

On the handling of atomic anisotropic displacement parameters

R.W. Grosse-Kunstleve & P.D. Adams

*Lawrence Berkeley National Laboratory, One Cyclotron Road, Mail Stop 4-230,
Berkeley, California 94720, U.S.A.*

E-mail: RWGrosse-Kunstleve@lbl.gov

Synopsis

Theoretical and practical aspects of handling atomic anisotropic displacement parameters are discussed.

Abstract

A number of conventions for the representation of atomic anisotropic displacements are used in the literature and in crystallographic programs. Here we review the commonly used conventions, with a special emphasis on their application in macromolecular crystallography. We then describe a new software toolbox for the handling of the various representations of atomic anisotropic displacement parameters and their interconversion. All algorithms are integrated into the freely available Computational Crystallography Toolbox.

1. Introduction

When high-resolution diffraction data are available, crystal structures are often modeled using atomic anisotropic displacement parameters (ADPs) to describe thermal motion and possible static displacive disorder. In the field of small-molecule crystallography ADPs have been routinely used for decades (Trueblood et al., 1996). Until recently, the use of ADPs in the refinement of macromolecular structures has been rare. However, due to many improvements in data collection technology, high-resolution data that enable the refinement of ADPs are becoming available for macromolecules at an increased rate. This is reflected by a rapid growth in the number of structures with ADPs deposited in the Protein Data Bank (PDB) (Berman et al., 2000) (Fig. 1).

As we show in section 2, a number of conventions for the representation of ADPs are currently in use. A comprehensive overview is given by Trueblood et al. (1996). These different conventions pose a significant potential for confusion. For example, the ADPs found in files that follow the *PDB format* (<http://www.rcsb.org/pdb/info.html>) follow a different convention than the ADPs found in files that follow the *mmCIF format* (<http://pdb.rutgers.edu/mmcif/>). If this is not taken into account, there exists a possibility that the ADPs can be misinterpreted, leading to incorrect analysis of a structure.

Therefore a library for the conversion between the different representations is a valuable tool. We have implemented such a library by adding the ADP toolbox (*adptbx*) to the Computational Crystallography toolbox (*cctbx*) (Grosse-Kunstleve et al., 2002). In addition to the conversions the library facilitates the computation of Debye-Waller factors, the handling of symmetry restrictions, and the determination of the eigenvalues and eigenvectors of anisotropic displacement ellipsoids. The *cctbx* is organized as a freely available *Open Source* library of reusable, object-oriented software components for crystallographic computations.

In general, the nomenclature that was adopted for the implementation of the *adptbx* follows the recommendations of the *IUCr Subcommittee on Atomic Displacement Parameter Nomenclature* (Trueblood et al., 1996). In the following sections we will review the commonly used conventions for ADPs in the literature, computer programs and databases. This is followed by a description of the *adptbx*. Further documentation is available online (<http://cctbx.sourceforge.net/>).

2. Commonly used conventions for the representation of atomic anisotropic displacement parameters

The mean-square displacements that define the probability density functions of atomic displacements are commonly parameterized as a trivariate Gaussian. The effect of the atomic displacements enters into the structure factor calculation as the *Debye-Waller factor* $T(h)$, where h is the Miller index (given as a column vector) of a Bragg reflection. The fundamental expression for $T(h)$ is (e.g. Trueblood et al., 1996)

$$T(h) = \exp(-2\pi^2 \langle (h \cdot u)^2 \rangle), \quad (\text{Eq. 1})$$

where u is a Gaussian mean-square displacement matrix. Eq. 1 takes on different forms depending on the basis vectors to which the diffraction and displacement vectors are referred. With a few exceptions (e.g. Bricogne 2001, section 1.3.4), diffraction vectors are almost universally referred to the basis of the crystallographic reciprocal lattice. This is also the convention assumed in the rest of this paper. If the Gaussian mean-square displacement matrix is referred to the same basis, Eq. (1) takes on the form

$$T(h) = \exp(-2\pi^2 h^t U^* h), \quad (\text{Eq. 2})$$

where h^t is the transposed of h (i.e. a row vector), and U^* is a symmetric, contravariant second rank tensor (Trueblood et al., 1996), typically represented as a symmetric 3x3 matrix of real numbers. This definition of U^* is equivalent to that of Giacovazzo (1992). (Bricogne (2001) uses the notation \mathbf{Q} for the same tensor.) Obviously, the representation of ADPs with respect to the reciprocal lattice can be used directly in the structure factor calculation. It is also the most suitable representation for establishing compatibility with the site-symmetry for special positions (see section 3.3).

