Acta Cryst. (1993). B**49**, 910–928

Symmetry-Modified Conformational Mapping and Classification of the Medium Rings from Crystallographic Data. I. Cycloheptane

BY FRANK H. ALLEN*

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

AND JUDITH A. K. HOWARD* AND NIGEL A. PITCHFORD

Department of Chemistry, University of Durham, South Road, Durham DH1 3LE, England

(*Received* 16 *November* 1992; *accepted* 17 *May* 1993)

Abstract

Examples of (*a*) non-bridged and (*b*) bridged cycloheptane rings have been retrieved exhaustively from organic molecules in the Cambridge Structural Database. The 101 fragments of type (*a*) and 310 fragments of type (*b*) have been used to fill, by symmetry expansion, the hyperdimensional conformational space spanned by the intra-annular torsion angles. The resulting distributions have been examined by use of symmetry-adapted deformation coordinates, principal-component analysis (PCA) and symmetry-modified cluster analysis. The four-dimensional nature of conformational space is confirmed. The two independent pseudorotation pathways (chair–twist-chair and boat–twist-boat) can be seen to follow helical tracks on mutually orthogonal toroidal surfaces in four-dimensional space as predicted theoretically. The PCA yields four principal components, as two degenerate pairs, that can be correlated directly with the symmetry-adapted deformation modes. Cluster analysis of the conformations of non-bridged rings shows that the symmetrical low-energy forms are dominant (89%) with overall chair (C):twist-chair (TC):boat (B):twist-boat (TB) ratios of 30:48:10:1, although twelve C conformers are induced by small-ring fusion. For the bridged rings, C and B conformers dominate, with C:TC:B:TB ratios of 85:36:106:15, and distorted conformations are more frequently observed.

Introduction

A knowledge of the potential-energy hypersurface for a chemical fragment, in terms of global or local energy minima and of possible low-energy interconversion pathways, is of fundamental chemical interest. Nowadays, such knowledge is essential in three-dimensional model building using graphical and rule-based methods (see *e.g.* Dolata, Leach & Prout, 1987; Wippke & Hahn, 1988). Here, it is common practice to obtain the required information by use of a variety of computational methods. However, these methods may either be too c.p.u. intensive, or may even be totally non-feasible due, for example, to a lack of suitable force-field parameters. An alternative approach is to analyse variations in geometry that occur in the available crystallographic observations of the fragment, since it is reasonable to assume that these observations will be closely aligned with the low-energy features of the relevant conformational hypersurface (see *e.g.* Dunitz & Bürgi, 1993).

With details of more than 100 000 small-molecule organo-carbon structures now available in the Cambridge Structural Database (CSD; Allen, Davies *et al.*, 1991), it is likely that several hundreds, even thousands, of crystallographic observations may be available for a given chemical fragment. Further, a significant number of parameters may be required to define the fragment geometry. In conformational studies, it is appropriate to define the geometry of each of the N_f fragments in terms of a set of N_t torsion angles (variations in bond lengths and valence angles are ignored since the data set is chemically homogeneous), and the raw data then form a multivariate matrix $G(N_f,N_t)$. Given this abundance of information, it has proved essential to use a variety of statistical and numerical methods to organize and view these large data sets, to recognize any patterns and relationships that may occur within them, and to give statistical confidence in the results obtained (Taylor & Allen, 1993).

Two numerical techniques have proved particularly useful in the analysis of multivariate data sets in crystallographic applications. Principal-component analysis (PCA; see *e.g.* Chatfield & Collins, 1980; Murray-Rust & Bland, 1978; Auf der Heyde, 1990) is used to reduce the dimensionality of the problem from N_t to some smaller value, say N_p, and to generate graphical mappings of the data set in this N_p space. Cluster analysis (see *e.g.* Everitt, 1980;

* Authors for correspondence.

Murray-Rust & Raftery, 1985; Taylor, 1986) attempts to dissect the data set so as to locate clusters of observations (fragments) that have closely similar intracluster geometries. Whilst the basic algorithms are directly applicable to asymmetric fragments, special modifications are necessary when dealing with the (many) chemically interesting fragments that exhibit topological symmetry (see *e.g.* Norskov-Lauritsen & Bürgi, 1985; Auf der Heyde & Bürgi, 1989*a*,*b*,*c*). These modifications involve either (*a*) the expansion of the original data set prior to a principal-component analysis to include all of the conformational variants dictated by the permutational symmetry group of the fragment (Allen, Doyle & Auf der Heyde, 1991), or (*b*) extensive modification of the algorithmic basis of the cluster-analysis routines (Allen, Doyle & Taylor, 1991*a*,*b*,*c*; Allen & Taylor, 1991).

Both of these symmetry-modification techniques are being introduced into the CSD software system (program *GSTAT*; Allen, Davies *et al.*, 1991) and were tested using trial data sets of four-, five- and six-membered rings. This short series of papers extends these experiments to the medium rings of size seven and eight, and uses all available crytallographic observations, rather than selected subsets. Part I is concerned with cycloheptane fragments having topological symmetry D_{7h}; eight-membered rings, and seven-membered rings for which symmetry is reduced by the presence of heteroatoms, or by exocyclic or endocyclic double bonds, will be considered in subsequent parts.

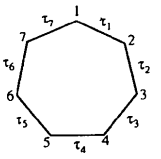
Conformational space for cycloheptane

Conformations of n-membered rings are usually described by the n intra-annular torsion angles, since these parameters are well understood and structurally informative. However, alternative descriptions of lower dimensionality are obtained by considering the out-of-plane atomic displacements (z_j) from a planar basis conformation of symmetry D_{nh} (Pickett & Strauss, 1970). The z_j transform according to those irreducible representations of D_{nh} that do not involve rotation or translation, and yield $n-3$ symmetry-adapted normal coordinates. These coordinates may be used to map the potential-energy hypersurface and to locate possible pseudorotational pathways. Later, Cremer & Pople (1975) defined the conformations of n-membered rings in terms of $n-3$ puckering coordinates relative to a unique mean plane through the ring. Recently, Boeyens & Evans (1989) have shown that these two $(n-3)$-dimensional descriptions of conformation are self-consistent, since the special conditions imposed by Cremer & Pople (1975) in defining the unique mean plane arise naturally from group-theoretical considerations.

Table 1. *Enumeration and permutations of cycloheptane torsion angles, together with numerical values* (°) *for the archetypal forms TC, C, TB, B* [(I)–(IV)] *obtained by minimum-energy calculations* (*Hendrickson*, 1967)

The coordinates θ, ρ (Bocian *et al.*, 1975) and q_2, φ_2; q_3, φ_3 (Cremer & Pople, 1975) are also given together with ΔE (kJ mol^{-1}): the energy difference from TC ($\Delta E = 0$) for C, TB and B conformers from Hendrickson (1967). Other (similar) values are given by Burkert & Allinger (1982).

(*a*) Enumeration



(*b*) Permutations

Forward							Reverse						
1	2	3	4	5	6	7	7	6	5	4	3	2	1
2	3	4	5	6	7	1	6	5	4	3	2	1	7
3	4	5	6	7	1	2	5	4	3	2	1	7	6
4	5	6	7	1	2	3	4	3	2	1	7	6	5
5	6	7	1	2	3	4	3	2	1	7	6	5	4
6	7	1	2	3	4	5	2	1	7	6	5	4	3
7	1	2	3	4	5	6	1	7	6	5	4	3	2

(*c*) Numerical values

Conf.	TC	C	TB	B
τ_1	−39.1	63.8	45.4	−57.5
τ_2	88.1	−83.5	−64.4	−30.9
τ_3	−72.3	66.1	−17.9	69.9
τ_4	54.3	0.0	74.6	0.0
τ_5	−72.3	−66.1	−17.9	−69.9
τ_6	88.1	83.5	−64.4	30.9
τ_7	−39.1	−63.8	45.4	57.5
θ (°)	50	54	~0	~0
ρ	0.80	0.74	1.15	1.15
q_2	0.52	0.43	1.15	1.15
φ_2*	90.0	0.0	270.0	0.0
q_3	0.61	0.60	0.0	0.0
φ_3*	90.0	180.0	—	—
ΔE	0.0	5.86	10.04	11.30

* See Table 2 for permutational equivalents for TC and C forms.

Other descriptions of ring pucker may also be obtained by using the torsion angles themselves as the out-of-plane deformation coordinates in a Fourier transformation (Altona, Geise & Romers, 1968; Altona & Sundaralingam, 1972).

For cycloheptane, the four sets of z_j displacements respectively transform according to the degenerate E_2'' and E_3'' representations of D_{7h}. Bocian, Pickett, Rounds & Strauss (1975) show that several descriptions of seven-membered ring conformation are equivalent to that given above. Two of these alternatives are (i) two amplitude–phase pairs ρ_2, φ_2, transforming as E_2'', and ρ_3, φ_3, transforming as E_3'', and given by

$$z_j = \rho_2\cos(2\pi 2j/7 + \varphi_2) + \rho_3\cos(2\pi 3j/7 + \varphi_3), \quad (1)$$

and (ii) a set of coordinates ρ, θ, φ_2, φ_3 (hereinafter referred to as BPRS coordinates) from

$$z_j = \rho[\cos\theta\cos(2\pi 2j/7 + \varphi_2) + \sin\theta\cos(2\pi 3j/7 + \varphi_3)], \quad (2)$$

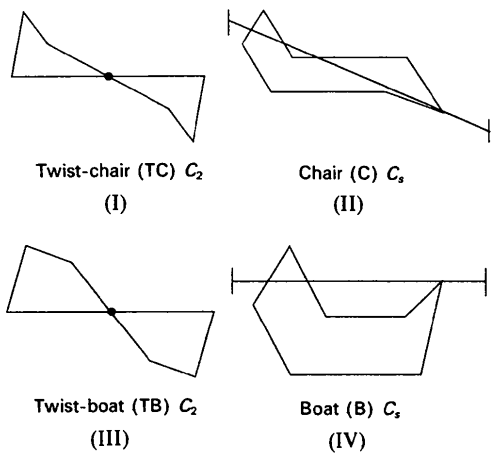
where the index $j = 1, 2, \ldots 7$ is the enumeration of the ring carbons in both (1) and (2). Cremer–Pople (1975) analysis also yields two amplitude–phase pairs q_2, φ_2; q_3, φ_3 (hereinafter referred to as CP coordinates), where the phases φ_2, φ_3 are identical to the BPRS values and q_2, q_3 are related to ρ, θ *via* the relationships (Boeyens & Evans, 1989; Boessenkool & Boeyens, 1980)

$$q_2 = \rho\cos\theta \quad \text{and} \quad q_3 = \rho\sin\theta. \tag{3}$$

Further, the BPRS coordinates ρ and θ may be expressed (Bocian *et al.*, 1975) as

$$\rho = (q_2^2 + q_3^2)^{1/2} \quad \theta = \tan^{-1}(q_3/q_2), \quad 0 \le \theta \le \pi/2. \tag{4}$$

Typical torsion angles (taken from Hendrickson, 1967), BPRS and CP coordinates for the four symmetrical low-energy cycloheptane conformers are collected in Table 1. These conformers (I)–(IV) have C_s (chair, boat) or C_2 (twist-chair, twist-boat) symmetry elements running through C1 and the midpoint of the C4—C5 bond. Thus, for each of (I)–(IV), seven conformational equivalents are generated by cyclic permutation of the atom labels. The number of equivalents rises to 14 when enantiomeric conformers are considered. For a general *asymmetric* cycloheptane, there are 14 possible cyclic permutations of atom labels (Table 1) leading to 28 possible equivalent conformations, each having a discrete sequence of torsion angles, and a discrete set of four-dimensional BPRS or CP coordinates which define the conformational hypersurface.



It is, of course, difficult to picture a four-dimensional surface in the three-dimensional minds eye. Bocian *et al.* (1975) and Bocian & Strauss (1977) use the torus of Fig. 1(*a*) to map the conformational equivalents of a given cycloheptane ring using the BPRS coordinates (see the 1977 paper for correct axial labelling). In this description, ρ specifies the degree of ring pucker, θ describes the amount of chair-like or boat-like character of the ring and φ_2, φ_3, represent the degree of twist in a given conformation. Thus, ρ, θ are constant for all 28 conformational equivalents of a particular ring and define the major ($c\rho\sin\theta = cq_3$ in CP coordinates) and minor ($\rho\cos\theta = q_2$) axes of the torus; φ_2, φ_3 position each equivalent on the toroidal surface. The arbitrary constant, c in Fig. 1, is employed (Bocian *et al.*, 1975) to avoid re-entrant tori in this visualization. It is also possible to construct a φ_2-dependent torus (major radius $c\rho\cos\theta = cq_2$, minor radius $\rho\sin\theta = q_3$) as depicted in Fig. 1(*b*). A set of cycloheptane rings may then be represented by a family of concentric tori in exactly the same way that a family of six-membered rings may be represented by a family of concentric spheres (see *e.g.* Allen & Taylor, 1991).

