JOURNAL OF
APPLIED
CRYSTALLOGRAPHY

ISSN: 1600-5767
journals.iucr.org/j

From atoms to bonds, angles and torsions: molecular metrics from crystal space, and two *Excel* implementations

Leslie Glasser

J. Appl. Cryst. (2020). **53**, 1101–1107

JOURNAL OF APPLIED CRYSTALLOGRAPHY

ISSN 1600-5767

From atoms to bonds, angles and torsions: molecular metrics from crystal space, and two *Excel* implementations

Leslie Glasser*

Curtin Institute for Computation, Discipline of Chemistry, Curtin University, GPO Box U1987, Perth, WA 6845, Australia. *Correspondence e-mail: l.glasser@curtin.edu.au

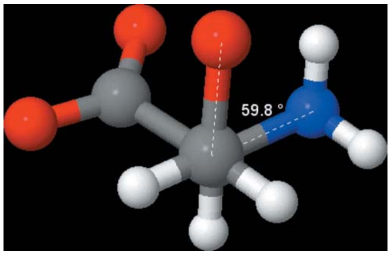
Received 11 March 2020
Accepted 1 June 2020

Edited by J. M. García-Ruiz, Instituto Andaluz de Ciencias de la Tierra, Granada, Spain

Keywords: metric matrices; bond angles; bond lengths; torsion angles; vectors.

Supporting information: this article has supporting information at journals.iucr.org/j

Values of molecular bond lengths, bond angles and (less frequently) bond torsion angles are readily available from databases, from crystallographic software, and/or from interactive molecular and crystal visualization programs such as *Jmol*. However, the methods used to calculate these values are less well known. In this paper, the computational methods are described in detail, and live *Excel* implementations, which permit readers to readily perform the calculations for their own molecular systems, are provided. The methods described apply to both fractional coordinates in crystal space and Cartesian coordinates in Euclidean space (space in which the geometric postulates of Euclid are valid) and are vector/matrix based. In their simplest computational form, they are applied as algebraic expansions which are summed. They are also available in matrix formulations, which are readily manipulated and calculated using the matrix functions of *Excel*. In particular, their general formulation as metric matrices is introduced. The methods in use are illustrated by a detailed example of the calculations. This contribution provides a significant practical application which can also act as motivation for the study of matrix mathematics with respect to its many uses in chemistry.



1. Introduction

Students and teachers of chemistry are familiar with the lengths of and angles between chemical bonds, and even with torsion angles defining molecular conformations. They are also familiar with the source of these data, which generally arise from diffraction experiments, principally with X-ray sources but also with electron or neutron sources. However, the methods of calculation which yield these values are less familiar, often hidden in crystallographic computational programs, and not generally accessible to the non-expert. The difficulties of these calculations are compounded by the fact that the experimental results are generally presented in crystal spaces, which best represent the crystalline symmetry but which are not readily manipulated by those only familiar with everyday orthogonal Euclidean space described using Cartesian coordinates. Even for students who have the opportunity to perform a full X-ray structure determination (Kantardjieff, 2010; Chapuis, 2011; Aldeborgh *et al.*, 2014; Gražulis *et al.*, 2015), it appears that the structural details are obtained from the computational software, without students having the opportunity of acquainting themselves with the calculation methods.

The fundamental data of crystallography, that is, chemical composition, unit-cell dimensions, space-group symmetry and fractional atom coordinates, are reported as the results of the diffraction experiments. These data are made widely available by publication both in the primary literature and, since the

https://doi.org/10.1107/S1600576720007311

1990s, via deposition as crystallographic information files (Hall *et al.*, 1991; http://www.iucr.org/resources/cif) (`*.cif`), which the major crystallographic journals currently require; the CIFs are subject to extensive checks (http://journals.iucr.org/services/cif/checkcif.html) in order to ensure their integrity. These files are freely available (Glasser, 2016) and large collections have been made into databases, which may be free or commercial, such as the Cambridge Crystallographic Database (CSD, for organic materials – now exceeding one million entries; Groom *et al.*, 2016; http://www.ccdc.cam.ac.uk/products/csd/), the Inorganic Crystal Structure Database (Zagorac *et al.*, 2019), the free Crystallography Open Database (Gražulis *et al.*, 2009), the free American Mineralogist Crystal Structure Database (Downs & Hall-Wallace, 2003), the free RCSB Protein Data Bank (PDB; Berman *et al.*, 2000) and so forth (Glasser, 2016). However, the CIF generally provides only the basic information, and the molecular metrics of bond lengths, bond angles and torsion angles around bonds may still need to be generated.

There are many molecular visualization programs which permit the user to import CIFs, display atoms, and select atom pairs to generate bond lengths, atom triplets to generate bond angles and even sequential atom quadruplets to generate bond torsions. Principal among these are the CSD's online *WebCSD* (Thomas *et al.*, 2010) with its free downloadable *Mercury* program (Macrae *et al.*, 2020; https://www.ccdc.cam.ac.uk/Community/csd-community/freemercury/), and the free downloadable *Jmol* (http://www.jmol.org/), *Avogadro* (https://avogadro.cc/), *VESTA* (Momma & Izumi, 2011) and *CrystalOgraph* (https://www.epfl.ch/schools/sb/research/iphys/teaching/crystallography/crystalograph/) programs (but the last provides only bond lengths).

