

Estimating temperature-dependent anisotropic hydrogen displacements with the invariom database and a new segmented rigid-body analysis program

Jens Lübben,^a Luc J. Bourhis^b and Birger Dittrich^{a,c*}

Received 2 July 2015

Accepted 27 September 2015

Edited by D. I. Svergun, European Molecular Biology Laboratory, Hamburg, Germany

Keywords: anisotropic displacement parameters; generalized invariom database; segmented rigid-body analysis.

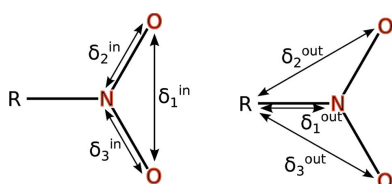
^aInstitut für Anorganische und Angewandte Chemie, Martin-Luther-King-Platz 6, 20146 Hamburg, Germany, ^bBruker-AXS SAS, 4 Allée Lorents, F-77447 Marne-la-Vallée, France, and ^cHeinrich-Heine Universität Düsseldorf, Anorganische Chemie und Strukturchemie, Universitätsstrasse 1, Gebäude 26.42.01.21, 40225 Düsseldorf, Germany. *Correspondence e-mail: dittrich@hhu.de

Invariom partitioning and notation are used to estimate anisotropic hydrogen displacements for incorporation in crystallographic refinement models. Optimized structures of the generalized invariom database and their frequency computations provide the information required: frequencies are converted to internal atomic displacements and combined with the results of a TLS (translation–libration–screw) fit of experimental non-hydrogen anisotropic displacement parameters to estimate those of H atoms. Comparison with TLS+ONIOM and neutron diffraction results for four example structures where high-resolution X-ray and neutron data are available show that electron density transferability rules established in the invariom approach are also suitable for streamlining the transfer of atomic vibrations. A new segmented-body TLS analysis program called *APD-Toolkit* has been coded to overcome technical limitations of the established program *THMA*. The influence of incorporating hydrogen anisotropic displacement parameters on conventional refinement is assessed.

1. Introduction

Improving the accuracy of structural information derived from conventional single-crystal X-ray diffraction (XRD) experiments has been the initial aim for transferring aspherical scattering factors (Brock *et al.*, 1991) and it remains the central aim of the generalized invariom database (GID; Dittrich *et al.*, 2013). Deriving molecular properties from aspherical electron density is another important and closely related aim, since such properties can only be as accurate as the best possible set of coordinates that can be refined from a given data set.

Anisotropic displacement parameters (ADPs) of atoms and their electron density distribution (EDD) are correlated (Hirshfeld, 1976). It has hence been shown to be beneficial to incorporate hydrogen ADPs for obtaining accurate EDDs by least-squares refinement of multipole parameters against high-resolution data (Madsen *et al.*, 2004). Since this must also hold true for parameters derived from conventional data sets, our recent efforts concern the estimation of temperature-dependent hydrogen ADPs (H-ADPs) for selected applications in structure determinations with the independent atom model (IAM) as well as refinement with aspherical scattering factors. This includes charge density (CD) studies and refinements of data sets of normal resolution ($\sin \theta/\lambda = 0.6 \text{ \AA}^{-1}$ or



OPEN ACCESS

$d = 0.84 \text{ \AA}$) using scattering-factor databases, of which the GID is one.¹

Scattering of H atoms is limited in reciprocal space owing to their comparably low scattering contribution and their missing core density. XRD therefore has limitations in locating their positions and displacements accurately (Cooper *et al.*, 2010), a situation that has already been improved considerably by scattering factor databases (Dittrich *et al.*, 2005) and Hirshfeld atom refinement (Jayatilaka & Dittrich, 2008; Capelli *et al.*, 2014), as comparisons with results from neutron diffraction have shown. These developments allow free refinement of hydrogen parameters but require good low-order data (Orben & Dittrich, 2014).

Several authors have proposed improved or alternative hydrogen treatment in XRD, mainly for CD work. With respect to positional parameters Hoser *et al.* (2009) recommend to only use low-order reflections from a high-resolution data set to determine the $X-H$ directions, and then elongate to average neutron diffraction values (Allen & Bruno, 2010), whereas we advocate the use of calculated positions and $X-H$ distances from the invariom database (Schürmann *et al.*, 2012), a procedure also applicable to data sets of low quality.

The *SHADE* (simple hydrogen anisotropic displacement estimator; Madsen, 2006) and *SHADE2* servers (Munshi *et al.*, 2008) can provide estimates of H-ADPs by combining a TLS (translation–libration–screw) fit (Schomaker & Trueblood, 1968) of the non-H atoms with average internal modes tabulated from neutron diffraction. The *SHADE2* server has established its usefulness in CD research.

Other ways to estimate ADPs for H atoms have been developed. Displacements can likewise be computed from spectroscopic data as implemented in the *SHADE3* server (Roversi & Destro, 2004; Madsen & Hoser, 2014). This idea was applied first by Hirshfeld & Hope (1980). One can also carry out theoretical optimizations of isolated molecular structures (Flaig *et al.*, 1998) or employ QM/MM cluster computations to retrieve the structure found in the crystal (Whitten & Spackman, 2006). Computed frequencies can subsequently be converted into internal atomic displacements, which are again combined with a TLS analysis² after appropriate scaling (Scott & Radom, 1996). Last but not least, full periodic computations, implemented in the *SHADE3* server, may also provide H-ADPs (Madsen *et al.*, 2013; Madsen & Hoser, 2014). However, all these approaches have disadvan-

tages: estimates derived from diffraction data do not take into account temperature dependence of the internal contribution of atomic vibration;³ neutron data for the *SHADE2* approach are not available for rare bonding environments; theoretical studies require high computational costs and are thus unsuited for conventional structure determinations. This is why we introduce a new approach based on the invariom database, combined with a new freely available TLS analysis program.

Our approach relies on the geometry-optimized model compounds in the invariom database.⁴ It covers a wide range of chemical environments in organic chemistry (Dittrich *et al.*, 2013) and now also facilitates aspherical-atom refinements of coordination compounds (Dittrich *et al.*, 2015). Earlier work is here extended by providing functionality to estimate H-ADPs relying on the established empirical rules of partitioning electron density with invarioms. The rules already allow one to separate and reconstruct molecular EDDs from fragments and now also provide estimates of the internal modes of vibration of a particular chemical environment.