The chair-like conformations (I) and (II) have $\rho \simeq 0.75$ and $\theta \simeq 60°$ and it is convenient to describe the chair–twist-chair interconversion in terms of φ_2, φ_3 alone. This path is the lowest energy pathway that connects the symmetry-equivalent conformations, *i.e.* there is minimum movement of φ_2, φ_3 in proceeding from a given chair form to its closest geometrical equivalent. Bocian *et al.* (1975) and Bocian & Strauss (1977) show that the chair–twist-chair pseudorotation follows a helical pathway on the surface of the torus such that if 'origin' chair (C) and twist-chair (TC) conformers are chosen at $\varphi_2^0(\mathrm{C}) = 0$, $\varphi_3^0(\mathrm{C}) = \pi$ and $\varphi_2^0(\mathrm{TC}) = \pi/2$, $\varphi_3^0(\mathrm{TC}) = \pi/2$, then permutationally isomeric conformations (k) will occur at

$$\varphi_2^k(\mathrm{C,\ TC}) = \varphi_2^0(\mathrm{C,\ TC}) + 3\pi k/7;$$
$$\varphi_3^k(\mathrm{C,\ TC}) = \varphi_3^0(\mathrm{C,\ TC}) + \pi k/7 \quad \text{for } k = 0 \text{ to } 13. \tag{5}$$

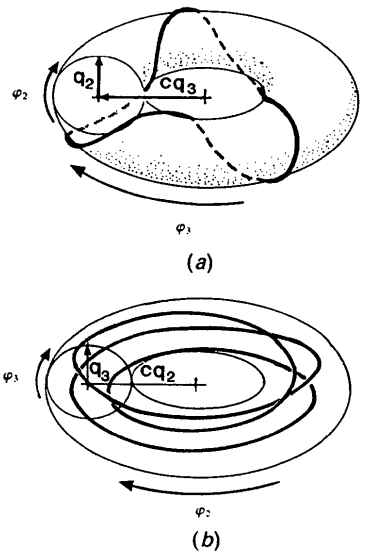


Fig. 1. (*a*) The φ_3 torus and (*b*) the φ_2 torus drawn using Cremer–Pople (1975) coordinates. The helical chair–twist-chair pseudorotation pathway is shown in (*a*), the boat–twist-boat pathway is shown in (*b*).

This places TC conformers midway between the C conformers at

$$\varphi_2^k(\mathrm{TC}) = \varphi_2^k(\mathrm{C}) + 3\pi/14; \quad \varphi_3^k(\mathrm{TC}) = \varphi_3^k(\mathrm{C}) + \pi/14. \tag{6}$$

The C–TC pathway is not isoenergetic since E(TC) is some 5.86 kJ mol^{-1} below that of E(C) (Hendrickson, 1967; see Table 1).

The boat family of conformers (III), (IV) have $\rho \simeq 1.10$ but θ is close to zero. Hence, the major radius of the torus of Fig. 1(*a*) is close to zero and the pseudorotation pathway is essentially a circle of radius ρ. The pathway is φ_2-dependent only, since φ_3 becomes indeterminate as θ approaches zero. If 'origin' boat (B) and twist-boat (TB) conformers are chosen at $\varphi_2^0(\mathrm{B}) = 0$ and $\varphi_2^0(\mathrm{TB}) = 3\pi/2$, then permutationally isomeric conformers (k) occur at

$$\varphi_2^k(\mathrm{B,\ TB}) = \varphi_2^0(\mathrm{B,\ TB}) + \pi k/7 \quad \text{for } k = 0 \text{ to } 13. \tag{7}$$

In practice, θ deviates slightly from zero for the minimum-energy forms and there is a small φ_3-dependency in the pseudorotation. Thus, the boat–twist-boat pseudorotation path may also be described as a helix, but winding around a very thin torus of major radius ρ (Fig. 1*b*) and normal to the chair–twist-chair helix in the four-dimensional hyperspace. The boat–twist-boat (B–TB) pathway is almost isoenergetic and some 10–11 kJ mol^{-1} (Hendrickson, 1967) above that of the chair–twist-chair (C–TC) pathway.

The 14 isomers of each of the C, TC, B, TB conformations that are defined by (5) and (7) correspond, of course, to the 14 unique permutations and inversions of the underlying torsion-angle sequences (see *e.g.* Table 1). This provides a method for the unique indentification of each permutational isomer, and its inverse, with reference to the C_s or C_2 symmetry element. Thus, for the fixed atomic and bond enumeration of Table 1, the torsional sequences shown correspond to the 'origin' conformers in each case. These conformers have the plane or axis of symmetry passing through C1 and the midpoint of the C4—C5 bond, the bond that carries the unique asymmetric torsion angle in each conformation. These isomers we denote as C^4, TC^4, B^4, TB^4 respectively, with their enantiomers at $\varphi_2 + \pi$, $\varphi_3 + \pi$ denoted as $C^{\bar{4}}$, $TC^{\bar{4}}$, $B^{\bar{4}}$, $TB^{\bar{4}}$. Permutation of the τ sequence merely moves the unique torsion angle to the six other possible bonds in turn, giving rise to generalized descriptors of the form C^n, TC^n, B^n, TB^n (and their enantiomers). For the C_2-symmetric TC, TB forms there is no ambiguity: positive descriptors TC^n, TB^n correspond to a positive value of the unique torsion angle, enantiomeric descriptors $TC^{\bar{n}}$, $TB^{\bar{n}}$ correspond to a negative unique torsion angle. For the C_s-symmetric C and B forms the definition depends on the signs of τ_3, τ_5 which flank the unique bond: if τ_3 is positive (hence τ_5 is negative) then positive descriptors, C^n, B^n, are assigned; for τ_3 negative (τ_5 positive) we assign $C^{\bar{n}}$, $B^{\bar{n}}$. The φ_2, φ_3 correspondences for the C/TC descriptors are presented and illustrated in Table 2 and Fig. 2, and are used to illustrate the pseudorotational pathways in Fig. 6. We note that this nomenclature differs from that proposed by Boessenkool & Boeyens (1980),

Table 2. *Values of φ_2, φ_3 (in units of $\pi/14$) for permutations/inversions of the chair and twist-chair conformations*

The column headed $\tau = 0$ indicates the position of the zero-valued torsion angle in the τ_1–τ_7 sequence for an ideal chair (see Table 1), and also gives the signs of the two torsion angles which flank this zero value; this yields the designators in the 'code' column (see text) which are then used in the visual display of pseudorotation itineraries in Figs. 2 and 6.

Chairs				Twist chairs		
φ_2	φ_3	$\tau = 0$	Code	φ_2	φ_3	Code
0	14	+4−	C^4	3	15	TC^1
6	16	−5+	$C^{\bar{5}}$	9	17	$TC^{\bar{2}}$
12	18	+6−	C^6	15	19	TC^3
18	20	−7+	$C^{\bar{7}}$	21	21	$TC^{\bar{4}}$
24	22	+1−	C^1	27	23	TC^5
2	24	−2+	$C^{\bar{2}}$	5	25	$TC^{\bar{6}}$
8	26	+3−	C^3	11	27	TC^7
14	0	−4+	$C^{\bar{4}}$	17	1	$TC^{\bar{1}}$
20	2	+5−	C^5	23	3	TC^2
26	4	−6+	$C^{\bar{6}}$	1	5	$TC^{\bar{3}}$
4	6	+7−	C^7	7	7	TC^4
10	8	−1+	$C^{\bar{1}}$	13	9	$TC^{\bar{5}}$
16	10	+2−	C^2	19	11	TC^6
22	12	−3+	$C^{\bar{3}}$	25	13	$TC^{\bar{7}}$

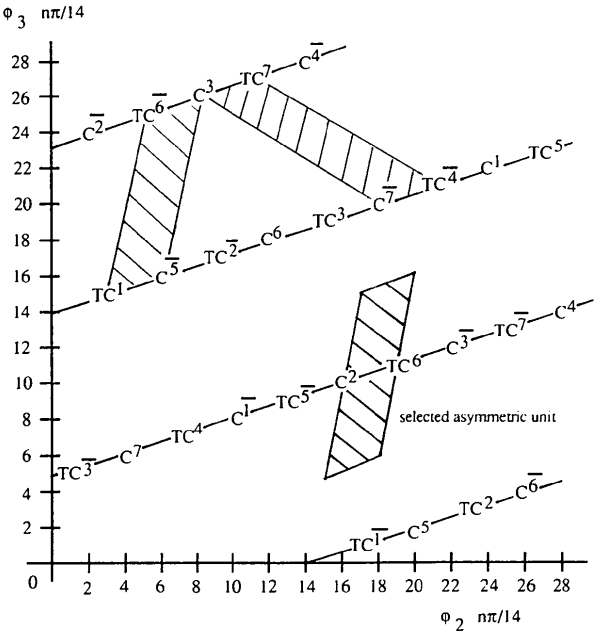


Fig. 2. The φ_2–φ_3 mapping of the chair–twist-chair pseudorotation pathway showing the 14 chair and 14 twist-chair symmetry variants whose coordinates and designators are given in Table 2.

who used the 'symmetry atom', rather than the 'symmetry bond' in their descriptors. However, we feel that a bond-centred descriptor is more natural, since torsional descriptions of conformation are themselves bond-centred.

The conformationally important C–TC pseudorotation itinerary is best illustrated (Bocian *et al.*, 1975; Bocian & Strauss, 1977) by the planar projection shown in Fig. 2, in which the lengths of the φ_2, φ_3 axes are respectively proportional to $\rho\cos\theta = q_2$ and $c\rho\sin\theta = cq_3$. The map is obtained by cross cutting the toroidal ring of Fig. 1(*a*) at $\varphi_3 = \pm 180°$, straightening to form a cylinder, slitting the cylinder lengthwise along $\varphi_2 = \pm 180°$ and flattening out to form Fig. 2. Indeed, it is convenient to *construct* the torus from the φ_2, φ_3 map, *e.g.* on tracing paper, by reversing this process: form a cylinder by glueing the $\varphi_2 = \pm 180°$ lines, then bend the cylinder into a torus and glue the $\varphi_3 = \pm 180°$ circular cross sections. It is also possible to form the φ-dependent torus of Fig. 1(*b*) by drawing the φ_2, φ_3 pseudorotation itinerary on axes proportional to $c\rho\cos\theta = cq_2$ and $\rho\sin\theta = q_3$, respectively: form the cylinder by glueing the $\varphi_3 = \pm 180°$ lines and form the torus by glueing the $\varphi_2 = \pm 180°$ circular cross sections. Both of these toroidal constructs were invaluable in visualizing and interpreting the principal component and Cremer–Pople mappings of conformational space described below.

The asymmetric unit of this conformational space must contain one of the possible 28 permutations/inversions for each ring, *i.e.* it must represent 1/28th of the total volume of the conformational space. Conversely, the entire space should be mapped by the operation of 28 symmetry elements of the D_{7h} topological point group acting on the chosen asymmetric unit. In this toroidal space, ρ is a size parameter, whilst θ must be allowed to range over all permissible values (0 to $\pi/2$). Hence, definition of an asymmetric unit depends upon the choice of suitable values for φ_2, φ_3, such that chair, twist-chair and their inverse conformers are fully represented. Three possible choices are indicated on Fig. 2. Each of these parallelograms has an area of $\pi^2/7$ rad^2, or 1/28th of the complete area ($4\pi^2$ rad^2) of the φ_2, φ_3 map.

Methodology

Version 4.5 (1 January 1992) of the Cambridge Structural Database System was used throughout for substructure search, coordinate retrieval and data analysis using the programs *QUEST* and *GSTAT* (Allen, Davies *et al.*, 1991). Local software amendments, not yet available in the distributed system, are summarized below together with precise definitions of search fragments.

Database search and retrieval

The chemical substructure studied here consisted of a ring of seven C_{sp^3} atoms connected by single bonds. The search was further constrained by use of the CSD bit-screen search mechanisms (Cambridge Structural Database, 1992) to locate entries with (*a*) atomic coordinates available, (*b*) no residual numerical errors following CSD check procedures, (*c*) no reported disorder in the crystal structure, (*d*) a crystallographic $R \le 0.10$, and which were (*e*) 'organic' compounds according to CSD definitions. Entries which contained very complex ring systems, *i.e.* those that failed CSD ring-identification procedures – less than 3% of the complete database, were also excluded *via* a SCREEN −620 setting.

Two searches were made: 7C1 in which entries containing bridged rings were avoided (SCREEN −622), and 7C2 which consisted entirely of entries containing bridged rings (SCREEN 622). This is a non-specific screen, *i.e.* it does not indicate *which* of the rings in a given molecule in bridged. Hence, chemical diagrams for subset 7C2 were examined and two entries which contained unbridged seven-membered rings were transferred to subset 7C1. As a result, 86 entries containing 101 unbridged cycloheptane fragments were included in data set 7C1 and 276 entries containing 310 bridged fragments were included in data set 7C2. CSD reference codes for data sets 7C1 and 7C2 are given in Table 3 and full citations have been deposited.*

Data analysis

The cycloheptane fragments were re-located in the crystallographic connection tables of data sets 7C1, 7C2 by use of program *GSTAT* using bond-length constraints (1.44–1.70 Å) to approximate a C_{sp^3}—C_{sp^3} bond length. It was further required that no atomic overlap should exist between multiple occurrences of the fragment in the same molecule. The conformation of each fragment which passed these criteria was described by the seven intra-annular torsion angles τ_1 to τ_7 (Table 1). The fragment location procedure of *GSTAT* will place each fragment in some arbitrary asymmetric unit of conformational space, hence it was necessary to fill this space according to the D_{7h} topological symmetry of the planar search fragment and the inverted conformations in three dimensions [see Allen, Doyle & Taylor (1991*a*), Allen & Taylor (1991) and Allen, Doyle & Auf der Heyde (1991) for a full discussion in the analogous case of cyclohexane]. For the conformational mapping described in

* Full literature citations have been deposited with the British Library Document Supply Centre as Supplementary Publication No. SUP 71142 (26 pp.). Copies may be obtained through The Technical Editor, International Union of Crystallography, 5 Abbey Square, Chester CH1 2HU, England.