It is the purpose of the current paper to introduce the reader to the methods of calculation of molecular metrics from crystal space (Dunitz, 1995; Julian, 2014). In order to allow these rather complex calculations to be readily performed by the reader, two separate live *Excel* implementations of the methods are supplied, in which users can easily insert their own data. One is a 'black-box' implementation of the BASIC code in Appendix I of Dunitz's book (Dunitz, 1995, pp. 495–497), while the second lays out the matrix calculations in detail, demonstrating the mathematical processes involved. These programs parallel an earlier implementation by the author using the proprietary *MathCad* software (Glasser, 1993).

We do not delve into the complexities of reciprocal spaces applicable in crystallographic calculations, since this extends into specialist applications of crystallography. However, it is of interest to note that the direct and reciprocal lattices are mutual Fourier transforms, with the momentum difference between incoming and diffracted X-rays of a crystal being a reciprocal lattice vector (https://www.doitpoms.ac.uk/tlplib/reciprocal_lattice/index.php).

2. Calculations using Cartesian coordinates

The atomic data in CIFs are listed in fractional coordinates, x, y, z, but it may be simpler to calculate the molecular metrics using Cartesian coordinates, X, Y, Z. (It is, however, possible to obtain the same results directly from the fractional coordinates, and this is considered in a subsequent section.)

2.1. Cartesian coordinates, *X*, *Y*, *Z*

The following matrix equation (McRee, 1993*a*,*b*; http://www.ruppweb.org/Xray/tutorial/Coordinate%20system%20transformation.htm) (easily applied using the *Excel* array function `MMULT`) transforms fractional coordinates, x, y, z, in crystal space into Cartesian coordinates, X, Y, Z, using the crystal cell constants a, b, c, α, β, γ:

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \begin{bmatrix} a & b\cos\gamma & c\cos\beta \\ 0 & b\sin\gamma & \dfrac{c(\cos\alpha - \cos\beta\cos\gamma)}{\sin\gamma} \\ 0 & 0 & \dfrac{V}{ab\sin\gamma} \end{bmatrix} \cdot \begin{pmatrix} x \\ y \\ z \end{pmatrix}, \tag{1a}$$

where the volume of the unit cell is

$$V = abc(1 - \cos^2\alpha - \cos^2\beta - \cos^2\gamma + 2\cos\alpha\cos\beta\cos\gamma)^{1/2}. \tag{1b}$$

The *OpenBabel Chemical Formatter* (O'Boyle *et al.*, 2011; http://www.cheminfo.org/Chemistry/Cheminformatics/FormatConverter/index.html) provides a convenient facility to convert CIFs directly to the Cartesian *XYZ* format.

The volume, V, may equivalently be obtained in matrix terms (using the *Excel* function `MDETERM`) as the square root of the determinant of the metric matrix, $\mathbf{G}$, which is introduced in Section 3.

2.2. Vectors

We use bold face, *e.g.* $\mathbf{v}$, to represent a vector, while $|\mathbf{v}|$ or italic v represents the magnitude (length – a scalar) of the vector. The italic form is used in algebraic expressions.

Vectors are described in terms of their coordinates along basis axes. For a general vector $\mathbf{p}_i$ pointing from the coordinate origin to a point i (*e.g.* an atom centre) along the crystal axes $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$,

$$\mathbf{p}_i = \begin{pmatrix} x_i\mathbf{a} \\ y_i\mathbf{b} \\ z_i\mathbf{c} \end{pmatrix} \rightarrow \begin{pmatrix} X_i \\ Y_i \\ Z_i \end{pmatrix}, \tag{2a}$$

where the arrow represents transformation through equation (1*a*) from coordinates in crystal space to Cartesian coordinates.

The vector length is calculated using the square root of the dot (or scalar) product function of the vector with itself:

$$|\mathbf{p}_i|^2 = \mathbf{p}_i \cdot \mathbf{p}_i = \begin{pmatrix} X_i \\ Y_i \\ Z_i \end{pmatrix} \cdot \begin{pmatrix} X_i \\ Y_i \\ Z_i \end{pmatrix} = X_i^2 + Y_i^2 + Z_i^2. \tag{2b}$$

For a general pair of vectors, $\mathbf{p}$ and $\mathbf{q}$ at an angle θ, the dot product yields a scalar value:

$$\mathbf{p}\cdot\mathbf{q} = |\mathbf{p}||\mathbf{q}|\cos\theta = \begin{pmatrix} X_p \\ Y_p \\ Z_p \end{pmatrix} \cdot \begin{pmatrix} X_q \\ Y_q \\ Z_q \end{pmatrix}$$
$$= X_pX_q + Y_pY_q + Z_pZ_q, \qquad (2c)$$

where each of these vectors is referenced to the coordinate origin. In *Excel*, the dot product can be simulated by the function SUMPRODUCT.

In the operation of the dot product function, the cosine first provides a projection of one vector onto the other; then the product function multiplies and sums the components together [equation (2*c*)]. In physics, for example, this may correspond to work as the product of force times distance in the same direction as determined by the angle θ between the vectors. When the two vectors are parallel (as in the determination of a bond length), θ is zero and the cosine multiplier has the value one.

