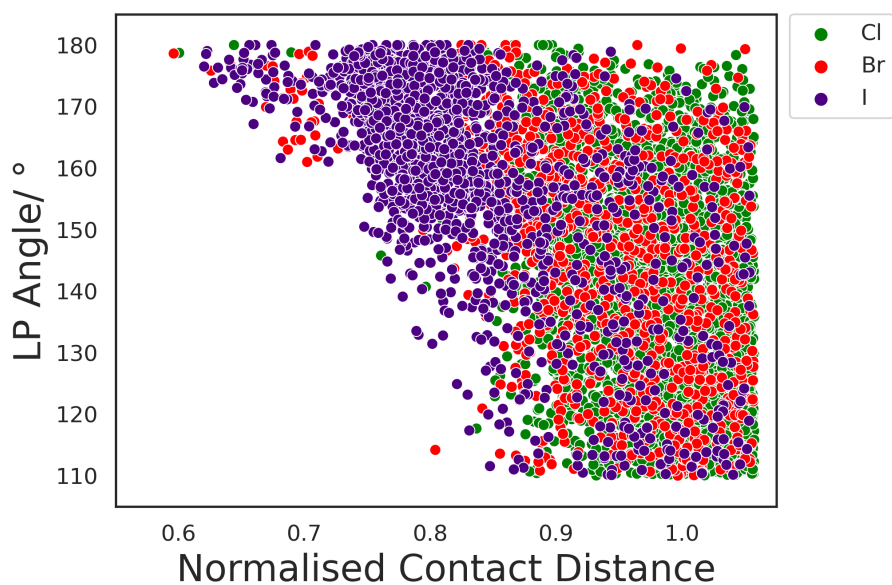


Figure 3.13: Scatter plot showing the distribution of normalised contact distance with respect to the lone-pair angle of halogen bonding contacts to sp^2 -hybridised nitrogen found in crystal structures in the CSD. Points colour coded by donor atom type.

Table 3.12: Summary of geometric characteristics of halogen (X) bonding contacts to sp^2 -hybridised nitrogen

Donor	Contacts	Mean Contact Distance/ Å	Mean Normalised Contact Distance	Mean σ -hole Angle/ °	Mean Lone-Pair Angle/ °
X	6002	3.245	0.913	158	148
I	2066	3.017	0.815	170	161
Br	1172	3.288	0.934	158	146
Cl	2764	3.397	0.976	149	140

3.4.2 Ch ... sp^2 -Nitrogen Contacts

For chalcogen bonding contacts to sp^2 -hybridised nitrogen, Table 3.11 shows an almost thirty five times increase in the number of contacts compared to chalcogen bonds to sp^3 -hybridised nitrogen (Table 3.1). The characteristic "wedge" shaped distribution associated with these plots of σ -hole against normalised contact distance, see Figure 3.14, shows a notable perturbation compared to the distribution seen with analogous halogen bonding contacts (see Figure 3.2). For selenium contacts the expected "wedge" shape is present, having a comparable distribution with bromine contacts to sp^2 -hybridised nitrogen. The contacts involving tellurium show an almost bimodal distribution, one distribution being located in similar σ -hole angle and normalised contact distance ranges to the selenium contacts. The second, more tightly packed distribution of points between 0.7 and 0.8 normalised contact distance have a σ -hole angle range of 150-160°. This second distribution of tellurium contacts are co-located with a distribution of sulfur contacts at a similar range for normalised contact distance, but at σ -hole angles of around 130°.

Similar trends can be seen in Figure 3.15 of the distributions of normalised contact distances with respect to the lone-pair angle. The selenium contacts show a similar "wedge" distribution to halogen bonding contacts with sp^2 -hybridised nitrogen, see Figure 3.13. The distribution of tellurium contacts has a similar spread compared to the selenium contacts, but there is a clustering of contacts around lone-pair angles of around 150° and normalised contact distances below 0.8. The main difference is that sulfur contacts appears to show little to no directionality with respect to the lone pair of the sp^2 -hybridised nitrogen.

The average geometric observables listed in Table 3.13 shows that the average normalised contact distance decreases with increasing polarisability of the chalcogen atom. Similarly, the σ -hole angle and the lone-pair angle increase from sulfur to tellurium. The selenium and tellurium contacts are very similar in terms of average σ -hole angle, with tellurium showing, on average, interactions with lone-pair angles closer to linearity.

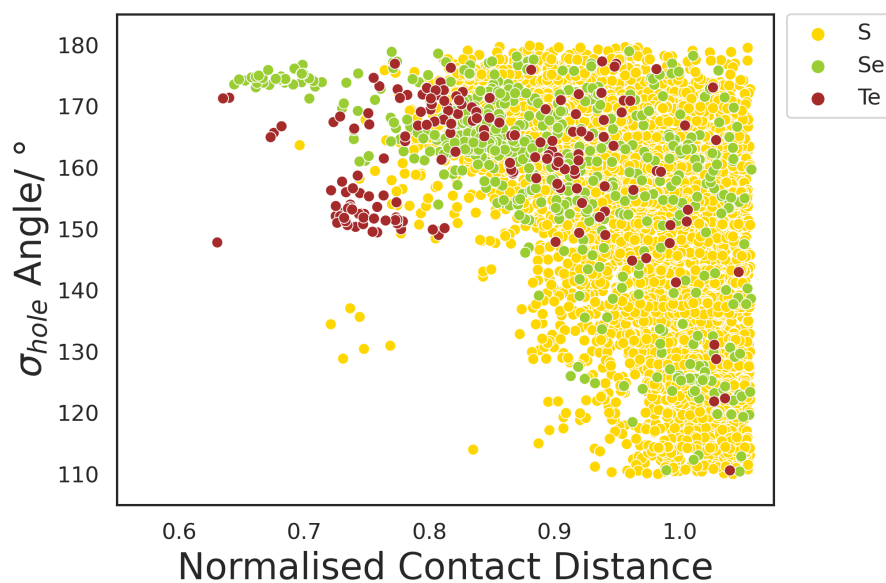


Figure 3.14: Scatter plot showing the distribution of normalised contact distance with respect to the σ -hole angle of chalcogen bonding contacts to sp^2 -hybridised nitrogen found in crystal structures in the CSD. Points colour coded by donor atom type.

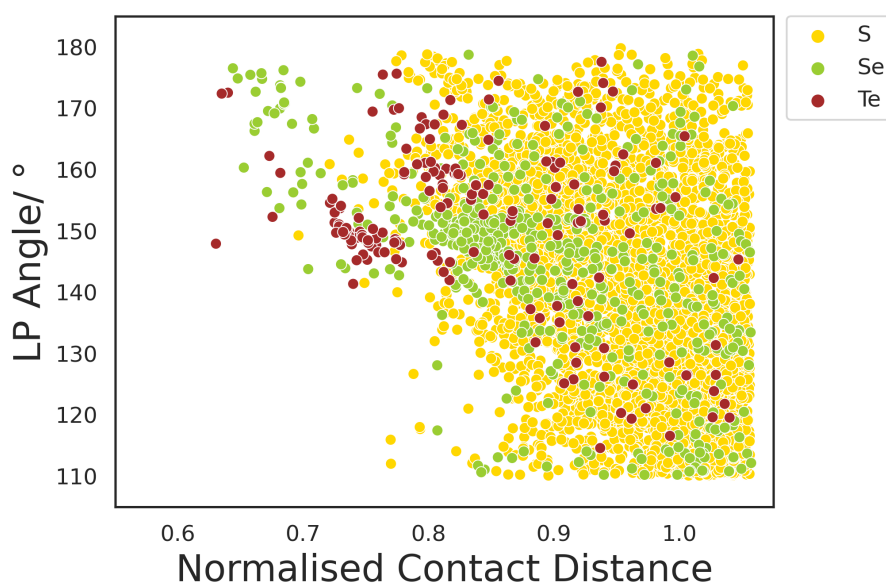


Figure 3.15: Scatter plot showing the distribution of normalised contact distance with respect to the lone-pair angle of chalcogen bonding contacts to sp^2 -hybridised nitrogen found in crystal structures in the CSD. Points colour coded by donor atom type.

Table 3.13: Summary of geometric characteristics of chalcogen (*Ch*) bonding contacts to sp^2 -hybridised nitrogen

Donor	Contacts	Mean Contact Distance/ Å	Mean Normalised Contact Distance	Mean σ -hole Angle/ °	Mean Lone-Pair Angle/ °
Ch	5302	3.349	0.944	153	140
Te	171	3.071	0.841	160	150
Se	586	3.097	0.890	159	145
S	4545	3.392	0.955	152	139

3.4.3 $Pn \cdots sp^2$ -Nitrogen Contacts

Table 3.11 shows that pnictogen bonding contacts involving sp^2 -hybridised nitrogen are approximately twice as numerous as to those involving sp^3 -hybridised nitrogen. Despite the scarcity of contacts, the distribution of normalised contact distance with respect to the σ -hole angle plotted in Figure 3.16 shows that the vast majority of the contacts have a σ -hole angle $> 140^\circ$, even contacts with distances beyond the sum of van der Waals radii.

Of the mean averages of geometric observables listed in in Table 3.14, antimony contacts make up the majority of the pnictogen bonding contacts, with an average σ -hole angle of 162° and a average normalised contact distance of 0.812. There are also a number of bismuth contacts that are involved in slightly longer contacts than the antimony contacts, but with similar σ -hole angles. Arsenic and phosphorus contacts to sp^2 -hybridised nitrogen are the least commonly occurring, with the arsenic contacts having, on average, shorter normalised contact distances than the phosphorus contacts. Both arsenic and phosphorus contacts have average normalised contact distances considerably longer than bismuth or antimony contacts.

The distribution of normalised contact distance with respect to the lone-pair angle plotted in Figure 3.17 show a clear trend that as the contacts get longer, the lone-pair angle decreases, with only one arsenic contact showing a somewhat linear lone-pair angle. In terms of the average trends shown in Table 3.14 of pnictogen bonds to sp^2 -hybridised nitrogen, the normalised contact distances follow the trend $Sb < Bi < As < P$ and the σ -hole and lone-pair angles follow the trend $P < As < Bi = Sb$. It is important to note that these are averages of a very small amount of data, so should be treated with caution when compared to other σ -hole interactions.

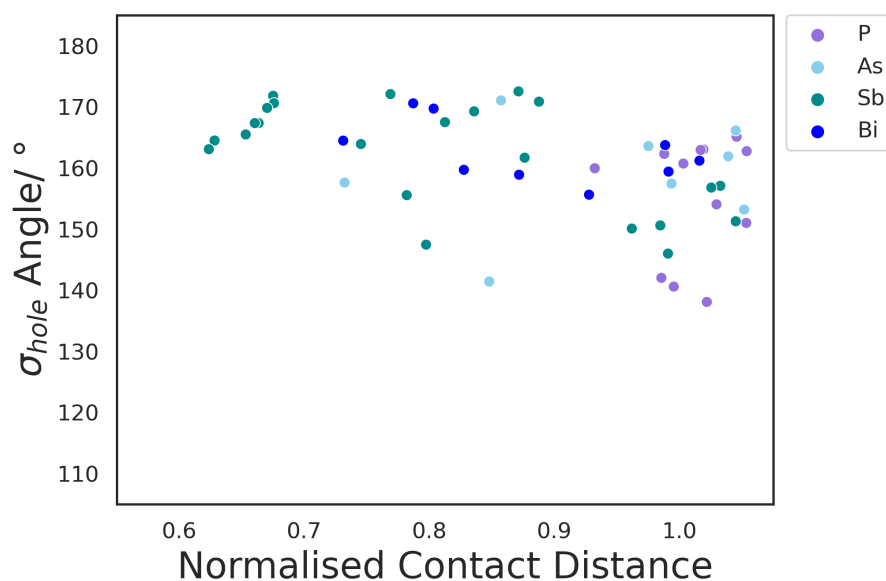


Figure 3.16: Scatter plot showing the distribution of normalised contact distance with respect to the σ -hole angle of pnictogen bonding contacts to sp^2 -hybridised nitrogen found in crystal structures in the CSD. Points colour coded by donor atom type.

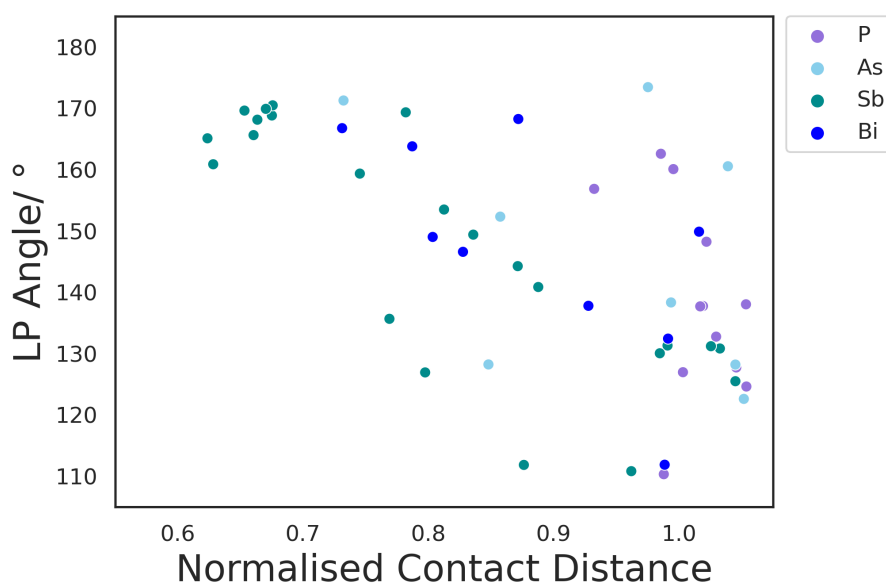


Figure 3.17: Scatter plot showing the distribution of normalised contact distance with respect to the lone-pair angle of pnictogen bonding contacts to sp^2 -hybridised nitrogen found in crystal structures in the CSD. Points colour coded by donor atom type.

Table 3.14: Summary of geometric characteristics of pnictogen (*Pn*) bonding contacts to sp^2 -hybridised nitrogen

Donor	Contacts	Mean Contact Distance/ Å	Mean Normalised Contact Distance	Mean σ -hole Angle/ °	Mean Lone-Pair Angle/ °
Pn	52	3.469	0.890	160	145
Bi	9	3.707	0.883	162	147
Sb	23	3.352	0.812	162	147
As	8	3.337	0.943	159	146
P	12	3.603	1.012	155	138

3.4.4 *Tt* ... sp^2 -Nitrogen Contacts

The number of tetrel bonding contacts to sp^2 -hybridised nitrogen, shown in Table 3.11, are much fewer than halogen and chalcogen bonding contacts to sp^2 -hybridised nitrogen, but much more than pnictogen bonding contacts, so more well-grounded analysis can be made. One notable aspect of these contacts is only a single example of germanium as a σ -hole donor.

The scatter plot of σ -hole angle against normalised contact distance in Figure 3.18 shows three distributions of contacts that show very different geometric characteristics. The first distribution involves a population of very short contacts with a range of 0.6-0.9 in normalised contact distance, and involve mainly tin, and a single lead contact, that are near linearity with respect to σ -hole angle. The second distribution of contacts has a comparable range in normalised contact distance, but the range of σ -hole angles deviates further from linearity, being 135-160°. These contacts make up the majority of the contacts involving lead but there are also a slightly smaller number of tin contacts. The final distribution is comprised of long contacts that have normalised contact distance greater than 0.95, which primarily consists of silicon contacts to sp^2 -hybridised nitrogen, with a small number of tin and lead contacts. This distribution shows large deviation from linearity with respect to the σ -hole angle, with angles ranging from 125-160°.

Overall the distribution of lone-pair angle against normalised contact distance, shown in Figure 3.19, shows very little directionality with respect to the lone-pair. However, the closest tin contacts have near linear lone-pair angles, but the closest lead contacts show no directionality. Silicon contacts are very long and show little to no directionality, with a broad range of lone-pair angles. The average values of the geometric criteria for tetrel bonding contacts to sp^2 -hybridised nitrogen are shown in Table 3.15. The trend in normalised contact distance is $\text{Sn} < \text{Pb} < \text{Si}$, which reflects the polarisability of the tetrel atom. The trend in average σ -hole angle is $\text{Si} < \text{Pb} < \text{Sn}$, and the trend in lone-pair angle is $\text{Pb} < \text{Si} < \text{Sn}$. Silicon contacts are the most numerous, followed by tin contacts, and then lead contacts, with germanium only having a single contact so cannot be averaged.

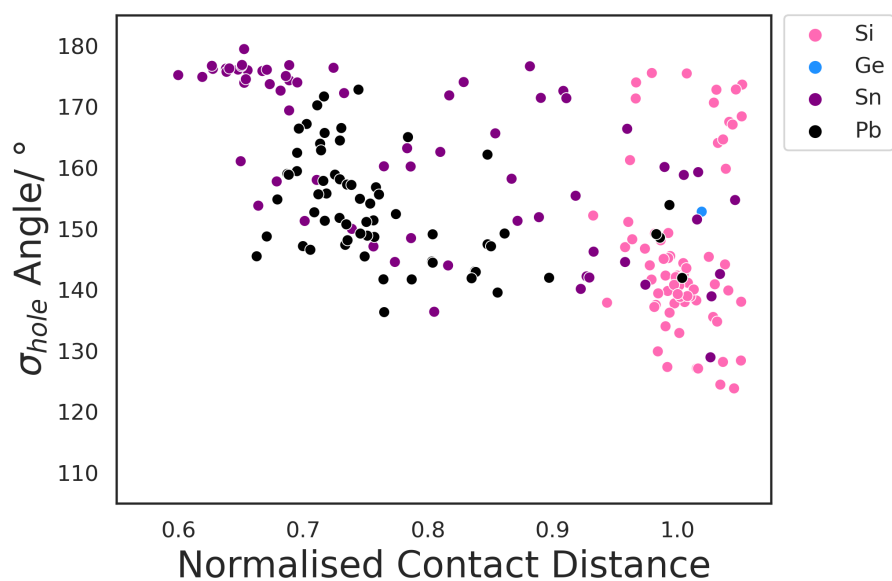


Figure 3.18: Scatter plot showing the distribution of normalised contact distance with respect to the σ -hole angle of tetrel bonding contacts to sp^2 -hybridised nitrogen found in crystal structures in the CSD. Points colour coded by donor atom type.

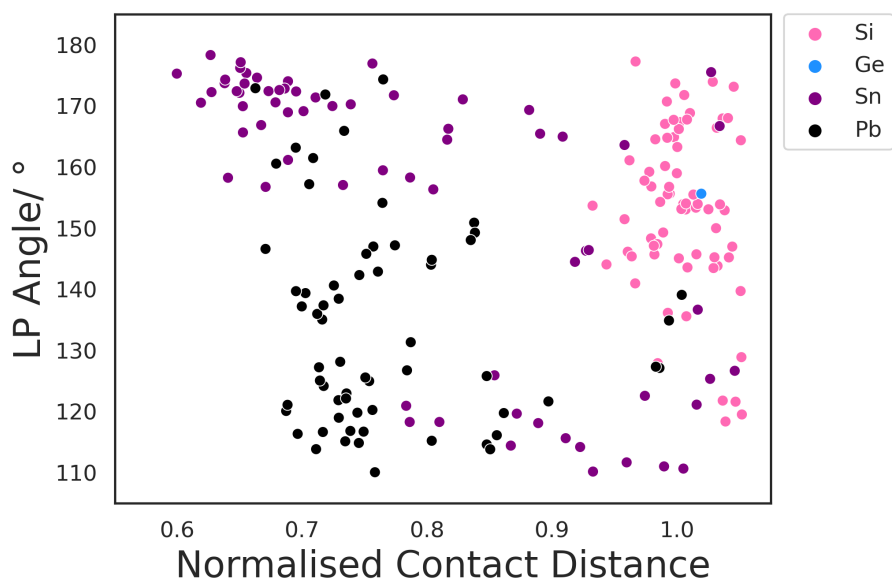


Figure 3.19: Scatter plot showing the distribution of normalised contact distance with respect to the lone-pair angle of tetrel bonding contacts to sp^2 -hybridised nitrogen found in crystal structures in the CSD. Points colour coded by donor atom type.

Table 3.15: Summary of geometric characteristics of tetrel (*Tt*) bonding contacts to sp^2 -hybridised nitrogen

Donor	Contacts	Mean Contact Distance/ Å	Mean Normalised Contact Distance	Mean σ -hole Angle/ °	Mean Lone-Pair Angle/ °
Tt	200	3.466	0.859	153	147
Pb	63	3.262	0.766	153	134
Sn	66	3.225	0.790	162	154
Ge	1	4.027	1.019	152	155
Si	70	3.868	1.005	145	153

3.4.5 Prevalence of $\sigma_{hole} \cdots sp^2$ -Nitrogen Contacts

The prevalence of crystals with the required moieties containing contacts indicative of σ -hole interactions to sp^2 -hybridised nitrogen was analysed and shown in Table 3.16. One first notices the large increase in number of structures containing both the required moieties and those containing contacts compared to σ -hole interactions to sp^3 -hybridised nitrogen (Table 3.6). For context, the prevalence of tetrel contacts to sp^2 -hybridised nitrogen is comparable to halogen bonding to sp^3 -hybridised nitrogen. This increased prevalence relative to sp^3 -hybridised nitrogen gives further indication that sp^2 -hybridised nitrogen Lewis bases are much more widely used in synthesising crystal structures, yet still don't see much use in pnictogen and tetrel bonding structures.

Table 3.16: Overall summary of the prevalence of contacts indicating σ -hole interactions to sp^2 -hybridised nitrogen compared to the number of unique crystal structures found in the CSD that contain the necessary moieties.

σ -hole Interaction Type	Structures with Moieties	Structures with Contact(s)	Prevalence
Halogen-bonding	27776	4055	14.60%
Chalcogen-bonding	13562	2853	21.04%
Pnictogen-bonding	855	33	3.86%
Tetrel-bonding	2843	124	4.36%

As shown in Table 3.17, the vast majority of the structures containing potential halogen bonding moieties are chlorine-containing species, followed by bromine containing species, and lastly iodine species. This trend differs in terms of the number of these moiety containing structures that contain a halogen bonding contact, following the trend in halogen bond donor of $Cl > I > Br$. The prevalence of total halogen bonding contacts in crystal structures containing the required halogen bond donor and acceptor moieties is 14.60%. Iodine contacts have, by far, the highest prevalence with 48.39% of moiety-containing structures displaying halogen bonding contacts. Bromine contacts, although fewer in number, have slightly higher prevalence than chlorine contacts.

Table 3.17: Summary of the prevalence of halogen (*X*) bonding contacts to sp^2 -hybridised nitrogen compared to the number of unique crystal structures found in the CSD that contain the necessary moieties

Donor	Structures with Motif	Structures with Contact(s)	Prevalence
X	27776	4055	14.60 %
I	2614	1265	48.39 %
Br	7511	863	11.49 %
Cl	18662	1959	10.50 %

Table 3.16 shows considerably fewer crystal structures containing both the required chalcogen bonding and sp^2 -hybridised nitrogen moieties than structures containing the halogen-bonding moiety. But, proportionately, there are more chalcogen bonding contacts to sp^2 -hybridised nitrogen than halogen bonding contacts, which is reflected in the higher prevalence in Table 3.18 which is around 6% higher than the analogous halogen bonding contacts. The vast majority of crystals containing chalcogen bonding moieties and those containing chalcogen bonding contacts, involve sulfur. These sulfur contacts have a prevalence of 19.24% which is nearly twice the prevalence of chlorine contacts to sp^2 -hybridised nitrogen, which is the most comparable halogen bonding motif. Tellurium contacts to sp^2 -hybridised nitrogen have a prevalence of 69.12% which is comparable to the iodine contacts to sp^2 -hybridised nitrogen. This is in contrast to selenium contacts to sp^2 -hybridised nitrogen which have over twice the prevalence compared to bromine contacts to sp^2 -hybridised nitrogen.

Table 3.18: Summary of the prevalence of chalcogen (*Ch*) bonding contacts to sp^2 -hybridised nitrogen compared to the number of unique crystal structures found in the CSD that contain the necessary moieties

Donor	Structures with Motif	Structures with Contact(s)	Prevalence
Ch	13562	2853	21.04 %
Te	136	94	69.12 %
Se	669	313	46.79 %
S	12814	2466	19.24 %

The crystal structures referred to in Table 3.16 containing the required moieties for pnictogen bonding with sp^2 -hybridised nitrogen are the least common, and also contain the fewest contacts indicating σ -hole interactions. For each of the different pnictogen-containing species, the prevalence is particularly low for phosphorus, and this affects the overall prevalence in Table 3.19. It is challenging to gain insight into pnictogen bonding interactions from this data, beyond suggesting that these structures are exotic and/or purposefully engineered proof-of-concept systems designed to show pnictogen bonding. The prevalence of tetrel bonding motifs containing sp^2 -hybridised nitrogen, shown in Table 3.20, reveals similar prevalence in comparison to pnictogen bonding structures. This is despite a greater number of both moiety-containing structures and contact-containing structures compared to the analogous pnictogen structures. The tin and lead contacts to sp^2 -hybridised nitrogen show the highest prevalence, with 16.33% and 50.00% respectively, which suggests that a number of

these structures may have been designed to form tetrel bonds. Germanium and sp^2 -hybridised nitrogen moiety containing structures are the least common, and show only one tetrel bonding contact, which suggests an absence of germanium chemistry directed towards tetrel bonding involving sp^2 -hybridised nitrogen moieties.

Table 3.19: Summary of the prevalence of pnictogen (Pn) bonding contacts to sp^2 -hybridised nitrogen compared to the number of unique crystal structures found in the CSD that contain the necessary moieties

Donor	Structures with Motif	Structures with Contact(s)	Prevalence
Pn	855	33	3.86 %
Bi	23	6	26.09 %
Sb	41	11	26.83 %
As	36	7	19.44 %
P	755	9	1.19 %

Table 3.20: Summary of the prevalence of tetrel (Tt) bonding contacts to sp^2 -hybridised nitrogen compared to the number of unique crystal structures found in the CSD that contain the necessary moieties

Donor	Structures with Motif	Structures with Contact(s)	Prevalence
Tt	2843	124	4.36 %
Pb	60	30	50.00 %
Sn	300	49	16.33 %
Ge	53	1	1.89 %
Si	2452	44	1.79 %

3.4.6 Notable sp^2 -Nitrogen Moieties

When inspecting this plethora of σ -hole interactions to sp^2 -hybridised nitrogen, some interesting moieties arose that provide some explanation to the distributions seen in some of the plots of angle measurements against normalised contact distances. A somewhat obvious finding was that the vast majority of sp^2 -hybridised nitrogen Lewis bases are in cyclic systems, which is not surprising given the hybridisation lends itself to participation of nitrogen in π -systems.

An interesting comparison can be seen with halogen and chalcogen bonding motifs, see Section 3.4.6, where halogen bonding contacts to sp^2 -hybridised nitrogen have moiety distributions biased towards 6-membered rings, whereas analogous chalcogen bonding contacts have moiety distributions skewed towards 5-membered rings. This can likely be attributed to aromaticity, for example thiazole, a five membered heterocycle containing sulfur and nitrogen, obeying Huckel's rule of 6π electrons.

Table 3.21: Numbers of halogen (*X*) and chalcogen (*Ch*) bonding contacts to various cycle sizes of cyclic sp^2 -hybridised nitrogen.

N-Heterocycle Size	X-bonding Contacts	Ch-bonding Contacts
3	2	1
4	7	1
5	1704	3782
6	3487	1064
7	27	38
8	35	34
9	0	0
10	6	1
11	0	0
12	11	1
Other	281	197

Distributions of normalised contact distances and σ -hole angles for halogen bonding contacts to five- and six-membered sp^2 -hybridised nitrogen heterocycles in Figure 3.20. These plots correspond to 1704 contacts to 5-membered nitrogen heterocycles, and 3487 contacts to 6-membered nitrogen heterocycles. The contacts involving 6-membered nitrogen Lewis bases are notably closer and show σ -hole angles closer to linearity. Similar trends are seen in the distributions of normalised contact distances and σ -hole angles for chalcogen bonding contacts to five- and six-membered sp^2 -hybridised nitrogen heterocycles, shown in Figure 3.21. The main difference is that for chalcogen bonding contacts there are more contacts involving 5-membered nitrogen heterocycles (3782) over 6-membered cycles (1064). This suggests an effect of the ring's nature on the Lewis basicity of the nitrogen, this could be due to a number of factors such as ring strain or other substituents in the heterocycle^{138 139}. Another potential explanation for the differing cycle size preferences between halogen and chalcogen bonding is the likelihood of molecular crystals where donor and acceptor moieties are part of the same cycle. As halogen-substituted aromatic groups are a popular halogen bond donor moiety (discussed further in Section 3.6) and chalcazoles are a popular chalcogen bonding moiety, one might expect a significant number of these self-interacting crystals.

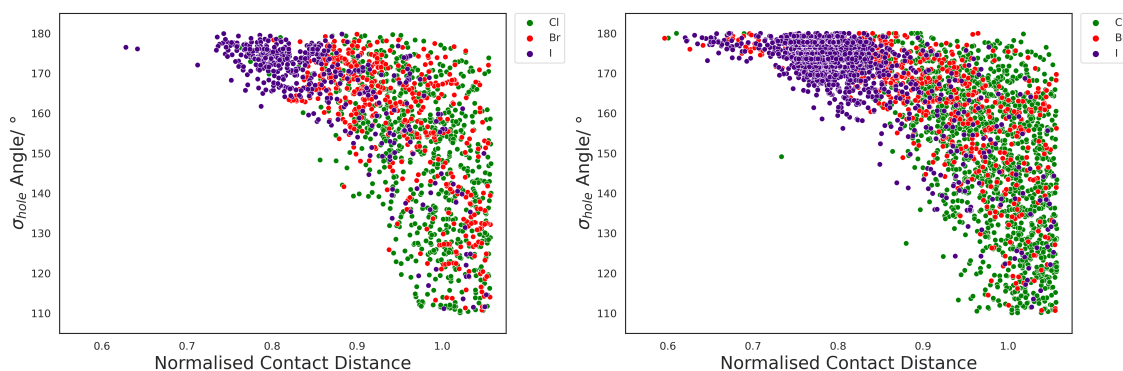


Figure 3.20: Scatter plot showing the distribution of normalised contact distance with respect to the σ -hole angle of halogen bonding contacts to 5-membered (left) and 6-membered (right) sp^2 -hybridised nitrogen heterocycles found in crystal structures in the CSD. Points colour coded by donor atom type.

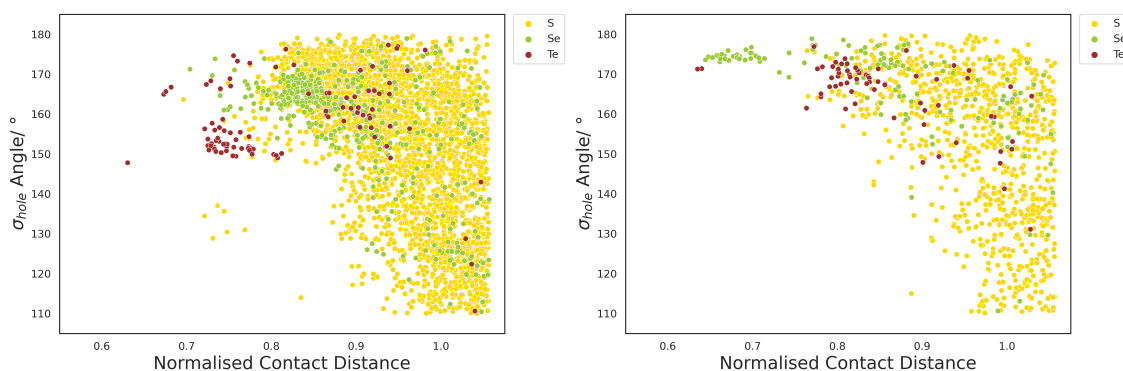


Figure 3.21: Scatter plot showing the distribution of normalised contact distance with respect to the σ -hole angle of chalcogen bonding contacts to 5-membered (left) and 6-membered (right) sp^2 -hybridised nitrogen heterocycles found in crystal structures in the CSD. Points colour coded by donor atom type.

On the topic of 5-membered heterocycles, off-linear σ -hole angles seen in chalcogen bonding contacts to sp^2 -hybridised nitrogen were noted in Section 3.4.2, see Figure 3.14. These groups correspond to the chalcazole moieties mentioned above, which are 5-membered heterocycles where different nitrogen atoms acts as both the electron-withdrawing group, creating the σ -hole, and the Lewis basic site. This leads to crystal structures containing these moieties forming inversion dimers that result in very close contacts, owing to multiple chalcogen bonds, but interaction geometries perturbed from linearity. An example is illustrated in Figure 3.22, the crystal structure of 1,2,5-telluradiazole-3,4-dicarbonitrile (CSD refcode **AREGEK**) which forms ribbon chains of inversion dimers¹⁴⁰.

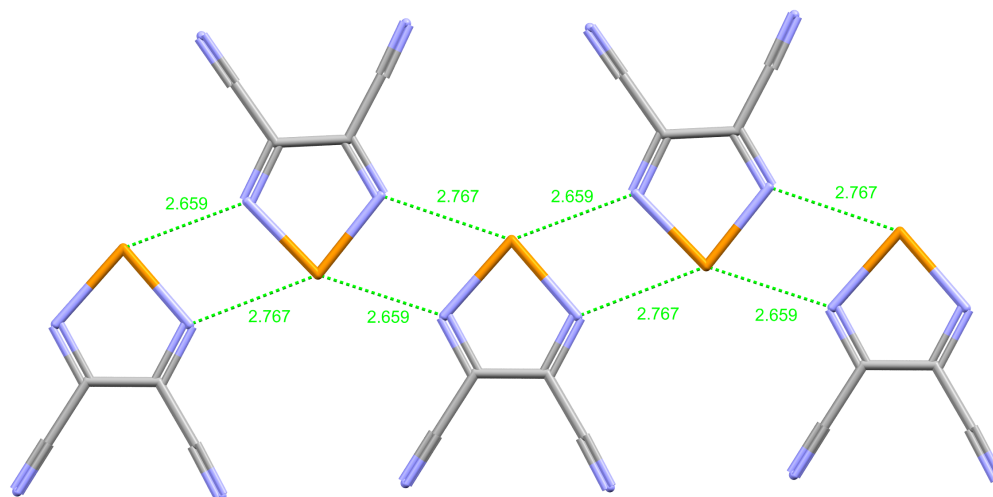


Figure 3.22: Crystal structure of 1,2,5-telluradiazole-3,4-dicarbonitrile, with the molecule showing both σ -hole donor and acceptor behaviour, with the chalcogen bonds represented as dotted lines. Te $\cdots$ N contact distances given in Ångström. CSD refcode: **AREGEK**.

3.5 *sp*-Hybridised Nitrogen

Lastly, we consider the σ -hole interactions to *sp*-hybridised nitrogen Lewis bases. The moieties presented in Section 3.5 comprise of primarily cyano groups, with a very small number of azide and diazonium groups present. Cyano and other *sp*-hybridised groups are linear in nature, and consequentially have very little steric bulk around the Lewis basic site. The lone pair of the nitrogen in these moieties are in an orbital with the most *s*-character, thus are the most contracted, meaning these acceptors are likely to be the least Lewis basic. There is also a π -system that the nitrogen participates in, which will offer an alternative acceptor site and affect the Lewis basicity of the lone-pair.

Table 3.22: Overall geometric summary of the contacts indicative of σ -hole interactions to *sp*-hybridised nitrogen

σ -hole Interaction Type	Contacts (Cyano)	Contacts (Azide/Diazonium)	Total Contacts	Mean Normalised Contact Distance	Mean σ -hole Angle/ °	Mean Lone-Pair Angle/ °
Halogen-bonding	1443	128	1571	0.930	158	137
Chalcogen-bonding	2124	31	2155	0.945	156	137
Pnictogen-bonding	38	3	41	0.871	162	140
Tetrel-bonding	97	9	106	0.883	167	141

3.5.1 $X \cdots sp$ -Nitrogen Contacts

Section 3.5 shows there are around twice as many halogen bonding contacts to *sp*-hybridised nitrogen than compared to those involving *sp*³-hybridised nitrogen (Table 3.1), but much fewer than the contacts involving *sp*²-hybridised nitrogen (Table 3.11). When subdivided by halogen species, as in Table 3.23, chlorine contacts are the most numerous, followed by bromine contacts, and iodine contacts are the fewest in number.

Out of the halogen bonding contacts to different hybridisation of nitrogen, those involving *sp*-hybridised nitrogen have the broadest "wedge" shaped distribution of normalised contact distances against σ -hole angle, shown in Figure 3.23. A clear relationship between the normalised contact distance and the σ -hole angle with respect to the different halogen bond donors can be seen in Table 3.23. This follows the trend in polarisability of the halogen, with normalised contact distance following the trend $I < Br < Cl$, and the σ -hole angle following the trend $I > Br > Cl$. The variation in average σ -hole angle is very small, with the average σ -hole angle of iodine contacts being only 6° larger than those involving chlorine. There is very little directionality seen for any halogen bonding contacts to *sp*-hybridised nitrogen with respect to the lone-pair angle, shown in Figure 3.24. Bromine and chlorine contacts to *sp*-hybridised nitrogen have an average lone-pair angle of 136° (19° standard deviation) and iodine has an average lone-pair angle of 140°(22° standard deviation).

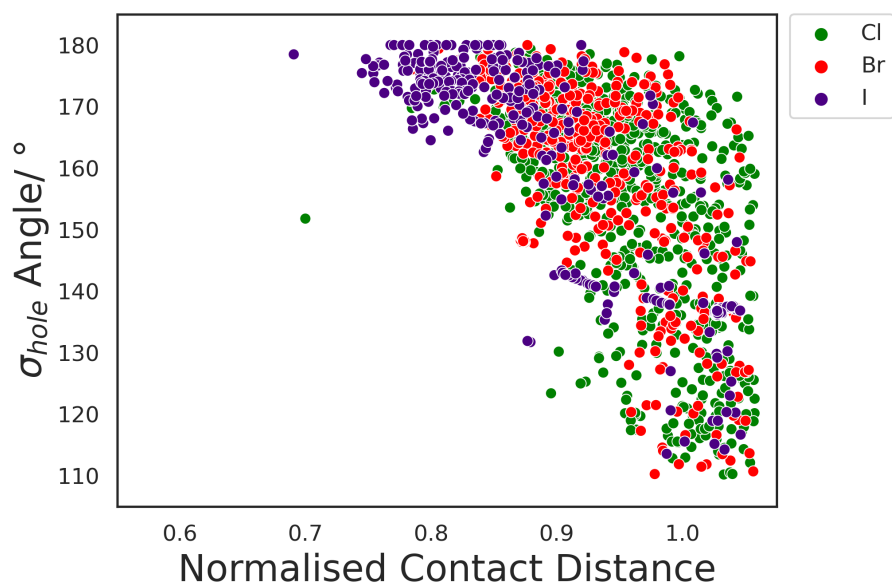


Figure 3.23: Scatter plot showing the distribution of normalised contact distance with respect to the σ -hole angle of halogen bonding contacts to *sp*-hybridised nitrogen found in crystal structures in the CSD. Points colour coded by donor atom type.

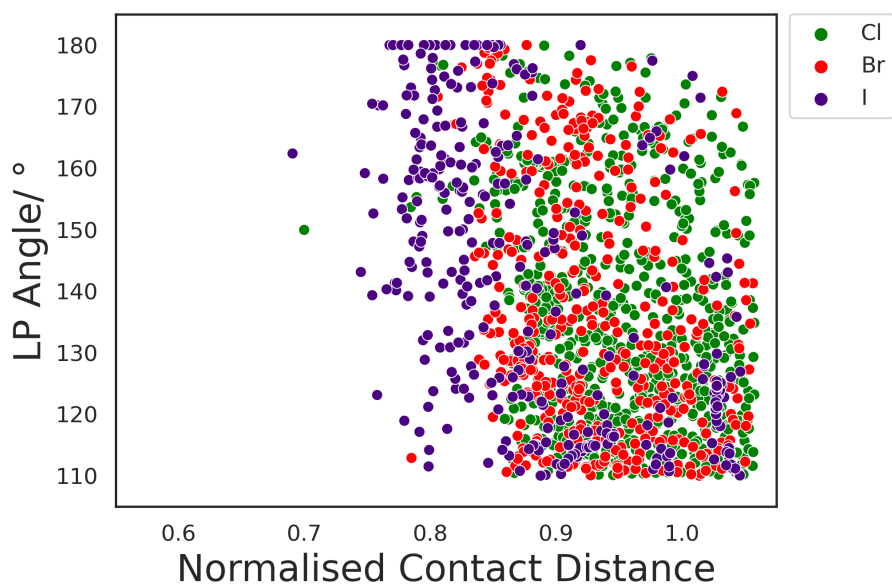


Figure 3.24: Scatter plot showing the distribution of normalised contact distance with respect to the lone-pair angle of halogen bonding contacts to *sp*-hybridised nitrogen found in crystal structures in the CSD. Points colour coded by donor atom type.

Table 3.23: Summary of geometric characteristics of halogen bonding contacts to *sp*-hybridised nitrogen

Donor	Contacts	Mean Contact Distance/ Å	Mean Normalised Contact Distance	Mean σ -hole Angle/ °	Mean Lone-Pair Angle/ °
X	1571	3.288	0.930	158	137
I	340	3.271	0.884	161	140
Br	455	3.270	0.929	159	136
Cl	776	3.305	0.950	155	136

3.5.2 Ch ··· *sp*-Nitrogen Contacts

The chalcogen bonding contacts to *sp*-hybridised nitrogen shown in Section 3.5 are around twice as numerous compared to halogen bonding contacts to *sp*-hybridised nitrogen.

When the contacts are subdivided by the identity of the chalcogen atom, as in Table 3.24, it can be seen that sulfur contacts are by far the most numerous, then a notable number of selenium contacts, lastly a small number of tellurium contacts. The distributions of σ -hole angle against normalised contact distance in Figure 3.25 show that the "wedge" shaped distribution has a greater spread in σ -hole angle than compared to the analogous halogen bonding contacts (see Figure 3.23).

The average geometric measurements in Table 3.24 show these contacts are on average longer and have σ -hole angles that deviate further from linearity than the analogous halogen bonding contacts to *sp*-hybridised nitrogen (see Table 3.23). The average normalised contact distances in Table 3.24 reflect the polarisability of the chalcogen bond donor, following the trend $\text{Te} < \text{Se} < \text{S}$. Selenium contacts have the σ -hole angles nearest linearity, with an average σ -hole angle of 163° which is larger on average than any of the average σ -hole angles involving halogen bonding contacts to *sp*-hybridised nitrogen (see Table 3.23).

The distributions of lone-pair angle against normalised contact distance are shown in Figure 3.26, and shows little to no overall directionality. Where the average lone-pair angles, listed in Table 3.24, show that although little directionality can be seen with respect to the location of the lone-pair the average lone-pair angle gets larger for the more polarisable chalcogens, following the trend $\text{Te} > \text{Se} > \text{S}$.

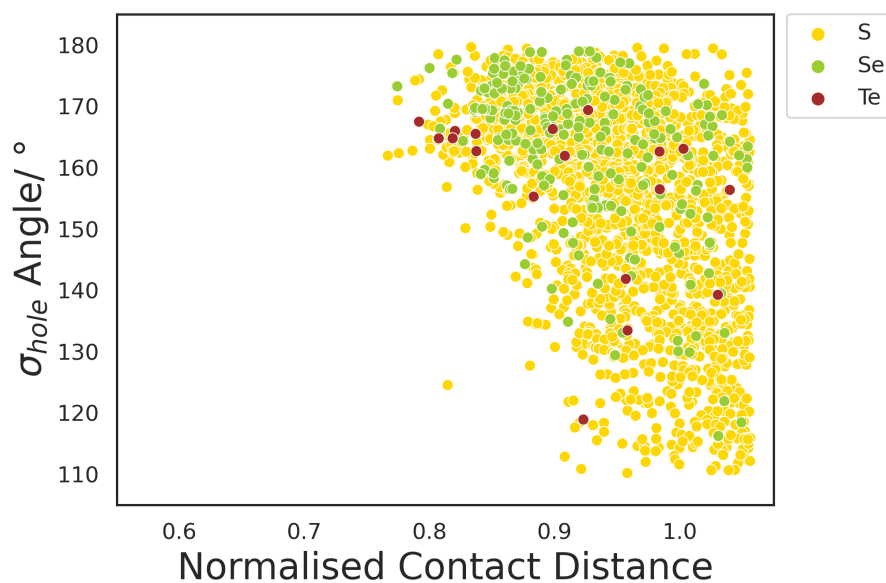


Figure 3.25: Scatter plot showing the distribution of normalised contact distance with respect to the σ -hole angle of chalcogen bonding contacts to *sp*-hybridised nitrogen found in crystal structures in the CSD. Points colour coded by donor atom type.

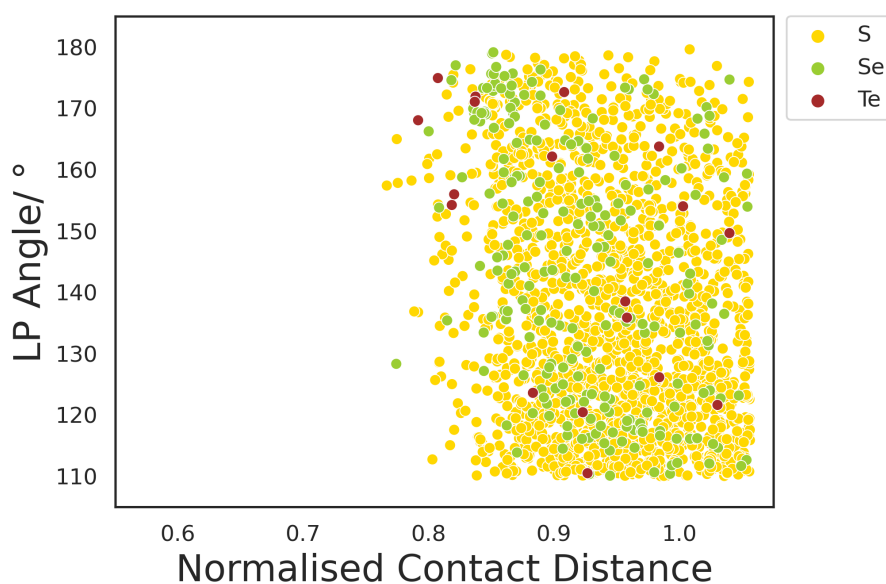


Figure 3.26: Scatter plot showing the distribution of normalised contact distance with respect to the lone-pair angle of chalcogen bonding contacts to *sp*-hybridised nitrogen found in crystal structures in the CSD. Points colour coded by donor atom type.

Table 3.24: Summary of geometric characteristics of chalcogen bonding contacts to *sp*-hybridised nitrogen

Donor	Contacts	Mean Contact Distance/ Å	Mean Normalised Contact Distance	Mean σ -hole Angle/ °	Mean Lone-Pair Angle/ °
Ch	2155	3.347	0.945	156	137
Te	18	3.327	0.912	156	148
Se	238	3.203	0.920	163	144
S	1899	3.365	0.948	155	136

3.5.3 *Pn* ... *sp*-Nitrogen Contacts

The pnictogen bonding contacts to *sp*-hybridised nitrogen, see Section 3.5, although few in number, show notable difference in geometric characteristics with halogen or chalcogen bonding contacts to *sp*-hybridised nitrogen. Once again these pnictogen contacts are few in number, meaning that one should account for likely selection bias when examining any trends posited here, as they are likely not representative of the whole scope of pnictogen bonding, just those found in this search.

The distributions of σ -hole angle against normalised contact distance, shown in Figure 3.27, show that the majority of the contacts, regardless of the pnictogen bond donor, have σ -hole angles greater than 145°. Halogen and chalcogen bonding contacts to *sp*-hybridised nitrogen show a very small proportion of those shorter than 0.8 normalised contact distance (4.5% & 0.4%, respectively) (see Figure 3.23 & Figure 3.25). In contrast, pnictogen bonding contacts to *sp*-hybridised nitrogen show a notable proportion of contacts (22%) that have normalised contact distances shorter than 0.8. The trend in average normalised contact distance, shown in Table 3.25, somewhat reflects the polarisability of the pnictogen atom, yet bismuth contacts are longer than antimony contacts on average, so the trend follows Sb < Bi < As < P.

The distributions of lone-pair angle against normalised contact distance in Figure 3.28 show some bias towards lone-pair angles closer to linearity as the normalised contact distance gets shorter. The average values detailed in Table 3.25 shows no notable trend with respect to the lone-pair angle and the identity of the pnictogen. Likewise, the average σ -hole angle varies little between the different pnictogen species, with arsenic contacts having the largest average σ -hole angle of 166°.

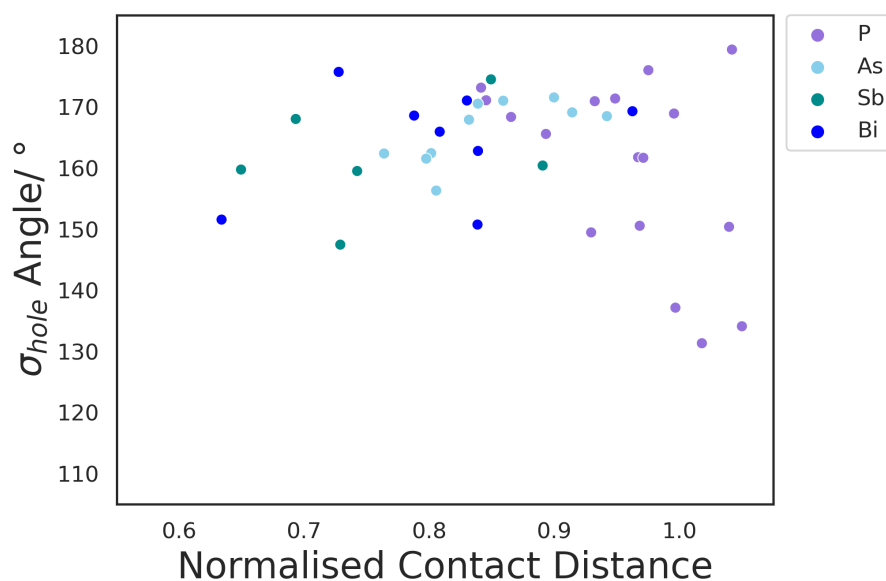


Figure 3.27: Scatter plot showing the distribution of normalised contact distance with respect to the σ -hole angle of pnictogen bonding contacts to *sp*-hybridised nitrogen found in crystal structures in the CSD. Points colour coded by donor atom type.

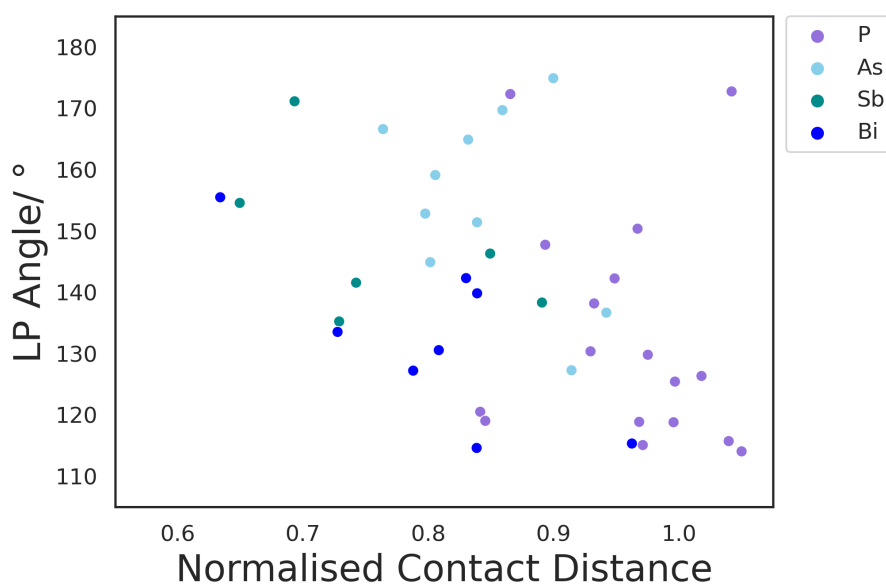


Figure 3.28: Scatter plot showing the distribution of normalised contact distance with respect to the lone-pair angle of pnictogen bonding contacts to *sp*-hybridised nitrogen found in crystal structures in the CSD. Points colour coded by donor atom type.

Table 3.25: Summary of geometric characteristics of pnictogen bonding contacts to *sp*-hybridised nitrogen

Donor	Contacts	Mean Contact Distance/ Å	Mean Normalised Contact Distance	Mean σ -hole Angle/ °	Mean Lone-Pair Angle/ °
Pn	41	3.260	0.871	162	140
Bi	8	3.374	0.803	164	132
Sb	6	3.134	0.759	161	147
As	10	2.993	0.845	166	154
P	17	3.409	0.958	160	132

3.5.4 *Tt* ... *sp*-Nitrogen Contacts

The tetrel bonding contacts to *sp*-hybridised nitrogen are somewhat similar to the pnictogen bonding contacts to *sp*-hybridised nitrogen presented in Section 3.5.3. In the distribution of σ -hole angle against the normalised contact distance, shown in Figure 3.29, there are a notable proportion of contacts with a normalised contact distance shorter than 0.8 (36%). The numbers of contacts in Section 3.5 shows that tetrel bonding contacts to *sp*-hybridised nitrogen over twice as numerous as the corresponding pnictogen bonding contacts, making geometric trends more reliable.

Unlike the pnictogen bonding contacts to *sp*-hybridised nitrogen, the distributions of σ -hole angle against normalised contact distance, shown in Figure 3.29, has a much clearer bias towards σ -hole angles near linearity. The majority of the near linear contacts, with σ -hole angles greater than 170° mainly involving tin and silicon tetrel bond donors. The silicon contacts have an average σ -hole angle shown in Table 3.26 of 171° but have a bimodal distribution of normalised contact distances. The first population have normalised contact distances around 0.8 and the population have normalised contact distances greater than 1.0. There are very few lead and germanium contacts, yet the latter makes up greatest number of germanium contacts out of the three hybridisations of nitrogen. Comparable to the other contacts indicating a σ -hole interaction to *sp*-hybridised nitrogen, there is little directionality with respect to the lone-pair, as shown in the distribution of lone-pair angles against normalised contact distance in Figure 3.30.

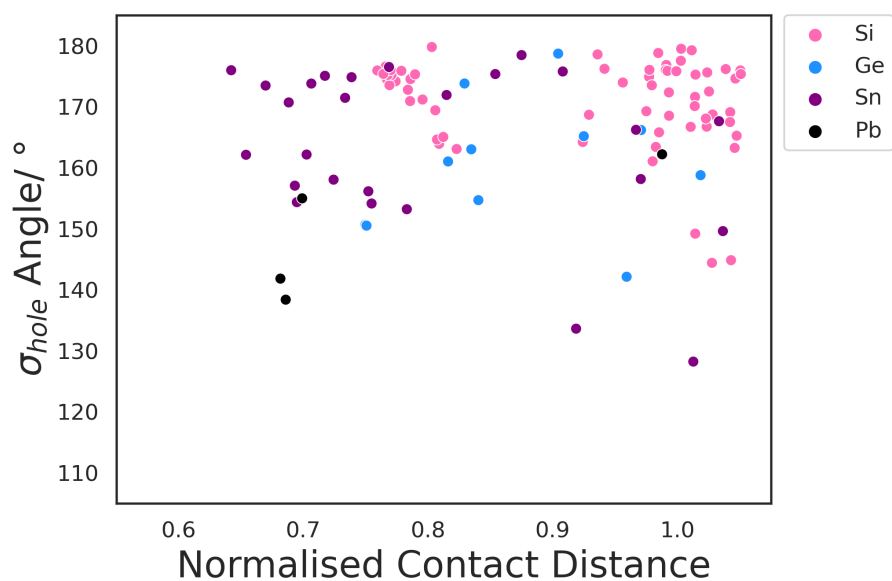


Figure 3.29: Scatter plot showing the distribution of normalised contact distance with respect to the σ -hole angle of tetrel bonding contacts to sp -hybridised nitrogen found in crystal structures in the CSD. Points colour coded by donor atom type.

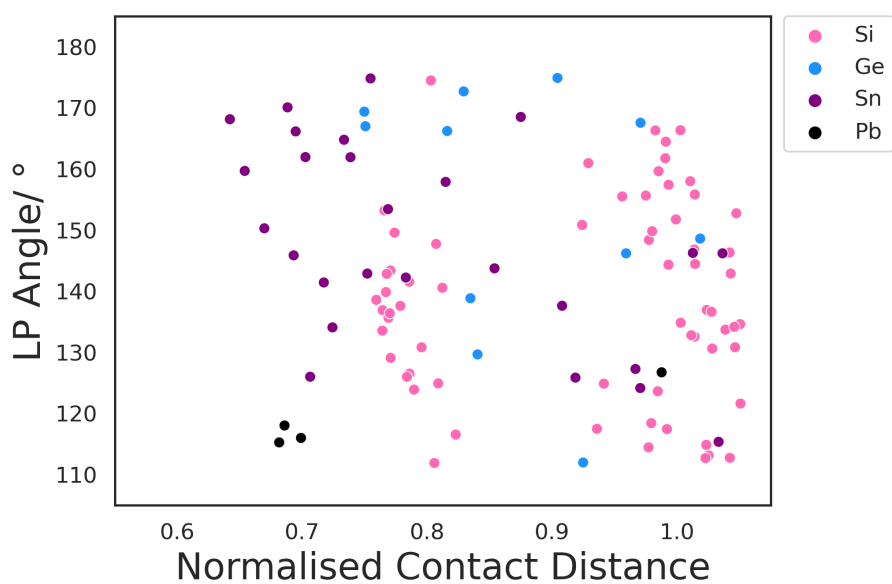


Figure 3.30: Scatter plot showing the distribution of normalised contact distance with respect to the lone-pair angle of tetrel bonding contacts to sp -hybridised nitrogen found in crystal structures in the CSD. Points colour coded by donor atom type.

Table 3.26: Summary of geometric characteristics of tetrel bonding contacts to *sp*-hybridised nitrogen

Donor	Contacts	Mean Contact Distance/ Å	Mean Normalised Contact Distance	Mean σ -hole Angle/ °	Mean Lone-Pair Angle/ °
Tt	106	3.466	0.883	167	141
Pb	4	3.252	0.763	149	119
Sn	26	3.266	0.800	163	148
Ge	11	3.447	0.873	160	153
Si	65	3.563	0.925	171	138

3.5.5 Prevalence of $\sigma_{hole} \cdots sp$ -Nitrogen Contacts

The prevalence of crystal structures with the required donor and acceptor moieties forming σ -hole interactions to *sp*-hybridised nitrogen was examined in Table 3.27.

In comparison to the prevalence of σ -hole interactions to *sp*²-hybridised nitrogen (see Table 3.16) there are far fewer moiety containing structures and contacts indicating σ -hole interactions. In comparison to σ -hole interactions to *sp*³-hybridised nitrogen (see Table 3.6), there are similar numbers of crystal structures containing the necessary moieties, but σ -hole contacts to *sp*-hybridised nitrogen have much greater prevalence than the contacts to *sp*³-hybridised nitrogen. The prevalence of pnictogen and tetrel bonding contacts to *sp*-hybridised nitrogen in Table 3.27 is the highest out of the pnictogen and tetrel bonding contacts to the other hybridisations of nitrogen.

Table 3.27: Overall summary of the prevalence of contacts indicating σ -hole interactions to *sp*-hybridised nitrogen compared to the number of unique crystal structures found in the CSD that contain the necessary moieties.

σ -hole Interaction Type	Structures with Moieties	Structures with Contact(s)	Prevalence
Halogen-bonding	5519	1032	18.70%
Chalcogen-bonding	3968	1157	29.16%
Pnictogen-bonding	236	29	12.29%
Tetrel-bonding	736	74	10.03%

The halogen bonding to *sp*-hybridised nitrogen motifs shown in Table 3.28 are the most numerous, with a prevalence of 18.70%. The prevalence of each halogen species increases with the polarisability of the halogen halogen bond donor, with the highest prevalence being that of iodine at 39.00%. The majority of both the total crystal structures and those containing halogen bonding contacts involve chlorine, despite it having the lowest percentage prevalence.

Chalcogen bonding to *sp*-hybridised nitrogen has a greater prevalence than for halogen bonding to *sp*-hybridised nitrogen, shown in Table 3.29, with an overall prevalence of 29.16%. Selenium and tellurium contacts have comparable prevalence, with selenium contacts having greater prevalence by around 5%. The number of tellurium and *sp*-hybridised nitrogen moiety containing structures is very small, with only 34 structures containing the required moieties. The vast majority of chalcogen bonding contacts to *sp*-hybridised nitrogen and structures containing the required moieties involve sulfur.

Sulfur chalcogen bonds to *sp*-hybridised nitrogen have a notable prevalence of 27.77%, which is nearly twice as prevalent as the analogous chlorine halogen bonds to *sp*-hybridised nitrogen that shows a comparable number of crystal structures containing the required moieties. This high prevalence of sulfur contacts within a large number of crystal structures suggests that this is a synthetically accessible and well-utilised chalcogen bonding motif.

Table 3.28: Summary of the prevalence of halogen bonding contacts to *sp*-hybridised nitrogen compared to the number of unique crystal structures found in the CSD that contain the necessary moieties

Donor	Structures with Motif	Structures with Contact(s)	Prevalence
X	5519	1032	18.70 %
I	482	188	39.00 %
Br	1808	326	18.03 %
Cl	3378	521	15.42 %

Table 3.29: Summary of the prevalence of halogen bonding contacts to *sp*-hybridised nitrogen compared to the number of unique crystal structures found in the CSD that contain the necessary moieties

Donor	Structures with Motif	Structures with Contact(s)	Prevalence
Ch	3968	1157	29.16 %
Te	34	13	38.24 %
Se	294	126	42.86 %
S	3691	1025	27.77 %

Pnictogen and tetrel bonding contacts to *sp*-hybridised nitrogen, shown in Table 3.30 and Table 3.31, respectively, have the highest prevalence out of pnictogen and tetrel contacts to nitrogen. Pnictogen bonding shows an overall prevalence of 12.29% and tetrel bonding shows an overall 10.03% prevalence. Phosphorus accounts for almost half of the pnictogen bonding contacts, thus it only shows a prevalence of 6.86%. Likewise, silicon makes up the majority of the tetrel bonding contacts, with a prevalence of 6.88%. A standout set of contacts are tin contacts to *sp*-hybridised nitrogen, with the contacts being more numerous than the phosphorus pnictogen bonding contacts, but the tin contacts have a substantially higher prevalence of 47.37%. High prevalences among a small number of crystal structures reinforces that these structures likely proof-of-concept structures designed to showcase pnictogen and tetrel bonding. Out of the pnictogen and tetrel bond donor containing crystal structures featuring nitrogen, those containing *sp*-hybridised nitrogen would seem to be the most suitable to form σ -hole interactions. This is likely due to the lack of steric bulk around a *sp*-hybridised nitrogen species, which makes them ideal Lewis bases to approach and interact closely with a sterically crowded σ -hole on a pnictogen or a tetrel.

Table 3.30: Summary of the prevalence of pnictogen bonding contacts to *sp*-hybridised nitrogen compared to the number of unique crystal structures found in the CSD that contain the necessary moieties

Donor	Structures with Motif	Structures with Contact(s)	Prevalence
Pn	236	29	12.29 %
Bi	8	5	62.50 %
Sb	11	5	45.45 %
As	17	5	29.41 %
P	204	14	6.86 %

Table 3.31: Summary of the prevalence of tetrel bonding contacts to *sp*-hybridised nitrogen compared to the number of unique crystal structures found in the CSD that contain the necessary moieties

Donor	Structures with Motif	Structures with Contact(s)	Prevalence
Tt	738	74	10.03 %
Pb	15	4	26.67 %
Sn	38	18	47.37 %
Ge	40	7	17.50 %
Si	654	45	6.88 %

3.5.6 Notable *sp*-Nitrogen Moieties

As mentioned at the beginning of Section 3.5, the vast majority of the *sp*-hybridised nitrogen σ -hole acceptors in the CSD are cyano groups, but some interesting trends can be observed when inspecting the element the cyano group is bonded to. There are a small number of halogen bonding contacts to *sp*-hybridised nitrogen, where the cyano group is coordinated to a metal centre, with the distribution of the σ -hole angle against normalised contact distance for these contacts shown in Figure 3.31. These contacts are among the shortest and have σ -hole angles approaching linearity, suggesting that the metal is making the cyano group a more effective σ -hole acceptor. This could be explained by several possible effects, such as the metal centre being anionic, or it could be due to π back-donation from the metal causing more electron density to be localised onto the cyano nitrogen¹⁴¹. This same effect can be seen in chalcogen bonding contacts, shown in Figure 3.32, but the contacts are longer and have σ -hole angles that deviate markedly from linearity. These metal-coordinated cyanides are not observed in close contact with pnictogen or tetrel bond donors however, suggesting a potential new area of study.

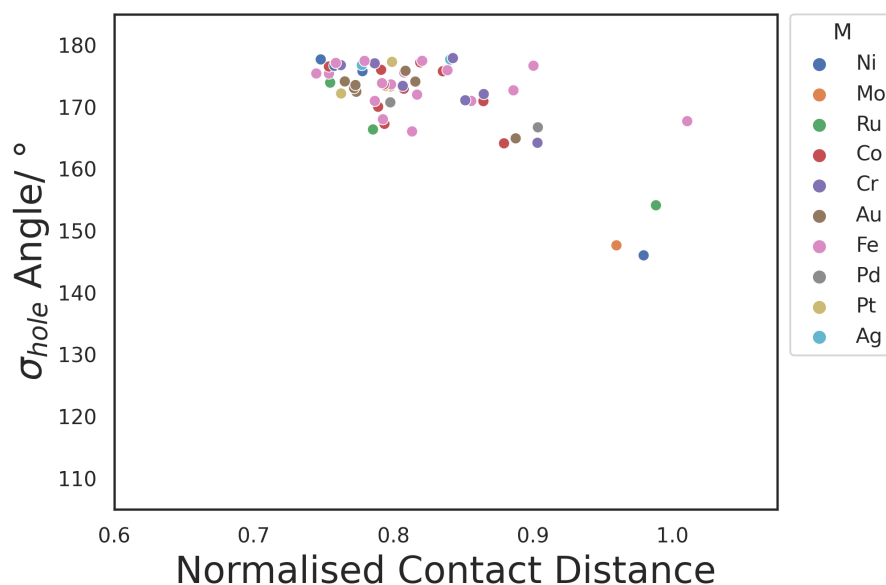


Figure 3.31: Scatter plot showing the distribution of normalised contact distance with respect to the σ -hole angle of halogen bonding contacts to cyano groups bound to metal centres found in crystal structures in the CSD. Points colour coded by metal centre.

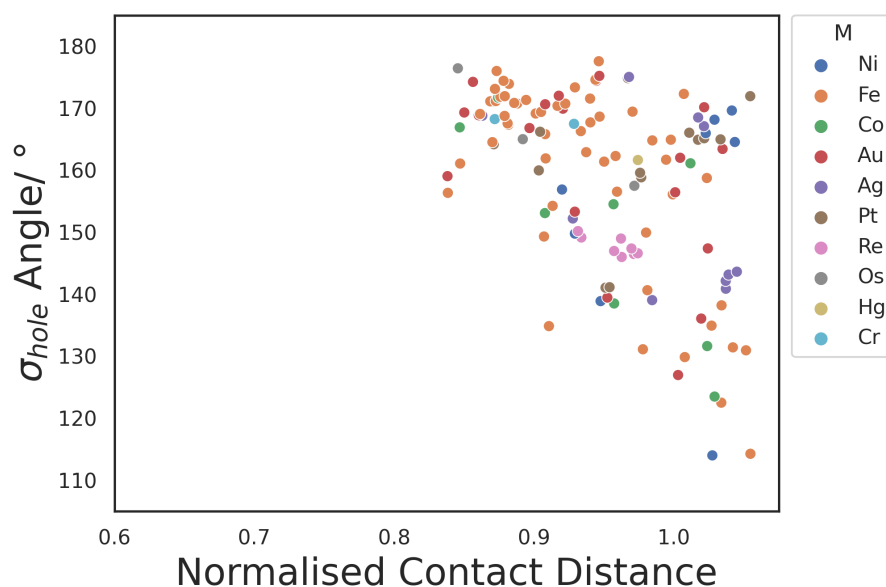


Figure 3.32: Scatter plot showing the distribution of normalised contact distance with respect to the σ -hole angle of chalcogen bonding contacts to cyano groups bound to metal centres found in crystal structures in the CSD. Points colour coded by metal centre.

3.6 Trends in the σ -hole Donor Atom

The trends seen in geometric characteristics of σ -hole interactions to the three hybridisations of nitrogen have varied with respect to the identity of the σ -hole donor. Literature suggests that more polarisable σ -hole donors more readily engage in stronger σ -hole interactions, due to electron-withdrawing groups being able to cause a greater charge anisotropy around

the σ -hole donor atom^{142,6}. This greater polarisability is often associated with larger atomic size, which will have an effect on the donor atom's tendency to get into close proximity to a Lewis basic site.

Among the halogen bond donors, with the average geometric observables with respect to the three nitrogen hybridisations listed in Table 3.2, Table 3.12, and Table 3.23; the shortness of the normalised contact distances reflect the polarisability of the donors with iodine having the shortest contacts, bromine longer contacts, and chlorine having the longest contacts. The closeness of the contacts is clearly linked to σ -hole angles approaching linearity, with the shorter iodine contacts having average σ -hole angles between 161-171°, the bromine contacts having average σ -hole angles between 157-162°, and chlorine contacts having σ -hole angles between 144-155°. There is a relationship seen between the σ -hole donor and the average lone-pair angles, but it is not as pronounced as the relationship with respect to the σ -hole angle. This is mainly seen in the halogen bonding contacts to sp^2 - and sp^3 -hybridised nitrogen, with the sp -hybridised nitrogen contacts showing the same average lone-pair angles for bromine and chlorine contacts, with only iodine contacts showing larger lone-pair angles.

An interesting comparison can be made to the so-called R_F values of Robin Taylor, which is the ratio of an interatomic contacts observed frequency of occurrence compared to the frequency expected at random (eg. determined solely by atomic exposed surface area)¹⁴³. This values give an indicator of the likelihood of a noncovalent interaction having a significant impact on crystal packing, with an R_F value of 1 meaning the interactions happen by chance, $R_F > 1$ indicating a significant attractive interaction, and $R_F < 1$ being repulsive. The R_F value of the C-I $\cdots$ N interaction is 6.5, which is comparable to typical strong hydrogen bonding motifs, the R_F for C-Br $\cdots$ N is much lower at 1.8, and the R_F for C-Cl $\cdots$ N is 0.9 which is slightly repulsive. This supports the findings in this chapter that iodine and bromine contacts with different hybridisations of nitrogen are significantly strong halogen bonds, whereas chlorine is most likely not forming significant halogen bonds. The only Cl $\cdots$ N motif with an R_F value > 1 is N₂C-Cl $\cdots$ N with an R_F of 1.5, this donor moiety corresponds to chlorine on sp^2 carbon bound to two nitrogen atoms, eg. 2-position of imidazole. This shows that there are circumstances where Cl $\cdots$ N motifs can be significantly strong halogen bonds, but require donor electron-withdrawing groups containing nitrogen, eliciting a greater charge anisotropy. Although no larger chalcogens, pnictogens, or any tetrrels are referred to in Taylor's work, there are a number of R_F values reported for sulfur and phosphorus. The only interactions of sulfur and phosphorus with nitrogen involve sulfur in an electron withdrawing environment such as thiazole or isothiazole, and three-coordinate phosphorus. The R_F value for S $\cdots$ N is 1.2 which is still very low, and the R_F value of P $\cdots$ N is 0.4 so would be considered a repulsive interaction. These low R_F values for the smaller σ -hole donors and their apparent weak σ -hole interactions are reflected in the distributions of normalised contact distance against both σ -hole angle and lone-pair angle within this chapter, as they show little to no directionality and very few close contacts.

In terms of the numbers of contacts, there are always more chlorine contacts than bromine contacts, yet the number of iodine contacts differs with the hybridisation of the nitrogen, with there being fewer contacts than bromine contacts to sp -hybridised nitrogen, yet more iodine contacts than bromine or chlorine contacts to sp^2 - and sp^3 -hybridised nitrogen. This

suggests that there is widespread utilisation of iodine motifs with sp^2 - and sp^3 -hybridised nitrogen to exploit halogen bonding in crystal structures^{144–146}. This is supported by the prevalence of iodine contacts to nitrogen being markedly greater than the prevalences of bromine and chlorine contacts, with bromine contacts being only slightly more prevalent than chlorine contacts.

Among the chalcogen bond donors, with the average geometric observables with respect to the three nitrogen hybridisations listed in Table 3.3, Table 3.13, and Table 3.24; there are significant differences in the chalcogen bonding contact's geometric characteristics than seen with the halogen bonding contacts. While much less commonly occurring than sulfur contacts, there are still a notable amount of selenium and tellurium contacts involving sp - and sp^2 -hybridised nitrogen, but very few involving sp^3 -hybridised nitrogen. The closeness of the normalised contact distances follows the trend in polarisability of the chalcogen, with tellurium contacts being the shortest, closely followed by selenium contacts, and then sulfur contacts being considerably longer on average. Selenium and tellurium contacts to nitrogen have very similar average σ -hole angles, around 155–160°, with selenium contacts being having slightly larger σ -hole angles on average, and sulfur contacts average σ -hole angle deviating further from linearity with a range between 147–155°. The average lone-pair angles for each nitrogen hybridisation is linked with increasing polarisability of the chalcogen atom, with tellurium contacts generally having largest lone-pair angles on average. The similarity between selenium and tellurium contacts involving nitrogen is also seen in the prevalences, detailed in Table 3.8, Table 3.18, and Table 3.29, of these interactions occurring in structures containing the required moieties. The prevalences of tellurium and selenium contacts with sp^3 -hybridised and sp -hybridised nitrogen are fairly consistent, with tellurium contacts only being substantially more prevalent with sp^2 -hybridised nitrogen. It is important to note that tellurium chemistry is generally much less common than selenium chemistry, shown in there being far fewer crystal structures containing tellurium bonding moieties.

Pnictogen and tetrel bonding contacts to nitrogen are much rarer than the analogous halogen and chalcogen bonding contacts. This makes the averages of the geometric observables of these contacts less reliable to compare. The most numerous pnictogen and tetrel bonding contacts for a single element involve phosphorus and silicon, the lightest of the pnictogen and tetrel bond donors. Pnictogens and tetrrels of the fifth period, antimony and tin respectively, generally have much shorter normalised contact distances and a greater tendency towards larger σ -hole angles.

3.6.1 Effect of Donor Electron-Withdrawing Group

So as not to neglect the nature of the σ -hole donor in this analysis, the substituents covalently bonded to the σ -hole donors that cause the charge anisotropy needed to create the σ -hole were recorded. The electron-withdrawing group is defined in this work as the substituent on the σ -hole donor that contains the potential σ -hole angle nearest linearity, an example is shown in Figure 3.33.



Figure 3.33: The location of the electron-withdrawing group (EWG) forming the σ -hole on a halogen bond donor (*left*), and a tetrel bond donor (*right*)

Nitrogen can play a different role in σ -hole interactions, other than as an acceptor, in being a strongly electron-withdrawing group capable of eliciting a strongly positive σ -hole. As shown in Figure 3.34, using the halogen bonding contacts to sp^2 -hybridised nitrogen as an example, all of the contacts below 0.7 normalised contact distance have σ -hole donors that are either dihalogens or have a nitrogen covalently bonded to the halogen σ -hole donor.

Similar distributions are seen in plots for chalcogen bonding contacts, which can be explained by the chalcazole chalcogen bond donor motif discussed in Section 3.4.6, with the distributions of σ -hole angle against normalised contact distance colour-coded by electron-withdrawing group for other σ -hole donors and other nitrogen hybridisations included in Appendix A.2.

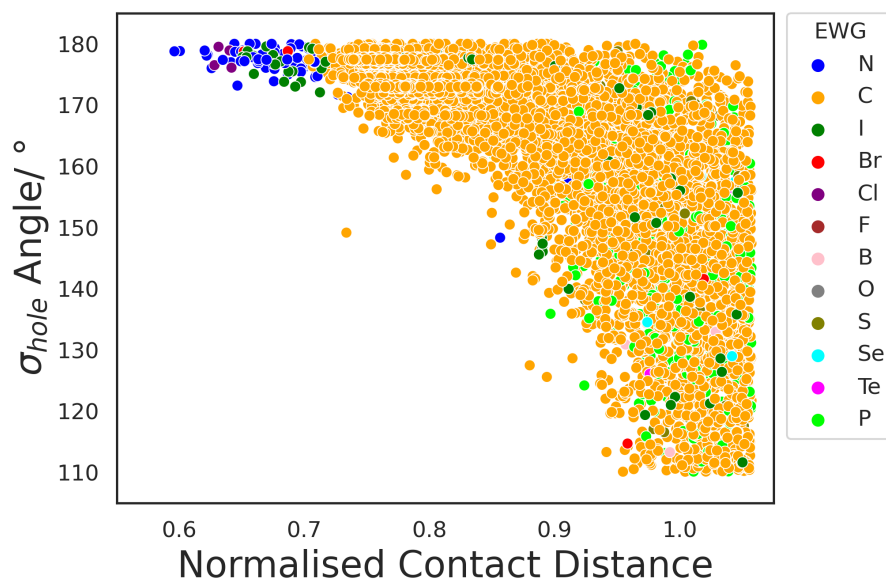


Figure 3.34: Scatter plot showing the distribution of normalised contact distance with respect to the σ -hole angle of halogen bonding contacts to sp^2 -hybridised nitrogen found in crystal structures in the CSD. Points colour coded by electron-withdrawing element

For halogen bonding contacts presented in Figure 3.34, an archetypal class of halogen bond donor that utilises a nitrogen as an electron-withdrawing substituent are N-halosuccinamides. The halogen-bonded complex of N-iodosuccinamide with pyridine (CSD refcode **OGIWOS**) is shown in Figure 3.35¹⁴⁷.

OGIWOS was part of an investigation by Espinosa *et al.* on the distinction between cocrystals and salts in crystalline halogen bonded systems¹⁴⁷. The halogen bond donor species used are N-iodosuccinamide and N-iodosaccharin, which are complexed with pyridine and substituted pyridines.

This change in bonding model can be expressed in the form $N-I \cdots N$ for the cocrystal or "close-to-neutral" structure, and $[N]^- \cdots [I-N]^+$ for the salt or "close-to-ionic" structure. This study offers insight into the halogen bonding continuum which draws parallels to hydrogen bonding. The "extreme" of hydrogen bonding results in protonation of the hydrogen bond acceptor, and analogously with this "extreme" of halogen bonding it results in the transfer of I^+ to the pyridine Lewis base, both of which result in an ionic mode of bonding.

In this study the N-iodosuccinamide complexes were determined to be "close-to-neutral" structures, whereas the complex of N-iodosaccharin and 4-(dimethylamino)pyridine (see CSD refcode **KUWHUH**, shown in Figure 3.36) was determined to be "close-to-ionic". This is supported by theoretical calculations and topological analysis of the electron density distribution indicating notable charge transfer between the N-iodosaccharin and 4-(dimethylamino)pyridine moieties. The iodine $\cdots$ nitrogen distance in **OGIWOS** is 2.430 Å (0.657 normalised contact distance), whereas in **KUWHUH** the analogous distance is 2.228 Å (0.602 if treated like a normalised contact distance). This is interesting as it shows, from an interatomic separation perspective, how little difference there is between these two modes of bonding, with a change of 0.202 Å (0.055 normalised contact distance) meaning a change from a "close-to-neutral" categorisation to "close-to-ionic".

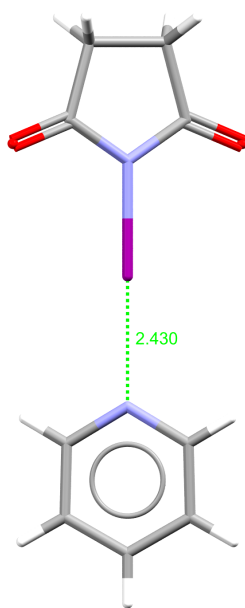


Figure 3.35: An example of a crystal structure with a halogen bonding contact involving a N-iodosuccinamide as an halogen bond donor and pyridine as a halogen bond acceptor. CSD refcode: **OGIWOS**

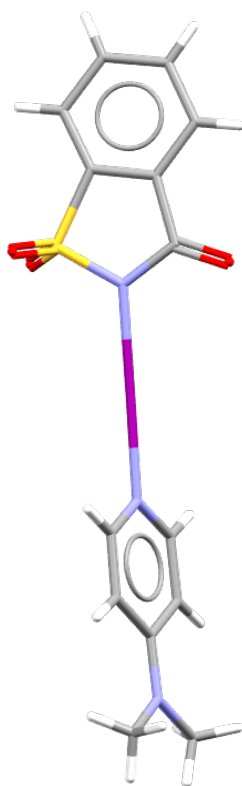


Figure 3.36: An example of a crystal structure involving a N-halosaccharin molecule that forms a salt, rather than a halogen bonded cocrystal with 4-(dimethylamino)pyridine. CSD refcode: **KUWHUH**

The majority of the σ -hole donors have a carbon-based substituent that is acting as an electron-withdrawing group. Although carbon is not an inherently electron-withdrawing element, the substituents on the carbon or the bonding environment of the carbon can result in an electron-withdrawing effect on the σ -hole donor atom. Although not resulting in contacts as short as those involving nitrogen- or halogen-based electron-withdrawing groups, there are clear trends to be seen with respect to the hybridisation of the carbon.

To illustrate this, Figure 3.37 shows that increased s -character of the carbon results in a distribution of σ -hole angles that are close to linearity and have normalised contact distances shorter than 0.85. This suggests that an sp -hybridised carbon EWG creates a stronger σ -hole, but it is a smaller number of contacts compared to those involving sp^2 - and sp^3 -hybridised carbon. This is likely to be due to the increased inductive effect of conjugation, where more of the electron density of the moiety is located in the π systems of the sp - and sp^2 -hybridised species, thus causing a greater anisotropy on the σ -hole donor.

This is supported by computational studies involving halogen bonding, with the magnitude of the σ -hole increasing with carbon electron-withdrawing group hybridisation in the order $C_{sp^3} < C_{sp^2} < C_{sp}$, and with studies involving hydrogen bonding following the same trend^{148 149}. Contacts involving sp^2 -hybridised and sp^3 -hybridised carbon EWGs show similar directionality but those involving sp^2 -hybridised EWGs are by far the most abundant. This is likely a reflection of the

widespread use of halogenated benzenes as halogen bond donors due to their tunable nature¹⁵⁰. See Appendix A.2 for the effect of different carbon electron-withdrawing group hybridisations on the distributions of σ -hole angle against normalised contact distance for halogen bonding contacts to sp^3 - and sp -hybridised nitrogen.

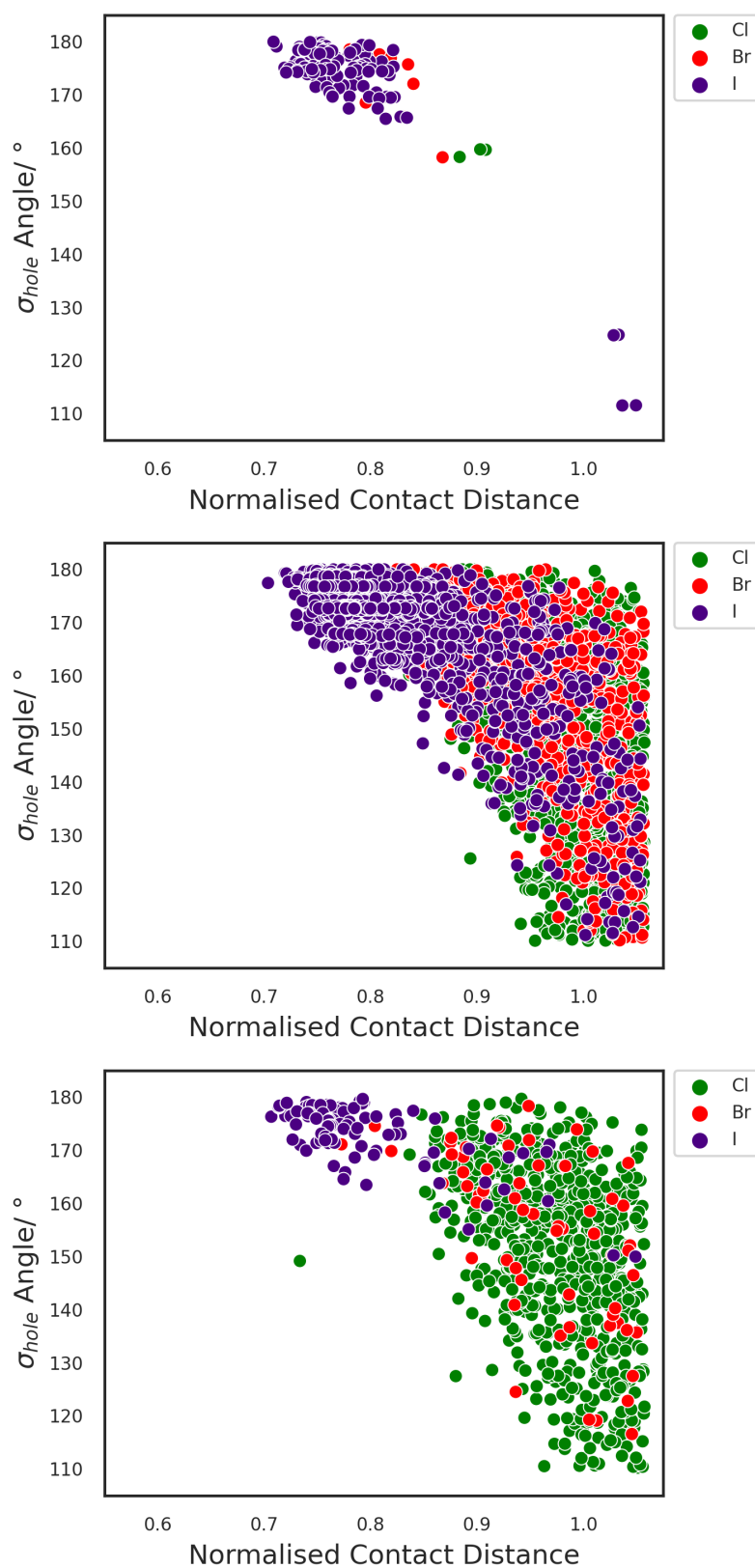


Figure 3.37: Scatter plot showing the distribution of normalised contact distance with respect to the σ -hole angle of halogen bonding contacts to sp^2 -hybridised nitrogen showing how the distribution changes with halogen bond donors containing a sp -hybridised carbon EWG (*top*), a sp^2 -hybridised carbon EWG (*middle*), and a sp^3 -hybridised carbon EWG (*bottom*). Points colour coded by donor atom

3.7 Summary

It is clear from the findings in this Chapter that nitrogen plays a crucial and pervasive role in forming σ -hole interactions, not just as a σ -hole acceptor, but sometimes also in causing the charge anisotropy to create a σ -hole when covalently bonded to the σ -hole donor. The hypothesised effects of varying the hybridisation of the nitrogen, detailed at the start of the Chapter, were confirmed by the geometric trends in contact distance and directionality with respect to the σ -hole and the lone pair on the nitrogen. For those looking to exploit these σ -hole interaction motifs involving nitrogen Lewis bases, the main takeaways from this chapter are as follows.

Lewis bases that involve a sp^3 -hybridised nitrogen are capable of forming some of the strongest interactions with σ -hole donors, but are more suited as acceptors in halogen bonds than in other σ -hole interactions. This is due to the steric bulk of a trivalent nitrogen hindering the approach of σ -hole donors with more than one substituent. This steric effect can be minimised by using less sterically hindered Lewis bases such as DABCO or quinuclidine, where the cage-like molecular structure holds the nitrogen substituents back, minimising steric interactions¹⁵¹. Nitrogen Lewis bases that are sp^2 -hybridised have σ -hole interactions comparable in closeness of contacts and directionality to analogous interactions involving sp^3 -hybridised nitrogen, but are able to participate in a wide range of σ -hole interactions owing to their planar nature which reduces steric clashes. These Lewis bases are also readily synthetically accessible through established nitrogen heterocycle chemistry and are consequently a widely utilised motif in these σ -hole interacting systems. Although sp -hybridised nitrogen Lewis bases would appear to display contacts that are relatively weak and less directional than those involving the other hybridisations of nitrogen, they offer some unique advantages. Being sterically unencumbered makes them ideal for interacting with sterically crowded σ -holes such as those found on pnictogen or tetrel species. There is also evidence of the strengthening impact of σ -hole interactions with cyano groups when coordinated to a metal centre.

Nitrogen is also a very effective electron-withdrawing group when bonded to the sigma-hole donor atom, causing a σ -hole that is very strongly interacting. Functional groups have been used with this in mind, such as N-halosuccinamides and chalczaoles. This electron-withdrawing nature can be so effective, however, that the σ -hole interaction begins to move beyond a noncovalent interactions towards something more covalent in nature.

The prevalence of σ -hole contacts to nitrogen in crystal structures that contain the required moieties for σ -hole interactions provides further insight. The higher prevalence of halogen and chalcogen bonding suggests that these interactions have moved beyond novel studies and proof-of-concept, and are actively being utilised in the creation of halogen and chalcogen bonded crystals. This assertion is supported by a recent review on σ -hole interactions¹⁵². In contrast pnictogen and tetrel bonding contacts are still uncommon, with low prevalence. This highlights the need for further experimental study of pnictogen and tetrel bonding interactions, potentially focussing on sp -hybridised nitrogen Lewis bases as acceptor moieties, to minimise unfavourable steric interactions. Given that metal coordination of cyano groups results in them being a more effective σ -hole acceptor, it poses the question of how this might effect other, differently hybridised nitrogen species. These metal coordinates nitrogen Lewis bases present an interesting area of future study and represent an under-represented motif in terms of σ -hole interactions.

