

Directional Preferences of Intermolecular Contacts to Hydrophobic Groups

JASON C. COLE, ROBIN TAYLOR* AND MARCEL L. VERDONK

Cambridge Crystallographic Data Centre, 12 Union Road, Cambridge CB2 1EZ, England.

E-mail: taylor@ccdc.cam.ac.uk

(Received 27 February 1998; accepted 29 June 1998)

Abstract

Analysis of data from the IsoStar library shows that many hydrophobic groups exhibit strikingly strong directional preferences in their intermolecular interactions. Specific directional interactions may occur because of the large quadrupole moments of many aromatic ring systems, the residual electropositive charge on most carbon-bound H atoms and the effects of polarization on soft hetero-ring atoms such as sulfur. In consequence, the binding of a hydrophobic group to a hydrophobic protein cavity is not simply a matter of matching complementary shapes. Directional preferences of nonbonded contacts to hydrophobic groups may need to be taken into account in parameterizing the next generation of protein–ligand docking programs.

1. Introduction

The ability to predict the binding orientation and affinity of a small molecule to a protein is of enormous importance in structure-based drug design and, potentially, in the rational design of combinatorial libraries. It has, therefore, been the subject of intensive research, which has led to the development of a number of computer programs for ligand docking and design, *e.g.* Kuntz *et al.* (1982); Rarey *et al.* (1996); Jones *et al.* (1997). At present, the best of these programs will reproduce the observed binding positions of perhaps 70–80% of protein-bound ligands in complexes taken from the Brookhaven Protein Data Bank (Bernstein *et al.*, 1977). Attempts to improve on this success rate are likely to focus on the development of better ‘scoring functions’, *i.e.* the empirical equations used to estimate protein–ligand binding energies. Typical scoring functions, such as those of Böhm (1994) and Jain (1996), include terms for ionic, hydrogen-bonding and hydrophobic protein–ligand interactions, together with terms that account for ligand conformational strain and (albeit crudely) entropic effects. In general, the ionic and hydrogen-bonding terms are directional in nature (*e.g.* hydrogen-bond energy varies with donor–hydrogen···acceptor angle). However, hydrophobic energies are usually estimated by measuring, explicitly or implicitly, the area of surface contact between hydrophobic parts of the

ligand and protein; *i.e.* hydrophobic contacts are assumed to be non-specific and non-directional.

Some justification for this assumption comes from model experiments showing an approximately linear relationship between the free energy of transfer of a small solute to a hydrophobic solvent and the burial of solvent-accessible hydrophobic surface area (Fersht & Serrano, 1993; Chothia, 1976; Eisenberg & McLachlan, 1986; Ooi *et al.*, 1987). However, as early as the mid-eighties, it was pointed out that specific edge-to-face interactions between phenylalanine residues lend stability to protein structures (Burley & Petsko, 1985). Attention was also drawn to the preference for phenylalanine···oxygen contacts to lie in the plane of the aromatic ring (Thomas *et al.*, 1982; Burley & Petsko, 1988). More recently, it was demonstrated that the large quadrupole moment of benzene leads to a stabilizing, and directional, cation··· π interaction (*e.g.* Verdonk, 1995; Dougherty, 1996). Thus, there is reason to believe that specific directional nonbonded interactions to hydrophobic groups may be of greater importance than recognized in most common protein–ligand scoring functions.

The recent development of a library of nonbonded interactions (IsoStar; Bruno *et al.*, 1997) affords the possibility of examining this question in more detail. Extensive analysis of the crystallographic data in this library, as reported below, indicates that many hydrophobic groups form highly directional nonbonded interactions. In consequence, different hydrophobic groups are likely to have different binding preferences, even if they have approximately the same size and shape.

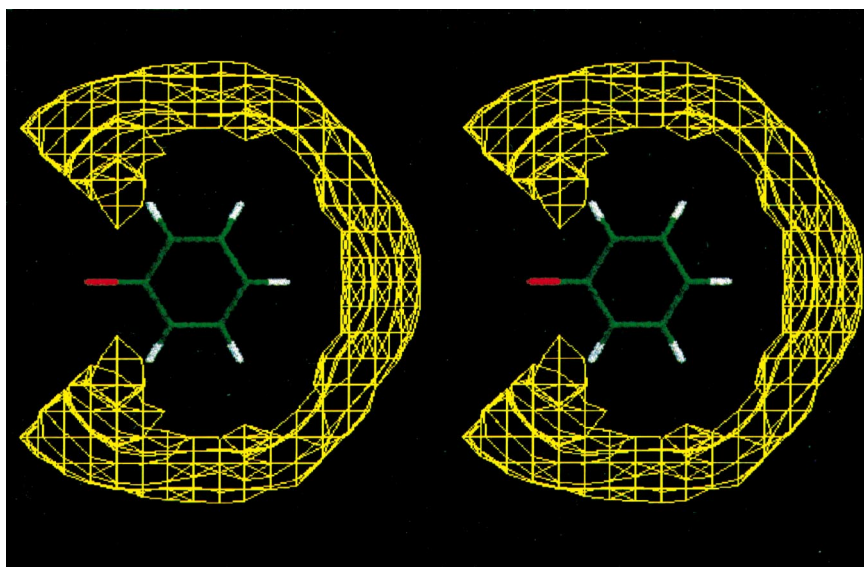
2. Experimental

The Cambridge Structural Database (CSD; Allen *et al.*, 1991) is probably the most comprehensive source of data on intermolecular contacts. Currently, the database contains the results of over 180 000 organic and organometallic crystal-structure analyses. Using the *QUEST* program (Cambridge Crystallographic Data Centre, 1992), it is possible to find all crystal structures in the CSD containing an intermolecular contact between two groups, *A* and *B*. By superimposing the *A*···*B* contacts

