Acta Cryst. (1973). A**29**, 148

Probability-Plot Comparison of Crystallographically Independent Molecules

BY WILSON H. DE CAMP*

Division of Biological Sciences, National Research Council of Canada, Ottawa, Canada K1A 0R6

(*Received* 14 *August* 1972; *accepted* 16 *October* 1972)

Molecules may be described in terms of chemical coordinates (bond lengths, bond angles and torsion angles) or bonded and non-bonded interatomic distances, as well as by crystallographic coordinates. For studies of molecular geometry, the former is more generally useful. Since random errors in interatomic distances should be normally distributed, they may be used in probability-plot analyses of differences in molecular geometry. Several examples of such analyses are discussed. The method is shown to be applicable to entire molecules or molecular fragments, even when the chemical or crystallographic environment is different. The technique is sensitive to minor changes in conformation which might otherwise be overlooked.

Abrahams & Keve (1971) introduced the application of probability-plot analysis to the comparison of independent measurements of crystallographic quantities. They showed that both normal and half-normal plots are of value, the former being applicable to differences between magnitudes, possibly modified by a scale factor (such as structure amplitudes), and the latter to signed quantitites (such as atomic positional parameters). However, a second set of structure-factor data on the same compound is necessary in either case.

Occasionally a molecule may be studied crystallographically in two different chemical or crystalline environments, yet it appears to have the same conformation in both forms. Although estimated standard deviations (e.s.d.'s) may be calculated for the various chemical coordinates (*i.e.* bond lengths, bond angles and torsion angles), the assumptions upon which they are based frequently cast doubt upon the reasonableness of the values. A method is described here by which probability-plot analysis may be applied to interatomic distances in comparing independently determined molecular geometries.

In general, a molecule of n atoms may be described by $3n-6$ chemical coordinates. Considering interatomic distances between non-bonded atoms to be equivalent to bond angles or torsion angles (Dunitz & Waser, 1972), comparisons may be made of the geometry of chemically identical molecules (or portions of molecules). If systematic error is absent, then the differences in their chemical coordinates or the equivalent interatomic distances should be normally distributed. Probability-plot analysis may then be applied to the distances between corresponding atoms without regard for differences in chemical of crystalline environment. For most organic molecules, no interatomic distance corresponding to any chemical coordinate will exceed three times the single-bond distance, so that the use of all interatomic distances to a limit of 4·65 Å seems appropriate. Inclusion of an adequate number of nonbonded distances appears to be important, however, since these values are most sensitive to conformational changes.

A computer program to carry out the analysis has been written in Fortran IV for the IBM 360/67 computer. It is readily adaptable to other machines, and may be obtained from the author. Several sets of published experimental results have been analysed with the program, and the results are discussed below.

Example 1

The crystal structure of 3-methyl-3-pyrazolin-5-one has been determined by De Camp & Stewart (1971). Independent solutions based on stationary-crystal stationary-counter and moving-crystal moving-counter measurements showed no apparent difference in molecular geometry. Half-normal probability plots are shown for the atomic positional parameters in Fig. 1(*a*), and for the interatomic distances to a limit of 4·65 Å in Fig. 1(*b*). The latter case involves a greater number of values, and thus would appear to be statistically more reliable. The essential similarity of the two plots is supported by the equations of the least-squares line through points for which δp, the half-normal order statistic defined by Abrahams & Keve (1971), is less than 2·0, thus excluding the extreme values. The slope of the lines suggests that the e.s.d.'s are slightly underestimated.

Example 2

Betaprodine (1,3-dimethyl-4-phenylpiperidin-4-yl propionate) has been studied both as the hydrochloride and hydrobromide (Ahmed & Barnes, 1963; Ahmed, Barnes & Masironi, 1963). On the basis of bond lengths and angles, and the dihedral angle between the phenyl and piperidine rings, no significant differences in conformation were reported. Fig. 1(*c*) is a half-normal probability plot comparison of the two molecules, using all interatomic distances $<4{\cdot}65$ Å. Although the central points lie nearly on a straight line, several extreme points deviate markedly from linearity, with two points being more than 15 units off the plot grid. Of the 31 values for which $\delta p > 3{\cdot}0$, 23 represent distances to atoms in the propionate group.

* N.R.C. Postdoctorate Fellow, 1970-1972.

Calculation of torsion angles showed that the carbonyl group was anticlinal to the terminal methyl in the hydrochloride, but synplanar to it in the hydrobromide. The analysis was repeated, eliminating all interatomic distances involving atoms of the propionate group, and the resulting half-normal plot is shown in Fig. 1(*d*). The linearity of the points is improved, and most of the extreme values have been eliminated. The remaining large δp's suggest a possible difference in conformation at the 3 or 4 positions of the piperidine ring, but no significant differences in torsion angles could be found. The slope of the least-squares line suggests that the e.s.d.'s for the interatomic distances are too small by a factor of about 2.

