

**B. Dittrich,\* C. B. Hübschle, K. Pröpper, F. Dietrich, T. Stolper and J. J. Holstein**

Institut für Anorganische Chemie der Universität  
Göttingen, Tammannstr. 4, Göttingen D-37077,  
Germany

Correspondence e-mail: [bdittri@gwdg.de](mailto:bdittri@gwdg.de)

## The generalized invariom database (GID)

Received 11 October 2012

Accepted 23 January 2013

Invarioms are aspherical atomic scattering factors that enable structure refinement of more accurate and more precise geometries than refinements with the conventional independent atom model (IAM). The use of single-crystal X-ray diffraction data of a resolution better than  $\sin \theta/\lambda = 0.6 \text{ \AA}^{-1}$  (or  $d = 0.83 \text{ \AA}$ ) is recommended. The invariom scattering-factor database contains transferable pseudoatom parameters of the Hansen–Coppens multipole model and associated local atomic coordinate systems. Parameters were derived from geometry optimizations of suitable model compounds, whose IUPAC names are also contained in the database. Correct scattering-factor assignment and orientation reproduces molecular electron density to a good approximation. Molecular properties can hence be derived directly from the electron-density model. Coverage of chemical environments in the invariom database has been extended from the original amino acids, proteins and nucleic acid structures to many other environments encountered in organic chemistry. With over 2750 entries it now covers a wide sample of general organic chemistry involving the elements H, C, N and O, and to a lesser extent F, Si, S, P and Cl. With respect to the earlier version of the database, the main modification concerns scattering-factor notation. Modifications improve ease of use and success rates of automatic geometry-based scattering-factor assignment, especially in condensed hetero-aromatic ring systems, making the approach well suited to replace the IAM for structures of organic molecules.

### 1. Introduction

Databases of non-spherical scattering factors of the Hansen–Coppens multipole model (Hansen & Coppens, 1978) were developed with several clear aims in mind. For the invariom database the initial focus was on structure refinement, with the objective to obtain ‘accurate molecular structures’, fulfilling the description of accuracy discussed by Seiler (1992) and Hirshfeld (1992). Early work on the subject showed that including non-spherical scattering factors of Hirshfeld’s aspherical-atom model derived from an experimental charge-density study on pyrene allowed the improvement of anisotropic displacement parameters (ADPs) and the figures of merit of two other polycyclic aromatic hydrocarbons, where only low-order data were available due to poor crystal quality (Brock *et al.*, 1991). These findings were later confirmed in studies on an octapeptide (Jelsch *et al.*, 1998). For such investigations the suitability of peptides with their 20 repeating building blocks, the naturally occurring genetically encoded amino acids, was recognized early on and was a major factor in creating the first scattering-factor database derived

from experimental charge-density studies (Pichon-Pesme *et al.*, 1995). Further improvements of this experimental approach led to the ELMAM (Zarychta *et al.*, 2007) and ELMAM2 databases (Domagała *et al.*, 2012). ELMAM has seen applications on a number of small-molecule (*e.g.* Domagała *et al.*, 2011; Dadda *et al.*, 2012) and protein structures (*e.g.* Housset *et al.*, 2000; Jelsch *et al.*, 2000). However, improving macromolecular structures remains a challenge (Guillot *et al.*, 2008; Pröpper *et al.*, 2013), irrespective of the database used.

The introduction of methodology used in a study on L-dopa (Howard *et al.*, 1992), *i.e.* calculation of the theoretical structure factors  $F(\mathbf{h})$  from isolated molecules placed in an artificial unit cell with lattice constants large enough to avoid interactions between individual molecules, enabled the more convenient, reproducible and easily extendable calculation of scattering-factor databases derived from theory without experimental uncertainty. A prerequisite for generating scattering-factor databases was that different conformations do not compromise the transferability (Koritsánszky *et al.*, 2002) of Hansen–Coppens' variety of 'pseudoatom' (Stewart, 1976) scattering factors. A discussion of the advantages and disadvantages of experimental or theoretical procedures in database development can be found in Pichon-Pesme *et al.* (2004) and Volkov, Koritsánszky, Li & Coppens (2004).

Two theoretical databases were introduced almost simultaneously soon after the possibility of database generation from theory emerged (Dittrich *et al.*, 2004; Volkov, Li, Koritsánszky & Coppens, 2004). Their application led to comparable results<sup>1</sup> (Johnas *et al.*, 2009; Bąk *et al.*, 2011), although they differed in underlying design decisions and the coverage of chemical environments. A central design question that needs to be addressed in all databases is the extent of transferability of an atom in a particular chemical environment. In the invariom formalism a set of empirical rules was established from analyzing theoretical calculations, taking into consideration the descriptive tools of Bader's quantum theory of atoms in molecules (QTAIM; Bader, 1990). Consequently, the invariom database is additive: new chemical environments can be added without changing earlier entries. A unique model compound is identified for each atom of interest in a particular crystal structure, thereby providing a defined chemical environment for each scattering factor. Model compounds for generating entries in the invariom database are geometry-optimized. In the University at Buffalo Database (UBDB; Dominiak *et al.*, 2007; Jarzemska & Dominiak, 2012) an averaging process provides information on the transferability of an atom. Scattering factors are derived from single-point energy calculations of experimentally determined structures where bond distances to H atoms are elongated to average neutron distances (Allen & Bruno, 2010). Hence, only the invariom database is entirely free of experimental input.

<sup>1</sup> Both theoretical databases show essentially the same performance in the structure refinement of the undecapeptide cyclosporin, the small-molecule pharmaceutical paracetamol and the dipeptide L-histidinyll-L-alanine dihydrate.

Another difference between scattering-factor databases concerns the treatment of H atoms. Geometry-optimization yields accurate bond distances to H atoms. These bond distances are included in the invariom database and can be retrieved as target values for restraints or riding-hydrogen constraints. Alternatively, aspherical scattering factors from the invariom database allow free refinement of H-atom positions (Dittrich *et al.*, 2007). Such refinement gives bond distances in good agreement with neutron diffraction (Dittrich *et al.*, 2005): accurate bond distances to H atoms can hence be obtained directly from single-crystal X-ray diffraction, despite recurrent claims to the contrary (Deringer *et al.*, 2012). This is arguably the most important contribution of such databases to the method of X-ray diffraction. In particular, this last feature is an advantage of theoretical over experimental databases, because in the latter usually only dipole population parameters can be reliably refined from experimental X-ray diffraction data, and bond distances to H atoms are still too short in refinements with bond-directed dipoles.<sup>2</sup> Nonetheless, an advantage of experimentally derived multipole parameters is that the average effect of hydrogen bonding, unfortunately limited by the capability of the multipole model to describe diffuse electron density (Dittrich *et al.*, 2012), is in principle included in the database entries.

Recently, another experimental database was introduced (Hathwar, Thakur, Dubey *et al.*, 2011; Hathwar, Thakur, Row & Desiraju, 2011), seeking to distinguish itself by focusing on scattering factors required in crystal engineering, and by including the intermolecular interactions between synthons. Most of the scattering factors in this SBFA (supramolecular synthon-based fragments approach) were already contained in the invariom database, and the feature of including the average effect of hydrogen bonding in synthons was already present in the ELMAM2 library. Since only a part of the effect of packing and hydrogen bonding can be successfully described by the multipole model (Koritsánszky *et al.*, 2012; Dittrich *et al.*, 2012), it remains to be seen whether databases specialized in particular functional groups or areas of chemistry can provide extra value.

Experimental verification has confirmed (Jelsch *et al.*, 1998; Dittrich *et al.*, 2005; Volkov *et al.*, 2007; Bąk *et al.*, 2011) that the main aim shared by all database developers, *i.e.* improving structure refinement, can be successfully achieved with all databases. At present, improving small-molecule structure refinement can be considered the most important application. Results from a number of studies show that better figures of merit and ADPs with improved physical significance can be routinely obtained when compared with refinements using the independent atom model (IAM; Dittrich, Munshi & Spackman, 2006; Kingsford-Adaboh *et al.*, 2006; Dittrich *et al.*, 2007; Volkov *et al.*, 2007). Standard deviations in derived parameters are reduced proportionally to the improvement in the *R*-factor.

<sup>2</sup> Modeling H atoms in experimental charge density studies is usually limited to  $l_{\max} = 2$  rather than  $l_{\max} = 4$  for other elements. This is due to the small scattering contribution of H atoms, which is limited in resolution, and due to convolution of electron density with thermal motion.

Being able to reduce standard deviations is important, *e.g.* in absolute-structure determination (Dittrich, Strumpel *et al.*, 2006), where the Flack parameter (Flack, 1983) is frequently used. In order to claim sufficient inversion-distinguishing power for enantiopure light-atom structures, reaching a value around zero with a  $\sigma$  level below 0.12 was recommended (Flack & Bernardinelli, 2000). Absolute-structure determinations with reduced standard deviations of the Flack parameter after invariom refinement have been reported in a number of studies (Albrecht *et al.*, 2010; Yadav *et al.*, 2010; Abdalla *et al.*, 2011; Talontsi *et al.*, 2012). These and earlier studies also confirmed the performance of scattering-factor databases in improving conventional structure determinations of organic molecules. Further examples are listed in §6. Consequently, we called for replacing the IAM altogether with the Hansen–Coppens multipole model in combination with the invariom database for organic structures (Dittrich *et al.*, 2009). Such a replacement has been – and will need to be – accompanied by continuous software development. The graphical user interface and auxiliary program *MoleCoolQt* (Hübschle & Dittrich, 2011) in combination with the *XD* suite (Koritsánszky *et al.*, 2003; Volkov *et al.*, 2006) provide a user-friendly program environment for this purpose. Support for handling the system files of the programs *MoPro* (Jelsch *et al.*, 2005) already exists and is planned for *JANA2006* (Petricek *et al.*, 2006). Limitations of current multipole-model least-squares refinement programs make further developments desirable. For this reason we are currently working on more convenient treatment of static disorder. The feasibility of high-throughput (*i.e.* measurement and modeling of hundreds of compounds in a short time; Schürmann *et al.*, 2012) shows that replacing the IAM has become a genuine possibility to be realised in the very near future.

