Title

Implementation of the riding hydrogen model in CCTBX to support the next generation of X-ray and neutron joint refinement in Phenix

Authors

Liebschner, Dorothee
Afonine, Pavel V
Urzhumtsev, Alexandre G
et al.

Publication Date

2020

DOI

10.1016/bs.mie.2020.01.007

Peer reviewed

HHS Public Access

Author manuscript

Methods Enzymol. Author manuscript; available in PMC 2020 October 20.

Published in final edited form as:

Methods Enzymol. 2020 ; 634: 177–199. doi:10.1016/bs.mie.2020.01.007.

Implementation of the riding hydrogen model in *CCTBX* to support the next generation of X-ray and neutron joint refinement in *Phenix*

Dorothee Liebschner[a,*], **Pavel V. Afonine**[a], **Alexandre G. Urzhumtsev**[b,c], **Paul D. Adams**[a,d]

[a]Molecular Biophysics and Integrated Bioimaging Division, Lawrence Berkeley National Laboratory (LBNL), Berkeley, CA, United States

[b]Centre for Integrative Biology, Institut de Génétique et de Biologie Moléculaire et Cellulaire, Illkirch, France

[c]Faculté des Sciences et Technologies, Université de Lorraine, Nancy, France

[d]Department of Bioengineering, University of California Berkeley, Berkeley, CA, United States

Abstract

A fundamental prerequisite for implementing new procedures of atomic model refinement against neutron diffraction data is the efficient handling of hydrogen atoms. The riding hydrogen model, which constrains hydrogen atom parameters to those of the non-hydrogen atoms, is a plausible parameterization for refinements. This work describes the implementation of the riding hydrogen model in the *Computational Crystallography Toolbox* and in *Phenix*. Riding hydrogen atoms can be found in several different configurations that are characterized by specific geometries. For each configuration, the hydrogen atom parameterization and the expressions for the gradients of refinement target function with respect to non-hydrogen parameters are described.

1. Introduction

X-ray crystallography is the leading method for obtaining three-dimensional structures of macromolecules. Oftentimes, hydrogen (H) atoms play an important biological role and may be the subject of a structural study, in particular when trying to understand enzyme mechanism. Accurate determination of the location of key H atoms therefore requires experimental information, but as hydrogen has a weak X-ray scattering cross section, its electron density is not observed unless very high quality, and very high-resolution data are available. Even then, it is not guaranteed that all H atoms will be observed: Petrova and Podjarny (2004) showed that only about half of the hydrogen atom positions can be determined experimentally with high resolution electron density maps of macromolecules. By contrast, neutron crystallography allows the observation of hydrogen atoms (or deuterium substitutes) at resolutions better than 3 Å (Ostermann, Tanaka, Engler, Niimura, & Parak, 2002; Shu, Ramakrishnan, & Schoenborn, 2000).

*Corresponding author: dcliebschner@lbl.gov.

The determination of macromolecular structures with neutron diffraction has two major complications. First, since H atoms have a similar scattering contribution as other atoms, they need to be refined individually in the atomic model. Also, as a result of partial deuteration, some H atom positions are shared with deuterium atoms (D) meaning that their relative occupancies need to be refined. Given that H atoms constitute about 50% of the atoms in a typical macromolecular structure, refining them individually adds a substantial number of refined parameters. Second, the experimental data for neutron diffraction data are often poorer than those of corresponding X-ray data: the low flux of available neutron beams and incoherent scattering of hydrogen (if present) lead to low signal-to-noise ratios, and the limited data-collection time on oversubscribed instruments result in low completeness, averaging about 80% (Liebschner, Afonine, Moriarty, Langan, & Adams, 2018). In contrast, the completeness of a typical X-ray dataset is expected to be greater than 95% (Dauter & Dauter, 2017).

With rare exceptions, X-ray data and the corresponding atomic models are almost always available prior to structure determination with neutrons. The availability of two sources of information, namely X-ray and neutron data, may help alleviate some of the aforementioned challenges of refinement against neutron data: the joint X-ray and neutron (joint XN) refinement method enables building more complete atomic models that include experimentally observed hydrogen and non-hydrogen atoms (Adams, Mustyakimov, Afonine, & Langan, 2009; Afonine et al., 2010; Coppens, 1967; Wlodawer, 1980; Wlodawer & Hendrickson, 1982; Wlodawer, Miller, & Sjölin, 1983). The currently used joint XN refinement method optimizes a single atomic model simultaneously against two datasets using a combined refinement target function:

$$T = w_x \bullet T_{\mathrm{X-ray}} + w_n \bullet T_{\mathrm{neutron}} + T_{\mathrm{restraints}} \quad (1)$$

where $T_{\text{X-ray}}$ and T_{neutron} are target functions relating the model and respective data, $T_{\text{restraints}}$ is a restraint term that adds a priori information, and w_{x} and w_{n} are empirical scale factors.

Using the refinement target (1) with a single atomic model relies on the assumption that the X-ray and neutron crystals and respective structures are isomorphous. This hypothesis is only approximate, because:

(a) If different crystals are used for the neutron and X-ray experiments, the unit cell parameters of the crystals may be different.

(b) Ordered and semi-ordered parts of the structures may be different.

(c) X-rays are scattered by electrons while neutrons are scattered by atomic nuclei. The datasets therefore convey different information about covalent X-H bond lengths: the electron distribution of the H atom, which has only one valence electron and no core electron, is shifted along the X-H bond away from the H atom nucleus toward atom X. Accordingly, the locations of H atom electron density peaks and nuclear scattering length density peaks are different and the X-H bond lengths appear about 10–20% shorter for X-rays (Allen, 1986; Allen & Bruno, 2010).

(d) The X-ray and neutron datasets may be collected under different conditions. For example, X-ray data are typically collected at cryogenic temperatures to mitigate radiation damage while neutron diffraction experiments are carried out at room temperature because neutrons do not damage the sample. Atomic models of structures derived from room-temperature experiments are likely to exhibit more variability due to thermal motion (Fraser et al., 2011).

(e) Data collection times are very different: X-ray data can be collected in minutes at synchrotron beamlines (Turkenburg & McAuley, 2013) while neutron data collections typically last days or weeks (Chen & Unkefer, 2017; Coates et al., 2015). Atomic models derived from diffraction data are averaged over space (unit cells of the crystal) and time (duration of the data collection). It may be that data collected over substantially different periods of time result in different models.

Analysis of atomic models derived from joint XN refinement (see Chapter “What are the current limits on determination of protonation state using neutron macromolecular crystallography?” by Liebschner et al.) shows that in practice there often are differences between the neutron and X-ray datasets (for the reasons just described). Therefore, it is desirable for a joint XN refinement algorithm to yield two models that are related to each other and contain H atoms; each model should fit their respective dataset and reflect its properties while at the same time it should benefit from the complementarity of the X-ray and neutron data. Implementation details of such an approach will be described elsewhere. In this chapter we describe a fundamental prerequisite for implementing such a procedure in the software package *Phenix* (Liebschner et al., 2019), namely efficient handling of H atoms.

The contribution of H atoms can be modeled in several ways. One possibility is to include the H atoms explicitly in the input model and to refine them independently, similar to the other non-H atoms. Alternatively, while also being modeled explicitly, their parameters can be constrained by those of the non-H atoms. This approach does not increase the number of parameters even when the model contains H atoms. This is because only the positions of the non-H atoms are refined explicitly while the positions of H atoms are recalculated based on geometric constraints. This riding hydrogen model (Busing & Levy, 1964; Sheldrick & Schneider, 1997) is almost universally used in X-ray refinements. It may also be a plausible parameterization for refinements against neutron data (Gruene, Hahn, Luebben, Meilleur, & Sheldrick, 2014), especially for low-resolution or poorly complete datasets. It is also conceivable to mix different hydrogen modeling approaches, such as refining H atoms with geometric degrees of freedom individually while treating the remaining H atoms as riding. Below we describe algorithmic details of the implementation of the riding hydrogen model in the Computational Crystallography Toolbox (*CCTBX*; Grosse-Kunstleve, Sauter, Moriarty, & Adams, 2002) and in *Phenix*.

2. Parameterizing the riding hydrogen atom model for typical geometrical configurations

2.1 Riding H definition

The concept of riding hydrogen model (‘riding H’ in what follows) relies on the fact that the coordinates of most (but not all) H atoms in proteins can be unambiguously expressed through the coordinates of their covalently connected non-H atoms. To define the riding H atom position, at least three neighboring non-H atoms are necessary. These are typically the ‘parent’ atom (A_0), to which the H atom is covalently bound, and two other atoms, that we note by A_1 and A_2, bound to the parent atom (neighbors). An example configuration of a hydrogen atom is shown in Fig. 1 (left). In this example, the H atom is in a plane with A_0, A_1 and A_2. This geometry can be found in the peptide unit or in aromatic rings.

