Journal of Computer-Aided Molecular Design, 10 (1996) 247–254
ESCOM

J-CAMD 335

Comparison of conformer distributions in the crystalline state with conformational energies calculated by ab initio techniques

Frank H. Allen[a], Stephanie E. Harris[b] and Robin Taylor[a,c,*]

[a]*Cambridge Crystallographic Data Centre, 12 Union Road, Cambridge CB2 IEZ, U.K.*
[b]*Department of Pharmaceutical Sciences, University of Aston, Birmingham B4 7ET, U.K.*
[c]*Zeneca Agrochemicals, Jealott's Hill Research Station, Bracknell, Berks RG12 6EY, U.K.*

Received 18 September 1995
Accepted 8 January 1996

Keywords: Ab initio; Cambridge Structural Database; Conformational energy

Summary

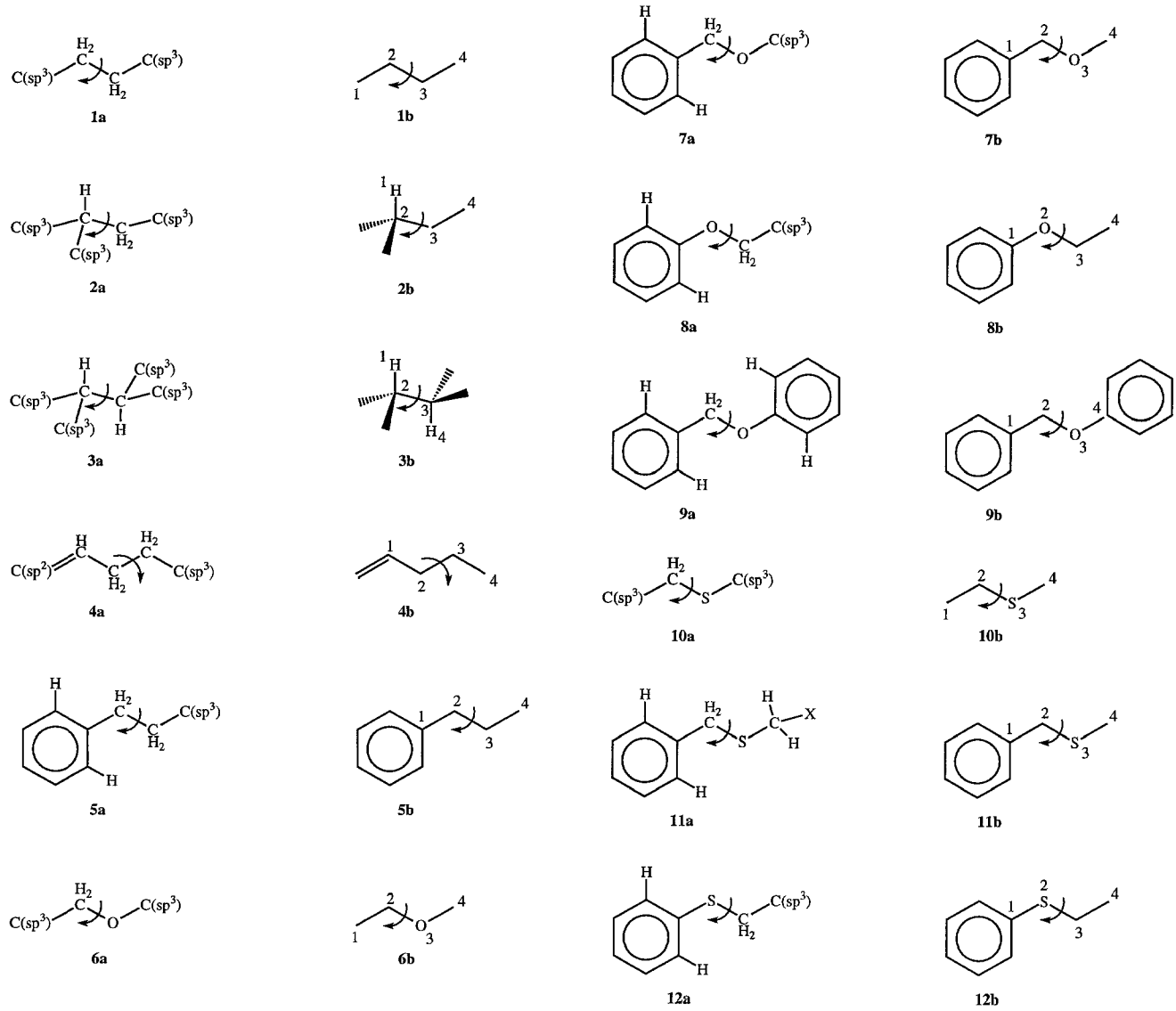
The conformational preferences of 12 molecular substructures in the crystalline state have been determined and compared with those predicted for relevant model compounds by ab initio molecular orbital calculations. Least-squares regression shows that there is a statistically significant correlation between the crystal-structure conformer distributions and the calculated potential-energy differences, even though the calculations relate to a gas-phase environment. Torsion angles associated with high strain energy (> 1 kcal mol^{-1}) appear to be very unusual in crystal structures and, in general, high-energy conformers are underrepresented in crystal structures compared with a gas-phase, room-temperature Boltzmann distribution. It is concluded that crystal packing effects rarely have a strong systematic effect on molecular conformations. Therefore, the conformational distribution of a molecular substructure in a series of related crystal structures is likely to be a good guide to the corresponding gas-phase potential energy surface.