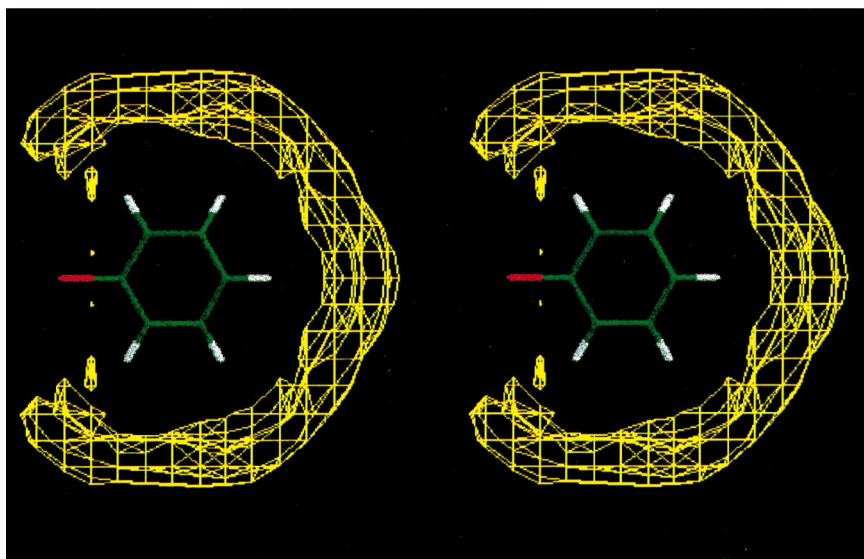
from the different CSD structures so that the *A* moieties are overlaid in a least-squares sense, a three-dimensional scatterplot can be produced showing the experimental distribution of *B* (the 'contact group') around *A* (the 'central group'). Using this method, a library of intermolecular interactions (IsoStar) has been constructed (Bruno *et al.*, 1997). It contains over 5000 scatterplots derived from the CSD, each showing a different type of *A*...*B* interaction. In addition, the library contains about 1500 scatterplots based on protein-ligand contacts observed in complexes from the Brookhaven Protein Data Bank.

Each scatterplot in IsoStar can be converted to a contoured surface which shows how the density of

contacts varies around the central group. In the case of CSD-based plots, these surfaces can be put on a 'normalized' scale (Verdonk *et al.*, 1998). This is performed by counting the number of contact groups in each unit volume around the central group, dividing by the number of contacts that would be expected by chance, and contouring on the resulting 'normalized' density values. The number of contacts expected by chance is estimated by computing the average density of contact groups in the CSD crystal structures from which the scatterplot was derived. Thus, a surface contoured at a value *y* encloses all regions in space where the density of contacts is at least *y* times greater than would be expected from purely stoichiometric factors. Except



(a)



(b)

Fig. 1. Contoured density surfaces showing the distribution of (a) chloride ions, (b) carbonyl O atoms.

where otherwise stated, all surfaces shown in this paper are based on CSD data and contoured at a level of 1.

3. Results

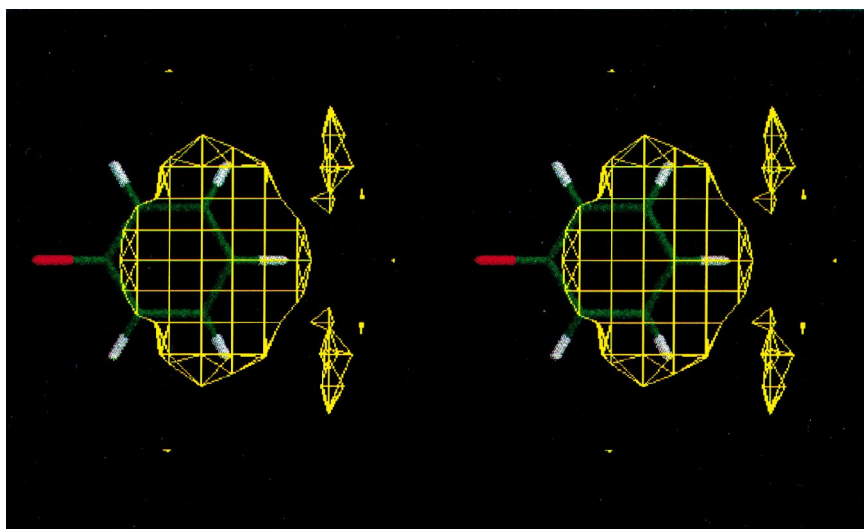
3.1. Contacts to phenyl

Phenyl, the archetypal hydrophobic group, forms highly directional nonbonded contacts. This is illustrated by the surfaces in Fig. 1, which show the density of X contacts around phenyl, where X = chloride, carbonyl oxygen, alkyl hydrogen (Csp^3-H) and aromatic hydrogen ($Car-H$). Quite clearly, contacts to chloride ions and carbonyl O atoms occur mainly around the ring

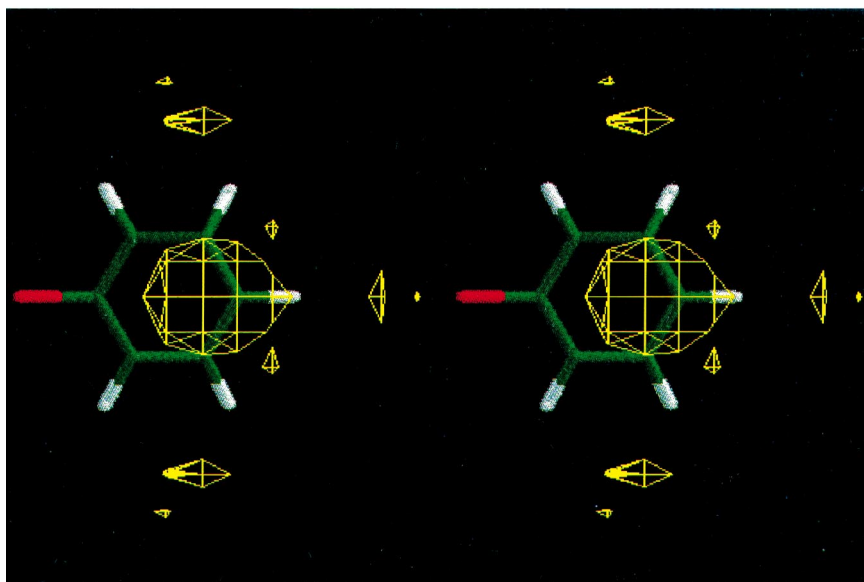
edges whereas those to the carbon-bound H atoms, which will generally be slightly electropositive, occur above the phenyl ring. The directional effects are strong and undoubtedly a consequence of the large quadrupole moment of benzene (Dougherty, 1996).