Riding H atoms can be found in several different configurations; for example, three non-H atoms may define simultaneously several H atoms; an example is the Hγ11-Hγ12-Hγ13 group in valine residue; all of its H atoms are bound to the Cγ atom. Other typical examples are shown in Fig. 2. In each configuration, the number of neighbors and their ideal bond and angle values may vary. The algorithmic challenge is to define the coordinates of the H atom from the coordinates of a group of its non-H neighbors for each configuration. It is desirable that the algorithm is independent of eventual imperfections in the geometry of non-H atoms as much as possible.

For each configuration the H atom position is parameterized using coordinates of (at least) three non-H neighbors and by using $\vec{r}_H = \vec{r}_0 + \vec{r}_{H0} = \vec{r}_0 + d_H \bullet \vec{u}_{H0}$, where $\vec{u}_{H0}$ is the unit vector along the A_0-H direction $\vec{r}_{H0}$ and d_H is the ideal H-A_0 bond length (Fig. 1). Since d_H is known for each kind of models (and is different for the X-ray and neutron data), the main task is to define $\vec{u}_{H0}$ from the set of non-H neighbors and ideal bond lengths and angles. A chain of transformations is developed for all possible configurations. Each step passes from a set of input parameters to a set of output parameters. This way, the first step in the chain of transformations uses the coordinates of atoms parameterizing the H atom and respective ideal bond lengths and angles. The last step of the chain of transformations provides $\vec{u}_{H0}$. This way, it is easier to apply the chain rule for gradient calculations (see Section 3 and Appendix) and the procedure can be carried out in the most efficient way (Lunin & Urzhumtsev, 1985).

We note that configurations described below are general and not specific to protein residues only.

2.2 Coplanar configuration

The first configuration describes the H atom linked to the N atom in the peptide unit (Fig. 2A) or the H atoms in aromatic rings. The theoretical configuration is represented in Fig. 1. In the ideal conformation, the unit vectors in the directions $A_0 - A_1\left(\vec{u}_{10}\right)$ and $A_0 - A_2\left(\vec{u}_{20}\right)$ form the angle α_0. Being coplanar, the unit vector $\vec{u}_{H0}$ can be expressed through these two

vectors as $\vec{u}_{H0} = a \bullet \vec{u}_{10} + b \bullet \vec{u}_{20}$ forms the ideal angles α_1 and α_2 with $\vec{u}_{10}$ and $\vec{u}_{20}$, respectively. We note $c_0 = \cos \alpha_0$, $c_1 = \cos \alpha_1$ and $c_2 = \cos \alpha_2$.

The coefficients a and b can be found from the conditions $a + b \bullet c_0 = c_1$, $a \bullet c_0 + b = c_2$ that give an analytic solution $a = \left(1 - {c_0}^2\right)^{-1}(c_1 - c_0 \bullet c_2)$, $b = \left(1 - {c_0}^2\right)^{-1}(c_2 - c_0 \bullet c_1)$. Thus, the values of parameters a and b for the ideal coplanar conformation can be calculated in advance.

The chain of transformations to obtain the H-atom position is the following:

(a) $\vec{r}_{10} = \vec{r}_1 - \vec{r}_0$, $\vec{r}_{20} = \vec{r}_2 - \vec{r}_0$

(b) $\vec{u}_{10}$ *is* normalized $\vec{r}_{10}$, $\vec{u}_{20}$ *is* normalized $\vec{r}_{20}$

(c) $\vec{r}_{H0} = a \bullet \vec{u}_{10} + b \bullet \vec{u}_{20}$

(d) $\vec{u}_{H0} = \left\|\vec{r}_{H0}\right\|^{-1}\vec{r}_{H0}$ with $\left\|\vec{r}_{H0}\right\| = \left(\vec{r}_{H0} \bullet \vec{r}_{H0}\right)^{1/2}$

This procedure positions the H atom in the plane and at a correct distance to its parent atom. We note that the input model does not necessarily possess an ideal geometry, which means that the angle between A_0-A_1 and A_0-A_2 is not necessarily equal to the ideal value α_0, and similarly for α_1 and α_2. However, the riding procedure uses the ideal angle values to calculate the coefficients a and b. This ensures that the H atom geometry is regularized to match the ideal values.

In summary, this configuration requires the following parameters: A_0, A_1, A_2, d_H, a, b.

2.3 2H-tetrahedral configuration

Examples of this configuration are the Hα1 and Hα2 atoms bound to the glycine Cα atom (Fig. 2B). The configuration is represented schematically in Fig. 3. The atoms H_1 and H_2 are at the given distance d_H from atom A_0. In the ideal configuration, they are symmetric with respect to the plane A_1-A_0-A_2 and form an angle H_1-A_0-H_2 equal to 2δ. The unit vector $\vec{u}_{H1}$ in the direction A_0-H_1 forms angles α_1 and α_2 with $\mathrm{A_0 - A_1}\left(\vec{u}_{10}\right)$ and $\mathrm{A_0 - A_2}\left(\vec{u}_{20}\right)$, respectively. If the coefficients a and b are defined as in Section 2.2, the vector $\vec{d} = a \bullet \vec{u}_{10} + b \bullet \vec{u}_{20}$ is collinear with the orthogonal projection of A_0-H_1 and A_0-H_2 onto the plane A_1-A_0-A_2.

Steps (a) and (b) are defined as in Section 2.2. The chain of additional transformations is as follows:

(c) $\vec{v} = \vec{u}_{10} \times \vec{u}_{20}$

(d) $\vec{v}_0$ *is* normalized $\vec{v}$

(e) $\vec{d} = a \bullet \vec{u}_{10} + b \bullet \vec{u}_{20}$

(f) $\vec{d}_0$ *is* normalized $\vec{d}$

(g) $\vec{u}_{H1} = \cos\delta \bullet \vec{d}_0 + \sin\delta \bullet \vec{v}_0$, $\vec{u}_{H2} = \cos\delta \bullet \vec{d}_0 - \sin\delta \bullet \vec{v}_0$

In summary, this configuration requires the following parameters: A_0, A_1, A_2, d_H, a, b, δ.

2.4 1H-tetrahedral configuration

An example is the H_α atom bound to the Cα atom of non-glycine residues (Fig. 2C). The configuration is represented schematically in Fig. 4. The H atom forms a tetrahedron with A_1, A_2 and A_3. The unit vectors $\vec{u}_{10}$, $\vec{u}_{20}$ and $\vec{u}_{30}$ are defined along the directions A_0-A_1, A_0-A_2 and A_0-A_3. The vector $\vec{r}_{H0}$ forms angles α_1, α_2, α_3 with the vectors $\vec{u}_{10}$, $\vec{u}_{20}$ and $\vec{u}_{30}$, respectively. $\vec{r}_{H0}$ can be expressed as a linear combination of these vectors, $\vec{r}_{H0} = a \bullet \vec{u}_{10} + b \bullet \vec{u}_{20} + c \bullet \vec{u}_{30}$. The coefficients can be found from the conditions $\vec{r}_{H0} \bullet \vec{u}_{10} = c_1$, $\vec{r}_{H0} \bullet \vec{u}_{20} = c_2$, $\vec{r}_{H0} \bullet \vec{u}_{30} = c_3$ where $c_1 = \cos\alpha_1$, $c_2 = \cos\alpha_2$, $c_3 = \cos\alpha_3$. If ω_{12}, ω_{23} and ω_{13} are angles between respective pairs of $\vec{u}_{10}$, $\vec{u}_{20}$ and $\vec{u}_{30}$, then these conditions become a system of linear equations,

$$\begin{cases} a + \cos\omega_{12} \bullet b + \cos\omega_{13} \bullet c = c_1 \\ \cos\omega_{12} \bullet a + b + \cos\omega_{23} \bullet c = c_2 \\ \cos\omega_{13} \bullet a + \cos\omega_{23} \bullet b + c = c_3 \end{cases}$$

The system can be solved analytically (Cramer, 1750; Kosinski, 2001), yielding the parameters a, b and c that can be calculated using ideal angles α_i and ω_{ij}.

This geometry is an example where four non-H atoms are used to parameterize the H atom. In principle, A_0 and any two atoms of A_1, A_2, A_3 are sufficient to determine the H atom position. However, if the conformation of the tetrahedron is distorted, this can lead to suboptimal H atom coordinates. The use of all three neighbors is therefore preferable.