All other commonly used conventions for the representations of ADPs are most conveniently defined through tensor transformations. Section 2.E of Giacovazzo (1992) gives a comprehensive collection of *Transformations of crystallographic quantities in rectilinear spaces*, including the transformation of “quadratic forms” (i.e. second rank tensors). For example, let A be an *orthogonalization matrix* that transforms the fractional coordinates x_{frac} with respect to a crystallographic basis system to coordinates x_{cart} with respect to a Cartesian basis system. Application of the transformation rules listed by Giacovazzo (1992) leads to

$$U_{cart} = A U^* A^t, \quad (\text{Eq. 3a})$$

$$U^* = A^{-1}U_{cart}(A^{-1})^t, \quad (\text{Eq. 3b})$$

where A^t is the transposed of A , A^{-1} is the inverse of A , and U_{cart} is the anisotropic displacement tensor with respect to the Cartesian basis. (It should be noted that Eq. 3 is using *matrix notation*. In contrast, Trueblood et al. (1996) use *tensor notation*. Compare Eq. (3a) in this paper with the similar Eq. (2.1.44) of Trueblood et al. (1996); see also Nye (1992), ch. IX.)

The ADPs found in files that follow the PDB format (*ANISOU* cards) define the elements of U_{cart} . The representation of ADPs with respect to a Cartesian basis system is useful because the eigenvalues of U_{cart} are directly equivalent to the mean-square atomic displacements in the principal directions of the corresponding ellipsoid (see section 3.4). In addition, the Cartesian basis system is typically the most useful for generating graphical representations of such ellipsoids (as performed, e.g., by the ORTEP program (Burnett & Johnson, 1996)).

In small-molecule crystallography, a third basis system is used almost universally for the representation of ADPs: the basis is formed by unit vectors that are parallel to the basis vectors of the crystallographic reciprocal lattice. To our knowledge, this *ad hoc* basis system is only used for the representation of ADPs. It is connected to the reciprocal lattice by the transformation matrix

$$N = \begin{pmatrix} a^* & 0 & 0 \\ 0 & b^* & 0 \\ 0 & 0 & c^* \end{pmatrix}, \quad (\text{Eq. 4})$$

where the diagonal elements of the matrix are the lengths of the basis vectors of the reciprocal lattice. The corresponding transformation laws for anisotropic displacement tensors are given by

$$U_{uvrs} = N^{-1}U^*(N^{-1})^t, \quad (\text{Eq. 5a})$$

$$U^* = NU_{uvrs}N^t, \quad (\text{Eq. 5b})$$

where U_{uvrs} is a tensor with respect to the “unit vectors parallel to the basis vectors of reciprocal space.” This convention for ADPs was adopted in the definitions of both the CIF (<http://www.iucr.org/iucr-top/cif/>) and the mmCIF dictionaries, where they are referred to as *aniso_U*. It also is the convention used by the SHELX programs (Sheldrick & Schneider, 1997), the XTAL suite (Hall et al., 2000), MOPRO (Guillot et al., 2001), and probably many other programs and databases. U_{uvrs} has the advantage that the elements of the tensor are quantitatively similar to the mean-square displacements. For unit cells with orthogonal basis vectors U_{uvrs} is exactly equivalent to U_{cart} . At the same time the computation of the Debye-Waller factor is relatively simple because N is a diagonal matrix. Substituting Eq. (5b) into Eq. (2) results in the familiar expression

$$T(hkl) = \exp(-2\pi^2 (h^2 (a^*)^2)U_{11} + k^2 (b^*)^2)U_{22} + l^2 (c^*)^2 U_{33} + 2hka^*b^*U_{12} + 2hla^*c^*U_{13} + 2klb^*c^*U_{23}), \quad (\text{Eq. 6})$$

where U_{ij} is the element in the i th row and the j th column of the tensor U_{uvrs} . This equation is found in many program descriptions and in the definition of the CIF dictionaries. Unfortunately, for unit cells with angles other than 90° the mean-square displacements are convoluted with the metric of the crystallographic basis. To determine the principal mean-square displacements for the general case it is therefore necessary to convert U_{uvrs} to U_{cart} . The conversion laws are easily obtained by combining Eqs. (3) and (5).