The vector length r_{ji} between points i and j, corresponding to a *bond length*, is obtained from the coordinate differences with the vector being multiplied by itself:

$$r_{ji}^2 = |\mathbf{r}_{ji}|^2 = \mathbf{r}_{ji}\cdot\mathbf{r}_{ji} = (X_j - X_i)^2 + (Y_j - Y_i)^2 + (Z_j - Z_i)^2$$
$$= \Delta X_{ji}^2 + \Delta Y_{ji}^2 + \Delta Z_{ji}^2. \qquad (2d)$$

Bond angles, θ, can be calculated using either the dot (scalar) product or the cross (vector) product function; the latter generates a pseudovector norm (defined below), **n**, orthogonal to the vector pair $\mathbf{r}_{ji}$ and $\mathbf{s}_{jk}$ which lie at an angle θ with respect to one another. The value of this vector norm **n** is only required if a torsion angle is to be calculated, as will be seen in the *Torsion* paragraph below.

The bond angle is calculated with respect to the atom sequence i—j—k with each bond vector, $\mathbf{r}_{ji}$ and $\mathbf{s}_{jk}$, referenced to the coordinates of the central atom j of the triplet; for example, $\Delta X_r = (X_j - X_i)$.

θ can be determined from the arc cosine of the dot product (Cockcroft, 2006):

$$rs = \mathbf{r}_{ji}\cdot\mathbf{s}_{jk} = |\mathbf{r}_{ji}||\mathbf{s}_{jk}|\cos\theta_{ijk} = \begin{pmatrix} \Delta X_r \\ \Delta Y_r \\ \Delta Z_r \end{pmatrix} \cdot \begin{pmatrix} \Delta X_s \\ \Delta Y_s \\ \Delta Z_s \end{pmatrix}$$
$$= \Delta X_r\Delta X_s + \Delta Y_r\Delta Y_s + \Delta Z_r\Delta Y_s. \qquad (3a)$$

Algebraically (Cockcroft, 2006)

$$rs = \mathbf{r}\cdot\mathbf{s} = |\mathbf{r}||\mathbf{s}|\cos\theta,$$

where

$$\mathbf{r} = \Delta x_r\mathbf{a} + \Delta y_r\mathbf{b} + \Delta z_r\mathbf{c} \quad \text{and} \quad \mathbf{s} = \Delta x_s\mathbf{a} + \Delta y_s\mathbf{b} + \Delta z_s\mathbf{c},$$

$$\begin{aligned}\mathbf{r}\cdot\mathbf{s} &= (\Delta x_r\mathbf{a} + \Delta y_r\mathbf{b} + \Delta z_r\mathbf{c})\cdot(\Delta x_s\mathbf{a} + \Delta y_s\mathbf{b} + \Delta z_s\mathbf{c}) \\ &= \Delta x_r\Delta x_s(\mathbf{a}\cdot\mathbf{a}) + \Delta y_r\Delta y_s(\mathbf{b}\cdot\mathbf{b}) + \Delta z_r\Delta z_s(\mathbf{c}\cdot\mathbf{c}) \\ &\quad + (\Delta y_r\Delta z_s + \Delta y_s\Delta z_r)(\mathbf{b}\cdot\mathbf{c}) \\ &\quad + (\Delta z_r\Delta x_s + \Delta z_s\Delta x_r)(\mathbf{c}\cdot\mathbf{a}) \\ &\quad + (\Delta x_r\Delta y_s + \Delta x_s\Delta y_r)(\mathbf{a}\cdot\mathbf{b}),\end{aligned}$$

and

$$\begin{aligned}\cos\theta = [&a^2\Delta x_r\Delta x_s + b^2\Delta y_r\Delta y_s + c^2\Delta z_r\Delta z_s \\ &+ bc\cos\alpha(\Delta y_r\Delta z_s + \Delta y_s\Delta z_r) \\ &+ ca\cos\beta(\Delta z_r\Delta x_s + \Delta z_s\Delta x_r) \\ &+ ab\cos\gamma(\Delta x_r\Delta y_s + \Delta x_s\Delta y_r)]/rs. \qquad (3b)\end{aligned}$$

Alternatively, the cross product, $\mathbf{r}\times\mathbf{s}$, yields the pseudovector, **n**, which is normal to the plane defined by **r** and **s**. The bond angle, θ_{ijk}, is determined using the arc sine function:

$$\mathbf{r}_{ji}\times\mathbf{s}_{jk} = |\mathbf{r}_{ji}||\mathbf{s}_{jk}|\sin\theta_{ijk}\ \mathbf{n}_{ijk}\begin{pmatrix} \Delta X_r \\ \Delta Y_r \\ \Delta Z_r \end{pmatrix} \times \begin{pmatrix} \Delta X_s \\ \Delta Y_s \\ \Delta Z_s \end{pmatrix}$$
$$= \begin{pmatrix} \Delta Y_r\Delta Z_s - \Delta Z_r\Delta Y_s \\ \Delta Z_r\Delta X_s - \Delta X_r\Delta Z_s \\ \Delta X_r\Delta Y_s - \Delta Y_r\Delta X_s \end{pmatrix}. \qquad (3c)$$

A pseudovector (or axial vector), being perpendicular to the plane of the three atoms forming the bond, changes sign when converted to its mirror image, so that the cross product is non-commutative:

$$\mathbf{r}\times\mathbf{s} = -\mathbf{s}\times\mathbf{r}. \qquad (3d)$$

The standard selection for the sign of the torsion is positive when **r** rotates to **s** according to the right-hand rule. The cross product, $\mathbf{r}\times\mathbf{s} = |\mathbf{r}||\mathbf{s}|\sin\theta$, is the signed area of the parallelogram bounded by the vectors **r** and **s**. Similarly, the (scalar) triple product, $\mathbf{a}\cdot\mathbf{b}\times\mathbf{c}$ (https://en.wikipedia.org/wiki/Triple_product), represents the signed volume, V, of the unit cell with cell constants a, b, c, α, β, γ [for the algebraic form of V, see equation (1*b*)]. *Excel* does not have a cross-product function, but a user-defined cross-product function is listed in the supplementary information file.