Introduction

Despite their importance in molecular modelling, theoretical methods of conformational analysis have significant limitations. Molecular mechanics suffers from incomplete parameterisation and gives unreliable results for polar molecules because of the inadequacy of classical formulae when applied to electrostatic interactions on a molecular scale [1]. Semi-empirical molecular orbital (MO) methods are also prone to incomplete parameterisation. In addition, each of the common semi-empirical Hamiltonians has known limitations; for example, MNDO is unable to reproduce hydrogen bonding accurately and AM1 gives poor results for some phosphorus-containing molecules [2]. Ab initio MO calculations are computationally intensive at the higher basis set levels, while more limited basis sets are often inadequate at representing, for example, hypervalent compounds and complexes with dative bonds [3]. Moreover, the Hartree–Fock Self-Consistent Field (HF/SCF) method is unable to reproduce the effects of electron correlation [4]. All theoretical methods are difficult to apply to metal complexes (although density functional techniques [5] show promise) and calculations generally pertain to an in vacuo situation and are thus of uncertain relevance to conformational preferences in condensed phases.

It is therefore useful to supplement theoretical conformational predictions by examining experimentally observed conformational preferences in the crystalline state. However, a common cause of concern is the fact that crystal packing forces may exert a strong effect on molecular conformations. This could be manifested in two ways. Firstly, packing forces in any given crystal structure might act to stabilise a high-energy molecular conformation. Secondly, and more insidiously, crystal packing effects in a series of related structures might act *systematically* to favour intramolecular geometries which, although low in energy, do not correspond to the gas-phase global minimum. The classic example of this is the biphenyl substructure, which tends to crystallise in an approximately planar geometry when there are no substituents in the ortho positions [6], despite the fact that the inter-ring

*To whom correspondence should be addressed.



Scheme 1. Substructures (**1a**–**12a**) for CSD searches and corresponding model compounds (**1b**–**12b**) for ab initio calculations.

dihedral angle of biphenyl itself is optimally 44° in the vapour phase [7]. Presumably, this occurs because planar biphenyls can pack very efficiently. Thus, in this case, the crystal-structure data give a misleading impression of conformational proclivities in the vapour phase.

On the other hand, there are several examples of good (albeit qualitative) correlations between crystal-structure geometry distributions and theoretical, gas-phase conformational energies [8–15]. For example, Gilli et al. [15] showed that crystal-structure geometries of the $R_1(X=)$-$C(sp^2)$-$N(sp^3)R_2R_3$ molecular fragment lie along a valley on the calculated (molecular mechanics) potential-energy surface. Klebe and Mietzner [16] have even exploited this sort of relationship for fast conformational searching, although they were at pains to emphasise the empirical nature of their procedure.

In summary, the evidence from the literature is equivocal: it is unclear whether the biphenyl case is a rare exception or merely one example of a common phenomenon. In this study, we have therefore attempted to characterise more fully the strength of correlation between crystal-structure conformer distributions and the potential-energy surfaces of isolated molecules. We have taken 12 molecular substructures (**1a**–**12a**) and determined their conformational preferences in the Cambridge Structural Database (CSD). These data have then been compared with the torsional energy profiles of the corresponding (and simplest possible) model compounds **1b**–**12b**, calculated by ab initio MO techniques.

