Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

A forward-looking suggestion for resolving the stereochemical restraints debate: ideal geometry functions

P. Andrew Karplus,[a]* Maxim V. Shapovalov,[b] Roland L. Dunbrack Jr[b] and Donald S. Berkholz[a]

[a]Department of Biochemistry and Biophysics, Oregon State University, Corvallis, OR 97331, USA, and [b]Institute for Cancer Research, Fox Chase Cancer Center, Philadelphia, PA 19111, USA

Correspondence e-mail: karplusp@science.oregonstate.edu

Received 19 January 2008
Accepted 21 January 2008

Jaskolski *et al.* (2007*a*) initiated a very important discussion (Jaskolski *et al.*, 2007*b*; Stec, 2007; Tickle, 2007) about the accuracy of ideal geometry targets and the appropriate stringency with which they should be obeyed at various resolutions. All of the discussants agree that protein structures determined at ultra-high resolution, which should be closer to the truth, tend to have larger r.m.s. deviations from ideality. This shows that real deviations from ideality do occur in true protein structures. Another point of agreement is that the real deviations from ideality are to some extent context dependent, so that the single target values used in current refinement programs are a simplification: 'The N–Cα–C valence angle has a wide spread and may have a bimodal distribution correlated with secondary structure', Jaskolski *et al.* (2007*a*); 'Recent results suggest that protein stereochemistry is context-dependent', Stec (2007); 'As they point out, the [deviations from ideality] will include real variations arising from the chemical environment', Tickle (2007); 'In such a situation, a single target, as used in the refinement programs, will not agree with any of the truly preferred values. This again suggests that the geometrical parameters of protein models should not be too tightly restrained to some predefined values', Jaskolski *et al.* (2007*b*).

We point out here that much of the context dependence of stereochemistry can be transformed from a frustrating reality that limits the accuracy of protein modeling to a feature that can instead enhance modeling accuracy. This transformation is possible because much of the context-dependent variation is not random but varies systematically with conformation. This systematic dependence was well documented in the mid-1990s, especially for the N—Cα—C bond angle, for which the expected value was seen to vary over a range of $\sim 10^\circ$ (Jiang *et al.*; Karplus, 1996; Schafer *et al.*, 1995). We are now updating these analyses using ultra-high resolution protein structures. These structures not only confirm the highly systematic variation of backbone bond angles with conformation (as illustrated for the N—Cα—C bond angle in Fig. 1), but they also reveal that the standard deviations are very low. For the N—Cα—C bond angle, the standard deviations in individual φ,ψ regions vary from 1.0 to 1.7° (Fig. 1*b*), much lower than the standard deviation of near 2.2° derived for the N—Cα—C bond angle in the population as a whole (see Table 3 of Jaskolski *et al.*, 2007*a*). Based on this result, it is clear that no single ideal target value can be appropriate for all of these conformations. However, it is also true that because each mean (or expected value) can be empirically discovered, this feature of protein structure can be accounted for by altering our refinement protocols to allow for ideal geometry targets that are dependent on context. In other words, we propose that more accurate refinement of protein structures at all resolution ranges will be obtained by moving beyond the 'single ideal value' paradigm to an 'ideal geometry function' paradigm (Schafer *et al.*, 1986) in which each restraint target value varies depending on the local conformation. We are now working to develop a set of empirical conformation-dependent 'ideal geometry functions' for backbone bond angles and lengths that can be incorporated into crystallographic refinement software (Berkholz & Karplus, unpublished work).

This work was partly supported by NSF grant MCB-9982727 (to PAK) and NIH grant P20-GM76222 (to RLD).

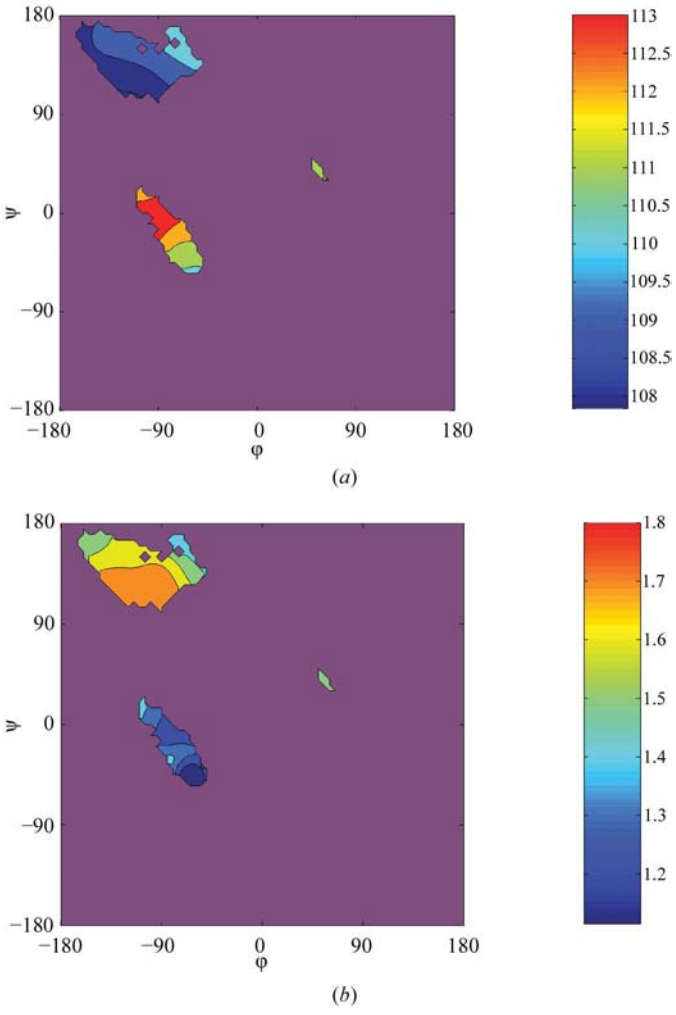

Figure 1
N—Cα—C bond angles vary systematically with conformation. (*a*) Average N—Cα—C backbone bond angles for well populated regions of the Ramachandran plot ($\geq$10 observations in a 5 $\times$ 5° bin) are displayed with coloring as indicated in the sidebar. The values were derived for general amino-acid residues (all but Gly, Pro, and residues preceding Pro) from a database of *ca* 19 000 residues in protein structures determined at 1.0 Å resolution or better. (*b*) The standard deviations of the distributions in each region of the Ramachandran plot are shown as in (*a*). The smoothed plots of the data were produced in Matlab using an adaptive kernel regression with a global κ of 25.

References

Jaskolski, M., Gilski, M., Dauter, Z. & Wlodawer, A. (2007*a*). *Acta Cryst.* D**63**, 611–620.
Jaskolski, M., Gilski, M., Dauter, Z. & Wlodawer, A. (2007*b*). *Acta Cryst.* D**63**, 1282–1283.
Jiang, X., Yu, C., Ming, C., Newton, S. Q., Paulus, E. P. & Schafer, L. (1997). *J. Mol. Struct.* **403**, 83–93.
Karplus, P. A. (1996). *Protein Sci.* **5**, 1406–1420.
Schafer, L., Cao, M. & Meadows, M. J. (1995). *Biopolymers*, **35**, 603–606.
Schafer, L., Ewbank, J. D., Klimkowski, V. J. & Siam, K. (1986). *J. Mol. Struct.* **135**, 141–158.
Stec, B. (2007). *Acta Cryst.* D**63**, 1113–1114.
Tickle, I. J. (2007). *Acta Cryst.* D**63**, 1274–1281.