Chapter 4

Oxygen as an Acceptor

4.1 Introduction to Oxygen Lewis Bases

Another key element capable of acting as an acceptor for σ -hole interactions is oxygen, which possesses two lone pairs in its saturated valence, allowing two potential acceptor sites for σ -hole interactions. There are also a wide variety of oxygen-containing functional groups, which will effect the nucleophilicity of the lone pairs, and their expected location as predicted by orbital hybridisation theory. It is expected from work pertaining to hydrogen bonding to oxygen and nitrogen systems, that oxygen will not be as effective a Lewis base in comparison to nitrogen, so the contacts should be generally longer than those in the previous chapter¹⁵³.

Once again comparison can be made to Robin Taylors R_F values, where for instance analogous Br/I contacts with N/O are both tabulated¹⁴³. The R_F values of $I \cdots O$ is 2.8, which is considerably lower than the $I \cdots N$ motif ($R_F = 5.4$), and $Br \cdots O$ has an R_F of 1.5 compared to a value of 1.8 for $Br \cdots N$. These lower R_F values for the halogen bonding contacts to oxygen suggest they are generally weaker, less significant interactions within a crystal structure. This is also seen in calculated interaction energies when oxygen and nitrogen are in competition as hydrogen bond acceptors¹⁵⁴. This is likely due to oxygen being more electronegative than nitrogen, meaning its lone-pairs are less available for donation than the single lone-pair on nitrogen.

In the literature there can be found a plethora of different cases where oxygen acts as an acceptor in noncovalent interactions. This includes multiple different functional groups, such as carbonyl compounds, for which halogen bonding has been utilised to activate the carbonyl in a Diels-Alder reaction¹⁵⁵. A similar approach has also been applied using tellurium and selenium chalcogen bonds to activate carbonyls in the Michael addition reaction¹⁵⁶. Halogen bonding interactions have been experimentally analysed via Raman and infrared spectroscopy between trifluoromethyl halides and dimethyl ether, an example of a divalent oxygen species. Similar halogen bonding to ether and alcohol moieties have been used in crystal engineering such as in the co-crystals of *o*-vanillin and diiodotetrafluorobenzenes¹⁵⁷. Zhao *et al.* has developed a methodology for activating of benzylic and allylic ethers via $Se \cdots O$ and $Se \cdots \pi$ chalcogen bonding interactions. Aromatic

N-oxides prove to be an important acceptor for halogen bonds, with several works involving pyridine-N-oxide and strong halogen bond donors^{158 159}. As many oxide species tend to be anionic, they have been utilised in explorations of σ -hole interactions for use in anion recognitions¹⁴². Little work exists with regards to halogen bonding interactions to metal carbonyls, but more comprehensive studies exist showcasing hydrogen bonding to metal carbonyls^{160 161 141}.

In contrast to nitrogen Lewis bases, oxygen Lewis bases have a much larger diversity of named functional groups. This means that much more granular trends can be analysed with respect to the acceptor moiety, such as the trends vast range of different carbonyl functional groups that exist, or the many forms of oxide ions. Given the many subdivisions that can be made in the analysis of σ -hole interactions to oxygen Lewis bases, the content of this chapter is more exploratory than the analysis with respect to nitrogen Lewis bases.

4.2 CSD Query Details

For efficiency in designing queries and post-processing of data using Python code, these oxygen Lewis bases were split up into pseudo-functional groups; organic carbonyls, divalent oxygen, oxides, and metal carbonyls, which includes terminal, di- and tri-bridging carbonyls. These can be further subdivided by grouping identities of substituent *R*-groups that are not conserved in the query, for example the two *R*-groups either side of the carbonyl carbon. To get complete specificity of all oxygen functional groups would require a query and associated code routine for every functional group, which as the reader will notice later in this chapter, would lead to a very sizeable codebase. This leads to some problematic imperfect divisions, such as an ether oxygen and the non-carbonyl oxygen of an ester or an anhydride being both retrieved by the divalent oxygen query. This also provides challenges in studying the directionality with respect to the lone pairs difficult if these lone pairs are predicted to be out of the plane of the functional group, like in dimethyl ether for instance. Through the use of angles defined between the plane of the functional group, and the plane through the donor $R-D_{\sigma_{hole}} \cdots O$ contact, this directionality can be explored in multiple dimensions.

The σ -hole donor moieties detailed in Figure 2.1 from Section 2.1.1 are combined with the various acceptor moieties shown in Figure 4.1 to form the query files used to complete these searches, with contact distance $d_{contact}$ defined between the D_{σ} donor atom (*X/Ch/Pn/Tt*) and the oxygen atom. The resulting 24 queries cover halogen, chalcogen, pnictogen, and tetrel contacts to organic carbonyls, divalent oxygens, oxides, terminal metal carbonyls, di-metal bridging carbonyls, and tri-metal bridging carbonyls.

This has led to the adoption of different descriptions for the directionality with respect to the lone-pair(s), which varies based on the different pseudo-functional groups used in these queries. These lone-pair angles are not constrained but are used as an exploratory measurement, but in many cases are expected to be around 120° following orbital hybridisation rules. For the divalent oxygen species (*R-O-R*) there needs to be two angles defined, relative to each of the two substituents, which are both analysed. For the organic carbonyl and divalent oxygen species, one can investigate the approach of the σ -hole donor relative to the plane of the oxygen-containing moiety by the use of a so-called "plane angle".

The interplanar "plane angle", defined in Figure 4.2, is measured between the "contact plane" and the "acceptor plane".

The "contact plane" is defined by the donor atom, the oxygen atom, and either the carbon of the carbonyl group or a centroid between the two R groups of the divalent moiety, and the "acceptor plane" and goes through the oxygen atom and both R groups on the acceptor component of the query. This angular measurement is used alongside the lone-pair angle(s) to probe whether the approach of a σ -hole donor is close to coplanarity or orthogonality to the acceptor moiety. Different bond types can be defined between two atoms within the CSD, these include: single, double, triple, quadruple, aromatic, polymeric, delocalised, and π . The bonds drawn in the below figure are only a representation, with several of the bonds present in the acceptor part of the queries having variable bond typing. The bonds in the figure that differ from the illustration are the two R-O-R bonds in the divalent oxygen query, which have possible bond types of single, double and delocalised; the oxide O-R bond can be of any type. For metal carbonyls the C-M bonds can be of any type, and the typing of the carbonyl in the terminal and di-bridging carbonyls can be doubly or triply bonded, while in the tri-bridging queries it can be singly, doubly, or triply bonded.

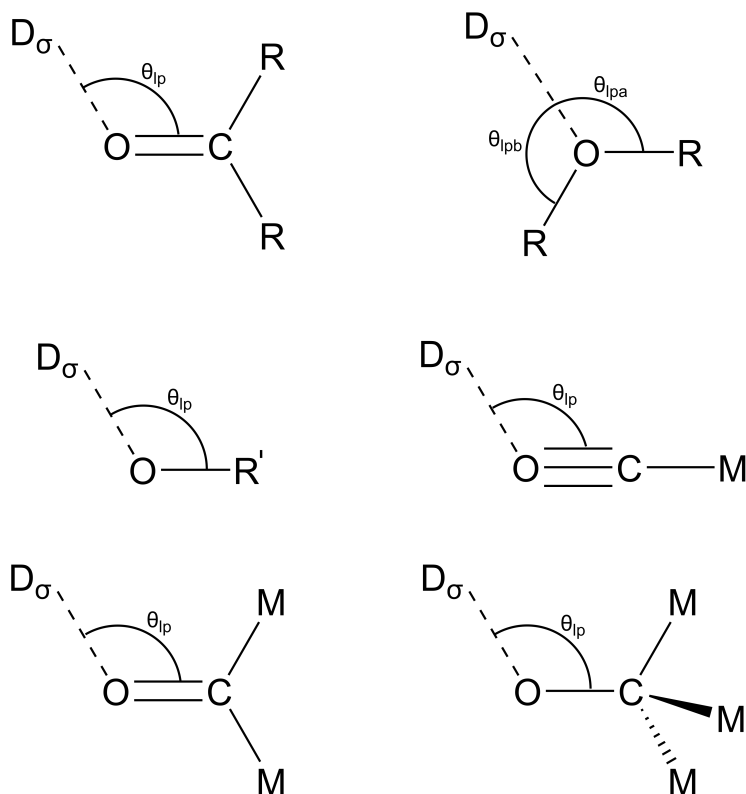


Figure 4.1: A representation of the various lone-pair angle measurements for the oxygen functional group components of the CSD search queries. R = H, C, N, O, Si, P, S, F, Cl, Br, I, Se, Te, B. R' = N, Ch, Tt, Pn, X, M. Ch = Chalcogen, Pn = Pnictogen, Tt = Tetrel, X = Halogen, M = Metal. Bond typing (eg. bond order) in the queries are allowed to be variable and should not be taken as limited to those shown, the figure is intended to convey the idealised functional group.

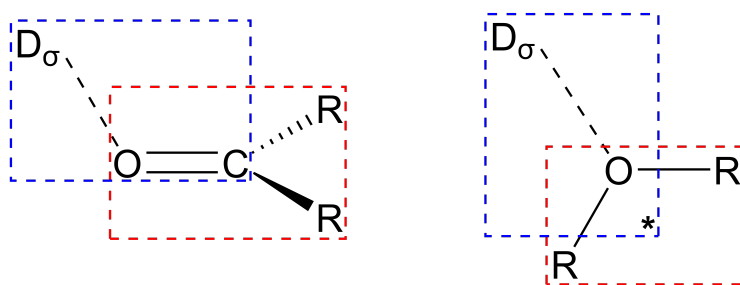


Figure 4.2: A representation of the two planes used to calculate the plane angle of the contact. The blue plane is the contact plane, which goes through the σ -hole donor atom, the oxygen atom, and either the carbon of the carbonyl group or a centroid between the two R groups of the divalent moiety. The red plane is the "acceptor plane" and goes through the oxygen atom and both R groups on the acceptor component of the query.

4.3 Organic Carbonyls

As to be expected, the number of contacts indicative of σ -hole interactions involving organic carbonyls is very large owing to their widespread use in organic chemistry. To hypothesise on the effects of different carbonyl substituents on the Lewis basicity of the carbonyl, we look to the work of Laurence and co-workers, which involves the formation of Lewis acid-base adducts involving various carbonyl compounds as the Lewis base and I_2 acting as the Lewis acid¹⁵³. This is most fortuitous, as this work is older than the definitions of halogen bonding or the σ -hole model itself, yet shows clear structural relationships for a range of halogen bonding systems. Laurence concludes that inductively withdrawing substituents on the carbonyl stabilises the lone pair, and π -donation causes the oxygen lone pair orbital to be less tightly bound. Laurence goes on to show that the least basic carbonyls were those substituted by halogens and that any π -donation of alkoxy or phenyl groups is overcome by their inductive effects, meaning these carbon and oxygen containing carbonyls follow the sequence in Lewis basicity ketones > esters and MeCOMe > PhCOMe > PhCOPh. The importance of π -donation is shown in the most Lewis basic of the carbonyl systems, these being amides and ureas in the sequence in Lewis basicity DEA (diethylacetamide) > TMU (tetramethylurea) > DMA (dimethylacetamide) > DMF(dimethylformamide) > DMBz (dimethylbenzamide) > TEU (tetraethylurea).

The lone-pair angles for these acceptor moieties are expected to favour an angle of around 120° , in the plane of the carbonyl, as predicted by sp^2 -hybridisation. This lone-pair angle is also expected to be coplanar with the carbonyl moiety, so plane angles (Figure 4.2) near 0° or 180° are anticipated. An overall summary of the geometric observables of contacts indicative of σ -hole interactions to organic carbonyl oxygen moieties is tabulated in Table 4.1.

Table 4.1: Overall geometric summary of the contacts indicating σ -hole interactions to organic carbonyl oxygen moieties

σ -hole Interaction Type	No. Contacts	Mean Normalised Contact Distance	Mean σ -hole Angle/ $^{\circ}$	Mean Lone-Pair Angle/ $^{\circ}$	Mean Plane Angle/ $^{\circ}$
Halogen-bonding	10823	0.969	155	124	48
Chalcogen-bonding	5497	0.977	154	125	48
Pnictogen-bonding	97	0.920	160	129	49
Tetrel-bonding	325	0.915	162	134	51

4.3.1 $X \cdots$ Organic Carbonyl Contacts

Out of the contacts summarised in Table 4.1, halogen bonding contacts are the most numerous, with a greater number than chalcogen, pnictogen, and tetrel bonding contacts to organic carbonyl moieties combined. The distribution of σ -hole angles against normalised contact distance for halogen bonding contacts to organic carbonyl moieties, presented in Figure 4.3, shows the characteristic "wedge" shape as seen with the contacts presented in Chapter 3. Very few of these contacts are shorter than 0.8 normalised contact distance and they comprise nearly entirely of iodine contacts.

As shown in the mean averaged geometric measurements shown in Table 4.2, the normalised contact distance decreases with the more polarisable halogen bond donor atom, following the trend $I < Br < Cl$. On a similar note the average σ -hole angle increases with the more polarisable halogen bond donor atom, following the trend $I > Br > Cl$. This trend is also seen with the lone-pair angle, with chlorine contacts having average lone-pair angles the closest to the expected 120° , the average angles get larger going from chlorine to bromine to iodine.

The distribution of lone-pair angles against normalised contact distance for halogen bonding contacts to organic carbonyl moieties, shown in Figure 4.4, and the distribution of plane angle against normalised contact distance, see Figure 4.5 were examined to analyse the directionality of the contact with respect to the lone-pair. One the clear preference for the closest contacts to adopt lone-pair angles of around 120° , and then a small number of contacts adopting plane angles nearer 0° , suggesting a coplanar approach from donor and acceptor moieties.

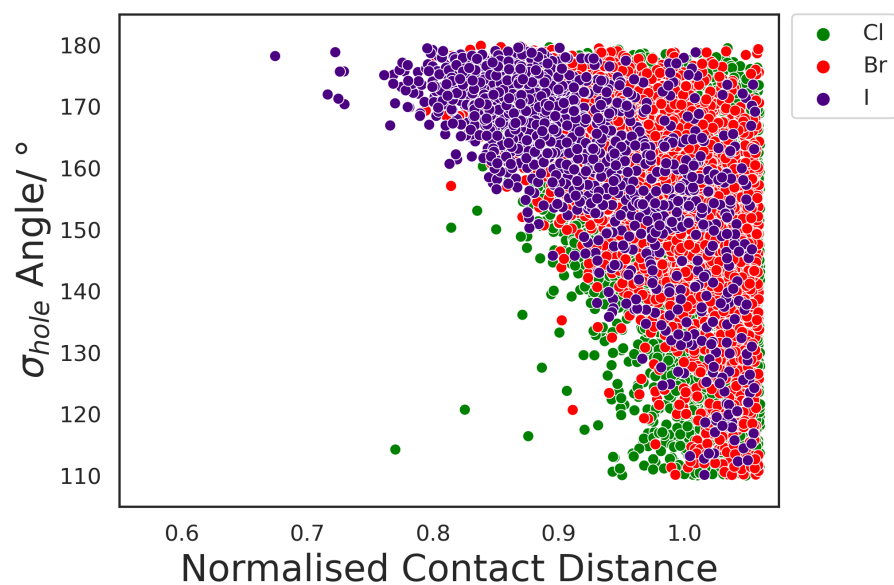


Figure 4.3: Scatter plots showing the distribution of normalised contact distance with respect to the σ -hole angle of halogen bonding contacts to organic carbonyls found in crystal structures in the CSD. Points colour coded by donor atom type

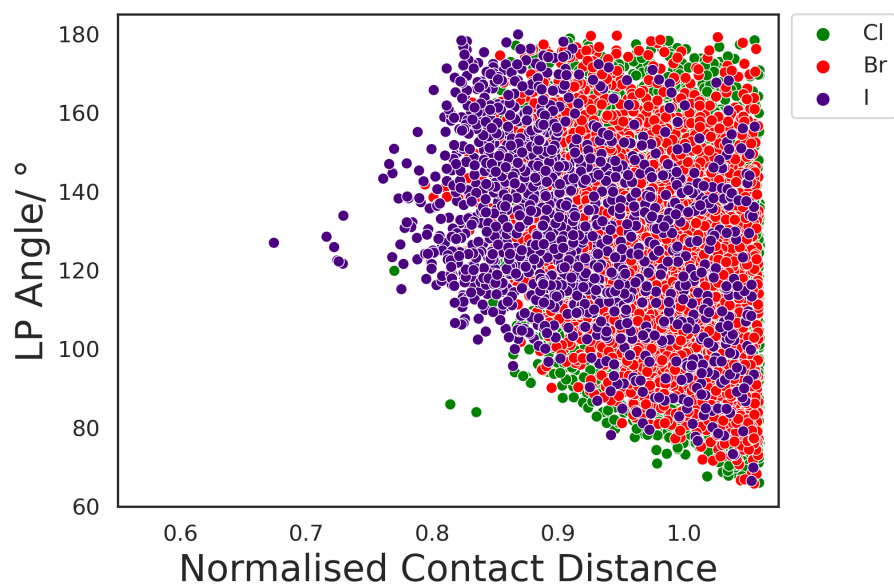


Figure 4.4: Scatter plots showing the distribution of normalised contact distance with respect to the lone-pair angle of halogen bonding contacts to organic carbonyls found in crystal structures in the CSD. Points colour coded by donor atom type

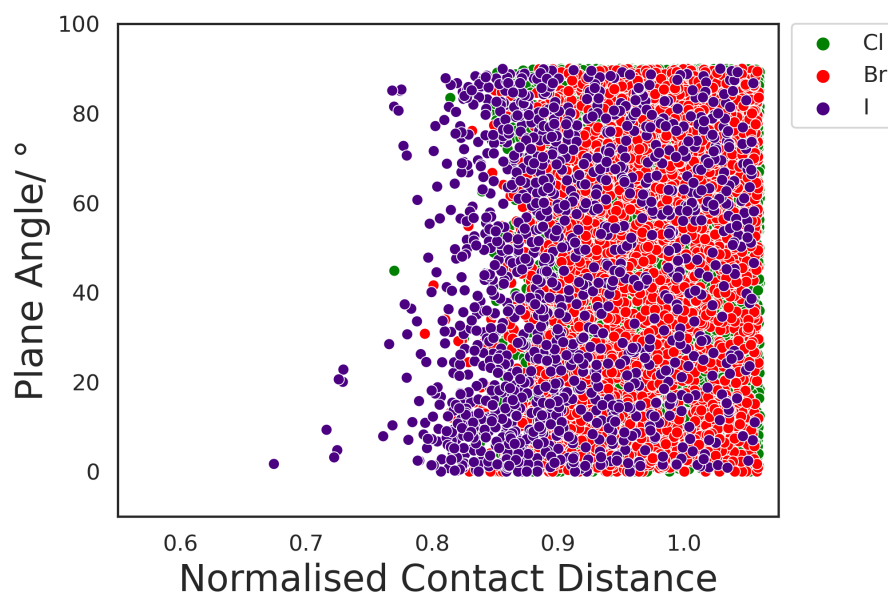


Figure 4.5: Scatter plots showing the distribution of normalised contact distance with respect to the plane angle angle of halogen bonding contacts to organic carbonyls found in crystal structures in the CSD. Points colour coded by donor atom type

Table 4.2: Summary of geometric characteristics of halogen bonding contacts to organic carbonyl oxygen

Donor	Contacts	Mean Contact Distance/ Å	Mean Normalised Contact Distance	Mean σ -hole Angle/ °	Mean Lone-Pair Angle/ °	Mean Plane Angle/ °
X	10823	3.252	0.969	155	124	48
I	1287	3.208	0.906	163	131	42
Br	3631	3.253	0.968	157	125	46
Cl	5905	3.261	0.982	152	121	50

By examining the combinations of R -groups either side of the carbonyl carbon $R_1C(=O)R_2$, a plethora of subdivisions can be made. This includes some more unusual compounds that would be impractical to consider separately, so for brevity only the following nine classes of carbonyls will be examined; ketones($R_1, R_2 = C, C$), aldehydes($R_1, R_2 = C, H$), amides($R_1, R_2 = C, N$), carboxyls($R_1, R_2 = C, O$), ureas($R_1, R_2 = N, N$), carbamates($R_1, R_2 = O, N$), formamides($R_1, R_2 = N, H$), thioesters($R_1, R_2 = C, S$), and acid chlorides($R_1, R_2 = C, Cl$). These subdivisions make up around 99% of the organic carbonyl contacts and the distributions of σ -hole angle against normalised contact distance of each subdivision is shown in Figure 4.6.

The most numerous of the carbonyl classes are carboxyls, amides, and ketones, with 3866, 3098, and 2644 contacts respectively, which is to be expected given how ubiquitous these functional groups are in organic chemistry. From visual inspection of these plots, in combination with the mean averages of the geometric measurements listed in Table 4.3, it can be seen that the distributions of these contacts are very similar to one another, suggesting that the carbonyl oxygen's ability as a σ -hole acceptor is fairly insensitive to the substituents on the carbonyl.

The contacts with the shortest normalised contact distance, largest σ -hole angles, and largest lone-pair angles involve amides as the halogen bond acceptor. These contacts, along with those involving ureas and formamides show that incorporating a nitrogen into the carbonyl group does slightly enhance the carbonyl oxygen's ability to act as a σ -hole acceptor, as shown in the work of Laurence¹⁵³. Interestingly, aldehydes seem to show a narrower spread to its "wedge" distribution, being fairly comparable qualitatively to the σ -hole angle directionality plot of the contacts to urea. It can be seen in the contacts to acid chlorides that there is a deactivating effect on the Lewis basicity of the carbonyl oxygen caused by the chlorine. This is likely due to chlorine being an electron withdrawing substituent and thus will remove electron density from the carbonyl system, resulting in less available lone-pairs on the oxygen. This structure-basicity relationship is also remarked upon in the work of Laurence and coworkers¹⁵³.

Table 4.3: Summary of geometric characteristics of halogen bonding contacts to different organic carbonyl functional groups

Functional Group	Contacts	Mean Contact Distance/ Å	Mean Normalised Contact Distance	Mean σ -hole Angle/ °	Mean Lone-Pair Angle/ °	Mean Plane Angle/ °
Organic Carbonyls	10823	3.252	0.969	155	124	48
Ketone	2644	3.260	0.972	155	125	45
Aldehyde	286	3.223	0.951	159	123	31
Amide	3098	3.237	0.965	156	123	49
Carboxyl	3866	3.264	0.972	155	124	49
Urea	388	3.218	0.959	156	121	53
Carbamate	319	3.279	0.976	153	122	58
Formamide	87	3.214	0.949	156	113	47
Thioester	35	3.228	0.960	157	124	43
Acid Chloride	26	3.290	0.987	152	125	28

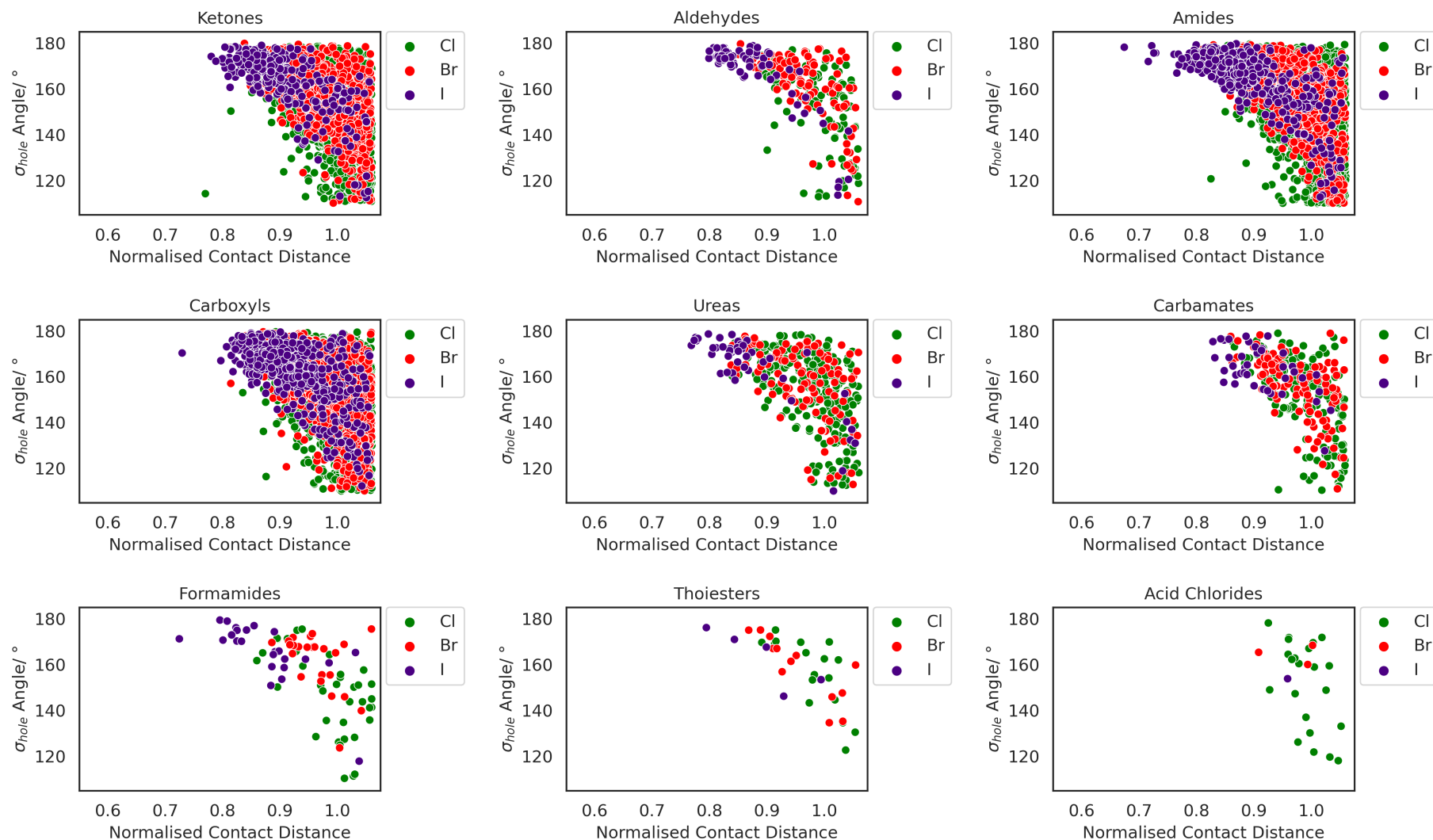


Figure 4.6: Scatter plots showing the distribution of normalised contact distance with respect to the σ -hole angle of halogen bonding contacts to various subdivisions of organic carbonyls found in crystal structures in the CSD. Points colour coded by donor atom type

4.3.2 Ch ... Organic Carbonyl Contacts

Moving on to chalcogen bonding contacts to organic carbonyls, some similarities can be seen with chalcogen bonding contacts to nitrogen Lewis bases. In comparison to the other contacts indicative of σ -hole interactions to organic carbonyl moieties shown in Table 4.1, chalcogen bonding contacts are the second most numerous, with around half as many as halogen bonding contacts.

From the averaged geometric measurements listed in Table 4.4, show that sulfur contains by far the most contacts, making up around 94% of the contacts, with a much smaller number of selenium and tellurium contacts. The average normalised contact distances get smaller for more polarisable chalcogens, following the trend $\text{Te} < \text{Se} < \text{S}$. Selenium and tellurium contacts have similar average σ -hole angles, whereas sulfur contacts have smaller average σ -hole angles. The average lone-pair angle is very consistent between the three chalcogens, with tellurium having an average nearest the expected value 120° . The average plane angles get larger as the polarisability of the chalcogen decreases, following the trend $\text{Te} < \text{Se} < \text{S}$.

Looking to the distributions of σ -hole angle against normalised contact distance, shown in Figure 4.7, selenium contacts make up the majority of the closest contacts with σ -hole angles nearest linearity. Tellurium contacts, which are much rarer, are longer and favour σ -hole angle nearer 160° . Sulfur, which dominates the chalcogen bonding contacts, shows notable directionality with respect to the σ -hole angle, but a group that stands out are some of the closest of the chalcogen bonding contacts to organic carbonyls. These very close sulfur contacts (< 0.8 normalised contact distance) have a range of σ -hole angles between 140 - 160° . These structures correspond to chalcazole molecules containing a carbonyl functional group, as discussed in Section 3.4.6, so the molecules are acting as both σ -hole donors and acceptors which perturbs the expected linear geometries of the contacts.

The distributions of lone-pair angle against normalised contact distance, shown in Figure 4.8, and the distribution of plane angle against normalised contact distance, see Figure 4.9, show similar directionality as halogen bonding contacts with respect to the lone-pair. The shortest contacts, especially those involving tellurium and selenium, show clear preference towards lone-pair angles of around 120° . Interestingly, some of the very short sulfur contacts seem to deviate from this directionality, and approach colinearly to the carbonyl $\text{C}=\text{O}$ bond, yet still having as close a contact as those that approach from the expected angle of the lone-pair. The only slight difference to the distribution of plane angles for halogen bonding, see Figure 4.5, is the shortest normalised contact distances have plane angles between 15 - 55° .

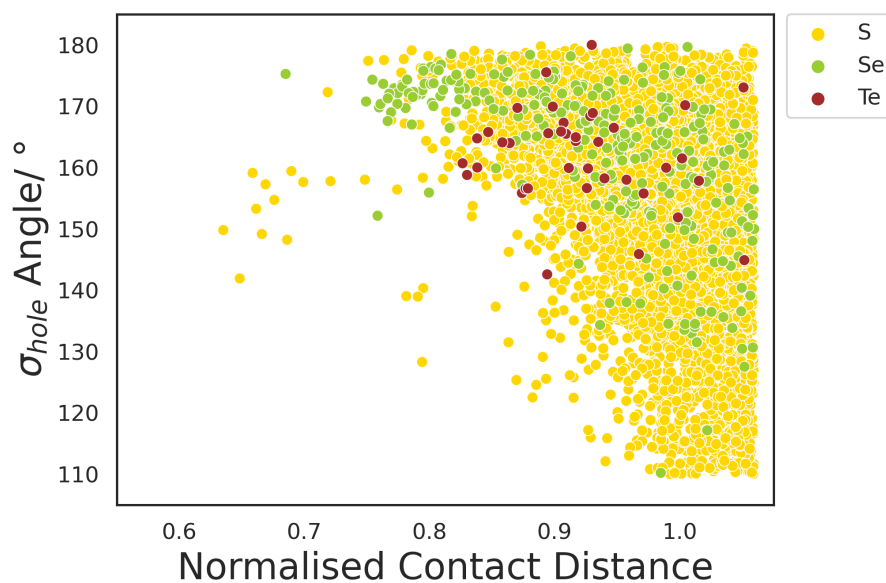


Figure 4.7: Scatter plots showing the distribution of normalised contact distance with respect to the σ -hole angle of chalcogen bonding contacts to organic carbonyls found in crystal structures in the CSD. Points colour coded by donor atom type

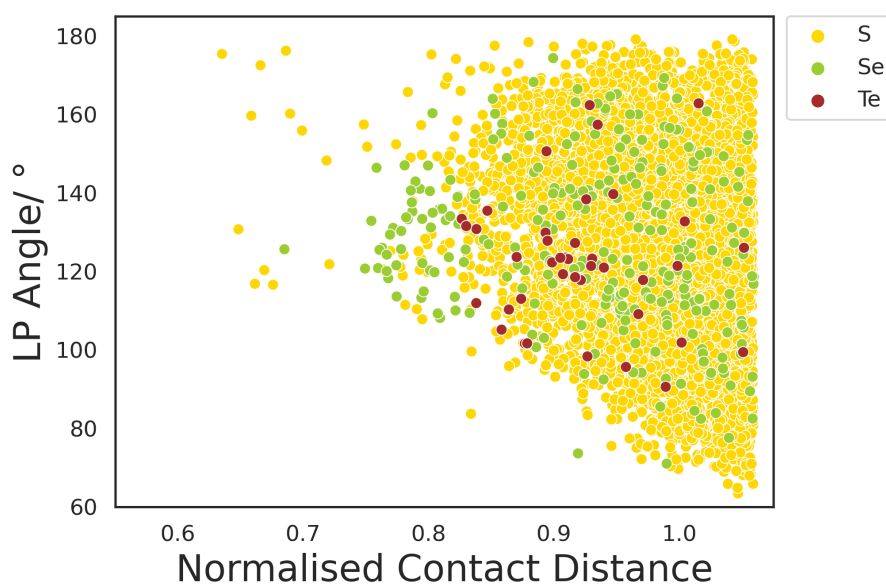


Figure 4.8: Scatter plots showing the distribution of normalised contact distance with respect to the lone-pair angle of chalcogen bonding contacts to organic carbonyls found in crystal structures in the CSD. Points colour coded by donor atom type

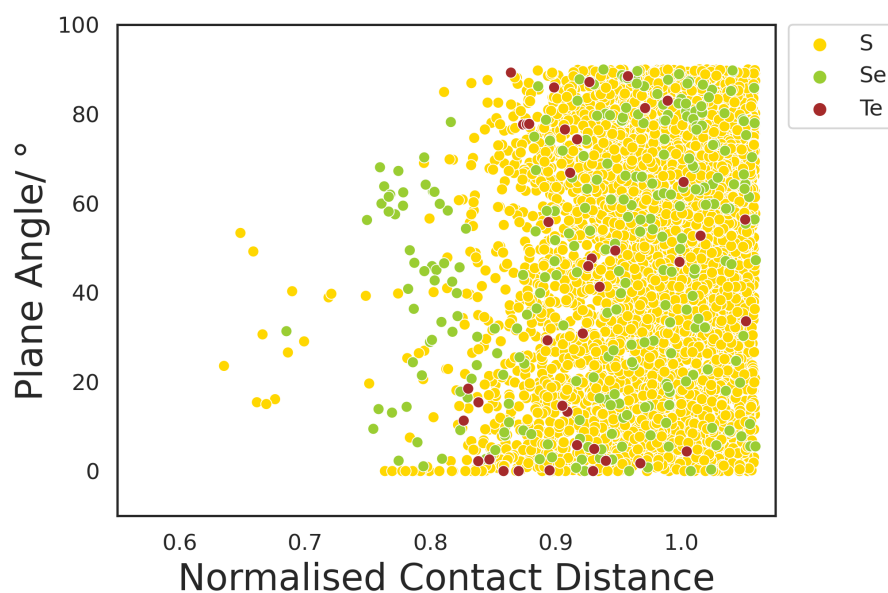


Figure 4.9: Scatter plots showing the distribution of normalised contact distance with respect to the plane angle of chalcogen bonding contacts to organic carbonyls found in crystal structures in the CSD. Points colour coded by donor atom type

Table 4.4: Summary of geometric characteristics of chalcogen bonding contacts to organic carbonyl oxygen

Donor	Contacts	Mean Contact Distance/ Å	Mean Normalised Contact Distance	Mean σ -hole Angle/ °	Mean Lone-Pair Angle/ °	Mean Plane Angle/ °
Ch	5497	3.310	0.977	154	125	48
Te	40	3.214	0.921	161	122	40
Se	280	3.105	0.935	162	126	46
S	5177	3.321	0.980	154	125	48

When the chalcogen bonding contacts to organic carbonyls are broken down into those involving different carbonyl functional groups, greater variation can be found than with halogen bonding contacts. There are an absence of any contacts involving acid chlorides, but an increased abundance of contacts involving dithiocarbonates ($R_1, R_2 = S, S$), carbamothioates ($R_1, R_2 = N, S$), and carbonothioates ($R_1, R_2 = O, S$). These subdivisions are shown in Figure 4.10 and tabulated in Table 4.5, showing some more distinct differences in the character of the interactions than those seen for halogen bonding. Contacts involving tellurium are comparatively rare, and occur mainly with ketones and amides. Likewise, the selenium contacts, although more abundant, mainly occur with ketones, amides, and carboxyls. This is likely a reflection of the greater challenges associated with involved with selenium and tellurium chemistry, which includes issues with toxicity and abundance^{162 163}.

Some of the closest contacts with σ -hole angles nearing linearity are selenium contacts to amides, and overall there is definitely a trend of amides and other nitrogen substituted carbonyls being involved in closer contacts with larger σ -hole angles. Of note are the sulfur contacts to ketones, which are very short and less linear with respect to σ -hole angle, which

correspond to the chalcazole structures mentioned previously.

Table 4.5: Summary of geometric characteristics of chalcogen bonding contacts to different organic carbonyl functional groups

Functional Group	Contacts	Mean Contact Distance/ Å	Mean Normalised Contact Distance	Mean σ -hole Angle/ °	Mean Lone-Pair Angle/ °	Mean Plane Angle/ °
Organic Carbonyls	5470	3.309	0.977	154	125	48
Ketone	1180	3.284	0.969	153	127	45
Aldehyde	82	3.336	0.984	152	116	43
Amide	1574	3.308	0.977	154	127	50
Carboxyl	1710	3.322	0.981	156	125	50
Urea	130	3.329	0.983	157	119	57
Carbamate	72	3.396	1.002	150	120	61
Formamide	60	3.218	0.952	159	117	44
Thioester	296	3.314	0.978	152	125	45
Dithiocarbonate	190	3.341	0.986	149	122	38
Carbamothioate	147	3.316	0.978	153	123	43
Carbonothioate	29	3.198	0.943	161	122	29

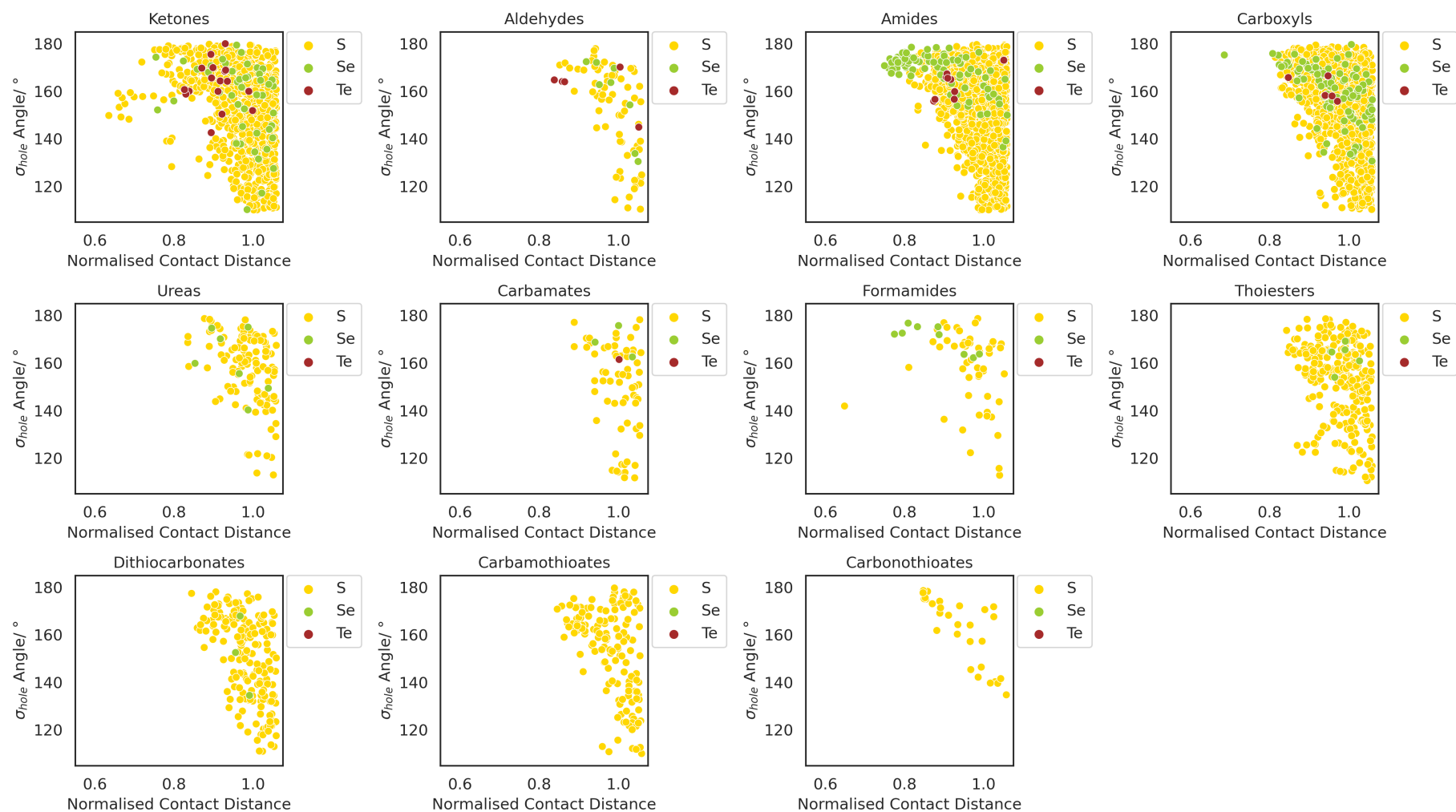


Figure 4.10: Scatter plots showing the distribution of normalised contact distance with respect to the σ -hole angle of chalcogen bonding contacts to organic carbonyls found in crystal structures in the CSD. Points colour coded by donor atom type

4.3.3 *Pn* ... Organic Carbonyl Contacts

Looking to the summary in Table 4.1, pnictogen bonding contacts to organic carbonyl oxygens are, far less abundant than halogen and chalcogen bonding contacts. The mean averages of the geometric measurements of these contacts is listed in Table 4.6, but the scarcity of contacts makes any averages less reliable for comparison. The shortest normalised contact distances involve antimony contacts, with an average normalised contact distance of 0.769. Interestingly, the average σ -hole angles are fairly consistent, but phosphorus contacts have the largest average σ -hole angle, despite the longest average contacts.

The distribution of σ -hole angle against normalised contact distance of pnictogen bonding contacts to organic carbonyl moieties, shown in Figure 4.11, show some directionality of these contacts. There is a greater proportion of overall pnictogen contacts to organic carbonyl moieties that have normalised contact distances less than 0.8, in comparison to the analogous halogen and chalcogen bonding contacts (see Figure 4.3 & Figure 4.7).

The distribution of lone-pair angle against normalised contact distance of pnictogen bonding contacts to organic carbonyl moieties, shown in Figure 4.12, shows a very clear bias towards lone-pair angles of 120° as the normalised contact distance decreases. There are fewer contacts at longer normalised contact distances that have lone-pair angles approaching linearity, unlike what is seen with halogen and chalcogen bonding (see Figure 4.4 & Figure 4.8). However there is little to no directionality with respect to the plane angle, shown in Figure 4.13, with the contacts showing similar likelihoods of for the full range of angles between coplanar and orthogonal.

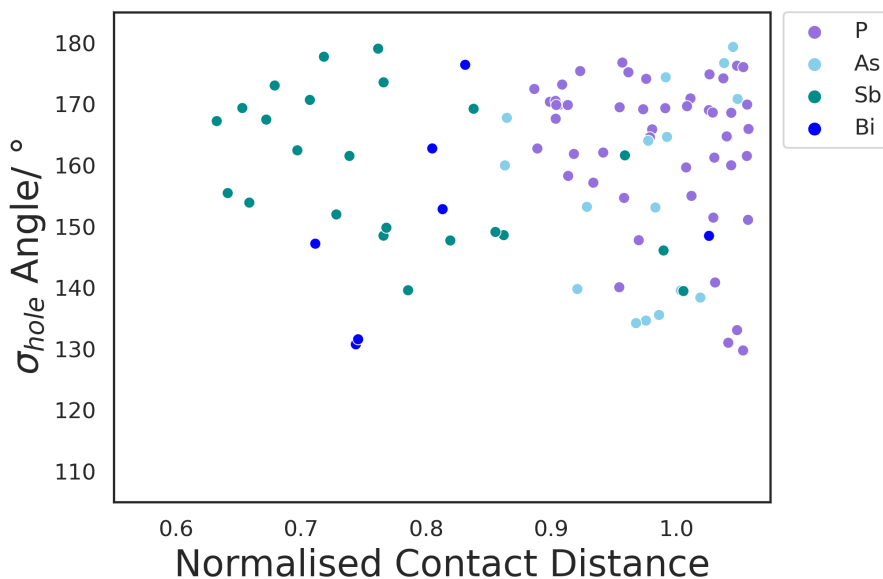


Figure 4.11: Scatter plots showing the distribution of normalised contact distance with respect to the σ -hole angle of pnictogen bonding contacts to organic carbonyls found in crystal structures in the CSD. Points colour coded by donor atom type

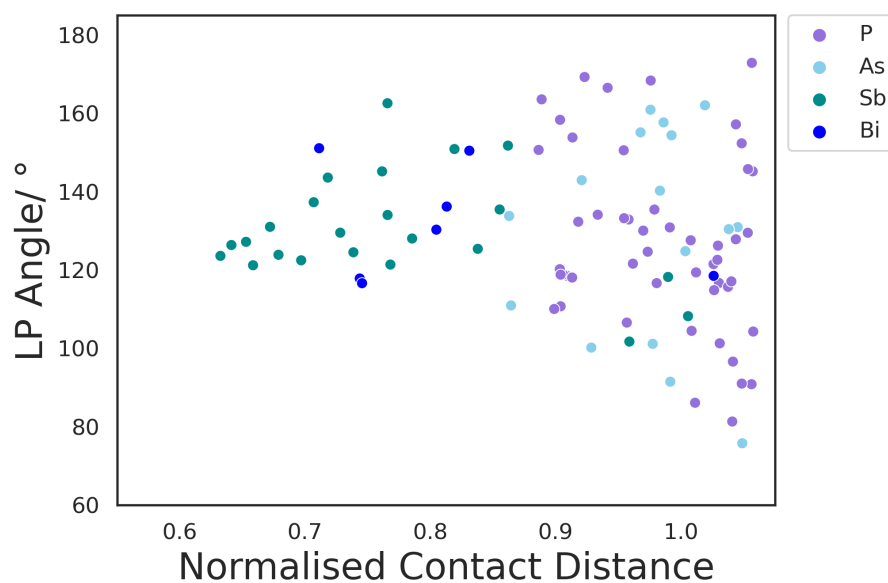


Figure 4.12: Scatter plots showing the distribution of normalised contact distance with respect to the lone-pair angle of pnictogen bonding contacts to organic carbonyls found in crystal structures in the CSD. Points colour coded by donor atom type

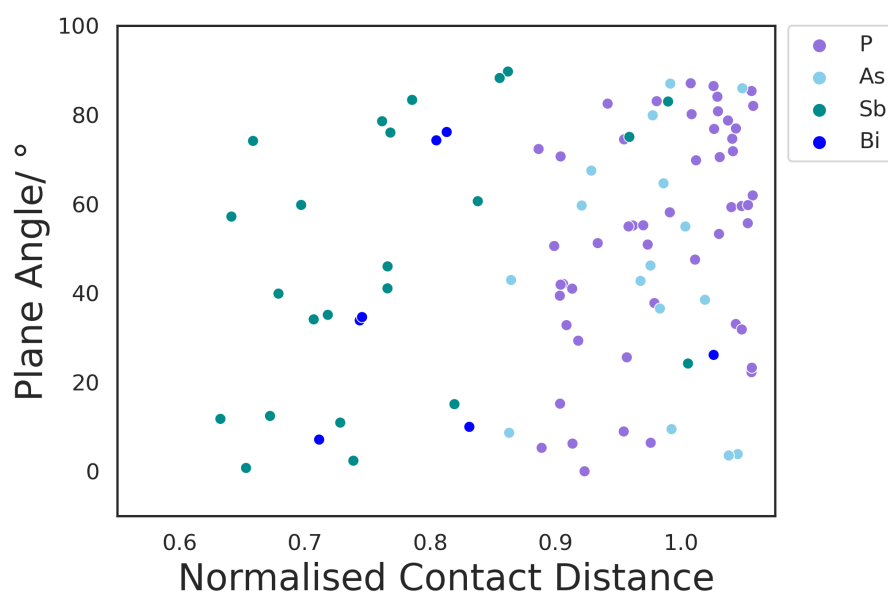


Figure 4.13: Scatter plots showing the distribution of normalised contact distance with respect to the plane angle angle of pnictogen bonding contacts to organic carbonyls found in crystal structures in the CSD. Points colour coded by donor atom type

Table 4.6: Summary of geometric characteristics of pnictogen bonding contacts to organic carbonyl oxygen

Donor	Contacts	Mean Contact Distance/ Å	Mean Normalised Contact Distance	Mean σ -hole Angle/ °	Mean Lone-Pair Angle/ °	Mean Plane Angle/ °
Pn	97	3.266	0.920	160	129	49
Bi	7	3.275	0.811	149	131	37
Sb	23	3.055	0.769	159	130	47
As	16	3.297	0.975	155	129	45
P	51	3.350	0.985	163	128	52

Congruent with a scarcity of contacts involving pnictogen bonding to organic carbonyls, there is also a lack of carbonyl functional group diversity within the contacts. Shown in the mean averages listed in Table 4.7 and the distributions of σ -hole angle against normalised contact distance Figure 4.14, the contacts are dominated by ketones, amides, carboxyls, and ureas. The data is sparse, and not representative of the whole chemical space of the interactions, however, the majority of the closer contacts involve ketone and carboxyl Lewis bases, rather than the nitrogen substituted amides and ureas. It is worth noting that the contacts not plotted separately due to only having single numbers of contacts, namely; carbamate, formamide, thioester, and dithiocarbonates, still show notable closeness of contact and directionality.

Table 4.7: Summary of geometric characteristics of pnictogen bonding contacts to different organic carbonyl functional groups

Functional Group	Contacts	Mean Contact Distance/ Å	Mean Normalised Contact Distance	Mean σ -hole Angle/ °	Mean Lone-Pair Angle/ °	Mean Plane Angle/ °
Organic Carbonyls	88	3.245	0.910	160	129	47
Ketone	13	3.099	0.835	164	129	51
Aldehyde	3	3.478	0.865	162	131	12
Amide	14	3.387	0.977	164	129	56
Carboxyl	39	3.214	0.908	157	129	43
Urea	13	3.270	0.933	158	132	52
Carbamate	2	3.493	1.027	168	122	85
Formamide	1	3.438	1.011	170	86	47
Thioester	1	2.873	0.711	147	151	7
Dithiocarbonate	2	3.156	0.795	150	140	50

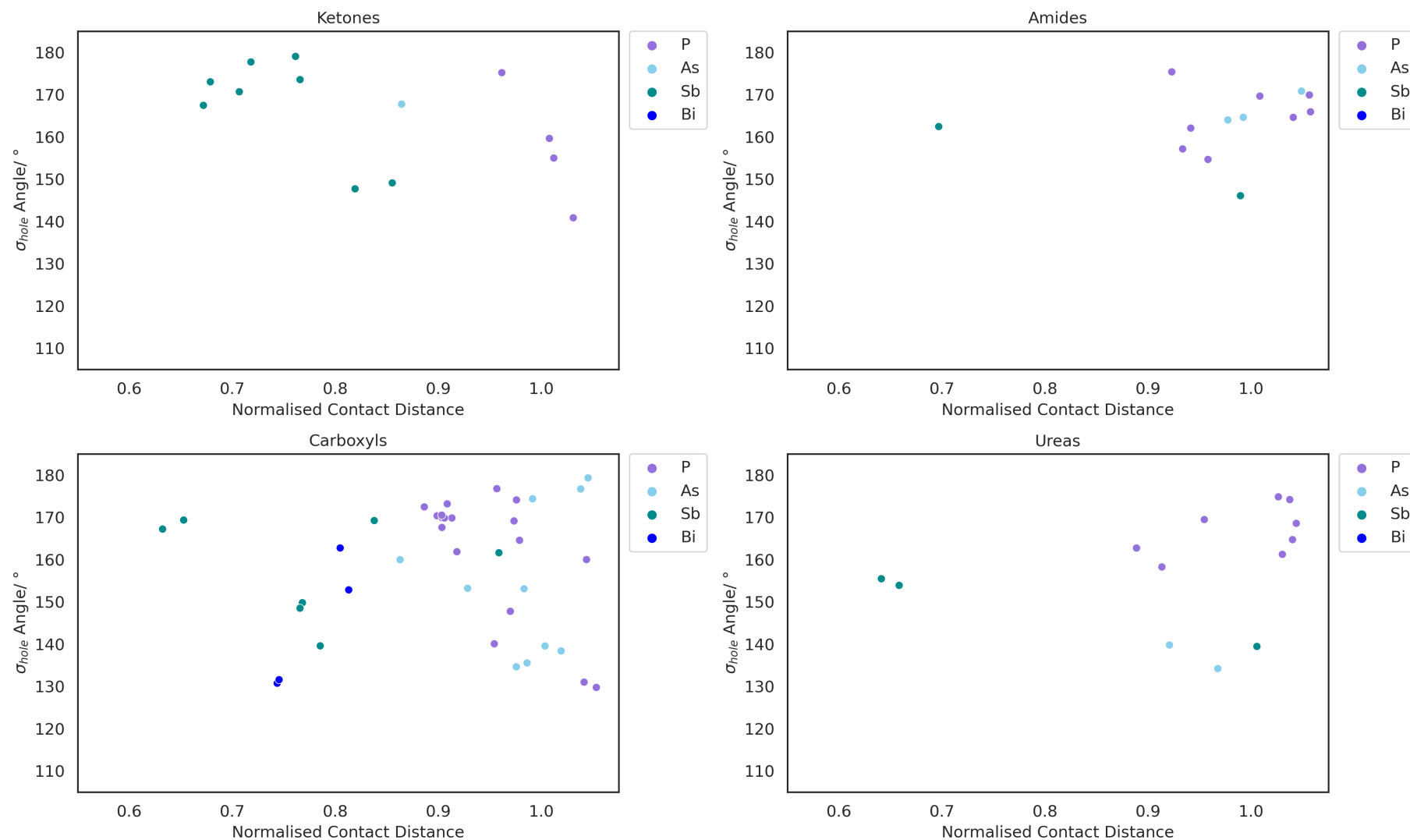


Figure 4.14: Scatter plots showing the distribution of normalised contact distance with respect to the σ -hole angle of pnictogen bonding contacts to organic carbonyls found in crystal structures in the CSD. Points colour coded by donor atom type

4.3.4 *Tt* ... Organic Carbonyl Contacts

As shown in Table 4.1, although not as abundant as halogen or chalcogen bonding, there is a larger number of contacts involving tetrel bonding to organic carbonyls than seen with pnictogen bonding contacts.

Distributions of σ -hole angle against normalised contact distance, shown in Figure 4.15, show a number of very short contacts with σ -hole angles approaching linearity. As seen with analogous pnictogen bonding contacts (see Figure 4.11), a marked proportion of the contacts have normalised contact distances shorter than 0.8. There is a population of tin contacts with very small normalised contact distances that have σ -hole angles with a range between 170-180°, suggesting a very strong tetrel bond.

Looking to the mean averages of the geometric measurements for tetrel bonding contacts to organic carbonyl moieties listed in Table 4.8, the average normalised contact distance decreases as the size of the tetrel atom increases, following the trend $\text{Pb} < \text{Sn} < \text{Ge} < \text{Si}$. Interestingly, the average σ -hole angles decreases with the size of the tetrel atom, following the trend $\text{Pb} < \text{Sn} < \text{Ge} < \text{Si}$. This is unusual as from the polarisabilities of tetrel atoms, see Table 1.4, tin is expected to be the strongest tetrel bond donor atom. Germanium contacts are very rare, but silicon contacts show an interesting characteristic of being rather long, with an average normalised contact distance of 1.008, yet have an average σ -hole angle of 167°.

Distributions of lone-pair angle against normalised contact distance, shown in Figure 4.16, show greater bias among the closest tetrel bonding contacts to lone-pair angles nearer linearity, rather than the expected angle of 120° determined by orbital hybridisation theory. The average tetrel bonding lone-pair angle in Table 4.8 is 134°. Distributions of plane angle against normalised contact distance, displayed in Figure 4.17, shows no trends in short or long contacts with respect to approach of the tetrel bond donor in or out of the plane of the carbonyl moiety.

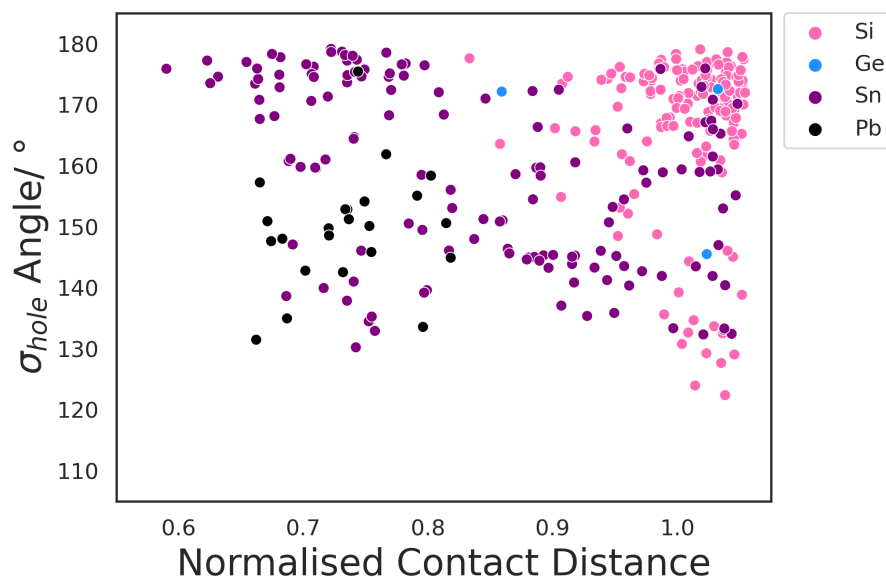


Figure 4.15: Scatter plots showing the distribution of normalised contact distance with respect to the σ -hole angle of tetrel bonding contacts to organic carbonyls found in crystal structures in the CSD. Points colour coded by donor atom type

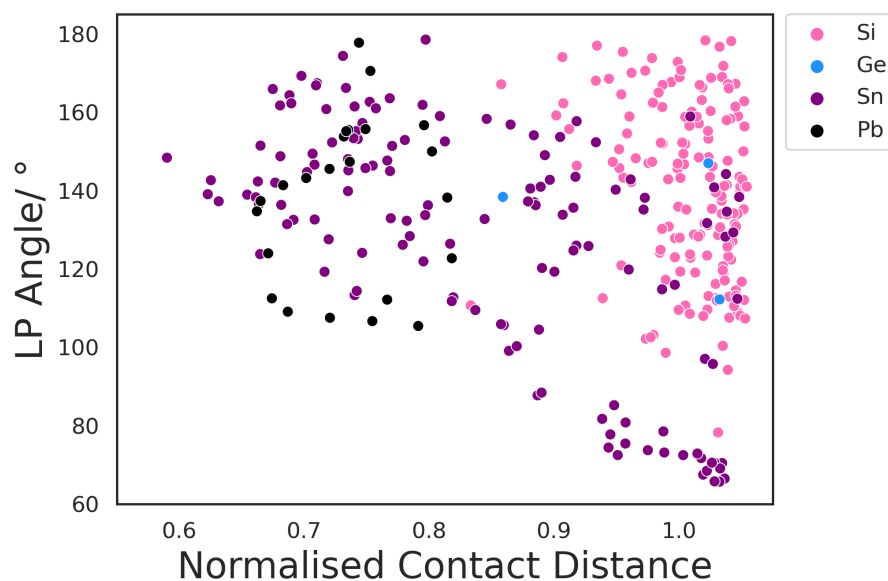


Figure 4.16: Scatter plots showing the distribution of normalised contact distance with respect to the lone-pair angle of tetrel bonding contacts to organic carbonyls found in crystal structures in the CSD. Points colour coded by donor atom type

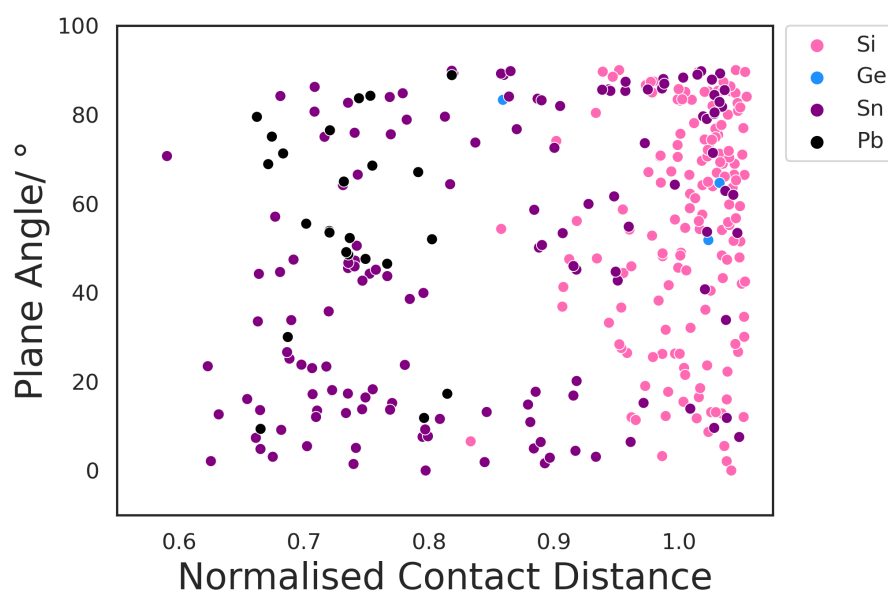


Figure 4.17: Scatter plots showing the distribution of normalised contact distance with respect to the plane angle angle of tetrel bonding contacts to organic carbonyls found in crystal structures in the CSD. Points colour coded by donor atom type

The most abundant and notable carbonyl functional groups are shown in Figure 4.18, which are ketones, amides, carboxyls, ureas, carbamates, and dithiocarbonates. Listed in the mean averages of geometric observables of tetrel bonding contacts for different carbonyl functional groups Table 4.9, the majority of the contacts involve carboxyl groups, and this is the population containing the widest spread of σ -hole angles, especially those involving lead. The geometric characteristics of silicon contacts do not seem to vary notably with the changing of functional groups, remaining very long but linear

with respect to the σ -hole angle. Published reports involving the close contacts involving tin and lead, are predominantly structure determination work, rather than as part of research into noncovalent interactions or any application, highlighting that these interactions are not finding widespread use^{164 165 166}.

Table 4.8: Summary of geometric characteristics of tetrel bonding contacts to organic carbonyl oxygen

Donor	Contacts	Mean Contact Distance/ Å	Mean Normalised Contact Distance	Mean σ -hole Angle/ °	Mean Lone-Pair Angle/ °	Mean Plane Angle/ °
Tt	325	3.484	0.915	162	134	51
Pb	24	3.012	0.735	149	137	56
Sn	141	3.300	0.842	158	128	46
Ge	3	3.683	0.972	163	132	66
Si	157	3.719	1.008	167	139	55

Table 4.9: Summary of geometric characteristics of tetrel bonding contacts to different organic carbonyl functional groups

Functional Group	Contacts	Mean Contact Distance/ Å	Mean Normalised Contact Distance	Mean σ -hole Angle/ °	Mean Lone-Pair Angle/ °	Mean Plane Angle/ °
Organic Carbonyls	320	3.482	0.914	161	134	52
Ketone	53	3.627	0.974	167	144	53
Aldehyde	6	3.708	0.990	162	130	56
Amide	55	3.404	0.896	168	133	60
Carboxyl	159	3.470	0.899	158	127	50
Urea	18	3.475	0.937	165	140	60
Carbamate	11	3.692	0.996	158	155	43
Formamide	7	3.509	0.923	155	134	40
Thioester	2	3.190	0.814	161	148	42
Dithiocarbonate	8	2.864	0.738	162	160	30

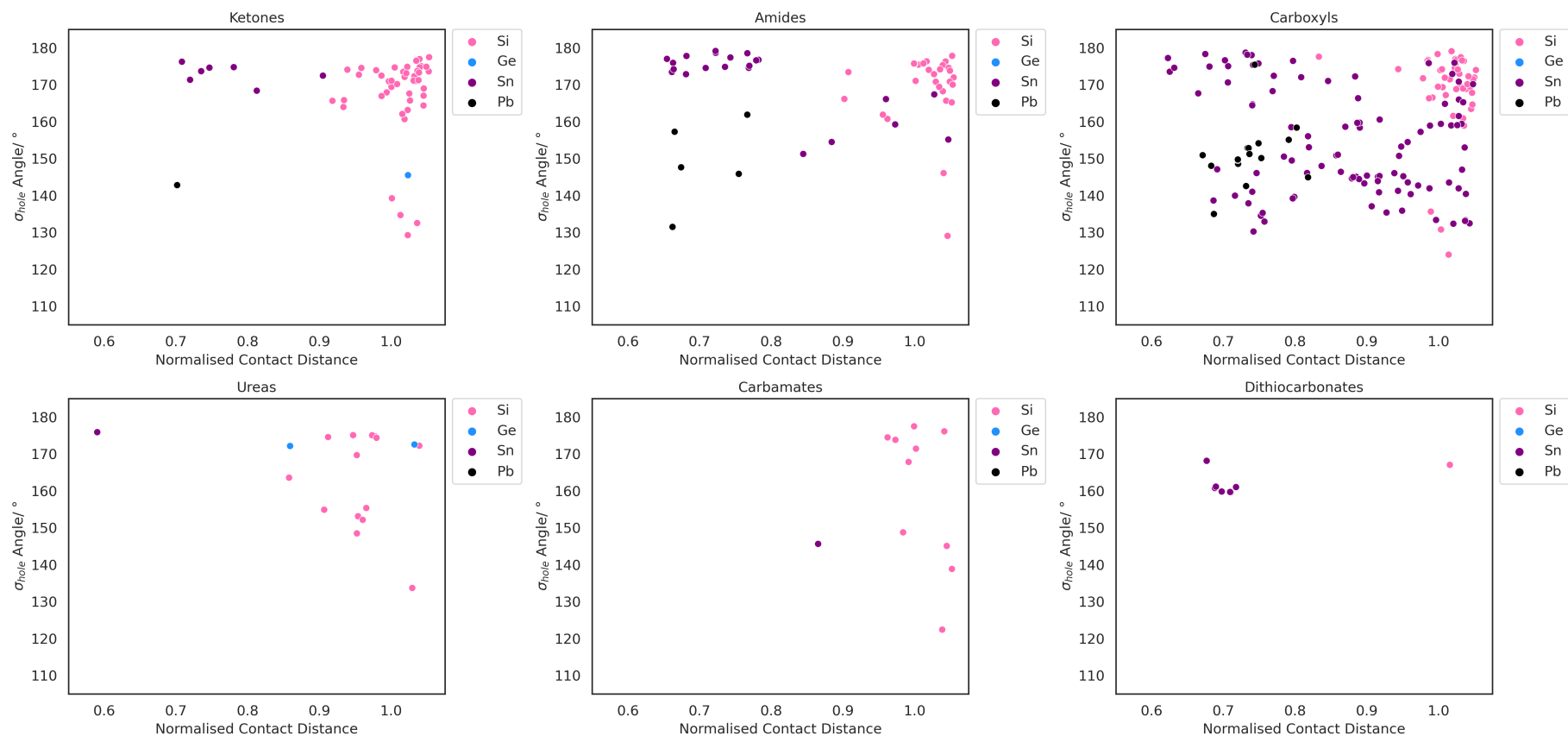


Figure 4.18: Scatter plots showing the distribution of normalised contact distance with respect to the σ -hole angle of tetrel bonding contacts to organic carbonyls found in crystal structures in the CSD. Points colour coded by donor atom type

4.3.5 Prevalence of $\sigma_{hole} \cdots$ Organic Carbonyl Oxygen Contacts

The prevalence of σ -hole interactions to organic carbonyls is detailed in Table 4.10. Crystal structures containing halogen bond donor and organic carbonyl moieties are the most numerous, along with containing the most contacts, thus the prevalence is 19.49%. Chalcogen bonding contacts have very similar prevalence, despite being less than half as numerous. Comparatively, pnictogen and tetrel bonding are much less prevalent, with prevalences of 10.03% and 4.68%, respectively. The prevalence of halogen and chalcogen bonding contacts, along with the close contacts with near-linear σ -hole angles shown in Figure 4.3 and Figure 4.7, suggests that these interactions are widely exploited in synthesis. In contrast, the low prevalences of pnictogen and tetrel bonding contacts suggests that these interactions are more likely to be incidental or proof-of-concept structures, rather than being deliberately used for a consistent application.

Table 4.10: Overall summary of the prevalence of contacts indicating σ -hole interactions to organic carbonyl oxygens compared to the number of unique crystal structures found in the CSD that contain the necessary moieties.

σ -hole Interaction Type	Structures with Moieties	Structures with Contact(s)	Prevalence
Halogen-bonding	43734	8518	19.48%
Chalcogen-bonding	17771	3682	21.73%
Pnictogen-bonding	648	65	10.03%
Tetrel-bonding	4508	211	4.68%

When examining the prevalence of halogen bonding contacts to organic carbonyl oxygens with respect to the different halogen bond donor atoms, shown in Table 4.11, the prevalence of the contacts increases with the polarisability of the halogen. There are considerably fewer structures containing iodine and organic carbonyl moieties than other halogen species, with only 3266 structures, yet 30.74% of these structures contain one or more halogen bonding contacts. This suggests that $I \cdots O=CR_2$ halogen bonding motifs are widely exploited when synthesising crystal structures. When separating the prevalence of halogen bonding contacts to organic carbonyl oxygens into more specific functional groups, presented in Table 4.12, shows fairly consistent prevalence between the most numerous carbonyl functional groups. The three moieties that make up the majority of the interactions; ketones, carboxyls, and amides, the prevalence is very similar - around 16% for each of the groups. Contacts involving aldehydes have a slightly higher prevalence of 22.96%, which along with the somewhat strong, directional contacts shown in Figure 4.6, suggests a potentially useful but somewhat underutilised halogen bonding motif. The prevalence of the less commonly occurring functional groups is harder to interpret due to the scarcity of data, making their comparison less reliable.

Table 4.11: Summary of the prevalence of halogen bonding contacts to organic carbonyls compared to the number of unique crystal structures found in the CSD that contain the necessary moieties

Donor	Structures with Moieties	Structures with Contact(s)	Prevalence
X	43734	8518	19.48%
I	3266	1004	30.74%
Br	15510	3030	19.54%
Cl	26385	4534	17.18%

Table 4.12: Summary of the prevalence of halogen bonding contacts to various carbonyl functional groups compared to the number of unique crystal structures found in the CSD that contain the necessary moieties

Functional Group	Structures with Moieties	Structures with Contact(s)	Prevalence
Organic Carbonyls	43734	8518	19.48%
Carboxyl	19283	3139	16.28%
Amide	15868	2532	15.96%
Ketone	12508	2083	16.65%
Urea	1936	317	16.37%
Carbamate	2190	271	12.37%
Aldehyde	932	214	22.96%
Formamide	503	62	12.33%
Dicarbonate	126	19	15.08%
Thioester	149	33	22.15%
Carbamothioate	80	13	16.25%
Formate	35	5	14.29%
Acid Chloride	28	16	57.14%
Dithiocarbonate	19	7	36.84%
Carbonothioate	10	0	N/A
Other	74	17	22.97%

When examining the prevalence of chalcogen bonding contacts to organic carbonyl oxygens with respect to the different chalcogen bond donor atoms, shown in Table 4.13, the prevalence of the contacts increases with the polarisability of the chalcogen. In contrast to the halogen bonding prevalence in Table 4.13, the prevalence of tellurium contacts (41.33%) is out of a very small pool of tellurium and organic carbonyl containing structures. Sulfur and selenium contacts have fairly similar prevalences, yet the sulfur contacts make up around 94% of the contacts and around 95% of the moiety-containing crystals.

The prevalence of sulfur contacts to organic carbonyls, along with the very close sulfur contacts to ketones in Figure 4.10, suggests that this chalcogen bonding motif shows notable use in crystal engineering. Likewise, the prevalence of selenium contacts to organic carbonyls, along with the close contacts to amides shown in Figure 4.10, suggests the use of this

motif specifically for its chalcogen bonding ability. In terms of chalcogen bonding to different organic carbonyl functional groups, see Table 4.14, there are some notably high prevalences among chalcogen bonding contacts with a non-negligible number of contacts. Functional groups such as thioesters, carbamothioates, dithiocarbonates, and carbonothioates show very high prevalence, and all contain sulfur as one of the substituents on the carbonyl. This suggests that these structures may contain chalcogen bonding interactions where the sulfur bonded to the carbonyl carbon could be acting as a σ -hole donor.

Table 4.13: Summary of the prevalence of chalcogen bonding contacts to organic carbonyls compared to the number of unique crystal structures found in the CSD that contain the necessary moieties

Donor	Structures with Moieties	Structures with Contact(s)	Prevalence
Ch	17771	3862	21.73%
Te	75	31	41.33%
Se	801	192	23.97%
S	16946	3645	21.51%

Table 4.14: Summary of the prevalence of chalcogen bonding contacts to various carbonyl functional groups compared to the number of unique crystal structures found in the CSD that contain the necessary moieties

Functional Group	Structures with Moieties	Structures with Contact(s)	Prevalence
Organic Carbonyls	17771	3862	21.73%
Carboxyl	7444	1291	17.34%
Amide	6701	1213	18.10%
Ketone	4265	765	17.94%
Urea	758	108	14.25%
Thioester	348	156	44.83%
Carbamate	803	65	8.09%
Formamide	335	48	14.33%
Aldehyde	346	66	19.08%
Carbamothioate	161	91	56.52%
Dithiocarbonate	116	104	89.66%
Carbonothioate	21	20	95.24%
Dicarbonate	33	2	6.06%
Formate	6	1	16.67%
Acid Chloride	5	0	N/A
Other	23	19	82.61%

The prevalence of pnictogen bonds to organic carbonyl moieties, shown in Table 4.15, correspond to primarily phosphorus contacts, which have prevalence lower than that of the overall pnictogen bond to organic carbonyl prevalence. The phosphorus contacts in Figure 4.11 are also very long and have smaller σ -hole angles. The closer contacts involving

the heavier pnictogens such as antimony have much higher prevalence, yet the crystals containing the required moieties are rare. In terms of functional groups, shown in Table 4.16, the divisions become even more sparse, highlighting the comparative lack of study of pnictogen bonding interactions to organic carbonyls compared to halogen and chalcogen bonding interactions with organic carbonyls.

Table 4.15: Summary of the prevalence of pnictogen bonding contacts to organic carbonyls compared to the number of unique crystal structures found in the CSD that contain the necessary moieties

Donor	Structures with Moieties	Structures with Contact(s)	Prevalence
Pn	648	65	10.03%
Bi	13	5	38.46%
Sb	24	12	50.00%
As	43	9	20.93%
P	568	39	6.87%

Table 4.16: Summary of the prevalence of pnictogen bonding contacts to various carbonyl functional groups compared to the number of unique crystal structures found in the CSD that contain the necessary moieties

Functional Group	Structures with Moieties	Structures with Contact(s)	Prevalence
Organic Carbonyls	648	65	10.03%
Carboxyl	281	24	8.54%
Amide	171	11	6.43%
Ketone	98	8	8.16%
Urea	42	8	19.05%
Aldehyde	32	3	9.38%
Carbamate	20	1	5.00%
Formamide	9	1	11.11%
Dicarbonate	3	0	N/A
Thioester	2	1	50.00%
Dithiocarbonate	1	1	100.00%
Formate	2	0	N/A
Other	28	7	25.00%

The prevalence of tetrel bonding contacts to organic carbonyl moieties, shown in Table 4.17, reveals the disparity between the prevalences of the different tetrel bond donor atoms. The majority of the contacts involve silicon, which has a very low prevalence, suggesting its unlikely use in tetrel bonding motifs. Germanium contacts show even lower prevalence, with only three structures containing $\text{Ge} \cdots \text{C}=\text{CR}_2$ contacts out of 125 moiety containing structures. Tin contacts show a notably higher prevalence with the second most contacts, and lead contacts show a very high prevalence yet it is out of a very small pool of lead and organic carbonyl containing structures. In terms of the tetrel bonding contacts to specific carbonyl functional groups, shown in Table 4.18, there aren't any notable functional groups that have a larger number of contacts

that have high prevalence.

Table 4.17: Summary of the prevalence of tetrel bonding contacts to organic carbonyls compared to the number of unique crystal structures found in the CSD that contain the necessary moieties

Donor	Structures with Moieties	Structures with Contact(s)	Prevalence
Tt	4508	211	4.68%
Pb	23	15	65.22%
Sn	239	66	27.62%
Ge	125	3	2.40%
Si	4142	127	3.07%

Table 4.18: Summary of the prevalence of tetrel bonding contacts to various carbonyl functional groups compared to the number of unique crystal structures found in the CSD that contain the necessary moieties

Functional Group	Structures with Moieties	Structures with Contact(s)	Prevalence
Organic Carbonyls	4508	211	4.68%
Carboxyl	2046	80	3.91%
Ketone	1313	48	3.66%
Amide	1085	44	4.06%
Carbamate	446	11	2.47%
Urea	158	9	5.70%
Aldehyde	127	6	4.72%
Formamide	35	7	20.00%
Dicarbonate	37	2	5.41%
Thioester	17	2	11.76%
Dithiocarbonate	7	3	42.86%
Formate	8	0	N/A
Carbonothioate	3	0	N/A
Carbamothioate	7	1	14.29%
Other	18	1	5.56%

4.3.6 Notable Organic Carbonyl Moieties

During the data analysis of this chapter, some notable σ -hole interactions motifs were noticed, some familiar and covered in Chapter 3, and some new and interesting crystal structures.

In terms of halogen bonding structures, much remains the same, with N-iodosuccinimides which contain two carbonyls as part of the five-membered ring, making up the shortest more linear contacts seen in the amide plot in Figure 4.6. The closest contact in the carboxyl plot in in Figure 4.6 refers to CSD refcode **NUZHIC**, which involves halogen bonded triazolium salts¹⁶⁷. The closest formamide contact is the crystal structure determination of N,N-diiodoformamide from

1974, which is interesting to see a strongly halogen bonding motif dating from long before the definition of halogen bonding¹⁶⁸. The very close yet off-linear chlorine contact seen in the ketone plot of Figure 4.6 is a coincidental close contact due to the packing of the crystal structure in CSD refcode **NAFBUR** and the associated interactions, there is another longer halogen bonding contact that matches the σ -hole criteria better¹⁶⁹. Figures depicting these two contacts found in **NAFBUR** are included in Figure B.1.

For chalcogen bonding motifs, the very close yet off-linear contacts involving sulfur are a set of contacts corresponding to three structures with CSD refcodes; **EPIVEF**, **FIDYUP**, and **LILRIJ**^{170–172}. The structures involve bisdithiazolyl molecules and exhibit interesting properties, such as semiconductivity and antiferromagnetism. As this bisdithiazolyl is a chalcazole-like derivative, there is a competing chalcogen bond interaction involving the nitrogen in the heterocycle, which distorts the linearity of both interactions, but results in a very close set of contacts. This is highlighted in Figure 4.19, which shows the molecular configuration of **FIDYUP**.

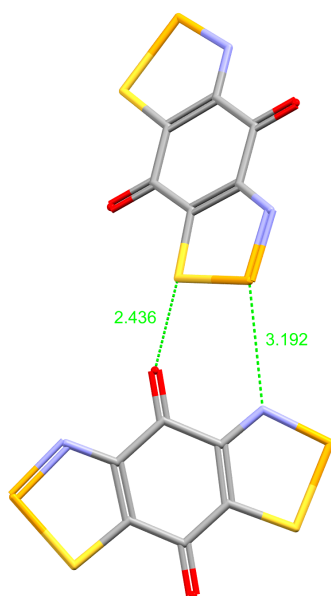


Figure 4.19: An example of a crystal structure with a chalcogen bonding contact involving a bisdithiazolyl molecule as a chalcogen bond donor and acceptor. CSD refcode: **FIDYUP**

Interestingly, similar chalcazole like motifs are seen involving selenium, which make up the close contacts with σ -hole angles near 180° in the amide plot of Figure 4.10. These correspond to isoselenazolone-based structures, which are analogues of an organoselenium drug used for its anti-inflammatory, anti-oxidant, and cytoprotective effects distributed under the generic name Ebselen¹⁷³. These chalcazole contacts likely have σ -hole angles closer to linearity than seen in the chalcazole contacts mentioned in Section 3.4.6 due to the nitrogen in the heterocycle not having an available lone-pair¹⁷⁴. An example of an Ebselen analogue is illustrated in Figure 4.20 which corresponds to CSD refcode **EDIHEF**.

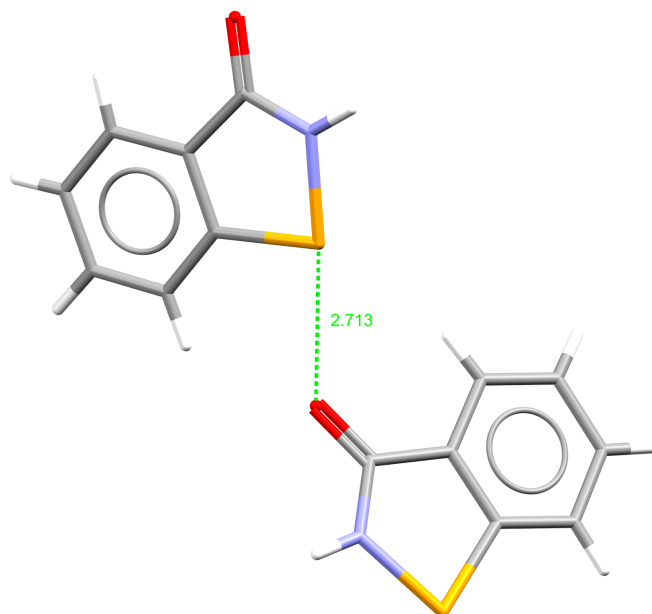


Figure 4.20: An example of a crystal structure with a chalcogen bonding contact involving an Ebselen analogue as an chalcogen bond donor and acceptor. CSD refcode: **EDIHEF**

The majority of the close pnictogen bonding structures involve antimony trifluoride, also known as Swart's reagent. This compound has use in inorganic and organofluorine chemistry as a fluorinating agent. The crystals involving complexes between SbF_3 and carbonyl compounds are mostly structure determination papers, an example is shown in Figure 4.21 of **UREAAF**¹⁷⁵.

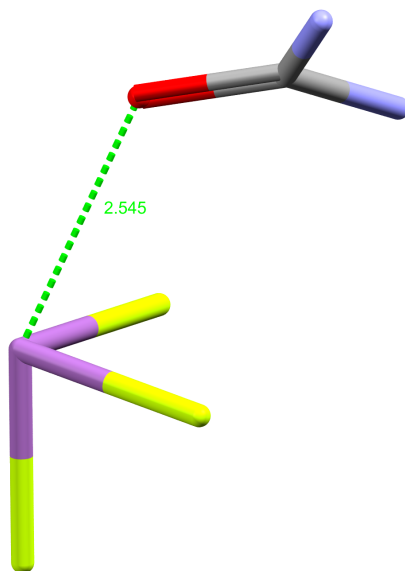


Figure 4.21: An example of a crystal structure with a pnictogen bonding contact involving SbF_3 as an pnictogen bond donor and urea as an acceptor. CSD refcode: **UREAAF**

Some of very close, tetrel bonding contacts with σ -hole angles near linearity involve organotin(IV) complexes, which are

of interest for their anti-tumour activity along with other potential uses in various fields of chemistry¹⁷⁶. These feature a tin ester or thioester and an imide component, which is highlighted in Figure 4.22, showing CSD refcode **AGEPOS**¹⁷⁷.

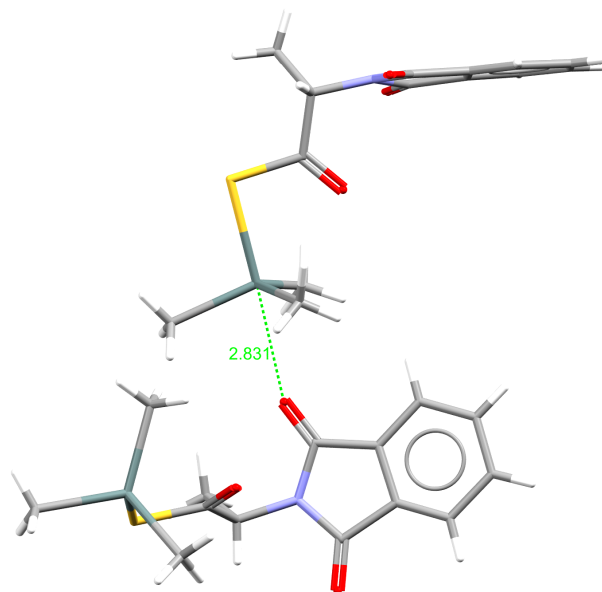


Figure 4.22: An example of a crystal structure with a tetrel bonding contact involving a tin thioester structure as an tetrel bond donor. CSD refcode: **AGEPOS**

4.4 Divalent Oxygen

Given the breadth of functional groups this subdivision encompasses, it only seems appropriate to refer to the group of σ -hole acceptors (R-O-R) as divalent oxygens. These groups, of the form R_1OR_2 , covers not just sp^3 -hybridised species such as water or alcohols, but also sp^2 -hybridised species such as furans. These can be distinguished from sp^3 -hybridised ethers by examining the valency of the R_1 and R_2 groups either side of the ether, to distinguish further. The R_1OR_2 query where $R_1, R_2 = C$ can be divided into imperfect subdivisions based on the hybridisation of each of the carbon atoms, with ethers being both sp^3 -hybridised and furans being both sp^2 -hybridised. This is imperfect due to the small number of anhydrides such as phthalic anhydride, being present in the furan subdivision due to meeting the hybridisation criteria, but there is a very small number of crystal structures containing anhydrides, let alone those involving σ -hole interactions. The ethers with mixed hybridisation, dubbed the "Mixed Ether/Ester" grouping, proves to be more problematic, as it contains both ester oxygens and aromatic systems with an alkoxy functional group such as alkyl phenyl ethers, along with a likely plethora of less common functional groups that have differing hybridisations around an ether moiety.

To hypothesise possible trends in the effectiveness of these divalent oxygen moieties to accept σ -hole interactions, we look to the work of Rauk *et al.* which is an *ab initio* study of the proton and BF_3 affinities of oxygen containing compounds¹⁷⁸. The work of Rauk also involves carbonyl compounds so it offers an interesting comparison with the previous section. This, along with the studies of Ventura and co-workers, suggests that water, alcohols, and sp^3 -hybridised ethers are less effective

Lewis bases than carbonyl compounds¹⁷⁹. The work of Rauk suggests that ring strain can make ethers more Lewis basic, yet furan is a much less effective Lewis base, likely due to the delocalisation of the aromatic system.

An overall summary of the geometric observables of contacts indicative of σ -hole interactions to divalent oxygen moieties is tabulated in Table 4.19.

Table 4.19: Overall geometric summary of the contacts indicating σ -hole interactions to divalent oxygen moieties

σ -hole Interaction Type	No. Contacts	Mean Normalised Contact Distance	Mean σ -hole Angle/ °	Mean Lone-Pair Angle A/ °	Mean Lone-Pair Angle B/ °	Mean Plane Angle/°
Halogen-bonding	10006	0.990	149	106	110	58
Chalcogen-bonding	3821	0.998	149	104	107	56
Pnictogen-bonding	185	0.926	153	108	105	59
Tetrel-bonding	252	0.976	152	95	102	50

4.4.1 X ... Divalent Oxygen Contacts

Out of the contacts summarised in Table 4.19, halogen bonding contacts are the most numerous, with a greater number than chalcogen, pnictogen, and tetrel bonding contacts to organic carbonyl moieties combined.

The distribution of σ -hole angles against normalised contact distance for halogen bonding contacts to organic carbonyl moieties, presented in Figure 4.23, shows the characteristic "wedge" shape, albeit with a broader spread than seen with halogen bonding contacts to organic carbonyls (see Figure 4.3). There can be seen an incongruous collected of contacts that are shorter than 0.8 normalised contact distance and they comprise nearly entirely of iodine contacts. This cluster of iodine contacts around 0.7 normalised contact distance, which correspond to a high-pressure study of the co-crystal of strong halogen-bond donor ICl with 1,4-dioxane, see CSD refcode **WINLAK**¹⁸⁰. There is also a number of close chlorine contacts that have σ -hole angles that deviate substantially from linearity. These contacts can be attributed to contacts involving alcohols and water, shown in the distributions of σ -hole angle against normalised contact distance for individual divalent oxygen functional groups in Figure 4.26. The contacts are likely due to hydrogen bonding interactions, as both alcohols and water are hydrogen bond donors and acceptors, and chlorine is also an acceptor of hydrogen bonds⁸⁸.

The mean averages of the halogen bonding contacts to divalent oxygen, are listed in Table 4.20. The average σ -hole angle increases with more polarisable halogen bond donor atoms, but not by a large magnitude, with iodine contacts having an average σ -hole angle only 8° larger than chlorine contacts. The average normalised contact distance decreases with more polarisable halogen bond donor atoms, following the trend $I < Br < Cl$. The lone-pair angles deviate little with change of identity of the halogen bond donor atom. The average plane angle shows no trend with respect to the identity of the halogen atom, with very little deviation from average.

The distributions of the two lone-pair angles against the normalised contact distance are shown in Figure 4.24. These plots show a clear preference for lone-pair angles nearing 120° as the normalised contact distance decreases. The distributions of plane angle against normalised contact distance, see in Figure 4.25, shows some preference among the closest contacts

for an orthogonal approach of the contact with respect to the plane of the divalent oxygen molecule.