Estimation of H-ADPs thus allows further improvement of all those structures where chemical environments are covered by scattering factors of the GID. Moreover, estimated H-ADPs increase the choices in handling three common situations: (a) data of low quality can be better evaluated by reducing the number of refined parameters; (b) high-quality data of comparably low resolution are available; or (c) refinement of H-atom positions becomes an option when aspherical scattering factors and ADPs are kept fixed, thereby reaching better agreement with results from neutron diffraction and bond-length predictions of quantum chemistry.

Central to this work is the underlying development of a new segmented-body (Schomaker & Trueblood, 1998) TLS refinement program called *APD-Toolkit* (anisotropic proton displacement toolkit), which is introduced here.

2. Automatic segmented rigid-body analysis

A simple approach that can provide information on the coupling between internal and external displacements is to assume segmented rigid-body motion. Our implementation analyzes the shape of all measured ADPs and determines how attached rigid groups should be added to the otherwise rigid body to best fit the observed ADPs. After internal and external contributions are estimated, a displacement model for H atoms is then generated by adding both contributions. The well known Fortran77 program for TLS fits *THMA14c* (Schomaker & Trueblood, 1998) is limited to 230 atoms in the asymmetric unit and can only handle up to seven manually defined attached rigid groups. These limitations were our motivation to develop a more flexible solution. Our program was developed to estimate the ADPs of H atoms and will be discussed next.

³ Neglecting temperature dependence of internal vibrational modes leads to systematic errors in the ADPs of H atoms (Lübben *et al.*, 2014).

⁴ So far we have used the term 'invariom database' synonymously with 'invariom scattering factor database'. Since now further functionality is being added, the invariom database becomes a database of model compounds from which several properties are derived.

¹ Four scattering-factor databases currently exist: the 'supramolecular-synthon based fragments approach' (SBFA; Hathwar *et al.*, 2011), the 'experimental library multipolar atom model' (ELMAM2; Zarychta *et al.*, 2007; Domagała *et al.*, 2012) (both based on high-resolution experiments), the 'generalized invariom database' (GID; Dittrich *et al.*, 2006, 2013) and the 'University at Buffalo Databank' (UBDB2011; Dominiak *et al.*, 2007; Jarzemska & Dominiak, 2012) (the latter two based on theoretical DFT computations). All four rely on the established Hansen/Coppens multipole model (Hansen & Coppens, 1978) and can successfully be used to improve the accuracy and precision of least-squares structure refinements.

² In the context of this manuscript the term 'TLS analysis' is used for a post-refinement analysis of the atomic displacement parameters determined by the refinement program. The term 'TLS refinement' is not used to avoid confusion with procedures that apply a TLS model during the refinement process (Merritt, 1999).

2.1. Workflow of the program

The *APD-Toolkit* carries out the following steps:

- (1) Determination of invariom names of all atoms.
- (2) Calculation of internal displacement parameters from *GAUSSIAN* (Frisch *et al.*, 2013) output files and caching of results⁵ for subsequent applications.
- (3) Transformation of internal ADPs to the crystal coordinate system.
- (4) Calculation of the difference between observed and calculated internal ADPs for all non-H atoms to remove contamination of the TLS parameters with internal ADPs.
- (5) Determination of a suitable segmentation model for the segmented rigid-body analysis.
- (6) Computation of a physically meaningful set of TLS+ARG (attached rigid group) parameters describing ADP differences.
- (7) Computation of external ADPs for all H atoms based on the TLS+ARG parameters and the atomic coordinates.
- (8) Estimation of H-ADPs by adding internal and external contributions.

2.2. Automated rigid-body segmentation

APD-Toolkit automatically analyzes the shape of non-H-atom ADPs to obtain a suitable segmentation model for a segmented rigid-body analysis. In contrast to similar procedures in protein refinement (Painter & Merritt, 2006) the method implemented analyzes the refined model to find a physically plausible segmentation model instead of finding the model that minimizes the $R_1(F)$ value.

The procedure works as follows. In a first step all single bonds in the molecule are flagged as potential axes for vibrations along a torsion angle (Blom & Haaland, 1985). For every potential axis the molecule is then separated into two parts that are connected only by one bond representing the rotation axis. The smaller of the two groups is considered to be the attached rigid group (Schomaker & Trueblood, 1998); the larger one is the rigid body. In a next step the difference of ADPs (δ_i^{in}) of a pair of bonded atoms in the direction of the connecting vector is computed for all atom pairs within the attached rigid group. In addition, the corresponding value (δ_i^{out}) is determined for all atom pairs where one atom is part of the attached rigid group and the other atoms are part of the rigid body. For every potential rotation axis the 'rigidity index' Ω is then determined, as defined in equation (1) and illustrated in Fig. 1. If Ω is negative, the implied attached rigid group is accepted. The expression of Ω is purely empirical.

⁵ Frequency information is extracted from all *GAUSSIAN* output files and stored in a database linking normal modes to the model compound including its geometry. This reduces the amount of data to store from several hundred gigabytes to about 100 MB. In a second step, when the method is applied to a structural model, the database is loaded and displacement parameters are calculated for all model compounds at the given temperature. The results are stored in a second database file now containing all internal ADPs at a certain temperature. The file is about 4 MB in size. During every subsequent application of the method it is checked whether the database already contains ADPs for the current temperature and ADPs are only recomputed if the method is applied at this temperature for the first time. Otherwise the cached results are used.

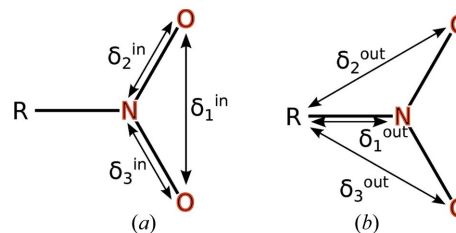


Figure 1

Illustration of the rigidity index: the NO_2 group attached to a molecule R is considered rigid if the average of δ_i^{out} is twice as large as the average value of δ_i^{in} .

