Acta Cryst. (1983). B**39**, 457–468

Molecular Geometry of Substituted Benzene Derivatives. IV.* Analysis of Variance in Monosubstituted Benzene Rings

By Aldo Domenicano

Istituto di Chimica Farmaceutica e Tossicologica, Università di Roma, Città Universitaria, 00185 *Roma, and*
Istituto di Strutturistica Chimica del CNR 'Giordano Giacomello', 00016 *Monterotondo Stazione, Italy*

Peter Murray-Rust†

*Department of Chemistry, The University of Stirling, Stirling FK*9 4*LA, Scotland*

and Alessandro Vaciago

*Chemical Crystallography Laboratory, The University of Oxford, Oxford OX*1 3*PD, England*

(*Received* 11 *September* 1981; *accepted* 4 *January* 1983)

Abstract

The variance of the geometry of monosubstituted benzene rings has been analysed systematically from a sample of 199 rings, representing a wide spectrum of substituents. The sample is composed of structures determined accurately by X-ray crystallography, neutron crystallography and microwave spectroscopy. Part of the analysis has been carried out using symmetry coordinates based on the irreducible representations of the D_{6h} point group. The application of bivariate analysis to these coordinates shows that the angular variance of monosubstituted benzene rings is fully described by two orthogonal components of distortion, involving angular changes in different ratios. The distortion that accounts for most of the variance is a simultaneous change of the internal angles at the *ipso*, *ortho* and (to a minor extent) *para* positions of the ring, and is definitely related to the σ electronegativity of the substituent. The second distortion involves mainly the internal angles at the *meta* and *para* positions and is controlled to a large extent by the π-donor/acceptor power of the substituent. The separation of the two distortions is not based on chemical assumptions, but stems directly from the statistical analysis of the data. When bond-length variation is included in the model, factor analysis reveals the existence of three orthogonal components of distortion. These account for *ca* 90% of the variance, the remaining 10% being essentially due to random experimental error. The first distortion, accounting for *ca* 40% of the variance, now also involves a change of the C_{ipso}—C_{ortho} bond length. The second distortion accounts for *ca* 20% of the variance and consists, as before, solely of angular changes. The third distortion is a simultaneous decrease (increase) of all the C—C bond lengths of the ring. Although statistically significant, it appears to arise primarily from uncorrected thermal motion and inhomogeneity of the sample.

1. Introduction

The effect of substituents on the geometry of monosubstituted benzene rings has been extensively analysed in our laboratories. From the experimental results available in 1974, it was concluded that although there was considerable variance in bond lengths, most of it was due to experimental error, particularly the lack of an accurate correction for thermal-motion effects (Domenicano, Vaciago & Coulson, 1975*a*). In contrast, it was found that variance in angles was related to the nature of the substituents and in particular that the angle at the *ipso* C atom, α, correlates well with the electronegativity of the substituent (Domenicano *et al.*, 1975*a,b*; Domenicano, Mazzeo & Vaciago, 1976). Further experimental results led to the observation that certain substituents can also affect the angle at the *para* C atom, δ, and that this is related to perturbations caused by the substituent in the π system of the ring (Domenicano & Vaciago, 1979; Domenicano & Murray-Rust, 1979).

With the aim of finding out how many independent modes of distortion may be present in monosubstituted benzene rings, we have applied bivariate and multivariate statistical analysis to the geometry of a carefully selected sample of benzene derivatives. The general approach has been outlined by Murray-Rust & Bland (1978) and Murray-Rust & Motherwell (1978*a,b*); here

---

[*] Part III: Domenicano & Vaciago (1979).

[†] Present address: Chemistry Department, Glaxo Group Research Ltd, Greenford Road, Greenford UB6 0HE, England.

0108-7681/83/040457-12$01.50

we concentrate on the application of bivariate correlation and factor analysis. We expect to determine which factors represent real structural effects rather than random error, and possibly to use them as indices of the effects.

2. The data set

2.1. *Selection of data*

The distortions of the benzene-ring geometry caused by substitution are generally rather small, amounting to no more than a few hundredths of an ångström unit for bond lengths and a few degrees for bond angles. They can thus be determined only through very accurate experiments, and this should be taken into account when planning a statistical analysis. Moreover, when the distortions involved are less than 0·01 Å and 1° (which is the order of magnitude of some of the systematic errors inherent in some techniques of structure determination), it is highly desirable to have independent experiments to provide a check.

In a statistical analysis the selection of data is of critical importance. Where a complete population cannot be analysed, a sample is taken whose selection should be as unbiased as possible. Ideally we should choose a set of compounds where the substituents are representative of the set of all possible substituents but this is clearly impossible. Compounds in the literature do not form a random set; they have been selected by the original authors because they are interesting, they are stable, they crystallize well and for other practical reasons. Thus *N*-substituted amino groups are very frequent whilst strongly electropositive substituents are rare.

The molecular structures of hundreds of monosubstituted benzene derivatives have so far been determined, mainly by X-ray crystallography. This large mass of information is stored in the files of the Cambridge Crystallographic Data Centre (CCDC) (Allen *et al.*, 1979) and is accessible to automatic retrieval and analysis by standard procedures (Murray-Rust & Motherwell, 1978*a*). It would be straightforward to form a data set by retrieving all monosubstituted benzene rings from the CCDC file but most of these would have to be discarded because of low precision. We have therefore compiled (with the help, where necessary, of the CCDC file) a carefully selected sample of high-precision structures,* attempting to cover the range of substituents as evenly as possible. Given that the crystallographic literature itself is biased, we believe that our sample gives as good a picture as is possible of today's knowledge of distortions in monosubstituted benzene rings.

All entries in our data set have been included only after careful analysis of the original literature. Both the accuracy and precision of published results have been checked. Crystal structures showing evidence of disorder or very high thermal motion in the benzene ring have been discarded. Structures in which the contribution of the H atoms to the scattering model used in the refinement was not taken into account have also been discarded. Whenever necessary (*e.g.* if the internal bond angles of the ring did not add up to 720°) the geometry has been checked from the published atomic coordinates, and e.s.d.'s of molecular parameters have been recalculated from the published e.s.d.'s of atomic coordinates. Occasionally it has been clear that the authors were rather optimistic in their assessment of precision; structures of this type have been excluded. Structures with $\sigma(\mathrm{C{-}C{-}C}) > 0{\cdot}3^\circ$ have not normally been considered for inclusion in the data set; the threshold has sometimes been raised to 0·4° to allow the inclusion of chemically important substituents.

We have included structures determined by one or more of the following techniques: (i) gas-phase structures determined by the 'substitution' method of microwave spectroscopy; (ii) structures determined by (room-temperature) neutron crystallography; (iii) structures determined by low-temperature X-ray crystallography; (iv) structures determined by room-temperature X-ray crystallography. Gas-phase electron diffraction (unless combined with microwave spectroscopy) does not generally allow a very accurate measurement of the distortions of the benzene ring that occur in monosubstituted derivatives.* Structures determined by this technique have not, therefore, been included. A molecule was included more than once, if possible, when its structure had been determined by different techniques or different authors.

The distortions of the benzene ring caused by substitution are known to be dependent upon the nature and properties of the substituent, an important factor being the row of the periodic table to which the first atom of the substituent belongs (Domenicano *et al.*, 1975*a*,*b*; Domenicano, Hargittai & Schultz, 1980). We have therefore carried out separate analyses for monosubstituted rings with first- and second-row substituents.†

The part of the data set pertaining to first-row substituents has 149 entries and includes four micro-

---

* No structures reported after mid-1980 have been considered.

* Gas-phase electron diffraction is, however, a very valuable technique for measuring accurately the distortions of symmetrically *para*-disubstituted benzene rings, a class of molecules not otherwise accessible to gas-phase structural studies (Domenicano, Schultz, Kolonits & Hargittai, 1979; Schultz, Hargittai & Domenicano, 1980; Domenicano, Schultz & Hargittai, 1982).