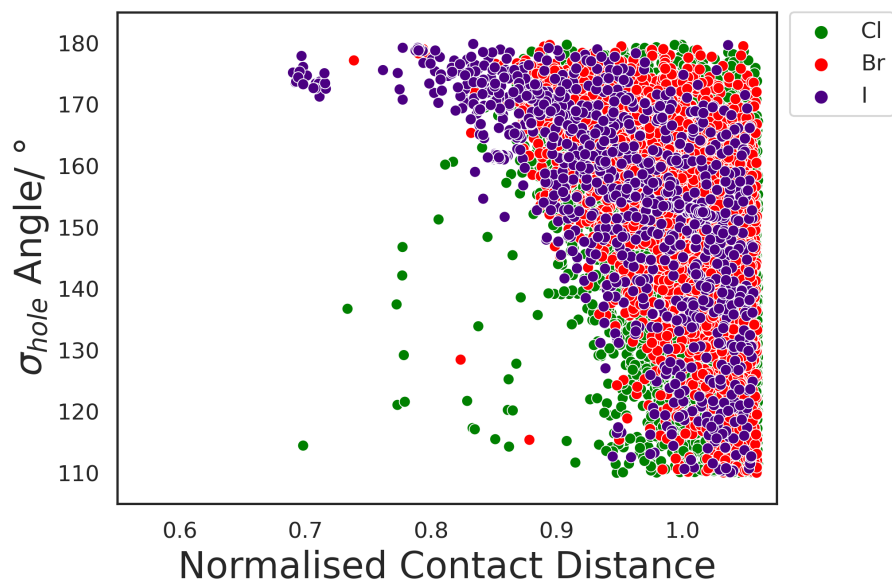


Figure 4.23: Scatter plots showing the distribution of normalised contact distance with respect to the σ -hole angle of halogen bonding contacts to divalent oxygens found in crystal structures in the CSD. Points colour coded by donor atom type

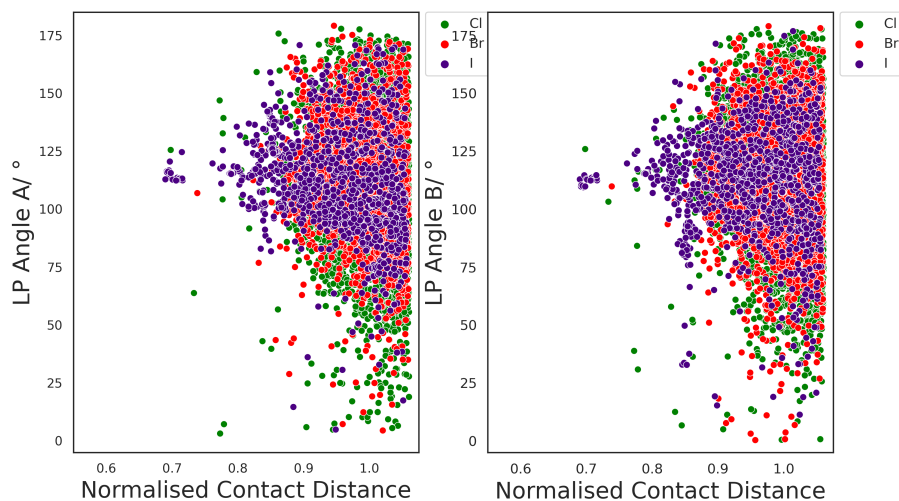


Figure 4.24: Scatter plots showing the distribution of normalised contact distance with respect to both lone-pair angles of halogen bonding contacts to divalent oxygens found in crystal structures in the CSD. Points colour coded by donor atom type

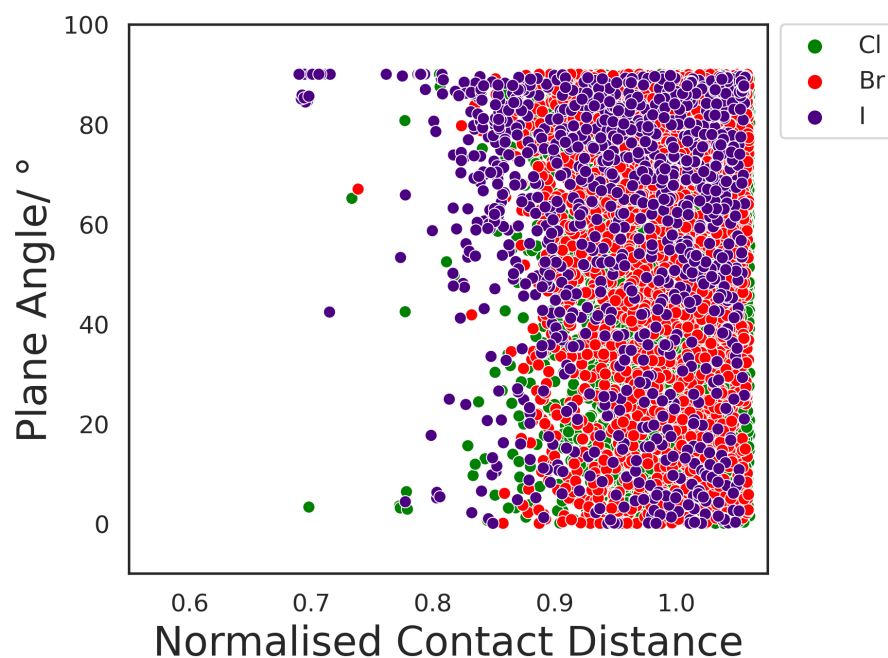


Figure 4.25: Scatter plots showing the distribution of normalised contact distance with the plane angle of halogen bonding contacts to divalent oxygens found in crystal structures in the CSD. Points colour coded by donor atom type

Table 4.20: Summary of geometric characteristics of halogen bonding contacts to divalent oxygen

Donor	Contacts	Mean Contact Distance/ Å	Mean Normalised Contact Distance	Mean σ -hole Angle/ °	Mean Lone-Pair Angle A/ °	Mean Lone-Pair Angle B/ °	Mean Plane Angle/ °
X	10006	3.319	0.990	149	106	110	58
I	1057	3.355	0.948	154	111	111	59
Br	2830	3.320	0.988	152	108	111	57
Cl	6119	3.313	0.998	146	104	110	58

The trends among different divalent oxygen functional groups were examined by subdivision of the data presented in Figure 4.23. The distributions of σ -hole angle against normalised contact distance for the different functional groups are presented in Figure 4.26, and mean averages of the geometric measurements are listed in Table 4.21.

These functional groups are subdivided from R_1OR_2 , and are as follows; ethers ($R_1, R_2 = C_{sp^3}, C_{sp^3}$), furans ($R_1, R_2 = C_{sp^2}, C_{sp^2}$), mixed ether/ester ($R_1, R_2 = C_{sp^3}, C_{sp^2}$), alcohols ($R_1, R_2 = C, H$), waters ($R_1, R_2 = H, H$), oxyphosphoranes ($R_1, R_2 = C, P$), oxylamines ($R_1, R_2 = C, N$), aminoboronates ($R_1, R_2 = B, N$), boronic esters ($R_1, R_2 = B, C$), hydroxyimines ($R_1, R_2 = H, N$), sulfonates ($R_1, R_2 = C, S$), silicates ($R_1, R_2 = C, Si$), phosphinous acids ($R_1, R_2 = H, P$), furazans ($R_1, R_2 = N, N$), boronic acids ($R_1, R_2 = B, H$), and peroxides ($R_1, R_2 = C, O$). The naming of these functional groups are chosen as a best representation of what this moiety describes, and in many case does not exclusively contain the named functional group.

The most frequently occurring contacts involve, unsurprisingly, mixed ether/esters, alcohols, ethers, furans, and water. It is clear from the mixed ether/ester and furan data that conjugation reduces the effectiveness of the oxygen as an acceptor

for halogen bonding, with the contacts being longer and have larger σ -hole angles compared to contacts involving ethers and alcohols. There is an notable number of contacts involving oxyphosphoranes, which predominantly involve chlorine as the σ -hole donor and have noticeable directionality with respect to the σ -hole angle. A similar distribution is seen with analogous oxylamine acceptors, but with a greater proportion of bromine and iodine contacts.

Table 4.21: Summary of geometric characteristics of halogen bonding contacts to different divalent oxygen functional groups

Functional Group	Contacts	Mean Contact Distance/ Å	Mean Normalised Contact Distance	Mean σ -hole Angle/ °	Mean Lone-Pair Angle A/ °	Mean Lone-Pair Angle B/ °	Mean Plane Angle/ °
Divalent Oxygen	9929	3.319	0.990	148	106	110	58
Mixed Ether/Ester	3565	3.333	0.995	148	103	114	61
Alcohol	2569	3.308	0.985	149	108	106	53
Ether	1358	3.271	0.973	153	111	115	61
Furan	625	3.350	1.000	144	110	108	61
Water	557	3.346	0.996	145	97	97	44
Oxyphosphorane	401	3.317	0.998	146	110	115	60
Oxylamine	202	3.340	1.000	148	115	112	50
Aminoboronate	191	3.316	0.997	144	96	106	70
Hydroxyimino	111	3.291	0.983	146	106	111	50
Boronic Ester	104	3.391	1.005	151	113	112	60
Sulfonate	61	3.358	1.007	140	112	109	61
Silicate	50	3.333	1.000	146	106	107	66
Phosphinous Acid	46	3.279	0.977	155	100	121	61
Furazan	45	3.363	1.003	137	113	113	42
Boronic Acid	31	3.434	1.020	147	88	104	67
Peroxide	13	3.300	0.985	155	114	114	64

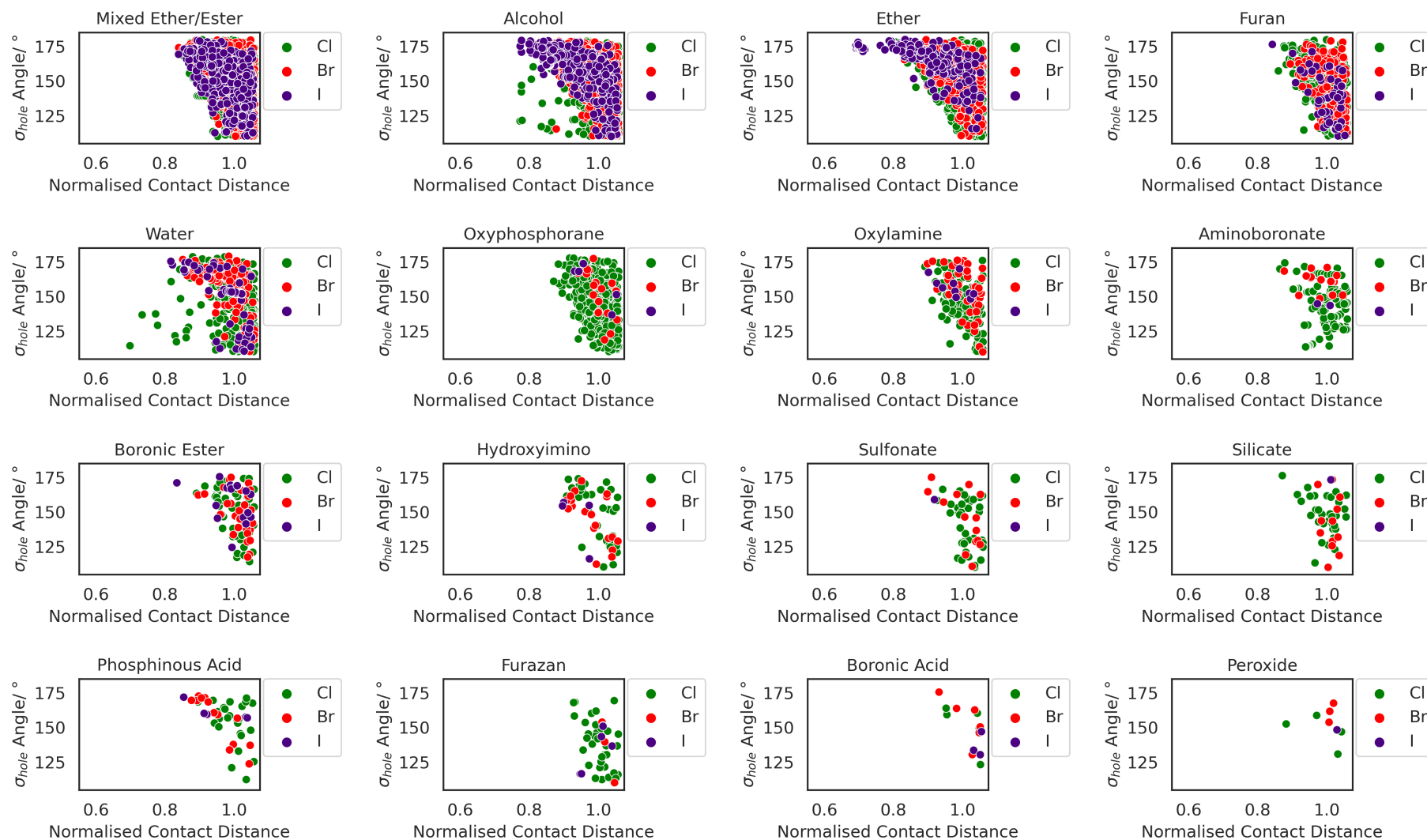


Figure 4.26: Distributions with respect to σ -hole angle of normalised halogen bonding contacts to divalent oxygen functional groups found in crystal structures in the CSD. Points colour coded by donor atom type

4.4.2 Ch ... Divalent Oxygen Contacts

Table 4.19 shows that chalcogen bonding contacts are the second most numerous σ -hole contact to divalent oxygen Lewis bases.

The distribution of σ -hole angles against normalised contact distance, shown in Figure 4.27, has a similar profile to those seen for analogous halogen bonding contacts (see Figure 4.23). The main difference is the lack of a population of near linear contacts around 0.7 normalised contact distance. There are a small number of close sulfur contacts that deviate substantially from linearity, suggesting a similar role to chlorine mentioned in Section 4.4.1. Sulfur and selenium contacts show similar distributions of σ -hole angle, with the less numerous selenium contacts having proportionally fewer σ -hole angles that deviate substantially from linearity. The much rarer tellurium contacts have very few σ -hole angles greater than 170° .

The mean averages of geometric measurements for chalcogen bonding contacts to divalent oxygen Lewis bases is listed in Table 4.22. Sulfur contacts make up the vast majority of the contacts, and as such mirror the averages across the chalcogens. The average normalised contact distance decreases for more polarisable chalcogen bond donor atoms, following the trend $\text{Te} < \text{Se} < \text{S}$. The largest average σ -hole angles are seen in selenium contacts, followed by tellurium contacts and then sulfur contacts. In contrast to the chalcogen bonding contacts to organic carbonyl acceptors, see Table 4.4, the average σ -hole angles are smaller across all donors. Both the lone-pair angles increase with greater polarisability of the chalcogen bond donor, but are much smaller than the lone-pair angles for the chalcogen bonding contacts to organic carbonyls, see Table 4.4.

The distribution of the two lone-pair angles against normalised contact distance is shown in Figure 4.28. While there is a slight tendency for contacts with shorter normalised contact distances to have lone-pair angles nearer 120° the spread is much broader than seen with the analogous halogen bonding contacts, see Figure 4.24. The distribution of plane angle against the normalised contact distance, see Figure 4.29, shows that selenium and tellurium contacts show a preference to orthogonality to the plane of the divalent oxygen moiety, but sulfur contacts show little to no preference.

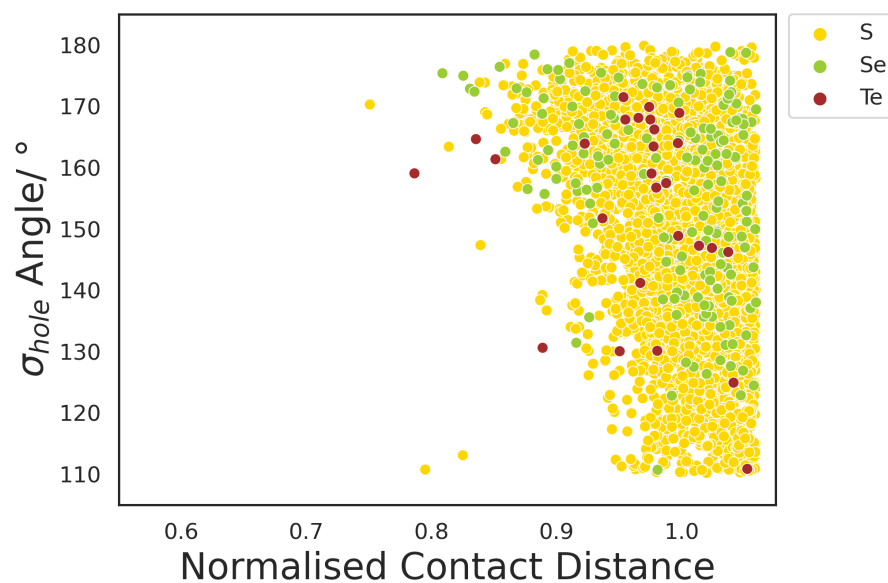


Figure 4.27: Scatter plots showing the distribution of normalised contact distance with respect to the σ -hole angle of chalcogen bonding contacts to divalent oxygens found in crystal structures in the CSD. Points colour coded by donor atom type

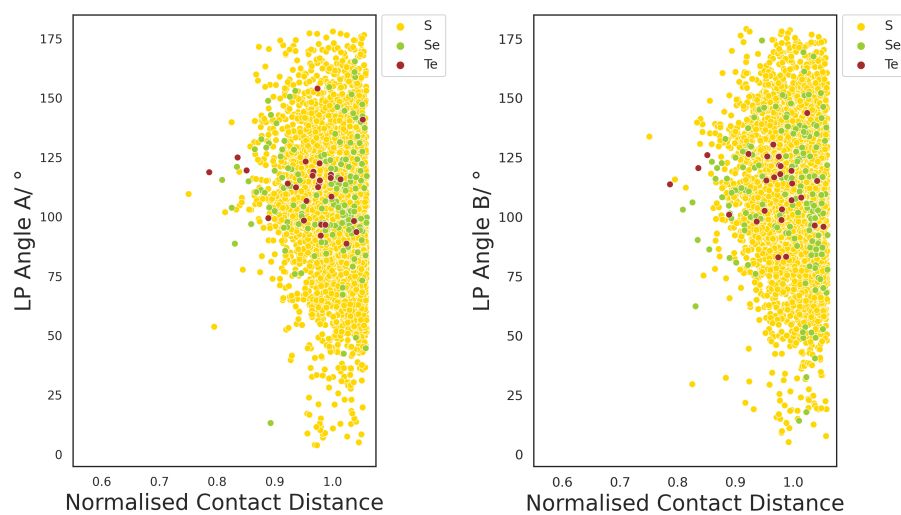


Figure 4.28: Scatter plots showing the distribution of normalised contact distance with respect to both lone-pair angles of chalcogen bonding contacts to divalent oxygens found in crystal structures in the CSD. Points colour coded by donor atom type

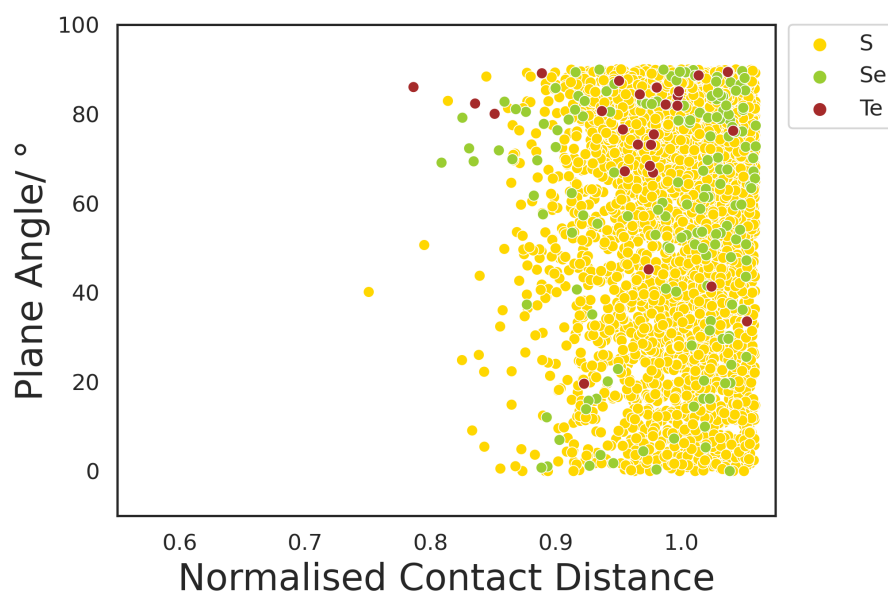


Figure 4.29: Scatter plots showing the distribution of normalised contact distance with the plane angle of chalcogen bonding contacts to divalent oxygens found in crystal structures in the CSD. Points colour coded by donor atom type

Table 4.22: Summary of geometric characteristics of chalcogen bonding contacts to divalent oxygen

Donor	Contacts	Mean Contact Distance/ Å	Mean Normalised Contact Distance	Mean σ -hole Angle/ °	Mean Lone-Pair Angle A/ °	Mean Lone-Pair Angle B/ °	Mean Plane Angle/ °
Ch	3821	3.381	0.998	149	104	107	56
Te	27	3.361	0.963	153	112	112	73
Se	163	3.281	0.988	156	109	108	58
S	3631	3.386	0.999	149	103	107	55

The trends among different divalent oxygen functional groups were examined by subdivision of the data presented in Figure 4.27. The distributions of σ -hole angle against normalised contact distance for the different functional groups are presented in Figure 4.30, and mean averages of the geometric measurements are listed in Table 4.23.

The most directional and close contacts are those involving water and alcohols as acceptors, with alcohol contacts having a greater number of selenium contacts, yet the contacts involving water show similar closeness but with a greater proportion of sulfur contacts. The majority of the tellurium contacts are found in the ether and mixed ether/ester subgroups, with similar σ -hole angles in each, but being closer contacts with ether ($R_1, R_2 = C_{sp^3}, C_{sp^3}$) Lewis bases rather than mixed ester/ether Lewis bases ($R_1, R_2 = C_{sp^3}, C_{sp^2}$).

There are far fewer close contacts involving sulfur with σ -hole angles that deviate substantially from linearity than the analogous chlorine contacts mentioned in Section 4.4.1 which were associated with contacts involving water and alcohols. These contacts were suggested to be due to the formation of hydrogen bonds within the crystal, and potentially the chlorine as a hydrogen bond acceptor. This could explain the lack of contacts as sulfur is a much weaker hydrogen bond acceptor than chlorine or oxygen due to its reduced electronegativity¹⁸¹.

Table 4.23: Summary of geometric characteristics of chalcogen bonding contacts to different divalent oxygen functional groups

Functional Group	Contacts	Mean Contact Distance/ Å	Mean Normalised Contact Distance	Mean σ -hole Angle/ °	Mean Lone-Pair Angle A/ °	Mean Lone-Pair Angle B/ °	Mean Plane Angle/ °
Divalent Oxygen	3799	3.382	0.998	149	104	107	56
Mixed Ether/Ester	1279	3.391	1.001	148	106	112	61
Alcohol	1141	3.372	0.995	150	102	105	50
Water	578	3.388	1.000	151	96	96	50
Ether	362	3.343	0.987	153	111	113	62
Furan	181	3.420	1.010	146	107	107	65
Oxylamine	57	3.399	1.003	150	112	102	58
Phosphinous Acid	44	3.333	0.983	147	94	115	52
Oxyphosphorane	34	3.395	1.002	143	97	112	69
Hydroxyimino	33	3.403	1.004	146	113	99	55
Boronic Ester	32	3.460	1.021	146	99	120	50
Aminoboronate	18	3.397	1.002	147	107	99	65
Sulfurous Acid	14	3.366	0.995	138	105	122	52
Sulfonate	10	3.435	1.013	151	104	107	53
Furazan	10	3.352	0.989	139	103	133	42
Silicate	6	3.409	1.006	155	99	117	60

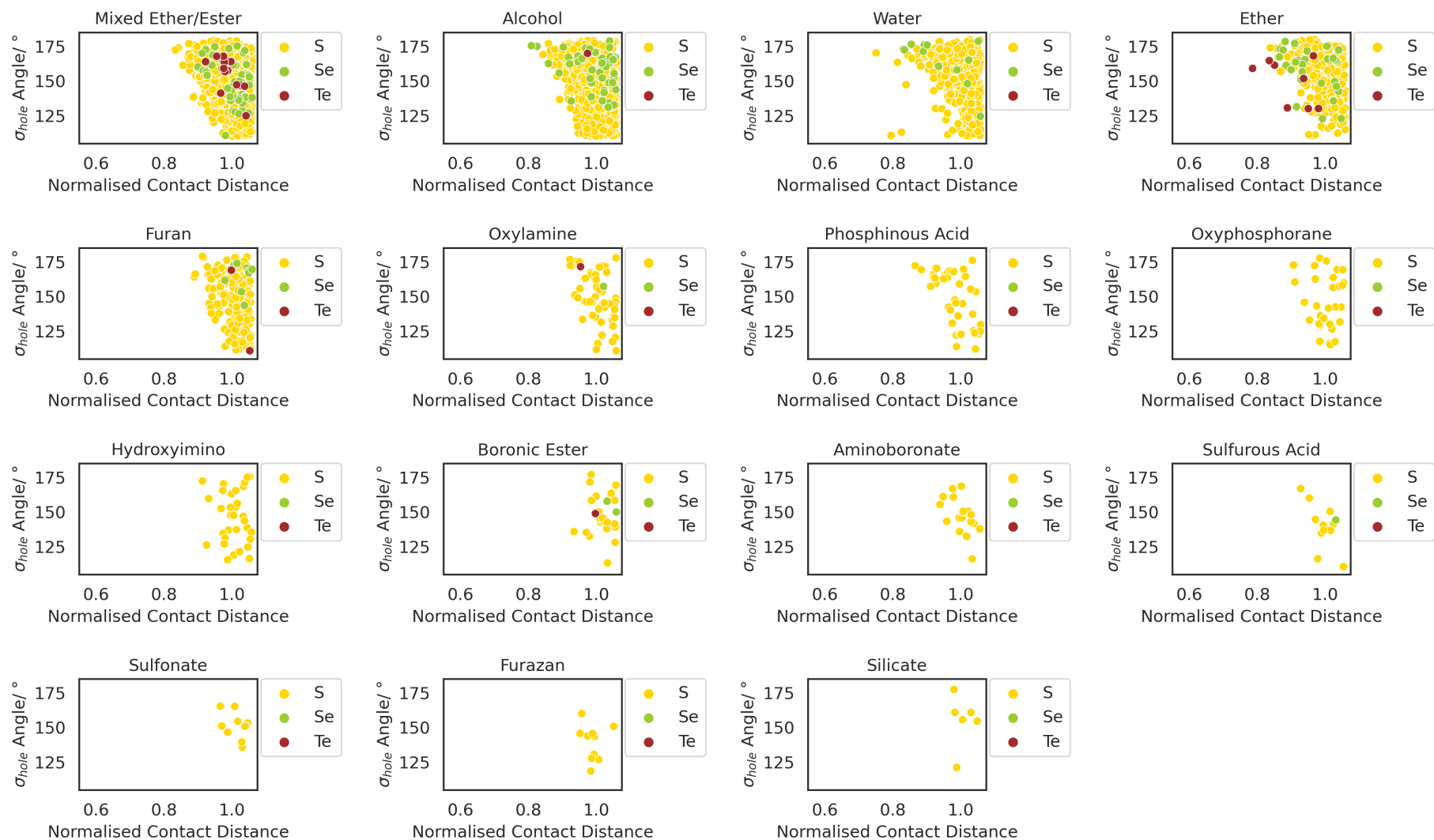


Figure 4.30: Scatter plots showing the distribution of normalised contact distance with respect to the σ -hole angle of chalcogen bonding contacts to divalent oxygen functional groups found in crystal structures in the CSD. Points colour coded by donor atom type

4.4.3 *Pn* ... Divalent Oxygen Contacts

Table 4.19 shows that pnictogen bonding is the least numerous of the σ -hole interactions with divalent oxygen Lewis bases. The distribution of σ -hole angle against normalised contact distance is shown in Figure 4.31. In comparison to the analogous halogen and chalcogen bonding contacts to divalent oxygen moieties (see Figure 4.23 & Figure 4.27), there is a markedly greater proportion of the contacts that have normalised contact distances less than 0.8. The directionality of these contacts appears to be biased towards 160-170° rather than 180°. Interestingly, the closest contacts which comprise mainly of antimony and bismuth contacts have a somewhat bimodal distribution, with one population having a range in normalised contact distance between 0.70-0.80 and a σ -hole angle between 155-170°, and the second population adopts a range in normalised contact distance between 0.75-0.85 and a σ -hole angle between 130-140°.

The mean averages of the geometric measurements for pnictogen bonding to divalent oxygen moieties is listed in Table 4.24. The average normalised contact distances are smaller than the analogous halogen and chalcogen bonding contacts, but are of a much smaller number of structures, which poses similar questions about selection bias as previously discussed. The trend in average normalised contact distances with respect to the pnictogen atom follows $\text{Sb} < \text{Bi} < \text{As} < \text{P}$, which somewhat reflects the polarisability of the pnictogen except that bismuth has longer normalised contact distances than antimony on average. The average σ -hole angles vary little with different pnictogen species, but interestingly the largest σ -hole angles involve arsenic. The average lone-pair angles vary little with respect to the pnictogen species, but the average plane angles are markedly larger for antimony contacts in comparison to the other pnictogen species.

The distributions of the lone-pair angles against normalised contact distance of pnictogen bonding contacts to divalent oxygen moieties is shown in Figure 4.32, and show a clear preference for the closest contacts to adopt a lone-pair angle between 100-130°. There are a selection of primarily phosphorus contacts that have lone-pair angles below 30°. The distributions of plane angle against normalised contact distance of pnictogen bonding contacts to divalent oxygen moieties is shown in Figure 4.33. There can be seen a clear trend towards orthogonality for contacts with a normalised contact distance less than 0.9.

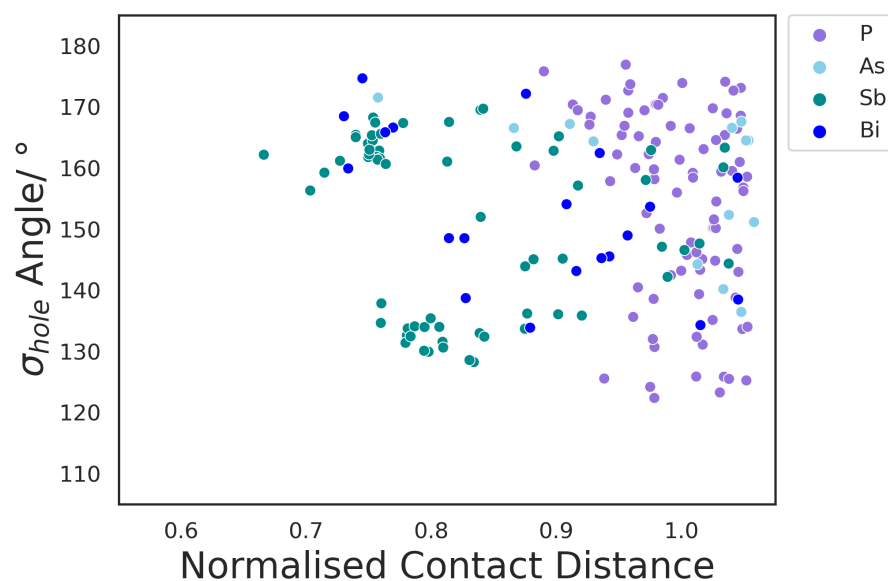


Figure 4.31: Scatter plots showing the distribution of normalised contact distance with respect to the σ -hole angle of pnictogen bonding contacts to divalent oxygens found in crystal structures in the CSD. Points colour coded by donor atom type

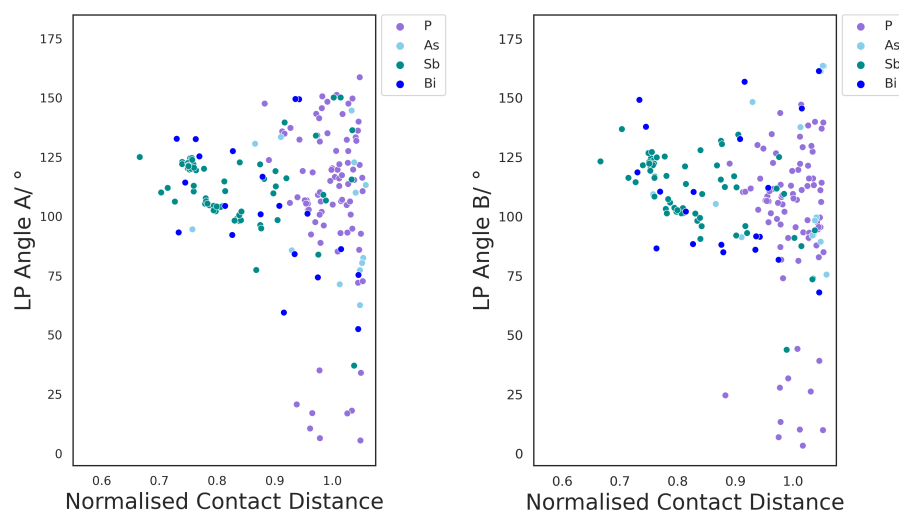


Figure 4.32: Scatter plots showing the distribution of normalised contact distance with respect to both lone-pair angles of pnictogen bonding contacts to divalent oxygens found in crystal structures in the CSD. Points colour coded by donor atom type

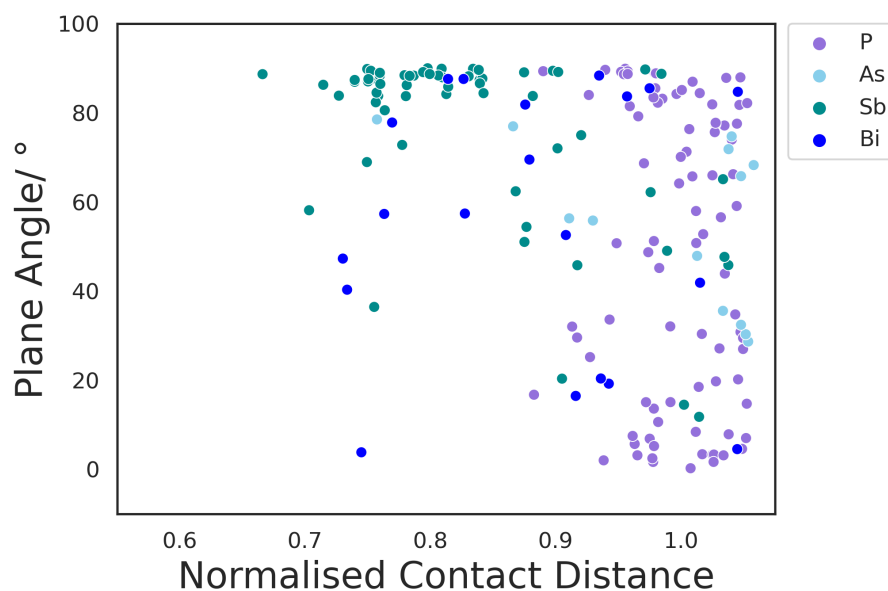


Figure 4.33: Scatter plots showing the distribution of normalised contact distance with the plane angle of pnictogen bonding contacts to divalent oxygens found in crystal structures in the CSD. Points colour coded by donor atom type

Table 4.24: Summary of geometric characteristics of pnictogen bonding contacts to divalent oxygen

Donor	Contacts	Mean Contact Distance/ Å	Mean Normalised Contact Distance	Mean σ -hole Angle/ °	Mean Lone-Pair Angle A/ °	Mean Lone-Pair Angle B/ °	Mean Plane Angle/ °
Pn	185	3.371	0.926	153	108	105	59
Bi	20	3.563	0.882	153	103	110	55
Sb	64	3.290	0.829	150	112	110	77
As	13	3.340	0.988	158	100	118	55
P	88	3.391	0.997	154	107	98	48

The pnictogen bonding contacts were compared on a functional group basis by subdividing the aforementioned data. The distributions of σ -hole angle against normalised contact distance for pnictogen bonding contacts to each functional group is shown in Figure 4.34, and the mean averages of these contacts is listed in Table 4.25.

The main group of interest are ethers, which contains the majority of the closest contacts. These close contacts can be attributed to complexes of SbF_3 with crown ethers¹⁸². This also explains bimodal distribution in antimony contacts mentioned earlier, as a pnictogen bonding contact with a near-linear σ -hole angle would result in the $\text{F-Sb}\cdots\text{O}$ angle of another fluorine substituent being within the search criteria detailed in Section 2.1.1. This means that these pnictogen bonds are in a way double-counted, with the contacts of larger σ -hole angle being the "true" σ -hole contact. This pnictogen bonding complex with a crown ether showcases the ability of pnictogen species to form multiple σ -hole interactions.

Table 4.25: Summary of geometric characteristics of pnictogen bonding contacts to different divalent oxygen functional groups

Functional Group	Contacts	Mean Contact Distance/ Å	Mean Normalised Contact Distance	Mean σ -hole Angle/ °	Mean Lone-Pair Angle A/ °	Mean Lone-Pair Angle B/ °	Mean Plane Angle/ °
Divalent Oxygen	180	3.367	0.923	153	108	105	59
Ether	56	3.158	0.814	153	115	114	75
Alcohol	34	3.418	0.993	146	88	92	36
Oxyphosphorane	30	3.438	0.992	159	119	110	57
Phosphate	20	3.365	0.990	156	109	112	64
Mixed Ether/Ester	19	3.603	0.933	149	124	97	58
Water	12	3.444	0.906	153	90	100	42
Furan	9	3.653	0.959	155	100	91	74

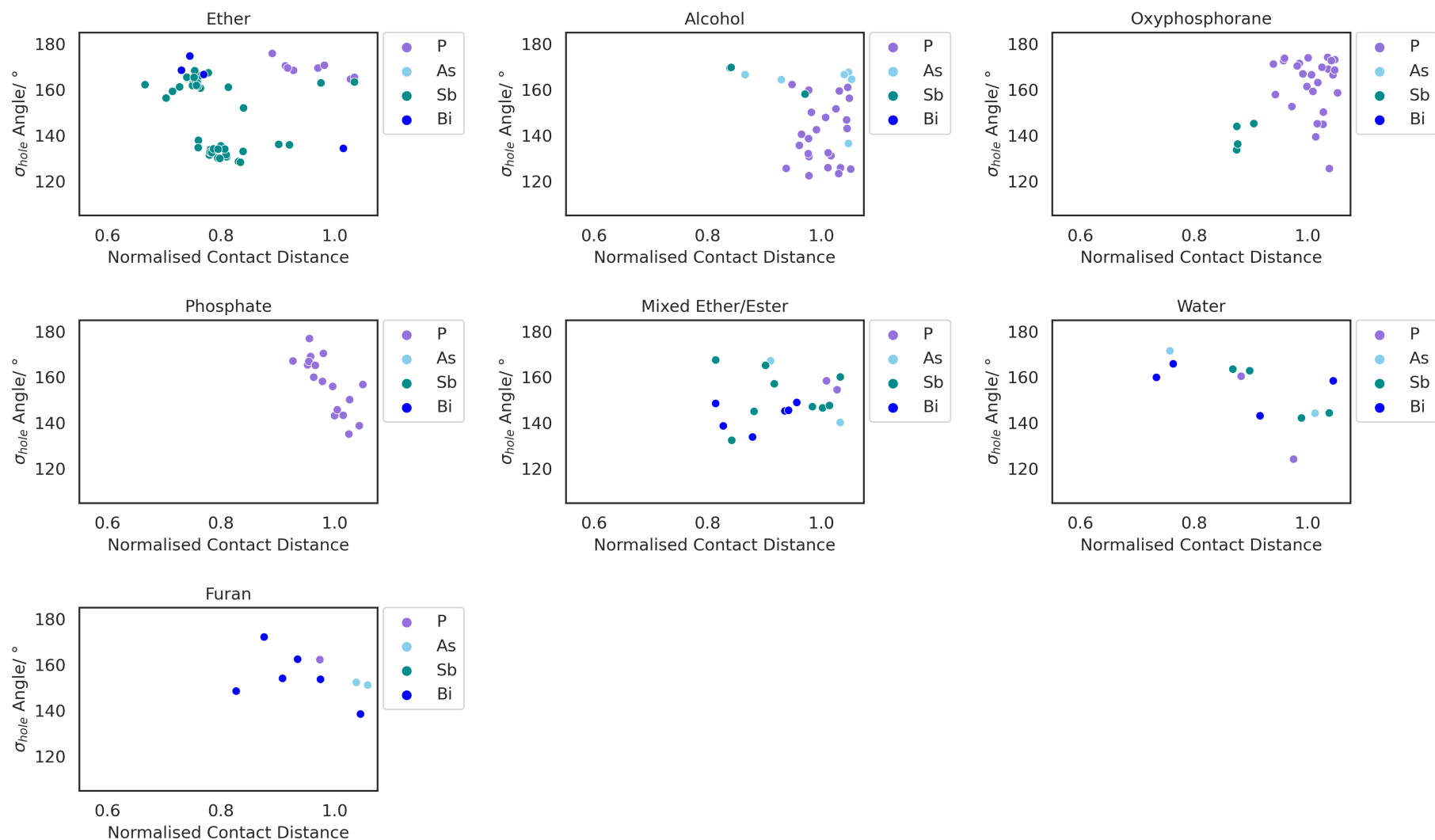


Figure 4.34: Scatter plots showing the distribution of normalised contact distance with respect to the σ -hole angle of pnictogen bonding contacts to divalent oxygen functional groups found in crystal structures in the CSD. Points colour coded by donor atom type

4.4.4 *Tt* ... Divalent Oxygen Contacts

Table 4.19 shows that tetrel bonding to divalent oxygen moieties is less numerous than the analogous halogen and chalcogen bonds, yet more numerous than the pnictogen bonds.

The distribution of σ -hole angle against the normalised contact distance for tetrel bonding contacts to divalent oxygen Lewis bases is shown in Figure 4.35. The silicon and germanium contacts are very long, with little to no relationship between the closeness of the contact and the σ -hole angle. The tin and lead contacts make up the only contacts with normalised contact distances less than 0.9, the difference between the two is lead contacts adopt σ -hole angles much further from linearity, whereas tin contacts are very close to linearity.

The mean averages of the geometric measurements of tetrel bonding contacts to divalent oxygen Lewis bases are listed in Table 4.26. The average normalised contact distance decreases with increasing atom size of the tetrel, following the trend $\text{Pb} < \text{Sn} < \text{Si} \approx \text{Ge}$. This is in contrast to the other tetrel bonding contacts analysed up to this point, with tin contacts expected to be closer than lead contacts due to greater polarisability (see Table 1.4). There is little trend with the average σ -hole angle over different tetrel bond donor atoms, with tin contacts having the largest angles. is little trend with the two average lone angles over different tetrel bond donor atoms, but the average plane angles are notably larger in the tin and lead contacts than the silicon or germanium contacts.

The distributions of the two lone-pair angles with respect to the normalised contact distance of tetrel bonding contacts to divalent oxygen Lewis bases is shown in Figure 4.36. There is clear preference for lone-pair angles around 120° with normalised contact distances around 120° . Little to no directionality with respect to the lone-pair angle is seen in silicon or germanium contacts. The distribution of plane angle against normalised contact distance for tetrel bonding contacts to divalent oxygen Lewis bases is shown in Figure 4.37. The closest tin contacts show notable bias towards the approach of the donor orthogonally to the plane of the divalent oxygen moiety, whereas the closest lead contacts, along with the silicon and germanium contacts show little to no directionality with respect to the plane angle.

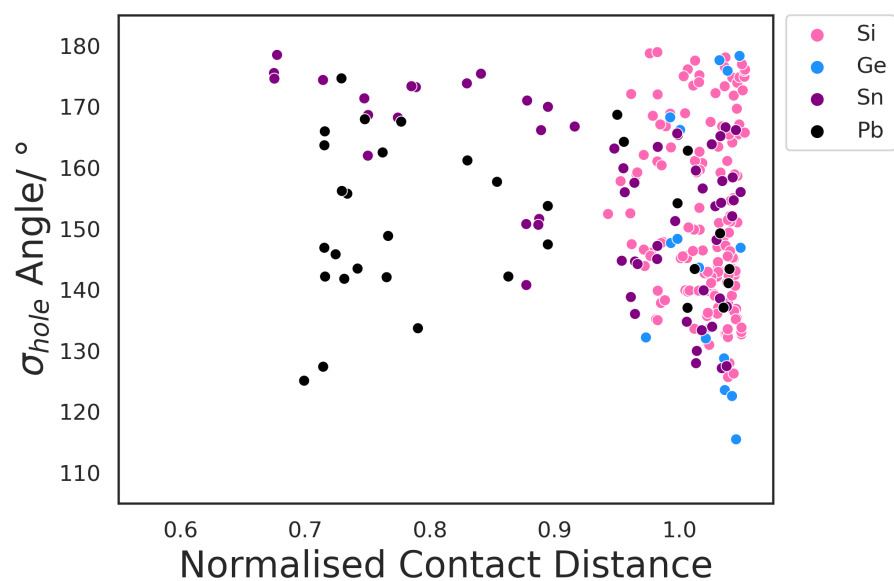


Figure 4.35: Scatter plots showing the distribution of normalised contact distance with respect to the σ -hole angle of tetrel bonding contacts to divalent oxygens found in crystal structures in the CSD. Points colour coded by donor atom type

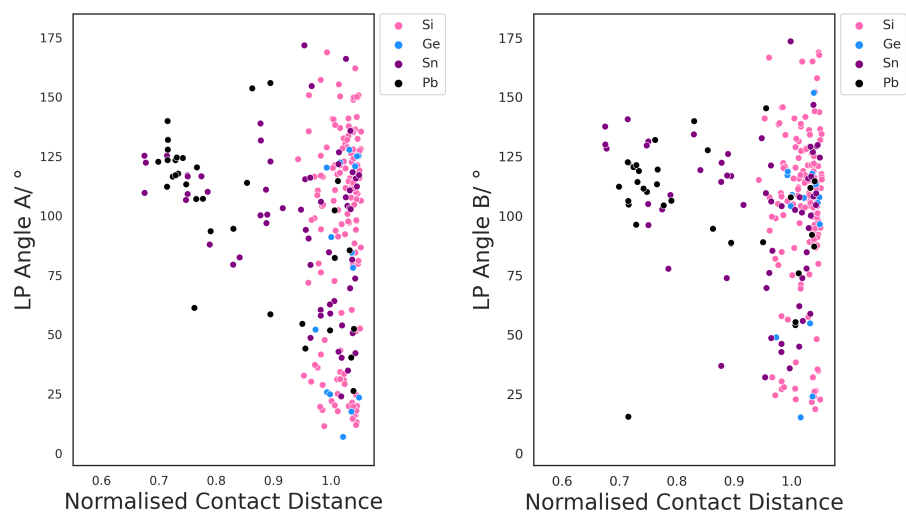


Figure 4.36: Scatter plots showing the distribution of normalised contact distance with respect to both lone-pair angles of tetrel bonding contacts to divalent oxygens found in crystal structures in the CSD. Points colour coded by donor atom type

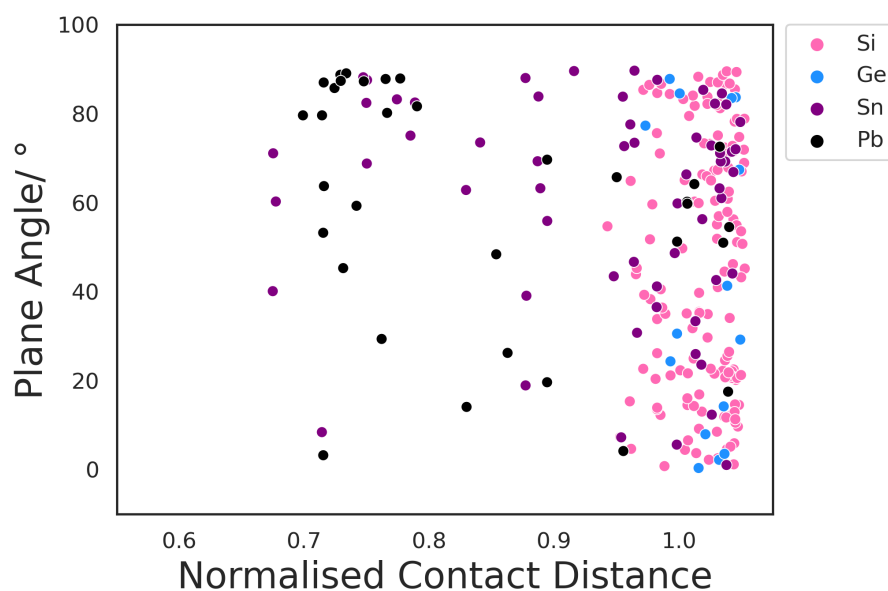


Figure 4.37: Scatter plots showing the distribution of normalised contact distance with the plane angle of tetrel bonding contacts to divalent oxygens found in crystal structures in the CSD. Points colour coded by donor atom type

The tetrel bonding contacts were compared on a functional group basis by subdividing the aforementioned data. The distributions of σ -hole angle against normalised contact distance for tetrel bonding contacts to each functional group is shown in Figure 4.38, and the mean averages of these contacts is listed in Table 4.27.

The most numerous functional groups involved in these tetrel bonds are alcohols, water, ethers, and mixed ether/ester subdivisions. Of note are the contacts involving ethers which have the most typically expected lone-pair angles and a plane angle nearer orthogonality, and the contacts involving water which have lone pair angles below 90° . The contacts with σ -hole angles nearest linearity are seen in the alcohol subplot of Figure 4.38, whereas the ether and mixed ether/ester subdivisions have lead contacts with σ -hole angles that deviate substantially from linearity.

Table 4.26: Summary of geometric characteristics of tetrel bonding contacts to divalent oxygen

Donor	Contacts	Mean Contact Distance/ Å	Mean Normalised Contact Distance	Mean σ -hole Angle/ °	Mean Lone-Pair Angle A/ °	Mean Lone-Pair Angle B/ °	Mean Plane Angle/ °
Tt	252	3.706	0.976	152	95	102	50
Pb	33	3.439	0.839	150	100	103	59
Sn	60	3.668	0.936	155	97	100	60
Ge	15	3.871	1.021	147	76	92	42
Si	144	3.765	1.020	152	95	103	44

Table 4.27: Summary of geometric characteristics of tetrel bonding contacts to different divalent oxygen functional groups

Functional Group	Contacts	Mean Contact Distance/ Å	Mean Normalised Contact Distance	Mean σ -hole Angle/ °	Mean Lone-Pair Angle A/ °	Mean Lone-Pair Angle B/ °	Mean Plane Angle/ °
Divalent Oxygen	249	3.705	0.976	152	94	101	49
Alcohol	85	3.693	0.980	155	89	102	43
Water	71	3.761	0.987	146	74	86	43
Ether	52	3.654	0.969	153	119	120	61
Mixed Ether/Ester	18	3.671	0.945	157	118	104	61
Boronic Acid	7	3.831	1.038	153	87	103	45
Oxyphosphorane	6	3.528	0.875	152	107	121	46
Hydroxylamine	4	3.564	0.924	152	114	96	73
Furan	3	4.018	1.030	150	111	76	65
Oxylamine	3	3.717	1.007	171	108	108	83

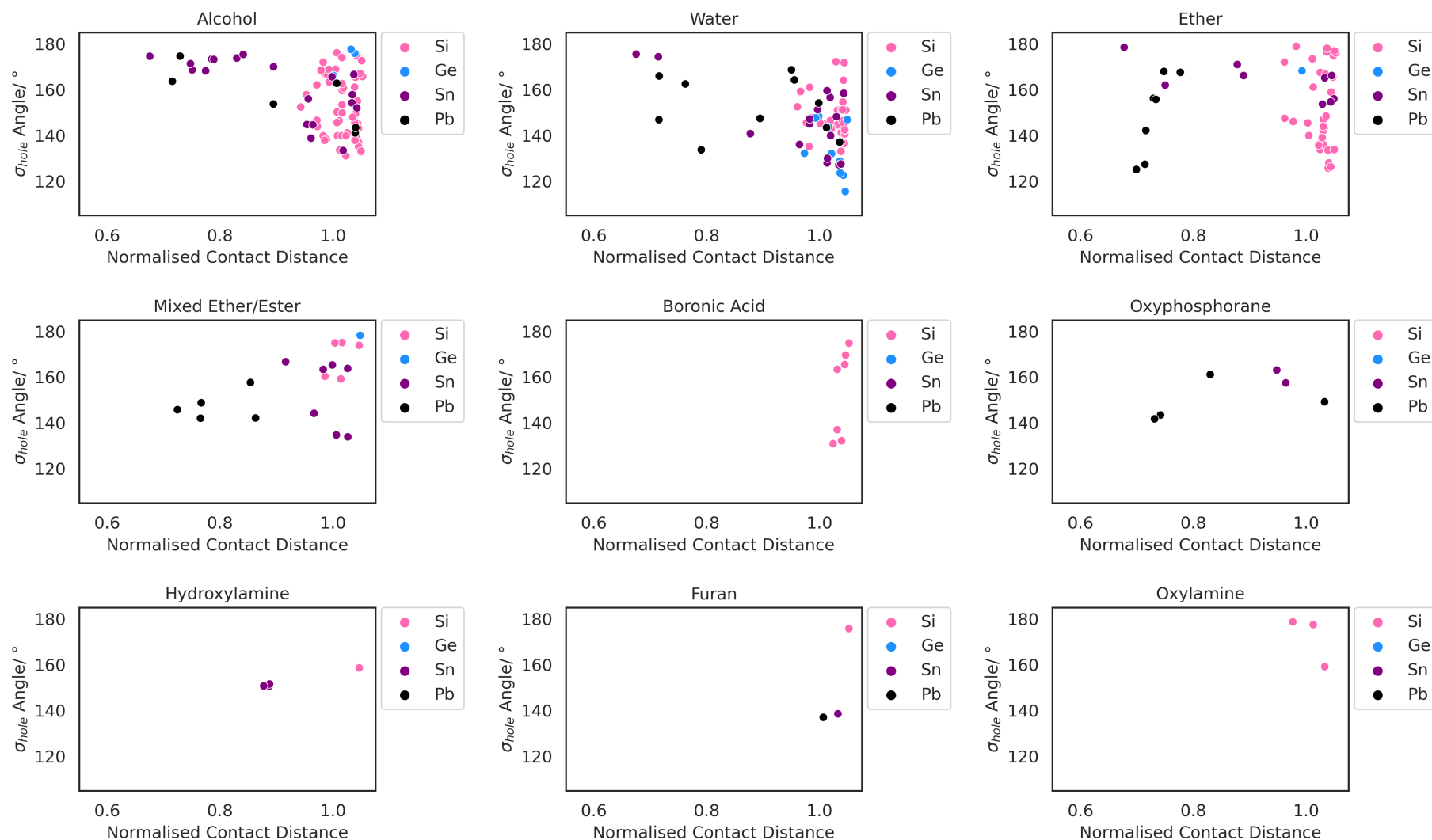


Figure 4.38: Scatter plots showing the distribution of normalised contact distance with respect to the σ -hole angle of tetrel bonding contacts to divalent oxygen functional groups found in crystal structures in the CSD. Points colour coded by donor atom type

4.4.5 Prevalence of σ_{hole} ··· Divalent Oxygen Contacts

The prevalence of σ -hole interactions to divalent oxygen Lewis bases is detailed in Table 4.10. Halogen and chalcogen bonding shows comparable prevalence, with bonding being slightly higher, whereas pnictogen bonding and tetrel bonding is substantially lower in prevalence. In comparison to σ -hole interactions to organic carbonyl Lewis bases (see Table 4.10), the prevalence of σ -hole interactions to divalent oxygen Lewis bases is considerably lower, despite having a much greater number of structures containing the required donor and acceptor moieties.

Table 4.28: Overall summary of the prevalence of contacts indicating σ -hole interactions to divalent oxygen moieties compared to the number of unique crystal structures found in the CSD that contain the necessary moieties.

σ -hole Interaction Type	Structures with Moieties	Structures with Contact(s)	Prevalence
Halogen-bonding	49171	7609	15.47%
Chalcogen-bonding	20661	2854	13.81%
Pnictogen-bonding	1705	104	6.10%
Tetrel-bonding	7273	186	2.56%

The prevalence of the halogen bonds to divalent oxygen with respect to the different halogen donor atoms is shown in Table 4.29.

The prevalence follows the trend of $I > Cl > Br$, which is interesting as it does not reflect the polarisability of the halogen atoms. It is important to note that the number of structures containing iodine and divalent oxygen moieties is much lower than those containing bromine or chlorine, which could lead to some selection bias. The increased prevalence of chlorine contacts over bromine contacts suggests that there may be competing use cases where chlorine and divalent oxygen moieties could be present in a crystal structure, such as the use of chlorinated solvents.

The number of structures containing the required moieties and the number of structures containing halogen bonding between the moieties decreases across the halogens going from chlorine to bromine to iodine. The prevalence of tetrel bonding to different divalent oxygen functional groups is shown in Table 4.21. The functional groups present in halogen bonding contacts in over 1000 unique crystal structures are mixed ether/esters, alcohols, and ethers, which have comparable prevalence to the average. Halogen bonding to furans have a markedly lower prevalence of 8.69% suggesting it is not a suitable halogen bonding motif. Aminoboronates and furazans show higher prevalence, albeit with fewer overall observations, suggesting nitrogen might have an enhancing effect on the availability of the lone pairs on divalent oxygen.

Table 4.29: Summary of the prevalence of halogen bonding contacts to divalent oxygens compared to the number of unique crystal structures found in the CSD that contain the necessary moieties

Donor	Structures with Moieties	Structures with Contact(s)	Prevalence
X	49171	7609	15.47%
I	3926	704	17.93%
Br	17097	2342	13.70%
Cl	29848	4591	15.38%

Table 4.30: Summary of the prevalence of halogen bonding contacts to various divalent oxygen functional groups compared to the number of unique crystal structures found in the CSD that contain the necessary moieties

Functional Group	Structures with Moieties	Structures with Contact(s)	Prevalence
Divalent Oxygens	49171	7609	15.47%
Mixed Ether/Ester	24841	3015	12.14%
Alcohol	14384	1962	13.64%
Ether	7781	1055	13.56%
Oxyphosphorane	1829	289	15.80%
Furan	5997	521	8.69%
Water	3157	407	12.89%
Boronic Ester	829	87	10.49%
Oxylamine	1663	166	9.98%
Aminoboronate	100	36	36.00%
Sulfonate	494	50	10.12%
Hydroxyimino	486	92	18.93%
Peroxide	130	10	7.69%
Phosphinous Acid	157	28	17.83%
Boronic Acid	160	24	15.00%
Furazan	106	30	28.30%
Sulforous Acid	39	6	15.38%
Other	987	85	8.61%

The prevalence of the chalcogen bonds to divalent oxygen with respect to the different chalcogen donor atoms is shown in Table 4.31. Sulfur dominates both the structures containing the required moieties, and those containing chalcogen bonding contacts, thus makes up the majority of the overall chalcogen bonding average prevalence. Selenium contacts are markedly lower in prevalence, with tellurium contacts having slightly lower prevalence than sulfur. The selenium and tellurium prevalence is out of a much smaller pool of crystal structures containing the required moieties.

The prevalence of chalcogen bonds to different divalent oxygen functional groups is shown in Table 4.32. The standout functional groups are water and alcohol, which have prevalences of 24.34% and 15.51% respectively, out of a large number of structures. Some high prevalence is seen with chalcogen bonding contacts to less common functional groups, such as aminoboronates (28.57%), phosphinous acids (45.28%), and sulfurous acids (32.35 %).

Table 4.31: Summary of the prevalence of chalcogen bonding contacts to divalent oxygens compared to the number of unique crystal structures found in the CSD that contain the necessary moieties

Donor	Structures with Moieties	Structures with Contact(s)	Prevalence
Ch	20661	2854	13.81%
Te	155	20	12.90%
Se	1151	112	9.73%
S	19450	2726	14.02%

Table 4.32: Summary of the prevalence of chalcogen bonding contacts to various divalent oxygen functional groups compared to the number of unique crystal structures found in the CSD that contain the necessary moieties

Functional Group	Structures with Moieties	Structures with Contact(s)	Prevalence
Divalent Oxygens	20661	2854	13.81%
Mixed Ether/Ester	10747	942	8.77%
Alcohol	5808	901	15.51%
Ether	2929	278	9.49%
Water	1738	423	24.34%
Furan	1948	156	8.01%
Oxyphosphorane	411	27	6.57%
Boronic Ester	255	24	9.41%
Aminoboronate	35	10	28.57%
Oxylamine	426	50	11.74%
Hydroxyimino	184	28	15.22%
Sulfonate	113	8	7.08%
Phosphinous Acid	53	24	45.28%
Boronic Acid	29	4	13.79%
Furazan	32	10	31.25%
Peroxide	18	0	N/A
Sulfurous Acid	34	11	32.35%
Other	244	19	7.79%

Pnictogen bonding is shown in Table 4.28 to have much lower prevalence than halogen or chalcogen bonding. Antimony and bismuth contacts have by far the highest prevalence compared to arsenic and phosphorus contacts. The most notable prevalence of pnictogen bonding to different divalent oxygen functional groups shown in Table 4.34 involve alcohols, with a prevalence of 13.46% out of 208 structures, which shows that this is a potentially suitable motif for pnictogen bonding.

Table 4.33: Summary of the prevalence of pnictogen bonding contacts to divalent oxygens compared to the number of unique crystal structures found in the CSD that contain the necessary moieties

Donor	Structures with Moieties	Structures with Contact(s)	Prevalence
Pn	1705	104	6.10%
Bi	49	15	30.61%
Sb	84	23	27.38%
As	121	10	8.26%
P	1452	56	3.86%

Table 4.34: Summary of the prevalence of pnictogen bonding contacts to various divalent oxygen functional groups compared to the number of unique crystal structures found in the CSD that contain the necessary moieties

Functional Group	Structures with Moieties	Structures with Contact(s)	Prevalence
Divalent Oxygens	1705	104	6.10%
Mixed Ether/Ester	648	15	2.31%
Oxyphosphorane	267	17	6.37%
Ether	398	21	5.28%
Furan	184	8	4.35%
Alcohol	208	28	13.46%
Boronic Ester	61	2	3.28%
Water	56	7	12.50%
Phosphinous Acid	8	2	25.00%
Aminoboronate	4	0	N/A
Sulfonate	7	0	N/A
Oxylamine	9	1	11.11%
Sulforous Acid	2	0	N/A
Boronic Acid	1	0	N/A
Furazan	1	0	N/A
Peroxide	1	0	N/A
Other	19	5	26.32%

Table 4.28 shows that tetrel bonding to divalent oxygen has the lowest prevalence of the σ -hole interactions involving divalent oxygen Lewis bases. This is due to the majority of the contacts are made up of silicon contacts, which have a very low prevalence of 1.69%. The lead contacts are the tetrel bond showing notable prevalence, how it is out of a very small number of crystal structures containing the required moieties. The most notable functional groups involved in tetrel bonding, with the prevalences shown in Table 4.36, are water, with a prevalence of 18.89%, and to a lesser extent alcohols with 4.17%.

Table 4.35: Summary of the prevalence of tetrel bonding contacts to divalent oxygens compared to the number of unique crystal structures found in the CSD that contain the necessary moieties

Donor	Structures with Moieties	Structures with Contact(s)	Prevalence
Tt	7273	186	2.56%
Pb	63	22	34.92%
Sn	493	45	9.13%
Ge	189	9	4.76%
Si	6565	111	1.69%

Table 4.36: Summary of the prevalence of tetrel bonding contacts to various divalent oxygen functional groups compared to the number of unique crystal structures found in the CSD that contain the necessary moieties

Functional Group	Structures with Moieties	Structures with Contact(s)	Prevalence
Divalent Oxygens	7273	186	2.56%
Mixed Ether/Ester	3252	16	0.49%
Ether	2530	41	1.62%
Alcohol	1656	69	4.17%
Oxyphosphorane	298	4	1.34%
Furan	448	3	0.67%
Boronic Ester	244	1	0.41%
Water	217	41	18.89%
Oxylamine	146	3	2.05%
Boronic Acid	81	7	8.64%
Sulfonate	86	0	N/A
Hydroxyimino	25	2	8.00%
Peroxide	34	1	2.94%
Phosphinous Acid	19	0	N/A
Sulforous Acid	1	0	N/A
Aminoboronate	3	0	N/A
Furazan	2	0	N/A
Other	56	1	1.79%

4.4.6 Notable Divalent Oxygen Functional Groups

The population of short contacts found in the distributions of σ -hole angle against normalised contact distance of halogen bonding contacts to ethers (see Figure 4.23 & Figure 4.26) are attributed to a high pressure study of the co-crystal of ICI and 1,4-dioxane (CSD refcode **WINLAK**) which is shown in Figure 4.39¹⁸⁰.

Inspection of other short halogen contacts to ethers with σ -hole angles near linearity shows a notable number of contacts involving morpholine derivatives as an acceptor. These are notably halogen bonding studies, indicating this to be a useful

acceptor moiety, especially as it has two acceptor sites with a nitrogen lone-pair also available, a prime example of this is the co-crystal of morpholine with 1,3, 5-triiodo-2,4,6-trifluorobenzene (CSD refcode **AJOGUE**), shown in Figure 4.40¹⁸³.

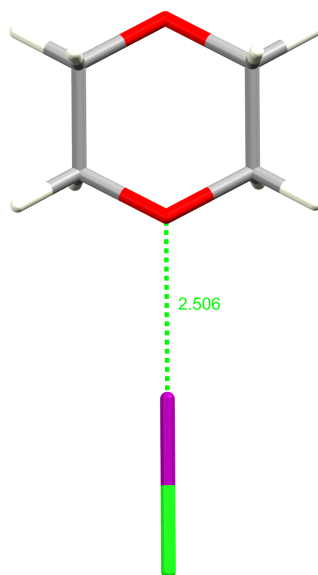


Figure 4.39: An example of a crystal structure with a halogen bonding contact involving an ether as a halogen bond acceptor. Distance given in Ångström. CSD refcode: **WINLAK**

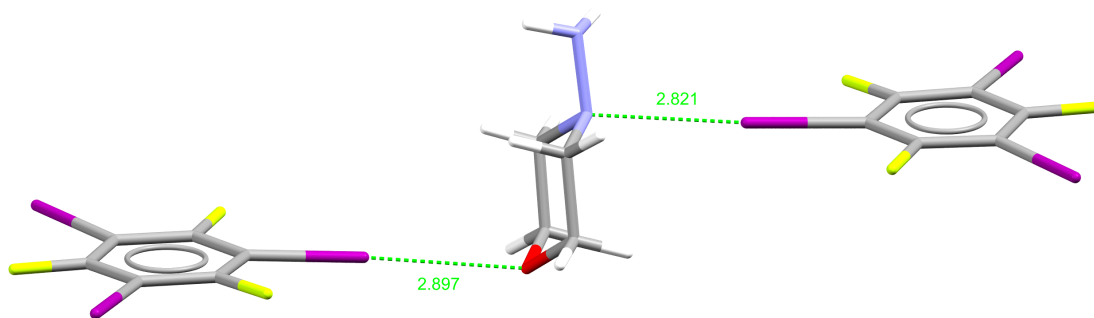


Figure 4.40: An example of a crystal structure with a halogen bonding contact involving morpholine as a halogen bond acceptor. Distance given in Ångström. CSD refcode: **AJOGUE**

An interesting set of motifs involving chalcogen bonding and pnictogen bonding involves crown ethers as the acceptor, which offers a plethora of suitable binding sites. A chalcogen bonding example can be seen in the co-crystal of a telluradiazole derivative with 18-crown-6 (CSD refcode **OGIMUP**) shown in Figure 4.41¹⁸⁴. A pnictogen bonding example can be seen

in the co-crystal of SbF_3 and dicyclohexano-18-crown-6 (CSD refcode **RIYMER**) shown in Figure 4.42¹⁸².

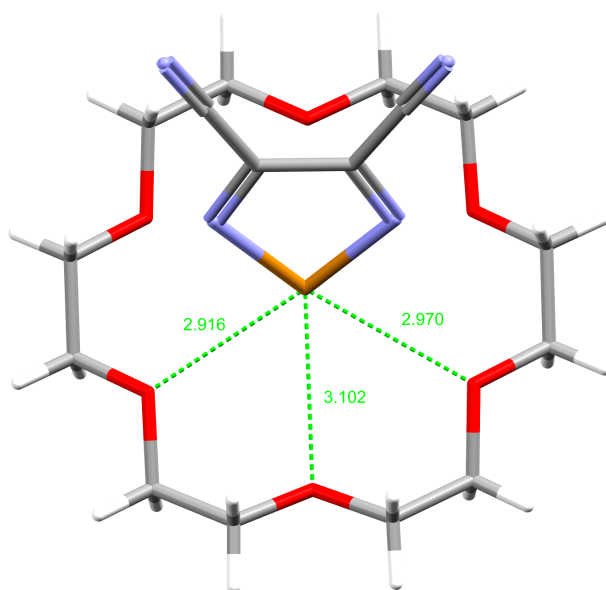


Figure 4.41: An example of a crystal structure with a chalcogen bonding contact involving crown ethers as a chalcogen bond acceptor. Distance given in Ångström. CSD refcode: **OGIMUP**

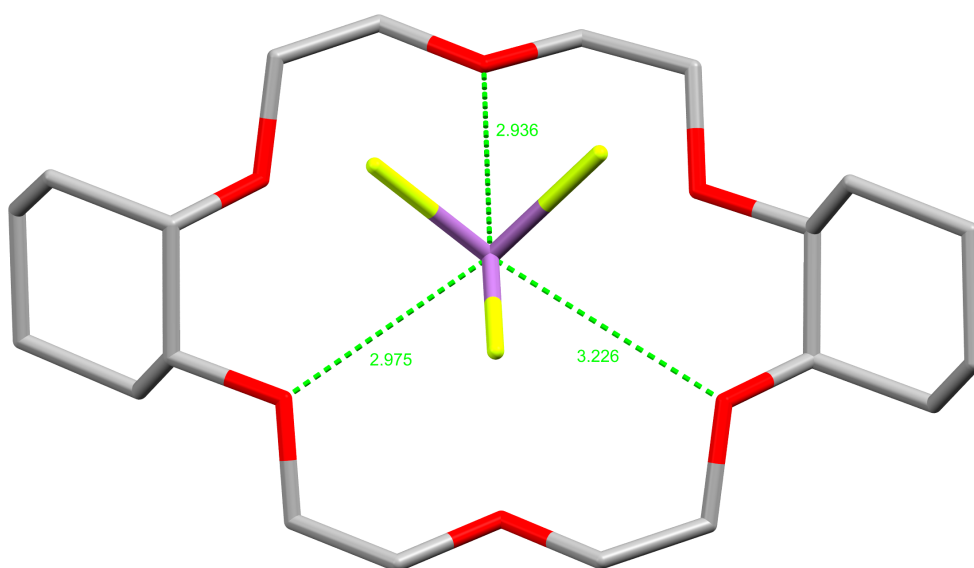


Figure 4.42: An example of a crystal structure with a pnictogen bonding contact involving crown ethers as a pnictogen bond acceptor. Distance given in Ångström. CSD refcode: **RIYMER**

4.5 Oxides

Oxides offer a plethora of options in terms of potential Lewis bases, with the element that the oxide is produced from having a key factor in its activity, along with a wide variety of potential bond orders and charges. The oxides studied in this work as potential σ -hole acceptors are terminal oxides of nitrogen, sulfur, phosphorus, halogens, and metals, in order to investigate the effect that the nature of the atom the oxide originates from on the Lewis basicity. Taking the form in the queries (see Section 4.2) as O-R', with R' = N, Ch, Tt, Pn, X, M. Ch = Chalcogen, Pn = Pnictogen, Tt = Tetrel, X = Halogen, M = Metal. Due to the scarcity of results involving oxides of chalcogens other than sulfur, pnictogens other than phosphorus, and tetrel oxides, the oxide subdivisions considered in this work can be defined by the groupings of O-R' = N, S, P, X, & M bonded by any bond type (see Section 2.1.1), it is only possible to explore the identity and valency of the element that the oxide is formed from, preventing the systematic investigation into the effects of the other substituents making up the oxide. This query means that a plane through the acceptor molecule cannot be defined, so the lone-pair directionality can only be explored with the single angle, but one would still expect a preference for a lone-pair angle of around 120°. The notable potential for oxides to be charged species is a key difference compared to the other σ -hole acceptors discussed in this work, and will likely show different character.

There exist several examples in the literature pertaining to halogen bonding interactions with N-oxides as the acceptor, such as pyridine N-oxide^{158 159}. These N-oxides have an additional lone-pair on the oxygen atom, resulting in a very nucleophilic O⁻-N⁺ system, which interacts strongly with halogen bond donors. Regarding oxides of sulfur, there has been investigation into the mechanism of absorption processes of sulfur dioxide, which involves hydrogen bonding¹⁸⁵. Halogen bonding involving phosphine oxides have been explored, along with similar studies concerning hydrogen bonding, showing them to be effective acceptors for these noncovalent interactions, capable of some large interaction energies^{186 187}. Hydrogen bond studies have been undertaken with perchlorates as the acceptor, which is the most common of the perhalate species¹⁸⁸. These groups are interesting from a hydrogen bonding perspective for their potential to form not just classical hydrogen bonds, but bifurcated and trifurcated hydrogen bonds also^{189 190}. CSD studies investigating halogen bonding to metal oxo groups have been undertaken by Brammer *et al.* where the acceptor atom for the halogen bond is a monoatomic oxygen bound to a transition metal centre (M=O)¹⁴¹. The structures found in this study by Brammer *et al.* were few in number but did illustrate the ability of these interactions to be exploited to link metal-oxo groups, but the authors note that some involve oxometallate anions substituted with halogenated organic groups, leading to a weaker halogen bond.

An important use case for σ -hole interactions involving oxides as acceptors involve the field of anion recognition. Championed by the work of Beer and Lim, halogen bonding and chalcogen bonding has been widely exploited to interact with a range of anions such as triflates, halides, and azides^{191 192}. The use of σ -hole interactions for anion recognition provides alternatives to hydrogen bonding for this purpose, with X $\cdots$ anion interactions rivalling or even surpassing hydrogen bonding analogues. These approaches for anion recognition offer enhanced selectivity due to the tunable nature of the σ -hole donors and the geometric control offered by higher valence σ -hole donors, such as chalcogen bonding effects^{142 193}. An overall summary of the geometric observables of contacts indicative of σ -hole interactions to terminal oxide oxygen

moieties is tabulated in Table 4.37.

Table 4.37: Overall geometric summary of the contacts indicating σ -hole interactions to terminal oxide oxygen moieties

σ -hole Interaction Type	No. Contacts	Mean Normalised Contact Distance	Mean σ -hole Angle/ $^{\circ}$	Mean Lone-Pair Angle/ $^{\circ}$
Halogen-bonding	8397	0.968	152	122
Chalcogen-bonding	4811	0.971	152	123
Pnictogen-bonding	394	0.927	157	127
Tetrel-bonding	450	0.911	160	126

4.5.1 X ... Oxide Oxygen Contacts

As shown in Table 4.37, halogen bonds are the most numerous of the σ -hole interactions involving oxide Lewis bases as acceptors.

The distribution of σ -hole angle against normalised contact distance for halogen bonding contacts to oxide oxygens is shown in Figure 4.43. The characteristic "wedge" shape of the distribution is once again present, with a shape closer to those found in the analogous plots in Chapter 3 than in the analogous plots in Section 4.3 or Section 4.4. This suggests an enhanced Lewis basicity compared to the organic carbonyl and divalent oxygen σ -hole acceptors, which is to be expected given oxide species are often anionic. There are a small number of chlorine contacts with normalised contact distances less than 0.85 that have σ -hole angles that deviate substantially from linearity. This suggests the presence of competing interactions such as hydrogen bonding, which chlorine can act as an acceptor for, that is perturbing the contact geometry. The mean averages of the geometric measurements of halogen bonding contacts to oxide oxygens is listed in Table 4.38. The average normalised contact distance decreases with more polarisable halogen bond donor atoms, following the trend $I < Br < Cl$. Likewise, the average σ -hole angle increases with more polarisable halogen bond donor atoms, following the trend $I > Br > Cl$. The average lone-pair angles vary little with different halogen bond donor atoms, only deviating 1° from the average 122° which is very close to the lone-pair angle predicted by orbital hybridisation theory.

The distribution of lone-pair angle against normalised contact distance for halogen bonding contacts to oxide oxygens is shown in Figure 4.44. There is a clear bias towards lone-pair angles of around 120° for the closest contacts with normalised contact distances less than 0.75. For contacts with normalised contact distances greater than 0.75 there are a greater number with lone-pair angles between 100 - 180° .

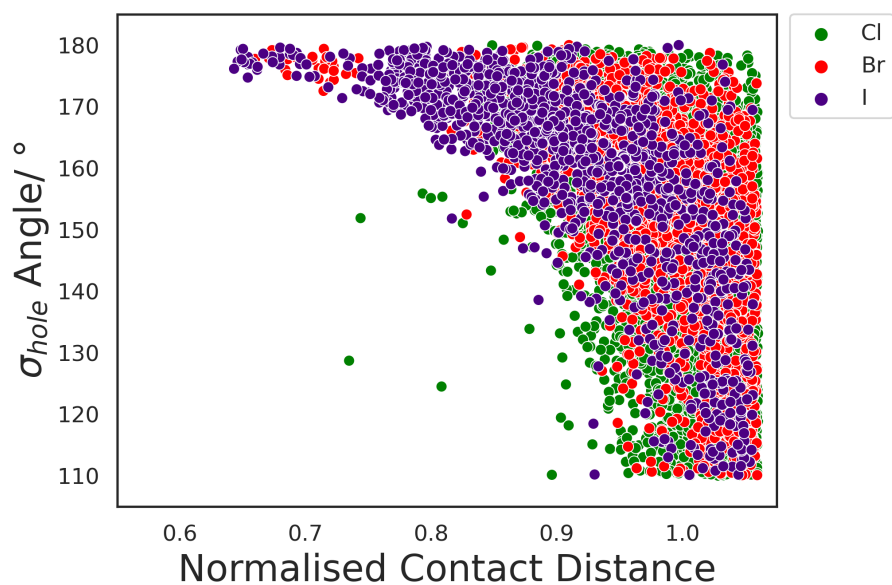


Figure 4.43: Scatter plots showing the distribution of normalised contact distance with respect to the σ -hole angle of halogen bonding contacts to oxide oxygens found in crystal structures in the CSD. Points colour coded by donor atom type

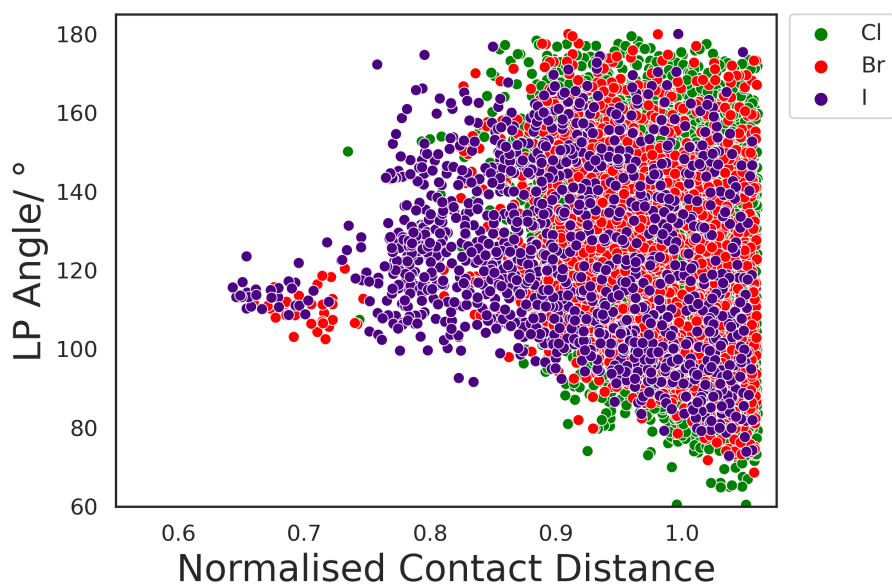


Figure 4.44: Scatter plots showing the distribution of normalised contact distance with respect to the lone-pair angle of halogen bonding contacts to oxide oxygens found in crystal structures in the CSD. Points colour coded by donor atom type

Table 4.38: Summary of geometric characteristics of halogen bonding contacts to oxide oxygens

Donor	Contacts	Mean Contact Distance/ Å	Mean Normalised Contact Distance	Mean σ -hole Angle/ °	Mean Lone-Pair Angle/ °
X	8397	3.256	0.968	152	122
I	1302	3.201	0.904	160	123
Br	2266	3.262	0.971	154	121
Cl	4829	3.269	0.985	149	123

The trends among different terminal oxides were examined by subdivision of the data presented in Figure 4.43. The distributions of σ -hole angle against normalised contact distance for halogen bonding to the different types of terminal oxide are presented in Figure 4.45, and those same distributions colour coded by the valency of each type of oxide is shown in Figure 4.46. The mean averages of the geometric measurements for halogen bonding contacts to each type of terminal oxide are listed in Table 4.39.

The majority of the halogen bonding contacts are to nitrogen oxides, with the distribution of σ -hole angle against normalised contact distance shown in Figure 4.45. These contacts account for the shortest contacts with the greatest tendency towards σ -hole angles near linearity of those shown in Figure 4.43. The valencies of these nitrogen oxides, plotted in Figure 4.46, shows the majority of these oxides are trivalent, with some divalent species showing no directionality with respect to the σ -hole angle. These divalent species correspond to nitrosyl ligands coordinated to a metal centre, analogous to metal carbonyls.

The next most numerous halogen bonding contacts are to sulfur oxides, with the distribution of σ -hole angle against normalised contact distance shown in Figure 4.45. These contacts do not have as many contacts with normalised contact distances less than 0.8 as with the analogous contacts to nitrogen oxides, yet still show clear directionality with respect to the σ -hole. The valencies of these sulfur oxides, plotted in Figure 4.46, shows that the majority of these sulfur oxides are tetravalent, with some trivalent oxides showing a greater proportion of closer contacts with σ -hole angles nearer linearity, which generally corresponds to sulfinyl or sulfinate moieties. There are also some contacts to divalent sulfur oxides which show a slight bias towards halogen bonding contacts with σ -hole angles between 160-170° which primarily involve sulfur dioxide ions.

The next most numerous halogen bonding contacts are to halogen oxides, with the distribution of σ -hole angle against normalised contact distance shown in Figure 4.45. These contacts have similar distributions of σ -hole angles against normalised contact distance with the halogen bonding contacts to metal oxides, see Figure 4.45. The halogen bonding contacts to halogen oxides primarily involve tetravalent perchlorate moieties shown in the distributions of σ -hole angles against normalised contact distance colour coded by valency in Figure 4.46 and by halogen oxide in Figure 4.47. This is likely due to the use of perchlorates as a counterion in crystal structures due to their poor coordination ability¹⁹⁴. There are also a very small number of contacts to trivalent chlorate moieties, which make up some of the closest halogen bonding

contacts to halogen oxides, but are too few in number to draw meaningful conclusions from.

The halogen bonding contacts to metal oxides involve a much greater range of metals and valencies of which the distributions of σ -hole angles against normalised contact distance colour coded by valency in Figure 4.46 and by metal oxide type in Figure 4.47. There are four notably shorter halogen bonding contacts to metal oxides with σ -hole angles greater than 165° that correspond to tetravalent rhenium oxides.

The least numerous halogen bonding contacts to terminal oxides involve phosphorus oxides, with the distribution of σ -hole angle against normalised contact distance shown in Figure 4.45. Shown in these distributions colour coded by the valency of the oxide (see Figure 4.46) the valencies of these phosphorus oxides are exclusively tetravalent and are likely to be phosphine oxides due to their widespread use in synthetic chemistry. These halogen bonding contacts to phosphorous oxides show a notably bias towards σ -hole angles greater than 160° for contacts with a normalised contact distance less than 0.9.

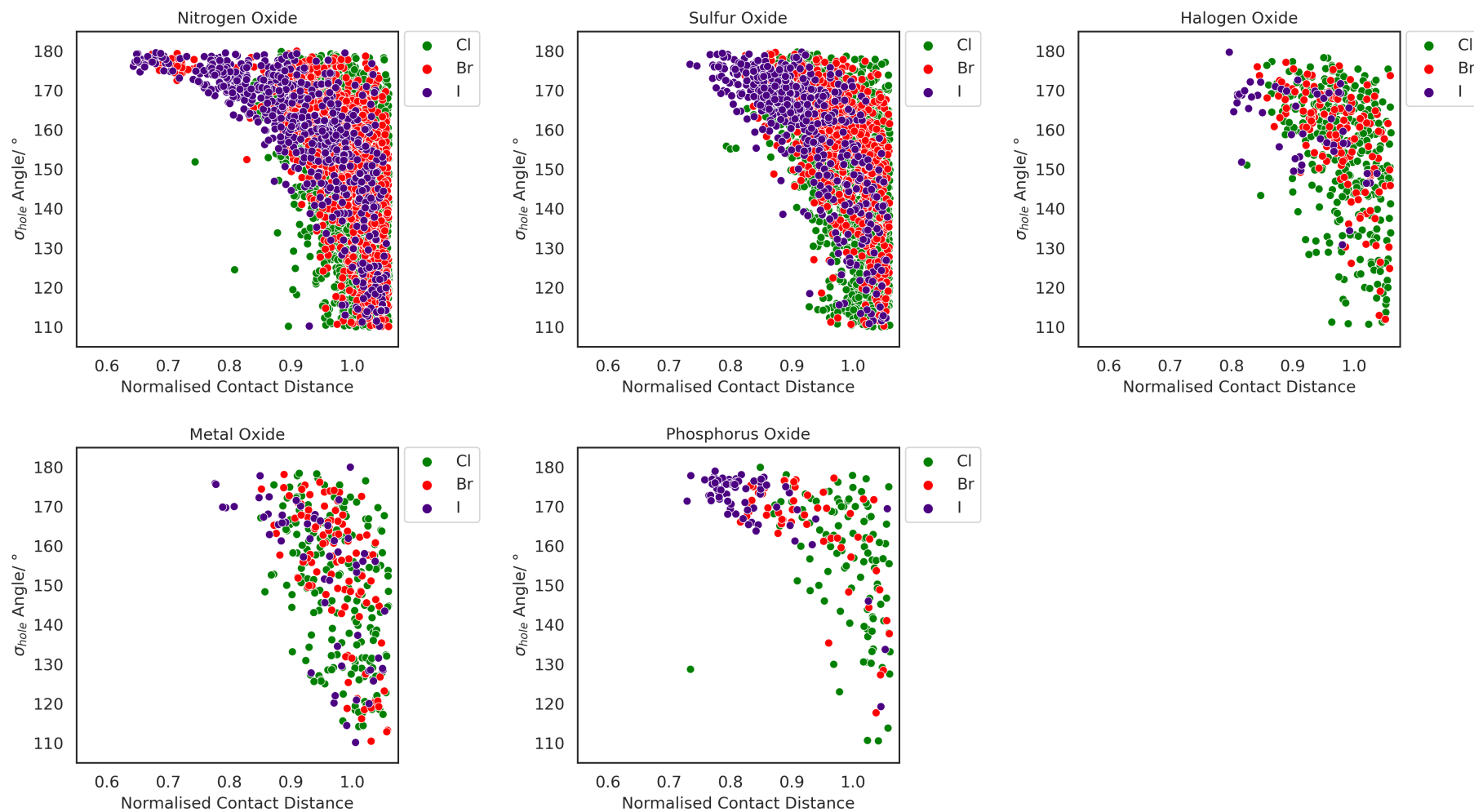


Figure 4.45: Scatter plots showing the distribution of normalised contact distance with respect to the σ -hole angle of halogen bonding contacts to various oxides found in crystal structures in the CSD. Points colour coded by oxide type

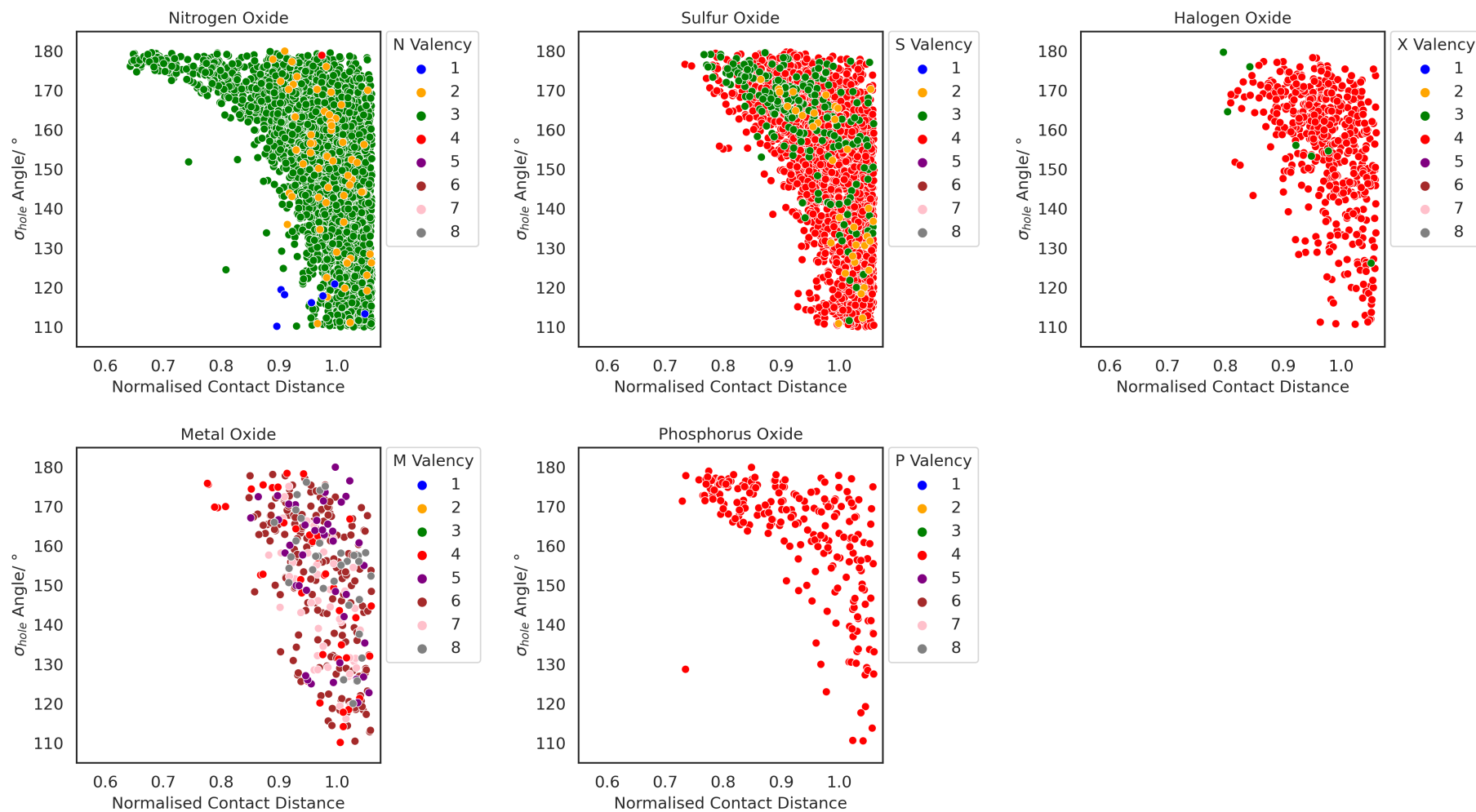


Figure 4.46: Scatter plots showing the distribution of normalised contact distance with respect to the σ -hole angle of halogen bonding contacts to various oxides found in crystal structures in the CSD. Points colour coded by valency of oxide element

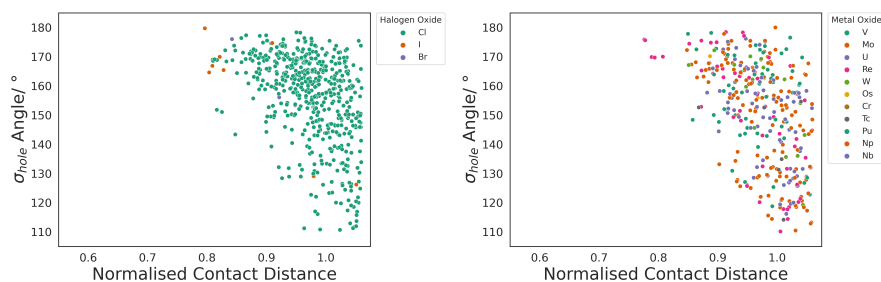


Figure 4.47: Scatter plots showing the distribution of normalised contact distance with respect to the σ -hole angle of halogen bonding contacts to halogen oxides (left), and metal oxides (right) found in crystal structures in the CSD. Points colour coded by halogen oxide atom, or metal oxide atom

Table 4.39: Summary of geometric characteristics of halogen bonding contacts to different classes of oxides

Functional Group	Contacts	Mean Contact Distance/ Å	Mean Normalised Contact Distance	Mean σ -hole Angle/ °	Mean Lone-Pair Angle/ °
Oxides	8397	3.256	0.968	152	122
Nitrogen Oxide	4232	3.257	0.968	151	120
Sulfur Oxide	3046	3.262	0.970	153	124
Halogen Oxide	538	3.259	0.973	156	126
Metal Oxide	353	3.269	0.971	151	125
Phosphorus Oxide	228	3.142	0.927	162	125

4.5.2 Ch ... Oxide Oxygen Contacts

Table 4.37 shows that chalcogen bonding contacts to terminal oxide oxygen Lewis bases are around half as numerous as the analogous halogen bonding contacts.

