

# New Applications of the Invariom Database

Dissertation

zur Erlangung des  
mathematisch-naturwissenschaftlichen Doktorgrades  
*Doctor rerum naturalium*  
der Georg-August-Universität Göttingen

vorgelegt von  
Claudia Manuela Wandtke  
geborene Orben  
aus Duisburg

Göttingen 2016

**Betreuungsausschuss:**

Prof. George M. Sheldrick, Ph.D., Institut für Anorganische Chemie.

PD Dr. Birger Dittrich, Institut für Anorganische Chemie und Strukturchemie, Heinrich Heine Universität Düsseldorf.

Dr. Tim Grüne, Paul Scherrer Institut, Schweiz.

**Mitglieder der Prüfungskommission:**

Referent: Prof. George M. Sheldrick, Ph.D., Institut für Anorganische Chemie.

Korreferent: PD Dr. Birger Dittrich, Institut für Anorganische Chemie und Strukturchemie, Heinrich Heine Universität Düsseldorf.

Weitere Mitglieder der Prüfungskommission:

Dr. Tim Grüne, Paul Scherrer Institut, Schweiz.

Prof. Dr. Dietmar Stalke, Institut für Anorganische Chemie.

Prof. Dr. Inke Siewert, Institut für Anorganische Chemie.

Dr. Heidrun Sowa, Abteilung Kristallographie, Fakultät für Geowissenschaften und Geographie.

Tag der Disputation: 12.12.2016



# Contents

<b>Abbreviations</b>	<b>v</b>
<b>1 General Information and Theoretical Background</b>	<b>1</b>
1.1 Crystallography . . . . .	1
1.1.1 Reflection intensities . . . . .	1
1.1.2 Least-squares refinement . . . . .	3
1.1.3 Residual factor . . . . .	4
1.2 Charge density . . . . .	4
1.2.1 Different charge density models . . . . .	5
1.2.2 Databases for transferable multipole parameters . . . . .	8
1.3 The Invariom Concept . . . . .	9
1.3.1 Assignment formalism . . . . .	9
1.3.2 Treatment of invariom model compounds . . . . .	11
1.3.3 Invariom history . . . . .	11
1.3.4 New invariom developments parallel to this thesis . . . . .	13
1.4 Scope of this thesis . . . . .	15
<b>2 Invariom database renewal</b>	<b>17</b>
2.1 Introduction . . . . .	17
2.1.1 Geometry Optimization . . . . .	17
2.1.2 Resolution . . . . .	18
2.1.3 Scattering factors . . . . .	18
2.1.4 New compounds . . . . .	19
2.2 Experiments and Results . . . . .	20
2.2.1 Performance of the new database in refinements . . . . .	20
2.2.2 Addition of new halogen model compounds . . . . .	21
2.2.3 New invariom names of double bonded oxygen for invariom refinements	22
<b>3 Identifying metals in coordination compounds</b>	<b>25</b>
3.1 Introduction . . . . .	25
3.1.1 Isomorphism . . . . .	26
3.2 The invariom like approach . . . . .	27
3.2.1 Metal atoms on special positions . . . . .	28
3.2.2 Complexes with multiple possible electronic configurations . . . . .	28
3.3 Experiments and Results . . . . .	29
3.3.1 Procedural details . . . . .	29
3.3.2 Case studies . . . . .	32
3.4 Summary and perspective . . . . .	56

<b>4</b>	<b>Invariom point charges</b>	<b>59</b>
4.1	Introduction . . . . .	59
4.1.1	Motivation . . . . .	59
4.1.2	Test set of angiogenesis inhibitor molecules . . . . .	60
4.1.3	Point charges . . . . .	60
4.2	Methods . . . . .	65
4.2.1	Invariom point charges . . . . .	65
4.2.2	Electrostatic potential . . . . .	67
4.2.3	Programs for application . . . . .	69
4.3	Experiments and Results . . . . .	73
4.3.1	Improving classification of O2c, the example of three homotriptides	73
4.3.2	Internal self-consistency test . . . . .	75
4.3.3	Charge averaging for the whole database – a statistical perspective	78
4.3.4	ESP uncertainties caused by point-charge deviations . . . . .	81
4.3.5	Comparison of molecular ESP . . . . .	84
4.3.6	Application examples . . . . .	88
4.4	Summary and discussion . . . . .	91
<b>5</b>	<b>Invarioms and aspherical modeling for SHELXL – bond-oriented deformation density</b>	<b>93</b>
5.1	Introduction . . . . .	93
5.2	The concept of bond-oriented deformation density (BODD) . . . . .	94
5.2.1	Bonding electron density (BEDE) . . . . .	94
5.2.2	Lone pair electron density (LONE) . . . . .	95
5.2.3	Density of pi-bonds . . . . .	98
5.2.4	Overall goal . . . . .	98
5.2.5	Comparison to previous modeling of aspherical density by dummy atoms . . . . .	99
5.3	Experiments and Results . . . . .	101
5.3.1	A first look at BODD in SHELXL and comparison to the multipole model in XD . . . . .	101
5.3.2	Chemically meaningful parameters . . . . .	104
5.3.3	Transferability study I: aromatic carbon . . . . .	106
5.3.4	Transferability study II: from several small molecules to a larger one	109
5.3.5	Transfer to and refinement of a structure against experimental data	119
5.3.6	Refining BEDE and LONE parameters against experimental data .	124
5.3.7	Comparison of bond-lengths uncertainties and displacement . . . .	126
5.3.8	Comparison to a model refined against neutron data . . . . .	127
5.3.9	Statistical study . . . . .	129
5.3.10	Application to identifying metal atoms . . . . .	135
5.4	Discussion . . . . .	136
5.4.1	Summary . . . . .	136
5.4.2	Perspective . . . . .	137

<b>6 Conclusion</b>	<b>139</b>
6.1 Summary . . . . .	139
6.2 Outlook . . . . .	141
<b>7 Appendix</b>	<b>143</b>
7.1 Appendix A – Identification of metal atoms . . . . .	143
7.2 Appendix B – Invariom point charges . . . . .	146
7.3 Appendix C – Invarioms and aspherical modeling for SHELXL refinements .	152
7.4 Appendix D – List of programs . . . . .	162
<b>Bibliography</b>	<b>167</b>
<b>List of Figures</b>	<b>183</b>
<b>Acknowledgments</b>	<b>187</b>
<b>Scientific contributions</b>	<b>189</b>
<b>Curriculum Vitae</b>	<b>191</b>



# List of abbreviations

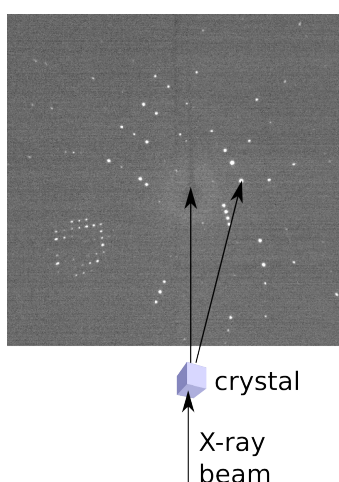
ADP	anisotropic displacement parameters
AIM	Atoms In Molecules
asu	asymmetric unit
au	atomic units
BCC	Bond-Charge Correction
BODD	Bond Oriented Deformation Density
BEDE	bonding electron density
DFT	Density Functional Theory
CSD	Cambridge Structural Database
e	electrons
e	elementary charge (equal to the electric charge of a proton.)
ED	electron density
EDD	electron density distribution
EN	electronegativities
ESP	electrostatic potential
F	structure factor
FT	<i>Fourier</i> transformation
GID	Generalized Invariom Database
HF	Hartree-Fock
hs	high-spin
IAM	independent atom model
invariom	invariant atom
IUCr	International Union of Crystallography

JT	Jahn-Teller
ls	low-spin
LONE	lone pair electron density
MBADNP	methylbenzylaminodinitropyridine
MK	<i>Merz and Kolman</i>
NN	nearest neighbor atoms
NNN	next-nearest neighbor atoms
ORTEP	Oak Ridge Thermal-Ellipsoid Plot
QM	quantum mechanics
R	residual factor
RESP	restrained fit to the electrostatic potential
$\rho$	electron density function
RMS	root-mean square difference
RRMS	relative root-mean square difference
SCF	self-consistent field
STDEV	standard deviation
TPACM4	Transferable Partial Atomic Charge Model
UBDB	University of Buffalo DataBase
XL	SHELXL
XRD	X-ray diffraction

# 1 General Information and Theoretical Background

## 1.1 Crystallography

Single crystal X-ray structure determinations provide highly reliable evidence for atomic connectivity and chirality of new molecules. Its popularity in chemistry is explained by the method's power to elucidate even unexpected three-dimensional molecular structures with great certainty.



**Figure 1.1:** Basic concept of single crystal X-ray diffraction.

In 1912 Max von Laue first thought about the diffraction of X-rays by crystals.<sup>[1]</sup> While he primarily discovered the phenomenon itself, the discovery led William Lawrence Bragg and his father William Henry Bragg to solve the first crystal structure in 1915.<sup>[2]</sup> Consequently it has been possible to determine distances between atoms in minerals for more than 100 years. Subsequently, crystals of more complex salts and molecular crystals were investigated by single crystal X-ray diffraction (XRD) providing atomic coordinates.<sup>[3]</sup> The development for interpretation of the reflection data by *Fourier* series around 1930<sup>[4]</sup> laid the grounds for the assignment of elements based on the intensity of reflections. Precise intensity measurements additionally allowed the modeling of atomic displacements in crystals that were attributed to vibration or disorder.<sup>[2]</sup> At present, it is possible to determine protein structures of molecules as heavy as 98 Megadalton.<sup>[5]</sup>

### 1.1.1 Reflection intensities

The interference of light waves with a point lattice is described by the *Bragg* equation. The X-rays, however, interact with electrons, which are not located at grid positions in a real crystal; they are distributed around the atomic nuclei. The atoms themselves are usually dispersed within the unit cell.<sup>1</sup> These deviations from an ideal point grid lead to additional interference, reducing the intensity of the reflections. The higher the reflection order the stronger this effect becomes. Since the intensities are proportional to the square of the structure factor ( $F$ ) at scattering vector  $\vec{s}$ , the intensities are related to the sum of

---

<sup>1</sup>except for crystals of metals

scattering contributions  $f$  (atomic form factor) of each atom  $j$  in a crystal's unit cell:

$$F(\vec{s}) = \sum_j^N f_j(\vec{s}) \exp(2\pi i \vec{s} \cdot \vec{r}_j) , \quad (1.1)$$

where  $f_j$  is the atomic form factor and  $\vec{s}$  is the vector from the origin of the reciprocal lattice to a reflection, and  $\vec{r}_j$  is the position of the atom  $j$  in the unit cell.

The structure factor phases  $i\vec{s} \cdot \vec{r}_j$  are not measurable, because the complex part of  $F$  is lost during squaring of  $F$  for obtaining reflection intensities. Therefore, the information that can be analytically deduced from the experiment is not complete. This phenomenon is known as the crystallographic phase problem. In every structure determination there is a step called 'phasing' or 'solving a structure' which addresses this problem. Once the problem is solved and a good set of starting phases has been determined, further refinement of the structure model, atomic positions and atom types, follows.

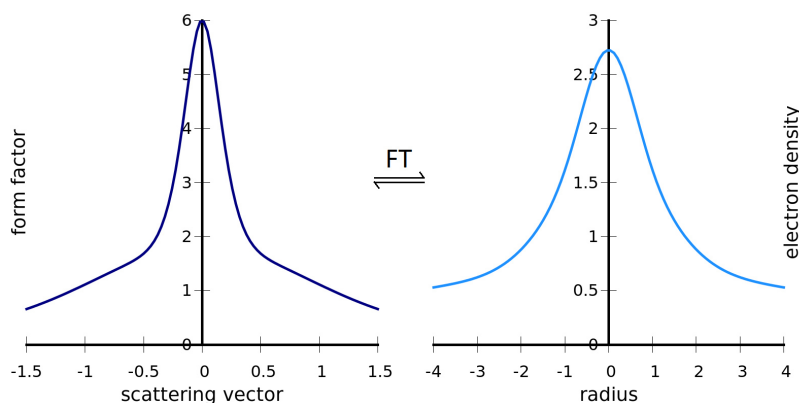
### 1.1.1.1 Atomic form factor

This leaves the question how the atomic form factor is described. Since single crystal XRD is mathematically described by *Fourier* transformation<sup>2</sup> (FT) the atomic scattering factor relates to the electron density (ED) in the following way:

$$f(\vec{s}) = \text{FT}^{-1}[\rho(\vec{r})] = \int_V \rho(\vec{r}) \exp(2\pi i \vec{s} \cdot \vec{r}) d\vec{r} \quad (1.2)$$

and is visualized in Figure 1.2. Most frequently  $\rho(\vec{r})$  is an element-specific, spherically symmetric function derived from isolated and independent atoms in the gas phase. Application of the *Fourier* transform of those functions for the ED in order to model the crystallographic data is known as the independent atom model (IAM).

Additionally  $f_j(\vec{s})$  includes wavelength ( $\lambda$ ) dependent contributions for anomalous scat-



**Figure 1.2:** *Fourier* transformation of the spherical ED of carbon to give its atomic form factor.

<sup>2</sup>This is only true if detectors are far from the diffracting crystal in order to observe *Fraunhofer* diffraction, but is always the case for common diffraction experiments.<sup>[6]</sup>



tering<sup>[7-9]</sup>:

$$f_j(\vec{s}, \lambda) = f(\vec{s}) + f'(\lambda) + f''(\lambda) \quad (1.3)$$

in which  $f'$  and  $f''$  are the real and imaginary dispersion corrections to the form factor.

### 1.1.1.2 Displacement parameters

With the advancement of methods and measurement techniques, displacements of atoms from their mean positions in the crystal lattice were observed and included in the model. The factor describing thermal motion and disorder of atoms is the *Debye-Waller* factor  $T$ :

$$T = e^{-B(\frac{\sin(\Theta)}{\lambda})^2} \quad \text{with } B = 8\pi^2 U \quad (1.4)$$

where  $U$  is the displacement factor in  $\text{\AA}^2$ . Thus, an extended description of an atomic scattering factor is given by:

$$f_j^B(\vec{s}) = f_j(\vec{s}) \cdot e^{-8\pi^2 U(\frac{\sin(\Theta)}{\lambda})^2} \quad (1.5)$$

$U$  can be isotropic or anisotropic, in the latter case  $U$  is a symmetric 3x3 tensor with six independent parameters  $U_{ij}$ . If the reflection data to parameter ratio permits, the six anisotropic displacement parameters (ADP) are usually refined for non-hydrogen atoms. The relatively high number of parameters refined for anisotropic modeling of the displacement renders them susceptible to shortcomings of data and model. Thus, they are also indicators for the quality of a structure refinement.

## 1.1.2 Least-squares refinement

Structure refinement means the adjustment of model parameters to optimize the agreement of structure factors calculated from the model ( $F_c$ ) and observed ones ( $F_o$ ). The optimized parameters usually include atomic coordinates, atomic displacement parameters, the overall scale factor and site occupancy factors for disordered parts of a molecule. In small-molecule crystallography refinement is achieved by minimizing the sum of the squared differences for each reflection.<sup>[9,10]</sup> The method for determination of the smallest squared differences is called least-squares fitting. Differences are calculated from either structure factor amplitudes ( $F$ ) or their square ( $F^2$ ).

Each squared difference can be weighted statistically by  $w = \frac{1}{\sigma^2}$ , where  $\sigma$  is the error of each reflection. This only applies to refinements against  $F^2$ . Weighting factors for refinements against  $F$  are more difficult to estimate<sup>[10]</sup> and negative values of  $F$  have to be set to zero or small positive values. Therefore, at present most refinement programs recommend refinements against  $F^2$ , in which

$$M = \sum w(F_o^2 - F_c^2) \quad (1.6)$$

is minimized.<sup>[8,10]</sup> During this thesis least-squares refinements were performed by two programs: SHELXL<sup>[11]</sup> and XD<sup>[12]</sup>. Mathematical details on the minimization of the non-linear least-squares problem can be found in the textbook by Giacovazzo *et al.*<sup>[8]</sup>

Refinements are commonly iterative procedures in which each cycle should improve the fit of the model to the data. Each new model supplies phases for the observed structure factors allowing computation of ED maps derived from the data and comparing them to those completely derived from the model. The difference can be visualized and is called difference density or residual density.

The most critical point of a refinement is a descriptor of model quality to judge if changes are an improvement. While internally  $M$  is the number which is optimized, other residual factors provide more comparable numbers to judge model quality.

### 1.1.3 Residual factor

By far the most popular way to assess the quality of a structural model is a descriptor for the difference between calculated and observed structure factors,  $F_c$  and  $F_o$ , known as residual factor, R1 and R(F).<sup>[10]</sup> It is based on the absolute, unweighted values of  $F$ :

$$R1 = R(F) = \frac{\sum ||F_o| - |F_c||}{\sum |F_o|} \quad (1.7)$$

The selection of reflections for this comparison influences the results considerably and should therefore be referenced. In any case, the smaller the value of R1 the better the fit of the model to the data. Values for R1 of organic molecules are usually higher than 1.8 % due to limitations of the IAM used for calculating  $F_c$ .

Hence, a more accurate description of the electron density distribution(EDD) can improve the fit to high-quality data. *Vice versa* can high-quality data yield information about the EDD. More realistic descriptions of the EDD are one topic of charge density research.

## 1.2 Charge density

This branch of crystallography is interested in, among other topics, the best possible description of the EDD of crystal structures under optimal experimental conditions. The difference between charge density and electron density (ED) lies in the inclusion of nuclear displacements in experimental studies of the charge density compared to an evaluation of a rather static, theoretical EDD. In principle, charge density studies yield EDD from experiment that are complimentary to those from theory and, hence, are compared frequently.<sup>[13]</sup> A requirement for such a comparison is a thorough evaluation of the experimental data quality. Therefore, one aim of charge density research is to minimize errors in the X-ray data collection and to reduce the data as well as possible<sup>[14-16]</sup> in order to be able trust the data in comparison to theoretical results.<sup>[17]</sup>

The descriptors used to access the resulting EDD of charge density studies are diverse: the theory of Atoms In Molecules (AIM)<sup>[18]</sup>, source function<sup>[19,20]</sup> and electron localization index (ELI-D)<sup>[21-23]</sup> are only some common concepts. They are usually applied to better understand the bonding situation of the molecules studied. Reviews<sup>[24-28]</sup> give an overview of the diversity of molecules as well as questions addressed and answered by charge density. This thesis's topic is mostly related to the models applied in charge density research and therefore they will be the focus in the following sections.

## 1.2.1 Different charge density models

Several charge density models were developed throughout the 1970's; the multipole model prevails and will be discussed later. Initially the focus was on simple point charge models.

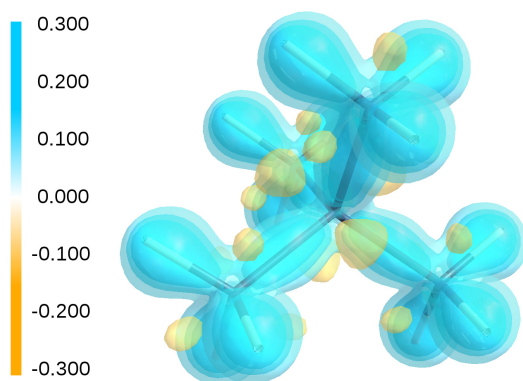
### 1.2.1.1 Dummy bond atoms

As early as 1960<sup>[29]</sup> the bonding ED was included in a model for XRD data by placing half an electron in the middle of a bond in diamond, as suggested theoretically by Ewald and Hönl in 1936.<sup>[30]</sup> Hellner additionally introduced a division of the structure factor into contributions from core electrons, bonding and non-bonding valence electrons in 1977.<sup>[31]</sup> The valence electrons were modeled by three dimensional *Gaussian* functions, also called 'charge clouds', which were placed in the middle of each bond and inherited  $B_{ij}$  parameters of the next two atoms. Next, populations and ADP were refined for each 'charge cloud'. Comparisons to theoretical calculations were presented together with Scheringer.<sup>[32]</sup> Within the same publication difference density maps were described, where the IAM ED was subtracted from the molecular ED, which later became known as the deformation density map (Figure 1.3).<sup>[33,34]</sup>

Around 1980<sup>[35]</sup> a comparison to the multipole model (see Section 1.2.1.5) showed advantages for models including octupoles and hexadecapoles especially for bonds to heteroatoms. Still, the simplicity of the model is appealing and many years later, in 2004 and 2007, attempts to model the bonding ED of proteins by 'dummy bond atoms' were presented.<sup>[36,37]</sup> Moreover, Jelsch *et al.* recently investigated modeling the bonding and lone pair ED by dummy atoms with respect to the electrostatic potential and also compared to dummy bond parameters transferred from a database.<sup>[38,39]</sup>

### 1.2.1.2 Bader's quantum theory of Atoms In Molecules

The quantum theory of Atoms In Molecules (AIM)<sup>[18]</sup> is more a method for analyzing an obtained molecular EDD than modeling it. Similar to the other models in charge density it contains a formalism to partition the molecular ED into atomic contributions. Here those contributions are called atomic basins and they are deduced from first and second derivatives of the EDD with respect to spacial changes in all directions. Such an analysis yields bond and other critical points, atomic volumes, atomic charges and valence shell charge



**Figure 1.3:** Deformation density of isopentane, obtained by subtracting the IAM EDD from the molecular EDD modeled by multipoles. Calculation with XDPROP<sup>[12]</sup> and visualization with MOLECOOLQT.<sup>[40]</sup>

concentrations as descriptors of EDD. Thereby existence, localization and characterization of bonds and non-bonding free electron pairs can be investigated.<sup>[28]</sup>

### 1.2.1.3 Hirshfeld atoms

Hirshfeld proposed to partition<sup>[41]</sup> the molecular ED by a stockholder approach, in which the atomic density  $\rho_a^{b.a.}$  of a bonded atom (*b.a.*) is given by the stock  $w_a$  it holds of the molecular ED,  $\rho^{mol}(\vec{r})$ , at a given point  $\vec{r}$ :

$$\rho_a^{b.a.}(\vec{r}) = w_a(\vec{r})\rho^{mol}(\vec{r}) . \quad (1.8)$$

The share at a given point,  $w_a(\vec{r})$ , corresponds to the part of the promolecular ED,  $\rho^{pro}(\vec{r})$ , that is modeled by the atomic ED  $\rho^{at}(\vec{r})$  from the IAM:

$$w_a(\vec{r}) = \frac{\rho_a^{at}(\vec{r})}{\rho^{pro}(\vec{r})} \quad (1.9)$$

where the promolecular ED  $\rho^{pro}(\vec{r})$  is the sum of all atomic IAM ED contributions,  $\rho^{at}(\vec{r})$ :

$$\rho^{pro}(\vec{r}) = \sum_b \rho_b^{at}(\vec{r}) . \quad (1.10)$$

Here  $b$  accounts for all atoms in a molecule, including  $a$ . Partitioning of the real molecular ED by these stocks yields aspherical bonded atomic EDD,  $\rho_a^{b.a.}$ .

This model works well and is also applied in order to create *Hirshfeld* surfaces for display and analysis by partitioning crystal structures into molecular contributions instead of molecular ones into atomic parts. A *Hirshfeld* surface is defined by a fixed  $w_a(\vec{r})$  of usually 0.5.<sup>[42,43]</sup>

Such a '*Hirshfeld* atom refinement' starts with a preliminary structure refinement. Next the ED of this structure model is calculated by quantum mechanics to obtain the aspherical ED of the bonded atoms,  $\rho_a^{b.a.}$ . Positions and displacements of those atomic ED fragments can then be refined against the diffraction data.<sup>[44]</sup> If  $\rho_a^{b.a.}$  is iteratively obtained for each newly refined model, such a refinement leads to results in good agreement with those from neutron studies.<sup>[45]</sup> Limits are larger molecules or many fragments in the asymmetric unit due to calculation times required. Treatment of polymeric structures is especially challenging and still under development.

### 1.2.1.4 X-ray constrained wave function fitting

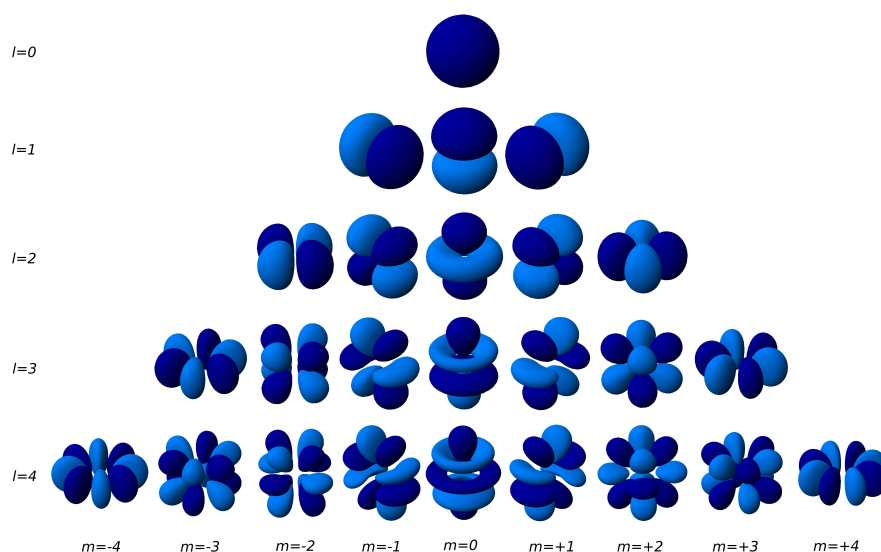
Developed by the same researchers as '*Hirshfeld* atom refinement', 'X-ray constrained wave function fitting' is a method that works with the molecular ED directly, without partitioning. Wave-function coefficients are optimized in order to obtain the wave-function fitting  $F_o$  that at the same time minimizes the energy of a quantum mechanics (QM) calculation.<sup>[46]</sup> Building upon this work of Jayatilaka *et al.*, Genoni restrained the molecular orbital's extension.<sup>[47]</sup> The by this method obtained 'extremely localized molecular orbitals' (ELMOs) were assembled in a database,<sup>[48]</sup> whose transferability was shown to be similar to those of 'pseudoatom' databases<sup>[49]</sup> based on the multipole model.

### 1.2.1.5 Multipole formalism

At present, the *Hansen* and *Coppens* multipole model<sup>[50]</sup> is the most frequently applied method to include asphericity in structure models. With this model the electron density distribution (EDD) of an atom is described via a multipolar expansion for the deformation of the valence shell:

$$\rho_{\text{atom}}(\vec{r}) = P_c \rho_{\text{core}}(\vec{r}) + P_v \kappa^3 \rho_{\text{val}}(\kappa \vec{r}) + \sum_{l=0}^{l_{\text{max}}} (\kappa')^3 R_l(\kappa' \vec{r}) \cdot \sum_{m=-l}^l P_{lm} Y_{lm}(\vartheta, \varphi). \quad (1.11)$$

$P$  represents the population of a special ED function  $\rho$ . The ED is split into three parts: core density, valence density and spherical harmonics  $Y_{lm}$  (Figure 1.4) multiplied by a radial function  $R_l$ . As an extension of the earlier model by Stewart,<sup>[51]</sup> the valence density and the radial function are each multiplied by a contraction parameter  $\kappa$ . For  $\kappa$  higher than one, a function is contracted. The higher  $l$  the more multipoles are included in the model, therefore if  $l = 0$ ,  $\kappa = 1$  and  $P_{\text{val}}$  equal to the number of valence electrons, the ED is practically identical to the IAM. Due to the spatial overlap of atoms in close proximity to each other the method is a way of partitioning molecular ED into 'fuzzy' atoms,<sup>[24]</sup> which differentiates it from AIM theory. Occasionally, description by multipoles is referred to as 'pseudoatom' model because the center of the fragment is still at atomic positions.



**Figure 1.4:** Representation of spherical harmonics up to  $l = 4$ . Created with Orbital Viewer.<sup>[52]</sup>

A big advantage of the multipole formalism is that the corresponding expression in reciprocal space is similar and includes the same populations:

$$f_j(\vec{s}) = P_{j,c} f_{j,\text{core}}(\vec{s}) + P_{j,v} \kappa^3 f_{j,\text{val}}(\vec{s}/\kappa) + \sum_{l=0}^{l_{\text{max}}} \sum_{m=-l}^l P_{lm} f_{lm}(\vec{s}/\kappa'). \quad (1.12)$$

The multipole scattering factors,  $f_{lm}(\vec{s})$ , are the orientation-dependent *Fourier-Bessel* transforms of the spherical harmonic functions.<sup>[53,54]</sup>

The coordinate systems of the spherical harmonics should be oriented in such a way that an atom's bonding situation and its symmetry can be described best. Usually only those multipole populations are allowed to be non-zero that are reasonable for a given symmetry on an atom's bonding environment. Similar atoms can be treated with the same populations in order to save parameters. Still, the multipole model introduces many new parameters, which can easily double their total number. Therefore, high resolution data ( $d < 0.50 \text{ \AA}$ ) are necessary for refinement of multipole populations. For the description of structures with data of lower resolution, multipole parameters have been tabulated in databases and can be transferred to a structure of interest. This way atomic asphericity is described by multipole parameters, but since they are fixed to their transferred values, the number of parameters refined is the same as for the corresponding IAM refinement.

### 1.2.2 Databases for transferable multipole parameters

Transfer of aspherical scattering factors between different structures refined against experimental data was first mentioned by Brock *et al.* in 1991.<sup>[55]</sup> Transfer from refinements against theoretical data was reported in 2002.<sup>[56]</sup> Building upon these ideas several experimental<sup>[57–60]</sup> and theoretical<sup>[61,62]</sup> databases of aspherical scattering factors were developed. In the beginning they mainly contained transferable aspherical atomic ED fragments for oligopeptides while later versions extended coverage to other organic structure motifs.

In addition to allowing an improved model without adding parameters to the refinement, the advantages of libraries for the description of asphericity include:

- deconvolution of thermal motion from valence ED,<sup>[57,59,60,63–67]</sup>
- improvement of the discrepancy R factors,
- more precise and accurate molecular geometry,<sup>[61,64,68,69]</sup>
- increasing precision of the *Flack* parameter<sup>[70]</sup> for absolute structure determination.<sup>[71]</sup>

Advantages and disadvantages of experimental versus theoretical databases have been discussed controversially.<sup>[72,73]</sup> The elimination of experimental errors and the uncomplicated addition of seemingly unlimited types of pseudoatoms is the biggest advantage of theoretical databases.<sup>[63]</sup> Detailed comparisons of different libraries for pseudoatom models have been presented.<sup>[74,75]</sup>

The database most similar to the invariom database is the University of Buffalo DataBase (UBDB).<sup>[76]</sup> While the UBDB averages over a "*family of chemically unique pseudoatoms derived from the theoretical densities of a number of small molecules*"<sup>[63]</sup>, the multipole parameters of a given invariom are taken from a unique model compound. Additionally, geometries for invariom model compounds are optimized by Density Functional Theory (DFT), whereas for UBDB DFT single-point calculations are performed for molecules from the Cambridge Structural Database (CSD).<sup>[77]</sup> In both cases the ED from the theoretical computations is projected onto the multipole model. In the UBDB new atom-types are 'spawned', depending on deviations during averaging of the multipole populations,<sup>[78]</sup>

whereas the invariom classification is based upon established empirical rules for transferability without averaging. This latter approach has the advantage that previous entries are not changed when new compounds are added to the library.

Invariom fragments have been shown to reproduce the molecular ED within  $0.05 \text{ e } \text{\AA}^{-3}$  and molecular electrostatic potentials could be reproduced equally well from those invariom pseudoatoms.<sup>[79]</sup> The invariom database, its extension and new applications are the focus of this thesis.

## 1.3 The Invariom Concept

The invariom database is a collection of small organic compounds, so called model compounds, whose EDD was projected onto the multipole model. From which model compound the multipole populations are transferred to a molecule of interest is decided for each atom individually. The decision is derived from an atom's bonding environment.

### 1.3.1 Assignment formalism

Scattering-factor assignment is based on the principle that pseudoatoms are invariant upon transfer between different molecules, if an atom has the same neighbors and connections to them in both molecules. This is where the name invariom comes from, it is a short form for '**invariant atoms**'.<sup>[80]</sup> Hence, atomic bonding environments need to be classified. As mentioned in Section 1.2.2, the scattering factor for an invariom is derived from one molecule. Since many molecules in the database can have atoms in equal bonding situations and therefore the same invariom classification, criteria to decide which of the model compounds the scattering factor is transferred from have been established.

#### 1.3.1.1 Criteria for invariom model compounds

Each scattering factor comes from the model compound which also contains the invariom. If more than one model compound contains an invariom there are priority rules which prefer

- the smaller molecule (with a lower number of atoms),
- the one with the most hydrogen and then carbon atoms, preferring oxygen before nitrogen before boron substituents,
- neutral molecules,
- the lower dipole moment
- and if these criteria do not suffice the lower energy of the self-consistent-field computation.

Hence, there is always exactly one model compound associated to an invariom name for scattering factors. This association is illustrated in Figure 1.5.

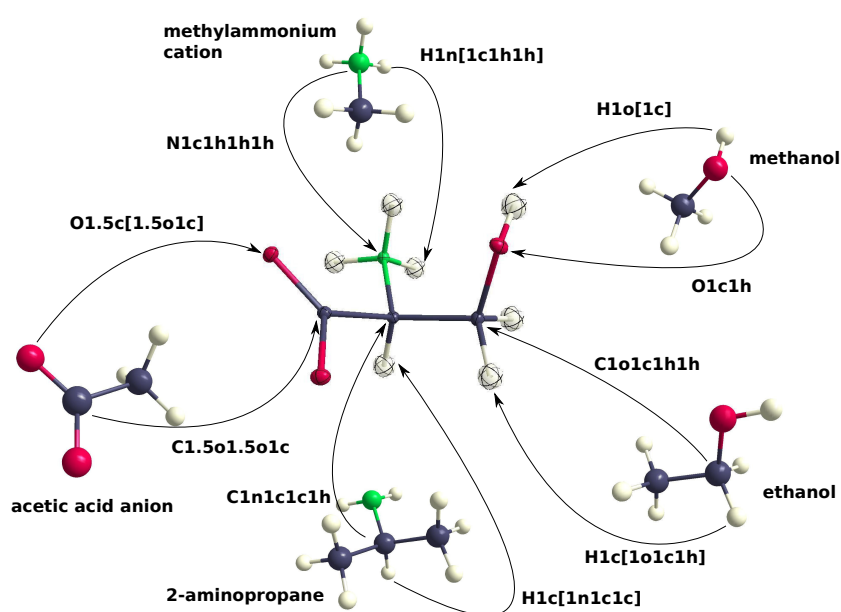


Figure 1.5: Invariom assignment for serine with invariom names and model compounds.

### 1.3.1.2 Invariom Classification

The classification of an atom's bonding environment into invarioms means that as equally classified atoms within the nearest-neighbor approximation have scattering factors which are invariant upon transfer between different molecules. The connectivity of an atom is characterized by an invariom name, sometimes also referred to as invariom string.<sup>[81]</sup> Invariom names commence with the element of the atom of interest, and contain the element of and bond type to neighboring atoms. Bonds are categorized into single, delocalized, double and triple bonds. Those descriptors are assigned on the basis of the bond-distinguishing parameter,  $\chi$ .<sup>[82,83]</sup> This characteristic number is derived from bond lengths,  $d$ , relative to the atomic covalent radii,  $r_c$ , corrected by a term for electronegativities (EN) according to Allred and Rochow:<sup>[84]</sup>

$$\chi = [r_c(\text{atom1}) + r_c(\text{atom2}) - 0.08 \cdot |\Delta(\text{EN})|] - d \quad (1.13)$$

Exemplary invariom names are shown for the amino acid serine in Figure 1.5.

Later developments<sup>[81]</sup> led to an inclusion of ring size in the invariom name for aromatic systems and next-neighbor environments for aromatic and double bonds. Delocalization in rings is indicated by a '#' and the number of atoms in the planar ring behind it. If an atom is part of two condensed delocalized rings (as in naphthalene), both numbers follow the '#'. Bond orders for the connection between an atom and an aromatic ring are replaced by an '@' and the ring size<sup>3</sup>.

Moreover, planar and non-planar nitrogen atoms are distinguished in the notation by an '=' in front of the invariom name. Their planarity is evaluated by:

$$PV_{I=3} = (\vec{n}_1 \times \vec{n}_2) \cdot (\vec{n}_1 \times \vec{n}_3) \cdot (\vec{n}_2 \times \vec{n}_3) \quad (1.14)$$

<sup>3</sup>An example of this extended notation is shown in Figure 5.15.



yielding values of one for three planar normalized bond vectors  $\vec{n}$ . If  $PV_{I=3} < 0.8$  the three bond vectors are not considered planar.

Invarioms can be chiral (although this occurs less frequently than chiral atoms appear, due to the limited shell of atoms considered). In this case the invariom name is preceded by an 'R-' or 'S-' according to Cahn, Ingold and Prelog.<sup>[85]</sup>

Whenever necessary the new invariom names include a second sphere of neighbor atoms in brackets behind the atom they are connected to. The special role of hydrogen atoms is discussed in Section 4.2.1.3. As early as 2004<sup>[80]</sup> next-nearest neighbors were included in invariom names and model compounds for atoms involved in mesomeric systems and delocalized double bonds.

### 1.3.2 Treatment of invariom model compounds

All model compounds of the invariom database are set up in the same way. In the beginning their geometry is optimized and the molecular EDD obtained from the wave function, which then is placed in an artificial unit cell with  $a = b = c$  and  $\alpha = \beta = \gamma = 90^\circ$  of space group  $P\bar{1}$  in order to yield theoretical diffraction data upon *Fourier* transformation.<sup>[86,87]</sup> Finally, the multipole parameters are refined against those theoretical diffraction data as illustrated in Figure 1.6.

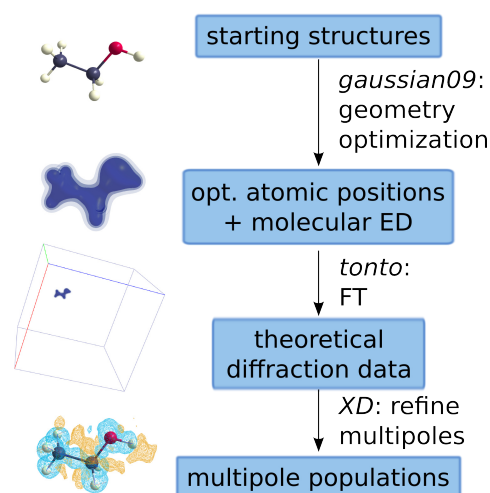


Figure 1.6: General treatment of invariom model compounds.

### 1.3.3 Invariom history

#### 1.3.3.1 Presentation of idea, first tests and a first database

In the first years of the invariom method<sup>[80]</sup> the theoretical ED of model compounds was computed with B3LYP/6-311++G(3df,3pd). When demonstrating the invariom concept for the examples of zwitterionic tri(L-valine) and terbogrel, rescaling of monopole populations after multipole parameter transfer to achieve neutrality (Section 1.3.3.2) was mentioned. In 2005 Dittrich *et al.*<sup>[88]</sup> described the procedure in more detail and investigated resolution and temperature dependence of invariom, IAM and refined multipole models of D,L-serine. Additionally, different basis sets for the theoretical ED were compared. The cubic cells used for generating the theoretical structure factors in the database had a cell length of 15 Å and the average resolution of the theoretical data was  $\sin(\theta)/\lambda = 1.15 \text{ \AA}^{-1}$ . Populations were refined for multipoles up to the hexadecapole level. In contrast to today's invariom names, they included a '+' or '-' at the end indicating the charge of the corresponding model compound.

One year later an invariom database for the description of peptides was released.<sup>[61]</sup> The publication explained the bond-distinguishing parameter closely and presented a list of the

37 model compounds necessary to describe all possible proteinogenic amino-acids and their combination in peptides. The database was validated by comparing refinement results for four amino-acid structures. Furthermore these refinements showed that the improvement with respect to the IAM model is more pronounced in low temperature studies and for higher resolution. Residual factors as well as residual densities could be improved for all the structures investigated.

Another investigation in the same year highlighted the improvement of absolute structure determination for compounds containing only light atoms<sup>[71]</sup> upon modeling with invariom scattering factors. Invarioms improved the *Flack* parameter or its standard uncertainty.<sup>[70]</sup> In contrast, the newer parameter from *Parsons'* quotients,<sup>[89]</sup> which is similar to the traditional *Flack* parameter, is less affected by the lack of a model for the bonding ED. Simultaneously the publication introduced that model compounds and thereby the invariom name of hydrogen atoms were to include next-nearest neighbor atoms. Furthermore, the basis set changed to D95++(3df,3pd)<sup>[90]</sup> and the theoretical cubic cell was increased to 30 Å.

### 1.3.3.2 InvariomTool, electroneutrality and further experiences

INVARIOMTOOL, a preprocessor program for invariom assignment, was introduced via a separate publication in 2007<sup>[91]</sup> although it was used since 2004 in development versions. The program automates the invariom density modeling by assigning each atom in a structure an invariom name and transferring the corresponding entry of the database, with all the multipole populations, to the molecule of interest. Neutrality and orientation of the local atomic coordinate systems are handled automatically. In order to achieve neutrality several procedures are implemented:

1. Addition of the average difference between the sum of transferred monopole populations and the sum of neutral monopole populations, divided by the number of all atoms, to each atom.
2. The same as before but only for hydrogen atoms.
3. The difference per atom is weighted by the difference of the EN of the element from the average EN. In case of a charge deficiency an atom with a higher EN is added a higher amount of (negative) charge than an atom of lower EN.

While the last option was recommended for the calculation of dipoles, the default treatment for organic molecules containing a considerable number of hydrogen atoms is, until today, option two.

Further studies emphasized the advantages of invariom-model refinements: geometries of invariom refinements resembled those of multipole refinements, notably the bond lengths to hydrogen atoms<sup>[69]</sup> and ADP were less contaminated by the aspherical density.<sup>[65]</sup> Considering the suitability factor of Stevens and Coppens<sup>[92]</sup> a lower decrease in the residual factor was observed and expected for changes from IAM to invariom models if "heavier elements" (in this case chloride) were part of the structure,<sup>[69]</sup> since their core electron description is unchanged in the model and has a considerable influence.

### 1.3.3.3 The generalized invariom database

A great leap forward was the development of the Generalized Invariom Database (GID).<sup>[81]</sup> Mainly the scattering factor notation was improved to ease treatment of aromatic systems including condensed rings and heteroaromatic structures, partly based on previous comparative studies.<sup>[93]</sup> The GID was introduced as extension of the earlier invariom database for peptides.<sup>[61]</sup> Thenceforward, the scattering factor formalism included description of three-membered rings and extended delocalized ring systems.

In the new database for atoms heavier than carbon  $\kappa'$  was refined in addition to  $\kappa$ . Moreover, it was stated: 'We now use a full sphere of data up to a resolution of  $1.2 \text{ \AA}^{-1}$ .', but considering that only half a sphere would be enough and that most data range for  $h$  and  $k$  from -40 to 40 and from 0 to 40 for  $l$  this was probably merely a plan at the time. Element coverage expanded to third row elements, which was especially useful for phosphorus, sulfur and chlorine compounds regularly encountered in organic chemistry.

Possible applications for calculations of electrostatic potentials, electrostatic interaction energies and molecular dipole moments were suggested and had been presented shortly before.<sup>[79]</sup> The GID had and has been applied in several studies of more general small molecules<sup>[94-97]</sup> and simultaneously the first macromolecular compound was modelled by invarioms.<sup>[98]</sup>

In order to treat metal-organic compounds a method involving molecule-specific databases of scattering factors was discussed and finally presented in 2015.<sup>[99]</sup> This last project includes some work of this thesis and the therein presented procedure will be discussed in detail in Chapter 3.2.

## 1.3.4 New invariom developments parallel to this thesis

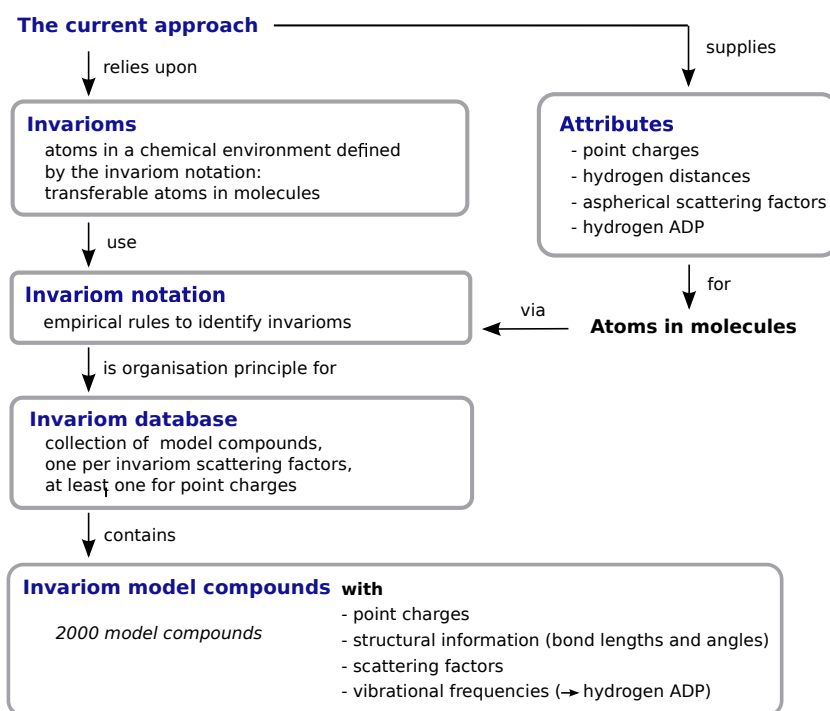
### 1.3.4.1 From extended functionality to more precise terminology

During this thesis new attributes (Section 1.4) were deduced from the database and assigned on the basis of invarioms. Therefore the terminology around the invariom concept had to be defined more precisely. Figure 1.7 shows an overview of the new invariom functionality and its vocabulary.

So far, an invariom corresponded to an ED fragment of an atom in a specific bonding environment which was transferable between a specific model compound and all other compounds with the same bonding environment. This environment was characterized by the invariom name. Therefore, an invariom as defined by the invariom string could be associated to one model compound, a set of multipole population parameters and/or just an atom in a specific bonding environment.

With the extension of the database to new applications, an invariom does not necessarily correspond to an ED fragment anymore. The ED fragment is only one of many properties which are assigned to an invariom. The invariom is merely an atom in a defined chemical bonding environment.

Now the term 'invariom database' means the collection of model compounds in contrast to the individual attribute tables that assign attribute values to each invariom name. The tables of attributes may be described as 'invariom charge database', 'invariom scattering



**Figure 1.7:** Overview of the invariom database, its functionality and terminology.<sup>[100]</sup> ©reprinted with permission of Wiley.

factor database', 'invariom hydrogen ADP' and 'hydrogen distance database'. These are the four attributes which can be derived from the invariom database at the moment.

#### 1.3.4.2 Invariom association to model compound

The unambiguous relation between invariom name and invariom model compound so far does not extend to all new attributes, although it is a big advantage of the invariom formalism for scattering factors and one distinguishing characteristic of the invariom scattering factors compared to other libraries. Unfortunately, as discussed during this thesis, invariom point charges benefit considerably from averaging this invariom attribute for several model compounds. So, for this special property an invariom is not directly associated to one model compound. For all other invariom attribute databases, though, the original assignment still holds true.

## 1.4 Scope of this thesis

The first aim of this thesis is the complete renewal of the database with a basis set that extends to elements of the fourth period. Next to including bromine compounds in the database, this will allow homogeneous treatment of all atoms in a metal-organic crystal structures by an invariom-like approach and the addition of further attributes to the database.

The renewed database had found its first application in a project to identify metal atoms in coordination compounds.<sup>[99]</sup> The therein developed method will be the basis for case studies in this thesis on crystal structure pairs from the literature. True isomorphism of the structures to be investigated can in some cases be excluded on the basis of the deposited XRD data. For those cases the application of aspherical scattering factors shall identify which of two possible metals is the correct central atom. Additionally, the method will be tested for cases in which the datasets are not identical but similar, so that the question whether two compounds are indeed isomorphous will be investigated by the invariom-like approach. The influence of inferior data quality on the method will be examined, too.

The second major project of this thesis will depend even more on the renewal of the database. Its aim is the addition of point charges as an attribute that can be transferred via the invariom classification from the invariom database. The development of invariom point charges will be one step towards fully automating force-field parameterization for molecular dynamics<sup>[101]</sup> simulations of organic molecules. Such molecular-dynamics simulations will allow a correct inclusion of dynamic disorder in crystallographic models. Moreover, point charges facilitate the way to a representation of the electrostatic potential (ESP). Molecular ESPs from invariom point charges will be compared those from other point charges and methods.

The aim of the last project within this thesis will be a pilot study for incorporating a description of the deformation density in models refined with the program SHELXL. This is a meaningful topic, because it became apparent that the change to a more sophisticated and hence more complex program for inclusion of asphericity is a major hindrance for many scientists and therefore limits wider application. Hence only few researches apply invariom scattering factors or similar descriptions of the bonding and lone pair ED. In order to avoid the most complex and error prone part of the multipole modeling, a bond oriented deformation density will be discussed. If the new model's transferability from the invariom database can be established, the same advantages as for the multipole scattering factor databases will be accessible from the most commonly applied program for structure determination, SHELXL. By improving accessibility, more scientific projects will be able to profit from the invariom database. So the aim of this pilot is to pave the way for a new attribute and thereby another new application of the invariom database.

Overall, the invariom database is renewed, metal atoms in coordination compounds are identified by aspherical scattering factors, invariom point charges are developed and invarioms as well as aspherical modeling for SHELXL are investigated. All of these projects either give an example, introduce or prepare the way for new applications of the invariom database.



## 2 Invariom database renewal

### 2.1 Introduction

In order to extend the properties provided by the invariom database to the four attributes listed in Figure 1.7 each model compound needed additional treatment. While multipole populations and distances to hydrogen atoms are present for each model compound, the atomic charges selected as invariom point charges had to be derived from the molecular wave function. The hydrogen ADP (at least the part due to internal vibration) require vibrational frequencies,<sup>[102]</sup> which can easily be obtained from the *Hessian* matrix, a side product of geometry optimization. The *Hessian* matrix includes the derivatives of the energy with respect to changes of atomic coordinates. Analytical determination of those gradients from a self-consistent field (SCF) calculation at a fixed geometry (single-point calculation) would be an alternative. Due to starting the geometry optimization from an already optimized one of another basis set, changes of atomic coordinates were minimal and not too time-consuming. Reoptimizing the geometry with a new basis set and functional yields a consistent treatment for compounds added due to the new basis set like bromine compounds and those already in the database. Therefore, a new geometry optimization was performed for each of the invariom model compounds.

#### 2.1.1 Geometry Optimization

Since the functional B3LYP used so far is old and specific to its implementation in the program GAUSSIAN,<sup>[103]</sup> it was time to change to a more up to date functional. Differences between scattering factors from different functionals are minimal, though. The Minnesota functional M06<sup>[104]</sup> was developed to cover organic compounds as well as metal-organic coordination compounds. It is as general as B3LYP but more state of the art and in contrast to B3LYP implemented consistently in popular QM programs. M06 is a hybrid meta-exchange correlation functional with 27 % Hartree-Fock (HF) exchange.

The basis set D95++(3df,3pd)<sup>[90]</sup> only allows the computation of elements up to argon for our database. Moreover, computation of the theoretical structure data does not work for effective core potentials. Hence, a new basis set was required in order to be able to include heavier elements like bromine in the database and enable treatment of metal containing compounds with the same basis set as used in the database. Application of the same basis set as for metal-organic compounds permits homogeneous treatment of metal, ligand and solvent in crystal structures when the metal center is treated by an invariom like approach, but a molecule-specific database.

Ahlrichs *et al.*<sup>[105]</sup> introduced a series of new basis sets without effective core potentials for elements up to krypton including transition metals up to zinc in 2003. Two years later two improved versions of these basis sets including slightly more diffuse functions in the

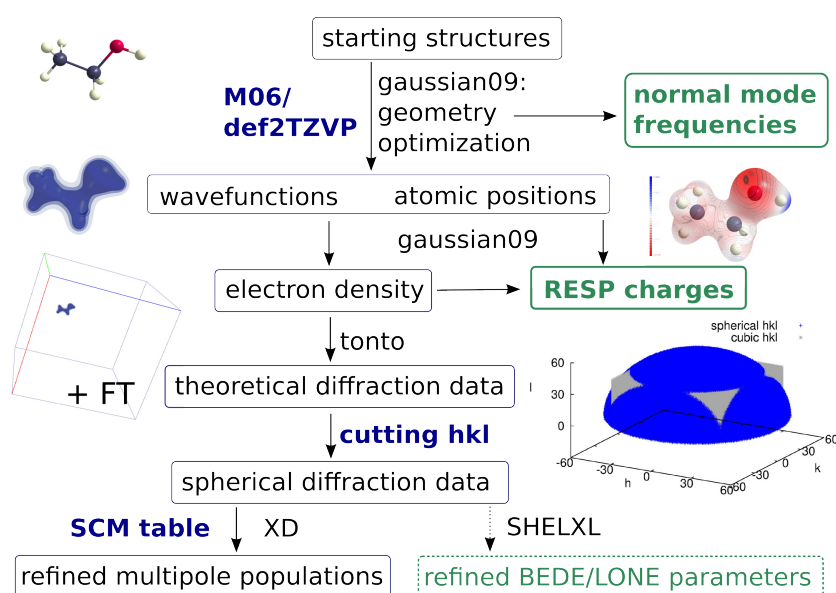
contracted triple-zeta valence basis for second to fourth row elements were presented,<sup>[106]</sup> with different degrees of polarization. The authors recommended the triple-zeta basis set without extra polarization functions for DFT calculations. Therefore, the def2TZVP basis set was chosen for the new geometry optimizations of the invariom model compounds.

## 2.1.2 Resolution

Another point of improvement was the resolution of theoretical scattering factors. The half cube of reflection data up to absolute  $h, k, l$  values of 40 was increased and cut to a half sphere where  $h, k, l$  values reach a maximum of 69 each (Figure 2.1). This way a well defined resolution of  $1.15 \text{ \AA}^{-1}$  ( $0.43 \text{ \AA}$ ) was achieved. In principle, a higher resolution would be possible but due to the application in which resolutions higher than  $1.15 \text{ \AA}^{-1}$  are the exception, it is not considered beneficial. The spherical shape instead of a cube should also allow the same level of information independent of a molecule's orientation in the theoretical cell.

## 2.1.3 Scattering factors

Discussions with fellow charge density researchers had raised the interest in different atomic scattering factor tables than those derived from HF *Slater*-type orbitals (STO) of Clementi and Roetti (CR).<sup>[107,108]</sup> Therefore, in the refinement of multipole parameters against the new theoretical diffraction data scattering factors derived from STO-Dirac-Fock atomic relativistic wave functions by Su, Coppens and Macchi (SCM)<sup>[109,110]</sup> were applied. The latter atomic form factors have no physically inadequate constant term in addition to the sum of six resolution dependent functions. For light-atom structures there is no difference between the two options, but for heavier atoms the newer SCM scattering factors provide



**Figure 2.1:** Additional treatment for every model compound in the database during makeover. Changes are highlighted in blue while new properties are marked green.



a better fit. An overview of the new model-compound treatment during database renewal is displayed in Figure 2.1.

#### **2.1.4 New compounds**

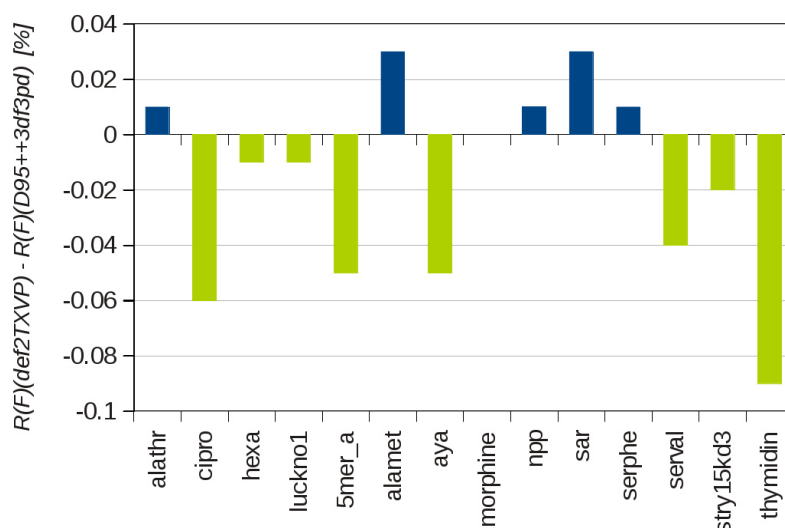
Since the newer basis set allowed more elements, many new model compounds became possible. The inclusion of bromine containing compounds was easily accomplished (see Section 2.2.2) and useful for modeling more organic molecules. Therefore, a systematic investigation of all the halogen compounds in the database allowed the addition of all those model compounds which were necessary to have each of the halogen model compounds with fluorine, chlorine and bromine.

## 2.2 Experiments and Results

### 2.2.1 Performance of the new database in refinements

The scattering factors of the new database from the def2TZVP basis set should perform similar to those from the GID,<sup>1</sup> otherwise the new database would only be useful for other properties than scattering factors. At the same time, obtaining a good multipole model refined against the new data was the most challenging part of the database renewal. In general the local atomic coordinate systems, the symmetry of refined multipoles and  $\kappa$  values could be taken from the previous refinements. Some coordinate systems, however are oriented by the help of dummy atoms, whose coordinates were marginally different now, but since the dummy atom coordinates are stored in the same file as the other copied information they were also taken from the previous refinement. The relative position of the old dummy atoms to the new atoms, which changed slightly during geometry optimization and sometimes just with respect to the molecules orientation in the unit cell, did not fit anymore. The dummy atoms are usually set up by linear combinations of bond vectors, which for new model compounds have to be generated by hand for unknown invarioms in the first refinement. But in this case usually the invarioms were already in the old database, hence INVARIOMTOOL was modified to also print out a master file for XD, in which only the dummy atoms are adjusted.

Refinements of fourteen crystal structures were compared for the old and new invarioms, as a test if the new scattering factor database performs as well as the old one. This test was of special interest, since in the old database the 1600 model compounds had been manually inspected before addition to the library over the course of time, whereas for the new database manual inspection was only performed for especially challenging cases.



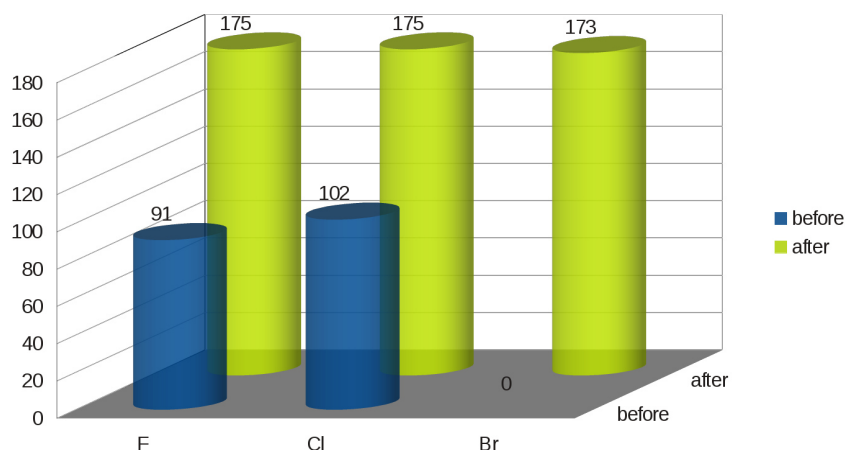
**Figure 2.2:**  $R(F)$  difference for invariom refinements with scattering factors from the old (D95++3df3pd) and new database (def2TZVP) for structures that have been used for benchmarking of invariom refinements before. Improvements are colored green.

<sup>1</sup>which were deduced for basis set D95++3df3pd.

Refinements were performed with XD Version 6.03<sup>[12]</sup> against  $F$  with riding hydrogen atoms. Neutrality was achieved by the default option of correcting only hydrogen monopole populations and the invariom names were the same for refinements with both databases. Residual factors,  $R(F)$ , for  $F_o > 2\sigma(F_o)$  were compared. Figure 2.2 shows the results where  $R(F)$  of the old database was subtracted from  $R(F)$  of the new database, so that an improvement is shown by differences below zero. The residual factor for morphine stayed the same, so that no difference was observed.

Overall, more structures improved upon modeling with scattering factors from the new database and the absolute value of improvement was higher than that of the worsening. The average change is -0.017 %. Therefore, this test lead to the conclusion that the new invariom scattering factor database is on average as good as the old one.

### 2.2.2 Addition of new halogen model compounds

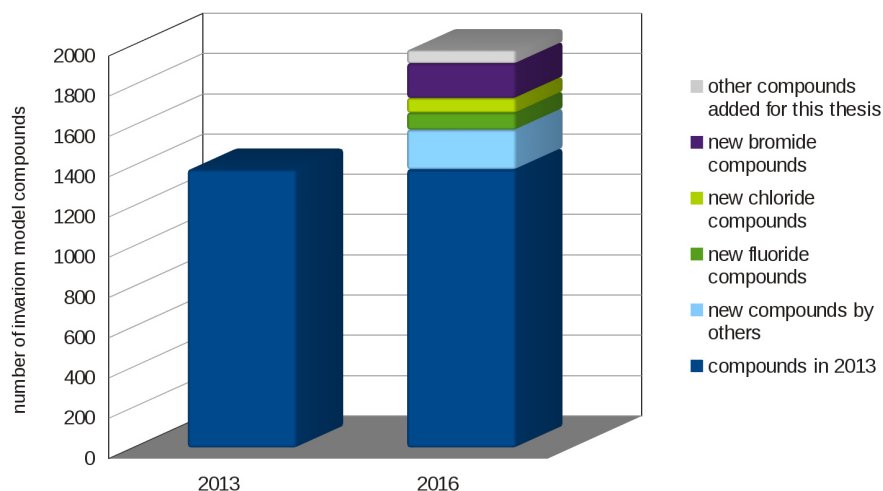


**Figure 2.3:** Number of halogen compounds in the invariom database before and after the bromine extension project.

Before adding all chlorine compounds as bromine ones, the chlorine compounds were compared to the fluorine ones. Ideally every model compound with a chlorine atom should also be present with a fluorine atom and *vice versa*; so that in order to extend the database to include bromine, all chlorine containing molecules would have to be copied and reoptimized, placed in a cell, its ED Fourier transformed and the multipoles refined with correct orientation and symmetry, with bromine instead of chlorine.

Results of the comparison between fluorine and chlorine compounds yielded 84 missing fluorine compounds and 73 missing chloromolecules. Automatic comparison of lists in which only the strings 'chlor' and 'fluor' were substituted needed careful evaluation, since trivial names were not equivalent for different halogens: e.g. chloroform was stored as trifluoromethane for the fluorine compound, and 1.1.1-trifluoroethane had been named trichloroethane for the other element. After filling in missing chlorine and fluorine compounds (157 molecules), the bromine compounds were added. The halogen compound numbers before and after the extension are shown in Figure 2.4, where the bromine count is lower by two because of doubly halogenated compounds of mixed kind: 1-bromo-3-

fluorobenzene, 1-bromo-1.1-dichloroethane where simple substitution of element symbols would have caused wrong results.



**Figure 2.4:** Number of model compounds in the database at start and end of this thesis.

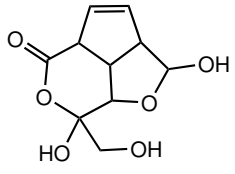
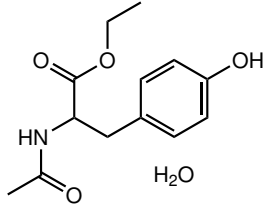
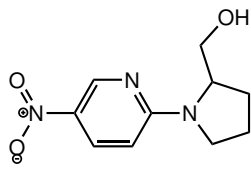
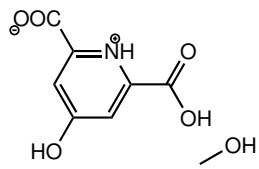
Overall, the number of database compounds increased from around 1400 to almost 2000, also due to other model compounds added during the time of the thesis. This growth of the database is visualized in Figure 2.4. It has grown by 40 %, of which more than half was due to the addition of the halogen compounds. Considering that usually each additional compound needs some manual initialization before the automatic treatment and manual setting of local coordinate systems, definition of symmetry and kappa assignment, this automatic way of improving database completeness was a great success. Otherwise model compounds are usually added only when their absence is detected. So the chance for experiencing a missing invariom should have decreased, and the work for manual addition of those compounds was also forgone. Moreover, for the charge project (see Section 4) where the charges are averaged from several model compounds, a higher number of model compounds improves charge transferability.

### 2.2.3 New invariom names of double bonded oxygen for invariom refinements

Another improvement concerning the classification of terminal double bonded oxygen atoms also relates to the database in a wider sense. So far multipole parameters for the invariom O2c had always been transferred from formaldehyde and O2n from aminoxide. As expected next-nearest neighbours (NNN) were found to influence the ESP from the invariom charges considerably (see Section 4.3.1) for carbonyl oxygen atoms. Thus, a more differentiated classification of those atoms improved the transferability of invariom point charges. Hence, atoms that are two bonds away are now included in the invariom name for charges so that there are 19 new invariom names instead of the one before.

If this finer classification would also improve the performance of scattering factors was examined for those molecules of the database test set which contained an invariom 'O2c' or 'O2n' by the old nomenclature. The results are shown in Table 2.1.

**Table 2.1:** Comparison of R1 for invariom refinements with scattering factors assigned by different invariom names for double bonded oxygen atoms.

structure code	structural formula	R1 for		new model compound
		O2x	O2x[NNN]	
luckno		1.48 %	1.48 %	acetic acid
nactyr		1.04 %	1.02 %	acetic acid
npp		1.88 %	1.87 %	2-nitrobenzeneamine
eg3095		2.56 %	2.55 %	benzoic acid

For three of the four test structures scattering factors transferred based on the elongated invariom names improved the fit to the XRD data. Therefore, this change in invariom names was also implemented in INVARIOMTOOL.



# 3 Identifying metals in coordination compounds

## 3.1 Introduction

Structure determination from single crystal XRD has become very fast in the past decades, allowing in easy cases data collection, structure solution, refinement and preparation for publication in *Acta Crystallographica E* within one day.<sup>[111]</sup> Reasons for this are improvements in instrumentation and software. Data collection has been accelerated by introduction of area detectors, which nowadays can be operated in continuous readout mode. Software has automated most steps and is easier to use, so that small-molecule structures are solved within seconds. Thus, the number of crystal structures published each year has increased exponentially, as shown by the statistics of the Cambridge Structural Database (CSD),<sup>[77,112]</sup> where most of the published structures are deposited. Since

"the number of experienced crystallographers dedicated to single-crystal studies has certainly not increased in proportion to the number of reported studies",<sup>[111]</sup>

automated validation of completeness, quality and correctness is required before crystal structures are submitted to the CSD. Incorrect structures cause problems especially for research that relies on them.

Numerous studies employ information from the CSD.<sup>[113]</sup> Most of them are of statistical nature,<sup>[113]</sup> derive properties,<sup>[114]</sup> investigate methods<sup>[115]</sup> or are simply based on selected structures obtained from the database.<sup>[116]</sup> According to the Cambridge Crystallographic Data Center (CCDC) homepage 17 publications were based upon information from the CSD in the first nine month of 2016.

Completeness and quality of a structure are usually ensured by automatic structure validation through CheckCIF.<sup>[111]</sup> Validating the correctness of a structure is more challenging. Erroneous structures in which hydrogen atoms are either missing or misplaced, and obviously incorrectly assigned atom-types can be identified by specific indicators.<sup>[111]</sup> The information for those indicators is deduced from the structure models. Investigating reflection data and comparing reflection files from different compounds can yield indications to possible fraud.<sup>[117-120]</sup> Such comparisons of different data sets can reveal that two probably isomorphous structures have reflection data deviating only by a scale factor, implying a linear correlation if both data are plotted against each other. In such cases only one of the structures can be correct, but automatic methods so far can not tell which of the compared structures is the correct one, if in principle both are chemically possible. Coordination compounds have a rather high flexibility concerning the geometry of ligands. It requires a chemists knowledge and experience to discern which of two structures with different metal centers is more likely. Likelihood, however, is a weak argument when judging other scientists

work. Therefore, a method delivering proof of the correct metal in otherwise isomorphous models is required.

This chapter of the thesis applies a method for identifying the metal atom that compares the ability to fit XRD data for models that include atomic asphericity. Since valence density strongly affects low order data,<sup>[30]</sup> a better description of bonding ED is also a model improvement for data collected to standard resolution (0.83 Å). The same region of data is affected by the difference between metals that have a similar number of electrons. Additionally, a better model improves crystallographic phases and hence, the general distinguishing power between different models. The method applied in this thesis to distinguish metals in crystal structures of coordination compounds by aspherical scattering factors was developed and validated for complexes, which had been synthesized in-house with different metals as central atom.<sup>[99]</sup>

In this project eleven pairs of crystal structures<sup>[121–142]</sup> from the CSD with pairwise the same cell and compound geometry but different metals as central atom were investigated. The crystal structure pairs were identified by Jim Simpson and Matthias Weil. In some cases the reflection data sets were not the same, but isomorphism was still questioned due to similar cell and molecule geometries.

