Journal of Computer-Aided Molecular Design, 8 (1994) 257–272
ESCOM

J-CAMD 228

Shape information from a critical point analysis of calculated electron density maps: Application to DNA–drug systems

Laurence Leherte* and Frank H. Allen

Cambridge Crystallographic Data Centre, 12 Union Road, Cambridge CB2 1EZ, U.K.

Received 10 May 1993
Accepted 28 July 1993

Key words: Molecular shape analysis; Critical point analysis; Electron density maps; DNA–drug interactions

SUMMARY

A computational method is described for mapping the volume within the DNA double helix accessible to the groove-binding antibiotic netropsin. Topological critical point analysis is used to locate maxima in electron density maps reconstructed from crystallographically determined atomic coordinates. The peaks obtained in this way are represented as ellipsoids with axes related to local curvature of the electron density function. Combining the ellipsoids produces a single electron density function which can be probed to estimate effective volumes of the interacting species. Close complementarity between host and ligand in this example shows the method to give a good representation of the electron density function at various resolutions. At the atomic level, the ellipsoid method gives results which are in close agreement with those from the conventional spherical van der Waals approach.

INTRODUCTION

Modelling of molecular recognition processes usually involves a combination of purely steric interactions with intermolecular potentials resulting from hydrophobic and electrostatic effects, and from hydrogen bonding. Defining steric properties requires the shapes and volumes of the partner molecules to be represented as surfaces. For example, the solvent-accessible portion of a molecule represented as van der Waals spheres can be determined by considering the locus of the centre of a spherical probe of appropriate size sweeping the surface [1]. Derived methods have also been proposed. Del Carpio et al. [2] have represented atoms as sets of points defining an icosahedral body embedded in each van der Waals sphere for the identification of cavities of proteins carrying a ligand molecule. In order to determine the optimum interaction between two molecules com-

*To whom correspondence should be addressed at Laboratoire de Physico-Chimie Informatique, Facultés Universitaires Notre-Dame de la Paix, Rue de Bruxelles 61, B-5000 Namur, Belgium.

posed of hard spheres, Santavy and Kypr [3] defined the maximal interface area by fitting projections of the molecular surfaces onto grids and minimizing the surface–surface distances.

The electron density associated with a system is a more direct starting point than the standard van der Waals representation of its component atoms for describing the volume accessible to other molecules. Electron density distributions, moreover, contain information about the anisotropy of the subunits within a molecule (atoms, functional groups, side chains, etc.), arising from chemical bonding and thermal motion. The aim of the present study is to demonstrate that analysis of the topology of an electron density distribution can provide information about molecular shapes in the crystalline state. Specifically, the topological method is used to investigate the complementarity between B-DNA dodecamer structures and the antibiotic netropsin. The method has, however, a wider application in the field of X-ray structure determination [4].

Systems under study

Complexes between B-DNA oligomers and distamycin A, netropsin and other drugs that bind in the minor groove [5] are particularly suitable as model systems for representing molecular shape. Crystal structures of a number of similar dodecamer B-DNA fragments and of B-DNA fragments complexed with drugs of this class (including netropsin), indicate that binding takes place without major change in the conformation of the drug or the DNA, and proceeds with displacement of water molecules from the minor groove of the double helix to enable formation of a number of hydrogen bonds and van der Waals contacts [6].

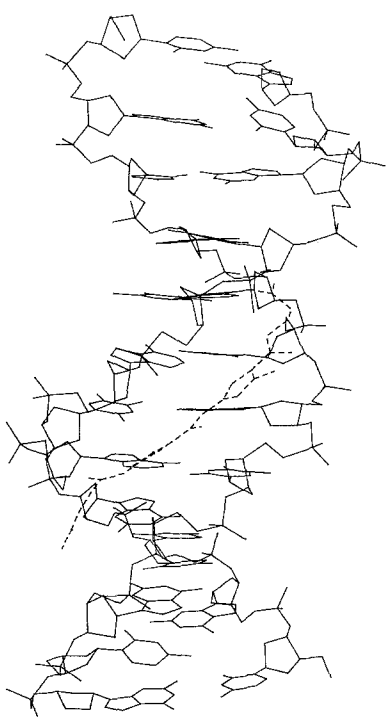

Fig. 1. DNA (solid lines)–netropsin (dashed lines) structure, obtained from the coordinates stored in the Nucleic Acid Database file GDL001.

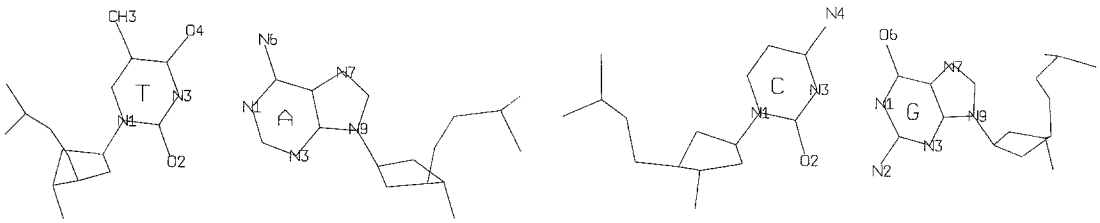

Fig. 2. Structure of the thymine-adenine and guanine-cytosine base pairs of GDL001.PDB.

Structural description of DNA

B-DNA is a double helix formed by two antiparallel strands coiled in a right-handed fashion about a common axis (Fig. 1) [7]. Each monomeric nucleotide unit consists of common deoxyribose and phosphate groups, which link to form the DNA backbone, and one of four heterocyclic side chains or bases. A characteristic property of the bases is their capacity to associate by hydrogen bonding into only two types of pair (Fig. 2), and this enables the highly specific recognition between individual strands that results in the familiar duplex structure. The double helix is composed of a central core of planar base pairs in a continuous stack, surrounded by two helical ridges, consisting of the sugar–phosphate backbones. The grooves separating these ridges are denoted major and minor to reflect the asymmetry of the glycosidic bonds joining the base and deoxyribose unit in each base pair, and allow access to the sequence-specific array of functional groups located only on the bases.

The geometry of the grooves is, therefore, relevant to molecular recognition processes involving DNA and is generally represented in terms of width and depth. Groove width is estimated as the shortest P_A–P_B distance across the groove (A and B denote the two strands), less 5.8 Å to allow for the radii of the phosphate groups themselves. The depth of the minor groove is the distance between phosphate and N2 of guanine, N3 of adenine or O2 of cytosine or thymine, less the sum of the appropriate van der Waals radii (r_O=1.4 Å, r_N=1.5 Å); that of the major groove is the distance between phosphate and O6 of guanine, N6 of adenine, O4 of thymine or N4 of cytosine, again minus the sum of the appropriate van der Waals radii.