The factor ε is used to control the weight between the rigidity of the ARG and the flexibility relative to the rest of the group. A value of $\varepsilon = 2$ gave the most reasonable results in our studies.

$$\Omega = \frac{\varepsilon}{n} \sum_{i=0}^n \delta_i^{\text{in}} - \frac{1}{n} \sum_{i=0}^n \delta_i^{\text{out}}. \quad (1)$$

The groups are cross-referenced after all rigid groups are assigned to ensure that they all consist of at least eight atoms. Smaller groups are discarded, since they would not allow stable optimization.⁶ After a suitable segmentation model is found, a least-squares optimization is carried out to find the optimal TLS+ARG parameters.

The procedure is applied for each molecule in cases where the asymmetric unit contains more than one molecule. Adaptation of the procedure for disordered compounds and molecules on special positions is planned only for future versions.

3. Results and discussion

3.1. Similarity of ADPs from TLS+ONIOM and TLS+INV⁷ estimates: similarity of $U(2)$ from TLS+ONIOM and TLS+INV

For initial validation, theoretical ADPs taken from the generalized invariom database were compared with those obtained from ONIOM computations (Svensson *et al.*, 1996) to assess the transferability of internal ADPs. Computations were performed with the B3LYP functional and the basis set combination 6-31G(d,p):3-21G, which has been shown to be a good compromise between computational requirements and the quality of results (Dittrich *et al.*, 2012). Internal ADPs are not compared directly since the internal parts of the ADPs encompass different parts of the overall displacements. Instead the values of internal displacements were combined

⁶ Theoretically two atoms, each providing six data points, should be sufficient since every attached rigid group adds seven parameters to the TLS model. The ADPs are, however, not completely independent of each other. Practical application shows a size of eight atoms to be a reasonable choice. If all potentially rigid groups consist of fewer than eight atoms, a rigid-body model is employed. Furthermore, TLS analysis fails if all atoms of a rigid group are placed on a conic section. Since this is likely for groups of six or fewer atoms, a group size of eight atoms avoids this problem in almost all cases. If a molecule consists of fewer than six atoms, the method cannot be applied and ADPs from the input file are used for the corresponding atoms.

⁷ INV stands for internal contributions computed from the generalized invariom database here.

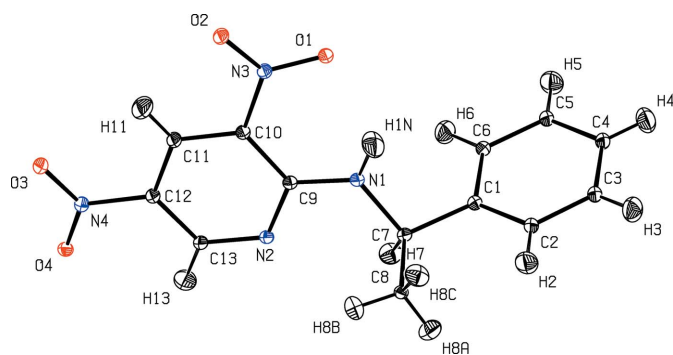


Figure 2
Structural model of methylbenzylaminodinitropyridine (MBADNP) at 20 K (Cole *et al.*, 2002) with ADPs estimated with the TLS+INV approach.

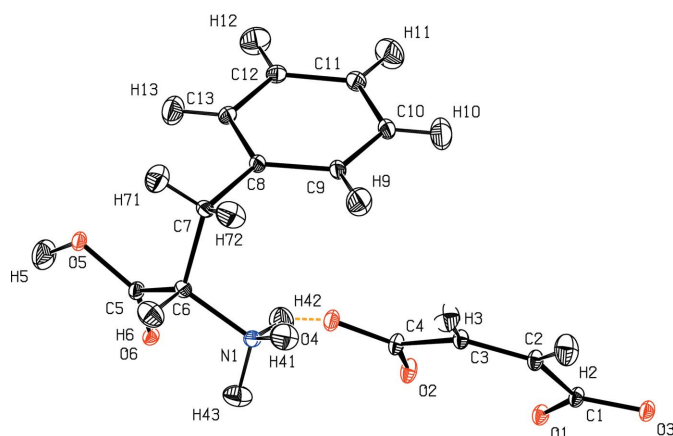


Figure 3
Structural model of L-phenylalaninium hydrogen maleate at 12 K (Woińska *et al.*, 2014) with ADPs estimated with the TLS+INV approach.

with external displacements derived from TLS analysis, analogously to §3.3. For estimating internal ADPs a low-frequency cutoff value of 200 cm⁻² was used (Madsen *et al.*, 2013).⁸

Structural models of four test structures obtained by the TLS+INV approach are shown in Figs. 2–5.⁹

To quantitatively compare ADPs obtained by different methods a procedure proposed by Whitten and Spackman was employed (Whitten & Spackman, 2006; Munshi *et al.*, 2008). This procedure determines the spacial overlap of two sets of ADPs. It yields a value of the comparison parameter (*S*) of zero if both ADPs are identical and a value of 100 if the ADPs do not overlap at all. *S* is computed as

$$S = 100 \left\{ 1 - \frac{2^{2/3} (\det U_1^{-1} U_2^{-1})^{1/4}}}{[\det(U_1^{-1} + U_2^{-1})]^{1/2}} \right\}. \quad (2)$$

⁸ Inaccuracies introduced in the estimation of internal displacements in the TLS+INV model are absorbed in the TLS part. We argue that the threshold separating internal and external displacements is arbitrary to some extent. Contributions to the displacements that are caused by vibrations with frequencies close to the threshold can be modeled by an internal and an external displacement model alike.

⁹ ONIOM computations for dimethylbiguanidiniumbishydrogensquarate did not converge and were therefore not included in the comparison.