3.2. Contacts to the π systems of aromatic rings

As might be expected, substitution of the phenyl ring by electron-withdrawing or donating groups alters the distribution of nonbonded contacts around the ring. Fig. 2 shows the density of alkyl H atoms (Csp^3-H) around p -methoxyphenyl, p -tolyl, p -nitrophenyl, pentafluorophenyl and 3,5-dinitrophenyl (*cf.* Fig. 1c). Since all the density surfaces are contoured at the same level (*viz.*

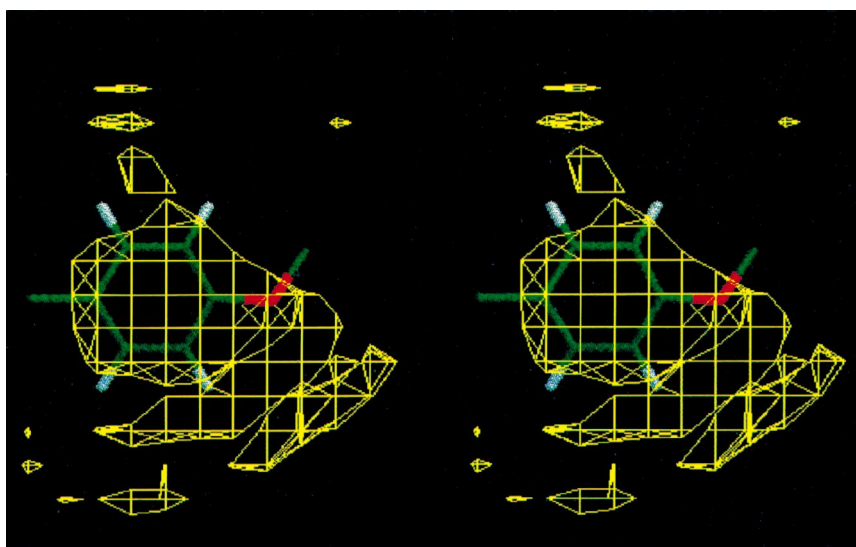


(c)

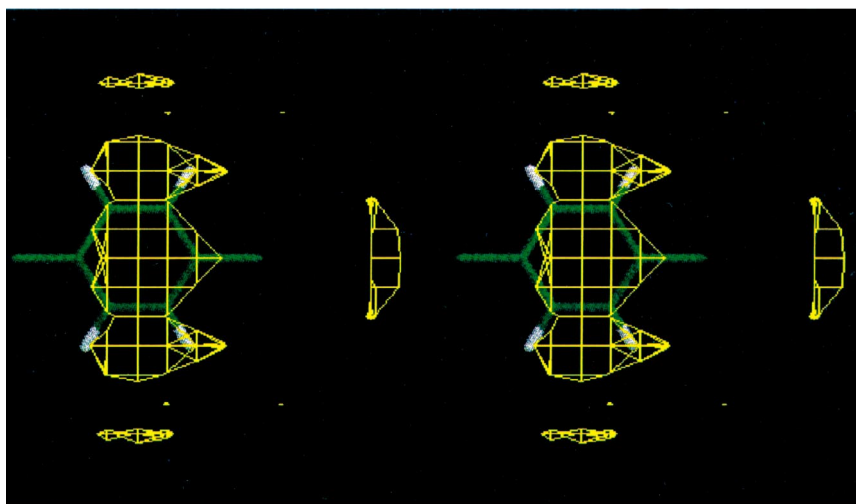


(d)

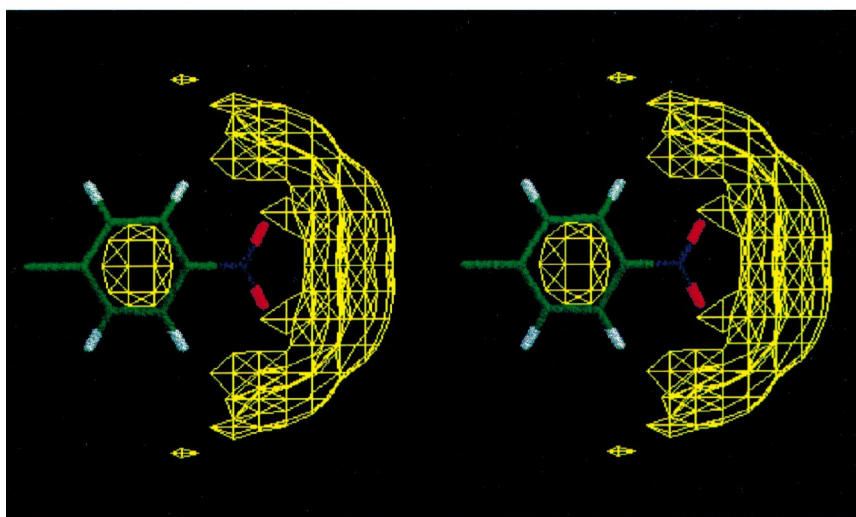
Fig. 1 (*cont.*) (c) alkyl H atoms and (d) aryl H atoms around phenyl rings.



(a)



(b)



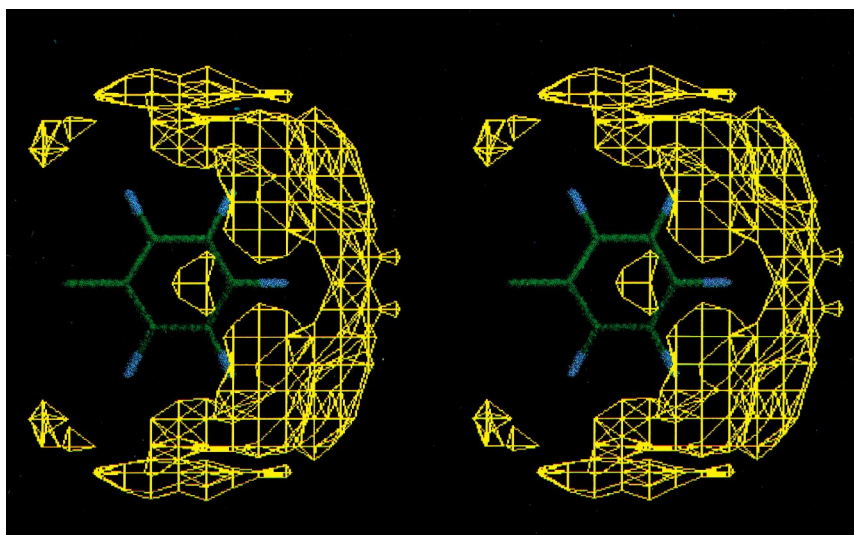
(c)

Fig. 2. Contoured density surfaces showing the distribution of alkyl H atoms around (a) *p*-methoxyphenyl, (b) *p*-tolyl, (c) *p*-nitrophenyl.