† Accurate structural studies of benzene derivatives with third- and fourth-row substituents are few and have not been considered.

wave structures,* five neutron diffraction structures, 36 low-temperature X-ray diffraction structures and 104 room-temperature X-ray diffraction structures. The part pertaining to second-row substituents has 50 entries and consists entirely of X-ray diffraction structures,* three of which were determined at low temperature.†

2.2. *Treatment of data*

In most solid-state structures in our data set the molecular parameters of the benzene ring showed almost exact C_{2v} symmetry, although this was not required by crystallographic constraints.‡ Small systematic distortions of the benzene ring from C_{2v} symmetry (but retaining planarity, *i.e.* C_s symmetry) have been shown to occur in *para*-substituted anisoles (Di Rienzo, Domenicano, Portalone & Vaciago, 1976). They have been interpreted in terms of asymmetric π donation from the methoxy group, caused by the inherent asymmetry of this substituent. A similar effect is definitely present in some of the molecules included in the data set, where a planar substituent (*e.g.* $-NHR$, $-CH{=}CHR$) interacts asymmetrically with the π-electron system of the ring. Because of the limited occurrence and modest extent of the deviations from axial symmetry (the difference between the values of the internal angles related by the $C_{ipso}\cdots C_{para}$ axis seldom exceeds 1°) we have treated all benzene rings in our data set as having idealized C_{2v} symmetry and averaged appropriate molecular parameters.§

Chemically and conformationally equivalent benzene rings within an asymmetric unit have also been averaged, so as to yield a single entry in the data set.

3. The parameters

3.1. *Internal coordinates*

To define the geometry of a C hexagon with C_{2v} symmetry only five parameters are necessary, although there may be three unequal bond lengths and four unequal angles, *i.e.* seven internal coordinates (Fig. 1). These coordinates are in fact linked by two independent equations of geometrical constraint, expressing the conditions of planarity and ring closure. One equation imposes the constancy of the sum of the angles:

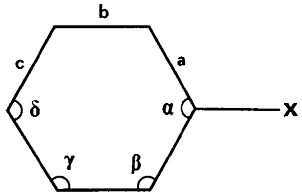


Fig. 1. Labelling of the C—C bonds and C—C—C angles in monosubstituted benzene rings (C_{2v} symmetry assumed).

$$\alpha + 2\beta + 2\gamma + \delta = 720°; \tag{1}$$

the other is non-linear and involves bond lengths as well as angles:

$$a \sin\frac{\alpha}{2} + b \sin\left(\beta + \frac{\alpha}{2} - 180°\right) = c \sin\frac{\delta}{2}. \tag{2}$$

In any statistical analysis of internal coordinates the equations of constraint will introduce correlation. Where the equation is linear a perfect correlation will be found. Thus in the present case $\alpha + \delta$ correlates *exactly* with $\beta + \gamma$. If the constraint is non-linear the correlation will be non-linear, but for small distortions of a planar hexagon from D_{6h} symmetry linear relationships give excellent approximations (Britton, 1977).

3.2. *Geometrical and statistical dependence*

In order to form an orthogonal basis for describing the geometry of monosubstituted benzene rings we might arbitrarily choose a subset of five *geometrically independent* internal coordinates, such as a, b, α, γ and δ. It is not clear, however, whether these are the 'best' internal coordinates to take, or whether some linear combination may not be more appropriate. It should also be realized that an inappropriate choice may bias the analysis. The method that retains most symmetry, and is thus equally 'fair' to each parameter, is to form *symmetry coordinates*. These parameters, discussed in detail in § 3.3, are linear combinations of internal coordinates and necessarily geometrically independent. They form an orthogonal basis for describing molecular geometry and the number of non-zero coordinates is the *geometrical dimensionality* of the problem (five in the present case).

Parameters which are geometrically independent, however, are not necessarily *statistically independent*. As we shall see later, excellent correlations are found between some geometrically orthogonal parameters in

---

* The microwave structures of nitrobenzene (Høg, 1971) and chlorobenzene (Michel, Nery, Nosberger & Roussy, 1976) are inaccurate in the *ipso* region of the ring (Schultz, Hargittai & Domenicano, 1980; Colapietro & Domenicano, 1982) and have not been included.

† The entire data set has been deposited with the British Library Lending Division as Supplementary Publication No. SUP 38317 (42 pp.). Copies may be obtained through The Executive Secretary, International Union of Crystallography, 5 Abbey Square, Chester CH1 2HU, England.

‡ C_{2v} symmetry was assumed in three out of the four gas-phase studies included in the data set.

§ A general distortion of a planar hexagon can indeed be represented exactly by two orthogonal parts: a distortion *from* C_{2v} symmetry and a distortion *retaining* C_{2v} symmetry. The averaging process discards the first component and reveals the second exactly. In most molecules in our data set, however, averaging simply eliminates some of the experimental error.

monosubstituted benzene rings. The *statistical dimensionality* of a problem can never be higher than the geometrical dimensionality, and in many cases may be considerably lower. The determination of statistical dimensionality is carried out by factor analysis (see § 7).

3.3. *Symmetry coordinates*

The use of symmetry coordinates parallels the decomposition of the harmonic vibrational force field into normal coordinates. In their analysis of the distortions of tetrahedral molecules, Murray-Rust, Bürgi & Dunitz (1978*a*,*b*) showed that the use of symmetry distortion coordinates transforming as the irreducible representations of the T_d point group led to a simple scheme for revealing geometrical correlations. Thus small distortions transforming as different (orthogonal) irreducible representations showed no correlation whereas those belonging to the same irreducible representation were often highly correlated. Bond-length and angle distortions transforming solely as T_2 were found to correlate well for several species of tetrahedral molecules. Although in that case the distortions were caused mainly by the crystal or molecular environment, we expect that the symmetry coordinates could also be useful in describing the small distortions of the benzene ring caused by substitution.

A full description of the in-plane distortions of benzene rings in terms of symmetry coordinates has been given elsewhere (Murray-Rust, 1982). In terms of six bond and six angle basis vectors (Fig. 2) we derive nine symmetry coordinates based on the irreducible representations of the D_{6h} point group:

$$\mathbf{D}_1\,(A_{1g}) = 6^{-1/2}(\mathbf{r}_{12} + \mathbf{r}_{23} + \mathbf{r}_{34} + \mathbf{r}_{45} + \mathbf{r}_{56} + \mathbf{r}_{61}) \quad (3)$$

$$\mathbf{D}_2\,(B_{2u}) = 6^{-1/2}(\mathbf{r}_{12} - \mathbf{r}_{23} + \mathbf{r}_{34} - \mathbf{r}_{45} + \mathbf{r}_{56} - \mathbf{r}_{61}) \quad (4)$$

$$\mathbf{D}_{3a}(E_{2g}) = 12^{-1/2}(-\mathbf{r}_{12} + 2\mathbf{r}_{23} - \mathbf{r}_{34} - \mathbf{r}_{45} + 2\mathbf{r}_{56} - \mathbf{r}_{61}) \quad (5)$$

$$\mathbf{D}_{3b}(E_{2g}) = 4^{-1/2}(-\mathbf{r}_{12} + \mathbf{r}_{34} - \mathbf{r}_{45} + \mathbf{r}_{61}) \quad (6)$$

$$\mathbf{D}_{4a}(E_{2g}) = 12^{-1/2}(2\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2 - \boldsymbol{\theta}_3 + 2\boldsymbol{\theta}_4 - \boldsymbol{\theta}_5 - \boldsymbol{\theta}_6) \quad (7)$$

$$\mathbf{D}_{4b}(E_{2g}) = 4^{-1/2}(\boldsymbol{\theta}_2 - \boldsymbol{\theta}_3 + \boldsymbol{\theta}_5 - \boldsymbol{\theta}_6) \quad (8)$$

$$\mathbf{D}_{5a}(E_{1u}) = 12^{-1/2}(2\boldsymbol{\theta}_1 + \boldsymbol{\theta}_2 - \boldsymbol{\theta}_3 - 2\boldsymbol{\theta}_4 - \boldsymbol{\theta}_5 + \boldsymbol{\theta}_6) \quad (9)$$

$$\mathbf{D}_{5b}(E_{1u}) = 4^{-1/2}(-\boldsymbol{\theta}_2 - \boldsymbol{\theta}_3 + \boldsymbol{\theta}_5 + \boldsymbol{\theta}_6) \quad (10)$$

$$\mathbf{D}_6\,(B_{1u}) = 6^{-1/2}(\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2 + \boldsymbol{\theta}_3 - \boldsymbol{\theta}_4 + \boldsymbol{\theta}_5 - \boldsymbol{\theta}_6). \quad (11)$$

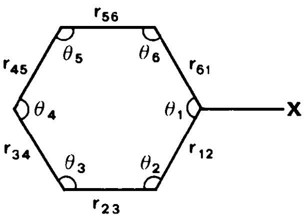


Fig. 2. The six bond and six angle basis vectors of a planar benzene ring.