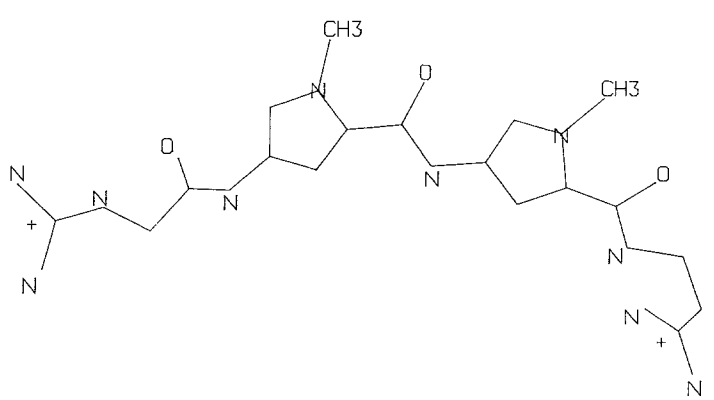

Fig. 3. Netropsin structure as bound to the dodecamer d(CGCGATATCGCG)$_2$. The coordinates are taken from file GDL001.PDB.

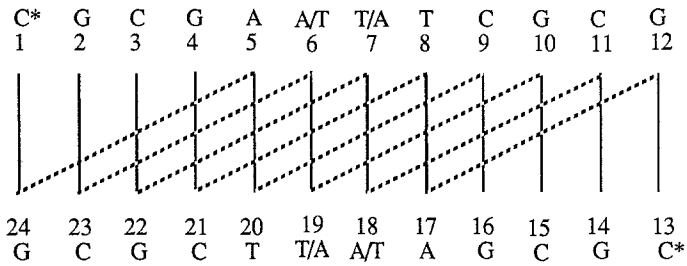

Fig. 4. Schematic representation of the two DNA strands as labelled in this work. Solid and dashed lines show base pairs and shortest P_A–P_B distances, respectively. *The phosphate groups of nucleotides 1 and 13 are not present in the structures.

Structural properties of netropsin

Netropsin, though essentially a linear molecule, is composed of pyrrole rings and planar amide, amidine and guanidine groups (Fig. 3). It adopts a crescent shape closely similar to that of the minor groove of B-DNA and has a length corresponding to four base pairs. Binding of netropsin is specific to A–T sequences. Efficient discrimination against guanine-containing regions is thought to be due to the potential steric clash with the 2-amino group (replaced by a hydrogen atom in the case of adenine) that projects into the minor groove [8]. Hydrogen bonding between the concave surface of the drug and the functional groups on the floor of the groove is, however, likely to be an important component of the interaction [9]. A positive charge on the netropsin molecule, due to protonation at neutral pH, facilitates binding to the generally negatively charged DNA, while MO calculations [10] have shown that an area of deep negative potential is located in the minor groove, which is suggested as ideal for burying a linear polycation.

COMPUTATIONAL METHODS

Topological analysis of electron density maps (EDMs)

We carry out the topological analysis of an electron density distribution ρ by looking at the so-called critical points of the function. By definition, these are the points $\mathbf{r}$ where the gradient $\nabla\rho(\mathbf{r})$ vanishes along each of the three main spatial directions. At such points, the sign of the second derivatives is a reflection of the electron density behaviour. Maxima and minima are associated with negative and positive second derivatives, respectively. Such an analysis requires the computation of the gradients and Hessians ($\mathbf{H}$) of the electron density. The eigenvalues of $\mathbf{H}$ provide information on the local curvature. More precisely, the Hessian matrix of a continuous 3D function such as the electron density, $\rho(\mathbf{r})$, has the form:

$$\mathbf{H}(\mathbf{r}) = \begin{vmatrix} \partial^2\rho/\partial x^2 & \partial^2\rho/\partial x\partial y & \partial^2\rho/\partial x\partial z \\ \partial^2\rho/\partial y\partial x & \partial^2\rho/\partial y^2 & \partial^2\rho/\partial y\partial z \\ \partial^2\rho/\partial z\partial x & \partial^2\rho/\partial z\partial y & \partial^2\rho/\partial z^2 \end{vmatrix} \quad (1)$$

This real and symmetric matrix can be diagonalised by finding a rotation of the original coordinate system that aligns the new coordinate axes with the three principal axes of the critical points:

$$\mathbf{H}'(\mathbf{r}) = \begin{vmatrix} \partial^2\rho/\partial x'^2 & 0 & 0 \\ 0 & \partial^2\rho/\partial y'^2 & 0 \\ 0 & 0 & \partial^2\rho/\partial z'^2 \end{vmatrix} \quad (2)$$

The three non-zero diagonal elements of **H'** are the eigenvalues of **H**. The rank of **H'** is the number of non-zero eigenvalues, and the signature, σ, is the algebraic sum of their signs. When the rank of **H** is 3, four cases are possible: $\sigma = -3$ corresponds to a local maximum or *peak*, i.e., the electron density function adopts maximum values along each of the three principal directions x', y' and z'; $\sigma = -1$ corresponds to a saddle point or *pass* and there are only two negative eigenvalues; $\sigma = +1$ corresponds to a saddle point or *pale*, characterized by only one negative eigenvalue and $\sigma = +3$ corresponds to a local minimum or *pit*, i.e., the electron density function adopts minimum values along each of the three principal directions. In topological terms, each atom is associated with a peak and the path traced between two atoms by the gradient vector gives rise to the definition of a chemical bond. Along such a path, the electron density decreases, reaches a minimum (i.e., the eigenvalue corresponding to the direction followed is positive), and then increases again. The path therefore crosses a pass [11]. The approach has been implemented in the computer program ORCRIT [12–14], which locates the critical points and their connectivity. The approach therefore reduces a 3D grid of real values, the so-called electron density map (EDM), into a graph, i.e., a set of points and their linkages. The major problem to be solved in this study is, therefore, to determine how local information associated with the critical points can lead to a 3D shape description. Considering only the presence of atoms or groups of atoms, our attention will focus on the peaks only.

Database structures

ORCRIT is written to analyse the EDMs of crystal structures. These are obtained either experimentally, e.g. from X-ray diffraction experiments, or can be calculated knowing the crystal structure of the systems of interest. In this work, the XTAL package [15] is applied to DNA–drug

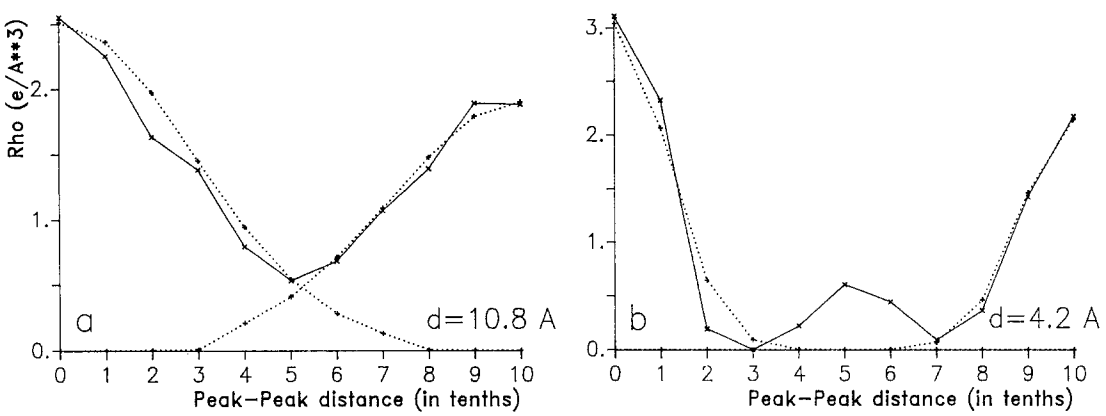

Fig. 5. Superimposition of the electron density values obtained using XTAL on a 3 Å resolution EDM, and its fitted Gaussian expression. Electron densities are represented along an interdistance vector defined by (a) two bonded peaks and (b) two nonbonded peaks. (···+···) = Gaussian fit; (—×—) = calculated.