The distribution of σ -hole angle against normalised contact distance for chalcogen bonding contacts to oxide Lewis bases is shown in Figure 4.48. The selenium and tellurium contacts show a more notable bias towards σ -hole angles nearer linearity as the normalised contact distance decreases. The closest selenium contacts adopt a σ -hole angle closer to linearity than the closest tellurium contacts which adopt a σ -hole angle nearer 160°. Sulfur contacts dominate the space of chalcogen bonding to oxides, and interestingly, make up the closest contacts with σ -hole angles that are greater than 170°, but has the broadest distribution of σ -hole angles.

The mean averages of the geometric measurements for chalcogen bonding contacts to terminal oxide oxygens is shown in Table 4.40. The average normalised contact distance decreases with more polarisable chalcogen bond donor atoms, following the trend $\text{Te} < \text{Se} < \text{S}$. The average σ -hole angles are very similar for selenium and tellurium contacts, but is notably smaller in sulfur contacts. The average lone-pair angles deviate little with respect to the different chalcogens, but the average lone-pair angle increases with increasing polarisability of the chalcogen.

The distribution of lone-pair angle against normalised contact distance for chalcogen bonding contacts to oxide Lewis bases is shown in Figure 4.49. For the closest contacts the range of lone-pair angles seems to be centered around the

expected 120° region, but has a range in angle of 100-140°. For normalised contact distances greater than 0.85 the range in lone-pair angles increases to between 90-180°.

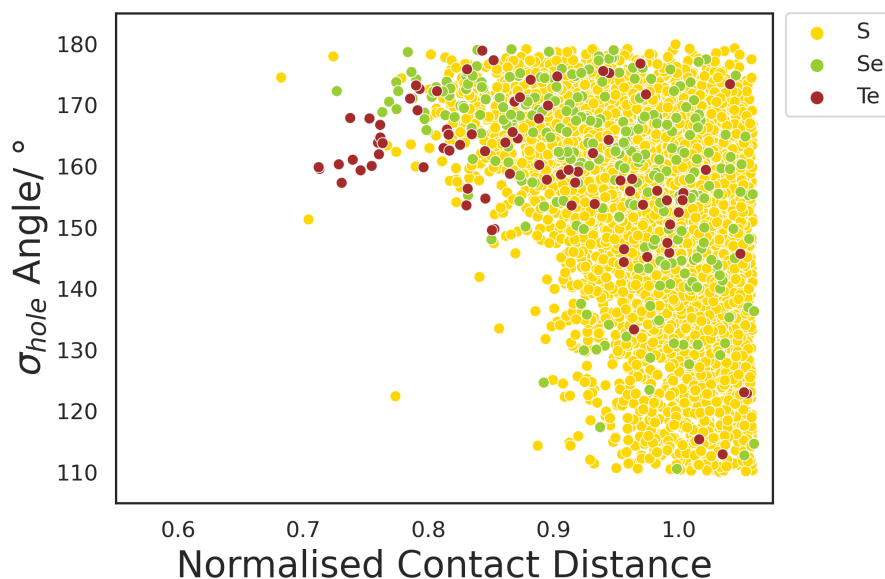


Figure 4.48: Scatter plots showing the distribution of normalised contact distance with respect to the σ -hole angle of chalcogen bonding contacts to oxide oxygens found in crystal structures in the CSD. Points colour coded by donor atom type

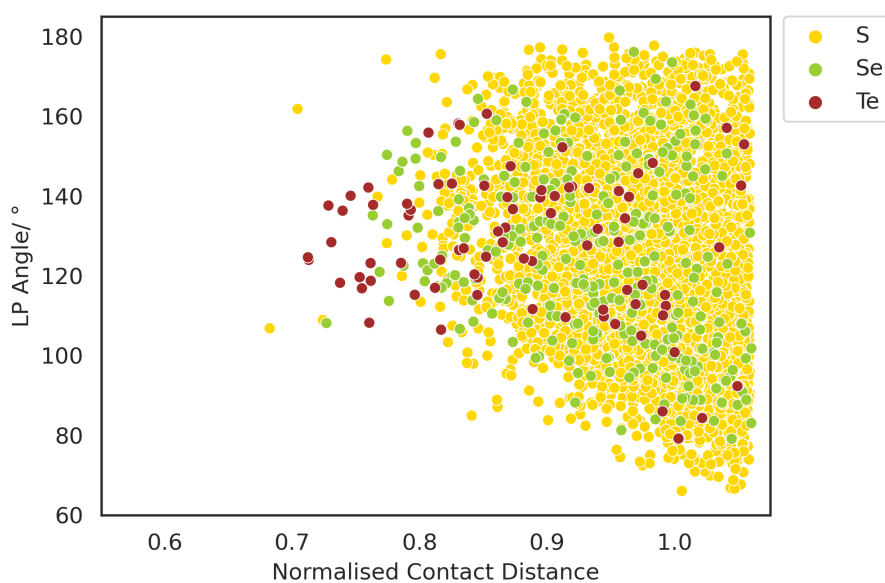


Figure 4.49: Scatter plots showing the distribution of normalised contact distance with respect to the lone-pair angle of chalcogen bonding contacts to oxide oxygens found in crystal structures in the CSD. Points colour coded by donor atom type

Table 4.40: Summary of geometric characteristics of chalcogen bonding contacts to oxide oxygens

Donor	Contacts	Mean Contact Distance/ Å	Mean Normalised Contact Distance	Mean σ -hole Angle/ °	Mean Lone-Pair Angle/ °
Ch	4811	3.291	0.971	152	123
Te	85	3.085	0.884	159	127
Se	280	3.107	0.936	160	124
S	4446	3.306	0.975	152	123

The trends among different terminal oxides were examined by subdivision of the data presented in Figure 4.48. The distributions of σ -hole angle against normalised contact distance for chalcogen bonding to the different types of terminal oxide are presented in Figure 4.50, and those same distributions colour coded by the valency of each type of oxide is shown in Figure 4.51. The mean averages of the geometric measurements for chalcogen bonding contacts to each type of terminal oxide are listed in Table 4.41. The trend in the number of contacts for each of the different types of oxides is the same as seen with the analogous halogen bonding contacts in Figure 4.45 & Table 4.39.

The distributions of σ -hole angle against normalised contact distance for chalcogen bonding contacts to nitrogen oxides and sulfur oxides in Figure 4.50 are somewhat similar. The main difference is that the closest contacts involve sulfur oxides and these include the very close sulfur contacts with σ -hole angles near linearity and the closest tellurium contacts with σ -hole angles around 160°. As seen in oxide valencies in the analogous halogen bonding contacts (see Figure 4.46), the contacts to nitrogen and sulfur oxides show similar trends in the valencies of the oxides, shown in Figure 4.51. The nitrogen oxides involved in chalcogen bonding are predominantly trivalent, with some contacts involving divalent nitrosyl moieties with little directionality with respect to the σ -hole angle. The chalcogen bonding contacts to sulfur oxides are predominantly tetravalent, with some tri- and divalent oxides occupying similar distributions of σ -hole angle and normalised contact distances.

For the chalcogen bonding contacts to halogen, metal, and phosphorus oxides, they occupy similar distributions, with the selenium contacts to metal oxides showing some clear relationship between σ -hole angles nearer linearity and shorter normalised contact distances. The chalcogen bonding contacts to halogen and phosphorus oxides involve exclusively tetravalent oxides, with Figure 4.52 showing primarily perchlorate ions involved in the chalcogen bonding contacts to halogen oxides. The chalcogen bonds to metal oxides generally involve a wider variety of metal oxide valence, shown in Figure 4.51, with the closest contacts with σ -hole angles nearing linearity comprising of tetravalent metal oxides, that Figure 4.52 reveals to be rhenium.

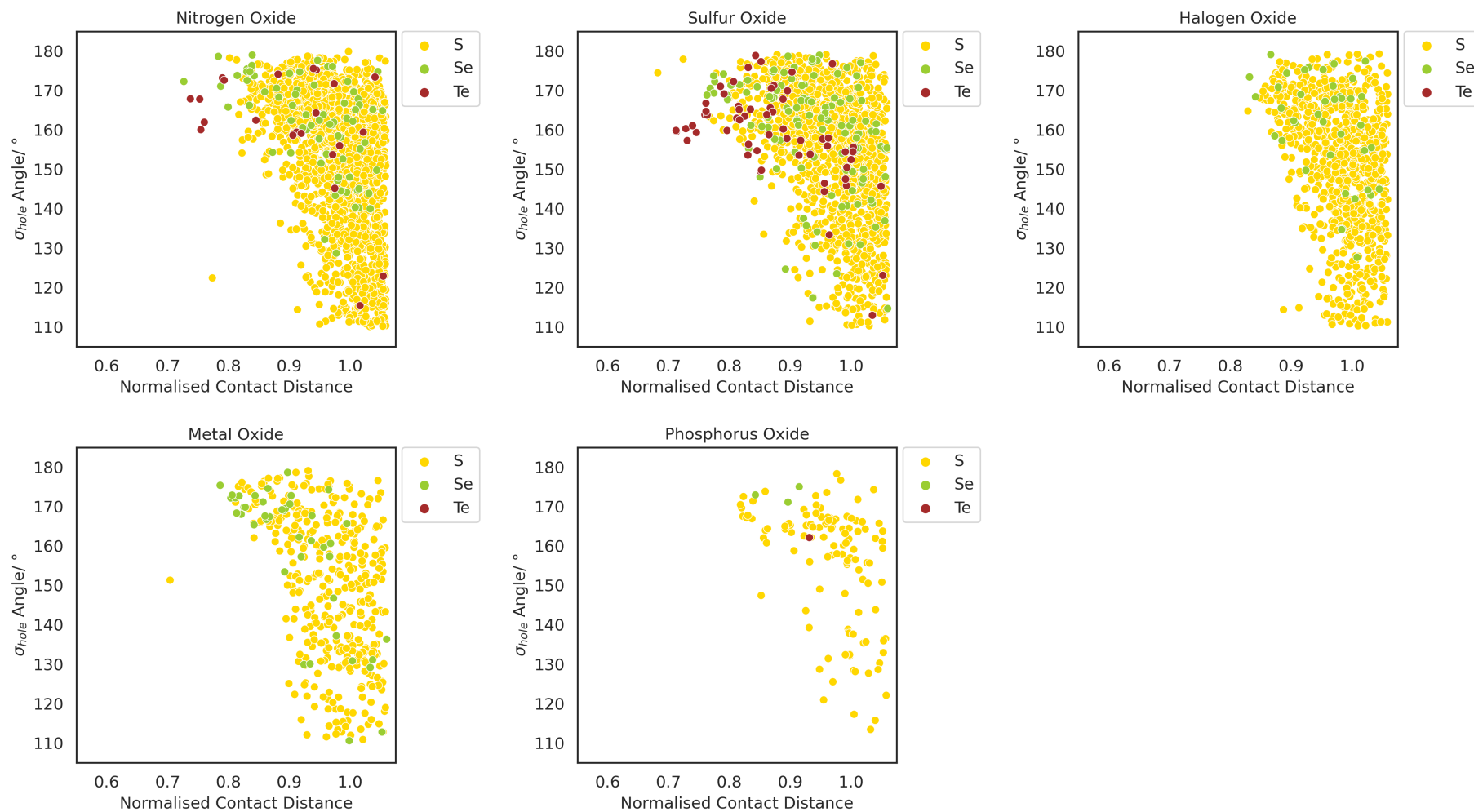


Figure 4.50: Scatter plots showing the distribution of normalised contact distance with respect to the σ -hole angle of chalcogen bonding contacts to various oxides found in crystal structures in the CSD. Points colour coded by donor atom

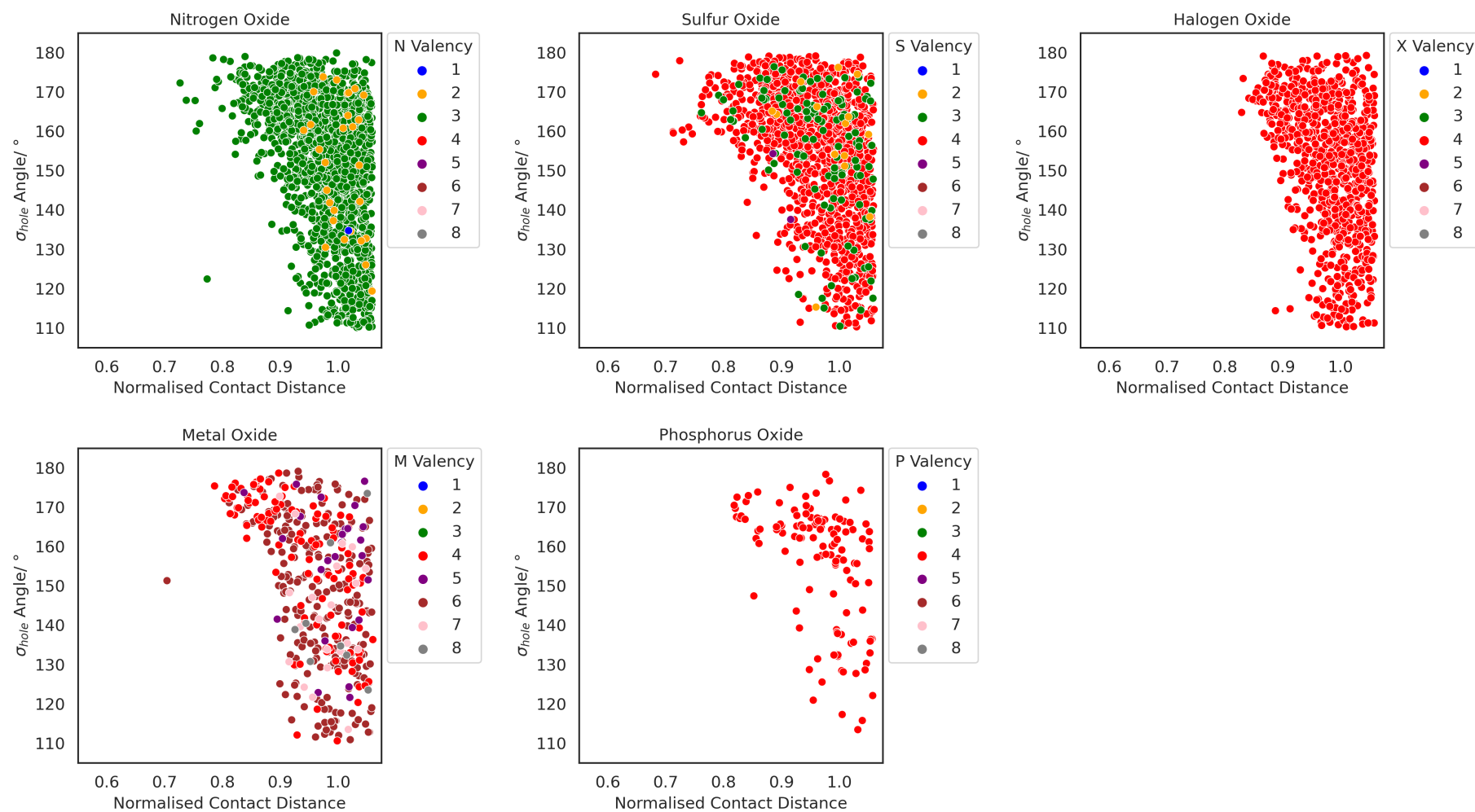


Figure 4.51: Scatter plots showing the distribution of normalised contact distance with respect to the σ -hole angle of chalcogen bonding contacts to various oxides found in crystal structures in the CSD. Points colour coded by valency of oxide element

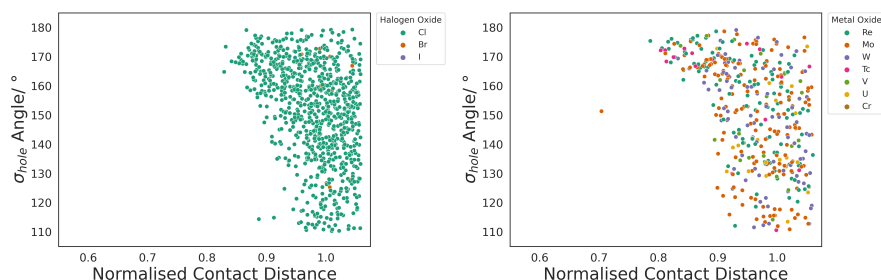


Figure 4.52: Scatter plots showing the distribution of normalised contact distance with respect to the σ -hole angle of chalcogen bonding contacts to halogen oxides (left), and metal oxides (right) found in crystal structures in the CSD. Points colour coded by halogen oxide atom, or metal oxide atom

Table 4.41: Summary of geometric characteristics of chalcogen bonding contacts to different classes of oxides

Functional Group	Contacts	Mean Contact Distance/ Å	Mean Normalised Contact Distance	Mean σ -hole Angle/ °	Mean Lone-Pair Angle/ °
Oxides	4811	3.291	0.971	152	123
Nitrogen Oxide	1827	3.316	0.978	152	122
Sulfur Oxide	1562	3.258	0.962	155	126
Halogen Oxide	864	3.321	0.981	151	120
Metal Oxide	434	3.248	0.960	150	120
Phosphorus Oxide	124	3.278	0.967	154	127

4.5.3 $Pn \cdots$ Oxide Oxygen Contacts

The least numerous of the σ -hole interactions to terminal oxide Lewis bases involve pnictogen bonding, see Table 4.37. The distribution of σ -hole angle against normalised contact distance for pnictogen bonding contacts to oxides is shown in Figure 4.53. A much broader form of the characteristic "wedge" shape for these plots is seen, with the range of σ -hole angles increasing from 145-180° at the shortest of normalised contact distances, to 130-180° at 0.9 normalised contact distance. The distribution seems to mainly involve a relationship between the normalised contact distance and the identity of the pnictogen donor species, with antimony contacts being the closest, followed by bismuth and then arsenic and phosphorus contacts being fairly similar to each other.

The mean averages of the geometric measurements for pnictogen bonding contacts to terminal oxide Lewis bases is listed in Table 4.42. The trend in average normalised contact distance with respect to the identity of the pnictogen atom is $Sb < Bi < As < P$. The average σ -hole angles varies little between pnictogen species, with phosphorus contacts having the largest σ -hole angles. There seems to be no relationship between the identity of the pnictogen bond donor atom and the lone-pair angle, with little variation in the lone-pair angles.

The distribution of lone-pair angle against normalised contact distance for pnictogen bonding contacts to oxides is shown in Figure 4.54. The range in lone-pair angle among the smallest normalised contact distances is 120-170° whereas there is a much broader range as the normalised contact distances get larger.

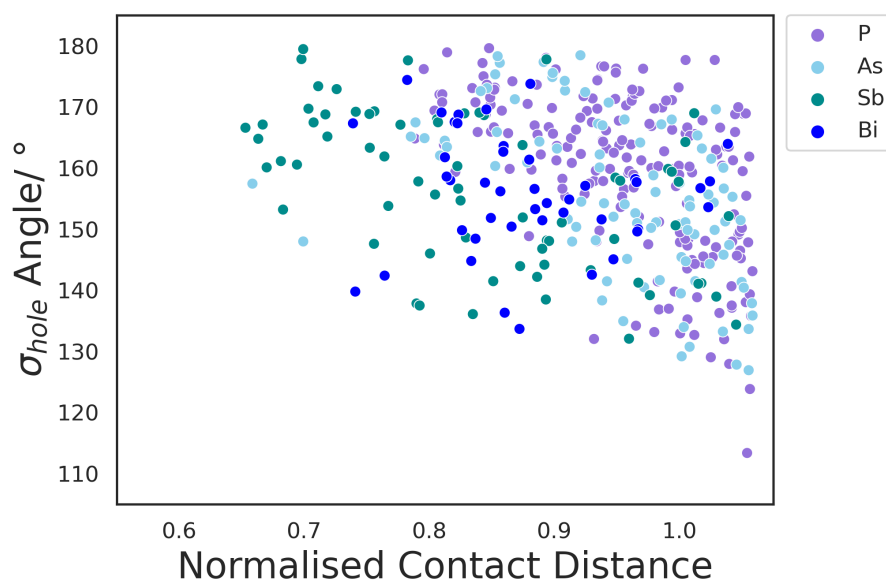


Figure 4.53: Scatter plots showing the distribution of normalised contact distance with respect to the σ -hole angle of pnictogen bonding contacts to oxide oxygens found in crystal structures in the CSD. Points colour coded by donor atom type

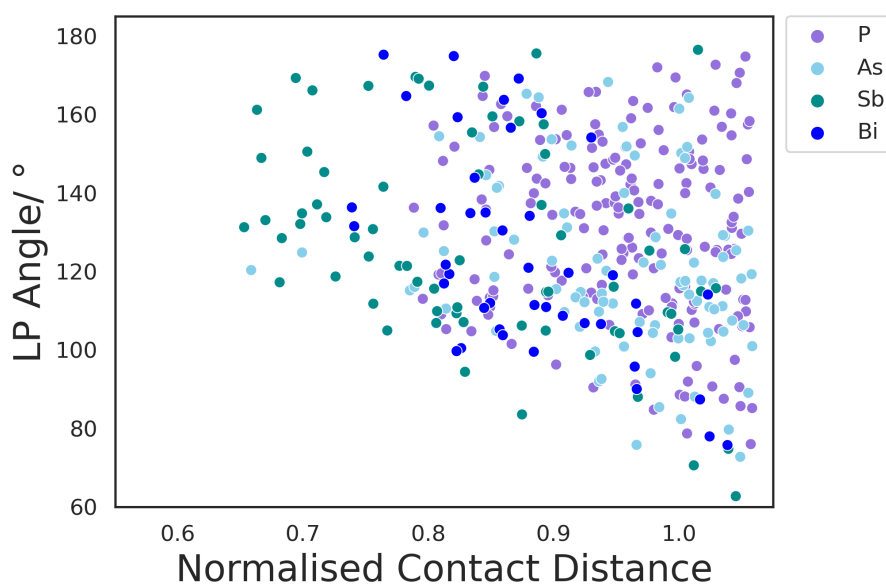


Figure 4.54: Scatter plots showing the distribution of normalised contact distance with respect to the lone-pair angle of pnictogen bonding contacts to oxide oxygens found in crystal structures in the CSD. Points colour coded by donor atom type

Table 4.42: Summary of geometric characteristics of pnictogen bonding contacts to oxide oxygens

Donor	Contacts	Mean Contact Distance/ Å	Mean Normalised Contact Distance	Mean σ -hole Angle/ °	Mean Lone-Pair Angle/ °
Pn	394	3.293	0.927	157	127
Bi	43	3.551	0.879	155	123
Sb	69	3.333	0.840	156	126
As	92	3.218	0.952	155	120
P	190	3.257	0.958	159	131

When comparing the pnictogen bonding contacts to different type of terminal oxide Lewis base, the different distributions of σ -hole angles against normalised contact distance for each oxide type is shown in Figure 4.55 and the mean averages for the geometric measurements are shown in Table 4.52. The majority of the contacts involve pnictogen bonding to sulfur oxides, that the same distributions colour coded by oxide valency in Figure 4.56 show to be tetravalent. The distributions of σ -hole angle against normalised contact distance that show the most typical "wedge" shape involve metal oxides and phosphorus oxides. The metal oxides involved are nearly exclusively pentavalent, as shown in Figure 4.56, and the distributions colour coded by metal oxide shown in Figure 4.57 reveal that these are pentavalent oxides of vanadium and molybdenum. The pnictogen bonding contacts to phosphorus oxides involve exclusively tetravalent phosphorus. Pnictogen bonding contacts to nitrogen oxides and halogen oxides are much less common and show little in the way of directionality with respect to the σ -hole angle.

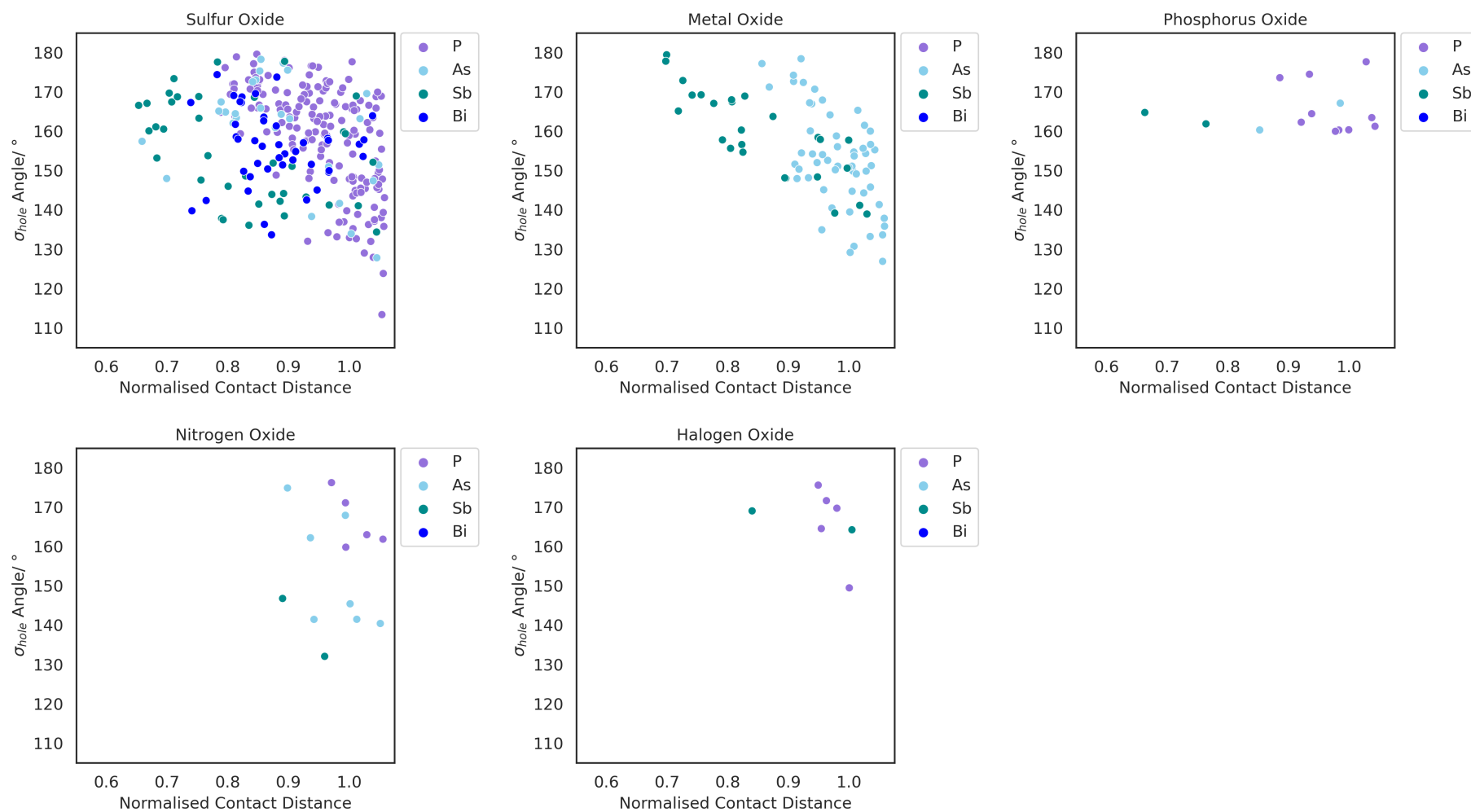


Figure 4.55: Scatter plots showing the distribution of normalised contact distance with respect to the σ -hole angle of pnictogen bonding contacts to various oxides found in crystal structures in the CSD. Points colour coded by oxide type

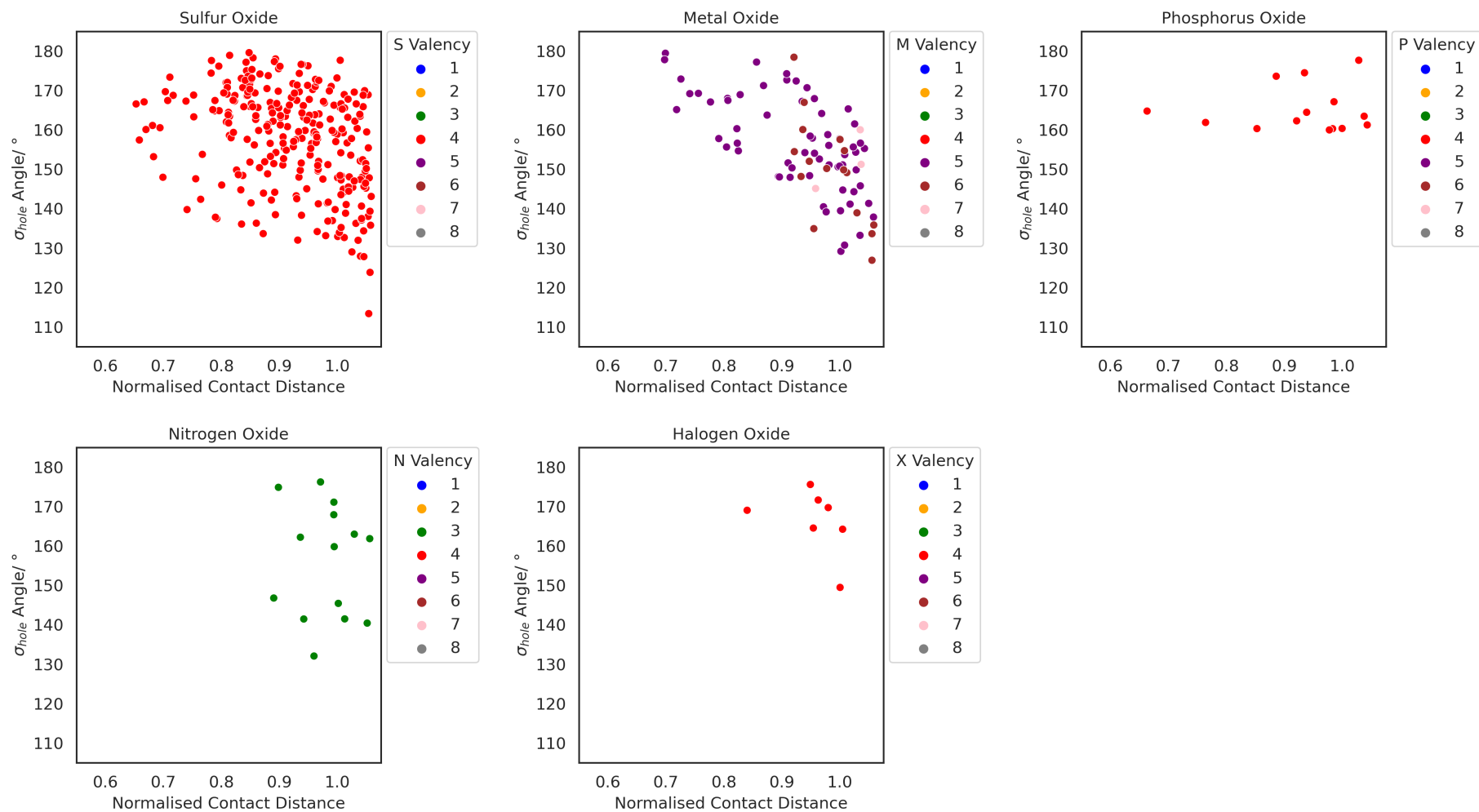


Figure 4.56: Scatter plots showing the distribution of normalised contact distance with respect to the σ -hole angle of pnictogen bonding contacts to various oxides found in crystal structures in the CSD. Points colour coded by valency of oxide element

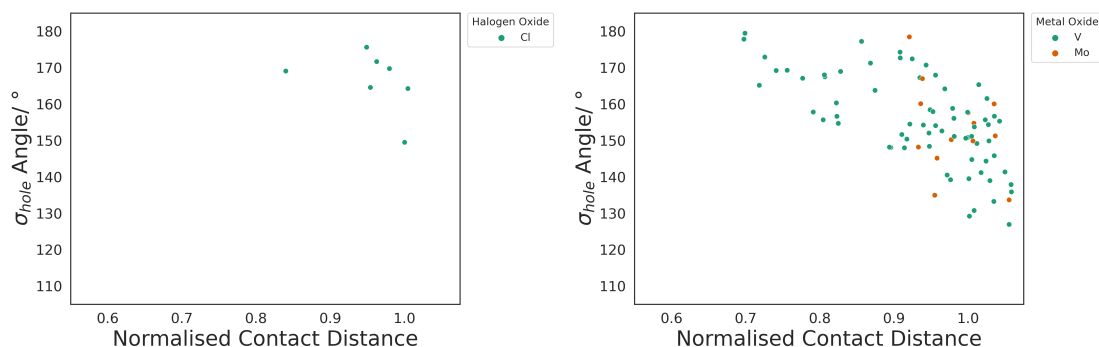


Figure 4.57: Scatter plots showing the distribution of normalised contact distance with respect to the σ -hole angle of pnictogen bonding contacts to halogen oxides (left), and metal oxides(right) found in crystal structures in the CSD. Points colour coded by halogen oxide atom, or metal oxide atom

Table 4.43: Summary of geometric characteristics of pnictogen bonding contacts to different classes of oxides

Functional Group	Contacts	Mean Contact Distance/ Å	Mean Normalised Contact Distance	Mean σ -hole Angle/ °	Mean Lone-Pair Angle/ °
Oxides	394	3.293	0.927	157	127
Sulfur Oxide	275	3.276	0.919	157	129
Metal Oxide	82	3.338	0.941	154	119
Phosphorus Oxide	16	3.223	0.934	165	134
Nitrogen Oxide	14	3.400	0.981	156	117
Halogen Oxide	7	3.400	0.956	166	129

4.5.4 *Tt* ... Oxide Oxygen Contacts

Table 4.37 shows that tetrel bonding contacts to terminal oxide Lewis bases are only slightly more numerous than the analogous pnictogen bonding contacts.

The distribution of σ -hole angle against normalised contact distance of pnictogen bonding contacts to terminal oxides is displaying in Figure 4.58. There exists a population of tin contacts with normalised contact distances ranging from 0.6-0.8 that have σ -hole angles near linearity. Silicon contacts seem to be biased towards σ -hole angles near linearity, but generally adopt normalised contact distances greater than the sum of van der Waals radii ($d_{norm} = 1.0$). Lead contacts to terminal oxides are generally very close, with normalised contact distances as low as 0.7, yet adopt a range of σ -hole angles from 130-180°. This population of lead contacts is also co-located with a distribution of mainly germanium and tin contacts, but also a small number of silicon contacts. This population of off-linear contacts means the aforementioned "wedge"-shaped distribution is not seen in this plot, this suggests that there could be other effects that are causing the close approach of predominantly lead atoms to oxide oxygens.

The mean averages of the geometric measurements of tetrel bonding contacts to terminal oxide Lewis bases is listed in Table 4.44. The average normalised contact distances decrease with increasing size of tetrel atom, following the trend Pb

$\text{Sn} < \text{Ge} < \text{Si}$. This is in contrast to what would be expected from a polarisability standpoint, with tin having a greater polarisability than lead (see Table 1.4), which is a major predictor for strong σ -hole interactions. Average σ -hole angles that approach linearity are only seen in contacts involving tin and silicon, with average σ -hole angles of 162° and 168° , respectively. The expected lone-pair angle of around 120° is seen with contacts involving germanium and tin, whereas silicon contacts adopt larger lone-pair angles, and lead contacts adopting smaller lone-pair angles.

The distribution of lone-pair angle against normalised contact distance of pnictogen bonding contacts to terminal oxides is shown in Figure 4.59. There is a clear preference for the closest contacts ($d_{\text{norm}} = 0.6$) for lone-pair angles of 120° yet for greater normalised contact distances the directionality is lost.

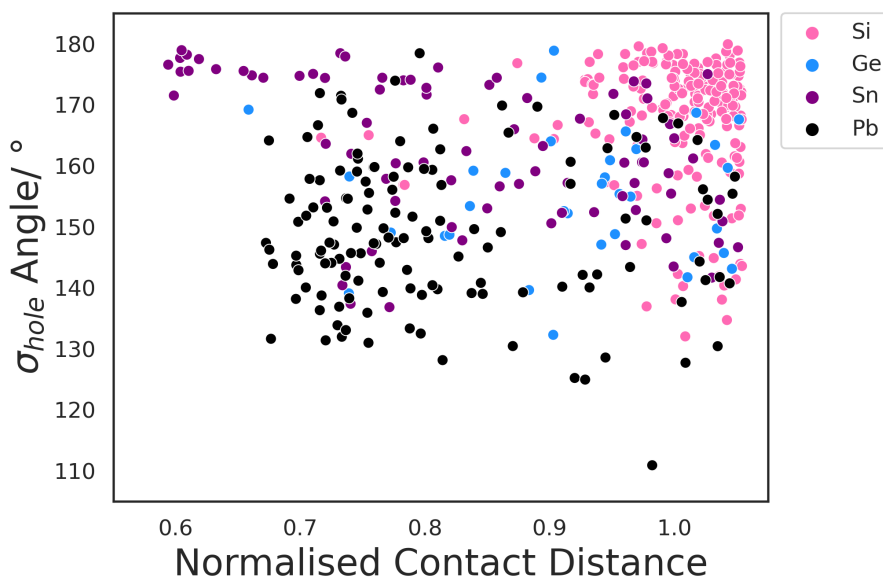


Figure 4.58: Scatter plots showing the distribution of normalised contact distance with respect to the σ -hole angle of tetrel bonding contacts to oxide oxygens found in crystal structures in the CSD. Points colour coded by donor atom type

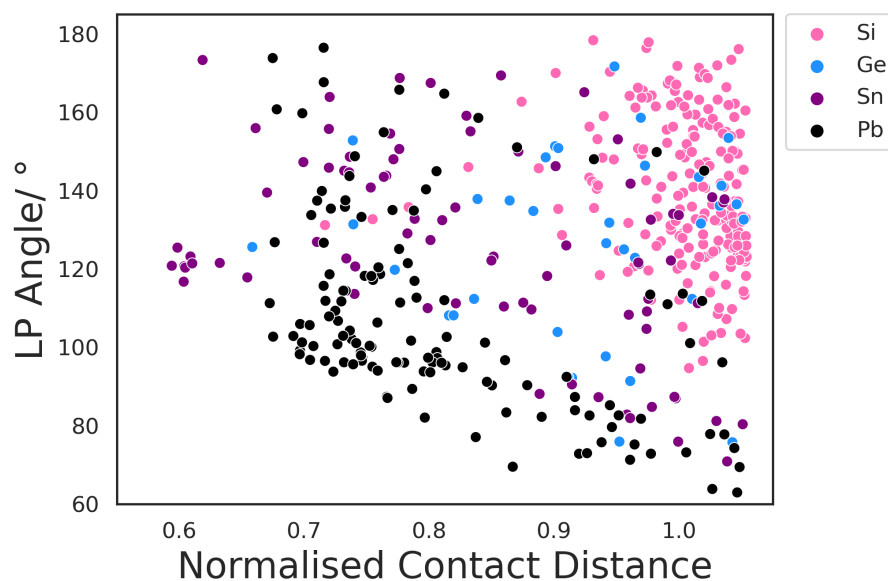


Figure 4.59: Scatter plots showing the distribution of normalised contact distance with respect to the lone-pair angle of tetrel bonding contacts to oxide oxygens found in crystal structures in the CSD. Points colour coded by donor atom type

Table 4.44: Summary of geometric characteristics of tetrel bonding contacts to oxide oxygens

Donor	Contacts	Mean Contact Distance/ Å	Mean Normalised Contact Distance	Mean σ -hole Angle/ °	Mean Lone-Pair Angle/ °
Tt	450	3.503	0.911	160	126
Pb	133	3.334	0.813	149	106
Sn	81	3.285	0.838	162	125
Ge	35	3.491	0.921	155	126
Si	201	3.705	1.004	168	139

The distributions of σ -hole angle against normalised contact distance of tetrel bonding contacts to different types of terminal oxides is shown in Figure 4.60. The mean averages of the geometric measurements for tetrel bonding contacts to different types of terminal oxides is listed in Table 4.45.

Tetrel bonding contacts to sulfur oxides make up the majority of the tetrel bonding contacts to oxides. These contacts to sulfur oxides along with tetrel bonding contacts to nitrogen oxides encompass the majority of the near linear contacts involving tin and silicon mentioned prior. The contacts to sulfur oxides contain close contacts that deviate substantially from linearity, which contains a variety of different tetrel atoms, whereas the close contacts to nitrogen oxides with σ -hole angles that deviate from linearity comprise entirely of lead contacts. The very short and highly linear contacts mentioned above correspond to tetrel bonding contacts to nitrogen oxides, more specifically from the work of Bryce and co-workers about the use of tetrel bonding of triphenyltin chloride to aromatic N-oxides in cocrystal engineering¹⁹⁵

The valencies of the sulfur oxides and nitrogen oxides involved in tetrel bonding contacts are shown in the same distributions

of σ -hole angle against normalised contact distance in Figure 4.61. The sulfur oxides are nearly entirely tetravalent, with a small number of trivalent sulfur oxides that include the two closest contacts within the plot. The nitrogen oxides are mainly trivalent, with two long contacts that involve divalent nitrogen which correspond to nitrosyl ligands.

For the tetrel bonding contacts to halogen oxides the distribution of σ -hole angle against normalised contact distance colour coded by oxide valency (see Figure 4.61) and colour coded by halogen oxide (see Figure 4.62) reveals that these contacts exclusively involve perchlorate moieties. These tetrel bonding contacts to perchlorates are dominated by close lead contacts with little directionality with respect to the σ -hole angle, showing σ -hole angles that deviate substantially from linearity. Tetrel bonding contacts to metal oxides and phosphorus oxides are much less numerous, and show little relationship between the σ -hole angle and shorter normalised contact distance, other than a single very close tin contact with a near linear σ -hole angle. This close tin contact to a phosphorus oxide corresponds to the crystal of 1-phenyl-1-oxo-4-methyl-4-chloro-1 λ^5 ,4-phosphastanninane which forms an inversion dimer of two Sn $\cdots$ O tetrel bonds, see Figure B.2¹⁹⁶.

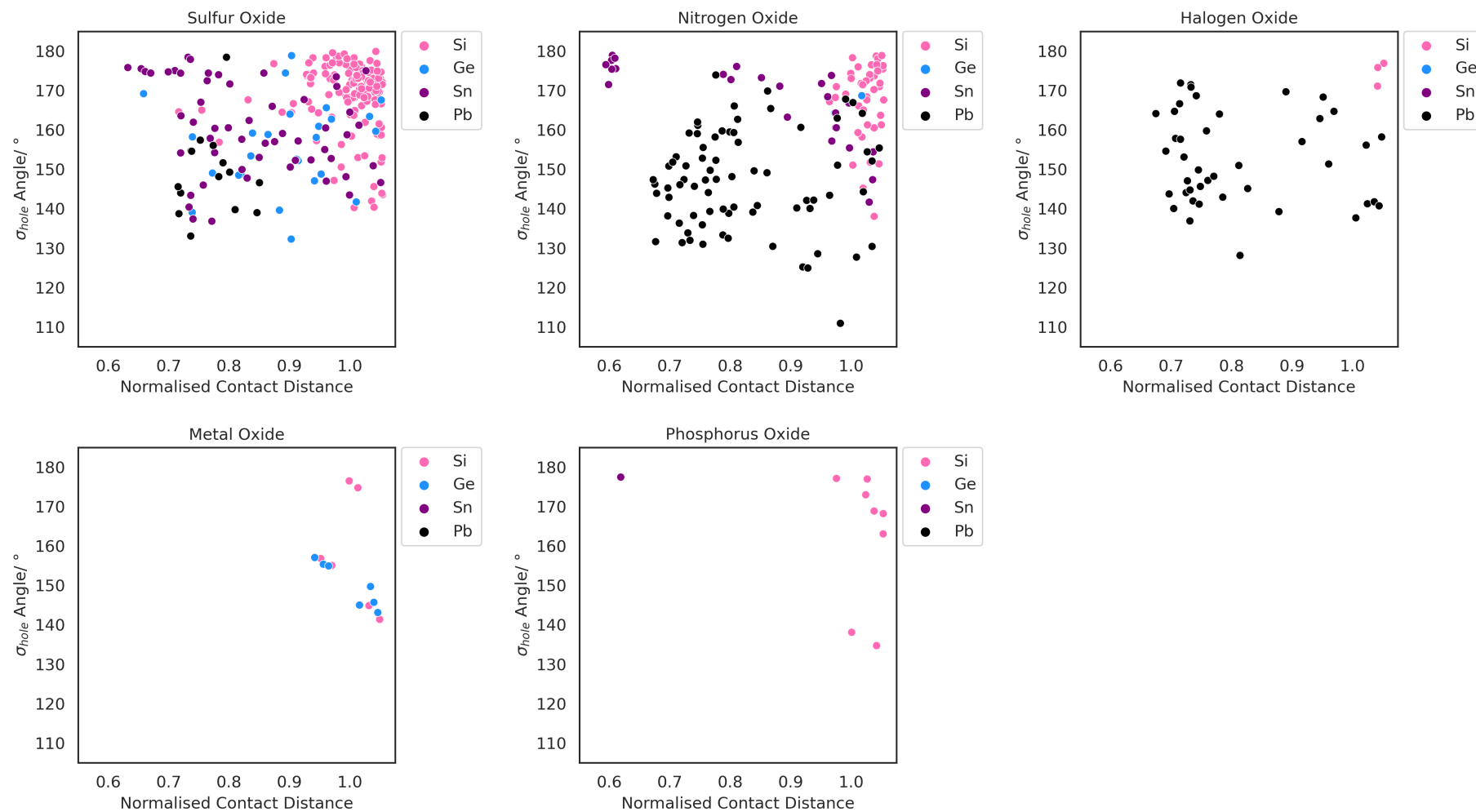


Figure 4.60: Scatter plots showing the distribution of normalised contact distance with respect to the σ -hole angle of tetrel bonding contacts to various oxides found in crystal structures in the CSD. Points colour coded by oxide type

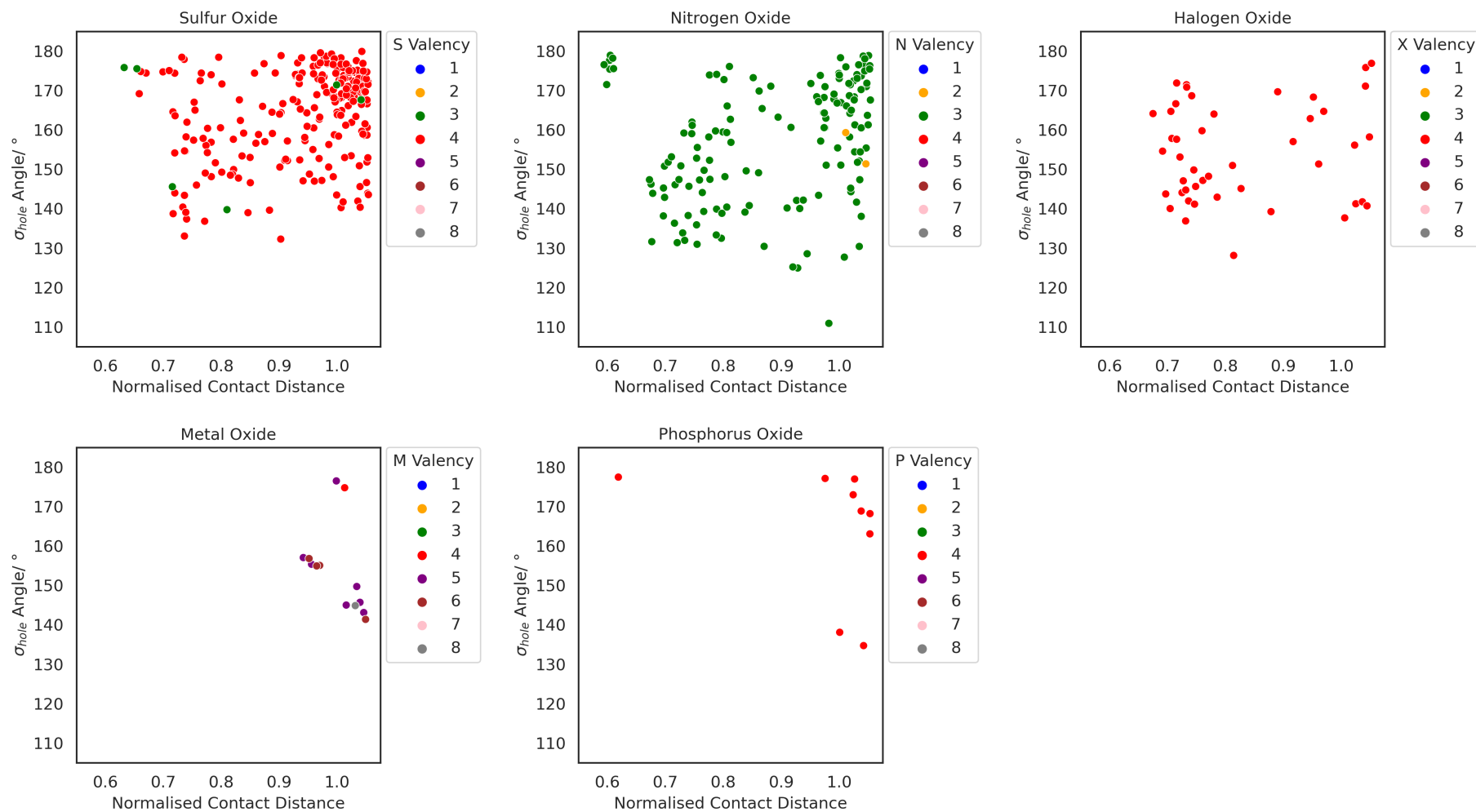


Figure 4.61: Scatter plots showing the distribution of normalised contact distance with respect to the σ -hole angle of tetrel bonding contacts to various oxides found in crystal structures in the CSD. Points colour coded by valency of oxide element

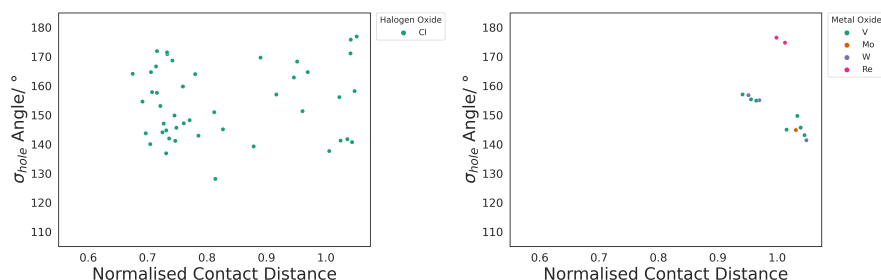


Figure 4.62: Scatter plots showing the distribution of normalised contact distance with respect to the σ -hole angle of tetrel bonding contacts to halogen oxides (left), and metal oxides(right) found in crystal structures in the CSD. Points colour coded by halogen oxide atom, or metal oxide atom

Table 4.45: Summary of geometric characteristics of tetrel bonding contacts to different classes of oxides

Functional Group	Contacts	Mean Contact Distance/ Å	Mean Normalised Contact Distance	Mean σ -hole Angle/ °	Mean Lone-Pair Angle/ °
Oxides	450	3.503	0.911	160	126
Sulfur Oxide	237	3.521	0.934	164	134
Nitrogen Oxide	145	3.488	0.889	156	116
Halogen Oxide	46	3.363	0.827	154	108
Metal Oxide	13	3.747	1.001	153	132
Phosphorus Oxide	9	3.637	0.981	164	153

4.5.5 Prevalence of $\sigma_{hole} \cdots$ Oxide Oxygen Contacts

The prevalence of σ -hole interactions to terminal oxide Lewis bases is detailed in Table 4.46. The prevalence of σ -hole interactions to oxides is considerably higher than the prevalence of σ -hole interactions to organic carbonyl and divalent oxygen σ -hole acceptors (see Table 4.10 & Table 4.28) and generally have higher prevalence than the σ -hole interactions to any of the hybridisations of nitrogen (see Table 3.6, Table 3.16, & Table 3.27).

Table 4.46: Overall summary of the prevalence of contacts indicating σ -hole interactions to terminal oxide moieties compared to the number of unique crystal structures found in the CSD that contain the necessary moieties.

σ -hole Interaction Type	Structures with Moieties	Structures with Contact(s)	Prevalence
Halogen-bonding	18716	5488	29.32%
Chalcogen-bonding	7796	2521	32.34%
Pnictogen-bonding	727	197	27.10%
Tetrel-bonding	2465	248	10.06%

The prevalence of halogen bonding to oxide Lewis bases, divided into each of the halogen bond donor atoms is shown in Table 4.47. The prevalence of the halogen bond increases with more polarisable halogen bond donor atoms, following the trend $\text{Cl} < \text{Br} < \text{I}$. The prevalence of chlorine and bromine halogen bonds to oxides are somewhat comparable, whereas the prevalence of iodine halogen bonds to oxide σ -hole acceptors is over 50%. This high prevalence along with the close

contacts with near linear σ -hole angles involving iodine shown in the distributions of σ -hole angle against normalised contact distance in Figure 4.43, suggests that iodine halogen bonding to terminal oxide moieties is a useful halogen bonding motif. The prevalence of halogen bonding to oxide Lewis bases, divided into each of different types of terminal oxide is shown in Table 4.48. The two types of oxide that are involved in halogen bonds with prevalences notably higher than the overall average are nitrogen oxides and halogen oxides, with prevalences of 38.75% and 34.68%, respectively.

Table 4.47: Summary of the prevalence of halogen bonding contacts to oxide oxygens compared to the number of unique crystal structures found in the CSD that contain the necessary moieties

Donor	Structures with Moieties	Structures with Contact(s)	Prevalence
X	18716	5488	29.32%
I	1502	753	50.13%
Br	5394	1591	29.50%
Cl	12378	3180	25.69%

Table 4.48: Summary of the prevalence of halogen bonding contacts to various classes of oxide compared to the number of unique crystal structures found in the CSD that contain the necessary moieties

Functional Group	Structures with Moieties	Structures with Contact(s)	Prevalence
Oxides	18716	5488	29.32%
Sulfur Oxide	8825	2134	24.18%
Nitrogen Oxide	6753	2617	38.75%
Halogen Oxide	966	335	34.68%
Metal Oxide	1174	246	20.95%
Phosphorus Oxide	1633	187	11.45%
Other	29	6	20.69%

Table 4.46 shows that chalcogen bonds to terminal oxide Lewis bases have the highest prevalence of the σ -hole interactions to oxides. The prevalence of chalcogen bonding to oxide Lewis bases, divided into each of the chalcogen bond donor atoms is shown in Table 4.49. Sulfur dominates the crystals containing the required chalcogen bond donor and oxide moieties and those that contain chalcogen bonding. Selenium contacts, while much less common has a comparable prevalence to sulfur contacts, but is much lower in prevalence than the tellurium contacts. It is important to note that the high prevalence of tellurium contacts is out of a very small number of crystal structures containing the necessary moieties (45). The prevalence of chalcogen bonding to oxide Lewis bases, divided into each of different types of terminal oxide is shown in Table 4.50. The standout in terms of prevalence are chalcogen bonds to halogen oxides, which is unusual as the distributions of σ -hole angle against normalised contact distance for chalcogen bonding to halogen oxides (see Figure 4.50) shows generally longer contacts with little relationship between σ -hole angles approaching linearity and shorter normalised contact distances.

Table 4.49: Summary of the prevalence of chalcogen bonding contacts to oxide oxygens compared to the number of unique crystal structures found in the CSD that contain the necessary moieties

Donor	Structures with Moieties	Structures with Contact(s)	Prevalence
Ch	7796	2521	32.34%
Te	45	26	57.78%
Se	455	160	35.16%
S	7317	2342	32.01%

Table 4.50: Summary of the prevalence of chalcogen bonding contacts to various classes of oxide compared to the number of unique crystal structures found in the CSD that contain the necessary moieties

Functional Group	Structures with Moieties	Structures with Contact(s)	Prevalence
Oxides	7796	2521	32.34%
Nitrogen Oxide	3058	993	32.47%
Sulfur Oxide	3438	974	28.33%
Halogen Oxide	565	328	58.05%
Metal Oxide	457	152	33.26%
Phosphorus Oxide	497	92	18.51%
Other	9	5	55.56%

Table 4.46 shows that pnictogen bonds to terminal oxide Lewis bases have the comparable prevalence to halogen and chalcogen bonds to oxides, despite being much less numerous. The prevalence pnictogen bonding contacts to oxide σ -hole acceptors with respect to the different pnictogen species is shown in Table 4.51, and with respect to the different types of oxide in Table 4.52. Other than phosphorus contacts to oxides, the heavier pnictogen bond donors have higher prevalence than the overall average, yet are prevalences from a small pool of structures containing the required moieties. The standout terminal oxide σ -hole acceptors are metal and sulfur oxides, with prevalences of 26.52% and 33.26%, respectively. This high prevalence of pnictogen bonds to oxides out of a small number of structures, along with the close contacts shown in the distributions of σ -hole angle against normalised contact distance shown in Figure 4.53 suggest that pnictogen bonding to oxides are proof-of-concept structures or the interactions occur coincidentally.

Table 4.51: Summary of the prevalence of pnictogen bonding contacts to oxide oxygens compared to the number of unique crystal structures found in the CSD that contain the necessary moieties

Donor	Structures with Moieties	Structures with Contact(s)	Prevalence
Pn	727	197	27.10%
Bi	43	29	67.44%
Sb	71	30	42.25%
As	110	39	35.45%
P	507	100	19.72%

Table 4.52: Summary of the prevalence of pnictogen bonding contacts to various classes of oxide compared to the number of unique crystal structures found in the CSD that contain the necessary moieties

Functional Group	Structures with Moieties	Structures with Contact(s)	Prevalence
Oxides	727	197	27.10%
Metal Oxide	132	35	26.52%
Sulfur Oxide	427	142	33.26%
Nitrogen Oxide	60	8	13.33%
Phosphorus Oxide	96	8	8.33%
Halogen Oxide	17	5	29.41%
Other	1	0	N/A

Tetrel bonds to terminal oxide σ -hole acceptors prove to be the least prevalent out of the σ -hole interactions to oxides, see Table 4.46. The prevalence of tetrel bonds to terminal oxide Lewis bases with respect to the different tetrel species is shown in Table 4.53, and with respect to the different types of terminal oxide in Table 4.54.

This low overall prevalence can be attributed to the large number of silicon containing structures and those containing tetrel bonds, which dominate. Germanium tetrel bonds to oxides are much the same as the overall average prevalence, while tin and lead contacts have considerably higher prevalence, with 38.89% and 71.43%, respectively. It is important to note that, other than silicon containing crystals, there are a very small number of crystal structures containing the required tetrel bond donor and terminal oxide σ -hole acceptor moieties, which makes the values less reliable to compare against more numerous σ -hole interaction motifs. Of the prevalences of tetrel bonds to different types of oxide in Table 4.54 the standout are tetrel bonds to halogen oxides with a prevalence of 42.86%.

This high prevalence along with the close lead contacts to halogen oxides seen in the distributions of σ -hole angle against normalised contact distance in Figure 4.60 suggests that there are consistently used motifs that results in an attractive interaction between lead centres and oxide moieties. The substantial deviation from linearity of the σ -hole angles suggests other factors than tetrel bonding could be involved, and on inspection of crystal structures containing these lead contacts to perchlorates, they seem to mainly involve tetravalent Pb^{2+} species that adopt hemidirected geometries rather than

tetrahedral, suggesting the ionic interaction is perturbing the tetrel bonds. An example of this type of geometry can be seen in Figure 4.68.

Table 4.53: Summary of the prevalence of tetrel bonding contacts to oxide oxygens compared to the number of unique crystal structures found in the CSD that contain the necessary moieties

Donor	Structures with Moieties	Structures with Contact(s)	Prevalence
Tt	2465	248	10.06%
Pb	42	30	71.43%
Sn	126	49	38.89%
Ge	147	16	10.88%
Si	2168	153	7.06%

Table 4.54: Summary of the prevalence of tetrel bonding contacts to various classes of oxide compared to the number of unique crystal structures found in the CSD that contain the necessary moieties

Functional Group	Structures with Moieties	Structures with Contact(s)	Prevalence
Oxides	2465	248	10.06%
Metal Oxide	513	10	1.95%
Sulfur Oxide	1262	148	11.73%
Nitrogen Oxide	548	68	12.41%
Phosphorus Oxide	173	8	4.62%
Halogen Oxide	28	12	42.86%
Other	15	3	20.00%

4.5.6 Notable Oxide Functional Groups

Over the course of Section 4.5, there have been contacts found within the CSD that show notably strong σ -hole interaction character, with very close contacts and σ -hole angles near linearity. For halogen bonding contacts, the shortest contacts with the largest σ -hole angles correspond to interactions with nitrogen oxides. These contacts mostly involve aromatic N-oxides such as pyridine N-oxide paired with strong halogen bond donors such as N-halosuccinamides or I₂ to produce a very strong halogen bond. An example shown in Figure 4.63 of CSD refcode **ZUYZOK** corresponds to complex of N-iodosuccinamide with 4-phenylpyridine N-oxide, as part of a study by Rissanen and co-workers into very strong halogen bonds¹⁹⁷. Significantly strong halogen bonding contacts are also formed with sulfur oxides, with the strongest involving tetravalent sulfur oxide anions such as sulfonates or sulfates, an example of which is shown in Figure 4.64 of CSD refcode **GEPFUF**. This work by Huber *et al.* relates to the field of anion recognition, in this case using haloimidazolium salts and the acceptor is trifluoromethanesulfonate also known as triflate¹⁹⁸. These haloimidazolium functional groups have seen widespread use in anion recognition, and are examples of charge-assisted halogen bonds, being also responsible for close contacts to other anions such as phosphate and even perrhenate^{199 200}.

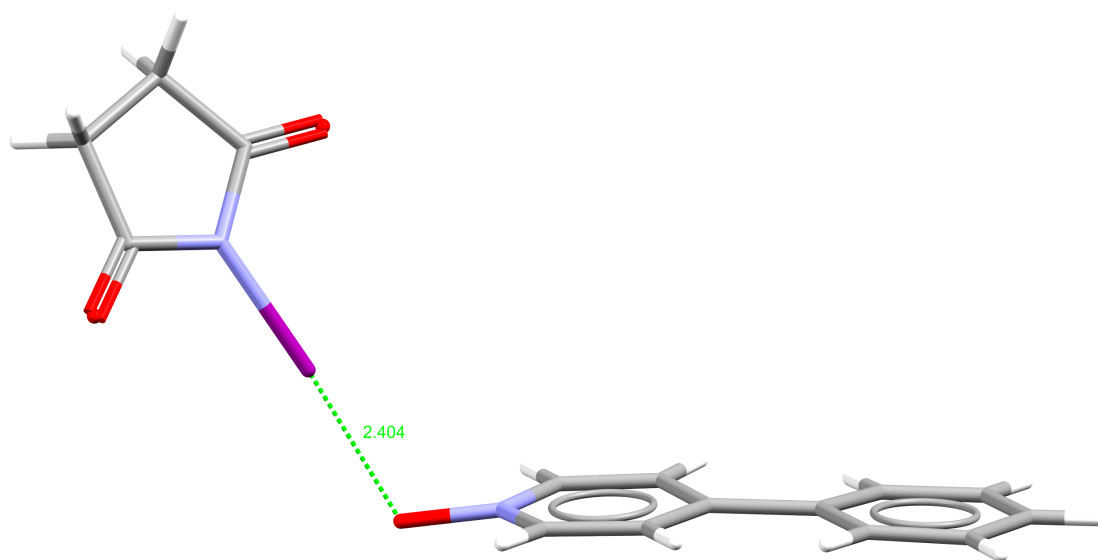


Figure 4.63: An example of a crystal structure with a halogen bonding contact involving N-iodosuccinimide and 4-phenylpyridine 1-oxide as the halogen bond acceptor. Distance given in Ångström. CSD refcode: **ZUYZOK**

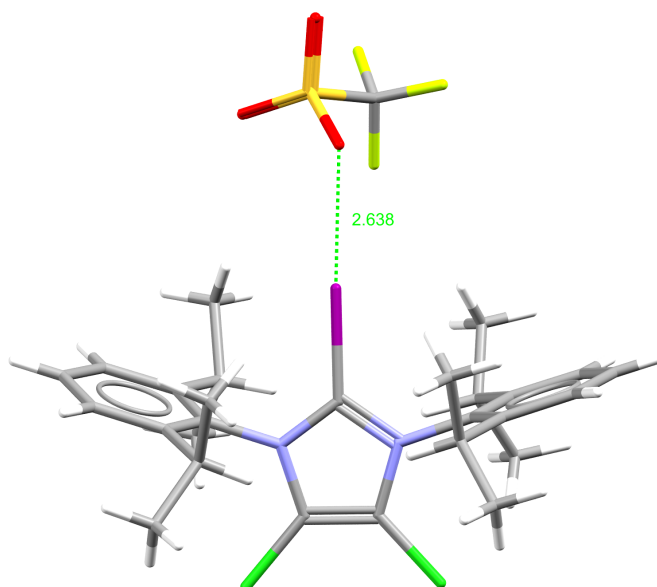


Figure 4.64: An example of a crystal structure with a halogen bonding contact involving iodoimidazolium and triflate as a halogen bond acceptor. Distance given in Ångström. CSD refcode: **GEPFUF**

Similar approaches to anion recognition have been applied using chalcogen bond donors, for example to recognise and precipitate perrhenate and pertechnetate ions (ReO_4^- and TcO_4^-)²⁰¹. An example of this is shown in Figure 4.65 which features a complex of a selenodiazolium cation and a technetate anion forming a charge-assisted chalcogen bond (CSD refcode **YEJYUL**). The close tellurium contacts to sulfur oxides that show σ -hole angles of around 160° primarily involve novel tellurium dications, with their behaviour being similar to N-heterocyclic carbenes by having a formally empty p -orbital

and a lone-pair. An example of this is shown in Figure 4.66 of a complex of one of these dicationic tellurium species with triflate²⁰².

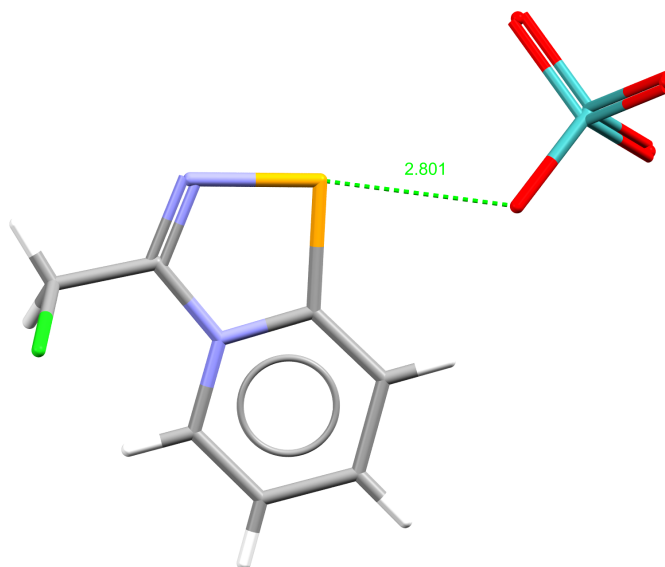


Figure 4.65: An example of a crystal structure with a chalcogen bonding contact involving 3-(chloromethyl)[1,2,4]selenadiazolo[4,5-a]pyridin-4-ium and a pertechnetate anion as a halogen bond acceptor. CSD refcode: **YEJYUL**

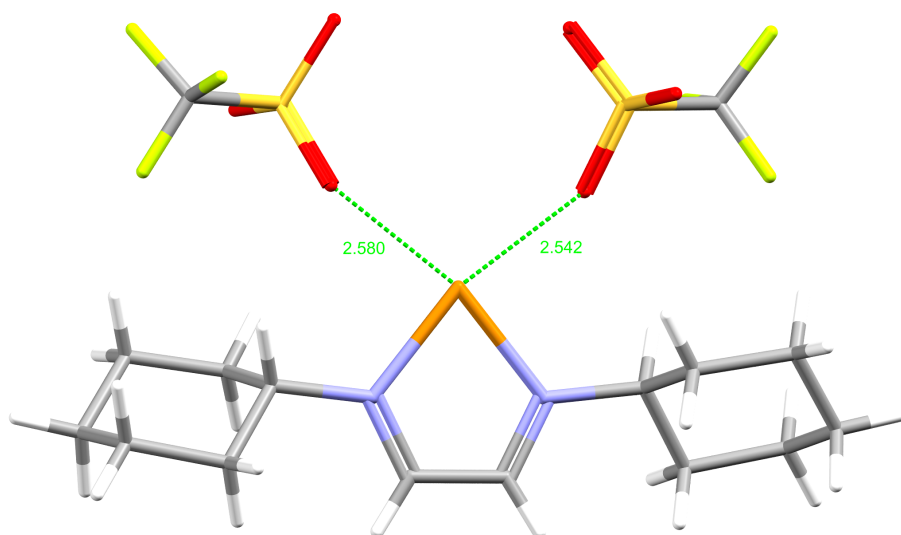


Figure 4.66: An example of a crystal structure with a chalcogen bonding contact involving (1,4-dicyclohexyl-1,4-diazabuta-1,3-diene)-tellurium(II) and a triflate anion as a halogen bond acceptor. CSD refcode: **LAKDIM**

Pnictogen bonding contacts show an interesting number of contacts to metal oxides, which correspond to an unusual set of structures featuring clusters of pnictogen polyoxovanadates. These are a type of polyoxometalate (POM) which are of

notable interest in supramolecular and materials chemistry, among other fields, owing to their interesting electronic and magnetic properties^{203–205}. An example of this is the structure of an antimony derivative of the $[V_{18}O_{42}]$ family of POMs shown in Figure 4.67 (CSD Refcode: **PEKGES**)²⁰⁶. This POM, and others like it show molecular magnetic behaviour, potentially allowing advances in future battery technology and materials²⁰⁷.

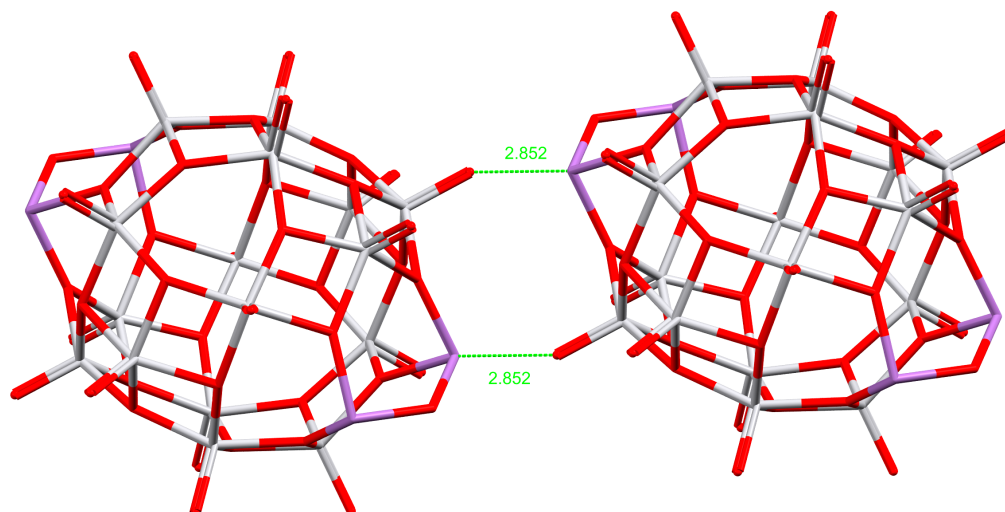


Figure 4.67: An example of a crystal structure with a pnictogen bonding contact involving an antimony derivative of $[V_{18}O_{42}]$ polyoxometallate, where the antimony moiety is the pnictogen bond donor and the $O=V$ moiety is the σ -hole acceptor. CSD refcode: **PEKGES**

Following from the close tetrel bonding complexes involving lead and halogen oxides what have σ -hole angles that deviate from linearity mentioned in Section 4.5.5, these can be attributed to macrocyclic lead compounds, with a range of aims such as a fluorescence behaviour or as precursor/template for further macrocycles^{208 209}. An example involving this interaction the Pb(II)-azamacrocyclic with a perchlorate anion (CSD refcode **QANDER**) shown in Figure 4.68, which notably is a potential explosive²¹⁰.

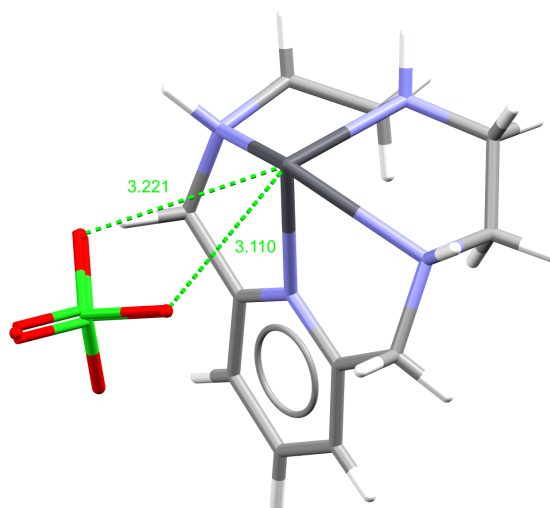


Figure 4.68: An example of a crystal structure with a tetrel bonding contact involving a lead macrocycle with hemidirected Pb coordination geometry and perchlorate. CSD refcode: **QANDER**

4.6 Metal Carbonyls

The research into metal carbonyls is a key part in the field of organometallic chemistry, displaying a wide variety of function and use. Carbonyls act as a π -acceptor ligand, thus the identity of the metal, its oxidation state, and remaining coordination sphere should impact the carbonyl's ability to act as an acceptor for σ -hole interactions. Another factor that can effect the Lewis basicity of a carbonyl ligand is back donation from the d -orbitals of the metal into the π^* -orbitals of the carbonyl ligand, which weakens the C-O bond. Carbonyls not only exist in a terminal M-C $\equiv$ O bonding mode, but can also bridge two or three metal centres, leading to increased oxygen basicity²¹¹.

Halogen and hydrogen bonding contacts have been observed in crystal structures of metal carbonyls, including both finite and extended network assemblies^{141 212}. The work of Brammer *et al.* indicates that the approach of the halogen or hydrogen donor atom have X/H $\cdots$ O $\equiv$ C angles, referred to the lone-pair angles in this manuscript, closer to 90°, which they attribute to the involvement of the π -system of the carbonyl acting as the acceptor. In this work the different bridging modes of metal carbonyls and how it related to σ -hole interactions are explored, but these systems are much less common, so the majority of the focus is placed on terminal metal carbonyls. Within the terminal metal carbonyl species, the identity of and the number of bonded substituents on the metal are investigated to see if any trends arise in Lewis basicity. It is important to note that the number of bonded substituents is not an exact substitute for valency, with hapticity being treated by the CSD as multiple bonds rather than as one. An example of this is a cyclopentadienyl ligand with η^5 hapticity would be counted as 5 bonds, so the iron in a ferrocene molecule would have 10 bonds. This makes any trends associated with the number of connected atoms to the metal centre as defined by the CSD less reliable,

due to it not being a like-for-like with coordination number and is not a descriptor of geometry. An overall summary of the geometric observables of contacts indicative of σ -hole interactions to metal carbonyl oxygen moieties is tabulated in Table 4.55.

Table 4.55: Overall geometric summary of the contacts indicating σ -hole interactions to terminal oxide oxygen moieties

σ -hole Interaction Type	No. Contacts	Mean Normalised Contact Distance	Mean σ -hole Angle/ °	Mean Lone-Pair Angle/ °
Halogen-bonding	2790	1.005	144	113
Chalcogen-bonding	888	1.003	148	116
Pnictogen-bonding	70	0.999	152	126
Tetrel-bonding	231	1.006	168	135

4.6.1 X ... Metal Carbonyl Oxygen Contacts

The most numerous of the contacts shown in Table 4.55, halogen bonding contacts to metal carbonyl oxygens have an average normalised contact distance greater than 1.0, the smallest average σ -hole angle, and the smallest average lone-pair angle compared to the other σ -hole interactions to metal carbonyls.

The distribution of σ -hole angle against normalised contact distance for halogen bonding contacts to metal carbonyl moieties is shown in Figure 4.69 and the mean averages of the geometric measurements of the contacts shown in Table 4.56 & Table 4.57. There are very few contacts with normalised contact distances less than 0.9, and very little relationship between shorter normalised contact distances and σ -hole angles closer to linearity. There are some notable chlorine contacts with normalised contact distances less than 0.9, but with σ -hole angles that deviate substantially from linearity. These contacts correspond to two instances of dichloromethane and one instance of chloroform as solvent within metal carbonyl crystal structures, which mean the σ -hole angle of these contacts are more easily perturbed from linearity to satisfy crystal packing requirements. The majority of the contacts in Figure 4.69 correspond to terminal carbonyl ligands, but with a small number of di- and tri-bridging metal carbonyls which seem to show a more notable relationship between σ -hole angles closer to linearity and shorter normalised contact distances. The mean averages of the geometric measurements of halogen bonding contacts to metal carbonyl moieties shows very little change in average normalised contact distance, average σ -hole angle, and average lone-pair angle with respect to changing halogen bond donor species.

The distribution of lone-pair angle against normalised contact distance for halogen bonding contacts to metal carbonyl moieties is shown in Figure 4.70. There is some bias towards the halogen bonding contacts approaching at lone-pair angles nearer 90° rather than the expected 120° in agreement with Brammer *et al.*¹⁴¹. The aforementioned very close chlorine contacts adopt lone-pair angles very close to 90° suggesting alignment with the π -system of the carbonyl ligand rather than the lone-pair.

The effect that the nature of the metal species and its bonding environment has on terminal metal carbonyls can be seen in the distributions of σ -hole angle against normalised contact distance in Figure 4.71 and the mean averages of the

geometric measurements in Table 4.58. There is a wide variety of different metals and coordination spheres, with the most frequently occurring contacts involving ruthenium, iron, and rhenium. The majority of the metal carbonyls feature a metal with six bonded substituents, suggesting them to be predominantly octahedral. The shortest average normalised halogen bonding contacts to carbonyl ligands involve those bound to rhenium and molybdenum, with average normalised contact distances of 0.997 and 0.995 respectively.

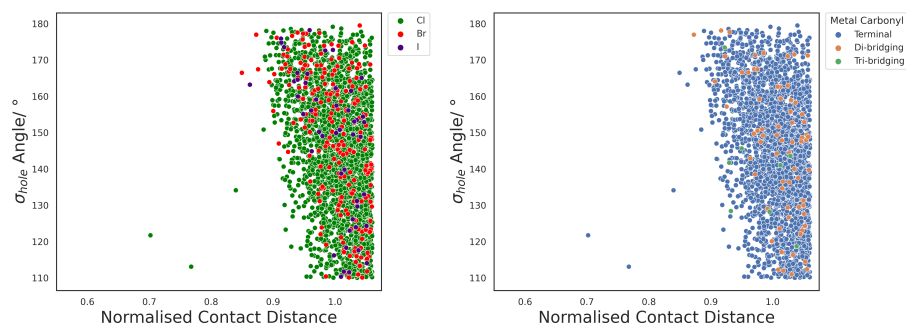


Figure 4.69: Scatter plots showing the distribution of normalised contact distance with respect to the σ -hole angle of halogen bonding contacts to terminal, di-bridging, and tri-bridging metal carbonyls found in crystal structures in the CSD. Points colour coded by donor atom type (left), and the bridging mode of the carbonyl (right)

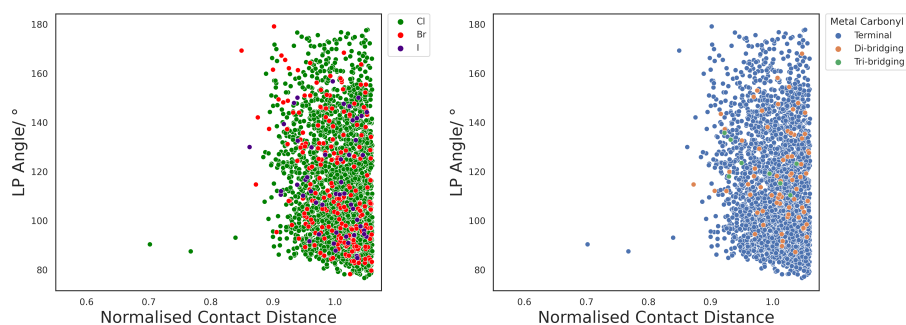


Figure 4.70: Scatter plots showing the distribution of normalised contact distance with respect to the lone-pair angle of halogen bonding contacts to terminal, di-bridging, and tri-bridging metal carbonyls found in crystal structures in the CSD. Points colour coded by donor atom type (left), and the bridging mode of the carbonyl (right)

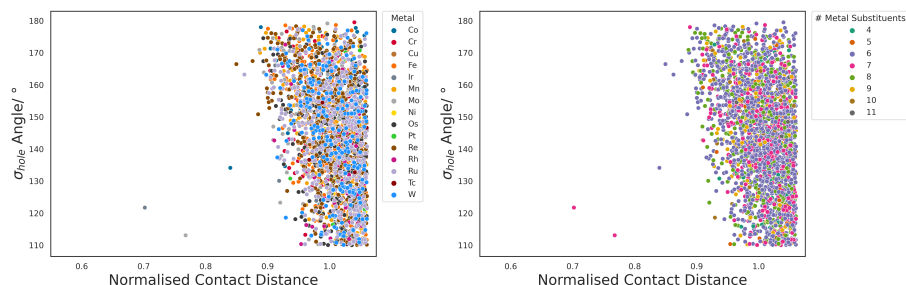


Figure 4.71: Scatter plots showing the distribution of normalised contact distance with respect to the lone-pair angle of halogen bonding contacts to terminal metal carbonyls found in crystal structures in the CSD. Points colour coded by the metal that the carbonyl is coordinated to (left) and the number of substituents coordinated to the metal (right).

Table 4.56: Summary of geometric characteristics of halogen bonding contacts to metal carbonyl oxygens

Donor	Contacts	Mean Contact Distance/ Å	Mean Normalised Contact Distance	Mean σ -hole Angle/ °	Mean Lone-Pair Angle/ °
X	2790	3.343	1.005	144	113
I	49	3.526	0.996	147	116
Br	217	3.359	1.000	148	115
Cl	2524	3.339	1.006	144	112

Table 4.57: Summary of geometric characteristics of halogen bonding contacts to terminal, di-bridging, and tri-bridging metal carbonyls

Functional Group	Contacts	Mean Contact Distance/ Å	Mean Normalised Contact Distance	Mean σ -hole Angle/ °	Mean Lone-Pair Angle/ °
Metal Carbonyls	2790	3.343	1.005	144	113
Terminal	2712	3.343	1.005	144	112
Di-bridging	70	3.355	1.010	145	119
Tri-bridging	8	3.241	0.976	140	122

Table 4.58: Summary of geometric characteristics of halogen bonding contacts to different terminal metal carbonyls

Functional Group	Contacts	Mean Contact Distance/ Å	Mean Normalised Contact Distance	Mean σ -hole Angle/ °	Mean Lone-Pair Angle/ °
Metal Carbonyls	2712	3.343	1.005	144	112
Ru-Carbonyl	488	3.346	1.007	143	113
Fe-Carbonyl	468	3.359	1.009	144	112
Re-Carbonyl	422	3.319	0.997	146	111
Os-Carbonyl	318	3.351	1.009	141	110
Mn-Carbonyl	248	3.365	1.007	146	111
W-Carbonyl	212	3.343	1.005	145	114
Mo-Carbonyl	177	3.310	0.995	146	114
Cr-Carbonyl	136	3.334	1.002	147	114
Co-Carbonyl	101	3.356	1.009	145	117
Rh-Carbonyl	61	3.346	1.006	139	121
Ir-Carbonyl	55	3.328	1.002	142	116
Pt-Carbonyl	14	3.369	1.015	142	117
Ni-Carbonyl	5	3.418	1.030	134	114
Tc-Carbonyl	5	3.370	1.015	137	121
Cu-Carbonyl	2	3.444	1.037	136	113

4.6.2 Ch ··· Metal Carbonyl Oxygen Contacts

Table 4.55 shows chalcogen bonding contacts to metal carbonyls are considerably less numerous than the analogous halogen bonding contacts.

The distribution of σ -hole angle against normalised contact distance of chalcogen bonding contacts to metal carbonyls is shown in Figure 4.72 and the mean averages of the geometric measurements of the contacts shown in Table 4.59 & Table 4.60. There are a very small number of selenium and tellurium contacts in comparison to the sulfur contacts to metal carbonyls. There are very few contacts with normalised contact distances less than 0.9, and very little relationship between shorter normalised contact distances and σ -hole angles closer to linearity. The majority of the contacts in Figure 4.72 correspond to terminal carbonyl ligands and a small number of di-bridging metal carbonyls which do not seem to show any relationship between σ -hole angle and normalised contact distance.

The distribution of lone-pair angle against normalised contact distance for chalcogen bonding contacts to metal carbonyl moieties is shown in Figure 4.73. There is some bias towards the chalcogen bonding contacts approaching at lone-pair angles nearer 90° rather than the expected 120° . Once again there is a plethora of metal carbonyl species involved in these contacts. The most numerous metal carbonyl species are those bound to rhenium and tungsten, and the most common coordination sphere of the metal centre is 6 bonds, as shown in Figure 4.74. This is likely a reflection of the use of octahedral tungsten and rhenium carbonyls as organometallic precursors^{213 214}.

