Journal of Computer-Aided Molecular Design, 8 (1994) 731–750
ESCOM

J-CAMD 267

Finding potential DNA-binding compounds by using molecular shape

Peter D.J. Grootenhuis*, Diana C. Roe, Peter A. Kollman and Irwin D. Kuntz**

Department of Pharmaceutical Chemistry, University of California, San Francisco, CA 94143-0446, U.S.A.

Received 21 February 1994
Accepted 23 May 1994

Key words: DNA docking; Database search; Selective DNA binding

SUMMARY

For the first time a general shape-search docking algorithm (DOCK) has been applied to the minor and major grooves of A-, B- and Z-type DNA dodecamers and to an intercalation site in a B-DNA-type hexamer. Both experimentally and theoretically derived geometries for the various DNA fragments were used. The DOCK searches were carried out on a subset of the Cambridge Crystallographic Database, consisting of almost 10 000 molecules. One of the molecules that scored best in terms of the DOCK algorithm was CC-1065, a potent antitumor agent known to (covalently) bind the AT-rich parts of the minor groove of B-DNA. Several known DNA-binding agents also scored highly. Molecules with shapes complementary to A-, B- and Z-type DNA were indicated by DOCK. In addition, compounds were extracted from the database that might be selective for the GC-rich regions of the minor groove of B-DNA. Many of the compounds in the present study may serve as a starting point for further molecular design of novel DNA-binding ligands.

INTRODUCTION

The double helix structure of DNA results in the occurrence of three types of potential receptor sites for ligands [1], i.e., the major and minor grooves [2], and intercalation sites [3]. Since the characteristics with respect to shape, electrostatics, hydrophobicity, etc. of these receptor sites differ substantially, ligands are known that interact selectively with one of the sites. Examples include: netropsin, a typical minor groove binder; repressor proteins that bind to the major groove; and intercalators, such as acridine, that bind between base pairs. Some ligands,

*Present address: Organon International, P.O. Box 20, NL-5340 BH Oss, The Netherlands.
**To whom correspondence should be addressed.
Supplementary Material available from the corresponding authors: Input (PREP) files for netropsin, CC-1065 and compound **10**.

however, combine the substructures needed for interaction with the different receptor sites of DNA. Daunomycin and adriamycin, for example, consist of a planar chromophore that intercalates between the base pairs and a sugar moiety which positions itself in the minor groove [4]. Nogalamycin [4b] intercalates and positions other functional groups in both major and minor grooves.

Although the electrostatic and hydrophobic profile of the various receptor regions of DNA may differ a great deal, it is clear that among other factors like the role of solvation/desolvation, conformational changes, etc., the molecular shape is an essential factor in understanding the varying selectivity of ligands, as exemplified in the work of Pyle et al. [5]. In order to further test this hypothesis and to generate ideas for novel potential DNA-binding ligands, we decided to apply the shape-scoring version of the DOCK software [6] to DNA [7]. The DOCK software is based on an algorithm that is capable of searching databases of three-dimensional structures of compounds and evaluating the shape complementarity between these compounds and a receptor site. In a standard application of DOCK, the 3D structure of the receptor site has to be known. DOCK has been applied to the active sites of a number of proteins, e.g., papain, carbonic anhydrase [6b] and HIV protease [6c], and some interesting novel classes of potential inhibitors were found.

Of course there are a number of serious limitations to the shape-scoring version of the DOCK software used in the present paper: (i) no electrostatic or hydrophobic properties are taken into account; (ii) the occurrence of hydrogen bond-donating or -accepting groups is neglected in the searches; and (iii) unless more conformations of a compound are included in the database, no conformational flexibility can be studied. However, DOCK works in a completely unbiased way and the molecular frameworks that are selected may serve as a starting point for further design and molecular modelling. In addition, we note that work is in progress on DOCK, aimed at taking away the limitations mentioned above [8].

In the present paper we study groove binding and intercalation of ligands interacting with DNA. We chose the d(CGCGAATTCGCG)$_2$ dodecamer since the crystal structure of its complex with netropsin [9] has been solved and a number of spectroscopic and thermodynamic data are available [10]. For comparison we also generated standard geometries [11] for the A, B and Z forms. The effect of the base-pair sequence was evaluated by studying the B-type d(CGCGCGCGCGCG)$_2$ duplex. In our studies on intercalation the d(CGTACG)$_2$ hexamer in a model-built (B-type) conformation, based on the X-ray structure of its complex with daunomycin, was used [12]. These sequences possess practical sizes, i.e., they are large enough for most DNA binders to interact but not too large for doing molecular mechanics follow-up calculations. Our goals were: (i) to select novel potential minor groove binding and intercalating ligands from a database; (ii) to study the effects of conformational and base-pair changes in the DNA sequences on the selection of ligands; and (iii) to evaluate the role of shape complementarity, among other factors, in the orientation of ligands in the minor groove.

