Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

An introduction to stereochemical restraints

Philip R. Evans

MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, England

Correspondence e-mail: pre@mrc-lmb.cam.ac.uk

Received 7 April 2006
Accepted 1 November 2006

At the resolution available from most macromolecular crystals, the X-ray data alone are insufficient to lead to a chemically reasonable structure, so stereochemical restraints are essential. These usually restrain bond lengths, bond angles, planes and chiral volumes. The definition of these restraints and where the values come from are described. A dictionary entry contains information about the atom types, their connectivity and all the appropriate restraints. Torsion angles are not usually restrained, but they do have optimum values. In the special case of flexible five- and six-membered rings, including pentose and hexose sugars, the ring pucker is defined by combinations of torsion angles and the pucker affects the position of substituents.

1. Introduction

One of the most confusing aspects of the refinement of macromolecules for novices (and frequently also for experienced crystallographers) is the use of stereochemical restraints. For proteins, and to a large extent for polynucleotides, the libraries distributed with refinement programs are usually adequate, but for small-molecule ligands it often falls to the user to construct a suitable 'dictionary' file describing the stereochemistry of the ligand. This paper describes briefly the main items in such dictionaries, both for components of macromolecules and for small ligands, and notes a few considerations which should be borne in mind. No attempt is made to give full references, since most of this is textbook chemistry, nor are the various tools for creating dictionaries described: for this, see Kleywegt (2006).

Stereochemical restraints are needed for refinement of macromolecules because the resolution is usually insufficient to define the positions of individual atoms with sufficient precision, with typically around 1.2–5 observations per non-H atom (for resolutions around 3–2 Å), rather than the ∼80 observations per atom at 0.8 Å resolution available for small molecules. In the absence of restraints, refinement would lead to a very distorted model. Even with atomic resolution data, as for small molecules, there may be disordered regions where stereochemical restraints are essential to give a sensible model. We know a great deal about the stereochemistry of organic molecules and this information may be considered as prior knowledge in the refinement process.

2. Types of stereochemical restraint and their uses

Stereochemical restraints may be used in refinement by adding what are essentially additional observations to the penalty function, typically as a quadratic penalty,

$$\text{Penalty} = \sum \text{weight}(\text{ObservedValue} - \text{IdealValue})^2$$

where weight = $1/\sigma(\text{Value})^2$ and the total penalty to be minimized is summed over all restraint pseudo-observations. The types of restraint used are: bond lengths, bond angles, planes, chirality, torsion angles, nonbonded interactions ('bumps'), noncrystallographic symmetry and *B* factors. The last three of these are not discussed here as they are generally set globally for all residue types.

In order to use these restraints, we need to know the ideal or target values for each parameter and an estimate of the likely error (σ) to provide the weight, *i.e.* how much penalty to give to a particular deviation from the target. The main primary source for values of bond lengths and angles is from accurate crystal structures of small molecules, most conveniently collected and distributed (for money) by the Cambridge Crystallographic Data Centre (CCDC; Allen, 2002). Some of the various computational tools listed by Kleywegt (2006) use data abstracted from the original structures by recognition of atom types, bond types or matched fragments. Commonly, the structure is not obtainable for the molecule of interest itself, so it needs to be constructed from appropriate fragments; even if the structure has been determined, the parameters may not be as reliable as those from an average of related structures.

2.1. Bond lengths

Bond lengths depend primarily on the atom types and the bond order. Typical values are: C—C, 1.51 Å; C═C, 1.33 Å; C≡C, 1.18 Å; C—O, 1.42 Å. However, bond lengths are not independent of neighbouring bonds and in general are a function of the distribution of electrons throughout the whole molecule. Delocalized systems have intermediate bond lengths and in particular it is important that equivalent bonds have equal target bond lengths. For example, all bond lengths in benzene are 1.39 ± 0.014 Å; a charged carboxyl group has both C—O bonds equal at 1.25 ± 0.015 Å, while a protonated carboxyl has different C—O bond lengths of 1.22 and 1.30 Å; similarly, a triply charged phosphate ion PO_4^{3-} has all P—O bonds equal at 1.51 Å. It is important to recognize partially delocalized and conjugated systems: the central bond in the diene —CH═CH—CH═CH— is intermediate in order and length between single and double, ∼1.43 Å, and the atoms will all lie in a plane.