## 2. Improvements in the pseudoatom description

Despite the success of the Hansen–Coppens and Stewart varieties of rigid pseudoatom models in high-accuracy single-crystal diffraction throughout the last decades (Coppens, 1997; Koritsánszky & Coppens, 2001; Stalke, 2011), shortcomings in the model have become increasingly apparent. The analysis of Koritsánszky *et al.* (2012) (based on theoretical calculations) and the report by Fischer *et al.* (2011) (based on experimental data to extremely high resolution) show that core polarization should also be taken into account for high-accuracy work. Concerning valence electron density, the limited flexibility of the radial functions can be problematic, especially for diffuse electron density (Volkov & Coppens, 2001; Dittrich *et al.*, 2012). Most interesting current developments are increasing the order  $l$  of the multipole expansion (Volkov *et al.*, 2009), customization and tabulation of radial functions for a particular chemical environment (Koritsánszky *et al.*, 2012) and direct-space rather than reciprocal-space fitting for projections onto the pseudoatom model. Alternatively, a basis-set description as it is used in quantum chemistry allows all of these problems to also be solved (Jayatilaka, 1998; Jayatilaka & Grimwood, 2001). A combination termed ‘X-ray wave-

function refinement’ (Grabowsky *et al.*, 2012) of X-ray restrained wavefunction fitting with Hirshfeld-atom refinement (Jayatilaka & Dittrich, 2008) is currently the best option for high-resolution data. We hope that these developments will also benefit future releases of scattering-factor databases. Nevertheless, the well-tested multipole model can already improve most of those structure refinements that currently rely on the dated but highly successful IAM. The Hansen–Coppens multipole model has been well tested over the last 30 years and provides an excellent compromise between the number of parameters and accuracy. It can describe the transferable part of the valence electron density reasonably well and at the current stage seems the right model for replacing the IAM for structures of organic molecules.

## 3. Applications of transferable electron-density fragments

The second application of scattering-factor databases, in addition to structure refinement, is to allow the computationally efficient calculation of comparably accurate molecular properties. These can be directly derived from the aspherical electron-density distribution. Property calculation is especially relevant for larger molecules of biological importance (Dominiak *et al.*, 2009; Dittrich *et al.*, 2010) or whole series of related molecules (Holstein *et al.*, 2012), *i.e.* for cases where the computational effort needs to be minimized. Most prominent applications of such calculations initially focused on the electrostatic potential, which has been obtained for macromolecules such as aldose reductase (Guillot *et al.*, 2008), neuraminidase (Dominiak *et al.*, 2009) and trichotoxin A50E (Dittrich *et al.*, 2010). Further studies are under way. Other properties, which are currently under critical study, include the electrostatic interaction energy (Abramov *et al.*, 2000*a,b*; Li *et al.*, 2002, 2006; Volkov, Koritsánszky & Coppens, 2004; Spackman, 2007; Bouhmaida *et al.*, 2009) and the molecular dipole moment (Spackman, 1992; Spackman *et al.*, 2007; Poulain-Paul *et al.*, 2012; Dittrich & Jayatilaka, 2012). An open question is whether the approximations set in the Hansen–Coppens multipole model (frozen core, order of the expansion  $l_{\max} = 4$ ,  $m$ -independent radial functions) permit an accuracy to be reached that is high enough to obtain these properties from a reciprocal-space fit to structure factors in a reliable manner (Bąk *et al.*, 2011; Jarzemska & Dominiak, 2012), and how modifications in the pseudoatom description can improve the situation. Further methodological studies on these questions are required.

Another interesting application of database-derived molecular density distributions is the evaluation of hydrogen-bond energies, and empirical relationships have been derived and exploited (Abramov, 1997; Espinosa *et al.*, 1998, 2001). Although such evaluations might only reliably provide relative energies, for example for polymorphic or epimeric structures where data were measured at the same temperature on the same diffractometer, *e.g.* Nelyubina *et al.* (2010) and Madsen *et al.* (2011), results show that electron density provided by the Hansen–Coppens multipole model can yield a number of

interesting results beyond bond lengths and angles. Providing such information should help to establish non-spherical scattering-factor databases in the refinement and modelling of small-molecule and high-resolution macromolecular structures.

#### 4. The generalized invariom database (GID)

The generalized invariom database is an extension of the invariom database for amino acids, oligopeptides and protein molecules (Dittrich, Hübschle *et al.*, 2006). Most of the central concepts, like the choice of suitable model compounds that are geometry-optimized to provide an electron density distribution, and the principle using an invariom name as a descriptor that characterizes local chemical environments, were kept unchanged. A central modification to the former release of the database is the evolution of scattering-factor notation. It was first introduced at a Sagamore conference in 2009 and has been thoroughly tested since. Usage of the new notation increases the reliability of scattering-factor assignment as it is required for automated use. Directly related to scattering-factor notation (explained in §4.2) are the empirical rules (details in §4.1) that ensure transferability of electron density. These rules were also modified. Features and improvements of the generalized invariom database, of empirical rules and model-compound selection will be illustrated and discussed below.

##### 4.1. Empirical rules for ensuring pseudoatom transferability

Empirical rules on transferability were initially derived from evaluation of quantum chemical calculations on small organic molecules. Calculations were evaluated in terms of atomic volumes and charges defined according to Bader's QTAIM (Bader, 1990). Transferability can be assumed when charges agree within a certain standard deviation, as shown for some examples in Luger & Dittrich (2007). The study of transferability of theoretical QTAIM atomic charges and volumes was complemented by comparison of bond topological properties from experimental electron densities (Dittrich *et al.*, 2002; Rödel *et al.*, 2006; Grabowsky *et al.*, 2009). Underlying atomic multipole population parameters show a very good agreement in a similar chemical environment. We found a similar agreement when we 'projected' quantum chemical electron density onto the Hansen–Coppens multipole model.

A recent study (Woińska & Dominiak, 2011) has complemented these results. Transferability within three partitioning schemes was investigated: Bader's QTAIM, Hirshfelds stockholder partitioning (Hirshfeld, 1977) and Hansen–Coppens' multipole model. It was confirmed that fuzzy boundary partitioning schemes perform better than Bader's discrete boundary scheme; stockholder partitioning leads to the lowest standard deviations.

Another prerequisite for pseudoatom transferability, apart from agreement of QTAIM charges and volumes (or of multipole populations or other suitable descriptors), is a

negligible local difference in geometry in terms of bond distance, character and angles between bonded atoms. Bond strength and character can be quantified by Bader's QTAIM *via* topological analysis of electron density and Laplacians at the bond-critical points.<sup>3</sup> Hence only when the local atomic environment is similar in terms of these descriptors can we expect transferability of electron density using the rigid pseudoatom representation. This is usually implied when the terminology 'similar chemical environment' is used.

It was observed that many chemical environments show a high degree of transferability when the atoms involved have identical nearest neighbors. This holds for first, second and third row atoms in single-, double- and triple-bond environments, where electron density is localized.<sup>4</sup> Surprisingly, even a heteroaromatic ring system like thymidine, where delocalization of electron density over the whole ring is expected, can be modeled quite well with only nearest neighbors for the non-H atoms (Hübschle *et al.*, 2008). This observation provides the basis for our first empirical rule for predicting transferability of electron-density fragments in real space: in a single-bond environment we can generate a model compound comprising the atom of interest and its nearest neighbors, saturated by H atoms.<sup>5</sup> An important objective in establishing empirical transferability rules is to keep the number of possible fragments as small as possible. Equally important is to ensure that the approximation of reconstructing a molecular electron density from pseudoatom fragments remains as accurate as possible. Although shared two-center two-electron bonds between atoms occur most frequently, many more complex bonding situations exist. Different levels of complexity to generate the smallest, but yet well suited model compounds were therefore established. These rules are listed below. They ensure a high degree of transferability, providing an acceptable compromise between accuracy, and the number of fragments and model compounds required.

(i) Single-bond environment {as in ethanol, where only single bonds occur,  $\chi \leq 0.09$  [for a definition of  $\chi$  see equation (1)] in §4.5}: the atom of interest and its bonded neighbors are included in generating the model compound. Next-nearest neighbor atoms in the model compound are omitted and replaced by H atoms.

(ii) Delocalized ('mesomeric') bonds with a  $\chi$  value between 0.09 and 0.183 (as in formamide): here next-nearest neighbors need to be considered. Again, atoms in a subsequent shell are replaced by H atoms.

(iii) H atoms also require next-nearest neighbors, since their electron density is easily perturbed.

(iv) In order to distinguish  $sp^3$ -hybridized atoms in three-membered fused ring systems from their counterparts in

<sup>3</sup> Torsion angles involving next-nearest neighbors only need to agree for delocalized bonding environments, but can differ without compromising transferability for single-bond environments.

<sup>4</sup> Localization can be visualized using the electron localization function, ELF (Becke & Edgecombe, 1990), the more evolved electron localization indicator, ELI (Wagner *et al.*, 2007), or the source function (Gatti, 2005).

<sup>5</sup> Terminating valences with H atoms leads to smaller model compounds and hence also minimizes the computational effort of quantum chemical calculations.

unstrained chemical environments, three-membered rings are treated like atoms in delocalized systems (*i.e.* requiring next-nearest neighbors). This exception is not made for four-membered rings, where atoms are treated the same as normal  $sp^3$ -hybridized atoms.

(v) Double bonds with  $\chi \geq 0.183 \leq 0.27$  (as in ethene) and triple bonds with  $\chi \geq 0.27$  (as in acetylene): only nearest neighbors are considered. Next-nearest neighbors of the atom of interest are replaced by H atoms. Note that the presence of a triple bond can induce a mesomeric character in bonds adjacent to it, leading to a delocalized system where next-nearest neighbors are considered.

(vi) ‘Hypervalent’ elements Si, P, S and Cl themselves do not require special treatment different to the rules given above. However, atoms attached to hypervalent atoms need to carry information about their next-nearest neighbors, since these can differ at the site of the hypervalent atom, thereby influencing the electron-density distribution of their neighbors.

(vii) Extended delocalized ring systems (as in naphthalene): as in the formamide example, model compounds include a first and second shell of neighbors. If the atom of interest is part of an  $n$ -membered ring, the ring size  $n$  is maintained in the model compound. This rule also applies to fused ring systems.<sup>6</sup> A delocalized system is identified by considering whether each atom in the ring is planar (defined by the difference of the angles of the atom-neighbor vectors). For those rare heteroaromatic systems where several model compounds can be considered suitable, the best model compound is one that fulfills four criteria in the following sequence of priority: it has the least number of (1) atoms and (2) electrons while still maintaining planarity, is preferred to be (3) uncharged rather than charged and contains, if possible, (4) C and H atoms rather than N or O. This last rule required programming of a suitable algorithm, that calculates a score for each model compound from the formula sum. In case model compounds have the same element composition, the lowest-energy isomer containing the invariom is considered to be the best model compound.

The rules mentioned above ensure that there is really only one best-suited model compound for a particular invariom. Chemical environments can be correctly described by – and are closely connected to – the scattering-factor notation explained in §4.2. The rules might appear complicated at first, but are easy to use in real life, requiring only a modest amount of practice.

Care has to be taken for some N atoms, where the local geometry and energy difference between planar and pyramidal geometry is small, whereas single-bond environments such as present in *e.g.* sugars are trivial to handle. Non-trivial cases where transferability might be limited are discussed in §7.1.