Methods

Geometrical parameters

For each substructure or model compound, our attention focussed on the torsional preferences around the 'central' bond, indicated by curved arrows in the struc-

tures of **1a–12a** and **1b–12b** depicted in Scheme 1. In all of the model compounds, there is one symmetric torsional minimum and two symmetry-related asymmetric minima (we prefer the terms 'symmetric' and 'asymmetric' to 'anti' and 'gauche', because they are unambiguous for all the systems considered in this work). The geometry around the central bond is defined by the torsion angle 1-2-3-4 (the 'central torsion angle'), the atom numbers being indicated in structures **1b–12b** (Scheme 1). For each molecule/substructure, the central torsion angle is 180° at the symmetric minimum and around +60° and −60° at the two asymmetric minima.

CSD searches

Substructure searches, geometry calculations and data analysis were performed in the CSD (v. 5.7, April 1994) [17], using the programs QUEST3D and VISTA [18]. Substructures were specified as shown in Scheme 1 for **1a–12a**. The sp^3 carbon atoms of **1a–10a** and **12a** and CH_2X of **11a** were constrained to be in acyclic environments. Bond length and bond angle constraints were imposed on the substructures as follows: **1a–5a**, **7a–12a**: $C(sp^3/aromatic)$-$C(sp^3)$ = 1.44–1.62 Å; **4a**: $C(sp^2)$-$C(sp^2)$ = 1.22–1.40 Å and $C(sp^2)$-$C(sp^3)$ = 1.42–1.60 Å; **6a**: $C(sp^3)$-$C(sp^3)$ = 1.42–1.60 Å; **6a–9a**: $C(sp^3)$-O = 1.34–1.50 Å; **8a**, **9a**: C(aromatic)-O = 1.30–1.46 Å; **10a–12a**: $C(sp^3)$-S = 1.75–1.93 Å; **12a**: C(aromatic)-S = 1.72–1.90 Å; **2a**,**3a**: $C(sp^3)$-C-$C(sp^3)$ = 100°–130°.

Secondary search constraints were imposed by use of the CSD bit-screen mechanism in order to locate only those entries that: (i) had error-free atomic coordinates, according to CSD criteria; (ii) had no reported disorder in the crystal structure; (ii) had a crystallographic R-factor <0.10; (iv) were classified in the CSD as 'organic' compounds; and (v) were not classified as polymeric, according to CSD criteria. The total numbers of crystallographically independent observations found for **1a–12a** were, respectively: 2865, 280, 17, 58, 57, 352, 71, 181, 31, 133, 13 and 11.

Absolute values of the torsion angles around the central bond were calculated for all the substructures, i.e., structures were inverted where necessary so that torsion angles fell in the range 0°–180°. For **2a** and **3a**, the required torsion angles involve hydrogen atoms and were estimated indirectly from the $C(sp^3)$-C-C-$C(sp^3)$ torsion angles; this procedure was necessary because hydrogen atom coordinates were not reported for all structures. Conformations were considered to be symmetric (asymmetric) if the central torsion angle was within 40° of 180° (60°) although, in practice, the vast majority were much closer than this to the idealised values.

Ab initio molecular orbital calculations

Ab initio MO calculations were performed on Silicon Graphics workstations, using the program GAMESS-UK [19] and the HF/SCF method. For each model compound, an initial torsional energy scan was performed in which the central torsion angle was rotated from 0° to 180° in 10° steps. At each point on the scan, all geometrical parameters were allowed to relax except for the central torsion angle. The STO-3G basis set was used.

The geometries of the asymmetric and symmetric conformers were then fully optimised at both the STO-3G and 3-21G levels. The central torsion angle invariably converged to 180°±0.1° in the symmetric conformers, but often diverged significantly from +60° in the asymmetric conformers. For molecules with more than one rotatable torsion (e.g. **7b**), several starting geometries were used to ensure that the lowest minima were found.

Finally, the energies of the symmetric and asymmetric conformers were computed at the 6-31G*//STO-3G and 6-31G*//3-21G levels. The results were used to calculate the energy difference between the two conformers, ΔE:

$$\Delta E = E_{asym} - E_{sym} \quad (1)$$

where E_{asym} is the lower of the 6-31G*//STO-3G and 6-31G*//3-21G energies of the asymmetric conformer, and a corresponding definition applies to E_{sym}.

Statistical calculations

Linear least-squares regressions were done using the partial least squares (PLS) function in the QSAR module of SYBYL [20]. In all cases, regressions were performed without cross-validation using one PLS component, making them formally equivalent to a simple linear regression.

Results and Discussion

Qualitative comparison of calculated energy curves and observed torsion angle distributions

Figure 1 shows histograms of the observed distributions in the CSD of the central torsion angles for **1a–12a**. In view of the symmetry of the substructures, absolute values of the torsion angles are used throughout. The histograms are superimposed on the calculated STO-3G torsional profiles of the corresponding model compounds **1b–12b**. All histograms are normalised to the same area and all torsional profiles use the same scale on the energy axis.