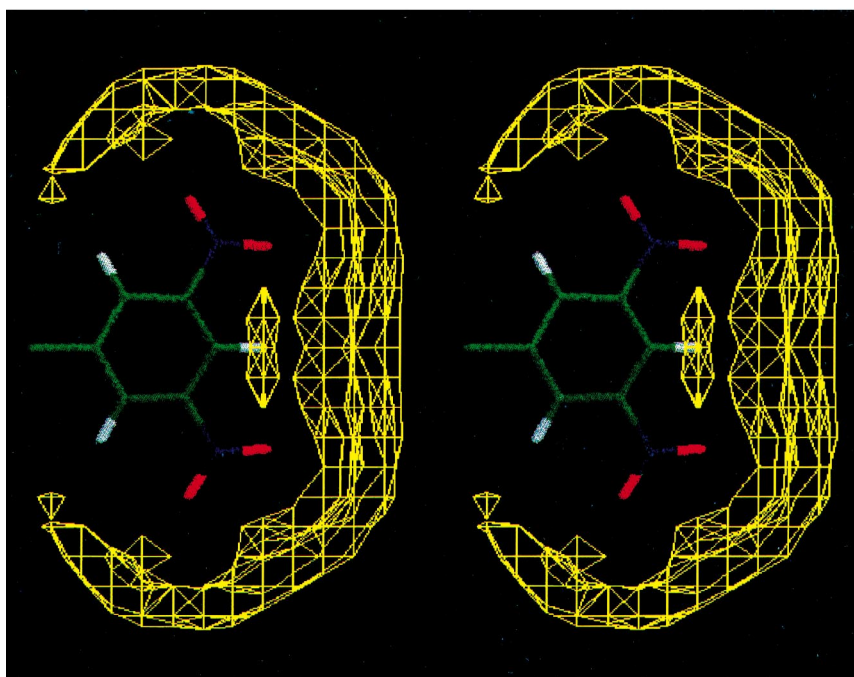
1), it is meaningful to compare the shape and extent of one surface with another. As the ring substituents become more electron withdrawing, the density of contacts to hydrogen above the aromatic ring decreases. Presumably, electron-withdrawing substituents deplete aromatic π density and hence make the regions above and below the ring less favourable, electrostatically, for carbon-bound H atoms.

Similar effects are produced by the introduction of hetero-atoms into aromatic rings. For example, the surfaces in Fig. 3 show the density distributions of alkyl

hydrogen contacts to indole and pyridine ring systems (*cf.* Fig. 1c). The electron-rich indole ring system forms a high density of contacts above and below the ring (note that this surface is contoured at a level of 2 instead of the more usual 1). The less electron-rich π system of pyridine forms fewer contacts to alkyl H atoms, and those that do occur tend to be displaced from the ring centre towards the hetero-nitrogen. Interestingly, the highest density of alkyl hydrogen contacts to isoxazole (plot not shown) occurs approximately above the middle of the N—O bond rather than above the centre of the ring,



(d)



(e)

Fig. 2 (*cont.*) (d) pentafluorophenyl and (e) 3,5-dinitrophenyl.