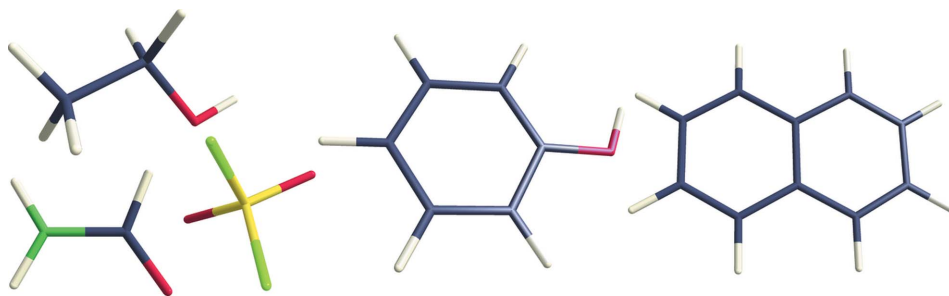
<sup>6</sup> In case the atom of interest or its neighbors are involved in a condensed ring system this rule might lead to relatively large model compounds. More importantly it fulfills the requirement to include ‘the whole mesomeric system’ (Luger & Dittrich, 2007).

## 4.2. Scattering-factor notation

In contrast to computer-interpretable notations for molecules like the Simplified Molecular Input Line Entry Specification (SMILES; Weininger, 1988) and InCHI, the notation of a scattering factor does not need to represent a molecule consisting of atoms bonded in a particular manner, but only requires to uniquely identify an atom in a particular chemical environment. A specialized scattering-factor notation was therefore developed in which the element type of the atom of interest commences a string of characters of the invariom name. Atoms in a single-bond environment provide the easiest case for such a name. Here the element type (in capitals) is followed by the bond order 1 and the element type of only the directly neighboring atoms, ordered by the number of electrons, *i.e.* the position in the periodic table of elements. Next-nearest neighbors are considered for H atoms, since their electron density is easily perturbed; next-nearest neighbors are added behind nearest neighbors in square brackets, like in a terminal methyl-group hydrogen: H1c[1c1h1h]. For atoms that are part of a delocalized chemical environment, next-nearest neighbors also need to be considered in this way. Delocalized bonds are found when the bond-distinguishing parameter  $\chi$  (Hübschle *et al.*, 2007) from the geometry-optimized structure [DFT, method/basis: B3LYP/D95++(3df,3pd)] has a value between 0.09 and 0.183. Formamide provides a fitting example: its C atom is called C2o1.5n[1h1h]1h. The O atom has a bond-distinguishing parameter above  $\chi = 0.183$ , and hence no next-nearest neighbors need to be considered for oxygen here; the invariom name for oxygen in formamide is O2c (being itself derived from the model compound formaldehyde). A triple bond is assigned for bonds where the bond-distinguishing parameter exceeds 0.27, *e.g.* for oxygen in carbon monoxide, which is called O3c.<sup>7</sup>

The highest level of complexity in the name is required for planar aromatic ring systems. Here the size of the ring is taken into account and given before the element type of the atom of interest. Delocalization is assumed for rings where atoms are planar. The need to distinguish the bond order for planar ring atoms therefore vanishes; bonding in delocalized planar rings is indicated by a ‘#’ symbol. For ordering of the neighbors in the string of the invariom name the delocalized #-bonds take precedence over single bonds. These aspects can be illustrated for phenol, where the invariom name for the C atom adjacent to the hydroxy group is 6-C#6c[#6c1h]#6c[#6c1h]1o. For atoms attached to planar rings (like the oxygen atom in phenol) the value of the bond-distinguishing parameter is likewise not taken into account and is replaced by the symbol ‘@’, giving O@6c1h for the phenol oxygen. Atoms taking part in several condensed rings carry this information in their invariom name by including the number of members of each ring (the ring size): for the central two C atoms in naphthalene the name is 66-C#66c[#6c#6c]#6c[#6c1h]#6c[#6c1h]. This notation also works well for heteroaromatic ring systems containing *e.g.* nitrogen.

<sup>7</sup> Note that for disordered crystal structures where split atoms are in close proximity, the analysis of geometry leads to meaningless invariom names.



**Figure 1**

Some of the example molecules used for explaining scattering factor notation and model-compound generation. From left to right: formamide, ethanol, sulfur difluoride, phenol and naphthalene. N atoms in green, O in red, C in dark blue, H in light grey and S in yellow.

N atoms may cause another complication: due to the small inversion barrier  $NR_3$  atoms can occur as either planar or pyramidal without much difference in the bond length of the bonds involved. We distinguish planar from pyramidal nitrogen by adding an equal sign in front of the invariom name as in  $=N1c1h1h$ . This sometimes leads to larger than expected and hence counterintuitive model compounds; only when both  $\chi$  and  $PV$  (for their definition see §4.6) match is the right invariom identified from its name (for examples, see the supplementary information<sup>8</sup>).

Chiral invarioms are a lot less common than chiral atoms, since we do not differentiate between neighboring atoms connected by a single bond. Consequently, for carbon only those invarioms are chiral that really have four neighbors with different element types or a different bonding situation involving next-nearest neighbors. When chiral invarioms appear they are assigned an ‘R-’ or ‘S-’ prefix following CIP (Cahn–Ingold–Prelog) notation (Prelog & Helmchen, 1982).

An example for an element that is ‘hypervalent’ and an atom attached to it is the O atom in  $SO_2F_2$ , which is different to that in  $SO_2$ . Here both invariom names take into consideration their next-nearest neighbors, again in square brackets, giving  $O2s[2o1f1f]$  and  $O2s[2o]$ .

The user can find transferability fulfilled to a very high degree where the same invariom name is assigned, which also holds for cases like pentachlorophenolate, as mentioned by Jarzemska & Dominiak (2012). All these rules are implemented in the computer programs *InvariomTool* (Hübschle *et al.*, 2007) and *MoleCoolQt* (Hübschle & Dittrich, 2011). Invariom names can be automatically generated from input geometry; more experienced users can also generate them by hand. Molecules discussed in this section are depicted in Fig. 1.

#### 4.3. Program development

Providing the *XD* suite preprocessor program *InvariomTool* (Hübschle *et al.*, 2007) was our initial attempt to facilitate least-squares refinement with database scattering

<sup>8</sup> Supplementary data for this paper are available from the IUCr electronic archives (Reference: SN5118). Services for accessing these data are described at the back of the journal.

factors (‘database application’). More recently, the functionality of *InvariomTool* has been incorporated into the program *MoleCoolQt* (Hübschle & Dittrich, 2011), a graphical user interface for the *XD* suite (Koritsánszky *et al.*, 2003; Volkov *et al.*, 2006) and the *MoPro* program (Guillot *et al.*, 2001; Jelsch *et al.*, 2005). Our aim was to automate scattering-factor assignment and orientation of the local atomic coordinate system to a high degree, *e.g.* also

for atoms in special positions. This latter feature is implemented only in *MoleCoolQt*. *MoleCoolQt* can also substantially facilitate modeling of disordered molecules, which will be described in a subsequent publication. Both programs are provided free of charge and are available for download (<http://www.molecoolqt.de/>).

#### 4.4. Missing invarioms

Despite continuous effort in calculating model compounds throughout the last 6 years we estimate that more than twice as many model compounds might be required for close to complete coverage of organic chemistry. Reaching an acceptable coverage when including the third-row elements Si, P, S and Cl will require even further efforts. We therefore offer a service to calculate model compounds to include missing invarioms starting from optimized geometries. Alternatively, for interested users we provide the tools that we developed for generating invariom-database entries, asking that new invarioms are shared with other users. These tools were developed for the Linux operating system, but could be compiled for other operating systems.

#### 4.5. Modifications and improvements in database generation and extension

In the current version of *InvariomTool* used to prepare input files for aspherical atom refinement, local atomic coordinate systems and chemical constraints still rely on the equation of Schomaker & Stevenson (1941), as improved by Blom & Haaland (1985) for the definition of a bond-distinguishing parameter  $\chi$  as in equation (1)

$$\chi = [r_c(\text{atom 1}) + r_c(\text{atom 2}) - 0.08 \cdot |\Delta(\text{EN})|] - d, \quad (1)$$

where EN is the electronegativity according to Allred & Rochow (1958), and  $r_c$  are the covalent radii of the respective atoms and  $d$  is the bond distance.

Concerning the reproduction of molecular electron density and the calculation of molecular properties we have found that a balanced increase of the sophistication of multipole-model flexibility gives minor improvements in fitting experi-

mental data; one shared  $\kappa'$  parameter for all atoms heavier than carbon is used in a consistent manner in the GID.<sup>9</sup> The most important modification in the extension of the database is a script-based semi-automatic generation procedure (see §4.6). Another improvement lies in the resolution of the simulated data. We now use a full sphere of data up to a resolution of  $1.2 \text{ \AA}^{-1}$ , whereas beforehand limiting indices of  $-40:40$ ,  $-40:40$  and  $0:40$  up to a resolution of  $1.15 \text{ \AA}^{-1}$  were used for  $h$ ,  $k$  and  $l$ , respectively. Yet a further modification of the procedure to generate scattering factors relates to the calculation of simulated structure factors from geometry-optimized molecules of structures containing third-row elements.

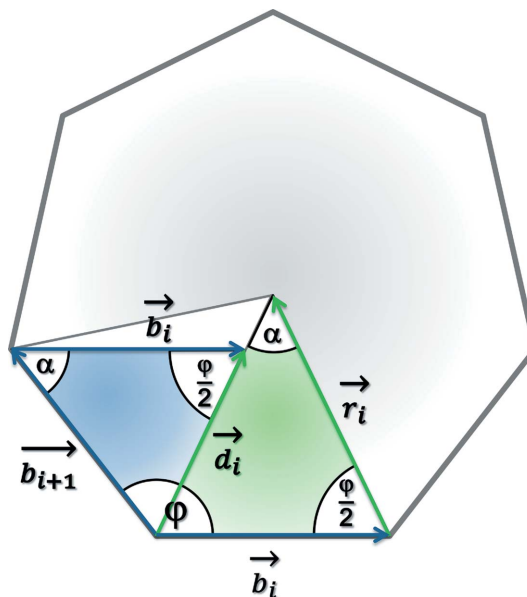
#### 4.6. Computational and procedural details, algorithms

We commence the procedure of generating invarioms by drawing a new model compound in the commercial program *ChemDraw*, where we generate the SMILES name (Weininger, 1988) of the compound. The open-source program *Avogadro* (Hanwell *et al.*, 2012) can interpret this string to give a starting geometry. The model compound is then initially geometry-optimized using the universal force field (UFF; Rappé *et al.*, 1992). With *Avogadro* we also generate input files for the quantum chemistry program *GAUSSIAN* (Frisch *et al.*, 2009). Geometry optimization in *GAUSSIAN* uses stringent convergence criteria (options 'very tight' and Grid = Ultra-Fine); frequency calculations are performed to ensure the global minimum is reached. The database is ordered by IUPAC names of the model compounds on the hard drive, and the name of the subdirectory is the basis set used. Next, the program *Tonto* (Jayatilaka & Grimwood, 2003) is used for analytical Fourier transform (Jayatilaka, 1994) of the real-space quantum chemical electron density to generate simulated<sup>10</sup> scattering factors for subsequent multipole projection. A utility program converts *GAUSSIAN* output into an *XD*-readable file. Both *GAUSSIAN* and *Tonto*, the programs for the most time-consuming steps, are parallelized. *InvariomTool* (Hübschle *et al.*, 2007) then preprocesses *XD* system files, already taking into account the existing entries in the database to generate a multipole model. It also changes the local-atomic coordinate systems and inserts multipoles to be refined in the process, based on the choices of local-atomic site symmetry for existing invarioms made in the database. Hence only new (missing) invarioms need to be manually assigned a local atomic coordinate system in the multipole refinement of new model compounds with simulated structure factors. A shell-script carries out all these tasks (*GAUSSIAN*, *Tonto* and *InvariomTool*). Overall scattering-factor generation requires very few manual steps.