Torsion angles, ϕ, are calculated as the twist around the central bond j—k of the two planes defined by the orthogonal

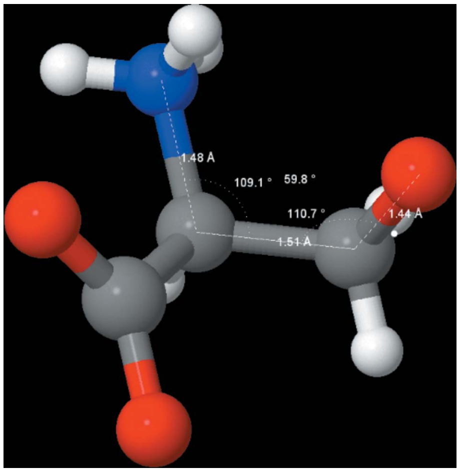


Figure 1
The three bond lengths (1.48, 1.51 and 1.44 Å), two bond angles (109.1 and 110.7°) and single torsion angle (+59.8°) for the atom sequence N—C—C—O(H) in L-serine. The diagram was prepared using *Jmol* and the CIF available as supporting information to the article by Moggach *et al.* (2005) (CSD-LSERIN11.cif from the CSD).

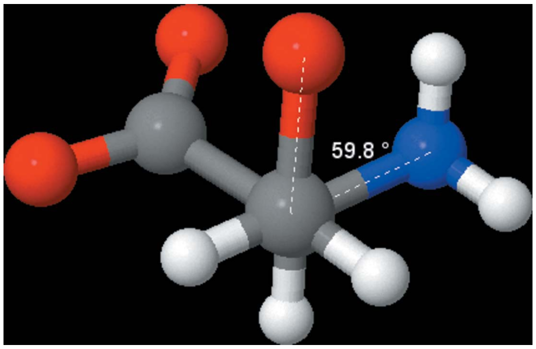


Figure 2
View down the atom sequence (H)O—C—C—N in L-serine, highlighting the right-hand screw twist around the central C—C bond relating the O atom to the N atom, yielding a torsion angle of +59.8°. The diagram was prepared using *Jmol* and the `CSD-LSERIN11.cif` file (Moggach *et al.*, 2005).

pseudovectors of the angle triplets $i{-}j{-}k$ and $j{-}k{-}l$ (see Figs. 1 and 2). Thus, we use the dot product of the pair of triplets to generate the cosine of the torsion angle, $\cos\phi_{ijkl}$, from which we evaluate ϕ_{ijkl} through the arc cosine function:

$$\begin{aligned}\phi_{ijkl} &= \cos^{-1}\left[\frac{(\mathbf{r}_{ji}\times\mathbf{r}_{jk})\cdot(\mathbf{r}_{kj}\times\mathbf{r}_{kl})}{|\mathbf{r}_{ji}\cdot\mathbf{r}_{jk}||\mathbf{r}_{kj}\cdot\mathbf{r}_{kl}|}\right]\\ &= \cos^{-1}\left[\frac{(\mathbf{r}_{ji}\times\mathbf{r}_{jk})\cdot(\mathbf{r}_{kj}\times\mathbf{r}_{kl})}{rs}\right].\end{aligned}\qquad (4a)$$

It remains to determine the sign of ϕ_{ijkl} (between −180 and +180°), which is accomplished in equation (4*b*) by calculating first the cross product of the orthogonal axial vectors of the two planes defining the bond angles (this cross product generates a new vector, now parallel to the planes) and then the scalar dot product of this vector with the vector representing the central bond $j{-}k$:

$$\text{sign of } (\mathbf{r}_{kj}\times\mathbf{r}_{lk})\cdot\mathbf{r}_{jk}. \qquad (4b)$$

By the right-hand rule, if these parallel vectors point in the same direction the torsion angle is positive, but it is negative if they point in opposite directions.

3. The metric matrix and crystal (vector) space

Converting from the fractional coordinates of crystal (vector) space to Cartesian coordinates of standard space is convenient, as shown in Section 4, but it takes no account of the symmetry of crystals. It is sometimes appropriate to perform geometry in the crystal space directly (for example, to generate symmetry-related atoms absent from the asymmetric unit data provided in the CIF) (De Graef & McHenry, 2011, 2012; https://dictionary.iucr.org/Asymmetric_unit). This introduces the scalar metric matrix, **G** (also known as the metric tensor, **g**, when it use extends to nonlinear processes and is no longer a scalar), which uses the six crystal constants, **a**, **b**, **c**, α, β, γ, as the basis vectors in evaluations of dot products:

$$\begin{aligned}r^2 &= \mathbf{r}\cdot\mathbf{r} = (\Delta x\mathbf{a}+\Delta y\mathbf{b}+\Delta z\mathbf{c})\cdot(\Delta x\mathbf{a}+\Delta y\mathbf{b}+\Delta z\mathbf{c})\\ &= \Delta x^2(\mathbf{a}\cdot\mathbf{a})+\Delta y^2(\mathbf{b}\cdot\mathbf{b})+\Delta z^2(\mathbf{c}\cdot\mathbf{c})+2\Delta y\Delta z(\mathbf{b}\cdot\mathbf{c})\\ &\quad+2\Delta z\Delta x(\mathbf{c}\cdot\mathbf{a})+2\Delta x\Delta y(\mathbf{a}\cdot\mathbf{b}).\end{aligned}\qquad (5a)$$