The chain of transformations is the following:

(a) $\vec{r}_{10} = \vec{r}_1 - \vec{r}_0$, $\vec{r}_{20} = \vec{r}_2 - \vec{r}_0$, $\vec{r}_{30} = \vec{r}_3 - \vec{r}_0$

(b) $\vec{u}_{10}$, $\vec{u}_{20}$, $\vec{u}_{30}$ are normalized vectors $\vec{r}_{10}$, $\vec{r}_{20}$, $\vec{r}_{30}$, respectively

(c) $\vec{r}_{H0} = a \bullet \vec{u}_{10} + b \bullet \vec{u}_{20} + c \bullet \vec{u}_{30}$

(d) $\vec{u}_{H0} = \|\vec{r}_{H0}\|^{-1}\vec{r}_{H0}$ with $\|\vec{r}_{H0}\| = \left(\vec{r}_{H0} \bullet \vec{r}_{H0}\right)^{1/2}$

In summary, this configuration requires the following parameters: A_0, A_1, A_2, A_3, d_H, a, b, c

2.5 Single H with rotational degree of freedom

An example is the Hη atom bound to the Oη atom of tyrosine (Fig. 2D), represented schematically in Fig. 5. The ideal angle between A_0-H and A_0-A_1 is α, and there is a

dihedral angle φ between the planes H-A_0-A_1 and A_0-A_1-B_1. We note $a = \cos(\pi - \alpha) = -\cos\alpha$, $b = \sin\alpha\cos\varphi$ and $c = \sin\alpha\sin\varphi$.

This configuration has one degree of freedom, the dihedral angle φ. Its value can be arbitrarily set to some a priori known value or according to the angle found in the input model, otherwise it can be optimized to minimize clashes or maximize hydrogen bonding interactions.

The chain of transformations is the following:

(a) $\vec{r}_{01} = \vec{r}_0 - \vec{r}_1$, $\vec{r}_{B1A1} = \vec{r}_{B1} - \vec{r}_1$—for this configuration we calculate vectors with respect to A_1

(b) $\vec{u}_1$ *is* normalized $\vec{r}_{01}$

(c) $\vec{v} = \vec{r}_{B1A1} - \left(\vec{r}_{B1A1} \bullet \vec{u}_1\right) \bullet \vec{u}_1$—vector in the plane A_0-A_1-B_1 and normal to $\vec{u}_1$

(d) $\vec{u}_2$ is normalized $\vec{v}$

(e) $\vec{u}_3 = \vec{u}_1 \times \vec{u}_2$

(f) $\vec{u}_{H0} = a \bullet \vec{u}_1 + b \bullet \vec{u}_2 + c \bullet \vec{u}_3$

In summary, this configuration requires the following parameters: A_0, A_1, B_1, d_H, a, b, c.

2.6 3H-propeller configuration

An example is the group of Hγ11, Hγ12 and Hγ13 atoms bound to Cγ1 of valine (Fig. 2E). In the ideal configuration, we note the hydrogen atoms H_1, H_2, H_3. Each atom forms an ideal angle α between A_0-H and A_0-A_1. A dihedral angle φ is defined between the planes H_1-A_0-A_1 and A_0-A_1-B_1. For H_2 and H_3 the dihedral angle is $\varphi + \frac{2\pi}{3}$ and $\varphi + \frac{4\pi}{3}$, respectively.

The chain of transformations and the schematic representation is similar to that for the free-rotation configuration, Section 2.5 and Fig. 5. The parameters are determined for the H-atom, H_1, that possesses the ideal dihedral angle. The parameters are identical for the other two H-atoms, with exception of the angle φ, which is $\varphi + n \bullet \frac{2\pi}{3} (n = 1, 2)$ for the second and third H-atom. Steps (a)–(e) are defined as in Section 2.5, step (f) is given by:

(f) $\vec{u}_{H1} = a \bullet \vec{u}_1 + b_1 \bullet \vec{u}_2 + c_1 \bullet \vec{u}_3$, $\vec{u}_{H2} = a \bullet \vec{u}_1 + b_2 \bullet \vec{u}_2 + c_2 \bullet \vec{u}_3$, $\vec{u}_{H3} = a \bullet \vec{u}_1 + b_3 \bullet \vec{u}_2 + c_3 \bullet \vec{u}_3$, where $a = -\cos\alpha$, $b_1 = \sin\alpha\cos\varphi$, $b_2 = \sin\alpha\cos\left(\varphi + \frac{2\pi}{3}\right)$, $b_3 = \sin\alpha\cos\left(\varphi + \frac{4\pi}{3}\right)$ and $c_1 = \sin\alpha\sin\varphi$, $c_2 = \sin\alpha\sin\left(\varphi + \frac{2\pi}{3}\right)$, $c_3 = \sin\alpha\sin\left(\varphi + \frac{4\pi}{3}\right)$.

In summary, this configuration requires the following parameters: A_0, A_1, B_1, d_H, a, b, c.

2.7 2H-planar configuration

An example is the Hε21 and Hε22 atoms bound to Nε2 of glutamine (Fig. 1F). Each of the two H-atoms forms an ideal angle α with A_0-A_1 and they are located in the plane formed by the neighboring non-H atoms. One of the H-atoms, H_1, forms the dihedral angle φ equal to 0 between the planes H_1-A_0-A_1 and A_0-A_1-B_1. For the second atom, H_2, this angle is equal to π.

To describe the chain of transformations we reuse the algorithm for the free-rotation configuration (Section 2.5). The parameters are determined for the H_1 atom with the ideal dihedral angle $\varphi = 0$. Step (f) for calculating the H atom coordinates becomes:

(f) $\vec{u}_{H1} = a \bullet \vec{u}_1 + b_1 \bullet \vec{u}_2 + c_1 \bullet \vec{u}_3$, $\vec{u}_{H2} = a \bullet \vec{u}_1 + b_2 \bullet \vec{u}_2 + c_2 \bullet \vec{u}_3$, where $a = -\cos\alpha$, $b_1 = \sin\alpha\cos\varphi$, $b_2 = \sin\alpha\cos(\varphi + \pi)$, $c_1 = \sin\alpha\sin\varphi$, $c_2 = \sin\alpha\sin(\varphi + \pi)$

If the angle φ is equal to zero, the coefficients can be simplified: $b_1 = \sin\alpha$, $b_2 = -\sin\alpha$ and $c_1 = c_2 = 0$. However, to keep this configuration versatile to allow for rotations of the H atom group around A_0-A_1, the general expression with φ is used.

In summary, this configuration requires the following parameters: A_0, A_1, B_1, d_H, a, b, c.

2.8 H atoms with a single parent

Some hydrogen atoms have a parent atom that has no further non-hydrogen neighbors covalently bound to it. Water molecules represent an example of this situation, as each H atom has only one non-H neighbor. It follows that the H atom positions cannot be expressed from connected non-H atoms, although it may be sometimes possible to parameterize the H atom using other interactions, such as hydrogen bonds and packing.

Therefore, the approach discussed in Sections 2.1–2.7, where the hydrogen position are redefined from information about covalent bonds and angles, cannot be used for H atoms that possess only a parent or a parent and only one neighbor.

This leaves two options. The first possibility is to refine such hydrogens as free (restrained) atoms leading to an increased number of parameters: at least three coordinates per hydrogen, e.g., by six for a water molecule. Another option is to introduce constraints and refine such groups as rigid using relevant *Phenix* options (Afonine, Grosse-Kunstleve, Urzhumtsev, & Adams, 2009). This increases the number of parameters by three for each rigid group regardless of the number of H atoms. In this work we focus on the parameterization of the non-water H atoms and do not discuss refinement of H atoms with a single parent here.

