

25.2. PROGRAMS IN WIDE USE

Table 25.2.6.1. Summary of expected values for stereochemical parameters in well resolved structures

Parameter	Old	New
% φ , ψ in core	>90.0%	>90.0%
χ_1 gauche ⁻	+64.1 ± 15.7°	+63.2 ± 11.4°
χ_1 trans	+183.6 ± 16.8°	+182.7 ± 13.1°
χ_1 gauche ⁺	-66.7 ± 15.0°	-66.0 ± 11.2°
χ_1 pooled standard deviation	±15.7°	±11.8°
χ_2 trans	+177.4 ± 18.5°	+177.2 ± 15.1°
χ_3 S—S bridge (left-handed)	-85.8 ± 10.7°	-84.8 ± 8.5°
χ_3 S—S bridge (right-handed)	+96.8 ± 14.8°	+92.2 ± 10.8°
Proline φ	-65.4 ± 11.2°	-64.6 ± 10.2°
α -Helix φ	-65.3 ± 11.9°	-65.5 ± 11.1°
α -Helix ψ	-39.4 ± 11.3°	-39.0 ± 9.8°
ω trans	+179.6 ± 4.7°	+179.5 ± 6.0°
C α —N—C'—C β (ζ) virtual torsion angle	+33.9 ± 3.5°	+34.2 ± 2.6°

Because the program requires only the 3D atomic coordinates of the structure, it can check the overall 'quality' of any model structure: whether derived experimentally by crystallography or NMR, or built by homology modelling. In the case of NMR-derived structures, it is useful to compare the protein geometry across the whole ensemble. An extended version of *PROCHECK*, called *PROCHECK-NMR*, is available for this purpose (Laskowski *et al.*, 1996), but will not be described here.

Note that *PROCHECK* only examines the geometrical properties of protein molecules; it ignores DNA/RNA and other non-protein molecules in the structure, except in so far as checking that the non-bonded contacts these make with the protein do not violate a fixed distance criterion.

25.2.6.2. The program

PROCHECK is in fact a suite of separate Fortran and C programs which are run successively *via* a shell script. The programs first

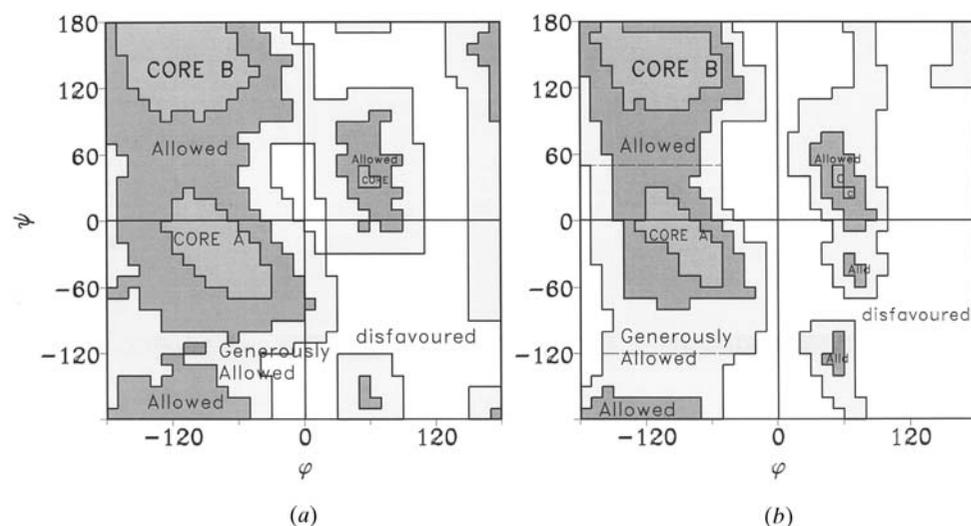


Fig. 25.2.6.1. *PROCHECK* Ramachandran plots showing the different regions, shaded according to how 'favourable' the φ - ψ combinations are, for (a) the original version of the program (1992) and (b) an updated version based on a more recent data set (1998) including more high-resolution structures. The 'core' and other favourable regions of the plot are more tightly compressed in the new version, with the white, disfavoured regions occupying more of the space.

'clean up' the input PDB file, relabelling certain side-chain atoms according to the IUPAC naming conventions (IUPAC-IUB Commission on Biochemical Nomenclature, 1970), then calculate all the protein's stereochemical parameters to compare them against the norms, and finally generate the PostScript output and a detailed residue-by-residue listing. Hydrogen and atoms with zero occupancy are omitted from the analyses and, where atoms are found in alternate conformations, only the highest-occupancy conformation is retained.

The source code for all the programs is available at <http://www.biochem.ucl.ac.uk/~roman/procheck>. It has also been incorporated into the *CCP4* suite of programs (Collaborative Computational Project, Number 4, 1994) at <http://www.dl.ac.uk/CCP/CCP4/main.html>, and can be run directly *via* the web from the Biotech Validation Server at <http://biotech.embl-ebi.ac.uk:8400/>.

25.2.6.3. The parameters

Table 25.2.6.1 shows the principal stereochemical parameters used by *PROCHECK*, based on the analysis of Morris *et al.* (1992), who looked for measures that are good indicators of protein quality. The table shows the original parameters together with a more up-to-date set derived from a more recent data set including a number of atomic resolution structures (*i.e.* those solved to 1.4 Å resolution or better).

For the most part, the parameters given in Table 25.2.6.1 are not included in standard refinement procedures and so are less likely to be biased by them. They can thus provide a largely independent and unbiased validation check on the geometry of each residue and hence point to regions of the protein structure that are genuinely unusual.

As more atomic resolution structures become available (Dauter *et al.*, 1997), these parameters will be improved. Because of their high data-to-parameter ratio, such structures can be refined using less strict restraints, and hence contain a smaller degree of bias in their geometrical properties – at least for the well ordered parts of the model. Such information moves us a step closer to an understanding of the 'true' geometrical and conformational properties of proteins in general and, one day, the target parameters will be derived exclusively from such structures.

PROCHECK also checks main-chain bond lengths and bond angles against the 'ideal' values given by the Engh & Huber (1991) analysis of small-molecule structures in the Cambridge Structural Database (CSD) (Allen *et al.*, 1979). Unlike the above parameters, these geometrical properties are usually restrained during refinement, and, furthermore, the Engh & Huber (1991) targets are the ones most commonly applied. Thus analyses of these values merely reflect the refinement protocol used and do not provide meaningful indicators of local or overall errors. However, the plots clearly show any wayward outliers which can nevertheless indicate problem regions in the structure.

25.2.6.4. Which parameters are best?

Possibly the most telling and useful of the 'quality' indicators for a protein model is the Ramachandran plot of residue φ - ψ torsion angles. This can often detect gross errors in the structure (Kleywegt & Jones, 1996*a,b*). In the original Ramachandran plot (Ramachandran *et al.*, 1963; Rama-