We collect the basis vectors into a metric matrix, **G**, where

$$\begin{aligned}\mathbf{G} &= \begin{pmatrix}\mathbf{a}\cdot\mathbf{a} & \mathbf{a}\cdot\mathbf{b} & \mathbf{a}\cdot\mathbf{c}\\ \mathbf{b}\cdot\mathbf{a} & \mathbf{b}\cdot\mathbf{b} & \mathbf{b}\cdot\mathbf{c}\\ \mathbf{c}\cdot\mathbf{a} & \mathbf{c}\cdot\mathbf{b} & \mathbf{c}\cdot\mathbf{c}\end{pmatrix}\\ &= \begin{pmatrix}a^2 & ab\cos\gamma & ac\cos\beta\\ ba\cos\gamma & b^2 & bc\cos\alpha\\ ca\cos\beta & cb\cos\alpha & c^2\end{pmatrix},\end{aligned}\qquad (5b)$$

so that

$$r^2 = |\mathbf{r}|^2 = \mathbf{r}\cdot\mathbf{r} = (\,\Delta x \quad \Delta y \quad \Delta z\,)\,\mathbf{G}\begin{pmatrix}\Delta x\\ \Delta y\\ \Delta z\end{pmatrix}.$$

Defining

$$\begin{pmatrix}\Delta x\\ \Delta y\\ \Delta z\end{pmatrix} = X,$$

so that $(\Delta x\ \ \Delta y\ \ \Delta z) = X^{\rm T}$, where the superscript T represents the transpose,

$$r^2 = |\mathbf{r}|^2 = \mathbf{r}\cdot\mathbf{r} = X^{\rm T}\mathbf{G}X,$$

$$\begin{aligned}r^2 &= a^2\Delta x^2 + b^2\Delta y^2 + c^2\Delta z^2 + 2bc\cos\alpha\Delta y\Delta z\\ &\quad + 2ca\cos\beta\Delta z\Delta x + 2ab\cos\gamma\Delta x\Delta y.\end{aligned}\qquad (5c)$$

As noted earlier, the volume of the unit cell is equal to the square root of the determinant of **G**, which can be found in *Excel* using the function `MDETERM`.

In computation, it is convenient to record the six terms in equation (5*c*) a^2, b^2, $c^2, \ldots, ab\cos\gamma$ individually, since they might be used repeatedly for both bond length and bond angle calculations (Cockcroft, 2006). In the algebraic expansion, equation (5*c*) provides the terms which need to be summed in order to calculate a bond length. However, it is simpler in *Excel* to calculate the metric matrix expression of equation (5*c*) directly using the `MMULT` array function, without requiring the six-term expansion.

The metric matrix can be used to calculate bond angles directly, without having to calculate bond lengths |**r**| and |**s**| independently, as follows:

$$\mathbf{r}\cdot\mathbf{s} = |\mathbf{r}||\mathbf{s}|\cos\theta = \begin{pmatrix}\Delta x_r\\ \Delta y_r\\ \Delta z_r\end{pmatrix}\cdot\begin{pmatrix}\Delta x_s\\ \Delta y_s\\ \Delta z_s\end{pmatrix}.$$

This relation can also be expressed in terms of the metric matrix **G**:

$$\mathbf{r}\cdot\mathbf{s} = |\mathbf{r}||\mathbf{s}|\cos\theta = \begin{pmatrix} \Delta x_r & \Delta y_r & \Delta z_r \\ \Delta x_s & \Delta y_s & \Delta z_s \end{pmatrix} \mathbf{G} \begin{pmatrix} \Delta x_r & \Delta x_s \\ \Delta y_r & \Delta y_s \\ \Delta z_r & \Delta z_s \end{pmatrix}$$

$$= (\mathbf{r} \quad \mathbf{s})\,\mathbf{G}\begin{pmatrix} \mathbf{r} \\ \mathbf{s} \end{pmatrix} = \begin{pmatrix} \mathbf{r}\cdot\mathbf{r} & \mathbf{s}\cdot\mathbf{r} \\ \mathbf{r}\cdot\mathbf{s} & \mathbf{s}\cdot\mathbf{s} \end{pmatrix}. \qquad (6)$$

Hence

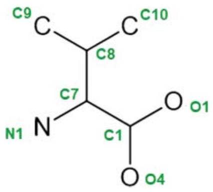
$$\theta = \cos^{-1}\left[\frac{\mathbf{r}\cdot\mathbf{s}}{(\mathbf{r}\cdot\mathbf{r})^{1/2}(\mathbf{s}\cdot\mathbf{s})^{1/2}}\right] = \cos^{-1}(\mathbf{r}\cdot\mathbf{s}/rs).$$

Torsion angle calculations are generally regarded as being too complex to be formulated with fractional coordinates, and are often performed using Cartesian coordinates, as discussed above. However, the BASIC program provided by Dunitz (1995) does not have this limitation.