3. Riding H: Refinement targets and their gradients

The aim of refinement is to change the parameters of a model (such as the atomic positions, $\vec{r}_0, \vec{r}_1, \vec{r}, \vec{r}_3, \ldots, \vec{r}_N$) to optimize a function of a set of observations that may be experimental data, some statistical information, stereochemical prior knowledge, etc. The fit between model parameters and the observations is commonly expressed by a target function. The lower the value of this function, the better the fit, usually resulting in an improved

model. Explicit modeling of hydrogen positions $\vec{r}_{H_0}, \vec{r}_{H_1}, \vec{r}_{H_2}, \ldots, \vec{r}_{H_M}$, may improve the macromolecular model and reduce the value of a target expressed through coordinates of the whole set of atoms:

$$R\left(\vec{r}_0, \vec{r}_1, \vec{r}_2, \vec{r}_3, \ldots, \vec{r}_N; \vec{r}_{H_0}, \vec{r}_{H_1}, \ldots, \vec{r}_{H_M};\ \text{other parameters}\right) \rightarrow \min \tag{2}$$

For simplicity, in what follows we consider only atomic coordinates. Refining the coordinates of H atoms independently increases the number of parameters, which can compromise the optimization procedure, especially when the number of experimental observations is low. Therefore, constraints can be applied in the form of riding hydrogens. For example, considering hydrogen H_0 at position $\vec{r}_{H_0}$ in formula (2) in a coplanar configuration defined by atoms A_0, A_1 and A_2 in the positions $\vec{r}_0$, $\vec{r}_1$ and $\vec{r}_2$ (Fig. 1), and doing similarly for other H atoms, reduces the target (2) to a function of coordinates of only non-hydrogen atoms

$$\begin{aligned} R_{\text{riding}}\left(\vec{r}_0, \vec{r}_1, \vec{r}_2, \vec{r}_3, \ldots, \vec{r}_N\right) = R\Big(\vec{r}_0, \vec{r}_1, \vec{r}_2, \vec{r}_3, \ldots, \vec{r}_N; \\ \vec{r}_{H_0}\left(\vec{r}_0, \vec{r}_1, \vec{r}_2\right), \vec{r}_{H_1}(\ldots)\ldots, \vec{r}_{H_M}(\ldots)\Big) \end{aligned} \tag{3}$$

In practice, target (3) is calculated in two steps:

(a) Determine the coordinates of each hydrogen atom through the coordinates of neighboring non-H atoms using one of the algorithms described above, e.g., that in Section 2.2:

$$\begin{aligned} \vec{r}_{H_0} = \vec{R}_{coplanar}\left(\vec{r}_0, \vec{r}_1, \vec{r}_2\right) = \Big(x_{H_0}\left(\vec{r}_0, \vec{r}_1, \vec{r}_2\right); y_{H_0}\left(\vec{r}_0, \vec{r}_1, \vec{r}_2\right); \\ z_{H_0}\left(\vec{r}_0, \vec{r}_1, \vec{r}_2\right)\Big) \end{aligned} \tag{4}$$

(b) Calculate target (2) using a full set of atoms, both non-H and H. Crystallographic programs minimize the target (3) by applying gradient methods (Booth, 1947; Cauchy, 1847). To calculate the gradient of the target efficiently, a two-step procedure is used that applies the chain rule and inverts the chain of transformations described in Section 2 (Lunin & Urzhumtsev, 1985). First, the gradient of the target is calculated with respect to all coordinates, as if all atoms (H and non-H) were independent:

$$\begin{aligned} &\frac{\partial R}{\partial x_0}, \frac{\partial R}{\partial y_0}, \frac{\partial R}{\partial z_0}, \frac{\partial R}{\partial x_1}, \frac{\partial R}{\partial y_1}, \frac{\partial R}{\partial z_1}, \ldots \frac{\partial R}{\partial z_N}; \frac{\partial R}{\partial x_{H_0}}, \frac{\partial R}{\partial y_{H_0}}, \frac{\partial R}{\partial z_{H_0}}, \ldots, \frac{\partial R}{\partial x_{H_M}}, \frac{\partial R}{\partial y_{H_M}}, \\ &\frac{\partial R}{\partial z_{H_M}} \end{aligned} \tag{5}$$

Then the gradient is recalculated with respect to the independent parameters, for example, in the case of atom H_0, expressed as in Eq. (4):

$$\frac{\partial R_{\text{riding}}}{\partial x_0} = \frac{\partial R}{\partial x_0} + \frac{\partial R}{\partial x_{H_0}}\frac{\partial x_{H_0}}{\partial x_0} + \frac{\partial R}{\partial y_{H_0}}\frac{\partial y_{H_0}}{\partial x_0} + \frac{\partial R}{\partial z_{H_0}}\frac{\partial z_{H_0}}{\partial x_0} \quad (6)$$

and similarly for all other coordinates. Each H atom contributes to the gradients of the non-H atoms that are used to describe its position. If the hydrogen group contains several H atoms, for example, as described in Sections 2.6 and 2.7, the derivatives of R_{riding} contain similar contributions from all of them:

$$\frac{\partial R_{\text{riding}}}{\partial x_0} = \frac{\partial R}{\partial x_0} + \frac{\partial R}{\partial x_{H1}}\frac{\partial x_{H1}}{\partial x_0} + \frac{\partial R}{\partial y_{H1}}\frac{\partial y_{H1}}{\partial x_0} + \frac{\partial R}{\partial z_{H1}}\frac{\partial z_{H1}}{\partial x_0} + \frac{\partial R}{\partial x_{H2}}\frac{\partial x_{H2}}{\partial x_0} + \frac{\partial R}{\partial y_{H2}}\frac{\partial y_{H2}}{\partial x_0} + \frac{\partial R}{\partial z_{H2}}\frac{\partial z_{H2}}{\partial x_0} + \ldots \quad (7)$$

Also, a non-H atom may be used to parameterize several hydrogen groups; then the gradient with respect to its coordinates is a sum of contributions from all these groups. The coordinates of the non-H atoms are refined; at each refinement iteration, the H atom position is updated from these new values of x_0, y_0, z_0, etc.

To simplify the calculation, approximations such as $\frac{\partial x_H}{\partial x_0} = \frac{\partial y_H}{\partial x_0} = \frac{\partial z_H}{\partial x_0} = 1$ and $\frac{\partial x_H}{\partial x_1} = \frac{\partial y_H}{\partial x_1} = \frac{\partial z_H}{\partial x_1} = 0$, can be made (Bourhis, Dolomanov, Gildea, Howard, & Puschmann, 2015). In our work, such approximations are not used and actually are not required. The transformation of the gradients based on the different H atom configurations is summarized in the Appendix.

4. Constructing the riding H model in *CCTBX*

4.1 Internal data structure for model and connectivity information

The *CCTBX* contains several entities describing an atomic model. One is the *hierarchy* (iotbx.pdb.hierarchy), which is a nested data structure describing the macromolecule at the model, chain, residue and atom level. The *hierarchy* is a container of atomic model information that makes it possible to access the atomic attributes of a model from files (such as mmCIF; Adams et al., 2019), such as chain identifiers, residue names, atom labels, coordinates and so on (Grosse-Kunstleve & Adams, 2010). While the *hierarchy* contains as much information as is available in the corresponding model file, it lacks any notion of atomic connectivity and covalent geometry. This information is available in another representation of the atomic model: the *cctbx.geometry_restraints* class (Grosse-Kunstleve, Afonine, & Adams, 2004). An object of this class is built from the *hierarchy* and from the information available in stereochemical libraries such as the Monomer Library (Engh & Huber, 1991; Vagin et al., 2004) or GeoStd (Moriarty & Adams, n.d., http://sourceforge.net/projects/geostd). The *cctbx.geometry_restraints* class is aware of atomic connectivity, it stores ideal and actual values of geometric parameters of the model (such as bonds, angles, torsions, planes, chiral volumes, and non-bonded interactions) and calculates functions to support refinement, such as the restraints target function and its derivatives.

Together, the coordinates of non-hydrogen atoms as well as the *hierarchy* and *cctbx.geometry_restraints* objects can be used to describe the configuration of hydrogen atoms in the model and calculate their riding model parameters (Sections 2.2–2.8) using ideal bond lengths and angles.

4.2 Building riding H from known connectivity

The riding H algorithm operates in two stages. First, the procedure determines the connectivity of H atoms in the input model. This means that for each H atom the information about atoms linked to each other (as well as corresponding bond lengths and angles) is accumulated for all neighbors defining its position. In the second step, this information is used to parameterize each H atom. This involves computing the coefficients described in Section 2 and checking if the input geometry is consistent with the assigned configuration. For example, if one H atom is missing in a propeller group, the information in the connectivity will be the same as for a 2H-planar group (such as in the ARG head group). Without checks, the riding H procedure would assume that the two H atoms are in a 2H-planar configuration. Such instances are therefore not parameterized as riding H because it is difficult to determine the actual configuration in this case.

When multiple conformations are present, determining the connectivity becomes complex. An H atom may then have several parents and neighbors, leading to a complicated network of connections. An example is shown in Fig. 6; residue i (right) adopts a double conformation and residue i + 1 (left) has a single conformation. To maintain planarity of the peptide unit that connects the two residues, the amide H atoms are also in a double conformation while the amide N atom is in a single conformation. For example, the H atom in conformation A has the parent atom N, which in turn is connected to the H atom in conformation B, the C atoms of both conformations A and B, and atom C_α in residue i + 1. To parameterize this H atom in a coplanar configuration, the correct subset of connections is needed. In this example, the atoms in conformation B are irrelevant for the atom H in conformation A. The assignment of parent and neighbor atoms in the presence of multiple conformations is achieved by exploiting the *hierarchy* object, which interprets the atomic model in terms of alternates (Grosse-Kunstleve & Adams, 2010).

