Acta Crystallographica Section B
Structural Science
ISSN 0108-7681

Using small molecule crystal structure data to obtain information about sulfonamide conformation

Andrew Parkin,[a]* Anna Collins,[a,b] Christopher J. Gilmore[a] and Chick C. Wilson[a]

[a]WestCHEM, Department of Chemistry, University of Glasgow, University Avenue, Glasgow G12 8QQ, Scotland, and [b]EaStCHEM, School of Chemistry, University of Edinburgh, Kings Buildings, West Mains Road, Edinburgh EH9 3JJ, Scotland

Correspondence e-mail: a.parkin@chem.gla.ac.uk

Received 2 August 2007
Accepted 6 December 2007

Understanding the conformations adopted by the sulfonamide group is essential to the understanding of the way that sulfa drugs act upon the body. The relative energies of these conformations in the solid state are estimated from the Cambridge Structural Database (CSD) using cluster analysis, and are used to confirm earlier findings that many high-level *ab initio* calculations do not reproduce the observed solid-state structure. These conformational studies have been extended to the adjacent torsion angles, and it has been shown that the sulfonamide group significantly affects the form adopted. The relative energies of the observed forms in the solid state have been estimated using data available in the CSD.

1. Introduction

The sulfonamide structural motif (Fig. 1*a*) is commonly observed in small molecules: 892 structures with three-dimensional-coordinates determined in the CSD, Version 5.28 (Allen, 2002) contain it in its neutral, 'native' protonated form, and a further 315 structures contain it deprotonated as a metal ligand. It is, however, somewhat unusual; high-level quantum mechanical and *ab initio* studies (Heyd *et al.*, 1997; Liang *et al.*, 1997; Nicholas *et al.*, 1991; Bindal *et al.*, 1990) predict a different energy minimum in the gas phase than that observed in the solid state (Fig. 2). Reviews of sulfonamide crystal structures retrieved from the CSD by Bock *et al.* (1998) show that the staggered conformer is observed for all unconstrained sulfonamides. Furthermore, it is not the case that the simpler sulfonamides are likely to adopt the calculated low-energy conformation, and the more complex ones another; the crystal structure of the simplest sulfonamide to contain a C—S—N—C torsion angle (and the simplest yet studied) also exhibits the staggered conformation (Higgs *et al.*, 2002).

There is, of course, no reason why the minimum-energy conformation in the gas phase should also be the minimum-energy conformation in the solid state. Extensive hydrogen bonding and C—H···O contacts as well as crystal-packing forces may influence which conformer is actually adopted in the solid state, and the gas-phase structure can also contain intramolecular hydrogen bonds. Nonetheless, this provides the structural chemist with a significant problem – how can reliable relative energies of the different conformers in the solid state be obtained? This is especially important for such problems as crystal structure prediction. One solution is to use the information available in the CSD. Bock *et al.* (1998) have previously shown how this might be done for simple sulfonamides. Here, these results are updated and these ideas extended using the ideas of structure correlation (Bürgi & Dunitz, 1994) to look at the difference in energies of the

extended conformations of the sulfonamide group, *i.e.* of atoms that are not directly bound to the sulfonamide; it is also shown how this chemical group can affect the conformer adopted. The concept of structure correlation is not novel, and there is a wealth of knowledge in the literature (for examples and in-depth coverage of this topic see Bürgi & Dunitz, 1994). We use these concepts here to obtain information about preferred conformations when the theoretical calculations are not applicable to the solid state.

The method used to study these geometries is cluster analysis, as implemented in the *dSNAP* software (Barr *et al.*, 2005). This has previously been successfully applied to conformational analysis problems in both organic (Barr *et al.*, 2005; Collins, Barr *et al.*, 2007) and inorganic (Parkin *et al.*, 2007) systems, as well as other geometry problems including complex intermolecular interactions (Parkin *et al.*, 2006; Collins, Parkin *et al.*, 2007). The methodology will not be detailed here, as it has been described in previous publications (Barr *et al.*, 2005); in brief, the method relies on computing a Minkowski distance matrix based on the total geometry of every hit fragment, before analysing the results using cluster analysis and metric multi-dimensional scaling (MMDS). The analysis tools are highly visual and interactive, with the principal representations of the data being in the form of a dendrogram and a three-dimensional MMDS plot.