Table 3. *CSD reference codes for structures used in conformational analysis of cycloheptane rings*

Data set 7C1						
AUTUMN	BUHXAD	DASPAT	FIYJUS	HYMINB	PLENOL10	STRAMB10
BAHIAL	CAZJUI	DIPHEP10	FOSFAU	HYMXON	PULCHL10	TAGHUE
BAHMIG	CEBBUG	DIPHIT10	FUGFUI	HYSTER	RADIAT	TDCOBZ
BAKVAK	CEWSAY10	DITYEK	GEXXOW	IPTDCN	RUDMOL	VALGEU
BERDAD	CEWSEC10	DIYKEB	GEYNAZ	JATGUG	SAJPAU	VALGIY
BEXFUF	CHBRBZ	FABFET	GUANLD	JEHMEO	SAJPEY	VAWBAW
BEXGAM	CIFBIC	FEGWAP	HAZFNO	JIDBON	SCHOHX10	VEKGAT
BIGVOC	CILXEA	FEGWAP01	HBCDEO	KAYVIP	SEGFUF	VEMMOP
BIKNIS	CPMOXZ	FEYDAO	HBCUDO	KAZNOO	SEHBOW	
BITENL10	CSNLRB	FGUOLD	HDAPTX	KEJPEU	SEZLUE	
BOLKUI	CUFVIII0	FIBSOY	HOXTHD	KEKXON	SILYIV	
BOLWUU	DACFOC	FIBSOY10	HXMXCT10	KIMKOG	SPDXLN	
BRMEXE10	DANNAH	FIWMIH	HYMINA	KRLAPY	SPICHB10	
Data set 7C2						
ABCOCX	BUXRER	CUTVIW	FAPKUC	GRAYAN	KIKZOT	SIFNOK
ACONTN10	BUXRER10	CUVTES	FAPLAJ	HFPCDN	KIKZOZ	SINXAO
ACONTO10	BZOTRP	CUWHOR	FEMJUC	HMCYDL10	LAPACN10	SINXES
ADHKAU	BZTRMS	CUXTOE	FEMKAJ	HMPCDO20	LEUCOA	STPHSN
AJMALN10	CAFCIV	CXMTUN	FEMKEN	HMSDCN10	LEUCOB	TCNONC
ANABSN10	CALKEF	DAFGAS	FEMKIR	HMTCDO	LYCOTD	TCTCDC
BABKEU	CALKEF10	DAFWAI	FEXKOI	IPCUND	MBOXPH	TRPHDT10
BAHCOC	CANHUU	DAKWUH	FEYSEH	JAJBEB	MBZAZL	TSBEYA
BAZFEN	CANLEI	DAPLAC10	FIDXUL	JAJVOF	MCUNDB	TZPHDO
BCUNDL	CANNIO	DATMOA	FIJYIG	JALFUX	MENTCN	VACZEE
BCYLON10	CAXPOG	DEBPAB	FIKXAY	JAPBIL	NHMENT	VAHWUW
BECKOJ	CDBOTN	DECZAM	FINLOO	JAPBOR	NMHXNT	VAHXAD
BEHCIA	CECKUQ	DELNUD10	FIVYEO	JAPLUH	NORCBR	VATJIJ10
BEHDUN	CEGZAP	DELPSC10	FIZPIN	JATWEG	OCTHEN	VATMAE
BEKGUT	CEKZIB	DELRUH	FIZVUF	JAYRUW	OHPHYL	VEDTIH
BEMOTD	CEXCIR	DEXWEI	FIZXUH	JAYZEO	OHRPRT	VEWKUD
BERLIT	CEXGER	DICKUV	FIZYAO	JENFAJ	PCCHLD	VEYPEU
BEXGUG	CHASBC10	DITCDC	FOLNAV	JEPMOG	PEIAJM10	VIGWUD
BEYMIB10	CHTDMS10	DITDUF	FOVSUE	JEPMUM	PHALDR	VIKNEI
BHPCHD10	CIJSET	DIWKOJ	FOVTIT	JIDNAL	PHMRNT	VIKYAP
BHYOSM	CIKSAQ	DIWLOK	FOWZAS	JIFRAR	PHYBAC	VIKYET
BIBXEP	CINZEY	DIWLUQ	FOWZAS10	KAMBUV	PIERIG	VIKYIX
BIGSIT	SITHUI	DLPHFL10	FOYPIS	KAMCAC	PIMARB	VIKYOD
BIRGEO	CITJAQ	DOBHIL	FUDHAN	KASNAT	PTSANO	VIKYUJ
BLONGA10	CIZWEN	DOFBEF	GADSIN	KAVDOA	SAJWIJ	VILCOI
BLONGF	CNADMT	DOFHUB	GAHGAX	KAVDUG	SAKDOX	VINVAP
BNORSO10	COCAIN10	DOKJUI	GAKPAJ	KAYSEI	SALTII	VINVET
BOBCIE	COCMEI	DONFER	GANCED	KECGOO	SALXOS	VUWZES
BOBCIE10	COCSES	DONHET	GANCIH	KEGJIP	SAMWAE	VUYYUJ
BODBEB	COCSES10	DOTHUP	GARCAD	KELYUV	SAPLOK	XMTDIB
BORRIJ	COLBAG	DUGKOF	GAWNOH	KERYIP	SAPMAX	COYFEB
BOTMOM	COLBEK	DUMRAE	GAYDIT	KETFIY	SARBIW	FATKAM
BOTNUT	COXNIM	DUMRAF	GAZSIJ	KETREG	SBLVES	FOJPID
BRLCTA	CPETLD	DUTNIP10	GEBLEE	KETRIK	SEGKOE	GRYNTX
BRLOFE10	CPRSBR	DUYREU	GERZIM	KETRUW	SEJPOM	SEMKOK
BRTRPN	CSNLRA	DUYRIY	GETCAJ	KETSAD	SEVFUU	SEMLAX
BTCUND	CTCDEC	DUZHAH	GIBJAC	KIDMIT	SEYZOL	
BUCHAI	CUCYII	ETCOGR10	GIGJEL	KIKYOS	SEZNUG	
BUHXEH	CUGLEV	EUPALB	GIHXAW	KIKZAF	SIFFIW	
BUMFEU	CUGLOF	FAJHAZ	GILYUV	KIKZIN	SIFFOC	

the next section, the initial data sets of 101 (7C1) and 310 (7C2) fragments were expanded to 2828 (7C1) and 8680 (7C2) fragments using an experimental *EXPAND* feature in *GSTAT*. This operates by applying a set of atomic permutations and coordinate inversions defined by the user. For each fragment in the expanded data set, the CP parameters, q_2, φ_2, q_3, φ_3 were generated by *GSTAT* and the BPRS coordinates, ρ, θ were calculated (4) using the *GSTAT* TRAnsform facility. Classification of conformations was performed using the symmetry-modified cluster-analysis algorithms described in detail elsewhere (Allen, Doyle & Taylor, 1991*a*,*b*,*c*; Allen & Johnson, 1991; Allen & Taylor, 1991).

Isolation of a single asymmetric unit

In this study, it has proved useful to display and examine the conformational relationships that exist between fragments that fall within a single asymmetric unit of the four-dimensional hyperspace. This unit may be isolated from an expanded data set of conformational descriptors, *e.g.* the CP or BPRS coordinates, by specifying appropriate limiting values for these descriptors. For the smaller carbocycles of size four, five or six, an asymmetric unit may be isolated, as 1/16, 1/20 or 1/24th of the respectively one-, two- or three-dimensional conformational spaces, by using the SELect command in *GSTAT* to define limiting values parallel to orthogonal descriptor axes. For seven-membered rings, the definition of the asymmetric unit (Fig. 2) is in terms of two non-orthogonal axes, neither of which is parallel to φ_2 or φ_3. To accommodate this situation *GSTAT* has been locally modified to permit the specification of four sets of x, y coordinates which define the corners of a non-orthogonal planar unit.

Fragments whose coordinates fall outside the specified unit are rejected. This philosophy will be upgraded to cover a non-orthogonal three-dimensional unit under the new *GSTAT* command BOX. However, in the present case x, y specifications in terms of φ_2, φ_3 only are required, since all values of ρ, θ within those limits define the four-dimensional asymmetric unit.

Conformational mapping

BPRS coordinate plots

The conformational diversity of data sets 7C1 and 7C2 was initially explored *via* plots derived directly from the BPRS coordinates of (2) as exemplified in Figs. 3 and 4. The symmetry-expanded φ_2–φ_3 map for 7C1 (Fig. 3*a*) has the structure expected from the schematics of Fig. 2. The C/TC pseudorotation pathway is the major continuous feature of Fig. 3(*a*), and density along this pathway rises to maxima at the TC positions. The φ_2–θ map (Fig. 3*b*) shows the clear θ separation of the C/TC ($\theta \simeq 30$–$90°$) and the B/TB ($\theta \simeq 0$–$30°$) conformers. These limiting θ values are used to segment Fig. 3(*a*) into two components in Figs. 3(*c*) ($\simeq$C/TC conformers) and 3(*d*) ($\simeq$B/TB conformers).

In Fig. 4, the scope of the mapping is restricted to an asymmetric unit only, as defined by BOX coordinates (in units of $\pi/14$) of 15,5; 18,6; 20,16; 17,15. Figs. 4(*a*), 4(*c*) and 4(*e*) are respectively φ_2–φ_3 and φ_2–θ plots and the θ histogram for data set 7C1; comparable plots for data set 7C2 are shown as Figs. 4(*b*), 4(*d*) and 4(*f*). The φ_2–θ plots (Figs. 4*c* and 4*d*) are particularly informative. For data set 7C1 (Fig. 4*c*), there is density leading from the dominant TC form at $\theta \simeq 50°$, $\varphi \simeq 240°$ through intermediate forms to a chair form at $\theta \simeq 60°$, $\varphi_2 \simeq 212°$. There is

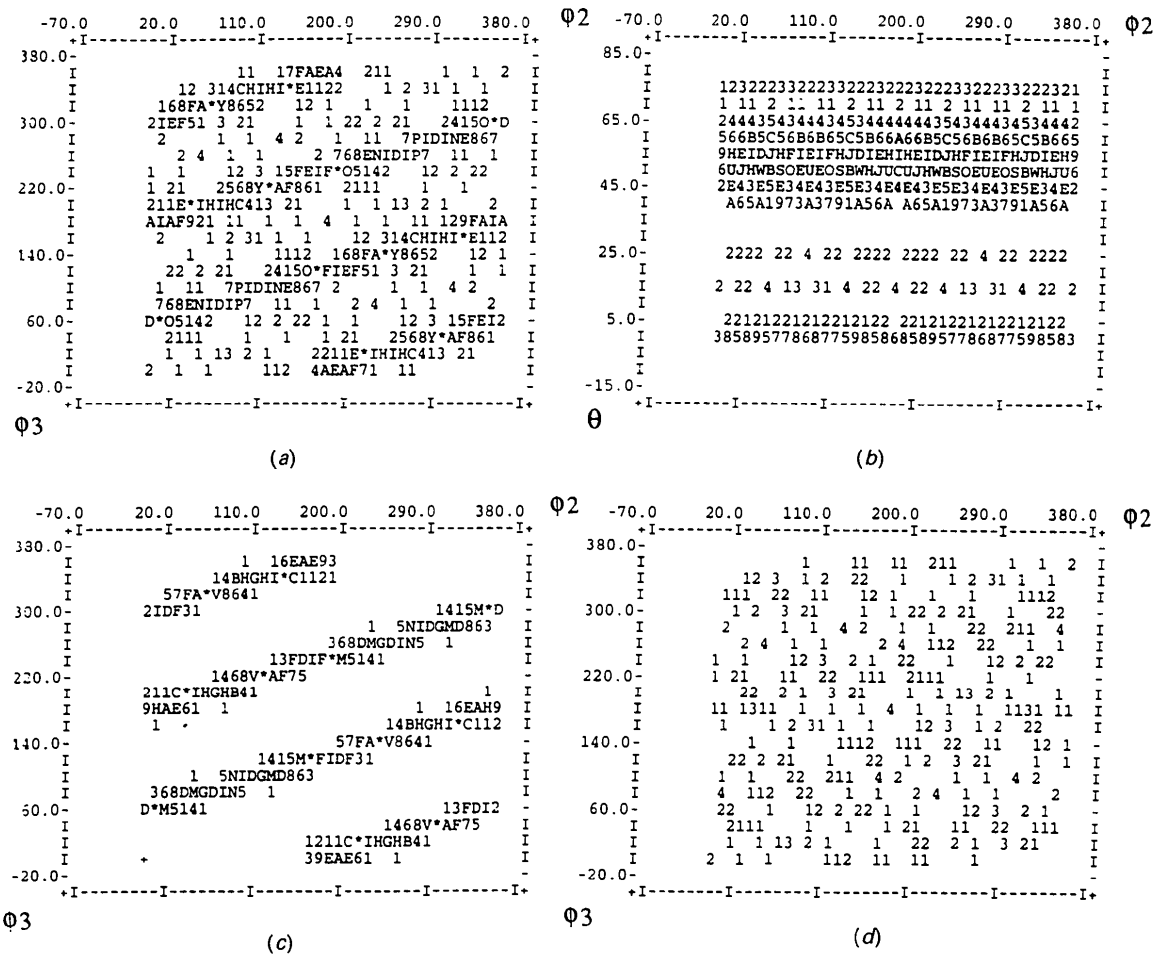


Fig. 3. Conformational plots from symmetry-expanded BPRS coordinates for data set 7C1. The φ_2–φ_3 plot (*a*) and the φ_2–θ plot (*b*) include all data points. The φ_2–φ_3 plot (*c*) covers rings having $30 \le \theta \le 90°$ *i.e.* chairs and twist-chairs, whilst the φ_2–φ_3 plot (*d*) covers rings with $0 \le \theta < 30°$ *i.e.* boats and twist-boats.

only a single small break in this transition at $\varphi_2 \simeq$ 228–234°. Very few fragments (15 out of 101 = 14.9%) do not fall on the C/TC pathway of Fig. 3(*c*), and only 11 (10.9%) may be classed as true B/TB conformers by virtue of θ values in the range 0–10° (Fig. 4*e*). For bridged rings of data sets 7C2 (Fig. 4*d*) the φ_2–θ conformational mapping is somewhat different. The higher energy C form is now dominant in the C/TC pathway, but with a smaller θ value ($\simeq$45–50°) than for the C forms of 7C1. The TC