4. An example calculation (for L-valine) using both algebraic expansions and vector methods

For this exercise we will demonstrate how the equations introduced in Sections 2 and 3 are implemented in practice, using the example of L-valine (Torii & Iitaka, 1970) with the data and results as appear in the Valine worksheet of the supplementary workbook `gj5247sup2.xlsx`.

Atom numbering:



Unit cell constants: $a = 9.71$, $b = 5.27$, $c = 12.06$ Å, $\alpha = 90$, $\beta = 90.8$, $\gamma = 90°$.

Transformation matrix for conversion from fractional coordinates, x, y, z, to Cartesian coordinates, X, Y, Z:

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \begin{bmatrix} a & b\cos\gamma & c\cos\beta \\ 0 & b\sin\gamma & \dfrac{c(\cos\alpha - \cos\beta\cos\gamma)}{\sin\gamma} \\ 0 & 0 & \dfrac{V}{ab\sin\gamma} \end{bmatrix} \cdot \begin{pmatrix} x \\ y \\ z \end{pmatrix}, \qquad (1a)$$

where the volume of the unit cell

$$V = abc(1 - \cos^2\alpha - \cos^2\beta - \cos^2\gamma + 2\cos\alpha\cos\beta\cos\gamma)^{1/2}, \qquad (1b)$$

$$V_{\rm cell} = 6.17071.$$

Note: in *Excel*, angles must be in radians, where `rads = degs*pi()/180`.

Algebraic result, after matrix multiplication:

Table 1
The input fractional atomic coordinates.

	C1	C7	C8	C9	C10	N1	O1	O4
x	−0.2234	−0.3649	−0.4025	−0.4574	−0.2757	−0.3753	−0.1295	−0.2084
y	−0.123	−0.0062	0.0412	−0.1998	0.1525	0.2332	0.0167	−0.3534
z	0.3635	0.3457	0.2224	0.1658	0.1597	0.4116	0.3996	0.3367

Table 2
The resulting Cartesian atomic coordinates.

X	−2.2304	−3.6014	−3.9457	−4.4693	−2.7039	−3.7135	−1.3247	−2.0803
Y	−0.6482	−0.0327	0.2171	−1.0529	0.8037	1.2290	0.0880	−1.8624
Z	4.3834	4.1687	2.6819	1.9994	1.9258	4.9634	4.8187	4.0602

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \begin{bmatrix} xa + yb\cos\gamma + zc\cos\beta \\ yb\sin\gamma + zc(\cos\alpha - \cos\beta\cos\gamma)/\sin\gamma \\ zV/ab\sin\gamma \end{bmatrix}. \qquad (1c)$$

The fractional atomic coordinates input from the `LVALIN.cif` file from the CSD are given in Table 1.

Coordinate transformation, from fractional to Cartesian, using atom C1 as an example:

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \begin{bmatrix} -0.2234 \times 9.71 + \cdots \\ -0.123 \times 5.27\sin(90 \times 3.142/180) + \cdots \\ 0.3635 \times 617.1/\cdots \end{bmatrix}$$

$$= \begin{pmatrix} -2.2304 \\ -0.6482 \\ 4.3834 \end{pmatrix}.$$

Alternatively, by direct matrix multiplication following substitutions into equation (1*a*)

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \begin{pmatrix} 9.71 & 0 & -0.1684 \\ 0 & 5.27 & 0 \\ 0 & 0 & 12.06 \end{pmatrix} \begin{pmatrix} -0.2234 \\ -0.123 \\ 0.3635 \end{pmatrix}$$

$$= \begin{pmatrix} -2.2304 \\ -0.6482 \\ 4.3834 \end{pmatrix}.$$

The resulting Cartesian atomic coordinates are given in Table 2.

Bond lengths [equation (2*d*) or (5*c*)]:

Using Cartesian coordinates,

$$\begin{aligned} r^2 &= |\mathbf{r}_{ji}|^2 = \mathbf{r}_{ji}\cdot\mathbf{r}_{ji} = |\mathbf{r}_{ji}||\mathbf{r}_{ji}| \\ &= (X_j - X_i)^2 + (Y_j - Y_i)^2 + (Z_j - Z_i)^2 \\ &= \Delta X_{ji}^2 + \Delta Y_{ji}^2 + \Delta Z_{ji}^2. \end{aligned} \qquad (2d)$$

Using fractional coordinates,

$$\begin{aligned} r^2 &= a^2\Delta x^2 + b^2\Delta y^2 + c^2\Delta z^2 + 2bc\cos\alpha\Delta y\Delta z \\ &\quad + 2ca\cos\beta\Delta z\Delta x + 2ab\cos\gamma\Delta x\Delta y. \end{aligned} \qquad (5c)$$

The resulting bond lengths are C8—C9 = 1.534 Å, C7—C8 = 1.547 Å and C1—O4 = 1.265 Å.

Bond angles [equations (3*a*) or (6)]:

Table 3
Cartesian coordinates for the O1—C1—C7 bond angle.