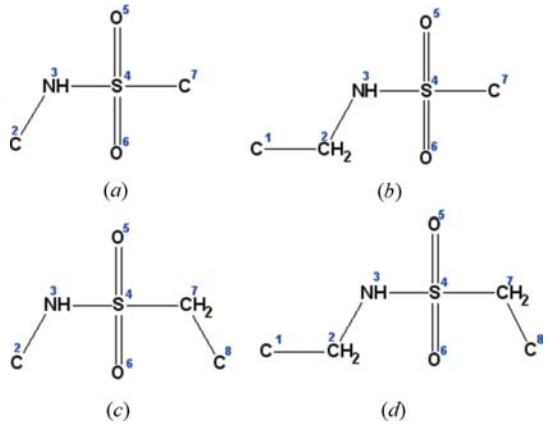

Figure 1
(*a*) The simple sulfonamide structural motif – this is substructure (i). (*b*) The sulfonamide motif extended on the nitrogen side (N-extended motif) – substructure (ii). (*c*) The sulfonamide motif extended on the sulfur side (S-extended motif) – substructure (iii). (*d*) The sulfonamide motif extended on both sides (both extended motif) – substructure (iv). The common numbering scheme used in the analysis is also shown.

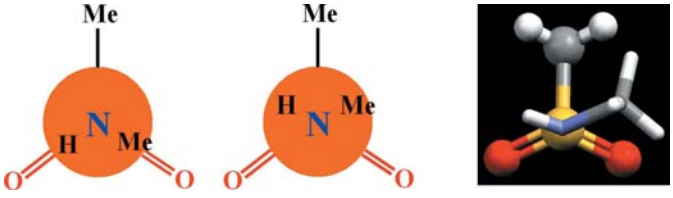

Figure 2
The theoretical minimum-energy structure for *N*-methylmethanesulfonamide (left), and the experimentally observed structure (CSD refcode MIYFEF; centre), as viewed down the N—S bond. On the right is the molecular structure taken directly from the crystal structure.

2. Experimental

Four separate sulfonamide fragments were defined as in Fig. 1. A search of the CSD (Version 5.28) yielded 1160 fragments from 892 structures for (i) (Fig. 1*a*); 160 fragments from 112 structures for (ii) (Fig. 1*b*); 32 fragments from 23 structures for (iii) (Fig. 1*c*); and a single fragment from a single structure for (iv) (Fig. 1*d*). Additional constraints applied to the search were that the S—N bond should be acyclic and that the three-dimensional coordinates of the structure were determined. In all cases the fragment possesses internal symmetry: atoms O5 and O6 are chemically identical and the two atoms are mapped onto the same region of conformation space by defining the C2—O5 distance to be greater than the C2—O6 distance.

The methods used in this paper employ the total geometry of the search fragment – that is all the interatomic distances and interatomic angles, not simply a single selected parameter or just the bonded parameters – and give excellent, well clustered results that are chemically interpretable. This geometry selection clearly includes a large amount of redundant information, but previous experience with this type of analysis has shown that this does not unduly affect the calculations and subsequent analysis. In the case of the basic