Although there is a freedom of choice in the analytical form of the E_{2g} distortions we prefer to separate bond and angle contributions. Because of the equations of constraint for planar-ring closure the E_{1u} distortion involves *linked* bond and angle changes. We have chosen for most of our analysis to represent it by angles. We could equally well express it by bond lengths:

$$\mathbf{D}'_{5a} = 4^{-1/2}(-\mathbf{r}_{12} + \mathbf{r}_{34} + \mathbf{r}_{45} - \mathbf{r}_{61}) \quad (12)$$

$$\mathbf{D}'_{5b} = 12^{-1/2}(\mathbf{r}_{12} + 2\mathbf{r}_{23} + \mathbf{r}_{34} - \mathbf{r}_{45} - 2\mathbf{r}_{56} - \mathbf{r}_{61}). \quad (13)$$

In this paper we shall consider only distorted rings which retain C_{2v} symmetry, *i.e.* for which:

$$|\mathbf{r}_{12}| = |\mathbf{r}_{61}| = a \quad (14)$$

$$|\mathbf{r}_{23}| = |\mathbf{r}_{56}| = b \quad (15)$$

$$|\mathbf{r}_{34}| = |\mathbf{r}_{45}| = c \quad (16)$$

$$|\boldsymbol{\theta}_1| = \alpha \quad (17)$$

$$|\boldsymbol{\theta}_2| = |\boldsymbol{\theta}_6| = \beta \quad (18)$$

$$|\boldsymbol{\theta}_3| = |\boldsymbol{\theta}_5| = \gamma \quad (19)$$

$$|\boldsymbol{\theta}_4| = \delta. \quad (20)$$

Substitution into equations (3)–(11) gives* $D_2 = D_{3b} = D_{4b} = D_{5b} = 0$, and the simplified expressions:

$$D_1 = (\tfrac{2}{3})^{1/2}(a + b + c) \quad (21)$$

$$D_3 = D_{3a} = 3^{-1/2}(-a + 2b - c) \quad (22)$$

$$D_4 = D_{4a} = 3^{-1/2}(\alpha - \beta - \gamma + \delta) \quad (23)$$

$$D_5 = D_{5a} = 3^{-1/2}(\alpha + \beta - \gamma - \delta) \quad (24)$$

$$D_6 = 6^{-1/2}(\alpha - 2\beta + 2\gamma - \delta). \quad (25)$$

In terms of bond lengths the E_{1u} distortion can now be written [from equation (12)]:

$$D'_5 = D'_{5a} = -a + c = D_5\langle r\rangle\,\pi/180 \quad (26)$$

(when D_5 is expressed in degrees and $\langle r\rangle$ is the average C—C bond length).

From the C—C bond length, r_o, and C—C—C angle, 120°, of a reference benzene ring of D_{6h} symmetry we can calculate the changes in internal coordinates:

$$\Delta a = a - r_o \quad (27)$$

$$\Delta b = b - r_o \quad (28)$$

---

* The distortion vectors, $\mathbf{D}_i$, can now be represented by scalars, D_i, and these are used from now on.

$$\Delta c = c - r_o \quad (29)$$

$$\Delta\alpha = \alpha - 120° \quad (30)$$

$$\Delta\beta = \beta - 120° \quad (31)$$

$$\Delta\gamma = \gamma - 120° \quad (32)$$

$$\Delta\delta = \delta - 120°. \quad (33)$$

Substitution of Δa, Δb, *etc.* into equations (22)–(25) does not change the numerical values of D_3, D_4, D_5 and D_6, which can thus also be regarded as *symmetry distortion coordinates*. To calculate the symmetry distortion coordinate corresponding to A_{1g} we have to know r_o:

$$\Delta D_1 = (\tfrac{2}{3})^{1/2}(a + b + c - 3r_o). \quad (34)$$

Since we do not have a value for r_o *a priori* (because of the heterogeneity of our sample, see § 4.2) we shall use D_1 from here on.

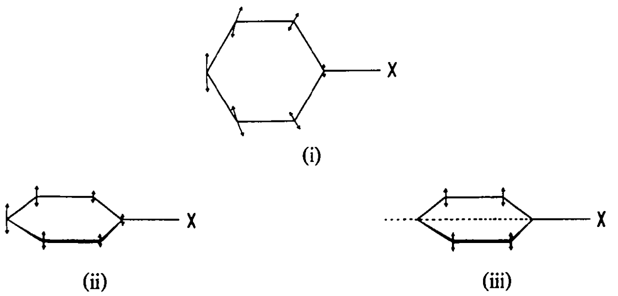
4. Systematic errors

Before discussing the results of the statistical analysis the effect of important systematic errors in the data set must be estimated. Two of these need specific discussion.

4.1. *Solid-state thermal motion*

This is by far the most important source of systematic error in our data set, which consists mainly of structural results obtained by room-temperature crystallography. We can visualize three types of libration of a phenyl group: they may be described as (i) yawing, (ii) pitching and (iii) rolling (Fig. 3).

The main effect of yawing and pitching is that atoms near the *para* position will have the highest thermal motion. Bond lengths will apparently be shortened, the effect being more pronounced for the *c* bonds and less pronounced for the *a* bonds. The changes in angles will probably be very much less (Wheatley, 1959; Domenicano *et al.*, 1975*a*). Since all bond lengths will tend to be reduced we may expect to observe a positive covariance between *a*, *b* and *c* due to yawing and pitching.



(i)

(ii) (iii)

Fig. 3. Different types of libration of a phenyl group: (i) yawing, (ii) pitching and (iii) rolling.

By contrast, rolling leads to a shortening of only *a* and *c* and also results in a decrease of α and δ and an increase of β and γ. This may cause some covariance between these parameters. Rolling is simply related to the symmetry coordinates, since the angle distortion transforms wholly as E_{2g} (*i.e.* D_4) and the bond distortion as $A_{1g} + E_{2g}$ ($D_1 + D_3$). This might lead to some covariance between D_3 and D_4.

4.2. *Heterogeneity of the data*

The data set used for first-row substituents is rather heterogeneous, since it includes structures determined by different techniques and hence with different systematic errors. Microwave spectroscopy yields accurate internuclear distances, essentially free from thermal-motion effects. In the *ipso* region of the benzene ring, however, the accuracy may be rather low, due to the proximity to the centre of gravity of the molecule. Internuclear distances obtained by neutron crystallography may be shortened to some extent by thermal-motion effects. Bond distances determined by X-ray crystallography do not correspond to internuclear separations, since the centroids of charge do not necessarily coincide with nuclear positions. Moreover, they are shortened to a varying extent by solid-state thermal motions.

We therefore expect to have four separate contributions to the variance of the data:

(1) – Variance from random errors;

(2) – Variance from systematic errors inherent to each technique of structure determination, particularly thermal-motion effects;

(3) – Variance from the different physical meaning of molecular parameters obtained by different techniques;

(4) – Variance from substituent-induced changes of molecular geometry.

The second contribution will be prominent in room-temperature diffraction studies. The relative magnitude of the third contribution is expected to be smaller than that of the second and can probably be neglected where angular variance is concerned. This contribution is irrelevant to second-row substituents, where the data set consists entirely of X-ray diffraction structures.

5. Univariate analysis

Univariate statistics* for the seven internal coordinates and five symmetry coordinates are presented in Tables

---

* All statistical calculations were performed using either the SPSS system on the CDC 7600 of the University of Manchester Regional Computing Centre or the OMNITAB system on the Univac 1100/22 of the University of Rome.

1 and 2, respectively. It is immediately clear that the standard deviations of many internal coordinates are much larger than the e.s.d.'s given by the original authors, which are on average about 0·003 Å for bond lengths and 0·2° for bond angles. Relative to the e.s.d.'s there appears to be a greater experimental variance in bond angles than bond lengths. It is noticeable how α varies four to five times as much as does δ, in keeping with previous observations (Domenicano & Vaciago, 1979). In contrast, bond-length variation is much less related to the position of the bond with respect to the substituent, being roughly the same for *a*, *b* and *c*.

The mean values of bond lengths decrease from *a* to *c*, which is largely due to thermal-motion effects, as discussed in § 4.1.