METHODS

Structural models

For the minor groove binding study, the X-ray structure of the d(CGCGAATTCGCG)$_2$• netropsin complex (Protein Data Bank code 6BNA) was used. Standard geometries for right-

handed A- and B-DNA with the same base sequence were generated using the NUCGEN module from the AMBER package [12]. These structures were subjected to 100 steps of conjugate gradient minimization (no counterions were added), using a united atom force field. For comparison, right-handed B- and left-handed Z-type DNA helical structures of $d(CGCGCGCGCGCG)_2$ were generated using the NUCSEQZ module from the AMBER package [12]. For the $d(CGTACG)_2$ sequence, coordinates were obtained by model building its complex with daunomycin in an intercalated position analogous to the X-ray structure, followed by energy refinement [4a].

Selection and docking of ligands

The DOCK* software package (version 1.01) was used for searching the database and docking the ligands into the grooves and intercalation sites of the DNA models studied. The database was a subset of the Cambridge Structural Database and consisted of 9777 molecules; X-ray structures of netropsin and daunomycin as observed in the complexes were included. This subset was obtained by clustering the database by a set of geometric (e.g. principal axes) and topologic (e.g. connectivity indices) parameters [13]. Unless indicated otherwise, the default parameters for DOCK were used. Solvent-accessible surfaces for the various DNA structures were generated by Connolly's MS program [14] as implemented in the MIDAS graphical software package [15]. The major groove of the A-DNA, the minor grooves of B- and Z-DNA, and the X-ray structure were characterized by clusters of 58, 91, 65 and 78 partly overlapping spheres. In our preliminary paper [7] we have shown that the molecular shape of the minor groove is characterized well by the overlapping sphere model by comparing DOCK-generated orientations of netropsin with the experimentally observed orientation. The intercalation site and part of the minor groove of the hexamer were described by 77 spheres. The minor groove of A-DNA and the major grooves of B- and Z-DNA could not be characterized well, because the grooves were either too shallow or too wide. All DOCK runs were performed with the variable NODLIM set to 4, which allows the most intensive sampling of orientational space in DOCK 1.01. DOCK calculates a score for each orientation as a decreasing exponential function of the atom–atom distance. The function is summed over all intermolecular contacts with atom separations less than 5 Å. The DOCK scores have been shown to correlate with intermolecular van der Waals interactions [16].

Molecular mechanics

The AMBER package [17], version 3.0 [18], was used for the molecular mechanics calculations. A residue-based 12 Å cutoff was applied to the distance-dependent dielectric $\varepsilon = r$ and van der Waals interactions. The 1–4 electrostatic and nonbonded interactions were scaled by a factor of two. For the molecular mechanics studies on the minor groove binder–DNA complexes, the all atom force field was used and counterions and, where appropriate, hydrogen atoms were included in calculated positions. The minimizations were continued until the rms of the energy gradient was less than 0.05 kcal/mol Å. Partial atomic charges for netropsin, CC-1065 and compound **10** were generated and have been deposited as Supplementary Material. In order to relax van der Waals contacts present in the standard DNA geometries, 100 steps of energy minimization were performed using the united atom force field. No counterions were added in this case.

*DOCK is available under license from the Regents, University of California, through the corresponding authors.

RESULTS

Results on known binding agents

In order to test the DOCK methodology as a tool for finding possible DNA-binding agents, we tested its ability to detect known DNA-binding agents. When we started this project, we were sure of only two compounds known to bind to DNA – Acridine Orange, and nogalamycin – in the subset of the CSD in this study. Therefore a set of 16 compounds – nine DNA intercalators and seven minor groove-binding agents – was collected from available resources, which included the Brookhaven PDB database, the full Cambridge Structural Database, and the Fine Chemicals Database (FCD) [23]. Compounds found only in the FCD had coordinates generated using CONCORD [24]. Two DOCK runs were performed: first the minor groove binders were docked to the minor groove of $d(CGCGAATTCGCG)_2$ in its crystallographic conformation bound with netropsin; second the intercalators were docked to the intercalation site of $d(CGTACG)_2$ in its crystallographic conformation bound with daunomycin. The DOCK scores of these compounds were compared to scores from DOCK search runs over the 10 000 compound subset of the CSD to the same two sites, to determine the relative rankings of the known DNA binders. The results are summarized in Table 1.