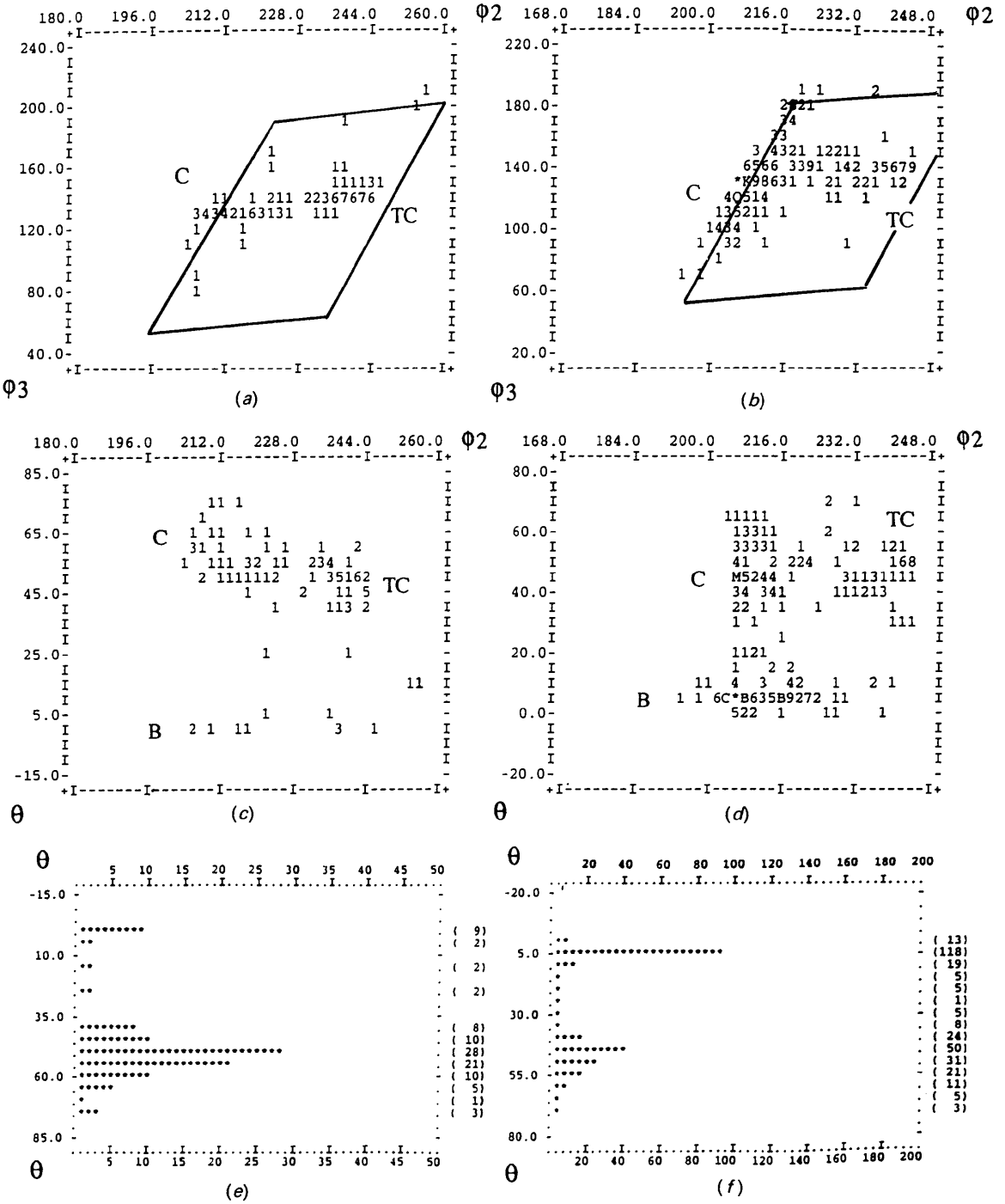


Fig. 4. Conformational plots (BPRS coordinates) of the asymmetric unit of conformational space for cycloheptane (see text). (*a*), (*c*), (*e*) Data set 7C1; (*b*), (*d*), (*f*) data set 7C2.

conformers in 7C2 appear at the expected θ value of *ca* 50°. A more obvious difference between Fig. 4(*c*) and Fig. 4(*d*) is the increase in B/TB conformers in data set 7C2, with 149 of the 310 fragments (48.1%) having θ in the range 0–10° (Fig. 4*f*). The plots of Fig. 4 were particularly useful in assessing the cluster-analysis results, which are described later.

Cremer–Pople (CP) coordinate plots

The two CP phase-amplitude pairs (q_2, φ_2), (q_3, φ_3) describe two circular pseudorotation pathways. These coordinates may be re-expressed in Cartesian form as CP1 = $q_3\sin\varphi_3$ (the set of points along the CP1 axis mapping a family of perfect TC conformers), CP2 = $q_3\cos\varphi_3$ (mapping a family of perfect chairs), CP3 = $q_2\sin\varphi_2$ (mapping a family of perfect TB conformers), and CP4 = $q_2\cos\varphi_2$ (mapping a family of perfect boats). The six possible two-dimensional scatterplots that may be obtained from these Cartesian values for the 2828 fragments of the symmetry-expanded data set 7C1 are shown in Fig. 5(*a*).

The true four-dimensional nature of conformational space for cycloheptane is apparent in the plots of Fig. 5(*a*), since all four of the CP axes are mutually perpendicular. The CP1–CP2 plot (top) may be obtained by projection along either of CP3 and CP4. The C/TC pathway forms the dense outer circle with the B/TB examples clustered in the centre of the plot. Two fragments with θ in the range 10–30° form an intermediate circle between these two extremes. With respect to the toroidal models of Fig. 1, the CP1–CP2 plot represents a view along the axis of a cylinder obtained by cutting and straightening the φ_2 torus of Fig. 1(*b*): the φ_3-dependent C/TC helical pathway now appears as a circle. This situation is reversed in the CP3–CP4 plot (Fig. 5*a*, bottom), obtained by projection along CP1 or CP2. The few B/TB conformers define a narrow band of density which surrounds a dense central region defined by the C/TC conformers. The B/TB examples do not define a perfect circle because of a 'tailing effect' caused by θ angles close to 0°, as discussed by Bocian *et al.* (1975) and by Bocian & Strauss (1977).

The four remaining plots of Fig. 5(*a*) represent two side views of the φ_2 torus and two of the φ_3 torus, each view of each pair being at 90° to the other. This can be confirmed by constructing the two tori from some transparent material (as described earlier) and observing projections of the helical pathways at various points around the major circumferences of the tori. In each plot, the B/TB pathway appears as a vertical central band of density which is crossed by orthogonal views of the C/TC helix.

Table 4. *Results of symmetry-modified principal-component analysis for data sets 7C1 and 7C2*

Variance analysis

Data set	No. of fragments	Variance (%) PC1	PC2	PC3	PC4	Total
7C1	2828	40.08	40.08	9.88	9.88	99.92
7C2	8680	26.94	26.94	22.86	22.86	99.60

PC loadings

	Data set 7C1 PC1	PC2	PC3	PC4	Data set 7C2 PC1	PC2	PC3	PC4
τ_1	55.4	11.1	3.8	27.8	−9.2	−51.5	4.7	−47.9
τ_2	−54.7	14.0	26.2	−9.8	−14.0	50.4	−47.8	6.1
τ_3	43.2	−36.4	−15.4	−23.4	34.5	−39.3	16.6	45.2
τ_4	−23.2	51.5	−19.4	20.3	−48.1	20.4	40.4	−26.2
τ_5	−1.5	−56.5	24.1	14.4	52.2	2.5	−34.5	−33.5
τ_6	25.8	50.2	8.7	−26.7	−46.0	−24.9	−25.0	41.1
τ_7	−45.1	−34.0	−27.9	−2.5	30.6	42.4	45.7	15.2
Sym.	C	TC	B/TB	B/TB	TC	C	B/TB	B/TB

Symmetry-modified principal-component analysis

PCA was performed on raw data matrices consisting of the seven torsion angles (τ_1–τ_7) for each of the 2828 (8680) fragments in the symmetry-expanded data sets 7C1 (7C2); results are collected in Table 4. Variance in the seven-dimensional torsional data set for 7C1 is fully (99.9%) accounted for by four PC's which occur as two degenerate pairs, as expected from the group-theoretical analysis of Bocian *et al.* (1975). Thus, PC1, PC2 each account for 40.1% of total variance, and PC3, PC4 each account for 9.9%. The PC loadings for 7C1 (Table 4) show that values for PC1 have approximate C_s symmetry and this axis is closely associated with chair conformers; values for PC2 have approximate C_2 symmetry and this axis is closely associated with a TC conformer at 90° in φ_3 from the PC1 chair. The loadings for PC3, PC4 are essentially asymmetric and are associated with intermediate conformations, separated by 90° in φ_2, on the B/TB pathway. Thus, the PCA for data set 7C1 indicates that some 19.8% of variance is associated with B/TB conformers. This value should be compared with the 15.9% observed in the BPRS plots of Fig. 4. However, C/TC conformations have some degree of B/TB character (and *vice versa*) and this factor is not taken into account in the BPRS plots based upon θ alone.

PCA of the symmetry-expanded data set 7C2 again reveals four PC's as two degenerate pairs, accounting for 99.6% of variance (Table 4), but with a variance distribution and PC loadings that differ from those for 7C1. Here PC1, PC2 are again associated with C/TC conformers, but with PC2 mapping a chair and PC1 mapping a twist-chair; PC3 and PC4 again map B/TB intermediates. More importantly, some 45.7% of total variance in the 7C2 data set is now associated with the B/TB (PC3, PC4) forms, a value which is again comparable to the 48.1% observed in the BPRS θ histogram (Fig. 4*f*). The PCA shows quite clearly that bridging of the

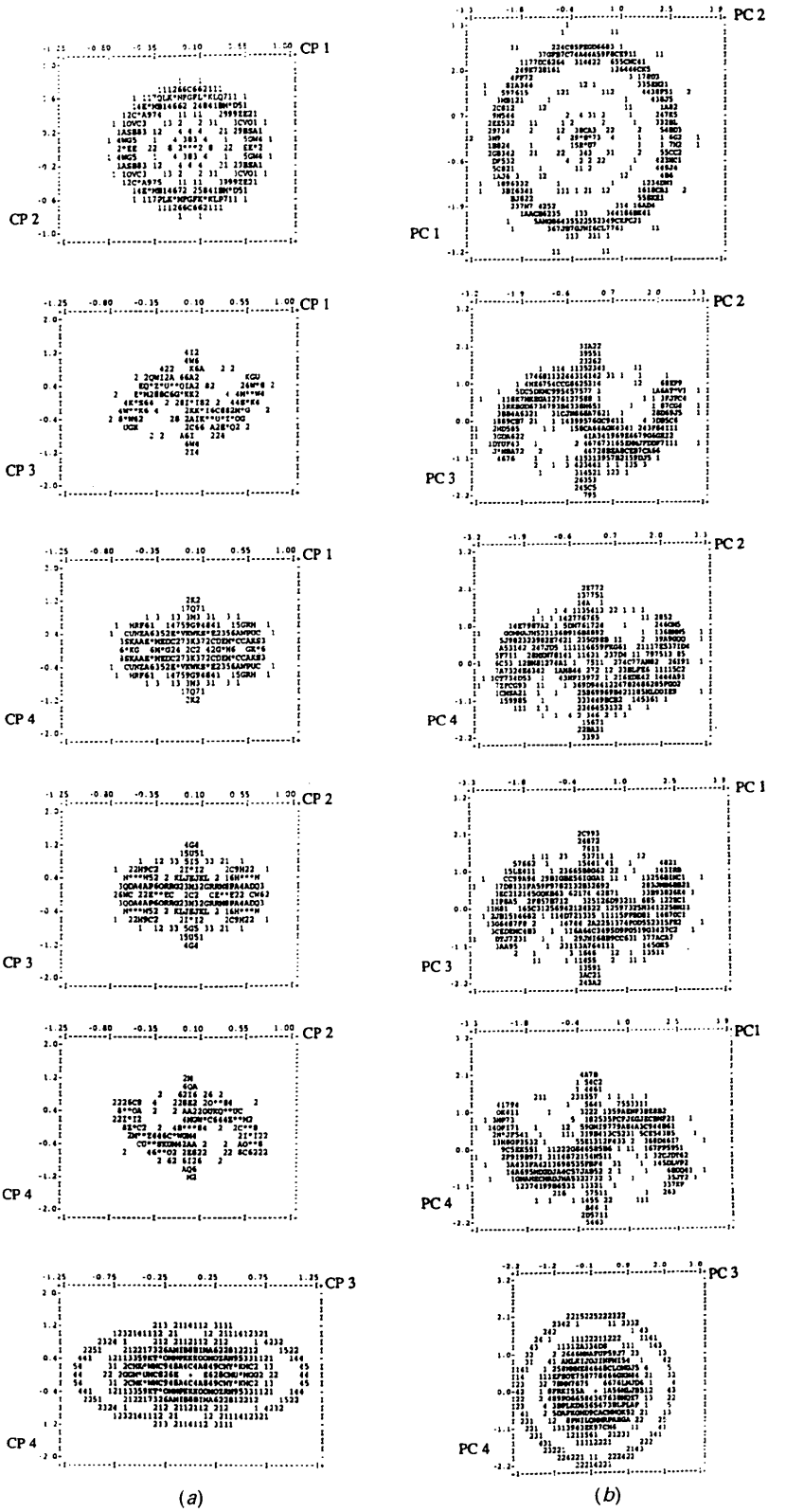


Fig. 5. Comparative two-dimensional scattergrams of (*a*) Cremer–Pople (1975) CP coordinates and (*b*) principal-component (PC) coordinates (or 'scores').

cycloheptane ring increases the proportion of B/TB conformers very significantly.