Concerning new algorithms, an important improvement in dealing with aromatic and strained three-membered rings was to extend the geometric criteria. Especially for modeling large

<sup>9</sup> For C atoms only  $\kappa$  but no  $\kappa'$  parameter seems to be required.

<sup>10</sup> We chose to use the term 'simulated' because we simulate an X-ray diffraction experiment. Furthermore, the term 'calculated structure factors' is already being used and well established for model structure factors, and using it would cause confusion.



**Figure 2**

The procedure for recognizing planar rings, exemplified for a seven-membered ring.

structures a computationally fast and elegant way to detect planar ring systems was required. Information on atomic planarity (only the nearest neighbors of an atom define whether it is planar or not) is now included to distinguish and assign invariom names. A planarity value<sup>11</sup> ( $PV$ ) is introduced for that purpose. It is calculated from the analysis of vectors between bonded atoms. These vectors  $n$  are first converted into a Cartesian frame and normalized. Next vector products are formed between all of them. The atom is planar when all of these resulting vectors point in the same direction. This can be probed by calculating scalar products between the resulting vectors and the maximal value that can be obtained is unity. This is summarized in equation (2)

$$PV = \prod_{i=1}^l \left( \prod_{j=i+1}^l \mathbf{n}_i \times \mathbf{n}_j \right). \quad (2)$$

Equation (3) illustrates how the planarity value is calculated in case of a chemical environment consisting of three covalent bonds

$$PV_{l=3} = (\mathbf{n}_1 \times \mathbf{n}_2) \cdot (\mathbf{n}_1 \times \mathbf{n}_3) \cdot (\mathbf{n}_2 \times \mathbf{n}_3). \quad (3)$$

For planar atoms or linear coordination  $PV$  therefore gives a value of unity, whereas for tetrahedral and octahedral environments it is zero. The information on atomic planarity is also used for identifying planar rings. A two-step 'atom casting' procedure reduces the number of numerical comparisons required: in the first step all planar atoms are listed; for each atom with at least two planar neighbors a potential midpoint is calculated. Fig. 2 illustrates how a midpoint of a potential  $n$  angle equilateral polygon is found.  $\mathbf{b}_i$  and  $\mathbf{b}_{i+1}$  are vectors originating from an atom to which two further atoms are

<sup>11</sup> Taking planarity into account was inspired by the procedure used in the UBDB (Dominiak *et al.*, 2007).

**Table 1**

List of model compounds for generating the invariom needed in the refinement of loperamide hydrochloride (Brüning *et al.*, 2012).

See Fig. 3 for the atomic naming scheme.

Atom	Invariom name	Model compound
Cl1	Cl@6c	Chlorobenzene
Cl2	Cl	Chloride
O1	O1c1h	Methanol
O2	O1.5c[1.5n1c]	Acetamide
N1	N1c1c1h	Trimethylammonium
N2	N1.5c[1.5o1c]1c1c	<i>N,N</i> -Dimethylacetamide
C1,18,24	6-C#6c[#6c1h]#6c[#6c1h]1c	Toluene
C2,6,19,23,25,29	6-C#6c[#6c1c]#6c[#6c1h]1h	Toluene
C3,5	6-C#6c[#6c1c]#6c[#6c1h]1h	Chlorobenzene
C4	6-C#6c[#6c1h]#6c[#6c1h]1c1	Chlorobenzene
C7	Cl@6c1c1c	2-Phenylpropan-2-ol
C8,11,13	Clc1c1h1h	Propane
C9,10,12	Cln1c1h1h	Ethylamine
C14	C@6c@6c1c1c	2,2-Diphenylpropane
C15	Cl.5o1.5n[1c1c]1c	<i>N,N</i> -Dimethylacetamide
C16,17	Cln1h1h1h	Methylamine
C20,21,22,26,27,28	6-C#6c[#6c1h]#6c[#6c1h]1h	Benzene
H1B	H1o[1c]	Methanol
H1A	H1n[1c1c1c]	Trimethylammonium
H2,3,5,6A,19–23A,25–29A	H@6c	Benzene
H8A,B,11A,B,13A,B	H1c[1c1c1h]	Propane
H9A,B,10A,B,12A,B	H1c[1n1c1h]	Ethylamine
H16A,B,C,17A,B,C	H1c[1n1h1h]	Methylamine

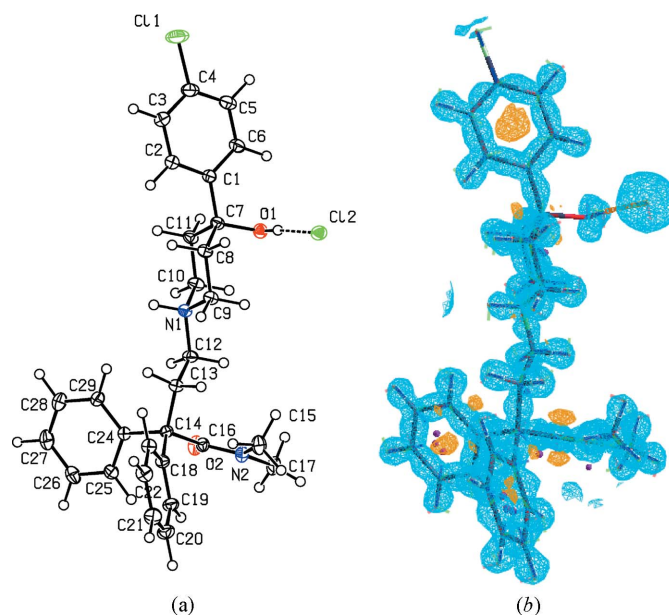
connected;  $\varphi$  is the angle between them. A vector  $\mathbf{d}_i = \mathbf{b}_i + \mathbf{b}_{i+1}$ , obtained from vector addition, points approximately in the direction of the polygon midpoint. To calculate a required length of  $\mathbf{d}_i$  we apply a scale factor  $s = |\mathbf{r}_i|/|\mathbf{d}_i|$ , which is, according to the sine law,  $s = \sin^2(\varphi/2)/\sin^2 \varphi$ . A planar ring then involves atoms within a threshold radius around the center thus obtained. The algorithm is capable of dealing with ring sizes from four to eight atoms, and assumes an equilateral polygon.

#### 4.7. Coordinate systems

Unlike IAM scattering factors, multipole-model scattering factors are not spherically symmetric and require a local atomic coordinate system for their orientation in space. In earlier publications we did not describe how we ensure correct and automated selection of such coordinate systems, while details on related developments for the ELMAM/ELMAM2 library were reported previously (Domagała & Jelsch, 2008). In the invariom approach, coordinate systems are based on matching connected element types and bond distances of model compounds, from which scattering-factor entries are generated. Relying on model compound and their distances gives us the flexibility we need to cover the variety of chemical bonding encountered, but requires, in contrast to generalized rules,<sup>12</sup> the individual model compound as a reference. Bond distances (translated into  $\chi$  values) and the element type of the atoms that the coordinate system vectors are pointing to, are stored together with the multipole parameters in the

<sup>12</sup> We have discussed *e.g.* using CIP rules to elegantly cover tetrahedral C atoms, but these would unfortunately not cover other bonding environments.

invariom database. Coordinate systems are transferred correctly, when the database  $\chi$  values (*i.e.* two chosen bond distances) can be matched well with those found in a real crystal structure. A unique coordinate system can often be defined by assigning the shortest bond to the first axis and the second shortest bond to the second axis; the third axis is calculated from a vector product of these two and the system is subsequently orthonormalized. We try to avoid less well defined H-atom positions in local coordinate systems; sometimes, like in O–H groups with *m*-symmetry, this is not possible. In case two equivalent bonds exist, *e.g.* in  $R_2$ –CH<sub>2</sub> groups, a dummy atom is generated, for example from the vector sum of the C–R vectors. *InvariomTool* does this for the most common chemical environments. To ensure that the correct coordinate system has indeed been transferred, *MoleCoolQt* can show coordinate axis directions (*x* in red, *y* in green, *z* in blue) and allows an additional visual inspection of the three-dimensional deformation density (see example in Fig. 3). Such inspection is currently only possible in combination with the *XD* suite of programs. In technical terms, the Fourier file *xd.fou* is read, and a deformation density is generated by fast-Fourier transform on the fly. In case the deformation electron density is misaligned, the user is made aware that the choice of coordinate system must be wrong – this presents a useful way of ensuring the choice of coordinate system especially for larger molecules. In case the automatic assignment or the dummy-atom generation fails, *MoleCoolQt* facilitates manual correction of the local atomic coordinate system.

**Figure 3**

(a) Atomic naming scheme for the anti-diarrhetic loperamide hydrochloride (Brüning *et al.*, 2012). H atom labels were omitted for clarity as they carry the name of their parent atoms. (b) Three-dimensional deformation electron density (0.1 e Å<sup>-3</sup> isosurface) from fast Fourier transform with *MoleCoolQt* that helps to validate the assignment of the local atomic coordinate systems on the click of the mouse.



Table 2

Figures-of-merit and experimental conditions for 32 invariom structure refinements and comparison with IAM refinements.