It has been noted (Colapietro, Domenicano, Hargittai, Riva di Sanseverino & Schultz, 1980) that because of the different meaning of bond distances determined by different techniques, X-ray interatomic distances in benzene derivatives are systematically shorter than distances determined by neutron crystallography or gas-phase electron diffraction. The difference seems to be fairly constant (*ca* 0·7%, *i.e.* about 0·01 Å in a C—C bond). The shortening of the mean C—C bond lengths in the present study [$\langle r^{\mathrm{I}}(\mathrm{C—C})\rangle$ = 1·385 Å,* $\langle r^{\mathrm{II}}(\mathrm{C—C})\rangle$ = 1·383 Å; r_g(C—C) for gaseous benzene is 1·399 Å (Tamagawa, Iijima & Kimura, 1976; Schultz, Kolonits & Hargittai, 1980)] can be explained by this effect and thermal motion.

Since the symmetry coordinates are normalized we can estimate from the observed variances the relative proportions of bond and angle variance in each coordinate. Thus much more angle variance is due to D_4 and D_6 than D_5 whereas more bond variance is due to D_1 than D_3 and D_5 (Table 2).

6. Bivariate analysis

6.1. *Internal coordinates*

Bivariate correlations were calculated for all combinations of internal coordinates; correlation coefficients are shown in Table 3.

Table 1. *Univariate statistics for internal coordinates*

Parameter	Range	Mean	Standard deviation	Variance
(*a*) First-row substituents				
a	1·367–1·411 Å	1·390 Å	0·0083 Å	0·000069 Å^2
b	1·371–1·397	1·387	0·0054	0·000029
c	1·362–1·399	1·379	0·0085	0·000072
α	111·8–123·4°	119·0°	1·99°	3·96°^{2}
β	117·6–124·6	120·3	1·19	1·42
γ	119·7–121·4	120·4	0·30	0·09
δ	118·2–121·2	119·6	0·52	0·27
(*b*) Second-row substituents				
a	1·376–1·402 Å	1·389 Å	0·0053 Å	0·000028 Å^2
b	1·372–1·400	1·386	0·0049	0·000024
c	1·359–1·385	1·373	0·0063	0·000040
α	114·4–122·5°	119·0°	1·52°	2·31°^{2}
β	117·9–123·0	120·2	0·94	0·88
γ	119·8–120·5	120·2	0·17	0·03
δ	119·5–121·0	120·1	0·34	0·12

Table 3. *Correlation coefficients* (×10^2) *for internal coordinates*

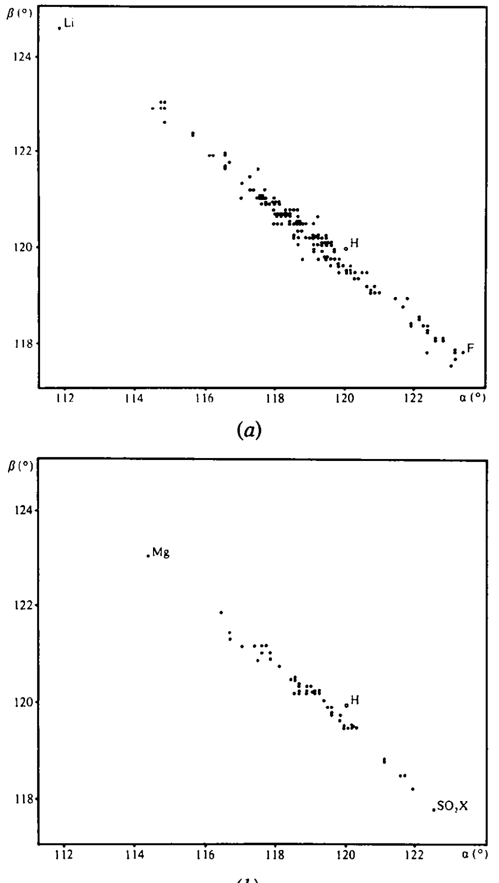
	a	*b*	*c*	α	β	γ
(*a*) First-row substituents						
b	33					
c	55	53				
α	−56	12	24			
β	55	−10	−19	−99·1		
γ	4	−5	−4	4	−14	
δ	−41	3	−1	64	−60	−64
(*b*) Second-row substituents						
b	10					
c	51	6				
α	−37	11	38			
β	36	−9	−32	−99·3		
γ	−29	−32	−2	41	−47	
δ	−8	38	12	66	−67	−22

Table 2. *Univariate statistics for symmetry coordinates*

Parameter	Range	Mean	Standard deviation	Variance	Percentage of bond variance	Percentage of angle variance
(*a*) First-row substituents						
D_1	3·364–3·428 Å	3·393 Å	0·0147 Å	0·000216 Å^2	64	—
D_3	−0·013–0·031	0·003	0·0077	0·000059	18	—
D_4	−8·43– 3·35°	−1·18°	2·05°	4·20°^{2}	—	58
D_5	−1·39– 0·58	−0·42	0·32	0·10	—	1
	−0·033–0·014 Å	−0·010 Å	0·0077 Å	0·000059 Å^2	18	—
D_6	−6·25–3·76°	−0·15°	1·72°	2·96°^{2}	—	41
(*b*) Second-row substituents						
D_1	3·366– 3·408 Å	3·386 Å	0·0094 Å	0·000088 Å^2	50	—
D_3	−0·021– 0·023	0·006	0·0077	0·000059	33	—
D_4	−5·14– 3·00°	−0·78°	1·52°	2·31°^{2}	—	55
D_5	−1·33– −0·23	−0·63	0·23	0·05	—	1
	−0·032– −0·006 Å	−0·015 Å	0·0055 Å	0·000030 Å^2	17	—
D_6	−4·61– 2·65°	−0·44°	1·37°	1·88°^{2}	—	44

---

* Here and throughout this paper the superscripts I and II denote quantities referring to first- and second-row substituents, respectively.



(a)

(b)

Fig. 4. Scattergrams of the angle β against α for monosubstituted benzene rings with (a) first-row substituents and (b) second-row substituents. The extreme data points of the scattergrams are identified as follows (for each molecule the code in the data set is also given): Li: di-μ-phenyl-bis[(*N*,*N*,*N*′,*N*′-tetramethylethylenediamine)lithium], LI01 (Thoennes & Weiss, 1978*a*); Mg: diphenyl(*N*,*N*,*N*′,*N*′-tetramethylethylenediamine)-magnesium, MG01 (Thoennes & Weiss, 1978*b*); F: fluorobenzene, F01 (Nygaard, Bojesen, Pedersen & Rastrup-Andersen, 1968); SO_2X: diphenyl disulphone, S10 (Kiers & Vos, 1972). Also shown is the reference position of benzene (H).

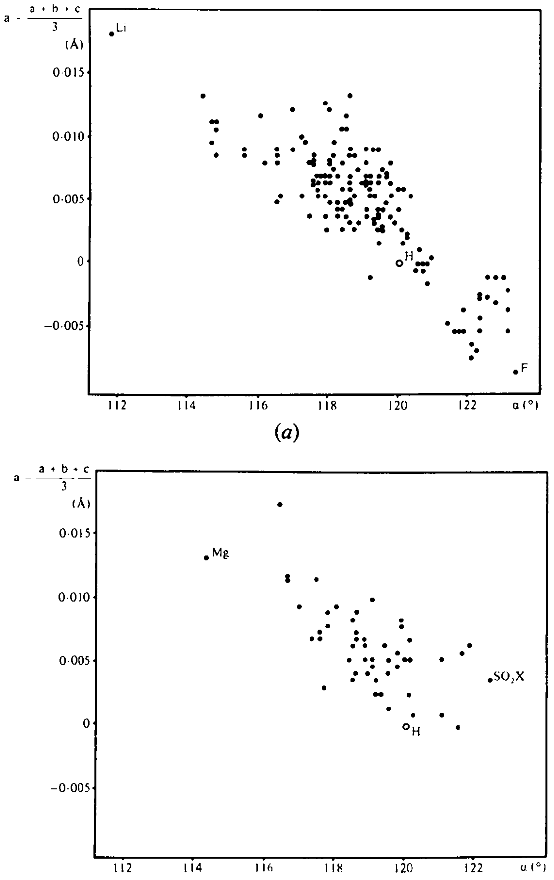
The rationalization of ring distortions in monosubstituted benzene derivatives in terms of the VSEPR model, as proposed by Domenicano *et al.* (1975*a*,*b*), implies that β and a should correlate with α. The correlation of β with α is indeed outstandingly good and perfectly linear for both parts of the data set (Fig. 4), with correlation coefficients $r^{\rm I} = -0{\cdot}991$ and $r^{\rm II} = -0{\cdot}993$. The equations of the regression lines are most conveniently expressed in terms of deviations from 120°:

$$\Delta\beta^{\rm I} = -0{\cdot}591\,(7)\,\Delta\alpha^{\rm I} - 0{\cdot}301\,(15)°* \qquad (35)$$

$$\Delta\beta^{\rm II} = -0{\cdot}615\,(11)\Delta\alpha^{\rm II} - 0{\cdot}384\,(19)°. \qquad (36)$$

The difference between the calculated coefficients for first- and second-row substituents is small and of marginal significance. Note that in both cases unsubstituted benzene ($\Delta\beta = \Delta\alpha = 0$) definitely lies off the regression line.