Metals are a particular problem since it is often not clear what the best model is: metals may vary in their oxidation state and coordination number and bond lengths (and angles) will vary with these. In proteins, metals may occur with unexpected or mixed oxidation states, which is confusing. The geometry of metal sites in proteins has been well discussed in a series of papers by Harding (1999, 2000, 2001, 2002, 2004).

Searches in the CCDC database for bond lengths of fragments give a distribution with a typical standard deviation of about 0.02–0.03 Å, which includes both the real variation and the experimental error. Note that there are often outliers which need to be excluded; many of these arise from disordered solvent molecules, showing that even for small molecules there is a case for using weak stereochemical restraints for those atoms which are not well defined from diffraction data.

2.2. Bond angles

Bond angles depend on the atom type and the number and type of bonded atoms. For C atoms, the canonical cases are the tetrahedral sp^3 carbon with angles of ∼109.5° and the planar triangular sp^2 carbon with angles of ∼120°. The angles will be different if the substituents are of different sizes; for instance, the C—C—C angle in —C—CH_2—C— is 113.4 ± 2.8°, larger than the tetrahedral value. Errors in bond angles are typically 2–3°.

Angles around metals are best left unrestrained, unless a regular tetrahedron or octahedron is expected.

2.3. Planes

Atoms should lie in or near a plane if they are attached to an sp^2 carbon (or equivalent) or in a delocalized aromatic or conjugated system. Examples include benzene rings and the peptide plane. Planarity restraints may be implemented either by minimizing the distance to the mean plane, with a typical standard deviation of around 0.02 Å, or by restraining a series of real torsion or 'improper' torsion angles (an improper torsion is a rotation around an axis between two atoms which are not bonded to each other; *e.g.* the planarity of the peptide carbonyl may be restrained by the torsion C—O—C^{α}—N′, where N′ belongs to the next residue).

2.4. Chirality

A tetrahedral atom with four different substituents is chiral; that is, it cannot be superimposed on its mirror image (there are also other causes of chirality). The sense of a chiral centre is denoted *R* or *S* as defined by the Cahn–Ingold–Prelog priority rules for the substituents (see, for example, the *REFMAC* documentation or Cahn *et al.*, 1966). Chirality may be restrained either as 'chiral volume' or by improper torsions. If the vectors from the central atom of the tetrahedron to the other atoms in order of priority are written as $\mathbf{v}_1$, $\mathbf{v}_2$, $\mathbf{v}_3$, then the chiral volume is defined as $\mathbf{v}_1 \cdot (\mathbf{v}_2 \times \mathbf{v}_3)$. In *REFMAC*, the target value is calculated from the ideal bond lengths and angles and the weights are derived from $\sigma = 0.2$ Å^3. The *REFMAC* dictionary also allows the ideal volume to be defined as 'positive', 'negative' (inverting the definition) or 'both' (accepting either *R* or *S*).

These definitions are easy to get wrong and there are incorrect examples in the PDB. Chirality restraints may also be used to keep nonchiral atoms pyramidal, either if two substituents are equal (*e.g.* Leu C^{γ}) or if there are only three substituents (*e.g.* tertiary amines), although strictly this should not be necessary if the bond angles are tightly restrained.

2.5. Torsion angles (dihedrals)

A torsion angle around a rotatable bond (2–3) is defined by four bonded atoms, 1–2–3–4, and is defined as the angle between $\mathbf{v}_{12} \times \mathbf{v}_{23}$ and $\mathbf{v}_{23} \times \mathbf{v}_{34}$, with a positive angle a

clockwise rotation of atom 4 away from the eclipsed position when viewed down the 2→3 bond. In general, it is neither necessary nor desirable to restrain torsion angles strongly, although weak restraints may be useful to produce a staggered conformation around single bonds. Their undesirability arises partly from their periodic nature, *i.e.* they repeat every 120 or 180°: this makes them behave badly in a minimization. Special cases arise in flexible closed five- and six-membered rings, in which the ring torsions are variable within a limited range, giving a variable ring pucker. Ring torsions should not normally be restrained, but they may need manual alteration during model building.