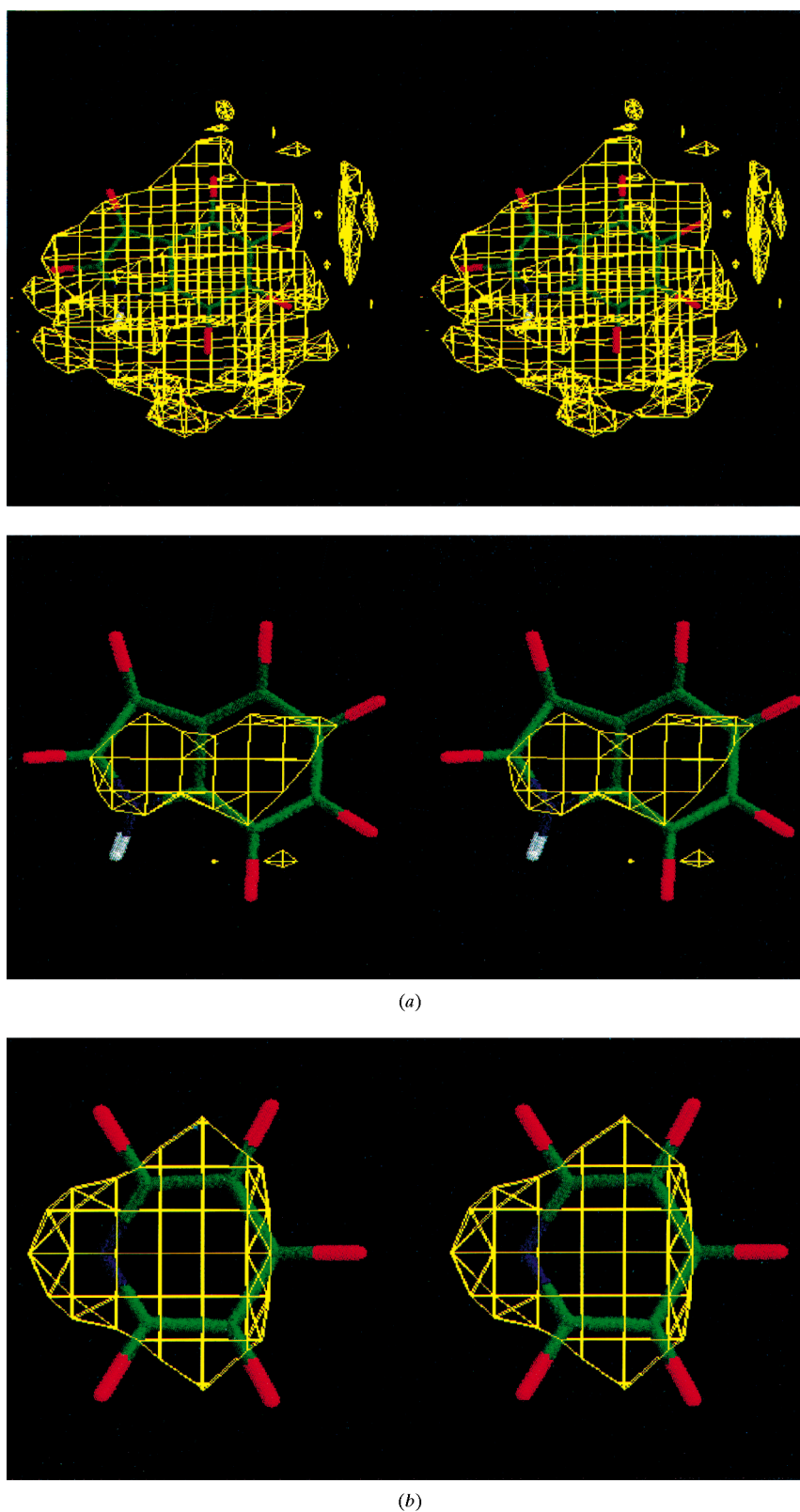


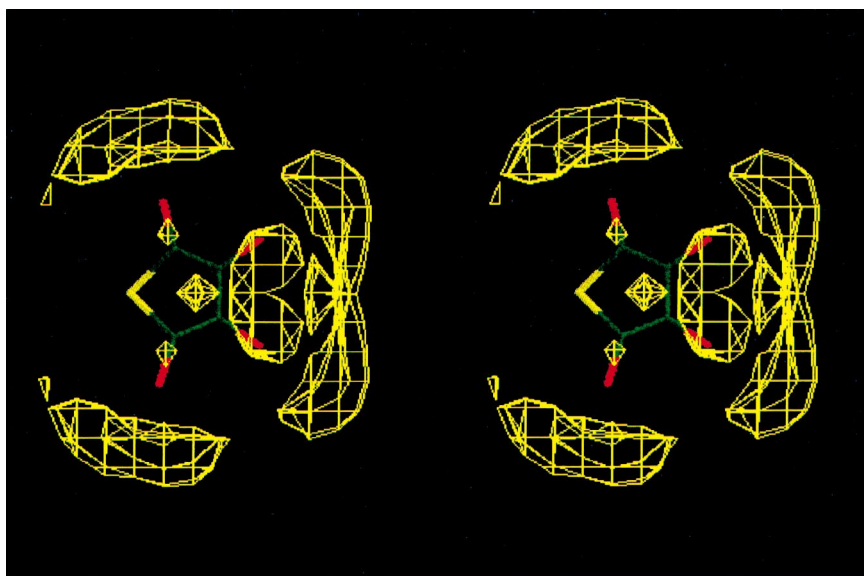
Fig. 3. Contoured density surfaces showing the distribution of alkyl H atoms around (a) indole (contoured at 2) and (b) pyridine.

although this observation is based on relatively few data. Because of its preference for forming contacts to H atoms above and below the ring plane, indole rarely forms stacking interactions with phenyl. In contrast, isoxazole, for example, does so commonly. Thus, in 109 contacts between phenyl and indole rings in the CSD (identified from IsoStar Version 1.0), interactions between the two are invariably edge-to-face, with the exception of only two structures (CSD reference codes HECJEE, Sivaraman *et al.*, 1994; PMPAIN, Cardellini *et al.*, 1980). In one of these (HECJEE), the phenyl ring is bonded to a sulfone group, so is more electron deficient than usual. Of 53 contacts between isoxazole and phenyl, several are convincing stacking interactions (*e.g.* in JIMZEK, Smith *et al.*, 1991; PIOXPY, Bovio & Locchi, 1972; SEYMIS, Verner *et al.*, 1990; VABXEB,

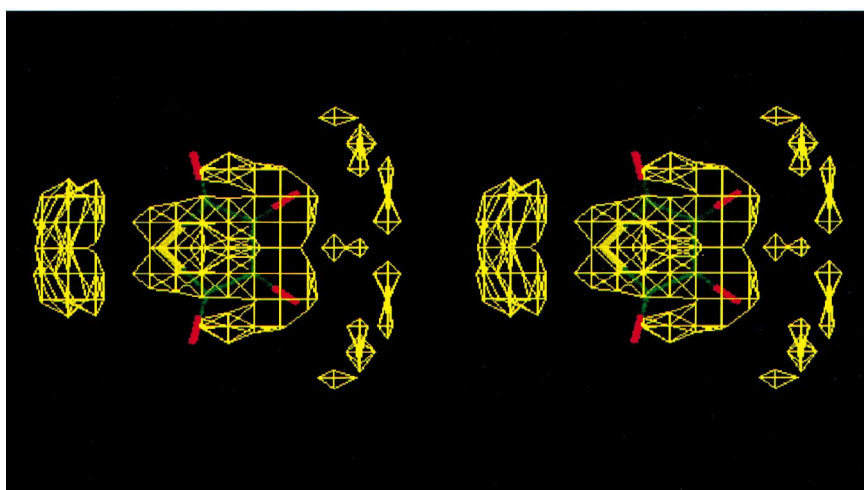
Marron *et al.*, 1988; VUTMEC, Yamauchi *et al.*, 1993; ZOXSIX, Madsen *et al.*, 1996).

3.3. Contacts around the edges of aromatic rings

Directionality is also seen in the contacts that occur around the edges of both substituted benzene rings and aromatic heterocycles. For example, Fig. 4 shows the density distributions of carbonyl O-atom and aryl hydrogen contacts to thiophene. Whereas O atoms predominate around most of the ring edge, this tendency is strikingly reversed in the vicinity of the S atom. Here, contacts to *Car*-H are more common, indicating that the ring is effectively more hydrophobic in this position. Similar effects are seen in substituted phenyl groups, *e.g.* in the vicinity of the F atom of *p*-fluor-



(a)



(b)

Fig. 4. Contoured density surfaces showing the distribution of (a) carbonyl O atoms and (b) aryl H atoms around thiophene.

ophenyl. Some heterocycles are also capable of forming specific nonbonded motifs. For example, the N and O atoms of isoxazole can form $\text{CH}\cdots\text{N}$ and $\text{CH}\cdots\text{O}$ interactions, respectively, to adjacent H atoms on a neighbouring, approximately coplanar, phenyl ring, resulting in a cyclic phenyl-isoxazole dimer. This is seen several times in the CSD, *e.g.* in YIBZEO (Ryng *et al.*, 1994).