TABLE 1
DNA AND DNA–DRUG STRUCTURES RETRIEVED FROM THE BROOKHAVEN PROTEIN DATABASE

Structure	Reference	Code
d(CGCGAATTCGCG)	18	1BNA
d(CGCGATATCGCG)–netropsin	19	GDL001
d(CGCGAATT^{Br}CGCG)–netropsin	20	GDLB05

systems whose crystal structures are available from the Brookhaven Protein Data Bank (PDB) [16] or the Nucleic Acids Database [17]. The maps are reconstructed for an asymmetric unit at various resolutions using the information stored in the PDB file, i.e., the atomic positions and the temperature factors. A medium resolution of 3 Å was selected first in order to study the global shape of the volume accessible to a drug molecule. In a second step, results for better resolved maps, at 1.5 Å resolution, were also considered in order to assess the validity of the method by comparing the results to a conventional vdW representation. The three structures used for the analyses are presented in Table 1. All structures belong to the space group $P2_12_12_1$ and have similar unit-cell parameters.

In this work, the nucleotide duplexes are labelled as in Fig. 4. Base pairing is represented using solid vertical lines while the shortest P_A–P_B distances depicting the minor groove width are shown by dashed lines. It is important to note that the shortest P_A–P_B distances are not the P_A–P_B distances of paired nucleotides.

Shape information from critical points

At the critical point locations, the three main curvatures of the electron density function are the eigenvalues of the Hessian matrix, constructed from the second derivatives. This is local information that can be transferred to the space surrounding the critical point concerned; hence it is possible to evaluate (or reconstruct) the 3D function in the close neighbourhood of each point. Each maximum of the electron density function, i.e. each peak, is considered as a centre of expansion of a Gaussian function and such a mathematical expression is fitted in order to define a volume around each peak, taking into account its three characteristic eigenvalues:

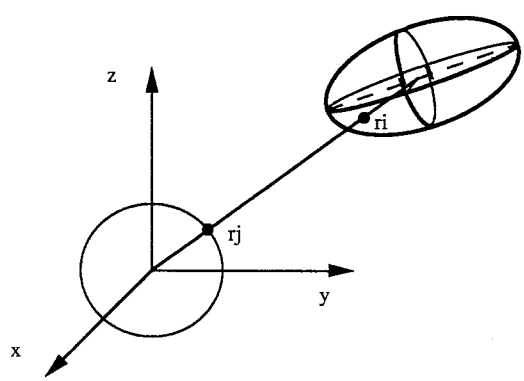

Fig. 6. Scheme of the interaction between a sphere probing the DNA accessible volume and an ellipsoid centred on a peak defined in the corresponding EDM.

$$\rho(r) = \rho(0)\exp(\alpha \mathbf{r}^T \mathbf{H} \mathbf{r} / \rho(0)) \tag{3}$$

In the frame ORCRIT, **H** is given in grid units [$e^-/(Å^3\ gu^2)$] and the parameter α is equal to 2.0. Since the three main eigenvalues for a peak are smaller than 0, the argument of the exponential function is also smaller than 0. Figures 5a and b show the agreement between the behaviour of the calculated electron density between two peaks obtained using XTAL and its fitted Gaussian approximation according to Eq. 3. The irregular shape of the 'experimental' electron density is due to the use of a crude (four-point) interpolation procedure.

In order to evaluate the volume associated with a particular peak, the exponential term of the Gaussian function has been integrated over the space within the frame of the ellipsoid:

$$\int_v \exp(\alpha \mathbf{r}^T \mathbf{H}' \mathbf{r} / \rho(0)) \tag{4}$$

which leads to the definition of an ellipsoid, characterized by three main axes r_X, r_Y and r_Z:

$$V = \frac{\pi^{3/2}\rho(0)^{3/2}}{2^{3/2} h_X^{1/2} h_Y^{1/2} h_Z^{1/2}} = \frac{4\pi}{3} r_X r_Y r_Z \tag{5}$$

and hence provides a method of representing shape anisotropy of critical points. We may extend this shape description by considering a larger structure as a set of ellipsoids and, hence, it would be useful to have another function that would reflect, or be the signature of, the whole set of individual volumes. Such a function already exists in studies of the interaction of small molecules within zeolitic systems, using a Lennard-Jones energy expression built on the vdW radii of the zeolite and adsorbate atoms [21]. In such a case, the interaction between two particles is orientation-independent. The total interaction between the host and an atom is expressed by the pair-potential approximation, where the dispersive interaction between two atoms i and j is proportional to the product of their polarizabilities ($\alpha_i.\alpha_j$), which is expressed in volume units:

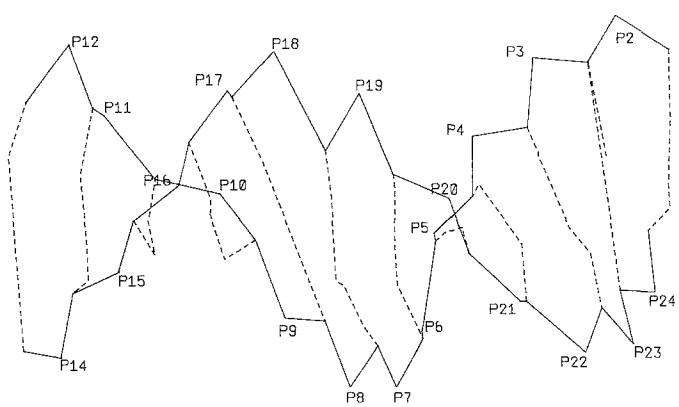

Fig. 7. Peak representation of the 1BNA PDB structure obtained from a topological analysis of the 3 Å resolution EDM using ORCRIT. Bonds between the peaks of the two DNA strands and bonds involving peaks of the base pairs are depicted by solid and dashed lines, respectively.

$$U = \sum_{i<j} U_{ij} \tag{6}$$

$$U_{ij} = -A_{ij} / r_{ij}^{6} + B_{ij} / r_{ij}^{12} \qquad \text{where } A_{ij} > 0 \text{ and } B_{ij} > 0 \tag{7}$$

$$A_{ij} \approx \alpha_i . \alpha_j \tag{8}$$

$$B_{ij} = A_{ij} (r_i + r_j)^6 \tag{9}$$

Here, r_{ij} is the separation distance between atoms i and j, and r_i and r_j are their vdW radii. In such a formula, the equilibrium distance between i and j is considered to be given by $2^{1/6}(r_i + r_j)$. Thus, overlap between two individual particles is forbidden.