Comparison of Cremer–Pople (CP) and principal-component (PC) mappings

The two-dimensional scatterplots derived from the PC coordinates of the symmetry-expanded 7C1 data set are illustrated in Fig. 5(*b*). Axial assignments in Fig. 5(*b*) are chosen so as to align chair, twist-chair, *etc.* conformations most closely with their CP equivalents: interchange of PC1/PC2 and PC3/PC4 is perfectly permissible because of their pairwise degeneracy (see below). The plots of Figs. 5(*a*), 5(*b*) illustrate an obvious functional equivalence between the four-dimensional PC description and the four-dimensional PC and BPRS descriptions. Similar equivalences are observed for the two- and three-dimensional conformational spaces of five- and six-membered carbocycles (Allen, Doyle & Auf der Heyde, 1991) and arise from the close mathematical analogies between PCA and normal coordinate analysis.

It is clear, however, that the PC plots of Fig. 5(*b*) are slightly rotated with respect to the CP plots of Fig. 5(*a*). This arises from the pairwise degeneracy of PC1/PC2 and PC3/PC4 so that their eigenvectors can adopt any orientation within the plane they define. The axial directions obtained in a given PCA are essentially accidental and depend upon round-off and similar numerical effects. The degeneracy explains, of course, the association of PC1 with chair conformations for data set 7C1, an association which is reversed in data set 7C2.

The degree of rotation of the PC axes with respect to the CP axes in our experiments can be deduced from the correlation coefficients, R(CP,PC), calculated from the CP and PC coordinates for each fragment. The R values for data set 7C1 are given in Table 5 (current *GSTAT* program limitations prevent their calculation for the 8680 fragments of data set 7C2), and the inverse cosines $\cos^{-1}[R(\mathrm{CP,PC})]$ indicate the relative orientations of the CP and PC axes. As expected, correlations of the form, R(CPn,PCm) for $n = 1, 2$, $m = 3, 4$ (and *vice versa*) are all zero, indicating the orthogonality of these axial pairs. However, Fig. 6(*a*) shows that PC1 is rotated by 1.52° from a CP chair axis at $\varphi_3 = 2\pi/7$ (51.4°) and PC2 is rotated by an identical angle from the orthogonal twist-chair axis at $\varphi_3 - 25\pi/14$ (321.4°). For the B/TB φ_2 pathway of Fig. 6(*b*), PC3 is rotated by 5.12° from a CP twist-boat axis, with PC4 rotated by 5.12° from an orthogonal boat axis. It is these arbitrary axial rotations that explain the asymmetric PC loadings of Table 4 and the minor comparative discrepancies in Fig. 5.

Rotation of the PC1, PC2 axes by 1.52° and of the PC3, PC4 axes by 5.12° will bring the PC axis set into alignment with one of the cokernel symmetry directions of the D_{7h} point group (see Murray-Rust, Bürgi & Dunitz, 1979). There are 14 such directions separated by rotations of $2\pi/14$ rad in Figs. 6(*a*) and 6(*b*). For our calculations, it is obviously possible to rotate PC1, PC2 by $(1.52 + 360n/14)°$, and PC3, PC4

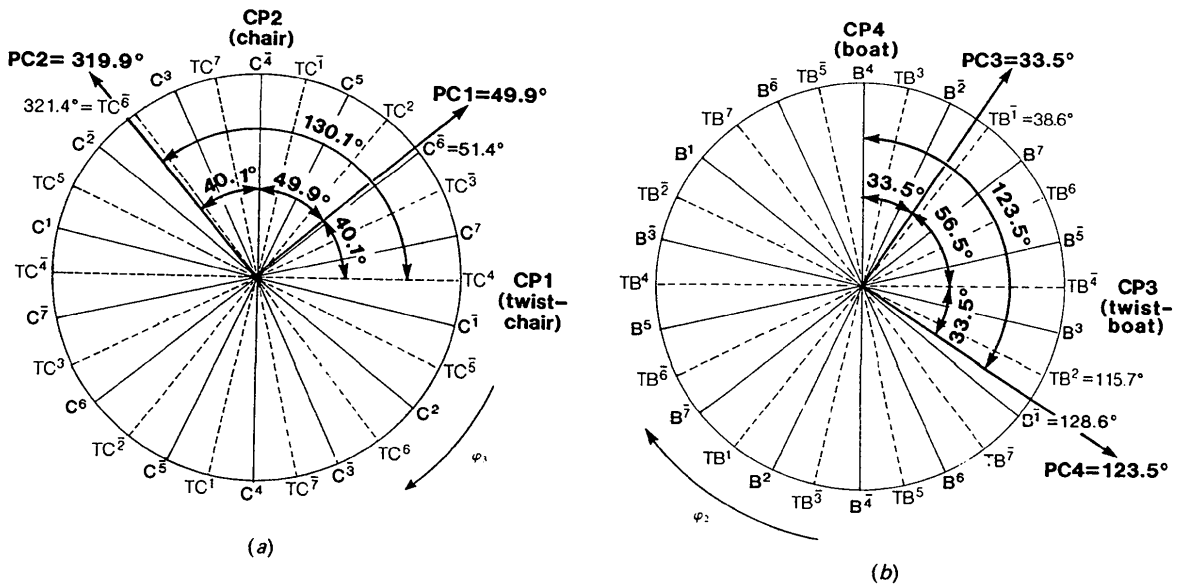


Fig. 6. Correlation of principal-component (PC) axes with the Cremer–Pople (1975) (CP) axes for (*a*) the chair–twist-chair pseudorotation pathway and (*b*) the boat–twist-boat pseudorotation pathway. Values of φ_3, φ_2 in (*a*), (*b*) for the ideal symmetry variants are given in Table 2.

Table 5. *Correlation coefficients $R(PC_n,CP_m)$ between principal-component scores for PC_n and Cremer–Pople coordinates (Cartesian form) for CP_m derived for data set 7C1*

PC_n		CP_m			
n		$m=1$	2	3	4
1	R	0.765	0.644	0.0	0.0
	$\cos^{-1}(R)$	40.1	49.9	90.0	90.0
2	R	−0.644	0.765	0.0	0.0
	$\cos^{-1}(R)$	130.1	40.1	90.0	90.0
3	R	0.0	0.0	0.552	0.834
	$\cos^{-1}(R)$	90.0	90.0	56.5	33.5
4	R	0.0	0.0	0.834	−0.552
	$\cos^{-1}(R)$	90.0	90.0	33.5	123.5

by $(5.12 + 360n/14)°$, so as to bring these axes into exact coincidence with the CP axes. Under these conditions, the correlation coefficients R(CP1,PC1) and R(CP3,PC3) will both be 1.0 and R(CP1,PC2) and R(CP3,PC4) with both be 0.0. Rotation of the PC eigenvectors to any of the cokernel directions will also generate PC loadings (eigenvalues) that have perfect C_s and C_2 symmetry.

Three-dimensional scatterplots of principal-component scores

In order to appreciate more clearly the axial relationships in four-dimensional space that lead to the two-dimensional scatterplots of Fig. 5, we have isolated two fragments, one an almost perfect chair and the other an almost perfect boat. The symmetry of these individual conformers generates 14 twofold degenerate equivalents in their symmetry expansions. PC scores for these C and B equivalents were input as sets of three coordinates to the plotting program *MacMoMo* (Dobler, 1990) on an Apple Macintosh computer. Pseudomolecules were constructed by joining adjacent equivalents for the two conformers so that various rotations of the pseudorotation pathways could be observed. The PC axes in this analysis of 56 fragments (28 isomers for each of two rings) were very closely aligned with a cokernel direction, so that further PC-axis rotations were deemed unnecessary.

In Fig. 7(*a*), the scores for PC1, PC2, PC3 are used as pseudocoordinates. The first five frames show how the PC2, PC1 plot of Fig. 5(*b*) transforms into the PC2, PC3 plot of Fig. 5(*b*). The chair pseudomolecule forms the outer 'circle' in the 0,0,0 position: a top view of the φ_3 helix of Fig. 1(*a*). The boat pseudomolecule forms the inner 'circle' at 0,0,0: a top view of the φ_2 helix of Fig. 1(*b*). Rotation through 90° about the PC2 axis brings the original projection axis, PC3, into the viewing plane to generate side views of both helices in frame five of Fig. 7(*a*). The PC2 rotations of frames 2, 3, 4 show how the B/TB helix in θ_2 rapidly elongates to an apparent vertical straight line as a result of the minimal dependency of this helix on φ_3 variations. Frames 6–9 of Fig. 7(*a*) show (right to left) the transformation of the PC2, PC3 plot to the PC1, PC3 plot *via* a 90° rotation about the PC3 axis. This is a rotation of the edge-on views of the φ_3 (C/TC) and φ_2 (B/TB) helices which are mutually perpendicular in this four-dimensional hyperspace.

Fig. 7(*b*) and Fig. 7(*c*) use the scores for PC2, PC3, PC4 and PC1, PC3, PC4, respectively. Both begin (at 0,0,0) with identical projections onto the PC3, PC4 plane along the mutually orthogonal PC2 (Fig. 7*b*) and PC1 (Fig. 7*c*) axes. The boat pseudomolecule forms the outer helical projection, and the chair pseudomolecule forms the inner helical projection, in these first frames. A rotation of 90° about PC4 in Fig. 7(*b*) then generates the PC2, PC4 plot (compare with Fig. 5*b*) and a rotation of 90° about PC4 in Fig. 7(*c*) generates the PC1, PC4 plot (again, compare with Fig. 5*b*). Thus, Fig. 7 provides a graphic illustration of the inter-relationships between the four mutually orthogonal axes which describe the conformational hypersurface of cycloheptane.

Classification of conformations

The conformational maps of a single asymmetric unit (Figs. 4*a*,*c*,*e*) show quite clearly that twist-chairs and chairs are the dominant conformers observed for the unbridged cycloheptanes of data set 7C1, together with a few B/TB examples. For data set 7C2, Figs. 4(*b*), 4(*d*) and 4(*f*) show increased population in the B/TB area, together with indications of some splitting of both the C and TC peaks into subgroups. Thus, we would expect to locate at least three conformational subgroups in data set 7C1 and probably more than three in data set 7C2 using the cluster-analysis algorithms incorporated into the development version of program *GSTAT* (Allen, Doyle & Taylor, 1991*a*,*b*,*c*).

For cycloheptane, we have used the symmetry-modified Jarvis–Patrick algorithm (Jarvis & Patrick, 1973; Allen, Doyle & Taylor, 1991*b*) to perform cluster analysis of the seven-dimensional torsional data sets 7C1 (101 fragments) and 7C2 (310 fragments). This algorithm calculates the dissimilarities D^n_{pq} between two fragments p and q using the torsion angles τ_i ($i = 1$–7) for each fragment, such that for $N_t = 7$

$$D^n_{pq} = \left[\sum_{i=1}^{N_t} (\Delta\tau_i)^n_{pq}\right]^{1/n}, \tag{8}$$

where $(\Delta\tau_i)_{pq}$ is the minimum value obtained from either

$$(\Delta\tau_i)_{pq} = |(\tau_i)_p - (\tau_i)_q|/180N_t \tag{8a}$$

or

$$(\Delta\tau_i)_{pq} = [360 - |(\tau_i)_p - (\tau_i)_q|]/180N_t, \tag{8b}$$

so as to take account of the torsional phase change at $\pm 180°$. The quantity n in (8) is an integer power; normally $n = 1$ (city block metric) or 2 (Euclidean metric). For symmetric fragments, D^n_{pq} is the minimum value obtained from (8) by holding $(\tau_i)_p$ constant, and then allowing $(\tau_i)_q$ to adopt all permutations (and their inversions) allowed by the topological symmetry. The D^n_{pq} are used to determine the K_{NN} nearest neighbours of each of the N_f fragments in the data set, within some limiting value (D_{lim}) for D^n_{pq}. The Jarvis–Patrick clustering criteria will then assign two fragments p, q to the same cluster if (*a*) p occurs in the NN table of q and *vice versa*, and (*b*) if a further K_{JP} fragments are common to the NN tables of both p and q. The Jarvis–Patrick method is an agglomerative hierarchical clustering algorithm that operates in a single pass and is economical of core storage. Further, it is less prone to some of the common problems, *e.g.* chaining, that affect other algorithms that employ the D^n_{pq} values directly. The algorithm has performed consistently well on trial data sets (Allen, Doyle & Taylor, 1991*c*) of known conformational complexity.

In common with most clustering methods, the Jarvis–Patrick algorithm cannot be described as fully automatic: different clustering structures are obtained for different settings of the power factor n, the limiting D^n_{pq} value, D_{lim}, the NN-table length K_{NN}, and the commonality criterion K_{JP}. In practice, a value for D_{lim} of -0.10 may be regarded as a constant, whilst variation of n (as 1 or 2) is not highly significant in dictating cluster formation. However, the composition and number of clusters formed *is* highly dependent upon the criterion K_{JP} that is chosen for a given value of K_{NN}: the ratio K_{JP}/K_{NN} is critical, and values close to 0.5 were found to generate chemically sensible clustering structures in the earlier work.

For cycloheptane, we have performed Jarvis–Patrick clustering with a variety of values for K_{JP} and K_{NN} using both normal values (1, 2) of the power factor n and with $D_{lim} = 0.06$. Relevant results, collected in Table 6, show that high values of K_{JP}/K_{NN} result in a large number of clusters (N_c) of low population, whilst lowering this ratio permits the coalescence of these original clusters into larger units. For data set 7C1, the results for $n = 1$ and $n = 2$ are quite similar, but N_c values are somewhat higher for $n = 1$ at a given value of K_{JP}/K_{NN}. The results of Table 6 are further discussed below in terms of the optimum clustering structure selected for this data set.