There is obvious complementarity between the histograms and the energy curves. The latter, as expected, have two minima, corresponding to the symmetric (anti) minimum at 180° and the asymmetric (g^+) minimum at approximately 60°. Correspondingly, the observed distributions are generally bimodal, with the crystal-structure conformations congregating in the potential energy wells. However, when the calculated barrier between two energy wells is low, the crystal-structure torsion angles span the minima (e.g. **10a**,**b**).

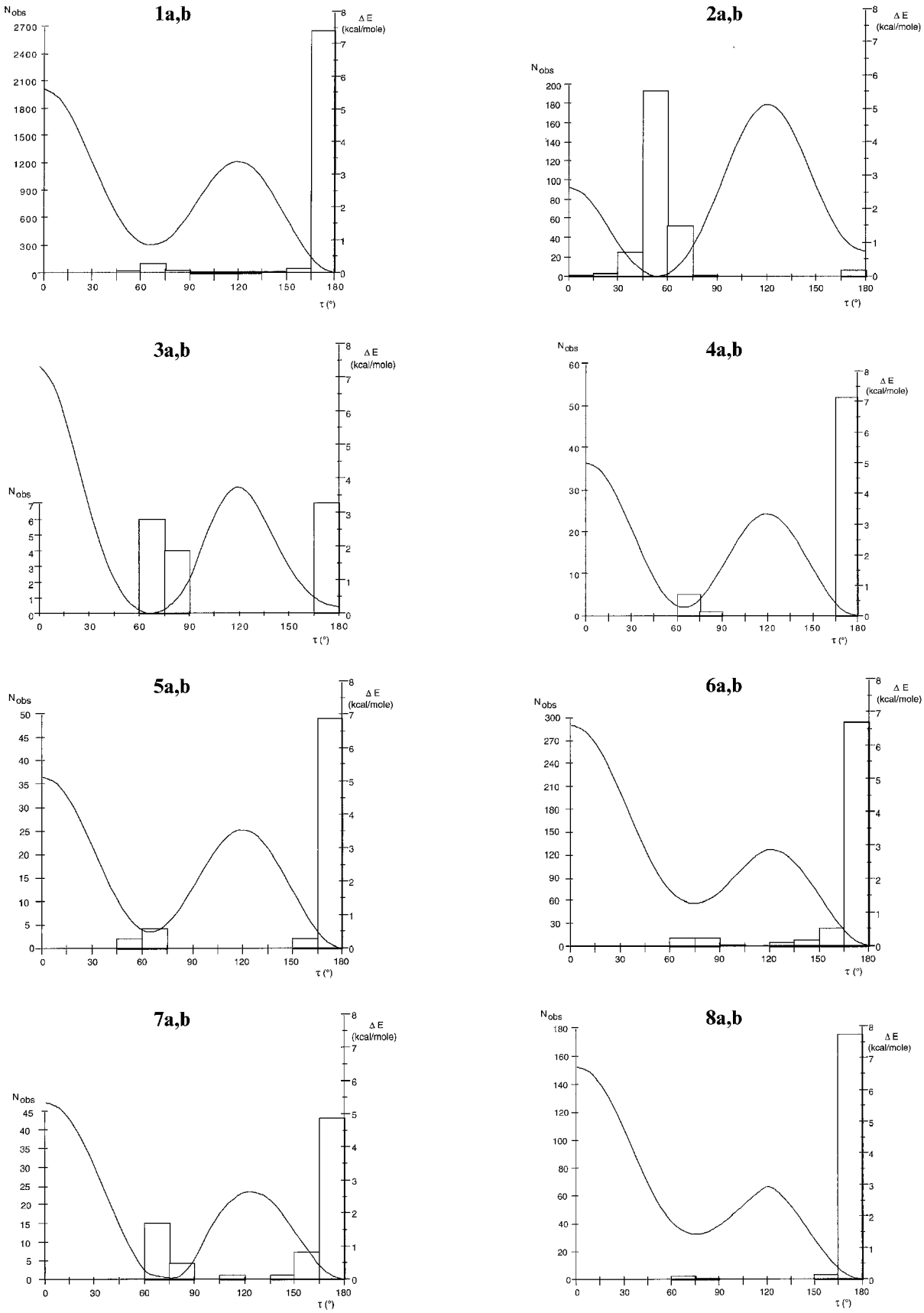


Fig. 1. Histograms of observed torsion angle distributions of **1a–12a,** with calculated STO-3G potential-energy curves of **1b–12b** superimposed All histograms are normalised to the same area and the same energy scale is used for all potential-energy curves.