#	Compound	Temperature (K)	$R_{IAM}^{(F)}$	$R_{inv}^{(F)}$	$\Delta_{IAM}$	$\Delta_{inv}$
1	( <i>E</i> )-1-[4-(Hexyloxy)phenyl]-3-(2-hydroxy-phenyl)prop-2-en-1-one <sup>a</sup>	100	0.0416	0.0264	0.37	0.14
2	Chelidamic acid methanol solvate <sup>b</sup>	173	0.0391	0.0275	0.29	0.20
3	5''-[( <i>E</i> )-2,3-Dichlorobenzylidene]-7'-(2,3-dichlorophenyl)-1''-methylspiro[acenaphthylene-1,5'-pyrrolo[1,2- <i>c</i> ][1,3]thiazole-6',3''-piperidine]-2,4''-dione <sup>c</sup>	293	0.0500	0.0434	0.41	0.52
4	5''-[( <i>E</i> )-4-Fluorobenzylidene]-7'-(4-fluorophenyl)-1''-methylspiro[acenaphthylene-1,5'-pyrrolo[1,2- <i>c</i> ][1,3]thiazole-6',3''piperidine]-2,4''-dione <sup>c</sup>	293	0.0441	0.0389	0.46	0.51
5	Baicalin nicotinamide (1/1) <sup>d</sup>	100	0.0703	0.0646	0.41	0.45
6	2-(1,4,7,10-Tetraazacyclododecan-1-yl)cyclohexan-1-ol (cyclyclen) <sup>e</sup>	173	0.0470	0.0429	0.29	0.31
7	2-(1 <i>H</i> -Indol-3-yl)-2-oxoacetamide <sup>f</sup>	90	0.0585	0.0525	0.43	0.34
8	<i>cis</i> -2-(2-Fluorophenyl)-3a,4,5,6,7,7a-hexahydroisindole-1,3-dione <sup>g</sup>	200	0.0456	0.0322	0.43	0.28
9	<i>cis</i> -2-(4-Fluorophenyl)-3a,4,5,6,7,7a-hexahydroisindole-1,3-dione <sup>g</sup>	200	0.0410	0.0318	0.22	0.24
10	5-Ethynyl-2'-deoxycytidine <sup>h</sup>	130	0.0301	0.0148	0.26	0.13
11	6-Chloro-3-methyl-1,4-diphenylpyrazolo[3,4- <i>b</i> ]pyridine-5-carbaldehyde <sup>i</sup>	120	0.0432	0.0335	0.42	0.45
12	6-Chloro-3-methyl-4-(4-methylphenyl)-1-phenylpyrazolo[3,4- <i>b</i> ]pyridine-5-carbaldehyde <sup>i</sup>	120	0.0535	0.0469	0.30	0.27
13	<i>N,N,N',N'</i> -Tetrabenzyl- <i>N''</i> -(2-chloro-2,2-difluoroacetyl)phosphoric triamide <sup>j</sup>	296	0.0491	0.0397	0.64	0.70
14	Benzoyl(hydroxyimino)acetonitrile 18-crown-6 water (2/1/4) <sup>k</sup>	213	0.0368	0.0327	0.18	0.16
15	<i>N,N</i> -Dibenzyl- <i>N'</i> -(furan-2-carbonyl)thiourea <sup>l</sup>	294	0.0401	0.0323	0.21	0.15
16	3-Phenylcoumarin <sup>m</sup>	100	0.0413	0.0305	0.25	0.19
17	3,4,6-Tri- <i>O</i> -acetyl-1,2- <i>O</i> -[1-(exoethoxy)ethylidene]- $\beta$ -D-manno-pyranose-0.11H <sub>2</sub> O <sup>n</sup>	153	0.0546	0.0531	0.35	0.38
18	4-[( <i>E</i> )-(4-Ethoxyphenyl)iminomethyl]phenol <sup>o</sup>	120	0.0299	0.0189	0.12	0.12
19	5 <i>H</i> -Dibenzo[ <i>b,e</i> ]diazepin-11(10 <i>H</i> )-one <sup>p</sup>	147	0.0521	0.0473	0.28	0.36
20	<i>N,N'</i> -Bis(2-methylphenyl)-2,2'-thiodibenzamide <sup>q</sup>	296	0.0352	0.0274	0.17	0.19
21	(3 <i>R</i> ,8 <i>R</i> )-2,2-Dimethyl-4,4,8,8-tetraphenyl-4,5,6,7,8,8a-hexahydro-3 <i>aH</i> -1,3-dioxolo-[4,5- <i>e</i> ][1,3]diazepin-6-one-0.33H <sub>2</sub> O <sup>r</sup>	173	0.0576	0.0442	0.30	0.31
22	(3 <i>aS</i> ,8 <i>aS</i> )-2,2-Dimethyl-4,4,8,8-tetraphenyl-4,5,6,7,8,8a-hexahydro-3 <i>aH</i> -1,3-dioxolo-[4,5- <i>e</i> ][1,3]diazepin-6-one-0.39H <sub>2</sub> O <sup>r</sup>	173	0.0494	0.0348	0.32	0.23
23	4-Cyano- <i>N</i> -(4-methoxyphenyl)-benzenesulfonamide <sup>s</sup>	293	0.0399	0.0359	0.23	0.30
24	<i>N</i> -(4-Methoxyphenyl)-4-(trifluoromethyl)benzenesulfonamide <sup>s</sup>	120	0.0597	0.0573	0.54	0.52
25	(3 <i>R</i> ,5 <i>S</i> ,5' <i>R</i> ,8 <i>R</i> ,9 <i>S</i> ,10 <i>S</i> ,13 <i>S</i> ,14 <i>S</i> )-10,13-dimethyl-5'-(2-methylpropyl)tetradecahydro-6' <i>H</i> -spiro[cyclopenta[ <i>a</i> ]phenanthrene-3,2'-[1,4]oxazinan]-6',17(2 <i>H</i> )-dione <sup>t</sup>	150	0.0390	0.0311	0.18	0.18
26	Methyl(2 <i>R</i> )-2-[(3 <i>R</i> ,5 <i>S</i> ,8 <i>R</i> ,9 <i>S</i> ,10 <i>S</i> ,13 <i>S</i> ,14 <i>S</i> )-10,13-dimethyl-2',17-dioxohexadecahydro-3' <i>H</i> -spiro[cyclopenta[ <i>a</i> ]phenanthrene-3,5'-[1,3]oxazolidin-3'-yl]]-4-methyl-pentanoate <sup>t</sup>	150	0.0372	0.0346	0.24	0.24
27	Loperamide hydrochloride <sup>u</sup>	169	0.0442	0.0340	0.41	0.39
28	Perindoprilate dimethyl sulfoxide hemisolvate <sup>v</sup>	100	0.0290	0.0216	0.28	0.36
29	Bis( <i>tert</i> -butyldimethylsilyl)(2,6-diisopropylphenyl)phosphane <sup>w</sup>	120	0.0310	0.0243	0.32	0.39
30	(2 <i>RS</i> ,4 <i>RS</i> )-7-Fluoro-2-(2-phenylethyl)-2,3,4,5-tetrahydro-1 <i>H</i> -1,4-epoxy-1-benzazepine <sup>x</sup>	120	0.0566	0.0443	0.26	0.24
31	1,3-Ninhydrin dihydrazone <sup>y</sup>	150	0.0404	0.0337	0.23	0.16
32	Isoquinolin-5-amine <sup>z</sup>	150	0.0392	0.0348	0.26	0.21

References: (a) Fadzillah *et al.* (2012); (b) Tutughamiasro *et al.* (2012); (c) Suresh *et al.* (2012); (d) Sowa *et al.* (2012); (e) de Sousa *et al.* (2012); (f) Sonar *et al.* (2012); (g) Smith & Wermuth (2012); (h) Seela *et al.* (2012); (i) Quiroga *et al.* (2012); (j) Pourayoubi *et al.* (2012); (k) Ponomarova & Domasevitch (2012); (l) Pérez *et al.* (2012); (m) Matos *et al.* (2012); (n) Liu *et al.* (2012); (o) Khalaji *et al.* (2012); (p) Keller *et al.* (2012); (q) Helliwell *et al.* (2012); (r) Gherase *et al.* (2012); (s) Gelbrich *et al.* (2012); (t) Djigoue *et al.* (2012); (u) Brüning *et al.* (2012); (v) Bojarska *et al.* (2012); (w) Boeae & Taghavikish (2012); (x) Blanco *et al.* (2012); (y) Blake *et al.* (2012); (z) Atria *et al.* (2012). The residual electron density is given in e Å<sup>-3</sup>. Refinements were initiated with *SHELXL* (Sheldrick, 2008) and were repeated in *XD2006* (Volkov *et al.*, 2006). These refinements on *F* used *F* > 3 $\sigma$ *F* and a weighting scheme of 1/ $\sigma^2$ , except when the *SHELXL* *R* factor was not reached in *XD<sub>IAM</sub>*, in which case a *SHELXL*-type weighting scheme was chosen to match it. None of the structures exceeds a resolution of 0.84 Å<sup>-1</sup> in  $\sin \theta/\lambda_{\max}$ . Invariom refinements are denoted 'inv'. Structures 14 and 15 were refined against *F*<sup>2</sup>. † Structures were measured with Cu *K* $\alpha$  radiation.

## 5. Database generation

Like the task of generating model compounds, the task to update and extend the database has been automated. A perl script selects the most suitable model compound based on its molecular formula for each invariom, and extracts the relevant information from the file structure. Hence, in case an incorrect (e.g. a too large) model compound has initially been calculated, and a better suited one is added at a later stage, the next version of the database, which we continuously extend, will contain the correct entry.

This procedure also facilitates tracking which model compound has been used for extracting a particular invariom; *InvariomTool* writes a '.descent' file that lists all atoms, invariom names and their parent compounds for a particular structure. Its content is listed in Table 1 for the example structure 27, loperamide hydrochloride, from §6. The atomic

naming scheme of the structure is given in Fig. 3. In a refinement of a real crystal structure such information can hence easily be retrieved and we recommend providing it to ensure reproducibility. Tables of invarioms and parent compounds for all structures modeled in §6 are provided in the supplementary material.

## 6. Validation and testing of the database

To illustrate the capabilities of the *GID* in structural work with conventional X-ray diffraction data we have tested the new database on a number of structures that we downloaded (including intensity data) from the home page of the journals *Acta Crystallographica* Sections C and E. This approach was already taken for validating the former version of the database (Dittrich, Hübschle *et al.*, 2006). We report here representative results for 32 selected structures in Table 2. These structures

represent a sample of the total number of structures tested, and they cover a variety of bonding situations commonly encountered in organic chemistry. They include single, double and triple bonds, neutral and charged species, chiral invarioms, planar and non-planar nitrogen, bonding in condensed rings with and without delocalization, and bonds to S, P as well as Si atoms. The molecules studied contain a nucleic acid in the form of a chemically modified cytidine, spiro compounds, a sugar, two steroids and several pharmaceutically active molecules. On the experimental side a variety of temperatures were encountered; radiation was either Cu  $K\alpha$  or Mo  $K\alpha$ . Only non-disordered structures were selected.