By analogy, a 'signature potential' has been constructed from the lengths obtained for the ellipsoid axes and is therefore orientation-dependent. In the frame of this work, the polarisabilities α_i and α_j have been replaced by the volumes V_i and V_j, obtained by integrating the corresponding Gaussian functions (Eq. 5), and the radius r_i is the radius of the ellipsoid along the interdistance vector i–j. The radius r_j is selected by the user and is called the probe radius (Fig. 6). The idea of using a 'hypothetical' potential energy function in order to determine the optimal steric location of a guest molecule has also been developed by Kuntz et al. [22]. These authors have simplified the overlap energy between two molecules to a contribution depending upon the vdW radii of the interacting atoms and their separation distance. However, considering a Lennard-Jones-type potential allows us to emphasize the effect of global curvature of the neighbourhood, e.g., a cavity leading to more attractive energies.

RESULTS AND DISCUSSION

Critical point representation of molecular structures

The critical point representation of a DNA structure in a medium-resolution EDM shows two simplified strands and the base pairs as sets of two or three peaks. Each backbone strand is a zig-zag chain composed of peaks (P, R) which are alternately close to the phosphate atoms (within 0.3 Å) or surrounded by the ribose atoms (Fig. 7). The zig-zag chains are regular and characterised by R-P-R angles that are smaller than those for P-R-P (Table 2). In Table 2, the high angle values $a_{i\text{-}i\text{-}1\text{-}i\text{-}2}$ suggest a quasi-linear conformation for the two structures, while the deviations of the torsion values $t_{i\text{-}i\text{-}1\text{-}i\text{-}2\text{-}i\text{-}3}$ from ± 180° reflect distorted planar structures, corresponding to a helical conformation.

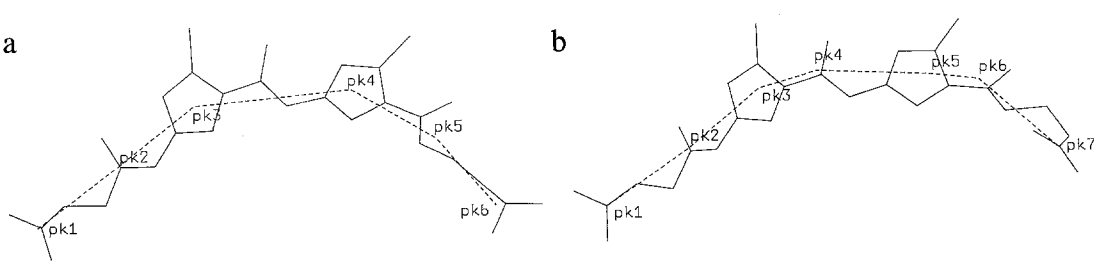

Fig. 8. Superimposition of the netropsin geometry (solid lines) and its critical point representation (dashed lines), determined by a critical point analysis of the 3 Å resolution EDMs of (a) GDLB05 and (b) GDL001, obtained using ORCRIT.

TABLE 2
GEOMETRY OF THE PEAK MOTIFS OF THE 1BNA STRUCTURE (Fig. 7) DETERMINED FROM A CRITICAL POINT ANALYSIS OF THE 3 Å RESOLUTION EDM USING ORCRIT

Chain A	$d_{i\text{-}i-1}$	$a_{i\text{-}i-1\text{-}i-2}$	$t_{i\text{-}i-1\text{-}i-2\text{-}i-3}$	Chain B	$d_{i\text{-}i-1}$	$a_{i\text{-}i-1\text{-}i-2}$	$t_{i\text{-}i-1\text{-}i-2\text{-}i-3}$
R1				R13			
P2	3.685			P14	3.223		
R2	3.791	100.047		R14	3.756	91.821	
P3	3.644	122.784	166.471	P15	3.910	120.536	171.996
R3	4.186	87.823	−149.634	R15	4.013	87.060	−144.582
P4	4.068	102.685	−177.451	P16	3.948	106.461	−177.420
R4	3.896	87.760	−163.999	R16	3.726	93.595	−176.711
P5	3.749	122.486	−174.975	P17	3.799	141.018	172.399
R5	5.314	67.597	−129.238	R17	3.659	95.461	−146.321
P6	5.991	74.160	−144.580	P18	3.979	123.829	178.472
R6	3.586	70.865	179.934	R18	5.891	71.713	−103.723
P7	3.804	129.387	155.815	P19	6.306	65.159	−150.902
R7	4.007	95.664	−144.884	R19	4.387	59.985	178.069
P8	3.733	109.460	−174.040	P20	3.969	107.080	179.475
R8	3.908	87.019	−167.842	R20	3.842	88.752	−168.432
P9	3.912	121.964	−175.130	P21	3.431	126.281	−179.972
R9	4.000	91.138	−161.857	R21	4.077	93.166	−153.518
P10	3.892	112.642	−164.432	P22	4.070	99.056	178.224
R10	4.200	86.407	−133.916	R22	4.246	85.921	−147.273
P11	3.796	111.254	−165.963	P23	3.749	111.185	−154.974
R11	4.168	89.656	154.350	R23	3.995	85.156	168.947
P12	3.676	124.761	−158.288	P24	4.110	114.609	164.946
R12	4.180	80.690	−137.357	R24	3.852	88.406	−154.218

R and P stand for ribose- and phosphate-centred peaks, respectively. Each distance value $d_{i\text{-}i-1}$ has been computed between the peak of row i (row where the value is given) and row i−1. The interpretation of the bond angle values $a_{i\text{-}i-1\text{-}i-2}$ and the torsion angles $t_{i\text{-}i-1\text{-}i-2\text{-}i-3}$ is made on a similar basis.

The netropsin molecule, when present in the complex, is similarly reduced to a small set of peaks (about six) which appear near electron-rich centres such as the amide groups and pyrrole rings (Fig. 8). The geometry of netropsin and its critical point representation, as adopted in structure GDLB05, are given in Table 3, and are compared with the geometry of the centroids of the shortest R–R segments. The comparison in Table 3 is based on the fact that, as described later, there is a tight fit between netropsin and the minor groove, causing it to be located on the minor groove central axis. Thus, from a critical point representation it is possible to derive information about the directionality of the skeleton of the electron density distribution.

At a resolution of 1.5 Å, the peaks of electron density are close to the atomic positions. As the calculation of a well-resolved EDM is also a function of the temperature factors, these must be set to zero in order to generate such maps. A comparison between the critical point representation and the vdW model at atomic resolution is presented later.