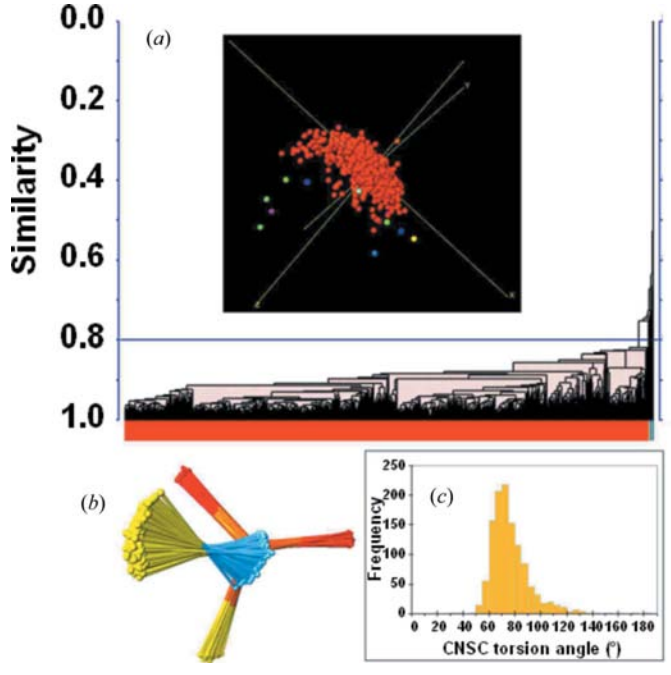

Figure 3
(*a*) Dendrogram and MMDS plot (inset) of the basic sulfonamide fragment [substructure (i)]. (*b*) Geometries of the main cluster group (*A*) superimposed using Procrustes methods; the fragments are clearly part of a continuous distribution. (*c*) Histogram of the distribution of the CNSC torsion angle.

sulfonamide fragment a single torsion angle may well provide the majority of information in which the interest lies in this case. However, as the fragments studied move to the larger, more flexible extended structures, the methods involving a full definition of the geometric parameters become increasingly powerful, quickly and efficiently revealing the differences of interest. It is noted here that any torsion angle changes are observed indirectly in *dSNAP*, as they are not included in the clustering calculation.

The interpretation of data using *dSNAP* relies heavily on reading dendrograms and MMDS plots. Each fragment in a dendrogram is represented by one of the boxes arranged along the bottom of the plot (*e.g.* Fig. 3*a*). 'Tie bars' (horizontal lines) link the fragments together according to the calculated similarity between the fragments. The vertical axis is a similarity scale, with a similarity of 0 at the top and 1.0 at the bottom. Samples joined near the top of the dendrogram are much less similar than those that join near the bottom. The fragments in a cluster, defined by the cut-level (a solid purple horizontal line; Fig. 3*a*), are identically coloured. This representation allows rapid comparison of the fragment geometries and their similarity, both within an individual cluster and within the dataset as a whole.

MMDS plotting is used independently of dendrograms; each point in this space represents a single fragment. The fragments are plotted as spheres (*e.g.* Fig. 3*a*), so that similar fragments lie close to each other and highly dissimilar fragments are large distances apart. More detailed information on reading dendrograms and MMDS plots is available as supplementary information[1] and in previous publications (*e.g.* Barr *et al.*, 2005).

3. Results and discussion

3.1. The basic sulfonamide fragment

The conformations observed for the basic sulfonamide fragment (where the N—S bond is defined as acyclic) reflect the results of Bock *et al.* (1998). The dendrogram and MMDS plot (Fig. 3*a*) clearly show a single main group *A* (colour red) with a few outlying fragments. This main group contains 930 of the 941 fragments, and these can clearly be seen to be *cis* in nature, with the N—C bond almost eclipsing the S═O bond (Fig. 3*b*). Of the outlying fragments, only one (CSD refcode YABMIX) is truly *trans* in nature, but a brief glance at the S═O (1.42 and 1.78 Å, expected ∼ 1.42 Å) and S—C (1.42 Å, expected ∼ 1.78 Å) bond lengths clearly shows that the atoms have been mis-assigned, and this fragment should fall squarely in the middle of the principal group. Only two other structures (CSD refcodes SLFNMC20 and WAFGAM) have a C—N—S—C torsion angle $\geq 140^\circ$, although neither of these are particularly close to a *trans* geometry; it is therefore not possible to estimate the solid-state energy difference between this conformation and the *cis* conformation, other than to say

[1] Supplementary data for this paper are available from the IUCr electronic archives (Reference: GP5019). Services for accessing these data are described at the back of the journal.

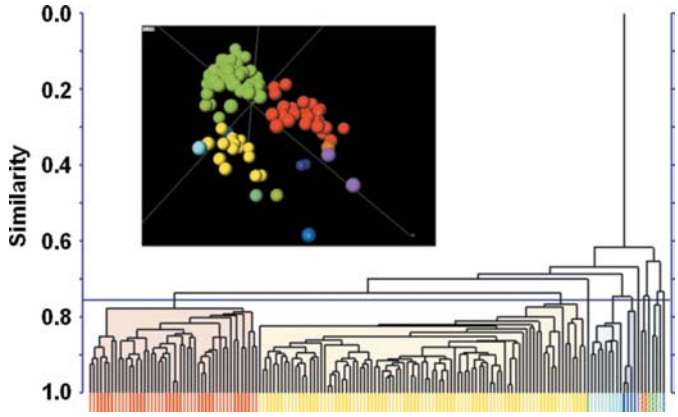

Figure 4
Dendrogram and MMDS plot (inset) from the *dSNAP* clustering calculation on fragment (ii).