Alternatively to the procedure described and applied here, it would of course be better to apply other chemical analyses to identify the correct metal atom, if the compound is at hand. But for this project only the deposited crystal structure data was available. Thankfully not only the models but also reflection data were deposited, otherwise the method could not be applied to identify the correct structure. A synthetic approach of trying to synthesize and recrystallize each of the possible complexes and redetermination of each crystal structure would show which of the structures can be reproduced. However, doubts concerning the reason for non-reproducible structures would have to be resolved and discovery of new polymorphs or co-crystals would also hinder the conclusion of such a synthetic approach.

#### 3.1.1 Isomorphism

For those compounds where the XRD data sets were not basically equal, real isomorphism was theoretically possible. Before discussing this subject, the terms isomorphism and isostructuralism shall be specified. The definition of isostructural crystals as given by the IUCr Online Dictionary of Crystallography is:

"Two crystals are said to be isostructural if they have the same structure, but not necessarily the same cell dimensions nor the same chemical composition, and with a 'comparable' variability in the atomic coordinates to that of the cell dimensions and chemical composition."<sup>[143]</sup>

The definition for isomorphous crystals by the same reference is:

"Two crystals are said to be isomorphous if (a) both have the same space group and unit-cell dimensions and (b) the types and the positions of atoms in both are the same except for a replacement of one or more atoms in one structure with different types of atoms in the other (diadochy), such as heavy atoms, or the presence of one or more additional atoms in one of them (isomorphous addition)."<sup>[143]</sup>



From those definitions isomorphism is the more precise description for the pairs of structures investigated, since they have the same cell, space group and just differ by one atom type. Hence, the isomorphism discussed here concerns diadochy.

Due to more degrees of freedom in the ligand, true isomorphism occurs less frequently when the ligand size increases. In complexes of *3d*-metals different coordination geometries are favored for different numbers of electrons and thus different electronic configurations. Bond distances can change due to e.g. Jahn-Teller (JT) splitting.<sup>[144]</sup> Therefore, unit cells and molecular geometry are usually not the same, so that true isomorphism occurs rarely. If isomorphous structures are found for two complexes, their chemical properties are commonly compared further, leading to a topic of interest for chemical rather than crystallographic journals.

Several cases<sup>1</sup> for isomorphous structures of octahedrally coordinated manganese(II) and cobalt(II) are known.<sup>[145–147]</sup> In an example of isomorphous octahedral nickel(II) and cobalt(II) complexes<sup>[148]</sup> a striking geometrical difference between the compounds is that the largest angle between two oxygens atoms is once  $152.08(6)^\circ$  and once  $157.17(5)^\circ$ , while the bond distances differ by around  $0.08 \text{ \AA}$ . The changes between bond lengths of nickel and cobalt are smaller ( $0.03 \text{ \AA}$ ),<sup>[145]</sup> but bond angles differ by up to three degrees. Between the cobalt and manganese complexes the bond lengths differ by values between  $0.05$  and  $0.3 \text{ \AA}$ .<sup>[146]</sup> Cell parameters were observed to change by  $0.03 \text{ \AA}$ <sup>[145]</sup> to  $0.3 \text{ \AA}$ <sup>[146]</sup> and even  $0.4 \text{ \AA}$ .<sup>[147]</sup> These numbers indicate the considerable extent of structural changes related to exchange of metal atoms in isomorphous structures.

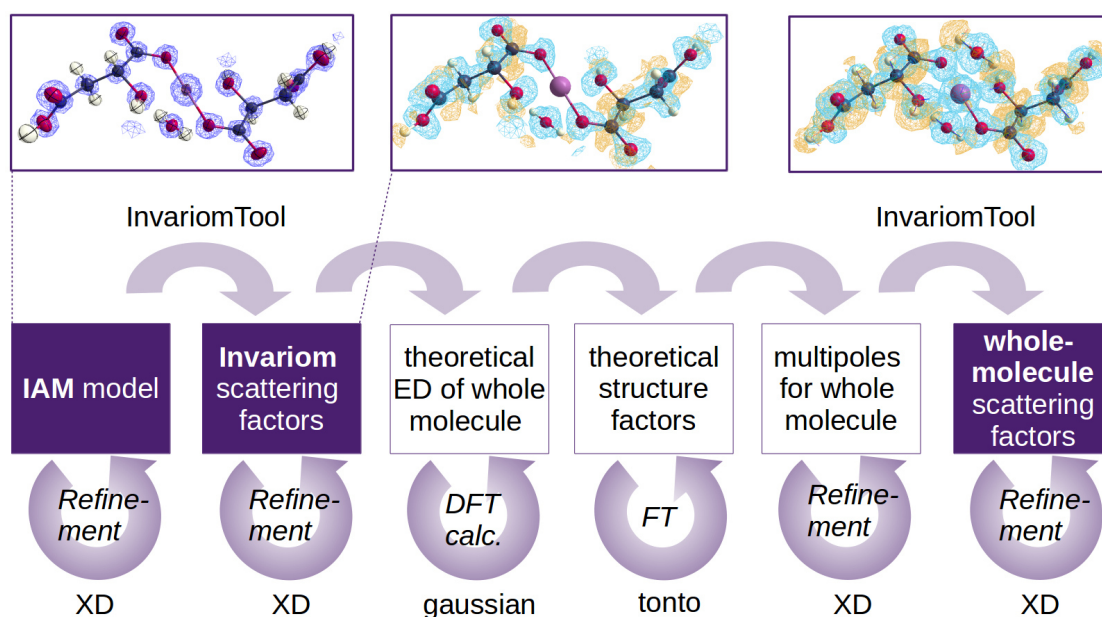
Additional chemical analytics should confirm the different metal atoms, if differences between 'isomorphous' complexes in crystal structure models are less or not even significant, especially if the coordination geometry is uncommon for one of the metals. Otherwise the isomorphism is suspicious even if XRD data are collected in different experiments, since the frequently applied IAM can not identify the correct metal atom.

## 3.2 The invariom like approach

In principle identification of the correct metal atom in pairs of structures was accomplished by an invariom refinement. However, the aspherical scattering factors could be transferred from the invariom database only for the ligand, because the database is restricted to organic molecules. Ignoring the metal during invariom classification, all atoms except the metal could be modeled by multipoles. In order to acquire the aspherical scattering factors for the metal center and consider the coordination bond from the donating ligand atoms, a DFT single-point calculation for the geometry obtained after invariom refinement was performed. The thus obtained molecular ED, as a square of the wave function, was projected onto the multipole model. For technical reasons this projection was performed via a detour through reciprocal space, as for all model compounds in the invariom data base. Next, the aspherical scattering factors obtained from the fit to the ED of the 'whole molecule' were used in a refinement of coordinates and ADP against the deposited experimental reflection data. The approach is illustrated in Figure 3.1.

This complete procedure was performed for both metal atoms considered, and each model

<sup>1</sup>Unfortunately all structures mentioned in this paragraph were deposited without reflection data.



**Figure 3.1:** Graphical representation of the 'whole-molecule' work flow. Boxes with a purple background represent the models refined against XRD data from experiment, while the white boxes are purely computational.

was refined against both deposited data sets. Next, IAM, invariom and 'whole-molecule' models were compared for each data set and metal. Figures of merit could be compared well, since the number of parameters refined was the same for each of the three models. The multipole parameters were kept at the values refined against the theoretical data.

### 3.2.1 Metal atoms on special positions

In half of the investigated structures the metal atom crystallized on a special position. In those cases the molecule has to be completed before quantum chemical computation of the molecular ED. Although a recent study<sup>[149]</sup> suggested calculation of the crystallographic unit cell, equal results are obtained by completing the molecule before continuing with the work flow of standard invariom model compounds. This proceeding requests that symmetrically equivalent atoms are treated equally in the multipole refinements against the theoretical data. Moreover, for those atoms on a special position in the real crystal structure only those multipoles are populated that were in accordance with the special site symmetry.

### 3.2.2 Complexes with multiple possible electronic configurations

For the calculation of molecular wave functions the spin state of the metals is relevant. For all investigated compounds spin states are approximated by considerations according to ligand field theory (LFT). Since LFT is a rather crude approximation and molecular wave functions are easily accessible via DFT computations, energies from QM calculations of high-spin (hs) and low-spin (ls) state are compared for nickel and cobalt complexes.

## 3.3 Experiments and Results

Each of the eleven cases studied was treated by the same procedure, which was described in general in Section 3.2.

### 3.3.1 Procedural details

#### 3.3.1.1 File setup

Initially crystallographic information and structure factors were downloaded from the CSD, and instruction files for SHELXL<sup>[11]</sup> were generated with PLATON.<sup>[111]</sup> The structure factors were then converted to SHELX format HKLF 4 by the small utility program FCF2HKL.<sup>2</sup> If not mentioned otherwise, the XRD data were collected at room temperature with molybdenum  $K\alpha$  radiation. Usually the reflection files excluded systematically absent reflections and contained negative intensities.

#### 3.3.1.2 SHELXL IAM refinement

The IAM refinements were repeated for the two metals of each pair. Scattering factors of neutral atoms were applied in all cases. For selected structures ionic scattering factors were applied by manually copying the coefficients from the International Tables of Crystallography.<sup>[150]</sup> Small changes of the residual factors were observed, but although the tendency of improvement for the correct metal atom could be noticed, the metal atom could not be identified clearly. In each case the weighting scheme was adjusted until convergence was reached.

#### 3.3.1.3 XD IAM refinement

The transition to the XD program package<sup>[12]</sup> (version 6.03) was facilitated by the packages subroutine XDINI. Corresponding to the first box of Figure 3.1 the IAM refinement was repeated for each metal with the subprogram XDLSM. After fixing selected atomic coordinates to cope with floating origins in space groups with only translational symmetry elements in one direction, the model was refined against F. Moreover, ADPs of atoms on special positions were constrained according to the position's site symmetry. Hydrogen atoms were treated by the riding atom model and their displacement parameters fixed to 1.2 and 1.5 times the parent atom's  $U_{eq}$ .<sup>3</sup> Differently from the refinement with SHELXL, the weighting scheme was not refined with XD but kept purely statistical ( $w = \frac{1}{\sigma^2}$ ), because adjustment of weights should follow correct element assignments only and in this study element assignment is investigated. That adjustment of weights would not probably do more harm than good was indicated by unnaturally high second weighting factors in the SHELXL refinements.

Element name, number of valence electrons in the  $d$ -orbitals and dispersion corrections needed adjustment in the files read by XD, in order to change the metal atom for the

<sup>2</sup>written by G. M. Sheldrick, Georg-August University Göttingen, 2014.

<sup>3</sup>All of those details are also true for refinements with SHELXL but did not need to be taken care of manually.

refinements with this program. In all refinements discussed in this chapter which were carried out with XD the scattering factors of the ionized metal were used and valence electrons as well as names were set accordingly in the necessary files.

#### 3.3.1.4 Invariom refinement

The Hansen and Coppens<sup>[50]</sup> multipole parameters were transferred from the new def2TZVP Invariom database (Section 2) for ligand, solvent and counter ion atoms with INVARIOM-TOOL.<sup>[91]</sup> The method for achieving the molecular charge of interest was the default option three, which adjusts the monopole population according to the electronegativities of the elements. The model including those aspherical scattering factors was refined against F with XDLSM. Hydrogen positions were set to theoretical bond lengths and constrained to ride on their parent atoms.

#### 3.3.1.5 Metal-complex specific theoretical ED

A single-point calculation of the metal complex with the coordinates after invariom refinement yielded a molecular wave function. The program GAUSSIAN09<sup>[103]</sup> was employed for DFT calculation with the functional M06<sup>[104]</sup> and basis set def2TZVP.<sup>[106]</sup> For the pairs of metal-complex structures investigated, the coordinates from invariom refinement against the data set that yielded the best refinement result were used. For cases containing atoms on a special position, the molecule was completed for the single point calculation. The missing cartesian coordinates were calculated with PLATON. The spin multiplicity, which had to be set for the single-point calculation, was deduced from ligand field theory at first. If in principle high- and low-spin states were possible, both were considered by DFT calculations to confirm the favored multiplicity.

For the investigated structures the Self-Consistent Field (SCF) energies at the experimental geometries confirmed the high-spin states for cobalt(II) and nickel(II) in octahedral complexes as the favored electronic configurations (Table 7.1 in Appendix A). In the two cases of tetrahedral coordination (see Table 7.2 in Appendix A) nickel(II) was energetically lower in its high-spin state for both cases. Cobalt(II) was only considered once for a tetrahedral complex, for which the high-spin state originally deduced from crystal field theory was also confirmed by quantum-mechanical calculation. For the first three of four square-planar complexes the low-spin state was favored for both metals (Appendix A, Table 7.3), so that for the last one the DFT calculation was only performed for the low-spin state to generate the scattering factors for the 'whole molecule'.

For those cases treated first,<sup>4</sup> scattering factors were generated and compared for both spin states in addition to the different metals. The comparisons, however, did not yield any valuable information, since with exception of one case, the different spin states affected merely the incorrect metal. Therefore, later refinements were performed only for the energetically preferred spin states.

---

<sup>4</sup>The chronological order differed from the numbering scheme, which was chosen only for better comprehensibility.

### 3.3.1.6 Generation of theoretical structure factors

The ED of each molecule was placed at fractional coordinates of approximately  $\frac{1}{4}, \frac{1}{4}, \frac{1}{4}$  in an artificial, cubic unit cell with  $a = 30 \text{ \AA}$ . To avoid problems with flexible phases in non-centrosymmetric space groups, the space group was set to  $P\bar{1}$ . With the exception of  $P1$  it required the least amount of molecules to be considered. In contrast to  $P1$ ,  $P\bar{1}$  does not require the complete *Ewald* sphere to be computed. Since  $P\bar{1}$  is centrosymmetric *Friedel* pairs are identical and the wavelength dependent parts of the atomic scattering contributions were not included in the calculated structure factors. The theoretical structure factors were calculated from  $h = k = -40$  and  $l = 0$  to  $h = k = l = 40$ . No thermal smearing or extinction was added, so that the following refinement step concentrated on the asphericity of the ED.

### 3.3.1.7 Refinement of aspherical models against theoretical data

Refinement of aspherical scattering factors against the theoretical data was carried out with the program XDLSM (XD 6.03). The choice which multipoles were refined was based on local molecular site symmetry up to next-nearest neighbors. The primary axis of local coordinate systems for the atoms engaged in a dative bond to the metal center were oriented towards the metal. Contraction parameters,  $\kappa$ , were refined for all atoms, while for atoms heavier than carbon  $\kappa'$  was refined, too. Refinements were performed against F in as many cycles as necessary to achieve convergence as low as a maximal shift/s.u. of  $10^{-5}$ .

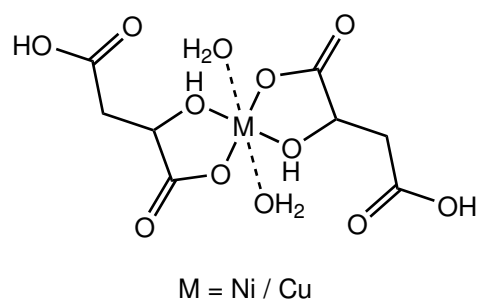
### 3.3.1.8 Refinement of the aspherical model of the 'whole molecule'

The multipole parameters refined for the 'whole molecule' against the theoretical data were stored in a molecule-specific invariom database by the developer version of INVARIOM-TOOL. With this file INVARIOMTOOL could transfer those parameters to the experimental structure models. All other settings for the refinement of the 'whole-molecule' scattering factors against the experimental data (last box in Figure 3.1) were equal to those of the invariom refinements.

### 3.3.1.9 Geometry optimization

In a few cases a geometry optimization was performed in addition to the single-point calculation to observe the changes upon geometry relaxation. DFT energies by themselves can not be compared between different metals for the complexes since the number of electrons and thus the total SCF energy are different. The relative change upon geometry relaxation however, could indicate that for one metal the coordinates from the crystal structure are at or close to an energetic minimum, whereas for the other one a different atomic arrangement would be preferred. Hence, in some selected cases, the findings from the invariom-like approach are supported by geometry optimizations. For single-point as well as geometry optimization calculations the SCF convergence criterion was  $10^{-8}$  au and the integration grid pruned (99,950). Functional and theory were consistently M06/def2TZVP.

## 3.3.2 Case studies

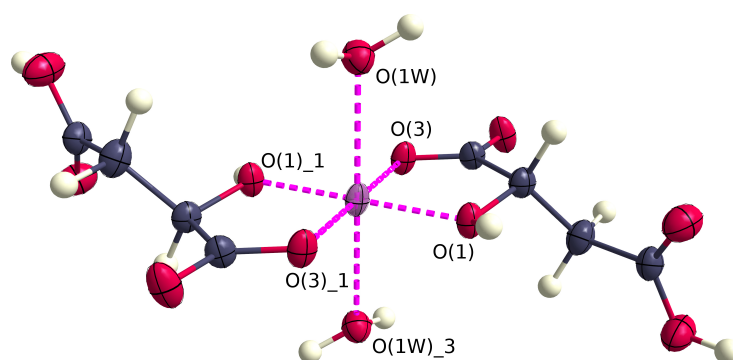
3.3.2.1 Pair 1: Diaquabis(malato- $\kappa^2O^1, O^2$ )nickel(II)/copper(II)

Pair	1	
	a <sup>[121]</sup>	b <sup>[122]</sup>
IUCr code	dn2151	hb2526
CSD code	BESNOC01	XILXIA
CCDC No.	647183	664185
Space group	$P2_1/c$	
Peculiarity	M on special position ( $\bar{1}$ )	
Coord. geom.	square planar	
Metal ion	$\text{Cu}^{2+}$	$\text{Ni}^{2+}$
Electron config.	$[\text{Ar}]4s^03d^9$	$[\text{Ar}]4s^03d^8$
Spin multiplicity	2	3

**Table 3.1:** Structural formula and selected crystallographic and chemical information of pair 1.

In this case of diaquabis(malato- $\kappa^2O^1, O^2$ ) complexes the metals nickel(II)<sup>[122]</sup> and copper(II)<sup>[121]</sup> were considered. The metal atom sits on an inversion center, so the asymmetric unit contained only one ligand and water molecule. The two malato ligands coordinate via two oxygen atoms, O(1) and O(3). Additionally, there are two water molecules coordinating axially from a longer distance. While O(1) has a distance of 1.9556 Å and O(3) of 1.9123 Å the oxygen water molecule was 2.5192 Å away.<sup>5</sup> An ORTEP type plot including the relevant atom names is given in Figure 3.2.

Data set a<sup>[121]</sup> contains reflections up to a higher resolution and an intensity that is constantly 1.094 times that of reflections in data set b<sup>[122]</sup>. The unit cell parameters are identical although the number of reflections used for cell measurement is different according to the crystallographic information. While a was supposedly measured at 293(2) K, b is stated to have been measured at 298(2) K. Nonetheless, the reflection data is the same. Hence, only the correct metal had to be identified.



**Figure 3.2:** Thermal ellipsoid plot at a probability of 50% of pair 1 with Cu after refinement of the whole-molecule scattering factors against data set a. Symmetry equivalent ligand and water molecule are also displayed.

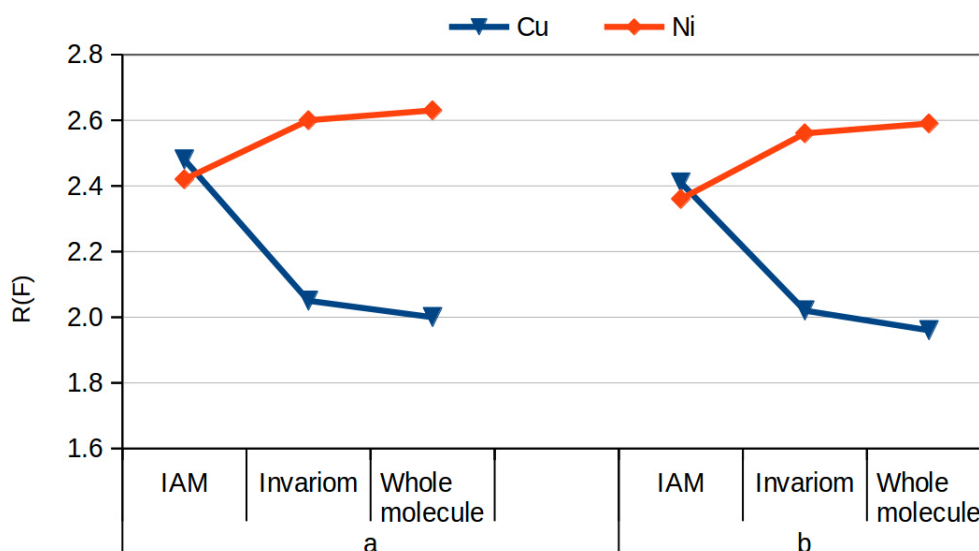
<sup>5</sup>Distances were taken from the final whole molecule copper model refined against data set a.

**Chemical reasoning** As discussed before the complex's coordination geometry is an elongated octahedron. This represents a strong argument for copper(II), since it is of electron configuration  $d^9$ , which unlike the  $d^8$  configuration of nickel(II) profits from JT splitting of orbitals. Therefore, basic orbital consideration already suggested copper as the correct metal.

**Table 3.2:** Selected computational and refinement results for pair 1.

	Cu	Ni
E(UM06) crystal geometry [au]	-2856.475	-2724.308
R(F) against theo. Data [%]	0.47	0.47
R(F) whole molecule a [%]	<b>2.00</b>	2.63
R(F) whole molecule b [%]	<b>1.96</b>	2.59

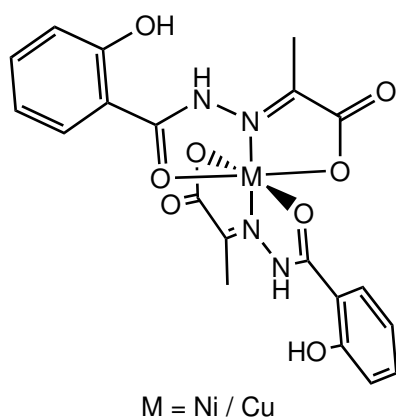
**Refinement results** Crystallographically speaking, this case is a typical example where in the IAM the heavier metal fits the reflection data worse (see Table 3.2). Upon invariom modeling, the fit for copper improves and the gap between nickel and copper increases further with aspherical-atom modeling for the whole molecule. Figure 3.3 visualizes that with increasing model quality the capability of nickel to fit the data gets worse than for copper. Therefore, it was proven that copper is the correct metal.



**Figure 3.3:** Comparison of R(F) for the refinements of the different metal atoms with different ED models against the two data sets of pair 1.

## 3.3.2.2 Pair 2:

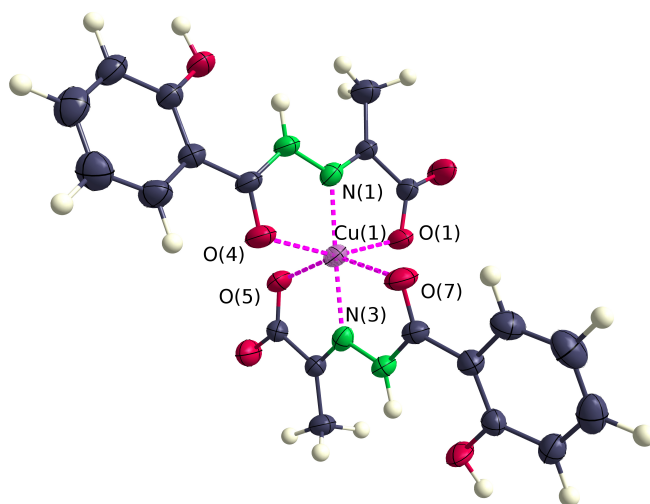
## Bis[2-(2-hydroxybenzoylhydrazono)propionato]nickel(II)/copper(II)



Pair	2	
	a <sup>[123]</sup>	b <sup>[124]</sup>
IUCr code	dn2235	hy2076
CSD code	WINVOF	XILWUL
CCDC No.	669120	663160
Space group	$P\bar{1}$	
Peculiarity	3 water in asu	
Coord. geom.	distorted octahedral	
Metal ion	Ni <sup>2+</sup>	Cu <sup>2+</sup>
Electron config.	[Ar]4s <sup>0</sup> 3d <sup>8</sup>	[Ar]4s <sup>0</sup> 3d <sup>9</sup>
Spin multiplicity	3	2

**Table 3.3:** Structural formula and selected crystallographic and chemical information of pair 2.

This octahedral complex **2** is similar to case **1**. The metals considered are copper(II)<sup>[124]</sup> and nickel(II)<sup>[123]</sup>, as before. In addition to the complex, three water molecules crystallized in the same unit cell, one of them forms a hydrogen bond to O(4). Otherwise the hydrogen bonds connect the water molecules among each other and those oxygen atoms of the complex not involved in metal coordination.



**Figure 3.4:** Thermal ellipsoid plot at a probability of 50 % of pair **2** with Cu after refinement of the whole-molecule scattering factors against data set **b**. Water molecules are omitted for clarity.

**Chemical reasoning** A tetragonal distortion of the octahedron could in this case be in principle caused by strain in the ligand, so JT splitting would not necessarily be the only explanation if the coordinating nitrogen atoms aligned to the principal tetragonal axis.



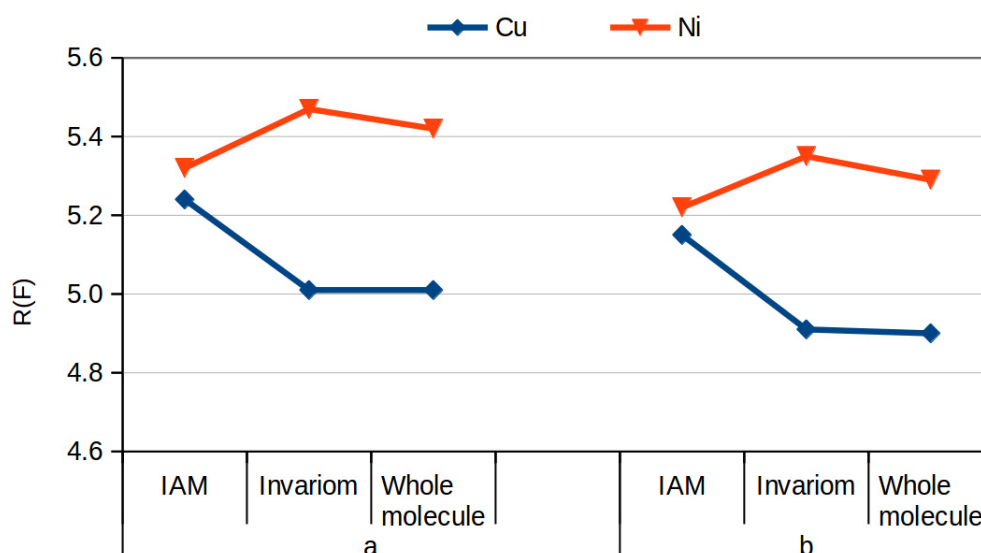
**Table 3.4:** Selected bond lengths for **2b** with copper after refinement with whole-molecule scattering factors.

Cu(1) - N(1)	1.929(3)
Cu(1) - N(3)	1.974(3)
Cu(1) - O(1)	2.042(3)
Cu(1) - O(4)	2.090(3)
Cu(1) - O(5)	2.208(3)
Cu(1) - O(7)	2.275(3)

An investigation of the bond distances around the central atom (see Table 3.4) revealed that the bonds to O(5) and O(7) are longer than the bonds to oxygen atoms O(1) and O(4). O(1) and O(4) are not chemically equivalent but O(1) is similar to O(5) (Figure 3.4). The bonds to the oxygen atoms of one ligand oppose each other, therefore, the longer bonds to O(7) and O(5) correspond to an elongating JT distortion of the octahedral complex. A distortion caused by JT is reasonable for stabilizing the  $d^9$  electron configuration of the copper ion, less so for the nickel ion ( $d^8$ ). The bond lengths differences were

already visible in the standard structure model in SHELXL.

**Crystallographic aspects and refinement results** The data for pair 2 is the same in both structures, differing only by a scale factor. Therefore, it was enough to prove via the deposited XRD data that copper is the correct metal atom. Figure 3.5 shows the evolution of  $R(F)$  with increasing aspherical modeling. The usual trend of improvement for the heavier metal can be seen, and a change for the worse for the lighter one. In contrast to most other cases, copper already fitted better in the IAM model. Especially the introduction of invariom scattering factors increased the distinguishing power of the diffraction data. Upon inclusion of multipole parameters for the metal atom the  $R(F)$  for the nickel model also decreased, which is perfectly reasonable, since previously was not modeled ED is modeled now. The difference between the 'whole-molecule' scattering factors of 5.42 % and 5.01 % for a and b, respectively (Table 3.5), unambiguously proves that copper yields the better fit to the XRD data.

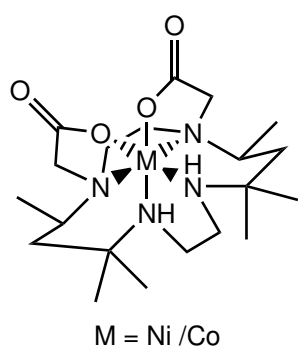


**Figure 3.5:** Comparison of  $R(F)$  for the refinements of the different metal atoms with different ED models against the two data sets of pair 2.

**Table 3.5:** Selected computational and refinement results for pair 2.

	Ni	Cu
DFT energy [au]	-3101.82	-3233.97
R(F) against theor. data	0.52	0.47
R(F) whole molecule a	5.42	<b>5.01</b>
R(F) whole molecule b	5.29	<b>4.90</b>

### 3.3.2.3 Pair 3: (5,5,7,12,12,14-Hexamethyl-1,4,8,11-tetraazacyclotetradecane-1,8-diacetato)nickel(II)/cobalt(II)



complex	3	
	a <sup>[125]</sup>	b <sup>[126]</sup>
IUCr code	ci6607	hb6232
CSD code	TAQSUA	CASHIO
CCDC No.	282277	282363
Space group	$P2_1/n$	
Peculiarity	4 water in asu	
Coord. geom.	octahedral	
Metal ion	Ni <sup>2+</sup>	Co <sup>2+</sup>
Electron config.	[Ar]4s <sup>0</sup> 3d <sup>8</sup>	[Ar]4s <sup>0</sup> 3d <sup>7</sup>
Spin multiplicity	3	2(ls) 4(hs)

**Table 3.6:** Structural formula and selected crystallographic and chemical information of pair 3.

This pair of structures included an octahedral nickel(II)<sup>[125]</sup> or cobalt(II)<sup>[126]</sup> complex with one hexadentate, chelating ligand. The ligand donates ED via four neutral nitrogen atoms and two negative oxygen atoms that are part of acetate groups. The compound crystallized in space group  $P2_1/n$  and there are four non-coordinating water molecules in the asymmetric unit. In structure **b** one of them was modeled as partially occupied. The data sets are not completely proportional to each other and their unit cells differ significantly, too (Table 7.4, Appendix A). Therefore, not only which metal atom is present had to be investigated, but also, if the metal atoms are the same or different in both structures.

**Refinement results I** In the IAM model, including the ionic scattering factors and according electron counts for the metal atoms, cobalt led to a lower R(F) of 5.83 % than nickel with 5.94 % for data set **a**. For data set **b** the absolute residual factors are higher, but the difference for the metals is similar. Upon introduction of the invariom scattering factors the relation changed for both data sets. For nickel as metal invariom modeling led to a significant improvement of R(F), whereas for cobalt it improved less for data set **a** and worsened the fit for data set **b** (Figure 3.6).

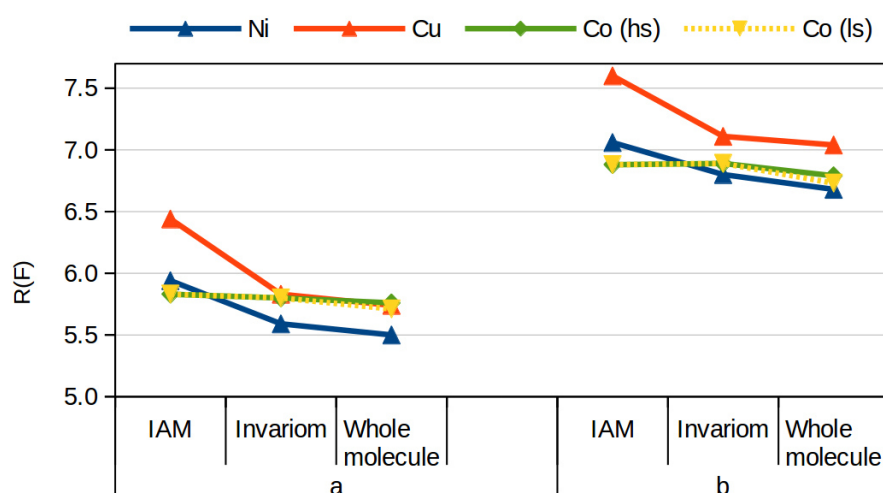
**Spin state** Before moving on to the 'whole-molecule' results, the input parameters for the DFT computation are briefly discussed here. Since the energy comparison for different

spin states (see Appendix A, Figure 7.1) strongly confirmed the deduction from crystal field theory that nickel has a spin multiplicity of three in this octahedral coordination environment only this spin state was evaluated further.

For cobalt the high-spin state is also energetically favored, at least in a gas phase calculation of the isolated molecule, but the ED of the low-spin state was also projected onto the multipole model and compared to the experimental data, because the energy difference between the spin states was an order of magnitude smaller than for nickel.

**Refinement results II** For data set **a**, which was published as nickel complex, introduction of aspherical scattering factors for the whole molecule improved the density fitting further. Although the improvement is similar for both metals, nickel clearly is a better model for the data (Figure 3.6).

For data set **b** the general data quality is worse. Different water models were tested, none improved the result considerably. Hence, improvements seemed small compared to the overall  $R(F)$  value of this data set. Still, nickel could be identified to fit the ED best. Since the water molecules are less well organized in this crystal, another part of the model than the coordination center might have been dominating the quality of the fit and thus rendering the method of aspherical scattering factors for metal atom identification less effective. A good structure model is a prerequisite and this case demonstrates the limit of the method.



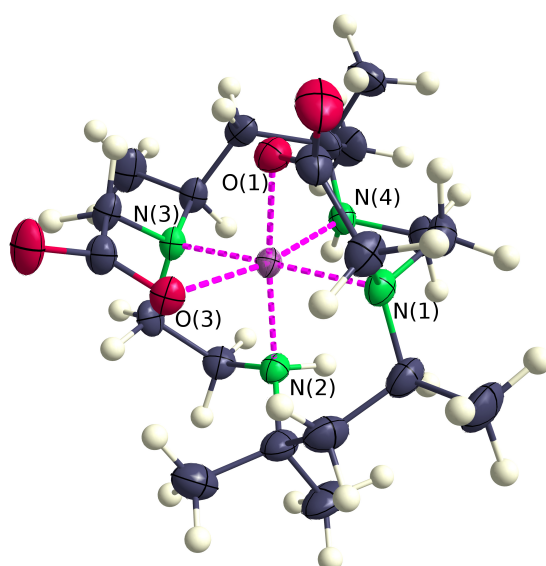
**Figure 3.6:** Comparison of  $R(F)$  for the refinements of the different metal atoms with different ED models against the two data sets of pair 3.

**Comparison to copper** Since all of the investigated structure pairs so far yielded the conclusion upon invariom treatment that the heavier metal is the correct one, the comparison with copper for this structure shows that the preference of the heavier element is not a systematic error, but indeed the result of the method. The comparison of cobalt, nickel and copper shows that invariom scattering factors improve the copper structure, too, but the copper model agreed definitely worse with the data than the nickel model.

**Table 3.7:** Selected computational and refinement results for pair 3.

	Ni(hs)	Co(hs)	Co(ls)	Cu
E(HF) crystal geometry [au]	-2812.97	-2687.41	-2687.34	-2945.10
R(F) against theo. Data [%]	0.48	0.49	0.51	0.48
R(F) whole molecule a [%]	<b>5.50</b>	5.76	5.71	5.74
R(F) whole molecule b [%]	<b>6.68</b>	6.79	6.73	7.04

**Energy considerations** Since some DFT calculations had been performed during the procedure a theoretical backing for the decision on the correct metal in data set **b** was considered helpful. Therefore, the relative change upon relaxation of geometry was also taken into consideration. The nickel structure of the invariom model for **a** was already at an energy minimum, while cobalt can be minimized by several kJ/mol (Table 7.5 in Appendix A). The crystal structure atomic coordinates are farthest away from those of the optimized low-spin cobalt complex and not at the minimum for cobalt high-spin state either. This was expected for data set **a**, where nickel had been determined to be the best fitting metal, while for **b** the situation was less clear. Especially, since the difference between low- and high-spin cobalt indicates a better fit for the low-spin state although it energetically appears as less stable in the gas phase. The difference between R1 of different spin states is almost as large as the difference between R1(all) of the Co(ls) model and the Ni model for data set **b**. Therefore, the support of nickel as the correct metal by the significantly larger change in energy upon geometry relaxation for cobalt in either spin state than for nickel was reassuring.

**Figure 3.7:** Thermal ellipsoid plot at a probability of 50 % of pair 3 with Ni after refinement of the whole-molecule scattering factors against data set **a**. Water molecules and hydrogen atoms are omitted for clarity.

**Geometry evaluation** As Table 3.8 shows the largest deviation from the mean bond length to nickel was found for N(3) and N(1) which are opposing each other (Figure 3.7), one bond longer, one shorter than average. N(2) and N(4) are the other two chemically equal nitrogen atoms and are both opposite of an oxygen atom [O(1) to N(2)]. Both bond pairs do not show a considerable difference in the distance between the opposing ligating atoms. Therefore, no static JT distortion is observed. The thermal ellipsoids of the coordinating atoms displayed in Figure 3.7 show no elongation in the direction of the dative bonds, either. This further indicates nickel as metal ion opposed to cobalt which, in its low-spin state, would benefit from splitting of  $e_g$  orbitals, whereas in the high-spin state the splitting of  $t_{2g}$  orbitals would be slightly favored.

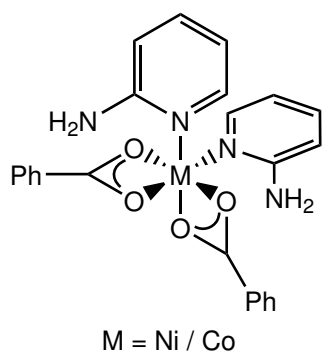
Comparison to the bond lengths from geometry optimizations shows a slightly stronger shortening of especially the metal-N(3) bond for cobalt than for nickel. The two bonds to N(2) and N(4) also shortened in the geometry optimizations but to a similar extent for both metals.

**Table 3.8:** Selected bond lengths for pair **3a** with nickel after refinement with whole-molecule scattering factors.

atom pairs	crystal	geom opt Ni	geom opt Co ls
Ni(1) - N(1)	2.113(2)	2.11141	2.11011
Ni(1) - N(2)	2.102(2)	2.09189	2.09184
Ni(1) - N(3)	2.094(2)	2.08105	2.07935
Ni(1) - N(4)	2.103(2)	2.09449	2.09437
Ni(1) - O(1)	2.107(2)	2.10836	2.10779
Ni(1) - O(3)	2.113(2)	2.11409	2.11345

**Conclusion** In summary, chemical reasoning consulting bond lengths, energetic considerations and refinement with aspherical scattering factors, confirm nickel as the correct metal in both structures.

## 3.3.2.4 Pair 4: Bis(2-aminopyridine)dibenzoatocobalt(II)/nickel(II)

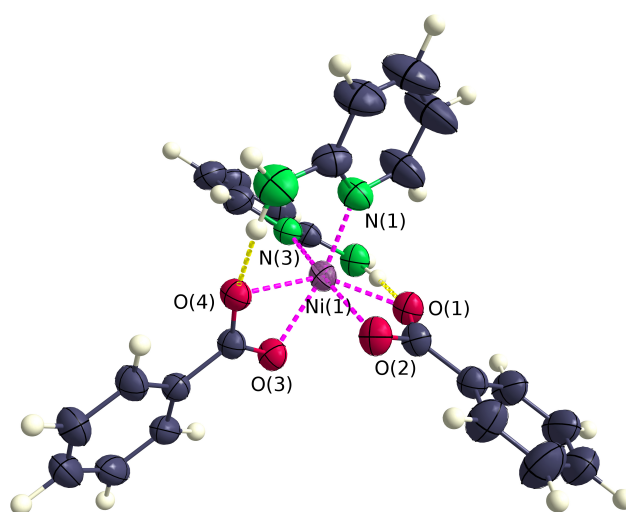


Pair	4	
	a <sup>[127]</sup>	b <sup>[128]</sup>
IUCr code	lh6101	hb2030
CSD code	OLOJOO	IDOKOC
CCDC No.	225650	608593
Space group	C2/c	
Peculiarity	-	
Coord. geom.	distorted octahedral	
Metal ion	Ni <sup>2+</sup>	Co <sup>2+</sup>
Electron config.	[Ar]4s <sup>0</sup> 3d <sup>8</sup>	[Ar]4s <sup>0</sup> 3d <sup>7</sup>
Spin multiplicity	3	2, 4

**Table 3.9:** Structural formula and selected crystallographic and chemical information of pair 4.

In the octahedral bis(2-aminopyridine)dibenzoatometal(II) complex, **4**, the pyridine ligands are in *cis* configuration and the negative charge in the two anionic benzoate ligands is distributed over all coordinating oxygen atoms. The crystal structure can contain either nickel or cobalt as central atom, since data sets **a**<sup>[127]</sup> and **b**<sup>[128]</sup> are identical. The unit cells reported are also identical, although a different number of reflections was used for cell determination (**b**: 2530, **a**: 19350). The number of reflections for refinement was the same in both files and the only difference is that the reflection intensity of data set **a** is 1.083 times as intense as in data set **b**.

**Chemical reasoning and spin state** In an octahedral environment both metals, cobalt(II) and nickel(II) are similarly plausible. Spin states and JT distortions have been discussed for structure pair 3 (Section 3.3.2.3), where the same two metals were considered for an octahedral complex.



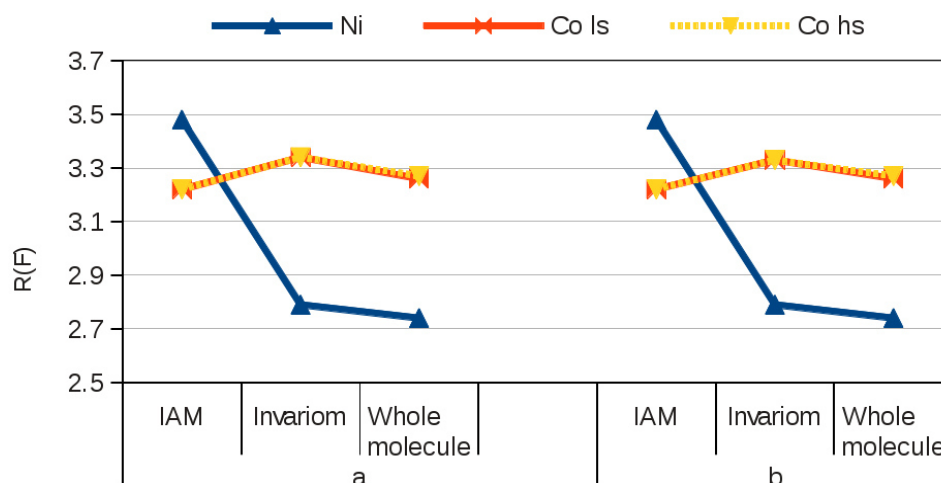
**Figure 3.8:** Thermal ellipsoid plot at a probability of 50 % of pair 4 with Ni after refinement of the whole-molecule scattering factors against data set **a**.

**Table 3.10:** Selected bond lengths from the final structure of pair 4.

Ni(1) - O(1)	2.1279(9)
Ni(1) - O(4)	2.0733(8)
Ni(1) - N(1)	2.0625(12)
Ni(1) - N(3)	2.0595(11)
Ni(1) - O(2)	2.1150(10)
Ni(1) - O(3)	2.1760(9)

Concerning the possibility of an JT distortion, the ellipsoids show no special elongation in the direction of the coordinating bond. But O(3) is farther away from the metal than O(2), although chemically equivalent. Similarly, O(1) is farther away than O(4) (see Table 3.10 and Figure 3.8). The bonds to O(2) and O(3) both oppose the donating nitrogen atoms and are on average longer than the ones opposing another oxygen atom. In general the arrangement is such that the opposing pair of O(2) and N(3) has both times the closest distance to the center among the equivalent atoms. So a JT deformation is possible. However,

the JT effect would have to be dynamic in this case. The different bond lengths could be caused by a slight inequality in the tilting of the two benzoate ligands, which would explain that the difference between the bond lengths of bonds to O(1) and O(4) is bigger than their deviation from the distance of O(2) to the central atom. Hence, there was no clear preference from a look at the atomic coordinates and a comparison using the diffraction data was required.



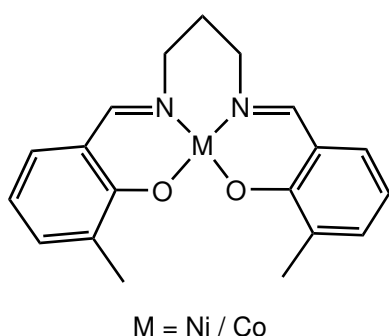
**Figure 3.9:** Comparison of  $R(F)$  for the refinements of the different metal atoms with different ED models against the two data sets of pair 4.

**Refinement results** In the IAM refinements the usual pattern that the lighter atom yielded a better fit to the data was observed. As displayed in Figure 3.9 this changes considerably for the invariom refinement.  $R(F)$  dropped from 3.48 % to 2.79 % for nickel(II) while it increased from 3.22 % to 3.34 % for cobalt(II). Both models profit from the inclusion of aspherical modeling around the central atom, but the improved models showed a clearly better fit for nickel(II) than for cobalt(II) in either spin state (Table 3.11).

**Table 3.11:** Selected computational and refinement results for pair 4.

	Ni	Co(hs)	Co(ls)
E(HF) crystal geometry [au]	-2955.783	<b>-2830.223</b>	-2830.202
R(F) against theo. Data [%]	0.45	0.48	0.46
R(F) whole molecule a [%]	<b>2.74</b>	3.27	3.26
R(F) whole molecule b [%]	<b>2.74</b>	3.27	3.26

### 3.3.2.5 Pair 5: [*N,N'*-Bis(3-methylsalicylidene)-1,3-propanediaminato]-nickel(II)/cobalt(II)



Pair	5	
	a <sup>[129]</sup>	b <sup>[130]</sup>
IUCr code	ci6619	sg6048
CSD code	TAQVEN	SAYZAU
CCDC No.	282297	296650
Space group	<i>A2<sub>1</sub>am</i>	
Peculiarity	M and C10 on special position	
Coord. geom.	square planar	
Metal ion	Ni <sup>2+</sup>	Co <sup>2+</sup>
Electron config.	[Ar]4s <sup>0</sup> 3d <sup>8</sup>	[Ar]4s <sup>0</sup> 3d <sup>7</sup>
Spin multiplicity	1(ls) (3 hs)	2(ls) (4 hs)

**Table 3.12:** Structural formula and selected crystallographic and chemical information of pair 5.

Structure **5a**<sup>[129]</sup> was published in 2005 with nickel as metal ion and structure **5b**<sup>[130]</sup> with cobalt. Otherwise the structures seem isomorphous. They contain a square-planar complex with a mirror symmetric tetra-dentate ligand that coordinates via two anionic oxygen and two nitrogen atoms. The two data sets differ by more than a scale factor, but isomorphism was questionable, because the two unit cells are identical within uncertainty limits.

**Chemical reasoning** Nickel(II) usually forms square-planar complexes, while cobalt(II) commonly assembles tetrahedral or octahedral coordination geometries. This may be the reason why from all the structures evaluated this was the only square-planar one, in which cobalt was one of the possible metals. In a CSD search for complexes that at least contain the ligand and either nickel or cobalt as central atom, cobalt has at least one more coordinative bond, whereas several crystal structures containing nickel in square-planar complexes with very similar ligand were found. An entry for the exact same structure but with an additional methanol molecule exists for nickel (TAQTUB). Therefore, nickel is more likely the correct central atom.

**Spin state** Theoretically the spin state of cobalt(II) in square-planar complexes would most likely be low spin. An explicit DFT calculation at the geometry obtained from invariom refinement against data set a shows (Table 3.13) that the low-spin complex for cobalt is



**Table 3.13:** Selected computational and refinement results for pair 5.

	Ni(ls)	Ni(hs)	Co(ls)	Co(hs)
E(HF) crystal geometry [au]	<b>-2504.0531</b>	-2504.0105	<b>-2378.4821</b>	-2378.4417
R(F) against theo. Data [%]	0.42	0.46	0.45	0.47
R(F) whole molecule a [%]	<b>3.34</b>	3.36	3.51	3.62
R(F) whole molecule b [%]	<b>3.39</b>	3.40	3.58	3.67

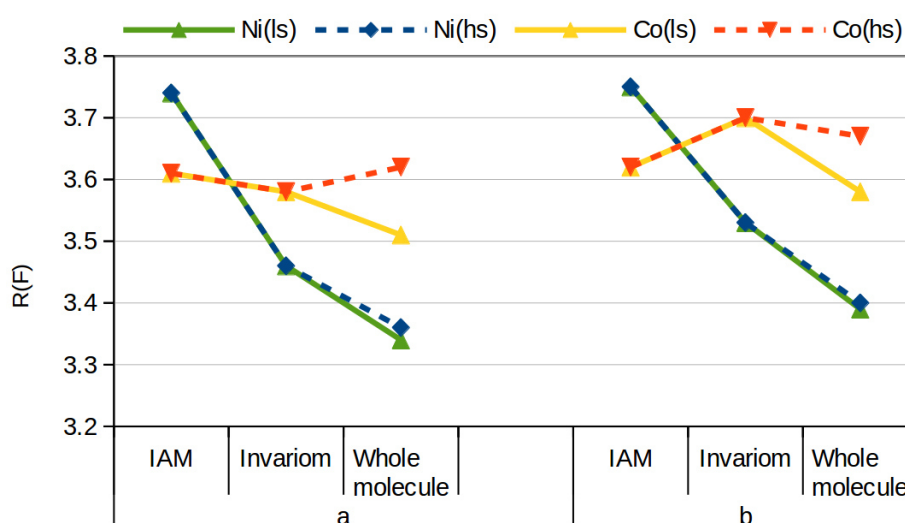
by 0.0404 Hartree (106kJ/mol) more stable in the gas phase. According to the equivalent calculation for nickel, low-spin configuration seems also energetically lower.

**Crystallographic refinement** From a crystallographic point of view the non-standard space group setting is noteworthy and two atoms are on a special position of symmetry  $m$ . Thus their  $z$  coordinates were not refined and  $U_{23}$  as well as  $U_{13}$  were constrained to zero. The space group furthermore requires fixing of  $x$  and  $y$  coordinates for one atom in order to prevent the floating origin from becoming a problem during refinement in XD.

SHELXL refinements as well as refinements of IAM in XD yielded a lower R-factor for cobalt than for nickel. Refinement with scattering factors from the invariom database reversed the relation of the R(F) values between cobalt and nickel, now favoring nickel in both data sets (Figure 3.10).

Refinement against the theoretical data showed no mirror symmetry for the donating oxygen atom. The multipoles at the metal were chosen according to the mirror symmetry at the atom. Cobalt modeled the theoretical data worse than nickel by refinement of otherwise equivalent multipole populations. For both metals the low-spin state could be described better by the multipole model than the high-spin data (Table 3.13).

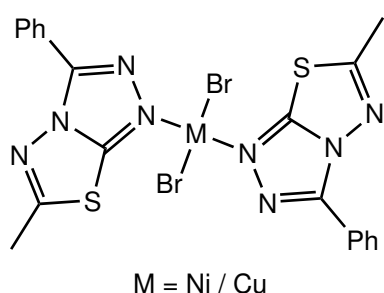
As shown in Figure 3.10 the 'whole-molecule' scattering factors refined against the experimental data of both structures improved the R(F) value of the invariom refinement for



**Figure 3.10:** Comparison of R(F) for the refinements of the different metal atoms with different ED models against the two data sets of pair 5.

nickel (hs and ls). For cobalt in the low-spin state the whole-molecule result improved with respect to the invariom result, too, but it stayed worse than for nickel by approximately the same difference in  $R(F)$ . High-spin cobalt yielded the worst result of all 'whole-molecule' refinements. The multipole parameters for low-spin nickel yielded the best results but the gap between the spin-states is small. For cobalt spin considerations were no longer of interest, since both crystals clearly contain nickel.

### 3.3.2.6 Pair 6: Dibromidobis(2-methyl-5-phenyl-*s*-triazolo[3,4-*b*][1,3,4]thiadiazole- $\kappa$ N)nickel(II)/copper(II)



Pair	6	
	a <sup>[131]</sup>	b <sup>[132]</sup>
IUCr code	bg2086	cf2132
CSD code	KIKBEM	XILXEW
CCDC No.	660111	664184
Space group	$P\bar{1}$	
Peculiarity	M on special position	
Coord. geom.	square planar	
Metal ion	Ni <sup>2+</sup>	Cu <sup>2+</sup>
Electron config.	[Ar]4s <sup>0</sup> 3d <sup>8</sup>	[Ar]4s <sup>0</sup> 3d <sup>9</sup>
Spin multiplicity	1(ls) (3 hs)	2

**Table 3.14:** Structural formula and selected crystallographic and chemical information of pair 6.

This sixth pair of coordination compounds is a square-planar nickel(II) or cobalt(II) complex with two bromide anions as ligands and two thiadiazolo ligands, which coordinate via a nitrogen atom. The data sets of the two structures **a**<sup>[131]</sup> and **b**<sup>[132]</sup> are identical (**b** is 103.09 times as intense as **a**) and unit cell parameters are equal except for one angle uncertainty. Thus, only one metal could be present and determination of the correct metal is the purpose of the investigation in this case.

**Refinement I** Interestingly, the metal atom with more electrons yielded the better  $R(F)$  in IAM refinement already. Therefore, the assignment of copper instead of nickel was more likely from the beginning.

**Chemical reasoning** Chemical reasoning could not argue against one of the metals. While bromide anions are weak ligands, nitrogen is a strong-field ligand. Accordingly copper could have square-planar geometry. A search of the CSD yielded that copper(II) coordinated by two bromide ions and two nitrogen donating ligands can be found in both square-planar and tetrahedral coordination geometries.

Nickel in square-planar geometry can be in a high or low-spin state. The HF energy of the low-spin state was lower by 0.023 Hartree (60 kJ/mol) when comparing single-point energies for the refined geometries after invariom modeling (see Table 3.15). Both spin states were included in the whole molecule procedure to investigate their influence on the aspherical model. The high-spin state could be described better by the multipole model,

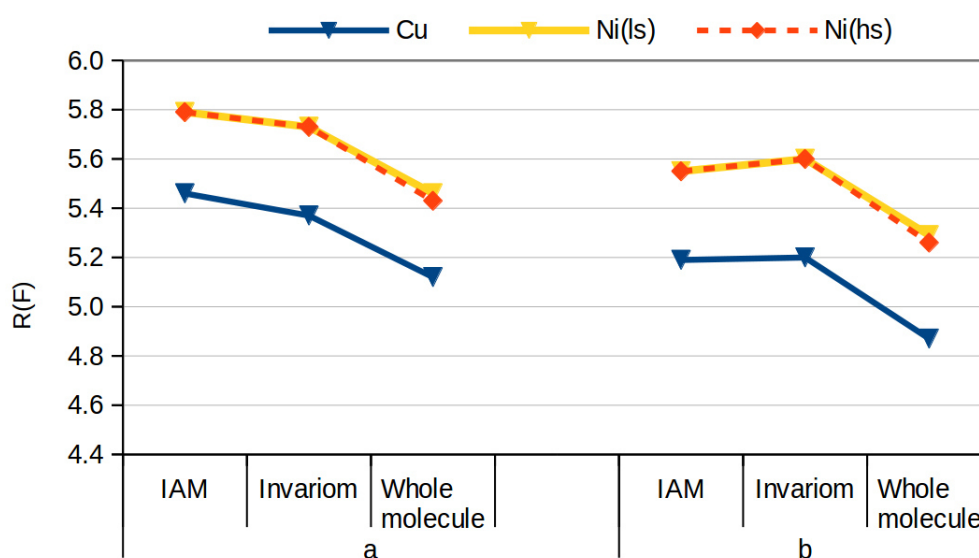
which was different to the observation for pair 5; so there is no general rule concerning multipole modeling of different spin states.

**Table 3.15:** Selected computational and refinement results for pair 6.

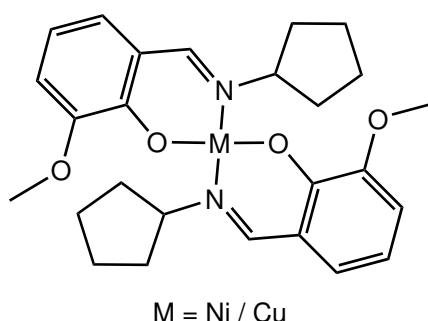
	Cu	Ni(hs)	Ni(ls)
E(HF) crystal geometry [au]	-8794.271	-8662.208	<b>-8662.231</b>
R(F) against theo. Data [%]	0.26	0.23	0.25
R(F) whole molecule a [%]	<b>5.12</b>	5.43	5.46
R(F) whole molecule b [%]	<b>4.87</b>	5.26	5.29

**Refinement results** Upon introduction of invariom scattering factors for the ligand the difference between the refinement results of the two metals stayed constant (Figure 3.11). Aspherical scattering factors for the whole molecule improved the fits to both data sets for both metals further, but the copper complex clearly remained the superior model for the data. That the high-spin state of nickel yielded better results than in the low-spin state can be explained by the population of similar orbitals for eight electrons, if two are unpaired, to nine electrons with one unpaired electron. In the low-spin case the highest occupied molecular orbital is different for nickel from the one in copper, so that the fit was worse. But compared to the overall value of R(F), the gap due to the spin state is insignificant. The lower R(F) values for the second data set originated in the smaller number of reflections therein. Data set **b** contains less reflections, whereas **a** includes reflections up to a resolution of 0.77 Å.

The two models including multipoles led to the conclusion that copper is the correct metal in this case.



**Figure 3.11:** Comparison of R(F) for the refinements of the different metal atoms with different ED models against the two data sets of pair 6.

3.3.2.7 Pair 7: Bis[6-(cyclopentyliminomethyl)-2-methoxypheno-  
lato]nickel(II)/copper(II)

Pair	7	
	a <sup>[133]</sup>	b <sup>[134]</sup>
IUCr code	cf2085	gk2077
CSD code	SEZFIN	YIDRUZ
CCDC No.	640293	650575
Space group	<i>Pbca</i>	
Peculiarity	M on special position ( $\bar{1}$ )	
coord. geom.	square planar	
Metal ion	Ni <sup>2+</sup>	Cu <sup>2+</sup>
Electron config.	[Ar]4s <sup>0</sup> 3d <sup>8</sup>	[Ar]4s <sup>0</sup> 3d <sup>9</sup>
Spin multiplicity	1(ls) (3 hs)	2

**Table 3.16:** Structural formula and selected crystallographic and chemical information of pair 7.

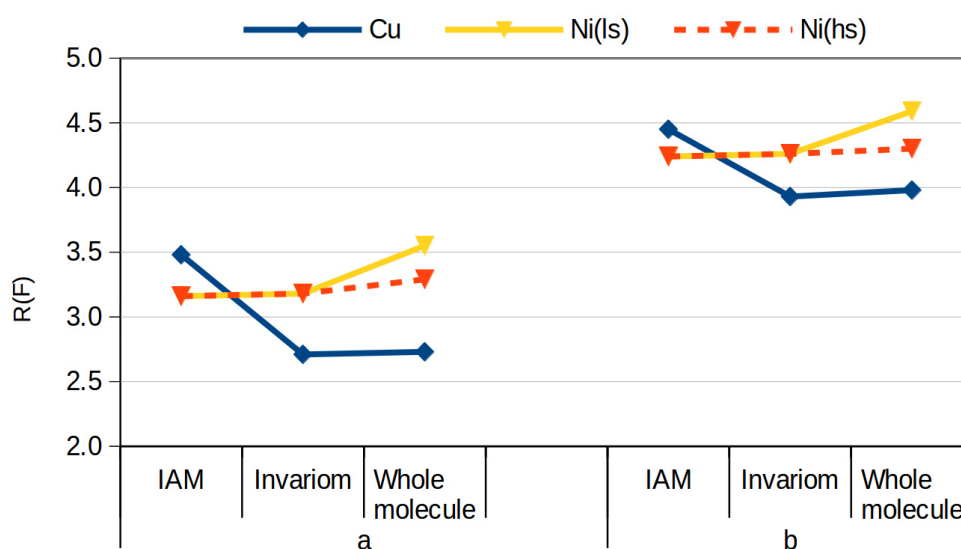
Complex pair 7 is similar to pair 6. It is also square-planar, crystallized with the metal on an inversion center and was published as a nickel<sup>[133]</sup> and a copper<sup>[134]</sup> compound. The reflection data, however, are non-identical for **a** and **b**. Hence, it was possible for both structures to contain different metals. As Table 3.17 shows, the unit cell parameters *b* and *c* differ. The unit cell volume of **a** is greater than the one of **b**, but **a** was reported to have been measured at a temperature 5 K higher than **b**. This temperature difference could explain the variation and prohibits a direct comparison of unit cell parameters. The crystal quality of **a** seemed higher, since more reflections were measured, cell e.s.d.s were lower and figures of merit (except the Goof) reported were superior in comparison to data set **b**.

**Table 3.17:** Cell parameters for the two datasets of pair 7.

Cell	<i>a</i>	<i>b</i>	<i>c</i>	T [K]
<b>a</b>	12.0620(10)	10.8110(10)	17.887(2)	298(2)
<b>b</b>	12.060(2)	10.8025(18)	17.863(3)	293(2)

**Refinement results** Comparing the IAM models for both metals, nickel(II) yielded a lower R(F) than copper for both data sets (see Figure 3.12). Upon inclusion of invariom scattering factors, the model with the heavier element, copper, improved considerably in both cases. The improvement was stronger for data set **a** but was also observed for the other data set. Including aspherical scattering factors for the metals provided some technical difficulties and in the end no further improvement was observed for either metal, but both data sets were modeled much better by copper as central atom. For data set **a** the difference to nickel in either spin state increased, and although the gap between the models with the different metals was smaller for **b**, it was still significant.

Energetically (Table 3.18) the low-spin state of nickel would be favored in this coordination geometry, but since nickel did not turn out to be correct in either case, the spin



**Figure 3.12:** Comparison of  $R(F)$  for the refinements of the different metal atoms with different ED models against the two data sets of pair 7.

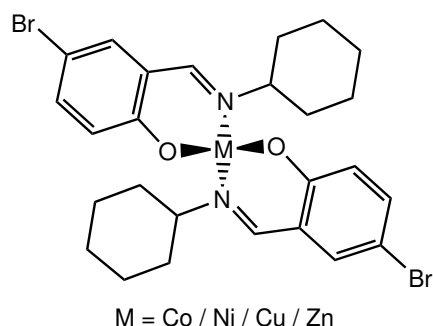
state's influence on refinement was not of special interest. It can however be explained in the same way as for pair 6.

Overall, it could be determined that copper was the metal in both structures. Although the data sets differed by more than a scale factor, aspherical scattering factors showed clearly that copper fitted better to both sets of diffraction data than nickel.

**Table 3.18:** Selected computational and refinement results for pair 7.

	Cu	Ni(hs)	Ni(Is)
E(HF) crystal geometry [au]	-3060.527	-2928.348	-2928.362
R(F) against theo. Data [%]	0.49	0.46	0.52
R(F) whole molecule a [%]	<b>2.73</b>	3.29	3.55
R(F) whole molecule b [%]	<b>3.98</b>	4.30	4.59

## 3.3.2.8 Pair 8/9: Bis[4-bromo-2-(cyclohexyliminomethyl)phenolato]-cobalt(II)/nickel(II)/copper(II)/zinc(II)



Pairs	8		9	
	a <sup>[136]</sup>	b <sup>[135]</sup>	a <sup>[137]</sup>	b <sup>[138]</sup>
IUCr code	ci6575	ci2197	ob6467	ci6692
CCDC No.	274365	629307	263535	281782
Space group	<i>Pbca</i>			
Peculiarity	<i>none</i>			
Coord. geom.	tetrahedral			
Metal ion	Ni <sup>2+</sup>	Co <sup>2+</sup>	Cu <sup>2+</sup>	Zn <sup>2+</sup>
Spin multiplicity	3	2, 4	2	1

**Table 3.19:** Structural formula and selected crystallographic and chemical information of pair 8 and 9.

In this case four isomorphous bis[4-bromo-2-(cyclohexyliminomethyl)phenolato]-complexes of the 3d-metals cobalt(II),<sup>[135]</sup> nickel(II),<sup>[136]</sup> copper(II)<sup>[137]</sup> and zinc(II)<sup>[138]</sup> were investigated. Their single crystal XRD data sets were all different from each other. In contrast the unit cell constants for 8a and 8b are the same and for 9a and b they differ only by 0.2 % (see Table 3.20).

**Table 3.20:** Cell parameters in Å for pair 8 and 9.

cell	a	b	c
8a	14.979(3)	13.609(3)	25.164(5)
8b	14.9790(10)	13.6090(10)	25.1640(10)
9a	14.9960(10)	13.5970(10)	25.156(2)
9b	14.9830(10)	13.5870(10)	25.143(2)
max. Δ	0.017	0.022	0.021
max. Δ [%]	0.1	0.2	0.1

**Chemical reasoning** The complex(es) crystallized in a tetrahedral coordination geometry as a racemate in space group *Pbca*. At first sight it seemed very unlikely for all four metal ions to crystallize in the same coordination despite the fact that they differ by up to three electrons.

While zinc is commonly found in tetrahedral coordination geometry, copper usually forms JT distorted octahedra or square planar complexes<sup>[151]</sup> if the ligands are weak field and thus induce a small ligand-field splitting. But strong-field ligands, leading to a large ligand-field splitting, as well as bulky ligands can prompt a more tetrahedral arrangement. Whereas amines are strong field ligands, yielding a high ligand field splitting, the hydroxide anion is a weak field ligand. Therefore, no clear conclusions could be derived concerning copper from ligand field theory.

There is an example with a voluminous ligand<sup>[152]</sup> and distant bromide atoms as fifth and sixth coordinating atoms that completed a JT distorted octahedron where the inner ligating atoms' arrangement resembled a tetrahedron. Regular tetrahedrons of copper(II)

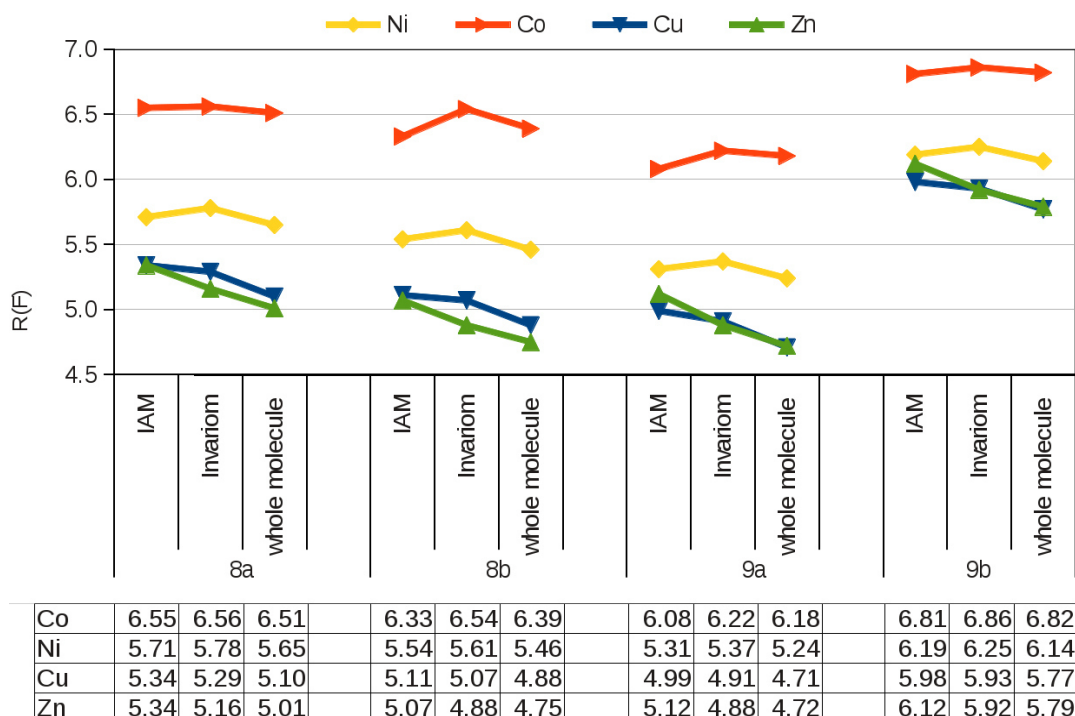
**Table 3.21:** Bond angles for pair **8** and **9** from the zinc model with aspherical scattering factors for the whole molecule.

angle	O1 Zn O2	N1 Zn N2	O1 Zn N1	O1 Zn N2	O2 Zn N1	O2 Zn N2
8a	119.65(11)	122.46(12)	93.74(11)	113.76(12)	113.81(12)	95.48(11)
8b	119.86(10)	122.42(11)	93.84(10)	113.71(10)	113.75(11)	95.36(10)
9a	119.94(10)	122.40(10)	93.66(10)	113.85(10)	113.74(10)	95.36(10)
9b	119.82(13)	122.47(13)	93.72(13)	113.93(13)	113.70(13)	95.30(13)

are not stable<sup>[151]</sup> due to the JT effect of the  $t_2$  orbitals. But looking at the bond angles listed in Table 3.21 the tetrahedral coordination is not perfect in the complexes **8** and **9** either. No additional coordination contacts at longer distances are present in the structure model, only hydrogen atoms at 3.02 Å and 3.71 Å distance, the closer one belonging to the cyclohexyl group.