Shape representation of molecular structures

The minor groove width is described by considering the volume of each peak. Table 4 and Fig. 9a report the shortest P_A–P_B distances and the radii associated with the peaks P_A and P_B. Those distances, minus the radius value obtained from their ellipsoid representation, give values ranging

TABLE 3
GEOMETRY OF THE PEAK MOTIFS OF NETROPSIN IN GDLB05 (Fig. 8a) DETERMINED FROM A CRITICAL POINT ANALYSIS OF THE 3 Å RESOLUTION EDM USING ORCRIT

Netropsin	d_{i-i-1}	$a_{i-i-1-i-2}$	$t_{i-i-1-i-2-i-3}$	Centroids	d_{i-i-1}	$a_{i-i-1-i-2}$	$t_{i-i-1-i-2-i-3}$
				R3–R24			
Pk1				R4–R23	4.353		
Pk2	3.430			R5–R22	4.706	143.299	
Pk3	3.996	154.792		R6–R21	4.901	157.903	82.571
Pk4	5.720	142.055	−66.419	R7–R20	5.193	150.628	−13.152
Pk5	3.651	148.915	27.642	R8–R19	5.339	151.107	39.238
Pk6	3.678	174.284	96.865	R9–R18	4.711	162.350	−7.288
				R10–R17	5.095	127.076	60.740
				R11–R16	4.121	154.689	−132.019
				R12–R15	4.668	131.779	−177.192

Comparison is made with the centroid geometries calculated from the shortest distances between the ribose peaks of GDLB05.

from 10 Å at the ends of the oligomer to values of ca. 4–5 Å at the centre, leaving just enough room to fit in the netropsin molecule. An indication of the radii associated with the netropsin peaks is given in Table 5. The largest radii, R_Z, correspond to the highest eigenvalues (the least negative ones) which describe the electron density behaviour between connected peaks within the netropsin molecular plane. Therefore, the radii that have to be considered to describe the size of the netropsin molecule are R_X and R_Y. The minor groove accessible space is even more restrained between the ribose peaks (Fig. 9b), where the minimum available distance reaches 4 Å only for structures GDLB05 and GDL001 and the smallest minor groove width for structure 1BNA is only 1 Å. This means that the interaction with a drug molecule induces a local rearrangement of the DNA structure, without involving changes in space-group symmetry.

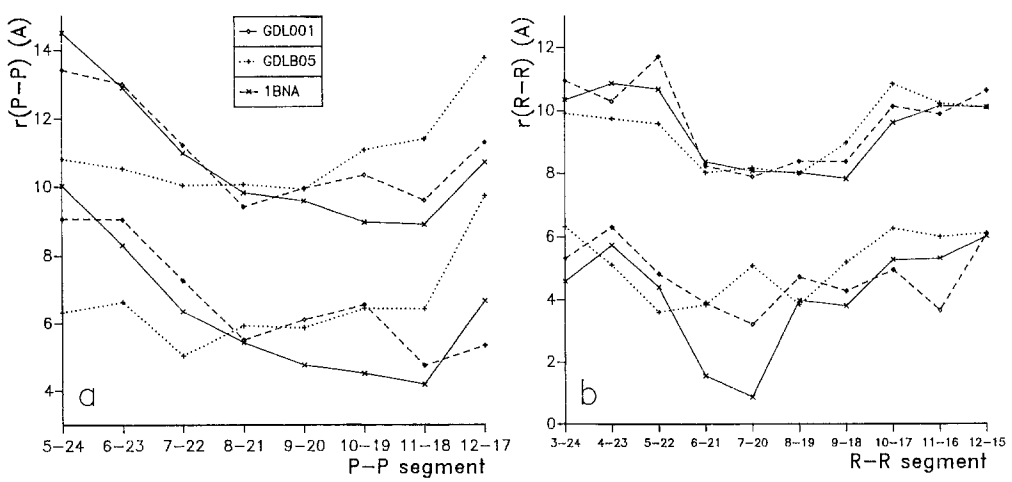

Fig. 9. Shortest (a) P_A–P_B and (b) R_A–R_B distances between peaks (upper part of the graphs) as obtained from the topological analysis of 3 Å resolution EDMs, using ORCRIT. The lower parts of the graphs give the same distances, less the radii defined by the ellipsoid representation.

TABLE 4
RADII OF THE ELLIPSOIDS ASSOCIATED WITH THE PEAKS REPRESENTING THE PHOSPHATE GROUPS (P_A AND P_B) AND P_A–P_B DISTANCES IN STRUCTURES 1BNA, GDLB05 AND GDL001

	$d(P_A–P_B)$ (Å)	P_A radius (Å)	P_B radius (Å)	$d(r_A–r_B)$ (Å)
1BNA				
$P_5–P_{24}$	14.50	2.20	2.28	10.02
$P_6–P_{23}$	12.89	2.50	2.11	8.29
$P_7–P_{22}$	10.98	2.44	2.19	6.35
$P_8–P_{21}$	9.82	2.21	2.18	5.43
$P_9–P_{20}$	9.58	2.40	2.43	4.76
$P_{10}–P_{19}$	8.97	2.23	2.22	4.51
$P_{11}–P_{18}$	8.90	2.34	2.37	4.19
$P_{12}–P_{17}$	10.72	2.02	2.03	6.67
GDLB05				
$P_5–P_{24}$	10.80	1.91	2.58	6.32
$P_6–P_{23}$	10.53	1.91	1.98	6.63
$P_7–P_{22}$	10.03	2.27	2.73	5.04
$P_8–P_{21}$	10.06	2.18	1.94	5.93
$P_9–P_{20}$	9.92	2.09	1.96	5.87
$P_{10}–P_{19}$	11.16	1.85	1.87	7.43
$P_{11}–P_{18}$	11.42	2.64	2.34	6.43
$P_{12}–P_{17}$	13.78	2.20	1.85	9.74
GDL001				
$P_5–P_{24}$	13.41	2.65	1.72	9.05
$P_6–P_{23}$	12.99	2.00	1.97	9.03
$P_7–P_{22}$	11.22	1.93	2.03	7.26
$P_8–P_{21}$	9.41	1.89	2.01	5.51
$P_9–P_{20}$	9.96	2.16	1.69	6.11
$P_{10}–P_{19}$	10.34	2.05	1.72	6.56
$P_{11}–P_{18}$	9.60	2.30	2.55	4.74
$P_{12}–P_{17}$	11.32	2.80	3.17	5.35

Values were determined from EDMs at 3 Å resolution, obtained using ORCRIT. An illustration is presented in Fig. 9a.