The algorithms that build the connectivity of H atoms and determine the parameterization are written in Python. This step typically takes 10% of the computing time necessary to generate the *hierarchy* and *cctbx.geometry_restraints* objects. The algorithms that rebuild H atoms and calculate the gradients are written in C++. They are typically an order of magnitude faster than calculating the target function (less than 0.1 s in most cases).

5. Summary

This work outlines future directions of the synergetic use of neutron and X-ray diffraction data and describes the implementation of the riding hydrogen model as a fundamental prerequisite for it. We argue that the traditional joint X-ray and neutron refinement approach that yields one atomic model meant to fit both datasets is inadequate. Instead, we suggest refining two atomic models, each one fitting the respective dataset, while also benefitting from the expectation that these models should be similar.

Riding H atoms can be found in several different configurations that are not specific to protein residues. For each configuration, the H atom position is parameterized using coordinates from a set of non-H neighbors and ideal bond lengths and angles. We successfully implemented the riding hydrogen model for typical H atom geometries. The parameterization and the expressions for the gradients of refinement targets with respect to the independent parameters are described for each configuration.

The riding H option is available in *CCTBX* and *Phenix* since release 1.15.

Acknowledgments

This work was supported by the US National Institutes of Health (NIH) (grant P01GM063210), the Phenix Industrial Consortium and the NIH–funded (grant R01GM071939) Macromolecular Neutron Consortium between Oak Ridge National Laboratory and Lawrence Berkeley National Laboratory. This work was supported in part by the US Department of Energy under Contract No. DE-AC02-05CH1123. A.U. acknowledges the support and the use of resources of the French Infrastructure for Integrated Structural Biology FRISBI ANR-10-INBS-05 and of Instruct-ERIC.

Appendix.: Gradient calculation for riding-H atoms

A.1 Chain rule and common steps

A.1.1 Gradient components

Each riding H group consists of one or several H atoms at the positions $\vec{r}_{Hm} = (x_{Hm}, y_{Hm}, z_{Hm})$, $m = 0, 1, \ldots, M$, and several non-hydrogen atoms at the positions $\vec{r}_k = (x_k, y_k, z_k)$, $k = 0, 1, \ldots, K$ including the 'parent' atom with $k = 0$ to which the hydrogen atoms of the group are bound. We suppose that we know the components of the gradient of the target R with respect to all atoms:

$$\vec{\nabla}_{Hm} R = \left(\frac{\partial R}{\partial x_{Hm}}, \frac{\partial R}{\partial y_{Hm}}, \frac{\partial R}{\partial z_{Hm}}\right); \vec{\nabla}_{rk} R = \left(\frac{\partial R}{\partial x_k}, \frac{\partial R}{\partial y_k}, \frac{\partial R}{\partial z_k}\right)$$

The gradient $\vec{\nabla}_{rk} R_{riding}$ of the target with respect to independent parameters (coordinates of non-hydrogen atoms) is a sum of $\vec{\nabla}_{rk} R$ with the contributions $\vec{\nabla}_{rkH} R$, recalculated from $\vec{\nabla}_{Hm} R$, possibly from several H atoms and several H groups.

Calculating these contributions $\vec{\nabla}_{rkH} R$ requires inversion of the formulae used to express the riding H atoms. Below we present calculation steps for each H atom configuration. There are some common expressions, so they are elaborated once and later only referred to.

A.1.2 Riding hydrogen position

The last step for all configurations consists in calculating the hydrogen position at the distance d_H along the unit vector $\vec{u}_{H0} = (x_{uH0}, y_{uH0}, z_{uH0})$ that starts from the parent non-H atom A_0 with coordinates $\vec{r}_0 = (x_0, y_0, z_0)$:

$$\vec{r}_H = \vec{r}_0 + d_H \vec{u}_{H0}$$

For the gradient calculation, we start from the gradient components $\vec{\nabla}_H R$ supposed to be known. Then

$$\vec{\nabla}_{uH0} R = d_H \vec{\nabla}_H R$$

$$\vec{\nabla}_{r0H} R = \vec{\nabla}_H R$$

In the second expression, the subscript *r0H* is used to indicate the contribution to the coordinates of the 'parent' atom A_0 from the riding H; we shall distinguish this from the contribution $\vec{\nabla}_{r0} R$ calculated directly from the target as explained above.

A.1.3 Vector normalization

Normalization of a vector $\vec{r} = (x, y, z)$ gives a new vector

$$\vec{u} = (x_u(x,y,z), y_u(x,y,z), z_u(x,y,z)) = \|.\|^{-1}(x,y,z)$$

where $\|.\| = \left(x^2 + y^2 + z^2\right)^{1/2}$. Let $\vec{\nabla}_u R = \left(\frac{\partial R}{\partial x_u}, \frac{\partial R}{\partial y_u}, \frac{\partial R}{\partial z_u}\right)$ be known. Then the gradient $\vec{\nabla}_r R$ with respect to (*x, y, z*) has the *x*-component

$$\frac{\partial R}{\partial x} = \frac{\partial R}{\partial x_u}\frac{\partial x_u}{\partial x} + \frac{\partial R}{\partial y_u}\frac{\partial y_u}{\partial x} + \frac{\partial R}{\partial z_u}\frac{\partial y_u}{\partial x} = \|.\|^{-1}\frac{\partial R}{\partial x_u} - x\|.\|^{-3}\left(\vec{r} \bullet \vec{\nabla}_u R\right)$$

and similar expressions for $\frac{\partial R}{\partial y}$ and $\frac{\partial R}{\partial z}$. Or in vector form:

$$\vec{\nabla}_r R = \|.\|^{-1}\vec{\nabla}_u R - \|.\|^{-3}\left(\vec{r} \bullet \vec{\nabla}_u R\right)\vec{r}$$

A.1.4 Cross product

The vector product of $\vec{r}_1 = (x_1, y_1, z_1)$ and $\vec{r}_2 = (x_2, y_2, z_2)$ gives

$$\vec{v} = \vec{r}_1 \times \vec{r}_2 = (x_v, y_v, z_v) = (y_1 z_2 - z_1 y_2, z_1 x_2 - x_1 z_2, x_1 y_2 - y_1 x_2)$$

Let $\vec{\nabla}_v R = \left(\frac{\partial R}{\partial x_v}, \frac{\partial R}{\partial y_v}, \frac{\partial R}{\partial z_v}\right)$ be known. Then

$$\vec{\nabla}_{r1} R = \vec{r}_2 \times \vec{\nabla}_v R,$$

$$\overrightarrow{\nabla}_{r2}R = \overrightarrow{\nabla}_{v}R \times \overrightarrow{r}_1$$

A.2 Gradient for particular riding H configurations

The next sections explain how to calculate the gradient of the target function with respect to independent coordinates, starting from the values given in Section A.1.1. For each configuration, we explicitly list the steps to which we apply the chain rule. Generally, we invert the steps described for each geometry, starting from the last and moving backward.

A.2.1 Coplanar configuration

For definition of parameters, see Section 2.2.

We start from the last step (d) and move backward to (a) after getting $\overrightarrow{\nabla}_{uH0}R$ and $\overrightarrow{\nabla}_{r0H}R$ from $\overrightarrow{\nabla}_{H}R$ according to Section A.1.2.

(d) obtain $\overrightarrow{\nabla}_{rH0}R$ from $\overrightarrow{\nabla}_{uH0}R$ according to Section A.1.3

(c) $\overrightarrow{\nabla}_{u10}R = a\overrightarrow{\nabla}_{rH0}R$, $\overrightarrow{\nabla}_{u20}R = b\overrightarrow{\nabla}_{rH0}R$

(b) obtain $\overrightarrow{\nabla}_{r10}R$ from $\overrightarrow{\nabla}_{u10}R$ and $\overrightarrow{\nabla}_{r20}R$ from $\overrightarrow{\nabla}_{u20}R$ according to Section A.1.3

(a) $\overrightarrow{\nabla}_{r1}R_{riding} = \overrightarrow{\nabla}_{r10}R$, $\overrightarrow{\nabla}_{r2}R_{riding} = \overrightarrow{\nabla}_{r20}R\ \overrightarrow{\nabla}_{r0}R_{riding} = \overrightarrow{\nabla}_{r0H}R - \overrightarrow{\nabla}_{r10}R - \overrightarrow{\nabla}_{r20}R$

A.2.2 2H-tetrahedral configuration

For definition of parameters, see Section 2.3.

We start from getting $\overrightarrow{\nabla}_{uH0}R$ and $\overrightarrow{\nabla}_{r0H}R$ from $\overrightarrow{\nabla}_{H}R$ according to Section A.1.2.