Since, according to Cinčić and Kaitner<sup>[153]</sup>, Schiff-base ligands are known to form almost tetrahedral coordination geometries for copper(II), chemical reasoning alone could not exclude copper(II) as the metal center.

**Spin state** The spin state of cobalt and nickel can be either high-spin as in tetrahedral or low-spin as in square-planar geometries, since the coordination geometry is in between tetrahedral and square-planar. DFT results show a preference of high spin for cobalt and nickel in single-point energy calculations at experimental molecular geometries.

**Figure 3.13:** Comparison of R(F) for the refinements of the different metal atoms with different ED models against the data sets of pairs **8** and **9**.

**Refinement results** Refinements with all four metals for each of the four data sets were performed. For each data set cobalt fitted least well and nickel did not fit well either. Copper and zinc yielded the best residuals. From experience with other structure pairs the better fit of heavier atoms in the IAM, in spite of their formally positive charge, was a first indication to the incorrectness of the lighter element.

Upon aspherical modeling of the ligands the cobalt and nickel models did not improve, but those with copper and zinc did. Zinc improved the most which in the case of **8a** and **b** led to a lower R(F) for zinc than copper, since in the IAM both fitted almost equally well. For **9a** and **b** the formerly worse fitting zinc became almost as good as copper.

Inclusion of multipoles for the central atom in the models led to better modeling of the ED throughout. For **8a** and **b** zinc stayed the best fitting metal. The two data sets of **9** were less distinctive concerning copper and zinc. **9a** and **9b** behaved similarly as did **8a** and **b**. Considering the unit cell constants, pair **8** and **9** could still differ.

**Considering isomorphism and geometrical aspects** Bond lengths for bonds involving the metal atom do not differ significantly as shown in Table 3.22. Very few isostructural Schiff-base complexes are known in the literature<sup>[153–155]</sup>, especially none for zinc and copper, rendering isomorphism of copper and zinc unlikely. Since two isostructural complexes with only minor geometry changes (around 0.01 Å for bonds to nitrogen<sup>[154]</sup>) was found for metals differing by two (cobalt and copper<sup>[154]</sup>) and three electrons (cobalt and zinc<sup>[153]</sup>), copper could not be definitely excluded as a possibility here.

**Table 3.22:** Bond distances for pair **8** and **9** from the zinc model with aspherical scattering factors for the whole molecule.

	Zn(1) O(1)	Zn(1) O(2)	Zn(1) N(1)	Zn(1) N(2)
<b>8a</b>	1.914(3)	1.913(3)	2.028(3)	2.023(3)
<b>8b</b>	1.913(2)	1.915(2)	2.025(3)	2.026(3)
<b>9a</b>	1.913(2)	1.916(2)	2.032(3)	2.031(3)
<b>9b</b>	1.914(3)	1.914(3)	2.029(3)	2.033(3)

**Energy considerations** In order to get a better idea if **9** contains copper or zinc the geometry for the two complexes was optimized by the same DFT method as the single-point calculation for acquiring the molecular ED. The gain in energy upon geometry relaxation was bigger for copper comparing to both starting geometries from structure **8** and **9** (Table 3.23<sup>6</sup>). Using this information, zinc theoretically fits better to the geometry determined by refinement of the invariom models against the XRD data.

**Conclusion** In summary, refinement results show that structures **8a** and **8b** contain zinc and definitely not cobalt or nickel as the metal center. For structure **9a** and **9b** data

<sup>6</sup>The results for starting from geometries **b** instead of the presented results for **a** yield even smaller differences between data sets **8** and **9**.

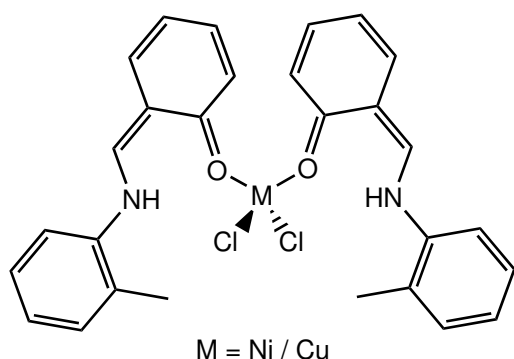


**Table 3.23:** Comparison of single point SCF energies and those after geometry optimization for pair **8** and **9** with copper and zinc.

Structure	<b>8</b>		<b>9</b>	
	Cu	Zn	Cu	Zn
Starting geometry	-8057.0633	-8195.9862	-8057.0629	-8195.9856
Optimized geometry	-8057.0819	-8195.9933	-8057.0819	-8195.9933
Ratio	1.00000230	1.00000087	1.00000235	1.00000093
Change in %	0.00023%	0.00009%	0.00023%	0.00009%
Difference in au	-0.0185	-0.0071	-0.0189	-0.0076
Change in kJ/mol	-48.7	-18.6	-49.7	-20.0

quality was not high enough to distinguish between copper and zinc. Judging from the refinement, cell parameter, bond angles and lengths comparisons, it was very likely that both data sets for **9** contain the same element. In combination with what is known about copper complexes and isomorphism the probability for **9** to also contain zinc is high. The outcome of QM calculations confirmed this and led to a strong indication that zinc is the metal center for all of the data sets.

## 3.3.2.9 Pair 10:

Dichlorobis[2-(*o*-tolyliminomethyl)phenolato]nickel(II)/copper(II)

Pair	10	
	a <sup>[139]</sup>	b <sup>[140]</sup>
IUCr code	cf6444	is2046
CSD code	TAQREJ	ADIFID
CCDC No.	282269	613739
Space group	<i>C</i> 2/ <i>c</i>	
Peculiarity	M on special position (2)	
Coord. geom.	tetrahedral	
Metal ion	Cu <sup>2+</sup>	Ni <sup>2+</sup>
Electron config.	[Ar]4s <sup>0</sup> 3d <sup>9</sup>	[Ar]4s <sup>0</sup> 3d <sup>8</sup>
Spin multiplicity	2	3

**Table 3.24:** Structural formula and selected crystallographic and chemical information of pair 10.

Pair 10 is the second tetrahedral complex investigated and the metal is coordinated by two chloride anions and two oxygen atoms, which are part of an almost completely planar ligand. The complex with either nickel(II)<sup>[140]</sup> or copper(II)<sup>[139]</sup> crystallized with the metal atom on a twofold rotation axis. The reflection data are identical, except for a lower resolution of **b**; so only one metal can be correct.<sup>7</sup>

**Chemical reasoning** Past results and chemical knowledge implicate that nickel(II) is more likely to be found in a tetrahedral coordination geometry than copper(II). There have been cases of close-to-tetrahedral copper(II) complexes,<sup>[155,156]</sup> however, rendering a tetrahedron a "commonly observed geometry"<sup>[10]</sup> for copper. Multidentate ligands or counterions as in Cs<sub>2</sub>[CuCl<sub>4</sub>]<sup>[157]</sup> can promote at least pseudo-tetrahedral arrangements, hence, copper is chemically possible.

Bond angles for the 'whole-molecule' structure with nickel refined against dataset **a** are shown in Table 3.25<sup>8</sup>. The bond angles indicate a larger deviation from a square-planar coordination geometry than from a tetrahedral one, but they neither belong to a perfect tetrahedron.

**Refinement results** For this structure refinement results were not particularly good. With R(F) higher than 6.5% for the higher resolution data set, the models had the worst fit to the XRD data of all cases investigated. Only the fit for **3b** was worse, but that was due to several disordered water molecules. Here no real reason for the residual factor being

<sup>7</sup>Something else that came to attention when studying the compound was the placement of a double bond at the nitrogen in the structural formula. As it is presented in the literature<sup>[139,140]</sup> there should at least be a positive charge at the -NH=. A better representation of the structure is the formula depicted in Table 3.24, since although the bond drawn as a double bond has a length of 1.42566 Å, which is shorter than a common aromatic bond, the following bond to the nitrogen atom is even closer to a common carbon-nitrogen double bond (1.29 Å<sup>[10]</sup>) with a length of 1.29 Å. Additionally, the carbon-oxygen bond is with 1.297 Å definitely closer to a double bond (1.21 Å<sup>[10]</sup>) than a single bond (1.43 Å<sup>[10]</sup>).

<sup>8</sup>For atom labels see the ORTEP type plot in Appendix A, Figure 7.1

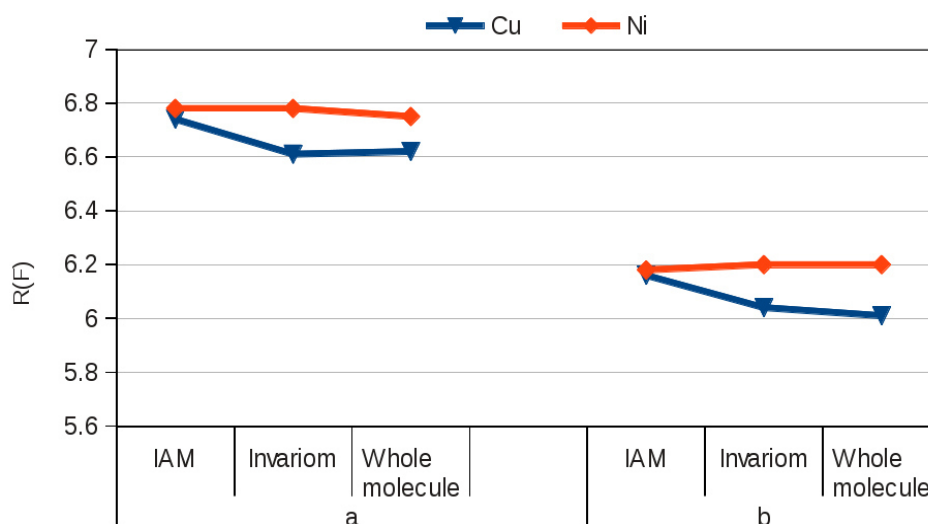
**Table 3.25:** Selected bond angles from the final structure of pair 10.

Cl(1) - Ni(1) - Cl(1)_1	110.22(9)
Cl(1) - Ni(1) - O(1)	106.25(10)
Cl(1) - Ni(1) - O(1)_1	115.76(11)
O(1) - Ni(1) - O(1)_1	102.7(2)

so high was found. The highest difference density peaks were located around the metal atom and since the residual density map was featureless besides the region around the metal center, there was no disorder. The data is simply very noisy.

Still, an invariom refinement led to a more improved fit for copper than for nickel. The inclusion of aspherical scattering factors for the coordination center had quite different effects for set **a** and **b**, although they do not differ in the low order data. This raised the question how accurately the inner data were determined. The fit to the data was best for copper(II) in both cases, so that at this point copper seems the correct metal.

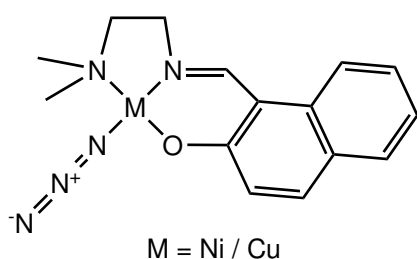
Since copper is not the likeliest element for this tetrahedral complex, IAM and invariom refinements were also performed for zinc, resulting in a poorer fit than for copper.



**Figure 3.14:** Comparison of R(F) for the refinements of the different metal atoms with different ED models against the two data sets of pair 10.

**Results from theoretical investigation** Since the crystallographic results did not agree well with chemical intuition and the data quality is low, the relative energies of both complexes were determined for atomic coordinates from the invariom refinements and geometry optimized ones. The energy change upon geometry relaxation was identical for high-spin nickel and cobalt, so that in this case only a slight preference of copper could be derived from the experimental data. However, the trust in the XRD data of this compound is low.

## 3.3.2.10 Pair 11: Azido[1-(isobutylaminomethyliminomethyl)-2-naphtholato]-nickel(II)/copper(II)



Pair	11	
	a <sup>[141]</sup>	b <sup>[142]</sup>
IUCr code	su6164	sj2024
CSD code	FEYXEN	TEJMAX
CCDC No.	263585	608486
Space group	<i>Pbca</i>	
Peculiarity	slight disorder of <i>en</i> ligand	
Coord. geom.	square planar	
Metal ion	Ni <sup>2+</sup>	Cu <sup>2+</sup>
Electron config.	[Ar]4s <sup>0</sup> 3d <sup>8</sup>	[Ar]4s <sup>0</sup> 3d <sup>9</sup>
Spin multiplicity	3	2

**Table 3.26:** Structural formula and selected crystallographic and chemical information of pair 11.

For the azido[1-(isobutylaminomethyliminomethyl)-2-naphtholato] complex shown in Table 3.26 two structures, one with nickel(II)<sup>[141]</sup> and one with copper(II)<sup>[142]</sup>, are stored in the CSD. Both are square-planar complexes with a slightly disordered ethylenediamine ligand. The data sets are different from each other, but the unit cells are very similar (Table 3.27). A comparison of bond lengths and angles (Table 3.28) did not reveal any significant differences between the two models refined against the different data sets either (see Figure 7.2 in Appendix A for atom names). Therefore, it is very likely that only one metal is correct. Copper(II) is the more likely metal center for square-planar coordination geometries and in this case the coordination geometry was not enforced by the ligand, since the azide could move out of the plane.

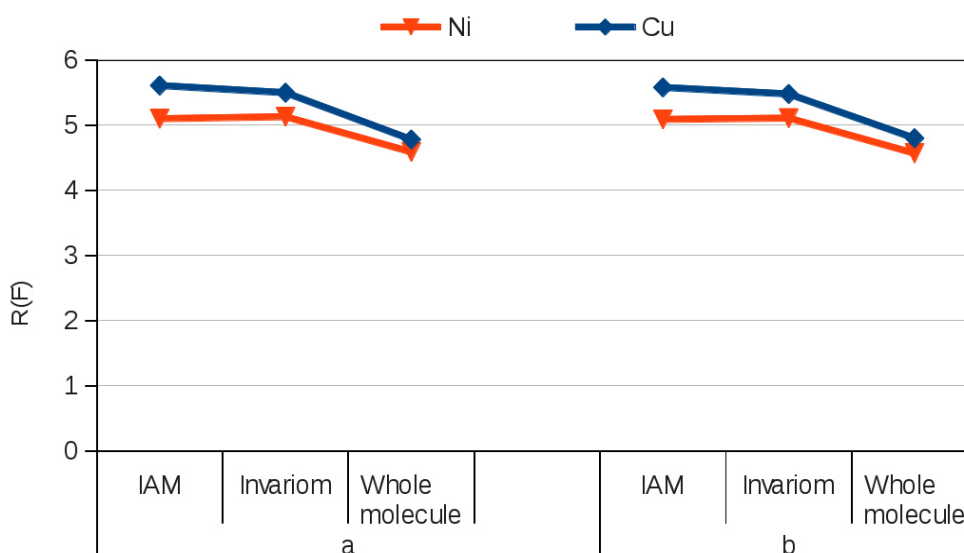
**Table 3.27:** Cell parameters in Å for pair 11 **a** and **b**, both measured at 298(2) K.

	<i>a</i>	<i>b</i>	<i>c</i>
<b>a</b>	7.5760(10)	13.300(2)	30.306(2)
<b>b</b>	7.5680(10)	13.3060(10)	30.280(2)
Δ [Å]	0.008	0.006	0.026
Δ [%]	0.1	0.04	0.8

**Refinement results** This structure pair was problematic due to its disordered ethylene group. The second conformation of the disorder was populated by less than 5%, but it hindered the correct element assignment from refinement results. The disorder could only be modeled by several strong restraints in *SHELXL*, which could not be treated properly in *XD*. Since the disorder accounted for less than one electron, aspherical refinements without description of the disorder were attempted with *XD*.

However, hardly any improvement could be seen upon invariom modeling of the ligand environment as illustrated in Figure 3.15. Although the model containing copper, which was the poorer IAM model, improved more than the one with nickel, the latter remained superior in fitting the XRD data. The same trend was observed for the aspherical modeling

around the metal center. Since this part of the model is farther away from the disorder, the improvement was stronger than for the invariom model. Again, the copper model improved more than the nickel model, but the influence was not enough to show an obvious preference for copper.



**Figure 3.15:** Comparison of R(F) for the refinement of the different metal atoms with different ED models against the two data sets of pair **11**.

Disorder clearly is a problem for aspherical model performance in distinguishing two metals. Hence, this case presented an example for a limit of this method.

Overall, the two data sets behaved very similar during the different refinements, supporting the hypothesis that only one metal is correct. The trends of the refinements suggest copper as the better model if the disorder would not damp improvement. Unfortunately, the evidence based upon the diffraction data is not strong enough to prove one structure as unambiguously correct.

**Table 3.28:** Selected bond lengths and angles from the final structures of **11 a** and **b**.

a		b	
Ni(1) - O(1)	1.8195(17)	Cu(1) - O(1)	1.8193(17)
Ni(1) - N(1)	1.846(2)	Cu(1) - N(1)	1.845(2)
Ni(1) - N(2)	1.958(2)	Cu(1) - N(2)	1.959(2)
Ni(1) - N(3)	1.905(2)	Cu(1) - N(3)	1.905(2)
O(1) - Ni(1) - N(1)	93.85(9)	O(1) - Cu(1) - N(1)	93.90(9)
O(1) - Ni(1) - N(2)	178.90(9)	O(1) - Cu(1) - N(2)	178.83(9)
O(1) - Ni(1) - N(3)	89.50(9)	O(1) - Cu(1) - N(3)	89.46(9)
N(1) - Ni(1) - N(2)	87.05(9)	N(1) - Cu(1) - N(2)	87.03(9)
N(1) - Ni(1) - N(3)	176.12(10)	N(1) - Cu(1) - N(3)	176.12(10)
N(2) - Ni(1) - N(3)	89.57(10)	N(2) - Cu(1) - N(3)	89.59(10)

## 3.4 Summary and perspective

In summary, eleven pairs of crystal structures of coordination compounds were modeled by aspherical scattering factors to identify the correct metal atom. In seven cases the identification from a fit to XRD data was successful and a definite result was obtained.

Limitations of the invariom-like approach were explored by investigations of three more difficult structures. Data quality does not have to be excellent, since for all successful cases room temperature measurements were sufficient. In case of very noisy data, however, results are not definite enough to compete with chemical probability, as in the case of pair **10**: the result from the fit to XRD data suggested copper(II) for a tetrahedral complex, but since the data quality was low, only a direction for further inquiries was suggested.

Case **11** showed the limits concerning disorder: the compound was too disordered in the crystal to use the diffraction data as evidence.

Lastly, for the quartet of similar structures, two of the four metals could be excluded by evaluating the fit of the models to the data. For two of those structures zinc could be identified as correct, but for the remaining two the final suggestion for the correct metal was derived from energy considerations, by comparing the changes in energy upon relaxation of the crystal geometry. Similar to case **10** data quality was a slightly limiting factor.

Table 3.29 summarizes the results for the different coordination geometries.

In conclusion, it was demonstrated that invarioms are a helpful tool to identify the correct metal atom in structures of at least moderate data quality and resolution<sup>9</sup>. In contrast, the information of an IAM refinement was usually not enough for the differentiation of the metals. Creating and using aspherical scattering factors for the whole molecule, including the metal center, could increase the model quality further, but it was seldom necessary for identification of the metal atom.

Therefore, future investigations of dubious structure pairs should employ scattering factors from the invariom database for the ligand. Only in cases where the results are not sufficiently convincing, treatment of the whole molecule with aspherical scattering factors would be worth the extra effort. An example for such a case, in which almost no improvement was observed upon invariom modeling is case **6b**. In most cases, however, invariom modeling should improve the model enough to distinguish between two metal atoms.

---

<sup>9</sup>of around 0.83 Å.

**Table 3.29:** Overview of structure pairs investigated and the results after invariom-like treatment.

pair	coordination geometry	reference	metal	data	invariom conclusion	remarks
1	square-planar	[121]	Cu	identical	Cu correct	M on special position
		[122]	Ni			
2	octahedral	[123]	Ni	identical	Cu correct	3 water in asym. Unit
		[124]	Cu			
3	octahedral	[125]	Ni	not identical	Ni correct	also compared to copper 4 water in asym. Unit
		[126]	Co			
4	octahedral	[127]	Ni	identical	Ni correct	
		[128]	Co			
5	square-planar	[129]	Ni	not identical	Ni correct	M and C10 on special position space group A21am
		[130]	Co			
6	square-planar	[131]	Ni	identical	Cu correct	M on special position
		[132]	Cu			
7	square-planar	[133]	Ni	not identical	Cu correct	M on special position
		[134]	Cu			
8	tetrahedral	[136]	Ni	not identical	most likely Zn	isomorphous to <b>9</b>
		[135]	Co			
9	tetrahedral	[137]	Cu	not identical	Cu or Zn	isomorphous to <b>8</b>
		[138]	Zn			
10	tetrahedral	[139]	Cu	identical	Cu fits better	below average data quality, different resolutions M on special position
		[140]	Ni			
11	square-planar	[141]	Ni	not identical	ambiguous	identification prevented by ligand disorder (hinting at Cu)
		[142]	Cu			





# 4 Invariom point charges

*Many parts of this chapter were published as a scientific article in ChemPhysChem, therefore several pictures are reprinted with permission from Wiley and most content is similar to the one in the publication.*<sup>[100]</sup>

## 4.1 Introduction

### 4.1.1 Motivation

This chapter reports the development, application and validation of look-up point charges for reproducing the electrostatic potential (ESP) of molecules. In computational chemistry, modeling and medicinal chemistry the molecular ESP has become an established tool.<sup>[158–161]</sup>

Look-up point charges should be transferable between atoms of the same local chemical environment in different molecules. This work relies on the invariom criteria for classification of transferable atom-types.<sup>[80,81]</sup> The collection of geometry-optimized model compounds within the invariom database, invariom notation and empirical transferability rules of the invariom formalism are the foundation for the development of these look-up point charges.

This project's aim is to get one step closer to fully automate force-field parameterization for molecular dynamics (MD)<sup>[101]</sup> simulations of organic molecules. Such MD simulations could for example allow a correct inclusion of disorder in crystallographic model refinement. An example of rotational disorder that would benefit from such modeling can be found in the literature.<sup>[162]</sup> PLATON's 'squeeze'<sup>[163]</sup> procedure is currently regularly applied for modeling disorder, but models based on more information are preferable. Hence there is a strong motivation to incorporate MD results in crystal structure models. This overall goal is the reason for the development of transferable invariom point charges that reproduce the molecular ESP.

These new transferable point charges will be applied to and validated for nineteen angiogenesis inhibitor molecules, which are test cases of broader pharmaceutical relevance. Their Lewis-structures are displayed in Table 4.1 and they show the wide variety of functional groups included, which makes the molecules a suitable test set.

Towards the end of this work a way to rapidly obtain a graphical representation of a molecular ESPs based on the invariom charges assigned will be described for atomic models from structural data. During this procedure, which involves point-charge assignment, it is necessary to modify the bond distances to hydrogen atoms, if the source of the atomic coordinates is a crystal structure model. Although single-crystal XRD usually provides atomic positions very well, those of hydrogen atoms deviate systematically from the theoretical ones and positions from neutron diffraction.<sup>[164]</sup> Models from invariom multipole and *Hirshfeld*-atom refinements are exceptions to this.<sup>[44,45,88]</sup> A program for quickly calculating

the molecular ESP and generating a rapid and informative visualization, starting from a *pqr* file<sup>1</sup> is presented. Its source code relies on that of MOLECOOLQT.<sup>[40]</sup>

### 4.1.2 Test set of angiogenesis inhibitor molecules

As a test set for validating the invariom point-charge procedure, 19 angiogenesis inhibitors were chosen, due to their wide variation of functional groups. They are displayed in Table 4.1.

Angiogenesis inhibitors prevent blood vessel growth and since cancer cells multiply at a faster speed than normal cells, they are good candidates for possible anti-cancer drug molecules.<sup>[165]</sup> Usually the angiogenesis inhibitor drug has to be combined with other chemotherapeutic drugs that can actually kill and not only limit growth of tumor cells. For treatment of different types of cancer a variety of substances have been developed. Their mechanisms of action fall into different classes. Two big categories can be distinguished: monoclonal antibodies like the popular drug *bevacizumab* and smaller organic molecules. In this study the focus was on small-molecule inhibitors that prevent functioning of vascular-endothelial growth factor receptor tyrosin kinase.<sup>[166]</sup> The reason why monoclonal antibodies were avoided was their size, which rendered them unsuitable for performing quantum chemical computations and thus the procedures used for validation of the developed point charges.

### 4.1.3 Point charges

Atomic charges in a molecule are not a clearly defined property.<sup>[167]</sup> Nevertheless, atomic point charges are a widely used concept in chemically or biologically motivated computational modeling. This is mostly due to their ease of use and conceptual simplicity.

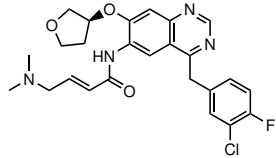
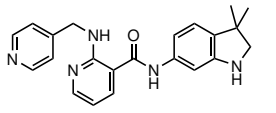
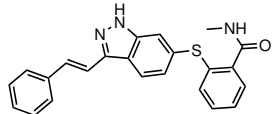
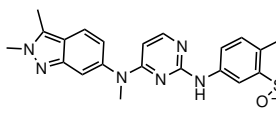
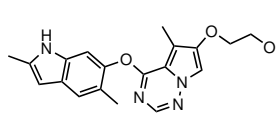
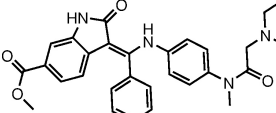
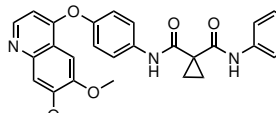
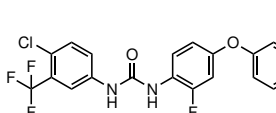
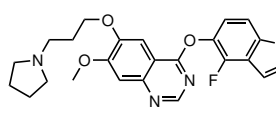
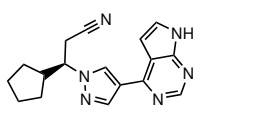
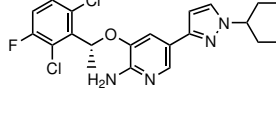
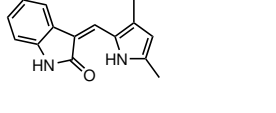
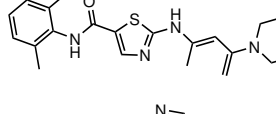
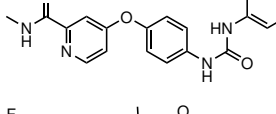
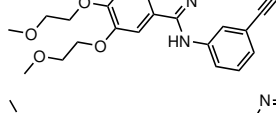
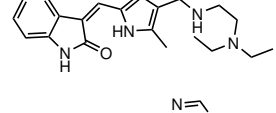
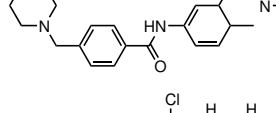
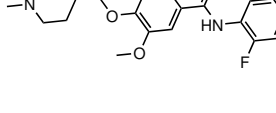
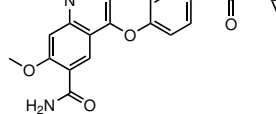
*Bader* charges<sup>[18]</sup> are one of the most rigorously defined atomic charges in quantum chemistry. They are not as basis-set dependent as *Mulliken* charges,<sup>[168,169]</sup> an advantage shared by charges from natural population analysis.<sup>[170]</sup> Alternatively, *Hirshfeld* partitioning<sup>[41]</sup> is an equally elegant way of obtaining charges. However, the disadvantage of all these approaches lies in the requirement of having a molecular EDD available. Usually the source of the EDD is a QM calculation except for *Bader* charges. Like Monopole- $\kappa$  charges, *Bader* charges can be derived from XRD experiments and have therefore found use in crystallography.<sup>[171]</sup>

Several procedures<sup>[172-176]</sup> have been established to obtain atomic point charges for force-field parametrization. They are efficient, well-tested and robust. Alongside other force-field parameters those charges have been used for molecular mechanics and dynamics simulations, in programs like AMBER<sup>[177]</sup>, CHARMM<sup>[178]</sup> and GROMACS.<sup>[179]</sup> Such point charges, however, still require at least a semi-empirical computation of the whole molecule of interest if it is not a common amino acid, nucleotide, carbohydrate or lipid. Thus the size of systems that are possible to study is limited. Partitioning a system into smaller fragments more amenable to computation is required as soon as systems become larger (>150 atoms), e.g. metalloproteins or supramolecular structure assemblies.

---

<sup>1</sup>*pqr* files are similar in format to protein database files,<sup>[5]</sup> but the column of site occupancy factors is replaced by atomic charges; the column with the Wilson *B* factor contains the *van-der-Waals* radii.

**Table 4.1:** Overview of small-molecule angiogenesis inhibitors used as test-set. API stands for active pharmaceutical ingredient. © reprinted with permission from Wiley.

API	Molecular formula	API	Molecular formula
afatinib		motesanib	
axitinib		pazopanib	
brivanib		nintedanib	
cabozantinib		regorafenib	
cediranib		ruxolitinib	
crizotinib		semaxanib	
dasatinib		sorafenib	
erlotinib		sunitinib	
imatinib		vandetanib	
lenvatinib			

Therefore a look-up table of atomic point charges fitted to a molecular ESP<sup>[173]</sup> (Section 4.1.3.1) was recently developed.<sup>[180]</sup> Look-up tables offer direct access to charges, thus reducing computing time and overcoming the limitation in the size of a system. This general idea of the Transferable Partial Atomic Charge Model (TPACM4) is followed, but not the procedural details. Atomic point charges derived according to *Merz and Kolman*<sup>[181]</sup> (MK) from a fit to the ESP of the model compounds in the generalized invariom database<sup>[81]</sup> are presented here.

### 4.1.3.1 Point charges for force-field applications

The parameter set 'amberff'<sup>[182]</sup> provides very specific charges for peptides, sugars and nucleotides for the popular molecular dynamics program AMBER.<sup>[177]</sup> The same applies to the CHARMM additive all-atom force field<sup>[178]</sup> which additionally contains parameters for lipids.<sup>[183]</sup> Charges for unparameterized compounds can be computed by several suggested procedures:

- An established way is the computation of RESP<sup>[173]</sup> charges. It is the same method which was applied for the parameterization of 'amberff'. A HF/6-31G(d) quantum chemical computation is required for the complete target molecule, followed by a calculation of the molecular ESP from the obtained wave function. The atomic point charges are then fitted to reproduce this ESP by a restrained least-squares algorithm. The restraints ensure that the inner, less well defined atomic charges stay close to neutral. After the quantum-mechanics program has calculated the ESP from the wave function, the program ANTECHAMBER<sup>[184]</sup> can fit the RESP charges, while the unrestrained MK method<sup>[181,185]</sup> is implemented in many programs, e.g. in GAUSSIAN09<sup>[103]</sup>. The disadvantage of this approach is the computation and time demand for the study of large molecules.<sup>2</sup>
- Bond-Charge Correction (BCC) is a semi-empirical method introduced by Bayly *et al.*<sup>[176]</sup> in 2000. This procedure derives molecule specific atomic point charges from AM1<sup>[187]</sup> semi-empirical point charges<sup>[188,189]</sup> via scaling by bond-charge correction factors from a table. They are designed to reproduce the RESP charges at a lower computational cost.<sup>3</sup>
- The general AMBER force field (GAFF)<sup>[191]</sup> mainly provides all other force field parameters, because its use for charges has been superseded by the two methods above.

The computing time required for those two options depends on molecular size. In the interest of procedure speed up, the look-up table TPACM4 for RESP charges was developed.<sup>[180]</sup> Its implementation is similar to the website of PDB2PQR.<sup>[192,193]</sup> Both work on a web server that converts a *pdb* file, into a *pqr* file. Typically the molecules for TPACM4 are organic. Another database designed to store RESP and MK point charges is REDDB.<sup>[194]</sup> Differently to other methods described above, the main purpose of REDDB is just the storage and not

---

<sup>2</sup>HF as well as DFT methods asymptotically scale as order  $N^2$  for large systems with  $N$  being the number of basis functions.<sup>[186]</sup>

<sup>3</sup>Semi-empirical methods are in general a thousand times faster than DFT, but still "several orders of magnitude slower than molecular-mechanics treatments".<sup>[190]</sup>

assignment of atomic point charges to atoms in a molecule of biological interest. Moreover, it contains less than 200 molecules and fragments as of September 2016.<sup>[195]</sup>

#### 4.1.3.2 ESP-reproducing point charges in crystallography

One of the main goals of crystallography is, to provide structure models with accurate atomic coordinates. Once this is achieved, understanding inter- and intramolecular interactions within the crystal lattice often is the new focus of interest. In small-molecule structure models<sup>4</sup> such an interaction analysis is based on geometric criteria. Adding electrostatic information is logically the next step, since electrostatic forces have a longer range than *van-der-Waals* interactions. Therefore, electrostatic complementarity is discussed in several studies<sup>[196–199]</sup> and it would be very useful to facilitate rapid access to ESPs for models from single-crystal X-ray structure determination.

The ESP is frequently discussed in the context of recognition processes involving biological macromolecules.<sup>[197,200–202]</sup> There are several charge sets and programs that calculate and/or display the molecular ESP for these structures.

#### Point charges in macromolecular crystallography

The ESP generated by the APBS<sup>[203]</sup> plug-in to PYMOL<sup>[204]</sup> is discussed often for protein structures. The charges APBS reads have usually been assigned by the program PDB2PQR<sup>[192,193]</sup> and are communicated via a *pqr* file. In PDB2PQR users can specify different charges to be assigned: PARSE<sup>[205,206]</sup>, AMBER<sup>[182]</sup>, CHARMM<sup>[178]</sup>, PEOEPB<sup>[207]</sup>, SWANSON<sup>[208]</sup> and TYL06<sup>[209]</sup>. In spite of this flexibility only charges for nucleic acids and proteins can be assigned automatically. Otherwise a file (mol2 format) which includes the atomic point charges has to be supplied.

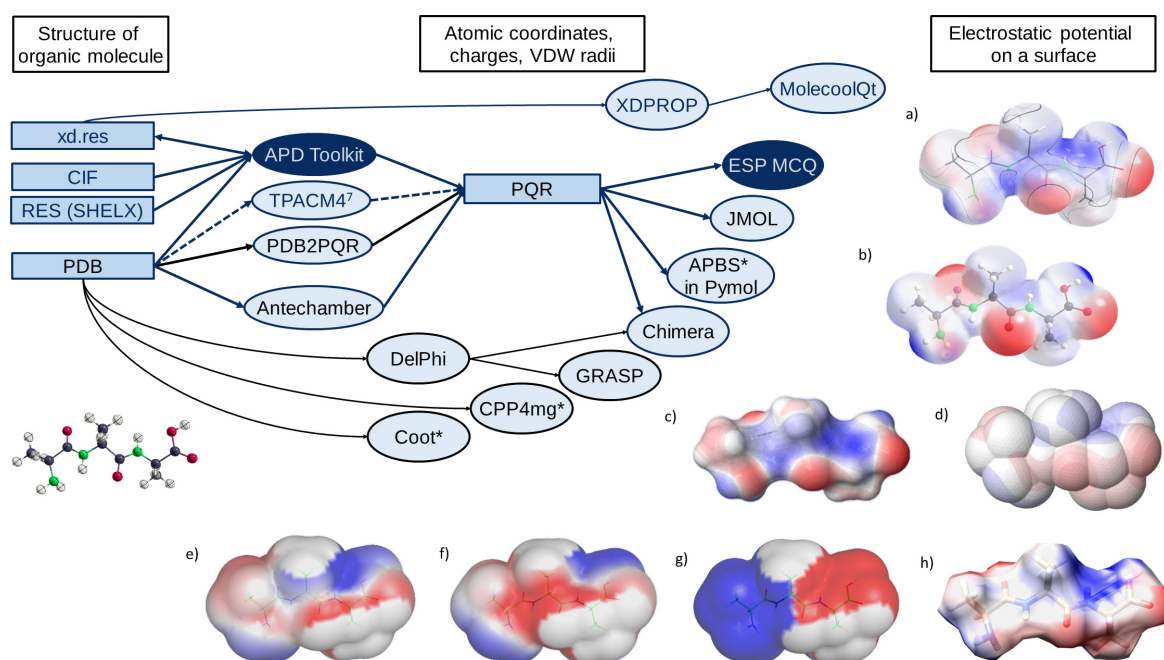
COOT<sup>[210]</sup> and CCP4MG<sup>[211]</sup> contain functions to assign atomic charges for standard proteins and nucleic acids as well. They can also map an ESP onto a molecular surface (a graphical overview is given in Figure 4.1). Regrettably, with respect to charges, neither PDB2PQR nor the CCP4 programs<sup>[212]</sup> or COOT cover charges for small-molecules such as ligands, co-factors or other pharmaceutically active molecules.

#### ESP modeling in charge density

Charge density is worth special mention<sup>[24]</sup>, since this branch of crystallography focuses on modeling the molecular charge-density distribution that best fits high-resolution<sup>5</sup> XRD data. Thus the subsequent ESP can be obtained from comparably demanding experiments<sup>[214]</sup> and least-squares refinements<sup>[50,51]</sup> of multipole populations. However, minor disorder can already lead to misleading results.<sup>[79]</sup> Therefore, accurate ESPs can only be obtained for non-disordered crystal structures. Moreover, attention should be given to possible additional polarization for molecules in a crystal compared to their gas-phase counterparts.<sup>[215,216]</sup>

<sup>4</sup>from data with a resolution of  $d \approx 0.83 \text{ \AA}$  or  $\sin \theta / \lambda \approx 0.6 \text{ \AA}^{-1}$

<sup>5</sup> $d \leq 0.5 \text{ \AA}$  or  $\sin \theta / \lambda \geq 1.0 \text{ \AA}$



**Figure 4.1:** Overview of ways to an ESP mapped on a molecular surface for trialanine. Blue boxes indicate file formats and ellipsoids programs. \* symbolizes a *Poisson-Boltzmann* potential<sup>[213]</sup> that models solvents and ions in the periphery instead of an isolated molecule. The different ESP were generated as follows: a) Invariom point charges via APD-TOOLKIT on a VDW·1.5 surface in ESP\_MCQ, b) AM1-BCC charges from ANTECHAMBER on a VDW·1.5 surface in ESP\_MCQ, c) Invariom point charges on a solvent accessible surface in JMOL, d) Invariom point charges on a VDW+2.4 Å surface in JMOL, e) Invariom point charges on a normal surface in APBS in PYMOL, f) TPACM4 charges on a normal surface in APBS in Pymol, g) Amberff charges via PDB2PQR on the default surface in APBS (PYMOL) different protonation, h) Invariom point charges on a solvent excluded surface in CHIMERA.

More sophisticated methods to study interactions in a crystal<sup>[217–219]</sup> have become available in computational crystallography. Both require QM wave functions and thus considerable expertise. In contrast, a molecular ESP can easily be generated from atomic point charges, which should be more useful for a broader group of scientists.

### Invariom point charge project

Therefore, this project's overall aim is :

1. introducing a database of 'invariom point charges' that reproduce the ESP obtained from gas-phase quantum mechanics,
2. reporting validation and application of these atomic charges by a comparative study for a test set of angiogenesis inhibitor molecules, and
3. providing a tool that allows assignment of these charges to molecules with atomic coordinates from sources like molecular mechanics or single crystal XRD.

## 4.2 Methods

### 4.2.1 Invariom point charges

#### 4.2.1.1 Atomic point charges from invariom-database model compounds

Invariom point charges are derived from the invariom database of currently around 2000 small organic model compounds. Each one of these molecules has been geometrically optimized with the program GAUSSIAN09<sup>[103]</sup> by applying the M06<sup>[104]</sup> DFT functional and the def2TZVP<sup>[104,106]</sup> basis set. The MK charge fitting implementation in GAUSSIAN09 yielded the atomic point charges for each model compound. The charges of those atoms sharing the same invariom name within the database were averaged. Compared to the alternative of transferring the charges from one model compound only as is done for invariom scattering factors, transferability of MK point charges improved considerably upon averaging.

#### Averaging charges

This transferability improvement upon averaging of the MK point charges has its cause in the strong influence of conformational variation on the MK charges. For generating scattering factors averaging is not necessary in the invariom database, since their independence of molecular conformation<sup>[56]</sup> and transferability have been shown on many occasions.<sup>[61,65,69,71,79,88,93–97]</sup>

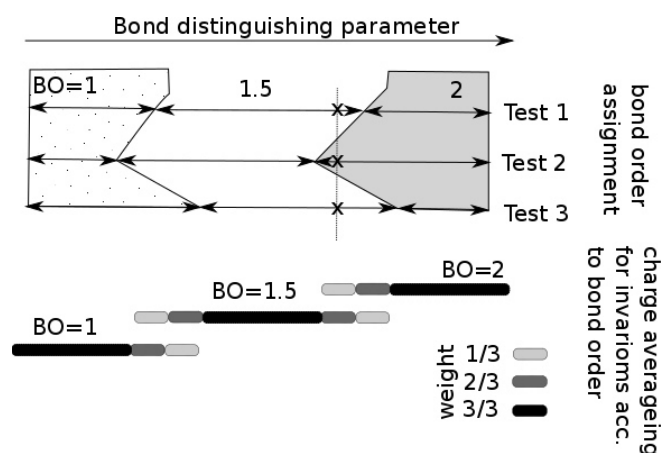
Database averaging is disadvantageous, since addition of new model compounds causes small changes to previous entries in the attribute databases. However, usually these changes are negligible and will lead to a continuous improvement of the database, since either redundancy is increased or new entries are provided.

Another scattering-factor database, the University of Buffalo Database, UBDB<sup>[76]</sup>, averages also the multipole populations it provides. Since it is otherwise similar to the invariom scattering factor database it could also be evaluated for the transfer of point charges. For experimental scattering-factor databases<sup>[58]</sup> point charges are not as easily derived, because the tools for fitting charges to the molecular ESP from experiment are missing. Additionally, the charges would be influenced by polarization due to crystal effects.

#### Alternative invariom names

In the process of developing and validating the invariom assignment for the MK charges, invariom names were improved by the generation of alternative names via variation of the threshold criteria. This change increased the number of chemical environments that are similar and therefore averaged. Hence, more atomic point charges have the same invariom name and contribute to one mean invariom point-charge. Furthermore, many formerly missing entries are added to the database.

Invariom classification for bond distances close to threshold values of the bond-distinguishing parameter  $\chi$ <sup>[91]</sup> is treated flexibly, but charges associated to atoms of such ambivalent cases get a lower weight during averaging (see Figure 4.2). The assignment of a different bond order has the potential to create an invariom not present in the database yet, thus the database becomes more complete (by 562 additional invariom names at the time of



**Figure 4.2:** Schematic illustration of alternative invariom names and their influence on averaging of point charges. The exemplary bond distinguishing parameter<sup>[91]</sup> shown is assigned a weight of two thirds for the invariom with a bond order (BO) of 1.5, and one third for the invariom with a bond order of two. © reprinted with permission from Wiley.

development). Hence the new alternative invariom names, which are also looked up if an invariom is missing for a molecule under study, result in a more robust charge assignment.

#### 4.2.1.2 Correction for electroneutrality and treatment of ionic compounds

Since the sum of point charges transferred from other molecules might deviate slightly from electroneutrality, they need correction after transfer to the molecule of interest. If molecules within a crystal structure model can be assumed to be neutral (e.g. when the protonation state suggests a non-ionic form or the crystal contains only one molecule), each molecule in the asymmetric unit is set to be neutral<sup>6</sup>. An average difference is added to each atom's charge<sup>7</sup> to ensure the correct molecular charge.

In order to model ionic compounds, molecular charges can be targeted<sup>[97]</sup> instead of electroneutrality. Knowing the 'real' molecular charge within a crystal of molecular ions is not straight forward. In the common case where only atomic coordinates are available, the practical solution for a chemist is to assume only charge transfer of whole electrons. This approximation is rather crude, but agrees with chemical intuition and has crystallographically been shown to yield good results for single crystal XRD of organic molecules.<sup>[97]</sup> Charge transfer is discussed and investigated further in Section 4.3.5.3. 'True' molecular charges require more thorough studies: either experimental charge density or theoretical computations that take crystal packing into account.

#### 4.2.1.3 Hydrogen-atom treatment

The treatment of hydrogen atoms is important for reproducing molecular ESPs, since they are often close to the molecular periphery where the ESP has the highest relevance. Small

<sup>6</sup>If not mentioned otherwise, in this chapter of the thesis neutral compounds are discussed.

<sup>7</sup>While the suggestion for the correction of scattering factors is based on the low X-ray scattering power of hydrogen atoms, so that only adjusting hydrogen-atom charges is recommended, the case is opposite for reproducing ESPs where hydrogen-atom charges are crucial.



changes in hydrogen-atom charges are, therefore, highly relevant when computing molecular ESPs. The influence is discussed in more detail in Section 4.3.4.1 by randomly disturbing hydrogen or non-hydrogen atom charges. Due to the importance of hydrogen-atom charges the existing empirical rules of transferability have been improved.

In XRD, where the empirical transferability rules originate from, the role of hydrogen atoms is less crucial than for reproducing molecular ESP, because hydrogen atoms have the lowest X-ray scattering contribution of all elements.<sup>[220]</sup> Although next-nearest neighbors are part of most hydrogen invariom names, it is possible to increase the degree of classification without increasing the number of required model compounds as discussed in Section 4.3.3.2. For scattering factors this has not been necessary, since the fit of diffraction data is good already, but for charges a higher degree of classification could be beneficial.

Another indirect hydrogen atom related topic that is relevant for comparing ESPs when atomic coordinates are determined by XRD is the choice of exact hydrogen-atom positions, in particular their bond distances<sup>8</sup> Hydrogen-atom bond distances need elongation to theoretical bond lengths when they are obtained from XRD, since they are around 10% too short in the IAM.<sup>[114]</sup>

An established method<sup>[95]</sup> is the application of distances from geometry-optimized model compounds of the invariom database assigned via the invariom name.<sup>[81]</sup> In contrast to bond distances from neutron diffraction<sup>[114]</sup> this is preferable, because neutron distances are unavailable for many chemical environments and are challenging to acquire in such cases while extension of the invariom database is easier and faster than a neutron diffraction experiment.

## 4.2.2 Electrostatic potential

### 4.2.2.1 Procedures to compare molecular ESPs

A qualitative way of assessing negative and positive regions of an ESP is by visualization. In such representations the ESP is mapped via a color code to a molecular surface. Global numerical indicators describing ESPs are provided by *Poltzer* analysis<sup>[223]</sup>, which can yield further insight if similar molecules are investigated. The figures of merit, RMS (root-mean square difference, Eq. 4.1) and RRMS (relative root-mean square difference, Eq. 4.2), have been developed alongside the RESP fitting algorithm<sup>[173,224]</sup> as a quality indicator for the reproduction of molecular ESP,

$$\text{RMS} = \sqrt{\sum_i^m (V_i^c - V_i^0)^2 / n} \quad (4.1)$$

$$\text{RRMS} = \frac{\sqrt{\sum_i^m (V_i^c - V_i^0)^2 / n}}{\sqrt{\sum_i^m (V_i^0)^2 / n}} = \frac{\sqrt{\sum_i^m (V_i^c - V_i^0)^2}}{\sqrt{\sum_i^m (V_i^0)^2}} \quad (4.2)$$

where  $m$  represents the number of grid points and  $n$  the number of atoms in a molecule,  $V_i^0$  is the reference ESP and  $V_i^c$  the ESP of interest at grid point  $i$ . These two indicators

<sup>8</sup>For angles between hydrogen atoms models with positions calculated on the basis of SHELXL,<sup>[221]</sup> yielded surprisingly good results for the crystallographic R-factor when considering experimental diffraction data. Theoretically this holds true, too, when comparing the change in single-point energies, including results for structures from neutron diffraction.<sup>[222]</sup>

were also applied during this project to allow quantitative analysis.

Since the RMS sums over all grid points and divides by the number of atoms, the value depends on the number of grid points considered. The RRMS, however, is independent from grid settings and molecular size. The RMS is not directly correlated with the number of atoms, but the number of grid points considered per atom is higher for small molecules than for large ones. Hence, the RMS can only be applied to compare different potentials of the same grid characteristics and molecular structure, and the RRMS has a broader range of application in comparison of different molecules. While RRMS is dimensionless, the RMS has the unit of the property described by the grid.

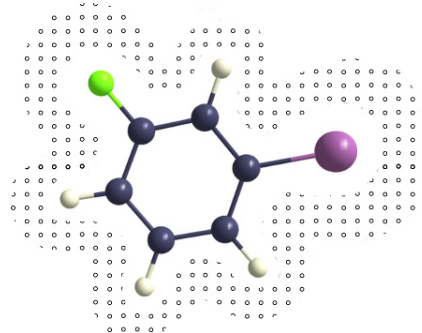
The program used for calculation of ESP grids writes out cubic grid files, but grid points in close proximity to the atoms are not of particular interest and would complicate summation due to their relatively high values. Therefore, only those grid points farther than a certain distance to the atoms in a molecule are considered for calculation of RMS and RRMS.<sup>[169,224]</sup> Grid points beyond a certain distance to the molecule do not add to the information that is already stored in those grid points closer to the molecule, thus also an upper distance limit is applied, leaving a molecular envelop of points evaluated as illustrated in Figure 4.3.

No routines for computing RMS, RRMS and selecting relevant grid points were available for the XD grid file format. Therefore the program MOLISO,<sup>[225]</sup> which could already read the corresponding files was changed in order to compute  $\sqrt{\sum_i^m (V_i^x)^2/n}$  for those grid points in a shell around the molecule. A schematic representation of how the program works is given in Appendix B (Figure 7.3). According to recommendations from literature<sup>[224]</sup> only grid points within 1.4 to 2.0 times the *van-der-Waals* radius of a molecule are considered in the summation. The ADDGRID utility of the XD package could easily set up files in which  $V_i^x = V_i^c - V_i^0$  which then could be read by the locally modified MOLISO in order to compute the RRMS.

If not stated otherwise, the reference ESP,  $V_i^0$ , was generated from MK charges fitted to the ESP from the wave function of the complete molecule (realMK). These realMK charges are assumed to yield the best molecular ESP possible by an atomic point-charge model. Thus they were referenced to, when different potentials  $V_i^c$  were compared. Atomic coordinates for the test-set molecules were obtained by DFT gas-phase optimizations (B3LYP/6-31G(d)<sup>[226,227]</sup> with GAUSSIAN09<sup>[103]</sup>) for the validations reported later.

#### 4.2.2.2 Computational details of ESP calculation

In this work ESPs were, if not mentioned otherwise, calculated via the method implemented in the program XDPROP<sup>[228]</sup>, which is part of the XD suite<sup>9</sup>.<sup>[12]</sup> For all these calculations



**Figure 4.3:** Molecular envelop of grid points for the evaluation of RMS and RRMS. © reprinted with permission from Wiley.

<sup>9</sup>The program is intended for application with the multipole formalism, thus point charges were converted to monopole populations fixing the  $\kappa$ -value to 1.0.

cubic grids with a point separation of 0.3 Å were chosen, because RMS values (see Section 4.2.2.1) depend on grid spacing as well as size and this setting allowed generation of a grid point density similar to the one used by Jakalian *et al.*<sup>[176]</sup> Visualization of ESPs was obtained by color mapping onto an ED isosurface by using the program MOLECOOLQT.<sup>[40]</sup> For the purpose of comparing ESPs for different sets of charges, the same isosurfaces of ED were obtained from the same IAM in which hydrogen atom positions had been adjusted to theoretical positions if necessary.

#### 4.2.2.3 A convenient procedure to get from atomic coordinates to an ESP

With the tools provided in the project an ESP can be mapped on a molecular surface of a model *e.g.* provided by conventional single crystal XRD within seconds. For a very fast route to display such an ESP coming directly from a crystallographic least-squares refinement the program SHELXLE<sup>[229]</sup> can be configured to call APD-TOOLKIT<sup>[230]</sup> and ESP\_MCQ as an external program. The first one will assign charges and write them to a *pqr* file. Then the program ESP\_MCQ, developed during this thesis (Section 4.2.3.2), or other programs as displayed in Figure 4.1 can read this *pqr* file, calculate the molecular ESP and map it to a *van-der-Waals*-like surface<sup>10</sup>. In ESP\_MCQ, users are allowed to specify their preferred colors, a factor for the *van-der-Waals* radii determining the distance of the surface to the atomic centers and optionally minimum and maximum of the ESP can be fixed for the color code. For solvent-accessible or hard sphere surfaces Jmol<sup>[231]</sup> provides an alternative quick way to a representation of the ESP.

### 4.2.3 Programs for application

#### 4.2.3.1 Automating the invariom-point-charge procedure: InvariomTool versus APD-Toolkit

In order to automatically assign charges to atoms in a molecule of interest, there are several prerequisites and steps required:

- For all molecules in the invariom database:
  - MK charges have to be derived for each atom (Chapter 2).
  - Each atom has to be classified by an invariom name.
  - Invariom names need to be matched with atomic charges for all atoms.
- Charges from all atoms of the database that have the same invariom name are to be averaged. Before discovering averaging as beneficial, invariom names had to be mapped to a specific model compound instead.
- For the molecule of interest:
  - Each atom has to be classified by an invariom name.
  - The transferred charges have to be adjusted in order to obtain a neutral molecule.

<sup>10</sup>The surface is only similar to a *van-der-Waals* surface, because boundaries between two or more in principle hard spheres are smoothed. Thus the surface appears more like an isosurface of ED.

Only the last block of functionality has to be performed by the user. All other steps are for generating the charge database only. Therefore, they could be accomplished by platform dependent programs and scripts. Every step from the first two points not related to invariom classification was performed by BASH<sup>[232]</sup> scripts, using AWK<sup>[233]</sup> for averaging.

Invariom classification had been implemented in three different tools already: INVARIOM-TOOL<sup>[91]</sup>, MOLECOOLQT<sup>[40]</sup> and APD-TOOLKIT<sup>[230]</sup>. While all atoms of the database already have an invariom name assigned by INVARIOMTOOL, extension of APD-TOOLKIT was easier and had the advantage of already incorporating some improved classification. However, it was crucial to have the same invariom classification in all the steps related to charges. Therefore, invariom names for the atoms in the database had to be redetermined with APD-TOOLKIT. Although differences between invarioms from APD-TOOLKIT and INVARIOMTOOL were only minor, procedures became much more robust upon sticking to one of the programs.

Differences between APD-TOOLKIT and INVARIOMTOOL included a different character for bonds of the order 1.5, a more consequent ordering of ligands according to bond order in APD-TOOLKIT and recognition of four-membered rings by APD-TOOLKIT. Additionally, APD-TOOLKIT can handle alternative invariom names automatically, which is a big advantage for the charge database, so that the advantage of INVARIOMTOOL concerning the differentiation of planar and non-planar nitrogen atoms was outperformed. Discrimination between 'R' and 'S' configuration missing in APD-TOOLKIT is not necessary for isotropic properties anyway. Therefore, APD-TOOLKIT was applied in all steps for invariom classification in this project.<sup>11</sup>

The first attempts of combining invariom names from the different programs provided valuable information about possible improvements of both programs. Since differences were found regularly, invariom classification in APD-TOOLKIT was adjusted by Jens Lübben, whenever necessary. Thereby, APD-TOOLKIT became simultaneously more reliable for its main purpose of assigning hydrogen ADP.

In the end, several functionalities were added to APD-TOOLKIT, of which only two plug-ins developed during this thesis were relevant for charge assignment by users:

**resp** This plug-in reads the atomic coordinates from files in XD format and writes a file called `xd.resp.inp` in the format of `xd.res` in which the monopole populations were adjusted to model the transferred charges. It was used for the validation presented in this thesis.

**pqr** This plug-in reads `pdb`, `cif` or `SHELX res/ins` files and writes a file in `pqr` format, requiring also the assignment of radii<sup>[234]</sup> according to the elements.

In both procedures molecules are assumed to be neutral and the charges are corrected automatically after charge transfer to yield neutral molecules. Molecular charges can in both cases be set by the keyword '*charged*' which triggers the program to ask for the charges of each molecule in the asymmetric unit.

Additional functions facilitated the conversion of MK charges for a given molecule to XD files, but were only intended for the validation study presented in the following sections.

---

<sup>11</sup>APD-TOOLKIT is written in PYTHON and can be downloaded from <https://github.com/jluebben/APD-toolkit>.

### 4.2.3.2 A program for calculating and visualizing molecular ESPs (ESP\_MCQ)

Once charges are assigned to a compound, the focus shifts to visualizing the result. The programs available for mapping an ESP to a molecular surface as shown in Figure 4.1 are mostly designed for macromolecules. Hence they are dedicated to display a big amount of atoms at once, so surface styles are quite different from what is used for visualization of small molecule potentials by charge-density researchers. In addition to the kind of surface displayed, the way the potential is calculated differs considerably.

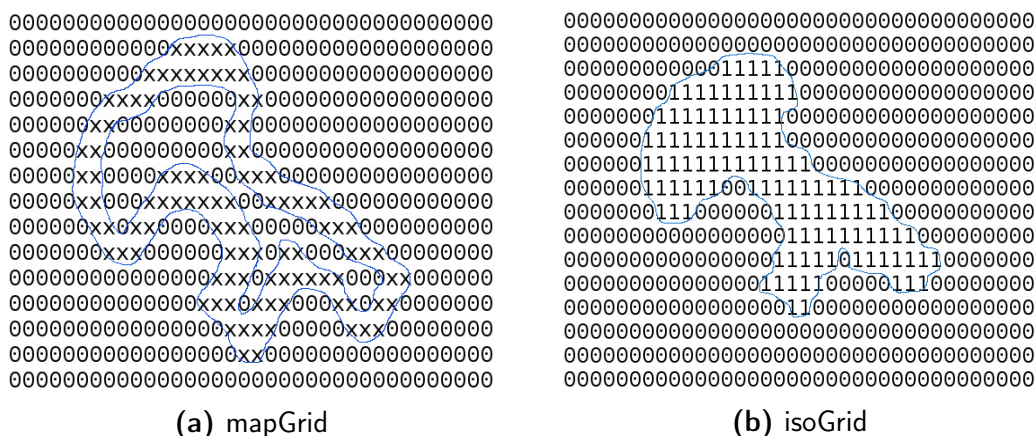


Figure 4.4: Schematic grid fillings within ESP\_MCQ.

With the aim of displaying molecular ESP as nicely as MOLECOOLQT<sup>[40]</sup> does when supplied with ESP and ED grid files calculated by XDPROP, ESP\_MCQ was developed. ESP visualization by MOLECOOLQT or MOLISO is only accessible to researchers capable of handling XD. Invariom point charges do not require this knowledge and should therefore grant direct access to a map of the ESP. In order to benefit from work already done by Christian Hübschle, his code of MOLECOOLQT revision 431 for graphical representation, preferences and options was used as the basis for the new program, ESP\_MCQ. Thus, all the new functions were written in the same language as the original program: C, C++ and Qt4.8.

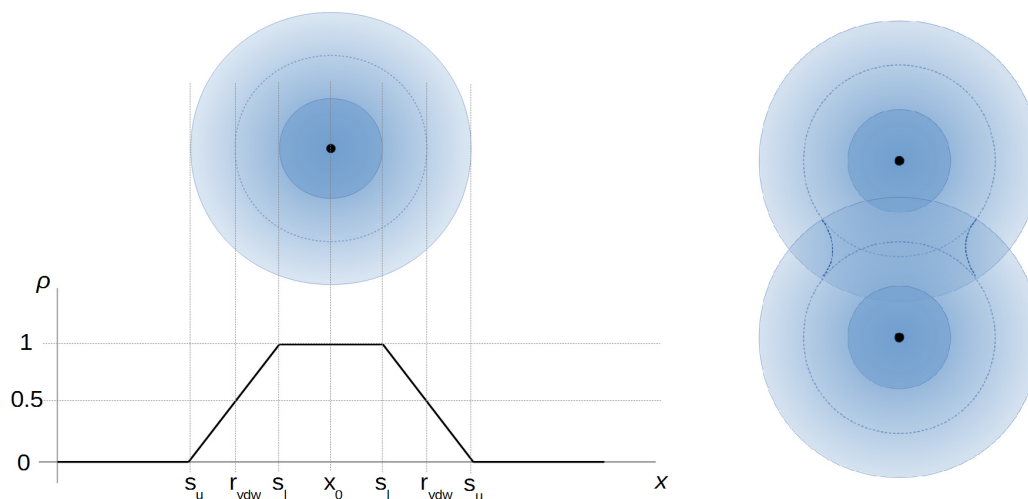
The program parts responsible for graphical display expect two cubic grids; one with the property to map (mapGrid) and the other one with the property which determines together with the isolevel the surface to be drawn (isoGrid). In the end only the value of mapGrid points close to the surface are relevant for the display. Therefore, the potential has only to be calculated for a small part of the grid (see Figure 4.4a). This saves time and is, in addition to the simple potential form, responsible for the increased speed compared to calculations with XDPROP. The potential for those points of the map grid that are relevant the potential is calculated by

$$V(i) = \frac{10^{10}}{4\pi\epsilon_0} \sum_j \frac{q_j}{d_j} \quad (4.3)$$

which yields the potential in units of e/Å<sup>3</sup> for a grid point  $i$ .

This leaves the isoGrid, usually an ED grid, to be set. The surface is drawn at a certain isolevel that can be specified. In order to quickly supply a grid resembling an ED grid, every

point within a certain distance to the atoms is set to one and all other grid points are set to zero as outlined in Figure 4.4b. This very simple procedure did not yield nice surfaces since gradients for calculating the surface normals had discontinuous directions. Hence the step from zero to one was substituted by a linear function of the distance to an atom as shown in Figure 4.5.



**Figure 4.5:** Quick method for filling the ED grid in order to map a property to a molecular surface and the smooth intersection between overlapping atoms.

This gradient has the additional advantage of smoothing the surface in areas where two hard spheres would create a hard line of intersecting spheres. By adding the EDDs of several atoms without dividing by the number of atoms contributing to the ED at one point the surface resembles an isosurface more than a model of hard spheres (see right side of Figure 4.5).

### Grid dimensioning

The most challenging task was to set-up the grid in such a way that it encloses the complete molecule but is not unnecessarily large. The developed process runs through the following steps:

1. Determination of a molecule's geometrical midpoint (or of another entity given in the file).
2. Determination of the maximum distance between two atoms of the molecule and addition of six Å.
3.
  - a) Reading the grid-point separation requested by the user, to determine the number of grid points in each direction.
  - b) Calculating the coordinates of the grid origin within the crystal lattice by subtracting half of the diagonal from the vector to the molecule's midpoint.
  - c) Calculation of atomic coordinates relative to the new grid.
4. Filling of isoGrid and mapGrid.

## 4.3 Experiments and Results

### 4.3.1 Improving classification of O2c, the example of three homotripeptides

#### 4.3.1.1 Aim and setup

Among the first experiments with invariom point charges were ESP comparisons for three homotripeptides. The ESPs for trialanine, triasparagine and tritryptophane were generated by three different sets of charges:

1. MK charges fitted to the ESP derived from the wave function of the whole molecule (realMK),
2. invariom point charges (assigned MK charges from model compounds),
3. amberff98 charges for a specific neutral amino acid.

At this stage automation of charge assignment was still to come and the whole experiment should only show if transfer of MK charges from invariom model compounds could in principle lead to satisfying results. Therefore, only qualitative analysis of graphical ESP representations was performed and charges were transferred from invariom model compounds without averaging over the whole database. Atomic coordinates came from a DFT geometry optimization with GAUSSIAN09<sup>[235]</sup> and the ED for surface generation was always derived from the IAM projected on the theoretical density.

If the ESPs from invariom point charges are close to those of the MK charges of the whole molecule, automation and more extensive studies would be goals worthwhile pursuing. The ESPs calculated from the amberff charges set a reference for a satisfying similarity, so that the ESPs from invariom point charges should not deviate stronger from the 'whole-molecule' ESP than the one from amberff point charges.

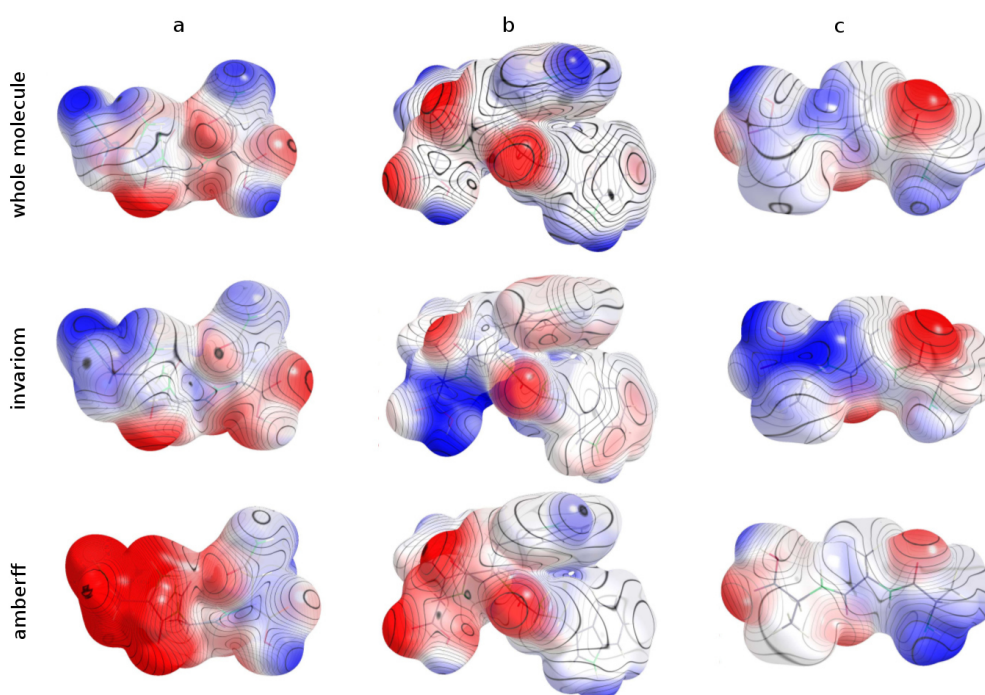
#### 4.3.1.2 Results

The ESPs for the three tripeptides are shown in Figure 4.6 for each of the charge sets tested. The results for the invariom transferred MK charges did not agree very well with those from charges fitted to the whole molecule, but compared to the difference between the whole molecule ESP and the ESP from the amberff charges, which are widely used for molecular dynamics simulations in biology, the difference to the whole molecule MK charges seemed within the limit of what can be used in simulations.

An interesting observation during manual transfer of invariom point charges was that formaldehyde served as model compound for all double bonded oxygen atoms. The origin for this was the very short invariom name of O2c for those oxygen atoms. Consideration of next-nearest neighbor atoms (NNN) significantly affects the charge of carbonyl oxygen atoms. The improvement due to a finer classification became apparent upon comparison of ESP from charges assigned on the basis of the short names as shown in Figure 4.7 and the longer invariom names (O2c[NNN]) as displayed in the middle of Figure 4.6.

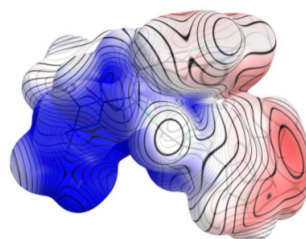
As discussed in Section 2.2.3 this stronger differentiation for O2c and O2n also improves assignment of scattering factors, although the improvement is less pronounced for figures





**Figure 4.6:** ESP derived from different point charges mapped to a *van-der-Waals* like isosurface at an ED of  $0.0067 \text{ e}/\text{\AA}^3$  with MOLECOOLQT<sup>[40]</sup>. The color code extremes were fixed to  $\pm 0.25 \text{ e}/\text{\AA}\epsilon$ . The molecules investigated were a) triasparagine, b) tritryptophane and c) trialanine.

**Figure 4.7:** ESP derived from invariom point charges of tri-L-tryptophane when all three double bonded oxygen atom charges had been transferred from formaldehyde as suggested by the invariom name O2c, instead of O2c[NNN].



of merit after refinement than for the difference in ESP from differently assigned point charges.

Result from *Poltzer* analysis<sup>[223]</sup> of each ESP (Table 7.6 in Appendix B) confirmed the insight from visual impression. Depending on the molecule, invariom point charges compared as well to the 'whole molecule' MK charges as the amberff98 charges.

Dipoles derived from the charges of the three homotriptides were additionally calculated with the program XDPROP. Those dipole moments derived from the charges fitted to the ESP of the whole molecule were better reproduced by the invariom point charges for each of the molecules than by the amberff charges (Table 7.7 in Appendix B).

Overall, this led to the conclusion that in principle point charges should be transferable from the invariom database. Hence, it was decided to automate and carry out a more extensive and quantitative study with more molecules.