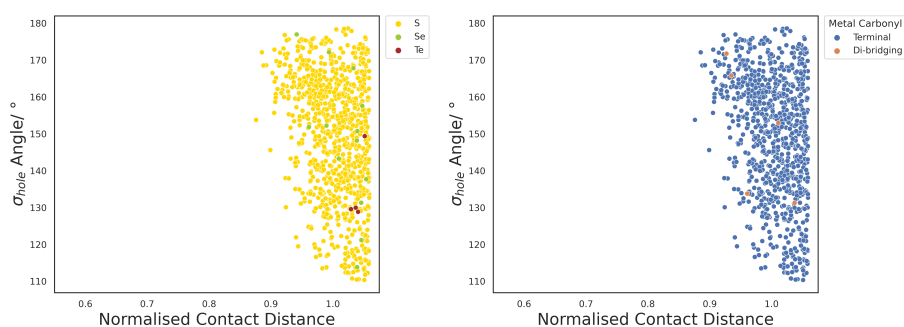


Figure 4.72: Scatter plots showing the distribution of normalised contact distance with respect to the σ -hole angle of chalcogen bonding contacts to terminal, di-bridging, and tri-bridging metal carbonyls found in crystal structures in the CSD. Points colour coded by donor atom type (left), and the bridging mode of the carbonyl (right)

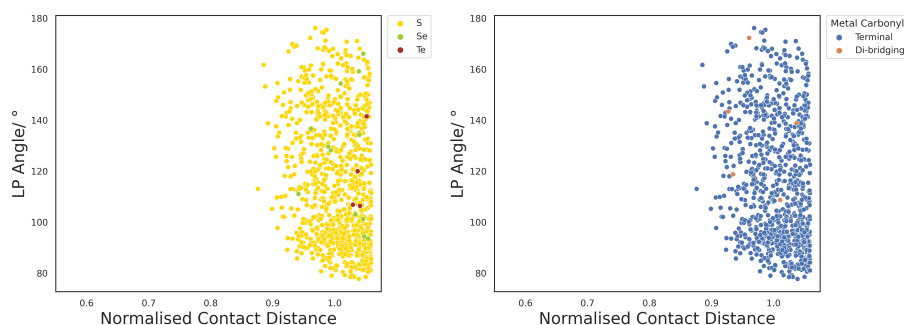


Figure 4.73: Scatter plots showing the distribution of normalised contact distance with respect to the lone-pair angle of chalcogen bonding contacts to terminal, di-bridging, and tri-bridging metal carbonyls found in crystal structures in the CSD. Points colour coded by donor atom type (left), and the bridging mode of the carbonyl (right)

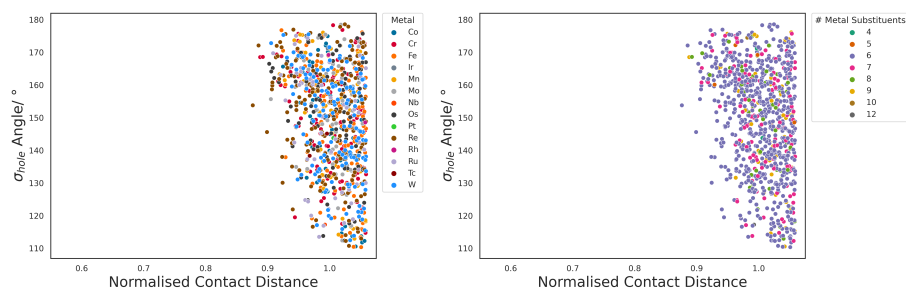


Figure 4.74: Scatter plots showing the distribution of normalised contact distance with respect to the lone-pair angle of chalcogen bonding contacts to terminal metal carbonyls found in crystal structures in the CSD. Points colour coded by the metal that the carbonyl is coordinated to (left) and the number of substituents coordinated to the metal (right).

Table 4.59: Summary of geometric characteristics of chalcogen bonding contacts to metal carbonyl oxygens

Donor	Contacts	Mean Contact Distance/ Å	Mean Normalised Contact Distance	Mean σ -hole Angle/ °	Mean Lone-Pair Angle/ °
Ch	888	3.401	1.003	148	116
Te	4	3.628	1.040	134	118
Se	13	3.381	1.018	148	126
S	871	3.400	1.003	149	116

Table 4.60: Summary of geometric characteristics of chalcogen bonding contacts to terminal, di-bridging, and tri-bridging metal carbonyls

Functional Group	Contacts	Mean Contact Distance/ Å	Mean Normalised Contact Distance	Mean σ -hole Angle/ °	Mean Lone-Pair Angle/ °
Metal Carbonyls	888	3.401	1.003	148	116
Terminal	883	3.402	1.004	148	116
Di-bridging	5	3.302	0.974	151	136

Table 4.61: Summary of geometric characteristics of chalcogen bonding contacts to different terminal metal carbonyls

Functional Group	Contacts	Mean Contact Distance/ Å	Mean Normalised Contact Distance	Mean σ -hole Angle/ °	Mean Lone-Pair Angle/ °
Metal Carbonyls	883	3.402	1.004	148	116
Re-Carbonyl	173	3.378	0.997	149	113
W-Carbonyl	140	3.413	1.007	147	114
Cr-Carbonyl	114	3.399	1.003	148	117
Fe-Carbonyl	114	3.432	1.012	147	118
Os-Carbonyl	106	3.397	1.002	151	113
Mn-Carbonyl	62	3.408	1.005	149	111
Co-Carbonyl	57	3.405	1.004	147	118
Ru-Carbonyl	56	3.386	0.999	147	126
Mo-Carbonyl	54	3.402	1.004	151	118
Rh-Carbonyl	2	3.484	1.028	158	100
Tc-Carbonyl	2	3.328	0.982	141	105
Ir-Carbonyl	1	3.380	0.997	171	156
Nb-Carbonyl	1	3.365	0.992	117	145
Pt-Carbonyl	1	3.410	1.006	143	142

4.6.3 *Pn* ... Metal Carbonyl Oxygen Contacts

Table 4.55 shows pnictogen bonding contacts to metal carbonyls the least numerous of the σ -hole interactions to carbonyl ligands.

The distribution of σ -hole angle against normalised contact distance of pnictogen bonding contacts to metal carbonyls is shown in Figure 4.75 and the mean averages of the geometric measurements of the contacts shown in Table 4.62 & Table 4.63. These pnictogen bonding contacts are predominantly phosphorus contacts to metal carbonyls. There are very few contacts with normalised contact distances less than 0.9, and very little relationship between shorter normalised contact distances and σ -hole angles closer to linearity. The majority of the contacts in Figure 4.75 correspond to terminal carbonyl ligands and a singular contact to a di-bridging metal carbonyl ligand shows little to no relationship between σ -hole angle and normalised contact distance. There are a small number of antimony and bismuth contacts that have normalised contact distances less than 0.9 with σ -hole angles approaching linearity.

The distribution of lone-pair angle against normalised contact distance for pnictogen bonding contacts to metal carbonyl moieties is shown in Figure 4.76. There are a small number of antimony and bismuth contacts that have normalised contact distances less than 0.9 and lone-pair angles around 120°. This suggests the greater likelihood of the lone-pair acting as the carbonyl ligands Lewis basic site over the π -system. The closest contacts with σ -hole angles approaching linearity involve carbonyl ligands bound to molybdenum and tungsten, with the molybdenum containing 7 bonds in its coordination sphere, while the tungsten carbonyls contain six or eight bonds in its coordination sphere.

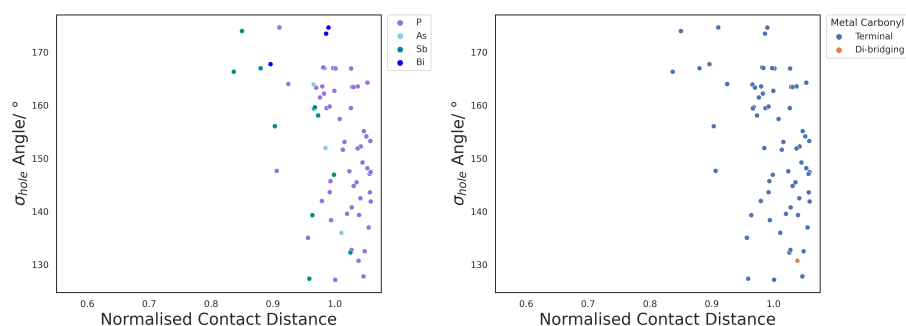


Figure 4.75: Scatter plots showing the distribution of normalised contact distance with respect to the σ -hole angle of pnictogen bonding contacts to terminal, di-bridging, and tri-bridging metal carbonyls found in crystal structures in the CSD. Points colour coded by donor atom type (left), and the bridging mode of the carbonyl (right)

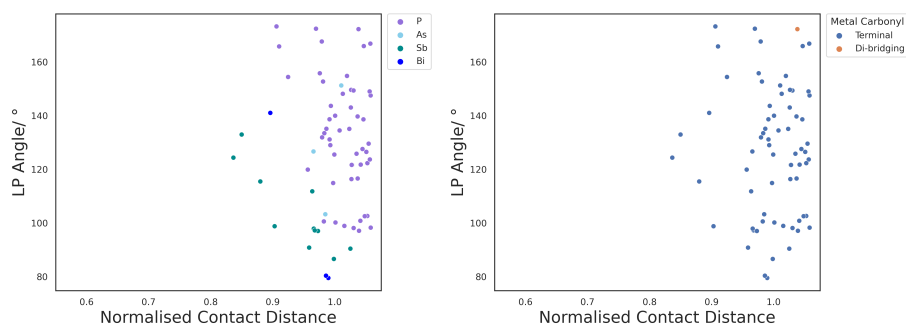


Figure 4.76: Scatter plots showing the distribution of normalised contact distance with respect to the lone-pair angle of pnictogen bonding contacts to terminal, di-bridging, and tri-bridging metal carbonyls found in crystal structures in the CSD. Points colour coded by donor atom type (left), and the bridging mode of the carbonyl (right)

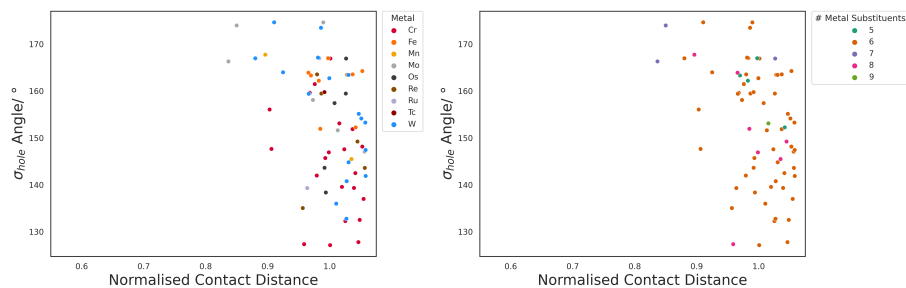


Figure 4.77: Scatter plots showing the distribution of normalised contact distance with respect to the lone-pair angle of pnictogen bonding contacts to terminal metal carbonyls found in crystal structures in the CSD. Points colour coded by the metal that the carbonyl is coordinated to (left) and the number of substituents coordinated to the metal (right).

Table 4.62: Summary of geometric characteristics of pnictogen bonding contacts to metal carbonyl oxygens

Donor	Contacts	Mean Contact Distance/ Å	Mean Normalised Contact Distance	Mean σ -hole Angle/ °	Mean Lone-Pair Angle/ °
Pn	70	3.505	0.999	152	126
Bi	3	3.868	0.957	171	100
Sb	11	3.726	0.939	153	103
As	3	3.337	0.987	150	127
P	53	3.448	1.014	151	132

Table 4.63: Summary of geometric characteristics of pnictogen bonding contacts to different terminal metal carbonyls

Functional Group	Contacts	Mean Contact Distance/ Å	Mean Normalised Contact Distance	Mean σ -hole Angle/ °	Mean Lone-Pair Angle/ °
Metal Carbonyls	69	3.505	0.998	152	125
Cr-Carbonyl	22	3.542	1.005	144	119
W-Carbonyl	17	3.507	1.003	155	131
Fe-Carbonyl	9	3.419	1.007	161	122
Mo-Carbonyl	7	3.563	0.958	162	123
Os-Carbonyl	5	3.432	1.009	153	140
Re-Carbonyl	5	3.417	1.005	150	127
Mn-Carbonyl	2	3.570	0.966	156	133
Ru-Carbonyl	1	3.827	0.964	139	111
Tc-Carbonyl	1	3.372	0.992	159	138

4.6.4 *Tt* ... Metal Carbonyl Oxygen Contacts

Table 4.55 shows tetrel bonding contacts to metal carbonyls are much less numerous than the analogous halogen or chalcogen bonding contacts, yet are more numerous than the analogous pnictogen bonding contacts.

The distribution of σ -hole angle against normalised contact distance of tetrel bonding contacts to metal carbonyls is shown in Figure 4.78 and the mean averages of the geometric measurements of the contacts shown in Table 4.64 & Table 4.65. Interestingly, the tetrel bonding contacts to metal carbonyl ligands show the most pronounced relationship between shorter normalised contact distances and σ -hole angles that approach linearity. The overall average σ -hole angle for tetrel bonding contacts to metal carbonyl ligands is 168°. The majority of the closest contacts involve tin contacts, but also some germanium and silicon contacts. The contacts in Figure 4.78 correspond to terminal carbonyl ligands and a singular contact to a di-bridging metal carbonyl ligand with a long contact with a σ -hole angle very near linearity.

The distribution of lone-pair angle against normalised contact distance for tetrel bonding contacts to metal carbonyl moieties is shown in Figure 4.79. There is a slight preference for the shortest tetrel bonding contacts to adopt lone-pair angles near 120°, get also with angles nearer linearity. This suggests the greater likelihood of the lone-pair acting as

the carbonyl ligands Lewis basic site over the π -system. The tetrel bonding contacts to different metal carbonyl species, shown in Figure 4.80 and Table 4.65, involve mainly carbonyl ligands bound to tungsten, with a notable number bound to chromium, iron, and molybdenum. The contacts to metal carbonyls with an average normalised contact distance shorter than the sum of van der Waals radii involve carbonyl ligands bound to iron and manganese.

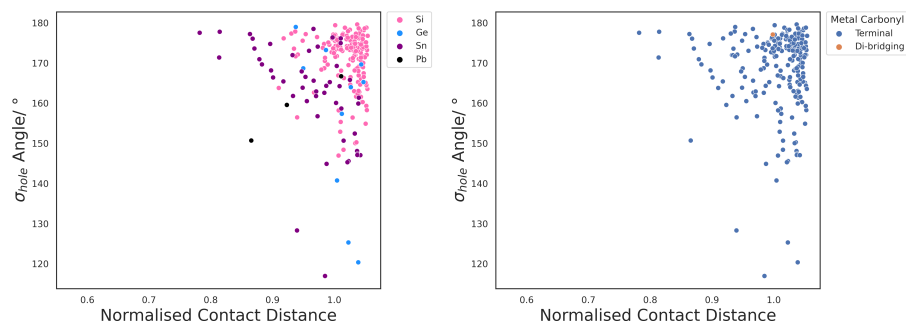


Figure 4.78: Scatter plots showing the distribution of normalised contact distance with respect to the σ -hole angle of tetrel bonding contacts to terminal, di-bridging, and tri-bridging metal carbonyls found in crystal structures in the CSD. Points colour coded by donor atom type (left), and the bridging mode of the carbonyl (right)

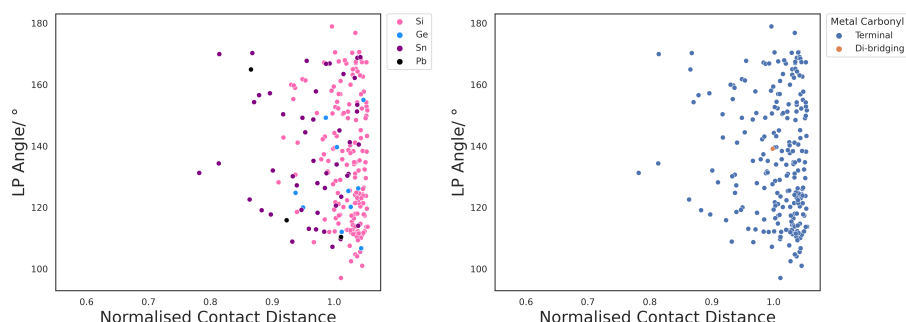


Figure 4.79: Scatter plots showing the distribution of normalised contact distance with respect to the lone-pair angle of tetrel bonding contacts to terminal, di-bridging, and tri-bridging metal carbonyls found in crystal structures in the CSD. Points colour coded by donor atom type (left), and the bridging mode of the carbonyl (right)

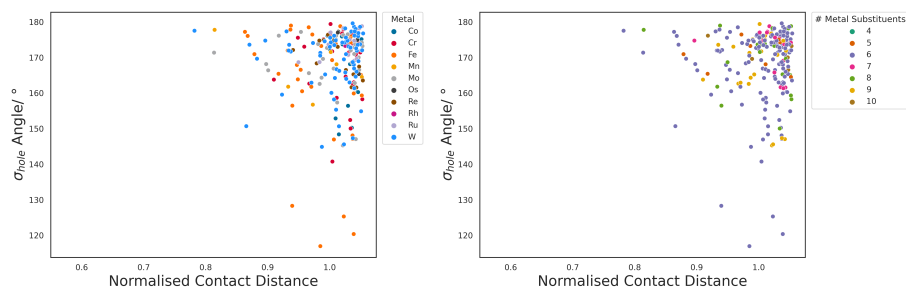


Figure 4.80: Scatter plots showing the distribution of normalised contact distance with respect to the lone-pair angle of tetrel bonding contacts to terminal metal carbonyls found in crystal structures in the CSD. Points colour coded by the metal that the carbonyl is coordinated to (left) and the number of substituents coordinated to the metal (right).

Table 4.64: Summary of geometric characteristics of tetrel bonding contacts to metal carbonyl oxygens

Donor	Contacts	Mean Contact Distance/ Å	Mean Normalised Contact Distance	Mean σ -hole Angle/ °	Mean Lone-Pair Angle/ °
Tt	231	3.769	1.006	168	135
Pb	3	3.827	0.933	159	130
Sn	49	3.769	0.961	162	137
Ge	10	3.817	1.007	156	127
Si	169	3.766	1.020	171	134

Table 4.65: Summary of geometric characteristics of tetrel bonding contacts to different terminal metal carbonyls

Functional Group	Contacts	Mean Contact Distance/ Å	Mean Normalised Contact Distance	Mean σ -hole Angle/ °	Mean Lone-Pair Angle/ °
Metal Carbonyls	230	3.770	1.006	168	135
W-Carbonyl	68	3.779	1.005	169	134
Cr-Carbonyl	35	3.783	1.019	169	131
Fe-Carbonyl	35	3.753	0.982	163	139
Mo-Carbonyl	31	3.765	1.002	169	132
Co-Carbonyl	19	3.786	1.026	167	135
Ru-Carbonyl	15	3.783	1.021	173	138
Re-Carbonyl	14	3.748	1.016	169	134
Mn-Carbonyl	7	3.669	0.970	168	144
Os-Carbonyl	5	3.783	1.025	173	144
Rh-Carbonyl	1	3.783	1.025	173	107

4.6.5 Prevalence of $\sigma_{hole} \cdots$ Metal Carbonyl Oxygen Contacts

Due to the extremely low number of σ -hole interactions to di- and tri-bridging metal carbonyls, the prevalence of σ -hole interactions will only be considered relative to structures containing terminal metal carbonyl ligands. The prevalence of σ -hole interactions to metal carbonyl σ -hole acceptors is detailed in Table 4.66. The prevalence of halogen and chalcogen bonding to metal carbonyl σ -hole acceptors is very high, even greater than the analogous contacts to terminal oxide σ -hole acceptors (see Table 4.46). The prevalence of pnictogen and tetrel bonding to metal carbonyl σ -hole acceptors is somewhat lower, with both around 8.8%.

Table 4.66: Overall summary of the prevalence of contacts indicating σ -hole interactions to metal carbonyl moieties compared to the number of unique crystal structures found in the CSD that contain the necessary moieties.

σ -hole Interaction Type	Structures with Moieties	Structures with Contact(s)	Prevalence
Halogen-bonding	4085	1677	41.05%
Chalcogen-bonding	1398	537	38.41%
Pnictogen-bonding	552	49	8.88%
Tetrel-bonding	2056	182	8.85%

Halogen bonding contacts show a trend in prevalence in Table 4.67 that is the inverse of the polarisability of the halogen element, being $I < Br < Cl$. This could be related to the common use of chlorinated solvents in organometallic chemistry, which are used to co-crystallise the metal carbonyl compound. The prevalence of halogen bonding to different metal carbonyl species is shown in Table 4.68. The most notable prevalence is seen with halogen bonds to carbonyl ligands bound to osmium, with a prevalence of 54.30%. For only a few metals does the prevalence of halogen-bond formation deviate substantially below the average prevalence of 41.05%

Table 4.67: Summary of the prevalence of halogen bonding contacts to terminal metal carbonyl oxygens compared to the number of unique crystal structures found in the CSD that contain the necessary moieties

Donor	Structures with Moieties	Structures with Contact(s)	Prevalence
X	4085	1677	41.05%
I	123	34	27.64%
Br	316	121	38.29%
Cl	3682	1526	41.44%

Table 4.68: Summary of the prevalence of halogen bonding contacts to different terminal metal carbonyls compared to the number of unique crystal structures found in the CSD that contain the necessary moieties

Functional Group	Structures with Moieties	Structures with Contact(s)	Prevalence
Metal Carbonyl	4085	1677	41.05%
Ru-Carbonyl	799	303	37.92%
Fe-Carbonyl	748	300	40.11%
Re-Carbonyl	519	258	49.71%
Os-Carbonyl	291	158	54.30%
Mn-Carbonyl	291	135	46.39%
W-Carbonyl	421	145	34.44%
Mo-Carbonyl	348	130	37.36%
Cr-Carbonyl	190	92	48.42%
Co-Carbonyl	174	66	37.93%
Rh-Carbonyl	235	47	20.00%
Ir-Carbonyl	222	44	19.82%
Pt-Carbonyl	34	13	38.24%
Ni-Carbonyl	12	2	16.67%
Tc-Carbonyl	9	5	55.56%
Cu-Carbonyl	12	2	16.67%
Other	14	0	N/A

The prevalence of chalcogen bonding to metal carbonyl Lewis bases, shown in Table 4.69, is dominated by sulfur. While making up the vast majority of the crystals containing the required moieties, and those showing chalcogen bonding

interactions, those involving sulfur show a very high prevalence. Chalcogen bonds involving carbonyl ligands bound to osmium show a prevalence greater than 50% that is notably above the overall average prevalence of chalcogen bonding to metal carbonyls. Chalcogen bonds involving carbonyl ligands bound to chromium also show a significant prevalence of 47.58%, seen in Table 4.70.

Table 4.69: Summary of the prevalence of chalcogen bonding contacts to terminal metal carbonyl oxygens compared to the number of unique crystal structures found in the CSD that contain the necessary moieties

Donor	Structures with Moieties	Structures with Contact(s)	Prevalence
Ch	1398	537	38.41%
Te	19	3	15.79%
Se	67	11	16.42%
S	1313	523	39.83%

Table 4.70: Summary of the prevalence of chalcogen bonding contacts to different terminal metal carbonyls compared to the number of unique crystal structures found in the CSD that contain the necessary moieties

Functional Group	Structures with Moieties	Structures with Contact(s)	Prevalence
Metal Carbonyl	1398	537	38.41%
Re-Carbonyl	231	98	42.42%
W-Carbonyl	208	83	39.90%
Fe-Carbonyl	280	81	28.93%
Cr-Carbonyl	124	59	47.58%
Os-Carbonyl	106	58	54.72%
Mn-Carbonyl	140	46	32.86%
Co-Carbonyl	78	34	43.59%
Ru-Carbonyl	114	33	28.95%
Mo-Carbonyl	101	38	37.62%
Rh-Carbonyl	19	2	10.53%
Tc-Carbonyl	2	2	100.00%
Ir-Carbonyl	14	1	7.14%
Nb-Carbonyl	1	1	100.00%
Pt-Carbonyl	1	1	100.00%
Other	6	0	N/A

Among pnictogen and tetrel bonds to metal carbonyls, already noted as having low overall prevalence (see Table 4.66). Of these σ -hole interactions, those involving antimony and tin have notably higher prevalence, albeit with only 10s of observations. The scarcity of data makes trends with respect to the polarisability of the σ -hole donor atom difficult to analyse. In terms of the metal carbonyl species, see Table 4.72 and Table 4.74, for both pnictogen and tetrel bonding contacts tungsten and rhenium carbonyls are the most prevalent, along with pnictogen contacts to osmium carbonyls

having notable prevalence.

Table 4.71: Summary of the prevalence of pnictogen bonding contacts to terminal metal carbonyl oxygens compared to the number of unique crystal structures found in the CSD that contain the necessary moieties

Donor	Structures with Moieties	Structures with Contact(s)	Prevalence
Pn	552	49	8.88%
Bi	3	3	100.00%
Sb	18	8	44.44%
As	31	2	6.45%
P	501	36	7.19%

Table 4.72: Summary of the prevalence of pnictogen bonding contacts to different terminal metal carbonyls compared to the number of unique crystal structures found in the CSD that contain the necessary moieties

Functional Group	Structures with Moieties	Structures with Contact(s)	Prevalence
Metal Carbonyl	552	49	8.88%
Cr-Carbonyl	87	13	14.94%
W-Carbonyl	142	12	8.45%
Fe-Carbonyl	106	5	4.72%
Mo-Carbonyl	65	6	9.23%
Re-Carbonyl	23	5	21.74%
Os-Carbonyl	30	4	13.33%
Mn-Carbonyl	30	2	6.67%
Ru-Carbonyl	13	1	7.69%
Tc-Carbonyl	3	1	33.33%
Other	61	0	N/A

Table 4.73: Summary of the prevalence of tetrel bonding contacts to terminal metal carbonyl oxygens compared to the number of unique crystal structures found in the CSD that contain the necessary moieties

Donor	Structures with Moieties	Structures with Contact(s)	Prevalence
Tt	2056	182	8.85%
Pb	2	1	50.00%
Sn	82	29	35.37%
Ge	45	8	17.78%
Si	1942	144	7.42%

Table 4.74: Summary of the prevalence of tetrel bonding contacts to different terminal metal carbonyls compared to the number of unique crystal structures found in the CSD that contain the necessary moieties

Functional Group	Structures with Moieties	Structures with Contact(s)	Prevalence
Metal Carbonyl	2056	182	8.85%
W-Carbonyl	422	55	13.03%
Cr-Carbonyl	311	27	8.68%
Fe-Carbonyl	500	26	5.20%
Mo-Carbonyl	202	20	9.90%
Co-Carbonyl	206	16	7.77%
Ru-Carbonyl	183	14	7.65%
Re-Carbonyl	56	12	21.43%
Mn-Carbonyl	64	6	9.38%
Os-Carbonyl	70	5	7.14%
Rh-Carbonyl	28	1	3.57%
Other	84	0	N/A

4.6.6 Notable Metal Carbonyl Functional Groups

As the σ -hole interactions to metal carbonyl ligands involve generally longer contacts with σ -hole angles that deviate substantially from linearity, there are few structures found within this CSD search that stand out. The recurrence of metal carbonyls such as those of rhenium, tungsten, and iron are due to their use in a wide variety of reactions and are readily commercially available, for example bromopentacarbonylrhenium(I), which is a useful precursor to rhenium complexes.

The crystal structure of $[\text{Re}(\eta^5\text{-C}_5\text{H}_4\text{I})(\text{CO})_3]$ (CSD refcode **DUKREH**) is reported as having "strong I...O interactions" (see Figure 4.81). This structure involves rhenium with three carbonyl groups and then a η^5 iodocyclopentadienyl ligand and is from the work of Romanov *et al.* with the aim of producing starting materials for new compound discovery²¹⁵.

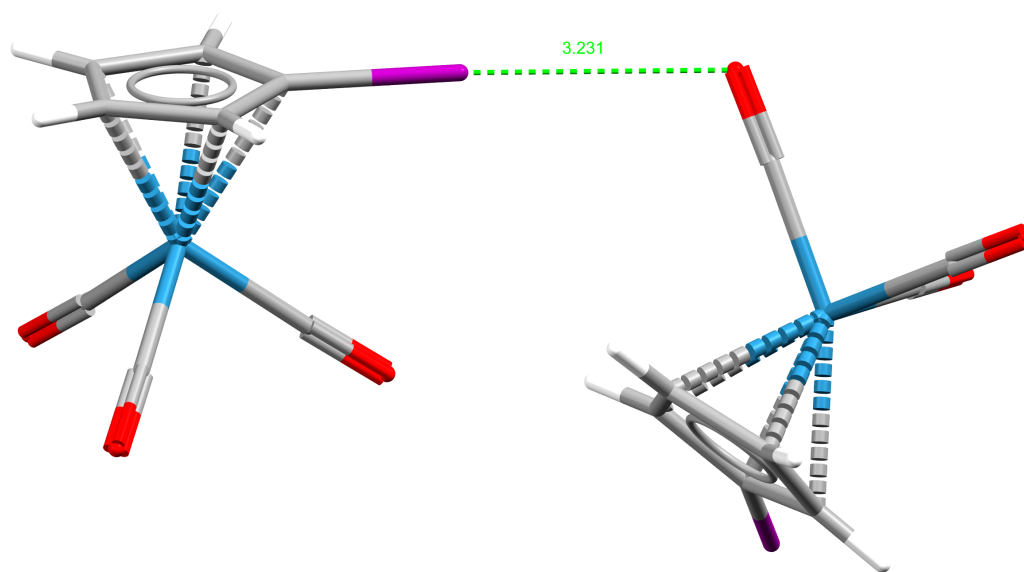


Figure 4.81: An example of a crystal structure with a halogen bonding contact $[\text{Re}(\eta^5\text{-C}_5\text{H}_4\text{I})(\text{CO})_3]$ with a halogen bond between the iodocyclopentadienyl ligand and the carbonyl bound to the rhenium. Distance given in Ångström. CSD refcode: **DUKREH**

A notable chalcogen bonding structure involving a rhenium carbonyl moiety is shown in Figure 4.82 of CSD refcode **EXUWOK**. This chalcogen contact has a normalised contact distance of 0.885, and is possibly enhanced by a very weak C–H hydrogen bond, between the carbonyl oxygen and the hydrogen on the benzyl moiety (2.619 Å - 0.970 normalised contact distance). This work by Gerber and co-workers involving benzothiazoles and their complexes with rhenium carbonyl cores, which is of interest for the synthesis of new Re and Tc radiotherapy and radioimaging agents²¹⁶.

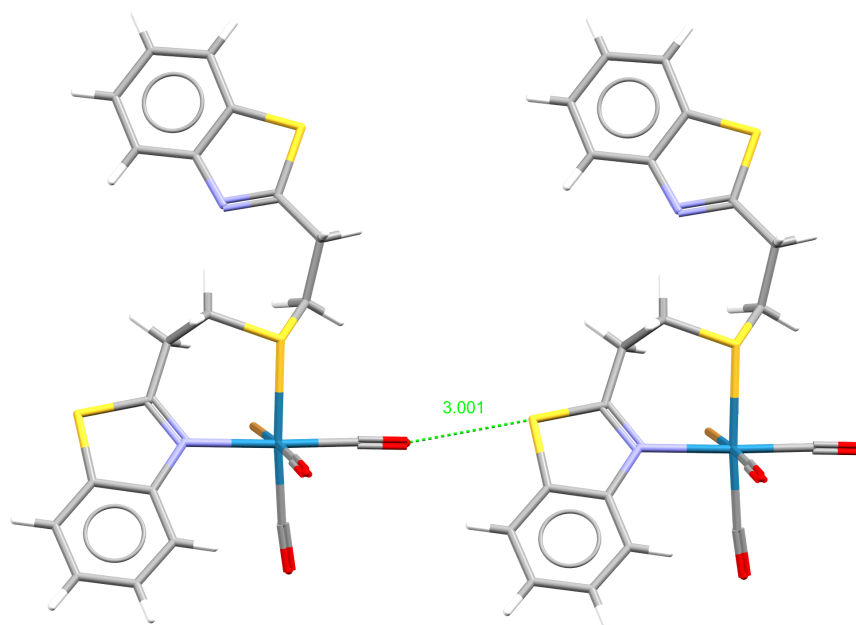


Figure 4.82: An example of a crystal structure of (bis(2-(1,3-Benzothiazol-2-yl)ethyl) sulfide)-bromo-tricarbonyl-rhenium with a chalcogen bonding contact between sulfur and the metal carbonyl oxygen. Distance given in Ångström. CSD refcode: **EXUWOK**

While pnictogen contacts are very scarce, an interesting structure is seen in CSD refcode **LAJGAI01** which involves a rhenium tris(2-pyridyl)phosphine scorpionate complex which shows a relatively close contact involving phosphorus which has a σ -hole angle of around 160° , which is characteristic of pnictogen bonding. However, these contacts are still rather long, with a normalised contact distance of 0.980, and are likely enhanced by weak C–H hydrogen bonding between the carbonyl and the pyridyl hydrogen ($2.767 \text{ \AA}$ - 1.024 normalised contact distance). These scorpionate complexes, especially Re(I) species such as this have interest in catalytic, polymer, and pharmaceutical fields, with this complex exhibiting thermochromic luminescence, and the structure is shown in Figure 4.83²¹⁷.

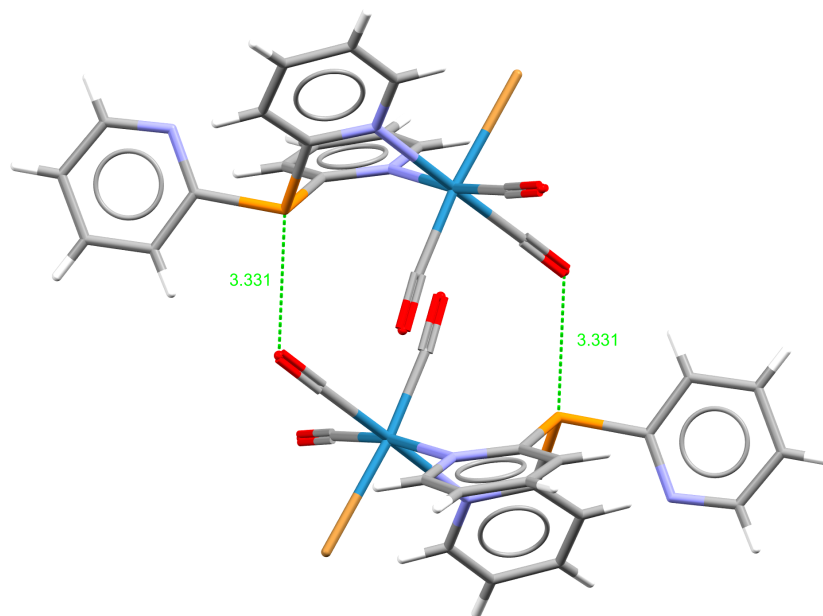


Figure 4.83: An example of a crystal structure with rhenium tris(2-pyridyl)phosphine scorpionate complex pnictogen bonding contacts between the phosphorus and the metal carbonyls. Distance given in Ångström. CSD refcode: **LAJGA101**

An interesting tetrel bonding complex involving a supramolecular complex containing both tin tetrel bond donor moieties and an iron tetracarbonyl moiety is shown in Figure 4.84 of CSD refcode **RELXAH**. This structure is from the work of Sita and coworkers on a main group cluster with potential to aid the synthesis of other transition metal and main group oxo complexes, such as the polyoxometallates(POM)s mentioned in Section 4.5.6²¹⁸.

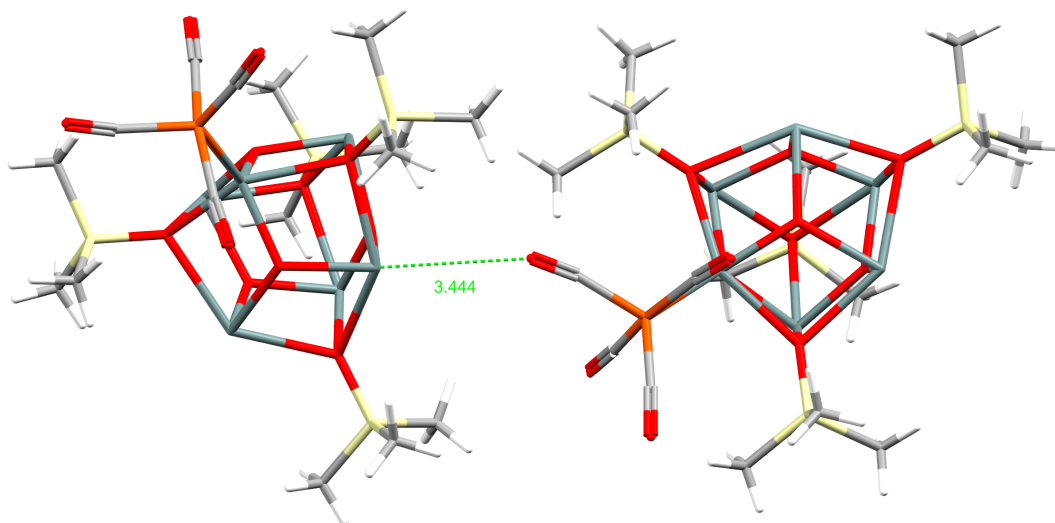


Figure 4.84: An example of a crystal structure of a tin polyoxometallate derivative bound to an iron carbonyl complex with a tetrel bonding contact between the tin and the carbonyl ligand. CSD refcode: **RELXAH**

4.7 Summary

Oxygen acting as an acceptor for σ -hole interactions shows a great deal more variety and complexity than nitrogen-based σ -hole acceptors. There is a great deal of impact that changes to both the immediate oxygen environment, and to the wider molecular system, have on the Lewis basicity of the oxygen lone-pairs. Overall, the contacts indicating σ -hole interactions within these crystal structures longer and have σ -hole angles that deviate further from linearity than the analogous contacts involving nitrogen Lewis bases. This is contrasted by a much greater number of contacts within the CSD between the σ -hole donor and these acceptor oxygen moieties. The most frequently occurring contacts follow the trend of organic carbonyls $\approx$ divalent oxygens $>$ oxides $>$ metal carbonyls and the same trend is followed in terms of increasing σ -hole angles.

Organic carbonyl oxygens show a great deal of variety with regards to the different carbonyl functional groups that can be found within the CSD. These different substituents, however, only slightly affects the geometry of the contacts, with the averages of the geometric measurements being fairly consistent. The main noticeable effect seen is that when nitrogen is incorporated into the carbonyl system, such as in amides, formamides, and ureas, there is a greater proportion of closer, contacts seen with σ -hole angles closer to linearity. Halogen bonding shows the greatest number of contacts to organic carbonyls, but chalcogen bonding contacts show slightly more variety in carbonyl functional groups. Halogen, chalcogen, and pnictogen bonding contacts tend to approach at the expected angle with respect to the lone-pair of 120° , but tetrel bonds favouring an angle closer to 140° . There is little to no preference for these contacts approaching coplanar or orthogonal to the plane of the carbonyl moiety. When the contacts distances are normalised with regards to the sum of their van der Waals radii, pnictogen and tetrel bonding contacts, while less abundant, show a larger proportion of contacts that are shorter than 0.8 normalised contact distance than halogen and chalcogen bonding contacts. These pnictogen and tetrel contacts are much less prevalent than the halogen and chalcogen bonding species, with 20% of the structures containing the required halogen and chalcogen bonding donor and acceptor moieties displaying one or more σ -hole interactions.

Divalent oxygen σ -hole acceptors show an even greater degree of variety of functional groups than seen with organic carbonyls, and a comparable number of contacts across the different σ -hole donors. This added variety also shows a more marked effect that the changing of the substituents around the oxygen has on its Lewis basicity. Alcohols and ethers are among the most numerous contacts that show a notable proportion of close contacts that are a strong tendency towards a linear σ -hole angle. Alcohols, and more so water, show a notable number of very close contacts to chlorine that have σ -hole angles closer to 90° than 180° . This is due to the ability of chlorine to act as a hydrogen bond acceptor, and thus these much stronger interactions are perturbing the weaker $\text{Cl} \cdots \text{O}$ halogen bond. This behaviour is seen less commonly with sulfur chalcogen bonds, which is a much weaker hydrogen bond acceptor, so there are fewer close contacts with σ -hole angle that deviate substantially from linearity, and none are seen with pnictogen and tetrel bonds to alcohols or water which do not act as hydrogen bond acceptors. Furans where the two carbons (C-O-C) are both sp^2 -hybridised participate as acceptors in σ -hole interactions that are notably longer and have σ -hole angles that deviate substantially

from linearity. This suggests that the delocalised system decreases the Lewis basicity of the oxygen. Halogen bonding contacts to divalent oxygen show a greater proportion of normalised contact distances less than 0.8 than the analogous chalcogen bonding contacts. The analogous pnictogen and tetrel bonding contacts, while few in number, have a greater proportion of normalised contact distances below this threshold. There is a very short bimodal distribution of antimony pnictogen bonding contacts which correspond to multiple pnictogen bonds to the oxygens within a crown ether. Tin shows some of the closest contacts, being σ -hole angles that are near linear, to alcohol moieties, with lead contacts being closer but having σ -hole angles much further from linearity while maintaining lone-pair directionality. In contrast to organic carbonyls, there is some bias towards the approach of the σ -hole donor atom orthogonally to the plane of the acceptor molecule, while still adopting lone-pair angles of around 120° . This is likely due to divalent oxygens being more sterically crowded than carbonyls, due to carbonyl oxygens having an extra atom separation from the bulk of the molecule. The prevalence of σ -hole interactions to divalent oxygen Lewis bases are slightly lower than for organic carbonyls, with halogen and chalcogen bonding contacts being found in 15% of structures containing the required halogen and chalcogen bond donor and acceptor moieties, and much less for pnictogen and tetrel bonding contacts.

Oxide σ -hole acceptor species are likely to be charged species, which is likely to have a considerable effect on the Lewis basicity of the lone pairs on the oxygen. This is shown in the notably closer contacts seen compared to the organic carbonyls and divalent oxygen species, with the distributions of σ -hole angles against normalised contact distances being more comparable to those seen in the nitrogen σ -hole acceptors. For halogen bonding complexes, N-oxides such as pyridine N-oxide are widely used with strong halogen bond donors such as N-iodosuccinamide, with normalised contact distance as low as 0.65 and σ -hole angles very close to linearity. While halogen bonding contacts to sulfur oxides are more common than those to phosphorus oxides, the phosphorus contacts show σ -hole angles closer to linearity among the closest contacts. Halogen bonds to metal and halogen oxides are longer and have a much broader range of σ -hole angles. Chalcogen bonds to oxides are generally longer compared to the analogous halogen bonds and have σ -hole angles that deviate further from linearity. Chalcogen bonds to sulfur oxides show similar geometric characteristics as those as with nitrogen oxides. Halogen and chalcogen bonds are found with some slightly closer and more linear contacts to perrhenate and pertechnetate, along with other anions such as triflate, which is due to the application of these interactions in the field of anion recognitions. Pnictogen bonding contacts to oxides have short normalised contact distances, but show reduced directionality with respect to the σ -hole angle, with a broad distribution of angles between 135 - 180° . Tetrel bonding contacts to oxide σ -hole acceptors show some of the closest and most linear interactions involving tin, while lead tetrel bonds adopt σ -hole angles further from 180° . Lone-pair angles among halogen and chalcogen bonds to oxides show strong bias towards 120° for contacts below 0.75 normalised contact distance, whereas pnictogen and tetrel bonding contacts only show this bias for contacts nearing 0.6 normalised contact distance, with a notable number of contacts with a bias towards the linear approach of the donor to the oxygen. The prevalence of these σ -hole interactions to oxide Lewis bases are some of the highest seen involving oxygen Lewis bases, with between 27-32% prevalence shown for halogen, chalcogen, and pnictogen bonds, while tetrel bonding has a lower prevalence of around 10%.

Lastly, metal carbonyl oxygens are drastically reduced in their Lewis basicity, with the lone-pair angle distributions indicating that σ -hole donors interact with the π -system of the carbonyl over the lone-pairs on the oxygen. For halogen, chalcogen, and pnictogen bonds to metal carbonyl σ -hole acceptors there are a small proportion of contacts with a normalised contact distance shorter than 0.9. There is little relationship between the closer normalised contact distances and σ -hole angles nearer linearity for halogen, chalcogen, and pnictogen bonding contacts to metal carbonyls. The vast majority of the structures involve terminal metal carbonyls, with only halogen bonding contacts showing a notable number of di-bridging metal carbonyls, which do show a slight bias towards larger σ -hole angles. This suggests that depleting the π -system of the carbonyl increases the Lewis basicity of the oxygen lone-pairs. Tetrel bonding to metal carbonyl σ -hole acceptors have the largest average σ -hole angle compared to the other σ -hole interactions with metal carbonyls. This includes a notable proportion of contacts with normalised contact distance less than 0.9. A commonly occurring metal carbonyl moiety that acts as an acceptor for σ -hole interactions are carbonyl ligands bound to rhenium. This is likely due to the commercially available bromopentacarbonylrhenium(I) reagent that is used as a precursor to a wide range rhenium carbonyl complexes. Overall, oxygen Lewis bases offer a highly synthetically accessible acceptor motif for σ -hole interactions, with σ -hole donors able to approach from a wide range of directions with respect to the oxygen atom multiple lone-pairs present. In comparison to nitrogen Lewis bases as σ -hole acceptors, there is a much greater variety in functional groups and how this effects the Lewis basicity of the moiety. Organic carbonyls offer consistent behaviour across the functional group range, with nitrogen containing groups such as amides showing lightly increased Lewis basicity, and a reduced steric bulk than divalent oxygens, which varies more in Lewis basicity with different substituents, but this also entails a increased variety of functional groups available. Oxides are the most effective σ -hole acceptor out of the oxygen Lewis bases, due to an abundance of anionic species, forming contacts similar in strength to those seen with nitrogen Lewis bases. Halogen and chalcogen bonding interactions to various oxides have been exploited for use in anion recognition, which explains its abundance of short contacts. Metal carbonyls prove to be ineffective σ -hole acceptors, except for a few notable exceptions involving tetrel bonds, where there is interest in new materials involving tin and metal carbonyl species, but none recognised for their tetrel bonding ability.

It is important to note that the large number of oxide Lewis bases that are anionic make them unsuitable as a direct like-for-like comparison with the neutral nitrogen Lewis bases. Thus on comparison of the σ -hole interactions to organic carbonyl or divalent oxygen moieties with the analogous σ -hole interactions with the various hybridisations of nitrogen discussed in Chapter 3, more reliable contrast can be made. The mean averages of normalised contact distance and σ -hole angle of these contacts can be found for organic carbonyl and divalent oxygen moieties in Table 4.1 & Table 4.19, and for the three hybridisations of nitrogen in Table 3.1, Table 3.11, & Section 3.5.

Generally the normalised contact distance of σ -hole interactions to oxygen Lewis bases are larger than the analogous interactions to nitrogen Lewis bases. For halogen and chalcogen bonding the average normalised contact distance for contacts to each hybridisation of nitrogen are lower than that of the analogous contacts to oxygen Lewis bases and show larger average σ -hole angles. The pnictogen and tetrel bonding contacts to divalent oxygen and organic carbonyl Lewis

bases have smaller average normalised contact distances than the analogous σ -hole interactions to sp^3 -hybridised nitrogen, but are longer than the average normalised contact distances of pnictogen and tetrel bonds to sp^2 - & sp -hybridised nitrogen. The average σ -hole angles of pnictogen and tetrel bonds are consistent between interactions with both oxygen and nitrogen σ -hole acceptors. From this one can conclude that oxygen Lewis bases present a wide range of potentially useful σ -hole acceptor moieties with huge scope for tuning the strength and geometry of the σ -hole interactions. This character is offset by the reduced strength of σ -hole interactions involving oxygen Lewis base σ -hole acceptors in comparison to those involving nitrogen Lewis bases.

Chapter 5

Machine Learning Prediction of σ -hole Interaction Energies

5.1 Overview

5.1.1 Introduction to Machine Learning in Chemistry

Machine learning (ML) is a form of artificial intelligence (AI), with the difference being that AI is more focused on the ability of a computers to exhibit human-like behaviour and intelligence, whereas ML involves the application of statistical algorithms to automatically improve at a task when provided with experience. Machine learning can be most simply described by a oft-repeated quote from the computer scientist Tom M. Mitchell: "A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P if its performance at tasks in T , as measured by P , improves with experience E ".²¹⁹ Using this analogy, to allow this ML model to learn, it must be trained for its specific task.

The process of training involves providing the ML model with a dataset to extract insight from to improve its performance at the task. There are various different approaches to training a model, which can be subdivided into supervised, unsupervised, and reinforcement learning. Supervised learning approaches use a training dataset of example input features and the outputs associated with these features. This allows the model to map the relationship between the input features and the outputs in, for example, regression or classification algorithms. Unsupervised learning involves no input features, with no labels associated, with the aim of the model being to find trends or commonalities within the dataset and execute a process based on these groupings. Examples of unsupervised learning are algorithms for dimensionality reduction such as principal component analysis²²⁰. Unsupervised learning is also a key tool in molecular simulation, showcased in a detailed review by Glielmo *et al.*²²¹ Reinforcement learning is the most different in its training approach, with these models acting as a so-called "agent" that affects an "environment" to achieve a "goal", the decisions made by this agent are evaluated

and rewarded or "punished" depending on its progress towards the goal. Reinforcement learning models are used in more complex simulations or, for example, self-driving cars^{222–224}.

ML has had a transformative impact on the chemical sciences, with use cases such as pharmaceutical drug discovery, reaction optimisation, and enhancing computational chemistry calculations^{225–227}. The field of ML approaches to chemistry has seen a great deal of hype, yet also notable critique with regards to application²²⁸. When presented with a task that, if attempted using traditional approaches, would prove to be unfeasible in terms of time or expense, a machine learning approach presents an attractive solution.

This work aims to produce a ML model that, when provided with a contact indicative of a σ -hole interaction from the CSD search results presented in Chapter 3, Chapter 4 and other searches, can extract features from this contact to rapidly approximate the interaction energy between the σ -hole donor and acceptor moieties. This will require a model dataset of a diverse range of σ -hole interactions, with geometric measurements and high-accuracy interaction energies. A suitable ML model will need to be selected along with the most relevant descriptors for the model to employ in its prediction of interaction energies.

5.1.2 General Considerations for ML Applications

In order to produce a model that provides accurate predictions to the task at hand, great care must be taken from a very early stage, long before the model is constructed and trained. A ML model is only as good as the data it is trained on, thus the first step is providing it with a dataset that is highly accurate and fully represents the task that the model is tasked to solve. When constructing a ML model, one could be tempted to provide the model with as many variables in the training set as possible, yet this is rarely beneficial in practice. Care should be taken so as not to introduce redundant variables as this will reduce the model's ability to generalise, along with increasing complexity and thus computational cost. When one trains a ML model, the dataset is split into a training set, and a testing set, the first of which is used to fit the model, and the second is not "seen" by the model and so is used to test its performance. The size of split and stratification of these sets are important to ensure that both sets are representative of each other and the dataset as a whole, to prevent the model being overfit to a small part of the dataset.

Another consideration to make is the training or fitting of the model, in other words the best possible match the model makes to the data provided. Two common problems that can occur with this are underfitting and overfitting, which affects the ability of the model to generalise when provided with new, unseen data. Underfitting occurs when the model lacks the complexity to learn the important trends in relationships within the training data, meaning it performs poorly on both seen and unseen data. Causes of underfitting can be from unsuitable model selection (e.g. linear model for nonlinear data), insufficient training, or lack of feature completeness. Overfitting occurs when the model has learnt too much from the data in the training set, incorporating trends found in noise or less relevant features. This results in very good performance for the training set, yet poor performance on unseen data outside of its training set. This can be caused by poor choice of model (e.g. a deep neural network for a linear relationship), over-training, feature complexity, or insufficient diversity

in training data. Getting the "best" fit of a model, the goal of machine learning, is to find a balance so that the model is complex enough to capture the desired trends in the data, while not being so sensitive as to capture irrelevant noise in the data.

5.1.3 ML Approaches to Predicting Noncovalent Interactions

There have been a variety of reported approaches to rapidly estimate the interaction energy of a noncovalently interacting molecules using more easily attainable data. There are the simple statistical models such as those of Legon and co-workers mentioned in Section 1.5, but also some recent ML approaches for predicting interaction energies of noncovalent interactions such as hydrogen bonding and halogen bonding^{89 90}.

An approach from the Sherrill group is the CLIFF kernel model, which is a **C**omponent-based machine-**L**earned **I**ntermolecular **F**orce **F**ield²²⁹. Aimed for use in pharmaceutically relevant systems, it is trained on a very large generated database of small drug like molecules and their dimers. The model is fit to match high-level symmetry-adapted perturbation theory (SAPT) interaction energies and energetic contributions. So not only is the interaction energy determined, parameters for electrostatics, exchange, dispersion, and induction are obtained, which gives insight into the nature of the interaction. This model performs very well on a wide variety of validation sets, albeit with some drawbacks, such as inaccurate treatment of exchange repulsion.

Machine learning predictions of hydrogen bonds have been made, such as free energies (ΔG) of hydrogen bond complexes predicted using a Support Vector Machine (SVM) model by Glavatskikh *et al.* trained on a large dataset of hydrogen bond complexes where the ΔG values were experimentally determined at 298 K in CCl_4 ²³⁰. The descriptors used in this model were ISIDA (In **S**ilico design and **D**ata **A**nalysis) fragment descriptors, which represents subgraphs of the molecule and contain topological, atomic, and bonding information²³¹. Another support vector model was produced by Bose *et al.* to predict interaction energies of rigid water clusters. This model uses geometric observations as descriptors, such as the reciprocal of the interatomic distances and intermolecular distances, and is trained on two and three body classical simulations of a water decamer²³².

The work of Devore and Shuford showcases several different approaches as a proof of concept for ML property determination of halogen bonding²³³. This work involved 1210 symmetrical substituted halobenzene/haloethynylbenzene halogen bond donors and their complexes with ammonia. Three ML models were compared to first classify the halogen bond donors by their principle halogen atom and then two different regression approaches were used to predict the potential of the σ -hole ($V_{S,max}$), the binding energy (E_{bind}), and the X $\cdots$ N bond local force constant ($k_{X\cdots N}^a$). The two regression models involved a "data-based" approach and a molecular fingerprint approach. The "data-based" approach used DFT derived features selected for the aforementioned classification model, while the molecular fingerprints used were generated from simplified molecular-input line-entry system (SMILES) of the halogen bond donor²³⁴. This study investigates a rather small range of halogen bonding character, being limited to only ammonia as the acceptor and the halogen bond donors all being halogenated benzene derivatives, but predicts binding energies comparable to those calculated using DFT.

5.2 The SHEF1552 σ -hole dataset

The work covered in Section 5.2 is included in the to-be-submitted manuscript "The SHEF1552 σ -hole data set: Insights into intermolecular interactions involving chalcogens, halogens, pnictogens, and tetrrels. Harry W. Nash, J. Grant Hill," A draft of this manuscript is provided in Appendix C.3.

Considering the existing datasets relevant to σ -hole interactions: the most well known are subsets of the GMTKN55 benchmark database, the HAL59 set of halogenated dimers, and the PNICO23 set of pnictogen...pnictogen dimers²³⁵. These two datasets do not exclusively contain σ -hole interactions, but also includes various other noncovalent interactions. The CHAL336 set prioritises chalcogen bonding interactions, but this not only includes σ -hole but also the similar π -hole interactions²³⁶. A more recent dataset relevant to this work is the SH250x10 subset of the Non-Covalent Interactions Atlas project, with is comprised of 250 complexes containing halogen, chalcogen, or pnictogen bonds²³⁷. The "x10" component of the dataset name refers to a ten-point dissociation curve made around the equilibrium separation of each noncovalent bond. The σ -hole donor elements using in this dataset only include chlorine, bromine, iodine, sulfur, selenium, phosphorus, and arsenic; which is more limited than the σ -hole donor elements studied in this thesis. Another issue with the SH250x10 set is the inconsistency of acceptors used between the different classes of σ -hole interactions, which makes comparison between the classes of σ -hole interactions within the set problematic. All of the datasets mentioned above were produced to benchmark the accuracy of various DFT functionals and other quantum chemistry methods, not for training ML models. With the aim of producing a training set for predicting σ -hole interaction energies we constructed the **Sigma-Hole Energies and additional Factors** (SHEF1552) dataset. This comprises of 1552 complexes constructed from 82 σ -hole donors (21 *X*, 21 *Ch*, 20 *Pn*, 20 *Tt*) and 19 σ -hole acceptors. The σ -hole acceptors chosen were ammonia, phosphine, hydrogen cyanide, acetone, thioacetone, trimethylamine, aniline, furan, pyridine, dioxane, water, acetonitrile, thiophene, imidazole, trimethylphosphine, trimethylphosphine sulfide, trimethylphosphine oxide, nitromethane, and pyridine-N-oxide. This results in a total of 1558 combinations (399 *X*-bonding, 399 *Ch*-bonding, 380 *Pn*-bonding, 380 *Tt*-bonding), but six structures were removed after geometry optimisation due to the structures not adopting geometries consistent with σ -hole interactions. These complexes were chosen so as to cover a broad range of σ -hole interaction types, interaction energies, σ -hole angles, and donor-acceptor distances, while also being representative of the structures found within the CSD searches such as those in Chapter 3 and Chapter 4.

The SHEF1552 dataset was constructed so that the σ -hole of the donor atom lines up with the lone-pair of the acceptor atom, and then the resulting complexes were optimised using density functional theory with the M06-2X exchange correlation functional, including the D3(0) dispersion correction, and the def2-TZVPD basis set^{118 120 238 239}. The optimisations and DFT interaction energies were performed using ORCA version 5.0.4²⁴⁰. The σ -hole angle and normalised donor-acceptor interatomic separation (analogous to the normalised contact distances mentioned in Chapter 3 and Chapter 4) were obtained from these optimised complexes.

In order to calculate accurate interaction energies for these optimised complexes with a large number of electrons, a form of localexplicitly correlated coupled cluster was utilised. Thus using the MOLPRO package of *ab initio* programs (version

2023.2), counterpoise-corrected interaction energies were calculated at the PNO-LCCSD(T*)-F12b level of theory^{241 242}. As can be seen in Figure 5.1 the distribution of normalised donor-acceptor interatomic separation against the σ -hole angle is comparable to many of the analogous distributions from CSD data shown in Chapter 3 and Chapter 4. Halogen and tetrel bonding complexes show the most linear σ -hole angles, while pnictogen bonding complexes favour a σ -hole angle around 170°, and chalcogen bonding complexes have the broadest range in σ -hole angles, see Table 1 of Appendix C.3. The normalised distances of these complexes range from 0.6, denoting very strong complexes with significant covalent character, and >1.0 which are weaker and likely dispersion dominated. The trends in interaction energies against σ -hole angle and normalised donor-acceptor interatomic separation are shown in Figure 5.2, drawing interesting parallels to the trends in Figure 5.1. The strongest σ -hole interactions are associated with near linear σ -hole angles, but some of the strongest pnictogen bonds have σ -hole angles around 170°. At normalised donor-acceptor separations less than 0.8 there are interaction energies that begin to get stronger (-15 kcal/mol) and then approaching 0.6 normalised separation the interaction energies can reach up to -40 kcal/mol, giving further suggestion of covalent character.

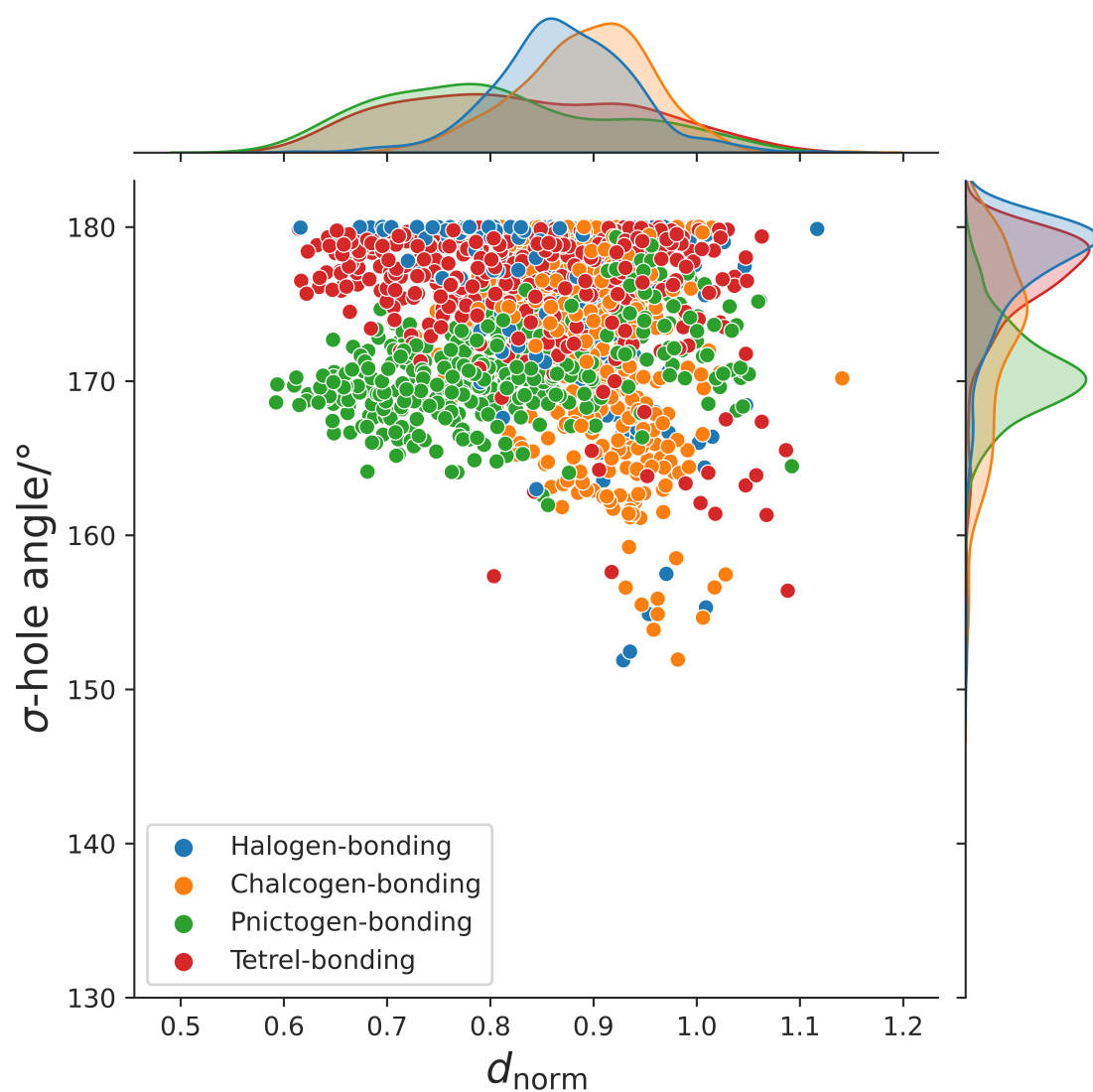


Figure 5.1: Distributions of σ -hole angle against normalised donor-acceptor interatomic distances of the optimised σ -hole interacting complexes in the SHEF1552 dataset. The complexes were optimised at the M06-2X-D3(0)/def2-TZVPD level of theory. Points colour coded by σ -hole interaction type. Each interaction type is also shown as kernel density estimate plots grouped by σ -hole interaction type.

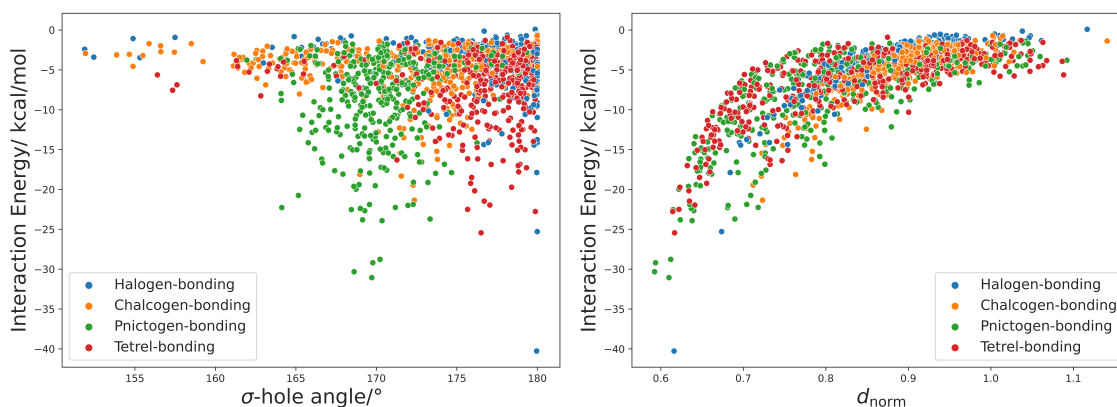


Figure 5.2: Distributions of σ -hole angle against interaction energy (left), and normalised donor-acceptor interatomic distance against interaction energy(right) of the optimised σ -hole interacting complexes in the SHEF1552 dataset. The complexes were optimised at the M06-2X-D3(0)/def2-TZVPD level of theory, and interaction energies were calculated at the PNO-LCCSD(T*)-F12b level of theory. Points colour coded by σ -hole interaction type.

5.2.1 Train-Test Split

The training and test sets for these models comprise of a 75:25 split of the SHEF1552 dataset that is stratified by the type of σ -hole interaction. This ensures that there are proportional numbers of $X/Ch/Pn/Tt$ bonds in the train and test sets. This results in a training set of 1164 complexes, comprising of 299 halogen bonding (106 Cl, 97 Br, 96 I), 297 chalcogen bonding (92 S, 105 Se, 100 Te), 183 pnictogen bonding (68 P, 69 As, 76 Sb, 70 Bi), and 285 tetrel-bonding (72 Si, 71 Ge, 71 Sn, 71 Pb) complexes. The test set contains 388 structures, with 99 halogen bonding (27 Cl, 35 Br, 37 I), 99 chalcogen bonding (38 S, 28 Se, 33 Te), 95 pnictogen bonding (25 P, 26 As, 19 Sb, 25 Bi), and 95 tetrel-bonding (23 Si, 24 Ge, 24 Sn, 24 Pb) complexes. The average geometric and energetic trends of the training and test sets are shown in Table 5.1 and Table 5.2 respectively.

Table 5.1: Summary of the number of complexes and mean averages of interaction energies (E_{int}), σ -hole angle, and normalised contact distances within the training set

σ -hole Interaction Type	No. Complexes	E_{int} / kcal/mol	σ -hole Angle/ $^{\circ}$	d_{norm}
Halogen-bonding	299	-4.032	177	0.871
Chalcogen-bonding	297	-5.172	172	0.889
Pnictogen-bonding	283	-8.730	171	0.800
Tetrel-bonding	285	-6.825	177	0.827

Table 5.2: Summary of the number of complexes and mean averages of interaction energies (E_{int}), σ -hole angle, and normalised contact distances within the test set

σ -hole Interaction Type	No. Complexes	E_{int} / kcal/mol	σ -hole Angle/ $^{\circ}$	d_{norm}
Halogen-bonding	99	-3.758	177	0.874
Chalcogen-bonding	99	-4.257	173	0.901
Pnictogen-bonding	95	-7.354	170	0.819
Tetrel-bonding	95	-6.857	176	0.831

The distributions of normalised contact distance and σ -hole angle are compared between the training and test sets in Figure 5.3 and Figure 5.4. The distributions of interaction energy against σ -hole angle and normalised contact distance are compared between the training and test sets in Figure 5.5 and Figure 5.6. Despite having one third as many structures as the training set, the test set shows analogous character in its distributions. The main disparities between the test set and the training set occur in the extrema of angles, normalised distances, and interaction energies, which contain the least concentration of data points. For example, interaction energies stronger than -15 kcal/mol are few in number and thus there will be differing sampling of the various σ -hole interaction types within this range. Similar issues of sampling arise in the normalised contact distances greater than 1.0 due to a scarcity of data points at this extreme.

Overall the test set is quantitatively comparable to the training set, with the mean averages of the geometric parameters and interaction energies only varying by small amounts. The distributions of normalised contact distance and σ -hole angle in the training set are qualitatively comparable to the distributions from the CSD presented in Chapter 3 & Chapter 4. The main difference present is the lack of interactions with $d_{norm} > 1.0$ that have σ -hole angles that deviate markedly from linearity. This is due to the training set being made of gas-phase optimised molecular structures, as opposed to experimental crystal structures where these longer range interactions are likely to be perturbed by other packing interactions.

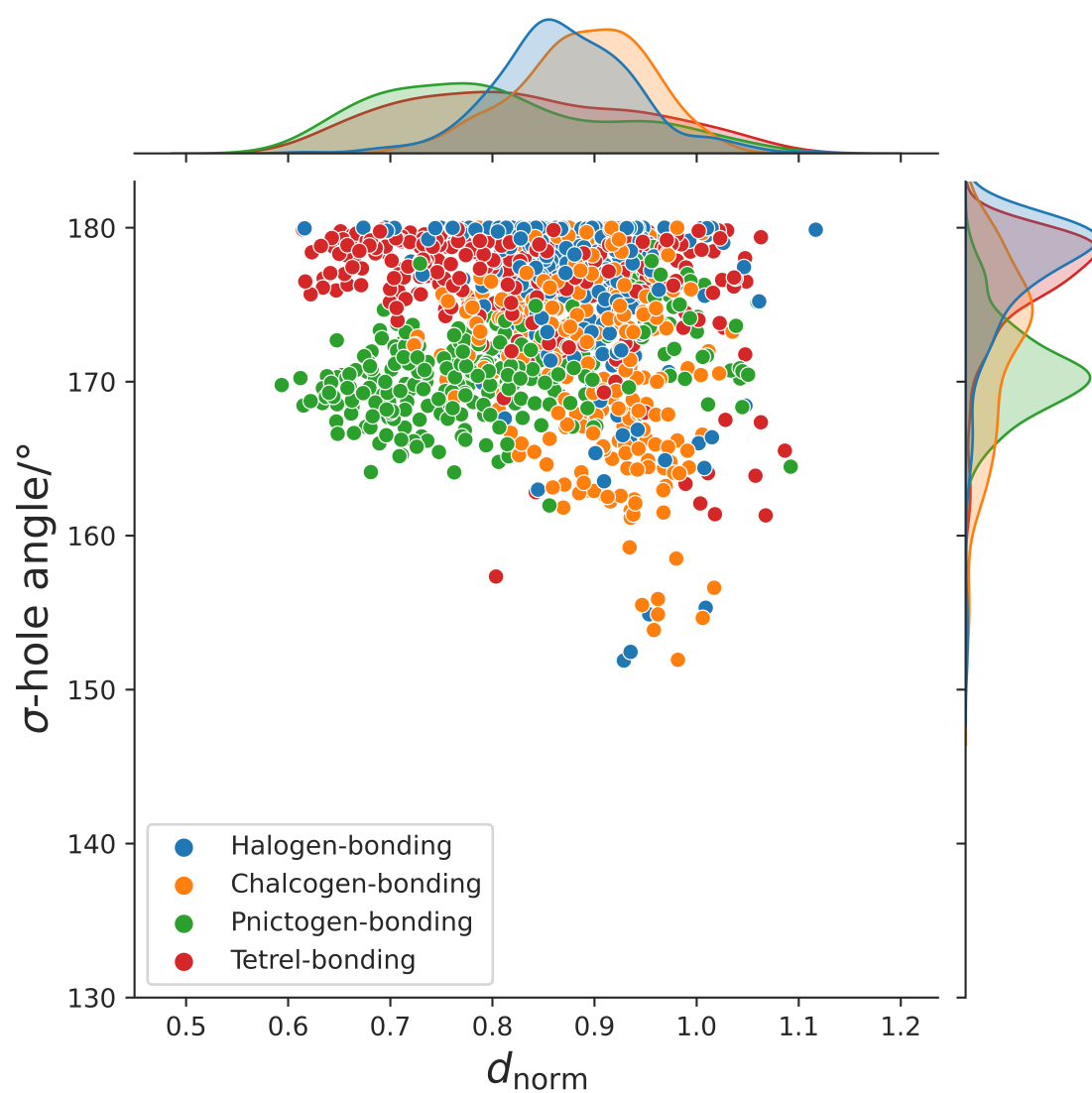


Figure 5.3: Distributions of σ -hole angle against normalised donor-acceptor interatomic distances of the optimised σ -hole interacting complexes in the training set. The complexes were optimised at the M06-2X-D3(0)/def2-TZVPD level of theory. Points colour coded by σ -hole interaction type. Each interaction type is also shown as kernel density estimate plots grouped by σ -hole interaction type.

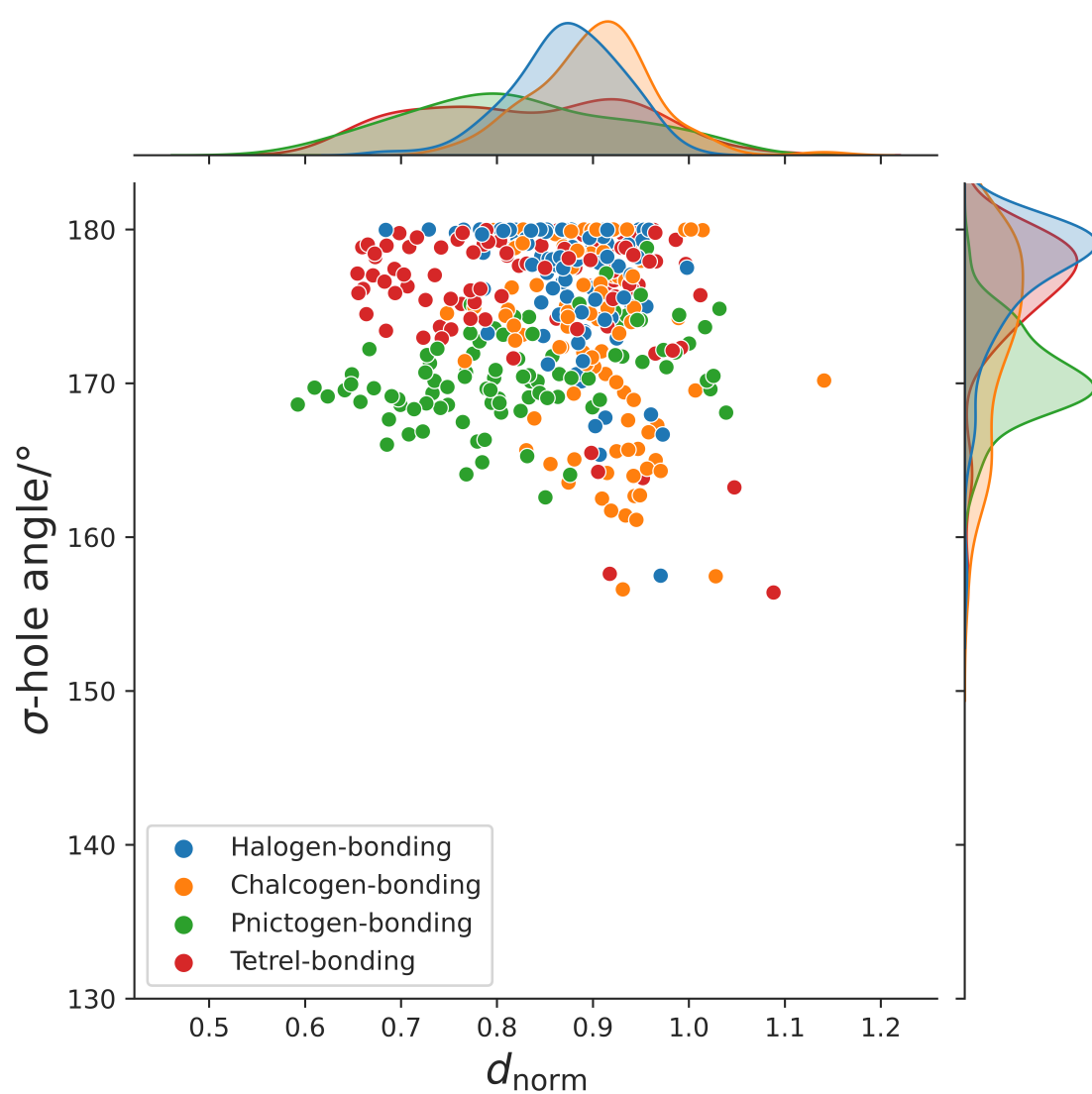


Figure 5.4: Distributions of σ -hole angle against normalised donor-acceptor interatomic distances of the optimised σ -hole interacting complexes in the test set. The complexes were optimised at the M06-2X-D3(0)/def2-TZVPD level of theory. Points colour coded by σ -hole interaction type. Each interaction type is also shown as kernel density estimate plots grouped by σ -hole interaction type.

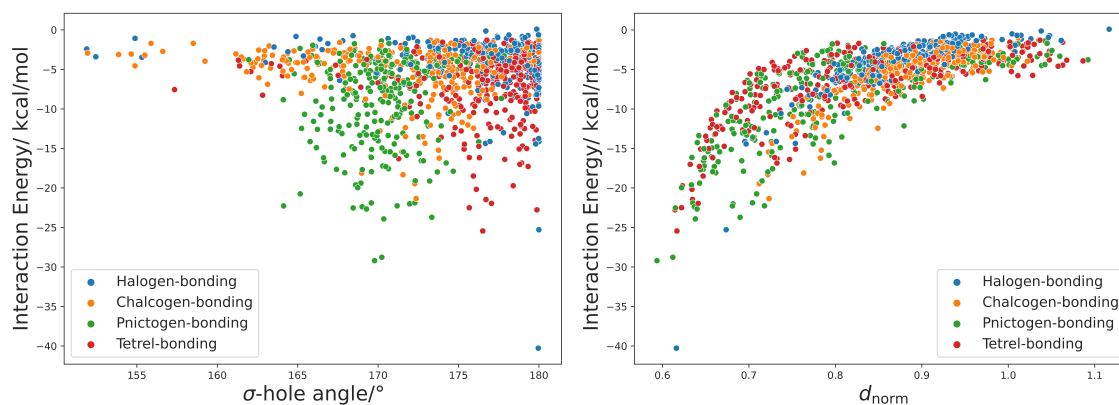


Figure 5.5: Distributions of σ -hole angle against interaction energy (left), and normalised donor-acceptor interatomic distance against interaction energy(right) of the optimised σ -hole interacting complexes in training set. The complexes were optimised at the M06-2X-D3(0)/def2-TZVPD level of theory, and interaction energies were calculated at the PNO-LCCSD(T*)-F12b level of theory. Points colour coded by σ -hole interaction type.

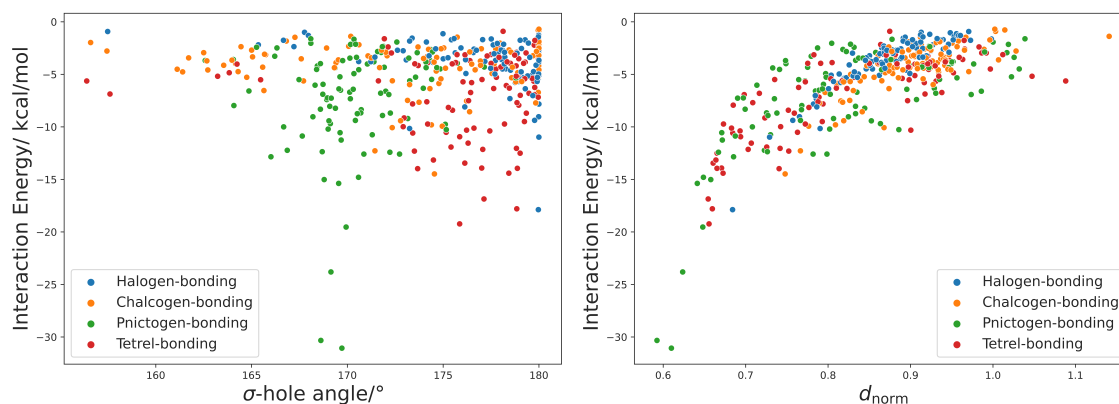


Figure 5.6: Distributions of σ -hole angle against interaction energy (left), and normalised donor-acceptor interatomic distance against interaction energy(right) of the optimised σ -hole interacting complexes in test set. The complexes were optimised at the M06-2X-D3(0)/def2-TZVPD level of theory, and interaction energies were calculated at the PNO-LCCSD(T*)-F12b level of theory. Points colour coded by σ -hole interaction type.

It is also important to test the performance of those chosen DFT interaction energy calculation method, as this method will be used to calculate interaction energies for structures from the CSD (see Section 5.5.2). The performance of M06-2X(D3-ZERO)/def2-TZVPD against the benchmark LCCSD(T*)-F12b/cc-pVDZ-F12 interaction energies for the training set and test set are tabulated in Table 5.3. Plots of these two interaction energy calculation methods for the training set and test are shown in Figure 5.7 and Figure 5.8, respectively. The metrics of root mean squared deviation, coefficient of determination, and Pearson correlation coefficient show good agreement between the two methods, showing that this combination of exchange-correlation functional and basis set is a suitable method to validate the performance of a ML model against for larger systems where the use of wavefunction-based methods such as LCCSD(T*)-F12b is unfeasible due to computational cost.

Table 5.3: Comparison of the DFT (M06-2X-D3(0)/def2-TZVPD) against the benchmark LCCSD(T*)-F12b interaction energies in the training and test sets. Metrics include root mean square deviation (RMSD), mean average deviation (MAD), standard deviation of errors (SD), coefficient of determination (R^2), and Pearson correlation coefficient (ρ). Interaction energies are in kcal/mol.

Set	RMSD	MAD	S.D.	R^2	ρ
Training	1.04	0.76	0.85	0.96	0.99
Test	1.02	0.74	0.86	0.95	0.99

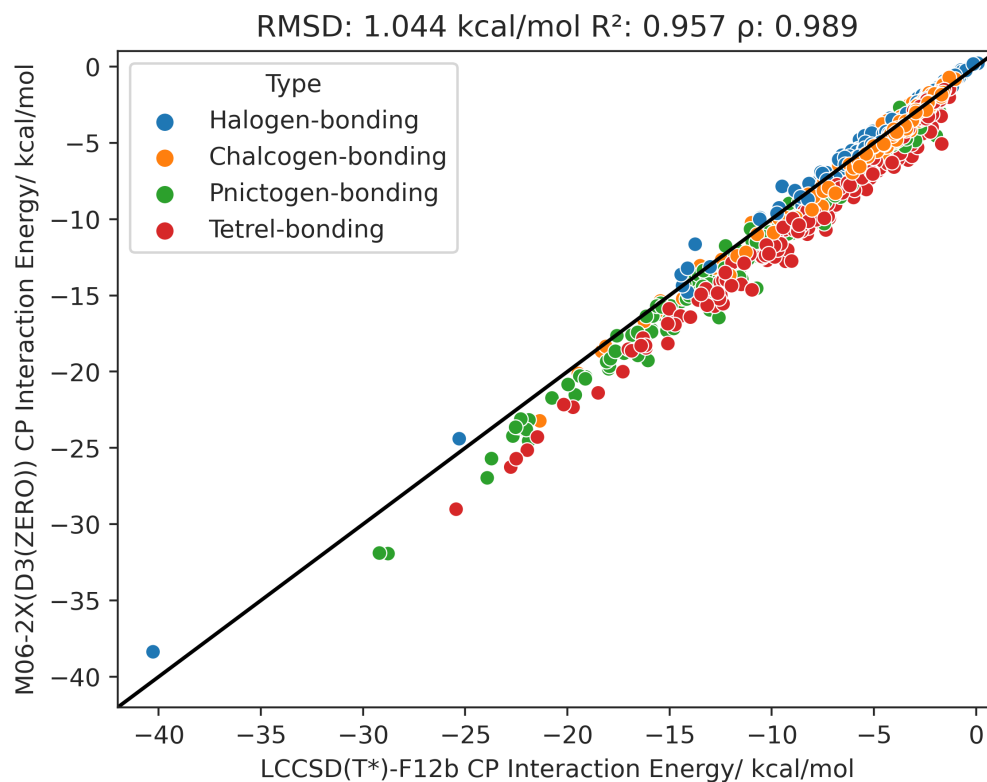


Figure 5.7: Performance of M06-2X-D3(0) compared to LCCSD(T*)-F12b for calculating σ -hole interaction energies in the training set. Metrics include root mean squared deviation (RMSD), coefficient of determination (R^2), and Pearson correlation coefficient (ρ). Points are colour coded by σ -hole interaction type over a line of $y = x$.

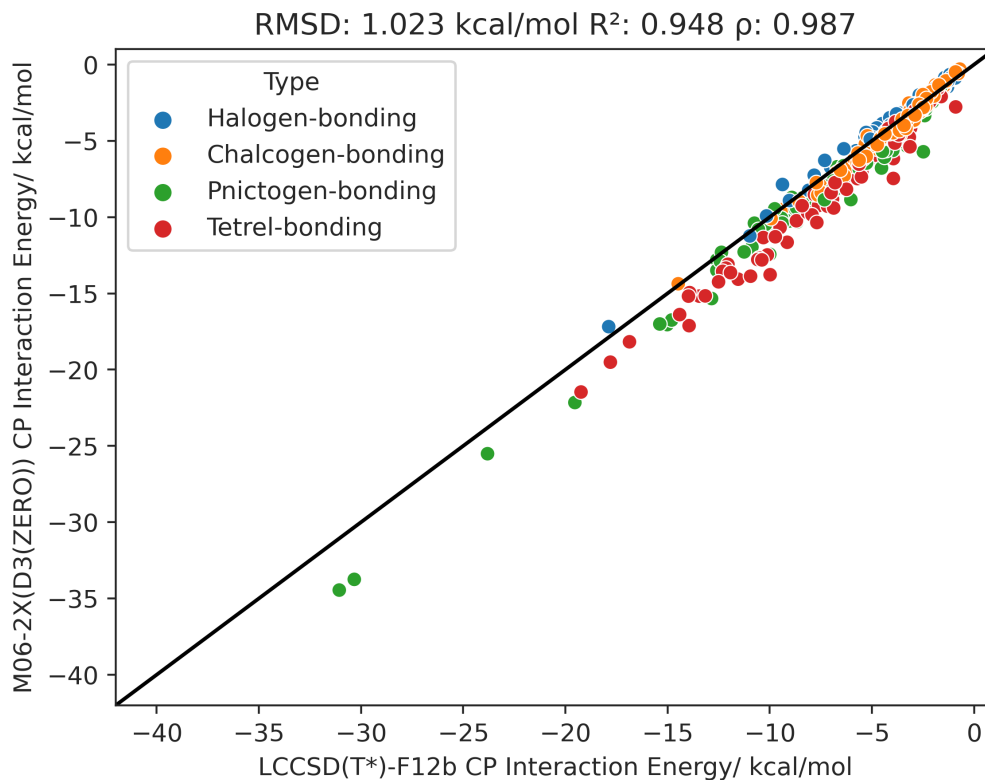


Figure 5.8: Performance of M06-2X-D3(0) compared to LCCSD(T*)-F12b for calculating σ -hole interaction energies in the test set. Metrics include root mean squared deviation (RMSD), coefficient of determination (R^2), and Pearson correlation coefficient (ρ). Points are colour coded by σ -hole interaction type over a line of $y = x$.

5.3 Descriptors

In order to represent the σ -hole interaction complexes and their makeup in a way that a machine learning model can interpret and learn from, one needs to select features, or descriptors, to do so. For chemistry applications, one approach to descriptors may be based on experimental or computed properties of the studied system, for example electronegativities or dipole polarisabilities of elements. A different approach is to provide some sort of molecular representation in a format that is readily machine-readable. This includes methods such as molecular graphs, where the molecule is constructed from nodes (atoms) and edges (bonds) with further descriptors assigned to each, but using a graph format lends itself to complex ML approaches such as graph neural networks. Arising from similar origins, an incredibly popular approach to molecular representation is **Simplified Molecular Input Line Entry System** (SMILES) notation²³⁴. This is a text- or string-based representation of a molecular system, where atomic symbols are used along with other alphanumeric characters to represent bonding or hybridisation. Some example SMILES strings for simple molecules include: "C=C" (ethene), "[CH2]=O" (formaldehyde), and "C1=CC=CC=C1" (benzene). Other similar string-based approaches include InChI/InChIKey, DeepSMILES, and SELFIES^{243–245}. These methods of molecular representation derive geometries based on graph connectivities and basic bonding rules associated with p -block elements, resulting in idealised conformations

that are coherent for the majority of organic chemistry. However, there is no way of defining bond lengths or angles, so one cannot describe molecules that deviate from ideal, and likewise cannot handle the complex bonding modes seen in metal-organic chemistry.

Structures found within the CSD are experimentally determined and include a massive diversity of chemical bonding and makeup, requiring molecular representation that is sensitive to non-ideal geometries and a wide range of chemistry. A useful approach that is used in applications such as database substructure similarity searching involves a method of molecular representation called molecular fingerprinting.

5.3.1 Molecular Fingerprints

Molecular fingerprints are a useful machine-readable format due to their vectorised nature, which captures the atomic configuration within the molecule. There are various methods to convert a molecule into bit vector format, with most only utilising the 2D graph of the molecule, but some methods can account for 3D structures. The main methods for creating fingerprints are substructure key fingerprints, topological fingerprints, and circular fingerprints²⁴⁶.

The bit vectors in substructure key fingerprints are determined by the presence of substructures within the molecule corresponding to certain structural keys. The choice of these structural keys depends on the use case for the fingerprints, as bits will not be encoded if the substructures are not present within the molecule and as such would not be represented. Thus, the length of these bit vectors can vary based on the structural keys used along with their presence in the molecule in question, which does not happen with other types of fingerprints. Examples of typically used substructure keys based fingerprints include: **M**olecular **A**ccess **S**ystem (MACCS), PubChem fingerprints, and **B**arnard **C**hemical **I**nformation Ltd. (BCI) fingerprints^{247–249}.

Topological fingerprints, or path-based fingerprints, encode bits based on fragmented paths along bonds within the molecule, then hashing all of these paths to form the fingerprint. This means that all the fingerprints, regardless of the molecule, are the same length and can be adjusted as needed. The process of hashing means that any single bit within the fingerprint cannot be traced back to a single feature of the molecule, meaning they are not human-readable. These fingerprints are typically vectors of 2048 bits, but can be longer in some cases and use different connectivity paths for the bit encoding. Examples of such fingerprints include: the RDKit fingerprint, atom-pair fingerprint, and topological-torsion fingerprint^{250–252}.

Lastly, circular fingerprints are a variant of hashed topological fingerprints, but rather than exploring paths within the molecular graph, the environment of a given radius around each atom centre is encoded into the bit vector. This makes these fingerprints unsuitable for substructure searching, but makes them ideal for similarity searching between full structures. There are various types of circular fingerprint, such as Molprint2D or Functional-Class Fingerprints, but the archetypal circular fingerprints are Extended-Connectivity Fingerprints, detailed below. An extensive overview and comparison of many different molecular fingerprints can be found in a review on their use in similarity searching by Cereto-Massagué and co-workers²⁴⁶.