The selection of an optimum clustering structure was guided by *a priori* chemical knowledge and by examination of the various statistical descriptors (Allen & Johnson, 1991) that summarize the distribu-

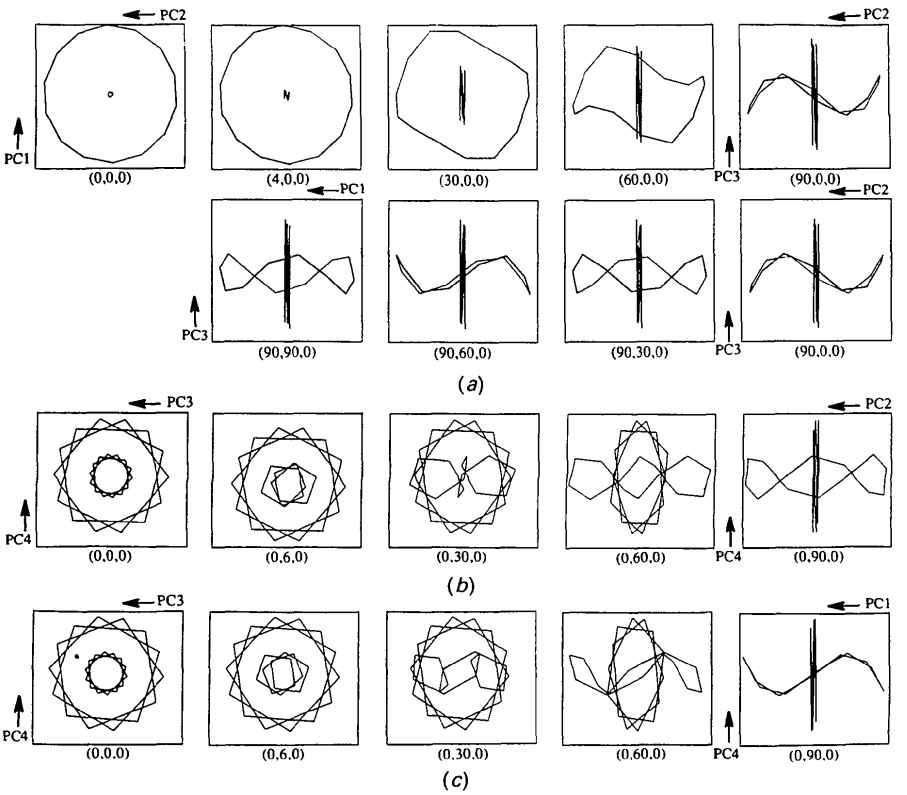


Fig. 7. Views of the chair–twist-chair and boat–twist-boat helical pseudorotation pathways in the four-dimensional principal-component (PC) space. The starting points and the rotations applied in (*a*), (*b*), (*c*) are fully described in the text.

Table 6. *Summary results for Jarvis–Patrick clustering of cycloheptane conformations in data set 7C1 using torsional descriptors*

Full definitions of the variables n, K_{NN}, K_{JP} are in the text. N_c is the number of clusters formed with population $N_p \geq 4$. ΣN_p is the total number of fragments covered by those clusters, with a percentage figure (out of 101 fragments in 7C1) in parentheses. An asterisk signifies a chemically acceptable result.

				$n = 1$		$n = 2$	
	K_{NN}	K_{JP}	K_{JP}/K_{NN}	N_c	ΣN_p (%)	N_c	ΣN_p (%)
1	11	7	0.636	6	82 (81.2)	6	82 (81.2)
2	11	6	0.546	4	89 (88.1)	4*	87 (86.1)
3	11	5	0.455	2	91 (90.1)	3	56 (55.4)
4	10	7	0.700	9	57 (56.4)	8	56 (55.4)
5	10	6	0.600	6	84 (83.2)	7	83 (82.2)
6	10	5	0.500	5	90 (89.1)	4*	88 (87.1)
7	10	4	0.400	3	95 (94.1)	3	94 (93.1)
8	9	6	0.666	8	65 (64.4)	8	64 (63.4)
9	9	5	0.555	6	83 (82.2)	6	82 (81.2)
10	9	4	0.444	3*	91 (90.1)	3	92 (91.1)
11	8	4	0.500	6	85 (84.2)	5	82 (81.2)
12	8	3	0.375	4	91 (90.1)	3	92 (91.1)
13	7	4	0.571	9	68 (67.3)	9	75 (74.3)
14	7	3	0.428	4*	83 (82.1)	8	92 (91.1)
15	7	2	0.286	4	92 (91.1)	4	93 (92.1)

tions of τ_1–τ_7 in each cluster. The results for $K_{NN} = 10$, $K_{JP} = 5$, $n = 2$ and $D_{lim} = 0.06$ were selected as optimum for data set 7C1. Mean torsion angles and statistical descriptors for the four major clusters are in Table 7(*a*).

The 101 fragments in the asymmetric unit of 7C1 are classified as 38 twist-chairs (cluster 1), 30 chairs (2), ten distorted twist-chairs (3) and ten distorted boats (4). However, the centroids of clusters 1 and 2 are sufficiently close to (respectively) a twofold axis and a mirror plane in conformational space, so that these clusters coalesce with their symmetry-related counterparts using a distance criterion ($D_{max} = 1.0$) described by Allen & Taylor (1991). Results of Table 7(*a*) relate to the coalesced clusters of 76 and 60 fragments, respectively, for which symmetrized torsion-angle means can be reported. The twist-chair and boat conformers of clusters 3 and 4 are sufficiently distorted to prevent coalescence with symmetry-related counterparts. Of the remaining 13 fragments, only one adopts a true twist-boat conformation, whilst two fragments having 1,3,5-fusion to three-membered rings adopt a conformation with zero torsion angles about the three fusion bonds. The remaining ten fragments adopt conformations that are heavily distorted variants of the C/TC forms described in Table 7(*a*).

A survey of cluster composition for the trial runs of Table 6 shows that results from runs 2 ($n = 2$) and 14 ($n = 1$) are almost identical to the selected optimal results of run 6 ($n = 2$). In run 10 ($n = 1$) twist-chair clusters 1 and 3 form a single large cluster. In other runs with increasing K_{JP}/K_{NN} ratios the increasing N_c values are as a result of a progressive splitting of the two main C/TC clusters (1 and 2) into smaller units. A feature of all of these runs is the omnipresence of

Table 7. *Jarvis–Patrick clustering of torsional data sets 7C1 and 7C2*

N_c is the cluster number, N_p is the cluster population, $\bar{\tau}_i$ (i = 1–7) are mean torsion angles (°) with their circular e.s.d.'s in parentheses, $\bar{R}_{min}$ and $\bar{R}_{max}$ are the upper and lower limits of the circular concentration values for $\bar{\tau}_i$ distributions, and D_{max} is the maximum intracluster dissimilarity (deg^2) of any fragment from the cluster centroid defined by the $\bar{\tau}_i$. Symmetry-coalesced clusters (see text) are marked (*). Full details of circular statistical descriptors are in Allen & Johnson (1991). Conformational descriptors (Conf.) are as in the text: superscript + or + + indicates increased puckering from the minimum-energy form, superscript *d* indicates an asymmetric distortion.

(*a*) Data set 7C1

N_c	1*	2*	3	4
N_p	76	60	10	10
τ_1	−84.8 (1.0)	−62.9 (1.4)	−81.7 (0.8)	−73.6 (1.0)
τ_2	71.2 (0.6)	80.9 (1.1)	57.7 (1.0)	−2.8 (1.9)
τ_3	−54.2 (1.0)	−63.8 (1.1)	−37.3 (1.2)	66.7 (1.0)
τ_4	71.2 (0.6)	63.8 (1.1)	60.8 (1.6)	−23.3 (1.7)
τ_5	−84.8 (1.0)	−80.9 (1.1)	−92.5 (1.5)	−62.8 (1.6)
τ_6	37.3 (1.5)	62.9 (1.4)	46.6 (1.0)	50.8 (1.4)
τ_7	37.3 (1.5)	0.0	36.1 (1.2)	40.3 (1.4)
$\bar{R}_{min}$	0.988	0.988	0.996	0.994
$\bar{R}_{max}$	0.998	0.995	0.999	0.999
D_{max}	31.1	30.7	13.7	16.2
Conf.	TC	C	TC^d	B^d

(*b*) Data set 7C2

N_c	1	2	3	4	5
N_p	116*	90*	39	50*	30*
τ_1	0.0	0.0	−14.9 (1.2)	0.0	−44.4 (0.8)
τ_2	−87.9 (0.3)	−86.9 (1.6)	−74.7 (1.0)	−103.9 (0.1)	−44.4 (0.8)
τ_3	53.1 (0.5)	104.6 (1.0)	56.8 (0.5)	127.5 (0.2)	124.6 (1.0)
τ_4	42.5 (0.9)	−59.6 (1.9)	38.0 (0.6)	−52.5 (0.2)	−88.1 (1.2)
τ_5	−42.5 (0.9)	59.6 (1.9)	−40.5 (0.6)	52.5 (0.2)	57.4 (1.6)
τ_6	−53.1 (0.5)	−104.6 (1.0)	−51.1 (0.6)	−127.5 (0.2)	−88.1 (1.2)
τ_7	87.9 (0.3)	86.9 (1.6)	100.1 (1.0)	103.9 (0.1)	124.6 (1.0)
$\bar{R}_{min}$	0.993	0.969	0.992	1.000	0.994
$\bar{R}_{max}$	0.999	0.993	0.998	1.000	0.999
D_{max}	35.2	53.6	24.9	3.4	25.1
Conf.	B	C$^+$	B^{+d}	C^{++}	TC$^+$

N_c	6	7	8	9	10
N_p	30*	30*	14	9	7
τ_1	−34.8 (0.7)	0.0	−33.5 (1.7)	−13.2 (0.3)	−31.9 (2.5)
τ_2	−60.6 (0.7)	−62.7 (18)	−48.3 (1.4)	−90.7 (0.4)	−59.2 (2.6)
τ_3	53.5 (0.7)	79.8 (1.2)	85.0 (2.3)	88.2 (2.2)	121.6 (1.4)
τ_4	53.5 (0.7)	−65.4 (1.0)	−62.2 (1.9)	33.6 (2.1)	−69.8 (0.8)
τ_5	−60.6 (0.7)	65.4 (1.0)	57.4 (1.3)	−48.8 (1.2)	42.6 (1.9)
τ_6	−34.8 (0.7)	−79.8 (1.2)	−81.4 (2.1)	−46.3 (1.3)	−80.6 (2.4)
τ_7	115.0 (0.5)	62.7 (1.8)	88.9 (2.0)	119.4 (0.4)	115.8 (1.9)
$\bar{R}_{min}$	0.999	0.989	0.989	0.993	0.993
D_{max}	0.999	0.998	0.996	1.000	0.999
D_{max}	14.3	25.7	27.7	19.1	20.1
Conf.	TB	C	TC	B^{++d}	TC^{++d}

the distorted twist-chair cluster 3 with $N_p = 10$ fragments. Lowering the K_{JP}/K_{NN} ratio from the optimum value of *ca* 0.5 merely leads to a coalescence of the boat and twist-chair conformations.

An additional simple graphical method of examining the conformational complexity of data set 7C1 is provided by its 'conformational spectrum' (Allen, 1990) shown in Fig. 8(*a*). Here we plot D^n_{pq} values from (8) with $n = 1$. The values are minimized as before by allowing fragments q to adopt all permutations/inversions of torsion angles. Now, however, we assign the torsion angles of the most representative chair conformation from Table 7(*a*) to the static fragment p in (8). Thus D^n_{pq} = DSIM, plotted in Fig. 8(*a*) represents the minimum distance of each fragment from a fixed chair position in conformational space. For data set 7C1, Fig. 8(*a*) shows a

'chair' peak close to DSIM = 0.0 followed by a broad 'twist-chair' peak at DSIM in the 80–150° range. Cluster 3 is, in fact, a small shoulder of ten distorted twist-chairs at the high end of this range. The 'boat' peak at DSIM > 300.0° is clearly demarcated.

The conformational spectrum for the 310 fragments of the larger bridged-ring data set 7C2, is shown in Fig. 8(*b*). The fixed reference chair conformation was defined exactly as for Fig. 8(*a*). This spectrum has very few points close to the normal chair position at DSIM = 0.0°. Major peaks occur as a doublet at 80–150°, a singlet at ~200°, a broad band at ~300–360° and a final sharp peak at ~400°. A brief inspection of the torsion-angle listing showed that the doublet arises because of normal twist-chairs and highly puckered chair conformers. The sharp 200° singlet arises primarily from highly puckered twist-chairs, whilst the singlet at 400° is due to twist-boats. The dominant broad band (300–360°) is, as expected, due to boat conformations showing varied degrees of puckering. Thus, the spectrum provides a rapid visual overview of the conformational complexity of 7C2, in terms of both the conformations present and their relative abundances, prior to any clustering runs. For 7C2, we would expect the cluster analysis to dissect the data set into at least six major subgroups.

A number of Jarvis–Patrick clusterings of 7C2 were performed using the Euclidean metric ($n = 2$ in 8), $D_{lim} = 0.06$, and with values of K_{JP} and K_{NN} chosen such that $K_{JP}/K_{NN} \simeq 0.5$. Runs at K_{JP}/K_{NN} = 5/10 and 6/12 led to co-clustering of some chair and twist-chair forms, whilst a 7/12 ratio yielded 13 clusters with $N_p \geq 4$ of which some pairs of clusters had almost identical geometries. The ten-cluster structure (all N_p values ≥ 7) obtained with a 7/13 ratio was selected as optimal for this data set, and mean torsion angles and associated statistical descriptors are given in Table 7(*b*). The ten clusters encompass 242 (78.1%) fragments. Additionally there were three clusters with $N_p = 3$, two with N_P = 2 and 55 singletons. Cluster coalescence across symmetry positions, within a $D_{max} = 1.0$ criterion (Allen & Taylor, 1991), is observed for six of the clusters. Thus, although many of the conformations of Table 7(*b*) represent highly puckered forms, the puckering retains the C_2 or C_s symmetry of the basic archetype.