Example 3

Both monoclinic and rhombohedral forms have been found for β-promedol alcohol (1,2,5-trimethyl-4-phenylpiperidin-4-ol), and structures have been reported for both forms (De Camp & Ahmed, 1972 *a*, *b*). of the two forms, using all interatomic distances < 4.65 Å (excluding distances to hydrogen atoms). Six of the seven extreme points represent distances to the hydroxyl oxygen atom. Elimination of all distances to this atom resulted in the plot shown in Fig. 1(*f*). The minor difference between the two molecules was a shift of the hydroxyl oxygen atom by about 0·02 Å, as explained in the original paper by differences in hydrogen bonding. The remaining data support the hypothesis that there is little difference in conformation between the two crystalline forms. As in Example 2, the e.s.d.'s appear to be underestimated by a factor of about 2.

Example 4

The crystal structure of procaine has been determined as the hydrochloride salt (Beall, Herdklotz & Sass, 1970; Dexter, 1972) and as a complex with bis-*p*-nitrophenyl phosphate (Sax, Pletcher & Gustaffson, 1970). Dexter concluded that the differences between the two hydrochloride structures were mostly due to random errors, but that there were significant differences in conformation between the hydrochloride and Fig. 1(*e*) is a half-normal probability plot comparison the complex. The change in conformation of the diethylamino group was clearly significant. The conformation of the remainder of the molecule was studied, elimin-