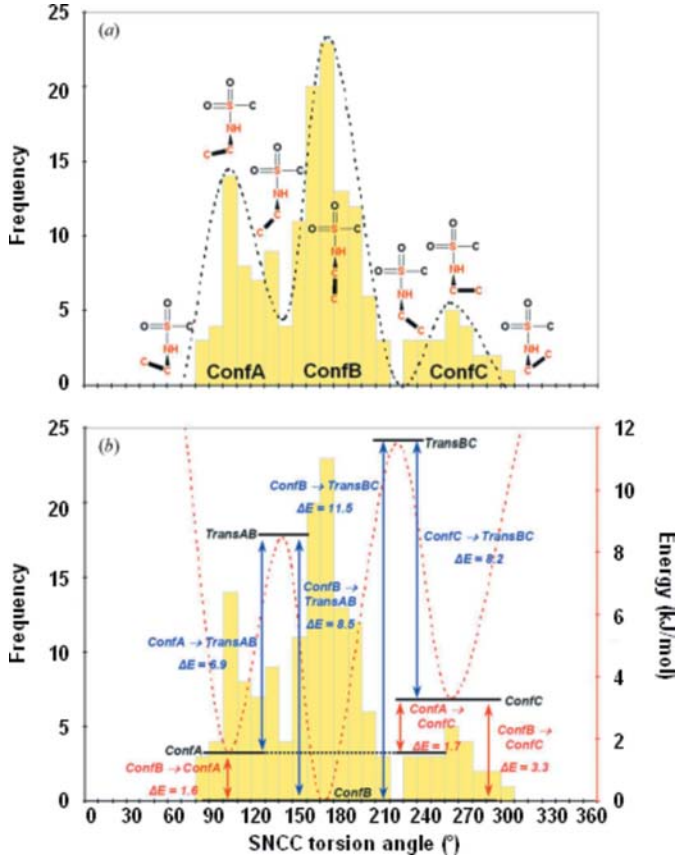

Figure 5
The distribution of the SNCC torsion angle in N-extended sulfonamide fragments observed in the CSD. (*a*) The black dotted line indicates the three approximate Gaussian distributions, labelled *ConfA*, *ConfB* and *ConfC* from left to right, with schematic structures illustrating the value for the torsion angle as if it were viewed down the C—N bond. On these schemes the atoms in the torsion of interest are shown in orange. (*b*) The red dotted line indicates the relative energies of the conformers mapped onto this distribution, as calculated using the probabilities from the CSD. The energies between the three conformations are shown in red, and the energies between these conformations and the transition states *TransAB* and *TransBC* are shown in blue.

it is clearly quite large. For the remainder of the work in this paper the C—S—N—C torsion angle was constrained to lie between 0 and 180° (and therefore the C—N—S—C torsion angle lies between −180 and 0°), thereby mapping all structures on to the same region of conformational space.

3.2. Extended N-side

A *dSNAP* clustering calculation on the 160 fragments (from 112 structures) extracted from the CSD containing substructure (ii) gives a dendrogram with seven poorly resolved clusters (Fig. 4), and a plot of the distribution of the S—N—C—C torsion angle (Fig. 5) suggests why this might be the case. A clustering calculation works most efficiently in the case of clearly distinct clusters and the distributions for the S—N—C—C torsion angle in Fig. 5 are relatively broad, as is the distribution for the C—S—N—C torsion angle (Fig. 3). Thus, although the clustering calculation works reasonably well at separating distinctly different geometries, in this case the broad, ill-defined distributions suggest that cluster analysis is not the ideal analysis tool. In such a case it is more useful to examine the torsion angle independently to see if meaningful information about the conformation can be obtained.