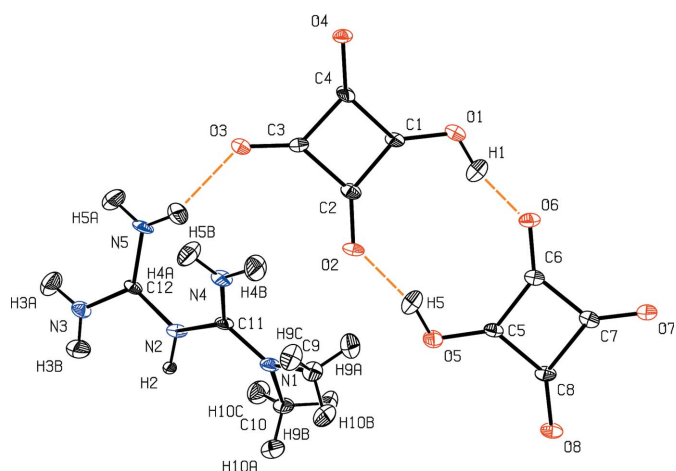


Figure 4
Structural model of dimethylbiguanidiniumbishydrogensquarate at 130 K (Šerb *et al.*, 2014) with ADPs estimated with the TLS+INV approach.

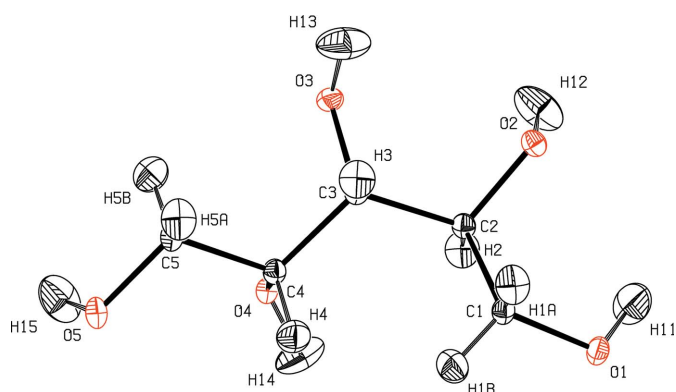


Figure 5
Structural model of xylitol at 122 K (Madsen *et al.*, 2003) with ADPs estimated with the TLS+INV approach.

Tables 1–3 indicate that the agreement between the two methods depends on whether or not an H atom is involved in hydrogen bonding. In these cases the ONIOM estimate is more realistic since the bonding interactions, which are omitted in the TLS+INV approach, add forces to the H atoms that counteract vibrational movement. For those atoms not involved in hydrogen bonding the agreement is good, especially in cases where the asymmetric unit is described as one overall rigid body. This is supported by the very small discrepancies seen in the structures of MBADNP and xylitol. In these cases the non-hydrogen-bonded atoms have nearly identical ADPs. When the asymmetric unit content is more flexible or contains more than one molecule the agreement becomes less good, as evident in the structure of L-phenylalaninium hydrogen maleate. Since the TLS+INV approach does not include intermolecular interactions it predicts larger ADPs than the ONIOM model, which approximates these interactions. Slightly larger differences seen for methyl-group H atoms can also be explained by intermolecular interactions: while the rotational movement of the methyl group around a C–X (*X* not hydrogen) single bond usually has a discrete minimum for an isolated molecule, intermolecular interactions

Table 1

Comparison of TLS+INV derived ADPs with TLS+ONIOM derived ADPs of MBADNP.

Label	<i>S</i>	Label	<i>S</i>
H11	0.31	H5	0.02
H13	0.07	H6	0.03
H1N	1.56	H7	0.02
H2	0.04	H8A	0.75
H3	0.03	H8B	0.52
H4	0.08	H8C	0.57
<i><H></i>	0.33		

Table 2

Comparison of TLS+INV derived ADPs with TLS+ONIOM derived ADPs of L-phenylalaninium hydrogen maleate.

Label	<i>S</i>	Label	<i>S</i>
H10	1.52	H42	1.36
H11	1.41	H43	1.22
H12	1.43	H5	5.13
H13	1.97	H6	1.47
H2	0.64	H71	1.85
H3	1.57	H72	2.54
H41	4.15	H9	1.81
<i><H></i>	2.00		

Table 3

Comparison of TLS+INV derived ADPs with TLS+ONIOM derived ADPs of xylitol.

Label	<i>S</i>	Label	<i>S</i>
H11	3.51	H1B	0.81
H12	9.58	H2	0.33
H13	3.84	H3	0.20
H14	13.60	H4	0.52
H15	10.52	H5A	0.37
H1A	0.80	H5B	0.57
<i><H></i>	3.74		

can lead to flattening of the potential, thus reducing the force required for rotating these groups.

Overall, the differences between the two methods are of the same order of magnitude as the differences seen between the estimated models and neutron diffraction derived models discussed below. We therefore argue that the TLS+INV method is an equivalent and easier to apply substitute for the computationally more demanding TLS+ONIOM approach. Empirical corrections for hydrogen bonding could be added at a later stage.

3.2. Temperature dependence of relative U_{iso} values

Accounting for the measurement temperature when calculating the internal contributions to the ADPs avoids systematic errors that otherwise would affect data sets collected at low temperatures, especially below 100 K (Lübben *et al.*, 2014). That the temperature-dependent behavior is well reproduced in the TLS+INV approach introduced here is shown by comparing the temperature dependence of U_{iso} values obtained by the TLS+INV model with those determined from neutron diffraction studies and from

ONIOM cluster computations in the same manner as in our earlier work (Fig. 6).

These results are in very good agreement with those of Lübben *et al.* (2014) and reproduce the temperature dependence. Additionally, the TLS+INV approach is able to estimate unbiased ADPs in cases where H atoms are disordered. Since the invariom approach relies on non-interacting molecules in the gas phase, displacement parameters of H atoms involved in hydrogen bonding are less well estimated.