We find an average reduction of the  $R$  factor of 0.81%. The positive residual electron density is often reduced as well. However, when heavy atoms are present, and an analytical absorption correction might not have been applied, the highest peak can also increase. Even then, the r.m.s. of positive and negative residual electron density is usually reduced. Moreover, average values of the Hirshfeld test (Hirshfeld, 1976) are always smaller than for the IAM refinements for all structures given (values not tabulated). Re-refinement with invarioms also allows structures to be validated. An earlier example of electron-density validation was provided by Holstein *et al.* (2010): residual density makes overlooked disorder more obvious and manipulated datasets (or those of low quality) can be identified. An overview of the 32 structures and their  $R$  factors is given in Table 2.

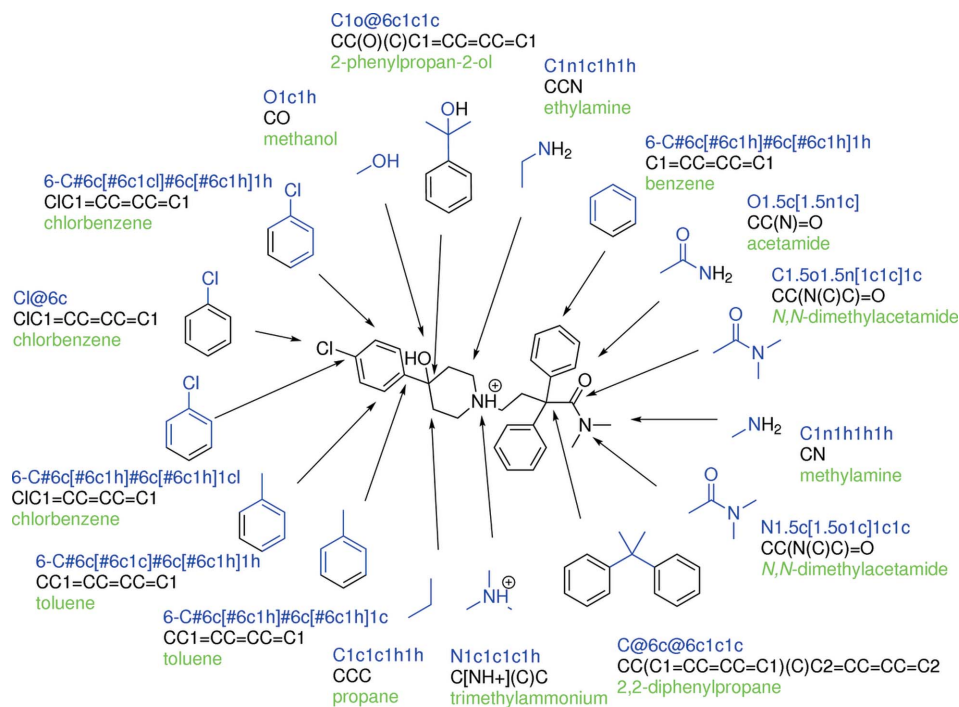
## 7. Holstein plots

Experience with these and other molecules has shown that in case the calculation of new model compounds is required, a systematic analysis of the chemical environments of each atom in a real crystal structure is recommended before embarking on time-consuming calculations. A correct choice of the best model compound for each invariom can be ensured by drawing the molecular structure surrounded by the model compounds for each atom (or at least those that require more complicated model compounds). Such plots should also contain the IUPAC name, the SMILES string (to be able to generate the respective model-compound starting geometries on the computer) and the invariom name assigned to each atom. Such plots were first introduced in a recent publication (Holstein *et al.*, 2012); one of them is included in Fig. 4. Many more Holstein plots for most of the compounds studied in §6 and Table 2 can be found in the supplementary material.

### 7.1. Current limitations and future developments

A central aim for the present version of the invariom database was to achieve a good coverage of the most frequent chemical environments in organic chemistry, including the elements C, H, N and O. While testing the example structures in §6 and numerous other structures has shown that a high reliability has been reached, some bonding environments require close attention of the user. One example would be the

N atom as part of the two condensed rings in  $\beta$ -lactam antibiotics, which is forced into a pyramidal conformation due to the fused ring system. Including only nearest (or next-nearest) neighbors to generate model compounds does not necessarily take this case into account. In general, special care is required every time a difference in local experimental and optimized geometry of the model compounds occurs. Fortunately, strained geometries are often accompanied by changes of the bond-distinguishing parameter  $\chi$ , so that often our assignment is nevertheless reliable. We therefore recommend visualization of the deformation density in *MoleCoolQt* to locate potential problems: when the deformation density does not show the expected features in the midpoint of the bonds – although the coordinate system has been assigned correctly – it is conceivable that a tailor-made scattering factor has to be generated to resolve a particular problem case. In the majority of cases involving



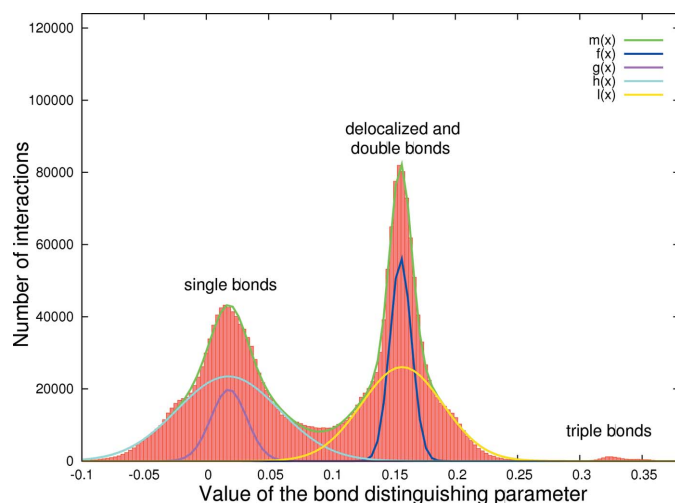
**Figure 4**

Exemplary Holstein plot for the anti-diarrhetic loperamide hydrochloride (Brüning *et al.*, 2012), example structure 27 in Table 2. The blue color helps to visualize the next-nearest neighbor sphere of an atom of interest. Invariom, IUPAC names and SMILES strings of model compounds used to generate the invariom are given.

first- and second-row elements (where usually only *s*- and *p*-orbitals are involved in chemical bonding) the present approach is sufficient.

A problem in practical modeling and invariom assignment is data precision. While geometry optimization gives reliable and reproducible predictions, experimental data of average quality (more frequent for room-temperature data with enhanced atomic motion) can lead to differences in the bond-distinguishing parameter that is used to assign scattering factors and coordinate systems. Small changes in bond distances can therefore sometimes give a rather different invariom name. This can happen *e.g.* for nitrogen-containing compounds where a  $\chi$  value indicating a delocalized bond requires the inclusion of next-nearest neighbors: if the bond is found to be shorter than expected, leading to a double bond in the invariom name, next-nearest neighbors will not be present. On the contrary when theory predicts a double bond (where only nearest neighbors are included) and the bond appears to be delocalized because it is found to be too long, nearest neighbors will be missing. Fig. 5 shows the distribution of  $\chi$  values from an evaluation of selected compounds from the Cambridge Structural Database: while a clear distinction between single, double and triple bonds is straightforward, there is no minimum in the region between delocalized and double bonds, as can be seen from the fitted Gaussian functions.

Our solution to the problem was to implement a user-friendly routine in *MoleCoolQt* (Hübschle & Dittrich, 2011) that allows bond-distinguishing parameters to be changed in a graphical pop-up window with the middle-mouse button. When the user changes the  $\chi$  value, all other invarioms affected are changed as well unless an individual name is fixed; feedback is provided on whether the modified values generated an invariom that is present in the database. Changing a single  $\chi$  value often resolves several such problems at once.



**Figure 5**  
Evaluation of values for the bond-distinguishing parameter  $\chi$  for a sample of the Cambridge Structural Database. While single, double and triple bonds can be distinguished, delocalized bonds merge with single and with double bonds.

Inaccurate data sets where bond lengths do not allow  $\chi$  to be calculated correctly can also lead to problems setting up local atomic coordinate systems, requiring manual correction. When two bonds should be but cannot be distinguished based on their length, *InvariomTool* and *MoleCoolQt* cannot assign the correct coordinate system automatically. In such (rare) cases, a manual change of the coordinate system can then lead to a better model. This is, however, a technical problem of finding the right algorithm and not a problem of the invariom approach in general.

Examples where the current invariom notation will require extension or future modification are compounds containing boron icosahedra with three-membered rings, or metal-organic compounds with  $\eta^5$  bonds as present in cyclopentadienyl ligands. For boron icosahedra we plan to use our former notation where rings were not yet taken into account, listing only single bonds. Therefore, metal-organic compounds and compounds containing boron icosahedra are not covered at the present stage.

Problems with transferability can also occur for Si—O—Si single bonds, where the bond angle is known to be flexible, probably due to the fact that the chemical interaction is predominantly ionic. Dative metal-organic bonds also show the same geometric flexibility, and bonds involving *d*-orbitals likewise do not seem to fulfill the requirement of a transferable local atomic geometry well enough for a useful generalization. However, all of these cases can already be handled by Hirshfeld-atom refinement (Jayatilaka & Dittrich, 2008), where the whole molecular wavefunction is used to derive aspherical atomic scattering factors. Hirshfeld-atom refinement can also be used for conventional data sets. Nevertheless, we will further investigate such cases and only after considerable experience is gained will an extension of the invariom approach be attempted for such bonding situations.

We use density functional theory DFT (the B3LYP functional) and the basis set D95++(3df,3pd) (Dunning, 1970) for our geometry optimizations. Re-optimization of the model compounds with a different basis set will be required to facilitate future extension of the invariom database to inorganic compounds including 3*d* metals, since the basis-set presently used is limited to elements up to krypton. Since there are many other challenges in modeling coordinative bonds in inorganic compounds we support only molecules containing the abovementioned elements at the current stage.

## 8. Conclusion

The invariom and other scattering-factor databases rely on decades of work and experience in the development and application of charge-density methodology. They can be applied in general small-molecule work and to larger molecules of biological interest. Property calculation of structures of normal resolution, as they are obtained in thousands of cases every year, is computationally undemanding and well established. It could be a routine outcome of a structure analysis. This paper reports on the introduction and application of a revised and extended version of the invariom data-

base, which now covers structures of organic molecules. Empirical rules for model-compound selection and the invariom notation were modified. Geometry optimizations of over 1300 model compounds were performed, resulting in tabulation of more than 2750 scattering-factor entries of valence-density environments of transferable pseudoatoms ('invarioms'). The database hence covers a respectable range of chemical environments, including main-block elements of the first, second and third row. The evolution of invariom notation is central to the generalized invariom database (GID): scattering-factor assignment based on geometry and the automation of most parts of the procedure have thereby been improved considerably.