## 4.3.2 Internal self-consistency test

### 4.3.2.1 Aim and setup

The first test with the nineteen inhibitor molecules was aimed at providing a reference for the upper limit of quality for invariom point charge transfer and at helping to identify problems with the automated procedure. Therefore, point charges fitted to the molecular ESP of one specific molecule were merged if they contained the same invariom name (as discussed in Section 4.3.2.2). Subsequent charge reassignment via invarioms, ESP calculation and RRMS evaluation were carried out for several basis sets and different ways of classifying hydrogen atoms. The assessment concerned A) individual point-charge comparisons for two molecules and B) ESP evaluation for the complete test set of nineteen angiogenesis inhibitors described in Section 4.1.2.

### 4.3.2.2 A: Results and discussion of point charge differences

Correct implementation of all steps involved in point charge transfer was tested by setting up a special charge database which contained only charges of one angiogenesis inhibitor at a time. Next, those charges were 'transferred' back to the exact inhibitor they came from. Compared to the original charges fitted for every atom against the ESP of the whole molecule (realMK), difference should only appear for atoms with invariom names that occur more than once within a molecule. The charges of atoms with the same invariom name are averaged before reassignment. Therefore the differences show how well charges of the same invariom name agree. The results of this internal self-consistency test are reported here for imatinib and cediranib.

Indeed, differences between these specially merged charges and those for the whole molecule appeared as expected only for those charges of invarioms that were found more than once in a molecule. Average differences between the realMK and the charges merged according to invariom name are given in Table 4.2 for imatinib and cediranib.

	imatinib			cediranib		
	C	H	H	C	H	H
invname		old	new		old	new
av. diff. (merged)	0.090	0.035	0.020	0.083	0.020	0.018

**Table 4.2:** Average difference of point charges from a fit to B3LYP/6-31G(d) ESP and point charges merged within the molecule according to invariom classification for the compounds imatinib and cediranib separately. The unit is  $e$  (elementary charge) for all charges and differences.

15 of 37 non-hydrogen atoms of imatinib had invariom names that occurred more than once; for cediranib 10 out of 33 did. This was the expected result, so the charge procedures had been implemented correctly.

The average difference between the averaged point charges for the eight carbon atoms of imatinib was 0.090  $e$  and 0.083  $e$  for cediranib. Hydrogen atoms with the same invariom name were more frequent, and more of them belonged to the same name. The average of absolute differences for hydrogen atoms in imatinib is 0.035  $e$ , which was significantly more than the 0.020  $e$  for cediranib. A closer look at imatinib revealed that most of the hydrogen

atoms were attached to an aromatic ring and thus classified as H@6c. Consequently, improvement of transferability seemed possible for those hydrogen atoms.

Inclusion of next nearest neighbor atoms in invariom classification for hydrogen atoms at aromatic rings should improve charge transferability. In the end an even more extensive classification by elongating the invariom name of hydrogen atoms by that of the atoms they are bonded to (see Section 4.2.1.3) was applied. This way the model compound for the directly bonded parent atom also supplies the property for the hydrogen atom, leading to a higher degree of classification by getting the most out of the model compounds already in the database. Application of this new nomenclature as classification decreased differences between the charges merged from 0.035 e to 0.020 e for imatinib. Although the decrease in difference of the merged charges is smaller for cediranib (0.020 e to 0.018 e) the new hydrogen atom treatment was beneficial in this case, too.

Hence the performance in reproducing molecular ESP is also expected to improve upon this more differentiated treatment of hydrogen atoms.

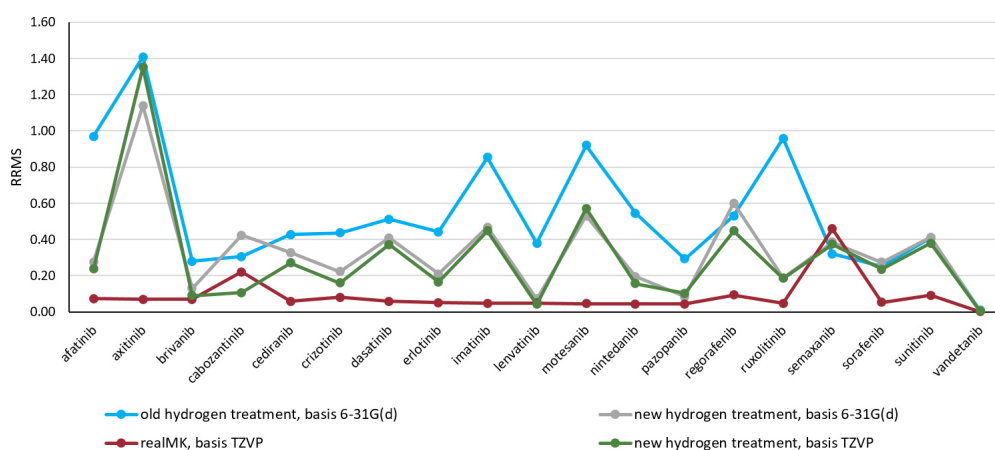
#### 4.3.2.3 B: Results and discussion of ESP evaluation

How averaging of charges with the same invariom name within each agniogenesis inhibitor molecule affected the ESP was investigated by comparing the RRMS of the resulting ESP with respect to the ESP calculated from the MK charges fitted to the ESP of the whole molecule (realMK). The basis set for computation of the ESP against which the charges were fitted was computed by three different basis sets: 6-31g(d), TZVP and def2TZVP. Thus the influence of different basis sets could be evaluated in addition to loss of information upon merging of charges according to invariom classification.

When the point charges were averaged for the original invariom names<sup>[81]</sup> the resulting ESP had an averaged RRMS of 0.55. The new hydrogen invariom names improved the RRMS to an average of 0.34. This demonstrated how sensitive the RRMS is towards choices in atom classification. Thus the RRMS is a suitable criterion for evaluating transferability.

Out of all the molecules studied axitinib turned out to be the most problematic molecule in this internal test (Figure 4.8). Its functional groups are a thioether, a carbon carbon double bond and annulated six- and five-membered rings. This specific variety led to several disagreements, but the largest differences were observed for carbon and hydrogen atoms of six-membered aromatic systems; one such ring contains a nitrogen atom, and the other one is linked via a sulfur atom to the rest of the molecule. The high RRMS for this molecule could be due to averaging, or it could be an extremely unlikely but possible statistical coincidence. Later on it was found that one of the charges was an outlier compared to the charges of all the atoms of the same invariom within the whole invariom database. Such outliers will have a small influence after averaging the charges of an invariom for the complete charge database, but since here one of the outliers is part of the reference the RRMS is unusually high. Hence, invariom point charges derived from the complete database can, like in this case, perform even better than expected from this internal test.

Also noteworthy was vandetanib. It had an especially low RRMS throughout, because the absolute potential values were in general higher than for the other molecules. Thus the RMS of the ESP was 20 times as much as for the average molecule, but the RMS of the potential differences was about the same as those of the other molecules. The cause



**Figure 4.8:** Detailed RRMS results for the internal test of all investigated angiogenesis inhibitors with respect to the 'whole molecule' ESP from 6-31g. Old and new hydrogen invariom names are compared and a different basis set was included to get an estimate of the error upon basis set change. ©reprinted with permission of Wiley.

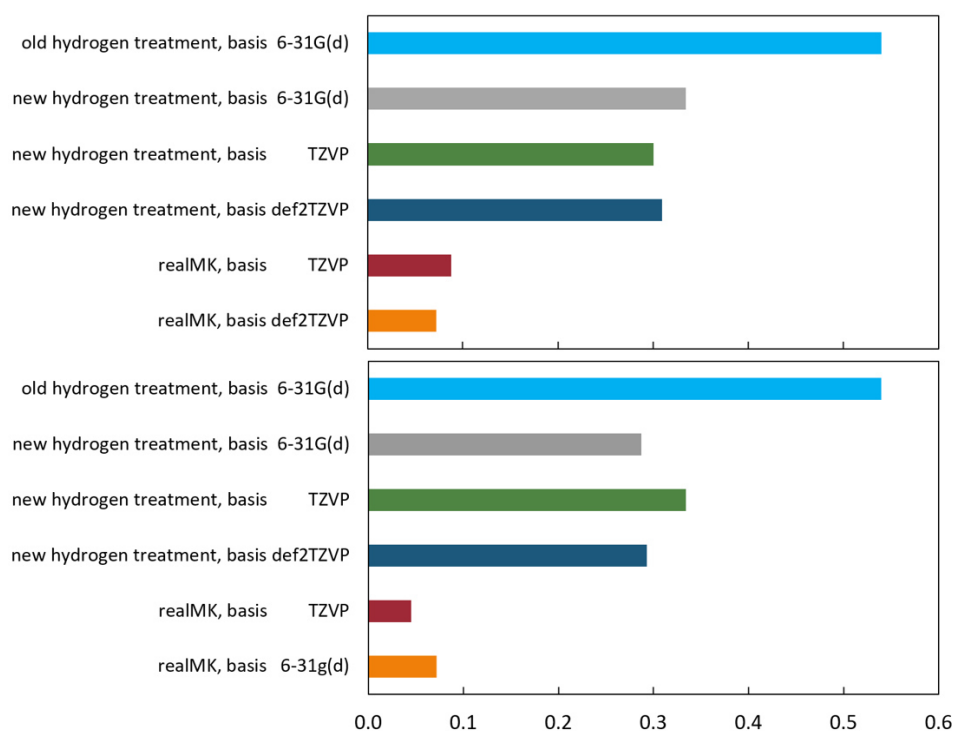
for the higher potential and its RMS was the bromine substituent, which was only present in vandetanib. Bromine has a high *van-der-Waals* radius, which led to a larger number of potential points in the sum for the RMS. Simultaneously bromine had a high charge and thereby additionally created high ESP values.

**Basis set influence** The same procedure was applied to investigate differences between basis sets 6-31g(d), TZVP and def2TZVP. So different sets of charges fitted to the ESP derived from a wave function of the whole molecule (no invariom averaging at all, further referred to as realMK) were studied for the different basis sets as well as transferred from model compounds that were treated with different basis sets.

The RRMS for simply changing the basis was approximately 0.09 (Figure 4.9). Figure 4.8 shows that mainly two molecules have especially different charges from fits to different basis sets: semaxanib and cabozantinib. The most deviating charges in cabozantinib are found for atoms in condensed rings and those in close non-bonding interactions (with other parts of the molecule). Additional diffuse basis functions could be the cause for the differences between the basis sets, since they could influence the ESP and thus yield different charges in interacting regions. Those different charges then reproduced the deviating ESP which led to higher RRMS numbers. This hypothesis was in agreement with the low RRMS for the test with TZVP, where the same functional groups were averaged for different non-bonding interactions. In semaxanib an intramolecular hydrogen bond was formed, but the charge differences were concentrated at the condensed five-membered ring attached to an oxygen atom. So this particular region might be especially sensitive to the choice of basis set.

In general, however, differences due to changing the basis set were minor. Therefore, employing def2TZVP instead of the frequently used 6-31G(d) basis set for deriving point charges from invariom model compounds was no problem. The basis set influence was smaller than the approximation of look-up point charges instead of molecule specific ones.

Figure 4.9 also shows that charges fitted to an ESP from a def2TZVP wave function are not as transferable as those from TZVP. This tendency was also observed for charges from



**Figure 4.9:** RRMS results from the internal invariom averaging tests, averaged for 19 angiogenesis inhibitors. Old and new hydrogen invariom names are compared. Moreover, a different basis set was tested to get an estimate of the error upon basis set change. realMK shows the RRMS upon basis set change without invariom averaging. On the upper chart the reference is the realMK with 6-31g(d) basis set while on the lower one realMK with def2TZVP is the reference. ©reprinted with permission of Wiley.

the invariom database. This could originate in the higher capability of the larger basis set, def2TZVP, to describe individual molecular features better and would yield a wider spread of charges to be averaged. Therefore, the following tests rely on TZVP invariom point charges.

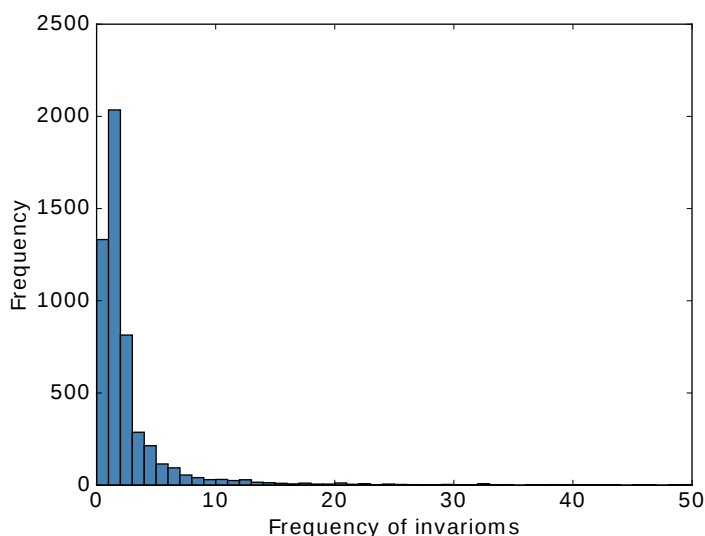
### 4.3.3 Charge averaging for the whole database – a statistical perspective

Before looking at the statistics of charge averaging, a general overview about the database statistics itself is given. For the scattering factors statistical evaluations were not as relevant, since the number of invarioms was all that mattered, but upon averaging, the information became relevant.

#### 4.3.3.1 General database statistics

In 2016 the database contained more than 5000 different invarioms when applying the new hydrogen notation from almost 2000 model compounds.

Several invarioms were part of many model compounds. Since the database contained more than 27000 atoms in total (in average 14 atoms per molecule), each invariom would



**Figure 4.10:** Frequency distribution of invariom frequencies within the invariom database (cut at 40 occurrences). ©reprinted with permission of Wiley.

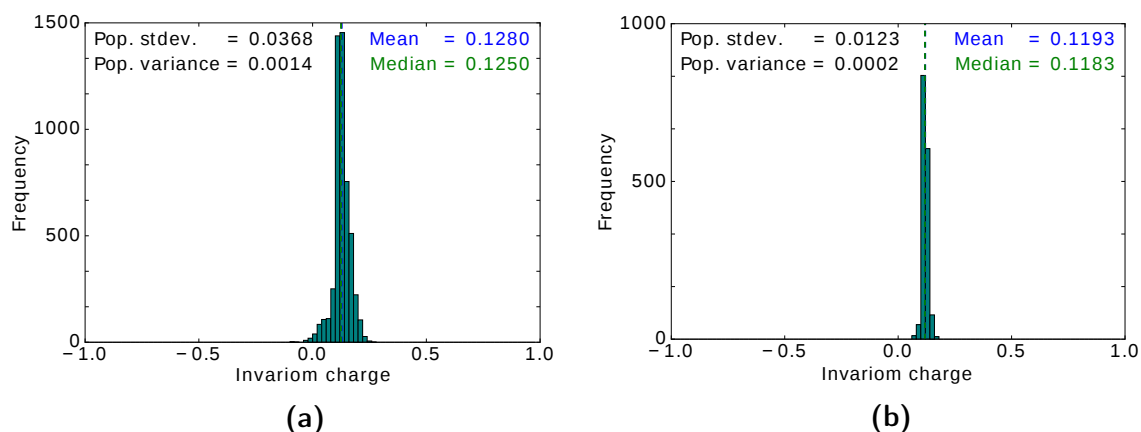
be expected to appear five times in the database for an even distribution. Figure 4.10 shows the real frequency distribution of invarioms in the database. The first bar ( $0 \leq x < 1$ ) represents those invarioms affected by the alternative invariom name procedure. The second bar is the highest, which means that most invarioms occur only once. This is due to the way the database is designed: each time an invariom name is found to be missing, the smallest possible model compound for a given invariom is added.<sup>[81]</sup> But there are also invarioms which come from many model compounds. The most frequent invarioms, their charges and counts are listed in Table 7.8 in the Appendix, showing that several invarioms occur more than 200 time in the database.

Non-integer counts were caused by the weighting scheme, which downweighted invarioms with bond orders assigned on the basis of bond lengths close to a threshold. Hence the importance of alternative invariom names is indicated by the first bar in Figure 4.10. At least 562 invarioms have a count of one third and thus would not be in the charge database without alternative invariom naming.

#### 4.3.3.2 Investigation of new hydrogen atom treatment

An evaluation of the point charges of invariom H@6c in the whole invariom database yielded a spread of 0.037 e (Figure 4.11a). Extended invariom notation led to 343 new invariom names replacing H@6c. The charges for each of the new names showed a smaller spread. The charge histogram for the most frequent hydrogen invariom among them, a hydrogen atom at a phenyl ring with no special next-nearest neighbor substituents, is shown in Figure 4.11b.

It should be noted that no model compounds had to be added for this new hydrogen invariom classification, since the model compounds required had already been computed for the parent atoms of the hydrogen atoms. Therefore, this scheme is suited to become a new option for the transfer of other invariom properties, which may be relevant when



**Figure 4.11:** Statistics of the most frequent hydrogen invariom. Frequency distribution of all charges in the invariom database (a) before and (b) after extending invariom notation. ©reprinted with permission of Wiley.

highest accuracy is required.<sup>12</sup> The more differentiated treatment of hydrogen atoms also improved the performance in reproducing molecular ESP as is shown in Section 4.3.2.3.

#### 4.3.3.3 Frequency distribution of selected invariom charges

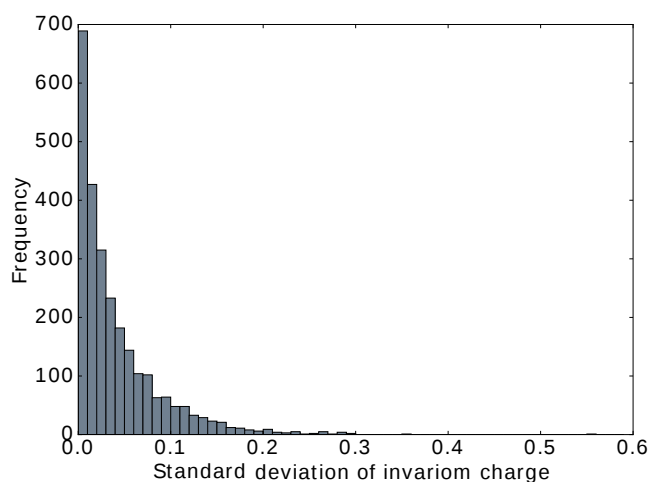
Frequency distributions of charges belonging to selected specific invarioms are shown in Figure 7.6 and Figure 7.7 in Appendix B, in which the counts appear tripled because of the weighting for alternative invarioms.

Non-hydrogen charges have a broader distribution, caused by their higher distance to the molecular surface at which the ESP was fitted by the charges. Hence the spread of charges for atoms farther away from the molecular surface was expected to be higher than for atoms closer to the surface. Overall the peaks have a narrow distribution, except for N@6c1h1h. This invariom for a nitrogen atom next to an aromatic system has no defined state of hybridization in APD-Toolkit so far; the atom could be  $sp^2$ - or  $sp^3$ -hybridized. In the  $sp^3$  hybridized form a lone pair is close to the surface on one side while in the  $sp^2$  hybridized form two regions are close to a negative charge, which could lead to the broader spread. The recognition of planarity is a feature still to be implemented, but will of course be applied to the charges once APD-TOOLKIT has been changed. However, all charges for the two nitrogen atoms shown are more negative than  $-0.5 e$ , so that the direction of electrostatic interaction stays similar for the width of the distribution.

#### 4.3.3.4 Standard deviations of invariom charges

The frequency distribution of standard deviations (STDEV) of all invarioms that occur more than once in the invariom database were summarized in Figure 4.12. The STDEV for merging of half of the charges was less than  $0.026 e$ .

<sup>12</sup>For the reasons described in Section 4.2.1.3 using more than next-nearest neighbors for hydrogen is not necessary to obtain the best least-squares fit to X-ray diffraction data. For computing atomic displacement parameters for hydrogen atoms, though, it is useful as well.



**Figure 4.12:** Frequency distribution of STDEV for those invarioms of the database that occur more than once. The median is 0.026 e. ©reprinted with permission of Wiley.

#### 4.3.4 ESP uncertainties caused by point-charge deviations

##### 4.3.4.1 Sensitivity of ESP towards general changes in point charges

For this test each database charge was disturbed by a random amount sampled from a *Gaussian* distribution whose mean was zero. Varied spreads of the distribution in the range of the STDEV were evaluated: 0.10, 0.050, 0.025 and 0.010 e. The thus disturbed charges were assigned to one molecule of the test set, sunitinib, and the resulting ESP calculated with XDFPROP. When those ESP grids were subtracted from the ESP grid calculated from the undisturbed charges, the RMS of the difference grids was investigated. 180 random disturbance experiments were run for each spread and the results are displayed in Figure 4.13.

The figures show that the RMS is highly sensitive to charge deviations. In order to obtain the RRMS from the RMS, the RMS has to be divided by 0.7539 e/Å, the RMS of the original unperturbed ESP. RRMS medians are summarized in Table 7.9 in the Appendix.

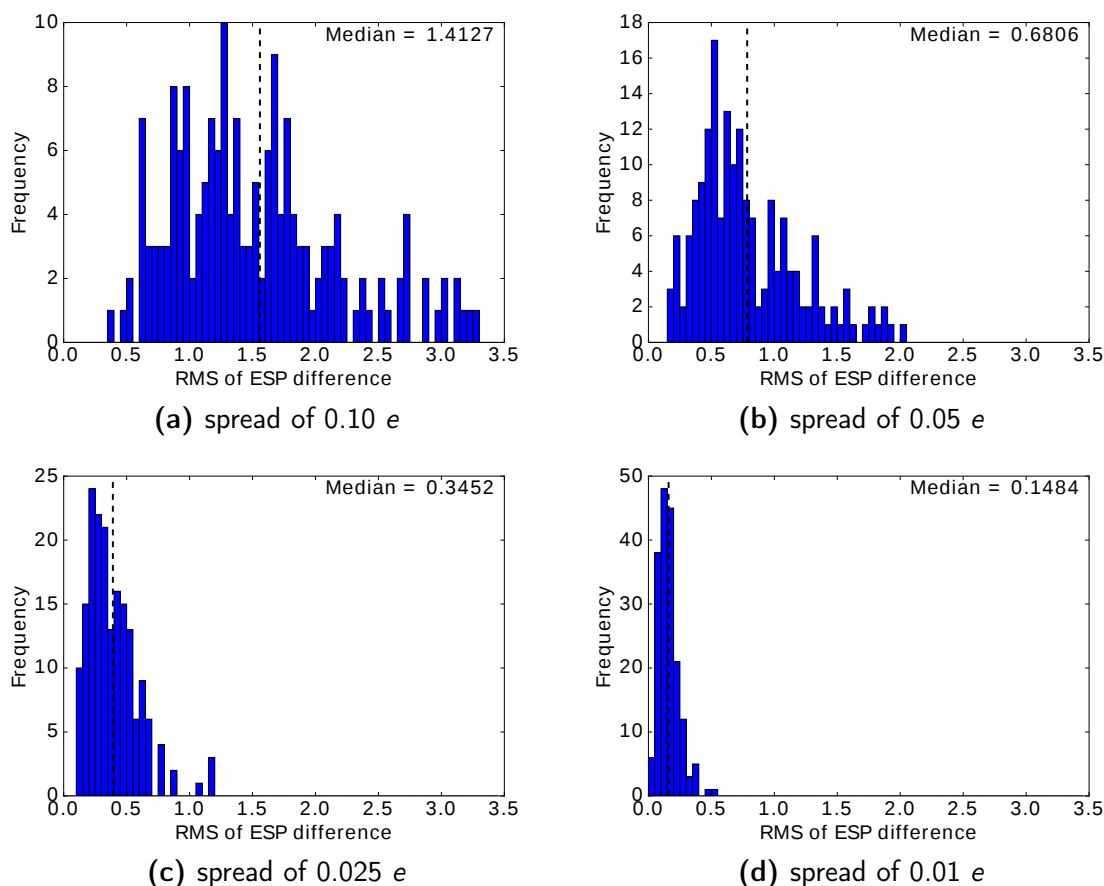
The spread of 0.025 e is closest to the median of STDEVs (see Figure 4.12) so for this an additional similar experiment was evaluated, that disturbed a) hydrogen charges only and b) non-hydrogen charges only.<sup>13</sup> Results are shown in Figure 4.14. As expected changes of hydrogen atoms charges had more impact on the ESP at a certain distance of the molecule. Hence the broader spread of charges for averaging of non-hydrogen invarioms is not necessarily limiting the ability of invariom charges to reproduce molecular ESP.

These results led to the design of the next more extensive and elaborate study.

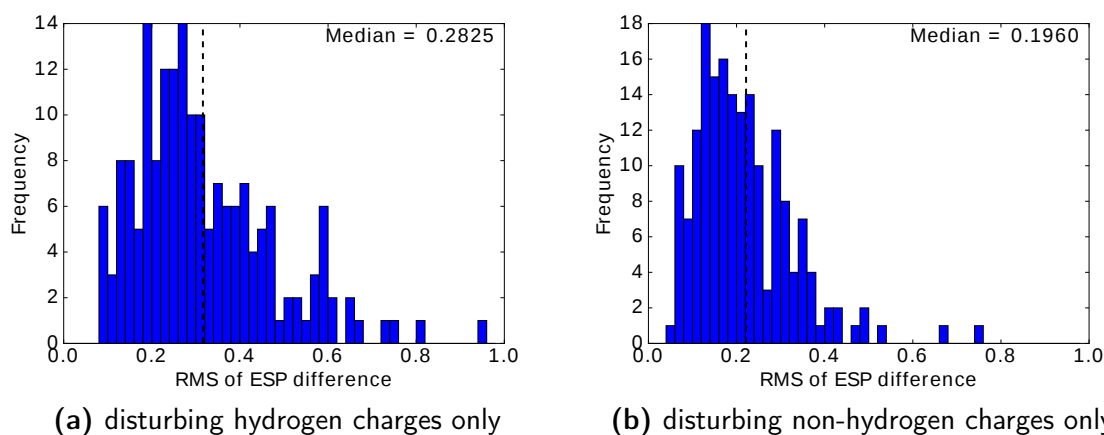
##### 4.3.4.2 Influence of charge standard deviations on the ESP

Combining the test described above with the standard deviation (STDEV) of the charges in the database yielded an approximation of error propagation to the ESP and the RMS as its

<sup>13</sup>At the time 1310 invarioms of all 5328 ones corresponded to a hydrogen atom.

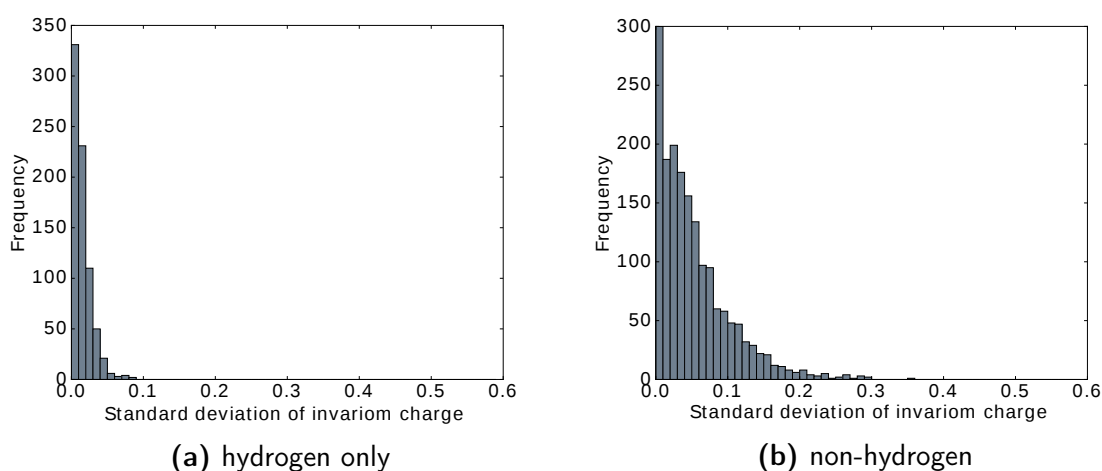


**Figure 4.13:** Frequency distributions of RMS of the ESP in  $e/\text{\AA}$  with randomly disturbed charges of different spreads of the random distribution minus the ESP from undisturbed charges. ©reprinted with permission of Wiley.



**Figure 4.14:** RMS frequency distributions for the difference between the ESP from randomly disturbed and undisturbed charges. The spread of disturbance was 0.025 e each time. ©reprinted with permission of Wiley.



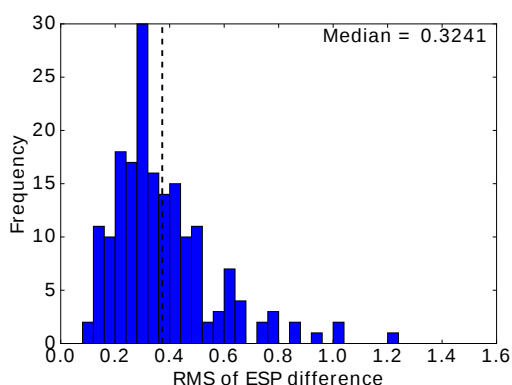


**Figure 4.15:** Frequency distribution of STDEV for those invarioms of the database that occur more than once. The STDEV median of the hydrogen charges is 0.012  $e$  and for non-hydrogen charges 0.040  $e$ . ©reprinted with permission from Wiley.

evaluation tool. As shown before the median STDEV of the database is 0.026  $e$ . Therefore, the test above for sunitinib with a spread of 0.025  $e$  yielded only an approximation for the performance of look-up charges in reproducing molecular ESP.

In this test the STDEV of hydrogen charges was used as spread for the random changes to hydrogen entries in the charge database and similarly the STDEV of non-hydrogen charges to non-hydrogen entries. Histograms of invariom charge STDEVs differentiated by hydrogen or non-hydrogen are displayed in Figure 4.15.

The medians for the STDEV distributions were set as spreads for the random disturbance of each charge in the database for all angiogenesis inhibitors. 250 such random disturbance experiments were run. This number of runs was sufficient since already after 180 runs only the third decimal place changed. Since for sunitinib (after 181 tests, for comparability) the median is only 0.32 (Figure 4.16), the application of an equal average spread for all charges



**Figure 4.16:** Frequency distributions of RMS of the ESP in  $e/\text{\AA}$  with randomly disturbed charges of the random distribution minus the ESP from undisturbed charges. The spread of the disturbance was the median of the STDEV from averaging equal invarioms in the database. ©reprinted with permission from Wiley.

independent of element overestimates the error of RMS and RRMS. The results for all 250 runs for all angiogenesis inhibitors are summarized in Table 4.3.

**Table 4.3:** Median values for RMS and RRMS of perturbed ( $_{per}$ ) and unperturbed ( $_{unper}$ ) ESP

molecule	$RMS_{unper}$ [e/Å]	$RMS_{per}$ [e/Å] (median)	$RMS_{per-unper}$ [e/Å] (median)	RRMS
afatinib	1.1430	1.1645	0.33	0.29
axitinib	0.3526	0.4416	0.27	0.77
brivatnib	1.2409	1.2732	0.33	0.26
cabozantinib	0.8756	0.9049	0.34	0.39
cediranib	1.3552	1.3765	0.36	0.27
crizotinib	0.5657	0.6052	0.28	0.50
dasatinib	0.8313	0.8943	0.34	0.41
erlotinib	0.7349	0.7810	0.33	0.45
imatinib	1.6463	1.6289	0.41	0.25
lenvatinib	0.7949	0.8403	0.32	0.40
motesanib	0.6058	0.6369	0.31	0.52
nintedanib	1.2849	1.2879	0.44	0.34
pazopanib	0.6728	0.7497	0.30	0.44
regorafenib	0.8089	0.9023	0.38	0.46
ruxolitinib	0.8086	0.8320	0.27	0.33
semaxinib	0.3918	0.4586	0.23	0.60
sorafenib	0.7877	0.8718	0.37	0.48
sunitinib	0.7539	0.8135	0.32	0.43
vandetanib	19.6261	19.6367	0.32	0.02
average	1.857	1.900	0.33	0.40

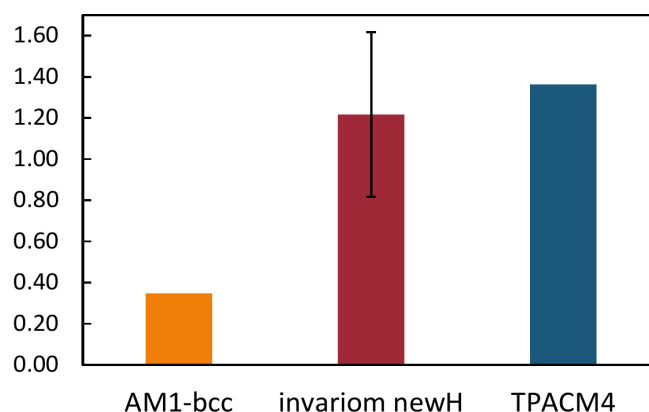
On average statistical charge errors led to an RMS error of 0.33 e/Å and for the RRMS to 0.40. The numbers determined in this experiment were applied as an approximation of ESP uncertainty in the comparative study presented next.

## 4.3.5 Comparison of molecular ESP

### 4.3.5.1 Point charges versus wave function

In general there is a limitation to the reproducibility of a molecular ESP by a point-charge model. The ESP charge fitting procedure<sup>[181]</sup> implemented in GAUSSIAN minimizes the difference between the ESP from the point charge model and the molecular wave functions. This difference commonly had an RRMS between 0.05 and 0.15 for the molecules of the test set. The values depended on the basis set; for 6-31G(d) basis the RRMS is lower by 0.02 compared to TZVP and def2TZVP, whose RRMS were very similar. Charges fitted to the ESP from def2TZVP systematically showed stronger deviations from zero than for those from TZVP.

Reference ESP (from B3LYP/6-31g(d) and M06/def2TZVP alike) were best reproduced by database charges from fits to an ESP of the TZVP basis. Therefore, only those charges will be discussed for the following comparisons. However, the def2TZVP basis reproduces automatization energies<sup>[106]</sup> better, which is the reason why both, M06/TZVP and



**Figure 4.17:** Comparison of RRMS results averaged for 19 angiogenesis inhibitors for ESP calculated from molecule specific point charges (AM1-BCC) and look-up tables (invariom and TPACM4). ©reprinted with permission from Wiley.

M06/def2TZVP, point charge databases will be provided. Independent of the difference between them, both invariom data sets perform better than TPACM4 charges, which is discussed next.

#### 4.3.5.2 Performance of different point charges

The two conceptually similar transferable point charges: invariom and TPACM4 were compared with respect to their performance in reproducing the molecular ESP of the 19 angiogenesis inhibitor molecules in the test set. In all cases the ESP from charges fitted directly to the B3LYP/6-31G(d) ESP (realMK) was employed as reference. Figure 4.17 shows the resulting average RRMS (see Eq. 4.2) while molecule-specific information is given in Appendix B (Figure 7.5).

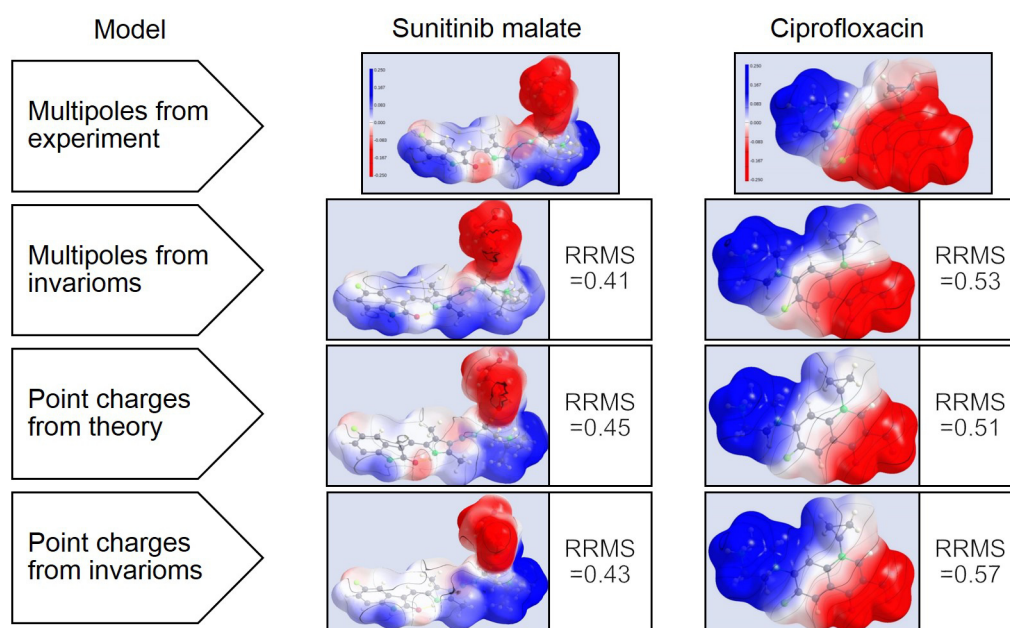
In order to give a perspective on improvement upon additional computational effort, performance of AM1-BCC charges is compared, too. As expected, those molecule specific charges reproduce the ESP created by realMK charges better than transferable atomic point charges. The deviation of the ESP from realMK charges is lower for invariom point charges than for TPACM4 charges, which are similar in time and effort invested.<sup>14</sup>

#### 4.3.5.3 Comparison to results from experimental charge-density studies

Sunitinib, one molecule of the test-set, was part of an experimental charge-density investigation in 2014<sup>[236]</sup> in which the compound was crystallized as a cocrystal with malate. The high resolution XRD data for sunitinib malate were kindly provided by the authors. This allowed a comparison of ESP from multipoles refined against experimental data to ESP from aspherical invariom scattering factors, from invariom point charges and point charges obtained from a fit to the theoretical ESP.<sup>15</sup> The equivalent comparison was also carried

<sup>14</sup>Formerly used hydrogen invariom names perform similar to TPACM4 point charges. When the local chemical environment was extended for hydrogen invarioms, better RRMS values were observed. Thus for further comparisons the extended invariom notation was applied.

<sup>15</sup>Specific experimental details are given in Section 7.2 of the Appendix.



**Figure 4.18:** ESPs of sunitinib malate and ciprofloxacin from different theoretical approaches compared to results from refinement of multipole parameters against experimental, high resolution X-ray data. The potentials are mapped by a color scale from  $-0.25$  to  $+0.25$   $e/(\text{\AA}\epsilon)$  onto an ED isosurface ( $0.0067$   $e/\text{\AA}^3$ ). Atomic positions were obtained by charge density refinement in both cases. ©reprinted with permission from Wiley.

out for structure models of the neutral compound ciprofloxacin, for which experimental charge density data were also available.<sup>[79]</sup> This way one salt and one neutral structure were investigated.

Several molecules were in the asymmetric unit of each crystal structure and could be identified by the charge-assignment program. For ciprofloxacin each of them was adjusted to be neutral. For the salt molecular charges were specified. The charge transfer from cationic sunitinib to the malate anion was estimated as one electron; and this charge is distributed between all atoms in both ions. This rather crude approximation yielded surprisingly good results in crystallographic refinements.<sup>[97]</sup> More importantly, the experimental charge density study<sup>[236]</sup> supports full charge transfer in sunitinib as discussed further in Section 4.3.5.4.

Maps of the different resulting ESP on the molecular electron-density isosurface are shown in Figure 4.18. The comparison shows similar features independent of the model. The vertically depicted malate of the sunitinib structure has a predominantly negative ESP, and the region of sunitinib which it is closest to is the one with the largest positive charge. For ciprofloxacin the zwitterionic nature of the molecule is preserved in all models. Moreover, all ciprofloxacin models compared exhibit a similar extent of the neutral region.

A comparison of the ESP from those point charges that best reproduce the ESP obtained from QM to the ESP from the experimental multipole model yielded RRMS values of 0.45 and 0.51. The values for the comparisons of the other models to the one from experimental multipoles are given in Figure 4.18 and they show that both invariom models (point charges and multipoles) deviate similarly from the experimental reference to the ESP from theoretical charges. In conclusion invariom point charges provide an acceptable

estimate of the molecular ESP. Additionally, it was observed that an aspherical description of the electron density distribution is not required to reproduce molecular ESP.

#### 4.3.5.4 Electroneutrality correction and ionic molecules

Different electroneutrality corrections had only minor impact on the ESP, since atomic charges were usually each corrected by less than 0.25 % of the valence electrons of a neutral molecule. Charged molecules can, in principle, have slightly higher corrections, but in case of sunitinib malate the corrections stayed below 0.1% for molecular charges between 0.0 and 1.4 e.<sup>16</sup> Ruxolitinib was the molecule that out of the 19 neutral angiogenesis inhibitors required the highest electroneutrality correction with 0.31 % of the sum of valence electrons for TZVP and 0.36% for def2TZVP invariom point charges.<sup>17</sup>

The topic of ionic molecules is of special interest in crystals, since the true charge transferred between anion and cation does not have to be a whole electron. Therefore different amounts of charge transfer were studied further at the example of sunitinib malate.

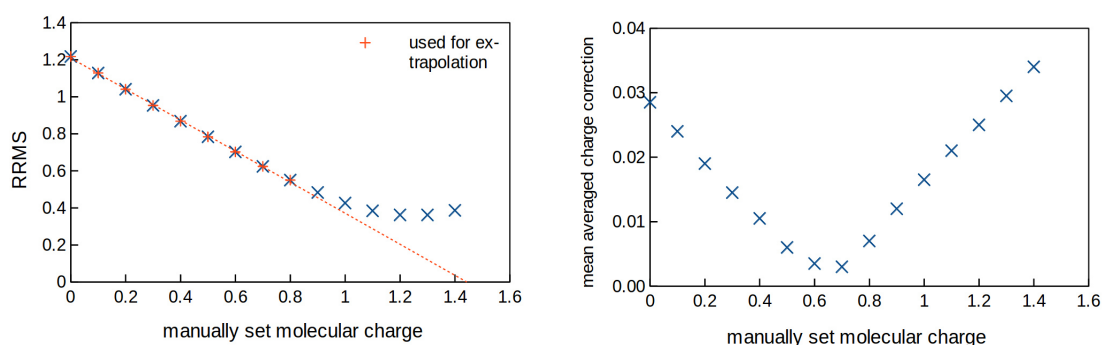
Figure 4.19a demonstrates the influence of the molecular charge assumed for sunitinib malate on the ESP. A chemist's knowledge of common functional groups and their charges (e.g. an anionic carboxylic acid group or a nitrogen with four bonds) allowed approximation by integer charges. For sunitinib malate the thus assumed transfer of one electron was in accordance with the charges refined for multipoles<sup>18</sup>, where the refined multipole populations suggest a charge transfer of 1.05 e. When molecular charges between 0.0 and  $\pm 1.4$  were set for invariom point-charge assignment and were compared to the molecular ESP, a charge transfer of 1.2 and 1.3 electrons performed best. Assuming neutral molecules in the case of this ionic salt lead to a strong deviation from the ESP derived from the charge density model. Since the difference of the ESP is small for 1.0 to 1.3 transferred electrons, the approximation of one transferred electron, as suggested by chemical intuition, seemed to be a valid method.

When chemical intuition can not predict reasonable molecular charges from the structural formula, the application limit of look-up point charge is reached. For a charge transfer of 0.7 electrons from cation to anion, the atomic invariom charges have to be corrected the least as shown in Figure 4.19b. A transfer without any correction (from summing the uncorrected database point charges) suggested an anion charge of  $-0.667$  e and a cationic one of  $+0.496$  e. Therefore, simple summation could be a possible alternative for future applications, since it provided at least a hint into a direction of the molecular charge transfer. However, MK charges are not the most reliable tool for testing this due to their conformation-dependence. Monopole population parameters of the multipole model would probably perform better for this task. However, knowledge of hydrogen-atom positions is required for invariom classification. Hence enough information for a chemist to deduce the molecular charge should be present in all cases for which invariom point charges can be applied.

<sup>16</sup>A part of the ionic character is often already accounted for, since some invarioms of a charged molecule are usually transferred from ionic model compounds.

<sup>17</sup>On average the correction of atomic charges is below 6 % of their absolute charge for neutral compounds.

<sup>18</sup>The only restraint used was an overall neutrality restraint.



(a) RRMS with respect to the ESP of refined multipoles for differently set molecular charges. (b) Comparison of mean average charge correction per atom.

**Figure 4.19:** Investigations of the charge transfer between sunitinib and malate. ©reprinted with permission from Wiley.

Full charge transfer could become less accurate if the ions are soft (with very diffuse atoms) like  $I_3^-$  and when electronegativity differences between cation and anion are small.

### 4.3.6 Application examples

Figure 4.20 shows four organic molecules, for which APD-TOOLKIT was used to assign invariom point charges. Afterwards ESP\_MCQ was run in order to compute and display the ESP mapped to a smoothed surface at a distance of twice the *van-der-Waals* radii. The input files came from a standard IAM refinement with SHELXL, and either the res or cif file were read by APD-TOOLKIT, in which the plug-in, pqr, elongated bonds to hydrogen atoms to theoretical bond lengths from the invariom database and assigned the invariom point charges. These charges were written to a pqr file, together with radii<sup>[234]</sup> also assigned by the pqr plug-in to a file called 'resp.pqr'. Upon start of ESP\_MCQ this file had to be loaded, and after color and surface settings had been adjusted, the given maps of the ESP on the molecular surfaces were obtained.

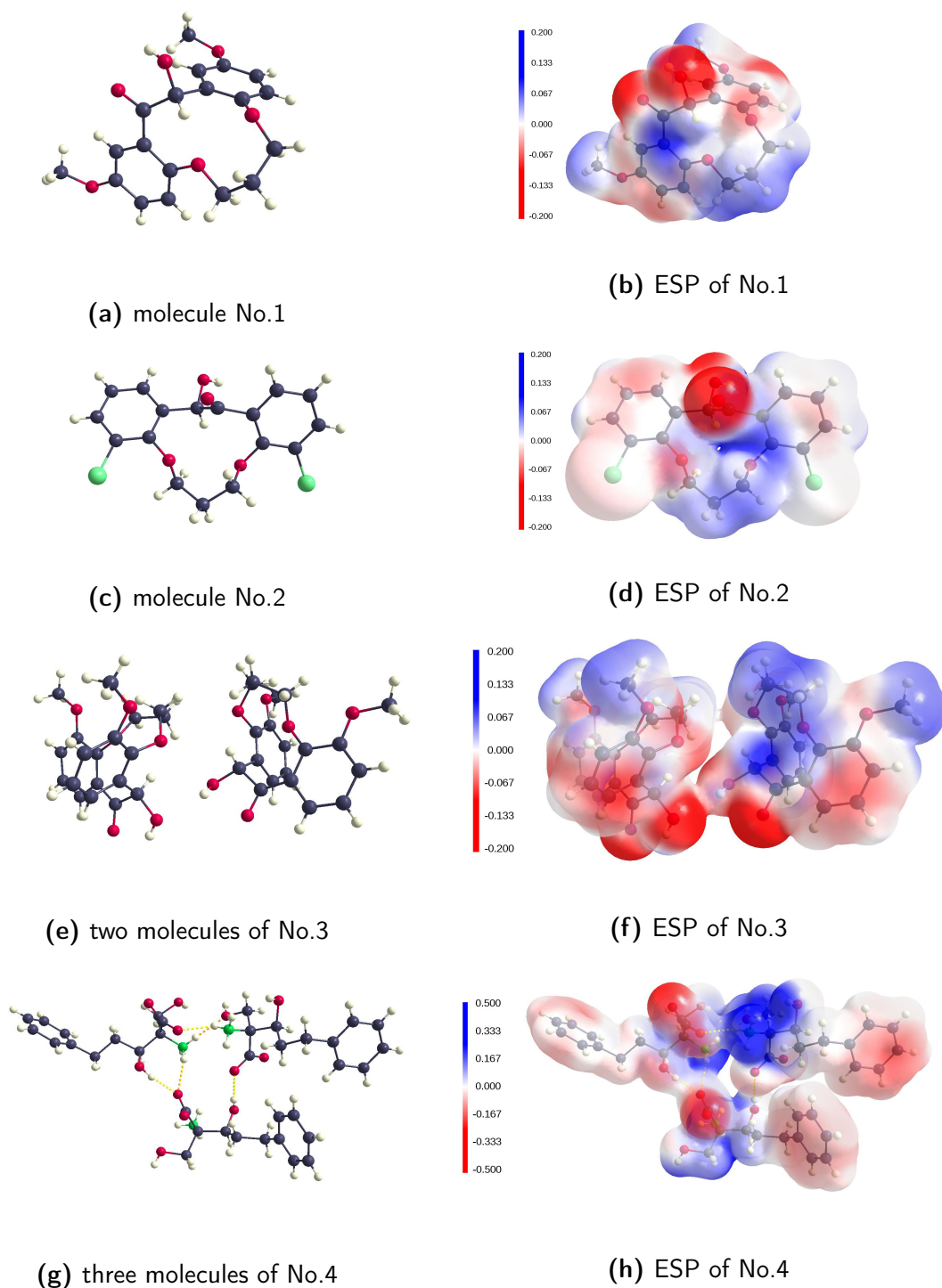
The compounds were synthesized by Giovanna Petrillo in the group of Isabel Usón and the main aim of crystal structure determination in this project was to determine their absolute structure. They serve as nice practical examples containing different numbers of molecules per asymmetric unit. For compound No.1 and No.2 only one molecule per asymmetric unit was found. Interestingly, the charge maxima except for the carbonyl and hydroxy groups were more pronounced in the first molecule than in the second. The chlorine substituent instead of a methoxy group is the only difference to compound No.3 and the first and third molecule differ by the position of the methoxy group. In each of them the propylene part of the molecule is a region of slightly positive potential while methoxy groups naturally have a different potential than chlorine substituents.

The closest interaction between the two molecules of No.3 shown was along a hydrogen bond, of which the partially positive charge at the hydrogen atom led to a slight blue taint on the surface. The neighboring oxygen atoms led to a negative potential and attractive interactions try to minimize the potential, which can be seen from the closer-to-zero po-

tential close to the hydrogen atom within a hydrogen bond than for those of molecules No.1 and 2, where the blue coloring close to the hydrogen atoms of the hydroxy groups is stronger. The hydroxy groups of the first two molecules could due to crystal packing still be involved in hydrogen bonding, but the potential is calculated considering only those atoms displayed.

The last compound crystallized with three molecules in the asymmetric unit and is a zwitterionic amino acid. The positively and negatively charged parts of the molecules interact with each other so that a network of strong hydrogen bonds is formed. The minima and maxima for the color mapping were adjusted to  $\pm 0.50 e/(\text{\AA}\epsilon)$  to emphasize the highly charged regions. Those show the difference to those areas with less charge like in phenyl rings. An alternative method to increase the amount of white surface area would be to adjust the preferences for the colors to be less intensive for medium charged regions.

Overall, those examples demonstrated the successful application of invariom point charges and that with the tools provided the creation of ESPs mapped on a smooth surface is facilitated.



**Figure 4.20:** Invariom point charge assignment, quick ESP computation and display of the asymmetric unit of compounds that were determined during the PhD in collaboration with synthetic work by Giovanna Petrillo in the group of Isabel Usón. Most of the color extrema were fixed to  $\pm 0.20 e/(\text{\AA}\epsilon)$ , except for (h) where they were fixed to  $\pm 0.50 e/(\text{\AA}\epsilon)$  due to its zwitterionic nature.



## 4.4 Summary and discussion

Atomic point charges were fitted to the molecular ESP for all molecules of the invariom database. Averaging those MK charges for equal invarioms yielded the invariom charge database. In combination with plug-ins for APD-TOOLKIT those charges can be assigned to organic molecules of interest. A tool for quick visualization of the resulting ESP on a smooth molecular surface was presented.

An internal charge transferability test showed that for hydrogen atoms, especially for those attached to an aromatic system, a higher degree of differentiation by inclusion of a third shell of neighbors would improve charge transferability and thus reproducibility of ESPs. Differences between charges of different hydrogen atoms that belong to one invariom significantly decreased upon transferring hydrogen-atom charges from the model compounds of the directly bonded parent atom. This approach did not require any new model compounds to be added to the database.

The ability of invariom point charges to reproduce molecular ESPs was validated for a test set of nineteen angiogenesis inhibitor molecules by comparison to several other atomic point charges: a) directly from a fit to B3LYP/6-31G(d) potentials, b) semi-empirical AM1-BCC charges and c) conceptually similar TPACM4 look-up charges. Based on the RRMS introduced by Williams *et al.* as criterion, both sets of look-up charges reproduced the theoretical potential less well than AM1-BCC charges, which however require molecule specific computations. Invariom point charges outperformed TPACM4 charges.

Comparison to ESPs from multipoles refined against experimental data also supported the suitability of the invariom point charges to quickly estimate molecular ESPs. Differences between experiment and theory were similar, independent of the kind of theory employed: MK point charges fitted to the molecular wave function, invariom multipoles or invariom point charges.

In conclusion, the here presented transferable point charges are entirely sufficient for estimating ESPs of organic molecules.



# 5 Invarioms and aspherical modeling for SHELXL – bond-oriented deformation density

## 5.1 Introduction

The transfer of aspherical electron-density fragments has been shown to work well in many cases<sup>[55–62]</sup> independent of the database used. Advantages of such a procedure include not only better figures of merit, but more precise bond distances due to less contamination of the ADP by bonding ED<sup>[57,59,60,63–67]</sup>. Although the refinement with transferred scattering factors does not require additional parameters and could therefore be applied to routine structures, applications are currently only performed by a small user group.

The main reason is that a different refinement program than the standard one for IAM is necessary to model asphericity. Refinement with *Hansen-Coppens* multipole scattering factors requires the orientation of the multipoles by local coordinate systems, which is implemented in XD2006<sup>[12]</sup>, MOPRO<sup>[68]</sup>, WINXPRO<sup>[237]</sup> and JANA<sup>[238]</sup>. This step of setting the atomic coordinate systems according to local symmetry is crucial and far from fail-safe. This makes two hurdles which entail each other. They pose a major challenge for refinement with multipole scattering factors, for those who do not use the programs regularly. This is the reason why databases containing aspherical density fragments are mainly used to supply starting values in charge density research. Researchers from this field need to handle the programs and have a deeper knowledge of orientation of local coordinate system anyway. They do not need to learn all the technicalities for the sole purpose of improving bond precision, ADP and figures of merit, but have more reasons get accustomed to the software.

Hence, it is reasonable to assume that there are many scientific projects that would benefit from a method to incorporate a description of the bonding and lone pair EDD similar to refinements with invariom scattering factors, if local coordinate systems and additional complex programs are not required.

Since the ED to be described is located in the direction of bonds, their use as orientation for placement of additional functions recommends itself. For the orthogonal coordinate systems of the *Hansen-Coppens* multipole model this usually works only for the first bond. Despite the orthogonality for the description of atomic orbitals by spherical harmonics, bonds often have geometrical arrangements resembling tetrahedrons or triangles – at least in organic molecules. Those geometries are commonly described by hybridization, thus the linear combination of atomic orbitals. Thinking intuitively the transition from molecular

geometry to orthogonal systems and back is a detour, which complicates the concepts of models for the bonding ED.

In the here presented approach bonds will directly be used as orientation for the placement of deformation functions. This is the reason for calling the new model 'bond-oriented deformation density' (BODD) throughout this chapter<sup>1</sup>.

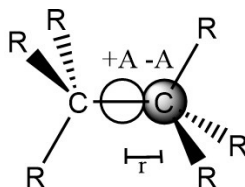
## 5.2 The concept of bond-oriented deformation density (BODD)

The ED not described adequately by the IAM is on the one hand the ED of bonds and on the other hand the ED at positions of lone pairs. Therefore, the concept of BODD is divided in two, parts one for modeling the bonding electron density (BEDE) and one for modeling the lone pair electron density (LONE).

### 5.2.1 Bonding electron density (BEDE)

The idea is based on density deformation functions similar to dipoles. They should add ED in the direction of the bond and in order to keep the overall electron count correct, subtract ED at the atomic positions. If, for example, a carbon atom with four bonds is described this way the deformation will look as presented in Figure 5.1, but the concept shown will be applied to each bond from both directions.

**Figure 5.1:** Basic concept of BEDE instruction for a standard bond.



As the most simple function to add and subtract ED, *Gaussian* functions

$$f(x) = A e^{-B(\sin \frac{\theta}{\lambda})^2} \quad (5.1)$$

will be positioned on the bond and at the atomic position; so there are two functions with position,  $r$ , amplitude  $A$  and spread  $B$  each. This amounts to six parameters per bonding direction of an atom. Because ED should not simply be added, the amplitude  $A$  at the atomic position is the negative of the other amplitude, reducing the number of parameters by one. The position of the *Gaussian* function is fixed to the atomic coordinates, which leads to only four parameters per bond.

This description via BODD circumvents the coordinate-system problem and can be applied with any refinement program that allows the user to set spherical scattering factors and free variables that can be restrained or constrained. SHELXL offer all those features.

For the purpose of this project a test version of SHELXL kindly provided by Prof. G. M. Sheldrick was used. It makes position calculation via an external program and restrain-

<sup>1</sup>A name suggested by G. M. Sheldrick to adequately convey the concept.

ing of free variables (FVAR) obsolete. This SHELXL version reads and interprets BEDE instructions. The instructions have the following syntax:

```
BEDE atom1 atom2 r A B1 B2
```

where:

atom1 is the atom the bond starts from,

$r$  the distance between the *Gaussian* function and atom1 along the bond to atom2 and

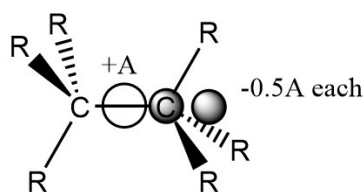
$A$  is the amplitude for the *Gaussian* function on the bond and the negative amplitude of the *Gaussian* function at atom1's position, as visualized by Figure 5.1.

$B1$  is the spread of the *Gaussian* function with  $+A$  and

$B2$  the spread of the *Gaussian* function with  $-A$ .

Both  $B$  values are multiplied with the displacement parameters of atom1. The last three parameters can be refined via free variables.

In order to deform the spherical ED as indicated by difference-density plots subtracting the IAM from the multipole model (Figure 1.3), ED should not only be subtracted from the atomic position but also opposite of the bond (as shown in Figure 5.2). If the amount



**Figure 5.2:** Basic positioning of functions along a standard bond.

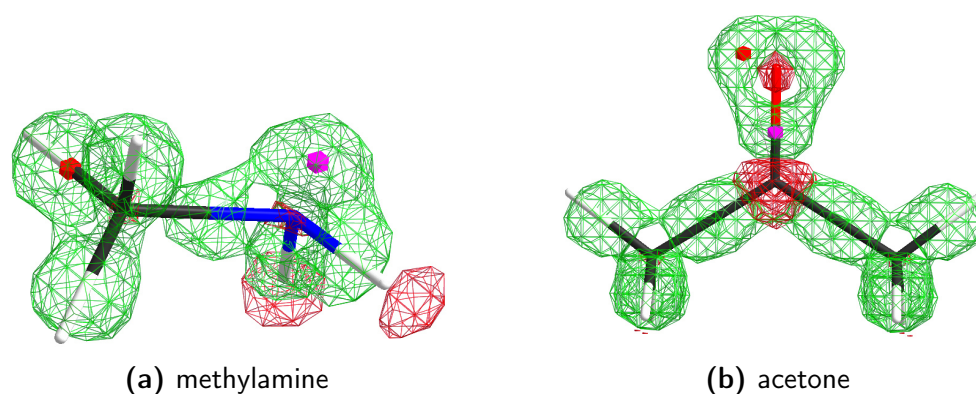
of subtracted ED should be divided by two or distributed differently is not decided at this point. Such a model can be set up in SHELXL by two BEDE instruction with the same *atom1* and *atom2*, where one dipole has a negative  $r$ .  $A$  for the instruction with negative  $r$  can then be set to be some part, e.g. half, of  $-A$  of the instruction with positive  $r$  or just be freely refined.

If a second BEDE instructions with exactly the same *atom1*, *atom2* and  $r$  is read by SHELXL, only the first one will be applied, which facilitates the use of general descriptors for bond between certain elements, for example.

### 5.2.2 Lone pair electron density (LONE)

The difference density map after IAM modeling of simulated diffraction data does not only have maxima on bonds but also at lone pair positions as shown in Figure 5.3.

The task of finding the direction for a function to model the lone pair is not as straight forward as using the bond direction. Fortunately there already is a mechanism in SHELXL to deduce the orientation of hydrogen atoms from the other bonds connected to an atom. For the BODD project application of a similar procedure to derive the orientation of lone pair density is suggested. SHELXL uses the number  $m$  to classify what kind of geometry an atom is in. An overview of all possible positions for lone pairs and their hydrogen atom



**Figure 5.3:** Residual density plot at  $\pm 0.07 \text{ e}/\text{\AA}^3$  for two molecules modeled by the IAM showing unmodeled ED for the lone pairs. Atomic positions and diffraction data up to  $0.5 \text{ \AA}$  were obtained theoretically.

analogous where applicable is presented in Figure 5.4. The syntax thus includes **m** instead of a second atom and in some cases an additional angle.

```
LONE m atom A B1 B2 r [angle]
```

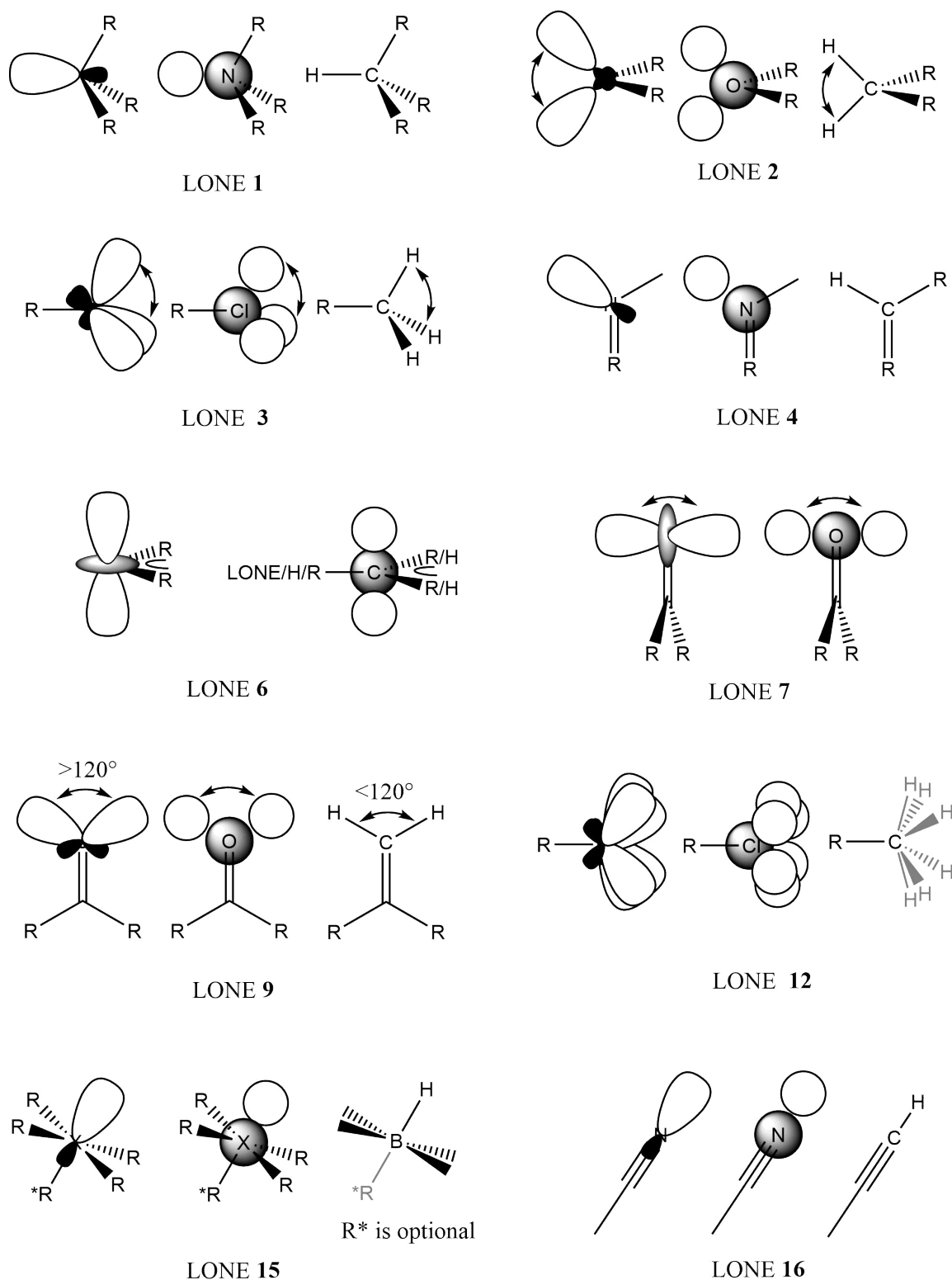
The angle applies only to **m**=2,3,7 and 9 and is in those cases fixed like it is for the hydrogen atoms. As known from the VSEPR theory<sup>[239,240]</sup> the angle may deviate from that of the ideal geometry due to valence shell repulsion.<sup>2</sup>

Similar to BEDE **B2** is the coefficient in the exponent of the subtracted *Gaussian* function. In those cases where more than one *Gaussian* function with amplitude *A* is created the *Gaussian* at the named atom will have an amplitude which is a multiple of *A*, so that the number of electrons stays balanced. For example, for **m**=2 the subtraction at the atom will be of  $2A$ . A special case is **m**=12, which corresponds to a disordered methyl group rotated from one another by  $60^\circ$ , thereby placing twelve half lone pairs in a circle, where the amplitude of the *Gaussian* function at the atom will be  $-6A$  with *A* being the height of each of the twelve *Gaussian* functions.

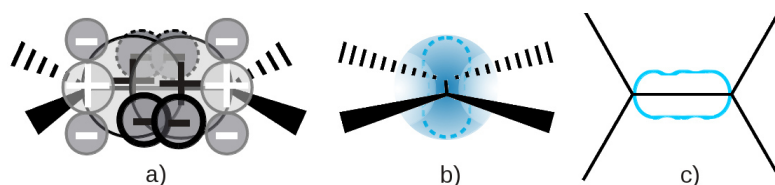
For **m**=6 and 7 there is no hydrogen analogue, because these LONE instructions are meant to be applied to  $\pi$ -bonds. While **m**=6 places *Gaussian* functions above and below the atom named in the direction perpendicular to the plane defined by two bonds of atoms (preferably to non-hydrogen atoms), **m**=7 does the same for terminal atoms where in order to define a plane the second bond is connected to the neighboring atom. **m**=9 is similar to **m**=7 but places the two *Gaussian* functions in the plane of the atoms instead of perpendicular to it.

The case of **m**=15 can be used in cases of lone pair placement for atoms with four or five bonds and one lone pair, for example  $\text{SF}_4$  or  $\text{BrF}_5$ .

<sup>2</sup>For the same reason angles between hydrogen atoms in methyl groups should be smaller than  $109.5^\circ$ , as supported by an investigation of those angles in molecules optimized by DFT (Figure 7.8 in Appendix C).



**Figure 5.4:** Concept and the different options *m* for the LONE instruction. The hydrogen atom treatment corresponding to the the same *m* is also shown, including angle expectations according to the VSEPR model.



**Figure 5.5:** Scheme for the modeling of  $\pi$  bonds with BEDE and LONE (a: side view). Visualization of how their combination can model elliptic ED along a double bond (b). (c) top view. In (a) density deformation from BEDE is lined black, while those from LONE are lined grey with white signs.

### 5.2.3 Density of pi-bonds

For the modeling of  $\pi$ -bonds a combination of several BEDE and LONE instructions is suggested as shown in Figure 5.5. Most importantly there is 'LONE 6' especially for this purpose. For aromatic systems at least  $A$  of LONE 6 will often refine to a negative value. The reason for the negative value lies in the combination of BEDE and LONE instructions applied for modeling of the double bond. The elliptic ED when looking along the bond is obtained by a comparably high  $A$  for BEDE along the bond, minus four negative amplitudes for BEDE in the negative bond direction of the bonds next to the double one. The latter BEDE functions subtract the ED in the nodal plane of the  $\pi$  orbital. Hence, LONE 6 just corrects the density orthogonal to that plane directly above and below the atom, where IAM and BEDE place too much ED.

In the case of a carbonyl bond there are no other bonds originating from the oxygen atom, so the subtraction of ED in the opposing bond direction has to come from the opposing direction of LONE 9.

### 5.2.4 Overall goal

With two BEDE instructions per bond direction, of which each has three to four refined parameters, the number of parameters can increase, even if  $B2$  is described uniformly for a structure, by twenty-four per atom<sup>3</sup>. Hence, convergence problems and overfitting are an issue if all parameters are refined against the measured diffraction data at once. Additionally, setting all the instructions up for every atom, assigning free variables, finding appropriate starting values and deciding which parameters to constrain is quite an effort.

These problems can be minimized if only a few very general BODDs are used, e.g. one BEDE instruction for all bonds between *element1* and *element2* for all element combinations present. But this would still require a lot of manual input, increase the number of refined parameters and not lead to the optimal model, if the data is not of high resolution ( $d < 0.50$  Å), low temperature and very good quality. These criteria are important for a good deconvolution of valence electron density and displacement parameters.