TABLE 1
DOCK SCORES AGAINST KNOWN DNA-BINDING COMPOUNDS[a]

Compound	Source[b]	Reference code	Atoms	Score[c]	Rank[d]
Intercalators					
Actinomycin D	CSD	BEJXET	92	–	–
Triostin A	CSD	CULVUA	76	–	–
Nogalamycin	CSD	BIDROV10	56	432	11
Adriamycin	PDB	1D12	39	381	73
Daunomycin	PDB	1D11	38	482	2
Ethidium bromide	CSD	ETHIDB	24	340	22
Acridine Orange	CSD	ACROZN	20	280	–
Proflavin	CSD	ACCYGB10	16	295	–
Minor groove binders					
Chromomycin	FCD	FCD 34129	82	–	–
Distamycin	PDB	2DND	35	319	8
Hoechst_33342	FCD	FCD11655	34	310	7
Hoechst_33258	FCD	FCD11656	32	298	12
Netropsin	PDB	1DNE	31	300	11
Berenil	PDB	1D63	21	217	–
DAPI	FCD	FCD11658	21	203	–

[a] DOCK scores are given for known DNA-binding compounds fitting the minor groove of $d(CGCGAATTCGCG)_2$ in the conformation observed in the crystalline complex with netropsin, and fitting the intercalation site of $d(CGTACG)_2$ in the conformation observed in its complex with daunomycin.

[b] CSD = Cambridge Structural Database, PDB = Brookhaven Protein Data Bank, FCD = Fine Chemicals Directory.

[c] No score means that no orientation was found with a positive DOCK score (i.e., all orientations tested physically intersected the DNA).

[d] Ranks are relative to scores from a search run of 10 000 compounds from a subset of the CSD. No rank means the compound was not within the top 2%.

For a purely shaped-based scoring system, the results are reasonably encouraging. DOCK ranked half the known binders in the top 1% of all the compounds. Several were in the top 20%. This is especially interesting, since many of these compounds were selected from the CSD and FCD, with no bias towards the binding conformation of the molecule.

In addition to examining the rankings, the orientation predicted by DOCK was compared to the crystallographic orientation for all the compounds for which a PDB structure was available, as shown in Fig. 2. For the most part, DOCK placed the compounds close to the correct crystallographic orientation, even in cases where we actually used the CSD or FCD compounds rather than the PDB ligand coordinates to dock the molecule to the site. The exception was nogalamycin, which was in the correct location, but rotated approximately 180°.

All the known binders that DOCK failed to rank highly were outside the range of 25–65 atoms. The larger compounds appear not to score well because precise placement is critical, and their orientation space was sampled insufficiently. When actinomycin D and triostin A were rerun using a more recent DOCK version (version 2), which allows more versatility of sampling orientations, DOCK was able to find reasonable orientations with the planar rings in the intercalation site. Chromomysin, in addition to generating difficulties due to the large size of the molecule, binds as a metal-coordinated dimer [19], which DOCK would not be able to predict. Since the subset of the CSD used for this study included no compounds having more than 60 atoms, it is expected that this undersampling is not a serious problem for the search runs performed. With the small compounds (< 25 atoms), DOCK predicts orientations which match the crystal structure for the known complexes. However, since DOCK v. 1.01 uses an atom-based scoring scheme (meant to correlate with the number of intermolecular van der Waals contacts), the scores are consistently lower for smaller compounds, which simply have fewer possible contacts than larger compounds. Thus, these compounds are not highly ranked.

In all, since so many of the DNA-binding compounds ranked highly based on shape score alone, this emphasizes the importance of shape as a criterion for DNA binding.

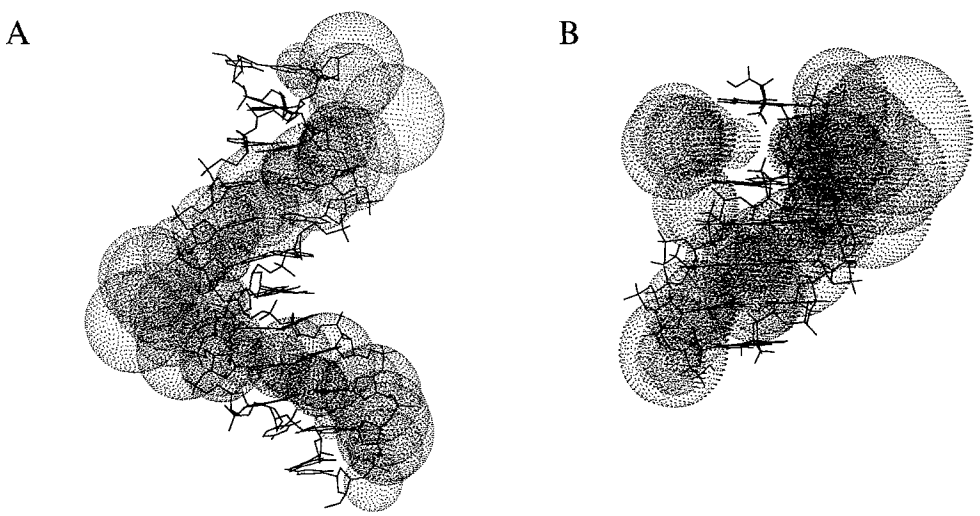


Fig. 1. Graphical representation of DOCK-generated spheres with (A) the crystal structure of the d(CGCGAATTCGCG)$_2$•netropsin complex and (B) the crystal structure of the d(CGTACG)$_2$ duplex.