Additional information about the minor groove width is obtained by looking at the regions where the interaction potential is less than zero (in arbitrary units). The radius of 2 Å selected for the probing sphere was based on the netropsin radii presented in Table 5. Figure 10 shows the Lennard-Jones potential interaction energy between the probe and the DNA structure along

TABLE 5
PRINCIPAL RADII OF THE ELLIPSOIDS ASSOCIATED WITH THE PEAKS REPRESENTING THE NETROPSIN STRUCTURE IN GDLB05 (Fig. 8a), AT 3 Å RESOLUTION, OBTAINED USING ORCRIT

Peak	R_X (Å)	R_Y (Å)	R_Z (Å)
1	1.61	2.11	2.30
2	1.49	2.19	2.59
3	1.41	2.18	2.76
4	1.42	2.13	3.49
5	1.45	1.92	3.13
6	1.70	2.25	3.18

each of the shortest P_A–P_B segments. It is observed that the 1BNA minor groove size (Fig. 10a), defined in terms of U = 0, shows deviations with respect to GDLB05 and GDL001 (Figs. 10b and c). One of the terminal segments, P_{12}–P_{17}, presents a region of large negative values which is interpreted as an opening for a guest molecule. This is not the case for the two other structures, whose minor groove extremities are less accessible while their widths show less variation. There is still a slight potential well at segment P_5–P_{24} in structure GDL001 (Fig. 10c), but the minor groove is completely closed in structure GDLB05 (Fig. 10b).

For each of the curves presented in Fig. 10, the minimum values of the potential energy along the minor groove for P_A–P_B segments are displayed in Fig. 11a. The corresponding values obtained for the R_A–R_B segments are presented in Fig. 11b. These are used here to obtain information concerning the minor groove length. Values are computed without taking into account the water and drug peaks, in order to clear the minor groove opening. Along the P_A–P_B (Fig. 11a) and R_A–R_B (Fig. 11b) segments, it is observed that minima are located near the A-T region. This minimum energy area is delimited by highly repulsive potential walls on either side of the groove. This suggests that a possible guest in the minor groove would have to fit its length and width

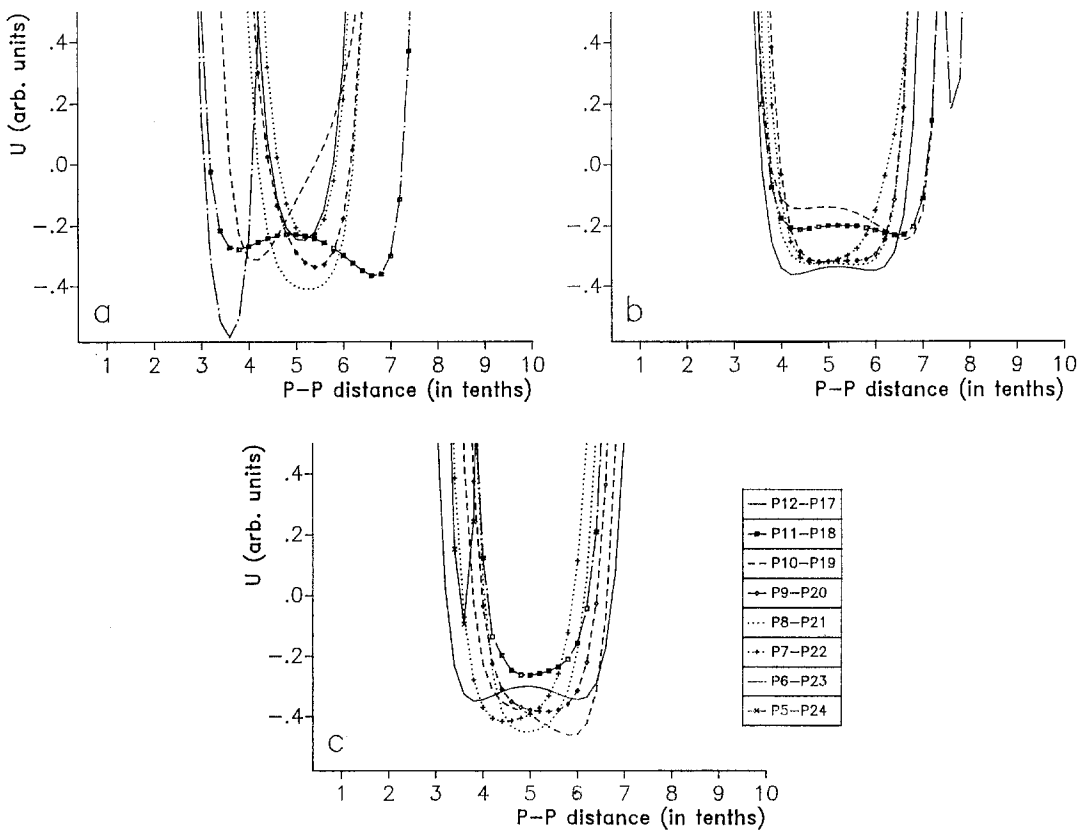

Fig. 10. Lennard-Jones potential energy (in arbitrary units) along the shortest P_A–P_B segments of (a) 1BNA; (b) GDLB05; and (c) GDL001, determined by a topological analysis of 3 Å resolution EDMs, using ORCRIT. The probing sphere radius is 2 Å.

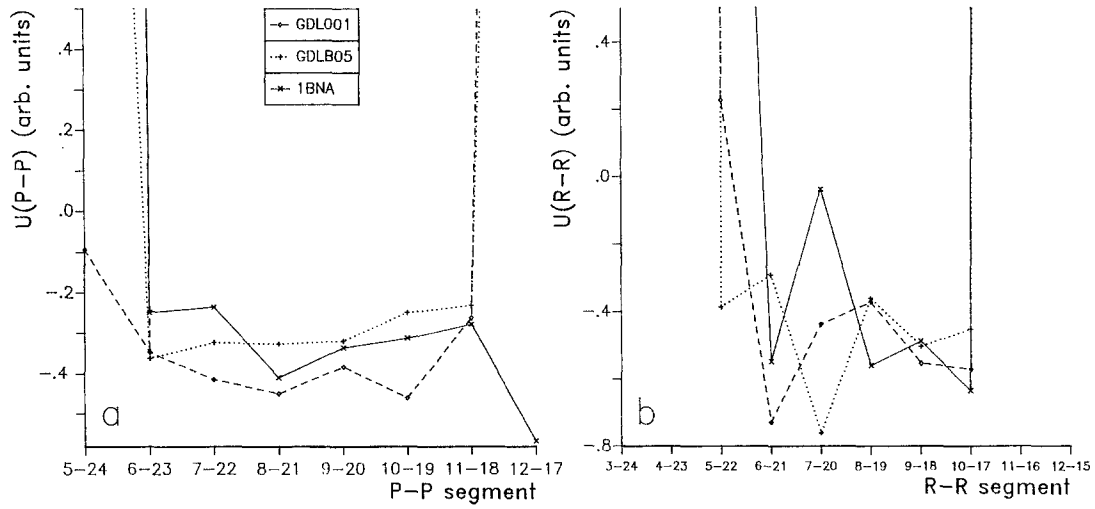

Fig. 11. Minimum Lennard-Jones potential energy (arbitrary units) between a 2 Å radius probing sphere and DNA structures along the minor groove at the level of (a) the P_A–P_B and (b) the R_A–R_B peaks determined by a topological analysis of 3 Å resolution EDMs, using ORCRIT.

exactly, or would need to undergo some conformational change. An estimation of the length of the minor groove is obtained by considering the mean $P_{A(B)}$–$P_{(A)B}$ and $R_{A(B)}$–$R_{(A)B}$ distances which are 6.7 and 5.5 Å, respectively. This corresponds to a length of 30 Å at the level of the phosphate groups and 27.5 Å at the level of the ribose rings.

