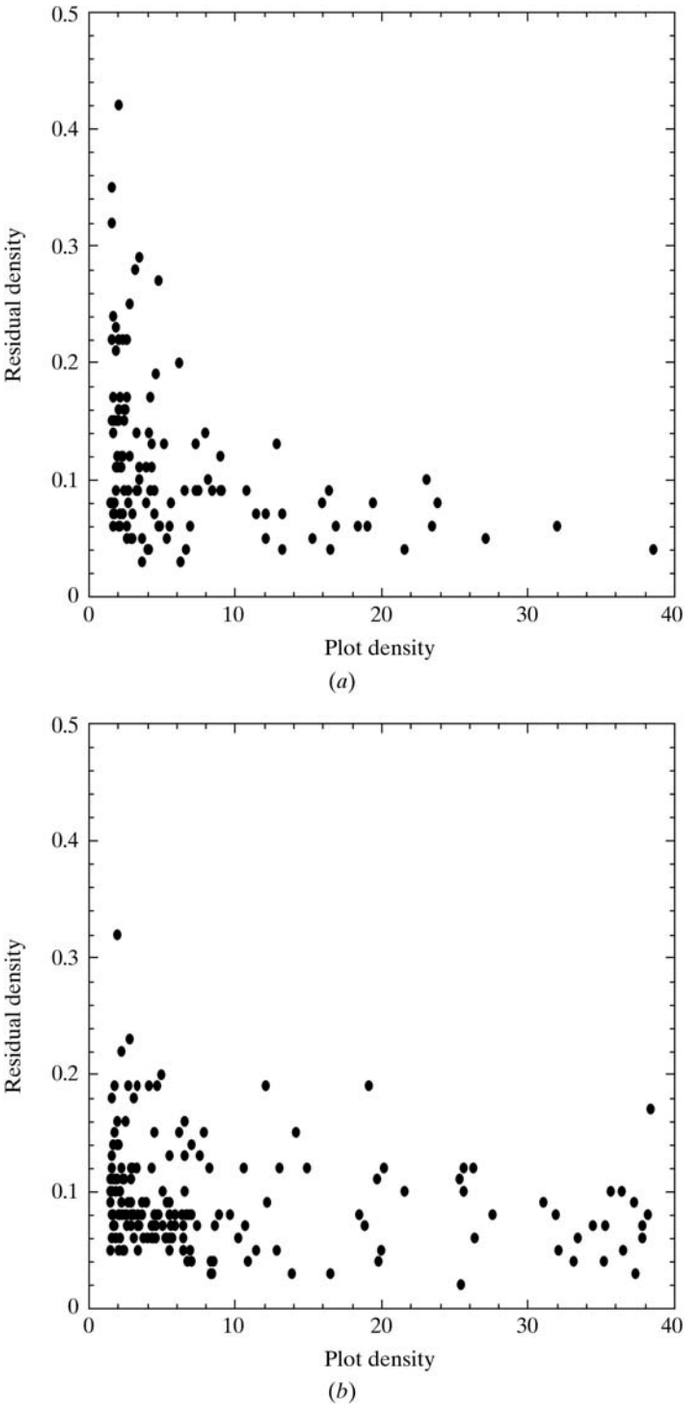

Fig. 22.4.5.4. Pairwise comparison of intermolecular-interaction density maps from the CSD and the PDB. Plots of *residual density* $|\rho(\text{CSD}) - \rho(\text{PDB})|$ *versus plot density*, *i.e.* the average density in the least dense situation (CSD or PDB), for situations where the protonation state of the central group is (*a*) unambiguous, and (*b*) ambiguous.

obviously caused by the more accurate calculation of the residual density. The 'true' residual density seems to be as low as about 6%.

Fig. 22.4.5.4(*b*) shows a similar graph, but now for those density maps in which the protonation state of the central group is ambiguous. As expected, the spread in the calculated residual densities is much higher, even for very dense plots. By comparing the density map from the PDB with the CSD maps for the different protonation states of the central group, the most frequent protonation state of this central group in the protein structures can be predicted. In Table 22.4.5.1, for example, the residual densities for protein carboxylic acid groups are shown, compared with the CSD plots of the neutral carboxylic acid and with those of the charged carboxylate anion. In all cases, the residual density is lower if the PDB map is compared with the CSD map for charged carboxylate anions. This indicates that the majority of glutamate and aspartate side chains are charged, which is consistent with other evidence.

Table 22.4.5.1. *Residual densities for carboxylic acid groups*

The PDB density maps are compared with the CSD maps for uncharged carboxylic acid and for charged carboxylate anions.

	Residual density (CCO_2H)	Residual density ($CCOO^-$)
Any (N,O,S)—H	0.06	0.04
Any N—H nitrogen	0.07	0.05
Any O—H oxygen	0.07	0.05
Non-donating oxygen	0.12	0.04
Carbonyl oxygen	0.13	0.07
Carbonyl carbon	0.12	0.04
Water oxygen	0.07	0.05
Any aliphatic C—H carbon	0.08	0.06

22.4.5.10. *Modelling applications that use CSD data*

Predicting binding modes of ligands at protein binding sites is a problem of paramount importance in drug design. One approach to this problem is to attempt to dock the ligand directly into the binding site. There are several protein–ligand docking programs available, *e.g. DOCK* (see Kuntz *et al.*, 1994), *GRID* (Goodford, 1985), *FLExX* and *FLExS* (Rarey *et al.*, 1996; Lemmen & Lengauer, 1997), and *GOLD* (Jones *et al.*, 1995, 1997). The docking program *GOLD*, developed by the University of Sheffield, Glaxo Wellcome and the CCDC, and which has the high docking success rate of 73%, uses a small torsion library, based on the data from the CSD, to explore the conformational space of the ligand. Its hydrogen-bond geometries and fitness functions are also partly based on CSD data. In the future, we intend to create a more direct link between the crystallographic data and the docking program, *via* IsoStar and the developing torsion library.

Another approach to the prediction of binding modes is to calculate the energy fields for different probes at each position of the binding site, for instance using the *GRID* program (Goodford, 1985). The resulting maps can be displayed as contoured surfaces which can assist in the prediction and understanding of binding modes of ligands. CCDC is developing a program called *SuperStar* (Verdonk *et al.*, 1999) which uses a similar approach to that of the *X-SITE* program (Singh *et al.*, 1991; Laskowski *et al.*, 1996). However, *SuperStar* uses non-bonded interaction data from the CSD rather than the protein side chain–side chain interaction data employed in *X-SITE*. Thus, for a given binding site and contact group (probe), *SuperStar* selects the appropriate scatter plots from the IsoStar library, superimposes the scatter plots on the relevant functional groups in the binding site, and transforms them into one composite probability map. Such maps can then, for example, be used to predict where certain functional groups are likely to interact with the binding site. The strength of *SuperStar* is that it is based entirely on experimental data (although this is also the cause of some limitations). The fields simply represent what has been observed in crystal strucures. We are currently verifying *SuperStar* on a test set of more than 100 protein–ligand complexes from the PDB and preliminary results are encouraging.

references