3.3. Comparison with results from neutron diffraction

Estimated ADPs were compared with ADPs refined against neutron diffraction data to further validate the TLS+INV method. A set of four structures where both high-resolution X-ray data and neutron data are available were taken from the literature (Cole *et al.*, 2002; Woińska *et al.*, 2014; Šerb *et al.*, 2014; Madsen *et al.*, 2003). A scaling model was fitted to each neutron data set to bring both sets of ADPs onto the same scale (Blessing, 1995). This was achieved by computing the set of parameters S_1 – S_7 in equation (3) that minimize the difference between ADPs of equivalent atoms in both models.

$$U_N^{\text{scaled}} = U_N^{\text{measured}} S_1 + \begin{pmatrix} S_2 & S_5 & S_6 \\ S_5 & S_3 & S_7 \\ S_6 & S_7 & S_4 \end{pmatrix}. \quad (3)$$

Modified TLS+INV results were then compared with those obtained with the *SHADE* server. One should note that the accuracy and absolute scale of H-ADPs remain unknown. While appropriate for atoms with similar mass, it has been shown that this scaling model does not yield accurate results when heavier atoms like iron are involved (Blessing, 1995). It is therefore reasonable to suspect that application of scaling parameters obtained by fitting against C and O atoms yields only rough estimates of hydrogen parameters. Absolute values of this comparison should therefore be interpreted with

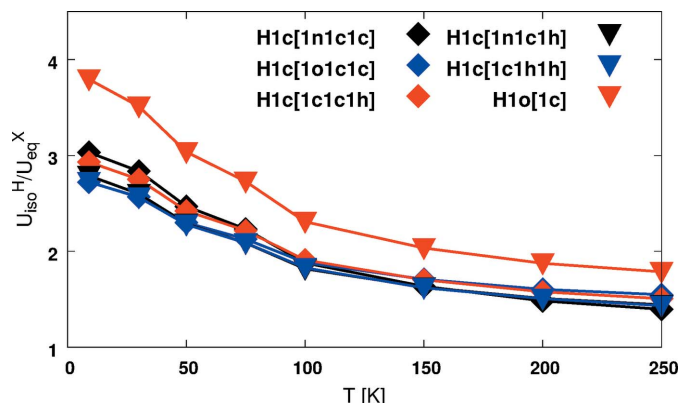


Figure 6
Temperature dependence of $U_{\text{iso}}^{\text{H}}/U_{\text{eq}}^{\text{X}}$ ratios obtained with the TLS+INV approach. The results are in very good agreement with those from our earlier study (Lübben *et al.*, 2014) reporting neutron and TLS+ONIOM results. Note: H atoms with the invariom name *H1c[1c1h1h]* are disordered and therefore appear larger when compared with the TLS+INV model. The H atom with the invariom name *H1o[1c]* is involved in hydrogen bonding, which is not accounted for in the TLS+INV model. Therefore, its $U_{\text{iso}}^{\text{H}}/U_{\text{eq}}^{\text{X}}$ ratio is systematically larger.

Table 4

Comparison of TLS+INV (S_{INV}) derived ADPs with SHADE (S_S) ADPs for the example of MBADNP.

Label	S_{INV}	S_S	Label	S_{INV}	S_S
H11	0.44	0.23	H5	0.75	0.28
H13	0.12	0.03	H6	1.17	0.27
H1N	1.35	0.39	H7	0.11	0.14
H2	0.17	0.09	H8A	1.76	1.30
H3	0.92	0.18	H8B	2.38	1.02
H4	0.17	0.14	H8C	2.21	0.90
$\langle H \rangle$	0.96	0.42			

Table 5

Comparison of TLS+INV (S_{INV}) derived ADPs with SHADE (S_S) ADPs for the example of L-phenylalaninium hydrogen maleate.

Label	S_{INV}	S_S	Label	S_{INV}	S_S
H10	3.84	0.52	H42	4.80	0.70
H11	3.19	0.61	H43	3.91	1.08
H12	2.31	0.52	H5	3.82	1.33
H13	4.10	1.49	H6	2.94	0.67
H2	2.05	1.05	H71	13.68	5.71
H3	2.27	0.67	H72	1.90	0.38
H41	4.57	0.73	H9	3.22	0.90
$\langle H \rangle$	3.30	1.17			

The large discrepancy for atom H71 is most likely due to ill-determined displacement parameters in the neutron refinement, as becomes obvious from visual inspection.

Table 6

Comparison of TLS+INV (S_{INV}) derived ADPs with SHADE (S_S) ADPs for the example of dimethylbiguanidiniumbishydrogensquarate.

Label	S_{INV}	S_S	Label	S_{INV}	S_S
H1	2.84	0.70	H4B	1.15	0.94
H10A	1.77	2.73	H5	4.04	0.37
H10B	2.15	3.97	H5A	0.73	0.06
H10C	1.60	2.42	H5B	0.70	0.13
H2	1.38	0.98	H9A	1.25	3.51
H3A	1.14	0.63	H9B	1.35	3.17
H3B	0.95	0.97	H9C	0.40	2.00
H4A	1.01	1.15			
$\langle H \rangle$	1.50	1.58			

Table 7

Comparison of TLS+INV (S_{INV}) derived ADPs with SHADE (S_S) ADPs for the example of xylitol.

Label	S_{INV}	S_S	Label	S_{INV}	S_S
H11	3.55	0.58	H1B	2.45	0.74
H12	2.85	0.49	H2	0.62	0.55
H13	3.76	0.24	H3	0.07	0.09
H14	1.92	0.91	H4	0.28	0.10
H15	2.47	0.41	H5A	3.41	1.68
H1A	2.46	0.78	H5B	2.97	1.83
$\langle H \rangle$	2.24	0.70			

caution. Concerning this problem, the SHADE server benefits from error cancellation: since the internal ADPs are obtained from neutron diffraction studies, the ADPs are already scaled appropriately for comparison with neutron diffraction results and possible systematic errors could be obscured.

The S values listed in Tables 4–7 quantify differences between the respective methods of estimation and ADPs from

Table 8

Temperature and resolution dependence of the improvements in the R value for a series of structure determinations.