For completeness we mention that Giacovazzo (1992) defines “ U ” with respect to the crystallographic basis. This leads to the relation $U = G U^* G$, where G is the metrical matrix of the crystallographic basis. However, we are not aware of any program or database that uses this convention for “ U ”. Therefore support for ADPs with respect to crystallographic bases is not included in the adptbx.

The representations of ADPs reviewed so far are all “pure” in the sense that they directly define a physical quantity, the mean-square displacement of an atom. The differences are exclusively in the choice of the basis system to which the ADPs refer. Unfortunately, these different representations are also sometimes convoluted with factors of π . The first example is the definition

$$\beta = 2\pi^2 U^* . \quad (\text{Eq. 7})$$

This definition is found throughout the literature. It is obviously motivated by the desire to optimize the computation of the Debye-Waller factor according to Eq. (2). β are also used by the SIR program (Altomare et al., 2000). The second example is the definition

$$B = 8\pi^2 U , \quad (\text{Eq. 8})$$

which is used in the definition of the CIF and mmCIF dictionaries. The factor $8\pi^2$ originates from the expression for the Debye-Waller factor for isotropic displacement parameters which is usually (e.g. Trueblood et al., 1996) defined as

$$T(h) = \exp(-8\pi^2 U (\sin \theta / \lambda)^2) , \quad (\text{Eq. 9})$$

where $(\sin \theta / \lambda)^2$ is a commonly used measure for the spacing between lattice planes perpendicular to the diffraction vector h . Presumably the definition of Eq. (8) is transferred from isotropic to anisotropic displacement parameters to make them “comparable.” However, this is highly artificial. Therefore the use of B is explicitly discouraged by Trueblood et al. (1996) and in the definitions of the CIF and mmCIF dictionaries. However, to be able to process input from sources that use B , support for this is included in the adptbx.

3. Overview of the adptbx and other related parts of the cctbx

3.1. Conversions

The adptbx supports the simple conversions between both isotropic and anisotropic “ U ” and “ B ” (Eq. 8), independently of the basis system that is used. Further supported are all possible conversions between U^* , U_{cart} , U_{uvrs} , and β . This results in a total of 12 conversion functions. However, only the conversions according to Eqs. (3) and (5) are generic implementations. The other conversions are implemented as nested function calls of the generic conversions. This minimizes the amount of source code, and the relations between the different representations of the ADPs are also easy to follow. It should also be noted that modern optimizing compilers generate code that is as efficient at runtime as manually coded specific conversion functions.

Eight additional functions are provided for the conversions between the anisotropic U_{cart} , U^* , U_{uvrs} , and β and the equivalent isotropic U_{iso} . The isotropic equivalent of the anisotropic displacement tensor is defined as the mean of the mean-square displacements in the principal directions of the anisotropic displacement ellipsoid, which is equivalent to the mean of the eigenvalues. In the Cartesian basis system the trace of the U_{cart} tensor is equivalent to the sum of the eigenvalues (see, e.g. p. 188 in Giacovazzo, 1992). Therefore the equivalent U_{iso} is defined by the simple relation

$$U_{iso} = \frac{1}{3} \text{trace}(U_{cart}). \quad (\text{Eq. 10})$$

The conversion from U_{iso} to U_{cart} is equally straightforward:

$$U_{cart} = \begin{pmatrix} U_{iso} & 0 & 0 \\ 0 & U_{iso} & 0 \\ 0 & 0 & U_{iso} \end{pmatrix} \quad (\text{Eq. 11})$$

The conversions between U^* , U_{uvrs} , and β and the equivalent U_{iso} are implemented as the combination of conversions between, e.g., U^* and U_{cart} , and U_{cart} and U_{iso} . In this way the complicated expressions for equivalent U_{iso} that are often found in the literature are avoided.

3.2. Computation of the Debye-Waller factor

Functions are provided for the computation of the Debye-Waller factor according to Eq. (1) given U^* , β , U_{cart} , U_{uvrs} , U_{iso} and B_{iso} . Only two functions are generic implementations (U^* , B_{iso}). The others are implemented as a conversion followed by a call of the suitable generic function.