Surprisingly, the correlation of a with α is very poor, especially with second-row substituents ($r^{\rm I} = -0{\cdot}56$, $r^{\rm II} = -0{\cdot}37$), presumably because most of the variance of a is due to thermal motion and inhomogeneity of the sample. If we assume that in any one structure thermal motion affects a, b and c in roughly equal amounts, the quantity $a' = a - (a + b + c)/3$ should give a much better indication of the structural variation of the C_{ipso}—C_{ortho} bond length. Indeed, the situation improves greatly when a' is used instead of a (Fig. 5: $r^{\rm I} = -0{\cdot}83$, $r^{\rm II} = -0{\cdot}68$), and this seems to be a promising way to reduce systematic errors like thermal motion and inhomogeneity of the sample. Further improvement is not achieved, however, when only low-temperature ($T < 200$ K) and gas-phase structures are used ($r^{\rm I} = -0{\cdot}83$ on 37 data points). These results confirm the



(a)

(b)

Fig. 5. Scattergrams of $a' = a - (a + b + c)/3$ against the angle α for monosubstituted benzene rings with (a) first-row substituents and (b) second-row substituents. The data points Li, Mg, F and SO_2X are identified in the caption of Fig. 4.

---

* The standard deviations of the coefficients are given in parentheses as units in the last digit.

Table 4. *Correlation coefficients* ($\times 10^2$) *for symmetry coordinates*

(*a*) First-row substituents

	D_1	D_3	D_4	D_5
D_3	−50			
D_4	−13	31		
D_5	8	16	80	
D_6	−11	27	90	76

(*b*) Second-row substituents

	D_1	D_3	D_4	D_5
D_3	−31			
D_4	11	9		
D_5	18	−16	66	
D_6	2	2	95	70

observation (Wheatley, 1959) that bond distances measured by crystal-structure analysis suffer more from systematic errors than do bond angles.

6.2. *Symmetry coordinates*

Bivariate correlations were calculated for all combinations of the five symmetry coordinates; correlation coefficients are shown in Table 4. Many of the correlation coefficients are essentially zero as expected. Surprisingly there is no correlation between D_3 and D_4, although these coordinates both transform as E_{2g} in D_{6h}. Even more surprisingly, there is a good correlation between the two angle coordinates D_4 and D_6 which transform as orthogonal distortions in D_{6h} ($r^{\mathrm{I}} = 0{\cdot}90$, $r^{\mathrm{II}} = 0{\cdot}95$). The variance of the symmetry coordinate D_5 is very much smaller than that of D_4 and D_6 (Table 2). Since D_5 correlates rather well with D_4 and D_6 (Table 4) the angular variance of monosubstituted benzene rings can be completely represented by the latter two coordinates.

Scattergrams of D_4 against D_6 are shown in Fig. 6. By correlating the position of the data points in the scattergram with the chemical nature of the substituent we can clearly see that most of the scatter is due to chemical variation rather than experimental error. Thus the most electropositive first-row substituent (Li) corresponds to the first point in the bottom left of Fig. 6(*a*), whereas the most electronegative (F) is one of the furthest points in the top right. The same situation occurs with second-row substituents (Fig. 6*b*).

The major axes of the distributions* are given by:

$$D_4^{\mathrm{I}} = 1{\cdot}20\,(4)\,D_6^{\mathrm{I}} - 0{\cdot}99\,(7) \qquad (37)$$

$$D_4^{\mathrm{II}} = 1{\cdot}11\,(5)\,D_6^{\mathrm{II}} - 0{\cdot}28\,(7). \qquad (38)$$

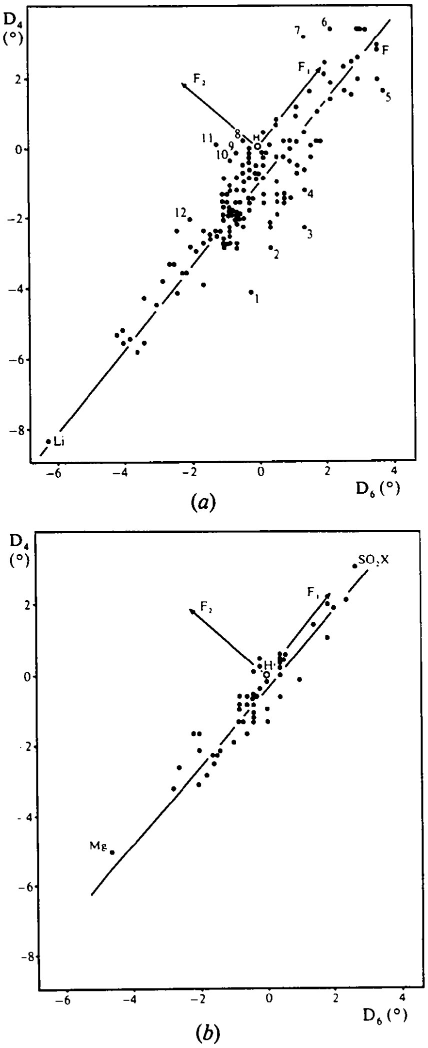


Fig. 6. Scattergrams of the symmetry coordinate D_4 against D_6 for monosubstituted benzene rings with (*a*) first-row substituents and (*b*) second-row substituents. The data points Li, Mg, F and SO_2X are identified in the catpion of Fig. 4; some major outliers are numbered sequentially and are identified in Table 5. Also shown are the major axis of the distributions and the axes of the F_1^{I} and F_2^{I} coordinate system (see text).

The slope of the two lines is essentially the same but the intercept differs widely. This confirms the observation that angular distortions in benzene derivatives depend upon the row of the periodic table to which the first atom of the substituent belongs (Domenicano *et al.*, 1975*a*,*b*; Domenicano *et al.*, 1980). Again the point representing unsubstituted benzene ($D_4 = D_6 = 0$) lies off the major axes. This comes as no surprise, as the symmetry of unsubstituted benzene, D_{6h}, is different from that assumed for all the substituted benzene rings used in the analysis, C_{2v}. The unique position of H in the periodic system should also, of course, be considered.

Since the major axes of the two plots have slopes of *ca* 1·15, we conclude that the most likely type of angular distortion is a combination of $D_6 + 1{\cdot}15D_4$. From equations (23) and (25) we see that this

---

* It is inappropriate to use regression here since we cannot regard one variable as dependent on the other.