3.4. Contacts to non-aromatic hydrophobic groups

Many nonaromatic hydrophobic groups also form directional interactions. A common cause of directionality is $\text{C}-\text{H}\cdots\text{O}$ hydrogen bonding (Taylor & Kennard, 1982; Desiraju, 1996). For example, the ethynyl group forms a high density of contacts to oxygen along the extension of the $\text{C}-\text{H}$ bond direction, whereas contacts to, *e.g.*, alkyl H atoms occur in a halo around the CC axis (Fig. 5). Analogous trends are seen for vinyl. Somewhat surprisingly, even the methyl group

shows directional preferences. For example, Fig. 6(a) shows the distribution of carbonyl O atoms around methyl. The highest density positions are approximately along the $\text{C}-\text{H}$ bond directions. In contrast, the distribution of *Car*-H H atoms around methyl shows maximum density approximately on the threefold axis of the methyl group (Fig. 6b). Some halogenated alkyl groups also show strong directional preferences, as illustrated by the density distribution of carbonyl oxygen contacts to dichloromethyl (Fig. 7).

3.5. Contacts to atoms in polarizing environments

Highly directional contacts to hydrophobic groups may also occur when a halogen substituent (Cl, Br or I; not F) is polarized by an electron-withdrawing environment. In this situation, short nonbonded contacts between the halogen and electronegative elements such as oxygen and nitrogen are well

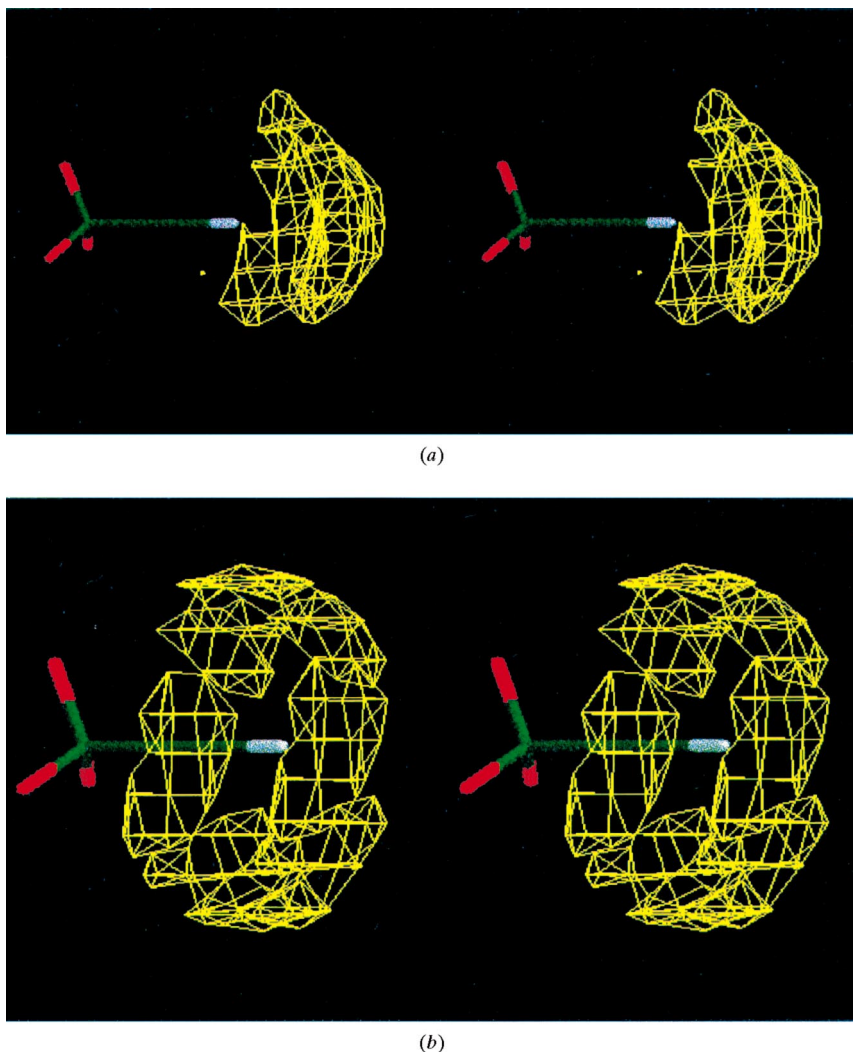


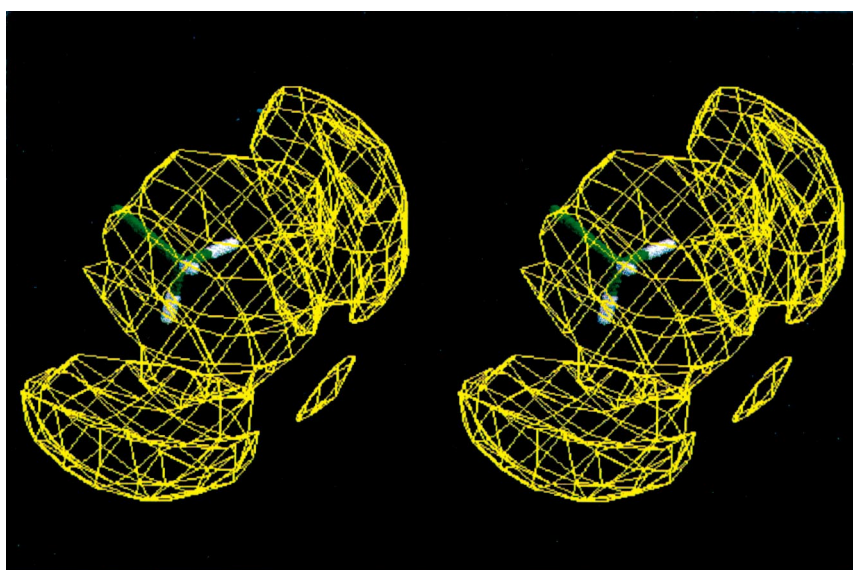
Fig. 5. Contoured density surfaces showing the distribution of (a) carbonyl O atoms and (b) alkyl H atoms around ethynyl groups bonded to tetrahedral C atoms (*i.e.* $\text{Csp}^3\text{-CCH}$)

known (Lommerse *et al.*, 1996). These contacts are strikingly directional, tending to occur with approximately linear C—Hal...O and C—Hal...N angles. The phenomenon is well illustrated by several *p*-bromobenzoate ester derivatives (*e.g.* CSANDR, Hope & Christensen, 1968; SPATLB, Gerwick *et al.*, 1980; SUHROC, Mori *et al.*, 1993; BBPROM, Abrahamsson, 1963; MURSBB, Paton & Paul, 1979), although many other illustrative series could have been chosen. Directional contacts between polarized S atoms (*e.g.* in sulfur-nitrogen heterocycles) and O atoms and N atoms are also common (Burling & Goldstein, 1992; Baalham, 1996); indeed, the directional properties of nonbonded contacts to sulfur formed the subject of one of the