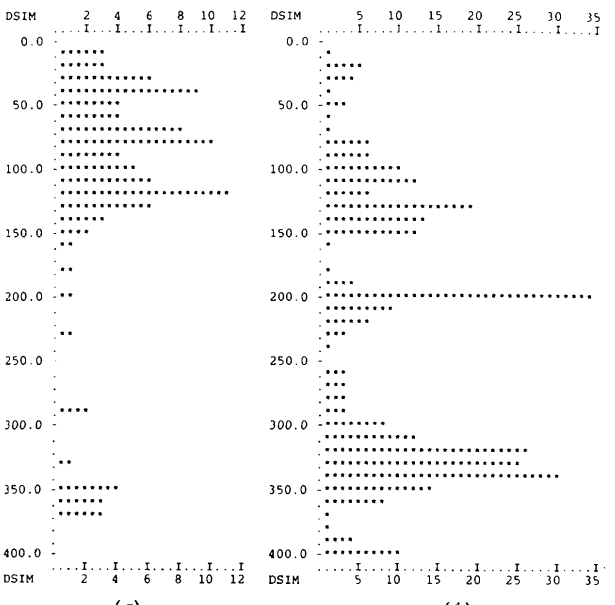


Fig. 8. Histograms of the minimized torsional dissimilarities (DSIM) of fragments in (*a*) data set 7C1 and (*b*) data set 7C2 from the torsion angles of the most representative chair conformation taken from cluster 2 of data set 7C1 (Table 7).

Typical coordinates for modelling applications

The cluster analysis package within *GSTAT* (Allen, Doyle & Taylor, 1991*c*) will locate the chemical fragment in each cluster which is closest to the cluster centroid. This is termed the 'most representative fragment' (MRF) of the cluster and the orthogonal coordinates for this fragment are of use in molecular modelling applications, since they represent a typical TC, C, B, TB or distorted conformation as observed in the accumulated crystallographic results. In Table 8, we present sets of MRF coordinates (referred to molecular axes) for the 14 clusters which are summarized in Tables 7(*a*) and 7(*b*). The conformations are illustrated graphically in Fig. 9.

Conformational variation and chemical environment

An acceptable classification of ring conformations leads inevitably to a consideration of the chemical environment of those fragments for which higher-energy conformations are observed. In this context, the results for data set 7C1 (Table 7*a*) are of more interest than those for data set 7C2 (Table 7*b*), since the latter encompasses a very wide range of environments, from simple bridging to cage formation of increasing complexity.

Non-bridged rings: data set 7C1

The chemical environments of cycloheptane fragments in this data set may be divided into three broad classes: (*a*) fused to small rings of size three or four (17 fragments), (*b*) fused to one or more five-membered rings (66 fragments), and (*c*) non-fused fragments or fragments fused to flexible rings of size greater than five atoms (18 fragments). The small rings of group (*a*) force the intra-annular cycloheptane torsion angle at the fusion bond to be close to

Table 8. *Orthogonal coordinates* (Å) *referred to molecular axes for the most representative fragment for each of the conformational clusters summarized in Table* 7

Each set of coordinates is ordered for C1–C7 in the enumeration shown in Table 1(*a*).

Data set 7C1

Cluster 1
Fragment 73 from KIMKOG

0.00000	0.00000	0.00000
1.51487	0.00000	0.00000
2.18141	1.36601	0.00000
2.37870	1.94067	1.40075
1.10784	2.41097	2.08581
−0.01962	1.38328	2.17856
−0.68148	1.06375	0.83655

Cluster 2
Fragment 28 from CUFVII10

0.00000	0.00000	0.00000
1.60477	0.00000	0.00000
2.33494	1.33460	0.00000
2.08713	2.19579	−1.26956
0.79047	2.98718	−1.25613
−0.50318	2.17122	−1.26496
−0.74229	1.32784	0.00082

Cluster 3
Fragment 82 from SAJPEY

0.00000	0.00000	0.00000
1.54766	0.00000	0.00000
2.03042	1.45320	0.00000
1.57473	2.37581	−1.17107
0.16042	2.29335	−1.73791
−0.26326	0.95919	−2.34454
−0.76263	−0.07631	−1.34485

Cluster 4
Fragment 58 from HYMINB

0.00000	0.00000	0.00000
1.55151	0.00000	0.00000
2.19854	1.40347	0.00000
1.28529	2.52860	−0.49351
0.06844	2.81084	0.42985
−0.12022	1.92323	1.63433
−0.67267	0.52646	1.27965

Data set 7C2

Cluster 1
Fragment 273 from SINXES

0.00000	0.00000	0.00000
1.56067	0.00000	0.00000
2.11660	1.42403	0.00000
1.38140	2.39355	0.96135
−0.13320	2.19235	0.83813
−0.63899	2.30382	−0.62580
−0.54260	0.84947	−1.18050

Cluster 2
Fragment 297 from VINVAP

0.00000	0.00000	0.00000
1.54735	0.00000	0.00000
1.94993	1.48414	0.00000
2.04872	1.88941	−1.49430
1.70404	0.61615	−2.29551
0.16942	0.67275	−2.48888
−0.35262	−0.37426	−1.46300

Cluster 3
Fragment 113 from DLPHFL10

0.00000	0.00000	0.00000
1.54482	0.00000	0.00000
1.93295	1.49593	0.00000
2.34675	2.01601	1.39530
1.27127	1.63060	2.43221
−0.14321	1.77647	1.89005
−0.38751	1.45872	0.38680

Cluster 4
Fragment 187 from JAPBIL

0.00000	0.00000	0.00000
1.54827	0.00000	0.00000
2.00729	1.46960	0.00000
2.98058	1.53847	1.16055
2.01083	0.93246	2.17780
1.55347	−0.36903	1.50983
0.00095	−0.36542	1.52592

Cluster 5
Fragment 181 from IPCUND

0.00000	0.00000	0.00000
1.56350	0.00000	0.00000
1.97812	1.46123	0.00000
0.97525	1.82890	−1.07990
1.10143	0.72003	−2.17490
−0.36298	0.18153	−2.33644
−0.37688	−0.88200	−1.24121

Cluster 6
Fragment 164 from GAYDIT

0.00000	0.00000	0.00000
1.54278	0.00000	0.00000
2.19822	1.37129	0.00000
1.66385	2.23518	−1.15581
1.83993	1.46168	−2.47930
0.95315	0.23359	−2.20881
−0.24975	0.78077	−1.33405

Cluster 7
Fragment 93 from CXMTUN

0.00000	0.00000	0.00000
1.54218	0.00000	0.00000
2.16806	1.40616	0.00000
2.27461	2.04582	−1.39324
0.99440	2.34990	−2.17346
−0.35930	2.04657	−1.51937
−0.62557	0.55848	−1.28848

Cluster 8
Fragment 287 from VEWKUD

0.00000	0.00000	0.00000
1.53443	0.00000	0.00000
1.88149	1.49424	0.00000
1.89275	2.13475	1.39811
0.59086	2.00357	2.25086
0.17211	0.56481	2.58384
−0.65152	−0.08030	1.40968

Cluster 9
Fragment 184 from JAJVOF

0.00000	0.00000	0.00000
1.54979	0.00000	0.00000
1.89723	1.51488	0.00000
1.91064	1.52060	1.55135
0.39870	1.86331	1.49115
0.50277	2.19587	−0.02515
−0.42762	1.24791	−0.74653

Cluster 10
Fragment 92 from CUXTOE

0.00000	0.00000	0.00000
1.51624	0.00000	0.00000
1.83048	1.55204	0.00000
2.67980	1.74778	−1.32032
1.83928	1.62201	−2.61076
1.03011	0.35142	−2.28970
−0.26262	0.67658	−1.42445

zero. Twelve of these fragments are classified as chairs (cluster 2 of Table 7*a*), whilst a further three are distorted chairs with τ(fusion) $\simeq 20°$. The last two have the 1,3,5 tri-fusion to three-membered rings referred to earlier.

Two structure types, (V), (VI), dominate subclass (*b*). In both cases, the O-containing five-ring contains points of unsaturation in the form of exocyclic keto or methylene groups, or has an endocyclic C=C double bond at (*); the other five-ring is usually, but not always, saturated. The seven-ring itself carries a variety of additional single substituents, *e.g.* Me,

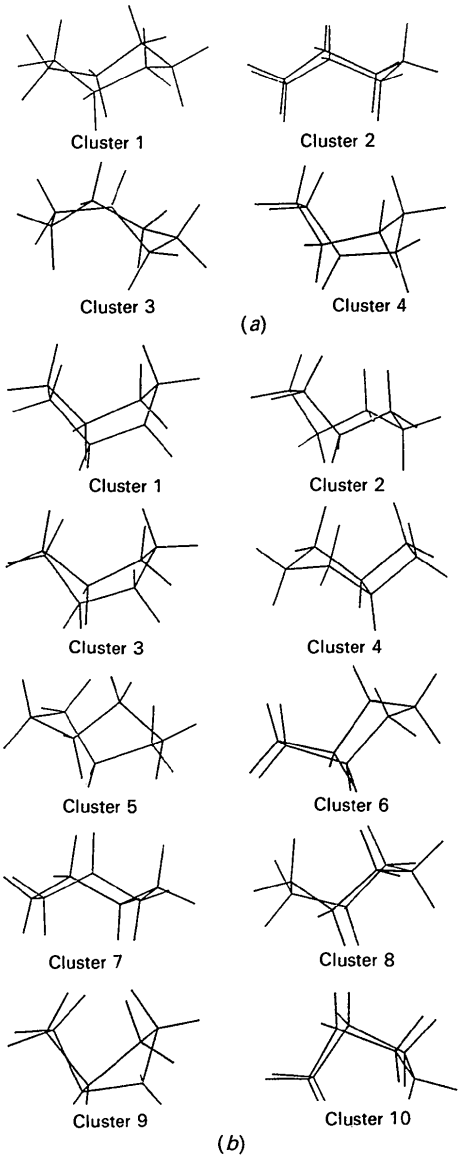
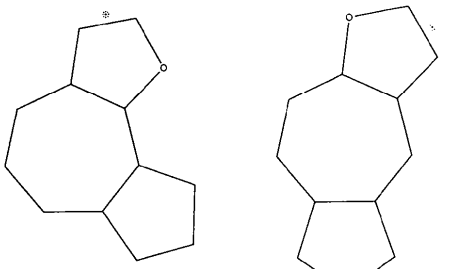


Fig. 9. Plots of the most representative fragments from conformational clusters identified for (*a*) data set 7C1 and (*b*) data set 7C2. Cluster numbers correspond to those in Table 7.

OH, halogen. The 24 fragments of type (V) are classified into 15 TC conformers and eight C conformers. It might have been expected that planarity in one (or both) of the five-rings would force a chair conformation. However this is not the case, since the seven-ring is flexible enough to adopt *cis* [τ(fusion) $\simeq 0°$] or *trans* [τ(fusion) $\simeq$ 80–90°] arrangements with respect to a planar five-ring. Analysis of fragments of type (VI) is even less predictive: of 28 fragments, 14 are in the TC class (clusters 1 and 3 of Table 7*a*), three are chairs (cluster 2) and six are boats (cluster 4), the remaining four fragments are some intermediate conformation between TC and C. Here, however, all of the boat-form seven-rings are *cis*-fused to planar five-rings.



(V) (VI)

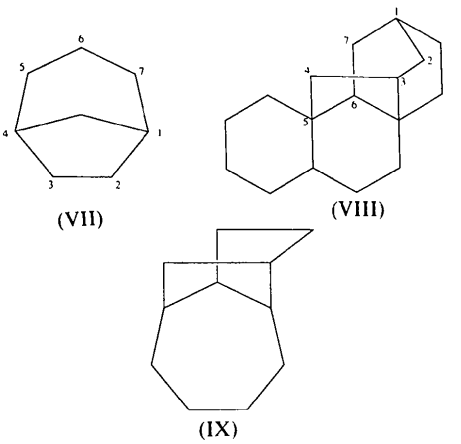
The 18 fragments of subclass (*c*), either non-fused or fused only to rings of size six or higher, divide into twelve TC conformers and four C conformers. The remaining two are sufficiently distorted so as to fall between TC and C forms.

This has been a visual analysis, an exercise in human pattern recognition and correlation. As such, it has not produced rules of thumb by which we may yet predict the conformation of a seven-ring in a given chemical environment, except for the case of small-ring fusion. However, if we ignore possible environmental effects in subsets (*b*) and (*c*), but exclude the C conformers enforced by small-ring fusion, then the ratio of TC:C:B:TB conformations arising on an *apparently* statistical basis is 48:18:10:1. These results show a 62% preference for the TC form over all other forms, and a 73% preference for the TC form in the C/TC pathway. These data compare well with the gas-phase electron-diffraction results for cycloheptane (Dillen & Geise, 1979) which were interpreted in terms of a TC/C mixture having a 76 (6)% abundance of the TC form.

Bridged rings: data set 7C2

Table 7(*b*) shows that exactly 50% (121) of the rings in clusters 1–10 adopt C/TC conformations and 50% are B/TB. The complete ratios C:TC:B:TB are 85 (35.1%):36 (14.8%):106 (43.8%):15 (6.2%). These ratios cannot, of course, be related to free-molecule structures, since they are simply a manifestation of the types of rigidified bridged rings studied so far by the crystallographic method.

The major cluster 1 contains normal boats, whilst the additional boat clusters 3 and 9 show increasing degrees of both twist and puckering. Cluster 1 is dominated by non-fused rings with simple single-atom 1,4-bridging, (VII). The deformations observed in clusters 3 and 9 may be ascribed to additional ring fusion(s) at cycloheptane bonds adjacent to the 1,4-bridge. The puckered twist-boats of cluster 6 occur in simple cages, of which structures of type (VIII) are typical. The normal chairs of cluster 7 are generally similar to (VII) but with two or more atoms in the bridge; fusion and/or cage formation then gives rise to the deformations observed in clusters 2 and 4 or, in some simple situations (*e.g.* IX) leads to the adoption of a normal twist-chair conformation. Further objective correlation of conformation with chemical environment is impossible, because of the wide variety of bridged/fused and cagelike situations that arise in data set 7C2.