(g) $\overrightarrow{\nabla}_{d0}R = (\cos\delta)\overrightarrow{\nabla}_{d0}R$, $\overrightarrow{\nabla}_{v0}R = (\sin\delta)\overrightarrow{\nabla}_{d0}R$

(f) $\overrightarrow{\nabla}_{d}R$ from $\overrightarrow{\nabla}_{d0}R$ according to Section A.1.3

(e) $\overrightarrow{\nabla}^{(1)}_{u10}R = a\overrightarrow{\nabla}_{d}R$ and $\overrightarrow{\nabla}^{(1)}_{u20}R = b\overrightarrow{\nabla}_{d}R$

(d) obtain $\overrightarrow{\nabla}_{v}R$ from $\overrightarrow{\nabla}_{v0}R$ according to Section A.1.3

(c) obtain $\overrightarrow{\nabla}^{(2)}_{u10}R$ and $\overrightarrow{\nabla}^{(2)}_{u20}$ from $\overrightarrow{\nabla}_{v}R$ according to Section A.1.4 and then calculate $\overrightarrow{\nabla}_{u10}R = \overrightarrow{\nabla}^{(1)}_{u10}R + \overrightarrow{\nabla}^{(2)}_{u10}R$ and $\overrightarrow{\nabla}_{u20}R = \overrightarrow{\nabla}^{(1)}_{u20}R + \overrightarrow{\nabla}^{(2)}_{u20}R$

(b) obtain $\overrightarrow{\nabla}_{r10}R$ from $\overrightarrow{\nabla}_{u10}R$ and $\overrightarrow{\nabla}_{r20}R$ from $\overrightarrow{\nabla}_{u20}R$ according to Section A.1.3

(a) $\vec{\nabla}_{r1}R_{riding} = \vec{\nabla}_{r10}R$, $\vec{\nabla}_{r2}R_{riding} = \vec{\nabla}_{r20}R\,\vec{\nabla}_{r0}R_{riding} = \vec{\nabla}_{r0H}R - \vec{\nabla}_{r10}R - \vec{\nabla}_{r20}R$

A.2.3 1H-tetrahedral configuration

For definition of parameters, see Section 2.4.

We start from getting $\vec{\nabla}_{uH0}R$ and $\vec{\nabla}_{r0H}R$ from $\vec{\nabla}_{H}R$ according to Section A.1.2.

(d) obtain $\vec{\nabla}_{rH0}R$ from $\vec{\nabla}_{uH0}R$ according to Section A.1.3

(c) $\vec{\nabla}_{u10}R = a\,\vec{\nabla}_{rH0}R$, $\vec{\nabla}_{u20}R = b\,\vec{\nabla}_{rH0}R$, $\vec{\nabla}_{u30}R = c\,\vec{\nabla}_{rH0}R$

(b) obtain $\vec{\nabla}_{r10}R$ from $\vec{\nabla}_{u10}R$, $\vec{\nabla}_{r20}R$, from $\vec{\nabla}_{u20}R$, $\vec{\nabla}_{r30}R$ from $\vec{\nabla}_{u30}R$ according to Section A.1.3

(a) $\vec{\nabla}_{r1}R_{riding} = \vec{\nabla}_{r10}R$, $\vec{\nabla}_{r2}R_{riding} = \vec{\nabla}_{r20}R$, $\vec{\nabla}_{r3}R_{riding} = \vec{\nabla}_{r30}R$, $\vec{\nabla}_{r0}R_{riding} = \vec{\nabla}_{r0H}R - \vec{\nabla}_{r10}R - \vec{\nabla}_{r20}R - \vec{\nabla}_{r30}R$

A.2.4 Single H with rotational degree of freedom

For definition of parameters, see Section 2.5.

We start from getting $\vec{\nabla}_{uH0}R$ and $\vec{\nabla}_{r0H}R$ from $\vec{\nabla}_{H}R$ according to Section A.1.2.

(f) $\vec{\nabla}^{(1)}_{u10}R = a\,\vec{\nabla}_{rH0}R$, $\vec{\nabla}^{(1)}_{u20}R = b\,\vec{\nabla}_{rH0}R$, $\vec{\nabla}_{u30}R = c\,\vec{\nabla}_{rH0}R$

(e) obtain $\vec{\nabla}^{(2)}_{u10}R$ and $\vec{\nabla}^{(2)}_{u20}R$ from $\vec{\nabla}_{u30}R$ according to Section A.1.4; superscript (2) is used to distinguish these derivatives from those calculated above $\vec{\nabla}^{(1)}_{u20}R$ and $\vec{\nabla}^{(1)}_{u10}R$; $\vec{\nabla}_{u2}R = \vec{\nabla}^{(1)}_{u20}R + \vec{\nabla}^{(2)}_{u20}R$;

(d) obtain $\vec{\nabla}_{v}R$ from $\vec{\nabla}_{u2}R$ according to Section A.1.2

(c) $\vec{\nabla}_{rBA}R = \vec{\nabla}_{v}R - \vec{u}_1\left(\vec{u}_1 \bullet \vec{\nabla}_{v}R\right)$; $\vec{\nabla}_{u1v}R = -\vec{\nabla}_{v}R\left(\vec{r}_{B1A1} \bullet \vec{u}_1\right) - \vec{r}_{B1A1}\left(\vec{u}_1 \bullet \vec{\nabla}_{v}R\right)$; $\vec{\nabla}_{u1}R = \vec{\nabla}^{(1)}_{u10}R + \vec{\nabla}^{(2)}_{u10}R + \vec{\nabla}_{u1v}R$

(b) obtain $\vec{\nabla}_{r01}R$ from $\vec{\nabla}_{u1}R$ according to Section A.1.2

(a) $\vec{\nabla}_{r1}R_{riding} = -\vec{\nabla}_{r01}R - \vec{\nabla}_{rBA}R$, $\vec{\nabla}_{rB1}R_{riding} = \vec{\nabla}_{rB1}R$, $\vec{\nabla}_{r0}R_{riding} = \vec{\nabla}_{r0H}R + \vec{\nabla}_{r01}R$

A.2.5 3H-propeller

For definition of parameters, see Section 2.6.

We start from getting $\overrightarrow{\nabla}_{uH1}R$, $\overrightarrow{\nabla}_{r0H1}R$, $\overrightarrow{\nabla}_{uH2}R$, $\overrightarrow{\nabla}_{r0H2}R$ and $\overrightarrow{\nabla}_{uH3}R$, $\overrightarrow{\nabla}_{r0H3}R$ from $\overrightarrow{\nabla}_{H1}R$, $\overrightarrow{\nabla}_{H2}R$, $\overrightarrow{\nabla}_{H3}R$, respectively, according to Section A.1.2. Then

(f) $\overrightarrow{\nabla}_{u10}R = a\sum_{m=1}^{3}\overrightarrow{\nabla}_{r0Hm}R$, $\overrightarrow{\nabla}_{u20}R = \sum_{m=1}^{3}b_m\overrightarrow{\nabla}_{r0Hm}R$, $\overrightarrow{\nabla}_{u30}R = \sum_{m=1}^{3}c_m\overrightarrow{\nabla}_{r0Hm}R$

Steps (e)–(a) repeat the steps from Section A.2.4 except the last expression for $\overrightarrow{\nabla}_{r0}R_{riding}$ that becomes

$$\overrightarrow{\nabla}_{r0}R_{riding} = \overrightarrow{\nabla}_{r01}R + \sum_{m=1}^{3}\overrightarrow{\nabla}_{r0Hm}R$$

A.2.6 2H-planar

For definition of parameters, see Section 2.7.

We start from getting $\overrightarrow{\nabla}_{uH1}R$, $\overrightarrow{\nabla}_{r0H1}R$ and $\overrightarrow{\nabla}_{uH2}R$, $\overrightarrow{\nabla}_{r0H2}R$ from $\overrightarrow{\nabla}_{H1}R$ and $\overrightarrow{\nabla}_{H2}R$, respectively, according to Section A.1.2. Then

(f) $\overrightarrow{\nabla}_{u10}R = a\left(\overrightarrow{\nabla}_{r0H1}R + \overrightarrow{\nabla}_{r0H2}R\right)$, $\overrightarrow{\nabla}_{u20}R = b\left(\overrightarrow{\nabla}_{r0H1}R + \overrightarrow{\nabla}_{r0H2}R\right)$, $\overrightarrow{\nabla}_{u30}R = c\left(\overrightarrow{\nabla}_{r0H1}R + \overrightarrow{\nabla}_{r0H2}R\right)$

Steps (e)–(a) repeat the steps from Section A.2.4 except the last expression for $\overrightarrow{\nabla}_{r0}R_{riding}$ that becomes

$$\overrightarrow{\nabla}_{r0}R_{riding} = \overrightarrow{\nabla}_{r01}R + \overrightarrow{\nabla}_{r0H1}R + \overrightarrow{\nabla}_{r0H2}R$$

References

Adams PD, Afonine PV, Baskaran K, Berman HM, Berrisford J, Bricogne G, et al. (2019). Announcing mandatory submission of PDBx/mmCIF format files for crystallographic depositions to the Protein Data Bank (PDB). Acta Crystallographica Section D: Structural Biology, 75, 451–454. [PubMed: 30988261]

Adams PD, Mustyakimov M, Afonine PV, & Langan P (2009). Generalized X-ray and neutron crystallographic analysis: More accurate and complete structures for biological macromolecules. Acta Crystallographica Section D: Structural Biology, 65, 567–573.