Extended Connectivity (Morgan) Fingerprints

Extended-Connectivity Fingerprints (ECFPs) were introduced by Rogers and Hahn, with the specific aim to capture molecular features relevant to molecular activity²⁵³. This type of circular fingerprint rapidly calculates a hashed bit vector that represents molecular structures by the circular topological neighbourhood of each atom centre. The features within this bit vector act as representations of particular substructures within the molecule, which gives the advantage of allowing the description of both the presence and absence of functionality. This form of molecular fingerprint has seen extensive use in molecular similarity searching, especially in the context of high-throughput computational drug discovery²⁵⁴.

These fingerprints are sometimes referred to as Morgan fingerprints due to them being derived from a variation of the Morgan algorithm²⁵⁵. The Morgan algorithm was designed for recognising isomorphic molecules, which are identical molecules with differing atom numbering. This is an iterative process that starts with assigning numeric identifiers to each atom, based on their inherent atomic properties. Then in the next iteration of the algorithm new identifiers are generated based on the surrounding identifiers from the previous iteration. These iterations are repeated until every atom has a unique identifier (symmetry allowing), with the final numbering scheme containing unique identifiers that are independent of the original atom numbering.

The changes made to this approach in the ECFP algorithm are that it terminates within a set number of iterations rather than until uniqueness, and replacing the computationally expensive recoding after each iteration to avoid mathematical overflow or "collisions" (where two identical atom environments can result in the same identifier) with a fast-to-compute hashing scheme. This hashing function converts the identifiers of the molecular fingerprint into a unique fixed length numerical representation that is deterministic. This maintains a consistent format for the fingerprints, and enables their storage and retrieval in a memory efficient manner.

The ECFP generation process can be summarised in a sequence of three steps:

1. Initial assignment stage - each atom is given an integer identifier generated by hashing properties e.g. atomic number, mass, charge.
2. Iterative neighbour dependent stage - each atom identifier is updated based on the identifiers of the nearest neighbours, including whether it is a structural duplicate of other features. This step is repeated a specified number of times.
3. Duplicate identifier removal stage - if the same feature occurs multiple times it is reduced to a single representative in the final feature list.

The number of times the iterative stage is repeated effectively expands the size of the information about the surrounding molecule that is encoded into each atom identifier. The standard number of iterations used for ECFPs is two, so each atom identifier contains information of the molecular system two bonds away from the atom. These final atom identifiers are then combined into a 2048 long vector, which forms the representation of the molecule.

5.3.2 Dimensionality Reduction

Descriptors such as molecular fingerprints, which are very long vectors, are treated by many machine learning models bitwise per feature. This means that using an ECFP of standard format would be treated as 2048 one-hot encoded features, which is a huge number of features for an ML model and will likely contribute to overfitting via the introduction of noise. Thus for the application of predicting interaction energies of σ -hole interactions, the irrelevant components of these fingerprints must be excluded to leave only the properties of the fingerprint relevant to the interaction. The approach to this comes under the umbrella of dimensionality reduction, where data that contains lots of features is transformed from this "high-dimensional" space into a smaller "low-dimensional" space while preserving important information. Along with potentially improving the fitting of an ML model, dimensionality reduction also speeds up computation.

A common approach to dimensionality reduction is Principle Component Analysis (PCA), which interprets the covariance of the features in the dataset²⁵⁶. Covariance is a representation of the relationship between two features, whether they increase/decrease together or against each other or have no relationship. The covariance of all of the features with each other is constructed into a covariance matrix **C**, shown in Equation (5.1).

$$\mathbf{C} = \frac{1}{n-1} \mathbf{X}^T \mathbf{X} \quad (5.1)$$

Where **X** is the mean centred data matrix (mean-subtracted data), and $\mathbf{X}^T$ is the transpose of **X** (rows and columns swapped), and n is the number of samples of each feature. **X** would have the dimensions $a \times b$ where a is the number of observations, and b is the number of features. In the case of applying this to ECFPs the standard length of these fingerprints are 2048 digits, this would define number of features b . The diagonal elements of **C** show the variance of each feature, and the off-directional elements show the covariance between feature pairs. The covariance matrix **C** is then framed as an eigenvalue problem [Equation (5.2)] where the eigenvectors v are the direction in which the data has the greatest variance (the principal component) and the eigenvalues λ are the magnitudes of the variance along those principal components.

$$\mathbf{C}v = \lambda v \quad (5.2)$$

These eigenvectors and eigenvalues are then sorted in descending order of eigenvalue, which allows the most principal components to be selected. The total variance can be obtained from the sum of all eigenvalues, then the fraction of the total variance for each principal component is given as an explained variance ratio. Thus, you can preserve the number of principal components that are responsible for the desired proportion for total variance, for example the selected principal components k that account for 95% of the total variance. Then the original centred data matrix **X** is projected onto the matrix of selected principal components $\mathbf{V}_k$, so the original data is now transformed to the lower dimensional space of the selected principle components, as shown in Equation (5.3).

$$\mathbf{X}_k = \mathbf{X} \cdot \mathbf{V}_k \quad (5.3)$$

5.3.3 Descriptor Selection for σ -hole Interaction Energy Prediction

Building on the concept of dimensionality reduction, different fingerprints were tested from those available in the RDKit Python package, such as ECFPs, RDKit fingerprints, atom-pairs fingerprints, and topological torsion fingerprints. The scikit-learn Python package's PCA function was used to map the original fingerprints to principal components accounting for 99% of the variance, which massively reduces the size of the fingerprint, as shown in Figure 5.9 for the example of ECFPs²⁵⁷. To complement these reduced fingerprints, the σ -hole angle and the normalised donor-acceptor interatomic separation are also provided as descriptors to provide some representation of the geometric characteristics of the complexes. It is important to note that the fingerprint will not contain any features that directly describe the σ -hole interaction, but will allow ML models to be able to find similarities between the complexes and predict an interaction energy. Similarly, the σ -hole and normalised distance descriptors will not give a full description of the geometric dependence of σ -hole interactions as they are all derived from optimised geometries. Full description of angular and distance distortion would require the training set to include bond dissociation curves and angular scans to be calculated for all of the complexes in the SHEF1552 set.

Each of the fingerprints was tested using a default XGBoost regression model to evaluate their performance²⁵⁸. The metrics for the prediction of the models being root mean squared deviation (RMSD), coefficient of determination (R^2), and Pearson correlation coefficient (ρ). Across all of the metrics shown in Table 5.4, it can be seen that extended connectivity fingerprints show the best performance, and as such was selected for use in the final model. Figure 5.11 compares the performance of this regression model using the different fingerprints against the PNO-LCCSD(T*)-F12b counterpoise-corrected interaction energy ground-truth values from the test set, and further fingerprint comparisons can be found in Appendix C.1.

Table 5.4: A comparison of root mean squared deviation (RMSD), coefficient of determination (R^2), and Pearson correlation coefficient (ρ) for different molecular fingerprints with a default XGBoost regression model for predicting σ -hole interaction energies.

Fingerprint	RMSD / kcal/mol	R^2	ρ
ECFP	1.181	0.896	0.955
RDKit	1.248	0.885	0.951
AtomPair	1.202	0.890	0.955
TopologicalTorsion	1.570	0.825	0.920

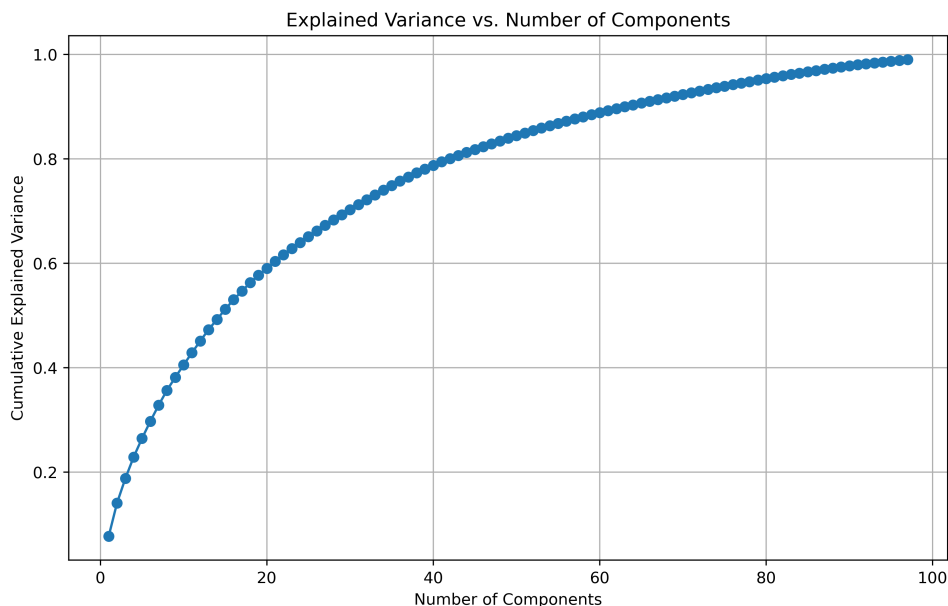


Figure 5.9: A plot of how each of the 97 principal components of the ECFP contribute to the cumulative explained variance across the SHEF1552 data set. The original size of the fingerprint is 2048 characters, and the objective of the PCA was to retain 99% of the variance in the fingerprints.

5.4 Models

It is clear that predicting σ -hole interaction energies is a regression problem, but there are a great deal of different algorithms to choose from to construct a model. The simplest and most widely used method of regression is linear regression, detailed in Equation (5.4), where y is the dependent variable, X is the independent variable, β_0 is the intercept, β_1 is the gradient, and ϵ is the error term. To fit the linear model to the data, β_0 and β_1 are varied to minimise the value of the error term ϵ , typically done through least-squares fitting method.

$$y = \beta_0 + \beta_1 X + \epsilon \quad (5.4)$$

A different set of approaches to regression problems involves decision trees, in which the model learns a series of decision rules based on the training data, then these rules are applied in prediction on unseen data. Each decision is made at a "node" that splits the set which eventually leads to a larger dendritic structure, giving the tree nomenclature, with a representation shown in Figure 5.10. The "branches" or subsection of the tree will eventually terminate into a "leaf" node where no decision is made, which is the final result. The issue with decision trees is that they are unstable, meaning that the sampling of the training data or the features used to make the decisions can produce drastically different results.

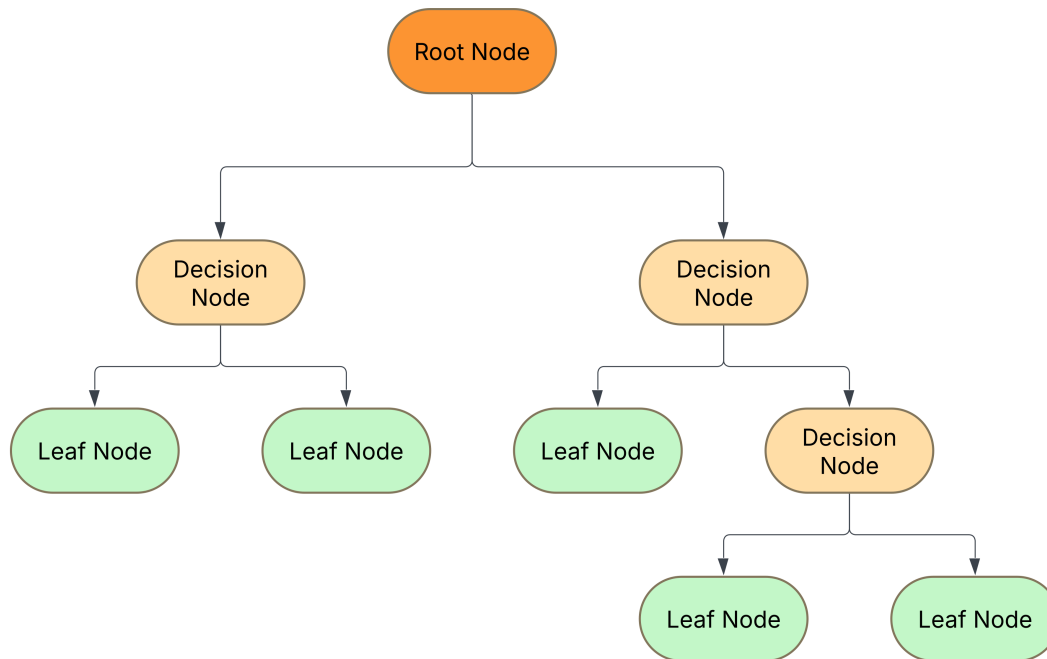


Figure 5.10: A representation of a decision tree used in a machine learning model.

Due to this weakness, decision tree models are rarely used in isolation, and instead used in ensemble learning methods. An example of this is a Random Forest model, where a collection of trees are independently trained²⁵⁹. The training for each decision tree is done using bootstrapped sampling, randomly sampling subsets of the overall training data, along with a random selection of features. Then the average prediction of the ensemble of trees (a forest), each of which has been trained on a different sample of the training data and different features, is taken to be the overall prediction. This mitigates the weakness of decision trees through the rationale that although one decision tree may be unreliable, the aggregate collective decision of an ensemble of trees will be more reliable. Another popular form of decision tree based ML model is gradient boosting, this also involves an ensemble of trees but works sequentially, rather than in parallel. New trees are added to the model to mitigate the error of the tree that came before, until the overall error is within an acceptable threshold. The sequential decision trees are trained on the "residuals" (errors) of the previous model, so it can determine what is missing from the previous model. The term gradient boosting comes from the fact that the algorithm utilises the gradient of the overall error function to make adjustments to the model to minimise this error, and the boosting comes from the sequential addition of the decision trees. A widely used form of gradient boosting is the XGBoost model, which stands for **eXtreme Gradient Boosting** and implements many beneficial regularisation, data-handling, and hyperparameter features²⁵⁸.

5.4.1 Model Selection for σ -hole Interaction Energy Prediction

The aforementioned regression models, linear regression, random forest regression, and XGBoost regression, were compared utilising the descriptors selected in Section 5.3.3. The linear regression and random forest models were constructed using the scikit-learn Python package, and the XGBoost regressor was constructed using the XGBoost Python package, with all models kept at their default settings^{257 258}. All of the models were trained on the same 75:25 train-test split of the SHEF1552 dataset detailed in Section 5.2.1.

The interaction energy predictions of these regression models were tested via three metrics; root mean squared deviation (RMSD), coefficient of determination (R^2), and Pearson correlation coefficient (ρ), which are presented in Table 5.5. From this testing the XGBoost regression model was selected for the final model to predict σ -hole interactions. The XGBoost model's performance for predicting interaction energies for the test set can be seen in the RMSD and correlation metrics shown in Figure 5.11, and further model comparisons can be found in Appendix C.2.

Table 5.5: A comparison of root mean squared deviation (RMSD), coefficient of determination (R^2), and Pearson correlation coefficient (ρ) for different regression models using PCA-treated ECFPs for predicting σ -hole interaction energies in the test set.

Fingerprint	RMSD/ kcal/mol	R^2	ρ
Linear	1.733	0.810	0.906
Random Forest	1.279	0.864	0.951
XGBoost	1.181	0.896	0.956

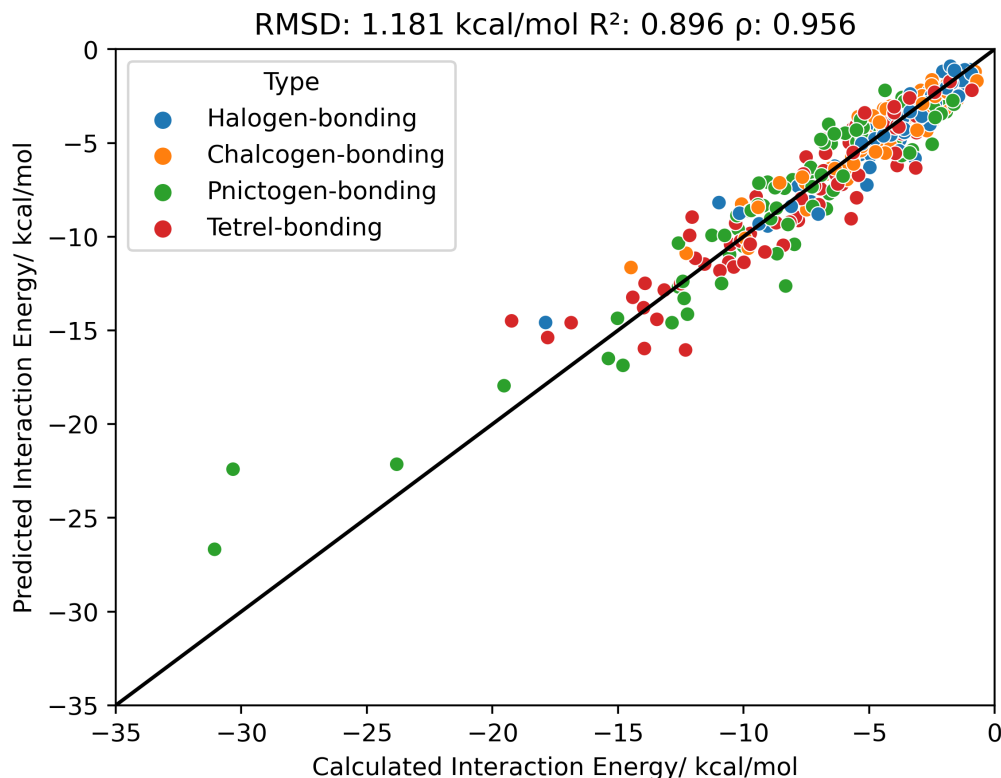


Figure 5.11: Performance of an XGBoost regression model with PCA-treated ECFPs for predicting σ -hole interaction energies within the test set. Metrics include root mean squared deviation (RMSD), coefficient of determination (R^2), and Pearson correlation coefficient (ρ). Points are colour coded by σ -hole interaction type over a line of $y = x$.

5.5 The σ -Hole Approximation Tool (σ HAT)

With the descriptors and model architecture determined, fine tuning of the model will maximise accuracy and prevent overfitting. This is done via optimising the settings and configuration pertaining to the function of the model itself, also called hyperparameters. Details for the σ HAT model can be found in Appendix C.3.

5.5.1 Fitting & Hyperparameter Optimisation

Hyperparameter optimisation for the XGBoost regression model was done via Bayesian search cross-validation using the scikit-optimize Python package²⁶⁰. The available hyperparameter space was defined, which can be found in Appendix C.3, and was optimised for five iterations, three cross validation folds and a negative root mean squared error function. Bayesian search cross validation is a highly efficient means of searching for the optimum hyperparameters for the model, with a standard grid search being a expensive step, especially when evaluating a large parameter space. Bayesian optimisation uses a Gaussian Process as a "surrogate" for the combinations of hyperparameters, giving a probabilistic model that gives an uncertainty estimate to allow efficient exploration of the parameter space. The hyperparameter optimisation step took 1 minute and 45 seconds and the only manually specified parameter was the early stopping rounds parameter which was

set to 4. The optimised hyperparameters are included in Appendix C.3.

The fitting of the optimised XGBoost model took 1.04 seconds, and terminated in 317 fitting cycles, also known as epochs. The minimisation of the root-mean-square error of the training and test sets are plotted against the fitting cycles in Figure 5.12. The training error and test error begin to plateau around 300 epochs and the small separation between the two errors suggests the model is well fit. This results in a model, hereby referred to as the σ -Hole Approximation Tool (σ HAT), with a test RMSD of 1.03 kcal/mol which is very close to the standard 1 kcal/mol required for "chemical accuracy"²⁶¹. The σ HAT predicted interaction energies compared to the DFT interaction energies from the test set, shown in Figure 5.13, display notable agreement between the predicted and calculated interaction energies, but underbinding can be seen in stronger interactions.

Table 5.6 presents a comparison summary for the performance of the σ HAT model and DFT against the benchmark interaction energies in the test set. The root-mean-squared deviation of the σ HAT predicted interaction energies is very comparable to that for the counterpoise-corrected M06-2X-D3(0)/def2-TZVPD interaction energies, having only 0.005 kcal/mol greater RMSD. Interestingly, the mean absolute deviation of the σ HAT predicted interaction energies is slightly smaller than for the DFT interaction energies, but the standard deviation of the errors is larger for the σ HAT interaction energies compared to the DFT interaction energies. In terms of R^2 and ρ DFT performs better than the σ HAT model for determining σ -hole interaction energies.

Overall, while the RMSD, MAD, and SD of the errors is comparable, the σ HAT model is marginally worse in terms of the coefficient of determination and Pearson correlation coefficient when it comes to determining σ -hole interaction energies than DFT. This is not surprising given DFT is a fully fledged quantum chemistry calculation, whereas the σ HAT model is a low-cost ML model. Table 5.7 shows a comparison of the summary statistics of σ HAT predicted, DFT, and the benchmark calculated interaction energies of the test set structures. This shows good agreement between the predicted interaction energies and the two methods of calculated interaction energies apart from the strongest σ -hole interactions, where the σ HAT model drastically underestimates the interaction energies, as seen in the difference in minimum interaction energies in Table 5.7. In contrast the DFT approach slightly overestimates these strongest interaction energies, along with slightly overbinding compared to LCCSD(T*)-F12b on average.

Table 5.6: Comparison of the performance of the σ HAT model and DFT (M06-2X-D3(0)/def2-TZVPD) against the benchmark LCCSD(T*)-F12b interaction energies in the test set. Metrics include root mean square deviation (RMSD), mean average deviation (MAD), standard deviation of errors (SD), coefficient of determination (R^2), and Pearson correlation coefficient (ρ). Interaction energies are in kcal/mol.

E_{int}	Determination	RMSD	MAD	S.D.	R^2	ρ
σ HAT		1.03	0.68	1.03	0.92	0.97
DFT		1.02	0.74	0.86	0.95	0.99

Table 5.7: Summary statistics of the interaction energies predicted by the σ HAT model and calculated via DFT (M06-2X-D3(0)/def2-TZVPD) and LCCSD(T*)-F12b for the structures within the test set. Metrics include the mean, standard deviation, minimum, and maximum interaction energies in kcal/mol.

E_{int}	Determination	Mean	S.D.	Min	Max
	σ HAT	-5.56	3.62	-23.95	-0.62
	DFT	-6.09	4.51	-34.45	-0.30
	LCCSD(T*)-F12b	-5.53	4.01	-31.06	-0.70

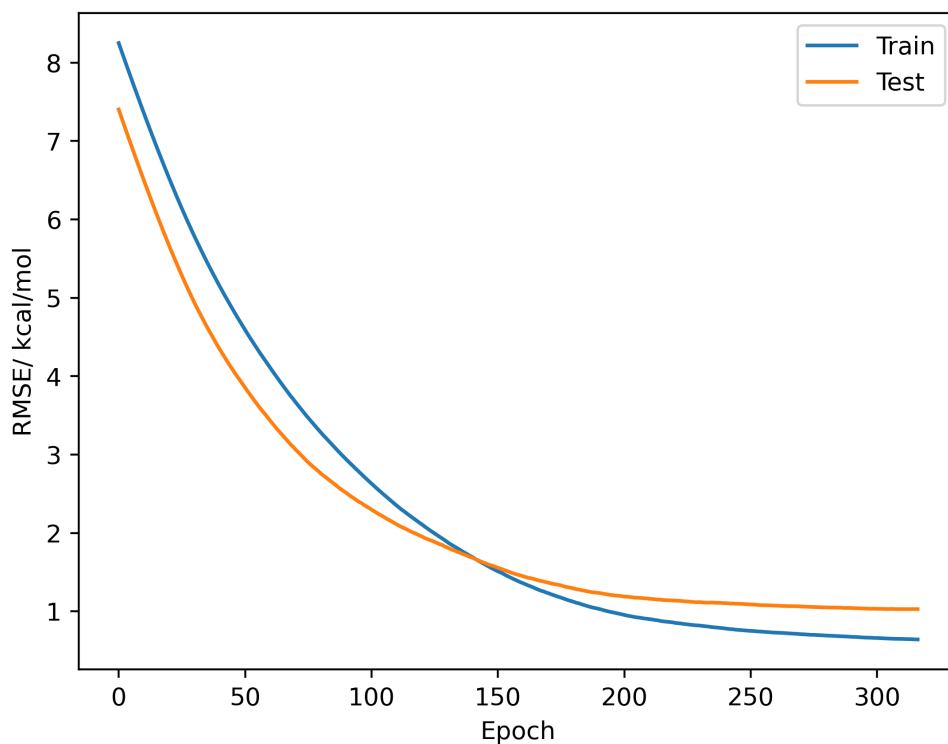


Figure 5.12: Plots of minimisation of the RMSE loss across the test and train sets as the σ HAT model was fit over 317 epochs to predict σ -hole interaction energies from the SHEF1552 set of sigma-hole complexes.

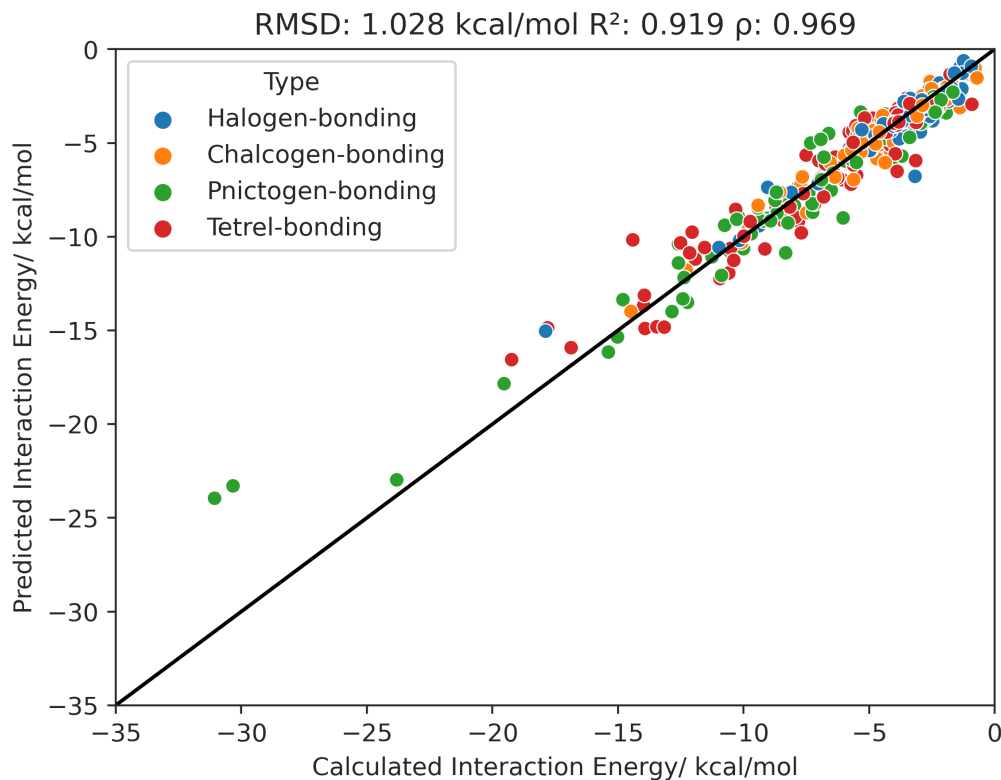


Figure 5.13: Performance of the σ HAT predicted σ -hole interaction energies compared to the LCCSD(T*)-F12b benchmark interaction energies in the test set. Metrics include root mean squared deviation (RMSD), coefficient of determination (R^2), and Pearson correlation coefficient (ρ). Points are colour coded by σ -hole interaction type and plotted over a line of $y = x$.

Feature Importance

In order to understand the main influences that the selected descriptors have on the performance of the σ HAT model for predicting σ -hole interaction energies, one can interpret the feature importances. The various importance measurements utilised within the XGBoost Python package are gain-based, cover-based, and split-based. The gain importance measure relates to the average of the error reduction of the model when the feature is used to split the data, thus a feature with a higher gain value would be a more important feature for increased prediction accuracy. The total gain is the feature's gain measure summed across all of the decision trees in the XGBoost model, showing the overall importance of the feature for reducing the error across the whole model, not just averaged over each decision tree. The cover importance measure corresponds to the average number of decision tree samples that involve a split based on the feature, showing how widely a feature is applied in model decisions. The total cover is the sum of the cover importance measures for each of the features, showing the total predictions over the whole XGBoost model that are affected by the feature. The split-based importance measure, more commonly referred to as weight, relates to the number of times a feature is used in all splits within all of the XGBoost decision trees. Thus larger values for these importance metrics indicate a greater feature importance, but from different perspectives.

The various feature importance measurements for the σ HAT model descriptors are shown in Table 5.8. The σ -hole angle and normalised contact distance features are examined singularly, whereas the fingerprint descriptor is interpreted both as a sum and an average of the importance measures for the 97 components of the PCA-reduced extended-connectivity fingerprints. In terms of the singular most important feature, the normalised contact distance has the greatest total gain, being greater than all the other features combined, and is the single descriptor that is involved in the most splits and samples across the σ HAT model. The σ -hole angle is similar or greater in importance to each individual fingerprint component feature, yet much less important than the normalised contact distance. The overall summed importance of the fingerprint descriptors is a major contributor to the σ HAT models predictive capability, which is encouraging as this is the means of encoding the molecular structure and suggests that this significantly affects the interaction energy predictions. One cannot consider the importance of each element of the PCA-reduced ECFP individually, as they possess no human interpretable meaning, given they are dimensionally reduced hashed bit vectors, there is no complete way of reverse-engineering the chemical significance of a fingerprint element.

Table 5.8: Feature importances (dimensionless) of the descriptors used in the σ HAT model to predict σ -hole interaction energies. The Fingerprint descriptor represents the importance values for the 97 components of the PCA-reduced extended-connectivity fingerprint.

Feature	Gain	Total Gain	Cover	Total Cover	Weight
σ -hole Angle	6.4	476.6	222.7	16703.0	75.0
Normalised Contact Distance	65.9	34263.1	451.2	234632.0	520.0
Fingerprint (Sum)	585.0	24034.6	23671.1	832107.0	3320.0
Fingerprint (Average)	6.0	247.8	244.0	8578.4	34.2

5.5.2 Validation Performance

The aim of constructing the σ HAT model is to predict interaction energies between the σ -hole donor and acceptor moieties within the CSD, such as those discussed in Chapter 3 and Chapter 4. There are a few potential issues with this, mainly that the optimal geometry of a gas phase complex will differ from the optimal geometry of the same complex in solid state, due to other packing interactions perturbing the system.

It is also not possible to resolve a single interaction from the greater crystal lattice, with there being multiple noncovalent interactions involved, such as dispersion, hydrogen bonding, π - π interactions and Coulombic forces. Attempts to isolate individual interactions have been made using electron density approaches, yet require electronic structure calculations along with associated costs, such as the PIXEL method²⁶². The approach used in this work is to take a molecular 'cutout' of the crystal around the contact indicating a σ -hole interaction and use this to calculate an interaction energy which will subsequently be used to validate the performance of the σ HAT model on unseen data that is significantly different to its training set.

Molecular Cutouts of Crystals

There are a variety of approaches to simplifying a large chemical system for the calculation of chemical or physical properties. These "fragmentation" or "partitioning" methods allow the computational investigation into systems that would otherwise be unfeasible, such as complex proteins, crystal structures, and materials²⁶³. The majority of these approaches are energy or electron density based, for example the Additive Fuzzy Density Fragmentation (AFDF) method^{264 265 266 267}. The AFDF method utilises a custom-made database of fragment electron densities, reflecting the system to be fragmented. This allows the calculation of electronic charge distributions within a larger molecule from these pre-calculated "fuzzy" fragment electron densities. These methods require datasets containing regularised systems, meaning the molecules need to be comprised of known functional groups. As the structures contained within the CSD are immeasurably diverse, one would need an equally diverse database of fragment electron densities.

Other approaches involve the cleaving of specific bonds within the system, for example the Molecular Fractionation with Conjugate Caps (MFCC) method of Zhang and co-workers^{268 269}. This approach is designed for accurate calculations of protein-molecule interactions, and is fractionated based on the peptide N-C bonds within the polypeptide chains. This approach is very effective but relies on the regular nature of proteins as the peptide bonding motif is conserved among all polypeptides. Developments on this methodology apply it to more complex organic systems, such as the methodology of Deev and Collins²⁷⁰. This works by fragmenting the system based on single bonds between organic functional groups, with the size of the fractions dictated by a 'level' system which extends how many bonds away the cleavage will occur. These cleaved atoms are capped with hydrogen atoms to saturate the valencies resulting from the broken bonds. This methodology was tested for the use in calculating nonbonded interactions within 96 organic molecule crystal structures obtained from the CSD²⁷¹. Readers interested in more information on molecular fragmentation are directed to the thorough review on energy-based methods by Collins and Bettens²⁷².

Due to the lack of consistency between structures found within the CSD, and with an aim to minimise the need for electronic structure calculations, a broadly applicable generalised approach to fragmentation must be implemented. This is achieved by the use of the CSD Python API's Crystal Module, where one can obtain the packing shell of crystal returned as a ".mol2" file of geometric Cartesian coordinates⁹⁷. In order to locate the longest of the σ -hole contacts within the packing shell, a packing shell size of 30 (contains 30 molecules) was selected. From this packing shell the intermolecular distances between potential σ -hole donor and acceptor atoms were measured to find the one matching the contact found in the corresponding CSD search. This allows the atomic indices within the geometry files of the σ -hole donor and acceptor atoms to be defined, which the cutout will be constructed upon. The cutout is made using a graph theory approach, with the molecule first being loaded into RDKit and then used to construct two Networkx (version 3.1) directed graphs, one originating from the donor atom, and the other from the acceptor atom^{250 273}. Each graph proceeds out five bonds from the origin atom and returns the indices of these atoms. These indices are then extracted from the geometry file of the packing shell and written to a new .xyz geometry file. The motivation for the five bond limit to the cutout is to describe a substituted halobenzene moiety, illustrated in Figure 5.14, as it is a commonly used as a halogen bond donor. The cutout

at this stage will likely contain atoms with unsaturated valencies, unless the crystal is made of small molecules, and as such OpenBabel (version 3.1.0) is utilised to add hydrogen atoms to the terminal nodes of the cutout graphs to satisfy the valency²⁷⁴. An example molecular cutout construction of CSD Refcode **WUNFET** is detailed in Figure 5.15.

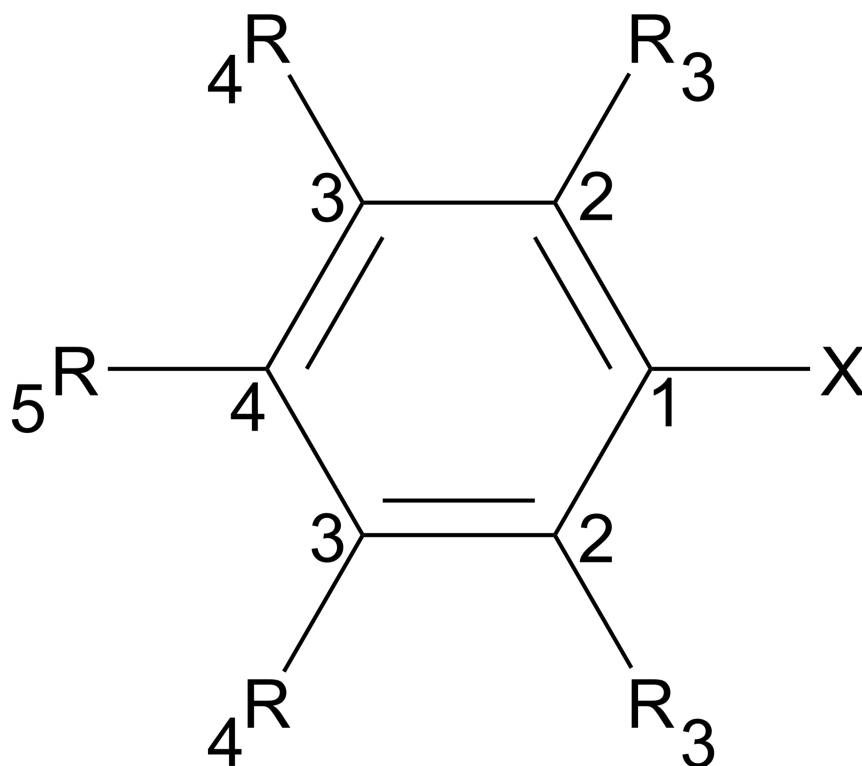


Figure 5.14: Example of the molecular graph that defines the σ -hole donor component of the molecular cutout of a crystal. The graph aims to retain a substituted halobenzene moiety, so includes the atoms five bonds away from the donor atom X. Another graph of the same construction is made, originating from the acceptor atom, and the atoms of these two graphs are combined into the molecular cutout, which is written to a new geometry file. Subscripts denote how many bonds away from the donor atom X each atom is.

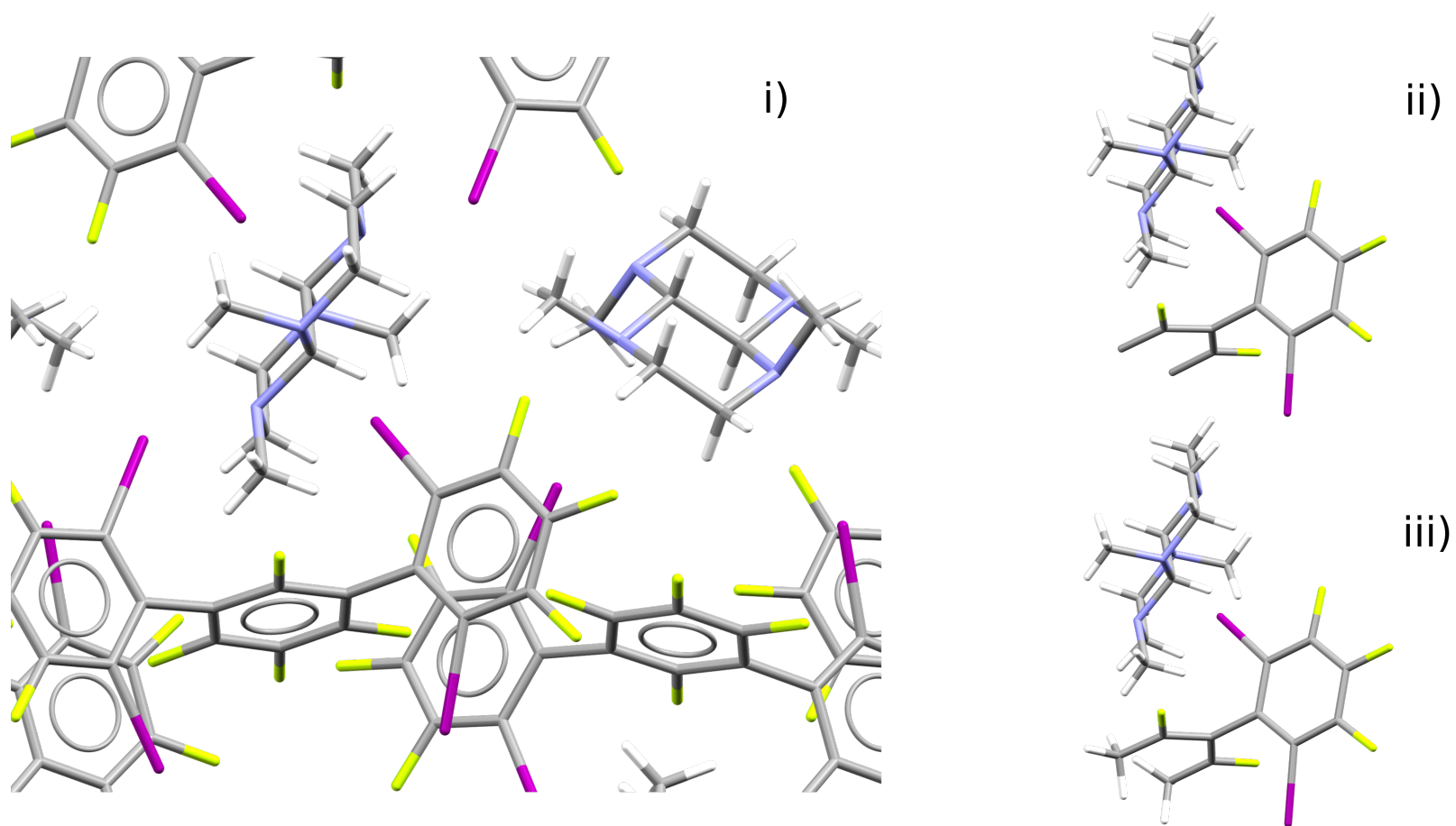


Figure 5.15: Illustration of the molecular cutout process for CSD Refcode **WUNFET**. i) Part of packing shell containing 30 molecules ii) Uncapped cutout iii) Capped cutout

The structures selected for this validation set are the hits resulting from the halogen bonding to sp^3 -hybridised nitrogen query, as detailed in Section 3.2, as they tend to show strong contacts that have σ -hole angles close to linearity. These short, linear contacts, along with the nature of halogens and the sp^3 -hybridised nitrogen acceptor, mean there are less likely to be other interactions in the vicinity of the donor/acceptor atoms that could contaminate the σ -hole interactions. There are several potential problems with this molecular cutout scheme, mainly owing to the diverse variety of chemistry available in the CSD that cannot be interpreted by cheminformatics tools such as RDKit and OpenBabel. Due to exotic bonding modes and/or transition metals, sometimes a valid molecule object cannot be constructed within these tools, may be constructed erroneously, or have valencies that cannot be saturated with hydrogens by OpenBabel. This leads to a moderate failure rate in achieving calculated interaction energies for these molecular cutouts, with 517 cutout structures (from a search of an older version of the CSD 2023.2.1).

These 517 cutouts undergo the same counterpoise corrected M06-2X-D3(0)/def2-TZVPD interaction energy calculations as the SHEF1552 data set. These 517 cutouts resulted in 346 valid DFT calculated interaction energies. The coupled cluster interaction energies are unable to be calculated on these cutouts as their much greater size would be prohibitively expensive, with the largest cutout containing 88 atoms compared to the largest structures in the SHEF1552 set containing 38 atoms. The cutouts are also used to calculate ECFPs, which are treated with the same PCA routine as in the training set. This, combined with the σ -hole angle and normalised contact distance is provided to the σ HAT model to predict interaction energies for each cutout. Structures are removed if the EFCP was not able to be initialised with the molecule, and if secondary interactions are found within the cutout. This was determined via a Python script written to identify any intermolecular distances, not including those between the σ -hole donor and acceptor, that are shorter than the sum of van der Waals radii. This process results in 309 valid cutouts with both a calculated and predicted interaction energy, along with ECFPs, which make up the validation set.

Results

The determination of descriptors and prediction of interaction energy for the 309 molecular cutouts took 0.4 seconds using a single 3.8 GHz core of a AMD RYZEN 7 5700G, which is negligible in comparison to calculating DFT interaction energies. The overall performance of the σ HAT model on the molecular cutout validation set is shown in Figure 5.16 as a scatterplot of predicted vs benchmark σ -hole interaction energies, with the RMSD of the predictions being roughly twice as large as that for the test set shown in Figure 5.13.

The summary statistics presented in Table 5.9 shows that the σ HAT predicted interaction energies are generally larger in magnitude compared to the DFT calculated interaction energies for the cutouts in the validation set. Interestingly, sign of the errors is reversed at either extremes of interaction energies, with the strongest DFT interaction energies being underestimated, and the weakest DFT interaction energies being overestimated. This is analogous to the summary statistical comparison of the σ HAT model and DFT interaction energies seen in the test set, see Table 5.7, in which DFT shows notable overbinding in comparison to benchmark LCCSD(T*)-F12b interaction energies.

Table 5.9: Summary statistics of the interaction energies predicted by the σ HAT model and calculated via DFT (M06-2X-D3(0)/def2-TZVPD) for the 309 molecular cutouts of CSD structures showing halogen bonding to sp^3 -hybridised nitrogen. Metrics include the mean, standard deviation, minimum, and maximum interaction energies in kcal/mol.

E_{int}	Determination	Mean	S.D.	Min	Max
	σ HAT	-6.87	3.65	-20.10	-1.84
	DFT	-7.27	4.43	-25.31	-0.16

A comparison of the model's performance based on the identity of the halogen acting as the σ -hole donor is shown in Table 5.10. The best performance is seen in the bromine halogen bonding species, with an RMSD of 1.550 kcal/mol which is well below the overall model RMSD, and R^2 and ρ values nearing 1.0. The performance with iodine halogen bond donors shows the worst RMSD, with a value of 2.257 kcal/mol, and as the most numerous halogen bond donor in the set it makes up the majority of the overall model RMSD of 2.018 kcal/mol. The coefficient of determination of the iodine complexes is less than half that of the bromine complexes, but still shows a notably high Pearson correlation coefficient. The worst overall performance is seen with the chlorine complexes, despite having a comparable RMSD to the bromine complexes, the Pearson correlation coefficient is nearer zero and the R^2 value is negative in value, showing that the performance is worse than simply predicting the mean of the interaction energy. This suggests that the bromine containing molecular cutouts from the CSD are more comparable to the training set and are thus described well, and while the model is capturing some of the underlying trends in the iodine contacts, there is some deficiencies in either the descriptors or training set.

The σ HAT model is not able to describe the underlying relationships within the chlorine containing molecular cutouts, suggesting that these contacts differ in character to those indicative of a σ -hole interaction. This difference in character is also seen in the DFT calculated interaction energies, when one compares the summary statistics of the σ HAT predicted and DFT calculated interaction energies shown in Table 5.11 & Table 5.12. For other halogen bonding contacts, a shorter normalised contact distance and a σ -hole angle nearer linearity would indicate a stronger interaction energy. The chlorine contacts in the validation set however, have consistent interaction energies regardless of the normalised contact distance or σ -hole angle, shown in both the σ HAT predicted and DFT calculated interaction energies. These interactions are the weakest, and thus show the largest relative error, along with little directionality with respect to the σ -hole angle. This, along with the poor R^2 and ρ values with respect to the σ HAT model for chlorine interactions, see Table 5.11, suggests these contacts are dissimilar to the chlorine halogen bonds found in the SHEF1552 set. This all gives evidence to the possibility that these chlorine contacts are unlikely to be caused by a σ -hole interaction and are likely to be caused by other packing factors within the crystal.

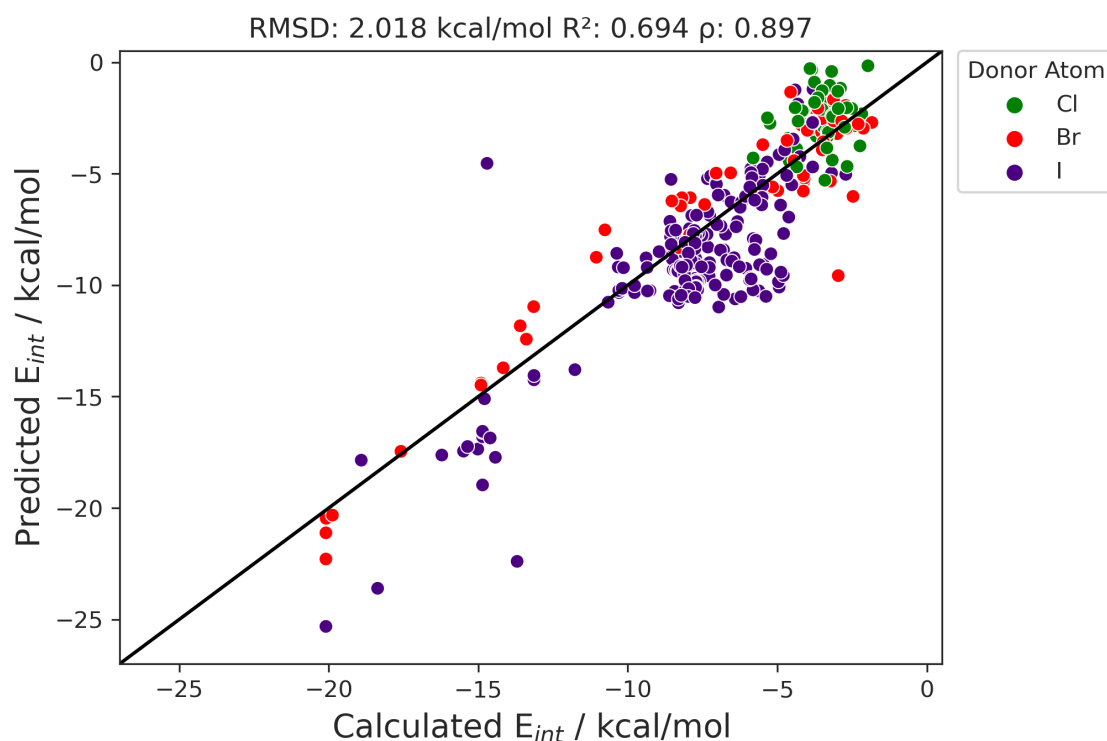


Figure 5.16: Performance of the σ HAT predicted interaction energies against DFT (M06-2x(D3(0))/def2-TZVPD) interaction energies of the 309 molecular cutouts showing halogen bonding to sp^3 -hybridised nitrogen. Metrics include root mean squared deviation (RMSD), coefficient of determination (R^2), and Pearson correlation coefficient (ρ). Points are colour coded by halogen bond donor atom and plotted over a line of $y = x$.

Table 5.10: A comparison of root mean squared deviation (RMSD), mean absolute deviation (MAD), standard deviation of errors (SD), minimum and maximum errors, coefficient of determination (R^2), and Pearson correlation coefficient (ρ) for the σ HAT model σ -hole interaction energy performance against DFT (M06-2X-D3(0)/def2-TZVPD) calculated values for the validation set. The 309 molecular cutouts showing halogen bonding to sp^3 -hybridised nitrogen that make up the validation set are subdivided by the different halogen bond donor atoms. Metrics represented with energies in kcal/mol.

Halogen	No. Structures	RMSD	MAD	SD	Min	Max	R^2	ρ
Cl	50	1.61	1.32	1.30	-3.63	2.01	-3.27	0.22
Br	71	1.55	1.13	1.52	-3.24	6.61	0.90	0.95
I	188	2.26	1.68	2.01	-10.17	8.69	0.42	0.84

Table 5.11: Summary statistics of the interaction energies predicted by the σ HAT model for the 309 molecular cutouts of CSD structures showing halogen bonding to sp^3 -hybridised nitrogen separated by halogen species. Metrics include the mean, standard deviation, minimum, and maximum interaction energies in kcal/mol.

Halogen	Mean	S.D.	Min	Max
Cl	-3.52	0.79	-5.81	-1.97
Br	-6.55	4.87	-20.10	-1.84
I	-7.89	2.98	-20.97	-2.71

Table 5.12: Summary statistics of the DFT (M06-2X-D3(0)/def2-TZVPD) calculated interaction energies for the 309 molecular cutouts of CSD structures showing halogen bonding to sp^3 -hybridised nitrogen separated by halogen species. Metrics include the mean, standard deviation, minimum, and maximum interaction energies in kcal/mol.

Halogen	Mean	S.D.	Min	Max
Cl	-2.55	1.21	-5.30	-1.97
Br	-6.21	4.97	-22.29	-1.34
I	-8.93	3.67	-25.31	-1.21

With regards to the normalised contact distance, plotted alongside interaction energies in Figure 5.17, the grouping of points is much closer in the predicted energies (*left*) compared to the calculated interaction energies (*right*). The normalised contact distance is the most important singular descriptor in the σ HAT model, as shown in Table 5.8, and this close grouping of predicted interaction energies as a function of this descriptor suggests that the model is potentially overfit to the normalised contact distance.

The performance of the σ HAT model relative to the identity of the halogen donor atom shown in Table 5.11 compared to the DFT interaction energies relative to the identity of the halogen donor atom in Table 5.12 proves to be an interesting comparison. The chlorine interaction energies have small standard deviations in both the predicted and DFT calculated interaction energies (0.787 and 1.214 kcal/mol, respectively). The trend in bromine DFT calculated interaction energies with normalised contact distance is mirrored very closely in the predicted interaction energies. There is some agreement between the predicted and DFT calculated interaction energies for the iodine complexes, with similar mean interaction energies between predicted and DFT calculated interaction energies, yet there is the greatest change in standard deviation between them. In the right panel of Figure 5.17 these deviations can be seen in the plots of normalised contact distance against interaction energy are mainly found in the 0.7-0.9 range of normalised contact distance.

The trends in the interaction energies with the σ -hole angle, plotted in Figure 5.18, show that once again chlorine complexes differ little in interaction energies over a wide range of σ -hole angles (111-175°). One would expect halogen bonding systems that have σ -hole angles closer to 180° to be more strongly interacting than a system with a σ -hole angle closer to 90°. This gives more evidence to the notion that these contacts are unlikely to be attributable to σ -hole interactions, and are likely close contacts due to other crystalline packing effects. Thus it makes it unsurprising that a model trained on σ -hole interactions would be unsuited to predicting these interaction energies. The distributions of interaction energies against σ -hole angle and normalised contact distance shown in Figure 5.18 & Figure 5.17 show that the geometric trends for bromine and iodine contacts are replicated rather well between predicted and calculated interaction energies. A notable issue is that there are no predicted interaction energies significantly stronger than -20 kcal/mol, suggesting a deficiency of very strong halogen bonding interactions within the training set. This is somewhat true, with the SHEF1552 set containing only two halogen bonding structures with LCCSD(T*)-F12b interaction energies stronger than -15 kcal/mol.

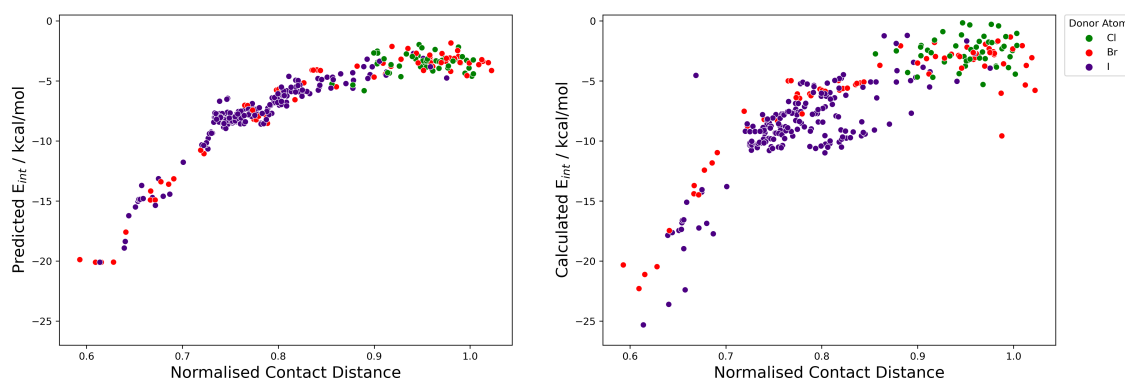


Figure 5.17: σ HAT predicted interaction energies (left) vs DFT (M06-2X-D3(0)/def2-TZVPD) interaction energies (right) against normalised contact distance for the 309 molecular cutouts showing halogen bonding to sp^3 -hybridised nitrogen. Points colour coded by halogen bond donor atom.

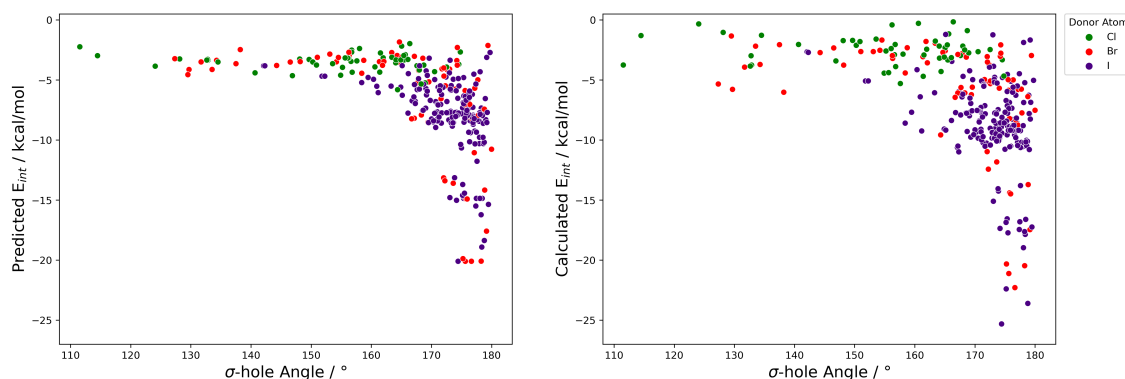


Figure 5.18: σ HAT predicted interaction energies (left) vs DFT (M06-2X-D3(0)/def2-TZVPD) interaction energies (right) against σ -hole angle for the 309 molecular cutouts showing halogen bonding to sp^3 -hybridised nitrogen. Points colour coded by halogen bond donor atom.

When examining the outlier structures where the absolute error of the predicted interaction energy is greater than 1.5 times the interquartile range, shown in Table 5.13, it can be seen that the majority of these structures contain iodine σ -hole donors. Upon inspection of the structures, they fall into two distinct groupings, one containing "cage"-like σ -hole acceptors such as DABCO for example CSD Refcode **HEKZOO**, **IBOZOS**, **IBOZUY**, **UZUNEL**^{275–277}. The other containing multiple halogen bonding interactions to large nitrogen heterocycles such as trans-decahydrotetraazapyrene shown in CSD Refcode **LORQUG**²⁷⁸. This group also contains the structures **WULPUR**, **WUNDUH**, **WUNFOD**, and **WUNGOE**, which are structures engineered to exhibit multipoint halogen bonds²⁷⁹. This indicates areas in the training set that can be improved upon, as analogous Lewis bases are not present in the training set. The interaction associated with CSD refcode **HMTNTI** has a DFT interaction energy that is weaker than the σ HAT predicted interaction energy, on inspection this is due to an error in the ORCA input where the geometry of the acceptor was not input, and as such is not a representative calculation for the system²⁸⁰. Other sources error include structures like **BEYQAX**, where hydrogens were erroneously omitted from the structure and coincidentally resulted in a neutral calculation²⁸¹. The structure of **NERFID**, which is of a halogen bond between diiodotetrafluorobenzene and a morpholine derivative, looks to be enhanced

in calculated interaction energy by a weak hydrogen bond to one of the fluorines, which can be seen in Figure B.5²⁸².

Table 5.13: Summary of geometric measurements, σ HAT predicted and DFT(M06-2X-D3(0)/def2-TZVPD) calculated interaction energies in kcal/mol of outlier structures from the 309 molecular cutouts showing halogen bonding to sp^3 -hybridised nitrogen making up the validation set.

Refcode	Donor	σ -hole Angle/ $^\circ$	Normalised Contact Distance	$E_{int}(\sigma\text{HAT})$	$E_{int}(\text{DFT})$
BEYQAX	Br	164	0.988	-2.96	-9.57
HEKZOO	I	178	0.656	-14.86	-18.97
HMTNTI	I	175	0.669	-14.71	-4.54
IBOZOS	I	179	0.640	-18.37	-23.60
IBOZUY	I	175	0.658	-13.70	-22.40
LORQUG	I	172	0.804	-5.62	-10.26
LORQUG	I	171	0.808	-5.81	-10.22
LORQUG	I	174	0.818	-4.94	-10.09
LORQUG	I	171	0.821	-4.96	-9.85
LORQUG	I	175	0.841	-4.80	-9.57
LORQUG	I	174	0.845	-4.88	-9.42
NERFID	I	167	0.794	-6.41	-10.62
UZUNEL	I	174	0.614	-20.10	-25.31
WULPUR	I	172	0.806	-6.22	-10.52
WULPUR	I	167	0.821	-5.38	-10.51
WUNDUH	I	167	0.803	-6.96	-10.98
WUNFOD	I	171	0.830	-5.88	-9.72
WUNGOE	I	169	0.843	-5.36	-9.29

5.6 Summary

In this work the σ HAT (σ -Hole Approximation Tool) model is presented; a simple, low-cost machine learning model was constructed and trained on the SHEF1552 data set of σ -hole interactions which, on the test set, performs comparably to density functional theory calculations. This model utilises the σ -hole angle, the normalised σ -hole donor-acceptor interatomic separation, and an extended connectivity molecular fingerprint of the complex which is treated with principle component analysis to reduce dimensionality. To apply this model to predicting interaction energies within close contacts found within the CSD indicative of σ -hole interactions, the first two descriptors are readily retrieved. The molecular fingerprints of these cutouts prove to be more challenging, as it can be difficult for cheminformatics software to be able to interpret the wide variety of bonding environments found in the CSD. These fingerprints were determined from molecular cutouts taken from the packing shells of the crystals containing the contact. The cutout was taken five bonds out from each of the donor and acceptor atoms and the valencies of the terminal atoms were saturated with hydrogens. The σ HAT model predictions of interaction energies for molecular cutouts taken from the halogen bonding searches to sp^3 -hybridised nitrogen have an RMSD of 2.018 kcal/mol. While larger than the 1.028 kcal/mol RMSD of the σ HAT predicted interaction energies for the training set, the systems are much more varied within the validation set. It is also important to note that the RMSD for the validation is with respect to DFT rather than LCCSD(T*)-F12b benchmark interaction energies for the test set. The performance of the σ HAT model on the bromine containing complexes is by far the best, followed by iodine contacts, and the chlorine containing structures are very poorly described by the model. This poor performance on the

chlorine containing structures is indicated by the calculated interaction energies of the structures, which does not show typical geometric relationship with the interaction energy that is characteristic of halogen bonding. The lower performance on the iodine containing structures can be explained by Lewis bases that are markedly different from those found in the training set.

Overall this proves to be an effective method of approximating σ -hole interaction energies incredibly quickly. The most time consuming step in determining interaction energies in crystal structures found in the CSD is generating the molecular cutouts, which took around 1 minute for 517 structures, in comparison to the prediction of interaction energies from the resulting 309 cutouts which took less than one second.

Chapter 6

Conclusions & Future Work

6.1 Conclusions

To conclude this thesis, the aims stated in Section 1.6 are reiterated below:

1. Identify relevant donor and acceptor moieties associated with σ -hole interactions. Define geometric and other constraints to enable identification of σ -hole interactions in crystal structures present in the CSD.
2. Use the CSD to determine geometric characteristics and trends for σ -hole interactions as a function of donor and acceptor characteristics to provide insight into the nature of the interactions.
3. Fit a predictive model for σ -hole interaction energies based on quantum-chemical calculations to enable rapid estimation of these energies with good accuracy. Apply this energy estimation to σ -hole interaction motifs found in the CSD.

In Chapters 2, 3, and 4, a methodology for searching the Cambridge Structural Database for close contacts within crystals indicative of σ -hole interactions involving nitrogen and oxygen Lewis bases is presented. This approach utilises ConQuest query files to define the donor and acceptor moieties and associated geometric definitions and measurements combined with the CSD Python API to automate, process, and present the results. This methodology is applicable to each new update of the CSD, so can automatically recognise new structures deposited into the CSD that match the criteria defined in the queries and post-processing. This approach, including the queries, along with the associated criteria in post-processing, clearly satisfy the first aim of this work.

Chapter 3 details the geometric trends found within contacts indicating σ -hole interactions to nitrogen Lewis bases. In order to form the closest contacts with σ -hole angles approaching linearity, one needs to complement a higher valency σ -hole donor moiety with a nitrogen Lewis base that has greater s -character, and vice versa. The shortest halogen bonding contacts with σ -hole angles approaching linearity are formed with sp^3 -hybridised nitrogen Lewis base, and on average show longer contacts with Lewis bases of increased p -character. Chalcogen bonding contacts are dominated by sp^2 -hybridised

systems, both in donor and acceptor moieties, but show similar geometric characteristics between contacts involving sp^2 -hybridised and sp -hybridised nitrogen. Pnictogen and tetrel bonding contacts, while rare, have the closest contacts with sp -hybridised nitrogen, along with σ -hole angles approaching linearity. This means that sp^3 -hybridised nitrogen Lewis bases are ideal acceptors for forming very strong halogen bonding complexes, but less useful for participating in pnictogen or tetrel bonding. Lewis bases with an sp^2 -hybridised nitrogen are more easily able to minimise steric effects due to their planar nature, and so are able to participate in a wide range of σ -hole interactions, while retaining a similar strength of interaction compared to sp^3 -hybridised nitrogen Lewis bases. Although many σ -hole interactions with sp -hybridised nitrogen Lewis bases as acceptors are somewhat weak, owing to their contracted lone-pair, the linear nature of these, primarily cyano groups allows them favourable approach to sterically crowded binding sites. This makes them ideally suited to the task of acting as an acceptor for pnictogen and tetrel bond donors. In addition, cyano groups coordinated to metal centres indicate a strengthening impact on σ -hole interactions, highlighting a potential new avenue of research that extends prior work focussing on halogen bonding involving cyanometallates¹²⁸.

Nitrogen also plays a key role as an electron withdrawing group for classes of σ -hole donors such as N-iodosuccinimide or selenazole derivatives, which make up some of the strongest σ -hole interactions identified in this work. In some cases, interactions involving these moieties can begin to approach the extreme of what one may consider to be noncovalent interactions. Other than nitrogen, a variety of elements are covalently bonded to σ -hole donors, the most common being carbon. It is clear that the hybridisation of such carbon atoms have an impact on the strength of the σ -hole interaction. Hybridisation with greater s -character for these carbons atom elicits a increased electron-withdrawing effect, resulting in stronger sigma-hole interactions, as inferred by their associated geometric characteristics.

Chapter 4 details the geometric trends found within contacts indicating σ -hole interactions to oxygen Lewis bases. To enable more detailed chemical interpretation, in this study, these σ -hole acceptors comprise of organic carbonyl, divalent oxygen (R-O-R), oxide (O-R), and metal carbonyl moieties. Other than the oxide Lewis bases, some of which are part of anionic moieties, the neutral oxygen Lewis bases are generally poorer σ -hole acceptors than the nitrogen Lewis bases. Organic carbonyls participate in marginally stronger σ -hole interactions than divalent oxygens, yet divalent oxygens are more tunable, with the varying of the R-O-R substituents affecting the Lewis basicity much more than varying the R-(C=O)-R substituents. Coordination of a carbonyl ligand to a metal shows a deactivating effect, with the contacts generally being longer and showing little to no relationship between shorter normalised contact distances and σ -hole angles approaching linearity. This is, however, consistent with analogous CSD analysis of hydrogen bonding in metallic carbonyls such as that by Desiraju and co-workers¹⁶¹. These investigations into the geometric characteristics for σ -hole interactions with nitrogen and oxygen σ -hole acceptors achieve the second aim of this work.

Of the results presented in Chapter 3 & Chapter 4, the reader might notice the lack of inclusion of the lightest halogen (F), chalcogen (O), pnictogen (N), and tetrel (C) atoms in the data. As mentioned previously, these atoms are very poor σ -hole donors, due to their small size and low polarisability, and thus are not considered. A strong argument can be made from the results of this work that the analogous elements in the third period of the Periodic Table; Cl, S, P, and Si, are also

very poor σ -hole donors. Of the contacts involving these atoms, those involving chlorine and sulfur are in great number, yet only a very small number show contact distances notably shorter than the sum of vdW radii and with little directionality with respect to the σ -hole angle. The less numerous phosphorus and silicon contacts are generally even longer, with a large proportion of contacts being greater than the sum of vdW radii, but show larger σ -hole angles than the equivalent halogens and chalcogens. As σ -hole interactions do not just cease to interact at a separation equal to the sum of van der Waals radii, these longer interactions cannot be discounted, yet the greater tendency towards linearity with respect to the σ -hole angle, and greater polarisability of Si/P compared to S/Cl gives weight to these longer pnictogen and tetrel contacts being caused by σ -hole interactions. There are reported cases of chlorine and sulfur contacts to N/O that are deemed to be significantly strong, such as in the work summarised in Robin Taylor's R_F values, yet these cases are much less commonly occurring, and require stronger electron withdrawing substituents on the Cl/S¹⁴³. It is with these factors in mind that one might conclude that chlorine and sulfur are not typical σ -hole donors, and should not be considered in further searches of the CSD for σ -hole interactions. On a similar, yet converse note, the contacts involving the heaviest elements, Pb and Bi, show short contacts yet there are questions as to whether they can be attributed to σ -hole interactions. These contacts have little directionality with respect to the σ -hole angle, indicating that these interactions may be due to other causes. As these are large metals, and that several of the Pb structures involve hemidirecting geometries, it might suggest that these contacts could be considered from a coordination standpoint, which posits an interesting area of future study.

Chapter 5 concerns the construction of a machine learning model to rapidly predict the energies of σ -hole interactions, with the aim of applying it to structures found from searches of the CSD. The σ -Hole Approximation Tool (σ HAT) is an XGBoost regression model that uses the σ -hole angle, normalised contact distance, and a PCA-reduced extended connectivity fingerprint as descriptors. This model was trained on a 75:25 train-test split of the SHEF1552 data set of high accuracy computationally calculated σ -hole interactions. The σ HAT model performs comparably to density functional theory calculations on the test set, with an RMSD of 1.03 kcal/mol while taking negligible time to compute. To validate the performance of the σ HAT model on experimentally derived structures found within the CSD, molecular cutouts must be taken from the crystal to create a molecular complex that can be treated with DFT to calculate interaction energies. The selected validation set was the examples of halogen bonding to sp^3 -hybridised nitrogen acceptors identified in CSD searches Chapter 3. The terminal atoms of these halogen bonding cutouts were capped with hydrogen atoms to saturate valencies and their interaction energies were calculated using counterpoise corrected M06-2X(D3ZERO)/def2-TZVPD. The descriptors from these cutouts were then calculated and used with the σ HAT model to predict interaction energies. The predicted interaction energies from the σ HAT model on the validation set has an RMSD of 2.02 kcal/mol, with varying levels of performance for each of the halogen donor atoms. The model performs best on the bromine-containing structures within the validation set, whereas the iodine-containing structures indicate some potential improvements to be made to the training set. The chlorine-containing structures are very poorly described by the σ HAT model, but their weak interaction energies and lack of directionality with respect to the σ -hole angle suggest that these contacts are more likely to be caused by other packing effects than a halogen bond. This poor fit for chlorine contacts with the σ HAT model gives

further suggestion that chlorine is an ineffective σ -hole donor. This approach proves to be a suitably accurate, low-cost means of predicting σ -hole interaction energies, completing the third aim of this work.

6.2 Future Work

CSD searches have been performed for σ -hole interactions to sulfur, phosphorus, and halogens acting as acceptor moieties but are not included in this thesis for brevity. It is hoped that these results will be analysed and interpreted with the same rigour and presented in future publications. These acceptor moieties require additional processing to investigate the interplay between their Lewis basicity and ability to act as σ -hole donors. An avenue of further study could be to investigate differing valencies of σ -hole donors, as this thesis only concerns σ -hole donors at saturated valencies. Another possibility could be to probe σ -hole interactions to π -systems and anions, would provide a more comprehensive description of the nature of sigma-hole interactions, but would considerably broaden the scope of the investigation. The results of all of these searches of the CSD could be then compiled into a subset of the CSD, analogous to the MOF subset which contains porous metal organic frameworks¹⁴.

It is apparent from the machine learning work covered in Chapter 5 that some improvements could be made. In terms of validation, the σ HAT model was only tested using halogen bonding structures obtained from the CSD, but the model is intended to predict chalcogen, pnictogen, and tetrel bonds also. Building on this more cutouts from a range of σ -hole donors and acceptors should be obtained, along with filtering out contacts that are more likely attributed to other effects than σ -hole interactions, to probe the σ HAT for weaknesses. Some additional donor and acceptor moieties could be implemented into the training set to produce more strongly interacting complexes, which are rather sparse in the SHEF1552 data set. Any further descriptor-based improvements made to the σ HAT model would likely require the development of a specialised molecular fingerprint that could encode the noncovalent interaction along with the molecular entities. An alternative to this would be a much more complex architecture such as a message-passing graph neural network which would utilise the molecular graphs of the complexes directly, along with describing the noncovalent interaction²⁸³. We believe that the implementation of the σ HAT model into the CSD software package would be of significant benefit to the analysis of crystal structures within a visualisation program such as Mercury. The σ HAT model could also be adapted architecturally to be a classification model, where instead of predicting an interaction energy the model could be used to recognise if a contact found in CSD searches such as those reported in this work can be primarily attributed to a σ -hole interaction. It could enlighten the user to potentially unnoticed σ -hole interactions within the structure, and inform them of its contribution to the overall lattice energy.

Appendices

Appendix A

CSD Geometric Data

A.1 Nitrogen Heterocycle Queries

In order to determine the different trends in the geometric measurements of σ -hole interactions to cyclic sp^2 -hybridised nitrogen Lewis bases, more specialised queries are needed. Combined with the donor moieties shown in Figure 2.1, the sp^2 -hybridised nitrogen moiety shown in Figure 3.1 is adapted into various different cycles from 3-membered to 12-membered cycles. An example of this is represented in Figure A.1 for a query pertaining to 6 membered heterocycles, where the nitrogen can only be bonded to two substituents, and all R atoms in the cycle must be cyclic, the bond types (see Section 2.1.1) can be any of single, double, aromatic, or delocalised.

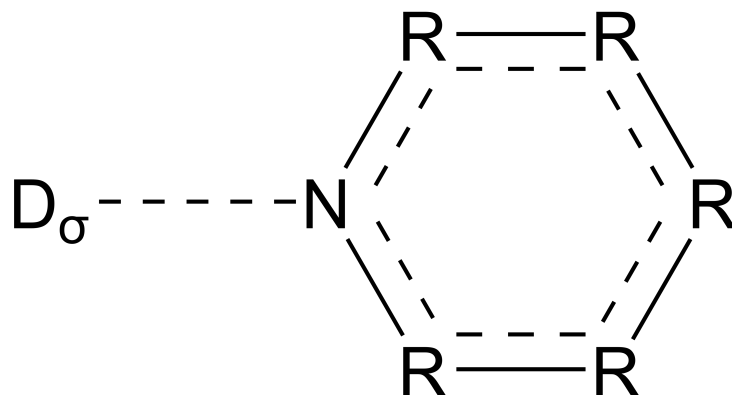


Figure A.1: A representation of the query construction for 6 membered nitrogen heterocycles where the sp^2 -hybridised nitrogen acts as a σ -hole acceptor. R = H, C, N, O, Si, P, S, F, Cl, Br, I, Se, Te, B.

A.2 Effect of Donor Electron-Withdrawing Group Continued

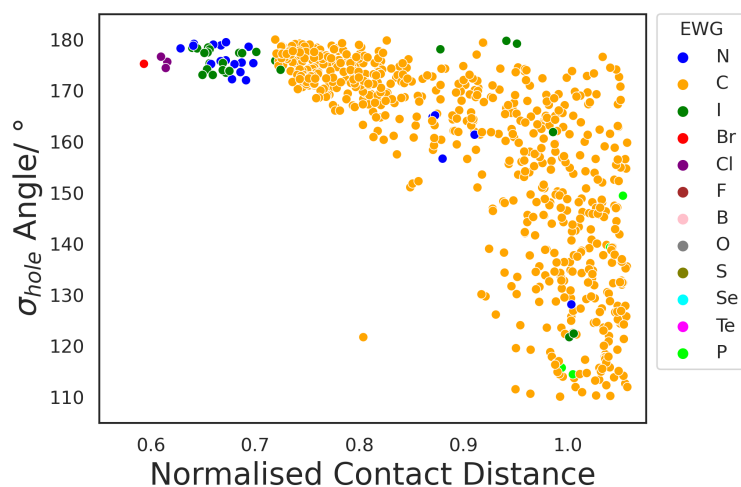


Figure A.2: Scatter plot showing the distribution of normalised contact distance with respect to the σ -hole angle of halogen bonding contacts to sp^3 -hybridised nitrogen found in crystal structures in the CSD. Points colour coded by electron-withdrawing element

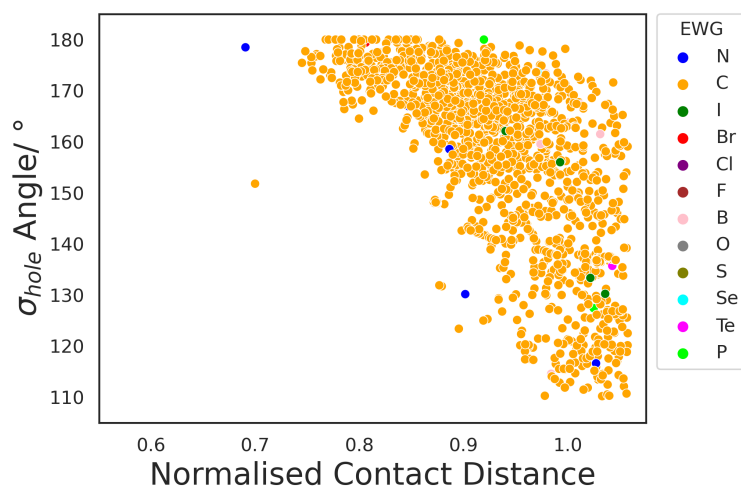


Figure A.3: Scatter plot showing the distribution of normalised contact distance with respect to the σ -hole angle of halogen bonding contacts to sp -hybridised nitrogen found in crystal structures in the CSD. Points colour coded by electron-withdrawing element

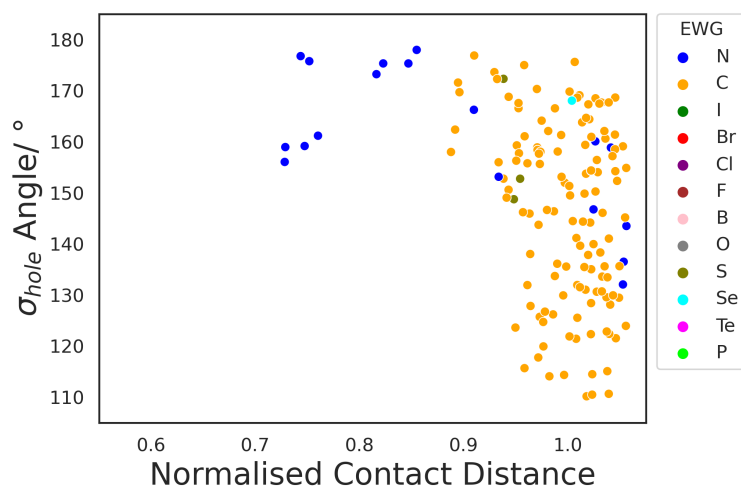


Figure A.4: Scatter plot showing the distribution of normalised contact distance with respect to the σ -hole angle of chalcogen bonding contacts to sp^3 -hybridised nitrogen found in crystal structures in the CSD. Points colour coded by electron-withdrawing element

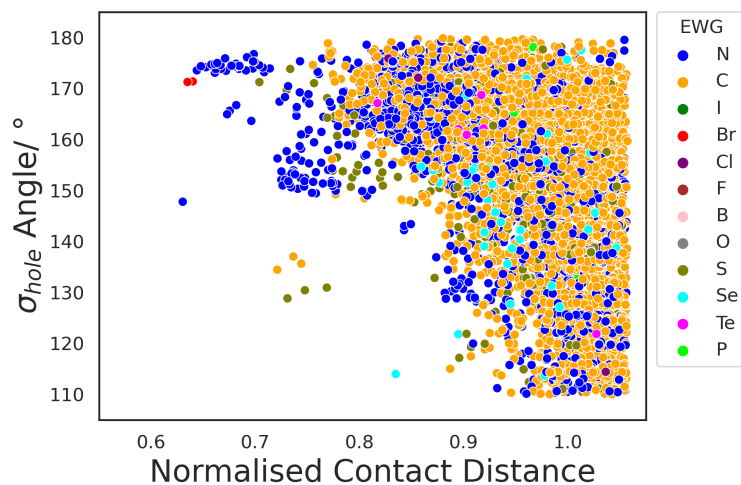


Figure A.5: Scatter plot showing the distribution of normalised contact distance with respect to the σ -hole angle of chalcogen bonding contacts to sp^2 -hybridised nitrogen found in crystal structures in the CSD. Points colour coded by electron-withdrawing element

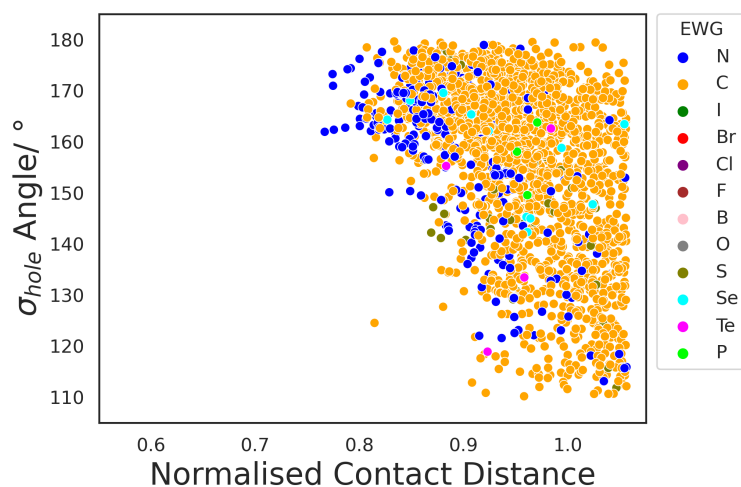


Figure A.6: Scatter plot showing the distribution of normalised contact distance with respect to the σ -hole angle of chalcogen bonding contacts to sp -hybridised nitrogen found in crystal structures in the CSD. Points colour coded by electron-withdrawing element

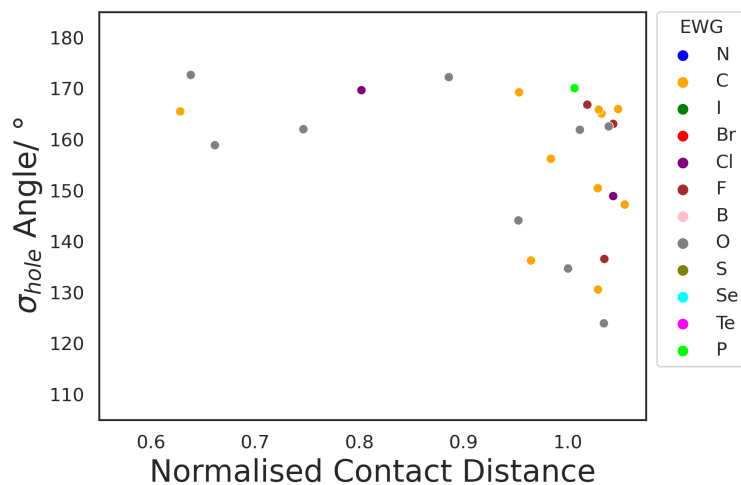


Figure A.7: Scatter plot showing the distribution of normalised contact distance with respect to the σ -hole angle of pnictogen bonding contacts to sp^3 -hybridised nitrogen found in crystal structures in the CSD. Points colour coded by electron-withdrawing element

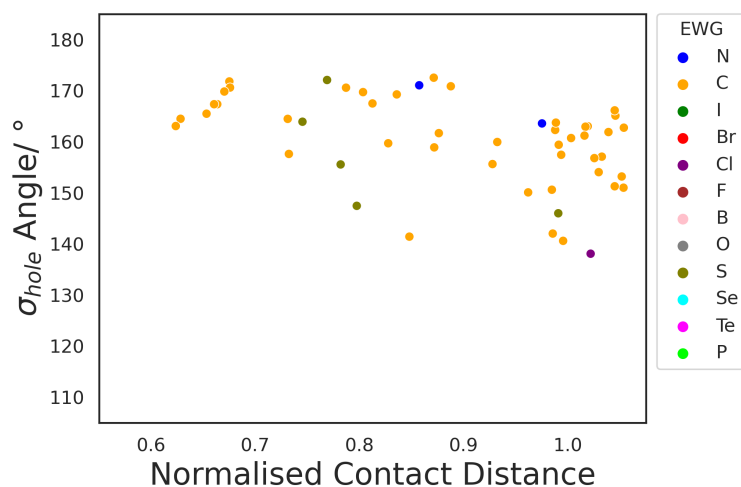


Figure A.8: Scatter plot showing the distribution of normalised contact distance with respect to the σ -hole angle of pnictogen bonding contacts to sp^2 -hybridised nitrogen found in crystal structures in the CSD. Points colour coded by electron-withdrawing element

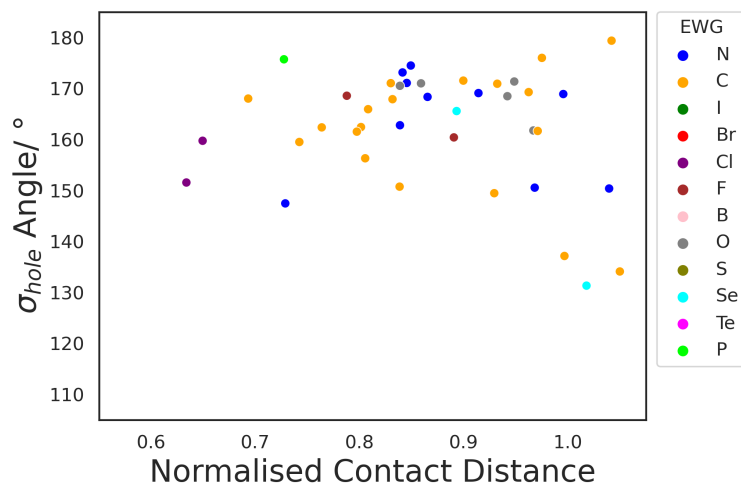


Figure A.9: Scatter plot showing the distribution of normalised contact distance with respect to the σ -hole angle of pnictogen bonding contacts to sp -hybridised nitrogen found in crystal structures in the CSD. Points colour coded by electron-withdrawing element

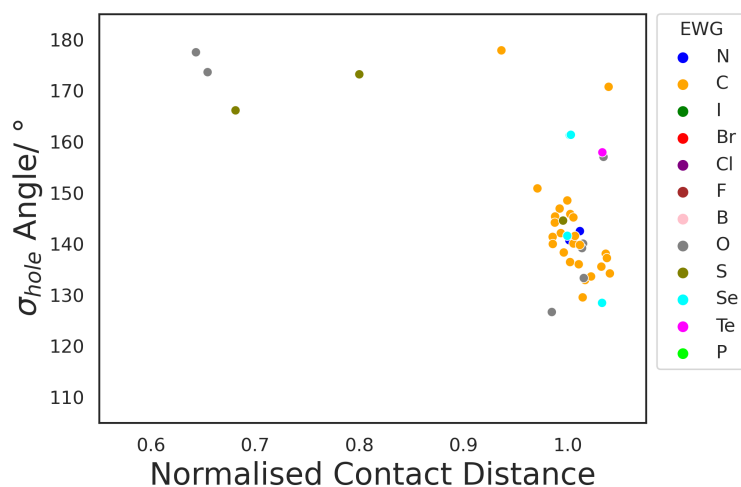


Figure A.10: Scatter plot showing the distribution of normalised contact distance with respect to the σ -hole angle of tetrel bonding contacts to sp^3 -hybridised nitrogen found in crystal structures in the CSD. Points colour coded by electron-withdrawing element

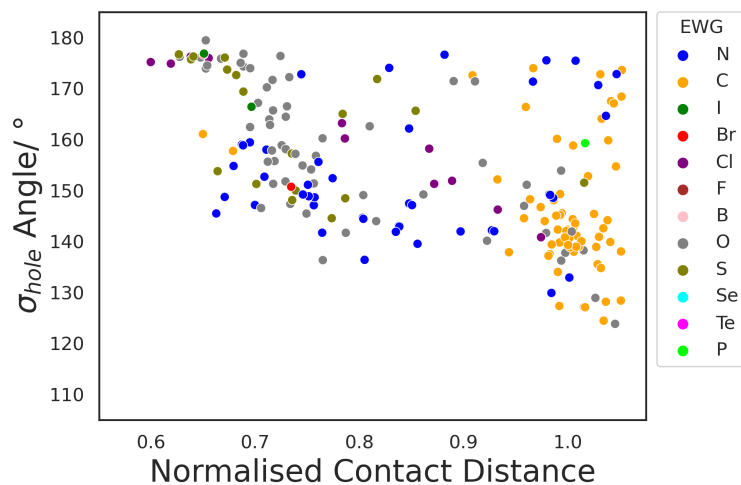


Figure A.11: Scatter plot showing the distribution of normalised contact distance with respect to the σ -hole angle of tetrel bonding contacts to sp^2 -hybridised nitrogen found in crystal structures in the CSD. Points colour coded by electron-withdrawing element

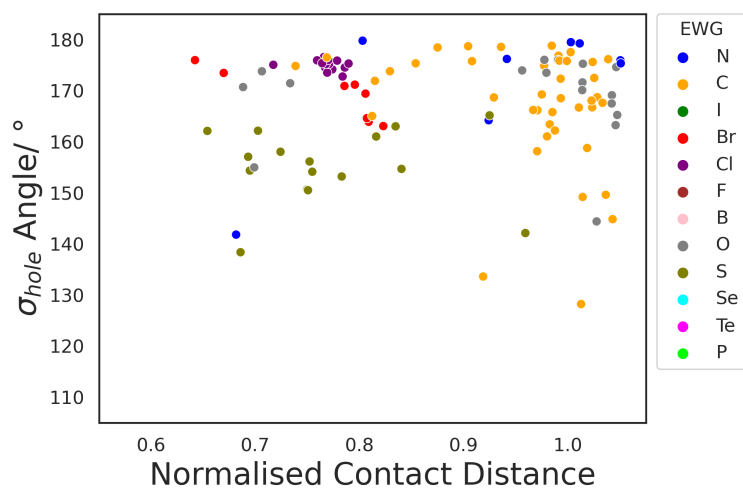


Figure A.12: Scatter plot showing the distribution of normalised contact distance with respect to the σ -hole angle of tetrel bonding contacts to sp -hybridised nitrogen found in crystal structures in the CSD. Points colour coded by electron-withdrawing element