The distribution of the conformers around the S—N—C—C torsion angle can be divided into three main groups, all of which can be approximately mapped to a Gaussian distribution (Fig. 5). If it is assumed that the three conformers are in thermal equilibrium, then conformer *ConfB*, the most frequently occurring form, can be assumed to be the minimum-energy form; all energies here are quoted relative to the lowest-energy form, set at 0 kJ mol^{-1}. Using Boltzmann statistics based on the relative populations of the conformers (and assuming that all the crystals were grown close to room temperature), conformer *ConfA* can be estimated to be approximately 1.6 kJ mol^{-1} higher in energy than conformer *ConfB*. Conformer *ConfC* can be estimated to be approximately 3.3 kJ mol^{-1} higher in energy than conformer *ConfB* and therefore around 1.7 kJ mol^{-1} higher in energy than conformer *ConfA*. Additionally, the broad distributions allow the energy of the reaction pathway from *ConfA* to *ConfB* and from *ConfB* to *ConfC* to be approximately mapped. In going from *ConfA* to *ConfB* the smallest distribution bin contains four fragments; the barrier energy for this transformation (*TransAB*) can therefore be estimated at around 8.5 kJ mol^{-1}. From *ConfB* to *ConfC* there is one distribution bin with no fragments; the barrier energy for this transformation (*TransBC)* can be estimated at slightly greater than 11.5 kJ mol^{-1}; if a smooth transition is assumed then the energy barrier might be around 11.5–12 kJ mol^{-1}. No estimate can be made for the rotation barrier from *ConfA* to *ConfC*, or whether or not there might be another (high energy: > 12 kJ mol^{-1}) energy minimum between these two forms, other than to say that the energy of any such form is likely to be greater than 12 kJ mol^{-1} higher than that of form *ConfB*. These energies are summarized in Fig. 5.

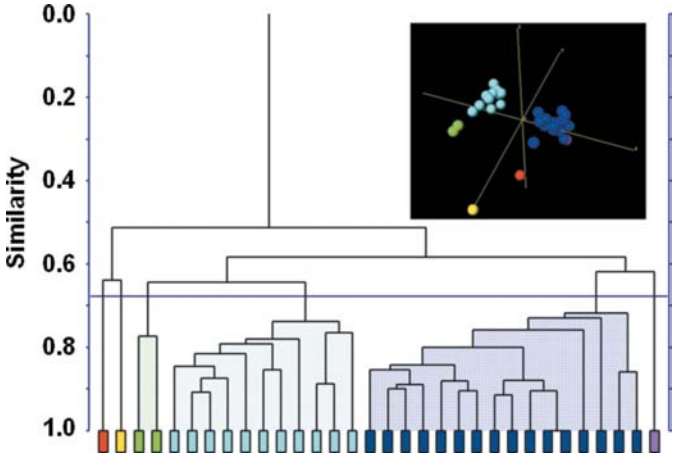

Figure 6
Dendrogram and MMDS plot (inset) from the *dSNAP* clustering calculation on substructure (iii).

3.3. Extended S-side

A *dSNAP* clustering calculation on the 32 fragments (from 23 structures) extracted from the CSD containing substructure (iii) gives a total of six clusters (Fig. 6), although only three very distinct distributions are observed in the torsion angle distribution (Fig. 7). In the dendrogram groups *A* and *B* (red and yellow) represent the fragments in conformer *ConfF*; groups *C* and *D* (green and pale blue) represent the fragments in conformer *ConfE*; and groups *E* and *F* (dark blue and