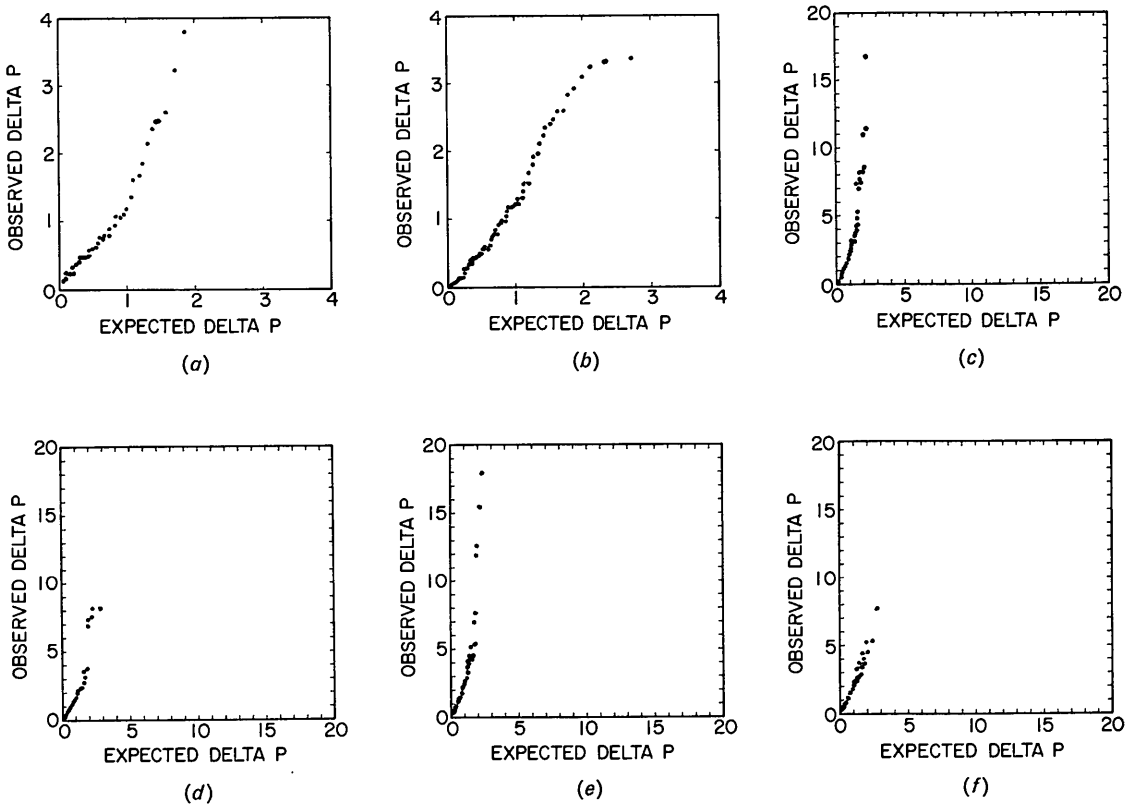

Fig. 1. Half-normal probability plots of independent molecular geometries. The three numbers given for each plot are: the number of values included, the slope of the least-squares straight line for points with $\delta p < 2{\cdot}0$, and the intercept of the least-squares line on the ordinate. (*a*) 3-methyl-3-pyrazolin-5-one, comparison of atomic positional parameters refined upon data measured by stationary-crystal stationary-counter and moving-crystal moving-counter methods, 39, 1·29, 0·01; (*b*) 3-methyl-3-pyrazolin-5-one, comparison of interatomic distances < 4·65 Å, 73, 1·40, −0·12; (*c*) betaprodine hydrobromide and hydrochloride, comparison of all interatomic distances < 4·65 Å, 114, 2·17, 0·02; (*d*) same as (*c*), but omitting all distances involving atoms in the propionate group, 67, 1·66, −0·03; (*e*) β-promedol alcohol, comparison of all interatomic distances < 4·65 Å (except to hydrogen atoms), 68, 2·10, 0·12; (*f*) same as (*e*), but omitting all distances to the hydroxyl oxygen atom, 59, 1·84, 0·12.

Table 1. *Probability-plot comparisons of procaine molecules*

Values compared	Half-normal plot: Maximum δp	Slope*	Intercept*	$\delta p > 3$	Normal plot: Range of δm	Slope*	Intercept*
13 bond lengths	3·35	2·42	−0·16	1	−3·35 to +2·11	2·03	−0·57
13 bond lengths 16 bond angles†	6·52	2·76	−0·21	7	−5·31 to +6·52	2·34	−0·36
13 bond lengths 16 bond angles† 4 torsion angles†	17·40	2·70	−0·13	8	−5·31 to +17·40	2·37	−0·43
13 bond lengths 16 bond angles† 15 torsion angles†	17·40	2·75	−0·12	13	−5·62 to +17·40	2·35	−0·48
51 interatomic distances <4·65 Å (hydrogens excluded)	17·96	3·06	−0·11	17	−7·35 to +17·96	2·68	−0·28

* Calculated for a least-squares fit of the central points (δp or $|\delta m| < 2$) to a straight line.
† Second- and third-nearest neighbour distances were used to define bond angles and torsion angles respectively.

ating distances to the carbon atoms of the ethyl groups. Both normal and half-normal plots were prepared using various subsets of the interatomic distances <4·65 Å as input data. The results are summarized in Table 1 in terms of the slope and intercept of the least-squares straight line through the central points of the plot.

Although Dexter concluded that the phenyl ring had significant quinonoid character, most of the extreme values are associated with distances to atoms in the alkylamino chain rather than the phenyl ring. In this part of the molecule, conformational differences are small. The high sensitivity of the method is also shown by the fact that the largest value of δm is associated with the N(1)···O(1) distance (using Dexter's notation). The corresponding torsion angle differs by only 8·5° between the hydrochloride and the complex.

The slope of the least-squares line tended to increase as more values were included, while the intercept approached zero. This suggests a greater underestimation of the e.s.d. for large interatomic distances, due to either a greater dependence of these values on the e.s.d.'s of the unit-cell dimensions (which were neglected in the present calculations) or a greater effect of a minor difference in conformation upon larger distances. In either event, the degree of underestimation of e.s.d.'s can be said to be at least 2.

Conclusion

Normal or half-normal probability plots based on differences between independently determined interatomic distances seem to be a highly sensitive method for analysis of conformational differences. The method may be applied to entire molecules or to molecular fragments, as long as the plot includes all appropriate interatomic distances. Because of the sensitivity of nonbonded distances to minor changes in conformation, a lack of extreme values of the error statistic appears to be conclusive evidence of two molecular conformations being identical.

I would like to express my thanks to Dr F. R. Ahmed and Dr S. C. Abrahams for their helpful suggestions and comments.

References

ABRAHAMS, S. C. & KEVE, E. T. (1971). *Acta Cryst.* A**27**, 157–165.
AHMED, F. R. & BARNES, W. H. (1963). *Acta Cryst.* **16**, 1249–1252.
AHMED, F. R., BARNES, W. H. & MASIRONI, L. DIM. (1963). *Acta Cryst.* **16**, 237–242.
BEALL, E., HERDKLOTZ, J. & SASS, R. L. (1970). *Biochem. Biophys. Res. Commun.* **39**, 329–334.
DE CAMP, W. H. & AHMED, F. R. (1972*a*). *Acta Cryst.* B**28**, 1796–1800.
DE CAMP, W. H. & AHMED, F. R. (1972*b*). *Acta Cryst.* B**29**, 3484–3489.
DE CAMP, W. H. & STEWART, J. M. (1971). *Acta Cryst.* B**27**, 1227–1232.
DEXTER, D. D. (1972). *Acta Cryst.* B**28**, 77–82.
DUNITZ, J. D. & WASER, J. (1972). *J. Amer. Chem. Soc.* **94**, 5645–5650.
SAX, M., PLETCHER, J. & GUSTAFFSON, R. (1970). *Acta Cryst.* B**26**, 114–124.