Structure code	$R_{TLS+INV}$	R_{riding}	δ_R	Resolution (Å)	T (K)
hb6948 (Fadzillah <i>et al.</i> , 2012)	0.0272	0.0278	+0.007	0.73	100
zj2091 (Matos <i>et al.</i> , 2012)	0.0300	0.0307	+0.007	0.83	100
eg3095 (Tutughamarso <i>et al.</i> , 2012)	0.0298	0.0301	+0.003	0.82	173
dt3014 (de Sousa <i>et al.</i> , 2012)	0.0533	0.0536	+0.003	0.80	173
yp3017 (Sonar <i>et al.</i> , 2012)	0.0529	0.0532	+0.003	0.83	90
fg3251 (Sowa <i>et al.</i> , 2012)	0.0580	0.0582	+0.002	0.81	100
bt5991 (Khalaji <i>et al.</i> , 2012)	0.0228	0.0230	+0.002	0.88	120
sh5011 (Madsen <i>et al.</i> , 2003)	0.0182	0.0182	+0.000	0.41	122
bi3042 (Liu <i>et al.</i> , 2012)	0.0474	0.0472	-0.002	0.73	153
fg3250 (Smith & Wermuth, 2012)	0.0316	0.0314	-0.002	0.81	293
fa3263 (Pérez <i>et al.</i> , 2012)	0.0441	0.0430	-0.011	0.77	293
fg3262 (Helliwell <i>et al.</i> , 2012)	0.0308	0.0280	-0.028	0.81	296

models refined against neutron diffraction data. The estimation methods are thereby not compared directly; instead their agreement with experimental data is given.

Overall, one can clearly see that the SHADE server estimates are closer to the model refined against neutron diffraction data. However, the overall differences between all three models are of the same order of magnitude. It should also be noted that the values depend on the applied refinement model that was used prior to TLS analysis. A variation of about 0.5–0.7 in the average S values was observed when the invariom aspherical-atom model rather than the IAM model was used. Such variations do not appear to be systematic and can lead to smaller as well as larger S values. Therefore we conclude that the overall standard deviations of the S values must be of the same order of magnitude.

For the case of xylitol it is noteworthy that parameterization of the SHADE server was initiated with neutron data for this compound. Hence it is expected that the SHADE server performs especially well for this structure.

3.4. Usability

The program *APD-Toolkit* was designed specifically to be easy to use. To demonstrate this a series of crystal structures were taken from the literature. The structures were re-refined with the invariom model and a subsequent TLS+INV treatment was applied. The TLS+INV application only requires one program call with output files from a previous refinement and no further input is needed. Currently *SHELXL*-style (Sheldrick, 2008) *.res* files, *XD*-style (<http://xd.chem.buffalo.edu/>) *.res* files, CIFs and PDB files (<http://www.rcsb.org/>) are supported. In Table 8 the effect of including the estimated H-ADPs on $R_1(F)$ was investigated. Respective refinement models use the same number of parameters.

Table 8 shows that the improvements in the $R_1(F)$ values are temperature dependent. We chose $R_1(F)$ (with units weights) for historic reasons and since unweighted $R_2(F)$ is not very meaningful. When non-hydrogen ADPs are large they increasingly deviate from the segmented rigid-body approximation, possibly because of anharmonic vibrational behavior (Zhurov *et al.*, 2011). Therefore TLS analysis may not provide

accurate estimates of the lattice vibrations, and H-ADPs appear unreasonable large. On the other hand, when the measurement temperature is low and the refined ADPs are reasonable, the model including the estimated ADPs fits the measured data better, therefore also providing a useful indicator of data quality in the low-order region.

4. Conclusion and outlook

The combination of segmented rigid-body analysis with information from geometry optimized model compounds allows one to rapidly estimate anisotropic hydrogen displacements from tabulated values in the TLS+INV approach introduced here. The invariom approach is thus being extended to predict not only aspherical scattering factors but also H-ADPs, all from one consistent model and notation. This is an important advantage over other scattering-factor databases.

The program *APD-Toolkit* provides an easy to use way of estimating these displacement parameters with an accuracy comparable to the TLS+ONIOM method without the need for extensive computations upon application. The software is a standalone alternative to the *SHADE* server, is freely available for download (<http://ewald.ac.chemie.uni-goettingen.de/programs.html>; <https://github.com/jluebben/APD-toolkit>) for various operating systems and can be easily adapted for other applications. The underlying TLS+ARG implementation can be combined with other software to generate segmented rigid-body models in an automatic fashion – without requiring specialized input-file formats or restrictions in system size.

APPENDIX A Transformations and TLS fit

A1. Coordinate transformation

The invariom database stores structural parameters like atomic positions and corresponding ADPs in a standardized format. These parameters need to be transformed to the crystal's native coordinate system upon application.

Atomic positions are stored in an artificial crystal coordinate system in fractional coordinates. The artificial cell is cubic with a cell length of 30 Å.

ADPs are obtained from frequency computations carried out with *GAUSSIAN* and stored in a Cartesian coordinate system.

If V is the unit-cell volume, the matrix M_{fc} is used to transform from fractional space to Cartesian space:

$$M_{fc} = \begin{bmatrix} a & b \cos(\gamma) & c \frac{\cos(\alpha) - \cos(\beta) \cos(\gamma)}{\sin(\gamma)} \\ 0 & b \sin(\gamma) & c \frac{\sin(\gamma)}{V} \\ 0 & 0 & c \frac{1}{\sin(\gamma)} \end{bmatrix}. \quad (4)$$

M_{cf} transforms from Cartesian to fractional systems:

$$M_{cf} = M_{fc}^{-1}. \quad (5)$$

If $M_{fc,inv}$ is M_{fc} with $a = b = c = 30$ Å and $\alpha = \beta = \gamma = 90^\circ$ and $M_{cf,cryst}$ is M_{cf} with the crystal's cell parameters, the atomic position of an atom in the invariom database v_{inv} in the crystal's coordinate system v_{cryst} can be computed as

$$v_{cryst} = M_{cf,cryst}(M_{fc,inv}v_{inv}). \quad (6)$$

The matrix representation of an ADP in Cartesian space,

$$U_{ij,cart} = \begin{pmatrix} U_{11} & U_{12} & U_{13} \\ U_{12} & U_{22} & U_{23} \\ U_{13} & U_{23} & U_{33} \end{pmatrix}, \quad (7)$$

is transferred to the crystal's coordinates system with

$$U_{ij,cryst} = N^{-1}U_{ij}^*(N^{-1})^T, \quad (8)$$

where

$$U_{ij}^* = M_{cf}U_{ij,cart}M_{cf}^T \quad (9)$$

and

$$N = \begin{pmatrix} a^{-1} & 0 & 0 \\ 0 & b^{-1} & 0 \\ 0 & 0 & c^{-1} \end{pmatrix}. \quad (10)$$

a , b and c are the cell constants of the crystal.