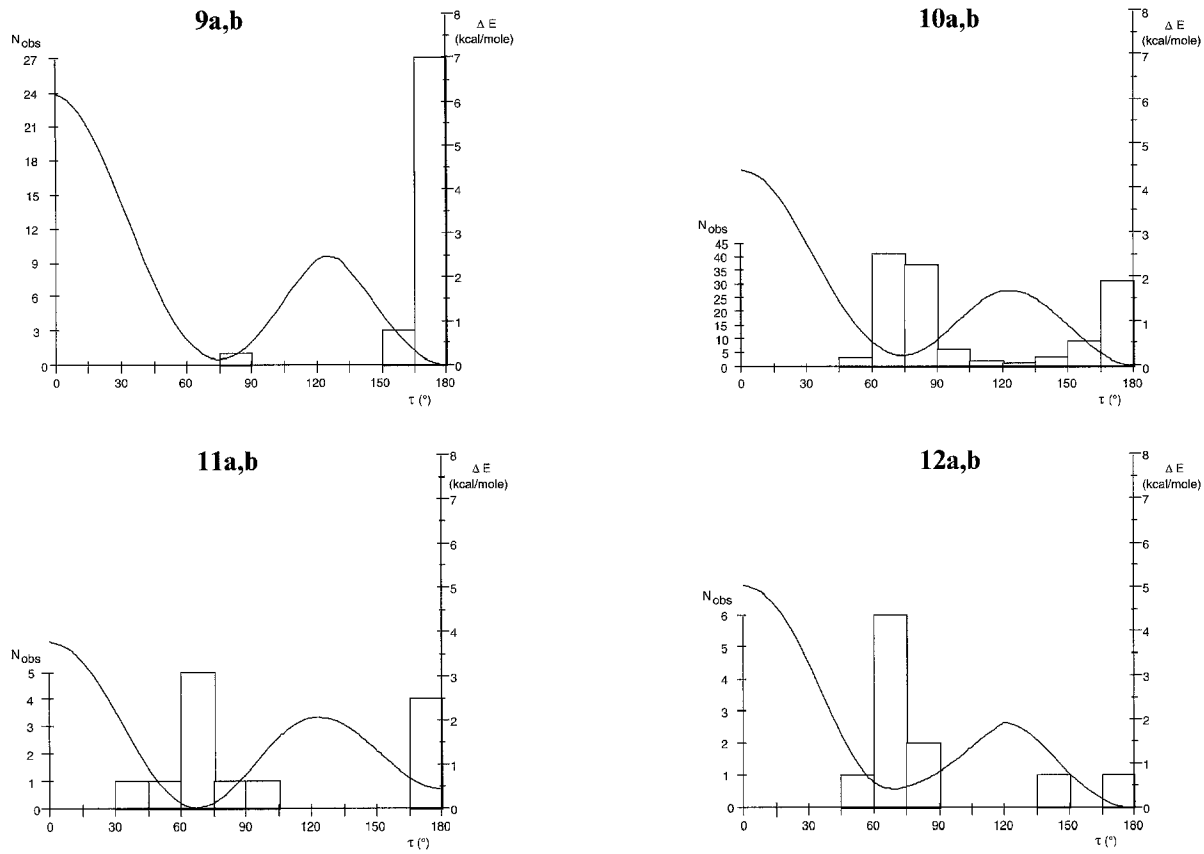


Fig. 1. (continued).

The most striking feature of Fig. 1 is the almost complete absence of highly strained torsions in the crystal structures. Based on the STO-3G energy curves, only 3% of the crystallographically observed torsion angles have strain energies exceeding 1 kcal mol^{-1}. This is particularly remarkable as the MO calculations were performed on the least sterically crowded model compounds possible, **1b–12b** being derived by adding hydrogen atoms to all unfilled valences of the corresponding substructures.

Comparison between observed and calculated torsion angles of asymmetric conformers

Table 1 summarises the optimised (3-21G) torsion angles of the asymmetric conformers of **1b–12b** (τ_{opt}), together with the average torsion angles of the asymmetric conformers of **1a–12a** in the CSD (τ_{av}). The two sets of values are plotted against each other in Fig. 2. Linear least-squares regression, forcing the intercept through the origin, gives:

$$\tau_{av} = 1.02 \cdot \tau_{opt} \quad (2)$$

with F(1,10) = 22.305 and statistical significance = 0.001, the Pearson correlation coefficient being 0.831. Thus, the average torsion angles of the asymmetric conformers from the CSD correlate very well with the calculated optimum values. A similar analysis for the symmetric conformers is precluded because of the discontinuity in torsion angles at 180°.