(VII) (VIII) (IX)

Mean fragment dimensions

Mean bond lengths and valence angles for the 76 fragments of clusters 1–4 (data set 7C1, Table 7) are collected in Table 9. In clusters 1 and 2, where conformational clusters coalesce about symmetry elements, averages are taken over both contributing permutational isomers to yield symmetrical bond-length and valence-angle sequences. The twelve fragments of cluster 2 that have chair conformations enforced by small-ring fusion are omitted. The mean geometry of Table 9, together with the mean torsion angles of Table 7, do not necessarily correspond to a feasible (closed) ring in three dimensions (Murray-Rust, Bürgi & Dunitz, 1978) by contrast to the coordinates of Table 8.

Table 9. *Mean bond lengths* (Å) *and valence angles* (°) *for the cycloheptane rings of data set 7C1*

Mean values (e.s.d.'s in parentheses) are cited for the four major conformational clusters of Table 7. Values for the symmetry-coalesced clusters (*, see text), are obtained over both contributing permutational isomers; the twelve fragments of cluster 2 that have chair conformations forced by small-ring fusion are omitted here. N_c is the cluster number, N_p is its population, *D*1–*D*7 are lengths of bonds (Å) identified by τ_1–τ_7 in Table 1, *A*1–*A*7 are valence angles (°) having atoms 1–7 (Table 1) as vertex. *H*1–*H*7 are valence angles computed by Hendrickson (1967) for TC, C and B forms using a fixed C—C bond length of 1.533 and C—H = 1.109 Å.

N_c	1	2	3	4	Mean
N_p	76	36	10	10	132
*D*1	1.533 (2)	1.526 (3)	1.522 (5)	1.514 (4)	1.529 (2)
*D*2	1.530 (2)	1.533 (3)	1.534 (5)	1.548 (3)	1.532 (2)
*D*3	1.528 (2)	1.521 (2)	1.533 (4)	1.532 (3)	1.527 (2)
*D*4	1.530 (2)	1.521 (2)	1.529 (4)	1.545 (3)	1.529 (2)
*D*5	1.533 (2)	1.533 (3)	1.540 (3)	1.548 (5)	1.535 (2)
*D*6	1.532 (3)	1.526 (3)	1.532 (5)	1.532 (4)	1.530 (2)
*D*7	1.532 (3)	1.546 (7)	1.530 (3)	1.543 (3)	1.536 (2)
⟨*D*1-*D*7⟩	1.531 (2)	1.529 (3)	1.531 (4)	1.537 (4)	1.531 (2)
*A*1	116.1 (3)	117.6 (5)	116.9 (5)	113.4 (3)	116.4 (3)
*A*2	114.5 (3)	115.3 (4)	116.7 (8)	117.7 (4)	115.1 (4)
*A*3	115.3 (3)	115.6 (5)	121.4 (6)	114.9 (4)	115.8 (4)
*A*4	115.3 (3)	114.4 (4)	116.6 (9)	115.4 (5)	115.1 (4)
*A*5	114.5 (3)	115.6 (5)	112.2 (8)	115.0 (4)	114.5 (4)
*A*6	116.1 (3)	115.3 (4)	116.8 (11)	117.3 (6)	116.0 (5)
*A*7	116.5 (4)	117.6 (5)	113.2 (5)	112.2 (6)	116.2 (5)
⟨*A*1-*A*7⟩	115.5 (3)	115.9 (5)	116.3 (7)	115.1 (4)	115.6 (4)
*H*1	115	118	115	115	
*H*2	113	115	113	116	
*H*3	115	114	115	116	
*H*4	115	115	115	115	
*H*5	113	114	113	115	
*H*6	115	115	115	115	
*H*7	116	118	116	115	

The mean bond length across all ring bonds is 1.531 (2) Å, almost identical to the 1.530 (2) Å cited by Allen *et al.* (1987) for 5777 C_{sp^3}—C_{sp^3} bonds averaged over all bond-substitution patterns. The value is also very close to the fixed value of 1.533 Å used by Hendrickson (1967) in his energy-minimization studies of the medium rings. Indeed, the results of Table 9 show that his predicted intra-annular valence angles for the TC and C forms are in close agreement with the experimental data currently available.

Concluding remarks

A basic feature of crystallographic research is to obtain knowledge of the size and shape of three-dimensional molecules and chemical fragments from raw data in the form of external coordinates referred to unit-cell axes. If enough data exists, this knowledge may be generalized and expressed in terms of typical dimensions or archetypal conformations. Essentially, this is a form of knowledge acquisition through pattern recognition and classification. In this paper we have had four objectives: (i) to apply a variety of novel classification techniques to conformational descriptions of cycloheptane, (ii) to interpret the results so as to provide a functional description of the conformational hypersurface and to locate highly populated regions on that hypersurface, (iii) to study, modify or improve the available methodologies for data analysis and (iv) to relate the conformational minima detected by these methods to the chemical environments of the cycloheptane fragments.

We have shown that the results of symmetry-modified principal component and cluster analyses are clearly interpretable in terms of *a priori* knowledge of conformational space for cycloheptane. The work has shown that symmetry expansion of data sets is an essential pre-requisite to a PCA for a symmetrical fragment. Indeed, even without the *a priori* knowledge the analysis would have indicated the four-dimensional nature of this space, and the twofold degeneracy of the axial pairs. The analysis has, however, pointed to the importance of isolating a single asymmetric unit of conformational space (from the fully symmetrized data set) as an adjunct to classification.

The Jarvis–Patrick clustering experiments have provided valuable insights concerning typical criteria to be employed in processing the nearest-neighbour table which underpins this algorithm. These insights indicate that a greater degree of automation of the pattern recognition process may be possible through (*a*) systematic computerized scanning of results obtained with small variations of the K_{JP}/K_{NN} ratio in the area of ~0.5, and (*b*) further assessment of the statistical descriptors generated for each cluster. Nevertheless, results from the current semi-automated methods were obtained rapidly and provide a satisfactory numerical classification of conformations. This work was aided by the inspection of simple conformational spectra and this technique will be further developed in the near future.

We thank Dr Olga Kennard FRS, Director of the CCDC, for her interest in this work, Drs Scott Rowland and Stephanie Garner of the CCDC for assistance with data processing, and Mrs Dee Hughes for preparing Figs. 1 and 6.

References

ALLEN, F. H. (1990). *Acta Cryst.* A**46**, C-139.

ALLEN, F. H., DAVIES, J. E., GALLOY, J. J., JOHNSON, O., KENNARD, O., MACRAE, C. F., MITCHELL, E. M., MITCHELL, G. F., SMITH, J. M. & WATSON, D. G. (1991). *J. Chem. Inf. Comput. Sci.* **31**, 187–204.

ALLEN, F. H., DOYLE, M. J. & AUF DER HEYDE, T. P. E. (1991). *Acta Cryst.* B**47**, 412–424.

ALLEN, F. H., DOYLE, M. J. & TAYLOR, R. (1991*a*). *Acta Cryst.* B**47**, 29–40.

ALLEN, F. H., DOYLE, M. J. & TAYLOR, R. (1991*b*). *Acta Cryst.* B**47**, 41–49.

ALLEN, F. H., DOYLE, M. J. & TAYLOR, R. (1991*c*). *Acta Cryst.* B**47**, 50–61.

ALLEN, F. H. & JOHNSON, O. (1991). *Acta Cryst.* B**47**, 62–67.

ALLEN, F. H., KENNARD, O., WATSON, D. G., ORPEN, A. G., BRAMMER, L. & TAYLOR, R. (1987). *J. Chem. Soc. Perkin Trans.* 2, pp. S1–S19.
ALLEN, F. H. & TAYLOR, R. (1991). *Acta Cryst.* B**47**, 404–412.
ALTONA, C., GEISE, H. J. & ROMERS, C. (1968). *Acta Cryst.* **24**, 13–32.
ALTONA, C. & SUNDARALINGAM, M. (1972). *J. Am. Chem. Soc.* **94**, 8205–8212.
AUF DER HEYDE, T. P. E. (1990). *J. Chem. Educ.* **67**, 461–469.
AUF DER HEYDE, T. P. E. & BÜRGI, H.-B. (1989*a*). *Inorg. Chem.* **28**, 3960–3969.
AUF DER HEYDE, T. P. E. & BÜRGI, H.-B. (1989*b*). *Inorg. Chem.* **28**, 3970–3981.
AUF DER HEYDE, T. P. E. & BÜRGI, H.-B. (1989*c*). *Inorg. Chem.* **28**, 3982–3989.
BOCIAN, D. F., PICKETT, H. M., ROUNDS, T. C. & STRAUSS, H. L. (1975). *J. Am. Chem. Soc.* **97**, 687–695.
BOCIAN, D. F. & STRAUSS, H. L. (1977). *J. Am. Chem. Soc.* **99**, 2876–2882.
BOESSENKOOL, I. K. & BOEYENS, J. C. A. (1980). *J. Cryst. Mol. Struct.* **10**, 11–18.
BOEYENS, J. C. A. & EVANS, D. G. (1989). *Acta Cryst.* B**45**, 577–581.
BURKERT, R. & ALLINGER, N. L. (1982). *Molecular Mechanics*, American Chemical Society Monograph No. 148. Washington, DC: American Chemical Society.
Cambridge Structural Database (1992). Version 4.6. Cambridge Crystallographic Data Centre, 12 Union Road, Cambridge, England.
CHATFIELD, C. & COLLINS, A. J. (1980). *Introduction to Multivariate Analysis*. London: Chapman and Hall.
CREMER, D. & POPLE, J. A. (1975). *J. Am. Chem. Soc.* **97**, 1354–1358.
DILLEN, J. & GEISE, H. J. (1979). *J. Chem. Phys.* **70**, 425–429.
DOBLER, M. (1990). *MacMoMo. Molecular Display and Modelling Program for Apple Macintosh Computers*. ETH, Zurich.
DOLATA, D. P., LEACH, A. R. & PROUT, C. K. (1987). *J. Comput. Aided Mol. Des.* **1**, 73–86.
DUNITZ, J. D. & BÜRGI, H.-B. (1993). Editors. *Structure Correlation*. Weinheim: VCH Publishers.
EVERITT, B. (1980). *Cluster Analysis*, 2nd ed. New York: Wiley.
HENDRICKSON, J. B. (1967). *J. Am. Chem. Soc.* **89**, 7047–7054.
JARVIS, R. A. & PATRICK, E. A. (1973). *IEEE Trans. Comput.* **22**, 1025–1034.
MURRAY-RUST, P. & BLAND, R. (1978). *Acta Cryst.* B**34**, 2527–2533.
MURRAY-RUST, P., BÜRGI, H.-B. & DUNITZ, J. D. (1978). *Acta Cryst.* B**34**, 1787–1793.
MURRAY-RUST, P., BÜRGI, H.-B. & DUNITZ, J. D. (1979). *Acta Cryst.* A**35**, 703–713.
MURRAY-RUST, P. & RAFTERY, J. (1985). *J. Mol. Graphics*, **3**, 50–59.
NORSKOV-LAURITSEN, L. & BÜRGI, H.-B. (1985). *J. Comput. Chem.* **6**, 216–228.
PICKETT, H. M. & STRAUSS, H. L. (1970). *J. Am. Chem. Soc.* **92**, 7281–7288.
TAYLOR, R. (1986). *J. Mol. Graphics*, **4**, 123–131.
TAYLOR, R. & ALLEN, F. H. (1993). In *Structure Correlation*, edited by J. D. DUNITZ & H.-B. BÜRGI. Weinheim: VCH Publishers.
WIPPKE, W. T. & HAHN, M. A. (1988). *Tetrahedron Comput. Methodol.* **1**, 141–153.

Book Reviews

Works intended for notice in this column should be sent direct to the Book-Review Editor (R. F. Bryan, Department of Chemistry, University of Virginia, McCormick Road, Charlottesville, Virginia 22901, USA). As far as practicable, books will be reviewed in a country different from that of publication.

Acta Cryst. (1993). B**49**, 928

Books Received

The following books have been received by the Editor. Brief and generally uncritical notices are given of works of marginal crystallographic interest; occasionally, a book of fundamental interest is included under this heading because of difficulty in finding a suitable reviewer without great delay.

Supramolecular architecture. Synthetic control in thin films and solids. ACS Symposium Series 499. Edited by THOMAS BEIN. Pp. xi + 444. Washington, DC: American Chemical Society, 1992. Price $97.95. ISBN 0-8412-2460-9. The volume derives from a symposium sponsored by the Division of Inorganic Chemistry at the American Chemical Society meeting held in Atlanta, GA, in April 1991. It contains 27 contributions in three main areas: two-dimensional assemblies, thin films; layered and low-dimensional structures; and three-dimensional assemblies and amorphous networks. Many of the contributions contain the results of X-ray structural studies and most are updated to early 1992.

Fullerenes. Synthesis, properties, and chemistry of large carbon clusters. ACS Symposium Series 481. Edited by GEORGE S. HAMMOND and VALERIE J. KUCK. Pp. xiii + 195. Washington, DC: American Chemical Society, 1992. Price $44.95. ISBN 0-8412-2182-0. This volume derives from a 'Fast Breaking Events Symposium' organized at the American Chemical Society meeting held in Atlanta, GA, in April 1991. It presents some of the earliest X-ray, NMR, mass spectrometric and thermal studies of fullerenes and their metal complexes. The contributions have been reviewed and most are updated to August 1991.

Proton conductors – solids, membranes and gels – materials and devices. By PHILLIPE COLOMBAN. Pp. xxxii + 581. Cambridge University Press, 1992. Price £75.00. ISBN 0-521-38317-X. The stated aim of this book is 'to give a comprehensive survey of the chemical and physical parameters governing proton conduction. It includes descriptions of the preparation, structures and properties of typical materials (glasses, crystals, ceramics, metals, organic and inorganic polymers) and of devices'.