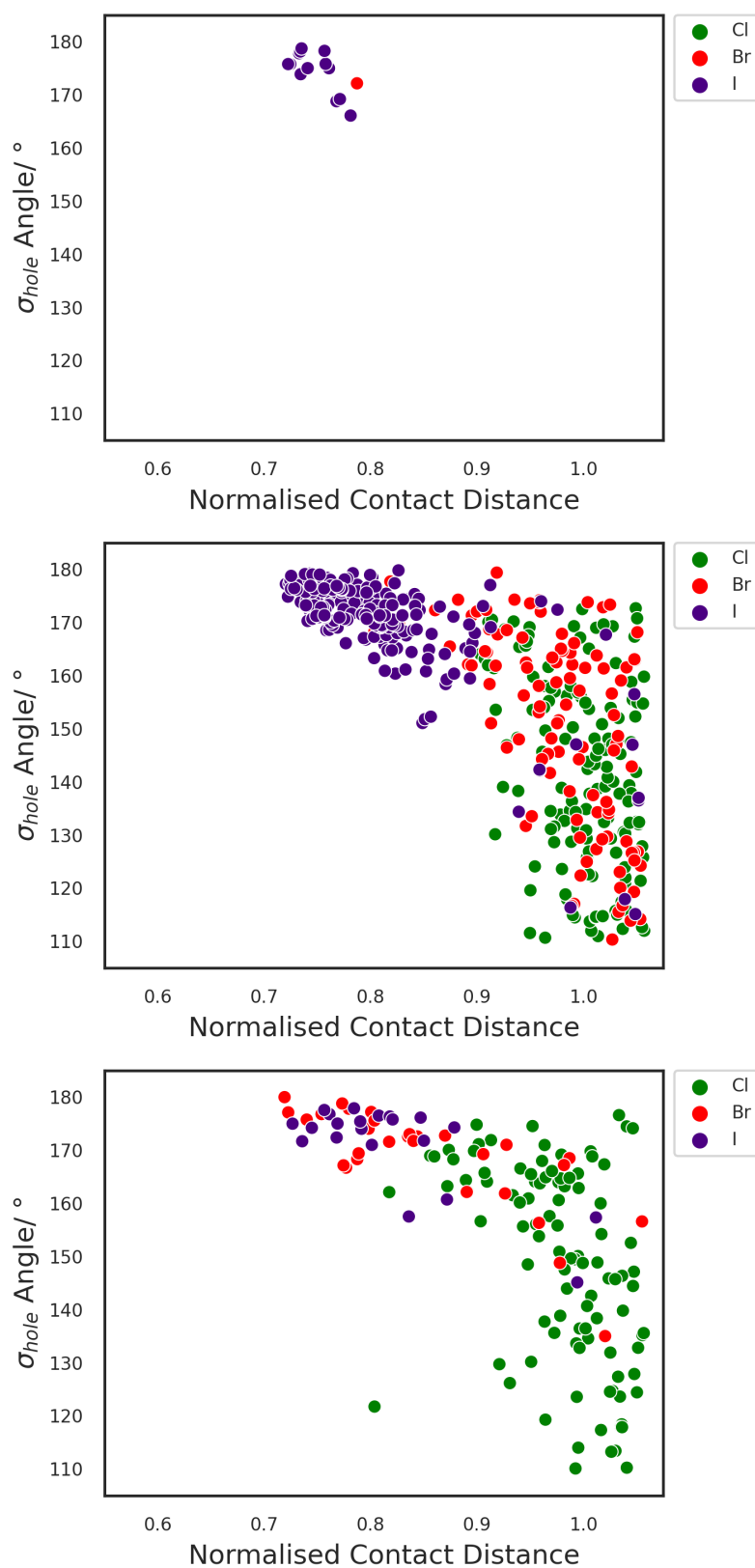


Figure A.13: Scatter plot showing the distribution of normalised contact distance with respect to the σ -hole angle of halogen bonding contacts to sp^3 -hybridised nitrogen showing how the distribution changes with halogen bond donors containing a sp -hybridised carbon EWG (*top*), a sp^2 -hybridised carbon EWG (*middle*), and a sp^3 -hybridised carbon EWG (*bottom*). Points colour coded by donor atom

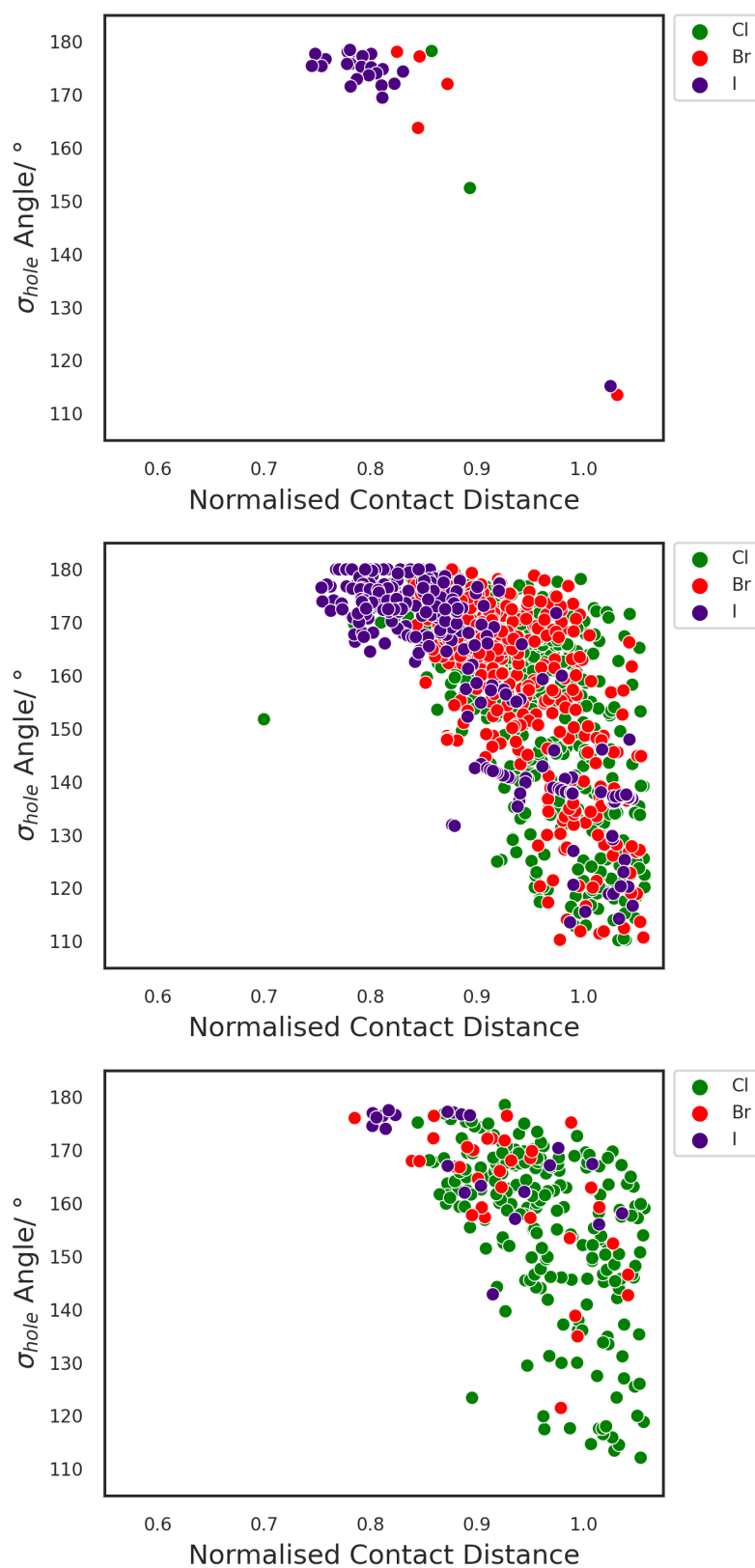


Figure A.14: Scatter plot showing the distribution of normalised contact distance with respect to the σ -hole angle of halogen bonding contacts to sp -hybridised nitrogen showing how the distribution changes with halogen bond donors containing a sp -hybridised carbon EWG (*top*), a sp^2 -hybridised carbon EWG (*middle*), and a sp^3 -hybridised carbon EWG (*bottom*). Points colour coded by donor atom

Appendix B

Crystal Structures

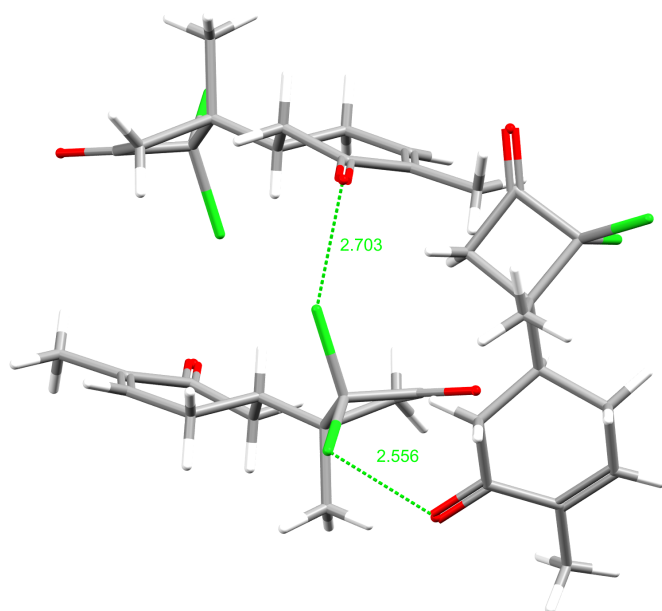


Figure B.1: An example of the crystal structure of (R,R)-(5R)-2-Methyl-5-(1-methyl-2,2-dichloro-3-oxocyclobutanyl)cyclohex-2-enone with two contacts that fall into the criteria for halogen bonding. The σ -hole angle of the Cl $\cdots$ O contact with $d=2.703$ is 150.31° and the σ -hole angle of the Cl $\cdots$ O contact with $d=2.556$ is 114.26° . Distance given in Ångström. CSD refcode: **NAFBUR**

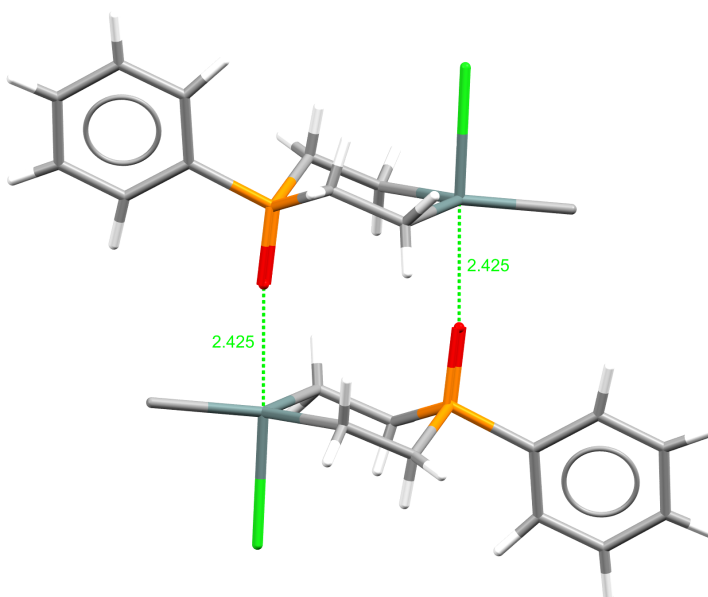


Figure B.2: An example of the crystal structure of 4-Chloro-4-methyl-1-oxo-1-phenyl-1 λ 5,4-phosphastanninane with a tetrel bonding contact forming an inversion dimer between the Sn and O. Distance given in Ångström. CSD refcode: **HAPLUF**

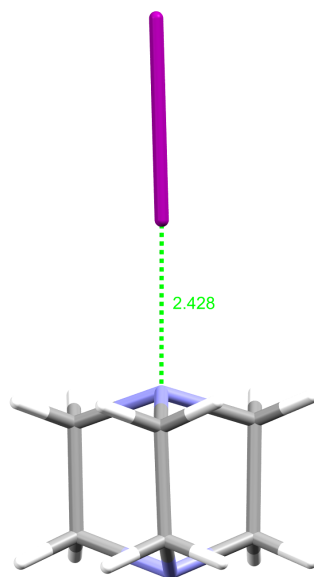


Figure B.3: An example of the crystal structure of diiodine and DABCO with a halogen bonding contact between the I and N. Distance given in Ångström. CSD refcode: **HEKZOO**

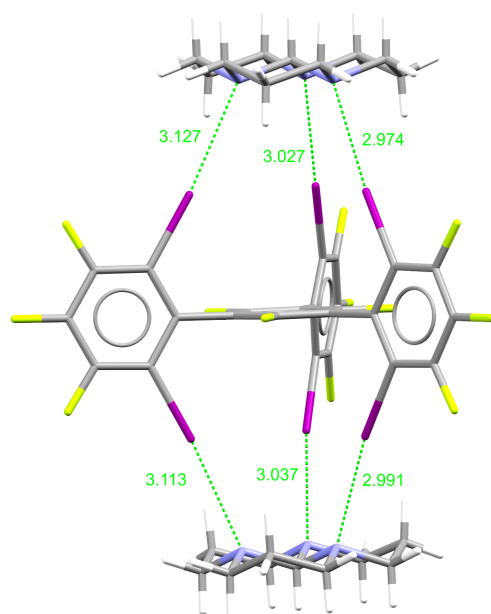


Figure B.4: An example of the co-crystal of 1,3,5-trifluoro-2,4,6-tris(3,4,5-trifluoro-2,6-diiodophenyl)benzene and bis(hexahydro-1H,4H,7H-3a,6a,9a-triazaphenalene with six I $\cdots$ N halogen bonding contacts. Distance given in Ångström. CSD refcode: **LORQUG**

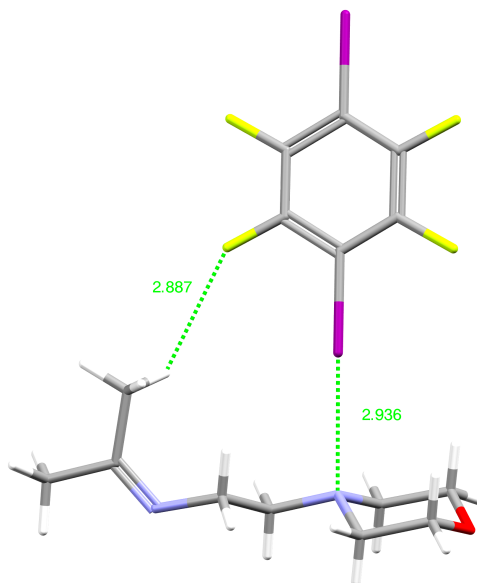


Figure B.5: An example of a molecular cutout taken of the crystal structure of N-[2-(morpholin-4-yl)ethyl]thian-4-imine 1,2,4,5-tetrafluoro-3,6-diiodobenzene which exhibits a I $\cdots$ N halogen bond, along with a potential weak H $\cdots$ F hydrogen bond. Distance given in Ångström. CSD refcode: **NERFID**

Appendix C

ML Work Appendices

C.1 Fingerprint Performance

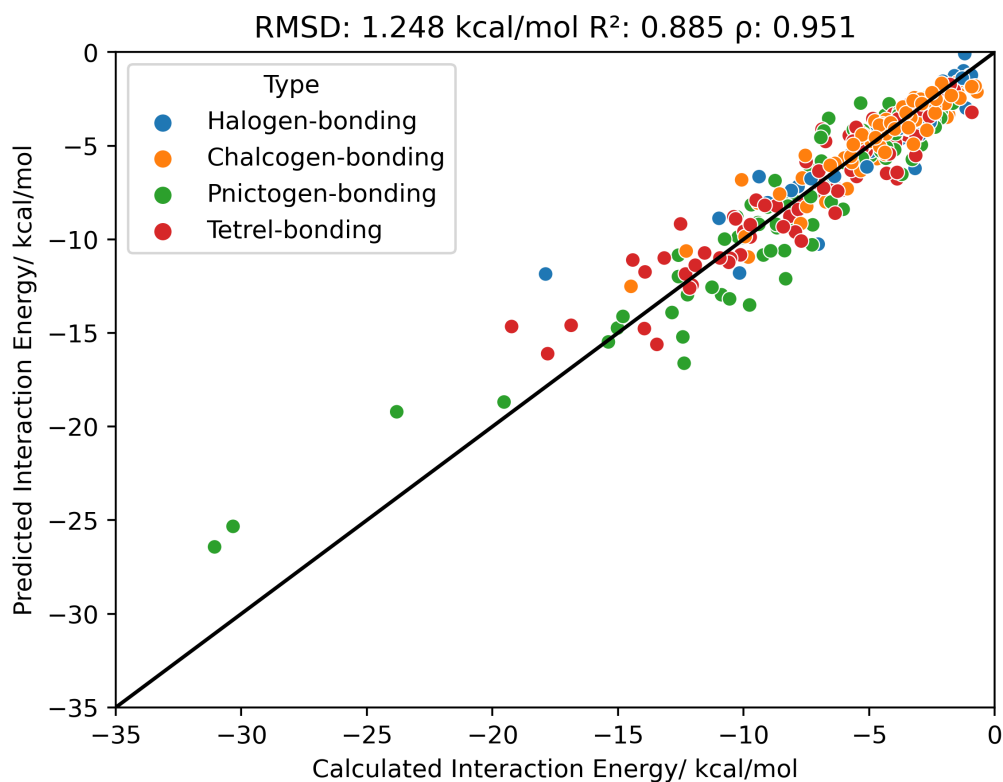


Figure C.1: Performance of PCA-treated RDKit fingerprint for predicting σ -hole interaction energies using a default XGBoost regression model. Metrics include root mean squared deviation (RMSD), coefficient of determination (R^2), and Pearson correlation coefficient (ρ). Points are colour coded by σ -hole interaction type over a line of $y = x$.

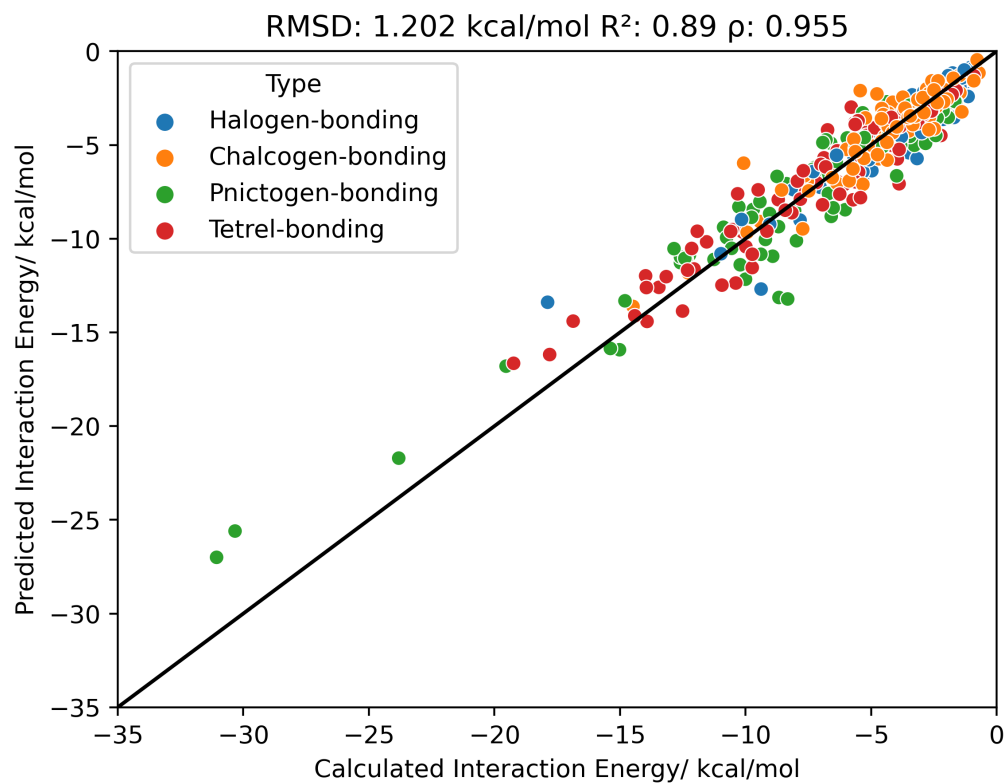


Figure C.2: Performance of PCA-treated atom pairs fingerprint for predicting σ -hole interaction energies using a default XGBoost regression model. Metrics include root mean squared deviation (RMSD), coefficient of determination (R^2), and Pearson correlation coefficient (ρ). Points are colour coded by σ -hole interaction type over a line of $y = x$.

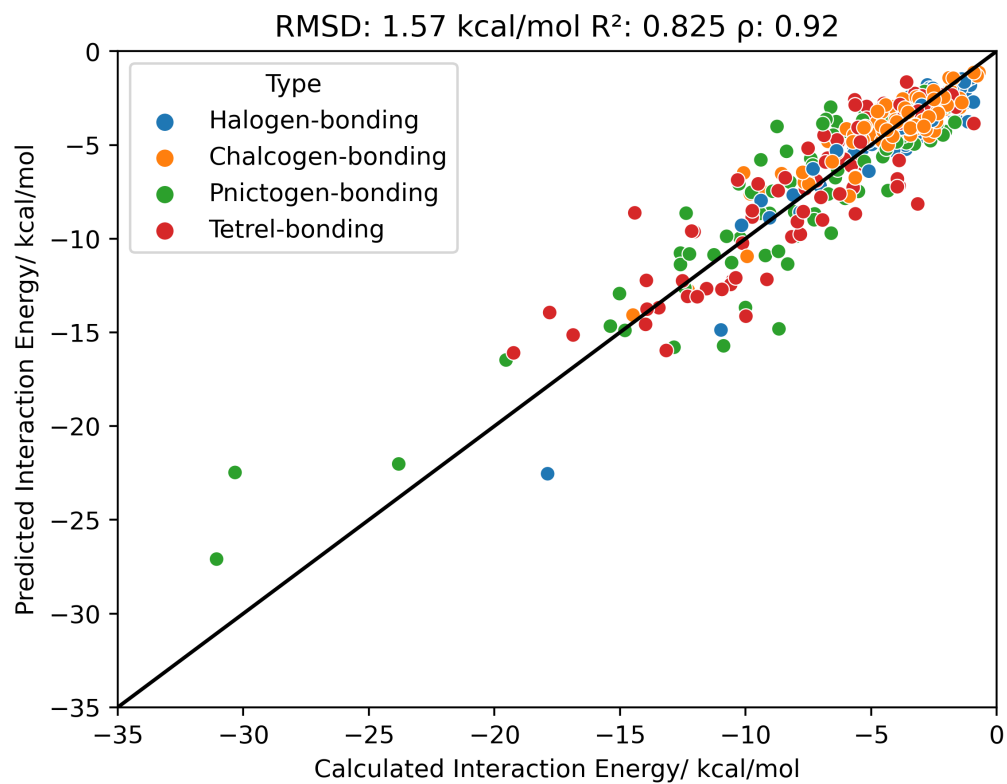


Figure C.3: Performance of PCA-treated topological torsion fingerprint for predicting σ -hole interaction energies using a default XGBoost regression model. Metrics include root mean squared deviation (RMSD), coefficient of determination (R^2), and Pearson correlation coefficient (ρ). Points are colour coded by σ -hole interaction type over a line of $y = x$.

C.2 Random Forest & Linear Regression Performance

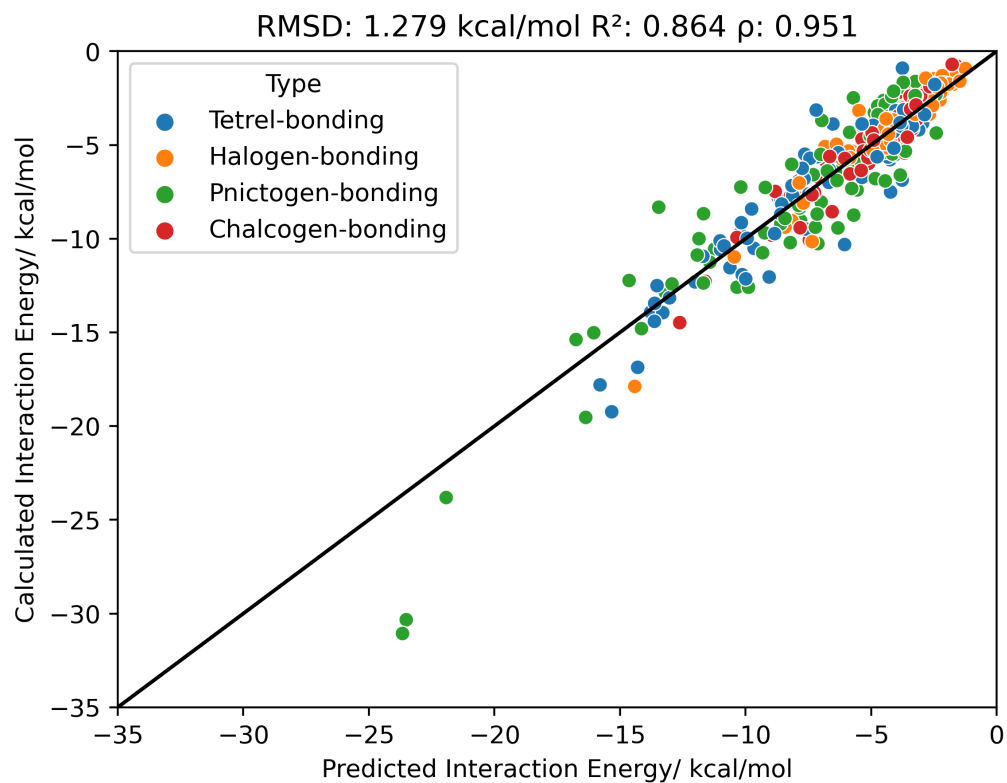


Figure C.4: Performance of a random forest regression model with PCA-treated ECFPs for predicting σ -hole interaction energies within the test set. Metrics include root mean squared deviation (RMSD), coefficient of determination (R^2), and Pearson correlation coefficient (ρ). Points are colour coded by σ -hole interaction type over a line of $y = x$.

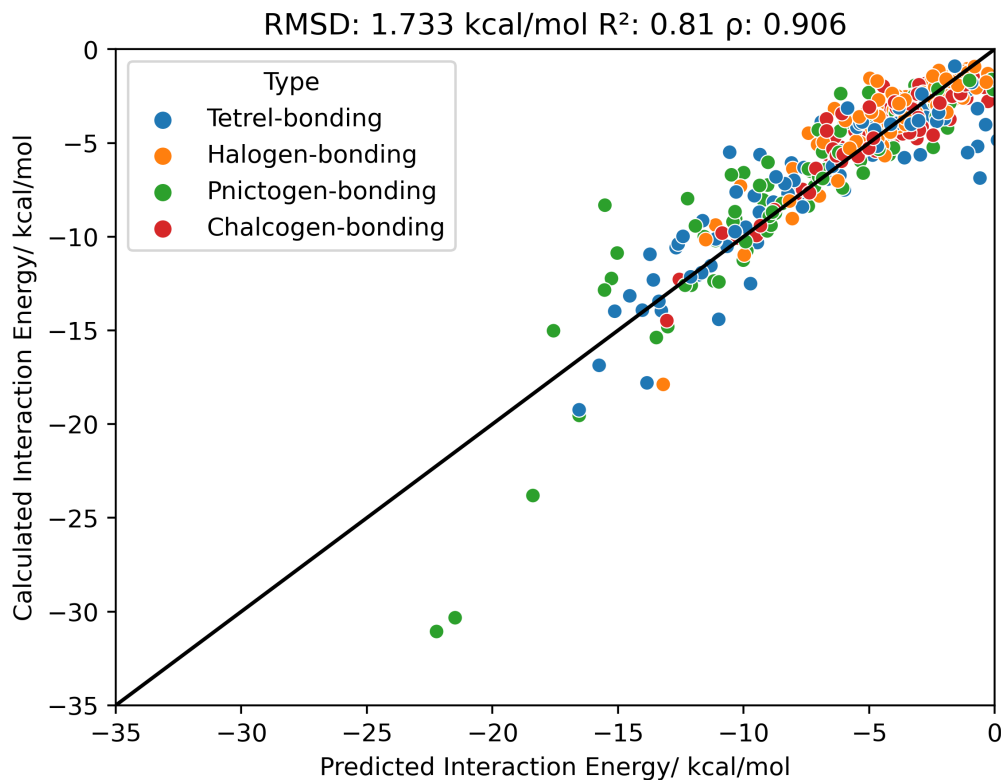


Figure C.5: Performance of a linear regression model with PCA-treated ECFPs for predicting σ -hole interaction energies within the test set. Metrics include root mean squared deviation (RMSD), coefficient of determination (R^2), and Pearson correlation coefficient (ρ). Points are colour coded by σ -hole interaction type over a line of $y = x$.

C.3 σ HAT Model Details

Listing C.1: Available hyperparameter space used with the BayesSearchCV function of scikit-learn to optimise the hyperparameters for the XGBoost regression model

```
params = {
    'learning_rate': Real(0.01, 1.0, 'log-uniform'),
    'n_estimators': Integer(50, 1000),
    'max_depth': Integer(1, 20),
    'min_child_weight': Integer(0, 50),
    'subsample': Real(0.5, 1.0),
    'colsample_bytree': Real(0.5, 1.0),
    'gamma': Real(0, 1.0, 'uniform'),
    'reg_alpha': Real(1e-5, 100, 'log-uniform'),
    'reg_lambda': Real(1e-5, 100, 'log-uniform'),
    'scale_pos_weight': Real(1e-6, 100, 'log-uniform'),
```

```

'base_score': Real(0.01, 1.0, 'uniform'),
'booster': Categorical(['gbtree', 'gblinear', 'dart']),
'n_jobs': Integer(1, 10),
'random_state': Integer(1, 1000),
'objective': Categorical(['reg:squarederror']),
'max_delta_step': Integer(0, 10),
'min_child_weight': Integer(1, 10),
'missing': Real(0, 1.0, 'uniform'),
'tree_method': Categorical(['auto', 'exact', 'approx', 'hist']),
'verbosity': Integer(0, 3)
}

```

Listing C.2: σ HAT XGBoost regression model optimised hyperparameters

```

XGBRegressor(base_score=0.8142720284737898, booster='gbtree',
colsample_bylevel=1, colsample_bynode=1,
colsample_bytree=0.799023538756064, enable_categorical=False,
gamma=0.8027647811563106, gpu_id=-1, importance_type=None,
interaction_constraints='', learning_rate=0.1111999664226395,
max_delta_step=1, max_depth=15, min_child_weight=9,
missing=0.9119288615068837, monotone_constraints='() ',
n_estimators=417, n_jobs=4, num_parallel_tree=1, predictor='auto',
random_state=611, reg_alpha=1.5275834840558614e-05,
reg_lambda=56.61143126682765,
scale_pos_weight=0.0037062124362106682, subsample=0.65704380873298,
tree_method='auto', validate_parameters=1, verbosity=2)

```

The SHEF1552 σ -hole data set: Insights into intermolecular interactions involving halogens, chalcogens, pnictogens, and tetrrels

Harry W. Nash¹ and J. Grant Hill^{1,2, a)}

¹⁾*School of Mathematical and Physical Sciences, University of Sheffield, Sheffield S3 7HF, U.K.*

²⁾*Centre for Machine Intelligence, University of Sheffield, Sheffield S10 2TN, U.K.*

The SHEF1552 data set presented here is a database of 1552 intermolecular σ -hole interactions encompassing halogen-, chalcogen-, pnictogen-, and tetrel-bonds. Each σ -hole donor molecule interacts with the same set of 19 Lewis base molecules, with explicitly correlated coupled cluster interaction energies and symmetry-adapted perturbation theory (SAPT) analysis computed for each complex. Detailed analysis of both geometric and energetic properties of the complexes allow trends to be identified. This includes that pnictogen-bonds typically have the strongest interactions, followed by tetrel-, chalcogen- and halogen-bonds. While halogen-, chalcogen- and tetrel-bonds all demonstrate a relationship where shorter intermolecular distances require σ -hole angles close to 180° , pnictogen-bonds display no clear relationship between these geometrical parameters. Using a cut-off for intermolecular interactions where the intermolecular distance must be shorter than the sum of the van der Waals radii would exclude approximately 5% of the structures in this data set, yet the majority of these longer-range interactions have an interaction energy stronger than -2 kcal mol⁻¹, indicating such cut-offs are too crude. SAPT analysis confirms that the underlying nature of the interactions is a subtle interplay of the different components. However, the strongest interactions are dominated by the electrostatic component, while the weakest can be classified as dispersion-dominated.

^{a)}Electronic mail: grant.hill@sheffield.ac.uk

I. INTRODUCTION

The concept of the σ -hole and the role it plays in the understanding of noncovalent interactions has affected the field profoundly,^{1–8} sparking huge interest and prompting many discussions on terminology and the use of simple models.^{9–14} A σ -hole is a region of lower electron density that is found along the extension of a covalent bond,^{1,2} usually resulting in a region of positive potential. This allows for simple visual analysis of molecular electrostatic potential maps to be used to rationalize the interaction between, for example, a dihalogen and ammonia, in a way that can be followed at the undergraduate level.³ An example molecular electrostatic potential for $\text{GeCN}(\text{Me})_3$ is shown in Figure 1, where the most positive region is clearly found along the extension of the cyano–Ge σ -bond. Further details on the σ -hole and electrostatic potentials in noncovalent interactions have been addressed in a large number of articles and reviews, and this information is not repeated here. Rather the interested reader is referred to Refs.^{4–8} and the references therein.

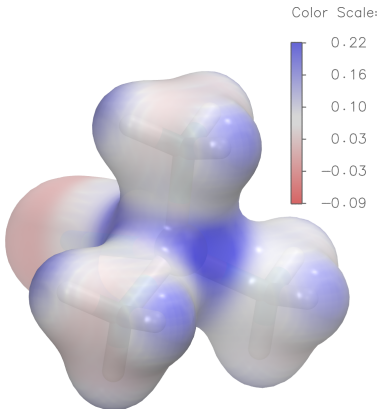


FIG. 1: Computed electrostatic potential (a.u.) on the 0.02 a.u. density isosurface of $\text{GeCN}(\text{Me})_3$. The cyano group is positioned to the rear of the molecule and the central Ge atom is visible as a sphere through the partially transparent surface. The blue disc-like area of positive potential along the extension of the cyano–Ge bond indicates the location of the σ -hole.

Perhaps the most well-established category of σ -hole interaction is the halogen bond,¹⁵ which the International Union of Pure and Applied Chemistry (IUPAC) define as: “*A halogen bond occurs when there is evidence of a net attractive interaction between an electrophilic region associated with a halogen atom in a molecular entity and a nucleophilic region in*

*another, or the same, molecular entity.*⁹ To establish a terminology to be used herein, the halogen atom can be referred to as the σ -hole donor, while the atom or region acting as a nucleophile is the σ -hole acceptor. We treat the atom/region being located within a molecular entity as being implied. Other groups of elements are known to act as donors in σ -hole interactions,¹⁶ including chalcogens,¹¹ pnictogens,¹² and tetrels,¹⁷ amongst others. We note that two recent perspectives have suggested that the naming of these interactions by their Periodic Table group and/or element is excessive, with simpler nomenclature such as σ -hole interactions preferred.^{13,14} While we agree with the principles driving such suggestions, in the present case classifying the interactions into Periodic Table groups enables trends, similarities, and divisions to be uncovered.

The IUPAC definition of the halogen bond also lists several features, including that the forces involved in the underlying nature of the interaction are “*primarily electrostatic, but polarization, charge transfer and dispersion contributions all play an important role.*”⁹ This definition also notes that the relative contribution of these terms differ between halogen bonds. As accurately describing all of these contributions is challenging for the majority of quantum chemical methods, and combining this with the strong interest in σ -hole interactions from both computational and experimental groups, it is perhaps unsurprising that a number of computational data sets have been published that provide optimized geometries and calculated interaction energies for σ -hole interactions. In the majority of cases these data sets are primarily targeted at benchmarking how well approximate computational chemistry methods describe the interactions.

To briefly introduce some of these existing data sets: The GMTKN55 benchmark database contains two relevant subsets;¹⁸ the PNICO23 set of pnictogen...pnictogen non-covalent interactions, which is based on the complexes investigated by Setiawan *et al.*,¹⁹ and the HAL59 set of noncovalent interactions involving halogenated molecules, which is based on an amalgamation of the XB51 and X40 data sets.^{20,21} We note that these PNICO23 and HAL59 sets include a variety of intermolecular interactions, not just those typically classified as σ -hole interactions. Bauzá *et al.* have proposed a set of 30 halogen, chalcogen and pnictogen bonding complexes with both neutral and anionic σ -hole acceptors, where the donors are typically highly fluorinated.²² Alkorta and Legon have also used a set of 132 intermolecular complexes to define a range of “reduced electrophilicities” of Lewis acids,²³ although these complexes included interactions involving alkali-metals and coinage-metals,

in addition to hydrogen bonds and halogen bonds. Meanwhile, the CHAL336 data set is focused on chalcogen bonding interactions and includes examples with either σ -holes or the somewhat related π -holes.²⁴ We note that the computational chemistry method benchmarking exercise conducted on CHAL336 concluded that the popular B3LYP and MP2 methods were not recommended for this type of interaction.

More recently, the SH250 $\times$ 10 data set of σ -hole interactions has been added to the wider Non-Covalent Interactions Atlas project.²⁵ This data set contains 250 complexes bound by halogen, chalcogen or pnictogen bonds, each providing a series of ten points along the respective dissociation curve. The σ -hole donor atoms are taken from the third and fourth period of the p -block of the Periodic Table, plus iodine, and combined with a variety of acceptors. It is important to note that the set of acceptors used across the halogen, chalcogen and pnictogen bonding complexes is not consistent, meaning that direct comparisons cannot always be drawn. The data set has been used to benchmark the performance of a large number of DFT functionals, however, it has demonstrated that modern semi-empirical quantum chemistry methods require re-parameterization to correctly describe the whole breadth of σ -hole interactions. The importance of reusing quantum chemistry benchmark sets, including those dedicated to intermolecular interactions, in the development of force fields for molecular simulations has also been recently highlighted.²⁶

Here we present the Sigma-Hole Energies and additional Factors data set of 1552 intermolecular complexes, denoted SHEF1552, which encompasses halogen, chalcogen, pnictogen and tetrel bonding. There are two major foci of the data set:

1. Discovery of insights into σ -hole interactions in terms of geometric and energetic properties, and the underlying nature of the interaction.
2. Relevance to both theoretical and experimental chemists/material scientists.

These points inform the choice of complexes investigated and are prioritized over, for example, the benchmarking of quantum chemical methods. The same set of acceptor molecules are paired with each donor molecule so that comparisons can be drawn between classes of interaction (halogen bond, chalcogen bond, etc.) and within each of those classes. The lightest elements from each class, that is, those from the second period of the Periodic Table, are not included as σ -hole donor atoms as they typically only form very weak interactions. In terms of the heavy elements from the sixth period, we only consider complexes where Pb or Bi act

as the donor atom. Both Po and At are highly radioactive and, while there may be some interest in the hypothetical complexes they may form, they are excluded here as they are less relevant to experimental studies. Similarly, we avoid adding large numbers of fluorine substituents to the donor molecules. It is well-known that these strong electron withdrawing groups enhance σ -hole interactions, but synthesis of such highly fluorinated compounds for experimental studies poses substantial experimental challenges and such species are less representative of those involved in interactions found in, for example, the Cambridge Structural Database (CSD). A more detailed specification of the complexes included in the data set is presented in Section II A and in the Supporting Information. Section II B details the computational methods used to determine the interaction energies of the complexes, while Section III presents an analysis of the results of the calculations on the complexes in the data set.

II. METHODS

A. Data set construction

With the aim of ensuring the data set is relevant to both experimental chemistry and computational studies, the choice of σ -hole donor molecules was partly informed by our experience of CSD searches for σ -hole interactions. Small neutral molecules that represent commonly-found motifs within the CSD searches were selected, with each of seven halogen-bonding donor motifs appearing three times (one each for Cl, Br and I), each of seven chalcogen-bonding donor motifs three times (S, Se and Te), each of five pnictogen-bonding donor motifs four times (P, As, Sb and Bi), and each of five tetrel-bonding donor motifs four times (Si, Ge, Sn and Pb). This resulted in a total of 82 σ -hole donor molecules (21 halogen-bonding, 21 chalcogen-bonding, 20 pnictogen-bonding, and 20 tetrel-bonding), with a full list of the molecules presented in Tables S1-S4 in the Supporting Information. All of these σ -hole donor molecules are then paired with a Lewis base / σ -hole acceptor, such that each donor is paired with each possible acceptor. The 19 Lewis bases selected are: ammonia; phosphine; hydrogen cyanide; acetone; thioacetone; trimethylamine; aniline; furan; pyridine; dioxane; water; acetonitrile; thiophene; imidazole; trimethylphosphine; trimethylphosphine sulfide; trimethylphosphine oxide; nitromethane; pyridine-N-oxide. Permuting over all of

these possibilities results in a total of 1558 complexes, comprised of 399 halogen bonding complexes, 399 chalcogen, 380 pnictogen, and 380 tetrel. After analysis of the optimized geometries, six of the complexes were removed as their structures did not appear to be consistent with σ -hole interactions. Further details are provided in Section III.

Each σ -hole donor molecule was constructed such that the extension of the σ -hole is aligned along the z-axis in the negative direction, while each acceptor/Lewis base was similarly constructed with the extension of the lone pair along the z-axis, but in the positive direction. The individual donor and acceptor geometries can then be concatenated such that the σ -hole and lone pair are aligned along the same axis. Although it is conceivable that this may not lead to the global minimum energy geometry for the resulting complex, it does meet the present objective of locating a σ -hole interaction. The initial separation of the σ -hole donor atom and the acceptor atom was set to 85% of the sum of the van der Waals radii of these interacting atoms.

The geometry of each σ -hole complex was subsequently optimized using the M06-2X exchange correlation functional with the D3 dispersion correction [zero damping, D3(0)],^{27,28} and the def2-TZVPD basis set.^{29,30} This polarized triple-zeta basis includes additional diffuse functions and small-core, relativistic effective core-potentials (ECPs) that replace 28 electrons for Sn, Sb, Te, and I, and 60 electrons for Pb and Bi.^{31,32} All geometry optimizations were carried out in ORCA version 5.0.4,³³ with tight SCF convergence criteria and the defgrid3 integration grid.³³ Density fitting was used in the evaluation of the Coulomb integrals, using the ‘universal’ def2 auxiliary Coulomb basis.³⁴ The M06-2X-D3(0) functional ranked as the third best exchange correlation functional for intermolecular interactions in the benchmarking of the GMTKN55 database.¹⁸ It also avoids the computationally expensive nonlocal van der Waals correction of the highest ranking functional. The M06-2X-D3(0) functional performed well for the SH250 data set,²⁵ where the root-mean-square error in the interaction energy relative to high-accuracy *ab initio* data was close to the best performing double hybrid functionals, and also in a test of various methods for computing the strength of halogen bond interactions.³⁵

The nature of each stationary point located was ascertained by calculating the harmonic vibrational frequencies at the same level of theory. If any imaginary frequencies were found, the structure was distorted along the associated normal mode(s) using a Python script that makes use of the cclib (version 1.8.1) library to parse ORCA output.^{36,37} The resulting

structures were then re-optimized. The optimized geometries are provided as a compressed archive of XYZ format files. After several cycles of geometry optimization, frequency analysis and normal-mode distortion, 30 structures still possessed low frequency imaginary modes, generally associated with propeller-like rotation of, for example, methyl groups not directly involved in the intermolecular interaction. As such, these structures were retained in the database and are listed in Table S5.

Two of the most important geometric properties are the intermolecular distance between the donor and acceptor atoms, and the angle between the two molecules within the complex. Herein, the σ -hole angle, θ , is defined as the angle formed by R-D $\cdots$ A, where D is the σ -hole donor atom, R forms a σ -bond to D such that the σ -hole is found along the extension of this bond, and A is the σ -hole acceptor atom. This angle is illustrated with the example of the NC(Me)₃Sn $\cdots$ NH₃ tetrel-bonding complex in Figure 2. The deviation from linearity of a given σ -hole interaction could be expressed as $(180 - \theta)^\circ$.

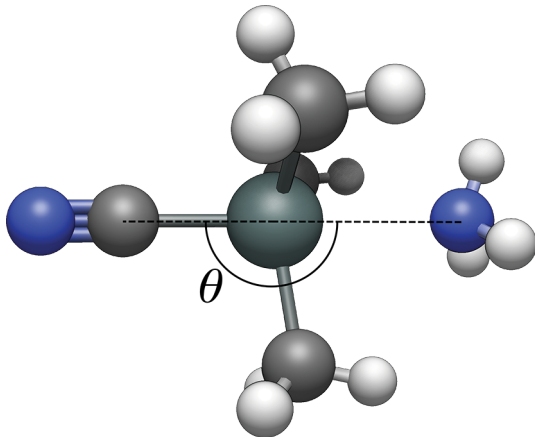


FIG. 2: Definition of the σ -hole angle (θ) of R-D $\cdots$ A, where D is the σ -hole donor atom and A is the acceptor atom. The complex shown is NC(Me)₃Sn $\cdots$ NH₃.

Normalized contact distances are often used in the analysis of interactions within crystal structures as they account for the inherent size of the atoms involved in terms of their van der Waals radii,^{38,39} allowing better comparison of geometries across interactions with different atom types. In the present work we use an analogous dimensionless, normalized distance, d_{norm} , calculated as:

$$d_{\text{norm}} = \frac{d}{r_{\text{D}}^{\text{vdW}} + r_{\text{A}}^{\text{vdW}}}$$

where d is the internuclear distance between the σ -hole donor and acceptors atoms, and $r_{\text{D}}^{\text{vdW}}$

and r_A^{vdW} are the van der Waals radii of the donor and acceptor atoms, respectively. The radii tabulated by Alvarez are used throughout.⁴⁰ Both the σ -hole angle and d_{norm} , along with interaction energies and other useful information, are provided for all complexes within a comma separated values (CSV) file as part of the Supporting Information.

B. Interaction energies

No attempt was made to incorporate the effects of spin-orbit coupling into the calculation of interaction energies or within geometry optimizations. Although “correcting” the interaction energies may be possible by using some form of composite computational scheme, accounting for this within the geometries of all complexes contained within this data set, at a suitable level of theory, is far beyond current, pragmatic computational power. This will affect the results of complexes containing heavier elements, although a recent study of chalcogen bonding suggests this would lead to a relatively small, though non-trivial, reduction in the strength of interaction.⁴¹

1. *Density functional theory*

Interaction energies at the optimized geometries were determined at the M06-2X-D3(0) level using the same basis sets, settings and software as for the geometry optimizations. In this supramolecular approach to interaction energies, the individual molecules were fixed in their interacting geometries and the counterpoise (CP) correction was used to account for basis set superposition error.⁴²

2. *Coupled cluster theory*

The explicitly correlated local coupled cluster methodology was also used to produce interaction energies at the DFT optimized geometries. More specifically, the PNO-LCCSD(T)-F12b method that uses pair natural orbitals was used,^{43–45} where the local correlation approach addresses the steep scaling of coupled cluster methods, and explicit correlation (F12) both accelerates convergence with respect to basis set size and partly corrects the errors introduced by the domain approximation in local correlation.⁴⁶ The use of approximation b in these F12 calculations is implicit herein. For elements from the first two rows of the

Periodic Table, the aug-cc-pVDZ-F12 basis set was used,^{47,48} while cc-pVDZ-F12 was used for hydrogen to prevent the basis set becoming linearly dependent. The aug-cc-pVDZ-PP-F12 basis set was used for the heavier p -block elements,^{49,50} with the ECP10MDF ECP for 4 p elements, ECP28MDF for 5 p , and ECP60MDF for 6 p .^{31,32,51,52} Herein, this mixed basis set shall be referred to as aVDZ-F12. It is notable that although this PNO-LCCSD(T)-F12/aVDZ-F12 level of theory only employs a double-zeta basis set, for the S66 benchmark set of Řezáč, Riley and Hobza,^{53,54} it produces interaction energies with a root-mean-square deviation (RMSD) of only 0.05 kcal mol⁻¹ and a maximum deviation of 0.11 kcal mol⁻¹,⁵⁵ making it an accurate and efficient choice for the relatively large number of complexes in the present investigation. All coupled cluster calculations were carried out in the Molpro package of *ab initio* programs, version 2023.2.^{56,57}

F12 calculations require three different types of auxiliary basis set (ABS) for integral evaluations. In the following, the Molpro notation for ABSs is used, where text after a forward slash denotes the context of the type of integrals approximated. The density-fitted Hartree-Fock calculations used the aug-cc-pV(T+d)Z/JKFit ABS for lighter elements,^{58,59} and def2-AQZVPP/JKFit for 4 p , 5 p , and 6 p elements.⁶⁰ In both cases, the exponents of the additional diffuse functions were determined by an even-tempered extension. The complementary auxiliary basis set (CABS) singles correction was added to the Hartree-Fock energies,⁶¹ and this was evaluated using the cc-pVDZ-F12+/OptRI ABS for lighter elements,⁶² and aug-cc-pVDZ-PP-F12/OptRI for the heavier p -block.⁵⁰ These CABS singles corrections were evaluated without local approximations. All of the remaining two-electron integrals in the PNO-LCCSD(T)-F12 calculations used the aug-cc-pVDZ(-PP)-F12/MP2Fit ABSs,^{50,63} with local density fitting.^{45,64} For the lighter elements, the aug-cc-pVDZ-F12/MP2Fit sets were produced by even-tempered extension of the cc-pVDZ-F12/MP2Fit sets with additional diffuse functions, and further details are provided as Supporting Information, along with the resulting exponents. The resolution-of-the-identity (RI) ABS used for the many-electron integrals that arise in F12 theory was the union of the orbital basis with aug-cc-pV(T+d)Z/JKFit for the lighter elements and def2-AQZVPP/JKFit for the heavier p -block. For ease of reference, Table S6 details which basis sets, orbital and auxiliary, are used for each element.

As the perturbative triples in PNO-LCCSD(T)-F12 are not directly affected by the explicit correlation, their convergence with respect to basis set size can be significantly slower than the singles and doubles. To address this, the triples contribution was scaled by the

ratio between the MP2-F12 correlation energy and the conventional MP2 correlation energy, which is commonly known as the (T*) energy.⁶⁵ To retain size consistency, the scaling factor was determined for each dimer complex and then held fixed for the respective monomers.⁶⁶ Tight domains, as defined by Ma and Werner,⁴⁶ are used in the PNO-LCCSD(T)-F12 calculations, along with the 3*A(FIX) *Ansatz* and F12b approximation.^{67,68}

In the context of halogen and chalcogen bonding, it has been shown that $(n - 1)d$ subvalence correlation is important for the calculation of accurate interaction energies with the heavier p-block elements.^{41,69} Herein, all coupled cluster calculations include the $(n - 1)d$ electrons in the correlation treatment, and the aug-cc-pVDZ-PP-F12 basis set does include functions specifically designed for this form of correlation. To provide a balance between valence-valence, core-valence and core-core correlation, pair specific geminal exponents are used in the F12 treatment for complexes containing the heavy p-block elements.⁷⁰ A geminal exponent of $1.0 a_0^{-1}$ was used for the valence-valence pairs, and $1.4 a_0^{-1}$ for all other pairs.⁷¹

Interaction energies at the following levels of theory, both with and without CP correction, are extracted from the explicitly correlated local coupled cluster calculations: Hartree-Fock with the CABS singles correction (HF+CABS), LMP2-F12, PNO-LCCSD(T)-F12b [further abbreviated as LCCSD(T)-F12], and PNO-LCCSD(T*)-F12b [LCCSD(T*)-F12].

3. *Symmetry-adapted perturbation theory*

Additional insights into the underlying nature of intermolecular interactions can be obtained through symmetry-adapted perturbation theory (SAPT).⁷² SAPT2 calculations were carried out using the PSI4 program package version 1.8.2,⁷³ with the def2-TZVPD basis set for all heavy atoms, and def2-TZVP for hydrogens. For the complexes $C_6H_5I \cdots NCCH_3$ and $Me_2N(Me)_2Sb \cdots$ dioxane the diffuse functions caused SCF convergence problems related to linear dependency, hence the def2-TZVP basis was used for all atoms for these two complexes. Density fitting in the HF parts of the calculation use the def2-QZVPP/JKFit ABS,⁶⁰ while the density fitting of the intramonomer correlation corrections use the def2-TZVP(D)/MP2Fit ABS.⁷⁴⁻⁷⁷ The individual SAPT terms are grouped into an electrostatic component (E_{elec}), an exchange component (E_{exch}), an induction component (E_{ind}), and a dispersion component (E_{disp}) as detailed in the so-called “chemist’s grouping” of Hohenstein and Sherrill.⁷⁸

III. RESULTS AND DISCUSSION

A. Geometrical Parameters

Analysis of the optimized geometries revealed that five structures had σ -hole angles that were surprisingly small ($< 150^\circ$), and one structure had an exceptionally long normalized distance ($d_{\text{norm}} = 1.21$). Visual inspection of the geometries indicated that these complexes were not interacting via a σ -hole and, after attempts to re-optimize the geometries led to the same minima, these complexes were removed from the data set. Further information on the composition and geometrical parameters of these complexes can be found in Table S7. An additional two complexes were found to have repulsive interaction energies at the CP-M06-2X-D3(0)/def2-TZVPD level of theory; the $\text{CH}_3\text{Cl} \cdots \text{NCH}$ interaction energy is $+0.21 \text{ kcal mol}^{-1}$ and $\text{CH}_3\text{Cl} \cdots \text{O}_2\text{NMe}$ is $+0.16 \text{ kcal mol}^{-1}$. At the CP-LCCSD(T*)-F12 level, the $\text{CH}_3\text{Cl} \cdots \text{O}_2\text{NMe}$ interaction becomes very weakly attractive at $-0.14 \text{ kcal mol}^{-1}$, while $\text{CH}_3\text{Cl} \cdots \text{NCH}$ remains slightly repulsive at $+0.09 \text{ kcal mol}^{-1}$. For the latter case, constrained potential energy scans along the internuclear distance coordinate indicate that there is some form of meta-stable minimum at the geometry presented within the data set. As the data set should contain examples of σ -hole interactions that span as full a range of interaction strengths as practicable, the decision was taken to retain both of these complexes.

Figure 3 shows d_{norm} and the σ -hole angle of all complexes in the data set, along with kernel density estimate plots of the distributions grouped by interaction type. To provide numerical detail on these distributions, Table I presents a statistical summary of d_{norm} and the σ -hole angles as the mean, standard deviation (S.D.), minimum value (Min), maximum value (Max), first quartile (Q1), and third quartile (Q3). In terms of general trends in normalized distances, pnictogen interactions are slightly shorter than tetrel interactions, while halogen and chalcogen interactions have similar average distances that are slightly longer still. Tetrel and pnictogen interactions have broader distributions of d_{norm} , with some bimodal character visible, but from the Min. and Max. normalized distances in Table I it can be seen that all four classes of interaction span a relatively wide and comparable range of values. In terms of σ -hole angle, tetrel- and halogen-bond interactions tend to be the closest to linearity, while chalcogen-bond interactions clearly display a broader distribution of angles. Pnictogen-bond interactions show a more clearly defined peak than those of

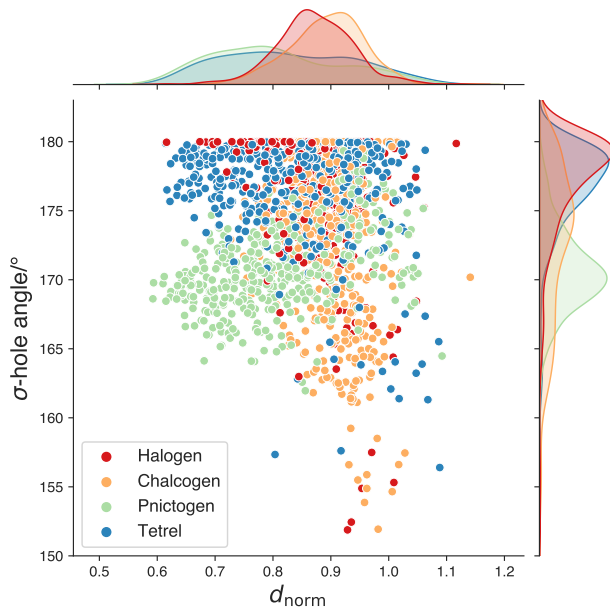


FIG. 3: Normalized distance (d_{norm}) and σ -hole angle from M06-2X-D3(0)/def2-TZVPD optimized geometries for the 1552 complexes in the data set. Also shown are kernel density estimate plots of the distribution of the normalized distances and σ -hole angles grouped by interaction type.

chalcogen-bonds, but with an average value of 170.6° there is a greater deviation from linearity than for halogen- or tetrel-bonds.

Although the geometries in this SHEF1552 data set are computational gas-phase structures, it is also possible to explore how representative they are of σ -hole interactions found more widely. To address this, searches of the Crystal Structure Database [CSD, database version 5.46 (November 2024)]^{79–81} from the Cambridge Crystallographic Data Centre have been carried out. The queries were constructed to find σ -hole interactions to so-called “divalent oxygens”, which take the form $\text{R}-\text{O}-\text{R}'$ and describe a wide range of potential σ -hole acceptors, including ethers, alcohols, water, furans, esters and so on. A close contact is defined in the query as any distance between the σ -hole donor atom and the oxygen atom that is less than the sum of the van der Waals radii plus $0.2 \text{ \AA}$. Further details of the CSD search construction are provided in the Supporting Information. The search results in a total of 14,264 “hits”, which are visualized in terms of d_{norm} and σ -hole angle in Figure S2. Figure S2(a) shows that these experimental structures have a wide range of σ -hole angles, and that the majority occur at large values of d_{norm} , suggesting some of these structures

TABLE I: Summary statistics of the normalized distances (d_{norm}) and σ -hole angles (in degrees) from M06-2X-D3(0)/def2-TZVPD optimized geometries grouped by interaction type.

d_{norm}	Tetrel	Pnictogen	Chalcogen	Halogen
Mean	0.8277	0.8049	0.8923	0.8718
S.D.	0.1129	0.1114	0.0632	0.0650
Min	0.6145	0.5925	0.7120	0.6161
Max	1.0882	1.0922	1.1409	1.1167
Q1	0.7337	0.7163	0.8556	0.8337
Q3	0.9212	0.8797	0.9355	0.9120
σ -hole angle	Tetrel	Pnictogen	Chalcogen	Halogen
Mean	176.4	170.6	172.2	176.9
S.D.	3.8	2.9	5.9	4.3
Min	156.4	162.0	151.9	151.9
Max	180.0	179.3	180.0	180.0
Q1	175.5	168.7	168.1	175.7
Q3	178.9	172.2	176.5	179.9

are likely to have the required atoms within relatively close proximity, but not necessarily engaged in a σ -hole interaction. To aid comparison of the SHEF1552 structures with those from the CSD search, we subsequently apply a relatively aggressive filter that retains only hits with a σ -hole angle greater than or equal to 150° , and those with a d_{norm} of less than or equal to 0.93, which results in 1,541 structures from the CSD. A scatterplot of d_{norm} and σ -hole angle for this revised set of CSD hits is also provided as Figure S2(b).

Although constraining the CSD search to hits with $d_{\text{norm}} \leq 0.93$ undoubtedly discards a number of σ -hole interactions, it is also clear that it removes many structures with large normalized distances between the atoms, and helps to remove many structures where the equivalent of the σ -hole angle is far from linearity. That the number of structures is similar for both this set of experimental CSD structures and SHEF1552 also facilitates the comparison between them. Figures 4 and 5 plot the distributions of d_{norm} and σ -hole angles,

respectively, as ridge plots with a separate “ridge” for each σ -hole interaction type. Each Figure includes a set of ridge plots for both the SHEF1552 data set and the filtered CSD search results, and it should be noted that because the CSD hits contain significantly more halogen- and chalcogen-bond interactions than pnictogen- and tetrel-bond, the SHEF1552 and CSD plots do not use the same normalization.

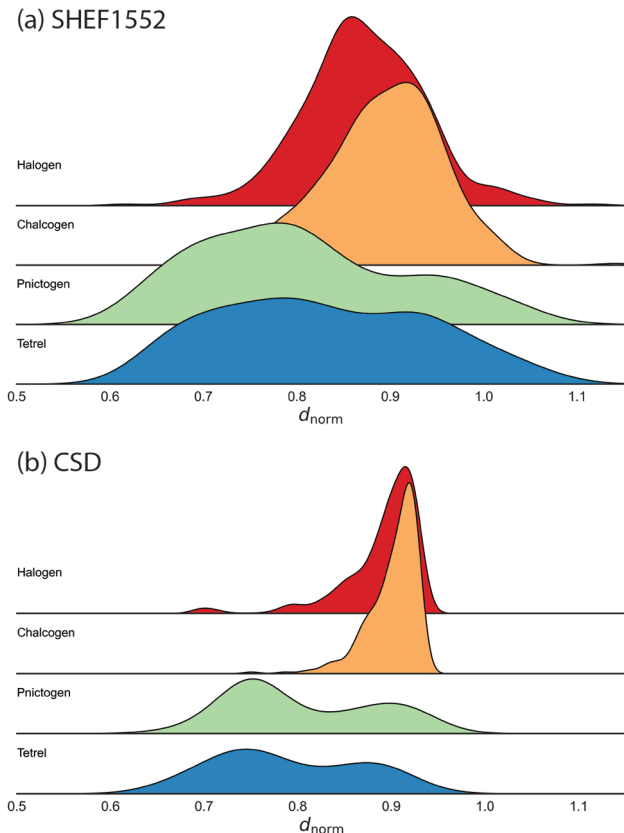


FIG. 4: Kernel density estimate plots, per interaction type, of the distribution of normalized distances (d_{norm}) found in: (a) the SHEF1552 data set of computational gas-phase structures, (b) a subset of the CSD search results for σ -hole interactions with divalent oxygen as the acceptor atom on the Lewis base.

The distributions of d_{norm} shown in Figure 4 demonstrate that the structures in SHEF1552 span a slightly wider range of normalized intermolecular distances than the crystal structures, but there are still strong similarities between the two. For example, the clearly bimodal distribution of distances for pnictogen and tetrel interactions are evident in both cases, and the peak with a lower d_{norm} has the larger density. In both cases the distributions for halogen and chalcogen interactions are much sharper with just a single obvious maxi-

mum. While the SHEF1552 distributions are significantly wider, it is important to recall that the CSD structures were constrained to those with $d_{\text{norm}} \leq 0.93$, which limits how wide the distributions can be.

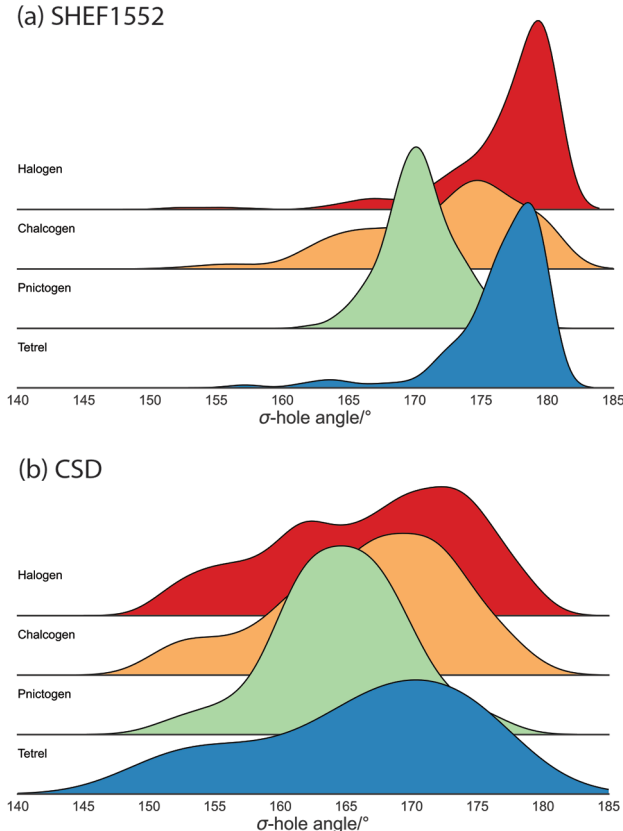


FIG. 5: Kernel density estimate plots, per interaction type, of the distribution of σ -hole angles found in: (a) the SHEF1552 data set of computational gas-phase structures, (b) a subset of the CSD search results for σ -hole interactions with divalent oxygen as the acceptor atom on the Lewis base.

In contrast to the d_{norm} distributions, Figure 5 shows that the CSD structures have consistently wider distributions of σ -hole angles than those in SHEF1552, with the latter being closer to 180° . There are still similarities between the two sets of distributions, with the relative ordering of the distribution maximum for each interaction type following the same order (pnictogen < chalcogen < tetrel < halogen). It is also notable that pnictogen-bond interactions have relatively narrow distributions in both cases. Bearing in mind this is a comparison between computational gas-phase and experimental crystal structures, Figures 4 and 5 combined provide some confidence that the structures in SHEF1552 are relevant to

experimental chemists in addition to theoreticians.

A commonly used criterion for the presence of an intermolecular interaction or noncovalent “bond” is that the distance between the two atoms involved is less than the sum of their van der Waals radii, such that $d_{\text{norm}} \leq 1.0$. In Figure 3 it can be clearly seen that a significant number ($\sim 5\%$) of the complexes in the data set have a $d_{\text{norm}} > 1.0$, yet many of these structures have σ -hole angles that are still close to 180° and, as discussed below, some have appreciable interaction energies that can be in excess of 5 kcal mol^{-1} . The use of $d_{\text{norm}} \leq 1.0$ as some form of “cut-off” for intermolecular interactions is known to be problematic,^{2,40,82,83} and in this data set would exclude a total of 82 possible σ -hole interactions, hence these structures are retained.

Figure 6 shows scatter plots of normalized distance and σ -hole angle for each type of interaction (halogen, chalcogen, pnictogen or tetrel), with each interaction type further subdivided by each donor atom. As may have been anticipated from Table I, the scatter plots clearly illustrate that each interaction type behaves somewhat differently, even though the same set of σ -hole acceptors are used throughout. For halogen, chalcogen and tetrel interactions, the scatter plots follow a similar “wedge”-shaped trend, where at short normalized distances the σ -hole angles are close to 180° (upper left of the plots). At longer distances the points then form an approximately diagonal “front”, such that each normalized distance appears to have a minimum σ -hole angle, and this minimum angle decreases as d_{norm} increases. These fronts are shown as dashed grey lines in Figure 6 and have been constructed entirely “by eye” in a qualitative fashion, similar to the work of Ostras *et al.* for a set of halogen bonding interactions to phosphine oxides.⁸⁴ In the present case, the gradient and positioning of this front varies between the types of σ -hole interaction and there are a small number of outliers. By contrast, a definitive relationship between d_{norm} and σ -hole angle for pnictogen-bond interactions in Figure 6 is more difficult to discern. The majority of observations do, however, resemble the majority of those found for tetrel-bond interactions, albeit with slightly smaller σ -hole angles. The pnictogen interactions where P or As are the σ -hole donor atom do, however, form two groups that are visible as a bimodal distribution; shorter and longer interactions in terms of d_{norm} .

Although the fronts highlighted in Figure 6 are mostly to guide the eye towards a qualitative relationship between d_{norm} and the minimum σ -hole angle, they may also be useful in identifying points of interest. For example, in the halogen interactions there is an inter-

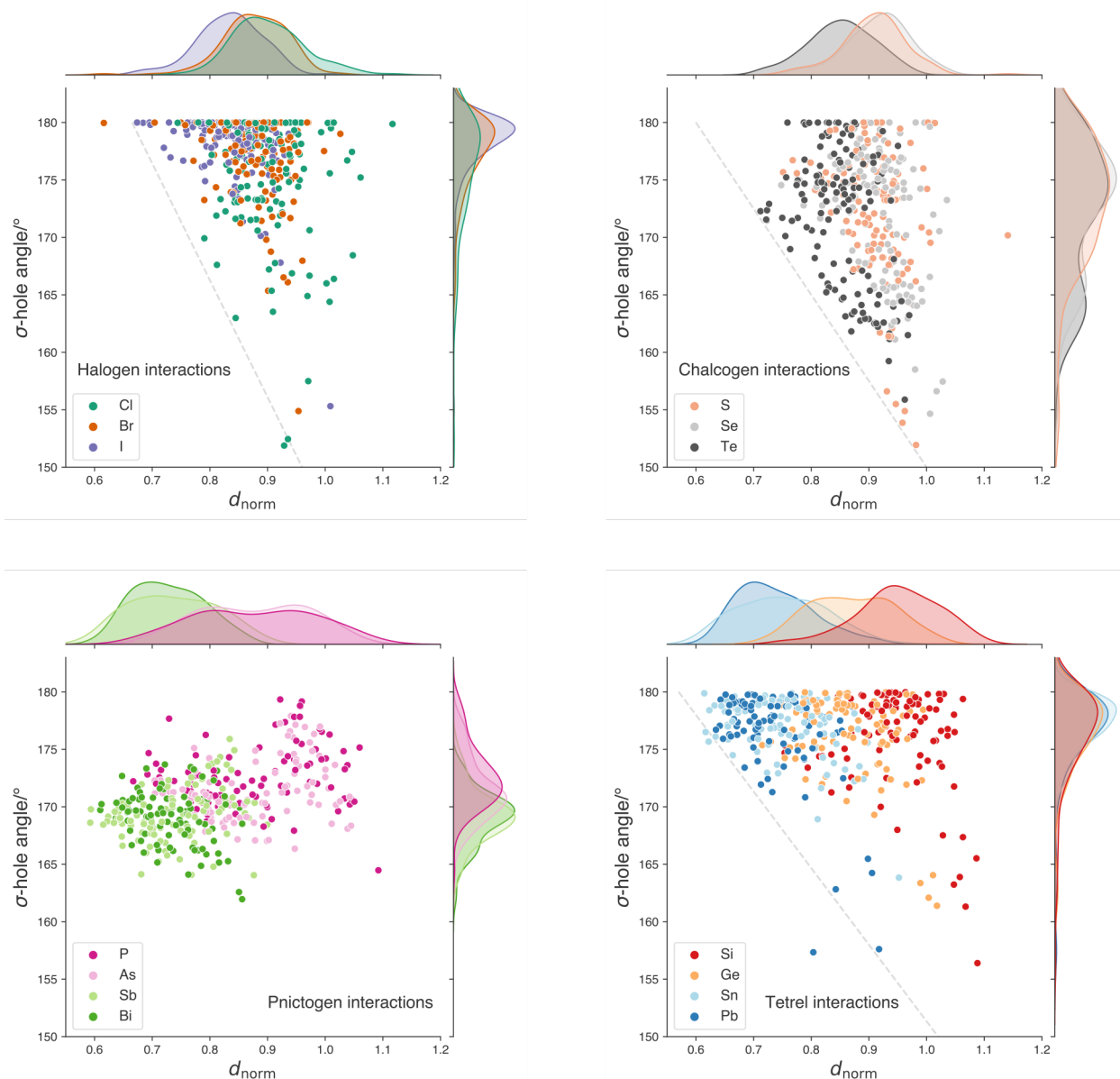


FIG. 6: Normalized distance (d_{norm}) and σ -hole angle (θ) from M06-2X-D3(0)/def2-TZVPD optimized geometries subdivided into halogen, chalcogen, pnictogen and tetrel interaction types. Each of these interaction types are also shown as kernel density estimate plots grouped by σ -hole donor atom.

action involving bromine that sits significantly to the left of the diagonal front, which may indicate an interaction that is stronger than expected. This point corresponds to the complex between bromosuccinimide and trimethylphosphine, which will be further discussed in the subsection III B on Interaction Energies. The somewhat arbitrary diagonal front for the chalcogen interactions also indicates that there may be area of interesting chemical space

that the data set does not fully explore. The upper left of the scatter plot has an area to the right of the diagonal front where no interactions are found. This suggests it may be possible to engineer chalcogen interactions that have shorter distances (while possessing a σ -hole angle close to linearity) than those from this data set, although such a qualitative picture provides no information on the exact nature of σ -hole donor and acceptor required to do so. Nor does it account for any additional factors, for example, steric effects, that may prevent interactions in this region.

The subdivisions of the plots in Figure 6 into σ -hole donor atom types provides insights into how the trends vary within a specific group of the Periodic Table. Summary statistics of these interactions in the SHEF1552 data set organised by σ -hole donor atom are presented in Tables S8 and S9. Starting with halogen-bond interactions, it is clear that those with iodine as the donor atom produce the shortest (normalized) intermolecular distances, and also the closest to linearity, which can be seen in the standard deviation/peak height of the σ -hole angle distribution. The distribution of bromine- and chlorine-based interaction distances are relatively similar, although the chlorine distribution has a longer tail corresponding to more interactions with a large d_{norm} . The distribution of σ -hole angles for bromine- and chlorine-interactions are also similar, with bromine being slightly closer to linearity. These trends are consistent with the halogen-bond strength following the polarizability of the halogen atom;^{15,85–87} $\text{Cl} < \text{Br} < \text{I}$.

The trends for the chalcogen-bond interactions are somewhat analogous to those for the halogens. The most polarizable element, tellurium, has the shorter d_{norm} values, while selenium and sulfur interactions have similar distributions. All three chalcogens show a somewhat bimodal distribution of σ -hole angles, leading to the widest spread of angles across all classes of interaction considered. In terms of normalized distance, the pnictogen interactions clearly form two groups, with phosphorous and arsenic possessing very similar distributions, and antimony and bismuth also having similar distributions, albeit at shorter normalized distances than those of the lighter pnictogens. The pnictogen σ -hole angles follow a general trend of $\text{Bi} \approx \text{Sb} < \text{As} < \text{P}$, but all of the distributions have long tails. Finally, for tetrel interactions, the scatter plot indicates that the shortest interactions are those involving Sn, followed by Pb, which does match the trend in polarizabilities ($\text{Sn} = 53$ a.u., $\text{Pb} = 47$ a.u.)⁸⁸, but the d_{norm} mean value for lead over the whole set is slightly shorter (0.7387) than that of tin (0.7503). The interactions involving germanium are somewhat longer, with silicon

longer again, which does match the trend one may expect from polarizability alone. The σ -hole angle distributions for the tetrels are very similar for all four donor atoms, although the directionality (σ -hole angle closer to 180°) does follow the trend in polarizability.

B. Interaction Energies

Whereas the normalized intermolecular distance provides an indirect guide to the strength of an intermolecular interaction, the computed interaction energy provides both a quantitative measure of the strength and reveals the relationship between the geometrical parameters and the strength of the interaction. Figure 7 plots the relationship between the CP-corrected LCCSD(T*)-F12 interaction energy and both d_{norm} [panel (a)] and the σ -hole angle [panel (b)]. As mentioned above, this coupled cluster method and basis set produced an RMSD of $0.05 \text{ kcal mol}^{-1}$ for the energies of the S66 set of intermolecular interactions,⁵⁵ and although the absolute performance on the systems considered within the SHEF1552 data set may vary, this level of accuracy for general intermolecular interactions lends extra confidence to its use in producing “benchmark” σ -hole interaction energies.

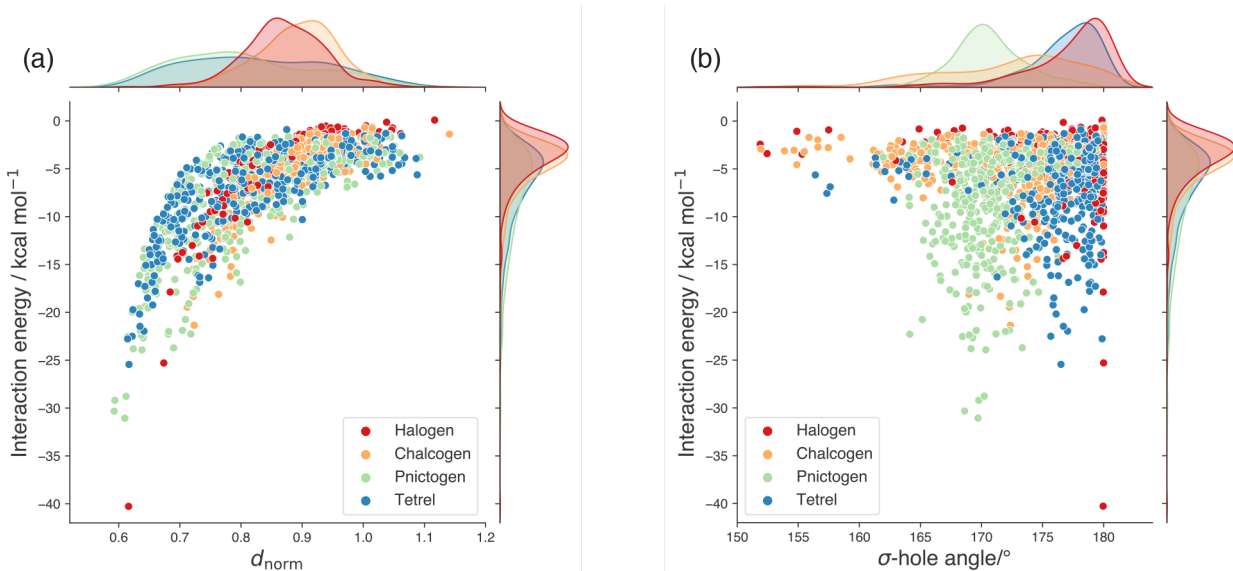


FIG. 7: Counterpoise-corrected LCCSD(T*)-F12 interaction energies plotted against: (a) normalized distance (d_{norm}); (b) σ -hole angle. All geometries were optimized at the M06-2X-D3(0)/def2-TZVPD level of theory. Also shown are kernel density estimate plots of the distributions grouped by interaction type.

The relationship between interaction energy and d_{norm} presented in Figure 7(a) shows that, unsurprisingly, there is a general trend that shorter normalized distances lead to stronger interactions. However, the overall distribution adopts a wedge- or boomerang-like shape whereby at longer distance the interaction energies gradually become stronger as d_{norm} decreases, but below a normalized distance of approximately 0.8 the strength of the interaction increases significantly more rapidly. More specifically, of the 1098 complexes with $d_{\text{norm}} \geq 0.8$, 1088 (99%) of those have an interaction energy greater (i.e., weaker) than or equal to $-10 \text{ kcal mol}^{-1}$. Meanwhile, 216 of the 454 complexes ($\sim 48\%$) with normalized distances shorter than 0.8 have interactions that are stronger than $-10 \text{ kcal mol}^{-1}$. Mean interaction strengths follow the trend halogen-bonds < chalcogen-bonds < tetrel-bonds < pnictogen-bonds, with numerical values of these and other summary statistics reported in Table S10. It is worth noting that if one used only the average normalized distance to judge the strength of interaction, it would appear that halogen-bond interactions are stronger than chalcogen-bonds, which is inverted relative to the mean of the interaction energies.

Figure 7(b) displays the variation of interaction energy with σ -hole angle. If the pnictogen-bond interactions are excluded (see Figure S3 for a version of this plot without any pnictogen-bond interactions), then once again the scatter plot shows a wedge-like distribution, with smaller σ -hole angles corresponding to weaker interactions whereas the stronger interactions tend to be close to linear in terms of this angle. As shown in Table I, the pnictogen-bond interactions tend to have a narrow distribution of σ -hole angles centered around 170.6° and there is seemingly little dependence of the interaction energy on σ -hole angle for this type of interaction. The trends in the σ -hole angle distribution for each interaction type can be qualitatively rationalized using VSEPR-type arguments by considering the σ -hole interaction as incipient formation of a new σ -bond. Halogen-bond interactions have a narrow distribution of angles that are close to 180° , which is due to the σ -hole being located along the elongation of the σ -bond that the halogen atom is engaged in, and the Lewis base typically aligns along this same axis (termed the σ -hole axis herein).^{1,15,86,89} Similarly, tetrel-bond interactions also have a narrow distributions of angles close to 180° . In this case the σ -hole axis will be an extension of the bond to the most electron withdrawing substituent attached to the tetrel center and, assuming sp^3 hybridization and a tetrahedral geometry, the three remaining substituents provide an additional steric factor that also favors σ -hole angles close to 180° .

Chalcogen-bond interactions exhibit a wide σ -hole angle distribution [Figure 7(b)]. This can be rationalised by considering the steric effect and the possibility of secondary interactions arising from the chalcogen substituent not positioned along the σ -hole axis. All chalcogens in the data set are divalent. For a trivalent pnictogen, VSEPR predicts a tetrahedral shape with a single lone-pair, resulting in a trigonal pyramidal geometry. If one imagines a Lewis base approaching along the σ -hole axis, it will also encounter repulsion from the pnictogen lone-pair that leads to a deviation from linearity. Large deviations from linearity are prevented due to steric interactions with the remaining two additional substituents, hence the σ -hole angle must fall into the narrow range of values around the mean of 170.6° mentioned above in context of Figure 7(b).

Individual plots of interaction energy against d_{norm} for the four interaction types are presented in Figure S4, with each type of interaction further grouped into σ -hole donor atom type. The summary statistics for the distributions of the interaction energies are also shown in Table S11. The energetic trends within each interaction type are analogous to those discussed above in terms of d_{norm} , with the trend in average interaction energies matching that in polarizability. One notable difference is that interactions of the lightest, least polarizable elements of the third period are noticeably weaker than those of the fourth period, which is less obvious when inspecting the normalized distances alone.

It is possible to fit a simple exponential function to the interaction energy (I.E.) data, with the resulting formula: $\text{I.E.} = -796.5 \times \exp(-5.97d_{\text{norm}})$. This function is plotted alongside a scatter plot of the interaction energy against d_{norm} as Figure S5, where it can be seen that this fit is more accurate for longer normalized distances and begins to perform poorly at shorter distance, mirroring the trends noted for Figure 7(a) above. Although more sophisticated models for predicting the strength of halogen bonding interactions already exist,^{90–93} it is interesting that this simple curve fitting to a single distance descriptor produces an RMSE of $2.55 \text{ kcal mol}^{-1}$, despite the large residuals for the strongest σ -hole interactions in this data set. A sixth-order polynomial fit to the same data results in a slightly reduced RMSE of $2.45 \text{ kcal mol}^{-1}$, but there is clear evidence of overfitting as it suggests that interactions become increasingly stronger with distance beyond $d_{\text{norm}} \approx 1.15$.

As mentioned above, it is relatively common to use $d_{\text{norm}} \leq 1.0$ as a crude criterion for the presence of an intermolecular interaction, and $\sim 5\%$ of the complexes in the SHEF1552 data set are longer than this cut-off. Figure 8 plots the LCCSD(T*)-F12 interaction energies

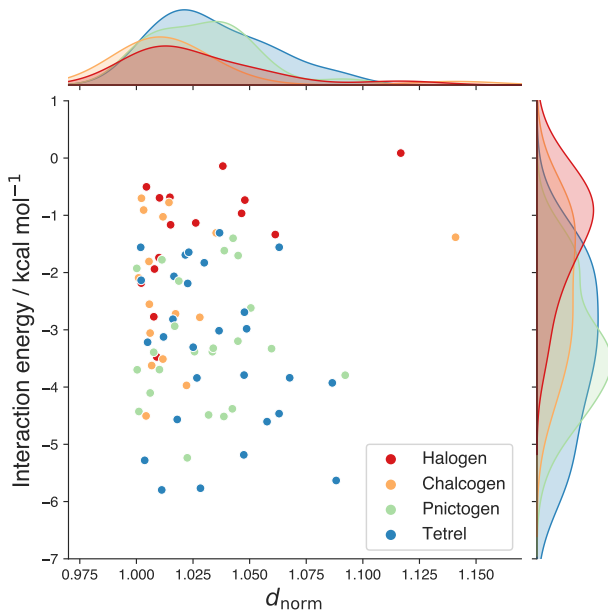


FIG. 8: Counterpoise-corrected LCCSD(T*)-F12 interaction energies for the complexes with $d_{\text{norm}} \geq 1.0$. All geometries were optimized at the M06-2X-D3(0)/def2-TZVPD level of theory. Also shown are kernel density estimate plots of the distributions grouped by interaction type.

of these 82 complexes against d_{norm} , and it can be seen that although some of the complexes possess interactions that are weaker than -1 kcal mol⁻¹, there are also six complexes with interactions between -5 and -6 kcal mol⁻¹. The mean interaction energy for these longer-ranged interactions is -2.74 kcal mol⁻¹, justifying their inclusion in the data set and adding extra evidence to the suggestion that using the $d_{\text{norm}} \leq 1.0$ cut-off excludes some structures that clearly involve attractive intermolecular interactions.^{2,40,82,83}

The strongest interaction within the data set, between N-bromosuccinimide and trimethylphosphine, has an interaction energy of -40.27 kcal mol⁻¹ and the M06-2X-D3(0)/def2-TZVPD optimized geometry is shown in Figure 9. N-bromosuccinimide is used in both electrophilic addition and substitution reactions,⁹⁶ hence interacting strongly with a soft Lewis base containing methyl groups is perhaps not surprising. Crystal structures of the N-halosuccinimides also indicate that N-bromosuccinimide and N-iodosuccinimide are strong halogen bond donors, while N-chlorosuccinimide is significantly weaker.^{97,98} The interaction energies for the N-iodosuccinimide $\cdots$ and N-chlorosuccinimide $\cdots$ trimethylphosphine complexes in SHEF1552 are -25.29 and -4.13 kcal mol⁻¹, respectively.

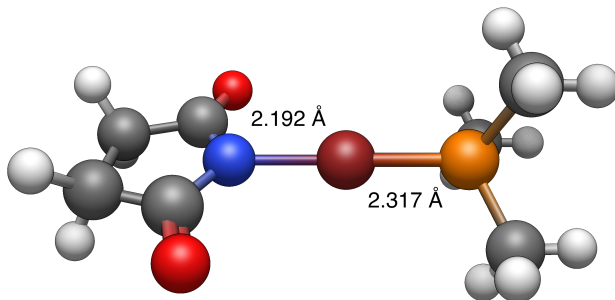


FIG. 9: The N-bromosuccinimide...trimethylphosphine complex (structure 1325 in the data set) optimized at the M06-2X-D3(0)/def2-TZVPD level of theory. The nitrogen-bromine (upper) and bromine-phosphorus (lower) interatomic distances are noted in the image. The molecular visualization software, Avogadro2,^{94,95} has drawn covalent bonds from the bromine to both the nitrogen of the succinimide and to the trimethylphosphine. These lines are based on the interatomic distance falling within a particular range.

To further assess the contact distance between N-bromosuccinimide and trimethylphosphine, the degree of interpenetration of the van der Waals crusts of the Br and P have been calculated using the index (p_{AB}) of Alvarez and co-workers:^{99,100}

$$p_{AB} = \frac{v_A + v_B - d_{AB}}{v_A + v_B - r_A - r_B} \times 100\%$$

where v_A and v_B are the van der Waals radii of atoms A and B, respectively, r_A and r_B are the covalent radii of the atoms, and d_{AB} is the interatomic distance between A and B. A p_{AB} of 0% corresponds to the two atoms being located at the sum of their van der Waals radii (or $d_{\text{norm}} = 1.0$), whereas 100% corresponds to the sum of the covalent radii. The value of p_{AB} for the N-bromosuccinimide...trimethylphosphine complex is 96.88%, indicating this is very close to a single covalent bond. A scatter plot of the interaction energy against p_{AB} for all complexes in the SHEF1552 data set is presented in Figure 10(a), where the overall distribution once again has a boomerang-like shape, although stronger complexes now appear to the lower right of the plot. Figure 10(b) shows the same points with a fifth order polynomial fit to the data, which has an RMSE of 2.26 kcal mol⁻¹. This improvement relative to the curve fitting based on d_{norm} suggests that p_{AB} may be a better descriptor for the strength/distance relationship than using the normalized distance alone.

Although extensive testing of the performance of different electronic structure methods

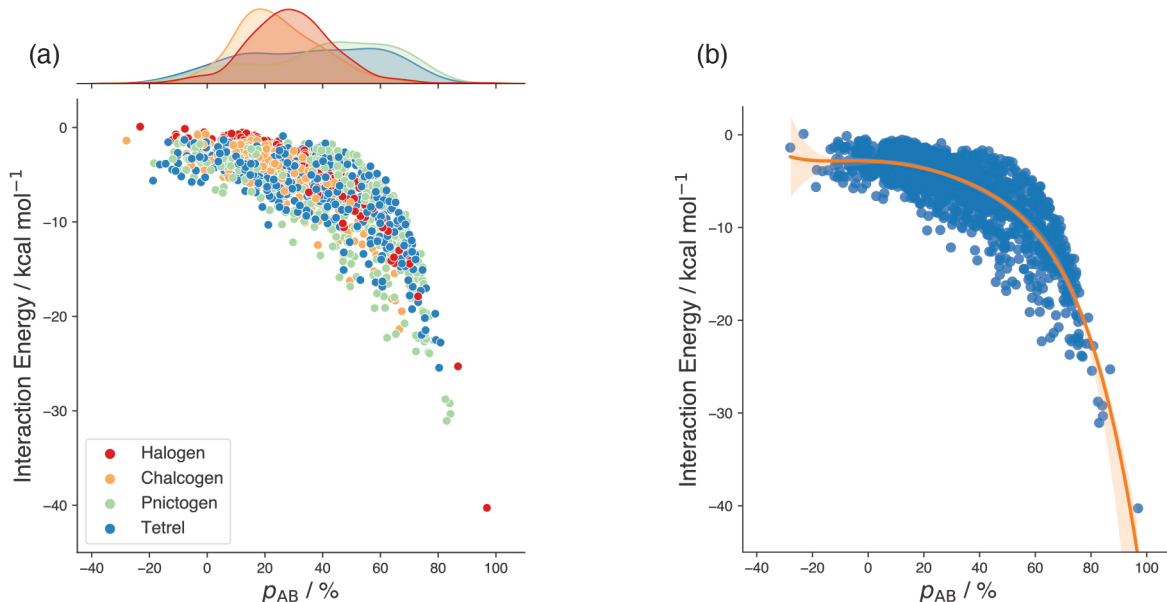


FIG. 10: Counterpoise-corrected LCCSD(T*)-F12 interaction energies plotted against the interpenetration of van der Waals crusts (p_{AB})^{99,100}. (a) indicates the type of interaction and includes kernel density estimate plots of the p_{AB} distributions grouped by interaction type. (b) a fifth order polynomial fit to the data (solid orange line) with the 95% confidence interval shown as translucent bands.

and basis sets for the σ -hole interactions in this data set is reserved for future investigations, a number of interaction energies are obtained indirectly as part of the calculations carried out in the present investigation, and hence they are briefly compared here. Specifically, the CP-LMP2-F12 energy can be obtained as part of the “benchmark” CP-LCCSD(T*)-F12 interaction energy calculations, SAPT0 energies are computed from a subset of the terms required for SAPT2, and the CP-M06-2X-D3(0) interaction energies have been calculated to reflect the M06-2X-D3(0) geometry optimization. Figure 11 shows violin plots of the error in interaction energy ($\Delta I.E.$), relative to the benchmark, calculated as the benchmark value minus the comparison method value. This ensures that positive values in the Figure indicate that a given method predicts that the interaction is stronger than the benchmark value (overbinding).