2.5.1. Five-membered rings. The simplest flexible five-membered ring is cyclopentane, but the most important for biology is ribose (and related sugars). The ring is continuously flexible and the pucker can be described in terms of a maximum torsion angle ν_{max} and the pseudorotation phase angle P (Altona & Sundaralingam, 1972; Saenger, 1983).The pseudorotation angle is a combination of the torsion angles around the ring, defined as $\tan P = [(\nu_4 + \nu_1) - (\nu_3 + \nu_0)]/[2\nu_2(\sin 36^\circ + \sin 72^\circ)]$, where ν_0–ν_4 are the ring torsions O4′—C1′, C1′—C2′ *etc*. There are two conformations of low energy, C3′-*endo* (^{3}E) and C2′-*endo* (^{2}E), with pseudorotation angles P around 18 and 162°, respectively, but distortion from these is not difficult and there is only a low energy barrier between them (Murray-Rust & Motherwell, 1978). For this reason, it is easy for a refinement program to interconvert these conformations.

As the ring pucker changes, the angle between substituents changes, so bulky substituents act as levers on the ribose and the position of the substituents determines the ring pucker. Thus, the angle between the C1′—N(base) bond and the C4′—C5′ bond changes from 91° in the C3′-*endo* conformation to 77° in C2′-*endo* (Fig. 1), so the ring pucker is usually unambiguously and automatically determined even at low resolution.

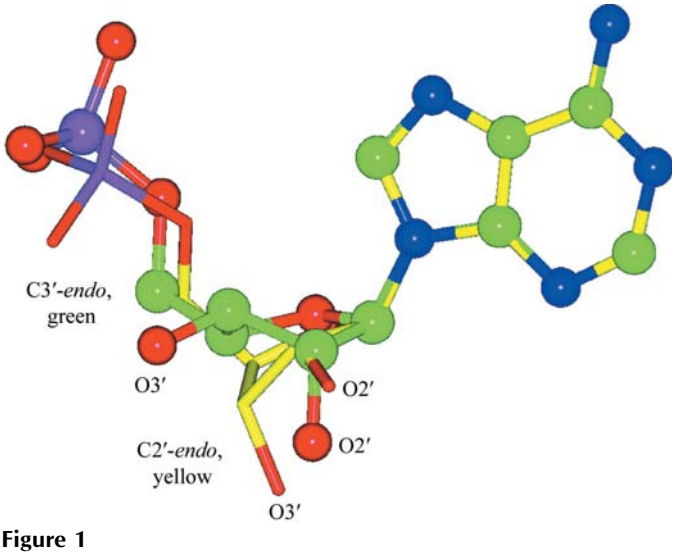

Figure 1
C2′-*endo* and C3′-*endo* conformations of AMP superimposed on the adenine ring, showing the difference in position of substituents on the C5′ and C3′ positions.

2.5.2. Six-membered rings. Flexible six-membered rings include cyclohexane derivatives, but most importantly pyranose sugars such as glucose and inositol derivatives. These rings can cycle through a series of conformations including the stable chair forms, through boat and twist-boat conformations that are less stable. In the chair forms, the substituents on each tetrahedral carbon can be divided into 'equatorial' substituents, in the approximate plane of the C atoms, and 'axial' substituents perpendicular to this plane. The ring can flip between two alternative chair forms, which interchanges axial and equatorial substituents. The molecule is most stable if the majority of bulky substituents are equatorial rather than axial, as this places them farther apart. For example, β-D-glucose, the most stable hexose sugar, has all its hydroxyl substituents equatorial. The ring-flipped form would have all hydroxyl substituents axial and is less stable (Fig. 2*a*). The pathway for the ring-flipping operation takes the molecule through higher energy boat conformations, so it is unlikely that a refinement procedure would flip an incorrectly puckered ring: this would probably have to be performed manually. This is not relevant to the construction of the dictionary, but it should be borne in mind during rebuilding.