Correlation between conformer ratios in the CSD and calculated energy differences

Table 2 gives the 6-31G* calculated energies of the symmetric and asymmetric conformers of the model com-

TABLE 1
CALCULATED AND OBSERVED AVERAGE TORSION ANGLES OF ASYMMETRIC CONFORMERS

Structure	τ_{opt} (°)	τ_{av} (°)	N_{obs}
1a,b	65.2	69.4	149
2a,b	55.7	54.1	269
3a,b	66.5	72.1	10
4a,b	67.3	70.7	6
5a,b	66.2	64.1	6
6a,b	75.3	79.0	22
7a,b	75.5	71.1	19
8a,b	78.6	74.2	3
9a,b	84.1	82.0	1
10a,b	72.9	76.8	87
11a,b	60.5	67.8	9
12a,b	66.4	67.8	9

τ_{opt} = calculated (3-21G) optimum torsion angle of asymmetric conformer. τ_{av} = observed mean torsion angle of asymmetric conformer in crystal structures. N_{obs} = number of crystallographic observations contributing to τ_{av}.

TABLE 2
AB INITIO (6-31G*) ENERGIES OF SYMMETRIC AND ASYMMETRIC CONFORMERS OF MODEL COMPOUNDS

Molecule	Symmetric conformer		Asymmetric conformer		ΔE
	Energy	Geometry	Energy	Geometry	
1b	−157.29807	3-21G	−157.29657	STO-3G	0.94
2b	−196.32999	3-21G	−196.33142	3-21G	−0.90
3b	−235.36262	3-21G	−235.36284	3-21G	−0.14
4b	−195.14064	3-21G	−195.14005	3-21G	0.37
5b	−347.80939	3-21G	−347.80844	3-21G	0.60
6b	−193.10246	3-21G	−193.09991	3-21G	1.60
7b	−383.61089	3-21G	−383.61132	3-21G	−0.27
8b	−383.62171	3-21G	−383.61900	3-21G	1.70
9b	−574.12978	3-21G	−574.12955	3-21G	0.14
10b	−515.76964	STO-3G	−515.76892	STO-3G	0.45
11b	−706.28015	STO-3G	−706.28008	STO-3G	0.04
12b	−706.27810	3-21G	−706.27781	3-21G	0.18

Energy is the energy of conformer (in hartrees) at the 6-31G*//3-21G or 6-31G*//STO-3G level. Geometry refers to the basis set used for geometry optimisation. $\Delta E = E_{asym} - E_{sym}$ (in kcal mol^{-1}).

pounds, together with the (asymmetric − symmetric) energy difference, ΔE. Table 3 lists ΔE alongside the ratio C:

$$C = N_{sym} / N_{asym} \quad (3)$$

where N_{sym} and N_{asym} are, respectively, the observed numbers of symmetric and asymmetric conformers in the CSD. Figure 3 shows a plot of ΔE against ln(C). Despite considerable scatter, a correlation is evident. This is confirmed by least-squares regression, which gives:

$$\Delta E = 0.20 + 0.23 \cdot \ln(C) \quad (4)$$

with F(1,10) = 12.146 and statistical significance = 0.006, the Pearson correlation coefficient being 0.740. Some of the scatter in Fig. 3 is probably due to errors in the calculated energy differences and, in this respect, it is worth noting that a few of the energy differences showed appreciable variation when the basis sets for geometry optimisation and for final energy calculation were varied from STO-3G through 3-21G to 6-31G*. In the most extreme example, the (asymmetric − symmetric) energy difference of **9b** varied from −1.9 to +0.3 kcal mol^{-1}. Another reason for scatter is that some of the C ratios are based on small numbers of crystal structures. In particular, there is only one example of an asymmetric conformer for **9a** and one of a symmetric conformer for **12a.** This makes the C values relatively ill determined for these systems. If they are omitted from the analysis, a least-squares regression on the remaining 10 observations gives:

$$\Delta E = 0.16 + 0.30 \cdot \ln(C) \quad (5)$$

with F(1,8) = 22.175 and statistical significance = 0.002, the Pearson correlation coefficient being 0.857. This least-squares line is shown in Fig. 3, with the two omitted observations being designated by circles rather than crosses.