	Bond **r** = O1—C1		Bond **s** = C7—C1
ΔX_r	0.9057	ΔX_s	−1.3710
ΔY_r	0.7362	ΔY_s	0.6155
ΔZ_r	0.4353	ΔZ_s	−0.2146
$r = \lvert\mathbf{r}\rvert$	**1.246**	$s = \lvert\mathbf{s}\rvert$	**1.518**

A bond angle is calculated with respect to the atom sequence i—j—k with each bond vector, $\mathbf{r}_{ji}$ and $\mathbf{s}_{jk}$, referenced to the coordinates of the central atom j of the triplet.

Consider the Cartesian coordinates for the bond angle O1—C1—C7 (Table 3), extracted from the supplementary Valine spreadsheet.

Using the scalar dot product with Cartesian coordinates, and expanding algebraically,

$$rs = \mathbf{r}_{ji}\cdot\mathbf{s}_{jk} = |\mathbf{r}_{ji}||\mathbf{s}_{jk}|\cos\theta_{ijk} = \begin{pmatrix}\Delta X_r\\ \Delta Y_r\\ \Delta Z_r\end{pmatrix}\cdot\begin{pmatrix}\Delta X_s\\ \Delta Y_s\\ \Delta Z_s\end{pmatrix}$$
$$= \Delta X_r\Delta X_s + \Delta Y_r\Delta Y_s + \Delta Z_r\Delta Z_s. \qquad (3a)$$

In our *Excel* Valine worksheet, the bond angle for O1—C1—C7 is calculated in cell J24 using the *Excel* expression

```
= ACOS(SUMPRODUCT(M24:M26,O24:O26)/(M27*O27)) = 117.80°
```

where M24:M26 and O24:O26 refer to the Cartesian coordinate differences: ΔX_r, ΔY_r, ΔZ_r for bond O1—C1 and ΔX_s, ΔY_s, ΔZ_s for bond C7—C1, normalized by dividing by the lengths of the respective bonds, M27 and O27. Note that the bond differences are all calculated with respect to the central atom, C1.

Expressing this calculation in matrix terms [equation (6)],

$$\mathbf{r} = \begin{pmatrix}0.9057\\ 0.7362\\ 0.4353\end{pmatrix}, \quad \mathbf{s} = \begin{pmatrix}-1.371\\ 0.6155\\ -0.2146\end{pmatrix}, \quad \theta = \cos^{-1}(\mathbf{r}\cdot\mathbf{s}/rs),$$

$$\theta = \cos^{-1}(-0.8820/1.246\times 1.518) = 117.8^\circ,$$

where $\mathbf{r}\cdot\mathbf{s}$ is determined in *Excel* as `SUMPRODUCT`(|**r**|,|**s**|).

Algebraically, with fractional coordinates,

$$\cos\theta = [a^2\Delta x_r\Delta x_s + b^2\Delta y_r\Delta y_s + c^2\Delta z_r\Delta z_s$$
$$+ bc\cos\alpha(\Delta y_r\Delta z_s + \Delta y_s\Delta z_r)$$
$$+ ca\cos\beta(\Delta z_r\Delta x_s + \Delta z_s\Delta x_r)$$
$$+ ab\cos\gamma(\Delta x_r\Delta y_s + \Delta x_s\Delta y_r)]/rs. \qquad (3b)$$

Alternatively, a bond angle can be found using the vector cross product:

$$\mathbf{r}_{ji}\times\mathbf{s}_{jk} = |\mathbf{r}_{ji}||\mathbf{s}_{jk}|\sin\theta_{ijk}\ \mathbf{n}_{ijk}\begin{pmatrix}\Delta X_r\\ \Delta Y_r\\ \Delta Z_r\end{pmatrix}\times\begin{pmatrix}\Delta X_s\\ \Delta Y_s\\ \Delta Z_s\end{pmatrix}$$
$$= \begin{pmatrix}\Delta Y_r\Delta Z_s - \Delta Z_r\Delta Y_s\\ \Delta Z_r\Delta X_s - \Delta X_r\Delta Z_s\\ \Delta X_r\Delta Y_s - \Delta Y_r\Delta X_s\end{pmatrix}. \qquad (3c)$$

The cross product can be determined using the user-defined function CVp (listed in the supplementary file) with the pair of bonds in Table 3. This generates a vector normal to the plane of the bond pair:

$$|\mathbf{r}_{ji}||\mathbf{s}_{jk}|\sin\theta_{ijk} = rs\sin\theta_{ijk} = \begin{pmatrix}\Delta X_r\\ \Delta Y_r\\ \Delta Z_r\end{pmatrix}\times\begin{pmatrix}\Delta X_s\\ \Delta Y_s\\ \Delta Z_s\end{pmatrix}$$
$$= \begin{pmatrix}X_{rs}\\ Y_{rs}\\ Z_{rs}\end{pmatrix} = \begin{pmatrix}-0.4259\\ -0.4024\\ 1.5668\end{pmatrix}.$$

The length of this vector is $|\mathbf{v}_{\rm rs}| = (X_{\rm rs}^2 + Y_{\rm rs}^2 + Z_{\rm rs}^2)^{1/2} = 1.673$. Hence

$$\theta_{rs} = \sin^{-1}(v_{rs}/rs) = \sin^{-1}[1.673/(1.246\times 1.518)]$$
$$= 1.085\ \text{rad} = 62.2\ \text{or}\ 117.8^\circ.$$

A choice needs to be made between the angle and its supplement. Note that the incorrect acute angle is also returned by the algebraic method unless the bond vectors are first referenced to the central atom j of the sequence ijk.