Since we consider the current Hansen–Coppens multiple model not sufficiently accurate for calculation of interaction energies, the generalized invariom database was not optimized for that purpose. We will provide continuous updates of the database with new scattering factors on our web site. Most of the software required for database extension is available free of charge. Interested users are encouraged to join the effort and to perform geometry optimizations of missing model compounds themselves. With these developments we think that – at least for the elements and the chemical environments contained in the database – replacing the independent atom model is straightforward; try it yourself.

This work was made possible by an Emmy Noether Fellowship of the DFG (Deutsche Forschungsgemeinschaft), grant DI921/3–1 and DI921/3–2, which is gratefully acknowledged. We thank Jens Lübber, Claudia Orben and Jannis Neugeboren for contributing invarioms and Peter Luger, Francesca Fabbiani, Ulli Englert, Ruimin Wang, Christian Lehmann, Daniel Kratzert, John Bacsa, Akmal Tojiboev, Yulia Nelyubina and Simon Grabowsky for testing the new database. We also thank George Sheldrick, Luc Bourhis and Dylan Jayatilaka for interesting discussions and Stefan Siemsen for corrections in the manuscript.

## References

- Abdalla, M. A., Yadav, P. P., Dittrich, B., Schäffler, A. & Laatsch, H. (2011). *Org. Lett.* **13**, 2156–2159.
- Abramov, Yu. A. (1997). *Acta Cryst.* **A53**, 264–272.
- Abramov, Y. A., Volkov, A., Wu, G. & Coppens, P. (2000a). *Acta Cryst.* **A56**, 585–591.
- Abramov, Y. A., Volkov, A., Wu, G. & Coppens, P. (2000b). *J. Phys. Chem. B*, **104**, 2183–2188.
- Albrecht, M., Borba, A., Barbu-Debus, K. L., Dittrich, B., Fausto, R., Grimme, S., Mahjoub, A., Nedic, M., Schmitt, U., Schrader, L., Suhm, M. A., Zehnacker-Rentien, A. & Zischang, J. (2010). *New J. Chem.* **34**, 1266–1285.
- Allen, F. H. & Bruno, I. J. (2010). *Acta Cryst.* **B66**, 380–386.
- Allred, A. L. & Rochow, E. G. (1958). *J. Inorg. Nucl. Chem.* **5**, 264–268.
- Atria, A. M., Garland, M. T. & Baggio, R. (2012). *Acta Cryst.* **C68**, o392–o394.
- Bader, R. F. W. (1990). *Atoms in Molecules: A Quantum Theory*, 1st ed. Oxford: Clarendon Press.
- Bak, J. M., Domagała, S., Hübschle, C., Jelsch, C., Dittrich, B. & Dominiak, P. M. (2011). *Acta Cryst.* **A67**, 141–153.
- Becke, A. D. & Edgecombe, K. E. (1990). *J. Chem. Phys.* **92**, 5397–5403.
- Blake, A. J., Chebude, Y., Tadesse, H. & Wondimu, B. (2012). *Acta Cryst.* **C68**, o362–o364.
- Blanco, M. C., Palma, A., Cobo, J. & Glidewell, C. (2012). *Acta Cryst.* **C68**, o195–o198.
- Blom, R. & Haaland, A. (1985). *J. Mol. Struct.* **128**, 21–27.
- Boeré, R. T. & Taghavikish, M. (2012). *Acta Cryst.* **C68**, o381–o382.
- Bojarska, J., Maniukiewicz, W., Sieroń, L., Fruziński, A., Kopczacki, P., Walczyński, K. & Remko, M. (2012). *Acta Cryst.* **C68**, o341–o343.
- Bouhmaida, N., Bonhomme, F., Guillot, B., Jelsch, C. & Ghermani, N. E. (2009). *Acta Cryst.* **B65**, 363–374.
- Brock, C. P., Dunitz, J. D. & Hirshfeld, F. L. (1991). *Acta Cryst.* **B47**, 789–797.
- Brüning, J., Podgorski, D., Alig, E., Bats, J. W. & Schmidt, M. U. (2012). *Acta Cryst.* **C68**, o111–o113.
- Coppens, P. (1997). *X-ray Charge Densities and Chemical Bonding*, No. 4, 1st ed. Oxford University Press.
- Dadda, N., Nassour, A., Guillot, B., Benali-Cherif, N. & Jelsch, C. (2012). *Acta Cryst.* **A68**, 452–463.
- Deringer, V. L., Hoepfner, V. & Dronskowski, R. (2012). *Cryst. Growth Des.* **12**, 1014–1021.
- Dittrich, B., Bond, C. S., Spackman, M. A. & Jayatilaka, D. (2010). *CrystEngComm*, **12**, 2419–2423.
- Dittrich, B., Hübschle, C. B., Luger, P. & Spackman, M. A. (2006). *Acta Cryst.* **D62**, 1325–1335.
- Dittrich, B., Hübschle, C. B., Messerschmidt, M., Kalinowski, R., Girnt, D. & Luger, P. (2005). *Acta Cryst.* **A61**, 314–320.
- Dittrich, B. & Jayatilaka, D. (2012). *Struct. Bond.* **147**, 27–46.
- Dittrich, B., Koritsánszky, T., Grosche, M., Scherer, W., Flaig, R., Wagner, A., Krane, H. G., Kessler, H., Riemer, C., Schreurs, A. M. M. & Luger, P. (2002). *Acta Cryst.* **B58**, 721–727.
- Dittrich, B., Koritsánszky, T. & Luger, P. (2004). *Angew. Chem. Int. Ed.* **43**, 2718–2721.
- Dittrich, B., Munshi, P. & Spackman, M. A. (2006). *Acta Cryst.* **C62**, o633–o635.
- Dittrich, B., Munshi, P. & Spackman, M. A. (2007). *Acta Cryst.* **B63**, 505–509.
- Dittrich, B., Strumpel, M., Schäfer, M., Spackman, M. A. & Koritsánszky, T. (2006). *Acta Cryst.* **A62**, 217–223.
- Dittrich, B., Sze, E., Holstein, J. J., Hübschle, C. B. & Jayatilaka, D. (2012). *Acta Cryst.* **A68**, 435–442.
- Dittrich, B., Weber, M., Kalinowski, R., Grabowsky, S., Hübschle, C. B. & Luger, P. (2009). *Acta Cryst.* **B65**, 749–756.
- Djigoue, G.-B., Simard, M., Kenmogne, L.-C. & Poirier, D. (2012). *Acta Cryst.* **C68**, o231–o234.
- Domagała, S., Fournier, B., Liebschner, D., Guillot, B. & Jelsch, C. (2012). *Acta Cryst.* **A68**, 337–351.
- Domagała, S. & Jelsch, C. (2008). *J. Appl. Cryst.* **41**, 1140–1149.
- Domagała, S., Munshi, P., Ahmed, M., Guillot, B. & Jelsch, C. (2011). *Acta Cryst.* **B67**, 63–78.
- Dominiak, P. M., Volkov, A., Dominiak, A. P., Jarzemska, K. N. & Coppens, P. (2009). *Acta Cryst.* **D65**, 485–499.
- Dominiak, P. M., Volkov, A., Li, X., Messerschmidt, M. & Coppens, P. (2007). *J. Chem. Theory Comput.* **2**, 232–247.
- Dunning, T. H. (1970). *J. Chem. Phys.* **53**, 2823–2833.
- Espinosa, E., Alkorta, I., Elguero, I. R. J. & Molins, E. (2001). *Chem. Phys. Lett.* **336**, 457–461.
- Espinosa, E., Molins, E. & Lecomte, C. (1998). *Chem. Phys. Lett.* **285**, 170–173.
- Fadzillah, S. M. H., Ngaini, Z., Hussain, H., Razak, I. A. & Asik, S. I. J. (2012). *Acta Cryst.* **E68**, o2909.
- Fischer, A., Tian, D., Scherer, W., Batke, K., Eickerling, G., Svendsen, H., Bindzus, N. & Iversen, B. (2011). *J. Phys. Chem. A*, **115**, 13061–13071.
- Flack, H. D. (1983). *Acta Cryst.* **A39**, 876–881.
- Flack, H. D. & Bernardinelli, G. (2000). *J. Appl. Cryst.* **33**, 1143–1148.