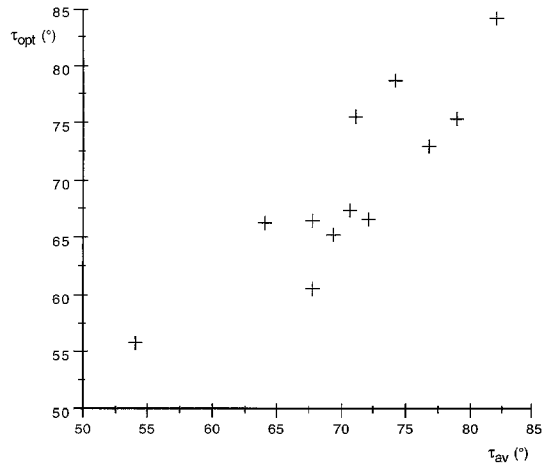


Fig. 2. Observed mean torsion angles of asymmetric conformers of **1a**–**12a** (τ_{av}) plotted against computed (3-21G) optimum torsion angles of asymmetric conformers of **1b**–**12b** (τ_{opt}).

TABLE 3
CRYSTAL-STRUCTURE CONFORMER RATIOS AND CALCULATED CONFORMATIONAL ENERGY DIFFERENCES

Structure	N_{sym}	N_{asym}	C	ln(C)	ΔE
1a,b	2708	149	18.17	2.900	0.94
2a,b	6	269	0.02	−3.803	−0.90
3a,b	7	10	0.70	−0.357	−0.14
4a,b	52	6	8.67	2.159	0.37
5a,b	51	6	8.50	2.140	0.60
6a,b	323	22	14.68	2.687	1.60
7a,b	51	19	2.68	0.987	−0.27
8a,b	178	3	59.33	4.083	1.70
9a,b	30	1	30.00	3.401	0.14
10a,b	41	87	0.47	−0.752	0.45
11a,b	4	9	0.44	−0.811	0.04
12a,b	1	9	0.11	−2.197	0.18

N_{sym} (N_{asym}) is the number of symmetric (asymmetric) conformers observed in the CSD. $C = N_{sym} / N_{asym}$. $\Delta E = E_{asym} - E_{sym}$ (in kcal mol^{-1}).

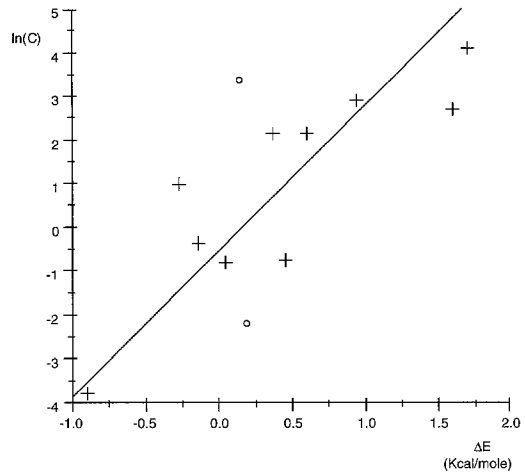


Fig. 3. Plot of ln(C) (Eq. 3) against the calculated energy difference $E_{asym} - E_{sym}$. Points based on only a few observations (corresponding to substructures **9a**, **12a**) are shown as circles rather than crosses. The least-squares line is fitted to the points shown as crosses.

Although Eq. 5 is a purely empirical relationship, with no theoretical justification [21], it is interesting to compare it with the expected gas-phase Boltzmann distribution of conformers. In the gas phase:

$$N_{sym} = c \cdot \exp[-E_{sym}/(RT)] \quad (6)$$

$$N_{asym^+} = N_{asym^-} = c \cdot \exp[-E_{asym}/(RT)] \quad (7)$$

where c is a constant, R is the gas constant, T is the temperature and N_{asym^+} and N_{asym^-} are the observed numbers of g^+ and g^- conformers, respectively. Summing the latter two quantities to get N_{asym} and combining Eqs. 6 and 7 gives:

$$N_{sym}/N_{asym} = \exp[-E_{sym}/RT] / 2 \cdot \exp[-E_{asym}/RT] \quad (8)$$

and thus:

$$\Delta E = RT\ln(2) + RT\ln(C) \quad (9)$$

Since Eq. 9 has the same functional form as Eq. 5, we can use both the intercept and the slope of Eq. 5 to estimate values for the 'temperature', T. The values obtained are 119 and 150 K. It is not correct to regard T as a true temperature. However, the result does reveal an important trend: high-energy conformers are underrepresented in crystal structures compared with a gas-phase Boltzmann distribution at room temperature.