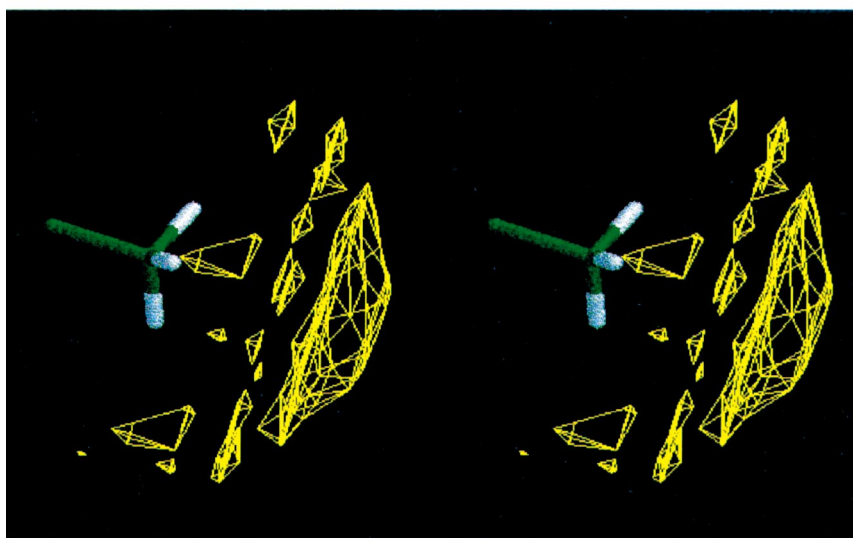
earliest analyses based on the CSD (Rosenfield *et al.*, 1977).

3.6. Contacts in protein–ligand complexes

Much less data are available from protein–ligand crystal structures. However, when comparisons can be made between contacts in small-molecule crystal structures and contacts between proteins and ligands, good concordance is usually seen. For example, Fig. 8 shows that ligand O atoms accumulate around the edges of phenylalanine rings (*cf.* Fig. 1*b*). Conversely, the C atoms of, *e.g.*, ligand alkyl CH groups cluster above and below tyrosine and tryptophan rings (and, in the latter



(a)



(b)

Fig. 6. Contoured density surfaces showing the distribution of (a) carbonyl O atoms and (b) aryl H atoms around methyl groups.

case, predominately over the six-ring rather than the five-ring).

4. Discussion

The results of this work show that the nonbonded contacts formed by a wide variety of hydrophobic groups, including some very common ones such as phenyl and indole, are strikingly directional. Three related but reasonably distinct reasons for directionality can be postulated. First, the substantial quadrupole moments of some aromatic systems lead to anisotropy in nonbonded contacts. Secondly, hetero-atoms and ring substituents, such as the sulfur of thiophene and the fluorine of *p*-fluorophenyl, may have very different

nonbonded preferences from ring CH's. This effect is sometimes emphasized or modified by intramolecular polarization effects, such as the presence of a strongly electron-withdrawing neighbour. Thirdly, CH groups, particularly when relatively acidic such as those in acetylenes, form directional CH \cdots O interactions and avoid contacts to electropositive atoms.

In consequence, different hydrophobic groups may be expected to show different binding properties to proteins. Finding a bioisosteric replacement of a hydrophobic substituent is, therefore, not simply a matter of size and shape. How important this is for the parameterization of protein-ligand docking and design programs remains to be seen. On the one hand, the energies of intermolecular interactions to hydrophobic

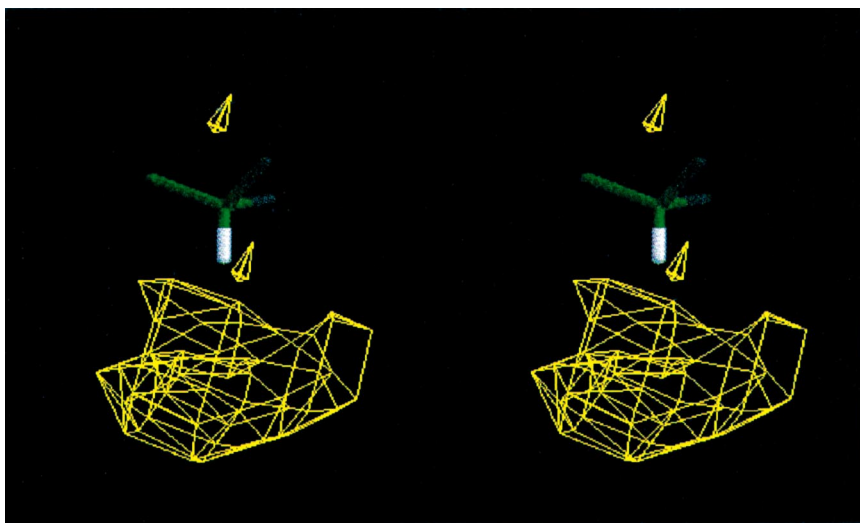


Fig. 7. Contoured density surface showing the distribution of carbonyl O atoms around dichloromethyl groups.

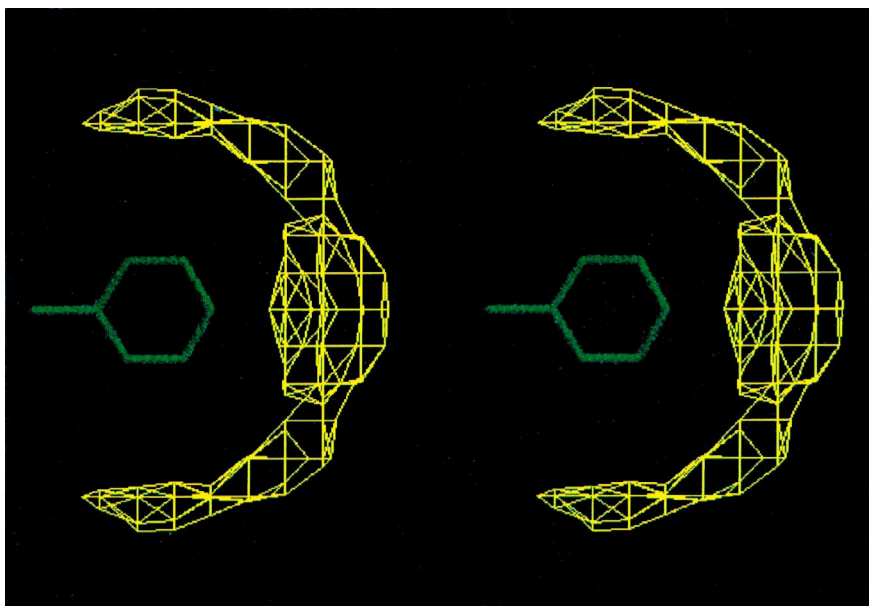


Fig. 8. Contoured density surface showing the distribution of ligand O atoms around phenylalanine residues in protein-ligand complexes from the Brookhaven Protein Data Bank.