Torsion angles are best found by the complex method of the Torsion worksheet, using equations (4*a*) and (4*b*), or by simple substitution into the `gj5247sup1.xlsx` workbook.

5. Conclusions

Procedures by which bond lengths, bond angles and torsion angles can be calculated from either Cartesian or fractional crystal coordinates, individually or by use of metric matrices, are illustrated and also demonstrated using live *Excel* spreadsheets. These procedures exemplify the otherwise hidden methods used in crystallographic and molecular visualization programs.

6. Supplementary files

Supporting information:

(i) An *Excel* macro-enabled workbook, `gj5247sup1.xls`, which calculates torsion angles, bond angles and bond lengths using fractional coordinate data inserted by the user into the worksheet. This *Excel* file contains a macro.

(ii) An *Excel* workbook, `gj5247sup2.xlsx`, consisting of four worksheets labelled SF6, Serine, Valine and Torsion, which lays out the matrix calculations involved in calculating molecular geometry from fractional coordinates. All the calculations are performed live, using standard *Excel* functions.

(iii) The gj5247sup3.pdf file describes the contents and operations within the four worksheets of gj5247sup2.xls, and describes a user-defined *Excel* cross-product function.

Acknowledgements

I thank both referees for their detailed and helpful corrections and comments which have saved me from much embarrassment. This contribution was inspired by the 'Molecular Geometry' analysis of J. K. Cockcroft (2006).

References

Aldeborgh, H., George, K., Howe, M., Lowman, H., Moustakas, H., Strunsky, N. & Tanski, J. M. (2014). *J. Chem. Crystallogr.* **44**, 70–81.

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. (2000). *Nucleic Acids Res.* **28**, 235–242.

Chapuis, G. (2011). *Crystallogr. Rev.* **17**, 187–204.

Cockcroft, J. K. (2006). *Molecular Geometry*, http://pd.chem.ucl.ac.uk/pdnn/refine2/geometry.htm.

De Graef, M. & McHenry, M. E. (2011). *Structure of Materials: Additional Material*, http://som.web.cmu.edu/frames.html.

De Graef, M. & McHenry, M. E. (2012). *Structure of Materials: An Introduction to Crystallography, Diffraction and Symmetry*, 2nd ed. Cambridge University Press.

Downs, R. T. & Hall-Wallace, M. (2003). *Am. Miner.* **88**, 247–250.

Dunitz, J. D. (1995). *X-ray Analysis and the Structure of Organic Molecules*, 2nd ed. Basel: Verlag Helvetica Chimica Acta.

Glasser, L. (1993). *Comput. Chem.* **17**, 107–108.

Glasser, L. (2016). *J. Chem. Educ.* **93**, 542–549.

Gražulis, S., Chateigner, D., Downs, R. T., Yokochi, A. F. T., Quirós, M., Lutterotti, L., Manakova, E., Butkus, J., Moeck, P. & Le Bail, A. (2009). *J. Appl. Cryst.* **42**, 726–729.

Gražulis, S., Sarjeant, A. A., Moeck, P., Stone-Sundberg, J., Snyder, T. J., Kaminsky, W., Oliver, A. G., Stern, C. L., Dawe, L. N., Rychkov, D. A., Losev, E. A., Boldyreva, E. V., Tanski, J. M., Bernstein, J., Rabeh, W. M. & Kantardjieff, K. A. (2015). *J. Appl. Cryst.* **48**, 1964–1975.

Groom, C. R., Bruno, I. J., Lightfoot, M. P. & Ward, S. C. (2016). *Acta Cryst.* B**72**, 171–179.

Hall, S. R., Allen, F. H. & Brown, I. D. (1991). *Acta Cryst.* A**47**, 655–685.

Julian, M. M. (2014). *Foundations of Crystallography with Computer Applications*, 2nd ed. Abingdon-on-Thames: CRC Press.

Kantardjieff, K. (2010). *J. Appl. Cryst.* **43**, 1276–1282.

Macrae, C. F., Sovago, I., Cottrell, S. J., Galek, P. T. A., McCabe, P., Pidcock, E., Platings, M., Shields, G. P., Stevens, J. S., Towler, M. & Wood, P. A. (2020). *J. Appl. Cryst.* **53**, 226–235.

McRee, D. E. (1993*a*). *Practical Protein Crystallography*. San Diego: Academic Press.

McRee, D. E. (1993*b*). *Practical Protein Crystallography*, https://www.sciencedirect.com/book/9780124860506/practical-protein-crystallography.

Moggach, S. A., Allan, D. R., Morrison, C. A., Parsons, S. & Sawyer, L. (2005). *Acta Cryst.* B**61**, 58–68.

Momma, K. & Izumi, F. (2011). *J. Appl. Cryst.* **44**, 1272–1276.

O'Boyle, N. M., Banck, M., James, C. A., Morley, C., Vandermeersch, T. & Hutchison, G. R. (2011). *J. Cheminform.* **3**, 33.

Thomas, I. R., Bruno, I. J., Cole, J. C., Macrae, C. F., Pidcock, E. & Wood, P. A. (2010). *J. Appl. Cryst.* **43**, 362–366.

Torii, K. & Iitaka, Y. (1970). *Acta Cryst.* B**26**, 1317–1326.

Zagorac, D., Müller, H., Ruehl, S., Zagorac, J. & Rehme, S. (2019). *J. Appl. Cryst.* **52**, 918–925.