A more precise depiction of the minor groove depth is obtained by plotting 2D potential energy maps, defined by planes containing the P_9–P_{20} vector and the central point between the R_8–R_{19} peaks (Figs. 12–14). This section has been selected since it crosses the minor groove near its centre and thus represents its narrowest part. In such maps, the potential energy values take into account the non-zero radius of the probing sphere. Therefore, when looking at the maps,

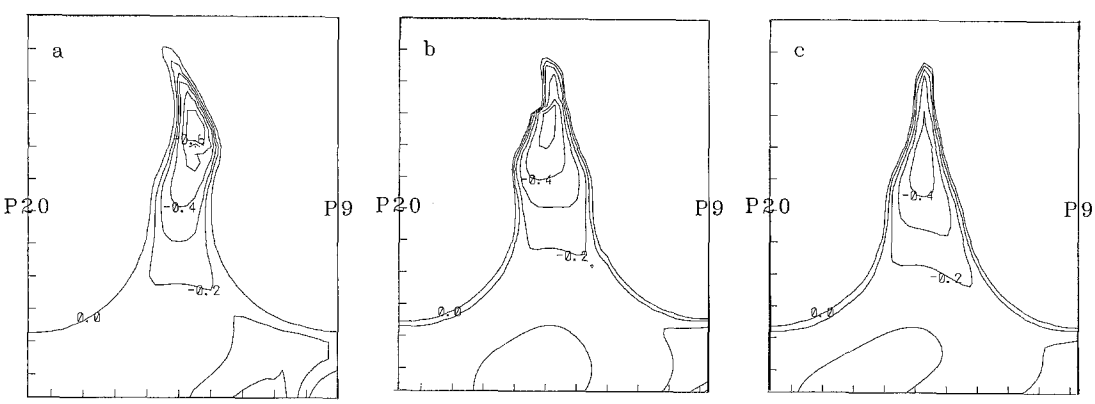

Fig. 12. 2D potential energy contour maps (step = 0.25 Å, map interval = 1 Å), defined by the P_9–P_{20} segment and the central point of the R_8–R_{19} segment of structures (a) 1BNA; (b) GDLB05; and (c) GDL001, using a Lennard-Jones potential. The interaction energy is computed between a 2 Å radius probing sphere and the DNA critical point representation obtained from a 3 Å resolution EDM, using ORCRIT. Positive energy contours are not represented.

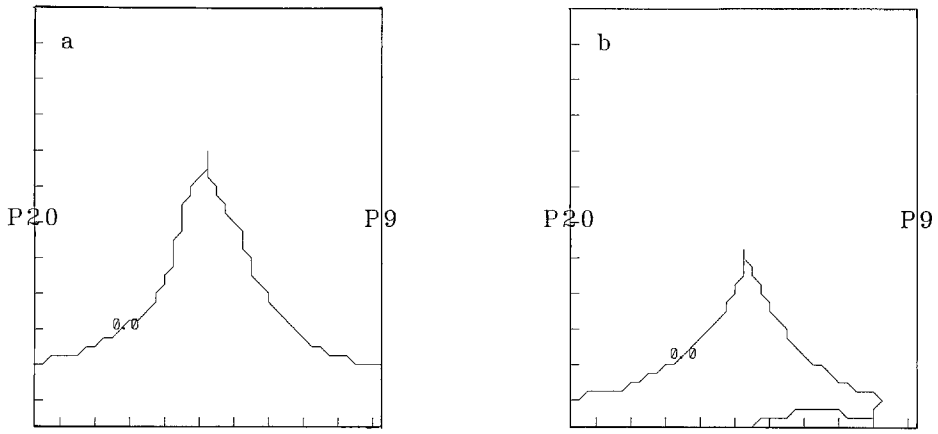

Fig. 13. 2D potential energy contour maps (step = 0.25 Å, map interval = 1 Å), defined by the P_9–P_{20} segment and the central point of the R_8–R_{19} segment of the structure GDL001, using a hard-core potential. The interaction energy is computed between (a) a 2 Å and (b) a 3 Å radius probing sphere and the DNA critical point representation obtained from a 3 Å resolution EDM, using ORCRIT. Positive energy contours are not represented.

the probing particle has to be considered as a point. Figure 12 shows that, near the floor of the minor groove, minima are present due to the highly confined space available to a host molecule. Figure 12a has been computed from the Lennard-Jones interaction between the 2 Å radius probe and the peaks associated with the 1BNA structure, represented at 3 Å resolution. Figures 12b and c correspond to the same cross-section defined for structures GDLB05 and GDL001, respectively. The geometry of the minor groove is presented in Table 6. The effective width, i.e., the region of negative potential energies, is measured at the P_9–P_{20} level, while the depth is determined from the top of the phosphate peaks delimited by U = 0 to the bottom of the minor groove. The structures to which the netropsin molecule is bound are again characterised by a slightly larger minor groove width than for the 1BNA structure (Figs. 12b and c). The same kind of map has been computed with a hard-core potential (Fig. 13) for structure GDL001, which defines a smaller minor groove by disallowing the influence of the neighbouring peaks, i.e., by neglecting the curvature effects. The access to the region at the level of the R_8–R_{19} centre is thus forbidden (compare Figs. 13a and 12c). The same map has also been computed with a probe radius of 3 Å

TABLE 6
MINOR GROOVE GEOMETRY OBTAINED FROM VARIOUS INTERACTION POTENTIALS BETWEEN A PROBING SPHERE AND THE DNA STRUCTURE DESCRIBED IN TERMS OF PEAKS, AT VARIOUS RESOLUTIONS, OBTAINED USING ORCRIT

DNA structure	Probe radius (Å)	Potential type	Resolution (Å)	Depth (Å)	'Effective' width (Å)	Figure
1BNA	2	LJ-peaks	3	8.8	1.8	12a
GDLB05	2	LJ-peaks	3	8.1	2.2	12b
GDL001	2	LJ-peaks	3	8.2	2.2	12c
GDL001	2	HC-peaks	3	5.5	1.2	13a
GDL001	3	HC-peaks	3	3.9	0.0	13b
GDL001	2	HC-peaks	1.5	8.3	1.1	14a
GDL001	2	HC-vdW	3	8.3	1.1	14b