Afonine PV, Grosse-Kunstleve RW, Urzhumtsev A, & Adams PD (2009). Automatic multiple-zone rigid-body refinement with a large convergence radius. Journal of Applied Crystallography, 42, 607–615. [PubMed: 19649324]

Afonine PV, Mustyakimov M, Grosse-Kunstleve RW, Moriarty NW, Langan P, & Adams PD (2010). Joint X-ray and neutron refinement with phenix.refine. Acta Crystallographica Section D: Structural Biology, 66, 1153–1163.

Allen FH (1986). A systematic pairwise comparison of geometric parameters obtained by X-ray and neutron diffraction. Acta Crystallographica Section B: Structural Science, Crystal Engineering and Materials, 42, 515–522.

Allen FH, & Bruno IJ (2010). Bond lengths in organic and metal-organic compounds revisited: X-H bond lengths from neutron diffraction data. Acta Crystallographica Section B: Structural Science, Crystal Engineering and Materials, 66, 380–386.

Booth AD (1947). Application of the method of steepest descents to X-ray structure analysis. Nature, 160, 196.

Bourhis LJ, Dolomanov OV, Gildea RJ, Howard JAK, & Puschmann H (2015). The anatomy of a comprehensive constrained, restrained refinement program for the modern computing environment - Olex2 dissected. Acta Crystallographica Section A: Foundations of Crystallography, 71, 59–75.

Busing WR, & Levy HA (1964). The effect of thermal motion on the estimation of bond lengths from diffraction measurements. Acta Crystallographica, 17, 142–146.

Cauchy AL (1847). Méthode générale pour la résolution des systèmes d'équations simultanées. Comptes Rendus de l'Academie des Sciences, 25, 536–538.

Chen JC-H, & Unkefer CJ (2017). Fifteen years of the protein crystallography station: The coming of age of macromolecular neutron crystallography. IUCrJ, 4, 72–86.

Coates L, Cuneo MJ, Frost MJ, He J, Weiss KL, Tomanicek SJ, et al. (2015). The macromolecular neutron diffractometer MaNDi at the spallation neutron source. Journal of Applied Crystallography, 48, 1302–1306.

Coppens P (1967). Comparative X-ray and neutron diffraction study of bonding effects in s-Triazine. Science, 158, 1577–1579. [PubMed: 17816628]

Cramer G (1750). Introduction à l'Analyse des lignes Courbes algébriques. Geneva: Europeana.

Dauter M, & Dauter Z (2017). Many ways to derivatize macromolecules and their crystals for phasing. Methods in Molecular Biology, 1607, 349–356. [PubMed: 28573580]

Engh RA, & Huber R (1991). Accurate bond and angle parameters for X-ray protein structure refinement. Acta Crystallographica Section A: Foundations and Advances, 47, 392–400.

Fraser JS, van den Bedem H, Samelson AJ, Lang PT, Holton JM, Echols N, et al. (2011). Accessing protein conformational ensembles using room-temperature X-ray crystallography. Proceedings of the National Academy of Sciences of the United States of America, 108, 16247–16252. [PubMed: 21918110]

Grosse-Kunstleve RW, & Adams PD (2010). cctbx PDB handling tools. Computational Crystallography Newsletter, 1, 4–11.

Grosse-Kunstleve RW, Afonine PV, & Adams PD (2004). cctbx news. IUCr Computer Communication Newsletter, 4, 19–36.

Grosse-Kunstleve RW, Sauter NK, Moriarty NW, & Adams PD (2002). The computational crystallography toolbox: Crystallographic algorithms in a reusable software framework. Journal of Applied Crystallography, 35, 126–136.

Gruene T, Hahn HW, Luebben AV, Meilleur F, & Sheldrick GM (2014). Refinement of macromolecular structures against neutron data with SHELXL2013. Journal of Applied Crystallography, 47, 462–466. [PubMed: 24587788]

Kosinski AA (2001). Cramer's rule is due to Cramer. Mathematics Magazine, 74, 310–312.

Liebschner D, Afonine PV, Baker ML, Bunkóczi G, Chen VB, Croll TI, et al. (2019). Macromolecular structure determination using X-rays, neutrons and electrons: Recent developments in Phenix. Acta Crystallographica Section D: Structural Biology, 75, 861–877. [PubMed: 31588918]

Liebschner D, Afonine PV, Moriarty NW, Langan P, & Adams PD (2018). Evaluation of models determined by neutron diffraction and proposed improvements to their validation and deposition. Acta Crystallographica Section D: Structural Biology, 74, 800–813. [PubMed: 30082516]

Lunin VY, & Urzhumtsev AG (1985). Program construction for macromolecule atomic model refinement based on the fast Fourier transform and fast differentiation algorithms. Acta Crystallographica Section A: Foundations of Crystallography, 41, 327–333.

Moriarty NW & Adams PD n.d. GeoStd https://sourceforge.net/projects/geostd/

Ostermann A, Tanaka I, Engler N, Niimura N, & Parak FG (2002). Hydrogen and deuterium in myoglobin as seen by a neutron structure determination at 1.5 Å resolution. Biophysical Chemistry, 95, 183–193. [PubMed: 12062378]

Petrova T, & Podjarny A (2004). Protein crystallography at subatomic resolution. Reports on Progress in Physics, 67, 1565–1605.

Sheldrick GM, & Schneider TR (1997). SHELXL: High-resolution refinement. Methods in Enzymology, 277, 319–343. [PubMed: 18488315]

Shu F, Ramakrishnan V, & Schoenborn BP (2000). Enhanced visibility of hydrogen atoms by neutron crystallography on fully deuterated myoglobin. Proceedings of the National Academy of Sciences of the United States of America, 97, 3872–3877. [PubMed: 10725379]

Turkenburg JP, & McAuley KE (2013). Data collection and processing. Acta Crystallographica Section D: Biological Crystallography, 69, 1193–1194. [PubMed: 23793144]

Vagin AA, Steiner RA, Lebedev AA, Potterton L, McNicholas S, Long F, et al. (2004). REFMAC5 dictionary: Organization of prior chemical knowledge and guidelines for its use. Acta Crystallographica Section D: Biological Crystallography, 60, 2184–2195. [PubMed: 15572771]

Wlodawer A (1980). Studies of ribonuclease-A by X-ray and neutron diffraction. Acta Crystallographica Section B: Structural Science, Crystal Engineering and Materials, 36, 1826–1831.

Wlodawer A, & Hendrickson WA (1982). A procedure for joint refinement of macromolecular structures with X-ray and neutron diffraction data from single crystals. Acta Crystallographica Section A: Foundations and Advances, 38, 239–247.

Wlodawer A, Miller M, & Sjölin L (1983). Active site of RNase: Neutron diffraction study of a complex with uridine vanadate, a transition-state analog. Proceedings of the National Academy of Sciences of the United States of America, 80, 3628–3631. [PubMed: 6574501]

Further reading

Blakeley MP, Ruiz F, Cachau R, Hazemann I, Meilleur F, Mitschler A, et al. (2008). Quantum model of catalysis based on a mobile proton revealed by subatomic x-ray and neutron diffraction studies of h-aldose reductase. Proceedings of the National Academy of Sciences of the United States of America, 105, 1844–1848. [PubMed: 18250329]

Fisher SJ, Wilkinson J, Henchman RH, & Helliwell JR (2009). An evaluation review of the prediction of protonation states in proteins versus crystallographic experiment. Crystallography Reviews, 15, 231–259.

Howard EI, Blakeley MP, Haertlein M, Haertlein IP, Mitschler A, Fisher SJ, et al. (2011). Neutron structure of type-III antifreeze protein allows the reconstruction of AFP–ice interface. Journal of Molecular Recognition, 24, 724–732. [PubMed: 21472814]

Howard EI, Guillot B, Blakeley MP, Haertlein M, Moulin M, Mitschler A, et al. (2016). High-resolution neutron and X-ray diffraction room-temperature studies of an H-FABP–oleic acid complex: Study of the internal water cluster and ligand binding by a transferred multipolar electron-density distribution. IUCrJ, 3, 115–126.