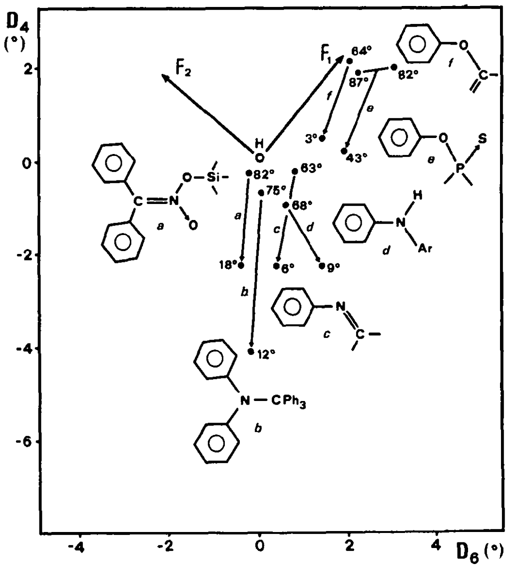


Fig. 7. Part of the D_4 *vs* D_6 scattergram for monosubstituted benzene rings with first-row substituents showing the effect of conjugation on the symmetry coordinates D_4 and D_6. Arrows connect data points having chemically identical (or very similar) planar substituents but markedly different twist angles, τ. In essentially all cases, decreasing τ causes a point to move downwards almost vertically. The geometries come from X-ray diffraction studies; τ values not reported by the original authors have been calculated from the given atomic and unit-cell parameters. The molecules involved are the following (the identification code in the data set is given in square brackets): (*a*) (*tert*-Butyl)dimethylsilyl ester of *aci*-nitrodiphenylmethane (Colvin, Beck, Bastani, Seebach, Kai & Dunitz, 1980: a low-temperature study). The two phenyl groups of the molecule have different conformations in the crystal, $\tau = 82°$ [C50] and $\tau = 18°$ [C51]. (*b*) (Diphenylamino)triphenylmethane (Hoekstra & Vos, 1975*a*,*b*: a low-temperature study). The two phenyl groups bonded to the N atom have different conformations in the crystal, $\tau = 75°$ [N28] and $\tau = 12°$ [N27]. (*c*) *N*-(Diphenylmethylene)aniline, $\tau = 63°$ [N36] (Tucker, Hoekstra, ten Cate & Vos, 1975) and '*N*-[3-phenyl-5-(1,2,3,4-oxatriazolio)]-phenylamide' ($C_{13}H_{10}N_4O$), $\tau = 6°$ [N35] (Ottersen, Christophersen & Treppendahl, 1975). Both studies have been carried out at low temperature. (*d*) Ammonium 8-anilino-1-naphthalenesulphonate hemihydrate (Cody & Hazel, 1977). The asymmetric unit contains two crystallographically independent molecules, having different conformations of the phenyl groups. The different orientation of the arrow with respect to all other cases is consistent with the change of hybridization of the N atom (from nearly tetrahedral when $\tau = 68°$ [N06] to nearly trigonal when $\tau = 9°$ [N05]). (*e*) Phenoxythiophosphoryl-dihydrazide, $\tau = 87°$ [O03] (Engelhardt, 1979), *trans*-dithiodihydrazidodimetaphosphoric acid diphenylester, $\tau = 82°$ [O04] (Engelhardt & Hartl, 1975), and 6,8-dimethyl-7-phenoxy-3-oxa-1,5,6,8-tetraaza-7-phosphabicyclo[3.3.1]nonane-7-thione, $\tau = 43°$ [O05] (Jaud, Galy, Kraemer, Majoral & Navech, 1980). (*f*) Diphenyl succinate, $\tau = 64°$ [O11] (Brisse, Molhant & Pérez, 1979) and 2-methyl-2-(2-phenoxyphenoxy)propionic acid, $\tau = 3°$ [O10] (Wägner & Malmros, 1979).

corresponds essentially to a simultaneous increase of α and decrease of β as the substituent becomes more electronegative, in agreement with previous observations (Domenicano *et al.*, 1975*a*,*b*).

Much of the variance of D_4 and D_6 is thus due to covariance and it might seem that the scatter perpendicular to the major axis of the D_4 *vs* D_6 distribution was due to experimental error. For example, high thermal motion of type (iii) (rolling) decreases D_4 without affecting D_6 so that a point representing a phenyl group librating in this manner will be displaced downwards along a vertical line.

However, in a study of monosubstituted benzene derivatives with potential π donors as substituents, it was shown that conjugation causes a peculiar type of ring distortion, involving all four internal angles to a comparable extent (Domenicano & Vaciago, 1979). This distortion is superimposed on that arising from the σ electronegativity of the substituent. The effect of conjugation on the symmetry coordinates D_4 and D_6 is shown in Fig. 7, where arrows connect data points corresponding to molecules having chemically identical (or very similar) planar substituents but markedly different twist angles and therefore different degrees of conjugation. In essentially all cases, decreasing a twist angle causes a point to move downwards almost vertically, *i.e.* decreases the value of D_4 but has only a minor effect on D_6.

If the chemical nature of the points below and to the right of the major axis of Fig. 6(*a*) is examined closely (Table 5) it can be seen that most of them correspond to π-donor substituents, *e.g.* $-NHR$ and $-NR_2$ groups nearly coplanar with the ring. Among the points to the left and above the axis we find some recognized π-acceptors. Thus, scatter perpendicular to the major axis, although much smaller than that along it, can be identified as a *real* structural variation, in part at least associated with the π-donor/acceptor properties of the substituents. And indeed with second-row substituents (Fig. 6*b*) the scatter perpendicular to the major axis is much less than with first-row substituents (Fig. 6*a*), which fits the idea that π effects are less marked with second-row elements.

The σ and π effects of the substituents, as measured from reactivity data, are generally accepted to be independent, *i.e.* they can be represented by orthogonal axes (Ehrenson, Brownlee & Taft, 1973). For our geometrical measurements the only *statistically independent* effects are the major and minor axes of the D_4 *vs* D_6 scattergram. Thus to describe distortions corresponding to these axes we can define two vectors, $\mathbf{F}_1^{\rm I}$ and $\mathbf{F}_2^{\rm I}$:

$$\mathbf{F}_1^{\rm I} = (2{\cdot}44)^{-1/2}(\mathbf{D}_6^{\rm I} + 1{\cdot}20\mathbf{D}_4^{\rm I}) \qquad (39)$$

$$\mathbf{F}_2^{\rm I} = (2{\cdot}44)^{-1/2}(-1{\cdot}20\mathbf{D}_6^{\rm I} + \mathbf{D}_4^{\rm I}). \qquad (40)$$

Using equations (7) and (11) we can express $\mathbf{F}_1^{\rm I}$ and $\mathbf{F}_2^{\rm I}$ in terms of our original basis vectors, $\boldsymbol{\theta}_i$, from which we can derive the ratios of the angular changes occurring in the two distortions, *i.e.*

$$(\mathbf{F}_1^{\rm I}) \quad \Delta\alpha{:}\Delta\beta{:}\Delta\gamma{:}\Delta\delta = 1{:}-0{\cdot}69{:}0{\cdot}06{:}0{\cdot}26 \qquad (41)$$

$$(\mathbf{F}_2^{\rm I}) \quad \Delta\alpha{:}\Delta\beta{:}\Delta\gamma{:}\Delta\delta = 0{\cdot}08{:}0{\cdot}19{:}-0{\cdot}73{:}1. \qquad (42)$$

Table 5. *Chemical identity of some major outliers in the D_4 vs D_6 scattergram for first-row substituents*

Identification number in Fig. 6(*a*)	Code in the data set	Compound	Reference	Twist angle of the benzene ring (°)
1	N27	(Diphenylamino)triphenylmethane	(*a*)	12*
2	N04	3-(Phenylamino)pyrrolidine-2,5-dione	(*b*)	18†
3	N05	Ammonium 8-anilino-1-naphthalenesulphonate hemihydrate	(*c*)	9‡
4	N07	5-Anilino-2-phenyl-2,3-dihydro-1*H*-pyrazolo[3,4-*d*]thiazol-3-one	(*d*)	25§†
5	O07	*N*,*N'*-Dibenzyl-*N*-(diphenoxyphosphoryl)thiourea	(*e*)	78¶†
6	N25	4-(3-Methyl-1-buten-2-yl)-3-phenylsydnone	(*f*)	55
7	N22	3-Phenyl-1,3-thiazolidine-2,4-dione	(*g*)	66
8	C66	Acetyl benzoyl peroxide	(*h*)	5
9	C47	Benzaldehyde azine	(*i*)	2†
10	C30	2,3-Diphenylcyclopropenone	(*j*)	3–11‖
11	C31	2,3-Diphenyl-2-cyclopropene-$\Delta^{1,\alpha}$-malononitrile	(*j*)	6#
12	B03	Phenylboronic acid	(*k*)	7,21‖

References: (*a*) Hoekstra & Vos (1975*a*,*b*). (*b*) Argay & Carstensen-Oeser (1973). (*c*) Cody & Hazel (1977). (*d*) Smith (1969). (*e*) Karolak-Wojciechowska, Wieczorek, Mikolajczyk, Kielbasiński, Struchkov & Antipin (1979). (*f*) Hašek, Obrda, Huml, Nešpůrek, Chojnacki & Šorm (1978). (*g*) Stanković & Andreetti (1979). (*h*) Karch, Koh, Whitsel & McBride (1975). (*i*) Mom & de With (1978). (*j*) Ammon (1973). (*k*) Rettig & Trotter (1977).