Conclusions

The results presented above show that conformational preferences in the crystalline state are significantly correlated with calculated conformational energy differences. However, high-energy conformers appear to be less common in crystal structures than would be expected in a room-temperature, gas-phase Boltzmann population. Thus, torsion angles associated with a strain energy > 1 kcal mol^{-1} in our calculations on **1b–12b** are observed in only 3% of the examples of the corresponding substructures **1a–12a** in the CSD. Furthermore, owing to our use of the least sterically hindered model compounds in the ab initio MO calculations, this figure is likely to be an overestimate of the actual percentage of highly strained torsion angles in crystal structures.

Despite their statistical significance, it is not suggested that equations such as Eqs. 4 and 5 be used to estimate conformational energy differences. There are three reasons. Firstly, Eq. 5 is a purely empirical relationship with no underlying theoretical foundation [21]. Secondly, the C ratio is likely to be ill determined if there are few crystallographic observations of the less common conformer. Thirdly, there is at least one well-documented example of a substructure with geometrical preferences that are different in the crystalline state compared to the gas phase, because of a systematic crystal packing effect (the ortho-unsubstituted biphenyl substructure, see Introduction). However, the present study suggests that such cases are rare: in general, the effects of crystal packing forces on molecular conformations are neither large nor systematic, so crystal-structure geometries will be a good indication of intrinsic conformational preferences.

The question arises: when can an exception to this general rule be expected? An obvious example is when a molecule is capable of strong intermolecular interactions, which may compete with intramolecular forces to induce a strained conformation. However, this possibility is not confined to small-molecule crystal structures but exists in any condensed medium, e.g. within a protein binding site. The other obvious situation is when a molecule can adopt a completely flat geometry with little increase in conformational strain energy, as exemplified by the ortho-unsubstituted biphenyl substructure [6].

Acknowledgements

We thank Dr. Russell Viner (Zeneca Agrochemicals, Bracknell, U.K.) for helpful discussions.

References

1 Allinger, N.L., Adv. Phys. Org. Chem., 13 (1976) 1.
2 Stewart, J.J.P., J. Comput.-Aided Mol. Design, 4 (1990) 1.
3 Hehre, W.J., Radom, L., Schleyer, P. v. R. and Pople, J.A., Ab initio Molecular Orbital Theory, Wiley, New York, NY, 1986.
4 Atkins, P.W., Molecular Quantum Mechanics, 2nd ed., Oxford University Press, Oxford, U.K., 1983.
5 Labanowski, J. and Andzelm, J. (Eds.) Density Functional Methods in Chemistry, Springer, Heidelberg, 1991.

6 Brock, C.P. and Minton, R.P., J. Am. Chem. Soc., 111 (1989) 4586.
7 Almenningen, A., Bastiansen, O., Fernholt, L., Cyvin, B.N., Cyvin, S.J. and Samdal, S., J. Mol. Struct., 128 (1985) 59.
8 Kalman, A., Czugler, M. and Argay, G., Acta Crystallogr., B37 (1981) 868.
9 Van Konigsveld, H. and Jansen, J.C., Acta Crystallogr., B40 (1984) 420.
10 Ashida, T., Tsunogae, Y., Tanaka, I. and Yamane, T., Acta Crystallogr., B43 (1987) 212.
11 Ianelli, S., Musatti, A., Nardelli, M., Benassi, R., Folli, U. and Taddei, F., J. Chem. Soc., Perkin Trans. 2, (1992) 49.
12 Setzer, W.N. and Bentrude, W.G., J. Org. Chem., 56 (1991) 7212.
13 Nyburg, S.C. and Faerman, C.H., J. Mol. Struct., 140 (1986) 347.
14 Mullaley, A. and Taylor, R., J. Comput.-Aided Mol. Design, 8 (1994) 135.
15 Gilli, G., Bertolasi, V., Bellucci, F. and Ferretti, V., J. Am. Chem. Soc., 108 (1986) 2420.
16 Klebe, G. and Mietzner, T., J. Comput.-Aided Mol. Design, 8 (1994) 583.
17 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. and Watson, D.G., J. Chem. Inf. Comput. Sci., 31 (1991) 187.
18 CSD User Manual, Cambridge Crystallographic Data Centre, Cambridge, U.K., 1992.
19 Guest, M.F. and Sherwood, P., GAMESS User's Guide, SERC Daresbury Laboratory, Daresbury, U.K., 1990.
20 SYBYL, Tripos Associates, St. Louis, MO.
21 Bürgi, H.B. and Dunitz, J.D., Acta Crystallogr., B44 (1988) 445.