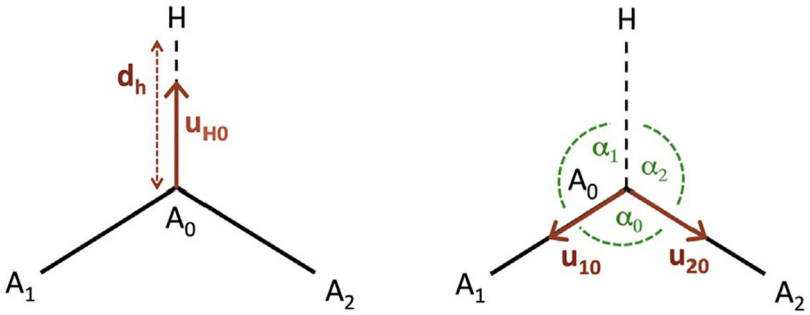

Fig. 1.

Example of a hydrogen atom H in a coplanar configuration. To improve visibility the parameters are shown in two figures. H is covalently bound to its parent atom A_0, the neighbors are A_1 and A_2.

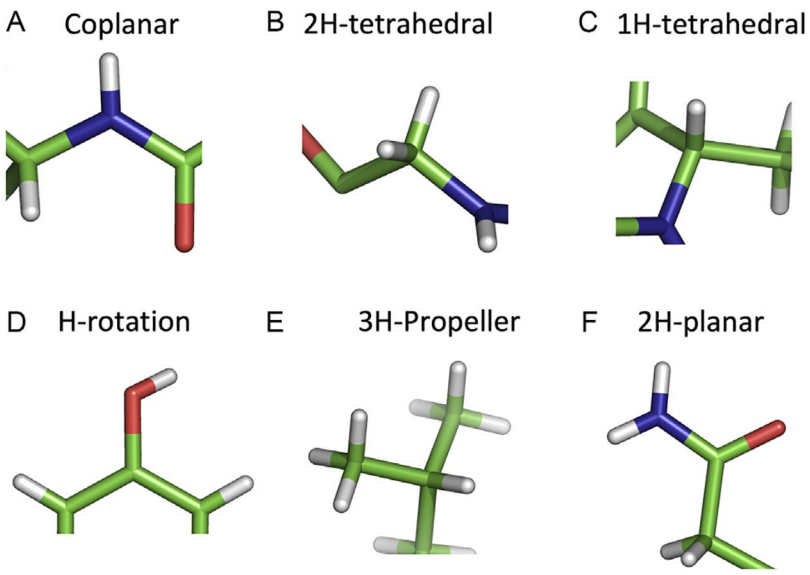

Fig. 2.

Examples of hydrogen atom configurations. While the examples shown in the figure refer to amino acid residues, the implemented riding-H algorithm is general and not specific to proteins only.

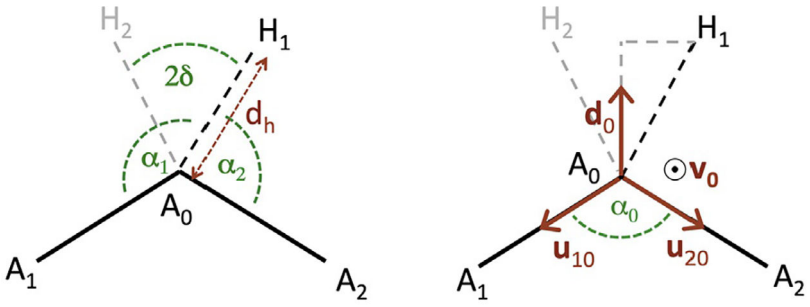

Fig. 3.

H-atoms in 2H-tetrahedral configuration. To improve visibility the parameters are shown on two figures. In the ideal configurations, the positions of H_1 and H_2 are symmetric with respect to the plane A_1-A_0-A_2. The vector d_0 is bisector of the angle H_1-A_0-H_2 and is in the plane A_1-A_0-A_2.

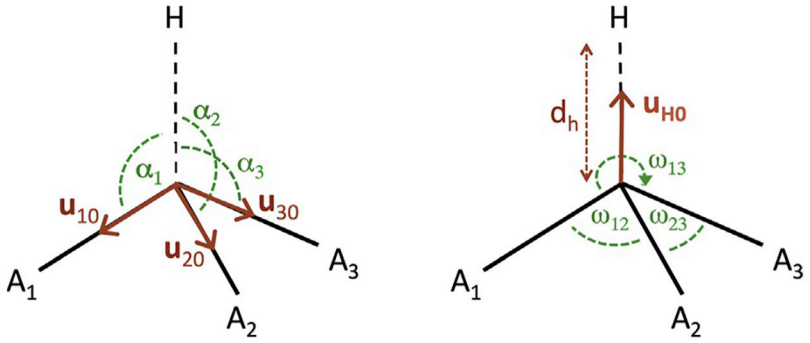

Fig. 4.

H atom in a 1H-tetrahedral position. The vector A_0-H makes angles α_1, α_2, α_3 with vectors A_0-A_1, A_0-A_2, A_0-A_3, which form angles ω_{12}, ω_{13}, ω_{23} between themselves.

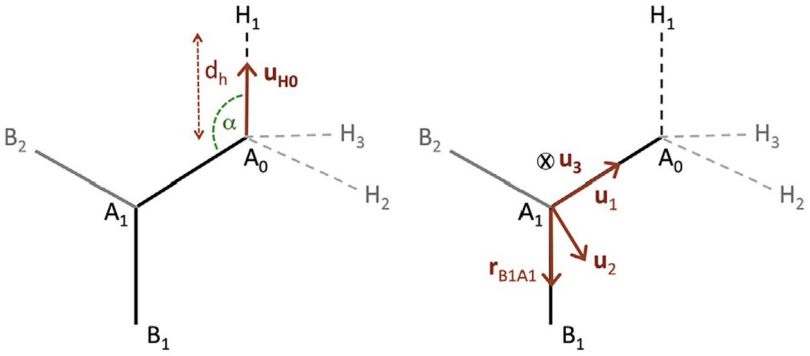

Fig. 5.

H-atom in the free-rotation configuration. Alternative positions are indicated with H_2 and H_3. Only one atom among B_1 and B_2 defines a dihedral angle with A_1-A_0-H_1 (in this example: B_1) and is therefore used to parameterize the H atom.

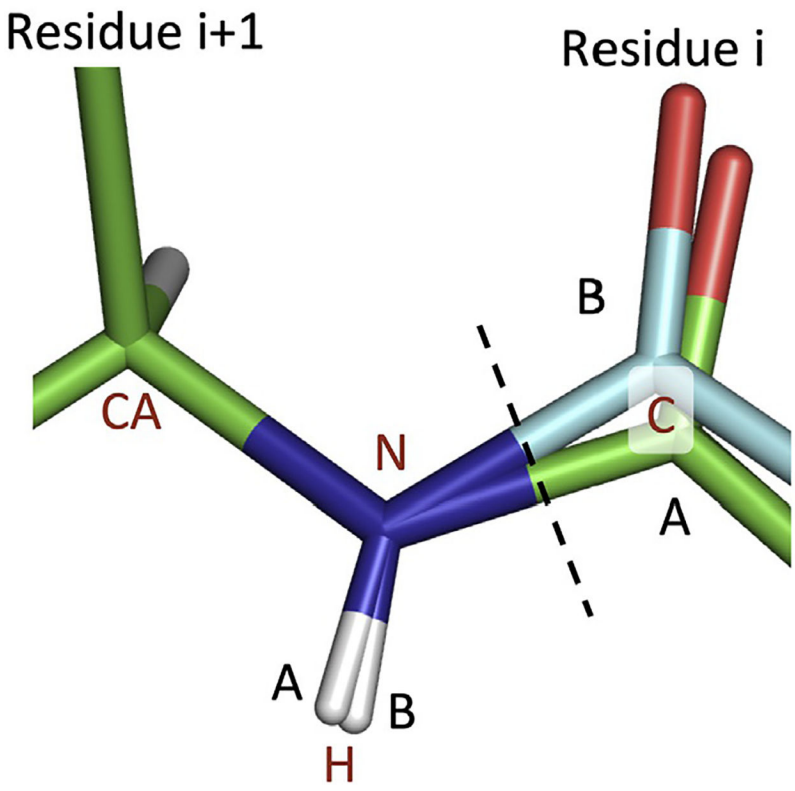

Fig. 6.

Illustration of connectivity when the H atom adopts a double conformation. Atom name labels are in orange, alternative conformation labels are in black.