A2. TLS+ARG fit

The TLS model describes the vibrational movement of a rigid atomic framework with 20 parameters in the form of the matrices T , L and S :

$$T = \begin{pmatrix} T_{11} & T_{12} & T_{13} \\ T_{12} & T_{22} & T_{23} \\ T_{13} & T_{23} & T_{33} \end{pmatrix}, \quad (11)$$

$$L = \begin{pmatrix} L_{11} & L_{12} & L_{13} \\ L_{12} & L_{22} & L_{23} \\ L_{13} & L_{23} & L_{33} \end{pmatrix}, \quad (12)$$

$$S = \begin{pmatrix} S_{11} & S_{12} & S_{13} \\ S_{21} & S_{22} & S_{23} \\ S_{31} & S_{32} & S_{33} \end{pmatrix}, \quad (13)$$

with

$$S_{11} + S_{22} + S_{33} = 0. \quad (14)$$

The six parameters of U_{ij} can be expressed with T , L , S (Merritt, 1999; Schomaker & Trueblood, 1998) and the fractional coordinates (x, y, z) as

$$\begin{aligned}
 U_{11}^{\text{TLS}} &= L_{22}z^2 + L_{33}y^2 - 2L_{23}yz + 2S_{21}z - 2S_{31}y + T_{11}, \\
 U_{22}^{\text{TLS}} &= L_{11}z^2 + L_{33}x^2 - 2L_{13}xz - 2S_{12}z + 2S_{32}x + T_{22}, \\
 U_{33}^{\text{TLS}} &= L_{11}z^2 + L_{33}x^2 - 2L_{12}xy - 2S_{23}x + 2S_{13}y + T_{33}, \\
 U_{12}^{\text{TLS}} &= -L_{33}xy + L_{23}xz + L_{13}yz - L_{12}z^2 \\
 &\quad + (S_{22} - S_{11})z + S_{31}x - S_{32}y + T_{12}, \\
 U_{13}^{\text{TLS}} &= -L_{22}xz + L_{23}xy - L_{13}y^2 + L_{12}yz \\
 &\quad + (S_{11} - S_{33})y + S_{23}z - S_{21}x + T_{13}, \\
 U_{23}^{\text{TLS}} &= -L_{11}yz - L_{23}x^2 + L_{31}xy + L_{12}xz \\
 &\quad + (S_{33} - S_{22})x + S_{12}y - S_{12}z + T_{23}.
 \end{aligned}
 \tag{15}$$

When U_{ij}^{meas} are the parameters obtained from structure refinement and n is the number of atoms, the set of parameters T , L and S are optimized to minimize

$$\sum_{k=1}^n (U_{ij,k}^{\text{meas}} - U_{ij,k}^{\text{TLS}})^2.
 \tag{16}$$

Each ARG rotating around the axis t adds seven parameters $A_1 \dots A_7$ to the U_{ij} description, yielding the following expressions (Schomaker & Trueblood, 1998) if the atom at position v is part of the ARG:

$$\begin{aligned}
 U_{11}^{\text{TLS+ARG}} &= U_{11}^{\text{TLS}} + V_1^2 A_1 + 2zV_1 A_3 - 2yV_1 A_4 + 2V_1 A_5, \\
 U_{22}^{\text{TLS+ARG}} &= U_{22}^{\text{TLS}} + V_2^2 A_1 - 2zV_2 A_2 + 2xV_2 A_4 + 2V_2 A_6, \\
 U_{33}^{\text{TLS+ARG}} &= U_{33}^{\text{TLS}} + V_3^2 A_1 + 2yV_3 A_2 - 2xV_3 A_3 + 2V_3 A_7, \\
 U_{12}^{\text{TLS+ARG}} &= U_{12}^{\text{TLS}} + V_1 V_2 A_1 - zV_1 A_2 + zV_2 A_3 \\
 &\quad + (xV_1 - yV_2)A_4 + V_2 A_5 + V_1 A_6, \\
 U_{13}^{\text{TLS+ARG}} &= U_{13}^{\text{TLS}} + V_1 V_3 A_1 + yV_1 A_2 + (zV_3 - xV_1)A_3 \\
 &\quad - yV_3 A_4 + V_3 A_5 + V_1 A_7, \\
 U_{23}^{\text{TLS+ARG}} &= U_{23}^{\text{TLS}} + V_2 V_3 A_1 + (yV_2 - zV_3)A_2 - xV_2 A_3 \\
 &\quad + xV_3 A_4 + V_3 A_6 + V_2 A_7,
 \end{aligned}
 \tag{17}$$

with

$$V = t \times R = (V_1, V_2, V_3)
 \tag{18}$$

and

$$R = v - P,
 \tag{19}$$

where P is the shortest distance between t and the origin.

For l ARGs, $20 + 7l$ parameters are determined analogously to equation (16) by minimizing the expression

$$\sum_{k=1}^n (U_{ij,k}^{\text{meas}} - U_{ij,k}^{\text{TLS+ARG}})^2.
 \tag{20}$$

Acknowledgements

BD thanks H. Puschmann (OlexSys) and J. A. K. Howard for encouragement in the early stages of the project. We thank the Deutsche Forschungsgemeinschaft DFG for funding within project DI 921/6-1.

References

Allen, F. H. & Bruno, I. J. (2010). *Acta Cryst.* **B66**, 380–386.