3.3. Handling of symmetry restrictions for special positions

For atoms on special positions, the *SiteSymmetry* class (Grosse-Kunstleve & Adams, 2002) in the *Space Group Toolbox* of the cctbx was extended to test whether a given anisotropic tensor U^* is compatible with the site symmetry. The anisotropic displacement ellipsoid must remain invariant under the application of each of the symmetry operations of the site-symmetry group. This leads to the condition

$$U^* = RU^*R^t, \quad (\text{Eq. 11})$$

where R is the rotation part of a given symmetry operation (see, e.g., Giacovazzo, 1992). A second function that was added to the *SiteSymmetry* class applies Wigner's theorem (Giacovazzo, 1992, p. 189) to derive a tensor

$$U_{inv}^* = \frac{1}{n} \sum_{s=1}^n R_s U^* R_s^t, \quad (\text{Eq. 11})$$

that is invariant under the n operations of the site-symmetry group even if U^* is not. This is useful to compensate for rounding errors due to limited input precision, and possibly to compensate for approximations and rounding errors in refinement algorithms.

3.4. Determination of eigenvalues

The eigenvalues of U_{cart} are directly equivalent to the mean-square atomic displacements in the principal directions of the anisotropic displacement ellipsoid. Since the anisotropic displacement tensor is a symmetric matrix, all eigenvalues are real. The eigenvalues λ are determined as the three real roots of the cubic equation

$$|U_{cart} - \lambda I| = 0, \quad (\text{Eq. 12})$$

where I is the identity matrix. The solutions are obtained analytically using Cardan's formula (see, e.g., Spiegel & Liu, 1998).

If any of the three eigenvalues is less than or equal to zero, the U_{cart} tensor is not *positive definite* and U_{cart} cannot be represented as an ellipsoid. Tests for this condition are included in the adptbx.

3.5. Determination of eigenvectors

The eigenvectors of U_{cart} define the principal directions of the anisotropic displacement ellipsoid. Since the anisotropic displacement tensor is a symmetric matrix, the eigenvectors can be chosen orthonormal. The eigenvectors are determined using a simple method of successive approximations specific for 3x3 matrices as outlined in Nye (1992), pp.165-168.

3.6. Software technology used

The adptbx is implemented in ISO C++ (International Standardization Organization et al., 1998). This programming language supports object-oriented design, namespaces, exception handling, and parameterization of types (templates). These features facilitate the design of libraries that are reusable and maintainable. For example, all functions in the cctbx are parameterized (templated) by the floating-point type. It is therefore possible to choose between single precision and double precision representations of ADPs. It is also possible to use multiple floating point precisions simultaneously in the same

program. Importantly, this is achieved without duplicating source code, and without impacting the runtime performance.

The `adptbx` may be used as a pure C++ library. However, for maximum flexibility, bindings to the high-level *Python* scripting language are provided. The benefits of this tight combination of a statically typed compiled language (C++) with a dynamically typed interpreted language (Python) are discussed in depth by Grosse-Kunstleve et al. (2002). A simple example script (*analyze_adp.py*) that demonstrates the use of the `adptbx` is provided online in the *examples* directory of the `cctbx` (<http://cctbx.sourceforge.net/>).

4. Conclusion

We hope that the discussion of commonly used conventions for the representation of ADPs in section 2 is a useful reference for crystallographers, in particular macromolecular crystallographers, who are unavoidably presented with at least two different representations: the PDB convention (U_{cart}) and the mmCIF convention (U_{wrs}).

We have introduced practical tools for the handling of ADPs, which are accessible through a scripting language (Python). This language was designed specifically for teaching programming concepts and the tools are therefore very approachable. Their integration into the `cctbx` creates a powerful resource for handling the diverse data representations used by crystallographic software. At the same time, their modular, highly reusable design, in combination with an Open Source license, makes them ideal for inclusion into large, integrated software systems.

Acknowledgements

We would like to thank S.R. Hall, G.N. Murshudov and G.M. Sheldrick for responding to our questions. The American Mineralogist Crystal Structure Database (<http://www.geo.arizona.edu/xtal-cgi/test>) that is maintained by R.T. Downs was very helpful for the testing of the `adptbx`. Our work was funded in part by the U.S. Department of Energy under Contract No. DE-AC03-76SF00098 and NIH/NIGMS under grant number 1P01GM063210-01.