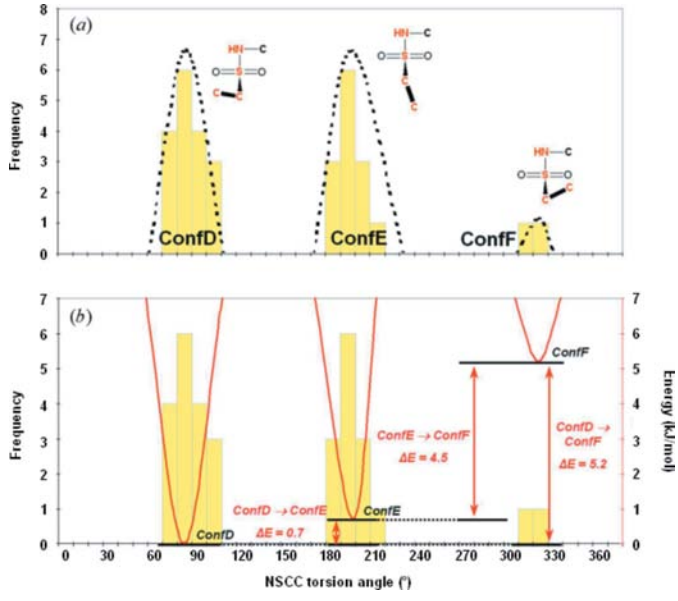

Figure 7
The distribution of the NSCC torsion angle in S-extended sulfonamide fragments observed in the CSD. (*a*) The black dotted line indicates the three approximate Gaussian distributions, labelled *ConfD*, *ConfE* and *ConfF* from left to right, with schematic structures illustrating the value for the torsion angle as if it were viewed down the C—S bond. On these schemes the atoms in the torsion of interest are shown in orange. (*b*) The red line indicates the relative energies of the conformers mapped onto this distribution, as calculated using the probabilities from the CSD. The energies between the three conformations *ConfD*, *ConfE* and *ConfF* are shown in red.

magenta) represent the fragments in conformer *ConfD*. The majority of the additional differences observed arise from variations in the basic sulfonamide (*i.e.* C—S—N—C) torsion angle; the scatterplot of the two torsion angles (Fig. 8) with the dendrogram colours mapped onto it clearly illustrates this. The only group not clearly separated out by this method is group *F*. The structure in this group (CSD refcode BIVTAB) has a number of slightly unusual bond lengths, with long C1—S2, C1—C7, S2—N3 bond lengths, and a short N3—C4 bond length. In all cases these values are only slightly outside the principal data ranges. Whether these unusual values are artefacts of poor data or real observations is difficult to tell in this case; however, the sum of these small differences results in a significant difference to the overall geometry of the fragment.

Although there are far fewer fragments in this dataset, it is again possible to examine the probability of the terminal C atom in the fragment adopting a particular conformation. On this occasion there are three very sharp distributions, with the most favourable being a *gauche* configuration, form *ConfD*. Using Boltzmann statistics form *ConfE* can be estimated to be 0.7 kJ mol^{-1} higher in energy, and form *ConfF* is estimated at a further 4.5 kJ mol^{-1} higher again. It is not possible to estimate the barriers to rotation, except to say that they are likely to be greater than 7 kJ mol^{-1} higher than the low-energy form *ConfD*. Given the sharpness of the distributions compared with those on the N-side of the functional group, these barriers might, however, be expected to be significantly greater than this.

3.4. Both extended

As there is only a single structure (CSD refcode ABEMOK) matching substructure (iv), it is impossible to draw any generalized conclusions about these extended geometries. In this structure the C—N—S—C torsion angle, at 82.7°, is slightly higher than the mode of the distribution of these angles; the N—S—C—C torsion angle lies in the region corresponding to the least common conformer (*ConfF*) with a value of 301.0°; and the S—N—C—C torsion angle places the structure in *ConfA*, but close to the transition conformation *TransAB* with a value of 128.5°.

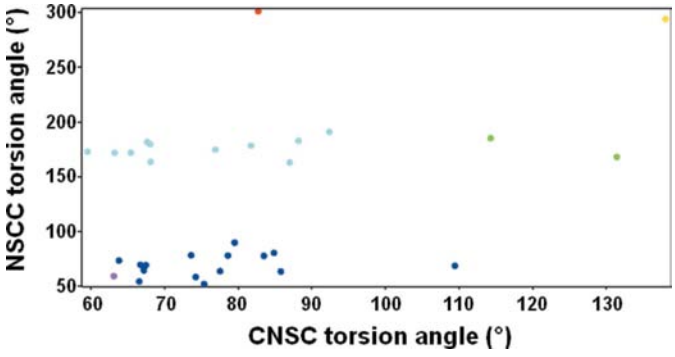

Figure 8
Plot of NSCC torsion angle against CNSC torsion angle for (iii), with dendrogram colours mapped onto the plot. The distinctions between all the groups except *F* (magenta) can clearly be seen.