All four methods typically predict interactions stronger than those from the coupled-cluster benchmark (Figure 11 and Table II) and both SAPT truncations produce some energies with substantial errors (e.g., up to 17.40 kcal mol⁻¹ for SAPT0). For the CP-

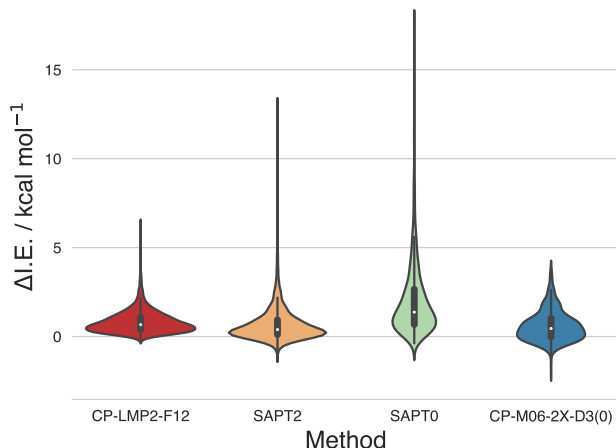


FIG. 11: Violin plots of the error in interaction energies, $\Delta\text{I.E.}$, (kcal mol^{-1}) relative to the CP-LCCSD(T*)-F12/aug-cc-pVDZ-F12 benchmark data. All geometries were optimized at the M06-2X-D3(0)/def2-TZVPD level of theory.

TABLE II: Summary statistics of the error in interaction energies, $\Delta\text{I.E.}$, (kcal mol^{-1}) relative to the CP-LCCSD(T*)-F12/aug-cc-pVDZ-F12 benchmark data.

$\Delta\text{I.E.}$	CP-LMP2-F12	SAPT2	SAPT0	CP-M06-2X-D3(0)
Mean	0.81	0.66	2.02	0.59
S.D.	0.61	0.91	2.07	0.85
Min	-0.11	-1.02	-0.38	-2.10
Max	6.30	12.99	17.40	3.87
Q1	0.38	0.11	0.67	-0.03
Q3	1.10	0.94	2.66	1.03

LMP2-F12, SAPT2 and SAPT0 methods, the maximum error occurs for the N-bromosuccinimide ··· trimethylphosphine complex highlighted as possessing significant covalent character above, and it is perhaps not surprising that SAPT methods do not work well for describing this type of interaction. If this structure is removed, the maximum errors become 3.96, 6.44 and 16.80 kcal mol^{-1} , for CP-LMP2-F12, SAPT2 and SAPT0, respectively. As one may anticipate, SAPT2 significantly outperforms SAPT0 and approaches CP-LMP2-F12 in terms of error relative to the benchmark. Meanwhile, the performance of CP-LMP2-F12 and CP-M06-2X-D3(0) are broadly similar; when taking the absolute value

of $\Delta I.E.$, the mean error for CP-LMP2-F12 is 0.81 kcal mol⁻¹, while that of CP-M06-2X-D3(0) is 0.76 kcal mol⁻¹. Additional insights into the performance of the different methods can be found by splitting the data into the four classes of σ -hole interaction under consideration in this work, with violin plots for halogen-, chalcogen-, pnictogen- and tetrel-bond interactions presented in Figures S6-S9. Briefly, CP-M06-2X-D3(0) generally outperforms CP-LMP2-F12 for halogen- and chalcogen-bond interactions, the two are roughly similar for pnictogen-bonds, and CP-M06-2X-D3(0) is slightly worse than CP-LMP2-F12 for tetrel-bond interactions.

C. SAPT Analysis

Using the “chemist’s grouping” at the SAPT2 level of theory, produces electrostatic, exchange, induction and dispersion components of the interaction, and each of these terms for every interaction in the data set is included in the CSV file. The reasonable agreement of the total SAPT2 interaction energy with the CP-LCCSD(T*)-F12/aug-cc-pVDZ-F12 benchmark lends some additional confidence to these energy component, although they should be used with caution for the strongest interactions, which have significant covalent character, such as the N-bromosuccinimide...trimethylphosphine complex. One application of SAPT analysis is to classify interactions into distinct types, as demonstrated by Gráfová *et al.* for the S22 data set of model intermolecular complexes.¹⁰¹ Their approach, which was based on the DFT-SAPT method (a related, but different approach to the SAPT used in this investigation) classified complexes as electrostatic-dominated, dispersion-dominated, or mixed-character, depending on the relative importance of the SAPT terms. For the present data set, the percentage contribution of the SAPT attractive energy components (Figure S10) shows that induction makes an important but not dominant contribution to the total interaction energy. The electrostatic component tends to have the greatest contribution, although in some cases dispersion contributes most.

For the classification of interactions herein, we adapt the DFT-SAPT approach of Gráfová *et al.* with minor modifications for SAPT. We also note, again following this previous work, that the classification is somewhat “rough” and cannot capture all of the qualitative differences between the types of interaction. Thus, we designate a complex as electrostatic-dominated type if $|E_{\text{elec}}|$ is equal to or greater than twice $|E_{\text{disp}}|$, or as dispersion-dominated

if $|E_{\text{disp}}|$ is equal to or greater than twice $|E_{\text{elec}}|$, or as mixed-character if none of the other criteria is met.

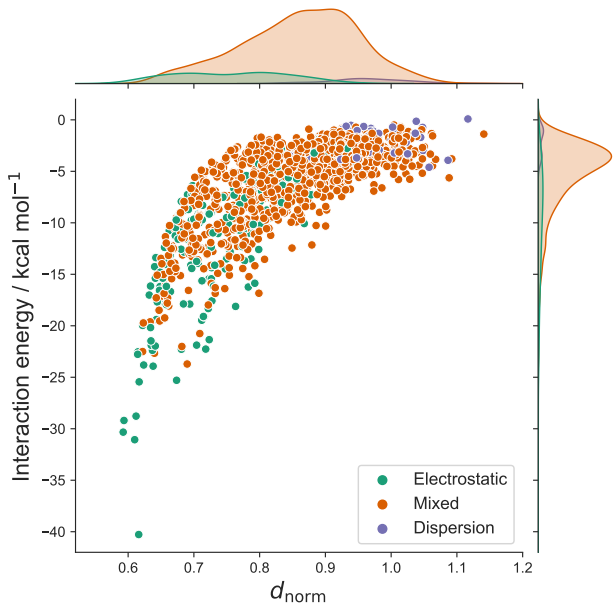


FIG. 12: Counterpoise-corrected LCCSD(T*)-F12 interaction energies plotted against the normalized distance (d_{norm}), with each point color-coded as electrostatic-dominated, dispersion-dominated, or mixed-character interaction type, as determined by SAPT analysis. All geometries were optimized at the M06-2X-D3(0)/def2-TZVPD level of theory.

Figure 12 displays the same scatterplot of CP-LCCSD(T*)-F12 interaction energies against d_{norm} as Figure 7(a), but with each point color-coded by their SAPT analysis classification into electrostatic-dominated, dispersion-dominated, or mixed-character interaction type. Although it can be seen that there is significant overlap between the distributions of the three classification types, in general the 56 complexes assigned as dispersion-dominated tend to have longer normalized distances and weaker interactions, the 214 electrostatic-dominated interactions are shorter/stronger, and the remaining 1282 mixed-character interactions sit somewhere between the two. This mixed character is consistent with the IUPAC definition published for halogen bonds “The attractive nature of the interaction is mainly due to electrostatic effects, but polarization, charge-transfer, and dispersion contributions all play an important role in causing interpenetration of van der Waals volumes”,⁹ which also holds for chalcogen and pnictogen bonds,^{11,12} and presumably any future definition of tetrel bonds. Scatterplots which show each of these electrostatic/dispersion/mixed classifications

sub-divided into σ -hole interaction type are presented in Figures S11–S13. We note that all types of σ -hole interaction are present in each classification.

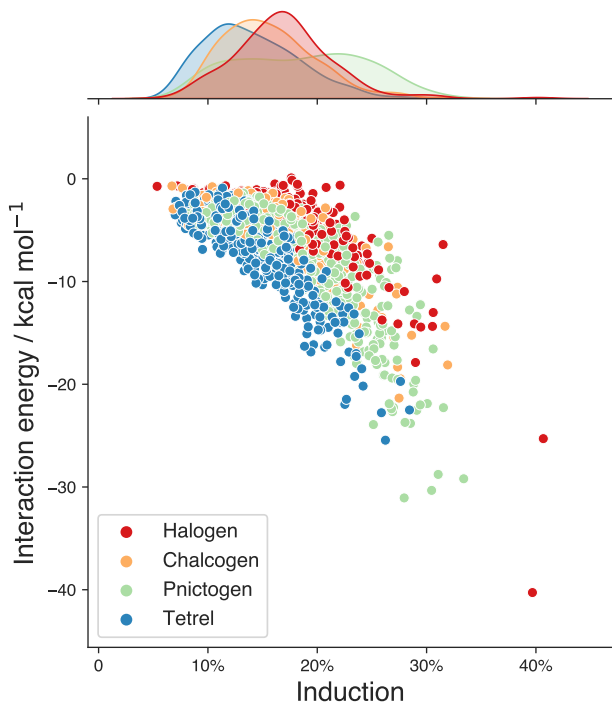


FIG. 13: Counterpoise corrected LCCSD(T*)-F12 interaction energies plotted against the percentage contribution of the induction component to the total of the attractive SAPT terms. Also shown are kernel density estimate plots of the induction contributions grouped by interaction type.

Although none of the interactions is dominated by the induction contribution, Figure 13 shows that, in general, stronger interactions have a greater contribution from induction. Indeed, complexes that display characteristics of significant covalent character, such as N-bromosuccinimide $\cdots$ trimethylphosphine have a large induction contribution to the interaction energy, although SAPT may be less reliable in cases where the interactions approach those of covalent chemical bonds. The maxima of the kernel density estimate distributions suggest that tetrel-bond interactions tend to have smaller induction contributions than chalcogen-bond interactions, with halogen-bond interactions having slightly more induction, typically, than either of those. The distribution for pnictogen-bond interactions is clearly bimodal, with some of these interactions having roughly similar amounts of induction character as tetrel-bond interactions, while others have substantially greater induction

contributions that are typically larger than those found in halogen-bond interactions.

The induction contribution to the interaction is further broken down into the electrostatic-dominated, dispersion-dominated and mixed-character classifications in Figures S14–S16. For the dispersion-dominated complexes there is no correlation between the interaction energy and the induction contribution, which is perhaps unsurprising. In contrast, the electrostatic-dominated complexes, which are typically the strongest, tend to have relatively large induction contributions. In this case, the distributions indicate that halogen-, chalcogen-, and tetrel-bond interactions have generally similar amounts of induction contribution, while pnictogen-bond interactions tend to have slightly more. The complexes classified as having mixed-character (Figure S16) have a relatively wide spread of induction contributions, with the distributions based on interaction type following a similar pattern as for the entire set of complexes shown in Figure 13.

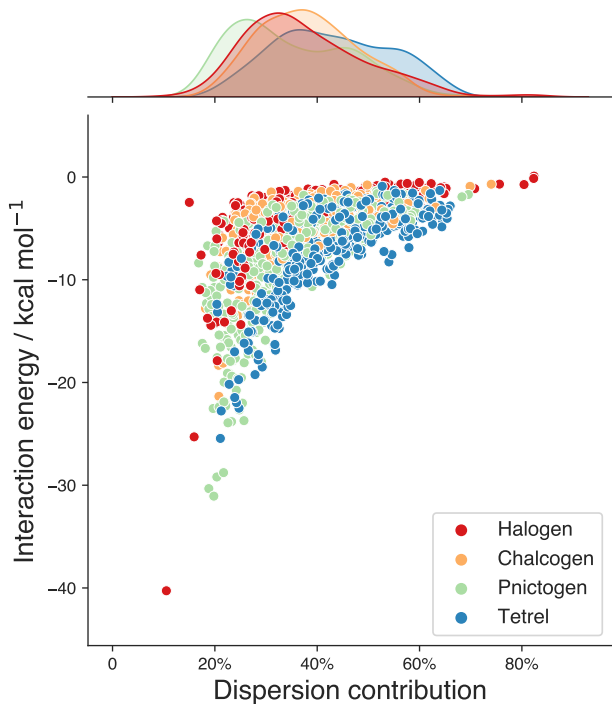


FIG. 14: Counterpoise-corrected LCCSD(T*)-F12 interaction energies plotted against the percentage contribution of the dispersion component to the total of the attractive SAPT terms. Also shown are kernel density estimate plots of the dispersion contributions grouped by interaction type.

Figure 14 displays a somewhat inverse relationship between the strength of the interaction

and the dispersion contribution, although this is clearly not linear and roughly resembles the boomerang-like shape observed for the relationship between interaction energy and d_{norm} in Figure 7(a). The distributions indicate that pnictogen-bonds tend to have the smallest dispersion contribution, followed by halogen, then chalcogen-bonds. Tetrel-bond interactions are bimodal in terms of the dispersion contribution, with one maximum close to that of chalcogen-bonds, and a second for weaker interactions with a high dispersion contribution.

Plots of the dispersion contributions for complexes classified as electrostatic-dominated and mixed-character are presented in Figures S17 and S18. The methodology used to classify electrostatic-dominated complexes guarantees that the dispersion contribution to these complexes must be relatively weak, and this is reflected in the scatter plot. There is visually little correlation between the interaction energy and dispersion contribution for this class of complexes, although it is clear that the strongest of the interactions have the smallest dispersion contributions. Turning to the mixed-character complexes, again a distribution resembling a boomerang-like shape is evident, with the stronger interactions also having smaller dispersion contributions. The dispersion-dominated complexes are not shown, but they (unsurprisingly) all have large dispersion contributions and are weak interactions.

IV. CONCLUSIONS

This study presents a data set of 1552 σ -hole interactions, specifically halogen-, chalcogen-, pnictogen- and tetrel-bonds, denoted Sigma-Hole Energies and additional Factors (SHEF1552). The elements included as σ -hole donor atoms are Cl, Br, I for the halogens, S, Se, Te for the chalcogens, P, As, Sb, Bi for the pnictogens, and Si, Ge, Sn and Pb for the tetrrels. By ensuring that each donor molecule interacts with the same set of 19 Lewis bases this facilitates comparison between the types of σ -hole interaction, and the data set as a whole covers a wider span of the relevant chemical space of σ -hole interactions than any other previous data set. The intermolecular complex geometries that form part of SHEF1552 have been optimized at the M06-2X-D3(0)/def2-TZVPD level, with counterpoise corrected interaction energies at the explicitly correlated PNO-LCCSD(T*)-F12b/aug-cc-pVDZ-F12 level. A SAPT analysis of the interaction energies in terms of electrostatic, induction, exchange, and dispersion components was also carried out.

Analysis of the data set uncovers a number of trends and physical insights, which we

summarise into the following major conclusions.

1. The qualitative trends observed in d_{norm} and the σ -hole angle for these gas-phase “computational” structures match those from a limited search of experimental crystal structures from the CSD. Pnictogen-bonds tend to have the shortest normalized distances, followed by tetrel-bonds, with halogen- and chalcogen-bonds being the longest and approximately similar in d_{norm} (halogen-bonds are slightly shorter). In terms of σ -hole angle, tetrel- and halogen-bonds are the closest to 180° , while chalcogen-bonds spans a broader range of angles. Pnictogen-bonding has a clear peak in its angle distribution at around 170° and, unlike the three other types of σ -hole interaction, has no obvious relationship between d_{norm} and the σ -hole angle.
2. Within each of the four types of σ -hole interaction considered, the strength of the interaction (both in terms of the interaction energy and d_{norm}) increases with the polarizability of the σ -hole donor atom. While this trend is somewhat expected, it is worth noting that Sn has a greater polarizability than Pb. Pnictogen-bonds have the strongest interactions on average, followed by tetrel-, chalcogen- and finally halogen-bonds.
3. Plotting the interaction energy against d_{norm} reveals a boomerang-like shape to the distribution, indicating a division between weaker and stronger complexes, which occurs in the region of $d_{\text{norm}} = 0.8$. If one uses the interpenetration of van der Waals crusts as opposed to d_{norm} ,¹⁰⁰ it is possible to fit a fifth order polynomial function to the plot with an RMSE of $2.26 \text{ kcal mol}^{-1}$.
4. Although $d_{\text{norm}} \leq 1.0$ is often used as a cut-off for intermolecular interactions, this would exclude around 5% of the structures in this data set. The mean interaction energy for these longer-range interactions is $-2.74 \text{ kcal mol}^{-1}$, adding extra evidence to the growing body of work suggesting that this cut-off is too crude.
5. In a limited benchmarking of electronic structure methods relative to the explicitly correlated coupled cluster benchmark, the MP2, M06-2X-D3(0), SAPT0 and SAPT2 methods all tend to produce interaction energies that are too strong. M06-2X-D3(0) has the lowest deviation from the benchmark at $0.59 \text{ kcal mol}^{-1}$. SAPT0 performs

very poorly for some complexes, although these are cases where there is likely to be some covalent character to the interaction.

6. The SAPT2 components of the interaction energy were used to classify the complexes into electrostatic-dominated, mixed character, and dispersion-dominated categories. While there is significant overlap between the interaction energies associated with these categories, the electrostatic-dominated complexes tend to be the strongest and dispersion-dominated the weakest. There is a subtle interplay between the different components, reflecting the broad nature of σ -hole interactions. However, the strongest interactions tend to have a greater contribution from induction, and there is something of a non-linear, inverse relationship between the dispersion component and the strength of the interaction.

In future work, the data set could be extended to include other descriptors of non-covalent interactions, such as those from the Quantum Theory of Atoms in Molecules or the Non-Covalent Interactions index,^{102–104} or to include geometries that perturb the complexes along the intermolecular distances and/or the σ -hole angle to further investigate how the interactions depend upon the geometrical parameters. One could also envisage the data set being used for extensive benchmarking of electronic structure methods, the training of machine learning models or parameterization of semi-empirical methods, or a host of other extensions and applications beyond those that we have thought of. It is our hope that by making the data set and the underlying XYZ geometries openly available, the community will be able to devise new methods and insights that enable σ -hole interactions to be used even more widely throughout chemistry.

CONFLICT OF INTEREST

The authors have no conflicts to disclose.

ACKNOWLEDGMENTS

J.G.H. thanks the UK Engineering and Physical Sciences Research Council (EPSRC) for supporting elements of this work through Grant No. EP/T027134/1. The Cambridge Crys-

tallographic Data Centre (CCDC) is thanked for contributions made to the postgraduate studentship awarded to H.W.N.

REFERENCES

- ¹T. Clark, M. Hennemann, J. S. Murray, and P. Politzer, “Halogen bonding: the σ -hole,” *J. Mol. Model.* **13**, 291–296 (2007).
- ²P. Politzer and J. S. Murray, “ σ -Hole Interactions: Perspectives and Misconceptions,” *Crystals* **7**, 212 (2017).
- ³J. S. Murray and P. Politzer, “The electrostatic potential: an overview,” *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **1**, 153–163 (2011).
- ⁴T. Clark, “ σ -Holes,” *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **3**, 13–20 (2013).
- ⁵M. H. Kolář and P. Hobza, “Computer Modeling of Halogen Bonds and Other σ -Hole Interactions,” *Chem. Rev.* **116**, 5155–5187 (2016).
- ⁶P. Politzer and J. S. Murray, “Electrostatics and Polarization in σ - and π -Hole Noncovalent Interactions: An Overview,” *ChemPhysChem* **21**, 579–588 (2020).
- ⁷N. Tarannam, R. Shukla, and S. Kozuch, “Yet another perspective on hole interactions,” *Phys. Chem. Chem. Phys.* **23**, 19948–19963 (2021).
- ⁸K. J. Donald, N. Pham, and P. Ravichandran, “Sigma Hole Potentials as Tools: Quantifying and Partitioning Substituent Effects,” *J. Phys. Chem. A* **127**, 10147–10158 (2023).
- ⁹G. R. Desiraju, P. Shing Ho, L. Kloo, A. C. Legon, R. Marquardt, P. Metrangolo, P. Politzer, G. Resnati, and K. Rissanen, “Definition of the halogen bond (IUPAC Recommendations 2013),” *Pure Appl. Chem.* **85**, 1711–1713 (2013).
- ¹⁰L. Brammer, “Halogen bonding, chalcogen bonding, pnictogen bonding, tetrel bonding: origins, current status and discussion,” *Faraday Discuss.* **203**, 485–507 (2017).
- ¹¹C. B. Aakeroy, D. L. Bryce, G. R. Desiraju, A. Frontera, A. C. Legon, F. Nicotra, K. Rissanen, S. Scheiner, G. Terraneo, P. Metrangolo, and G. Resnati, “Definition of the chalcogen bond (IUPAC Recommendations 2019),” *Pure Appl. Chem.* **91**, 1889–1892 (2019).
- ¹²G. Resnati, D. L. Bryce, G. R. Desiraju, A. Frontera, I. Krossing, A. C. Legon, P. Metrangolo, F. Nicotra, K. Rissanen, S. Scheiner, and G. Terraneo, “Definition of the pnictogen bond (IUPAC Recommendations 2023),” *Pure Appl. Chem.* **96**, 135–145 (2024).

- ¹³M. Savastano, “Ye Olde supramolecular chemistry, its modern rebranding and overarching trends in chemistry,” *Dalton Trans.* **53**, 1373–1392 (2024).
- ¹⁴R. Taylor, “Aerogen Bond, Halogen Bond, Chalcogen Bond, Pnictogen Bond, Tetrel Bond, Triel Bond . . . Why So Many Names?” *Cryst. Growth Des.* **24**, 4003–4012 (2024).
- ¹⁵G. Cavallo, P. Metrangolo, R. Milani, T. Pilati, A. Priimagi, G. Resnati, and G. Terraneo, “The Halogen Bond,” *Chem. Rev.* **116**, 2478–2601 (2016).
- ¹⁶L. Brammer, A. Peuronen, and T. M. Roseveare, “Halogen bonds, chalcogen bonds, pnictogen bonds, tetrel bonds and other σ -hole interactions: a snapshot of current progress,” *Acta Cryst.* **C79**, 204–216 (2023).
- ¹⁷P. R. Varadwaj, A. Varadwaj, H. M. Marques, and K. Yamashita, “Definition of the tetrel bond,” *CrystEngComm* **25**, 1411–1423 (2023).
- ¹⁸L. Goerigk, A. Hansen, C. Bauer, S. Ehrlich, A. Najibi, and S. Grimme, “A look at the density functional theory zoo with the advanced GMTKN55 database for general main group thermochemistry, kinetics and noncovalent interactions,” *Phys. Chem. Chem. Phys.* **19**, 32184–32215 (2017).
- ¹⁹D. Setiawan, E. Kraka, and D. Cremer, “Strength of the Pnictogen Bond in Complexes Involving Group Va Elements N, P, and As,” *J. Phys. Chem. A* **119**, 1642–1656 (2015).
- ²⁰S. Kozuch and J. M. L. Martin, “Halogen Bonds: Benchmarks and Theoretical Analysis,” *J. Chem. Theory Comput.* **9**, 1918–1931 (2013).
- ²¹J. Řezáč, K. E. Riley, and P. Hobza, “Benchmark Calculations of Noncovalent Interactions of Halogenated Molecules,” *J. Chem. Theory Comput.* **8**, 4285–4292 (2012).
- ²²A. Bauzá, I. Alkorta, A. Frontera, and J. Elguero, “On the Reliability of Pure and Hybrid DFT Methods for the Evaluation of Halogen, Chalcogen, and Pnictogen Bonds Involving Anionic and Neutral Electron Donors,” *J. Chem. Theory Comput.* **9**, 5201–5210 (2013).
- ²³I. Alkorta and A. Legon, “A reduced electrophilicity for simple Lewis acids A involved in non-covalent interactions with Lewis bases B,” *Phys. Chem. Chem. Phys.* **24**, 6856–6865 (2022).
- ²⁴N. Mehta, T. Fellowes, J. M. White, and L. Goerigk, “CHAL336 Benchmark Set: How Well Do Quantum-Chemical Methods Describe Chalcogen-Bonding Interactions?” *J. Chem. Theory Comput.* **17**, 2783–2806 (2021).
- ²⁵K. Kříž and J. Řezáč, “Non-covalent interactions atlas benchmark data sets 4: σ -hole interactions,” *Phys. Chem. Chem. Phys.* **24**, 14794–14804 (2022).

- ²⁶K. Kříž, L. Schmidt, A. T. Andersson, M.-M. Walz, and D. van der Spoel, “An Imbalance in the Force: The Need for Standardized Benchmarks for Molecular Simulation,” *J. Chem. Inf. Model.* **63**, 412–431 (2023).
- ²⁷Y. Zhao and D. G. Truhlar, “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.* **120**, 215–241 (2008).
- ²⁸S. Grimme, J. Antony, S. Ehrlich, and H. Krieg, “A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu,” *J. Chem. Phys.* **132**, 154104 (2010).
- ²⁹F. Weigend and R. Ahlrichs, “Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy,” *Phys. Chem. Chem. Phys.* **7**, 3297–3305 (2005).
- ³⁰D. Rappoport and F. Furche, “Property-optimized Gaussian basis sets for molecular response calculations,” *J. Chem. Phys.* **133**, 134105 (2010).
- ³¹B. Metz, H. Stoll, and M. Dolg, “Small-core multiconfiguration-Dirac-Hartree-Fock-adjusted pseudopotentials for post-d main group elements: Application to PbH and PbO,” *J. Chem. Phys.* **113**, 2563–2569 (2000).
- ³²K. A. Peterson, D. Figgen, E. Goll, H. Stoll, and M. Dolg, “Systematically convergent basis sets with relativistic pseudopotentials. II. Small-core pseudopotentials and correlation consistent basis sets for the post-d group 16–18 elements,” *J. Chem. Phys.* **119**, 11113–11123 (2003).
- ³³F. Neese, “Software update: The ORCA program system—Version 5.0,” *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **12**, e1606 (2022).
- ³⁴F. Weigend, “Accurate Coulomb-fitting basis sets for H to Rn,” *Phys. Chem. Chem. Phys.* **8**, 1057–1065 (2006).
- ³⁵L. P. Wolters, P. Schyman, M. J. Pavan, W. L. Jorgensen, F. M. Bickelhaupt, and S. Kozuch, “The many faces of halogen bonding: a review of theoretical models and methods,” *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **4**, 523–540 (2014).
- ³⁶<https://github.com/Sheffield-Theoretical-Chemistry/twizzler> (accessed Mar 14, 2025).

- ³⁷N. M. O’Boyle, A. L. Tenderholt, and K. M. Langner, “cclib: A library for package-independent computational chemistry algorithms,” *J. Comp. Chem.* **29**, 839–845 (2008).
- ³⁸J. P. M. Lommerse, A. J. Stone, R. Taylor, and F. H. Allen, “The Nature and Geometry of Intermolecular Interactions between Halogens and Oxygen or Nitrogen,” *J. Am. Chem. Soc.* **118**, 3108–3116 (1996).
- ³⁹M. A. Spackman and D. Jayatilaka, “Hirshfeld surface analysis,” *CrystEngComm* **11**, 19–32 (2009).
- ⁴⁰S. Alvarez, “A cartography of the van der Waals territories,” *Dalton Trans.* **42**, 8617–8636 (2013).
- ⁴¹N. Mehta and J. M. L. Martin, “Exploring the influence of $(n-1)d$ subvalence correlation and of spin-orbit coupling on chalcogen bonding,” *Phys. Chem. Chem. Phys.* **27**, 1274–1283 (2025).
- ⁴²S. F. Boys and F. Bernardi, “The calculation of small molecular interactions by the differences of separate total energies. Some procedures with reduced errors,” *Mol. Phys.* **19**, 553–566 (1970).
- ⁴³M. Schwilk, Q. Ma, C. Köppl, and H.-J. Werner, “Scalable Electron Correlation Methods. 3. Efficient and Accurate Parallel Local Coupled Cluster with Pair Natural Orbitals (PNO-LCCSD),” *J. Chem. Theory Comput.* **13**, 3650–3675 (2017).
- ⁴⁴Q. Ma, M. Schwilk, C. Köppl, and H.-J. Werner, “Scalable Electron Correlation Methods. 4. Parallel Explicitly Correlated Local Coupled Cluster with Pair Natural Orbitals (PNO-LCCSD-F12),” *J. Chem. Theory Comput.* **13**, 4871–4896 (2017).
- ⁴⁵Q. Ma and H.-J. Werner, “Scalable Electron Correlation Methods. 5. Parallel Perturbative Triples Correction for Explicitly Correlated Local Coupled Cluster with Pair Natural Orbitals,” *J. Chem. Theory Comput.* **14**, 198–215 (2018).
- ⁴⁶Q. Ma and H.-J. Werner, “Explicitly correlated local coupled-cluster methods using pair natural orbitals,” *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **8**, e1371 (2018).
- ⁴⁷K. A. Peterson, T. B. Adler, and H.-J. Werner, “Systematically convergent basis sets for explicitly correlated wavefunctions: The atoms H, He, B–Ne, and Al–Ar,” *J. Chem. Phys.* **128**, 084102 (2008).
- ⁴⁸N. Sylvetsky, M. K. Kesharwani, and J. M. L. Martin, “The aug-cc-pVnZ-F12 basis set family: Correlation consistent basis sets for explicitly correlated benchmark calculations on anions and noncovalent complexes,” *J. Chem. Phys.* **147**, 134106 (2017).

- ⁴⁹J. G. Hill and K. A. Peterson, "Correlation consistent basis sets for explicitly correlated wavefunctions: Pseudopotential-based basis sets for the post-d main group elements Ga–Rn," *J. Chem. Phys.* **141**, 094106 (2014).
- ⁵⁰J. G. Hill, "Diffuse basis functions for explicitly correlated calculations on the heavy p-block: aug-cc-pVnZ-PP-F12 sets for Ga–Kr, In–Xe, and Tl–Rn," *J. Chem. Phys.* **160**, 114102 (2024).
- ⁵¹K. A. Peterson, B. C. Shepler, D. Figgen, and H. Stoll, "On the Spectroscopic and Thermochemical Properties of ClO, BrO, IO, and Their Anions," *J. Phys. Chem. A* **110**, 13877–13883 (2006).
- ⁵²B. Metz, M. Schweizer, H. Stoll, M. Dolg, and W. Liu, "A small-core multiconfiguration Dirac–Hartree–Fock-adjusted pseudopotential for Tl - application to TlX ($X = \text{F, Cl, Br, I}$)," *Theor. Chem. Acc.* **104**, 22–28 (2000).
- ⁵³J. Řezáč, K. E. Riley, and P. Hobza, "S66: A Well-balanced Database of Benchmark Interaction Energies Relevant to Biomolecular Structures," *J. Chem. Theory Comput.* **7**, 2427–2438 (2011).
- ⁵⁴J. Řezáč, K. E. Riley, and P. Hobza, "Extensions of the S66 Data Set: More Accurate Interaction Energies and Angular-Displaced Nonequilibrium Geometries," *J. Chem. Theory Comput.* **7**, 3466–3470 (2011).
- ⁵⁵Q. Ma and H.-J. Werner, "Accurate Intermolecular Interaction Energies Using Explicitly Correlated Local Coupled Cluster Methods [PNO-LCCSD(T)-F12]," *J. Chem. Theory Comput.* **15**, 1044–1052 (2019).
- ⁵⁶H.-J. Werner, P. J. Knowles, F. R. Manby, J. A. Black, K. Doll, A. Hesselmann, D. Kats, A. Köhn, T. Korona, D. A. Kreplin, Q. Ma, T. F. Miller III, A. Mitrushenkov, K. A. Peterson, I. Polyak, G. Rauhut, and M. Sibaev, "The Molpro quantum chemistry package," *J. Chem. Phys.* **152**, 144107 (2020).
- ⁵⁷H.-J. Werner, P. J. Knowles, P. Celani, W. Györffy, A. Hesselmann, D. Kats, G. Knizia, A. Köhn, T. Korona, D. Kreplin, R. Lindh, Q. Ma, F. R. Manby, A. Mitrushenkov, G. Rauhut, M. Schütz, K. R. Shamasundar, T. B. Adler, R. D. Amos, S. J. Bennie, A. Bernhardsson, A. Berning, J. A. Black, P. J. Bygrave, R. Cimiraglia, D. L. Cooper, D. Coughtrie, M. J. O. Deegan, A. J. Dobbyn, K. Doll, M. Dornbach, F. Eckert, S. Erfort, E. Goll, C. Hampel, G. Hetzer, J. G. Hill, M. Hodges, T. Hrenar, G. Jansen, C. Köppl, C. Kollmar, S. J. R. Lee, Y. Liu, A. W. Lloyd, R. A. Mata, A. J. May, B. Mussard,

- S. J. McNicholas, W. Meyer, T. F. Miller III, M. E. Mura, A. Nicklass, D. P. O'Neill, P. Palmieri, D. Peng, K. A. Peterson, K. Pflüger, R. Pitzer, I. Polyak, M. Reiher, J. O. Richardson, J. B. Robinson, B. Schröder, M. Schwilk, T. Shiozaki, M. Sibaev, H. Stoll, A. J. Stone, R. Tarroni, T. Thorsteinsson, J. Toulouse, M. Wang, M. Welborn, and B. Ziegler, "MOLPRO, version 2023.2, a package of ab initio programs," (2023), see <http://www.molpro.net>.
- ⁵⁸F. Weigend, "A fully direct RI-HF algorithm: Implementation, optimised auxiliary basis sets, demonstration of accuracy and efficiency," *Phys. Chem. Chem. Phys.* **4**, 4285–4291 (2002).
- ⁵⁹H. W. Nash, R. A. Shaw, and J. G. Hill, "Correlation consistent auxiliary basis sets in density fitting Hartree–Fock: The atoms sodium through argon revisited," *J. Comput. Chem.* **44**, 1119–1128 (2023).
- ⁶⁰F. Weigend, "Hartree-Fock exchange fitting basis sets for H to Rn," *J. Comput. Chem.* **29**, 167–175 (2008).
- ⁶¹E. F. Valeev, "Improving on the resolution of the identity in linear R12 ab initio theories," *Chem. Phys. Lett.* **395**, 190–195 (2004).
- ⁶²R. A. Shaw and J. G. Hill, "Approaching the Hartree–Fock Limit through the Complementary Auxiliary Basis Set Singles Correction and Auxiliary Basis Sets," *J. Chem. Theory Comput.* **13**, 1691–1698 (2017).
- ⁶³S. Kritikou and J. G. Hill, "Auxiliary basis sets for density fitting in explicitly correlated calculations: The atoms H–Ar," *J. Chem. Theory Comput.* **11**, 5269–5276 (2015).
- ⁶⁴M. Schwilk, D. Usvyat, and H.-J. Werner, "Communication: Improved pair approximations in local coupled-cluster methods," *J. Chem. Phys.* **142**, 121102 (2015).
- ⁶⁵G. Knizia, T. B. Adler, and H.-J. Werner, "Simplified CCSD(T)-F12 methods: Theory and benchmarks," *J. Chem. Phys.* **130**, 054104 (2009).
- ⁶⁶O. Marchetti and H.-J. Werner, "Accurate Calculations of Intermolecular Interaction Energies Using Explicitly Correlated Coupled Cluster Wave Functions and a Dispersion-Weighted MP2 Method," *J. Phys. Chem. A* **113**, 11580–11585 (2009).
- ⁶⁷S. Ten-No, "Explicitly correlated second order perturbation theory: Introduction of a rational generator and numerical quadratures," *J. Chem. Phys.* **121**, 117–129 (2004).
- ⁶⁸S. Ten-No, "Initiation of explicitly correlated Slater-type geminal theory," *Chem. Phys. Lett.* **398**, 56–61 (2004).

- ⁶⁹M. K. Kesharwani, D. Manna, N. Sylvetsky, and J. M. L. Martin, “The X40×10 Halogen Bonding Benchmark Revisited: Surprising Importance of ($n-1$)d Subvalence Correlation,” *J. Phys. Chem. A* **122**, 2184–2197 (2018).
- ⁷⁰H.-J. Werner, G. Knizia, and F. R. Manby, “Explicitly correlated coupled cluster methods with pair-specific geminals,” *Mol. Phys.* **109**, 407–417 (2011).
- ⁷¹K. A. Peterson, C. Krause, H. Stoll, J. G. Hill, and H.-J. Werner, “Application of explicitly correlated coupled-cluster methods to molecules containing post-3d main group elements,” *Mol. Phys.* **109**, 2607–2623 (2011).
- ⁷²B. Jeziorski, R. Moszynski, and K. Szalewicz, “Perturbation Theory Approach to Intermolecular Potential Energy Surfaces of van der Waals Complexes,” *Chem. Rev.* **94**, 1887–1930 (1994).
- ⁷³D. G. A. Smith, L. A. Burns, A. C. Simmonett, R. M. Parrish, M. C. Schieber, R. Galvelis, P. Kraus, H. Kruse, R. Di Remigio, A. Alenaizan, A. M. James, S. Lehtola, J. P. Misiewicz, M. Scheurer, R. A. Shaw, J. B. Schriber, Y. Xie, Z. L. Glick, D. A. Sirianni, J. S. O’Brien, J. M. Waldrop, A. Kumar, E. G. Hohenstein, B. P. Pritchard, B. R. Brooks, I. Schaefer, Henry F., A. Y. Sokolov, K. Patkowski, I. DePrince, A. Eugene, U. Bozkaya, R. A. King, F. A. Evangelista, J. M. Turney, T. D. Crawford, and C. D. Sherrill, “PSI4 1.4: Open-source software for high-throughput quantum chemistry,” *J. Chem. Phys.* **152**, 184108 (2020).
- ⁷⁴E. G. Hohenstein and C. D. Sherrill, “Density fitting of intramonomer correlation effects in symmetry-adapted perturbation theory,” *J. Chem. Phys.* **133**, 014101 (2010).
- ⁷⁵F. Weigend, M. Häser, H. Patzelt, and R. Ahlrichs, “RI-MP2: optimized auxiliary basis sets and demonstration of efficiency,” *Chem. Phys. Lett.* **294**, 143–152 (1998).
- ⁷⁶F. Weigend, A. Köhn, and C. Hättig, “Efficient use of the correlation consistent basis sets in resolution of the identity MP2 calculations,” *J. Chem. Phys.* **116**, 3175–3183 (2002).
- ⁷⁷A. Hellweg and D. Rappoport, “Development of new auxiliary basis functions of the Karlsruhe segmented contracted basis sets including diffuse basis functions (def2-SVPD, def2-TZVPPD, and def2-QVPPD) for RI-MP2 and RI-CC calculations,” *Phys. Chem. Chem. Phys.* **17**, 1010–1017 (2015).
- ⁷⁸E. G. Hohenstein and C. D. Sherrill, “Wavefunction methods for noncovalent interactions,” *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2**, 304–326 (2012).

- ⁷⁹C. R. Groom, I. J. Bruno, M. P. Lightfoot, and S. C. Ward, "The Cambridge Structural Database," *Acta Crystallogr. B* **72**, 171–179 (2016).
- ⁸⁰I. J. Bruno, J. C. Cole, P. R. Edgington, M. Kessler, C. F. Macrae, P. McCabe, J. Pearson, and R. Taylor, "New software for searching the Cambridge Structural Database and visualizing crystal structures," *Acta Crystallogr. B* **58**, 389–397 (2002).
- ⁸¹R. A. Sykes, N. T. Johnson, C. J. Kingsbury, J. Harter, A. G. P. Maloney, I. J. Sugden, S. C. Ward, I. J. Bruno, S. A. Adcock, P. A. Wood, P. McCabe, A. A. Moldovan, F. Atkinson, I. Giangreco, and J. C. Cole, "What has scripting ever done for us? The CSD Python application programming interface (API)," *J. Appl. Cryst.* **57**, 1235–1250 (2024).
- ⁸²R. S. Rowland and R. Taylor, "Intermolecular Nonbonded Contact Distances in Organic Crystal Structures: Comparison with Distances Expected from van der Waals Radii," *J. Phys. Chem.* **100**, 7384–7391 (1996).
- ⁸³I. Dance, "Distance criteria for crystal packing analysis of supramolecular motifs," *New J. Chem.* **27**, 22–27 (2003).
- ⁸⁴A. S. Ostras, D. M. Ivanov, A. S. Novikov, and P. M. Tolstoy, "Phosphine Oxides as Spectroscopic Halogen Bond Descriptors: IR and NMR Correlations with Interatomic Distances and Complexation Energy," *Molecules* **25**, 1406 (2020).
- ⁸⁵P. Metrangolo, F. Meyer, T. Pilati, G. Resnati, and G. Terraneo, "Halogen Bonding in Supramolecular Chemistry," *Angew. Chem. Int. Ed.* **47**, 6114–6127 (2008).
- ⁸⁶P. Politzer, J. S. Murray, and T. Clark, "Halogen bonding: an electrostatically-driven highly directional noncovalent interaction," *Phys. Chem. Chem. Phys.* **12**, 7748–7757 (2010).
- ⁸⁷J. G. Hill and X. Hu, "Theoretical Insights into the Nature of Halogen Bonding in Prereactive Complexes," *Chem. Eur. J.* **19**, 3620–3628 (2013).
- ⁸⁸P. Schwerdtfeger and J. K. Nagle, "2018 Table of static dipole polarizabilities of the neutral elements in the periodic table," *Mol. Phys.* **117**, 1200–1225 (2019).
- ⁸⁹J. G. Hill and A. C. Legon, "On the directionality and non-linearity of halogen and hydrogen bonds," *Phys. Chem. Chem. Phys.* **17**, 858–867 (2015).
- ⁹⁰A. C. Legon, "A reduced radial potential energy function for the halogen bond and the hydrogen bond in complexes $B \cdots XY$ and $B \cdots HX$, where X and Y are halogen atoms," *Phys. Chem. Chem. Phys.* **16**, 12415–12421 (2014).

- ⁹¹A. C. Legon, "Correction: A reduced radial potential energy function for the halogen bond and the hydrogen bond in complexes $B \cdots XY$ and $B \cdots HX$, where X and Y are halogen atoms," *Phys. Chem. Chem. Phys.* **16**, 25199–25199 (2014).
- ⁹²R. A. Shaw and J. G. Hill, "A Simple Model for Halogen Bond Interaction Energies," *Inorganics* **7**, 19 (2019).
- ⁹³E. Engelage, D. Reinhard, and S. M. Huber, "Is There a Single Ideal Parameter for Halogen-Bonding-Based Lewis Acidity?" *Chem. Eur. J.* **26**, 3843–3861 (2020).
- ⁹⁴M. D. Hanwell, D. E. Curtis, D. C. Lonie, T. Vandermeersch, E. Zurek, and G. R. Hutchison, "Avogadro: an advanced semantic chemical editor, visualization, and analysis platform," *J. Cheminform.* **4**, 17 (2012).
- ⁹⁵"Avogadro2, version 1.99.0," <https://two.avogadro.cc> (accessed Mar 14, 2025).
- ⁹⁶J. H. Incremona and J. C. Martin, "N-bromosuccinimide. Mechanisms of allylic bromination and related reactions," *J. Am. Chem. Soc.* **92**, 627–634 (1970).
- ⁹⁷R. W. Troff, T. Mäkelä, F. Topić, A. Valkonen, K. Raatikainen, and K. Rissanen, "Alternative Motifs for Halogen Bonding," *Eur. J. Org. Chem.* **2013**, 1617–1637 (2013).
- ⁹⁸O. Makhotkina, J. Lieffrig, O. Jeannin, M. Fourmigué, E. Aubert, and E. Espinosa, "Cocrystal or Salt: Solid State-Controlled Iodine Shift in Crystalline Halogen-Bonded Systems," *Cryst. Growth Des.* **2015**, 3464–3473 (2015).
- ⁹⁹D. M. Gil, J. Echeverría, and S. Alvarez, "Tetramethylammonium Cation: Directionality and Covalency in Its Interactions with Halide Ions," *Inorg. Chem.* **61**, 9082–9095 (2022).
- ¹⁰⁰J. Echeverría and S. Alvarez, "The borderless world of chemical bonding across the van der Waals crust and the valence region," *Chem. Sci.* **14**, 11647–11688 (2023).
- ¹⁰¹L. Gráfová, M. Pitoňák, J. Řezáč, and P. Hobza, "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.* **6**, 2365–2376 (2010).
- ¹⁰²R. F. W. Bader, "A quantum theory of molecular structure and its applications," *Chem. Rev.* **91**, 893–928 (1991).
- ¹⁰³E. R. Johnson, S. Keinan, P. Mori-Sánchez, J. Contreras-García, A. J. Cohen, and W. Yang, "Revealing Noncovalent Interactions," *J. Am. Chem. Soc.* **132**, 6498–6506 (2010).
- ¹⁰⁴E. Pastorczak and C. Corminboeuf, "Perspective: Found in translation: Quantum chemical tools for grasping non-covalent interactions," *J. Chem. Phys.* **146**, 120901 (2017).

Bibliography

- [1] S. Sirimulla, J. B. Bailey, R. Vegesna and M. Narayan, *Journal of Chemical Information and Modeling*, 2013, **53**, 2781–2791.
- [2] S. Benz, A. I. Poblador-Bahamonde, N. Low-Ders and S. Matile, *Angewandte Chemie International Edition*, 2018, **57**, 5408–5412.
- [3] E. Corradi, S. V. Meille, M. T. Messina, P. Metrangolo and G. Resnati, *Angewandte Chemie - International Edition*, 2000, **39**, 1782–1786.
- [4] S. Zhu, C. Xing, W. Xu, G. Jin and Z. Li, *Crystal Growth and Design*, 2004, **4**, 53–56.
- [5] F. Allen, J. Lommerse, V. Hoy, J. Howard and G. Desiraju, *Acta Crystallographica Section B: Structural Science*, 1997, **53**, 1006–1016.
- [6] T. Clark, M. Hennemann, J. S. Murray and P. Politzer, *Journal of Molecular Modeling*, 2007, **13**, 291–296.
- [7] B. Mallada, A. Gallardo, M. Lamanec, B. De La Torre, V. Špirko, P. Hobza and P. Jelinek, *Science*, 2021, **374**, 863–867.
- [8] A. Frontera and A. Bauza, *International Journal of Molecular Sciences*, 2021, **22**, 12550.
- [9] S. J. Grabowski, *Physical Chemistry Chemical Physics*, 2017, **19**, 29742–29759.
- [10] K. Müller-Dethlefs and P. Hobza, *Chemical Reviews*, 2000, **100**, 143–167.
- [11] C. R. Groom, I. J. Bruno, M. P. Lightfoot and S. C. Ward, *Acta Crystallographica Section B*, 2016, **72**, 171–179.
- [12] M. P. Walsh, J. A. Barclay, C. S. Begg, J. Xuan, N. T. Johnson, J. C. Cole and M. O. Kitching, *JACS Au*, 2022, **2**, 2235–2250.
- [13] J. C. Cole and R. Taylor, *Crystal Growth & Design*, 2022, **22**, 1352–1364.
- [14] P. Z. Moghadam, A. Li, S. B. Wiggin, A. Tao, A. G. Maloney, P. A. Wood, S. C. Ward and D. Fairen-Jimenez, *Chemistry of Materials*, 2017, **29**, 2618–2625.
- [15] G. Cavallo, P. Metrangolo, R. Milani, T. Pilati, A. Priimagi, G. Resnati and G. Terraneo, *Chemical Reviews*, 2016, **116**, 2478–2601.
- [16] G. Cavallo, J. S. Murray, P. Politzer, T. Pilati, M. Ursini and G. Resnati, *IUCrJ*, 2017, **4**, 411–419.
- [17] I. Remsen and J. Norris, *American Chemistry Journal*, 1896, **18**, 90–95.
- [18] J. H. Miller and L. Andrews, *Inorganic Chemistry*, 1979, **18**, 988–992.

- [19] O. Hassel and K. O. Strømme, *Nature*, 1958, **182**, 1155–1156.
- [20] O. Hassel, K. O. Strømme, E. Hammarsten, C. G. Hedén, B. Malmgren and H. Palmstierna, *Acta Chemica Scandinavica*, 1959, **13**, 1781–1786.
- [21] T. Sakurai, M. Sundaralingam and G. A. Jeffrey, *Acta Crystallographica*, 1963, **16**, 354–363.
- [22] G. R. Desiraju and R. Parthasarathy, *Journal of the American Chemical Society*, 1989, **111**, 8725–8726.
- [23] P. Metrangolo and G. Resnati, *IUCrJ*, 2014, **1**, 5–7.
- [24] K. Eskandari and M. Lesani, *Chemistry - A European Journal*, 2015, **21**, 4739–4746.
- [25] N. Guo, R. Maurice, D. Teze, J. Graton, J. Champion, G. Montavon and N. Galland, *Nature Chemistry*, 2018, **10**, 428–434.
- [26] E. Rossi, M. De Santis, D. Sorbelli, L. Storchi, L. Belpassi and P. Belanzoni, *Physical Chemistry Chemical Physics*, 2020, **22**, 1897–1910.
- [27] J. G. Hill and X. Hu, *Chemistry – A European Journal*, 2013, **19**, 3620–3628.
- [28] P. Schwerdtfeger and J. K. Nagle, *Molecular Physics*, 2019, **117**, 1200–1225.
- [29] L. P. Wolters and F. M. Bickelhaupt, *ChemistryOpen*, 2012, **1**, 96–105.
- [30] C. C. Robertson, J. S. Wright, E. J. Carrington, R. N. Perutz, C. A. Hunter and L. Brammer, *Chemical Science*, 2017, **8**, 5392–5398.
- [31] A. Priimagi, G. Cavallo, P. Metrangolo and G. Resnati, *Accounts of Chemical Research*, 2013, **46**, 2686–2695.
- [32] A. C. Legon, *Angewandte Chemie International Edition*, 1999, **38**, 2686–2714.
- [33] G. R. Desiraju, P. S. Ho, L. Kloo, A. C. Legon, R. Marquardt, P. Metrangolo, P. Politzer, G. Resnati and K. Rissanen, *Pure and Applied Chemistry*, 2013, **85**, 1711–1713.
- [34] F. Heinen, E. Engelage, C. J. Cramer and S. M. Huber, *Journal of the American Chemical Society*, 2020, **142**, 8633–8640.
- [35] S. Scheiner and J. Lu, *Chemistry - A European Journal*, 2018, **24**, 8167–8177.
- [36] J. D. McCullough, *Journal of the American Chemical Society*, 1937, **59**, 789–794.
- [37] S. Husebye, *Acta Chemica Scandinavica*, 1967, **21**, 42–52.
- [38] N. Alcock, *N. Alcock*, Academic Press, 1972, vol. 15, pp. 1–58.
- [39] M. O. Mitchell, *Journal of Molecular Modeling*, 2017, **23**, 1–12.
- [40] B. Galmés, A. Juan-Bals, A. Frontera and G. Resnati, *Chemistry - A European Journal*, 2020, **26**, 4599–4606.
- [41] A. Franconetti and A. Frontera, *Dalton Transactions*, 2019, **48**, 11208–11216.
- [42] W. Zierkiewicz, R. Wysokiński, M. Michalczyk and S. Scheiner, *Physical Chemistry Chemical Physics*, 2019, **21**, 20829–20839.
- [43] R. M. Minyaev and V. I. Minkin, *Canadian Journal of Chemistry*, 1998, **76**, 776–788.
- [44] C. B. Aakeroy, D. L. Bryce, G. R. Desiraju, A. Frontera, A. C. Legon, F. Nicotra, K. Rissanen, S. Scheiner, G. Terraneo, P. Metrangolo and others, *Pure and Applied Chemistry*, 2019, **91**, 1889–1892.

- [45] K. Kříž, J. Fanfrlík and M. Lepšík, *Chemphyschem: A European Journal of Chemical Physics and Physical Chemistry*, 2018, **19**, 2540–2548.
- [46] O. Carugo and K. Djinović-Carugo, *Frontiers in Molecular Biosciences*, 2023, **10**, year.
- [47] O. I. Carugo, *Journal of Biomolecular Structure and Dynamics*, 2023, **41**, 9576–9582.
- [48] O. Carugo, G. Resnati and P. Metrangolo, *ACS Chemical Biology*, 2021, **16**, 1622–1627.
- [49] A. Varadwaj, P. R. Varadwaj, H. M. Marques and K. Yamashita, *Molecules*, 2022, **27**, 3421.
- [50] F. Carré, C. Chuit, R. J. Corriu, P. Monforte, N. K. Nayyar and C. Reyé, *Journal of Organometallic Chemistry*, 1995, **499**, 147–154.
- [51] V. M. Cangelosi, M. A. Pitt, W. J. Vickaryous, C. A. Allen, L. N. Zakharov and D. W. Johnson, *Crystal Growth and Design*, 2010, **10**, 3531–3536.
- [52] H. J. Trubenstein, S. Moaven, M. Vega, D. K. Unruh and A. F. Cozzolino, *New Journal of Chemistry*, 2019, **43**, 14305–14312.
- [53] R. Mokrai, J. Barrett, D. C. Apperley, A. S. Batsanov, Z. Benkő and D. Heift, *Chemistry - A European Journal*, 2019, **25**, 4017–4024.
- [54] S. Zahn, R. Frank, E. Hey-Hawkins and B. Kirchner, *Chemistry – A European Journal*, 2011, **17**, 6034–6038.
- [55] G. Resnati, D. L. Bryce, G. R. Desiraju, A. Frontera, I. Krossing, A. C. Legon, P. Metrangolo, F. Nicotra, K. Rissanen, S. Scheiner and G. Terraneo, *Pure and Applied Chemistry*, 2024, **96**, 135–145.
- [56] B. Ren, Y. Shi, Y. Lu, Z. Xu and H. Liu, *Computational and Theoretical Chemistry*, 2022, **1209**, 113636.
- [57] R. M. Gomila and A. Frontera, *CrystEngComm*, 2020, **22**, 7162–7169.
- [58] N. W. Mitzel and U. Losehand, *Angewandte Chemie International Edition in English*, 1997, **36**, 2807–2809.
- [59] P. Scilabra, V. Kumar, M. Ursini and G. Resnati, *Journal of Molecular Modeling*, 2018, **24**, 1–18.
- [60] H. Ullah, B. Twamley, A. Waseem, M. K. Rauf, M. N. Tahir, J. A. Platts and R. J. Baker, *Crystal Growth and Design*, 2017, **17**, 4021–4027.
- [61] S. González-Montiel, B. Flores-Chávez, J. G. Alvarado-Rodríguez, N. Andrade-López and J. A. Cogordan, *Polyhedron*, 2009, **28**, 467–472.
- [62] S. J. Grabowski, *Applied Organometallic Chemistry*, 2017, **31**, 3727.
- [63] A. Bauzá, D. Quiñero, P. M. Deyà and A. Frontera, *CrystEngComm*, 2013, **15**, 3137–3144.
- [64] S. Scheiner, *Physical Chemistry Chemical Physics*, 2021, **23**, 5702–5717.
- [65] S. Scheiner, *Journal of Physical Chemistry A*, 2017, **121**, 5561–5568.
- [66] K. J. Donald and M. Tawfik, *The Journal of Physical Chemistry A*, 2013, **117**, 14176–14183.
- [67] T. J. Mooibroek, *Molecules*, 2019, **24**, 3370.
- [68] S. J. Grabowski, *Molecules*, 2015, **20**, 11297–11316.
- [69] M. Michalczyk, W. Zierkiewicz and S. Scheiner, *ChemPhysChem*, 2018, **19**, 3122–3133.
- [70] Q. Yang, B. Zhou, Q. Li and S. Scheiner, *ChemPhysChem*, 2021, **22**, 481–487.

- [71] A. Frontera and A. Bauzá, *Physical Chemistry Chemical Physics*, 2017, **19**, 30063–30068.
- [72] A. Bauzá and A. Frontera, *Angewandte Chemie International Edition*, 2015, **54**, 7340–7343.
- [73] A. C. Legon, *Physical Chemistry Chemical Physics*, 2017, **19**, 14884–14896.
- [74] R. H. Crabtree, *Chemical Society Reviews*, 2017, **46**, 1720–1729.
- [75] S. M. Huber, J. D. Scanlon, E. Jimenez-Izal, J. M. Ugalde and I. Infante, *Physical Chemistry Chemical Physics*, 2013, **15**, 10350–10357.
- [76] J. G. Hill, *Physical Chemistry Chemical Physics*, 2014, **16**, 19137–19140.
- [77] R. Mulliken and W. Person, *Molecular complexes: a lecture and reprint volume*, Wiley, New York, 1969.
- [78] P. J. Costa, *Physical Sciences Reviews*, 2017, **2**, 20170136.
- [79] P. Politzer, J. S. Murray, T. Clark and G. Resnati, *Physical Chemistry Chemical Physics*, 2017, **19**, 32166–32178.
- [80] T. Brinck and A. N. Borrfors, *Journal of Molecular Modeling*, 2019, **25**, 1–9.
- [81] S. M. Huber, E. Jimenez-Izal, J. M. Ugalde and I. Infante, *Chemical Communications*, 2012, **48**, 7708–7710.
- [82] P. Su, Z. Tang and W. Wu, *WIREs Computational Molecular Science*, 2020, **10**, e1460.
- [83] M. J. S. Phipps, T. Fox, C. S. Tautermann and C.-K. Skylaris, *Chemical Society Reviews*, 2015, **44**, 3177–3211.
- [84] J. Řezáč and A. de la Lande, *Physical Chemistry Chemical Physics*, 2017, **19**, 791–803.
- [85] K. E. Riley, J. A. Platts, J. Řezáč, P. Hobza and J. G. Hill, *Journal of Physical Chemistry A*, 2012, **116**, 4159–4169.
- [86] S. J. Ang, A. M. Mak, M. B. Sullivan and M. W. Wong, *Physical Chemistry Chemical Physics*, 2018, **20**, 8685–8694.
- [87] J. M. Holthoff, R. Weiss, S. V. Rosokha and S. M. Huber, *Chemistry - A European Journal*, 2021, **27**, 16530–16542.
- [88] L. Brammer, E. A. Bruton and P. Sherwood, *Crystal Growth and Design*, 2001, **1**, 277–290.
- [89] A. C. Legon and D. J. Millen, *Journal of the American Chemical Society*, 1987, **109**, 356–358.
- [90] I. Alkorta and A. C. Legon, *Molecules*, 2017, **22**, 1786.
- [91] R. A. Shaw and J. G. Hill, *Inorganics*, 2019, **7**, 19.
- [92] J. A. Chisholm and S. Motherwell, *Journal of Applied Crystallography*, 2004, **37**, 331–334.
- [93] R. S. Rowland and R. Taylor, *The Journal of Physical Chemistry*, 1996, **100**, 7384–7391.
- [94] I. Dance, *New Journal of Chemistry*, 2003, **27**, 22–27.
- [95] J. P. Lommerse, A. J. Stone, R. Taylor and F. H. Allen, *Journal of the American Chemical Society*, 1996, **118**, 3108–3116.
- [96] I. J. Bruno, J. C. Cole, P. R. Edgington, M. Kessler, C. F. Macrae, P. McCabe, J. Pearson and R. Taylor, *Acta Crystallographica Section B: Structural Science*, 2002, **58**, 389–397.
- [97] R. A. Sykes, N. T. Johnson, C. J. Kingsbury, J. Harter, A. G. P. Maloney, I. J. Sugden, S. C. Ward, I. J. Bruno, S. A. Adcock, P. A. Wood, P. McCabe, A. A. Moldovan, F. Atkinson, I. Giangreco and J. C. Cole, *Journal of Applied Crystallography*, 2024, **57**, 1235–1250.
- [98] F. H. Allen, W. D. S. Motherwell, P. R. Raithby, G. P. Shields and R. Taylor, *New Journal of Chemistry*, 1999, **23**, 25–34.

- [99] C. B. Aakeröy, T. A. Evans, K. R. Seddon and I. Pálinkó, *New Journal of Chemistry*, 1999, **23**, 145–152.
- [100] H. S. Biswal, A. K. Sahu, B. Galmés, A. Frontera and D. Chopra, *ChemBioChem*, 2022, **23**, e202100498.
- [101] M. Mascal, *Journal of the Chemical Society, Perkin Transactions 2*, 1997, **0**, 1999–2001.
- [102] L. Pauling, *Journal of the American Chemical Society*, 1931, **53**, 3225–3237.
- [103] A. Bondi, *The Journal of Physical Chemistry*, 1964, **68**, 441–451.
- [104] P. Politzer and J. S. Murray, *Structural Chemistry*, 2021, **32**, 623–629.
- [105] R. S. Rowland and R. Taylor, *The Journal of Physical Chemistry*, 1996, **100**, 7384–7391.
- [106] S. S. Batsanov, *Inorganic Materials*, 2001, **37**, 871–885.
- [107] M. Mantina, A. C. Chamberlin, R. Valero, C. J. Cramer and D. G. Truhlar, *The Journal of Physical Chemistry A*, 2009, **113**, 5806–5812.
- [108] S. Alvarez, *Dalton Transactions*, 2013, **42**, 8617–8636.
- [109] A. Szabo and N. S. Ostlund, *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*, Courier Corporation, 2012.
- [110] F. Jensen, *Introduction to Computational Chemistry*, Wiley, Chichester, 2nd edn, 2013.
- [111] C. J. Cramer, *Essentials of computational chemistry: theories and models*, Wiley, Chichester, 2nd edn, 2004.
- [112] S. H. Vosko, L. Wilk and M. Nusair, *Canadian Journal of Physics*, 1980, **58**, year.
- [113] A. D. Becke, *Physical Review A*, 1988, **38**, 3098–3100.
- [114] A. D. Becke, *The Journal of Chemical Physics*, 1993, **98**, 5648–5652.
- [115] C. Lee, W. Yang and R. G. Parr, *Physical Review B*, 1988, **37**, 785–789.
- [116] P. J. Stephens, F. J. Devlin, C. F. Chabalowski and M. J. Frisch, *The Journal of Physical Chemistry*, 1994, **98**, 11623–11627.
- [117] Y. Zhao and D. G. Truhlar, *Journal of Chemical Physics*, 2006, **125**, 194101–194101.
- [118] Y. Zhao and D. G. Truhlar, *Theoretical Chemistry Accounts*, 2008, **120**, 215–241.
- [119] N. Mardirossian and M. Head-Gordon, *Molecular Physics*, 2017, **115**, 2315–2372.
- [120] S. Grimme, J. Antony, S. Ehrlich and H. Krieg, *The Journal of Chemical Physics*, 2010, **132**, 154104.
- [121] S. Grimme, J. G. Brandenburg, C. Bannwarth and A. Hansen, *The Journal of Chemical Physics*, 2015, **143**, 054107.
- [122] S. Grimme, S. Ehrlich and L. Goerigk, *Journal of Computational Chemistry*, 2011, **32**, 1456–1465.
- [123] S. Boys and F. Bernardi, *Molecular Physics*, 1970, **19**, 553–566.
- [124] X. W. Sheng, L. Mentel, O. V. Gritsenko and E. J. Baerends, *Journal of Computational Chemistry*, 2011, **32**, 2896–2901.
- [125] M. Gutowski, J. G. van Duijneveldt-van de Rijdt, J. H. van Lenthe and F. B. van Duijneveldt, *The Journal of Chemical Physics*, 1993, **98**, 4728–4737.
- [126] J.-Y. Le Questel, M. Berthelot and C. Laurence, *Journal of Physical Organic Chemistry*, 2000, **13**, 347–358.

- [127] R. B. Walsh, C. W. Padgett, P. Metrangolo, G. Resnati, T. W. Hanks and W. T. Pennington, *Crystal Growth and Design*, 2001, **1**, 165–175.
- [128] J. E. Ormond-Prout, P. Smart and L. Brammer, *Crystal Growth and Design*, 2012, **12**, 205–216.
- [129] I. Alkorta, J. Elguero and J. E. Del Bene, *ChemPhysChem*, 2018, **19**, 1886–1894.
- [130] Q. Yang, X. Yao and X. Yang, *ACS Omega*, 2020, **5**, 21300–21308.
- [131] A. Ebrahimi, M. Habibi-Khorassani and M. Doosti, *Chemical Physics Letters*, 2010, **491**, 11–16.
- [132] M. Liu, Q. Li, W. Li and J. Cheng, *Journal of Molecular Graphics & Modelling*, 2016, **65**, 35–42.
- [133] K. H. Møller, A. Kjaersgaard, A. S. Hansen, L. Du and H. G. Kjaergaard, *The Journal of Physical Chemistry A*, 2018, **122**, 3899–3908.
- [134] V. Kumar, Y. Xu and D. L. Bryce, *Chemistry – A European Journal*, 2020, **26**, 3275–3286.
- [135] V. Lippolis and C. Santi, *New Journal of Chemistry*, 2019, **43**, 11032–11033.
- [136] M. Arhangelskis, F. Topić, P. Hindle, R. Tran, A. J. Morris, D. Cinčić and T. Friščić, *Chemical Communications*, 2020, **56**, 8293–8296.
- [137] T. Fellowes and J. M. White, *CrystEngComm*, 2019, **21**, 1539–1542.
- [138] N. Radić, I. Despotović and R. Vianello, *Croatica Chemica Acta*, 2012, **85**, 495–504.
- [139] R. W. Alder, *Chemical Reviews*, 1989, **89**, 1215–1223.
- [140] A. F. Cozzolino, Q. Yang and I. Vargas-Baca, *Crystal Growth and Design*, 2010, **10**, 4959–4964.
- [141] L. Brammer, G. Mínguez Espallargas and S. Libri, *CrystEngComm*, 2008, **10**, 1712–1727.
- [142] J. Y. Lim and P. D. Beer, *Chem*, 2018, **4**, 731–783.
- [143] R. Taylor, *CrystEngComm*, 2014, **16**, 6852–6865.
- [144] N. Bedeković, T. Piteša, M. Eraković, V. Stilinović and D. Cinčić, *Crystal Growth & Design*, 2022, **22**, 2644–2653.
- [145] A. Iglesias-Reguant, J. Zielak-Milewska, T. Misiaszek, R. Zaleśny, J. M. Luis and B. Ośmiatowski, *The Journal of Organic Chemistry*, 2022, **87**, 15159–15165.
- [146] W. G. Marshall, R. H. Jones, K. S. Knight, J. Clews, R. J. Darton, W. Miller, S. J. Coles and M. B. Pitak, *CrystEngComm*, 2017, **19**, 5194–5201.
- [147] O. Makhotkina, J. Lieffrig, O. Jeannin, M. Fourmigué, E. Aubert and E. Espinosa, *Crystal Growth & Design*, 2015, **15**, 3464–3473.
- [148] F. Nunzi, D. Cesario, F. Tarantelli and L. Belpassi, *Chemical Physics Letters*, 2021, **771**, 138522.
- [149] M. Domagała and S. J. Grabowski, *The Journal of Physical Chemistry A*, 2005, **109**, 5683–5688.
- [150] K. E. Riley, J. S. Murray, J. Fanfrlík, J. Řezáč, R. J. Solá, M. C. Concha, F. M. Ramos and P. Politzer, *Journal of Molecular Modeling*, 2011, **17**, 3309–3318.
- [151] N. Chakraborty and A. K. Mitra, *Organic & Biomolecular Chemistry*, 2023, **21**, 6830–6880.
- [152] L. Brammer, A. Peuronen and T. M. Roseveare, *Acta Crystallographica Section C Structural Chemistry*, 2023, **79**, 204–216.

- [153] C. Laurence, G. Guiheneuf and B. Wojtkowiak, *Journal of the American Chemical Society*, 1979, **101**, 4793–4801.
- [154] H.-J. Böhm, S. Brode, U. Hesse and G. Klebe, *Chemistry – A European Journal*, 1996, **2**, 1509–1513.
- [155] S. H. Jungbauer, S. M. Walter, S. Schindler, L. Rout, F. Kniep and S. M. Huber, *Chemical Communications*, 2014, **50**, 6281–6284.
- [156] P. Wonner, T. Steinke, L. Vogel and S. M. Huber, *Chemistry – A European Journal*, 2020, **26**, 1258–1262.
- [157] M. Zbačnik, M. Pajski, V. Stilinović, M. Vitković and D. Cinčić, *CrystEngComm*, 2017, **19**, 5576–5582.
- [158] K.-N. Truong, J. M. Rautiainen, K. Rissanen and R. Puttreddy, *Crystal Growth & Design*, 2020, **20**, 5330–5337.
- [159] W. Borley, B. Watson, Y. P. Nizhnik, M. Zeller and S. V. Rosokha, *The Journal of Physical Chemistry A*, 2019, **123**, 7113–7123.
- [160] T. Watanabe and M. Uemura, *Chemical Communications*, 1998, 871–872.
- [161] D. Braga, F. Grepioni, K. Biradha, V. R. Pedireddi and G. R. Desiraju, *Journal of the American Chemical Society*, 1995, **117**, 3156–3166.
- [162] E. P. Painter, *Chemical Reviews*, 1941, **28**, 179–213.
- [163] D. Smith, G. Jenkin, D. Holwell and M. Keith, author, 2019.
- [164] F. E. Hahn, T. S. Dory, C. L. Barnes, D. Van der Helm and J. J. Zuckerman, *Organometallics*, 1983, **2**, 969–973.
- [165] M. Birkhahn, E. Glozba, W. Massa and J. Lorberth, *Journal of Organometallic Chemistry*, 1980, **192**, 171–176.
- [166] G. Mohammadnezhad, A. R. Ghanbarpour, M. M. Amini and S. W. Ng, *Acta Crystallographica. Section E, Structure Reports Online*, 2010, **66**, m946.
- [167] X. Xu, S. Huang, Z. Zhang, L. Cao and X. Yan, *Beilstein Journal of Organic Chemistry*, 2020, **16**, 78–87.
- [168] H. Pritzkow, *Monatshefte für Chemie / Chemical Monthly*, 1974, **105**, 621–628.
- [169] P. Dowd, W. Zhang and S. J. Geib, *Tetrahedron*, 1995, **51**, 3435–3454.
- [170] J. W. L. Wong, A. Mailman, S. M. Winter, C. M. Robertson, R. J. Holmberg, M. Murugesu, P. A. Dube and R. T. Oakley, *Chemical Communications*, 2013, **50**, 785–787.
- [171] X. Yu, A. Mailman, P. A. Dube, A. Assoud and R. T. Oakley, *Chemical Communications*, 2011, **47**, 4655–4657.
- [172] K. Lakin, A. A. Leitch, A. Assoud, W. Yong, J. Desmarais, J. S. Tse, S. Desgreniers, R. A. Secco and R. T. Oakley, *Inorganic Chemistry*, 2018, **57**, 4757–4770.
- [173] M. J. Laws, C. H. Schiesser, J. M. White and S.-L. Zheng, *Australian Journal of Chemistry*, 2000, **53**, 277–283.
- [174] K. P. Bhabak and G. Mugesh, *Chemistry – A European Journal*, 2007, **13**, 4594–4601.
- [175] M. Bourgault, R. Fourcade and G. Mascherpa, *Journal of Solid State Chemistry*, 1981, **36**, 214–220.
- [176] X. Xiao, Y. Li, Y. Dong, W. Li, K. Xu, N. Shi, X. Liu, J. Xie and P. Liu, *Journal of Molecular Structure*, 2017, **1130**, 901–908.
- [177] Z. H. Fard, C. Müller and S. Dehnen, *Zeitschrift für Anorganische und Allgemeine Chemie*, 2008, **634**, 1851–1856.
- [178] A. Rauk, I. R. Hunt and B. A. Keay, *The Journal of Organic Chemistry*, 1994, **59**, 6808–6816.

- [179] P. Saenz, R. E. Cachau, G. Seoane, M. Kieninger and O. N. Ventura, *The Journal of Physical Chemistry A*, 2006, **110**, 11734–11751.
- [180] R. H. Jones, C. L. Bull, K. S. Knight and W. G. Marshall, *CrystEngComm*, 2023, **25**, 4146–4156.
- [181] M. K. Krepps, S. Parkin and D. A. Atwood, *Crystal Growth & Design*, 2001, **1**, 291–297.
- [182] J. Lipkowski, M. S. Fonari, V. C. Kravtsov, Y. A. Simonov, E. V. Ganin and V. O. Gelmboldt, *Journal of Chemical Crystallography*, 1996, **26**, 823–833.
- [183] V. Nemec, T. Piteša, T. Friščić and D. Cinčić, *Crystal Growth & Design*, 2020, **20**, 3617–3624.
- [184] E. A. Chulanova, E. A. Radiush, I. K. Shundrina, I. Y. Bagryanskaya, N. A. Semenov, J. Beckmann, N. P. Gritsan and A. V. Zibarev, *Crystal Growth & Design*, 2020, **20**, 5868–5879.
- [185] J. Zhang, P. Zhang, F. Han, G. Chen, L. Zhang and X. Wei, *Industrial & Engineering Chemistry Research*, 2009, **48**, 1287–1291.
- [186] A. S. Ostras', D. M. Ivanov, A. S. Novikov and P. M. Tolstoy, *Molecules*, 2020, **25**, 1406.
- [187] S. Y. Oh, C. W. Nickels, F. Garcia, W. Jones and T. Friščić, *CrystEngComm*, 2012, **14**, 6110–6114.
- [188] K. Nieszporek, P. Podkościelny and J. Nieszporek, *Physical Chemistry Chemical Physics*, 2016, **18**, 5957–5963.
- [189] K. Nieszporek, P. Podkościelny and J. Nieszporek, *Physical Chemistry Chemical Physics*, 2016, **18**, 5957–5963.
- [190] I. Olovsson, *The Journal of Chemical Physics*, 1968, **49**, 1063–1067.
- [191] R. Tepper, B. Schulze, M. Jäger, C. Friebe, D. H. Scharf, H. Görls and U. S. Schubert, *The Journal of Organic Chemistry*, 2015, **80**, 3139–3150.
- [192] J. Y. C. Lim and P. D. Beer, *European Journal of Inorganic Chemistry*, 2017, **2017**, 220–224.
- [193] A. Brown and P. Beer, *Chemical Communications*, 2016, **52**, 8645–8658.
- [194] S. Vela, J. Jornet-Somoza, M. M. Turnbull, R. Feyerherm, J. J. Novoa and M. Deumal, *Inorganic Chemistry*, 2013, **52**, 12923–12932.
- [195] V. Kumar, C. Rodrigue and D. L. Bryce, *Crystal Growth & Design*, 2020, **20**, 2027–2034.
- [196] H. Weichmann and J. Meunier-Piret, *Organometallics*, 1993, **12**, 4097–4103.
- [197] R. Puttreddy, O. Jurček, S. Bhowmik, T. Mäkelä and K. Rissanen, *Chemical Communications*, 2016, **52**, 2338–2341.
- [198] N. Schulz, P. Sokkar, E. Engelage, S. Schindler, M. Erdelyi, E. Sanchez-Garcia and S. M. Huber, *Chemistry – A European Journal*, 2018, **24**, 3464–3473.
- [199] M. Cametti, K. Raatikainen, P. Metrangolo, T. Pilati, G. Terraneo and G. Resnati, *Organic & Biomolecular Chemistry*, 2012, **10**, 1329–1333.
- [200] S. Mondal, A. Rashid and P. Ghosh, *Journal of Organometallic Chemistry*, 2021, **952**, 122027.
- [201] A. A. Artemjev, A. P. Novikov, G. M. Burkin, A. A. Sapronov, A. S. Kubasov, V. G. Nenajdenko, V. N. Khrustalev, A. V. Borisov, A. A. Kirichuk, A. S. Kritchenkov, R. M. Gomila, A. Frontera and A. G. Tskhovrebov, *International Journal of Molecular Sciences*, 2022, **23**, 6372.
- [202] J. L. Dutton and P. J. Ragogna, *Chemistry – A European Journal*, 2010, **16**, 12454–12461.

- [203] K. Dashtian, S. Shahsavari, M. Usman, Y. Joseph, M. R. Ganjali, Z. Yin and M. Rahimi-Nasrabadi, *Coordination Chemistry Reviews*, 2024, **504**, 215644.
- [204] Y. Gao, M. Choudhari, G. K. Such and C. Ritchie, *Chemical Science*, 2022, **13**, 2510–2527.
- [205] N. I. Gumerova and A. Rompel, *Nature Reviews Chemistry*, 2018, **2**, 1–20.
- [206] R. Kiebach, C. Näther and W. Bensch, *Solid State Sciences*, 2006, **8**, 964–970.
- [207] R. Kiebach, C. Näther, P. Kögerler and W. Bensch, *Dalton Transactions*, 2007, 3221.
- [208] B. Dutta, B. Adhikary, P. Bag, U. Flörke and K. Nag, *Journal of the Chemical Society, Dalton Transactions*, 2002, 2760–2767.
- [209] M. Tadokoro, H. Sakiyama, N. Matsumoto, H. Okawa and S. Kida, *Bulletin of the Chemical Society of Japan*, 1990, **63**, 3337–3339.
- [210] A. Liberato, A. Aguinaco, M. P. Clares, E. Delgado-Pinar, J. Pitarch-Jarque, S. Blasco, M. G. Basallote, E. García-España and B. Verdejo, *Dalton Transactions*, 2017, **46**, 6645–6653.
- [211] D. Braga and F. Grepioni, *Accounts of Chemical Research*, 1997, **30**, 81–87.
- [212] G. R. Desiraju, *Accounts of Chemical Research*, 1996, **29**, 441–449.
- [213] S. G. Rosenberg, M. Barclay and D. H. Fairbrother, *Physical Chemistry Chemical Physics*, 2013, **15**, 4002–4015.
- [214] S.-Y. Cho and T. Mochida, *Inorganic Chemistry*, 2020, **59**, 847–853.
- [215] A. S. Romanov, J. M. Mulroy, M. Y. Antipin and T. V. Timofeeva, *Acta Crystallographica Section C Crystal Structure Communications*, 2009, **65**, m431–m435.
- [216] T. Gerber, K. Potgieter and P. Mayer, *Inorganic Chemistry Communications*, 2011, **14**, 1115–1118.
- [217] M. Y. Petyuk, A. S. Berezin, A. L. Gushchin, I. Y. Bagryanskaya, A. Y. Baranov and A. V. Artem'ev, *Inorganica Chimica Acta*, 2021, **516**, 120136.
- [218] L. R. Sita, R. Xi, G. P. A. Yap, L. M. Liable-Sands and A. L. Rheingold, *Journal of the American Chemical Society*, 1997, **119**, 756–760.
- [219] T. M. Mitchell, *Machine Learning*, McGraw-Hill, 1997.
- [220] A. Maćkiewicz and W. Ratajczak, *Computers & Geosciences*, 1993, **19**, 303–342.
- [221] A. Glielmo, B. E. Husic, A. Rodriguez, C. Clementi, F. Noé and A. Laio, *Chemical Reviews*, 2021, **121**, 9722–9758.
- [222] K.-i. Nomura, A. Mishra, T. Sang, R. K. Kalia, A. Nakano and P. Vashishta, *The Journal of Physical Chemistry Letters*, 2024, **15**, 5288–5294.
- [223] M. Martínez-Díaz and F. Soriguera, *Transportation Research Procedia*, 2018, **33**, 275–282.
- [224] S. Gow, M. Niranjana, S. Kanza and J. G. Frey, *Digital Discovery*, 2022, **1**, 551–567.
- [225] C. Arnold, *Nature Medicine*, 2023, **29**, 1292–1295.
- [226] Z. Zhou, X. Li and R. N. Zare, *ACS Central Science*, 2017, **3**, 1337–1344.
- [227] J. A. Keith, V. Vassilev-Galindo, B. Cheng, S. Chmiela, M. Gastegger, K.-R. Müller and A. Tkatchenko, *Chemical Reviews*, 2021, **121**, 9816–9872.

- [228] F. Oviedo, J. L. Ferres, T. Buonassisi and K. T. Butler, *Accounts of Materials Research*, 2022, **3**, 597–607.
- [229] J. B. Schriber, D. R. Nascimento, A. Koutsoukas, S. A. Spronk, D. L. Cheney and C. D. Sherrill, *The Journal of Chemical Physics*, 2021, **154**, 184110.
- [230] M. Glavatskikh, T. Madzhidov, V. Solov'ev, G. Marcou, D. Horvath and A. Varnek, *Molecular Informatics*, 2016, **35**, 629–638.
- [231] F. Ruggiu, G. Marcou, A. Varnek and D. Horvath, *Molecular Informatics*, 2010, **29**, 855–868.
- [232] S. Bose, D. Dhawan, S. Nandi, R. R. Sarkar and D. Ghosh, *Physical Chemistry Chemical Physics*, 2018, **20**, 22987–22996.
- [233] D. P. Devore and K. L. Shuford, *Journal of Chemical Information and Modeling*, 2024, **64**, 8201–8214.
- [234] D. Weininger, *Journal of Chemical Information and Computer Sciences*, 1988, **28**, 31–36.
- [235] L. Goerigk, A. Hansen, C. Bauer, S. Ehrlich, A. Najibi and S. Grimme, *Physical Chemistry Chemical Physics*, 2017, **19**, 32184–32215.
- [236] N. Mehta, T. Fellowes, J. M. White and L. Goerigk, *Journal of Chemical Theory and Computation*, 2021, **17**, 2783–2806.
- [237] K. Kříž and J. Řezáč, *Physical Chemistry Chemical Physics*, 2022, **24**, 14794–14804.
- [238] F. Weigend and R. Ahlrichs, *Physical Chemistry Chemical Physics*, 2005, **7**, 3297–3305.
- [239] D. Rappoport and F. Furche, *The Journal of Chemical Physics*, 2010, **133**, 134105.
- [240] F. Neese, *WIREs Computational Molecular Science*, 2022, **12**, e1606.
- [241] Q. Ma and H.-J. Werner, *WIREs Computational Molecular Science*, 2018, **8**, e1371.
- [242] G. Knizia, T. B. Adler and H.-J. Werner, *The Journal of Chemical Physics*, 2009, **130**, 054104.
- [243] S. R. Heller, A. McNaught, I. Pletnev, S. Stein and D. Tchekhovskoi, *Journal of Cheminformatics*, 2015, **7**, 23.
- [244] N. O'Boyle and A. Dalke, *DeepSMILES: An Adaptation of SMILES for Use in Machine-Learning of Chemical Structures*, 2018, <https://chemrxiv.org/engage/chemrxiv/article-details/60c73ed6567dfe7e5fec388d>.
- [245] M. Krenn, Q. Ai, S. Barthel, N. Carson, A. Frei, N. C. Frey, P. Friederich, T. Gaudin, A. A. Gayle, K. M. Jablonka, R. F. Lameiro, D. Lemm, A. Lo, S. M. Moosavi, J. M. Nápoles-Duarte, A. Nigam, R. Pollice, K. Rajan, U. Schatzschneider, P. Schwaller, M. Skreta, B. Smit, F. Strieth-Kalthoff, C. Sun, G. Tom, G. Falk von Rudorff, A. Wang, A. D. White, A. Young, R. Yu and A. Aspuru-Guzik, *Patterns*, 2022, **3**, 100588.
- [246] A. Cereto-Massagué, M. J. Ojeda, C. Valls, M. Mulero, S. Garcia-Vallvé and G. Pujadas, *Methods*, 2015, **71**, 58–63.
- [247] J. L. Durant, B. A. Leland, D. R. Henry and J. G. Nourse, *Journal of Chemical Information and Computer Sciences*, 2002, **42**, 1273–1280.
- [248] E. E. Bolton, Y. Wang, P. A. Thiessen and S. H. Bryant, *Annual Reports in Computational Chemistry*, Elsevier, 2008, vol. 4, pp. 217–241.
- [249] J. M. Barnard and G. M. Downs, *Journal of Chemical Information and Computer Sciences*, 1997, **37**, 141–142.
- [250] G. Landrum, P. Tosco, B. Kelley, Ric, sriniker, gedec, R. Vianello, D. Cosgrove, NadineSchneider, E. Kawashima,

- D. N, A. Dalke, G. Jones, B. Cole, M. Swain, S. Turk, AlexanderSavelyev, A. Vaucher, M. Wójcikowski, I. Take, D. Probst, V. F. Scalfani, K. Ujihara, g. godin, A. Pahl, F. Berenger, JLVarjo, jasondbiggs, strets123 and JP, *rdkit/rdkit: 2022_09_4 (Q3 2022) Release*, 2023, <https://zenodo.org/records/7541264>.
- [251] R. E. Carhart, D. H. Smith and R. Venkataraghavan, *Journal of Chemical Information and Computer Sciences*, 1985, **25**, 64–73.
- [252] R. Nilakantan, N. Bauman, J. S. Dixon and R. Venkataraghavan, *Journal of Chemical Information and Computer Sciences*, 1987, **27**, 82–85.
- [253] D. Rogers and M. Hahn, *Journal of Chemical Information and Modeling*, 2010, **50**, 742–754.
- [254] J. Yang, Y. Cai, K. Zhao, H. Xie and X. Chen, *Drug Discovery Today*, 2022, **27**, 103356.
- [255] H. L. Morgan, *Journal of Chemical Documentation*, 1965, **5**, 107–113.
- [256] K. Pearson, *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 1901, **2**, 559–572.
- [257] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, Bertrand Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot and E. Duchesnay, *Journal of Machine Learning Research*, 2011, **12**, 2825–2830.
- [258] T. Chen and C. Guestrin, Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, New York, NY, USA, 2016, pp. 785–794.
- [259] T. K. Ho, Proceedings of 3rd International Conference on Document Analysis and Recognition, 1995, pp. 278–282 vol.1.
- [260] T. Head, MechCoder, G. Louppe, I. Shcherbatyi, fcharras, Z. Vinícius, cmmalone, C. Schröder, nel215, N. Campos, T. Young, S. Cereda, T. Fan, rene rex, K. K. Shi, J. Schwabedal, carlosdanielcsantos, Hvass-Labs, M. Pak, SoManyUsernamesTaken, F. Callaway, L. Estève, L. Besson, M. Cherti, K. Pfannschmidt, F. Linzberger, C. Cauet, A. Gut, A. Mueller and A. Fabisch, *scikit-optimize/scikit-optimize: v0.5.2*, 2018, <https://zenodo.org/records/1207017>.
- [261] J. A. Pople, *Reviews of Modern Physics*, 1999, **71**, 1267–1274.
- [262] A. Gavezzotti, *Zeitschrift für Kristallographie*, 2005, **220**, 499–510.
- [263] J. Liu and X. He, *Physical Chemistry Chemical Physics*, 2020, **22**, 12341–12367.
- [264] P. D. Walker and P. G. Mezey, *Journal of the American Chemical Society*, 1993, **115**, 12423–12430.
- [265] P. G. Mezey, *Theoretical and Computational Chemistry*, Elsevier, 1999, vol. 6, pp. 593–612.
- [266] T. E. Exner and P. G. Mezey, *Journal of Computational Chemistry*, 2003, **24**, 1980–1986.
- [267] M. Habgood, *Molecular Simulation*, 2012, **38**, 1103–1113.
- [268] D. W. Zhang and J. Z. H. Zhang, *The Journal of Chemical Physics*, 2003, **119**, 3599–3605.
- [269] Y. Mei, C. Ji and J. Z. H. Zhang, *The Journal of Chemical Physics*, 2006, **125**, 094906.
- [270] V. Deev and M. A. Collins, *The Journal of Chemical Physics*, 2005, **122**, 154102.
- [271] M. A. Collins and V. A. Deev, *The Journal of Chemical Physics*, 2006, **125**, 104104.

- [272] M. A. Collins and R. P. A. Bettens, *Chemical Reviews*, 2015, **115**, 5607–5642.
- [273] A. A. Hagberg, D. A. Schult and P. J. Swart, *scipy*, 2008.
- [274] N. M. O’Boyle, M. Banck, C. A. James, C. Morley, T. Vandermeersch and G. R. Hutchison, *Journal of Cheminformatics*, 2011, **3**, 33.
- [275] A. Peuronen, A. Valkonen, M. Kortelainen, K. Rissanen and M. Lahtinen, *Crystal Growth & Design*, 2012, **12**, 4157–4169.
- [276] *Halogen Bonding of N-Halosuccinimides with Amines and Effects of Brønsted Acids in Quinuclidine-Catalyzed Halocyclizations - Li - 2021 - Helvetica Chimica Acta - Wiley Online Library*, <https://onlinelibrary.wiley.com/doi/10.1002/hlca.202100080>.
- [277] G. Anyfanti, A. Bauzá, L. Gentiluomo, J. Rodrigues, G. Portalone, A. Frontera, K. Rissanen and R. Puttreddy, *Frontiers in Chemistry*, 2021, **9**, year.
- [278] S. H. Jungbauer, D. Bulfield, F. Kniep, C. W. Lehmann, E. Herdtweck and S. M. Huber, *Journal of the American Chemical Society*, 2014, **136**, 16740–16743.
- [279] D. Bulfield, E. Engelage, L. Mancheski, J. Stoesser and S. M. Huber, *Chemistry – A European Journal*, 2020, **26**, 1567–1575.
- [280] H. Pritzkow, *Zeitschrift für anorganische und allgemeine Chemie*, 1974, **409**, 237–247.
- [281] T. Borowiak, M. Kubicki, W. Wysocka and A. Przybył, *Journal of Molecular Structure*, 1998, **442**, 103–113.
- [282] V. Nemec and D. Cinčić, *Crystal Growth & Design*, 2022, **22**, 5796–5801.
- [283] H. Cho, E. K. Lee and I. S. Choi, *Scientific Reports*, 2020, **10**, 21155.