krishnan & Ramachandran, 1965), the 'allowed' regions were defined on the basis of simulations of dipeptides. In the *PROCHECK* version, the different regions of the plot are defined on the basis of how densely they are populated with data points taken from a database of well refined protein structures. The regions are: core, allowed, generously allowed and disallowed.

The 'core' regions are particularly important; the points on the plot tend to converge towards these regions, and to cluster more tightly within them, as one goes from structures solved at low resolution to those solved at high resolution (Morris *et al.*, 1992). This trend has recently been confirmed by Wilson *et al.* (1998), who looked at the case of atomic resolution structures. It has also been analysed in terms of 'attractors' at the most favourable regions of the plot; as the resolution improves, so the points are drawn towards these attractors (Walther & Cohen, 1999).

Fig. 25.2.6.1 shows the original *PROCHECK* Ramachandran plot and a more up-to-date version. The original was based on all 462 structures known at that time (1989/90), while the more recent one, generated in 1998, is based on 1128 non-identical (*i.e.* having a sequence identity $<95\%$) structures. It can be seen that the second plot has core regions which are much tighter than the original, and this is primarily due to the increase in the number of very high resolution structures giving a more accurate representation of the tight clustering in the most favourable regions.

Another parameter that seems to be a particularly sensitive measure of quality is the standard uncertainty (s.u.) of the χ torsion angles. Morris *et al.* (1992) found that the average values of a protein's χ_1 and χ_2 torsion angles are well correlated with the resolution at which the protein structure was solved. Although the data set was a fairly small one, the conclusion was borne out when tested on a larger set of more recent structures, including some solved to atomic resolution (Wilson *et al.*, 1998). This measure, however, cannot be relied on where side-chain conformations are either restrained or heavily influenced by the use of rotamer libraries.

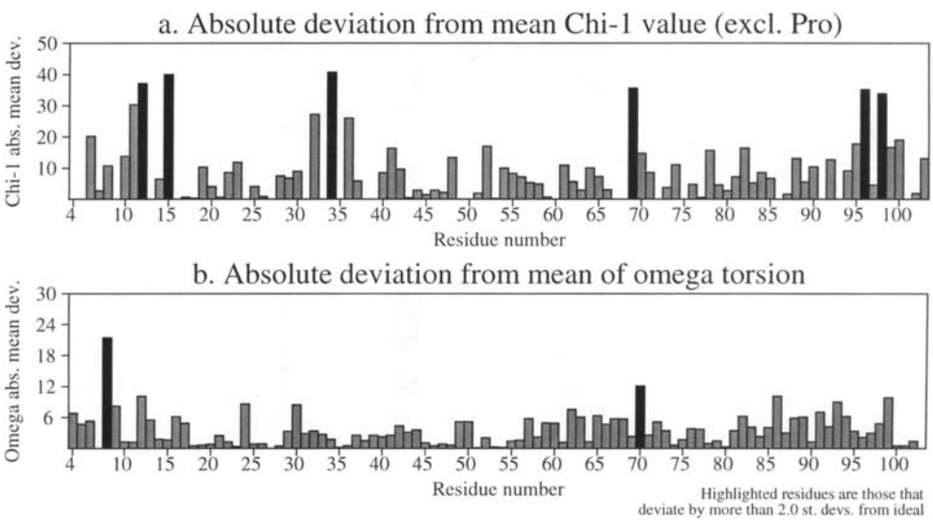

Fig. 25.2.6.2. Two of the residue-property plots generated by *PROCHECK*. The plots shown here are (*a*) the absolute deviation from the mean of the χ_1 torsion angle (excluding prolines) and (*b*) the absolute deviation from the mean of the ω torsion angle. Usually, three such plots are shown per page and can be selected from a set of 14 possible plots. On each graph, unusual values (usually those more than 2.0 standard deviations away from the 'ideal' mean value) are highlighted.

25.2.6.5. *Input*

The primary input to *PROCHECK* is the file containing the 3D coordinates of the protein structure to be processed. The file is

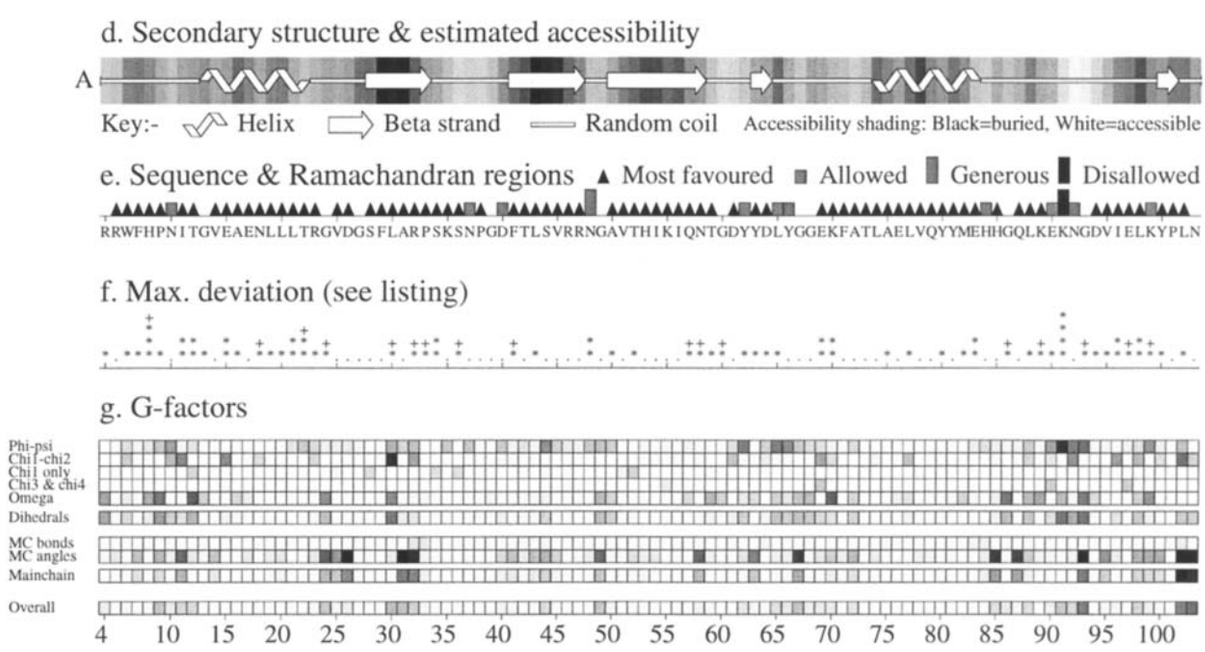

Fig. 25.2.6.3. Schematic plots of various residue-by-residue properties, showing (*d*) the protein secondary structure, with the shading behind it giving an approximation to each residue's accessibility, the darker the shading the more buried the residue; (*e*) the protein sequence plus markers identifying the region of the Ramachandran plot in which the residue is located; (*f*) a histogram of asterisks and plus signs showing each residue's maximum deviation from one of the ideal values, as shown on the residue-by-residue listing; and (*g*) the residue '*G* factor' values for various properties, where the darker the square the more 'unusual' the property.

Copyright © 2006 International Union of Crystallography

references