All of those problems can be prevented by transferring the BODD according to atom types from a database. Therefore, the goal is to code a preprocessor program that can be used easily. It shall classify the atoms according to their bonding environment, look up all the BEDE and LONE instructions for this classification and automatically write them

<sup>3</sup>Two BEDE instructions times three parameters times four bonds.



to an additional instruction file which is then included in the main SHELXL instruction file by '+filename.ins'. The entries in the database of the BODD can be obtained from refinements against simulated diffraction data to allow complete deconvolution from thermal parameters. In the end the parameters transferred to the molecule for the BODD are kept fixed at the theoretical values and the same number of parameters are refined against the experimental X-ray data as in an IAM refinement.

The overall aim is a method similar to the invariom scattering factor assignment but instead of multipole populations BEDE and LONE parameters are transferred. But from the application point of view, switching from SHELXL to XD should not be necessary anymore. The only action the user should have to perform is to request a BODD model and to run a refinement once the parameters are transferred. The thus refined model should include the valence ED and therefore yield lower uncertainties for atomic coordinates, bond lengths and angles. Thermal parameters should be lower on average, since they are deconvoluted from the bonding ED and their physical significance should increase.

### 5.2.5 Comparison to previous modeling of aspherical density by dummy atoms

Brill<sup>[29]</sup> was one of the first to include bonding ED in a model for XRD data. He showed the improvement of a "reliability factor" upon locating half an electron at the midpoint between bonded carbon atoms in the model of diamond. He applied a *Gaussian* function as form factor for this charge as suggested theoretically by Ewald and Hönl in 1936<sup>[30]</sup>. Those two researchers were already aware that by adding electrons on the bond, they had to subtract an equivalent number from the atomic form factor of the independent atom. Brill, Dietrich and Dierks also investigated the bonding electron-density distribution in decaborane in 1970<sup>[241]</sup>. Since more than two atoms had to be modeled in the asymmetric unit, coordinates and displacement parameters were derived from neutron diffraction, while the bonding ED was refined against X-ray data by modeling fractions of hydrogen atoms on the bonds and refining coordinates for those dummy atoms.

In 1977 Hellner introduced his "*Simple Refinement of Density Distributions of Bonding Electrons*"<sup>[31]</sup>. He divided the structure factor into contributions from core electrons, bonding valence electrons and non-bonding electrons. The bonding valence electrons were modeled by three dimensional *Gaussian* functions that inherited 10 % of the two bonded atoms'  $B_{ij}$  parameters. The center of the functions which were called 'charge clouds' was placed in the middle of each bond in a molecule. Populations and ADP were refined for each 'charge cloud' separately. Together with Mullen, Hellner presented a refinement of this model against experimental data for diborane in 1977<sup>[242]</sup> and later for urea<sup>[243]</sup> and thiourea<sup>[244]</sup>. The publication series on diborane continued: together with Scheringer<sup>[32]</sup> a comparison to QM calculations, including a detailed description of difference density maps between molecular and free-atom model, which later became known as the 'deformation-density map' was performed. Scheringer and Dietrich continued to apply the model to cyanuric acid<sup>[245]</sup>, for which they refined three smearing parameters for the charge clouds on bonds. The off-diagonal elements were set to zero and were oriented along the bond,

resulting in seven parameters per cloud at most<sup>4</sup>. The ADP were averaged between the two connected atoms or set to those of the according atom for lone pair charge clouds.

In 1980 Mullen also turned to model urea by the multipolar expansion, already much discussed before.<sup>[51]</sup> A comparison of the two models led to the conclusion that pseudoatom modeling of the valence ED is similar to refinement of quadrupoles with respect to residual factors and deformation-density maps. Both "*fail to describe the density in the C-O bond properly with the occurrence of a hole (negative peak) near O*"<sup>[35]</sup>. Lone pairs of oxygen are described differently with octupoles and hexadecapoles. This deficiency of not being able to model the hole close to oxygen could be different with the BEDE model, since it is designed for at least two pseudo-atoms per bond. The number of parameters is still lower in the BEDE model, since ADP are not refined for pseudo-atoms of the valence electron density. They are just modeled by one parameter scaling the displacement to that of the neighboring atom.

Many years later in 2004 and 2007 another attempt at modeling the bonding ED by pseudo-atoms in the valence density for bonds in proteins<sup>[36,37]</sup> was published. Afonine *et al.* applied dummy bond electrons, of which they refined occupancy and  $B_{iso}$ .

Moreover, Jelsch *et al.* recently investigated modeling the bonding and lone pair density by dummy atoms with respect to the electrostatic potential for small molecules and additionally compared the results to dummy bond parameters transferred from a database as well as a multipole model.<sup>[38,39]</sup> In contrast to the invariom database their database is based on experimental data, which limits the addition of new compounds and can be biased by systematic errors. The model applied by Jelsch *et al.* consisted of "*a superposition of real and virtual spherical atoms*"<sup>[38]</sup>. Positions, populations and contraction parameters were refined for real and virtual atoms. Contraction of lone pair virtual atoms turned out to be higher than for virtual atoms of the bonding ED. Positions of the virtual atoms were heavily restrained. Isotropic displacement parameters were relative to those of the real atoms.

The BODD model differs from all the models before by the use of two instead of one virtual atom, charge cloud or dummy atom per bond. Moreover, the density is a direct deformation of the spherical density, since each addition of ED is accompanied by a subtraction. Whereas most of the models before had also some means to keep the electron count correct, the subtraction opposite of the bonds is also new. The BODD concept is thus closer to the multipole model. Instead of requiring orthogonal coordinate systems, however, it keeps the advantage of easy bond orientation.

---

<sup>4</sup>One parameter for charge, three for position and three for the smearing tensor.

## 5.3 Experiments and Results

### 5.3.1 A first look at BODD in SHELXL and comparison to the multipole model in XD

#### 5.3.1.1 Aim

BODD parameters were refined against simulated XRD data to assess to which degree the residual density maps improve, as a first feasibility test.  $R1(\text{all})$  values were evaluated and compared to XD refinements using multipole parameters. In case  $R1 < 1.0\%$ , the experiment should be considered successful.

#### 5.3.1.2 Procedure

BEDE and LONE parameters for selected model compounds were refined against simulated diffraction data stored in the invariom database. Geometries and ED were derived from M06/def2TZVP calculations. In contrast to the usual reflection files for XD in the invariom database, an isotropic thermal motion of  $u_{\text{iso}}=0.01$  was added to the simulated diffraction data before refinement in SHELXL to avoid unnecessary warnings concerning non-positive definite ADP. The resolution for SHELXL refinements was set to  $0.50 \text{ \AA}$ .

The bonding and lone pair ED of the molecules was modeled by the new BEDE and LONE instructions. Parameters  $A$ ,  $B1$  and  $B2$  were refined via a least-squares algorithm against the simulated data, while  $r$  and the angle, where applicable, were changed and optimized manually. Atomic positions and displacement parameters were fixed to calculated values in all of the refinements, thus for the IAM model only the overall scale factor was refined.

The IAM model was refined first to show the improvement upon a spherical modeling. This initial refinement was performed with both programs, SHELXL and XD, to see within which tolerances the values between the different programs should be evaluated. For the seven molecules investigated in these first tests of the BEDE and LONE model the results are shown in Figure 5.10.

The molecules were chosen to incorporate different chemical groups. Cyclohexane is a simple aliphatic molecule in the database with just two different invariom names in it. Benzene is the archetype of an aromatic compound. The two fluorocompounds were investigated to evaluate the treatment with BEDE and LONE in the absence of hydrogen atoms, which are always special in refinements against X-ray data. Acetone and formaldehyde are considered because of their double bond and lone pairs at the oxygen. In furan the situation at the oxygen atom is not easy to model, either. More molecules were investigated in later experiments. Some interesting peculiarities of the first refinements with the new instructions are shown and explained below.

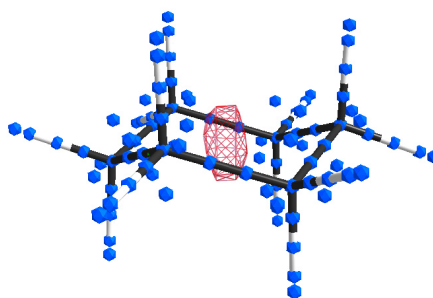
#### 5.3.1.3 Observations during the procedure

**Cyclohexane** contains only three different bonds or, without conformational differences, only two: those in between carbon atoms and those between carbon and hydrogen atoms. With two BEDE instructions per bond and direction, one with a positive and one with a negative  $r$ , the SHELXL instructions for the final refinement were:

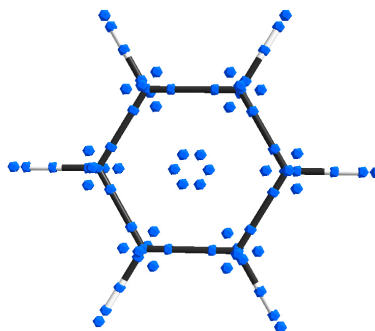
BEDE	\$C	\$C	0.55	41	21	31
BEDE	\$C	\$H	0.38	51	21	31
BEDE	\$H	\$C	0.36	61	21	31
BEDE	\$H	\$C	-0.20	71	21	31
BEDE	\$C	\$C	-0.40	81	21	31
BEDE	\$C	\$H	-0.55	91	21	31

and can be visualized as shown in Figure 5.6.

**Figure 5.6:** Residual density at  $\pm 0.02 \text{ e}/\text{\AA}^3$  of cyclohexane. Positions of BEDE are represented by blue icosahedra.



**Benzene** was modeled with BEDE instructions similar to those used for cyclohexane. The distances  $r$  for carbon-carbon bonds were reduced due to the shorter bonds. According to the bond strength  $A$  was higher, too. Since there was negative residual density in the middle of the aromatic ring, a third BEDE instruction with a negative  $r = -1.15$  was introduced to subtract this. The biggest difference to cyclohexane is the use of LONE 6 at the optimized distance of  $0.43 \text{ \AA}$ .  $A$  refined to  $-0.1059(2)$ . It was beneficial to refine  $B1$  by additional free variables for: LONE instructions, BEDE between carbon atoms and BEDE between carbon and hydrogen atoms. This way the original variable for  $B1$  remains for all BEDE instructions with a negative  $r$ . At this point the necessity of such a high level of differentiation for refinement against real, experimental data will not be discussed. But against simulated data with fixed atomic positions it was possible to refine several free variables for  $B1$ . It did not only lead to a drop in  $R1(\text{all})$  by  $0.1 \%$ , but also contributed to a very low level of residual density, as there is no residual density visible at a level  $0.02 \text{ e}/\text{\AA}^3$  in Figure 5.7.

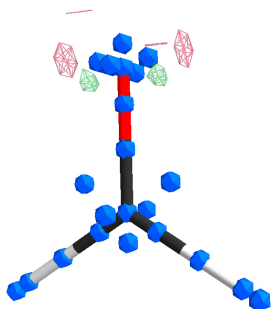


**Figure 5.7:** Residual density at  $\pm 0.02 \text{ e}/\text{\AA}^3$  of benzene. Positions of BEDE and LONE are represented by blue icosahedra.

**Formaldehyde** is similar to acetone and has, in addition to the LONE 6 instruction for the carbon atom, two LONE instructions for the oxygen:

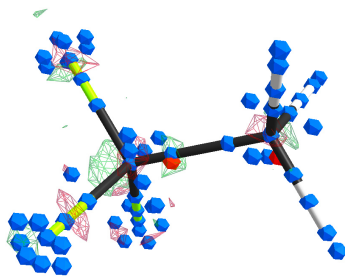
```
LONE 6 C2 41 121 31 0.4
LONE 7 O1 111 131 31 0.23 185
LONE 9 O1 101 141 31 0.23 147
```

One interesting observation for both carbonyl groups, in formaldehyde as well as in acetone, was that the amplitude of the BEDE function placed close to oxygen was negative. This was not a refinement error but could also be seen in the residual density before refinement. This observation can at least partly be attributed to the fact that BEDE values as well as the residual density are based on the IAM. Therefore negative amplitudes simply mean less density than modeled by all other surrounding functions, but not an altogether negative ED. The positive *Gaussian* function in the middle of the carbonyl bond has a height of 1.39 electrons. Similar features for carbon-oxygen bonds have been reported in the literature.<sup>[35]</sup>

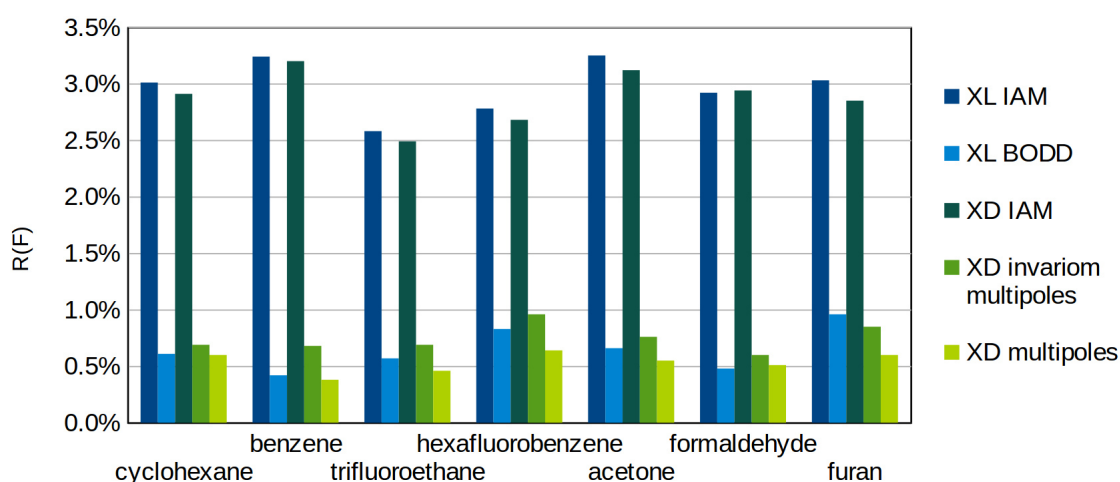


**Figure 5.8:** Residual density at  $\pm 0.02 \text{ e}/\text{\AA}^3$  of formaldehyde. Positions of BEDE and LONE are represented by blue icosahedra.

**Trifluoroethane** was modeled with the usual BEDE instructions. The bonds to fluorine were modeled with two *Gaussian* functions that are both placed closer to fluorine than to carbon (see Figure 5.9). The one positioned only  $0.26 \text{ \AA}$  from fluorine has an amplitude of  $-0.18$ . Even the one in the middle of the bond is negative. Additionally,  $0.12$  electrons were transferred from the carbon atom to each fluorine atom. The lone pairs of fluorine were modeled with LONE 12 at an angle of  $110^\circ$  and a distance of  $0.3 \text{ \AA}$ .



**Figure 5.9:** Residual density at  $\pm 0.02 \text{ e}/\text{\AA}^3$  of trifluoroethane. Positions of BEDE and LONE are represented by blue icosahedra.



**Figure 5.10:**  $R(F)$  ( $\equiv R1(\text{all})$ ) for different refinements against simulated data compared to BODD refinement with SHELXL for selected model compounds.

#### 5.3.1.4 Results

As shown in Figure 5.10,  $R1(\text{all})$  decreased considerably upon inclusion of asphericity in the model. While the refinements of IAM against simulated data yielded an  $R(F)$  of 2.5 to 3.0 % for a resolution of 0.50 Å, all aspherical models reached numbers well below 1.0 %. The SHELXL BODD bars are smaller than those of the XD invariom refinement, but taller than those where the multipole parameters were refined. During invariom refinement only one parameter was adjusted, exactly like in the IAM refinements. In BODD and multipole refinements more parameters were refined for the description of the bonding and lone pair ED.

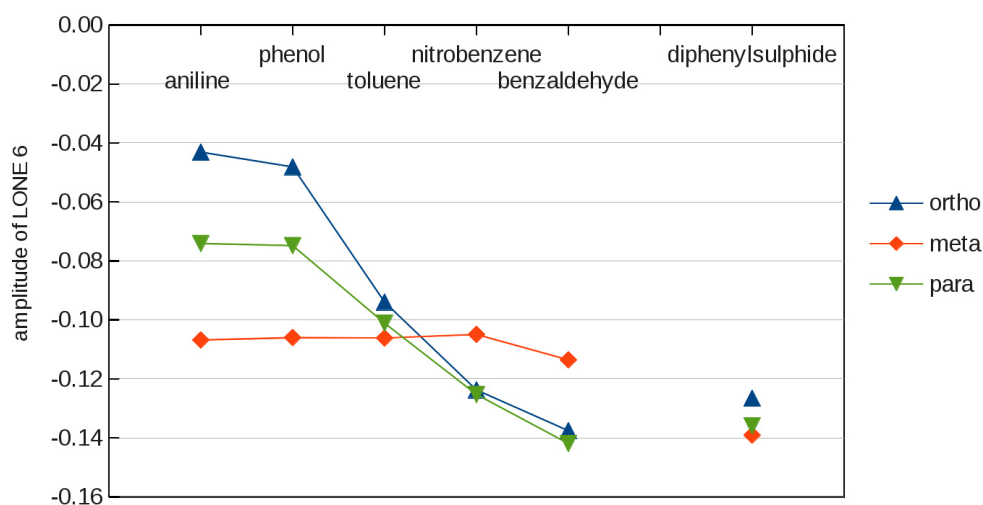
Furan provided some problems due to the lone pair density, which has its reason partly in the involvement of the p-orbitals in the  $\pi$ -system and the  $sp^2$ -orbital. This strict assignment of the two orbitals and modeling thereof seemed not completely adequate, and the five-membered ring also leads to slight strain on the system, that made it harder to fit by BODD (See Appendix C Figure 7.9).  $R(F)$  for the XD refinements of furan were, however, also above average, though not as much as for BODD.

Overall, the experiment was successful, since the goal was to reach  $R(F)$  values below 1.0 %. It is possible to model bonding and lone pair ED by BEDE and LONE, which decreased  $R1(\text{all})$  compared to IAM almost as much as the multipole model. Lone pairs and strained systems seem a challenge for modeling with BODD.

### 5.3.2 Chemically meaningful parameters

During the refinements of different aromatic compounds for the next chapter, 5.3.3, the question arose how many different variables should be used for LONE 6 of the carbon atoms in an aromatic ring. In order to find an answer, different variables in *ipso*, *ortho*, *meta* and *para* position to the substituent were compared for different functional groups.

In principle those substituents with similar mesomeric effect should yield similar values for the amplitude  $A$  of BEDE 6. The *ipso* atom may deviate from this trend due to the



**Figure 5.11:** LONE 6  $A$  values refined at a distance of  $0.43 \text{ \AA}$  from the corresponding atom for carbon atoms in aromatic systems with substituents of different mesomeric effects.

effects of the neighboring bond to very different atoms but *ortho*, *meta* and *para* could possibly show this trend. The LONE  $A$  values were refined with a distance of  $r=0.43 \text{ \AA}$  to the corresponding atoms. The resulting values of this comparison are shown in Figure 5.11.

For aniline and phenol, which have a positive mesomeric effect, less ED is subtracted from the  $\pi$ -bond ED above close to *ortho* and *para* carbon atoms. For nitrobenzene and benzaldehyde, which have a negative mesomeric effect (taking away  $\pi$ -bond ED), there is less  $\pi$ -bond ED modeled below and above the *ortho* and *para* carbon atoms. Toluene, which has no mesomeric effect, is in-between the two cases.

Diphenylsulphide served as an example for substitution by a third-row element and hence a weaker positive mesomeric effect than phenol. Due to the more diffuse valence electrons the effects on the  $\pi$ -system could be completely different here. Unfortunately, the refinement for diphenylsulphide led to an unreasonably high value for  $B2$  of two, which caused unmodeled ED at the atomic positions, so that all LONE 6 would have a reason to relocate density upon the atomic centers. This is also reflected in the amplitudes. All  $A$  for *ortho*, *meta* and *para* are more negative than those for the other investigated compounds. But the general relation of the data points of diphenylsulphide would arrange it between phenol and toluene, which seems chemically reasonable.

This comparison is not only an example that BODD parameters can describe chemical trends, but also that it can easily be related to basic chemical concepts. Their resemblance to hybridized orbitals is appealing and is inline with concepts used in organic chemistry.

The multipole approach starts from functions similar to the orthogonal atomic orbitals, which at least in this case are not interpreted as easily for organic chemistry. Considering populations of the multipoles resembling a  $d_{z^2}$  orbital (Q0 in XD) equivalent to the amplitudes of the LONE 6 description above did not yield the same trend as the one observed for LONE. As Figure 7.10 in Appendix C shows, only the population at the atom in *ortho* position changes in a similar manner. Assuming that a negative population will mean less ED in the lobes along the  $z$ -axis, which is oriented in a way that it is perpendicular to the aromatic carbons, the direction of the trend is also the same as for LONE. But the

populations for *meta* and *para* are similar. They were refined starting from a model where they had been restrained, but this was the case for the LONE parameters as well,

Since substitution reactions show similar preferences for *ortho* and *para* positions with respect to the electron withdrawing character of a group, it is more reasonable that *ortho* behaves similar to *para*, than *meta* and *para* behaving similarly.

Moreover, nitrobenzene shows very different  $d_{z^2}$  populations than all other systems, although inductive and mesomeric effects should be of a similar direction as in benzaldehyde. A direct interpretation of multipole populations rather than the overall ED is difficult.

### 5.3.3 Transferability study I: aromatic carbon

#### 5.3.3.1 Study design

BEDE and LONE parameters for selected, exemplary model compounds which contain carbon atoms with the invariom name  $6-C\#6c[\#6c1h]\#6c[\#6c1h]1h$  were refined as described in the previous section (5.3.1). The selected compounds were benzene, biphenyl, naphthalene<sup>5</sup>, toluene, phenol diphenylether, aniline, diphenylamine, styrene, benzaldehyde and pyridine. All BODD parameters from those compounds for invariom  $6-C\#6c[\#6c1h]\#6c[\#6c1h]1h$  and the bonded hydrogen atom were averaged and compared.

Next, BODD parameters for the general aromatic carbon atom, its hydrogen atom and bonds between them were set to fixed values for each compound in the study. Selection of the general aromatic carbon atoms, for which the values were averaged and set, varied. Additional to values averaged from all compounds, those from the invariom model compound benzene were transferred unchanged for comparison. So overall four transfers were carried out:

- parameters averaged from the selected model compounds for
  - the whole invariom name  $6-C\#6c[\#6c1h]\#6c[\#6c1h]1h$
  - the shortened invariom name  $6-C\#6c\#6c1h$  with nearest neighbors only (NN)
- parameters transferred from the invariom model compound benzene
  - to carbon atoms of the invariom name  $6-C\#6c[\#6c1h]\#6c[\#6c1h]1h$
  - to carbon atoms of the shortened invariom name  $6-C\#6c\#6c1h$  with nearest neighbors only (NN).

#### 5.3.3.2 Results from averaging

The mentioned model compounds contained 40 bonds between carbon invariom  $6-C\#6c[\#6c1h]\#6c[\#6c1h]1h$  and a hydrogen atom ( $H@6c$ ) and 56 bonds between two of the carbon atoms. BEDE parameter averaging for those bonds and LONE yielded:

---

<sup>5</sup>Naphthalene does not contain the invariom  $6-C\#6c[\#6c1h]\#6c[\#6c1h]1h$ , but atoms that are described in the shorter invariom name  $6-C\#6c\#6c1h$ , which was also tested.



BEDE \$C \$C	0.359(11)	1.12(8)	0.106(3)	0.88(10)
BEDE \$C \$C	-0.299(4)	-0.69(7)	0.113(6)	0.88(10)
BEDE \$C \$H	0.22(4)	0.54(9)	0.086(8)	0.87(10)
BEDE \$C \$H	-0.300	-0.86(8)	0.113(6)	0.87(10)
BEDE \$C \$H	-1.123(5)	0.041(4)	0.113(6)	0.87(10)
BEDE \$H \$C	0.400	0.57(3)	0.090(3)	0.87(10)
BEDE \$H \$C	-0.200	-0.210(12)	0.113(6)	0.87(10)
LONE 6 \$C	-0.100(15)	0.062(6)	0.87(10)	0.430

Values for LONE have been discussed in detail in Section 5.3.2. For the BEDE amplitudes *A* the standard deviation is around 10 %, and *B*2 is around ten times the value of *B*1.

For the nearest neighbor (NN) invariom names the results are derived from 108 carbon-carbon bonds, 68 carbon hydrogen bonds and 68 equivalent atoms for LONE. The standard deviations rose only slightly and the value changes are small:

BEDE \$C \$C	0.358(10)	1.13(8)	0.106(3)	0.87(9)
BEDE \$C \$C	-0.299(4)	-0.70(8)	0.113(6)	0.87(9)
BEDE \$C \$H	0.23(3)	0.53(9)	0.086(8)	0.86(9)
BEDE \$C \$H	-0.300	-0.86(8)	0.113(6)	0.86(9)
BEDE \$C \$H	-1.123(5)	0.039(4)	0.113(6)	0.86(9)
BEDE \$H \$C	0.400	0.57(3)	0.090(3)	0.87(9)
BEDE \$H \$C	-0.200	-0.210(13)	0.113(6)	0.87(9)
LONE 6 \$C	-0.094(22)	0.062(5)	0.87(9)	0.430

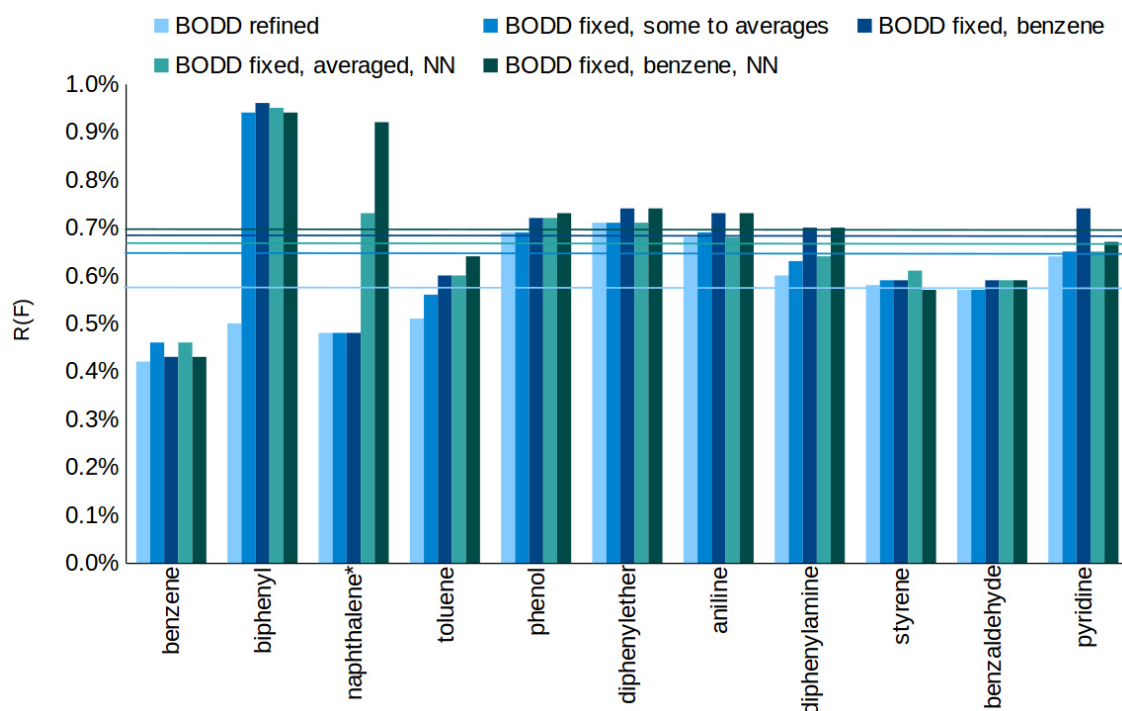
This suggests that at least for aromatic systems similar to the chosen ones shorter invariom names can be used if parameters are derived from combinations of invariom names. It also confirms that averaging is not necessary and that the empirical rules applied for the transfer from the invariom database could be relied upon.

### 5.3.3.3 Results from refinements

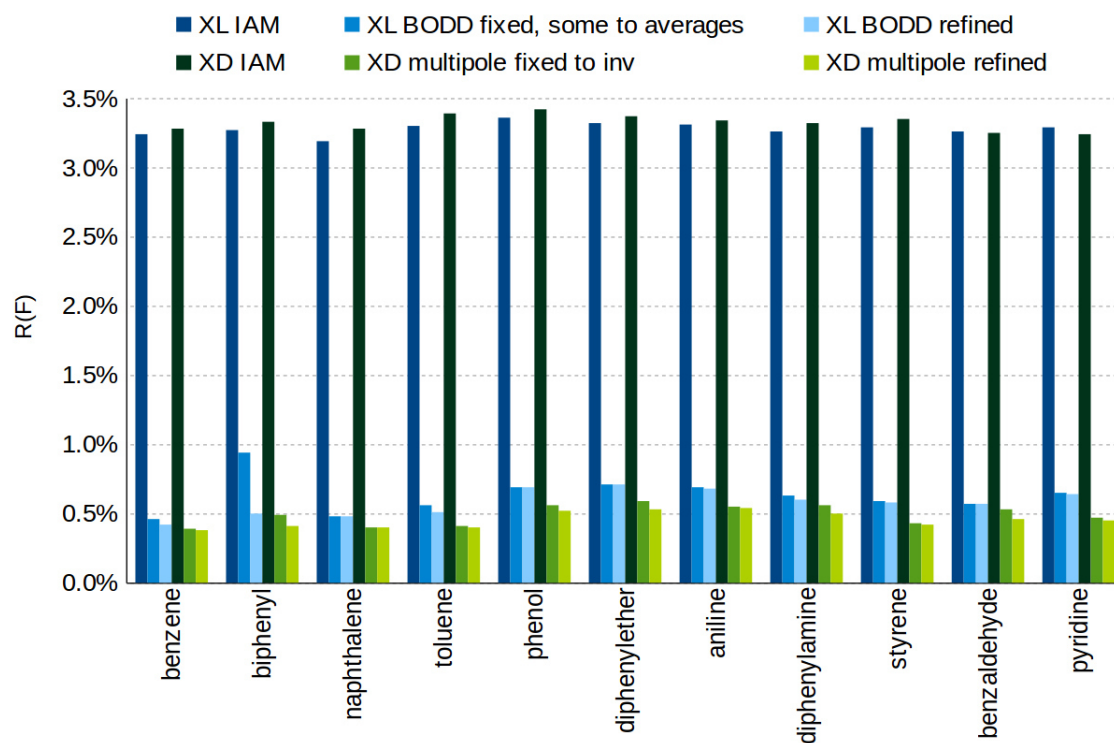
If used in a refinement those parameters averaged according to the long invariom name should yield the best result. Indeed this is what was observed for all of the model compounds as shown in Figure 5.12.

In the case of styrene the experiment showed that the refined BODD parameters via manual adjustment of the distances did lead to the optimal result, since the parameters transferred from benzene yielded an even better fit to the simulated data. Otherwise *R*(*F*) is, as expected, slightly higher when averaged parameters are applied compared to a free refinement.

On average the differences between values from benzene or averaged ones almost level out. For some compounds the benzene parameters (darkest bars, Figure 5.12) perform



**Figure 5.12:** Comparison of differently transferred BODD parameters in SHELXL refinements for each molecule and lines for the averaged values. \*Naphthalene contains only carbons of invariom 6-C#6c#6c1h not of the longer classification.



**Figure 5.13:** R(F) for refinements of transferred and refined BODD and multipole refinements with XD and SHELXL.

better, for others the averaged ones (medium dark bars). Naphthalene is quite the exception since it is only relevant for the NN test, but it is the molecule for which transferred parameters fit the worst. This is an indicator for still using long invariom names. In all other cases the shorter NN invariom names (greenish coloured bars) perform only slightly worse.

If the results for the averaging with respect to the long invariom names are compared to invariom refinements in XD and IAM refinements (Figure 5.13), the difference between the freely refined BODD parameters for invariom 6-C#6c[#6c1h]#6c[#6c1h]1h and the refined values is only minor and comparable to the difference between freely refined XD multipole parameters and invariom multipole parameters. Only in biphenyl the fixed BEDE parameters lead to an outlier in this series. For all molecules R(F) stays below 1 % and yields a considerable improvement to the IAM models.

### 5.3.4 Transferability study II: from several small molecules to a larger one

#### 5.3.4.1 The test molecule

The test molecule was chosen for a planned comparison of results from BODD refinement against XRD data to those from refinement against neutron diffraction data (see Section 5.3.5). The requirement of high-quality neutron diffraction data limited the number of possible test molecules to a small number.<sup>[236,246–248]</sup> The choice fell on methylbenzylamino-dinitropyridine (MBADNP, officially 3,5-dinitro-2-[1-phenyl-ethyl]aminopyridine),<sup>[248]</sup> due to a high diversity of chemical environments in the molecule (Figure 5.14). The data were measured at 20 K. The molecule is neutral and has only one possible donor group for strong hydrogen bonds. This reduces the number of potential problems. Once the method works for this first test case with a low number of expected issues, studies concerning the challenges of charged compounds and hydrogen bonded systems can be performed. Therefore, MBADNP was selected as subject for the following studies. The first studies will focus on simulated, ideal diffraction data; refinements against experimental data will

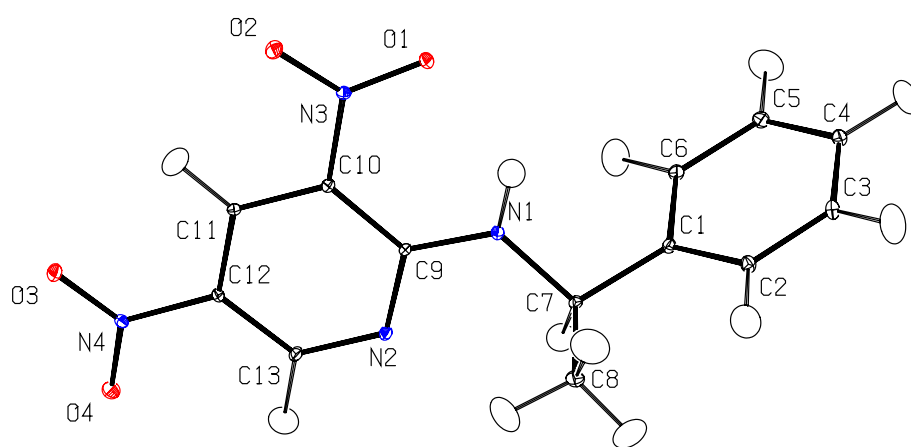


Figure 5.14: ORTEP representation of MBADNP refined against neutron data measured at 20 K.

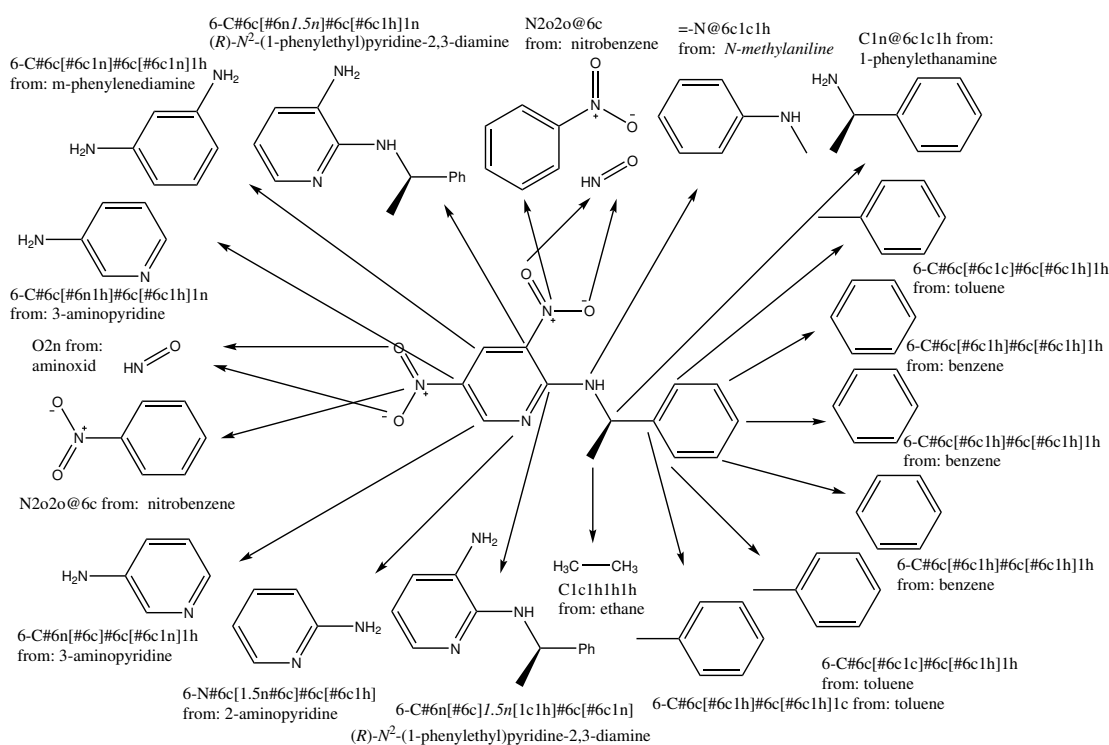
follow in Section 5.3.5.

### 5.3.4.2 Design of the study

The molecular wave function for MBADNP was calculated with M06/def2TZVP in GAUSSIAN09.<sup>[103]</sup> The three-dimensional coordinates were set to those from the model refined against neutron data. Using TONTO<sup>[249]</sup> the resulting ED was placed in a rectangular unit cell with  $a=b=c=30$  Å of symmetry  $P\bar{1}$ , like all model compounds of the invariom database, in order to receive simulated diffraction data up to a resolution of 0.50 Å. This static model was convoluted with an isotropic thermal motion of  $u_{\text{iso}}=0.01$  before the first refinement, which was a scale factor refinement with SHELXL of the IAM model and worked as reference refinement.

In all refinements of this study atomic positions and displacement parameters were fixed to theoretical values. Bonding and lone pair ED of the molecule was described by the new BODD model. In resemblance to refinements with invariom multipole scattering factors the BODD parameters were transferred from model compounds of the invariom database. The results of the refinement with the transferred but not further refined BEDE and LONE parameters then were compared to a refinement, where  $A$ ,  $B1$  and  $B2$  were adjusted<sup>6</sup>.

The model compounds from which the parameters were transferred were treated by the same procedure as described for the previous study (section 5.3.1). Selecting the model compounds needed for MBADNP was more challenging than expected.



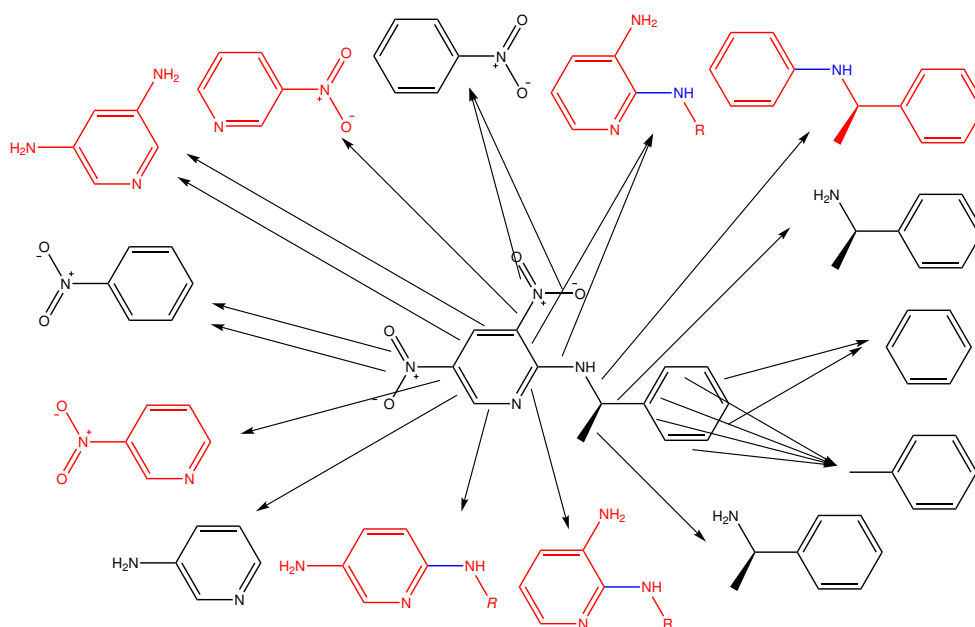
**Figure 5.15:** Model compounds for MBADNP assigned for each non-hydrogen atom according to invariom name.

<sup>6</sup>applying the transferred parameters as starting values for the least-squares refinement

### 5.3.4.3 Transfer of BEDE parameters – option 1: based on bond classification

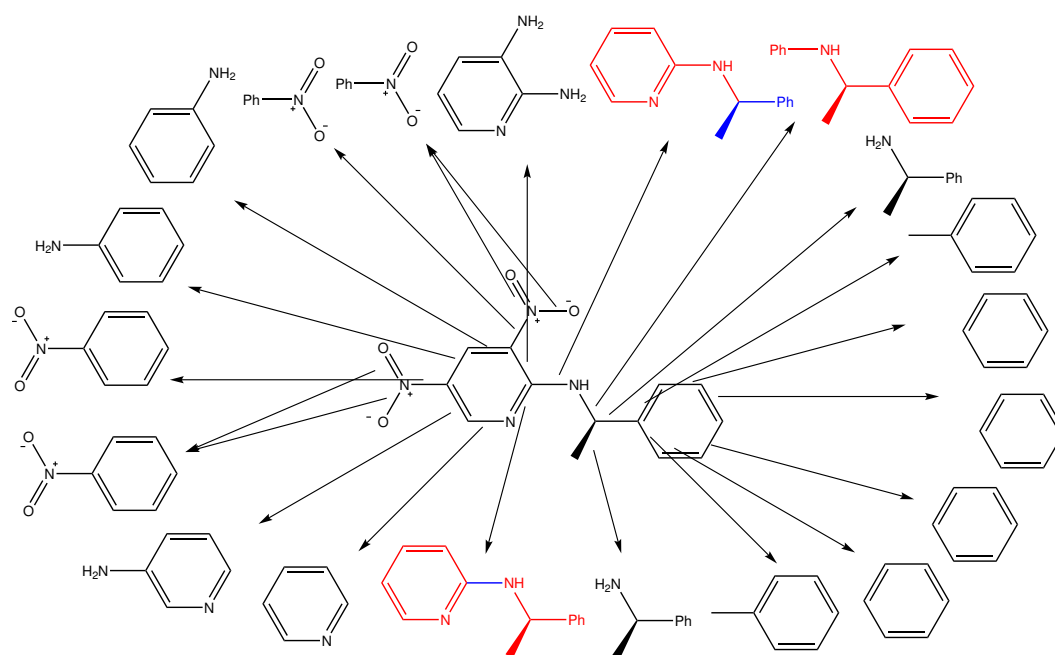
Standard invariom names for MBADNP and the thereby assigned model compounds as used for multipole scattering factor assignment are shown in a plot type as suggested by Julian Holstein<sup>[75]</sup> in Figure 5.15.

The problem is that at first sight BEDE is focused on bonds between two atoms rather than just a single atom, so that for a set of parameters for BEDE two invariom names have to be considered. One way to solve this problem would be to increase the model-compound size. An assignment of model compounds for each bond between non-hydrogen atoms is shown in Figure 5.16. All bonds connecting two invarioms from different model compounds have the potential to require a new model compound and those may not be present in the database, yet. The compounds necessary for the combination of two invarioms but not for a single invariom are colored in red in Figure 5.16.



**Figure 5.16:** Model compounds for MBADNP assigned for each bond between non-hydrogen atoms according to invariom name including next-nearest neighbors. Bond and nitrogen atom with a possibly different classification in the model compound are marked blue. Compounds not in the database are marked red.

One possible way around the many missing model compounds would be to shorten the invariom names by excluding next-nearest neighbors from the shell considered for assignment. (Figure 5.17) This yields smaller and thus fewer model compounds, but it does not resolve the problem for bonds which did not include any NNN before and came from two different model compounds. In MBADNP test molecule this procedure would require only two additional model compounds; in the database, however, 467 of 3486 invarioms do not include NNN. Those including an @ would probably not need a new compound if combined with this atom, but there are 234 invariom names without NNN and without @. Each of them can probably be combined with several other invarioms which would be missing from the database, so that too many bonding environments would have to be added to



**Figure 5.17:** Model compounds for MBADNP assigned for each bond between non-hydrogen atoms according to invariom name without next-nearest neighbors. Bond and nitrogen atom with a possibly different classification in the model compound are marked blue. Compounds not in the database are marked red.

the database. Therefore, even the shortened invariom names are unfeasible as a general concept.

For the purpose of this study the missing model compounds were treated as the model compounds of the database and refined with a BODD model in order to have a good transfer model to start with, just as a proof of principle in this pilot study.

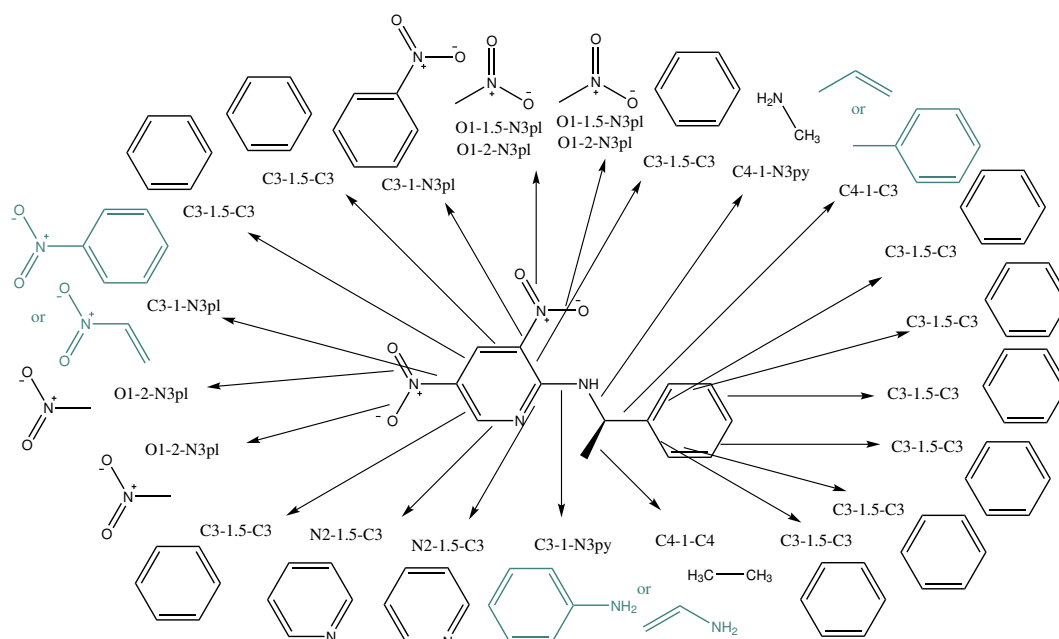
### New classification of BEDE parameters

For the future, a different classification of bonds is necessary to reduce the number of model compounds. One idea would be to use the bond order, the elements connected and the number of bonds the connected atoms form. Nitrogen atoms with three additional bonds need to be differentiated by planarity. Such a naming scheme and model compound assignment are shown for MBADNP as a Gedankenexperiment in Figure 5.18

Planar carbon atoms with three bonds could be found in aromatic systems or at a double bond. The different model compounds for ambiguous cases are highlighted in green in Figure 5.18. A single bond between an  $sp^3$ -hybridized carbon atom and an  $sp^2$ -hybridized carbon atom (C4-1-C3) could come from toluene or propene. The same applies if a planar or a non-planar nitrogen atom instead of an  $sp^3$ -hybridized carbon atom is involved. So in such cases transferability would need to be compared to decide on the best model compound.

#### 5.3.4.4 Transfer of BEDE parameters – option 2: based on atom classification

Instead of the completely new concept of a bond lookup table, there could also be a different way of archiving and assigning the BEDE parameters per invariom, which is closer to the



**Figure 5.18:** Model compounds for MBADNP assigned to each bond between non-hydrogen atoms according to very simplified classification of bonds and atoms.

original invariom method. Two approaches for storing, transferring and assigning the BEDE parameters are suggested:

**Approach A** BEDE and invariom name followed by

- parameters for shortest bond
- parameters for second shortest bond
- parameters for third shortest bond
- ... as many as the invariom has bonds

or alternatively:

**B** BEDE and invariom name followed by:

- parameters for first neighbor atom in invariom name
- parameters for second neighbor atom in invariom name
- parameters for third neighbor atom in invariom name
- ... as many as the invariom has bonds

This would not require new invariom names and the assignment of model compounds would be the same as for the 'traditional' invariom procedure (according to Figure 5.15). Approach A could have the problem that two almost equally long bonds could accidentally

be switched. This would not be a problem if the bonded atoms were similar, but otherwise this could lead to wrong assignments. Moreover, for bond distances that vary during refinement the assignment may depend on the refinement step after which the transfer is carried out. Hence, it would seem advisable to add some more details to the order than just the bond length to lead to a stable assignment mechanism. At least the element type should be integrated in the ordering mechanism.

Approach B is less dependent on bond length since it focuses on the invariom name. Therefore problems only occur upon changes of bond order. A technical challenge for approach B is the implementation, where ordering in the invariom name is not related to atom names anymore.

Although both alternative approaches, which stay closer to the original invariom concept, seem a good approach for the future, they require considerably more thought and programming. Prior to this the result of this pilot study needs to establish transferability as feasible. Therefore, the bond-based approach as summarized in Figure 5.17 is tested in this study on MBADNP.

#### 5.3.4.5 Transfer of LONE parameters via invarioms – the nitrogen problem

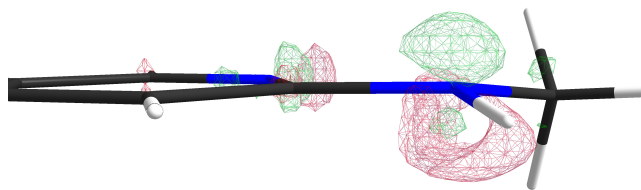
There are several kinds of nitrogen atoms with three bonds in neutral organic molecules: primary amines (as in ethylamine and aniline), secondary amines (as in dimethylamine and diphenylamine) and tertiary amines. Of course there are also ammonium cations, which in principle could be quaternary, but since ammonium cations have no lone pair at the nitrogen atom there is nothing to be modeled by LONE. Those nitrogen atoms involved in double and triple bonds are also easy to handle, since they have a fixed configuration.

Amines, however, can be either planar or not, depending on their hybridization state. While primary amines tend to be pyramidal, secondary amines can become planar in conjugation to aromatic systems or double bonds. Predicting planarity is not easy. Additionally the threshold of planarity is artificial and in the case of invariom classification empirically derived for the specific purpose.

Especially difficult is the problem of planarity for secondary amines, since the hydrogen atom positioning determines the geometry. The hydrogen atom parameters are the least well defined in a crystal structure refinement against X-ray data. Primary amines have the same problem but can usually be considered to be pyramidal as geometry optimizations of aniline, 2-aminopyridine, 3-aminopyridine, 2,3-diaminopyridine, ethylamine, methylamine and aminoethene suggest. Secondary amines between two  $sp^2$ -hybridized atoms as in diphenylamine are planar and can be modeled best by LONE 6, but if only one of the two carbon atoms is  $sp^2$ -hybridized the situation is ambiguous. Atom N1 in MBADNP is in such an ambiguous case.

Among the possible model compounds for N1 there are many molecules for which not one hybridization but a mixture would be the best description. An example is shown in Figure 5.19, where above and below the nitrogen atom of the secondary amine group unmodeled density was observed when no LONE instruction was applied to the atom of interest. But there is also too much density on one side and the amount of unmodeled density is less on the same side. So neither a description as  $sp^2$  via LONE 6 nor as  $sp^3$  via LONE 1 alone would be appropriate.





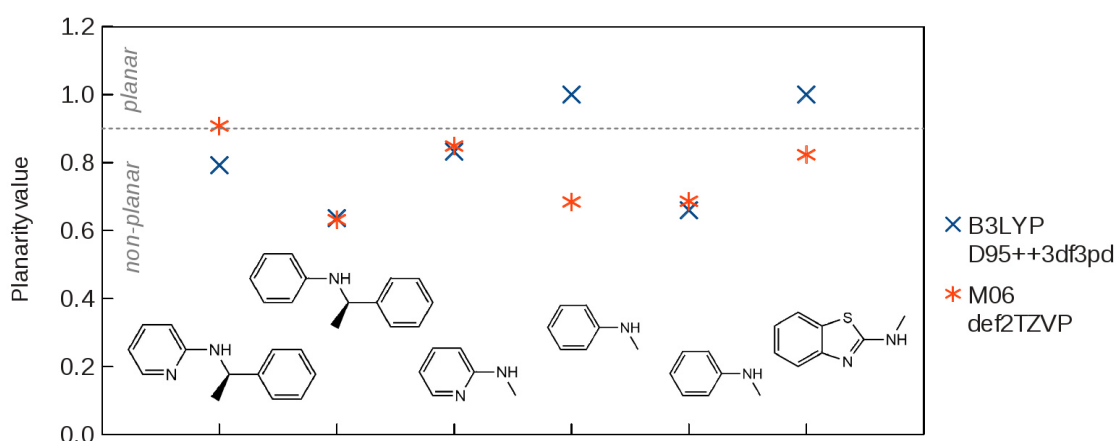
**Figure 5.19:** Residual density at the secondary nitrogen atom of *N*-methylpyridin-2-amine connected to one  $sp^2$  and one  $sp^3$  hybridized carbon atom. Density levels are shown at  $\pm 0.30 \text{ e}/\text{\AA}^3$ .

A look at several secondary amines which were possible model compounds for N1 in MBADNP revealed that the angle sum deviates from 360 to different small degrees, as shown in Table 7.10 in Appendix C and the results are summarized in Figure 5.20. The investigation of planarity for the model compounds for secondary amines, in which the nitrogen connects an  $sp^3$ -hybridized carbon atom to an  $sp^3$ -hybridized one, lead to the surprising observation that the preferred geometry depended on method and basis set used. But for *N*-methylaniline two geometry optimizations with the same basis set and theory<sup>7</sup> yielded one planar and one non-planar geometry at the nitrogen atom. This suggests that both positions are a local minimum. The non-planar molecule is of lower energy by around 0.5 kJ/mol. This accidental observation is interesting and good to know for further development.

In this transferability study results from M06/def2TZVP were used, where the two identical molecules were optimized to the same minimum of non-planarity, leaving only *R-N*-(1-phenylethyl)pyridin-2-amine as model compound for a planar nitrogen atom in this bonding environment.

Tertiary amines have a better defined geometry in crystal structures, but ambiguous cases concerning the threshold for planarity are also possible.

In the MBADNP model refined against neutron data N1 is planar. Due to the nature of neutron diffraction, which does not rely on the ED but the neutron scattering length, the position of the hydrogen atom is more reliable than from X-ray diffraction. This knowledge



**Figure 5.20:** Planarity value of selected secondary amines after geometry optimization with different basis sets.

<sup>7</sup>This twofold calculation was originally caused by two entries in the database of which one was listed under the trivial name methyl-phenyl-amine instead of *N*-methylaniline.

of hybridization of nitrogen atom N1 in MBADNP was used in this transferability study, since the major aim of the study is not the investigation of transferability details but to get an idea of transferability of the BEDE and LONE parameters in general.

**Directionality of bonds** The occurrence of planarity of a nitrogen atom is correlated to the detection of a meso bond to/from the nitrogen atom. This leads to another special case concerning invariom names. The bond of order 1.5 usually goes to the  $sp^2$  hybridized atom of an aromatic ring which is signaled by an @. The @ replaces the bond-order character, and thereby this information is neglected in the nitrogen-atom invariom name. But from the atom bonded to the nitrogen by the bond of order 1.5, which is part of the aromatic system, the name includes the correct bond order of 1.5. In our case this is relevant for atom C(9). Its invariom name of  $6-C\# 6n[\# 6c]1.5n[1c1h]\# 6c[\# 6c1n]$  includes the bond order information. In combination with planarity this is the other reason why in our case the unexpectedly large model compound *R-N*-(1-phenylethyl)pyridin-2-amine is chosen as model compound for BEDE and LONE parameters involving atom C9 or N1.

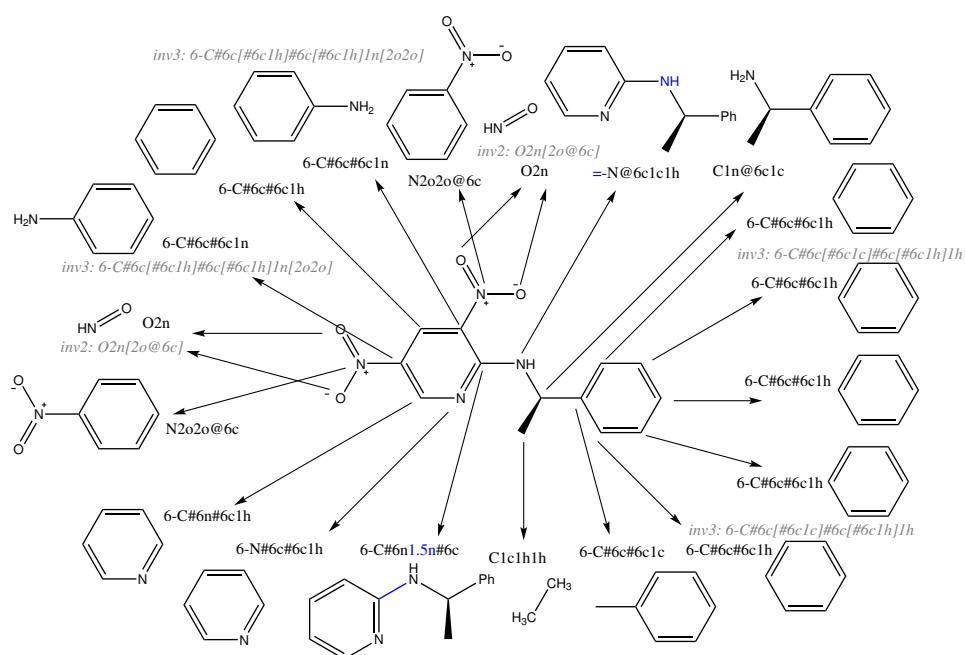
Changes caused by planarity of nitrogen or the bond order of 1.5 are marked blue in Figures 5.17 and 5.21, which show the model compounds from which the parameters were actually transferred from in this study. By knowing N1 to be planar the transfer of LONE 6 parameters became possible. Even if in the charge-density structure the hydrogen position is not that fixed, LONE 6 is not affected. Only LONE 1 is problematic in planar cases because it uses the sum of all three bond vectors, while LONE 6 only uses two bond vectors for a cross product preferably to non-hydrogen atoms. Therefore LONE 6 parameters were also transferred for N1 in MBADNP.

#### 5.3.4.6 Transfer of LONE parameters via invarioms – actual model compounds

For consistency LONE parameters were also transferred according to the invariom names without next-nearest neighbors in a first experiment (abbreviated with **inv1**) In principle LONE parameters could be transferred according to the standard invariom names (shown in Figure 5.15).

Especially for oxygen atoms with double bonds the transferred ED improves significantly upon considering next-nearest neighbors as discussed in earlier sections of this thesis. For this reason the model compound for LONE parameters of O1 to O4 were transferred from nitrobenzene in a second experiment (**inv2**). The corresponding longer invariom names are given in Figure 5.21 in grey. Interestingly only LONE 9 is required to model the lone pair ED in nitrobenzene while in aminoxide an additional LONE 7 command is necessary for  $\pi$ -bond modeling. This big difference suggests a better fit for parameters transferred from nitrobenzene.

Since LONE parameters of aromatic systems are quite sensitive to the substitution pattern and substituent, as shown in Section 5.3.2, assignment based on the standard invariom names could be beneficial. But an atom which seems especially problematic is the aromatic carbon in *ipso* position to the nitro groups. The LONE 6 amplitudes for this carbon atom are considerably different. In aniline the amplitude *A* is  $-0.163$  while in nitrobenzene  $A = 0.122$ . The difference is in agreement with the different mesomeric effects of amine



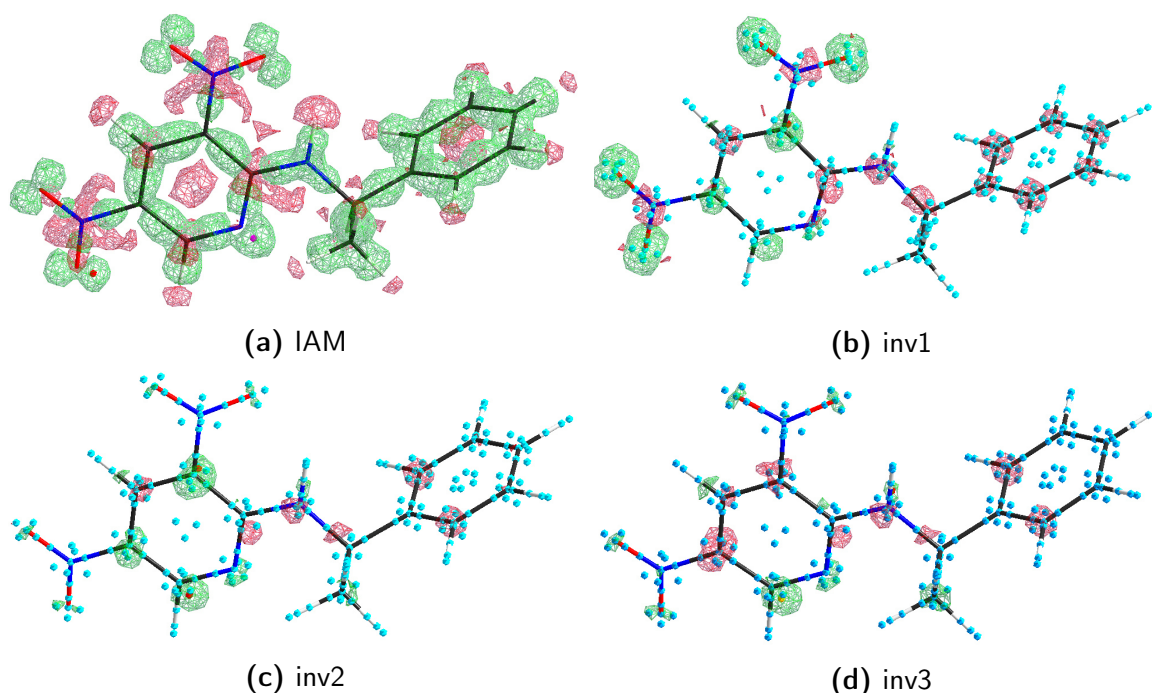
**Figure 5.21:** Invariom names without next-neighbor atoms. Assignments used for the transferability study with MBADNP. Model compounds for LONE parameters and for BEDE parameters of bonds from hydrogen atoms.

(+M) and nitro groups (-M). Unfortunately this issue would not be solved by the common invariom names including NNN for meso bonds since the bond to nitrogen is a single bond. Therefore in a third experiment (**inv3**) LONE parameters for carbon atoms C10 and C12 were transferred from nitrobenzene. In the same experiment C2 and C6 LONE parameters were taken from toluene according to the unabbreviated invariom names, because the results of the first and second experiment showed that these atoms could be modeled better.

In summary all invariom names and model compounds involved in the transfer of LONE parameters are displayed in Figure 5.21. The BEDE parameters for bonds involving a hydrogen atom include the invariom name of the parent atom and therefore are always taken from the same model compound as the LONE parameters.

### 5.3.4.7 Results

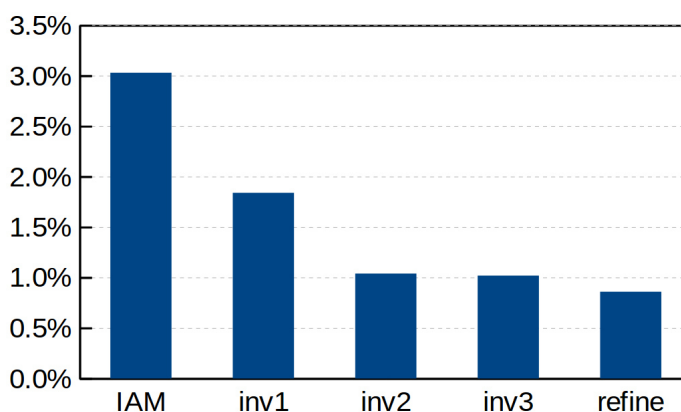
The residual density plots for the different transfer experiments after scale factor refinement against the simulated XRD data are shown in Figure 5.22 and R1(all) in Figure 5.23. As expected the first experiment with LONE parameters from aminoxide can be significantly improved by applying the parameters from aminobenzene. Then not only R1(all) drops from 1.84 % to 1.04 %, but also the residual density decreases, especially around the oxygen atoms. Describing C10 and C12 by LONE parameters from nitrobenzene instead of aniline yielded residual density of opposing sign, but led to just a slight drop of R1 to 1.02 %. Transferring LONE parameters for C2 and C6 from toluene instead of benzene yielded almost no change in residual density, which is due to very similar values transferred. Compared to the residual density of the IAM all transferred BEDE and LONE parameters considerably decreased the residual density and also R1(all), thus led to the conclusion that



**Figure 5.22:** Residual density maps at  $\pm 0.05 \text{ e}/\text{Å}^3$  for MBADNP against simulated data modeled by IAM and the transferred BEDE and LONE parameters. The transfer for inv2 differed from inv1 by nitrobenzene as model compound for O2n and for inv3 also the carbon atoms in *ipso* position to the nitro group were transferred from nitrobenzene.

transfer of BODD parameters is possible and a great improvement of the structural model.

Further refinement of the transferred parameters leads to minor improvements only (see Figure 5.23). Only parameters belonging to BEDE instructions modeling the density between non-hydrogen atoms were refined. The total number of parameters refined was 26 instead of one for the other refinements in the comparison. Considering the number of parameters saved, the BODD models with parameters transferred from invariom model compounds performed well. Hence, transferability from small to large molecules has been established for refinements against simulated data.



**Figure 5.23:** R1(all) for different models of MBADNP after scale factor refinement with SHELXL against simulated data. Only for the last bar BEDE parameters between non-hydrogen atoms were refined.

### 5.3.5 Transfer to and refinement of a structure against experimental data

#### 5.3.5.1 Data sets

MBADNP and a related compound were investigated in 2002 via a combined charge density and neutron study by Cole *et al.*<sup>[248]</sup> The crystallographic data for both is summarized in Table 7.11 in Appendix C. MBADNP crystallizes in the monoclinic space group  $P2_1$  with one molecule in the asymmetric unit. The data was collected at 20 K under helium cooling on a diffractometer in Durham and the neutron beamline D9 at ILL. The X-ray data has a resolution of 0.55 Å, which is not high by charge-density standards but above average. Neutron data allowed the refinement of hydrogen-atom positions and thereby identifying the planar geometry at the secondary amine. Space group and molecule are chiral, so the enantiomer depicted in the following models was chosen based upon the *Flack* parameter of the IAM refinement against X-ray, although with  $x = 0.092(575)$  for the classical *Flack* and  $x = -0.058(309)$  from *Parsons'* quotients the standard uncertainties are too high for absolute structure determination from anomalous dispersion.

#### 5.3.5.2 Procedure

The same BEDE and LONE parameters as in the transferability study II, were also transferred to the model refined against the experimental XRD data. Bond distances to hydrogen atoms were elongated to theoretical values, while their geometrical placement was geometrically ideal for all but the hydrogen atom of the amine. Hydrogen displacement parameters were isotropic and constrained to be 1.2 and 1.5 times their parent atoms  $U_{eq}$ . Coordinates and ADP of all non-hydrogen atoms and the overall scale factor were refined similar to a standard refinement.