- Frisch, M. J. *et al.* (2009). *GAUSSIAN09*, Revision A.02. Gaussian, Inc., Wallingford CT, USA.
- Gatti, C. (2005). *Z. Kristallogr.* **220**, 399–457.
- Gelbrich, T., Threlfall, T. L. & Hursthouse, M. B. (2012). *Acta Cryst.* **C68**, o421–o426.
- Gherase, D., Naubron, J.-V., Roussel, C. & Giorgi, M. (2012). *Acta Cryst.* **C68**, o247–o252.
- Grabowsky, S., Kalinowski, R., Weber, M., Förster, D., Paulmann, C. & Luger, P. (2009). *Acta Cryst.* **B65**, 488–501.
- Grabowsky, S., Luger, P., Buschmann, J., Schneider, T., Schirmeister, T., Sobolev, A. N. & Jayatilaka, D. (2012). *Angew. Chem. Int. Ed.* **51**, 6776–6779.
- Guillot, B., Jelsch, C., Podjarny, A. & Lecomte, C. (2008). *Acta Cryst.* **D64**, 567–588.
- Guillot, B., Viry, L., Guillot, R., Lecomte, C. & Jelsch, C. (2001). *J. Appl. Cryst.* **34**, 214–223.
- Hansen, N. K. & Coppens, P. (1978). *Acta Cryst.* **A34**, 909–921.
- Hanwell, M. D., Curtis, D. E., Lonie, D. C., Vandermeersch, T., Zurek, E. & Hutchison, G. R. (2012). *J. Cheminformatics*, **4**, 17.
- Hathwar, V. R., Thakur, T. S., Dubey, R., Pavan, M. S., Row, T. N. G. & Desiraju, G. R. (2011). *J. Phys. Chem. A*, **115**, 12852–12863.
- Hathwar, V. R., Thakur, T. S., Row, T. N. G. & Desiraju, G. R. (2011). *Cryst. Growth Des.* **11**, 616–623.
- Helliwell, M., Moosun, S., Bhowon, M. G., Jhaumeer-Laulloo, S. & Joule, J. A. (2012). *Acta Cryst.* **C68**, o387–o391.
- Hirshfeld, F. L. (1976). *Acta Cryst.* **A32**, 239–244.
- Hirshfeld, F. L. (1977). *Theoret. Chim. Acta*, **44**, 129–138.
- Hirshfeld, F. L. (1992). *Accurate Molecular Structures. Their Determination and Importance*, edited by A. Domenicano & I. Hargittai, pp. 237–269. Oxford University Press.
- Holstein, J. J., Hübschle, C. B. & Dittrich, B. (2012). *CrystEngComm*, **14**, 2520–2531.
- Holstein, J. J., Luger, P., Kalinowski, R., Mebs, S., Paulman, C. & Dittrich, B. (2010). *Acta Cryst.* **B66**, 568–577.
- Housset, D., Benabicha, F., Pichon-Pesme, V., Jelsch, C., Maierhofer, A., David, S., Fontecilla-Camps, J. C. & Lecomte, C. (2000). *Acta Cryst.* **D56**, 151–160.
- Howard, S. T., Hursthouse, M. B., Lehmann, C. W., Mallinson, P. R. & Frampton, C. S. (1992). *J. Chem. Phys.* **97**, 5616–5630.
- Hübschle, C. B. & Dittrich, B. (2011). *J. Appl. Cryst.* **44**, 238–240.
- Hübschle, C. B., Dittrich, B., Grabowsky, S., Messerschmidt, M. & Luger, P. (2008). *Acta Cryst.* **B64**, 363–374.
- Hübschle, C. B., Luger, P. & Dittrich, B. (2007). *J. Appl. Cryst.* **40**, 623–627.
- Jarzemska, K. N. & Dominiak, P. M. (2012). *Acta Cryst.* **A68**, 139–147.
- Jayatilaka, D. (1994). *Chem. Phys. Lett.* **230**, 228–230.
- Jayatilaka, D. (1998). *Phys. Rev. Lett.* **80**, 798–801.
- Jayatilaka, D. & Dittrich, B. (2008). *Acta Cryst.* **A64**, 383–393.
- Jayatilaka, D. & Grimwood, D. J. (2001). *Acta Cryst.* **A57**, 76–86.
- Jayatilaka, D. & Grimwood, D. J. (2003). *Comput. Sci.* **2660**, 142–151.
- Jelsch, C., Guillot, B., Lagoutte, A. & Lecomte, C. (2005). *J. Appl. Cryst.* **38**, 38–54.
- Jelsch, C., Pichon-Pesme, V., Lecomte, C. & Aubry, A. (1998). *Acta Cryst.* **D54**, 1306–1318.
- Jelsch, C., Teeter, M. M., Lamzin, V., Pichon-Pesme, V., Blessing, R. H. & Lecomte, C. (2000). *Proc. Natl Acad. Sci. USA*, **97**, 3171–3176.
- Johnas, S. K. J., Dittrich, B., Meents, A., Messerschmidt, M. & Weckert, E. F. (2009). *Acta Cryst.* **D65**, 284–293.
- Keller, M., Bhadbhade, M. M. & Read, R. W. (2012). *Acta Cryst.* **C68**, o240–o246.
- Khalaji, A. D., Fejfarová, K. & Dušek, M. (2012). *Acta Cryst.* **E68**, o2646.
- Kingsford-Adaboh, R., Dittrich, B., Hübschle, C. B., Gbewonyo, W. S. K., Okamoto, H., Kimura, M. & Ishida, H. (2006). *Acta Cryst.* **B62**, 843–849.
- Koritsánszky, T., Richter, T., Macchi, P., Volkov, A., Gatti, C., Howard, S., Mallinson, P. R., Farrugia, L., Su, Z. W. & Hansen, N. K. (2003). *XD*, Technical Report. Freie Universität Berlin, Germany.
- Koritsánszky, T., Volkov, A. & Chodkiewicz, C. (2012). *Struct. Bond.* **147**, 1–26.
- Koritsánszky, T., Volkov, A. & Coppens, P. (2002). *Acta Cryst.* **A58**, 464–472.
- Koritsánszky, T. S. & Coppens, P. (2001). *Chem. Rev.* **101**, 1583–1628.
- Li, X., Volkov, A. V., Szalewicz, K. & Coppens, P. (2006). *Acta Cryst.* **D62**, 639–647.
- Li, X., Wu, G., Abramov, Y. A., Volkov, A. V. & Coppens, P. (2002). *Proc. Natl Acad. Sci.* **99**, 12132–12137.
- Liu, Y.-L., Zou, P., Wu, H., Xie, M.-H. & Luo, S.-N. (2012). *Acta Cryst.* **C68**, o338–o340.
- Luger, P. & Dittrich, B. (2007). *The Quantum Theory of Atoms in Molecules*, edited by C. F. Matta & R. J. Boyd, pp. 317–342. Weinheim: Wiley VCH.
- Madsen, A. Ø., Mattson, R. & Larsen, S. (2011). *J. Phys. Chem. A*, **115**, 7794–7804.
- Matos, M. J., Santana, L. & Uriarte, E. (2012). *Acta Cryst.* **E68**, o2645.
- Nelyubina, Y. V., Glukhov, I. V., Antipin, M. Y. & Lyssenko, K. A. (2010). *Chem. Commun.* **46**, 3469–3471.
- Pérez, H., Corrêa, R. S., Plutín, A. M., O'Reilly, B. & Andrade, M. B. (2012). *Acta Cryst.* **C68**, o19–o22.
- Petricek, V., Dusek, M. & Palatinus, L. (2006). *JANA2006*. Institute of Physics, Praha, Czech Republic.
- Pichon-Pesme, V., Jelsch, C., Guillot, B. & Lecomte, C. (2004). *Acta Cryst.* **A60**, 204–208.
- Pichon-Pesme, V., Lecomte, C. & Lachekar, H. (1995). *J. Phys. Chem.* **99**, 6242–6250.
- Ponomarova, V. V. & Domasevitch, K. V. (2012). *Acta Cryst.* **C68**, o359–o361.
- Poulain-Paul, A., Nassour, A., Jelsch, C., Guillot, B., Kubicki, M. & Lecomte, C. (2012). *Acta Cryst.* **A68**, 715–728.
- Pourayoubi, M., Jasinski, J. P., Shoghpour Bayraq, S., Eshghi, H., Keeley, A. C., Bruno, G. & Amiri Rudbari, H. (2012). *Acta Cryst.* **C68**, o399–o404.
- Prelog, V. & Helmchen, G. (1982). *Angew. Chem. Int. Ed.* **21**, 567–583.
- Pröpper, K., Holstein, J. J., Hübschle, C. B., Bond, C. S. & Dittrich, B. (2013). *Acta Cryst.* Submitted.
- Quiroga, J., Díaz, Y., Cobo, J. & Glidewell, C. (2012). *Acta Cryst.* **C68**, o12–o18.
- Rappé, A. K., Casewit, C. J., Colwell, K. S., Goddard, W. A., Skid, W. M. & Bernstein, E. R. (1992). *J. Am. Chem. Soc.* **114**, 10024–10039.
- Rödel, E., Messerschmidt, M., Dittrich, B. & Luger, P. (2006). *Org. Biomol. Chem.* **4**, 475–481.
- Schomaker, V. & Stevenson, D. P. (1941). *J. Am. Chem. Soc.* **63**, 37–40.
- Schürmann, C. J., Pröpper, K., Wagner, T. & Dittrich, B. (2012). *Acta Cryst.* **B68**, 313–317.
- Seela, F., Mei, H., Xiong, H., Budow, S., Eickmeier, H. & Reuter, H. (2012). *Acta Cryst.* **C68**, o395–o398.
- Seiler, P. (1992). *Accurate Molecular Structures. Their Determination and Importance*, edited by A. Domenicano & I. Hargittai, pp. 170–198. Oxford University Press.
- Sheldrick, G. M. (2008). *Acta Cryst.* **A64**, 112–122.
- Smith, G. & Wermuth, U. D. (2012). *Acta Cryst.* **C68**, o253–o256.
- Sonar, V. N., Parkin, S. & Crooks, P. A. (2012). *Acta Cryst.* **C68**, o405–o407.
- Sousa, A. S. de, Sannasy, D., Fernandes, M. A. & Marques, H. M. (2012). *Acta Cryst.* **C68**, o383–o386.
- Sowa, M., Ślepokura, K. & Matczak-Jon, E. (2012). *Acta Cryst.* **C68**, o262–o265.
- Spackman, M. A. (1992). *Chem. Rev.* **92**, 1769–1797.
- Spackman, M. A. (2007). *Acta Cryst.* **A63**, 198–200.
- Spackman, M. A., Munshi, P. & Dittrich, B. (2007). *ChemPhysChem*, **8**, 2051–2063.

- Stalke, D. (2011). *Chem. Eur. J.* **17**, 9264–9278.
- Stewart, R. F. (1976). *Acta Cryst.* **A32**, 565–574.
- Suresh, J., Vishnupriya, R., Sivakumar, S., Kumar, R. R. & Athimoolam, S. (2012). *Acta Cryst.* **C68**, o257–o261.
- Talontsi, F. M., Kenla, T. J. N., Dittrich, B., Douanla-Meli, C. & Laatsch, H. (2012). *Planta Med.* **78**, 1020–1023.
- Tutughamiarso, M., Pisternick, T. & Egert, E. (2012). *Acta Cryst.* **C68**, o344–o350.
- Volkov, A. & Coppens, P. (2001). *Acta Cryst.* **A57**, 395–405.
- Volkov, A., Koritsánszky, T., Chodkiewicz, M. & King, H. F. (2009). *J. Comput. Chem.* **30**, 1379–1391.
- Volkov, A., Koritsánszky, T. & Coppens, P. (2004). *Chem. Phys. Lett.* **391**, 170–175.
- Volkov, A., Koritsánszky, T., Li, X. & Coppens, P. (2004). *Acta Cryst.* **A60**, 638–639.
- Volkov, A., Li, X., Koritsánszky, T. & Coppens, P. (2004). *J. Phys. Chem. A*, **108**, 4283–4300.
- Volkov, A., Macchi, P., Farrugia, L. J., Gatti, C., Mallinson, P., Richter, T. & Koritsánszky, T. (2006). *XD2006*. University at Buffalo, State University of New York, NY, USA; University of Milano, Italy; University of Glasgow, UK; CNRISTM, Milano, Italy; Middle Tennessee State University, TN, USA; Freie Universität, Berlin, Germany.
- Volkov, A., Messerschmidt, M. & Coppens, P. (2007). *Acta Cryst.* **D63**, 160–170.
- Wagner, F. R., Bezugly, V., Kohout, M. & Grin, Y. (2007). *Chem. Eur. J.* **13**, 5724.
- Weininger, D. (1988). *J. Chem. Inf. Comput. Sci.* **28**, 31–36.
- Woińska, M. & Dominiak, P. M. (2011). *J. Phys. Chem. A*, doi: 10.1021/jp204010v.
- Yadav, P. P., Nair, V., Dittrich, B., Schüffler, A. & Laatsch, H. (2010). *Org. Lett.* **12**, 3800–3803.
- Zarychta, B., Pichon-Pesme, V., Guillot, B., Lecomte, C. & Jelsch, C. (2007). *Acta Cryst.* **A63**, 108–125.