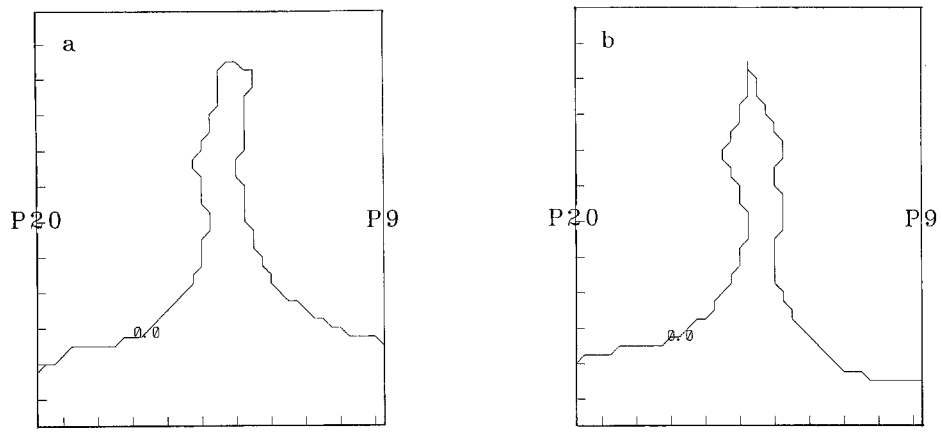

Fig. 14. 2D potential energy contour maps (step = 0.25 Å, map interval = 1 Å), defined by the P_9–P_{20} segment and the central point of the R_8–R_{19} segment of the structure GDL001, using a hard-core potential. The interaction energy is computed between a 2 Å radius probing sphere and (a) the DNA critical point representation obtained from a 1.5 Å resolution EDM, using ORCRIT, and (b) the DNA van der Waals spheres representation. Positive energy contours are not represented.

(Fig. 13b). It shows that the minor groove is not accessible at all to a guest molecule of this thickness. In order to determine the validity of the method, a comparison is shown between a map computed from a model at high resolution, i.e., an EDM reconstructed at a resolution of 1.5 Å (Fig. 14a), and the conventional vdW model (Fig. 14b). The two maps are similar, except at the level of the phosphate peaks P. This is due to the anisotropy of the critical point ellipsoids; the vdW representation considers spheres only.

CONCLUSIONS

Considering steric effects only, it has been shown in this work that:

(1) The global shape of a molecule can be described using local information associated with its centres of high electron density, i.e. its peaks, and expanded in terms of ellipsoids. The set of ellipsoids obtained is merged into one mathematical expression, depicting the close-contact interactions between the whole DNA structure and a probing sphere;

(2) The closest distances between netropsin and DNA, ≈ 4–5 Å, are those involving the high-density areas of netropsin (amide and pyrrole groups) and the ribose peaks. The shortest netropsin–base distances are slightly longer (5–6 Å). This observation would correspond to the presence of ribose–pyrrole interactions as suggested by Dickerson et al. [9];

(3) A change in the size of the minor groove between complexed and uncomplexed DNA can be detected;

(4) The minor groove length is limited by repulsive energy walls, appearing due to a loss of concavity when reaching the ends of the DNA structure;

(5) The shape analysis of the DNA structure therefore suggests that a guest molecule should have the following properties: (i) a length smaller than 27 Å, or highly flexible segments; (ii) a width not exceeding 4 Å, which corresponds to a single atom layer; and (iii) a global conformation similar to that of the groove walls.

Further work is necessary to generalise the method. This will include the study of systems for which no a priori knowledge about the binding site locations is available. From these, the development of a method to automatically match shapes described in terms of critical points may then be considered.

ACKNOWLEDGEMENTS

We thank Dr. Olga Kennard, FRS, and Dr. S.A. Salisbury for valuable discussions, and IBM Belgium and the Facultés Universitaires Notre-Dame de la Paix for the use of the Namur Scientific Computing Facility. Financial support to L.L. from the Belgian National Foundation for Scientific Research (FNRS) and the NATO Scientific Division is gratefully acknowledged.

REFERENCES

1 Connolly, M.L., J. Appl. Crystallogr., 16 (1983) 548.
2 Del Carpio, C.A., Takahashi, Y. and Sasaki, S.-I., J. Mol. Graphics, 11 (1993) 23.
3 Santavy, M. and Kypr, J., J. Mol. Graphics, 2 (1984) 47.
4 Fortier, S., Castleden, I., Glasgow, J., Conklin, D., Walmsley, C., Leherte, L. and Allen, F.H., Acta Crystallogr., D49 (1993) 168.
5 Kennard, O. and Hunter, W.N., Angew. Chem., Int. Ed. Engl., 30 (1991) 1254.
6 Kopka, M.L. and Larsen, T.A., In Propst, C.L. and Perun, T.J. (Eds.) Nucleic Acid Targeted Drug Design, Marcel Dekker, New York, NY, 1992, pp. 302–374.
7 Saenger, W., Principles of Nucleic Acid Structure, Springer, New York, NY, 1984.
8 Wang, A.H.-J. and Teng, M.-K., In Bugg, C.E. and Ealick, S.E. (Eds.) Crystallographic and Modelling Methods in Molecular Design, Springer, New York, NY, 1990, pp. 123–150.
9 Dickerson, R.E., Kopka, M.L. and Pjura, P.E., In Guschlbauer, W. and Saenger, W. (Eds.) DNA–Ligand Interactions: From Drugs to Proteins, Plenum Press, New York, NY, 1987, pp. 45–62.
10 Lavery, R. and Pullman, B., Int. J. Quantum Chem., 20 (1981) 259.
11 Bader, R.F.W., Atoms in Molecules, Clarendon Press, Oxford, 1990.
12 Johnson, C.K., Proceedings of the American Crystallographic Association Meeting 1976, Evanston, IL, Abstr. B1.
13 Johnson, C.K., Proceedings of the American Crystallographic Association Meeting 1977, Asilomar, CA, Abstr. JQ6.
14 Johnson, C.K., ORCRIT, The Oak Ridge Critical Point Network Program, Chemistry Division, Oak Ridge National Laboratory, Oak Ridge, TN, 1977.
15 Hall, S.R. and Stewart, J.M. (Eds.) XTAL 3.0 Users Manual, Universities of Western Australia and Maryland, 1990.
16 Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer, E.F., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T. and Tasumi, M., J. Mol. Biol., 112 (1978) 535.
17 Berman, H.M., Olson, W.K., Beveridge, D.L., Westbrook, J., Glebin, A., Demeny, T., Hsieh, S.-H., Srinivasan, A.R. and Schneider, B., Biophys. J., 63 (1992) 751.
18 Drew, H.R., Wing, R.M., Takano, T., Broka, C., Tanaka, S., Itakura, K. and Dickerson, R.E., Proc. Natl. Acad. Sci. USA, 78 (1981) 2179.
19 Coll, M., Aymami, J., Van der Marel, G.A., Van Boom, J.H., Rich, A. and Wang, A.H.-J., Biochemistry, 28 (1989) 310.
20 Kopka, M.L., Yoon, C., Goodsell, D., Pjura, P. and Dickerson, R.E., J. Mol. Biol., 183 (1985) 553.
21 Derouane, E.G., Leherte, L., Vercauteren, D.P., Lucas, A.A. and André, J.-M., J. Catalysis, 119 (1989) 266.
22 Kuntz, I.D., Blaney, J.M., Oatley, S.J., Langridge, R. and Ferrin, T.E., J. Mol. Biol., 161 (1982) 269.