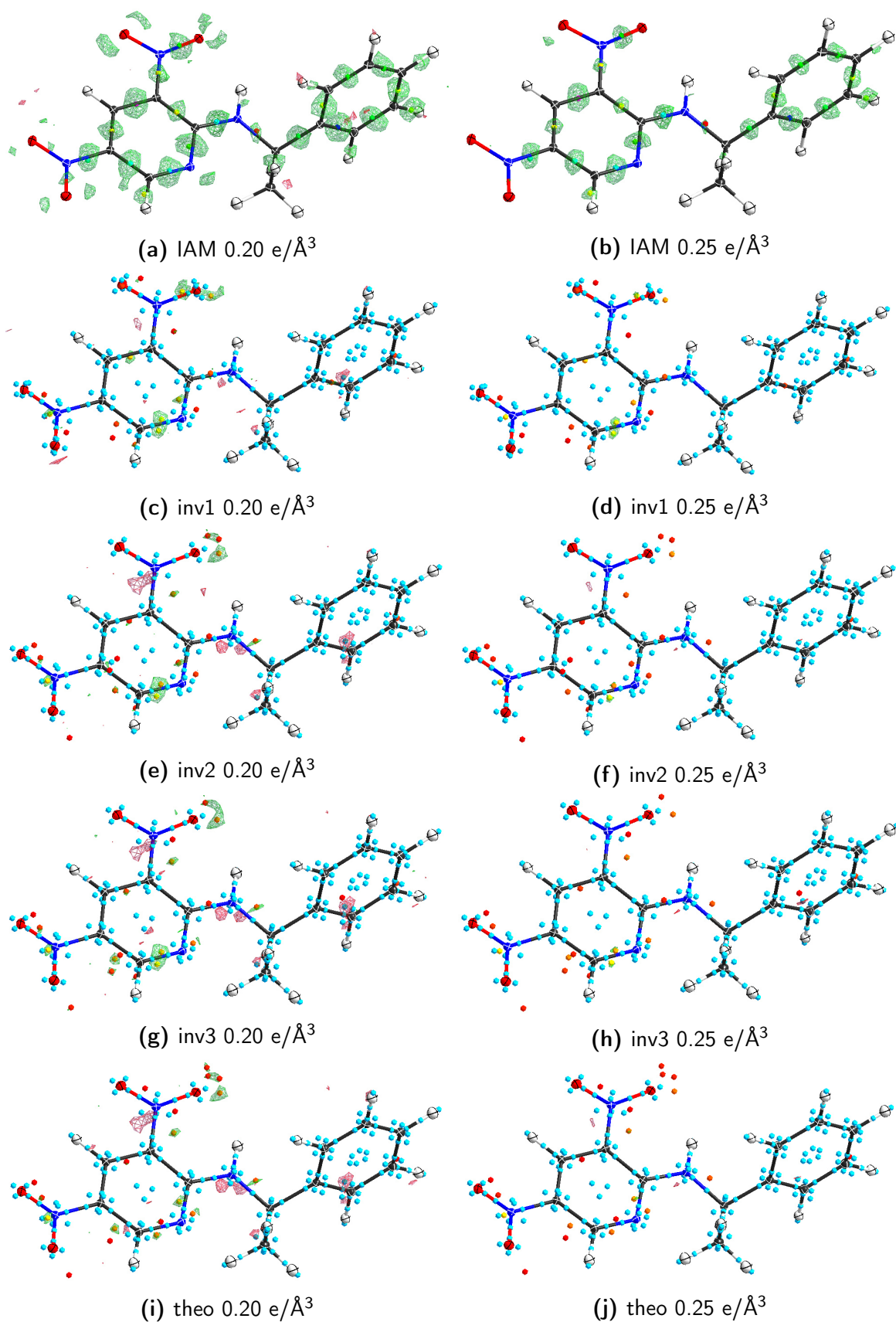
An equivalent refinement without BODD parameters, hence of a normal IAM, was performed simultaneously with the same hydrogen atom treatment as a reference. Every refinement was carried out until convergence, including adjustments of the weighting scheme.

#### 5.3.5.3 Results

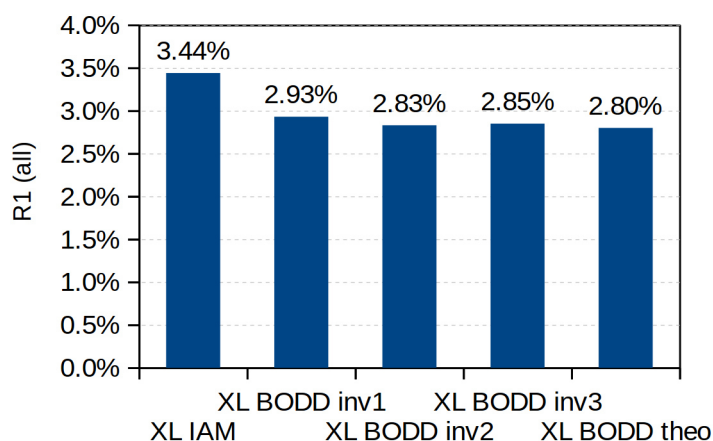
Surprisingly the figures of merit worsened upon introduction of BODD parameters in the model as can be seen in Figure 7.11 in Appendix C. The corresponding residual density maps displayed in Figure 7.12 (Appendix C) clearly show that the residual ED increased. However, upon refinement of one new parameter that scaled the transferred aspherical ED the residual density maps (Figure 5.24) as well as the figures of merit (Figure 5.25) improved.

Comparing the BODD results to equivalent XD refinements<sup>8</sup> as shown in Figure 5.26, the new model performs very well. Actually, the figure of merit is lower than for the multipole refinement, although distances to hydrogen atoms,  $U_{iso}$  constraints for hydrogen atoms, weighting scheme and the sigma threshold were transferred from SHELXL. The models in

<sup>8</sup>For all refinements systematically absent or negative reflections were not considered, the same two reflections omitted, the same first parameter of weighting scheme and the same hydrogen atom treatment (riding hydrogen model and  $U$  relative to non-hydrogen atom) were applied.



**Figure 5.24:** Comparison of residual density maps at  $\pm 0.20 \text{ e}/\text{\AA}^3$  and  $\pm 0.25 \text{ e}/\text{\AA}^3$  for different models refined against X-ray data of MBADNP. (Inverted, uncut and omitted the worst 2 reflections of model inv3 in all maps. With a global variable as multiplier for all BODD A values.)

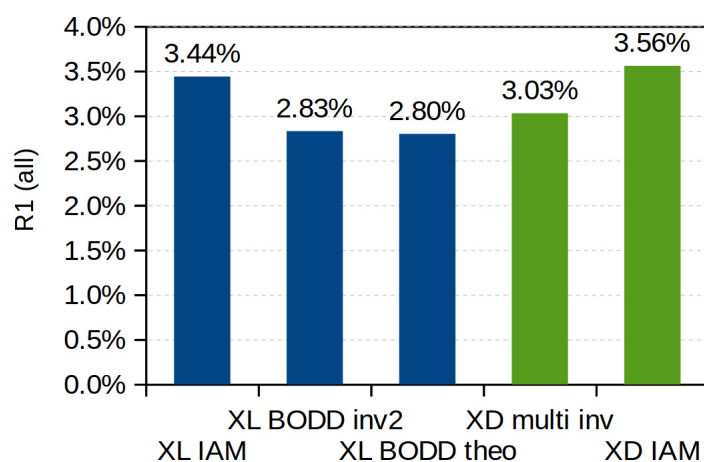


**Figure 5.25:** R1(all) for differently transferred BODD parameters for the refinement of MBADNP with SHELXL against experimental XRD data.

XD were also refined against  $F^2$  with floating origin restraints, but still the discrepancy between the SHELXL and XD residuals remains around 0.2 %. However, both asphericity models yield a similar improvement compared to IAM in both programs by 0.6 %.

The number of parameters refined with SHELXL was always higher by one due to the rotational freedom of the otherwise fixed methyl group. But this refined position was transferred to XD, so the additional parameter should not have been of significance. This left only the one additional parameter in the aspherical models in SHELXL that scaled the complete transferred ED (theory2experiment scale factor) for discussion. The scale factor scales the transferred amplitudes down by around 40 %. This extent of scaling was unexpected.

The *Flack* parameter and more importantly its uncertainty (Table 7.12, Appendix C) determined via the classical method decreased upon modeling of bond and lone pair ED as has been known for invariom refinements with the multipole model.<sup>[71]</sup> The improvement is small, since anomalous differences are small for molybdenum radiation. The *Flack* parameter determined from *Parsons'* quotients<sup>[89]</sup> usually yields values for the IAM model similar to those from the classical method after modeling aspherical ED. The systematic errors of omitting asphericity no longer influence the result, since the quotients are calculated from structure factors that are all derived from a similarly incomplete model. Therefore, the improvement of the classical *Flack* parameter only agrees with expectations.



**Figure 5.26:** R1(all) for different models of MBADNP, refined with SHELXL and XD comparing BODD with the multipole model.



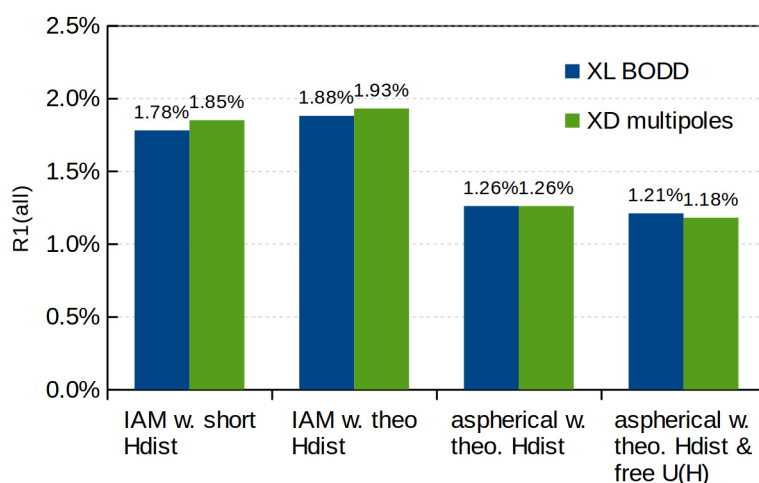
5.3.5.4 A 2<sup>nd</sup> experimental data set: 2-dimethylsulfuranylidene-indan (ylid)

In order to investigate the compound specificity of the unexpected theory2experiment scale factor, a second crystal structure of an organic molecule, ylid<sup>[250]</sup>, with good X-ray data was modeled by BODD. General crystallographic information on this data set can be found in Table 7.13 in Appendix C. In contrast to the data of MBADNP the new data set was measured at 100 K and has a resolution of  $d_{max}=0.5\text{\AA}$ .

Since mainly the scale factor for transition from theory to experiment was of interest, the choice of model compounds was unimportant. Therefore, BEDE and LONE parameters were all directly transferred from the complete 2-dimethylsulfuranylidene-indan (ylid) molecule treated in the same way as the other model compounds before. Hence, the refinement of the BODD model with the theoretical parameters against experimental data is most comparable to the 'XL BODD theo' model for MBADNP.

The data set for ylid showed extinction, so in the beginning the models were refined with and without extinction correction. Since the difference between comparisons of corrected and uncorrected models were negligible, all results displayed in Figure 5.27 include refined isotropic extinction models.

From the results it could be deduced that modeling of bonding and lone pair density improved R1 independent of asphericity model and refinement program. Figure 5.27 also includes a comparison of different treatments of hydrogen atoms. In all models the riding hydrogen model was applied, and the hydrogen atoms were placed in geometrically ideal orientations. An IAM refinement with the standard shortened bond lengths to hydrogen atoms was performed, but R1 values for an IAM in which the hydrogen atoms were placed at the theoretical bond lengths are also shown. Theoretical distances are preferable if parameters are transferred from compounds with a geometry obtained via quantum chemistry calculations. This is why hydrogen atoms in the aspherical models were placed at distances deduced from QM, corresponding to positions of the nuclei rather than the center of the ED. The EDD of the hydrogen atoms can be described by the asphericity model and hence bond lengths do not need to be artificially shortened any more, to model the shifted



**Figure 5.27:** R1(all) for different ylid models refined with SHELXL and XD comparing the BODD model with the multipole model and different kinds of hydrogen atom treatment.



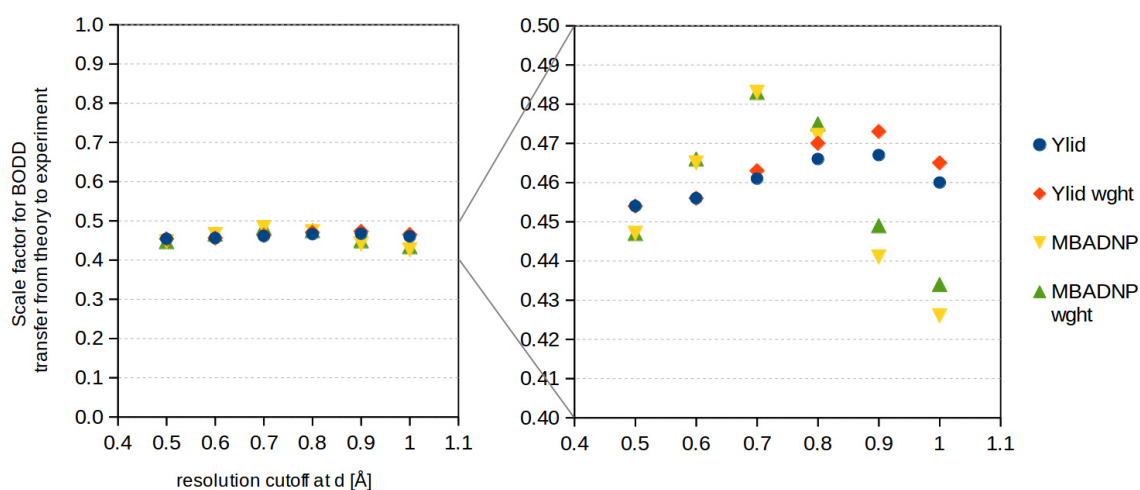
center of ED from the nucleic position. Further investigation concerning hydrogen atom positioning is given in Section 5.3.6. The lowest residuals were obtained, when isotropic displacement parameters for hydrogen atoms were refined. In contrast to the standard ratio of  $U_{\text{iso}}(\text{H})/U_{\text{eq}}(\text{parent atom})$  of 1.2 and 1.5, refinement of the displacement parameters allows adjustment to the temperature dependence of the ratio.<sup>[102]</sup> Although it should be noted that convergence of the additional displacement parameters was slow. Overall the figures of merit for ylid show that the BEDE and LONE model fitted the ED equally well as the multipole model, if a theory2experiment scale factor was refined.

### Investigation of the theory2experiment scale factor

The theory2experiment scale factor for this second compound is around 0.45. Although it is slightly higher than for MBADNP it is still far from one. Since the information about the bonding ED is mainly contained in the low-order reflections, their resolution dependency was investigated. The resulting plot of scale factors versus resolution cutoff is displayed in Figure 5.28.

Both compounds reach a maximum scale factor between 0.6 and 1.0 Å. Its value, however, is always between 0.40 and 0.50. The differences between the datasets are the temperature during data collection, the extinction correction for ylid and that ylid contains sulfur, a heavier element. Without the refinement of extinction the scale factor drops to around 0.42 for ylid at a resolution of 0.50 Å.

Free refinement of displacement parameters against the simulated data did change the amplitude only marginally, which excludes errors in the convolution of the artificial displacement parameter with the simulated data. Neither was the scale factor dependent on the  $U$ -value of the simulated  $hkl$ , nor does refining a separate scale factor for all hydrogen related BEDE instructions lead to a scale factor closer to one. Different weighting schemes in refinement against the experimental data influenced the scale factor, but not significantly more than the deviations caused by resolution cutoffs. So the reason for the positive effect of the scale factor remains elusive.



**Figure 5.28:** Scale factor for transfer of BODD from theory to experiment versus resolution of the datasets of MBADNP and ylid.

### 5.3.6 Refining BEDE and LONE parameters against experimental data

Instead of BODD modeling by parameters transferred from refinements against simulated data, the parameters could be refined directly against the experimental data. Therefore, such a free refinement with the lowest possible number of parameters was attempted.

The bonding ED for all carbon-carbon bonds was modeled with one parameter ( $A$ ) and an additional amplitude per bond to and between other elements, with one parameter for each direction. MBADNP contains atoms of the elements carbon, nitrogen, oxygen and hydrogen. The bonds to hydrogen were not considered in the first model, neither were negative values for  $r$ .  $B1$  was set to 0.1 for all BEDE and LONE instructions and  $B2$  to 0.7<sup>9</sup>. For this model the highest residual density is found on bonds between carbon and nitrogen. Hence, those bonds were given separate free variables according to the atoms bonded to the nitrogen atom. This resulted in thirteen parameters for the BODD, since the lone pair at the nitrogen in the aromatic system N2 was modeled by LONE 4, and since the oxygen atoms were also given one parameter for LONE 9. The distances were adjusted manually between 0.45 and 0.55 Å. Refinement of those parameters in addition to the usual atomic coordinates and displacement parameters was listed as 'refine' in Table 5.1 and is shown in Figure 5.29. Refining the  $B1$  value with an additional free variable lowered the  $R$  value by a small degree and decreased the residual density (see Figure 5.30).

#### Hydrogen-atom treatment

While the atoms in the two refinements so far were placed at standard crystallographic positions, it is common in charge density studies to elongate the bond lengths of hydrogen atoms to theoretical ones. The displacement of the main ED (observed by XRD) from this theoretical position of the hydrogen nucleus can then be modeled more accurately as bonding ED.

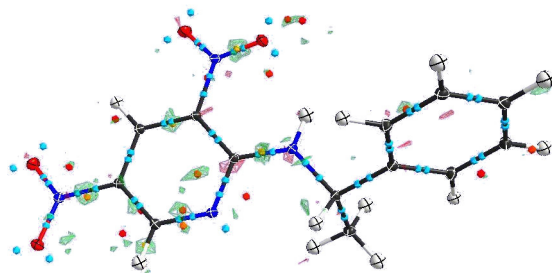
Accordingly, a mere elongation of the bonds to hydrogen atoms led to a worse fit of the XRD data, but if three additional parameters for BEDE instructions involving hydrogen were added, the model fitted the data equally well (Table 5.1 as well as Figures 5.31 and 5.32). The bonding ED was placed 0.3 Å along the bond to the hydrogen atoms and in the opposing direction from hydrogen to its neighbor atom at  $r = 0.2$ . Additionally, a negative *Gaussian* function was added to the model next to the hydrogen atoms facing away from the bond at a distance of 0.15 Å.

Refining more parameters, starting from the invariom values, led to better  $R$  values (down to an  $R1(\text{all})$  of 2.60 %) but after including 24 new BEDE and LONE parameters the density fit had not improved considerably compared to the starting model. On the contrary looking at the difference between highest peak and deepest hole in the residual density revealed an increase of the gap. Hence, free refinement of BODD parameters against experimental data should be performed very carefully, if necessary at all.

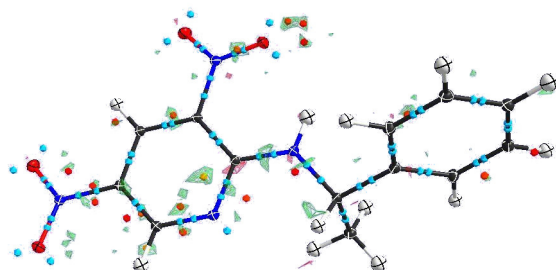
For investigative charge-density studies by crystallography, the multipole model will still be preferred. Hence, there is no need to compete with the well-established procedures of charge-density research, concerning the refinement against experimental diffraction data.

---

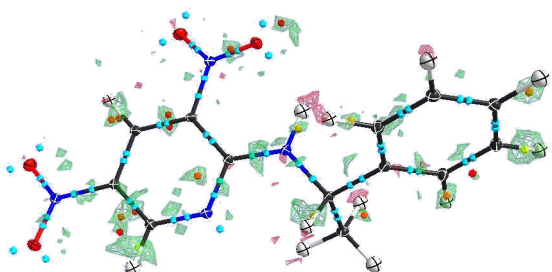
<sup>9</sup>0.8 yielded almost the same results, so the exact value is not very important.



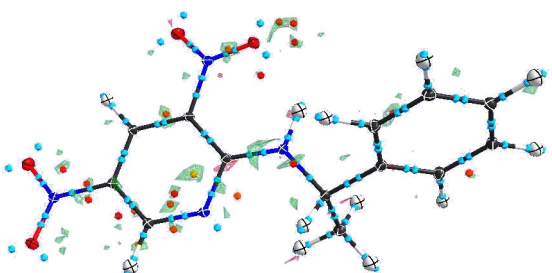
**Figure 5.29:** Residual density at  $0.15 \text{ e}/\text{\AA}^3$  of MBADNP after refining 13 BODD parameters for the bonds between non-hydrogen atoms while hydrogen atoms are placed at standard crystallographic positions for X-ray data and  $B1 = 0.1$  for all BEDE and LONE instructions.



**Figure 5.30:** Residual density at  $0.15 \text{ e}/\text{\AA}^3$  of MBADNP after refining 14 BODD parameters for the bonds between non-hydrogen atoms while hydrogen atoms are placed at standard crystallographic positions for X-ray data.



**Figure 5.31:** Residual density at  $0.15 \text{ e}/\text{\AA}^3$  of MBADNP after refining 14 BODD parameters for the bonds between non-hydrogen atoms while hydrogen atoms are placed at theoretical bond lengths.



**Figure 5.32:** Residual density at  $0.15 \text{ e}/\text{\AA}^3$  of MBADNP after refining 17 BODD parameters for all bonds, for those to hydrogen atoms even at  $r = -0.15$  from hydrogen atoms, while hydrogen atoms are placed at theoretical bond lengths.

**Table 5.1:** Refinement results for different hydrogen atom treatments for refining BEDE and LONE parameters against experimental X-ray data of MBADNP.<sup>[248]</sup>

	R1 [%]	wR2 [%]	max	min	Goof	R1(all) [%]
refine fixed B1 (Figure 5.29)	2.44	8.06	0.26	-0.25	1.120	2.96
refine (Figure 5.30)	2.36	7.91	0.25	-0.25	1.120	2.90
refine hdist (Figure 5.31)	2.65	8.82	0.39	-0.26	1.106	3.16
refine hdist BODD H (Figure 5.32)	2.35	7.87	0.25	-0.24	1.124	2.89

### 5.3.7 Comparison of bond-lengths uncertainties and displacement

The improvement of residual factors alone is not a scientific achievement. If the new model is accurate, results from single-crystal structure determination are expected to improve. The information of interest acquired by a single-crystal XRD experiment is the three dimensional atomic arrangement. Atomic coordinates allow derivation of connectivity and bond lengths. By modeling the bonding and lone-pair ED the parameters describing the anisotropic displacement are expected to improve, because they do not compensate the missing description of the aspherical ED by artificially modeling more movement than actually present. Although ADP are usually not of primary interest, they influence the precision of bond-length determination.

Tables 7.16 and 7.17 in Appendix C show that upon inclusion of asphericity in general the atomic displacement parameters,  $U_{\text{iso}}$  and  $U_{\text{equiv}}$ , were smaller. Evaluation of bond-lengths uncertainties of MBADNP and ylid (see Tables 7.14 and 7.15 in Appendix) showed a higher precision for those models with transferred BODD and invariom multipole parameters. The average uncertainty of bond lengths between non-hydrogen atoms in ylid was  $6.6 \cdot 10^{-3} \text{ \AA}$  in the IAM,  $5.0 \cdot 10^{-3} \text{ \AA}$  for the freely refined BODD parameters and  $3.8 \cdot 10^{-3} \text{ \AA}$  in the model BODD parameters transferred from refinement against simulated data.

Interestingly, for the free BODD refinement most  $U_{\text{equiv}}$  values of non hydrogen atoms did not decrease at all, while they did in case of the free BODD MBADNP refinement. This could have its origin in the globally refined  $B2$  (to a value of 0.18) in the case of ylid whereas for MBADNP just one  $B1$  parameter was refined for all BEDE and LONE instructions. This is a first indication that refining  $B2$  may lead to physically less meaningful results. Not refining it but setting it to 0.7 instead, as in the case of MBADNP, led to an increased R1 of 1.43 % instead of 1.37 % but to even a smaller average displacement than that of the model with BODD parameters transferred from refinement against simulated data. Therefore, it seems advisable to not refine  $B2$ , but simply set it to 0.7 or a similar number.

In general the comparison suggests that it is better to use BODD parameters transferred from refinement against simulated data than to refine them against the experimental data, even if the resolution is above average and would in principle allow the refinement of additional parameters.

### 5.3.8 Comparison to a model refined against neutron data

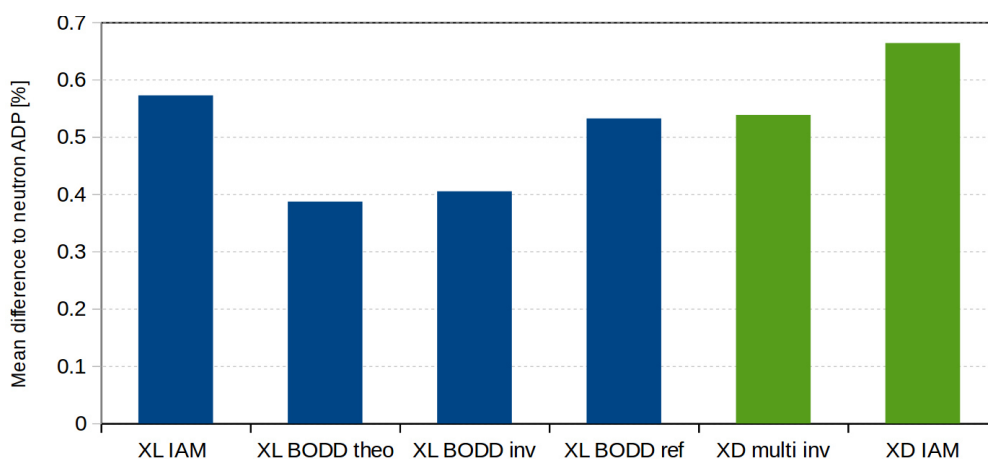
#### 5.3.8.1 Design of the study

In order to show, in addition to the higher precision, that the new ADPs are more accurate they are compared to ADPs refined against neutron data. Due to the scattering of neutrons from atomic cores, ADP refined against neutron data are deconvoluted from the asphericity of the ED.

The comparison includes the IAM models refined in SHELXL (XL IAM) and XD (XD IAM), the invariom multipole model (XD multi inv) and three different BODD models. The BODD model with parameters from invariom transfer number two (XL BODD inv2), with parameters from the refinement against simulated data of MBADNP (XL BODD theo) and those refined manually against the real data for non hydrogen atoms<sup>10</sup> (XL BODD ref). None of the models was refined with RIGU or any other restraint on the ADPs. Just the distances to hydrogen atoms were set to the theoretical bond lengths for the 'XD invariom multipole' model, the 'XL BODD inv' and 'XD BODD theo' model.

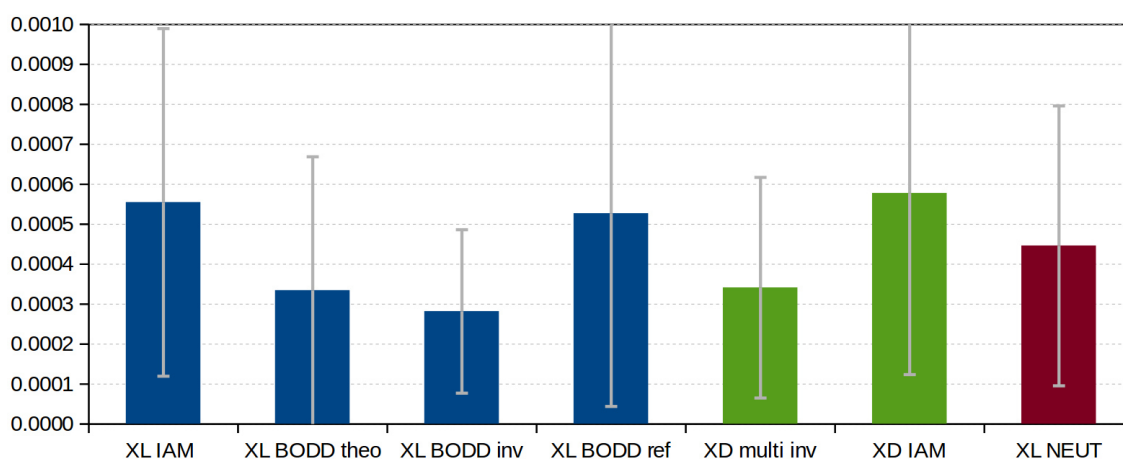
#### 5.3.8.2 Results

**Anisotropic displacement parameters** The results comparing the ADP for different models of MBADNP against the ones refined against the neutron data are shown in Figure 5.33. The mean ADP difference to the neutron model is smallest for the two BODD models with parameters transferred from theory either directly ('XL BODD theo') or from the model compounds assigned via invarioms (option two, 'XD BODD inv'). Interestingly, the application of fewer BODD instructions that are less specific but refined freely against the experimental data results in ADPs which agree worse with the neutron model, but similarly well as those from the invariom multipole refinement with XD. Regardless of the differences between the models for the valence density, the two IAM models deviate most from the model refined against neutron data.



**Figure 5.33:** Mean ADP difference between selected models and the one refined against neutron data for MBADNP. (Calculation was kindly automated by Jens Lübber.)

<sup>10</sup>Refinement of one global free *B*1 and hydrogen atoms at standard crystallographic positions



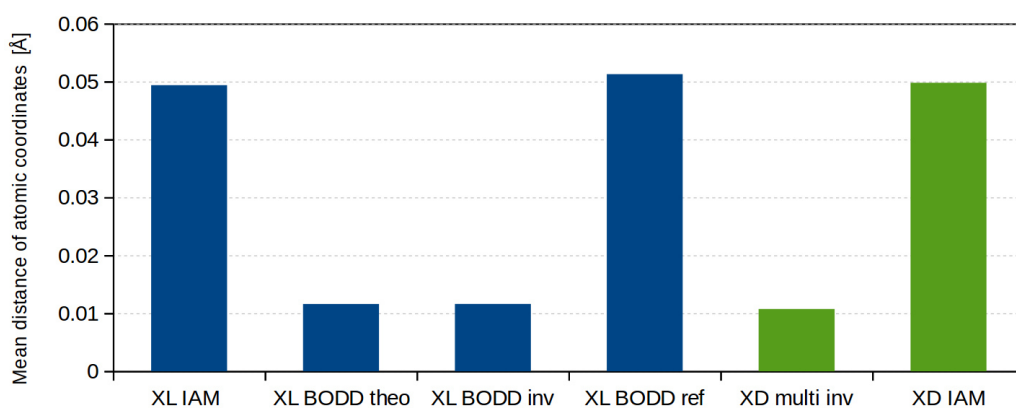
**Figure 5.34:** Mean *Hirshfeld* test results for different models incl. the model refined against neutron data for MBADNP derived from PLATON calculations for each bond. Error bars represent the variance for the different bonds in each model.

**Hirshfeld test** A look at the *Hirshfeld* test for all of the models including the one refined against neutron data (Figure 5.34) shows that it is fulfilled best for those asphericity models with parameters transferred from theory. The most significant improvement is observed for bonds within the nitro group (see Figure 7.13 in Appendix C for atom specific results), among which some would not have passed the five sigma test level in the IAM<sup>11</sup>.

The fact that the freely refined BODD parameters resulted in a worse *Hirshfeld* test result can originate in the placement of the BODD *Gaussian* functions farther away from the atoms than in refinements against simulated data. The residual density map upon which the decision of BODD placement was based was already biased by ADPs that were convoluted with the valence density. This way BODD mainly modeled the density not yet described by the contaminated ADPs, whereas the transferred theoretical BODD parameters were placed closer to the atoms due to the deconvolution of valence density from atomic displacement. Thus a free refinement of BODD led to less model improvement than the transfer of BODD parameters from a database or theory. Unfortunately, the model refined against neutron data has an average *Hirshfeld*-test result that is not quite as good as the one from the aspherical models, so that they cannot be relied upon as high-quality reference results.

**Atomic coordinates** Figure 5.35 displays the agreement of atomic coordinates for different X-ray models with those refined against neutron data. Though the comparison excludes hydrogen atoms asphericity models with transferred parameters and theoretical bond lengths agree best with the atomic coordinates of the neutron model. This supports the initial conclusion obtained from the standard uncertainties of the bond lengths (see Chapter 5.3.7), which are on average slightly lower for the invariom BODD model than for the BODD model refined against experimental data.

<sup>11</sup>This failure of the *Hirshfeld* test was observed independent of the application of RIGU as restraint in SHELXL refinement



**Figure 5.35:** Average distance between atomic coordinates of different models and the one refined against neutron data for MBADNP. (Calculation was kindly automated by Jens Lübber.)

**Conclusion** In summary this means that a refinement of BODD parameters may lead to a lower R value but this does not necessarily coincide with a better model. In the case of MBADNP models with BODD or multipole parameters from theory led to a more accurate model than a model in which BODD parameters were refined freely.

### 5.3.9 Statistical study

After successful BODD modeling of MBADNP and ylid by theoretical parameters there were in principle two possible options to get closer to the overall goal of adding BODD to a model via one click: transfer or calculation of BODD parameters.

For the first option many more model compounds, most likely all the model compounds of the invariom database, would need to be parametrized so that parameters could be transferred via a big look-up table based on a classification that would still have to be decided upon. For the second option dependencies of the BEDE and LONE parameters on other structural information could be analyzed to evaluate the possibility of parameter calculation via some simple equations instead of a look-up table. This evaluation of parameter dependencies requires several parametrized compounds as well. A first attempt to find a dependency for bonds between carbon atoms will be presented next, to evaluate feasibility of this approach.

The parameters already refined for 34 model compounds against simulated data so far (a list is given in Appendix C, Table 7.18) were investigated for relations between different bonding parameters. Even if the first option will be pursued further in the end, the results of such an investigation could supply starting values for the parametrization of all the compound in the invariom database.

#### 5.3.9.1 First exploratory data analysis

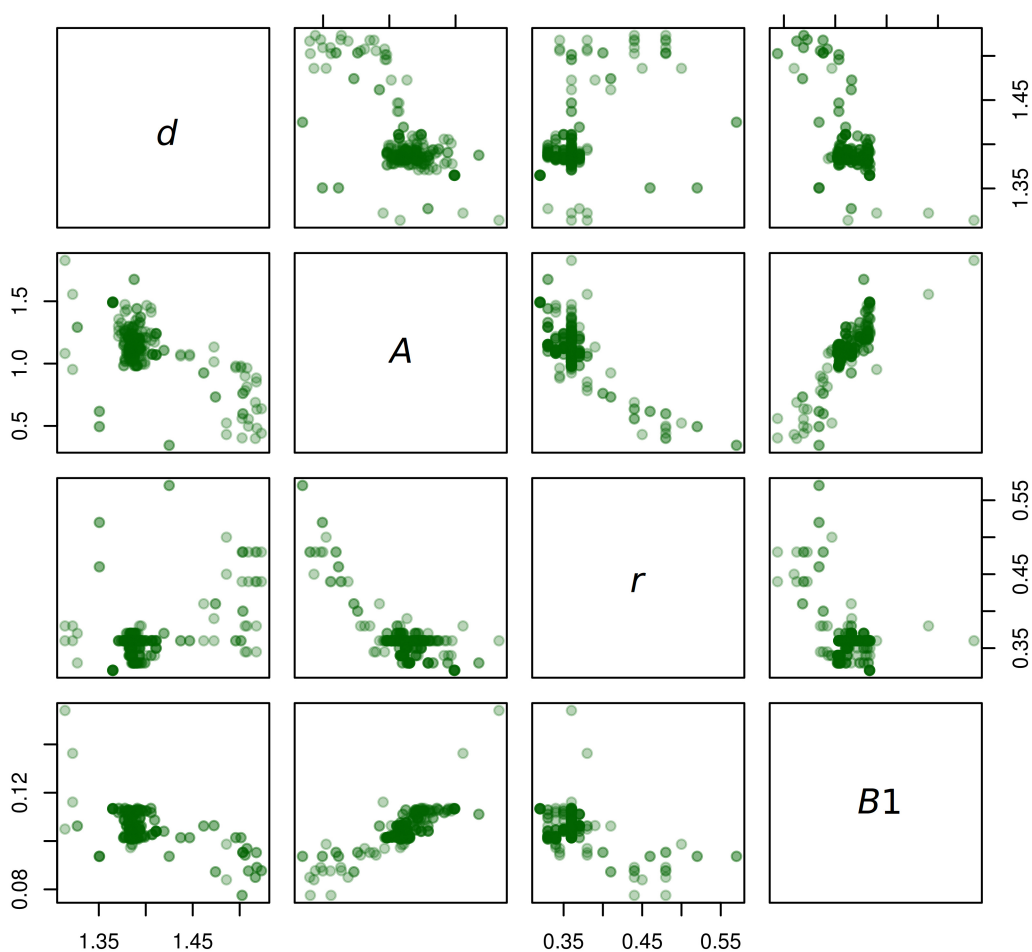
A correlation between  $A$  and the bond length of bonds between carbon atoms would be plausible, which therefore was the focus of the investigation. The hypothesis is that the shorter the bond, the higher the ED and thereby the higher the  $A$  value. This is based

on Pauling's correlation of bond length to bond strength,<sup>[251]</sup> which is underlying many relations of bond strength to bond length<sup>[252]</sup> as for example the bond valence theory.

A quick first look at a plot of  $A$  versus the bond distance,  $d$ , did not suggest a simple correlation. But since  $A$  is refined dependent on  $r$  of the BEDE instruction, the relation could be more complex. A matrix of scatter plots for the BEDE parameters of bonds between carbon atoms and positive  $r$  is shown in Figure 5.36. It reveals a more pronounced correlation between  $r$  and  $A$  and even a possibly linear relation between  $A$  and  $B1$ . The first observation is interesting, because it could be the reason why  $A$  versus the bond length shows no simple correlation. The relation between  $A$  and  $r$  seems hyperbolic and linear regression of  $1/A$  against  $r$  (Appendix C, Figure 7.14) fits quite well. Which is the reason for plotting  $1/A$  against the bond length and  $r$  in Figure 5.37.

It highlights the domination of  $r$  on  $A$  compared to the bond distance. Of course one could use the mean  $r$  in the fitted plane for obtaining an equation to calculate  $A$  from the bond distance, but the result (as shown in Figure 7.15 in Appendix C) was not convincing.

A better result may be obtained after eliminating the dependence on  $r$  by generating a new data set for which all model compounds investigated were refined with  $r = 0.36 \text{ \AA}$  for



**Figure 5.36:** Scatter plot matrix of BEDE parameters from bonds between carbon atoms and positive  $r$  collected from 34 model compounds, where  $d$  represents the bond length.



bonds between carbon atoms. Alternatively the ratio of  $r$  to the bond length could have been fixed, its mean is 0.2595 with a standard deviation 0.023 and variance of 0.0005.

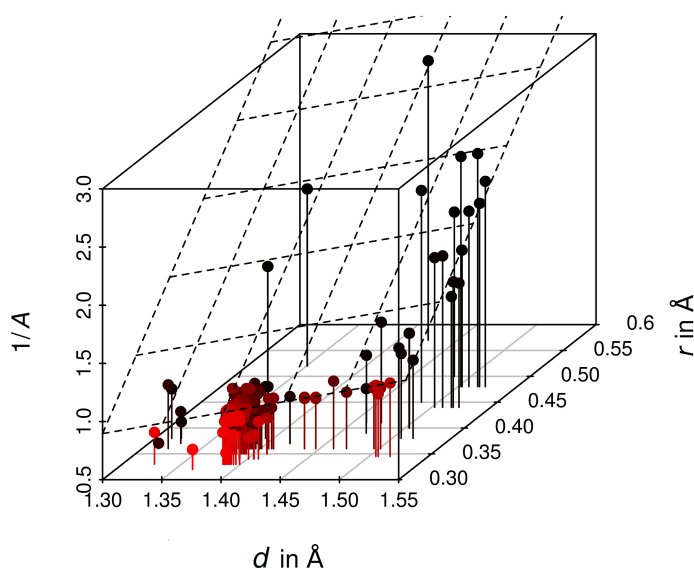
Investigations of BEDE parameters for bonds between carbon atoms with negative  $r$  and also for bonds between carbon and hydrogen in both directions did not reveal any correlations.

### 5.3.9.2 Analysis of bonds between carbon atoms with fixed $r$

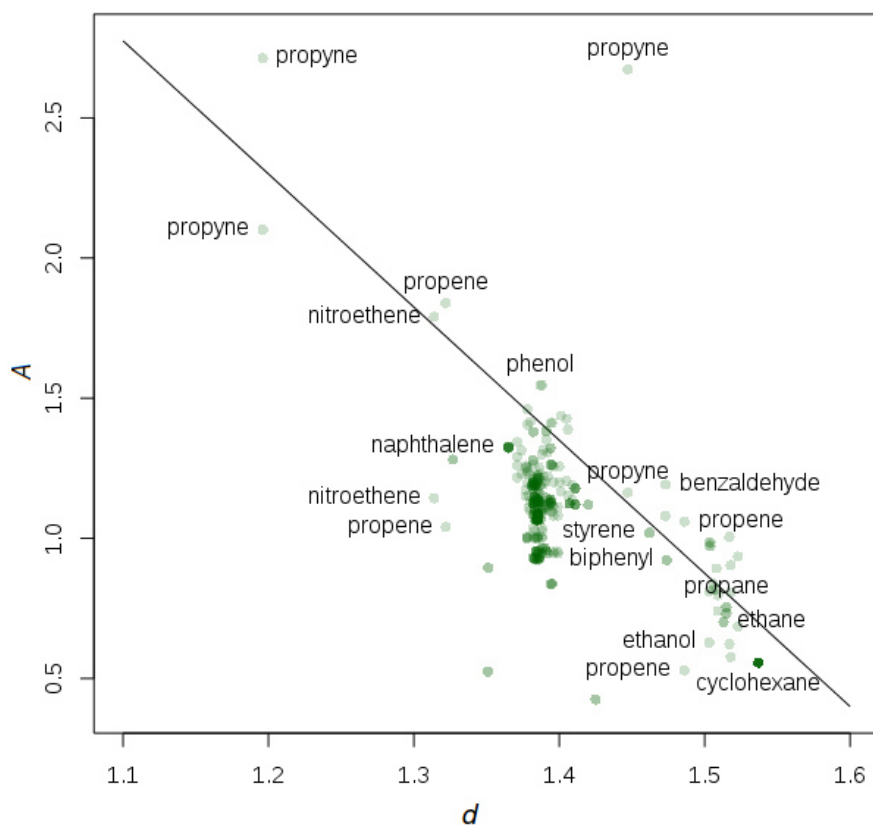
Propyne was added to the model compounds studied already for refinement with a fixed  $r$  of 0.36 Å. Fixing  $r$  allowed the investigation of  $A$  in dependence of the bond length without the correlation with  $r$ . This approach was preferred to the alternative of a fixed ratio of  $r$  to bond length, because for the additional compound propyne not the slightest hint of a fixed ratio between  $r$  and  $d$  was found and no clear relation between the two parameters was visible in the scatter plot matrix in Figure 5.36.

Noteworthy about the BODD refinement of propyne is that one  $A$  value for the single bond refined to 1.4, which is similarly high as  $A$  for the triple bond. In the following analysis of all the BEDE parameters for bonds between carbon atoms with a positive  $r$  (set to 0.36 Å) this was the strongest outlier.

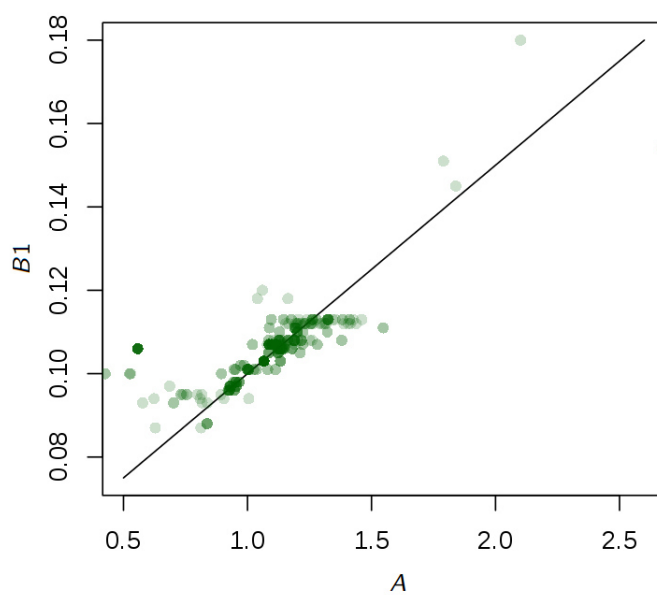
A plot of amplitude  $A$  versus the bond length is shown in Figure 5.38 and  $B1$  versus  $A$  is presented as Figure 5.39. The amplitudes  $A$  of the triple bond in propyne supply a valuable orientation for the relation on one parameter to the other. The linear regression is mostly determined by the huge number of points from the aromatic systems and is therefore unreasonable in this case. Hence, slope and intersection were deduced manually from certain points which should be close to the correlation between the parameters studied. In the case of parameter  $A$  versus bond length the resulting equation is  $A = -4.75 \cdot d + 8$ , which although not very precise, yields an approximation for the calculation of amplitude  $A$  for carbon bonds. The relation was applied to the two test structures to evaluate the effect on the figures of merit in comparison to other models. The results will be discussed in Section 5.3.9.4.



**Figure 5.37:** 3D plot of  $1/A$  (BEDE) versus  $r$  and bond length  $d$  from bonds between carbon atoms and  $r > 0$  from 34 model compounds. The regression plane suggests  $1/A = 1.6(2) \cdot d + 7.2(2) \cdot r - 3.9(3)$  with an adjusted  $R^2 = 0.847$ .



**Figure 5.38:** Parameter  $A$  versus bond length for bonds between carbon atoms for 31 compounds refined with  $r=0.36$  Å with a function roughly approximated manually as  $A = -4.75 \cdot d + 8$ .



**Figure 5.39:** Parameter  $B1$  versus  $A$  for bonds between carbon atoms for 31 compounds refined with  $r = 0.36$  with a function roughly approximated manually as  $B1 = 0.05 \cdot A + 0.05$ .

### 5.3.9.3 Analysis of bonds between carbon atoms with $r$ , $B1$ and $B2$ fixed

Since correlation between  $A$  and  $B1$  was still observed, the model compounds were refined once more with fixed  $B1$  and  $B2$  values for the BEDE instructions concerning the carbon bonds with a positive  $r$ .  $B1$  was set to 0.11 and  $B2$  to 0.81. Whereas 0.81 is the mean value deduced from the original refinement of the 34 model compounds, the mean value for  $B1$  is closer to 0.10 but since there are systematically fewer bonds of a high bond order, which would have a higher  $B1$ ,  $B1$  was fixed to 0.11. Thus the only parameter left to refine for the BEDE instructions of carbon bonds with positive  $r$  was  $A$  in this experiment.

Except for propyne, which could not be modeled well with free  $r$  and  $B$  parameters either, all model compounds still refined to R1 values below 1.0 %. Since  $B$  values were changed, several refinements required many cycles until convergence. Those molecules with only very few bond types sometimes required also the  $B2$  values of non-carbon bond related BEDE instructions to be fixed to 0.81 in order to prevent otherwise unreasonable results for  $B2$ . The number of relevant model compounds investigated was 33 (including propyne, but leaving out methanol and methylamine).

Unfortunately the fixed  $B2$  prohibited many amplitudes from increasing, so that no reasonable relation between  $A$  and the bond length could be deduced (see Figure 7.16, Appendix C). Therefore results from the previous investigations were applied to refinements against experimental data.

### 5.3.9.4 Applying statistical results to refinements against experimental data

The relation between  $B1$  and  $A$  could approximately be described by  $B1 = 0.05 \cdot A + 0.05$  as shown in Figure 5.39. The mean ratio of  $r$  over the bond distance was 0.26. Those approximate results from the statistical investigations were applied to model the ED between carbon atoms, by deriving all parameters except  $B2$  from the bond length,  $d$ :

$$r = 0.26 \cdot d$$

$$A = -4.75 \cdot d + 8.0 \quad (+20 \text{ for refinement of theory2experiment factor})$$

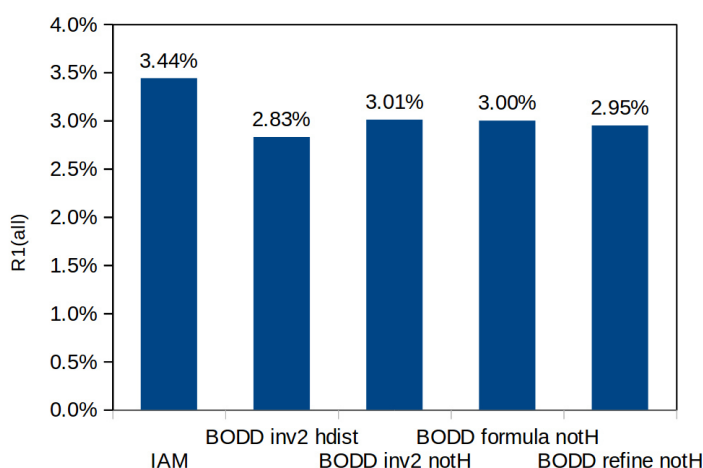
$$B1 = -0.238 \cdot d + 0.45$$

$B2$  was left at the value that was transferred from the refinement against simulated data, as were all the other BODD parameters for non-hydrogen atoms. No BEDE instructions involving hydrogen atoms were applied, while the distance to hydrogen atoms was refined.<sup>12</sup>

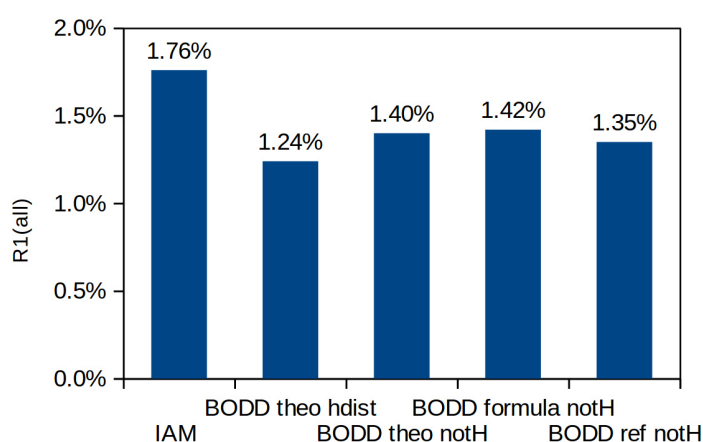
Application of the equation derived for bonds between carbon atoms to bonds between carbon and other elements like oxygen and sulfur is not reasonable, since bond lengths differ considerably. Therefore,  $A$  would become negative for the carbon-sulfur bond in ylid and higher than two for the carbon-oxygen bond. In order to calculate parameters for bonds to and between heteroatoms either different relations need to be derived, or electronegativity difference are to be included in the calculation. Differences in directionality for bonds between different atoms could also have a relevant influence.

<sup>12</sup>As shown earlier refining the distance to hydrogen atoms is closer to modeling the bonding ED than fixing them to the theoretical distances without any BEDE modeling.

**Figure 5.40:** R1 for refinements of MBADNP against simulated data with different models including one where the BEDE parameters for ED between carbon atoms were calculated by the derived formulas.



**Figure 5.41:** R1 for refinements of ylid against simulated data with different models including one where the BEDE parameters for ED between carbon atoms were calculated by the derived formulas.

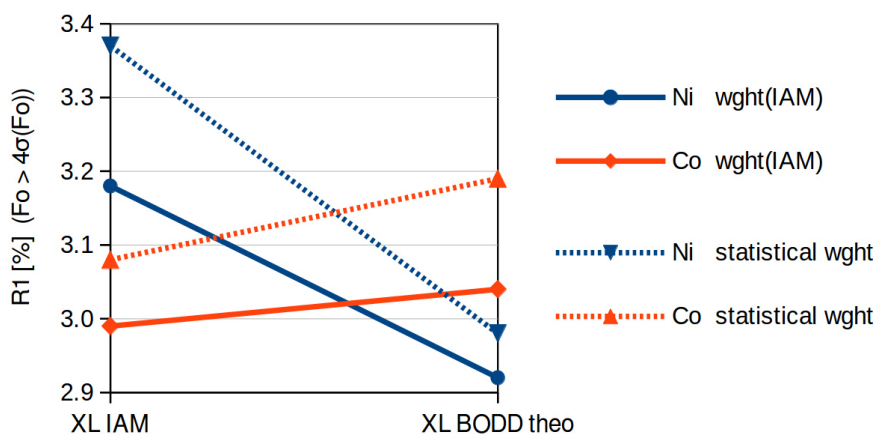


As Figures 5.40 and 5.41 show for MBADNP and ylid respectively, R1 increased only slightly for the model in which the parameters for the bonds between carbon atoms were calculated (formula) compared to the refinement where all parameters were transferred from theory. Of course refining parameters for bonds between non-hydrogen atoms yielded a better R1, but the best result is still obtained for the transfer of theoretical parameters and constraining the hydrogen atoms to theoretical bond lengths.

Including the number of parameters refined in addition to those applied in the IAM, the models with theoretical hydrogen distances included only one more parameter whereas six had to be added for refining the bond length to hydrogen atoms for ylid and nine for MBADNP. When BEDE parameters were refined, nine parameters were added for ylid and thirteen for MBADNP. This means that the lowest R1 actually corresponds to the lowest number of additional parameters, which shows the power of parameter transfer from theory.

### 5.3.10 Application to identifying metal atoms

The application of aspherical scattering factors to identify the correct metal atom in crystal structures of coordination compounds as presented in Chapter 3 for the multipole model should be possible with BODD, too. Therefore, BODD parameters were refined with SHELXL against the simulated diffraction data of pair 4 of Section 3.3.2.4. Only the ligand was modeled by BODD, since this resembles the application of parameters from the invariom database best. In order to ignore the bond to the metal atom during refinement of LONE parameters for the coordinating atoms the covalent radius of the metal was reduced artificially for the program.<sup>13</sup> Only 31 parameters were refined for the whole molecule, with *B2* set to 0.7 and one global *B1* variable that refined to about 0.07. A more detailed model would have been possible, but this crude refinement already lead to surprisingly good results upon transfer to the experimental structure model. The BODD parameters were transferred to the model refined against the experimental data and kept fixed. Accidentally, no theory2experiment scale factor was refined, and hydrogen atom distances were elongated to theoretical values.



**Figure 5.42:** Comparison of R1 for reflections stronger than  $4\sigma$  for the refinements of pair 4 with SHELXL with different metal atoms in an IAM and a BODD model.

As Figure 5.42 shows the refinement improved considerably for nickel while it worsened for cobalt. For refinements displayed by continuous lines the weighting scheme (wght) as suggested by SHELXL after IAM refinement of nickel was applied. Statistical weights lead to higher figures of merit in general, but the slopes upon inclusion of asphericity become steeper. The theoretical BODD parameters were transferred from simulated data of the nickel complex but usage of those from the cobalt complex just rendered an even slightly lower R1 for nickel of 2.90 % for the IAM weighting scheme and confirmed the results for cobalt. Modeling the central atom by scattering factors of the metal ions slightly increased the residual factors for IAM and BODD models with both metals to a similar extent.

This example shows how useful modeling of aspherical density by theoretical parameters can be. With the new BODD model in SHELXL a change of programs in order to apply multipole modeling is not necessary any more. BODD parameters from the invariom database should have comparable effects as invariom multipole parameters.

<sup>13</sup>Alternatively the SHELXL CONN instruction could have been applied to define the metal's connectivity.

## 5.4 Discussion

### 5.4.1 Summary

A new way to include aspherical atoms in crystal structure models refined with SHELXL was presented.

BEDE and LONE are new instructions for SHELXL which facilitate modeling of bonding and lone pair ED. Their main advantage compared to the multipole model is the use of bond orientations instead of local orthogonal coordinate systems. Thereby, the new model is closer to a chemist's intuition and named bond-oriented deformation density (BODD). The different procedure for assignment of directionality avoids problems during the orientation of local atomic coordinate systems.

For simple organic, small molecules BODD describes the molecular ED almost as well as the multipole model when refined with XD. Lone pair treatment is still inferior to the multipole model by judging from refinements against simulated data. A benefit of BEDE and LONE is the chemical intuitiveness as mentioned above. This agreement with basic chemistry knowledge is reflected by values of LONE parameters for substituted benzene molecules. LONE amplitude of carbon atoms in *ortho*, *meta* and *para* positions changes in accordance with the mesomeric effect of a certain substituent.

Since the goal of the new model is to facilitate the modeling of bonding and lone pair ED for single crystal X-ray diffraction data of standard quality, the overall aim is to refine BEDE and LONE parameters against simulated diffraction data of small, organic molecules and transfer them to structures of interest as is known for the multipole parameters of the invariom database. The proof of principle was successful. Transferability of BEDE and LONE parameters between different theoretical structures was demonstrated for the first time. Averaged parameters were transferred to different aromatic molecules and also from several small molecules to the larger molecule MBADNP. The transferability rules show similarities to those of multipole parameters as shown at the example of oxygen atoms in nitro groups. In general, transferability between theoretical molecules was established.

Successful transfer of BODD parameters from refinements against simulated data to models of experimental data required refinement of a scale factor for the transferred BODD. The thus refined models of MBADNP and ylid led to improved figures of merit and ADPs as well as smaller uncertainties of bond lengths and angles. For MBADNP a comparison to neutron data also suggested a higher accuracy in addition to the improved precision. The best result for transferred parameters was obtained by placement of hydrogen atoms at bond lengths obtained from theoretical calculations and modeling their bonds by BEDE instructions. Compared to the IAM it was also an improvement to only treat non-hydrogen atoms by BEDE and LONE, while the distances to hydrogen atoms were refined or set to standard X-ray bond length. This treatment proved useful if BEDE and LONE parameters were refined against experimental data, although deconvolution of aspherical ED and thermal motion was best for BODD parameters transferred from theory.

Furthermore, the new BODD model could be applied to identify the correct metal atom in coordination compounds by refinement against diffraction data, if the asphericity parameters are transferred from a refinement against simulated data in an invariom-like approach.

### 5.4.2 Perspective

In addition to the presented comparison of different models for XRD data of MBADNP to the model refined against neutron data, similar comparisons for other compounds would support the conclusions by validation against more than one neutron diffraction experiment.

The most important decision to be addressed, however, concerns the best classification of BODD parameters for their transfer from the invariom database. Invariom names are a possibility, for which then an appropriate technical way of parameter assignment has to be decided on. Different possibilities have been presented in this thesis. Another challenge is the refinement of ionic model compounds and a general treatment of nitrogen atoms in secondary amines between  $sp^2$ - and  $sp^3$ -hybridized carbon atoms, since their planarity can be ambiguous.

When those problems are solved, a preprocessor program like `INVARIOMTOOL` can be created to transfer BEDE and LONE parameters from the model compounds of the invariom database. Successful application of the tool to any model of an organic molecule requires BODD parameters for all invariom model compounds.

Instead of parameterizing the numerous model compounds, relations between BEDE and LONE parameters and the information already present in the model could probably be derived, as exemplary shown in a first attempt for BEDE parameters of bonds between carbon atoms. In order to follow this approach bonds between different atoms will have to be investigated with respect to their bond lengths, and the relation between different elements should be checked for correlation with electronegativity differences. Such relations would provide an elegant way of generating BODD parameters or starting values for refinements in the course of parameterizing all invariom model compounds. A transfer of parameters from the invariom database, however, should yield better results.

Before distribution of software that provides easy access to aspherical modeling with `SHELXL` to a broad range of scientist, the reason for the `theory2experiment` scale factor should be elucidated. Afterwards, the way to more precise and accurate structure modeling in `SHELXL` will be paved, so that many scientific projects can profit from improved crystallographic structure models.





# 6 Conclusion

## 6.1 Summary

In the previous chapters many new applications of the invariom database have been discussed.

Firstly, the invariom database itself was completely renewed and expanded. All model compounds of the database were treated with a newer functional and a basis set that allows invariom scattering factors to be obtained for elements as heavy as krypton. The new basis set allowed the inclusion of bromine compounds in the database, so for each model compound containing either chlorine or fluorine the equivalent bromine compound was added. Thus the database was not only renewed but also expanded. Overall, the collection of model compounds in the invariom database was increased by almost 400 molecules.

Secondly, additional properties transferable via invariom classification were introduced. The renewed geometry optimizations of each model compound yielded frequencies for prediction of hydrogen ADPs. A fit of atomic point charges to the electrostatic potential deduced from the molecular wave functions built the basis for invariom point charges. The development of these charges was a main topic of this thesis. The addition of bond-oriented deformation density for aspherical modeling with SHELXL transferable from the invariom database was prepared for future application in form of a pilot study.

The new attributes now transferable from the invariom database required a more strictly defined invariom terminology. Moreover, the thesis yielded two changes in invariom classification. During the atomic point charge project it turned out that double bonded oxygen atoms as well as hydrogen atoms should be classified to a higher degree. A way to achieve this without requiring additional model compounds has been established. While the hydrogen-invariom names were extended only for invariom point charges the new classification of oxygen atoms involved in a double bond also led to an improved performance of the invariom scattering factors.

Due to the more extensive basis set, the new database allowed homogeneous treatment of structures that contain coordination compounds with 3d-metals and co-crystallized other molecules. Aspherical scattering factors for coordination compounds were specifically tailored by an invariom-like approach. This approach for aspherical modeling of complexes was applied to a series of crystal structures, in order to clarify which of two possible metal atoms fits the X-ray diffraction data best. Although modeling of the whole coordination compound by multipoles often improved the fit to the data further, invariom modeling of only the ligand already allowed identification of the correct metal atom in most cases. This case study is a new example for a contribution to science by application of invarioms.

The development of invariom point charges was more methodical and included the development of computational tools for the transfer from the database to a molecule of interest.

The invariom model compound assignment was changed for the construction of an invariom point charge database. While for the other properties the attribute to a given invariom is derived from one model compound, the invariom point charges are averaged. Point charges of all atoms in the database that have a specific chemical bonding environment were averaged to yield the charge for the invariom defining this environment. Although this proceeding has not been typical for invarioms a clear improvement upon averaging of point charges was observed. The developed charges were validated by their ability to reproduce the electrostatic potential of a set of angiogenesis inhibitor molecules. Additionally, the molecular electrostatic potentials from the new charges were compared to those acquired from different point charges and those derived from multipole refinements against experimental diffraction data. Especially the last comparison showed that invariom point charges perform well at reproducing molecular electrostatic potentials. Tools for invariom point charge assignment and for quickly deriving as well as visualizing the electrostatic potential on a molecular surface were presented.

The pilot study on invarioms for aspherical modeling in SHELXL by bond-oriented deformation density was a great success. The bond-oriented deformation density model is a method by which the invariom database can be applied to improve refinements in SHELXL. The overall target is a preprocessor program that will perform a similar task as INVARIOMTOOL, but for bond-oriented deformation density instead of multipoles. Thus, access to the benefits of invariom modeling will be provided to a broader user group and, hence, many more scientific projects. Bond-oriented deformation density is an alternative model to the commonly used *Hansen* and *Coppens* multipole model and resembles charge cloud models developed earlier, but includes some improvements. The ability of this model to describe the aspherical density compared to that of the multipole model for refinements against simulated data has problems with constrained ring geometries and heavier elements; otherwise it performs almost similarly well. In combination with the conceptual simplicity and ease of application the model has the potential to become useful for improving routine structure determinations. The transferability of the new parameters between different structures refined against simulated data was demonstrated. Bond-oriented deformation density from theory improved structural models of experimental data upon refinement of one scale factor for the transferred density. The models including bond-oriented deformation density from theory yielded bond lengths of higher precision than models in which the parameters were refined against the experimental data or models with spherical scattering factors. Furthermore, different models of one molecule were compared to a model refined against neutron diffraction data. The comparison suggested higher accuracy of ADPs and atomic coordinates for models that describe the asphericity by parameters from invariom model compounds than for independent atom models. This pilot study shows that invariom parameters for aspherical modeling in SHELXL can and should be developed.

Modeling of coordination compounds with bond-oriented deformation density parameters refined against simulated X-ray diffraction data allowed the identification of the correct central metal atom. The results were similar to those of the project during which correct metal atoms were identified by modeling with multipole parameters from the invariom database. The difference is that this time the program SHELXL could be used directly to model bonding and lone pair electron density.

Overall, new applications for the invariom database were developed, interesting cases

were studied by special invariom treatment and a basis for better access to the improved modeling by aspherical scattering factors was provided.

## 6.2 Outlook

The extraction of geometrical restraints for highly differentiated atomic environments is one possible further application of the invariom database. Questions concerning classification of bonds instead of atoms similar to those of bond-oriented deformation density will have to be solved for this. Especially the derivation of restraints for even farther distances such as 1,3-distances for bond angles will be challenging.

The next step, however, for broadening the application of the invariom database is the parametrization of the complete invariom database for bond-oriented deformation density and introduction of a preprocessor program for parameter transfer. This will improve structural models refined with *SHELXL* without adding variables to the refinement. The application of bond-oriented deformation density from the invariom database should be as simple as possible to convince many users of the benefits this approach entails.



# 7 Appendix

## 7.1 Appendix A – Identification of metal atoms

**Table 7.1:** SCF energies in Hartree for the geometries refined with the invariom model for high- and low-spin cobalt(II) and nickel(II) octahedral complexes. Lower energies are marked in bold.

pair	Co(hs)	Co(ls)	difference (ls-hs)
3	<b>-2687.4064</b>	-2687.3897	0.0167
4	<b>-2830.2233</b>	-2830.2019	0.0214
	Ni(hs)	Ni(ls)	
1	<b>-2724.3078</b>	-2724.2855	0.0223
3	<b>-2812.9664</b>	-2812.8447	0.1217
4	<b>-2955.7825</b>	-2955.7155	0.0670

**Table 7.2:** SCF energies in Hartree for the geometries refined with the invariom model for high- and low-spin cobalt(II) and nickel(II) in tetrahedral complexes. Lower energies are marked in bold.

pair	Co(hs)	Co(ls)	difference (ls-hs)
8	<b>-7799.3570</b>	-7799.3111	0.0459
	Ni(hs)	Ni(ls)	
8	<b>-7924.9005</b>	-7924.8531	0.0473
10	<b>-3770.9427</b>	-3770.8689	0.0738

**Table 7.3:** SCF energies in Hartree for the geometries refined with the invariom model for high- and low-spin cobalt(II) and nickel(II) square-planar complexes. Lower energies are marked in bold.

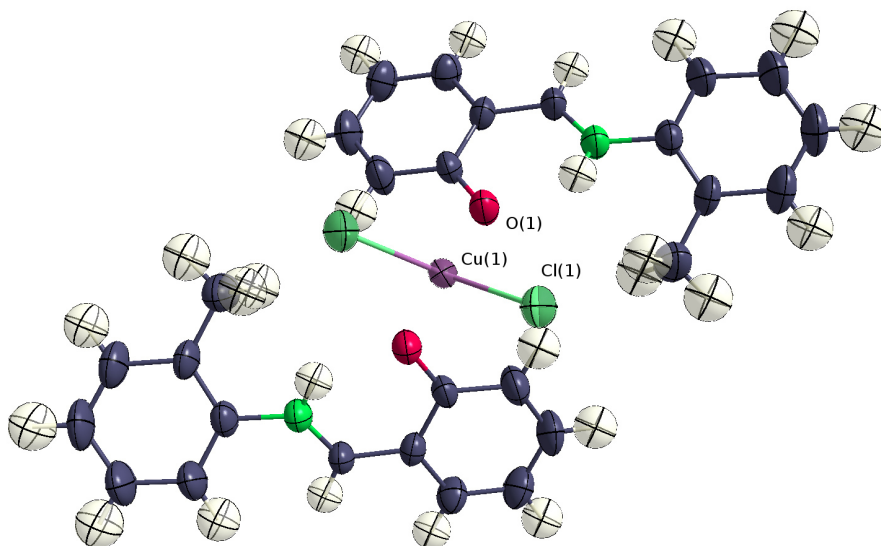
pair	Co(hs)	Co(ls)	difference (ls-hs)
5	-2378.4417	<b>-2378.4821</b>	-0.0404
	Ni(hs)	Ni(ls)	
5	-2504.0105	<b>-2504.0531</b>	-0.0426
6	-8662.2079	<b>-8662.2306</b>	-0.0227
7	-2928.3477	<b>-2928.3619</b>	-0.0142

**Table 7.4:** Cell parameters for both structures of pair **3** and their deviation.

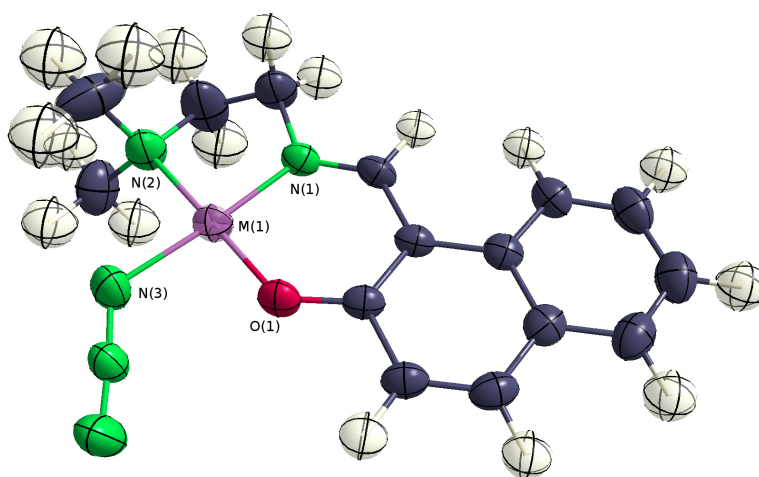
cell	<i>a</i>	<i>b</i>	<i>c</i>	$\beta$
<b>a</b>	9.1863	13.8653	19.987	93.604
<b>b</b>	9.1896	13.8731	20.001	93.641
e.s.d.	0.0006	0.0009	0.001	0.001
$\Delta$	0.0033	0.0078	0.014	0.037

**Table 7.5:** Energy change upon geometry relaxation for pair **3**, in atomic units if not indicated otherwise.

	Ni(ls)	Ni(hs)	Co(hs)	Co(ls)
E(HF) starting geometry <b>a</b>	-2812.8400	-2812.9664	-2687.4064	-2687.3370
E(HF) starting geometry <b>b</b>	-2812.8984	-2812.9658	-2687.4055	-2687.3903
E(HF) optimized geometry		-2812.9664	-2687.4320	-2687.4161
quotient <b>a</b>		1.0000000	1.0000095	1.0000294
quotient <b>b</b>		1.0000002	1.0000098	1.0000095798
change [%]		0.0000%	0.0010%	0.0029%
change opt-sp		0.0000	-0.0256	-0.0791
change [kJ/mol]		0	-67	-208



**Figure 7.1:** Thermal ellipsoid plot at a probability of 50 % of pair 10 with Cu after refinement of the whole-molecule scattering factors against data set a.



**Figure 7.2:** Thermal ellipsoid plot at a probability of 50 % of pair 11 with Ni after refinement of the whole-molecule scattering factors against data set a.