4. Conclusions

It is clear from this work that there is interdependence between adjacent torsion angles in the crystal structures studied, and that information about this is readily obtainable from the CSD. It has been shown that a much larger sample set backs up the observations of Bock *et al.* (1998) that sulfonamides rarely crystallize in their calculated low-energy structure. In such circumstances the CSD often contains sufficient information to allow an estimate of the energy differences between different conformers in the solid state. This has been illustrated by studying the effect of the sulfonamide group on α-torsion angles (*i.e.* torsion angles at one remove from the functional group of interest); it is likely that it would be equally possible to study β-torsion angles in such a way, provided that a statistically significant number of samples were observed to contain the fragment of interest.

The knowledge that the minimum-energy structure of sulfonamides in the gas phase might not be identical to that observed in the solid state might cause doubt in assumptions made about their solution conformation. This is important as the sulfonamide group is an extremely important biological functional group. Several antibiotics contain this group and the first antimicrobial drugs were all based on sulfonamides, becoming widely known as 'sulfa drugs'. Understanding the structural conformation of sulfonamides is thus vital for drug design; a recent paper (Senger *et al.*, 2007) used the results of calculations on sulfonamide conformations to explain an increase in the activity of a particular potential drug compound over another. Although their assumptions *may* be correct, the calculations as described use information from the work of Bindal *et al.* (1990) that has since been shown to contain erroneous assumptions about *N*-methylmethanesulfonamide in the solid state (Higgs *et al.*, 2002). It would certainly be equally valid to use conformational data obtained from the CSD, although the conformation observed in the solution state could differ substantially from both gas-phase calculations and from observed solid-state conformations. Intermolecular interactions, and hydrogen bonding in particular, might strongly influence the conformations adopted in the solid state or in solution, whereas intramolecular interactions and hydrogen bonding will dominate the gas-phase conformation. With this in mind, it might seem reasonable in such a case to be wary of using results solely derived from calculations to predict and explain interactions in the solution phase if they are not backed up by experimental observations. Equally, solid-state information extracted from the CSD should be compared with solution data before being used in these circumstances.

References

Allen, F. H. (2002). *Acta Cryst.* B**58**, 380–388.

Barr, G., Dong, W., Gilmore, C. J., Parkin, A. & Wilson, C. C. (2005). *J. Appl. Cryst.* **38**, 833–841.
Bindal, R. D., Golab, J. T. & Katzenellenbogen, J. A. (1990). *J. Am. Chem. Soc.* **112**, 7861–7868.
Bock, H., Nagel, N. & Naether, C. (1998). *Z. Naturforsch. B*, **53**, 1389–1400.
Bürgi, H.-B. & Dunitz, J. D. (1994). *Structure Correlation*, Vols. 1 and 2. New York: VCH Publishers, Inc.
Collins, A., Barr, G., Dong, W., Gilmore, C. J., Middlemiss, D. S., Parkin, A. & Wilson, C. C. (2007). *Acta Cryst.* B**63**, 469–476.
Collins, A., Parkin, A., Barr, G., Dong, W., Gilmore, C. J. & Wilson, C. C. (2007). *CrystEngComm*, **9**, 245–253.
Heyd, J., Thiel, W. & Weber, W. (1997). *J. Mol. Struct. Theochem.* **391**, 125–130.
Higgs, T. C., Parkin, A., Parsons, S. & Tasker, P. A. (2002). *Acta Cryst.* E**58**, o523–o525.
Liang, G., Bays, J. P. & Bowen, J. P. (1997). *J. Mol. Struct. Theochem.* **401**, 165–179.
Nicholas, J. B., Vance, R., Martin, E., Burke, B. J. & Hopfinger, A. J. (1991). *J. Phys. Chem.* **95**, 9803–9811.
Parkin, A., Barr, G., Collins, A., Dong, W., Gilmore, C. J., Tasker, P. A. & Wilson, C. C. (2007). *Acta Cryst.* B**63**, 612–620.
Parkin, A., Barr, G., Dong, W., Gilmore, C. J. & Wilson, C. C. (2006). *CrystEngComm*, **8**, 257–264.
Senger, S., Chan, C., Convery, M. A., Hubbard, J. A., Shah, G. P., Watson, N. S. & Young, R. J. (2007). *Bioorg. Med. Chem. Lett.* **17**, 2931–2934.