- Blessing, R. H. (1995). *Acta Cryst.* **B51**, 816–823.
- Blom, R. & Haaland, A. (1985). *J. Mol. Struct.* **128**, 21–27.
- Brock, C. P., Dunitz, J. D. & Hirshfeld, F. L. (1991). *Acta Cryst.* **B47**, 789–797.
- Capelli, S. C., Bürgi, H.-B., Dittrich, B., Grabowsky, S. & Jayatilaka, D. (2014). *IUCrJ*, **1**, 361–379.
- Cole, J. M., Goeta, A. E., Howard, J. A. K. & McIntyre, G. J. (2002). *Acta Cryst.* **B58**, 690–700.
- Cooper, R. I., Thompson, A. L. & Watkin, D. J. (2010). *J. Appl. Cryst.* **43**, 1100–1107.
- 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., Hübschle, C. B., Pröpper, K., Dietrich, F., Stolper, T. & Holstein, J. J. (2013). *Acta Cryst.* **B69**, 91–104.
- Dittrich, B., Pfitzenreuter, S. & Hübschle, C. B. (2012). *Acta Cryst.* **A68**, 110–116.
- Dittrich, B., Wandtke, C. M., Meents, A., Pröpper, K., Mondal, K. C., Samuel, P. P., Amin Sk, N., Singh, A. P., Roesky, H. W. & Sidhu, N. (2015). *Chem. Phys. Chem.* **16**, 412–419.
- Domagała, S., Fournier, B., Liebschner, D., Guillot, B. & Jelsch, C. (2012). *Acta Cryst.* **A68**, 337–351.
- Dominiak, P. M., Volkov, A., Li, X., Messerschmidt, M. & Coppens, P. (2007). *J. Chem. Theory Comput.* **3**, 232–247.
- Fadzillah, S. M. H., Ngaini, Z., Hussain, H., Razak, I. A. & Asik, S. I. J. (2012). *Acta Cryst.* **E68**, o2909.
- Flaig, R., Koritsánszky, T., Zobel, D. & Luger, P. (1998). *J. Am. Chem. Soc.* **120**, 2227–2238.
- Frisch, M. J. *et al.* (2013). *Gaussian 09*, Revision D.01. Technical Report, Gaussian Inc., Pittsburgh, PA, USA.
- Hansen, N. K. & Coppens, P. (1978). *Acta Cryst.* **A34**, 909–921.
- 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. & Hope, H. (1980). *Acta Cryst.* **B36**, 406–415.
- Hoser, A. A., Dominiak, P. M. & Woźniak, K. (2009). *Acta Cryst.* **A65**, 300–311.
- Jarzembska, K. N. & Dominiak, P. M. (2012). *Acta Cryst.* **A68**, 139–147.
- Jayatilaka, D. & Dittrich, B. (2008). *Acta Cryst.* **A64**, 383–393.
- Khalaji, A. D., Fejfarová, K. & Dušek, M. (2012). *Acta Cryst.* **E68**, o2646.
- Liu, Y.-L., Zou, P., Wu, H., Xie, M.-H. & Luo, S.-N. (2012). *Acta Cryst.* **C68**, o338–o340.
- Lübben, J., Volkmann, C., Grabowsky, S., Edwards, A., Morgenroth, W., Fabbiani, F. P. A., Sheldrick, G. M. & Dittrich, B. (2014). *Acta Cryst.* **A70**, 309–316.
- Madsen, A. Ø. (2006). *J. Appl. Cryst.* **39**, 757–758.
- Madsen, A. Ø., Civalleri, B., Ferrabone, M., Pascale, F. & Erba, A. (2013). *Acta Cryst.* **A69**, 309–321.
- Madsen, A. Ø. & Hoser, A. A. (2014). *J. Appl. Cryst.* **47**, 2100–2104.
- Madsen, A. Ø., Mason, S. & Larsen, S. (2003). *Acta Cryst.* **B59**, 653–663.
- Madsen, A. Ø., Sørensen, H. O., Flensburg, C., Stewart, R. F. & Larsen, S. (2004). *Acta Cryst.* **A60**, 550–561.
- Matos, M. J., Santana, L. & Uriarte, E. (2012). *Acta Cryst.* **E68**, o2645.
- Merritt, E. A. (1999). *Acta Cryst.* **D55**, 1109–1117.
- Munshi, P., Madsen, A. Ø., Spackman, M. A., Larsen, S. & Destro, R. (2008). *Acta Cryst.* **A64**, 465–475.
- Orben, C. M. & Dittrich, B. (2014). *Acta Cryst.* **C70**, 580–583.
- Painter, J. & Merritt, E. A. (2006). *Acta Cryst.* **D62**, 439–450.
- Pérez, H., Corrêa, R. S., Plutín, A. M., O'Reilly, B. & Andrade, M. B. (2012). *Acta Cryst.* **C68**, o19–o22.
- Roversi, P. & Destro, R. (2004). *Chem. Phys. Lett.* **386**, 472–478.
- Schomaker, V. & Trueblood, K. N. (1968). *Acta Cryst.* **B24**, 63–76.

- Schomaker, V. & Trueblood, K. N. (1998). *Acta Cryst.* **B54**, 507–514.
- Schürmann, C. J., Pröpper, K., Wagner, T. & Dittrich, B. (2012). *Acta Cryst.* **B68**, 313–317.
- Scott, A. P. & Radom, L. (1996). *J. Phys. Chem.* **100**, 16502–16513.
- Şerb, M.-D., Kalf, I. & Englert, U. (2014). *CrystEngComm*, **16**, 10631–10639.
- 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.
- Svensson, M., Humbel, S., Froese, R. D. J., Matsubara, T., Sieber, S. & Morokuma, K. (1996). *J. Phys. Chem.* **100**, 19357–19363.
- Tutughamiarso, M., Pisternick, T. & Egert, E. (2012). *Acta Cryst.* **C68**, o344–o350.
- Whitten, A. E. & Spackman, M. A. (2006). *Acta Cryst.* **B62**, 875–888.
- Woińska, M., Jayatilaka, D., Spackman, M. A., Edwards, A. J., Dominiak, P. M., Woźniak, K., Nishibori, E., Sugimoto, K. & Grabowsky, S. (2014). *Acta Cryst.* **A70**, 483–498.
- Zarychta, B., Pichon-Pesme, V., Guillot, B., Lecomte, C. & Jelsch, C. (2007). *Acta Cryst.* **A63**, 108–125.
- Zhurov, V. V., Zhurova, E. A., Stash, A. I. & Pinkerton, A. A. (2011). *Acta Cryst.* **A67**, 160–173.