## 7.2 Appendix B – Invariom point charges

Iterating over all points (flag=0)

– Iterating over all atoms (if flag not >1)

If dist < 1.4 vdw → flag = flag+2

If dist > 2.0 vdw → flag = flag\*1

Else → flag = 1

Test:

flag = 0  
outside  
mol. envelop

flag = 1  
on  
mol. envelop

flag >1  
Inside  
mol. envelop

add to sum

Cases:

a) far from any atom → 0

b) too close to one atom and  
too far from others → 2

within envelope of other atom → 3

c) too close to 2 atoms and  
too far otherwise → 4

In envelope of other atom → 5

In envelope of 2 atoms → 6

d) far from all atoms but

In range of one atom → 1

In range of 2 or more atoms → 1

**Figure 7.3:** How the program for RMS calculation selects grip points within distance between 1.4 and 2.0 Å of the molecule works in principle. All possible cases and their outcomes are listed on the right side.

	$V_{S,av.}^+$	$V_{S,av.}^-$	$\Pi$	$\sigma_{tot}^2$	$\nu$
L-Asn-L-Asn-L-Asn					
whole mol.	0.051	-0.062	0.057	0.0028	0.2442
invariom	0.047	<b>-0.057</b>	<b>0.051</b>	<b>0.0025</b>	<b>0.2274</b>
amber	0.054	-0.149	0.041	0.0191	0.0736
L-Trp-L-Trp-L-Trp					
whole mol.	0.036	-0.041	0.038	0.0022	0.2308
invariom	0.049	<b>-0.032</b>	0.042	0.0027	<b>0.1810</b>
amber	<b>0.029</b>	<b>-0.050</b>	<b>0.038</b>	<b>0.0024</b>	0.1505
L-Ala-L-Ala-L-Ala					
whole mol.	0.034	-0.052	0.039	0.0023	0.2221
invariom	0.050	<b>-0.049</b>	0.049	0.0033	<b>0.2447</b>
amber	<b>0.040</b>	-0.046	<b>0.041</b>	<b>0.0023</b>	0.2488

**Table 7.6:** Results from *Poltzer* analysis of the three homotripeptides' ESPs calculated from different atomic point charges. Values in better agreement with those from the 'whole-molecule' charges are printed bold.



	x	y	z	total
L-Asn-L-Asn-L-Asn				
whole mol.	-3.61	-2.38	-5.34	6.87
invarioms	<b>-5.58</b>	<b>-6.26</b>	<b>-3.84</b>	<b>9.22</b>
amber	24.47	8.17	-13.90	29.30
L-Trp-L-Trp-L-Trp				
whole mol.	-3.51	0.25	3.15	4.70
invarioms	<b>3.80</b>	-5.67	<b>4.08</b>	<b>7.95</b>
amber	-9.24	<b>2.28</b>	5.13	10.81
L-Ala-L-Ala-L-Ala				
whole mol.	0.66	0.72	0.52	1.11
invarioms	<b>5.43</b>	<b>0.68</b>	-0.76	<b>5.53</b>
amber	-5.89	1.54	<b>0.99</b>	6.16

**Table 7.7:** Dipol moments of the three homotriptides calculated from different atomic point charges. Values in better agreement with those from the 'whole-molecule' charges are printed bold.

**Listing 7.1:** BASH commands for using APD-TOOLKIT in combination with JMOL.

```
/path/to/APDToolkit -pqr
sh /path/to/jmol.sh resp.pqr -s /path/to/esp_script.spt -o
```

**Listing 7.2:** Setting for JMOL to display an nice ESP, stored in esp\_script.spt for starting the program by script.

```
set frank off; select all; hbonds off; spin off; wireframe off
; spacefill off; trace off; set ambient 40; set specpower
40; slab off; ribbons off; cartoons off; label off; monitor
off; set background white;
isosurface resolution 9 solvent 1.4 color absolute -.25 0.25
colorscheme rwb MAP mep ;
```

### Invariom names for new hydrogen atom treatment

In order to achieve a better classification of hydrogen atoms, especially those at aromatic systems, new hydrogen invariom names were introduced. The new names are implemented in the program APD-Toolkit<sup>[230]</sup>, but not in the program INVARIOMTOOL<sup>[91]</sup> yet.

To ensure downwards compatibility the old invariom names are supplemented by the invariom name of the adjacent atom linked by the character '&'. So exemplary new names are:

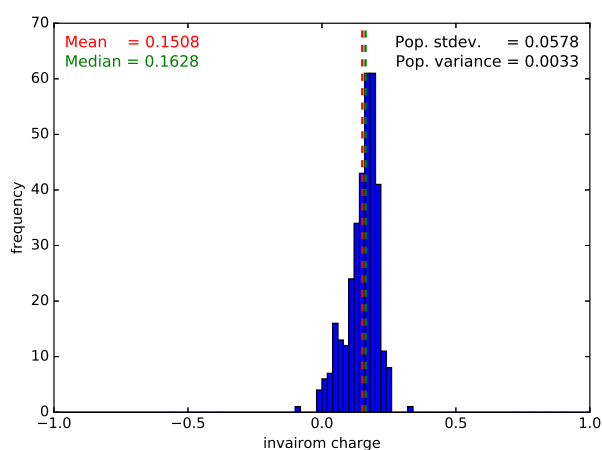
H@6c&6-C#6c[#6c1c]#6c[#6c1h]1h (from: toluene),  
H@6c&6-C#6c[#6c1n]#6c[#6c1h]1h (from: aniline),  
H@6c&6-C#6c[#6c1n]#6c[#6c1n]1h (from: m-phenylenediamine).

Table 7.8 contains three invarioms of hydrogen attached do an aromatic six-membered ring which in the old nomenclature would have been H@6c. Their charges of 0.126, 0.1316

**Table 7.8:** Most frequent invarioms within the invariom database with their charges and count.

invariom name	mean charge	frequency
6-C#6c[#6c1h]#6c[#6c1h]1c	0.1848	202.667
C@6c1h1h1h	-0.41041	267
C1c1h1h1h	-0.372984	273.667
6-C#6c[#6c1c]#6c[#6c1h]1h	-0.217738	307.333
H@6c&6-C# 6c[#6c1c]#6c[# 6c1h]1h	0.130544	307.333
H1n[@6c1h]&N@6c1h1h	0.352939	330
6-C#6c[#66c1h]#6c[#6c1h]1h	-0.10866	364
H@6c&6-C#6c[#66c1h]#6c[#6c1h]1h	0.125853	364
6-C#66c[#66c#6c]#6c[#6c1h]1h	-0.162577	365
H@6c&6-C#66c[#66c#6c]#6c[#6c1h]1h	0.13163	365
H1c[@5c1h1h]&C@5c1h1h1h	0.12691	483
H1c[1n1h1h]&C1n1h1h1h	0.10942	531
H1c[@6c1h1h]&C@6c1h1h1h	0.118041	801
H1c[1c1h1h]&C1c1h1h1h	0.107962	821
6-C#6c[#6c1h]#6c[#6c1h]1h	-0.104417	1582
H@6c&6-C#6c[#6c1h]#6c[#6c1h]1h	0.119284	1582

and 0.1193 differ insignificantly (the standard deviation of the last one being 0.013 as shown in Figure 7.6a). But there are less frequent examples closer to heteroatoms which differ significantly (Figure 7.4).

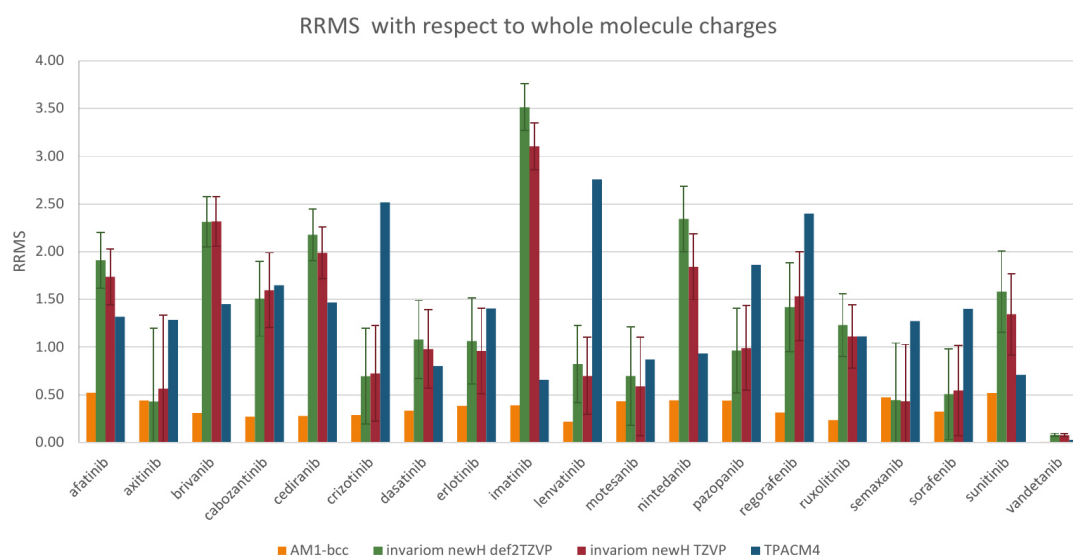
**Figure 7.4:** Frequency distribution of the 343 different new invariom charges that formerly were invariom H@6c.

Overall there are 343 different H@6c with the new names. If they are all averaged their frequency distribution is as shown in Figure 7.4. Comparison to the histogram of the most frequent invariom with the elongated name in Figure 7.6a shows considerably improved spread.

**Table 7.9:** Median values for RMS and RRMS of disturbed ESP of sunitinib (Section 4.3.4)

disturbance spread	charges changed	RMS [ $e/\text{\AA}$ ] median	RRMS median
0.100	all	1.413	1.874
0.050	all	0.681	0.903
0.025	all	0.345	0.458
0.010	all	0.148	0.197
0.025	hydrogen	0.283	0.375
0.025	non-hydrogen	0.196	0.260

## Comparison to other point charges

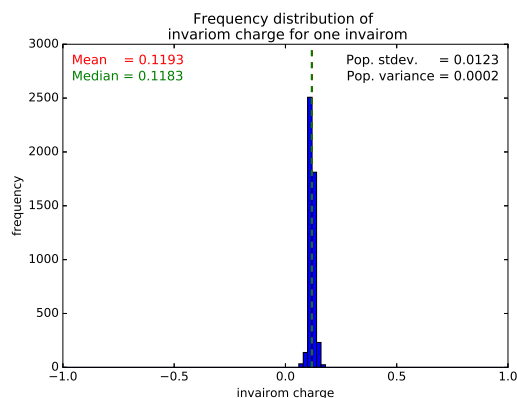
**Figure 7.5:** Detailed RRMS results for the comparison of different point charges for the investigated angiogenesis inhibitors.

## Experimental details for the comparison to aspherical refinement against experimental data

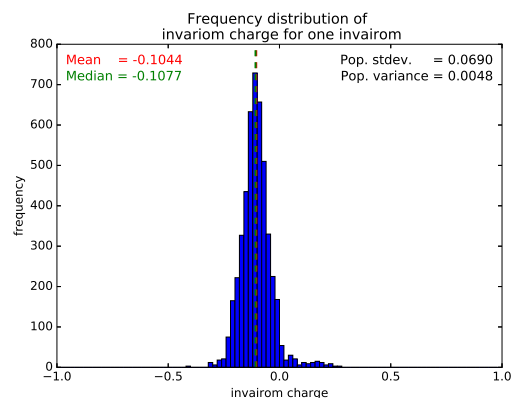
For this comparison the ESP was calculated by XD2006<sup>[12]</sup>. The electron density onto which it was mapped was calculated for the same grid also by XD2006. Mapping and graphical representation was done with MOLECOOLQT<sup>[40]</sup>. The geometry was always the one refined from the charge density data, and the electron density from a simple spherical atom model with Kappas of 1.00 for non-hydrogen atoms and 1.2 for hydrogen atoms. The electrostatic potentials from the different approaches were all derived from an adjusted multipole model. In the case of point charges the charges were set as monopole populations without deformation functions. The point charges from theory were generated

by GAUSSIAN09<sup>[103]</sup> MK population analysis after a single point calculation of the geometry known from experiment. Invariom multipoles were transferred with INVARIOMTOOL<sup>[91]</sup>.

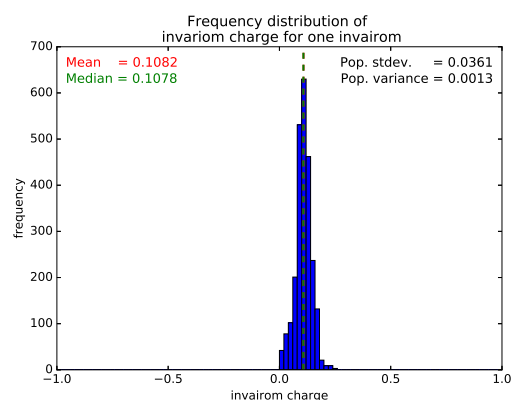
## Frequency distribution of selected invariom charges



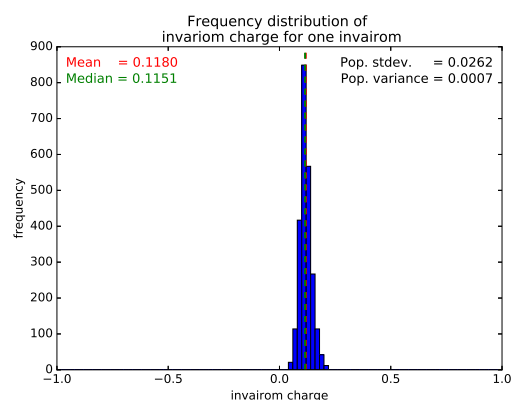
(a) H@6c&6-C#6c[#6c1h]#6c[#6c1h]1h



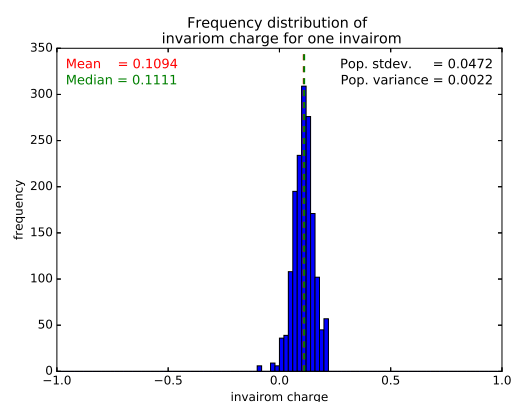
(b) 6-C#6c[#6c1h]#6c[#6c1h]1h



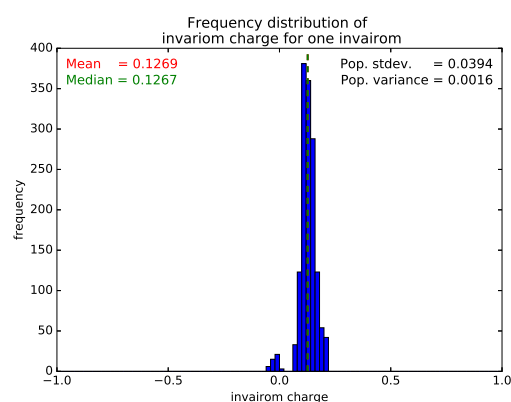
(c) H1c[1c1h1h]&C1c1h1h1h



(d) H1c[@6c1h1h]&C@6c1h1h1h

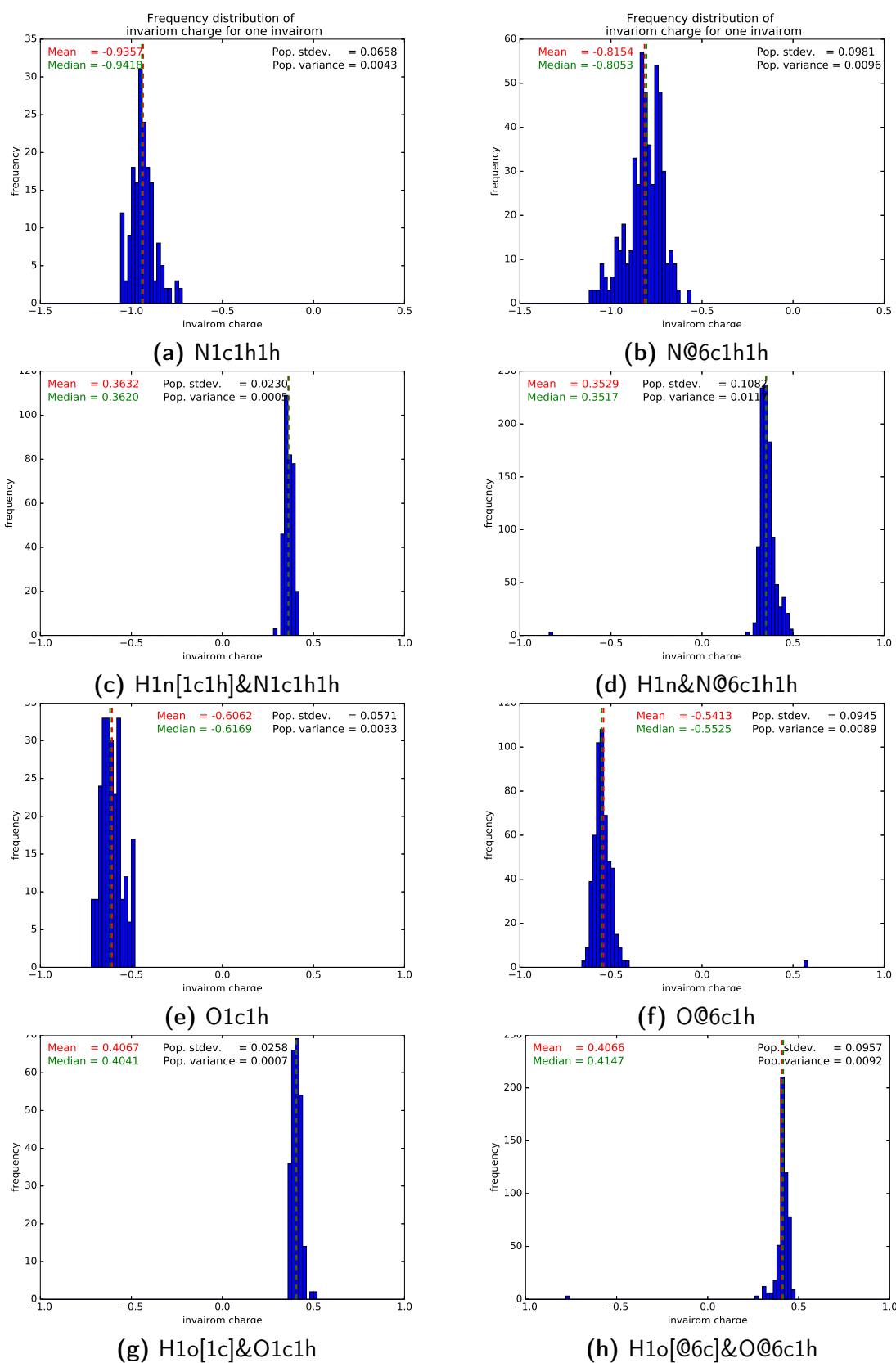


(e) H1c[1n1h1h]&C1n1h1h1h



(f) H1c[@5c1h1h]&C@5c1h1h1h

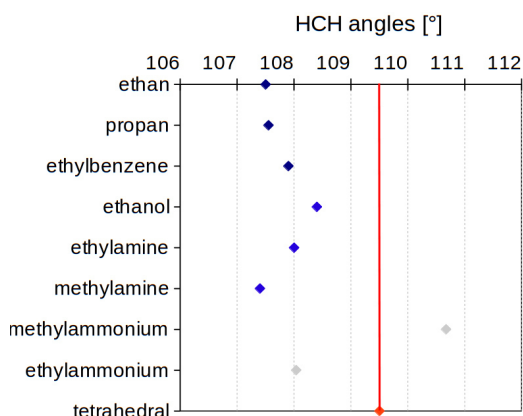
Figure 7.6: Frequency distributions of the charge of the most frequent invarioms.



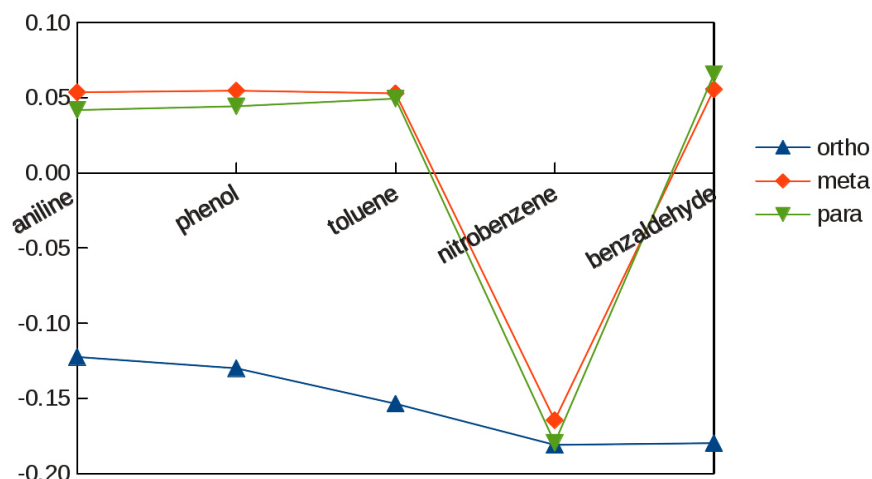
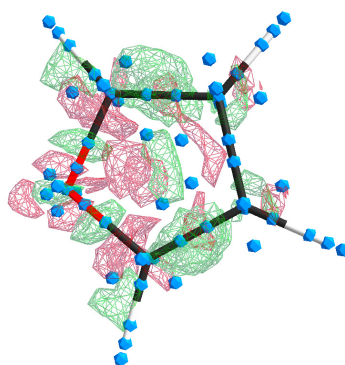
**Figure 7.7:** Frequency distributions of charge from exemplary non-carbon invarioms and their hydrogen invarioms.

## 7.3 Appendix C – Invarioms and aspherical modeling for SHELXL refinements

**Figure 7.8:** Bond angles between hydrogen atoms of selected molecules as calculated by DFT in comparison to the ideal tetrahedral angle.



**Figure 7.9:** Residual density at  $\pm 0.02 \text{ e}/\text{\AA}^3$  of furan with BEDE and LONE positions represented by blue icosahedra.



**Figure 7.10:** Populations of the  $d_{z^2}$  type quadrupole refined with XD for carbon atoms in aromatic systems with substituents of different mesomeric effects.

**Table 7.10:** Angle sums, planarity values and planarity of nitrogen atoms in selected secondary amines with bonds to one planar and one tetrahedral carbon atom.

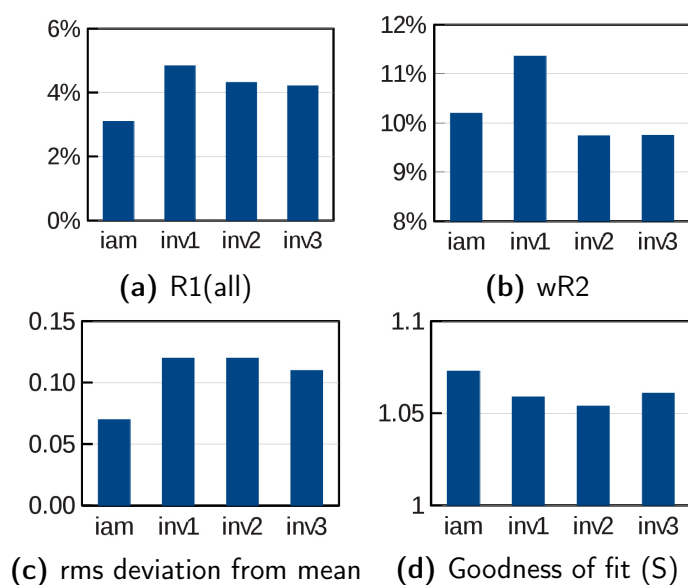
molecule	D95++3df3pd			def2TZVP		
	angle sum	planarity value	planar	dangle sum	planarity value	planar
R- <i>N</i> -(1-phenylethyl)pyridin-2-amine	356.25	0.792	no	358.41	0.908	yes
R- <i>N</i> -(1-phenylethyl)aniline	352.69	0.636	no	352.56	0.632	no
<i>N</i> -methylpyridin-2-amine	356.98	0.832	no	357.3	0.849	no
<i>N</i> -methylaniline	360.00	1.000	yes	353.81	0.684	no
Methyl-phenyl-amine	353.25	0.650	no	353.86	0.686	no
<i>N</i> -methylbenzo[d]thiazol-2-amine	360.01	1.000	yes	356.8	0.823	no

**Table 7.11:** Crystallographic data of MBADNP<sup>[248]</sup>.

	MBADNP (X-ray)	MBADNP (neutron)
formula	C <sub>13</sub> H <sub>12</sub> N <sub>4</sub> O <sub>4</sub>	C <sub>13</sub> H <sub>12</sub> N <sub>4</sub> O <sub>4</sub>
crystal system	monoclinic	monoclinic
space group, <i>Z</i>	<i>P</i> 2 <sub>1</sub> , 2	<i>P</i> 2 <sub>1</sub> , 2
cell constants		
<i>a</i> [Å]	8.352(3)	8.352(3)
<i>b</i> [Å]	8.570(3)	8.570(3)
<i>c</i> [Å]	8.909(4)	8.909(4)
$\beta$ [°]	93.98(2)	93.98(2)
crystal shape, color	block, yellow	hexagonal prism, yellow
crystal size [mm]	0.30 x 0.26 x 0.20	3.0 x 2.5 x 1.0
temperature [K]	20.0(2)	20.0(1)
radiation type, $\lambda$ [Å]	MoK $\alpha$ , 0.71073	neutron (steady state reactor, D9 at ILL)
$2\theta_{\max}$ [°]	79.94	93.12
$\sin \theta_{\max}/\lambda$ [Å <sup>-1</sup> ]	0.904	
$d_{\max}$ [Å]	0.553	
independent reflections	7873	3397
independent reflections with $I > 2\sigma(I)$	7612	
$R_{\text{int}}$	0.0203	0.0283

**Table 7.12:** Flack parameters for selected models of MBADNP refined against experimental data.

model	Flack <i>x</i>	Flack s.u.	Parsons <i>x</i>	Parsons s.u.
IAM S	0.092	0.575	-0.058	0.309
IAM hdist S	0.134	0.594	-0.063	0.309
BODD inv2 hdist	0.015	0.511	-0.079	0.310
BODD inv3 hdist	-0.006	0.512	-0.081	0.309
BODD theo hdist	0.001	0.508	-0.070	0.311

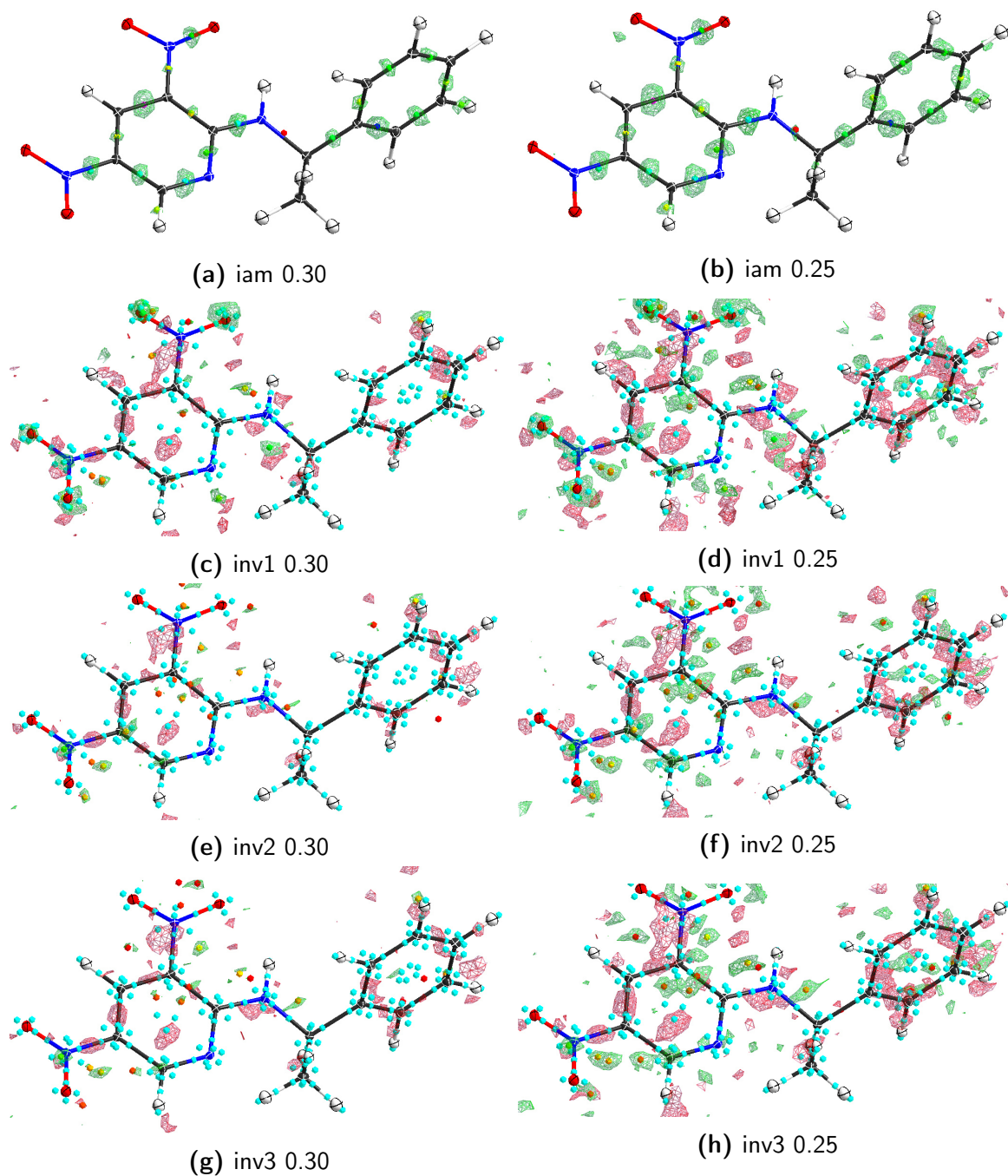


**Figure 7.11:** Comparison of figures of merit for different models refined against x-ray data of MBADNP.

**Table 7.13:** Crystallographic data of ylid.

formula	$C_{11}H_{10}O_2S$
crystal system	orthorhombic
space group, Z	$P2_12_12_1$ , 8
cell constants	
$a$ [Å]	5.8538(3)
$b$ [Å]	8.9284(5)
$c$ [Å]	18.3409(10)
temperature [K]	102(2)
radiation type, $\lambda$ [Å]	MoK $\alpha$ , 0.71073
$2\theta_{\max}$ [°]	90.628
$d_{\max}$ [Å]	0.50
independent reflections	8030
independent reflections with $I > 2\sigma(I)$	7949
$R_{\text{rim}}$	0.0290
$R_{\text{pim}}$	0.0042





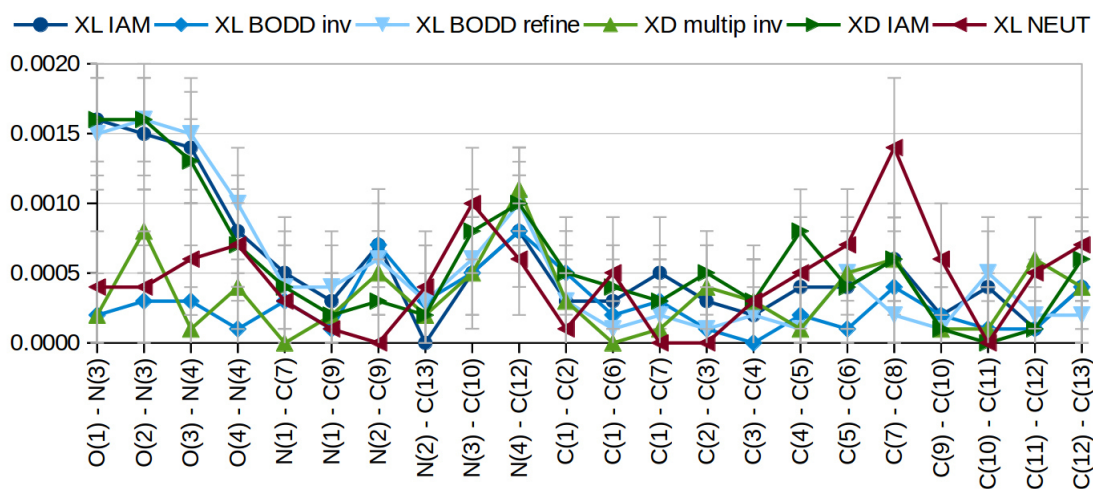
**Figure 7.12:** Comparison of residual electron density maps at  $\pm 0.25 \text{ e}/\text{\AA}^3$  and  $\pm 0.30 \text{ e}/\text{\AA}^3$  for different models refined against X-ray data of MBADNP. (Inverted, uncut and omitted the worst 2 reflections of model inv3 in all maps.)

Table 7.14: Bond lengths of different MBADNP models.

		IAM XL	inv2 BODD	free BODD	inv XD	IAM XD
O1	N3	1.2414(12)	1.2370(11)	1.2397(11)	1.2366(7)	1.2423(8)
O2	N3	1.2296(12)	1.2267(11)	1.2284(11)	1.2276(7)	1.2304(5)
O3	N4	1.2339(12)	1.2305(11)	1.2326(11)	1.2326(7)	1.2333(5)
O4	N4	1.2362(12)	1.2330(11)	1.2355(11)	1.2358(7)	1.2370(6)
N1	C9	1.3391(13)	1.3388(12)	1.3388(12)	1.3412(8)	1.3399(8)
N1	C7	1.4726(13)	1.4735(12)	1.4748(12)	1.4719(8)	1.4714(4)
N1	H1N	0.9071	1.004	0.865(19)	1.005	0.907059
N2	C13	1.3245(13)	1.3218(12)	1.3246(13)	1.3585(8)	1.3601(7)
N2	C9	1.3605(13)	1.3611(12)	1.3606(13)	1.3237(8)	1.3250(9)
N3	C10	1.4532(13)	1.4532(12)	1.4545(13)	1.4548(8)	1.4546(6)
N4	C12	1.4456(13)	1.4449(12)	1.4465(13)	1.4456(8)	1.4463(8)
C1	C2	1.3964(14)	1.3986(13)	1.3979(13)	1.3977(8)	1.3948(8)
C1	C6	1.4040(14)	1.4039(13)	1.4044(13)	1.4047(8)	1.4038(7)
C1	C7	1.5245(14)	1.5239(13)	1.5231(13)	1.5229(8)	1.5259(9)
C2	C3	1.4015(14)	1.4005(13)	1.4013(13)	1.4021(8)	1.4027(9)
C2	H2	0.95	1.084	0.95	1.083896	0.950134(3)
C3	C4	1.3920(15)	1.3934(13)	1.3943(14)	1.3923(9)	1.3887(6)
C3	H3	0.95	1.084	0.95	1.083897	0.94987(4)
C4	C5	1.4000(15)	1.4012(14)	1.4007(14)	1.4022(9)	1.4001(8)
C4	H4	0.95	1.084	0.95	1.083902	0.949985(8)
C5	C6	1.3913(15)	1.3914(13)	1.3922(13)	1.3915(8)	1.3914(9)
C5	H5	0.95	1.084	0.95	1.083898	0.949994
C6	H6	0.95	1.084	0.95	1.083898	0.950022
C7	C8	1.5290(14)	1.5294(13)	1.5310(13)	1.5302(9)	1.5278(7)
C7	H7	1	1.098	1	1.104009	0.999998
C8	H8A	0.98	1.091	0.98	1.091702	0.980029(17)
C8	H8B	0.98	1.091	0.98	1.091702	0.98004(2)
C8	H8C	0.98	1.091	0.98	1.091704	0.980019(17)
C9	C10	1.4370(14)	1.4353(12)	1.4360(13)	1.4341(8)	1.4359(5)
C10	C11	1.3802(14)	1.3821(13)	1.3811(13)	1.3809(8)	1.3805(9)
C11	C12	1.3841(14)	1.3856(13)	1.3853(13)	1.3855(8)	1.3824(7)
C11	H11	0.95	1.084	0.95	1.083898	0.949925
C12	C13	1.3999(14)	1.4017(13)	1.3997(13)	1.4004(8)	1.4000(5)
C13	H13	0.95	1.084	0.95	1.083897	0.950004(4)

**Table 7.15:** Bond lengths of different ylid models.

	IAM XL	theo BODD	free BODD	inv XD	IAM XD	
S1	C1	1.7115(5)	1.7110(3)	1.7103(3)	1.7113(2)	1.7119(3)
S1	C10	1.7896(5)	1.7887(3)	1.7897(4)	1.7893(2)	1.7900(4)
S1	C11	1.7980(6)	1.7968(4)	1.7979(4)	1.7974(3)	1.7977(4)
O1	C2	1.2376(6)	1.2355(4)	1.2383(5)	1.2356(3)	1.2380(4)
O2	C9	1.2327(6)	1.2312(4)	1.2336(5)	1.2311(3)	1.2329(5)
C1	C2	1.4359(6)	1.4365(4)	1.4358(5)	1.4362(3)	1.4354(5)
C1	C9	1.4424(6)	1.4428(4)	1.4423(5)	1.4432(3)	1.4422(5)
C2	C3	1.5021(6)	1.5021(4)	1.5018(5)	1.5020(3)	1.5022(5)
C3	C4	1.3839(7)	1.3838(4)	1.3845(5)	1.3848(3)	1.3839(5)
C3	C8	1.3964(7)	1.3977(4)	1.3975(5)	1.3961(3)	1.3948(5)
C4	C5	1.4028(8)	1.4036(5)	1.4046(6)	1.4027(4)	1.4007(6)
C4	H4	0.956(13)	1.084	0.907(9)	1.083898127	0.963419
C5	C6	1.3928(9)	1.3954(6)	1.3959(7)	1.3976(4)	1.3933(7)
C5	H5	0.970(13)	1.084	0.942(9)	1.083892904	0.965195
C6	C7	1.4066(8)	1.4053(5)	1.4075(6)	1.4045(4)	1.4088(6)
C6	H6	1.008(13)	1.084	0.950(9)	1.083906631	1.003359
C7	C8	1.3850(7)	1.3844(4)	1.3850(5)	1.3854(3)	1.3848(5)
C7	H7	0.944(13)	1.084	0.902(9)	1.08389951	0.941992
C8	C9	1.5068(7)	1.5069(4)	1.5063(5)	1.5065(3)	1.5084(5)
C10	H10A	0.930(8)	1.089	0.942(5)	1.088566096	0.932833
C10	H10B	0.930(7)	1.089	0.942(5)	1.088563161	0.932834
C10	H10C	0.930(7)	1.089	0.942(5)	1.088563836	0.932823
C11	H11A	0.936(8)	1.089	0.944(6)	1.088567803	0.942437
C11	H11B	0.936(8)	1.089	0.944(6)	1.088559565	0.942436
C11	H11C	0.936(8)	1.089	0.944(6)	1.088564805	0.942443

**Figure 7.13:** Hirshfeld test results per atom for different models for MBADNP.

**Table 7.16:** U(iso)/(equiv) of different MBADNP models.

	IAM XL	inv2 BODD	free BODD	inv XD	IAM XD
O1	0.01172(13)	0.01124(11)	0.01156(11)	0.01136	0.01171
O2	0.01228(13)	0.01176(12)	0.01208(12)	0.01188	0.01213
O3	0.01127(13)	0.01072(11)	0.01111(11)	0.01068	0.01116
O4	0.01113(12)	0.01062(11)	0.01097(11)	0.01062	0.01118
N1	0.00874(12)	0.00826(11)	0.00859(11)	0.00835	0.00869
N2	0.00886(12)	0.00846(11)	0.00868(11)	0.00839	0.00878
N3	0.00843(12)	0.00791(11)	0.00829(11)	0.00798	0.00843
N4	0.00825(12)	0.00780(11)	0.00814(11)	0.0079	0.0082
C1	0.00814(13)	0.00766(12)	0.00786(12)	0.00762	0.00801
C2	0.00944(14)	0.00916(12)	0.00920(12)	0.00904	0.0094
C3	0.01006(14)	0.00968(13)	0.00987(13)	0.00966	0.01002
C4	0.00993(14)	0.00951(13)	0.00968(13)	0.00947	0.0098
C5	0.01010(14)	0.00968(13)	0.00986(13)	0.00969	0.01014
C6	0.00939(14)	0.00904(12)	0.00909(13)	0.00897	0.00937
C7	0.00792(13)	0.00741(12)	0.00775(12)	0.0075	0.00789
C8	0.01031(15)	0.00982(13)	0.01017(13)	0.00995	0.01027
C9	0.00772(13)	0.00731(11)	0.00751(12)	0.00722	0.00768
C10	0.00785(13)	0.00734(12)	0.00768(12)	0.00731	0.00768
C11	0.00796(13)	0.00763(12)	0.00773(12)	0.00754	0.00794
C12	0.00801(13)	0.00754(12)	0.00777(12)	0.00762	0.00789
C13	0.00888(14)	0.00832(12)	0.00863(12)	0.0084	0.00888
H1N	0.01	0.01	0.01	0.0100(1)	0.0104(1)
H2	0.011	0.011	0.011	0.0108(1)	0.0113(1)
H3	0.012	0.012	0.012	0.0116(1)	0.0120(1)
H4	0.012	0.011	0.012	0.0114(1)	0.0118(1)
H5	0.012	0.012	0.012	0.0116(1)	0.0122(1)
H6	0.011	0.011	0.011	0.0108(1)	0.0112(1)
H7	0.01	0.009	0.009	0.0090(1)	0.0095(1)
H8A	0.015	0.015	0.015	0.0149(1)	0.0154(2)
H8B	0.015	0.015	0.015	0.0149(1)	0.0154(2)
H8C	0.015	0.015	0.015	0.0149(1)	0.0154(2)
H11	0.01	0.009	0.009	0.0090(1)	0.0095(1)
H13	0.011	0.01	0.01	0.0101(1)	0.0107(1)

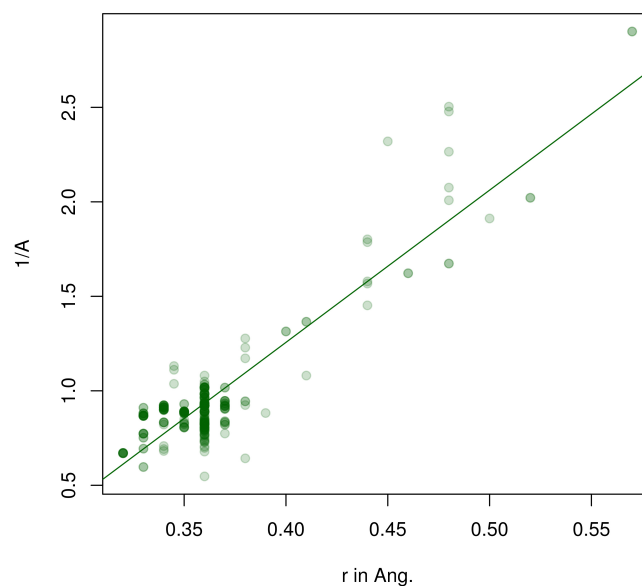
**Table 7.17:** U(iso)/(equiv) of different ylid models.

	XL IAM	theo BODD	free BODD	XD inv	XD IAM
S1	0.00922(2)	0.00878(1)	0.00935(2)	0.00914	0.00923
O1	0.01354(6)	0.01352(4)	0.01370(4)	0.01355	0.01361
O2	0.01425(6)	0.01429(4)	0.01443(5)	0.01438	0.01423
C1	0.00957(6)	0.00956(4)	0.00972(4)	0.00949	0.00958
C2	0.00931(6)	0.00926(4)	0.00944(4)	0.00922	0.00931
C3	0.00996(6)	0.00993(4)	0.01001(4)	0.00986	0.00994
C4	0.01318(7)	0.01315(4)	0.01330(5)	0.01306	0.01327
C5	0.01579(8)	0.01583(5)	0.01592(6)	0.0158	0.01584
C6	0.01624(8)	0.01622(5)	0.01627(6)	0.01619	0.01621
C7	0.01408(8)	0.01402(5)	0.01415(6)	0.01402	0.01424
C8	0.01025(6)	0.01023(4)	0.01033(4)	0.01018	0.01027
C9	0.00964(6)	0.00959(3)	0.00976(4)	0.00953	0.00962
C10	0.01314(7)	0.01317(4)	0.01335(5)	0.01313	0.01319
C11	0.01712(9)	0.01711(5)	0.01727(6)	0.01708	0.01722
H4	0.016	0.016	0.016	0.01567(4)	0.01593(6)
H5	0.019	0.019	0.019	0.01896(5)	0.01901(7)
H6	0.019	0.019	0.02	0.01943(5)	0.01945(8)
H7	0.017	0.017	0.017	0.01682(5)	0.01709(7)
H10A	0.02	0.02	0.02	0.01576(4)	0.01976(6)
H10B	0.02	0.02	0.02	0.01576(4)	0.01976(6)
H10C	0.02	0.02	0.02	0.01576(4)	0.01976(6)
H11A	0.026	0.026	0.026	0.02049(5)	0.02577(8)
H11B	0.026	0.026	0.026	0.02049(5)	0.02577(8)
H11C	0.026	0.026	0.026	0.02049(5)	0.02577(8)

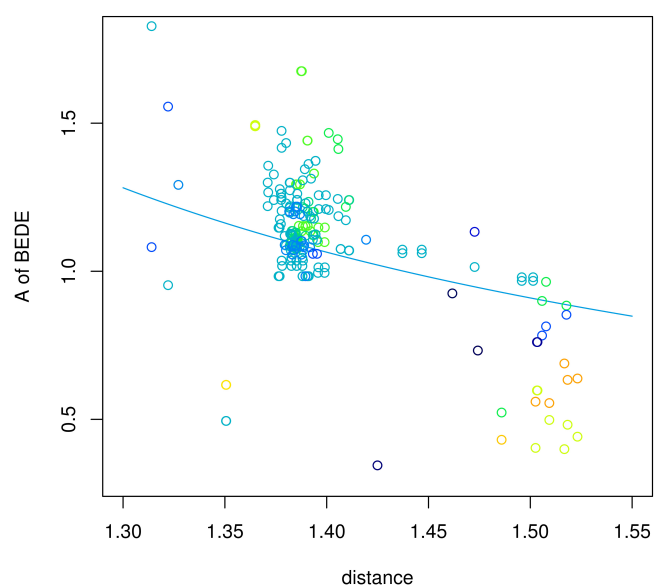
**Table 7.18:** List of molecules for which BEDE and LONE parameters were refined against theoretical data which were included in the statistical study.

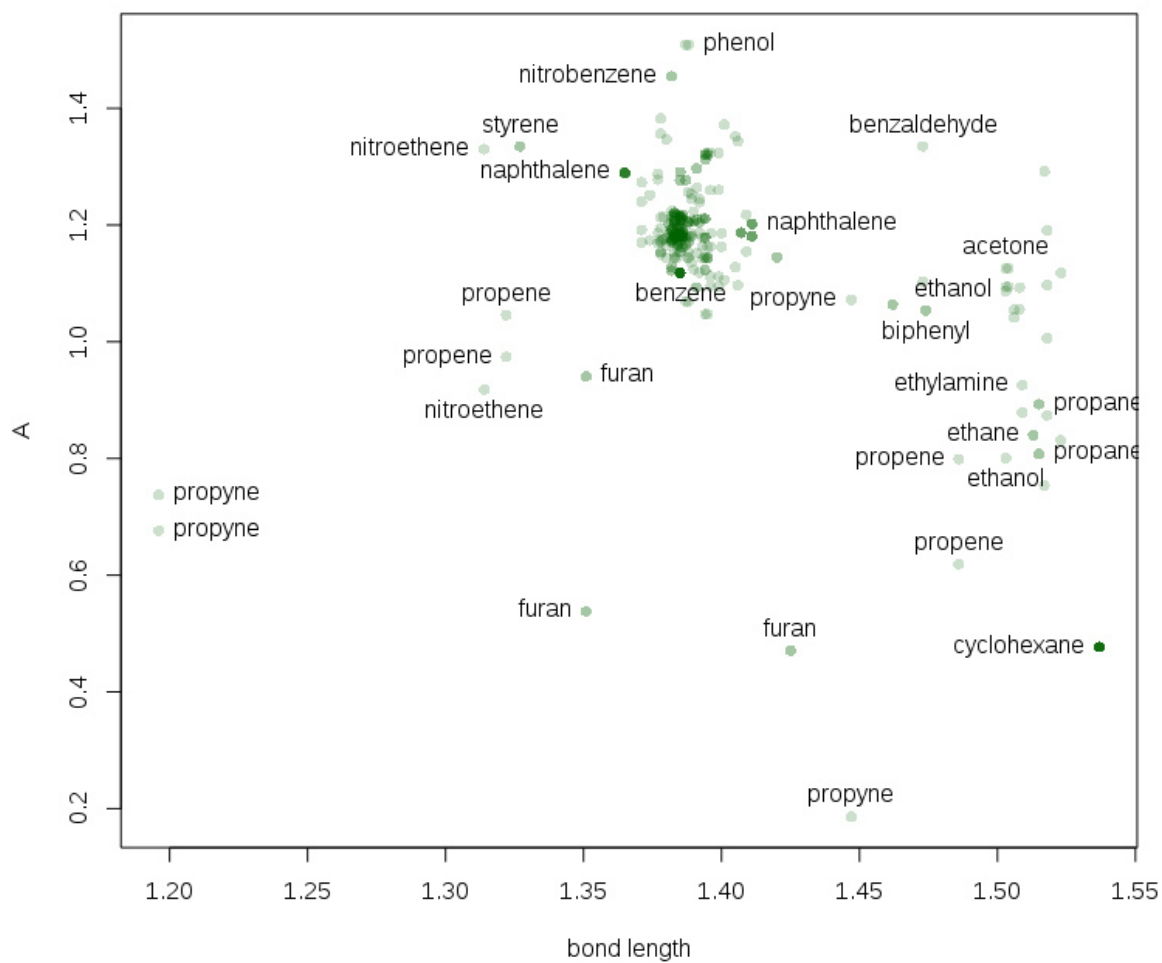
ethane	propane	N-methylaniline
benzene	propene	2-aminopyridine
naphthalene	cyclohexane	1-phenylethanamine
methylamine	ethylamine	m-phenylenediamine
methanol	ethanol	2,3-diaminopyridine
phenol	nitrobenzene	pyridine-2,5-diamine
biphenyl	acetone	methyl-phenyl-amine
benzaldehyde	furan	N-methylpyridin-2-amine
diphenylamine	nitroethene	R-N-(1-phenylethyl)aniline
pyridine	hexafluorobenzene	R-N-(1-phenylethyl)pyridin-2-amine
1-phenoxybenzene	styrene	2-dimethylsulfuranylidene-indan
aniline		

**Figure 7.14:** Scatter plot of  $1/A(\text{BEDE})$  versus  $r$  from bonds between carbon atoms and positive  $r$  from 34 model compounds. The regression line suggests  $1/A = 8.06(21)r - 1.97(8)$  with an adjusted  $R^2=0.818$ .



**Figure 7.15:** Plot of  $A$  (BEDE parameter) versus bond length from bonds between carbon atoms and  $r > 0$  from 34 model compounds. The curve drawn belongs to the equation  $A = 8.06(21) \cdot d - 1.97(8)$  derived from the 3D regression and a mean  $r$  of 0.36. The points are color coded according to their difference to the mean  $r$  value.





**Figure 7.16:** Parameter A versus bond length for bonds between carbon atoms for 31 compounds refined with  $r=0.36\text{\AA}$ ,  $B1=0.11$  and  $B2=0.81$ .

## 7.4 Appendix D – List of programs

Program name	Description, version, author and url:	Applied in this thesis for:
ANTECHAMBER	A set of auxiliary programs for molecular mechanic studies, part of AMBERTOOLS14, University of California, San Francisco. <a href="http://ambermd.org/#AmberTools">http://ambermd.org/#AmberTools</a>	Generation of AM1-BCC charges for angiogenesis inhibitors for comparison to invariom point charges
APBS	An adaptive <i>Poisson-Boltzmann</i> solver, <sup>[203]</sup> plugin for PYMOL, 2006, M. G. Lerner and H. A. Carlson, University of Michigan, Ann Arbor. <a href="http://www.poissonboltzmann.org/apbs/">http://www.poissonboltzmann.org/apbs/</a>	Calculation of <i>Poisson-Boltzmann</i> ESP of trialanine.
APD-TOOLKIT	A crystallography program for estimating hydrogen ADPs based on the invariom database and segmented rigid body analysis, development versions incl. the first ones in Python 3.5, by J. Lübben, University Göttingen. <a href="https://github.com/JLuebben/APD-Toolkit">https://github.com/JLuebben/APD-Toolkit</a>	Appended by several plugins to transfer invariom point charges, which are now distributed together with the program.
AVOGADRO	An open-source molecular builder and visualization tool, <sup>[253]</sup> version 1.0.3. <a href="http://avogadro.openmolecules.net/">http://avogadro.openmolecules.net/</a>	Generation of GAUSSIAN09 input files and pre-optimization of a molecules geometry by molecular mechanics.
UCSF CHIMERA	An extensible molecular modeling system, version 1.10.2, University of California, San Francisco. <a href="http://www.cgl.ucsf.edu/chimera/">http://www.cgl.ucsf.edu/chimera/</a>	Calculation and visualization of ESP of trialanine.
CONQUEST	The primary program for searching and retrieving information from the Cambridge Structural Database, Version 1.17, by Cambridge Crystallographic Data Center Cambridge, <a href="http://www.ccdc.cam.ac.uk">www.ccdc.cam.ac.uk</a>	Searching the Cambridge Structural Database <sup>[77]</sup>
FCF2HKL	G. M. Sheldrick, Georg-August University Göttingen, 2014.	Conversion of structure factors in fcf format to SHELX HKLF 4 format.



---

ESP_MCQ	A program for calculation and display of ESP on a molecular surface, C. M. Wandtke, developed during this thesis.	Calculation and display of ESP on a molecular surface starting from pqr files. Demonstrated for four exemplary molecules.
INVARIOMTOOL	A preprocessor program for aspherical atom modeling with XD using invarioms, <sup>[91]</sup> by C. B. Hübschle & B. Dittrich. <a href="http://ewald.ac.chemie.uni-goettingen.de/programs.html">http://ewald.ac.chemie.uni-goettingen.de/programs.html</a>	Transfer of invariom scattering factors and generation of invariom scattering factor databases. The code was slightly modified during this thesis.
GAUSSIAN09	A theoretical chemistry program for electronic structure modeling, rev. A02 and D01, <sup>[103,235]</sup> Gaussian, Inc., Pittsburgh PA.	DFT computations (geometry optimization and single-point), fitting of point charges to the ESP according to MK, computation of force constants and harmonic vibrational frequencies.
JMOL	An open-source Java viewer for chemical structures in 3D, <sup>[231]</sup> Version 14.2.4, by R. M. Hanson, <a href="http://jmol.sourceforge.net/">http://jmol.sourceforge.net/</a>	Calculation and visualization of the ESP of trialananine.
LAUE-SCRIPT	A Python library for crystallographic data processing, several development versions, by J. Lübben, University Göttingen, 2016. <a href="https://github.com/JLuebben/Laue-Script">https://github.com/JLuebben/Laue-Script</a>	Writing tools to read and analyze SHELXL BODD results.
MATPLOTLIB	A 2D graphics environment <sup>[254]</sup> by J. D. Hunter, Version 1.5.1, <a href="http://matplotlib.org">http://matplotlib.org</a>	In combination with Python 2.7.3 for statistical analyses of invariom point charges and their influence on the ESP.
MOLECOOLQT	A molecule viewer for charge density related science, <sup>[40]</sup> C. B. Hübschle, <a href="http://www.molecoolqt.de/">http://www.molecoolqt.de/</a>	Visualizations of ESP, deformation and difference density maps, adjusting local coordinate systems for multipole modeling in XD. Adjustment of invariom names and the code as basis for ESP_MCQ.

MOLISO	A program for visualization of properties on isosurfaces, <sup>[225]</sup> C. B. Hübschle, Freie Universität Berlin. <a href="http://www.moliso.de/index.html">http://www.moliso.de/index.html</a>	Adaptation of source code for the RMS-TOOL.
NUMPY	The fundamental package for scientific computing with Python, version 1.6.2, NumPy Developers, <a href="http://www.numpy.org/">http://www.numpy.org/</a>	Several small computations, especially for changing point charges by a random amount sampled from a <i>Gaussian</i> distribution.
ORBITAL VIEWER	A program for visualizing atomic orbitals, Version 1.04, D. Manthey, 2004. <a href="http://www.orbitals.com/orb/ov.htm">http://www.orbitals.com/orb/ov.htm</a>	Visualization of spherical harmonics for this thesis.
PDB2PQR	An automated pipeline for the setup of <i>Poisson-Boltzmann</i> electrostatic simulations, <sup>[192,193]</sup> Version 2.0.0, J. E. Nielsen, University College Dublin; N. A. Baker, Pacific Northwest National Laboratory, operated by Battelle Memorial Institute, Pacific Northwest Division for the U.S. Department Energy.; P. Czodrowski & G. Klebe, University of Marburg. <a href="https://sourceforge.net/projects/pdb2pqr/">https://sourceforge.net/projects/pdb2pqr/</a>	Assignment of point charges to trialanine.
PLATON	A multipurpose crystallographic tool, <sup>[111]</sup> version 240314, A. L. Spek, Utrecht University, Utrecht. <a href="http://www.cryst.chem.uu.nl/spek/-platon/">http://www.cryst.chem.uu.nl/spek/-platon/</a>	Computing cartesian coordinates for a whole molecule that sits on a special position, performing the <i>Hirshfeld</i> test and visualization of ADPs in an ORTEP of MBADNP.
PYMOL	Molecular graphics system, <sup>[204]</sup> Version 1.5.0.3. Copyright (c) Schroedinger, LLC. <a href="http://www.pymol.org">http://www.pymol.org</a>	Visualization of ESP on a molecular surface as calculated by APBS
R	A language and environment for statistical computing, R Core team (2012), R Foundation for Statistical Computing, Vienna, Austria. <a href="http://www.R-project.org">http://www.R-project.org</a>	For statistical analyses of BEDE parameters.
RMS-TOOL	A tool for computing the RMS for XD grid files. C. M. Wandtke	Summation of grid points within a shell around a molecule from files generated by XD.

---

SHELXL	A program for refining small molecules as well as macromolecules against single-crystal diffraction data, <sup>[11]</sup> versions 2013 till 2016 and further development versions, G. M. Sheldrick, University Göttingen. <a href="http://shelx.uni-goettingen.de/">http://shelx.uni-goettingen.de/</a>	IAM and BODD refinement of all structures discussed in this thesis with experimental data and those model compounds studied in the BODD project.
SHELXLE	A Qt graphical user interface for SHELXL, <sup>[229]</sup> several versions up to 758, by C. B. Hübschle, <a href="http://ewald.ac.chemie.uni-goettingen.de/shelx/eingabe.php">http://ewald.ac.chemie.uni-goettingen.de/shelx/eingabe.php</a>	Modeling and writing instructions for SHELXL, slightly modified for the display of BODD function positions.
SHELXT	A program for solving small molecule crystal structures. <sup>[221]</sup> by G. M. Sheldrick, University Göttingen. <a href="http://shelx.uni-goettingen.de/">http://shelx.uni-goettingen.de/</a>	Phasing for the crystal structures used as examples for the application of invariom point charges and their visualization.
TONTO	A fortran based object-oriented system for quantum chemistry and crystallography, <sup>[249]</sup> version 3.2 v. 4268, D. Jayatilaka. <a href="http://www.theochem.uwa.edu.au/tonto/">www.theochem.uwa.edu.au/tonto/</a>	Calculation of theoretical ('simulated') X-ray diffraction data starting from GAUSSIAN FChk files.
TPACM4 SERVER	Transferrable partial atomic charge model, B. Jayaram & Co-workers, Supercomputing Facility for Bioinformatics and Computational Biology, IIT Delhi. <a href="http://www.scfbio-iitd.res.in/software/drugdesign/charge.jsp">www.scfbio-iitd.res.in/software/drugdesign/charge.jsp</a>	Assignment of atomic point charges to angiogenesis inhibitor molecules for comparison with invariom point charges.
XD	A computer program package for multipole refinement, topological analysis of charge densities and evaluation of intermolecular energies from experimental or theoretical structure factors, <sup>[12]</sup> version 5.34, 6.03 and a special one for large molecules, T. Koritzanzky, P. Macchi, C. Gatti. L. J. Farrugia, P. R. Mallinson, A. Volkov, T. Richter. University at Buffalo, NY, USA; University of Milano, Italy; University of Glasgow, UK; CNR-ISTM, Milano, Italy; Middle Tennessee State University, TN, USA.	XDINI was used to convert files from SHELX format to XD format, XDLSM was applied for IAM and multipole model refinements against theoretical and experimental XRD data, XDPROP was employed to calculate ESP, ED and deformation density grids, XDGEOM summarized bond lengths and angles, ADDGRID was used for building differences between two property grids.



# Bibliography

- [1] M. Eckert, **2012**: *Max von Laue and the discovery of X-ray diffraction in 1912*, in *Ann. d. Phys. (Berlin)*, **5**, A83–A85.
- [2] W. L. Bragg, **1962**: *The growing Power of X-ray analysis*, in P. P. Ewald (editor), *Fifty Years of X-Ray Diffraction*, chapter 8, International Union of Crystallography, Uetrecht, Netherlands, 120–135.
- [3] J. M. Robertson, **1936**: *X-Ray Study of the Phthalocyanines. Part II. Quantitative Structure determination of the Metal-free Compound.*, in *J. Chem. Soc. (Resumed)*, 1195–1209.
- [4] W. L. Bragg, **1929**: *The Determination of Parameters in Crystal Structures by means of Fourier Series*, in *Proc. Royal Soc. A*, **123**, 537–559.
- [5] H. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. Bhat, H. Weissig, I. Shindyalov, P. Bourne, **2000**: *The Protein Data Bank*, in *Nucleic Acids Res.*, **28**, 235–242.
- [6] J. W. Goodman, **2005**: *Introduction to Fourier Optics*, Roberts and Company Publishers, 3rd edition.
- [7] W. Massa, **2009**: *Kristallstrukturbestimmung*, Teubner Studienbücher, Vieweg + Teubner, Wiesbaden, 6th edition.
- [8] C. Giacovazzo, H. L. Monaco, G. Artioli, D. Viterbo, M. Milanesio, G. Ferraris, G. Gilli, P. Gilli, G. Zanotti, M. Catti, **2011**: *Fundamentals of Crystallography*, number 15 in IUCr Texts on Crystallography, Oxford University Press, Oxford, 3rd edition.
- [9] B. Rupp, **2009**: *Biomolecular Crystallography: Principles, Practice and Application to Structural Biology*, Garland Science, New York.
- [10] P. Müller, R. Herbst-Irmer, A. Spek, T. Schneider, M. Sawaya, **2006**: *Crystal Structure Refinement: A Crystallographer's Guide to SHELXL*, Oxford University Press, New York, 1st edition.
- [11] G. M. Sheldrick, **2015**: *Crystal structure refinement with SHELXL*, in *Acta Cryst. C*, **71**, 3–8.
- [12] A. Volkov, P. Macchi, L. J. Farrugia, C. Gatti, P. Mallinson, T. Richter, T. Koritsánszky, **2006**: *XD2006 – A Computer Program Package for Multipole Refinement, Topological Analysis of Charge Densities and Evaluation of Intermolecular Energies from Experimental or Theoretical Structure Factors*.
- [13] P. Coppens, A. Volkov, **2004**: *The interplay between experiment and theory in charge-density analysis*, in *Acta Cryst. A*, **60**, 357–364.
- [14] R. Destro, L. Loconte, L. L. Presti, P. Roversi, R. Soave, **2004**: *On the role of data quality in experimental charge-density studies*, in *Acta Cryst. A*, **60**, 365–370.
- [15] K. Meindl, J. Henn, **2008**: *Foundations of residual-density analysis*, in *Acta Cryst. A*, **64**, 404–418.
- [16] B. Niepötter, R. Herbst-Irmer, D. Stalke, **2015**: *Empirical correction for resolution- and temperature dependent errors caused by factors such as thermal diffuse scattering*, in *J. Appl. Cryst.*, **48**, 1485–1497.
- [17] H. Wolf, M. R. V. Jørgensen, Y.-S. Chen, R. Herbst-Irmer, D. Stalke, **2015**: *Charge density investigations on [2,2]-paracyclo-phane – in data we trust*, in *Acta Cryst. B*, **71**, 10–19.

- [18] R. F. W. Bader, **1990**: *Atoms in Molecules: A Quantum Theory*, number 22 in The International Series of Monographs on Chemistry, Clarendon Press, Oxford, 1st edition.
- [19] R. F. W. Bader, C. Gatti, **1998**: *A Green's Function for the Density*, in *Chem. Phys. Lett.*, *287*, 233–238.
- [20] C. Gatti, L. Bertini, **2004**: *The local form of the source function as a fingerprint of strong and weak intra- and intermolecular interactions*, in *Acta Cryst. A*, *60*, 423–449.
- [21] B. Silvi, A. Savin, **1994**: *Classification of chemical bonds based on topological analysis of electron localization functions*, in *Nature*, *371*, 683–686.
- [22] M. Kohout, **2004**: *A measure of electron localizability*, in *Int. J. Quant. Chem.*, *97*, 651–658.
- [23] F. R. Wagner, M. Kohout, Y. Grin, **2008**: *Direct Space Decomposition of ELL-D: Interplay of Charge Density and Pair-Volume Function for Different Bonding Situations*, in *J. Phys. Chem. A*, *112*, 9814–9828.
- [24] T. S. Koritsánszky, P. Coppens, **2001**: *Chemical Application of X-ray Charge-Density Analysis*, in *Chem. Rev.*, *101*, 1583–1628.
- [25] P. Coppens, **2005**: *Charge Densities Come of Age*, in *Angew. Chem. Int. Ed.*, *44*, 6810–6811.
- [26] C. Gatti, **2005**: *Chemical bonding in crystals: new directions*, in *Z. Kristallogr.*, *220*, 399–457.
- [27] C. Gatti, P. Macchi (editors), **2012**: *Modern Charge-Density Analysis*, Springer, New York.
- [28] D. Stalke, **2011**: *Meaningful Structural Descriptors from Charge Density*, in *Chem. Eur. J.*, *17*, 9264–9278.
- [29] R. Brill, **1960**: *On the influence of the binding electrons on X-ray intensities*, in *Acta Cryst.*, *13*, 275.
- [30] P. P. Ewald, H. Hönl, **1936**: *Die Röntgeninterferenzen an Diamant als wellenmechanisches Problem. Teil I*, in *Ann. Phys. (Lpz.)*, *25*, 281–307.
- [31] E. Hellner, **1977**: *A Simple Refinement of Density Distributions of Bonding Electrons. I. A Description of the Proposed Method*, in *Acta Cryst. B*, *33*, 3813–3816.
- [32] C. Scheringer, D. Mullen, E. Hellner, **1977**: *A simple refinement of density distributions of bonding electrons. III. Experimental static electron densities for the diborane molecule.*, in *Acta Cryst. A*, *34*, 621–625.
- [33] F. L. Hirshfeld, **1971**: *Difference Densities by Least-Squares Refinement: Fumaramic Acid*, in *Acta Cryst. B*, *27*, 769–781.
- [34] W. H. E. Schwarz, K. Ruedenberg, L. Mensching, **1989**: *Chemical deformation densities. 1. Principles and formulation of quantitative determination*, in *J. Am. Chem. Soc.*, *111*, 6926–6933.
- [35] D. Mullen, **1980**: *Electron-Density Distribution in Urea. A Multipolar Expansion*, in *Acta Cryst. B*, *36*, 1610–1615.
- [36] P. V. Afonine, R. W. Grosse-Kunstleve, P. D. Adams, V. Y. Lunin, A. Urzhumtsev, **2007**: *On macromolecular refinement at subatomic resolution with interatomic scatterers*, in *Acta Cryst. D*, *63*, 1194–1197.
- [37] P. V. Afonine, V. Y. Lunin, N. Muzet, A. Urzhumtsev, **2004**: *On the possibility of the observation of valence electron density for individual bonds in proteins in conventional difference maps*, in *Acta Cryst. D*, *60*, 260–274.
- [38] N. Dadda, A. Nassour, B. Guillot, N. Benali-Cherif, C. Jelsch, **2012**: *Charge-density analysis and electrostatic properties of 2-carboxy-4-methylanilinium chloride monohydrate obtained using a multipolar and a spherical-charges model*, in *Acta Cryst. A*, *68*, 452–463.
- [39] A. Nassour, M. Kubicki, J. Wright, T. Borowiak, G. Dutkiewicz, C. Lecomte, C. Jelsch, **2014**: *Charge-density analysis using multipolar atom and spherical charge models: 2-methyl-*

- 1,3-cyclopentanedione, a compound displaying a resonance-assisted hydrogen bond*, in *Acta Cryst. B*, **70**, 197–211.
- [40] C. B. Hübschle, B. Dittrich, **2011**: *MoleCoolQt – A Molecule Viewer for Charge Density Related Science*, in *J. Appl. Cryst.*, **44**, 238–240.
- [41] F. L. Hirshfeld, **1977**: *Bonded-Atom Fragments for Describing Molecular Charge Densities*, in *Theoret. Chim. Acta (Berl.)*, **44**, 129–138.
- [42] M. A. Spackman, P. G. Byrom, **1997**: *A Novel Definition of a Molecule in a Crystal*, in *Chem. Phys. Lett.*, **267**, 215–220.
- [43] M. A. Spackman, D. Jayatilaka, **2009**: *Hirshfeld surface analysis*, in *CrystEngComm*, **11**, 19–32.
- [44] D. Jayatilaka, B. Dittrich, **2008**: *X-ray Structure Refinement Using Aspherical Atomic Density Functions from Quantum-mechanical Calculations*, in *Acta Cryst. A*, **64**, 383–393.
- [45] S. C. Capelli, H. B. Bürgi, B. Dittrich, S. Grabowsky, D. Jayatilaka, **2014**: *Hirshfeld Atom Refinement*, in *IUCrJ.*, **1**, 361–379.
- [46] D. J. Grimwood, I. Bytheway, D. Jayatilaka, **2003**: *Wave Functions Derived from Experiment. V. Investigation of Electron densities, Electrostatic Potentials, and Electron Localisation Functions for Noncentrosymmetric Crystals*, in *J. Comput. Chem.*, **24**, 470–483.
- [47] A. Genoni, **2013**: *X-ray Constrained Extremely Localized Molecular Orbitals: Theory and Critical Assessment of the New Technique*, in *J. Chem. Theory and Comput.*, **9**, 3004–3019.
- [48] B. Meyer, B. Guillot, M. F. Ruiz-Lopez, A. Genoni, **2016**: *Libraries of Extremely Localized Molecular Orbitals. 1. Model Molecules Approximation and Molecular Orbitals Transferability*, in *J. Chem. Theor. Comp.*, **12**, 1052–1067.
- [49] B. Meyer, B. Guillot, M. F. Ruiz-Lopez, C. Jelsch, A. Genoni, **2016**: *Libraries of Extremely Localized Molecular Orbitals. 2. Comparison with the Pseudoatoms Transferability*, in *J. Chem. Theor. Comp.*, **12**, 1068–1081.
- [50] N. K. Hansen, P. Coppens, **1978**: *Testing Aspherical Atom Refinements on Small-Molecule Data Sets*, in *Acta Cryst. A*, **34**, 909–921.
- [51] R. F. Stewart, **1976**: *Electron Population Analysis with Rigid Pseudoatoms*, in *Acta Cryst. A*, **32**, 565–574.
- [52] D. Manthey, **2004**: *Orbital Viewer, a program for visualizing atomic orbitals*, Technical report, Version 1.04.
- [53] P. Coppens, **1997**: *X-Ray Charge Densities and Chemical Bonding*, number 4 in IUCr Texts on Crystallography, Oxford University Press, Oxford, 1st edition.
- [54] V. G. Tsirelson, R. P. Ozerov, **1996**: *Electron Density and Bonding in Crystals. Principles, Theory and X-ray Diffraction Experiments in Solid State Physics and Chemistry.*, Institute of Physics Publishing, Bristol and Philadelphia.
- [55] C. P. Brock, J. D. Dunitz, F. L. Hirshfeld, **1991**: *Transferability of Deformation Densities among Related Molecules : Atomic Multipole Parameters from Perylene for Improved Estimation of Molecular Vibrations in Naphthalene and Anthracene*, in *Acta Cryst. B*, **47**, 789–797.
- [56] T. Koritsánszky, A. Volkov, P. Coppens, **2002**: *Aspherical-atom scattering factors from molecular wave functions. 1. Transferability and conformation dependence of atomic electron densities of peptides within the multipole formalism*, in *Acta Cryst. A*, **58**, 464–472.
- [57] V. Pichon-Pesme, C. Lecomte, H. Lachekar, **1995**: *On Building a Data Bank of Transferable Experimental Electron Density Parameters: Application to Polypeptides*, in *J. Phys. Chem.*, **99**, 6242–6250.