groups are undoubtedly small. On the other hand, they are clearly large enough to have a striking effect on the organization of molecules in condensed phases. One possibility is that binding *energies* may be comparatively uninfluenced by directional hydrophobic group interactions but that binding *orientations* will be significantly affected. This is a matter of speculation at present, but it is probably fair to say that the directional properties of nonbonded contacts to hydrophobic groups have been underestimated in rational drug design.

We thank our colleagues and collaborators in the IsoStar project, especially Ian Bruno, Jos Lommerse and Scott Rowland.

References

- Abrahamsson, S. (1963). *Acta Cryst.* **16**, 409–418.
- Allen, F. H., Davies, J. E., Galloy, J. J., Johnson, O., Kennard, O., Macrae, C. F., Mitchell, E. M., Mitchell, G. F., Smith, J. M. & Watson, D. G. (1991). *J. Chem. Inf. Comput. Sci.* **31**, 187–204.
- Baalham, C. A. (1996). MPhil thesis, University of Cambridge, UK.
- Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977). *J. Mol. Biol.* **112**, 535–542.
- Böhm, H.-J. (1994). *J. Comput. Aided Mol. Design*, **8**, 243–256.
- Bovio, B. & Locchi, S. (1972). *J. Cryst. Mol. Struct.* **2**, 89–97.
- Bruno, I. J., Cole, J. C., Lommerse, J. P. M., Rowland, R. S., Taylor, R. & Verdonk, M. L. (1997). *J. Comput. Aided Mol. Design*, **11**, 525–537.
- Burley, S. K. & Petsko, G. A. (1985). *Science*, **229**, 23–28.
- Burley, S. K. & Petsko, G. A. (1988). *Adv. Protein Chem.* **39**, 125–189.
- Burling, F. T. & Goldstein, B. M. (1992). *J. Am. Chem. Soc.* **114**, 2313–2320.
- Cambridge Crystallographic Data Centre (1992). *CSD User Manual*, Cambridge, UK.
- Cardellini, L., Tosi, G., Bocelli, G. & Musatti, A. (1980). *Cryst. Struct. Commun.* **9**, 233–237.
- Chothia, C. (1976). *J. Mol. Biol.* **105**, 1–14.
- Desiraju, G. R. (1996). *Acc. Chem. Res.* **29**, 441–449.
- Dougherty, D. A. (1996). *Science*, **271**, 163–168.
- Eisenberg, D. & McLachlan, A. D. (1986). *Nature (London)*, **319**, 199–203.
- Fersht, A. R. & Serrano, L. (1993). *Curr. Opin. Struct. Biol.* **3**, 75–83.
- Gerwick, W. H., Fenical, W., Van Engen, D. & Clardy, J. (1980). *J. Am. Chem. Soc.* **102**, 7991–7993.
- Hope, H. & Christensen, A. T. (1968). *Acta Cryst.* **B24**, 375–382.
- Jain, A. N. (1996). *J. Comput. Aided Mol. Design*, **10**, 427–440.
- Jones, G., Willett, P., Glen, R. C., Leach, A. R. & Taylor, R. (1997). *J. Mol. Biol.* **267**, 727–748.
- Kuntz, I. D., Blaney, J. M., Oatley, S. J., Langridge, R. & Ferrin, T. E. (1982). *J. Mol. Biol.* **161**, 269–288.
- Lommerse, J. P. M., Stone, A. J., Taylor, R. & Allen, F. H. (1996). *J. Am. Chem. Soc.* **118**, 3108–3116.
- Madsen, U., Frydenvang, K., Ebert, B., Johansen, T. N., Brehm, L. & Krosgaard-Larsen, P. (1996). *J. Med. Chem.* **39**, 183–190.
- Marron, B. E., Schlicksupp, L. & Natale, N. R. (1988). *J. Heterocycl. Chem.* **25**, 1067–1070.
- Mori, K., Aki, S. & Kido, M. (1993). *Liebigs Ann. Chem.* pp. 83–90.
- Ooi, T., Oobatake, M., Nemethy, G. & Scheraga, H. A. (1987). *Proc. Natl Acad. Sci. USA*, **84**, 3086–3090.
- Paton, W. F. & Paul, I. C. (1979). *Cryst. Struct. Commun.* **8**, 207–211.
- Rarey, M., Kramer, B., Lengauer, T. & Klebe, G. (1996). *J. Mol. Biol.* **261**, 470–489.
- Rosenfield, R. E. Jr, Parthasarathy, R. & Dunitz, J. D. (1977). *J. Am. Chem. Soc.* **99**, 4860–4862.
- Ryng, S., Machon, Z. & Glowiak, T. (1994). *J. Chem. Cryst.* **24**, 483–488.
- Sivaraman, J., Subramanian, K., Velmurugan, D., Subramanian, E. & Sadanandan, E. V. (1994). *Acta Cryst.* **C50**, 787–789.
- Smith, M. P., Mirzaei, Y. R., Natale, N. R., Scott, B. & Willett, R. D. (1991). *Acta Cryst.* **C47**, 1328–1330.
- Taylor, R. & Kennard, O. (1982). *J. Am. Chem. Soc.* **104**, 5063–5070.
- Thomas, K. A., Smith, G. M., Thomas, T. B. & Feldman, R. J. (1982). *Proc. Natl Acad. Sci. USA*, **79**, 4843–4847.
- Verdonk, M. L. (1995). PhD thesis, University of Utrecht, The Netherlands.
- Verdonk, M. L., Cole, J. C. & Taylor, R. (1998). In preparation.
- Verner, E. J., Oliver, B. J., Schlicksupp, L. & Natale, N. R. (1990). *Heterocycles*, **31**, 327–339.
- Yamauchi, M., Akiyama, S., Watanabe, T., Okamura, K. & Date, T. (1993). *J. Chem. Soc. Chem. Commun.* pp. 17–18.