* The benzene ring considered here is one of the two linked to the N atom.

† Calculated from the atomic and unit-cell parameters given in the original paper.

‡ The benzene ring considered here belongs to one of the two independent anions contained in the asymmetric unit.

§ The benzene ring considered here is that of the anilino moiety.

¶ The benzene ring considered here is linked to an O atom.

‖ The asymmetric unit contains two independent molecules.

The molecule has C_2 crystallographic symmetry.

If we take the origin of the F_1/F_2 coordinate system corresponding to an undistorted benzene ring (*i.e.* $D_4 = D_6 = 0$) then by substituting equations (23) and (25) we get the scalars:

$$F_1^{\rm I} = 0{\cdot}70\Delta\alpha - 0{\cdot}97\Delta\beta + 0{\cdot}08\Delta\gamma + 0{\cdot}18\Delta\delta \qquad (43)$$

$$F_2^{\rm I} = 0{\cdot}06\Delta\alpha + 0{\cdot}26\Delta\beta - 1{\cdot}00\Delta\gamma + 0{\cdot}68\Delta\delta. \qquad (44)$$

If the above interpretation is valid F_1 and F_2 should give measures of the σ electronegativity and π-donor/acceptor power of the substituent, respectively. The fact that increasing conjugation of a planar substituent with the ring causes the corresponding point in the D_4 *vs* D_6 scattergram to move downwards and not perpendicularly to the major axis (Fig. 7) is not surprising. We expect that both the σ and π effects of a planar substituent change continuously with the twist angle, since as the exocyclic bond shortens due to conjugation the σ effect of the substituent will change as well as the π. What is particularly interesting is that equations (43) and (44) give a measure of the two effects not for a particular substituent but for a particular conformation of this substituent.

The labels 'σ-electronegativity' and 'π-donor/acceptor power' are, of course, merely labels. Nevertheless, if we reify $F_1(F_2)$ in terms of $\sigma(\pi)$ effects we expect the values of $F_1(F_2)$ for a substituent to correlate with other measures of $\sigma(\pi)$ effects, such as the $\sigma_I(\sigma_R^0)$ parameters derived from reactivity data. We shall show elsewhere that there is good general agreement between our $F_1(F_2)$ values and $\sigma_I(\sigma_R^0)$ values, and that the measurement of the geometry of monosubstituted benzene rings may serve to derive estimates of the latter.

We must be careful, however, not to assume *a priori* that F_1 and F_2 will correspond *exactly* to real chemical effects. An uneven composition of the sample (caused in part by the selection of substituents available) may have biased the analysis, and it is possible that F_1 does not correspond exactly to the σ-electronegativity axis. Our measure of the π effect is even more sensitive to the reference system, due to the smaller spread of data along the minor axis.

7. Multivariate analysis

The bivariate analysis of symmetry coordinates for monosubstituted benzene rings gave a useful reduction in the dimensionality of the problem, but one that was slightly fortuitous. It happened that the angular variance could be completely represented by the symmetry coordinates D_4 and D_6. It remains to be seen whether the statistical dimensionality of the problem increases when bond-length variation is included in the analysis. We have thus applied factor analysis to the sample of 149 benzene rings with first-row substituents, for which the spectrum of electronic effects is wider and the experimental bond lengths more accurate.

The application of factor analysis to molecular geometry has been described in detail (Murray-Rust & Bland, 1978; Murray-Rust & Motherwell, 1978*b*) and the present example illustrates it well. Any number of internal parameters (not necessarily independent) are chosen (in the present case three bond lengths and four angles) and the matrix of correlation coefficients calculated. The matrix is then diagonalized and the

Table 6. *Factor analysis of internal coordinates for benzene rings with first-row substituents*

(*a*) Factors

Factor	Eigenvalue	Percentage of variance	Cumulative percentage of variance
1	2·94	42·0	42·0
2	1·91	27·3	69·3
3	1·40	20·1	89·4
4	0·54	7·7	97·1
5	0·20	2.9	100·0
6	0·00	—	—
7	0·00	—	—

(*b*) Components of internal coordinates in the first three factors. The units are standard deviations and by multiplying each component by the standard deviation of the relevant parameter (as given in Table 1*a*) the relative importance in Å or ° is obtained

	Factor 1		Factor 2		Factor 3	
Parameter	Component	Relative importance	Component	Relative importance	Component	Relative importance
a	0·71	0·0059 Å	0·59	0·0049 Å	−0·10	−0·0008 Å
b	0·00	0·0000	0·80	0·0043	−0·01	−0·0001
c	0·00	0·0000	0·92	0·0078	0·06	0·0005
α	−0·94	−1·87°	0·21	0·42°	0·25	0·50°
β	0·92	1·09	−0·17	−0·20	−0·34	−0·40
γ	0·20	0·06	−0·06	−0·02	0·97	0·29
δ	−0·81	−0·42	0·06	0·03	−0·53	−0·28

(*c*) As above, after rotation of the factor axes (Kaiser varimax criterion)

	Factor 1		Factor 2		Factor 3	
Parameter	Component	Relative importance	Component	Relative importance	Component	Relative importance
a	−0·65	−0·0054 Å	0·66	0·0055 Å	0·10	0·0008 Å
b	0·07	0·0004	0·80	0·0043	−0·05	−0·0003
c	0·10	0·0009	0·91	0·0077	0·01	0·0001
α	0·99	1·97°	0·12	0·24°	−0·07	−0·14°
β	−0·99	−1·18	−0·08	−0·10	−0·02	−0·02
γ	0·12	0·04	−0·03	−0·01	0·98	0·29
δ	0·60	0·31	−0·03	−0·02	−0·76	−0·40

eigenvectors weighted by the eigenvalues to give the factors. The number of non-zero factors is the dimensionality of the problem (which cannot here be more than five). The results obtained here are shown in Table 6. The problem is to determine how many of the five non-zero factors represent real structural effects rather than random error.

If we know how much of the variance is due to random error we can estimate the *maximum* amount of variance due to chemical effects. The average angular variance is about $1{\cdot}4^{\circ 2}$ (Table 1*a*) whereas the variance estimated from the authors' e.s.d.'s is about $0{\cdot}1^{\circ 2}$, which would suggest that about 90% of the angular variance is not due to random experimental error. A similar calculation for bond-length variance gives a lower figure, about 70%. However, an unknown amount of these percentages will be due to pseudo-variance (see Murray-Rust & Bland, 1978) from systematic errors such as thermal motion and the inhomogeneity of the sample. It is thus unlikely that factors (4) and (5) (which jointly explain 10·6% of the variance, Table 6*a*) represent important effects. This is confirmed by two empirical guidelines commonly used in the social sciences: (i) eigenvalues below 1·0 are usually unimportant (Kaiser's criterion) and (ii) a sharp drop in eigenvalue, followed by a lower rate of decrease for the remainder, indicates the end of useful factors (the 'scree' test). We shall therefore not consider factors (4) and (5) further.

The remaining three factors are now examined to see if they can be reified in terms of chemical effects. This is often easiest by rotating the factor axes (preserving orthogonality) until they lie as close as possible to some of the original internal parameters, when interpretation is sometimes much simpler. In the present case rotation gives a marked improvement for factors (2) and (3) since many of the factor components become essentially zero (Table 6*c*). Thus factor (2) is now solely concerned with bond-length variation, whilst factor (3) is almost entirely concerned with angle variation in the *para* region of the ring.

The three factors can now all be seen to be experimentally significant. Factor (1) corresponds to the *ipso* effect caused by the σ electronegativity of the substituent (Domenicano *et al.*, 1975*a*), which can be seen to involve not only α, β and a, but also δ.* As far as angular changes are concerned, factor (1) is virtually identical with the first effect derived from the D_4 *vs* D_6 scattergram (see equation 41). Factor (2) is simply the simultaneous increase (decrease) of all the C—C bond lengths and it is likely that most of it is due to uncorrected thermal motion and inhomogeneity of the sample. Factor (3) corresponds to π-donation/acceptance and involves primarily γ and δ. The relative angular changes are similar to those associated with the second of the two effects derived from the D_4 *vs* D_6 scattergram (equation 42).

8. Conclusions

The present study shows how a systematic analysis of variance applied to the geometry of related molecules can reveal the number, nature and relative importance of the different modes of distortion from a reference geometry. The statistical techniques are most useful when, as with monosubstituted benzene rings, many accurate data are available and the observed variance is several times higher than that expected from random errors.