- [58] S. Domagala, B. Fournier, D. Liebschner, B. Guillot, C. Jelsch, **2012**: *An improved experimental databank of transferable multipolar atom models - ELMAM2. Construction details and applications*, in *Acta Cryst. A*, *68*, 337–351.
- [59] C. Jelsch, V. Pichon-Pesme, C. Lecomte, A. Aubry, **1998**: *Transferability of Multipole Charge-Density Parameters: Application to Very High Resolution Oligopeptide and Protein Structures*, in *Acta Cryst. D*, *54*, 1306–1318.
- [60] B. Zarychta, V. Pichon-Pesme, B. Guillot, C. Lecomte, C. Jelsch, **2007**: *On the application of an experimental multipolar pseudo-atom library for accurate refinement of small-molecule and protein crystal structures*, in *Acta Cryst. A*, *63*, 108–125.
- [61] B. Dittrich, C. B. Hübschle, P. Luger, M. A. Spackman, **2006**: *Introduction and Validation of an Invariom Database for Amino Acid, Peptide and Protein Molecules*, in *Acta Cryst. D*, *62*, 1325–1335.
- [62] A. Volkov, X. Li, T. Koritsánzky, P. Coppens, **2004**: *Ab initio quality electrostatic atomic and molecular properties including intermolecular energies from a transferable theoretical pseudoatom databank*, in *J. Phys. Chem. A*, *108*, 4283–4300.
- [63] J. M. Bąk, P. M. Dominiak, C. C. Wilson, K. Woźniak, **2009**: *Experimental charge-density study of paracetamol – multipole refinement in the presence of a disordered methyl group*, in *Acta Cryst. A*, *65*, 490–500.
- [64] A. Volkov, M. Messerschmidt, P. Coppens, **2007**: *Improving the scattering factor formalism in protein refinement: application of the university at buffalo aspherical-atom databank to polypeptide structures*, in *Acta Cryst. D*, *63*, 160–170.
- [65] B. Dittrich, J. Warren, J. J. McKinnon, **2008**: *Improvement of anisotropic displacement parameters from invariom-model refinements for three L-hydroxylysine structures*, in *Acta Cryst. B*, *64*, 750–759.
- [66] B. Dittrich, C. B. Hübschle, J. J. Holstein, F. P. A. Fabbiani, **2009**: *Towards extracting the charge density from normal-resolution data*, in *J. Appl. Cryst.*, *42*, 1110–1121.
- [67] B. Dittrich, M. Weber, R. Kalinowski, S. Grabowsky, C. B. Hübschle, P. Luger, **2009**: *How to easily replace the independent atom model – the example of bergenin, a potential anti-HIV agent of traditional Asian medicine*, in *Acta Cryst. B*, *65*, 749–756.
- [68] C. Jelsch, B. Guillot, A. Lagoutte, C. Lecomte, **2005**: *Advances in protein and small-molecule charge-density refinement methods using MoPro*, in *J. Appl. Cryst.*, *38*, 38–54.
- [69] B. Dittrich, P. Munshi, M. A. Spackman, **2007**: *Re-determination and invariom model refinement of L-ornithine hydrochloride*, in *Acta Cryst. B*, *63*, 505–509.
- [70] H. D. Flack, **1983**: *On Enantiomorph-Polarity Estimation*, in *Acta Cryst. A*, *39*, 876–881.
- [71] B. Dittrich, M. Strümpel, T. Koritsánzky, M. Schäfer, M. A. Spackman, **2006**: *Invarioms for improved absolute structure determination of light-atom molecular structures*, in *Acta Cryst. A*, *62*, 217–223.
- [72] V. Pichon-Pesme, C. Jelsch, B. Guillot, C. Lecomte, **2004**: *A Comparison between experimental and theoretical aspherical-atom scattering factors for charge-density refinement of large molecules*, in *Acta Cryst. A*, *60*, 204–208.
- [73] A. Volkov, T. Koritsánzky, X. Li, P. Coppens, **2004**: *Response to the paper A comparison between experimental and theoretical aspherical-atom scattering factors for charge-density refinement of large molecules, by Pichon-Pesme, Jelsch, Guillot & Lecomte (2004)*, in *Acta Cryst. A*, *60*, 638–639.
- [74] J. M. Bąk, S. Domagala, C. Hübschle, C. Jelsch, B. Dittrich, P. M. Dominiak, **2011**: *Verification of the structural and electrostatic properties obtained by the use of the different pseudoatom databases.*, in *Acta Cryst. A*, *67*, 141–153.



- [75] J. J. Holstein, **2011**: *Untersuchung der Elektronendichte von Antibiotika in Bezug auf pharmakologische Wirksamkeit*, Ph.D. thesis, Georg-August-Universität Göttingen.
- [76] K. N. Jarzemska, P. M. Dominiak, **2012**: *New version of the theoretical databank of transferable aspherical pseudoatoms, UBDB2011 – towards nucleic acid modelling*, in *Acta Cryst. A*, *68*, 139–147.
- [77] C. R. Groom, I. J. Bruno, M. P. Lightfoot, S. Ward, **2016**: *The Cambridge Structural Database*, in *Acta Cryst. B*, *72*, 171–179.
- [78] P. M. Dominiak, A. Volkov, X. Li, M. Messerschmidt, P. Coppens, **2007**: *A Theoretical Databank of Transferable Aspherical Atoms and Its Application to Electrostatic Interaction Energy Calculations of Macromolecules*, in *J. Chem. Theory and Comput.*, *2*, 232–247.
- [79] J. J. Holstein, C. B. Hübschle, B. Dittrich, **2012**: *Electrostatic properties of nine fluoroquinolone antibiotics derived directly from their crystal structure refinements*, in *CrystEngComm*, 2520–2531.
- [80] B. Dittrich, T. Koritsánszky, P. Luger, **2004**: *A Simple Approach to Nonspherical Electron Densities by Using Invarioms*, in *Angew. Chem. Int. Ed.*, *43*, 2718–2721.
- [81] B. Dittrich, C. B. Hübschle, K. Pröpper, F. Dietrich, T. Stolper, J. J. Holstein, **2013**: *The generalized invariom database (GID)*, in *Acta Cryst. B*, *69*, 91–104.
- [82] V. Schomaker, D. P. Stevenson, **1941**: *Some Revisions of the Covalent Radii and the Additivity Rule for the Lengths of Partially Ionic Single Covalent Bonds*, in *J. Am. Chem. Soc.*, *63*, 37–40.
- [83] R. Blom, A. Haaland, **1985**: *A modification of the Schomaker-Stevenson rule for prediction of single bond distances*, in *J. Mol. Struct.*, *128*, 21–27.
- [84] A. L. Allred, E. G. Rochow, **1958**: *A Scale of Electronegativity Based on Electrostatic Force*, in *J. Inorg. Nucl. Chem.*, *5*, 264–268.
- [85] V. Prelog, G. Helmchen, **1982**: *Basic Principles of the CIP-System and Proposals for a Revision*, in *Angew. Chem. Int. Ed.*, *21*, 567–583.
- [86] G. S. Chandler, M. A. Spackman, **1978**: *Fourier transforms of Gaussian orbital products*, in *Acta Cryst. A*, *34*, 341–343.
- [87] D. Jayatilaka, **1994**: *Fourier transforms of property densities with Gaussian functions*, in *Chem. Phys. Lett.*, *230*, 228–230.
- [88] B. Dittrich, C. B. Hübschle, M. Messerschmidt, R. Kalinowski, D. Girnt, P. Luger, **2005**: *The invariom model and its application: Refinement of D,L-serine at different temperatures and resolution*, in *Acta Cryst. A*, *61*, 314–320.
- [89] S. Parsons, H. D. Flack, T. Wagner, **2013**: *Use of intensity quotients and differences in absolute structure refinement*, in *Acta Cryst. B*, *69*, 249–259.
- [90] J. Thom. H. Dunning, P. J. Hay, **1977**: *Gaussian Basis Sets for Molecular Calculations*, in H. F. S. III (editor), *Methods of Electronic Structure Theory*, chapter 1, Plenum Press, New York and London, 1–27.
- [91] C. B. Hübschle, P. Luger, B. Dittrich, **2007**: *Automation of invariom and of experimental charge density modelling of organic molecules with the preprocessor program InvariomTool*, in *J. Appl. Cryst.*, *40*, 623–627.
- [92] E. D. Stevens, P. Coppens, **1976**: *A Priori Estimates on the Errors in Experimental Electron Densities*, in *Acta Cryst. A*, *32*, 915–917.
- [93] C. B. Hübschle, B. Dittrich, S. Grabowsky, M. Messerschmidt, P. Luger, **2008**: *Comparative experimental electron density and electron localization function study of thymidine based on 20 K X-ray diffraction data*, in *Acta Cryst. B*, *64*, 363–374.

- [94] J. J. Holstein, P. Luger, R. Kalinowski, S. Mebs, C. Paulmann, B. Dittrich, **2010**: *Validation of experimental charge densities: refinement of the macrolide antibiotic roxithromycin*, in *Acta Cryst. B*, *66*, 568–577.
- [95] C. J. Schürmann, K. Pröpper, T. Wagner, B. Dittrich, **2012**: *Invariom modeling of cef-tazidime pentahydrate: molecular properties from a 200 second synchrotron microcrystal experiment*, in *Acta Cryst. B*, *68*, 313–317.
- [96] Y. V. Nelyubina, A. A. Korlyukov, K. A. Lyssenko, **2014**: *Probing Weak Intermolecular Interactions by Using the Invariom Approach: A Comparative Study of S-Tetrazine*, in *Chem. Eur. J.*, *20*, 6978–6984.
- [97] Y. V. Nelyubina, K. A. Lyssenko, **2015**: *Probing Ionic Crystals by the Invariom Approach: An Electron Density Study of Guanidinium Chloride and Carbonate*, in *Chem. Eur. J.*, *21*, 9733–9741.
- [98] K. Pröpper, J. J. Holstein, C. B. Hübschle, C. S. Bond, B. Dittrich, **2013**: *Invariom refinement of a new monoclinic solvate of thioestrepton at 0.64 Å resolution*, in *Acta Cryst. D*, *69*, 1530–1539.
- [99] B. Dittrich, C. M. Wandtke, A. Meents, K. Pröpper, K. C. Mondal, N. Amin, A. Singh, H. W. Roesky, N. Sidhu, **2015**: *Aspherical-Atom Modeling of Coordination Compounds by Single-Crystal X-ray Diffraction Allows the Correct Metal Atom To Be Identified*, in *ChemPhysChem*, *16*, 412–419.
- [100] C. M. Wandtke, J. Lübben, B. Dittrich, **2016**: *Molecular Electrostatic Potentials from Invariom Point Charges*, in *ChemPhysChem*, *17*, 2238–2246.
- [101] W. F. van Gunsteren, H. J. Berendsen, **1990**: *Computer Simulation of Molecular Dynamics: Methodology, Applications, and Perspectives in Chemistry*, in *Angew. Chem. Int. Ed.*, *29*, 992–1023.
- [102] J. Lübben, C. Volkmann, S. Grabowsky, A. Edwards, W. Morgenroth, F. P. A. Fabbiani, G. M. Sheldrick, B. Dittrich, **2014**: *On the temperature dependence of H-Usio in the riding hydrogen model*, in *Acta Cryst. A*, *70*, 309–316.
- [103] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, T. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. B. C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, O. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, D. J. Fox, **2013**: *Gaussian 09, Revision D.01, Technical report*, Gaussian, Inc., Pittsburgh PA.
- [104] Y. Zhao, D. G. Truhlar, **2008**: *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*, in *Theor. Chem. Acc.*, *120*, 215–241.
- [105] F. Weigend, F. Furche, R. Ahlrichs, **2003**: *Gaussian basis sets of quadruple zeta valence quality for atoms H–Kr*, in *J. Chem. Phys.*, *119*, 12753–12762.
- [106] F. Weigend, R. Ahlrichs, **2005**: *Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy*, in *Phys. Chem. Chem. Phys.*, *7*, 3297–3305.

- [107] E. Clementi, C. Roetti, **1974**: *Roothaan-Hartree-Fock Atomic Wavefunctions*, in *At. Data Nucl. Data Tables*, *14*, 177–478.
- [108] E. Clementi, D. L. Raimondi, **1963**: *Atomic Screening Constants from SCF Functions*, in *J. Chem. Phys.*, *38*, 2686–2689.
- [109] Z. Su, P. Coppens, **1998**: *Nonlinear Least-Squares Fitting of Numerical Relativistic Atomic Wave Functions by a Linear Combination of Slater-Type Functions for Atoms with  $Z = 1-36$* , in *Acta Cryst. A*, *54*, 646–652.
- [110] P. Macchi, P. Coppens, **2001**: *Relativistic Analytical Wave Functions and Scattering Factors For Neutral Atoms Beyond Kr and for All Chemically Important Ions Up to  $\Gamma^-$* , in *Acta Cryst. A*, *57*, 656–662.
- [111] A. L. Spek, **2009**: *Structure validation in chemical crystallography*, in *Acta Cryst. D*, *65*, 148–155.
- [112] F. H. Allen, S. Bellard, M. D. Brice, B. A. Cartwright, A. Doubleday, H. Higgs, T. Hummelink, B. G. Hummelink-Peters, O. Kennard, W. D. S. Motherwell, J. R. Rodgers, D. G. Watson, **1979**: *The Cambridge Crystallographic Data Centre: computer-Based Search, Retrieval, Analysis and Display of Information*, in *Acta Cryst. B*, *35*, 2331–2339.
- [113] F. H. Allen, W. D. S. Motherwell, **2002**: *Applications of the Cambridge Structural Database in Organic Chemistry and Crystal Chemistry*, in *Acta Cryst. B*, *58*, 407–422.
- [114] F. H. Allen, I. J. Bruno, **2010**: *Bond lengths in organic and metal-organic compounds revisited: X–H bond lengths from neutron diffraction data*, in *Acta Cryst. B*, *66*, 380–386.
- [115] P. G. Jones, **1984**: *The Determination of Absolute Structure. II. Absolute Configuration from the Cambridge Crystallographic Data Center Files for 1982*, in *Acta Cryst. A*, *40*, 663–668.
- [116] W. D. S. Motherwell, G. P. Shields, F. H. Allen, **2000**: *Graph-Set and Packing Analysis of Hydrogen-Bonded Networks in Polyamide Structures in the Cambridge Structural Database*, in *Acta Cryst. B*, *56*, 857–871.
- [117] W. T. A. Harrison, J. Simpson, M. Weil, **2010**: *Editorial*, in *Acta Cryst. E*, *66*, e1–e2.
- [118] H. Zhong, S.-H. Duan, Y.-P. Hong, M.-L. Li, Y.-Q. Liu, C.-J. Luo, Q.-Y. Luo, S.-Z. Xiao, H.-L. Xie, Y.-P. Xu, X.-M. Yang, X.-R. Zeng, Q. Y. Zhong, **2010**: *Retraction of articles by H. Zhong et al.*, in *Acta Cryst. E*, *66*, e11–e12.
- [119] T. Liu, Y.-X. Wang, Z.-W. Wang, Z.-P. Xie, J. Y. Zhu, **2010**: *Retraction of articles by T. Liu et al.*, in *Acta Cryst. E*, *66*, e13–e14.
- [120] I. E. Office, **2012**: *Retraction of Articles*, in *Acta Cryst. E*, *68*, e16.
- [121] X.-Y. Zhang, **2007**: *Diaquabis(malato- $\kappa^2 O^1, O^2$ )copper(II)*, in *Acta Cryst. E*, *63*, m1254–m1255.
- [122] H.-Q. Liu, **2007**: *Diaquabis(malato- $\kappa^2 O, O'$ )nickel(II)*, in *Acta Cryst. E*, *63*, m2470.
- [123] W.-P. Wu, F.-C. Zeng, Y. Wu, **2007**: *Bis[2-(2-hydroxybenzoylhydrazono)propionato]nickel(II) trihydrate*, in *Acta Cryst. E*, *63*, m2664.
- [124] F. Liu, W.-T. Wu, W.-P. Zhang, F.-Y. Chen, S.-Y. He, **2007**: *Bis(E)-2-[2-(2-hydroxybenzoyl)hydrazono]propanoato- $\kappa^3 O, N, O'$ copper(II) trihydrate*, in *Acta Cryst. E*, *63*, m2450–m2451.
- [125] S.-F. Wang, J.-Y. Xue, L. Shi, H.-L. Zhu, S. W. Ng, **2005**: *(5,5,7,12,12,14-Hexamethyl-1,4,8,11-tetraazacyclotetradecane-1,8-diacetato)nickel(II) tetrahydrate*, in *Acta Cryst. E*, *61*, m1481–m1483.
- [126] S.-F. Wang, B.-F. Ruan, H.-Q. Li, H.-L. Zhu, S. W. Ng, **2005**: *(5,5,7,12,12,14-Hexamethyl-1,4,8,11-tetraazacyclotetradecane-1,8-diacetato)cobalt(II) tetrahydrate*, in *Acta Cryst. E*, *61*, m1484–m1485.

- [127] H.-L. Zhu, S.-C. Shao, J.-L. Ma, X.-Y. Qiu, L. Sun, S. Yang, **2003**: *Dibenzoato-di(2-aminopyridine)nickel(II)*, in *Acta Cryst. E*, *59*, m843–m844.
- [128] W.-Z. Ju, R.-H. Jiao, P. Cao, R.-Q. Fang, **2006**: *Bis(2-aminopyridine)dibenzoatocobalt(II)*, in *Acta Cryst. E*, *62*, m1012–m1013.
- [129] Z.-L. You, **2005**: *[N,N'-Bis(3-methylsalicylidene)-1,3-propanediaminato]nickel(II)*, in *Acta Cryst. E*, *61*, m1637–m1638.
- [130] Y. Chen, **2006**: *6,6'-Dimethyl-2,2'-[propane-1,3-diylbis(nitrilomethylidene)]diphenolato-cobalt(II)*, in *Acta Cryst. E*, *62*, m144–m145.
- [131] L.-G. Wang, **2007**: *Dibromidobis(2-methyl-5-phenyl-s-triazolo[3,4-b][1,3,4]thiadiazole- $\kappa$ N)nickel(II)*, in *Acta Cryst. E*, *63*, m2345.
- [132] H.-Q. Liu, **2007**: *Dibromidobis(6-methyl-3-phenyl-s-triazolo[3,4-b][1,3,4]thiadiazole- $\kappa$ N<sup>1</sup>)copper(II)*, in *Acta Cryst. E*, *63*, m2466.
- [133] X.-F. Zhao, **2007**: *Bis[6-(cyclopentyliminomethyl)-2-methoxyphenolato]nickel(II)*, in *Acta Cryst. E*, *63*, m704–m705.
- [134] H.-Y. Hou, **2007**: *Bis2-[(E)-cyclopentyliminomethyl]-6-methoxyphenolatocopper(II)*, in *Acta Cryst. E*, *63*, m1766.
- [135] Q. Wang, X.-Y. Qiu, **2006**: *Bis[4-bromo-2-(cyclohexyliminomethyl)phenolato]cobalt(II)*, in *Acta Cryst. E*, *62*, m3000–m3001.
- [136] Y.-X. Sun, G. Yang, G. Chen, M. Sun, H.-L. Zhang, **2005**: *Bis(5-bromo-N-cyclohexylsalicylideneaminato)nickel(II): an orthorhombic polymorph*, in *Acta Cryst. E*, *61*, m1103–m1104.
- [137] D.-S. Yang, **2005**: *Bis[4-bromo-2-(cyclohexyliminomethyl)phenolato]copper(II)*, in *Acta Cryst. E*, *61*, m249–m250.
- [138] Z.-L. You, **2005**: *Bis[4-bromo-2-(cyclohexyliminomethyl)phenolato]zinc(II)*, in *Acta Cryst. E*, *61*, m2493–m2494.
- [139] Y.-M. Yang, **2005**: *Dichlorobis[2-(o-tolyyliminomethyl)phenolato]copper(II)*, in *Acta Cryst. E*, *61*, m1616–m1617.
- [140] X.-H. Liu, J.-H. Zeng, **2006**: *trans-Dichlorobis[2-(o-tolyyliminomethyl)phenolato- $\kappa$ O]nickel(II)*, in *Acta Cryst. E*, *62*, m1501–m1503.
- [141] Y.-X. Sun, G.-Z. Gao, R. Zhang, H.-X. Pei, **2005**: *Azido[1-(isobutylaminomethyl)iminomethyl]-2-naphtholato]nickel(II)*, in *Acta Cryst. E*, *61*, m397–m398.
- [142] Q.-Y. Zhu, Y.-J. Wei, F.-W. Wang, **2006**: *Azido1-[2-(dimethylamino)ethyliminomethyl]naphthalenolatocopper(II)*, in *Acta Cryst. E*, *62*, m983–m985.
- [143] Commission for Crystallographic Nomenclature of the International Union of Crystallography, **2016**, *Online Dictionary of Crystallography*, [http://reference.iucr.org/mediawiki/index.php?title=Main\\_Page&oldid=3145i](http://reference.iucr.org/mediawiki/index.php?title=Main_Page&oldid=3145i), visited 2016-08-19.
- [144] H. A. Jahn, E. Teller, **1937**: *Stability of Polyatomic Molecules in Degenerate Electronic States I – Orbital Degeneracy*, in *Proc. Roy. Soc. London*, 220–235.
- [145] W.-Y. Hunag, Z.-L. Chen, K. Wang, X. Zhou, F.-P. Liang, **2012**: *Synthesis, crystal structures, and magnetic properties of three isomorphous helical coordination polymers*, in *Transition Met. Chem.*, *37*, 291–296.
- [146] W.-C. Song, J. Tao, T.-L. Hu, Y.-F. Zeng, X.-H. Bu, **2011**: *Isomorphous tetrazolate MnII and CoII compounds built on  $\Delta$ -chain showing different magnetic behaviors*, in *Dalton Trans.*, *40*, 11955–11959.
- [147] Q.-X. Jia, Y.-Q. Wang, Q. Yue, Q.-L. Wang, E.-Q. Gao, **2008**: *Isomorphous CoII and MnII materials of tetrazolate-5-carboxylate with an unprecedented self-penetrating net and distinct magnetic behaviours*, in *Chem. Commun.*, 4894–4896.

- [148] M. A. Nadeem, M. Bhadbhade, R. Bircher, J. A. Stride, **2010**: *Controlled Synthesis of Isomorphous Coordination Polymers via in Situ Ligand Transformation Reaction: Crystal Structure, Thermal and Magnetic Properties*, in *Cryst. Growth Des.*, *10*, 4060–4067.
- [149] A. Thangavel, M. Wieliczko, C. Scarborough, B. Dittrich, J. Bacsá, **2015**: *An investigation of the electron density of a Jahn–Teller-distorted  $Cr^{II}$  cation: the crystal structure and charge density of hexakis(acetonitrile- $\kappa$ N)-chromium(II) bis(tetraphenylborate) acetonitrile disolvate  $[Cr(C_2H_3N)_6](C_{24}H_{20}B)_2 \cdot 2C_2H_3N$* , in *Acta Cryst. C*, *71*, 936–943.
- [150] E. Prince (editor), **2004**: *International Tables for X-ray Crystallography Volume C: Mathematical, Physical and Chemical Tables*, volume C, Kluwer Academic Publishers, Dordrecht, 3rd edition.
- [151] S. K. Hoffmann, J. Goslar, **1982**: *Crystal Field Theory and EPR Parameters in  $D_{2d}$  and  $C_{2v}$  Distorted Tetrahedral Copper(II) Complexes*, in *J. Solid State Chem.*, *44*, 343–353.
- [152] J. Costamagna, F. Caruso, J. Vargas, V. Manriquez, **1998**: *Planar and tetrahedral coordination in copper(II) complexes induced by bromo substitution*, in *Inorg. Chim. Acta*, *267*, 151–158.
- [153] D. Cinčić, B. Kaitner, **2011**: *Schiff base derived from 2-hydroxy-1-naphthaldehyde and liquid-assisted mechanochemical synthesis of its isostructural Cu(II) and Co(II) complexes.*, in *CrystEngComm*, *13*, 4351–4357.
- [154] M. Amirnasr, A. H. Mahmoudkhani, A. Gorji, S. Dahghanpour, H. R. Bijanzadeh, **2002**: *Cobalt(II), nickel(II), and zinc(II) complexes with bidentate  $N,N'$ -bis( $\beta$ -phenylcinnamaldehyde)-1,2-diiminoethane Schiff base: synthesis and structure*, in *Polyhedron*, *21*, 2733–2742.
- [155] L. Sacconi, M. Ciampolini, **1964**: *Pseudo-tetrahedral Structure of Some  $\alpha$ -Branched Copper(II) Chelates with Schiff Bases*, in *J. Chem. Soc. (Resumed)*, 276–280.
- [156] N. Kitajima, K. Fujisawa, Y. Moro-oka, **1990**: *Tetrahedral Copper(II) Complexes Supported by a Hindered Pyrazolylborane, Formation of the Thiolato Complex, Which Closely Mimics the Spectroscopic Characteristics of the Blue Copper Proteins*, in *J. Am. Chem. Soc.*, *112*, 3210–3212.
- [157] A. F. Holleman, N. Wiberg, **2007**: *Lehrbuch der Anorganischen Chemie*, Walter de Gruyter, Berlin, 102nd edition.
- [158] H. Weinstein, R. Osman, J. P. Green, S. Topiol, **1981**: *Chemical Applications of Atomic and Molecular Electrostatic Potentials*, chapter Electrostatic Potentials as Descriptors of Molecular Reactivity: The Basis for Some Successful Predictions of Biological Activity, Plenum Press, New York, 309–334.
- [159] K. A. Sharp, B. Honig, **1990**: *Electrostatic Interactions in Macromolecules: Theory and Applications*, in *Annu. Rev. Biophys. Biomol. Struct.*, *19*, 301–332.
- [160] P. Politzer, J. S. Murray, **1991**: *Reviews in Computational Chemistry II*, chapter Molecular Electrostatic Potentials and Chemical Reactivity, VCH, 273–312.
- [161] G. Náray-Szabo, G. G. Ferenczy, **1995**: *Molecular Electrostatics*, in *Chem. Rev.*, *95*, 829–847.
- [162] B. Dittrich, J. E. Warren, F. P. A. Fabbiani, W. Morgenroth, B. Corry, **2009**: *Temperature dependence of rotational disorder in a non-standard amino acid from X-ray crystallography and molecular dynamics simulation*, in *Phys. Chem. Chem. Phys.*, *11*, 2601–2609.
- [163] A. L. Spek, **2015**: *PLATON SQUEEZE: a tool for the calculation of the disordered solvent contribution to the calculated structure factors*, in *Acta Cryst. C*, *71*, 9–18.
- [164] R. I. Cooper, A. L. Thompson, D. J. Watkin, **2010**: *CRYSTALS enhancements: dealing with hydrogen atoms in refinement*, in *J. Appl. Cryst.*, *43*, 1100–1107.
- [165] E. C. Hayden, **2009**: *Cutting of cancer's supply lines*, in *Nature*, *458*, 686–687.

- [166] J. Folkman, **2007**: *Angiogenesis: an organizing principle for drug discovery?*, in *Nature Rev.*, *6*, 273–286.
- [167] K. B. Wiberg, P. R. Rablen, **1993**: *Comparison of Atomic Charges Derived via Different Procedures*, in *J. Comput. Chem.*, *14*, 1504–1518.
- [168] R. S. Mulliken, **1955**: *Electronic Population Analysis on LCAO-MO Molecular Wave Functions. I*, in *J. Chem. Phys.*, *23*, 1833–1840.
- [169] S. R. Cox, D. E. Williams, **1981**: *Representation of the Molecular Electrostatic Potential by Net Atomic Charge Model*, in *J. Comput. Chem.*, *2*, 304–323.
- [170] A. E. Reed, F. Weinhold, **1985**: *Natural localized molecular orbitals*, in *J. Phys. Chem.*, *83*, 1736–1740.
- [171] P. Coppens, T. N. G. Row, P. Leung, E. D. Stevens, P. J. Becker, Y. W. Yang, **1979**: *Net Atomic Charges and Molecular Dipole Moments from Spherical-Atom X-ray Refinements, and the Relation Between Atomic Charge and Shape*, in *Acta Cryst. A*, *35*, 63–72.
- [172] L. E. Chirlian, M. M. Francl, **1987**: *Atomic Charges Derived from Electrostatic Potentials: a Detailed Study*, in *J. Comput. Chem.*, *8*, 894–901.
- [173] C. I. Bayly, P. Cieplak, W. D. Cornell, P. A. Kollman, **1993**: *A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges: The RESP Model*, in *J. Phys. Chem.*, *97*, 10269–10280.
- [174] M. Miller Francl, C. Carey, L. E. Chirlian, D. M. Gange, **1996**: *Charges Fit to Electrostatic Potentials. II. Can Atomic Charges Be Unambiguously Fit to Electrostatic Potentials?*, in *J. Comput. Chem.*, *17*, 367–383.
- [175] E. Sigfridsson, U. Ryde, **1998**: *Comparison of Methods for Deriving Atomic Charges from the Electrostatic Potential and Moments*, in *J. Comput. Chem.*, *19*, 377–395.
- [176] A. Jakalian, B. L. Bush, D. B. Jack, C. I. Bayly, **2000**: *Fast, Efficient Generation of High-Quality Atomic Charges. AM1-BCC Model: I. Method*, in *J. Comput. Chem.*, *21*, 132–146.
- [177] D. Case, I. T.E. Cheatham, T. Darden, H. Gohlke, R. Luo, J. K.M. Merz, A. Onufriev, C. Simmerling, B. Wang, R. Woods, **2005**: *The AMBER Biomolecular Simulation Programs*, in *J. Comput. Chem.*, *26*, 1668–1688.
- [178] J. A. D. MacKerell, D. Bashford, M. Bellott, R. L. Dunbrack, Jr., J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin, M. Karplus, **1998**: *All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins*, in *J. Phys. Chem. B*, *102*, 3586–3616.
- [179] B. Hess, C. Kutzner, D. van der Spoel, E. Lindahl, **2008**: *GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation*, in *J. Chem. Theory Comp.*, *4*, 435–447.
- [180] G. Mukherjee, N. Patra, P. Narkua, B. Jayaram, **2011**: *A Fast Empirical GAFF Compatible Partial Atomic Charge Assignment Scheme for Modeling Interactions of Small Molecules with Biomolecular Targets*, in *J. Comput. Chem.*, *32*, 893–907.
- [181] B. H. Besler, K. M. Merz, P. A. Kollman, **1990**: *Atomic Charges Derived from Semiempirical Methods*, in *J. Comput. Chem.*, *11*, 431–439.
- [182] L. Wroblewska, J. Skolnick, **2007**: *Can a physics-based, all-atom potential find a protein's native structure among misfolded structures? I. Large scale AMBER benchmarking*, in *J. Comput. Chem.*, *28*, 2059–2066.
- [183] R. W. Pastor, A. D. M. Jr., **2011**: *Development of the CHARMM Force Field for Lipids*, in *J. Phys. Chem. Lett.*, *13*, 1526–1532.

- [184] J. Wang, W. Wang, P. A. Kollman, D. A. Case, **2006**: *Automatic atom type and bond type perception in molecular mechanical calculations*, in *J. Mol. Graph. Mod.*, *25*, 247–260.
- [185] U. C. Singh, P. A. Kollman, **1984**: *An Approach to Computing Electrostatic Charges for Molecules*, in *J. Comp. Chem.*, *5*, 129–145.
- [186] D. L. Strout, G. E. Scuseria, **1995**: *A quantitative study of the scaling properties of the Hartree-Fock method*, in *J. Phys. Chem.*, *102*, 8448–8452.
- [187] M. J. S. Dewar, E. G. Zoebisch, E. F. Healy, J. J. P. Stewart, **1985**: *Development and use of quantum mechanical molecular models. 76. AM1: a new general purpose quantum mechanical molecular model*, in *J. Am. Chem. Soc.*, *107*, 3902–3909.
- [188] J. W. Storer, D. J. Giesen, C. J. Cramer, D. G. Truhlar, **1995**: *Class IV charge models: A new semiempirical approach in quantum chemistry*, in *J. Comput. Aid. Mol. Des.*, 87–160.
- [189] J. Li, T. Zhu, C. J. Cramer, D. G. Truhlar, **1998**: *New Class IV Charge Model for extracting Accurate Partial Charges from Wavefunctions*, in *J. Phys Chem. A*, *102*, 1820–1831.
- [190] W. Thiel, **2014**: *Semiempirical quantum-chemical methods*, in *WIREs Comput. Mol. Sci.*, *4*, 145–157.
- [191] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Collman, D. A. Case, **2004**: *Development and testing of a general AMBER force field*, in *J. Comput. Chem.*, *25*, 1157–1174.
- [192] T. J. Dolinsky, J. E. Nielsen, J. A. McCammon, N. A. Baker, **2004**: *PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatic simulations*, in *Nucleic Acids Res.*, *32*, W665–W667.
- [193] T. J. Dolinsky, P. Czodrowski, H. Li, J. E. Nielsen, J. H. Jensen, G. Klebe, N. A. Baker, **2007**: *PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations*, in *Nucleic Acids Res.*, *35*, W522–W525.
- [194] F.-Y. Dupradeau, C. Cezard, R. Lelong, E. Stanislawiak, J. Pecher, J. Delepine, P. Cieplak, **2008**: *R.E.DD.B.: a database for RESP and ESP charge values and force field libraries*, in *Nucleic Acids Res., Database issue*, D360–D367.
- [195] E. Stanislawiak, G. Caron, **2016**, *RESP ESP charge DDataBase home page*, <http://upjv.q4md-forcefieldtools.org/REDDB/index.php>, visited 2016-10-17.
- [196] A. J. McCoy, V. C. Epa, P. M. Colman, **1997**: *Electrostatic Complementarity at Protein/Protein Interfaces*, in *J. Mol. Biol.*, *268*, 570–584.
- [197] N. Muzet, B. Guillot, C. Jelsch, E. Howard, C. Lecomte, **2003**: *Electrostatic complementarity in an aldose reductase complex from ultra-high-resolution crystallography and first principle calculations*, in *Proc. Nat. Acad. Sci. USA*, *100*, 8742–8747.
- [198] M. A. Spackman, J. J. McKinnon, D. Jayatilaka, **2008**: *Electrostatic potentials mapped on Hirshfeld surfaces provide direct insight into intermolecular interactions in crystals*, in *CrystEngComm*, *10*, 377–388.
- [199] M. W. Shi, A. N. Sobolev, T. Schirmeister, B. Engels, T. C. Schmidt, P. Luger, S. Mebs, B. Dittrich, Y.-S. Chen, J. M. Bak, D. Jayatilaka, C. S. Bond, M. Turner, S. G. Stewart, M. A. Spackman, S. Grabowsky, **2015**: *Electrostatic complementarity in pseudoreceptor modeling based on drug molecule crystal structures: the case of loxistatin acid (E64c)*, in *New J. Chem.*, *39*, 1628–1633.
- [200] M. Nix, C. Kaelin, R. Palomino, J. Miller, G. Barsh, G. Millhauser, **2015**: *Electrostatic Similarity Analysis of Human  $\beta$ -Defensin Binding in the Melanocortin System*, in *Biophys. J.*, *109*, 1946–1958.
- [201] A. C. Pan, D. W. Borhani, R. O. Dror, D. E. Shaw, **2013**: *Molecular determinants of drug–receptor binding kinetics*, in *Drug Discov. Today*, *18*, 667–673.
- [202] R. O. Dror, H. F. Green, C. Valant, D. W. Borhani, J. R. Valcourt, A. C. Pan, D. H. Arlow, M. Canals, J. R. Lane, R. Rahmani, J. B. Baell, P. M. Sexton, A. Christopoulos, D. E. Shaw,

- 2013: *Structural basis for modulation of a G-protein-coupled receptor by allosteric drugs*, in *Nature*, 503, 295–299.
- [203] N. A. Baker, D. Sept, S. Joseph, M. J. Holst, J. A. McCammon, 2001: *Electrostatics of nanosystems: application to microtubules and the ribosome*, in *Proc. Natl. Acad. Sci.*, 98, 10037–10041.
- [204] Schroedinger LLC: *The PyMol Molecular Graphics System, Version 1.5.0.4, Technical report*.
- [205] D. Sitkoff, K. a. Sharp, B. Honig, 1994: *Accurate Calculation of Hydration Free Energies Using macroscopic Solvent Models*, in *J. Phys. Chem.*, 98, 1978–1988.
- [206] C. L. Tang, E. Alexov, A. M. Pyle, B. Honig, 2007: *Calculation of pKas in RNA: on the structural origins and functional roles of protonated nucleotides*, in *J. Mol. Biol.*, 5, 21475–1496.
- [207] P. Czodrowki, I. Dramburg, C. A. Sotriffer, G. Klebe, 2006: *Development, validation, and application of adapted PEOE charges to estimate pKa values of functional groups in protein-ligand complexes*, in *Proteins*, 65, 424–437.
- [208] J. M. J. Swanson, J. A. Wagoner, N. A. Baker, J. A. McCammon, 2007: *Optimizing the Poisson Dielectric Boundary with Explicit Solvent Forces and Energies: Lessons Learned with Atom-Centered Dielectric Functions*, in *J. Chem. Theory and Comput.*, 3, 170–183.
- [209] C. Tan, L. Yang, R. Luo, 2006: *How well does Poisson-Boltzmann implicit solvent agree with explicit solvent? A quantitative analysis*, in *J. Phys. Chem. B*, 110, 18680–18687.
- [210] P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, 2010: *Features and development of Coot*, in *Acta Cryst. D*, 66, 486–501.
- [211] S. McNicholas, E. Potterton, K. S. Wilson, M. E. M. Noble, 2011: *Presenting your structures: the CCP4mg molecular-graphics software*, in *Acta Cryst. D*, 67, 386–3943.
- [212] M. D. Winn, C. C. Ballard, K. D. Cowtan, E. J. Dodson, P. Emsley, P. R. Evans, R. M. Keegan, E. B. Krissinel, A. G. W. Leslie, A. McCoy, S. J. McNicholas, G. N. Murshudov, N. S. Pannu, E. A. Potterton, H. R. Powell, R. J. Read, A. Vagin, K. S. Wilson, 2011: *Overview of the CCP4 suite and current developments*, in *Acta Cryst. D*, 67, 235–242.
- [213] F. Fogolari, A. Brigo, H. Molinari, 2002: *The Poisson–Boltzmann equation for biomolecular electrostatics: a tool for structural biology*, in *J. Mol. Recognit.*, 15, 377–392.
- [214] R. F. Stewart, B. M. Craven, 1993: *Molecular Electrostatic Potentials from Crystal Diffraction: The Neurotransmitter  $\gamma$ -Aminobutyric Acid*, in *Biophys. J.*, 65, 998–1005.
- [215] B. Dittrich, R. Flaig, T. Koritsánszky, H.-G. Krane, W. Morgenroth, P. Luger, 2000: *Topological Properties of the Peptid Bond in Glycyl-L-threonine Dihydrate Based on a Fast Synchrotron/CCD-Diffraction Experiment*, in *Chem. Eur. J.*, 6, 2582–2589.
- [216] S. T. Howard, M. B. Hursthouse, C. W. Lehmann, P. R. Mallinson, C. S. Frampton, 1992: *Experimental and Theoretical Study of the Charge Density in 2-Methyl-4-nitroaniline*, in *J. Chem. Phys.*, 97, 5616–5630.
- [217] R. Dovesi, R. Orlando, A. Erba, C. M. Z.-W. B. Civalleri, S. Casassa, L. Maschio, M. Ferrabone, M. D. L. Pierre, P. D’Arco, Y. N. M. Causa, M. Rerat, B. Kirtman, 2014: *CRYSTAL14: A program for the ab initio investigation of crystalline solids*, in *Int. J. Quantum. Chem.*, 114, 1287–1317.
- [218] M. J. Turner, S. P. Thomas, M. W. Shi, D. Jayatilaka, M. A. Spackman, 2015: *Energy frameworks: insights into interaction anisotropy and the mechanical properties of molecular crystals*, in *Chem. Comm.*, 51, 3735–3738.
- [219] M. J. Turner, S. Grabowski, D. Jayatilaka, M. A. Spackman, 2014: *Accurate and Efficient Model Energies for Exploring Intermolecular Interactions in Molecular Crystals*, in *J. Phys. Chem. Lett.*, 5, 4249–4255.



- [220] R. F. Stewart, E. R. Davidson, W. T. Simpson, **1965**: *Coherent X-Ray Scattering for the Hydrogen Atom in the Hydrogen Molecule*, in *J. Chem. Phys.*, **42**, 3175–3187.
- [221] G. M. Sheldrick, **2015**: *SHELXT - Integrated space-group and crystal structure determination*, in *Acta Cryst. A*, **71**, 3–8.
- [222] B. Dittrich, J. Lübben, S. Mebs, A. Wagner, P. Luger, R. Flaig, **2017**: *Accurate bond distances to hydrogen atoms from single-crystal x-ray diffraction by including estimated hydrogen ADPs and comparison to neutron and QM/MM benchmarks*, in *Eur. J. Chem.*, accepted for publication.
- [223] P. Politzer, J. S. Murray, Z. Preralta-Inga, **2001**: *Molecular Surface Electrostatic potentials in Relation to Noncovalent Interactions in Biological Systems*, in *Int. J. Quant. Chem.*, **85**, 676–684.
- [224] D. E. Williams, **1991**: *Net Atomic Charge and Multipole Models for ab Initio Molecular Electric Potential*, in K. B. Lipkowitz, D. B. Boyd (editors), *Reviews in Computational Chemistry II*, chapter 6, VCH, 219–271.
- [225] C. B. Hübschle, P. Luger, **2006**: *Moliso, a Program for Visualization of Properties on Isosurfaces*, in *J. Appl. Cryst.*, **39**, 901–904.
- [226] A. D. Becke, **1988**: *Density-Functional Exchange-Energy Approximation with Correct Asymptotic Behavior*, in *Phys. Rev. A*, **38**, 3098–3100.
- [227] M. M. Francl, W. J. Pietro, W. J. Hehre, J. S. Binkley, M. S. Gordon, D. J. DeFrees, J. A. Pople, **1982**: *Self-consistent molecular orbital methods. XXIII. A polarization-type basis set for second-row elements*, in *J. Chem. Phys.*, **77**, 3654–3665.
- [228] A. Volkov, H. F. King, P. Coppens, L. J. Farrugia, **2006**: *On the calculation of the electrostatic potential, electric field and electric field gradient from the aspherical pseudoatom model*, in *Acta Cryst. A*, **62**, 400–408.
- [229] C. B. Hübschle, G. M. Sheldrick, B. Dittrich, **2011**: *ShelXle: a Qt graphical user interface for SHELXL*, in *J. Appl. Cryst.*, **44**, 1281–1284.
- [230] J. Lübben, L. J. Bourhis, B. Dittrich, **2015**: *Estimating temperature-dependent anisotropic displacements with the invariom database and a new rigid-body analysis program*, in *J. Appl. Cryst.*, **48**, 1785–1793.
- [231] R. M. Hanson, **2010**: *Jmol - a paradigm shift in crystallographic visualization*, in *J. Appl. Cryst.*, **43**, 1250–1260.
- [232] I. Free Software Foundation, **2011**: *BASH - GNU Bourne-Again SHell*, Technical report, Free Software Foundation, Inc., [www.gnu.org/software/bash/](http://www.gnu.org/software/bash/).
- [233] A. V. Aho, B. W. Kernighan, P. J. Weinberger, **1979**: *Awk - a pattern scanning and processing language*, in *Software: Practice and Experience*, **9**, 267–279.
- [234] M. Mantina, A. C. Chamberlin, R. Valero, C. J. Cramer, D. G. Truhlar, **2009**: *Consistent van der Waals Radii for the Whole Main Group*, in *J. Phys. Chem.*, **19**, 5806–5812.
- [235] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels,

- O. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, D. J. Fox, **2009**: *Gaussian 09, Revision A.02, Technical report*, Gaussian, Inc., Wallingford CT.
- [236] M. Malinska, K. N. Jarzemska, A. M. Goral, A. Kutner, K. Woźniak, P. M. Dominiak, **2014**: *Sunitinib: from charge-density studies to interaction with proteins*, in *Acta Cryst. D*, **70**, 1257–1270.
- [237] A. Stash, V. Tsirelson, **2002**: *WinXPRO: A Program for Calculating Crystal and Molecular Properties Using Multipole Parameters of the Electron Density*, in *J. Appl. Cryst.*, **35**, 371–373.
- [238] V. Petricek, M. Dusek, L. Palatinus, **2006**: *JANA2006. The crystallographic computing system.*, *Technical report*, Institute of Physics, Praha, Czech Republic.
- [239] R. J. Gillespie, **1963**: *The Valence-Shell Electron-Pair Repulsion (VSEPR) Theory of Directed Valency*, in *J. of Chem. Edu.*, **6**, 295–301.
- [240] R. J. Gillespie, **1970**: *The Electron-Pair Repulsion Model for Molecular Geometry*, in *J. of Chem. Edu.*, **47**, 18–23.
- [241] R. Brill, H. Dietrich, H. Dierks, **1971**: *Die Verteilung der Bindungselektronen im Dekaboran-Molekül ( $B_{10}H_{14}$ )*, in *Acta Cryst. B*, **27**, 2003–2018.
- [242] D. Mullen, E. Hellner, **1977**: *A Simple Refinement of Density Distributions of Bonding Electrons. II. Bond Electron Density Distribution in Diborane*, in *Acta Cryst. B*, **33**, 3816–3822.
- [243] D. Mullen, E. Hellner, **1978**: *A Simple Refinement of Density Distributions of Bonding Electrons. V. Bond Electron Density Distribution in Urea,  $CO(NH_2)_2$ , at 123 K*, in *Acta Cryst. B*, **33**, 1624–1627.
- [244] D. Mullen, E. Hellner, **1978**: *A Simple Refinement of Density Distributions of Bonding Electrons. IX. Bond Electron Density Distribution in Thiourea,  $CS(NH_2)_2$ , at 123 K*, in *Acta Cryst. B*, **33**, 2789–2794.
- [245] H. Dietrich, C. Scheringer, H. Meyer, K.-W. Schulte, A. Schweig, **1979**: *Refinement of the molecular charge distribution in cyanuric acid. Comparison of observed and calculated electron densities. X. Theoretical static difference densities at experimental resolution and theoretical dynamic difference densities perpendicular to the molecular plane for cyanuric acid*, in *Acta Cryst. B*, **35**, 1191–1197.
- [246] A. Ø. Madsen, S. Mason, S. Larsen, **2003**: *A neutron diffraction study of xylitol: derivation of mean square internal vibrations for H atoms from a rigid-body description*, in *Acta Cryst. B*, **59**, 653–663.
- [247] M.-D. Şerb, I. Kalf, U. Englert, **2014**: *Biguanide and squaric acid as pH-dependent building blocks in crystal engineering*, in *CrystEngComm*, **16**, 10631–10639.
- [248] J. M. Cole, A. E. Goeta, J. A. K. Howard, G. J. McIntyre, **2002**: *X-ray and neutron diffraction studies of the non-linear optical compounds MBANP and MBADNP at 20 K: charge-density and hydrogen-bonding analyses*, in *Acta Cryst. B*, **58**, 690–700.
- [249] D. Jayatilaka, D. J. Grimwood, **2003**: *Tonto: A Fortran Based Object-Oriented System for Quantum Chemistry and Crystallography*, in *Computational Science - ICCS 2003*, 2660, 142–151.
- [250] I. A. Guzei, G. A. Bikzhanova, L. C. Spencer, T. V. Timofeeva, T. L. K. and Charles F. Campana, **2008**: *Polymorphism and History of 2-Dimethylsulfuranylidene-1,3-indanedione (YLID)*, in *Cryst. Growth Des.*, **8**, 2411–2418.
- [251] L. Pauling, **1948**: *The Nature of the Chemical Bond*, Cornell University Press, Ithaca, New York, 2nd edition.
- [252] I. D. Brown, R. R. Shannon, **1973**: *Empirical Bond-Strength–Bond-Length curves for Oxides*, in *Acta Cryst. A*, **29**, 266–282.

- [253] M. D. Hanwell, D. E. Curtis, D. C. Lonie, T. Vandermeersch, E. Zurek, G. R. Hutchison, **2012**: *Avogadro: An advanced semantic chemical editor, visualization, and analysis platform*, in *J. of Cheminform.*, *4*, 17.
- [254] J. D. Hunter, **2007**: *Matplotlib: A 2D graphics environment*, in *Comput. Sci. Eng.*, *9*, 90–95.



# List of Figures

1.1	Basic concept of single crystal X-ray diffraction. . . . .	1
1.2	<i>Fourier</i> transformation of the spherical ED of carbon to give its atomic form factor. . . . .	2
1.3	Deformation density of isopentane. . . . .	5
1.4	Representation of spherical harmonics up to $l = 4$ . . . . .	7
1.5	Invariom assignment for serine with invariom names and model compounds. . . . .	10
1.6	General treatment of invariom model compounds. . . . .	11
1.7	Overview of the invariom database, its functionality and terminology. . . . .	14
2.1	Additional treatment for every model compound in the database during makeover. . . . .	18
2.2	R(F) difference for invariom refinements with scattering factors from the old and new database. . . . .	20
2.3	Number of halogen compounds in the database before and after bromine addition. . . . .	21
2.4	Number of model compounds in the database at start and end of this thesis. . . . .	22
3.1	Graphical representation of the 'whole-molecule' work flow. . . . .	28
3.2	Thermal ellipsoid plot of pair <b>1</b> with Cu. . . . .	32
3.3	Comparison of R(F) for the refinements of the different metal atoms with different ED models against the two data sets of pair <b>1</b> . . . . .	33
3.4	Thermal ellipsoid plot of pair <b>2</b> with Cu. . . . .	34
3.5	Comparison of R(F) for the refinements of the different metal atoms with different ED models against the two data sets of pair <b>2</b> . . . . .	35
3.6	Comparison of R(F) for the refinements of the different metal atoms with different ED models against the two data sets of pair <b>3</b> . . . . .	37
3.7	Thermal ellipsoid plot of pair <b>3</b> with Ni. . . . .	38
3.8	Thermal ellipsoid plot of pair <b>4</b> with Ni. . . . .	40
3.9	Comparison of R(F) for the refinements of the different metal atoms with different ED models against the two data sets of pair <b>4</b> . . . . .	41
3.10	Comparison of R(F) for the refinements of the different metal atoms with different ED models against the two data sets of pair <b>5</b> . . . . .	43
3.11	Comparison of R(F) for the refinements of the different metal atoms with different ED models against the two data sets of pair <b>6</b> . . . . .	45
3.12	Comparison of R(F) for the refinements of the different metal atoms with different ED models against the two data sets of pair <b>7</b> . . . . .	47
3.13	Comparison of R(F) for the refinements of the different metal atoms with different ED models against the data sets of pairs <b>8</b> and <b>9</b> . . . . .	49
3.14	Comparison of R(F) for the refinements of the different metal atoms with different ED models against the two data sets of pair <b>10</b> . . . . .	53
3.15	Comparison of R(F) for the refinement of the different metal atoms with different ED models against the two data sets of pair <b>11</b> . . . . .	55
4.1	Overview of program flows to an ESP mapped on a molecular surface for trialanine. . . . .	64

## LIST OF FIGURES

---

4.2	Schematic illustration of alternative invariom names and their influence on averaging of point charges. . . . .	66
4.3	Molecular envelop of grid points for the evaluation of RMS and RRMS. . . . .	68
4.4	Schematic grid fillings within ESP_MCQ. . . . .	71
4.5	Quick method for filling the ED grid in ESP_MCQ. . . . .	72
4.6	ESP of three homotriptides derived from different point charges mapped to a molecular surface. . . . .	74
4.7	ESP derived from invariom point charges of tri-L-tryptophane when all three double bonded oxygen atom charges had been transferred as suggested by the invariom name O2c. . . . .	74
4.8	Detailed RRMS results of the internal test for each angiogenesis inhibitor. . . . .	77
4.9	Mean RRMS results from the internal invariom averaging tests. . . . .	78
4.10	Frequency distribution of invariom frequencies within the invariom database. . . . .	79
4.11	Frequency distribution for charges of the most frequent hydrogen invariom. . . . .	80
4.12	Frequency distribution of invariom point charge STDEV. . . . .	81
4.13	Frequency distributions of RMS of the ESP from charges disturbed by different random amounts. . . . .	82
4.14	Frequency distributions of RMS of the ESP from disturbing hydrogen and non-hydrogen atom charges. . . . .	82
4.15	Frequency distribution of STDEV for those invarioms of the database that occur more than once. . . . .	83
4.16	Frequency distributions of RMS of the ESP in e/Å with randomly disturbed charges of the random distribution minus the ESP from undisturbed charges. The spread of the disturbance was the median of the STDEV from averaging equal invarioms in the database. . . . .	83
4.17	Comparison of RRMS results averaged for 19 angiogenesis inhibitors for ESP calculated from molecule specific point charges (AM1-BCC) and look-up tables (invariom and TPACM4). . . . .	85
4.18	Comparison of ESPs from different theoretical approaches compared to experimental multipole-model results from refinement against high resolution X-ray data for sunitinib malate and ciprofloxacin. . . . .	86
4.19	Investigations of the charge transfer between sunitinib and malate. . . . .	88
4.20	Invariom point charge assignment, quick ESP computation and display of the asymmetric unit of compounds that were determined during the PhD. . . . .	90
5.1	Basic concept of BEDE instruction for a standard bond. . . . .	94
5.2	Basic positioning of functions along a standard bond. . . . .	95
5.3	Residual density for two molecules showing the unmodeled ED of the lone pairs. . . . .	96
5.4	Concept and options for the LONE instruction. . . . .	97
5.5	Scheme for the modeling of $\pi$ bonds with BEDE and LONE. . . . .	98
5.6	Residual density of cyclohexane after BODD modeling. . . . .	102
5.7	Residual density of benzene after BODD modeling. . . . .	102
5.8	Residual density of formaldehyde after BODD modeling. . . . .	103
5.9	Residual density of trifluoroethane after BODD modeling. . . . .	103
5.10	R(F) for different refinements against simulated data compared to BODD refinement with SHELXL for selected model compounds. . . . .	104
5.11	LONE 6 A values refined for carbon atoms in different aromatic systems. . . . .	105
5.12	Comparison of differently transferred BODD parameters. . . . .	108

5.13	R(F) for refinements of transferred and refined BODD and multipole refinements.	108
5.14	ORTEP representation of MBADNP.	109
5.15	Model compounds for MBADNP assigned according to invariom names.	110
5.16	Model compounds for MBADNP assigned for each bond according to invariom names including next-nearest neighbors.	111
5.17	Model compounds for MBADNP assigned for each bond according to invariom names without next-nearest neighbors.	112
5.18	Model compounds for MBADNP assigned to each bond according to very simplified classification of bonds and atoms.	113
5.19	Residual density at the secondary nitrogen atom of <i>N</i> -methylpyridin-2-amine.	115
5.20	Planarity value of selected secondary amines after geometry optimization with different basis sets.	115
5.21	Invariom assignments for the transferability study with MBADNP. Model compounds for LONE parameters and for BEDE parameters of bonds from hydrogen atoms.	117
5.22	Residual density maps for MBADNP against simulated data modeled by IAM and transferred BODD.	118
5.23	R1(all) for different models of MBADNP after scale factor refinement with SHELXL against simulated data.	118
5.24	Residual density maps for different models refined against experimental data of MBADNP.	120
5.25	R1(all) for differently transferred BODD parameters for the refinement of MBADNP with SHELXL against experimental XRD data.	121
5.26	R1(all) for different models of MBADNP, comparing BODD with the multipole model.	121
5.27	R1(all) for different ylid models comparing BODD with the multipole model and different kinds of hydrogen atom treatment.	122
5.28	Scale factor for transfer of BODD from theory to experiment versus resolution of the datasets of MBADNP and ylid.	123
5.29	Residual density of MBADNP after refining 13 BODD parameters.	125
5.30	Residual density of MBADNP after refining 14 BODD parameters, when hydrogen atoms are placed at standard crystallographic positions.	125
5.31	Residual density at $0.15 \text{ e}/\text{\AA}^3$ of MBADNP after refining 14 BODD parameter when hydrogen atoms are placed at theoretical bond lengths.	125
5.32	Residual density at $0.15 \text{ e}/\text{\AA}^3$ of MBADNP after refining 17 BODD parameters when hydrogen atoms are placed at theoretical bond lengths.	125
5.33	Mean ADP difference between selected models and the one refined against neutron data for MBADNP.	127
5.34	Mean <i>Hirshfeld</i> test results for different models incl. the model refined against neutron data for MBADNP.	128
5.35	Average distance between atomic coordinates of different models and the one refined against neutron data for MBADNP.	129
5.36	Scatter plot matrix of BEDE parameters from bonds between carbon atoms and positive $r$ collected from 34 model compounds.	130
5.37	3D plot of $1/A$ (BEDE) versus $r$ and bond length $d$ from bonds between carbon atoms and $r > 0$ from 34 model compounds.	131
5.38	Parameter $A$ versus bond length for bonds between carbon atoms for 31 compounds refined with $r=0.36 \text{ \AA}$ .	132

## LIST OF FIGURES

---

5.39	Parameter $B1$ versus $A$ for bonds between carbon atoms for 31 compounds refined with $r = 0.36$ .	132
5.40	R1 for refinements of MBADNP against simulated data with different models including one where the BEDE parameters for ED between carbon atoms were calculated.	134
5.41	R1 for refinements of ylid against simulated data with different models including one where the BEDE parameters for ED between carbon atoms were calculated.	134
5.42	Comparison of R1 for SHELXL refinements of pair <b>4</b> with different metal atoms and modeling by IAM and BODD.	135
7.1	Thermal ellipsoid plot of pair <b>10</b> with Cu.	145
7.2	Thermal ellipsoid plot of pair <b>11</b> with Ni.	145
7.3	Work flow for the selection of grip points within 1.4 to 2.0 Å to a molecule.	146
7.4	Frequency distribution of the 343 different new invariom charges that formerly were invariom H@6c.	148
7.5	Detailed RRMS results for the comparison of different point charges for the investigated angiogenesis inhibitors.	149
7.6	Frequency distributions of the charge of the most frequent invarioms.	150
7.7	Frequency distributions of charge from exemplary non-carbon invarioms and their hydrogen invarioms.	151
7.8	Bond angles between hydrogen atoms of selected molecules as calculated by DFT in comparison to the ideal tetrahedral angle.	152
7.9	Residual density of furan after BODD modeling.	152
7.10	Populations of the quadrupole with $m=0$ refined with XD for carbon atoms in different aromatic systems.	152
7.11	Comparison of figures of merit for different models refined against x-ray data of MBADNP.	154
7.12	Comparison of residual densities for different models of MBADNP refined against experimental X-ray data.	155
7.13	Hirshfeld test results per atom for different models for MBADNP.	157
7.14	Scatter plot of $1/A(\text{BEDE})$ versus $r$ from bonds between carbon atoms and positive $r$ from 34 model compounds.	160
7.15	Plot of $A$ (BEDE parameter) versus bond length from bonds between carbon atoms and $r>0$ from 34 model compounds color coded according to their difference to the mean $r$ value.	160
7.16	Parameter $A$ versus bond length for bonds between carbon atoms for 31 compounds refined with $r=0.36\text{Å}$ , $B1=0.11$ and $B2=0.81$ .	161



# Acknowledgments

First of all I thank my advisors Prof. George Sheldrick and PD Birger Dittrich for their supervision and support of my thesis. On the one hand I much appreciated the freedom to choose the focus of many projects myself. On the other hand I enjoyed working with and learning from two crystallographic experts. I am grateful for the opportunities to present my research at conferences all over the world and to travel to synchrotrons in Switzerland and Hamburg for data collection.

I also thank Tim Grüne, member of my advisory committee, and all the members of my examination board, who invested their time to survey my work.

Special thanks goes to my colleagues Jens and Anna Lübben. Besides sharing an office, we spent a lot of time together at lunch, during coffee breaks and an unforgettable trip through different states of America. I also want to acknowledge their critical and constructive suggestions for some chapters of this thesis. Jens cooperated on a lot of projects and helped immensely by developing programs. His advice for the implementation of my plugins to those was very valuable.

I also owe a big thanks to Hilke Wolf, Felix Engelhardt, Lennard Krause and Christian Schürmann for maintenance of and support for the collection of X-ray diffractometers of the Stalke group. Prof. Stalke is thanked for granting me access to his much valued single-crystal diffractometers. Scientific discussions with his afore mentioned group members and Regine Herbst-Irmer were very helpful, especially Lennard's suggestion on combining two of my thesis projects for one experiment.

I am happy that Julian Holstein and Christian Hübschle introduced me to the field as well as to the many programs of crystallography and charge density in particular.

I would like to thank the Evangelisches Studienwerk Villigst e.V. for their PhD scholarship, which granted me scientific and financial independence. Moreover, I am thankful to the Volkswagen Stiftung for further financial support of conference travels.

P. Dominiak and M. Malinska are acknowledged for the diffraction data of sunitinib malate. The data of ylid were kindly provided by Michael Ruf from Bruker AXS.

Finally I thank my husband Sebastian for his support during my time as a PhD student. I feel lucky to have an emphatic husband, who proofread my complete thesis and was especially caring as well as encouraging in the final phase of writing.



# Scientific contributions

## Publications

- C. M. Wandtke, J. Lübben, B. Dittrich, **2016**, Molecular Electrostatic Potentials from Invariom Point Charges in *ChemPhysChem* **17**, p. 2238–2246.
- W. M. Bloch, Y. Abe, J. J. Holstein, C. M. Wandtke, B. Dittrich, G. H. Clever, **2016**, Geometric Complementarity in Assembly and Guest Recognition of a Bent Heteroleptic cis-[Pd2LA2LB2] Coordination Cage in *J. Am. Chem. Soc.* **138**, 13750–13755.
- B. Dittrich, C. M. Wandtke, A. Meents, K. Pröpper, K. C. Mondal, P. P. Samuel, S. K. Amin, A. P. Singh, H. W. Roesky, N. Sidhu, **2015**, Aspherical-Atom Modeling of Coordination Compounds by Single-Crystal X-ray Diffraction allows the Correct Metal Atom to be Identified in *ChemPhysChem* **16**, p. 412–419.
- K. Hernandez, I. Zelen, G. Petrillo, I. Usón, C. M. Wandtke, J. Bujons, J. Joglar, T. Parella, P. Clapés, **2015**, Engineered L-Serine Hydroxymethyltransferase from *Streptococcus thermophilus* for the Synthesis of  $\alpha,\alpha$ -Dialkyl- $\alpha$ -Amino Acids in *Angew. Chem.* **125**, p. 6163–6165.
- Yan Li, K. C. Mondal, P.P. Samuel, H. Zhu, C. M. Orben, S. Panneerselvam, B. Dittrich, B. Schwederski, W. Kaim, T. Modla, D. Koley, H. W. Roesky, **2014**, C4 Cumulene and the Corresponding Air-Stable Radical Cation and Dication in *Angew. Chem.* **126**, p. 4252–4256.
- C. M. Orben, B. Dittrich, **2014**, Hydrogen ADPs with Cu  $K\alpha$  data? Invariom and Hirshfeld atom modelling of fluconazole in *Acta Cryst. C* **70**, p. 580–583.
- M. Müller, C. M. Orben, N. Schützenmeister, M. Schmidt, A. Leonov, U.M. Reinscheid, B. Dittrich, C. Griesinger, **2013**, The Absolute Configuration of (+)- and (–)-erythro-Mefloquine in *Angew. Chem.* **125**, p. 6163–6165.
- S. Meyer, C. M. Orben, S. Demeshko, S. Dechert, F. Meyer, **2011**, Synthesis and Characterization of Di- and Tetracarbene Iron(II) Complexes with Chelating N-Heterocyclic Carbene Ligands and Their Application in Aryl Grignard–Alkyl Halide Cross-Coupling in *Organometallics* **30**, p. 6692–6702.

## Oral and poster presentations at conferences

- 08/2016** Poster: 'Identifying the correct metal atom in pairs of crystal structures of coordination compounds by aspherical-atom refinement', *29<sup>th</sup> European Crystallographic Meeting (ECM)*, Basel, Switzerland.
- 07/2016** Invited talk: 'Identifying metal atoms in coordination compounds.' *Meeting of the American Crystallographic Association (ACA)*, Denver, USA. *Etter Student Lecturer Award*.
- 07/2016** Invited tutorial: 'Shelxle – a GUI for structure determination with SHELXL.' *SHELX Workshop before the Meeting of the American Crystallographic Association (ACA)*, Denver, USA.
- 08/2015** Poster: 'Comparing electrostatic potentials from experiment, point charges and theory.' *28<sup>th</sup> European Crystallographic Meeting (ECM)*, Rovinji, Croatia.
- 01/2015** Talk: 'Invariom point charges – a rapid way to the electrostatic potential.' *23<sup>rd</sup> Annual Meeting of the German Crystallographic Association for Crystallography (DGK)*, Göttingen, Germany.
- 08/2014** Poster: 'Invariom point charges' *Meeting of the International Union of Crystallography (IUCr)*, Montreal, Canada.
- 01/2014** Poster: 'Point charges for force-field parametrization from the invariom database.' *22<sup>nd</sup> Annual Meeting of the German Crystallographic Association for Crystallography (DGK)*, Berlin, Germany.
- 08/2013** Poster: 'Can one use invarioms to predict hydrogen-atom positions for X-ray refinement?' *28<sup>th</sup> European Crystallographic Meeting*, Warwick, England. *IUCr poster prize*.