A confusing case arises with D-*myo*-inositol hexakisphosphate. This natural stereoisomer has the phosphate group on the 2-carbon in an opposite conformation to that on the other five C atoms. There are thus two possible chair forms: with five axial and one equatorial phosphates or with five equatorial and one axial. The former is less stable in solution, but is the conformation which occurs in the crystal structure, presumably because of crystal contacts. When bound to

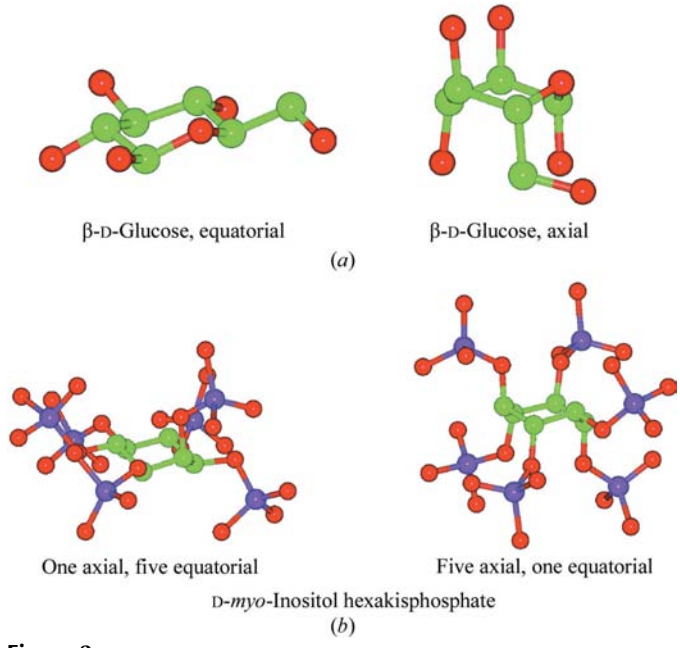

Figure 2
Axial and equatorial substituents on six-membered rings. (*a*) β-D-Glucose is most stable in the all-equatorial conformer. (*b*) D-*myo*-Inositol hexakisphosphate is largely axial in its crystal structure (CCSD code NAMIHP10), but when bound to a protein it is in the more stable largely equatorial form (bound to AP2; PDB code 1gw5).

proteins, *e.g.* the AP2 complex (Collins *et al.*, 2002), it is in the second (five equatorial, one axial) conformation (Fig. 2*b*).

3. Conclusions

Construction of dictionary entries is not always entirely straightforward and requires consideration of the chemistry of the molecule. All dictionary entries for small-molecule ligands, new or old, should be checked carefully to ensure that they are chemically sensible. Dictionaries have not always been constructed correctly: the PDB is not a reliable source of good stereochemistry.

References

Allen, F. H. (2002). *Acta Cryst.* B**58**, 380–388.
Altona, C. & Sundaralingam, M. (1972). *J. Am. Chem. Soc.* **100**, 2607–2613.
Cahn, R. S., Ingold, C. K. & Prelog, V. (1966). *Angew. Chem. Int. Ed. Engl.* **5**, 385–415.
Collins, B. M., McCoy, A. J., Kent, H. M., Evans, P. R. & Owen, D. J. (2002). *Cell*, **109**, 523–535.
Harding, M. (1999). *Acta Cryst.* D**55**, 1432–1443.
Harding, M. (2000). *Acta Cryst.* D**56**, 857–867.
Harding, M. (2001). *Acta Cryst.* D**57**, 401–411.
Harding, M. (2002). *Acta Cryst.* D**58**, 872–874.
Harding, M. (2004). *Acta Cryst.* D**60**, 849–859.
Kleywegt, G. (2006). *Acta Cryst.* D**63**, 94–100.
Murray-Rust, P. & Motherwell, S. (1978). *Acta Cryst.* B**34**, 2534–2546.
Saenger, W. (1983). *Principles of Nucleic Acid Structure.* Berlin: Springer–Verlag.