References

- Altomare, A., Giacobuzzo, C., Guagliardi, A., Moliterni, A.G.G., Rizzi, R. (2000). *Acta Cryst.* A56, 451-457.
- Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., Bourne, P.E. (2000). *Nucleic Acids Research*, 28, 235-242.
- Bricogne, G. (2001). *International Tables for Crystallography*, Vol. B, ch. 1.3. Dordrecht: Kluwer.
- Burnett, M.N. & Johnson, C.K. (1996). ORTEP-III: Oak Ridge Thermal Ellipsoid Plot Program for Crystal Structure Illustrations, Oak Ridge National Laboratory Report ORNL-6895 (<http://www.ornl.gov/ortep/>).
- C. Giacobuzzo (Ed.) (1992). *Fundamentals of Crystallography*, International Union of Crystallography, Oxford University Press.
- Grosse-Kunstleve, R.W., Sauter, N.K., Moriarty, N.W. & Adams, P.D. (2002). *Journal of Applied Crystallography*, accepted for publication.
- Grosse-Kunstleve, R.W. & Adams, P.D. (2002). *Acta Cryst. A*, accepted for publication.
- Guillot, B., Viry, L., Guillot, R., Lecomte, C. & Jelsch, C. (2001). *J. Appl. Cryst.* 34, 214-223.
- Hall, S.R., du Boulay, D.J. & Olthof-Hazekamp, R. (2000) Eds. Xtal3.7 System. University of Western Australia (<http://xtal.crystal.uwa.edu.au/>).
- International Standardization Organization (ISO), International Electrotechnical Commission (IEC), American National Standards Institute (ANSI), and Information Technology Industry Council (ITI) (1998). *International Standard ISO/IEC 14882*, 1st ed., Information Technology Industry Council, 1250 Eye Street NW, Washington, DC 20005 (also available at <http://webstore.ansi.org/>).
- Murshudov, G.N., Vagin, A.A. & Dodson, E.J. (1997). *Acta Cryst.* D53, 240-255.
- Nye, J.F. (1992). *Physical Properties of Crystals*. Oxford: Clarendon Press.
- Trueblood, K. N., Bürgi, H.-B., Burzlaff, H., Dunitz, J. D., Gramaccioni, C. M., Schulz, H. H., Shmueli, U. & Abrahams, S. C. (1996). *Acta Cryst.* A52, 770-781 (also available at <http://www.iucr.org/iucr-top/comm/cnom/adp/finrep/finrep.html>).
- Sheldrick, G.M. & Schneider, T.R. (1997). *Methods Enzymol.* 277, 319-343.
- Spiegel, M.R. & Liu, J. (1998). *Mathematical handbook of formulas and tables*. New York: McGraw-Hill.

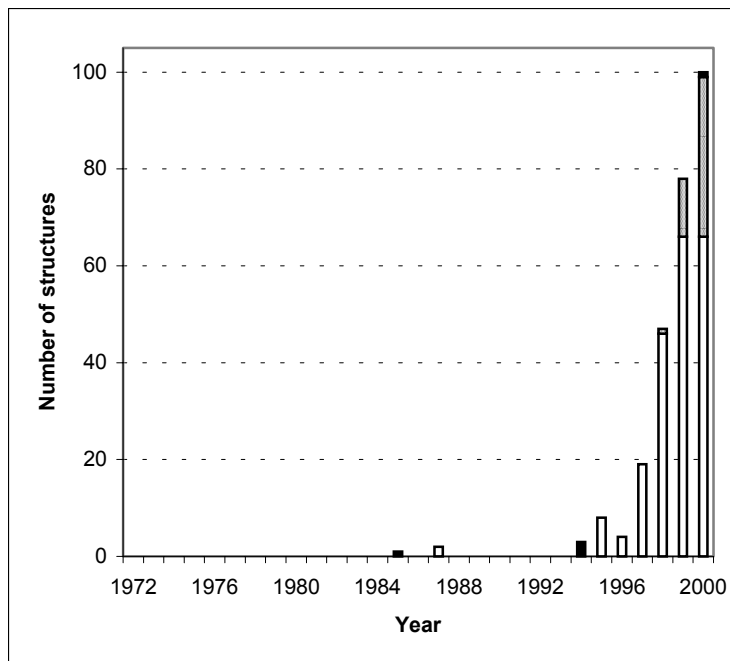


Figure 1: Histogram of structures that were deposited in the Protein Data Bank (PDB) with ANISOU cards, grouped by year, based on the PDB holdings on 16-Oct-2001. The dates used for the histogram are the days of deposition as found on the PDB HEADER card. The total number of structures with ANISOU cards is 306. 237 structures were refined with SHELX (white bars) (Sheldrick & Schneider, 1997), 64 structures with REFMAC (gray bars) (Murshudov et al., 1997), and five structures with three other programs (black bars) (see the references archived in the PDB under the access codes 7bna, 1etl, 1etm, 1etn, and 1ejg).