Multivariate analysis reveals that *three independent effects* account for about 90% of the variance in the geometry of monosubstituted benzene rings, the remaining 10% being essentially due to random

---

* In an attempt to reduce the effect of thermal motion and the inhomogeneity of the sample on bond lengths, factor analysis was also carried out on a', b' and c' (the differences of the C—C bond lengths from their mean value, see § 6.1) and α, β, γ and δ. After rotation, factor (1) contained also an appreciable component in c', but until better experimental data are available we cannot judge the significance of this.

experimental error. The effect which accounts for most of the variance (*ca* 40%) is a simultaneous change of the angles α, β and δ and of the length of the *a* bond, the largest angular change being at α, the smallest at δ. This effect is definitely related to the σ electronegativity of the substituent, and promises to be a useful indicator of this parameter. Another effect, accounting for *ca* 20% of the variance, involves mainly the γ and δ angles, and appears to be controlled to a large extent by the π-donor/acceptor power of the substituent. We emphasize that the separation of the two effects is not based on chemical assumptions, but stems directly from the statistical analysis of the data. The last effect, accounting for *ca* 30% of the variance, is a simultaneous increase (decrease) of all the C—C bond lengths of the ring. Although statistically significant, it appears to arise primarily from uncorrected thermal motion and inhomogeneity of the sample. Further progress in the analysis of bond-length variations will have to be deferred until more accurate data become available.

We thank Mrs C. Marciante for the preparation of the drawings. One of us (PM-R) thanks the British Council and the Italian CNR for a travel grant.

References

Allen, F. H., Bellard, S., Brice, M. D., Cartwright, B. A., Doubleday, A., Higgs, H., Hummelink, T., Hummelink-Peters, B. G., Kennard, O., Motherwell, W. D. S., Rodgers, J. R. & Watson, D. G. (1979). *Acta Cryst.* B**35**, 2331–2339.

Ammon, H. L. (1973). *J. Am. Chem. Soc.* **95**, 7093–7101.

Argay, Gy. & Carstensen-Oeser, E. (1973). *Acta Cryst.* B**29**, 1186–1190.

Brisse, F., Molhant, N. & Pérez, S. (1979). *Acta Cryst.* B**35**, 1825–1829.

Britton, D. (1977). *Acta Cryst.* B**33**, 3727–3729.

Cody, V. & Hazel, J. (1977). *J. Med. Chem.* **20**, 12–17.

Colapietro, M. & Domenicano, A. (1982). *Acta Cryst.* B**38**, 1953–1957.

Colapietro, M., Domenicano, A., Hargittai, I., Riva di Sanseverino, L. & Schultz, Gy. (1980). 6th Eur. Crystallogr. Meet., Barcelona, Spain. Abstr. p. 22.

Colvin, E. W., Beck, A. K., Bastani, B., Seebach, D., Kai, Y. & Dunitz, J. D. (1980). *Helv. Chim. Acta*, **63**, 697–710.

Di Rienzo, F., Domenicano, A., Portalone, G. & Vaciago, A. (1976). 2nd Yugosl.–Ital. Crystallogr. Conf., Dubrovnik, Yugoslavia. Abstr. pp. A102–A104.

Domenicano, A., Hargittai, I. & Schultz, Gy. (1980). 8th Austin Symp. Mol. Struct., Austin, USA. Abstr. pp. 17–21.

Domenicano, A., Mazzeo, P. & Vaciago, A. (1976). *Tetrahedron Lett.* pp. 1029–1032.

Domenicano, A. & Murray-Rust, P. (1979). *Tetrahedron Lett.* pp. 2283–2286.

Domenicano, A., Schultz, Gy. & Hargittai, I. (1982). *J. Mol. Struct.* **78**, 97–111.

Domenicano, A., Schultz, Gy., Kolonits, M. & Hargittai, I. (1979). *J. Mol. Struct.* **53**, 197–209.

Domenicano, A. & Vaciago, A. (1979). *Acta Cryst.* B**35**, 1382–1388.

Domenicano, A., Vaciago, A. & Coulson, C. A. (1975*a*). *Acta Cryst.* B**31**, 221–234.

Domenicano, A., Vaciago, A. & Coulson, C. A. (1975*b*). *Acta Cryst.* B**31**, 1630–1641.

Ehrenson, S., Brownlee, R. T. C. & Taft, R. W. (1973). *Prog. Phys. Org. Chem.* **10**, 1–80.

Engelhardt, U. (1979). *Acta Cryst.* B**35**, 3116–3119.

Engelhardt, U. & Hartl, H. (1975). *Acta Cryst.* B**31**, 2098–2105.

Hašek, J., Obrda, J., Huml, K., Nešpůrek, S., Chojnacki, H. & Šorm, M. (1978). *Acta Cryst.* B**34**, 2756–2759.

Hoekstra, A. & Vos, A. (1975*a*). *Acta Cryst.* B**31**, 1716–1721.

Hoekstra, A. & Vos, A. (1975*b*). *Acta Cryst.* B**31**, 1722–1729.

Høg, J. H. (1971). Thesis, Univ. of Copenhagen.

Jaud, J., Galy, J., Kraemer, R., Majoral, J. P. & Navech, J. (1980). *Acta Cryst.* B**36**, 869–872.

Karch, N. J., Koh, E. T., Whitsel, B. L. & McBride, J. M. (1975). *J. Am. Chem. Soc.* **97**, 6729–6743.

Karolak-Wojciechowska, J., Wieczorek, M., Mikołajczyk, M., Kiełbasiński, P., Struchkov, Yu. T. & Antipin, M. Y. (1979). *Acta Cryst.* B**35**, 877–881.

Kiers, C. T. & Vos, A. (1972). *Recl Trav. Chim. Pays-Bas*, **91**, 126–132.

Michel, F., Nery, H., Nosberger, P. & Roussy, G. (1976). *J. Mol. Struct.* **30**, 409–415.

Mom, V. & de With, G. (1978). *Acta Cryst.* B**34**, 2785–2789.

Murray-Rust, P. (1982). *Acta Cryst.* B**38**, 2818–2825.

Murray-Rust, P. & Bland, R. (1978). *Acta Cryst.* B**34**, 2527–2533.

Murray-Rust, P., Bürgi, H. B. & Dunitz, J. D. (1978*a*). *Acta Cryst.* B**34**, 1787–1793.

Murray-Rust, P., Bürgi, H. B. & Dunitz, J. D. (1978*b*). *Acta Cryst.* B**34**, 1793–1803.

Murray-Rust, P. & Motherwell, S. (1978*a*). *Acta Cryst.* B**34**, 2518–2526.

Murray-Rust, P. & Motherwell, S. (1978*b*). *Acta Cryst.* B**34**, 2534–2546.

Nygaard, L., Bojesen, I., Pedersen, T. & Rastrup-Andersen, J. (1968). *J. Mol. Struct.* **2**, 209–215.

Ottersen, T., Christophersen, C. & Treppendahl, S. (1975). *Acta Chem. Scand. Ser. A*, **29**, 45–50.

Rettig, S. J. & Trotter, J. (1977). *Can. J. Chem.* **55**, 3071–3075.

Schultz, Gy., Hargittai, I. & Domenicano, A. (1980). *J. Mol. Struct.* **68**, 281–292.

Schultz, Gy., Kolonits, M. & Hargittai, I. (1980). Personal communication.

Smith, D. L. (1969). *Acta Cryst.* B**25**, 625–632.

Stanković, S. & Andreetti, G. D. (1979). *Acta Cryst.* B**35**, 3078–3080.

Tamagawa, K., Iijima, T. & Kimura, M. (1976). *J. Mol. Struct.* **30**, 243–253.

Thoennes, D. & Weiss, E. (1978*a*). *Chem. Ber.* **111**, 3157–3161.

Thoennes, D. & Weiss, E. (1978*b*). *Chem. Ber.* **111**, 3381–3384.

Tucker, P. A., Hoekstra, A., ten Cate, J. M. & Vos, A. (1975). *Acta Cryst.* B**31**, 733–737.

Wägner, A. & Malmros, G. (1979). *Acta Cryst.* B**35**, 2222–2225.

Wheatley, P. J. (1959). *Struttura delle Molecole*, pp. 203–208. Roma: Accademia Nazionale dei Lincei.