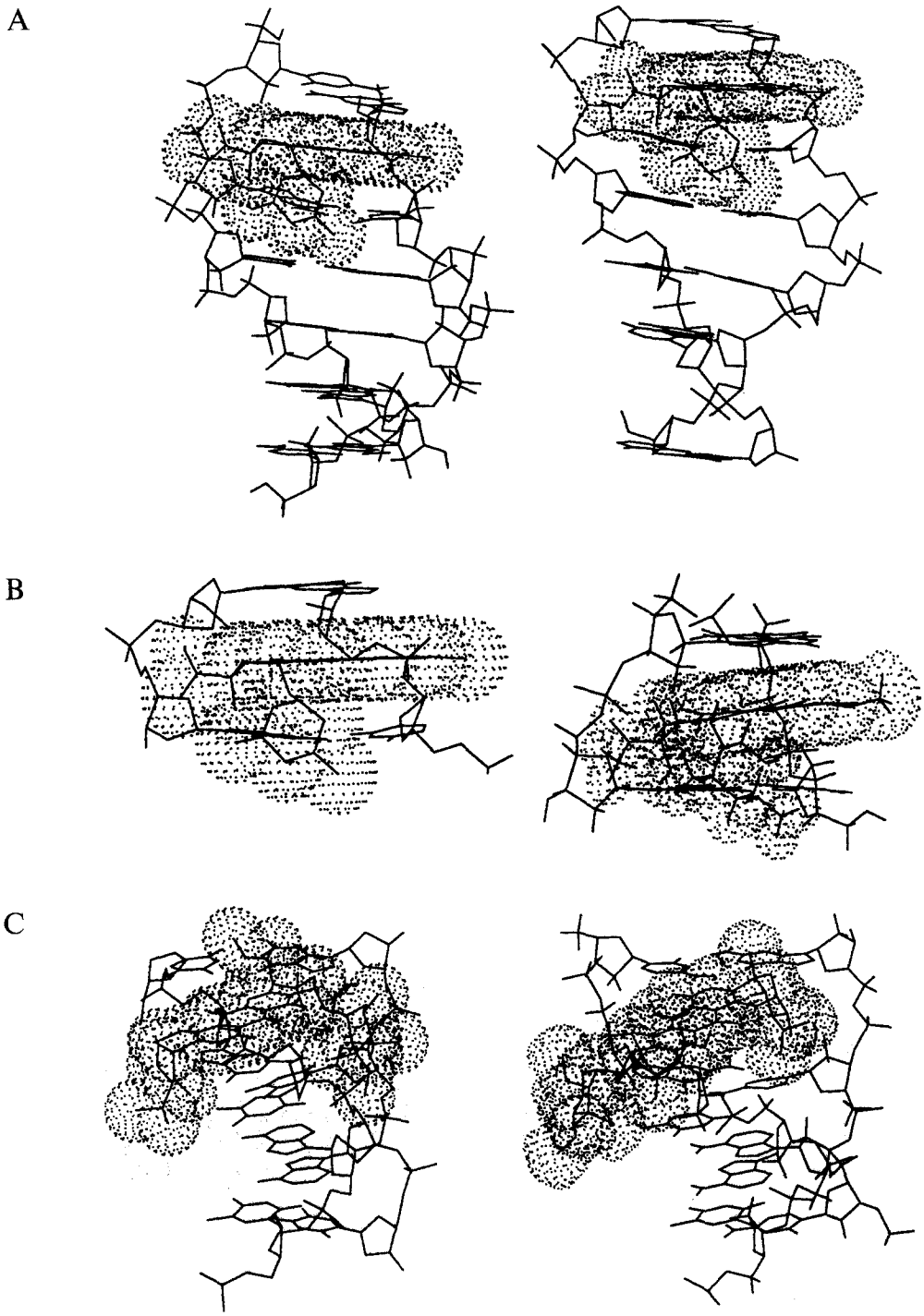


Fig. 2. Comparison of crystal complexes (left) with DOCK-generated orientations (right) for the following known DNA-binding compounds: (A) adriamycin; (B) daunomycin; (C) nogalamycin; (D) distamycin; (E) Hoechst_33258; (F) netropsin; and (G) berenil.

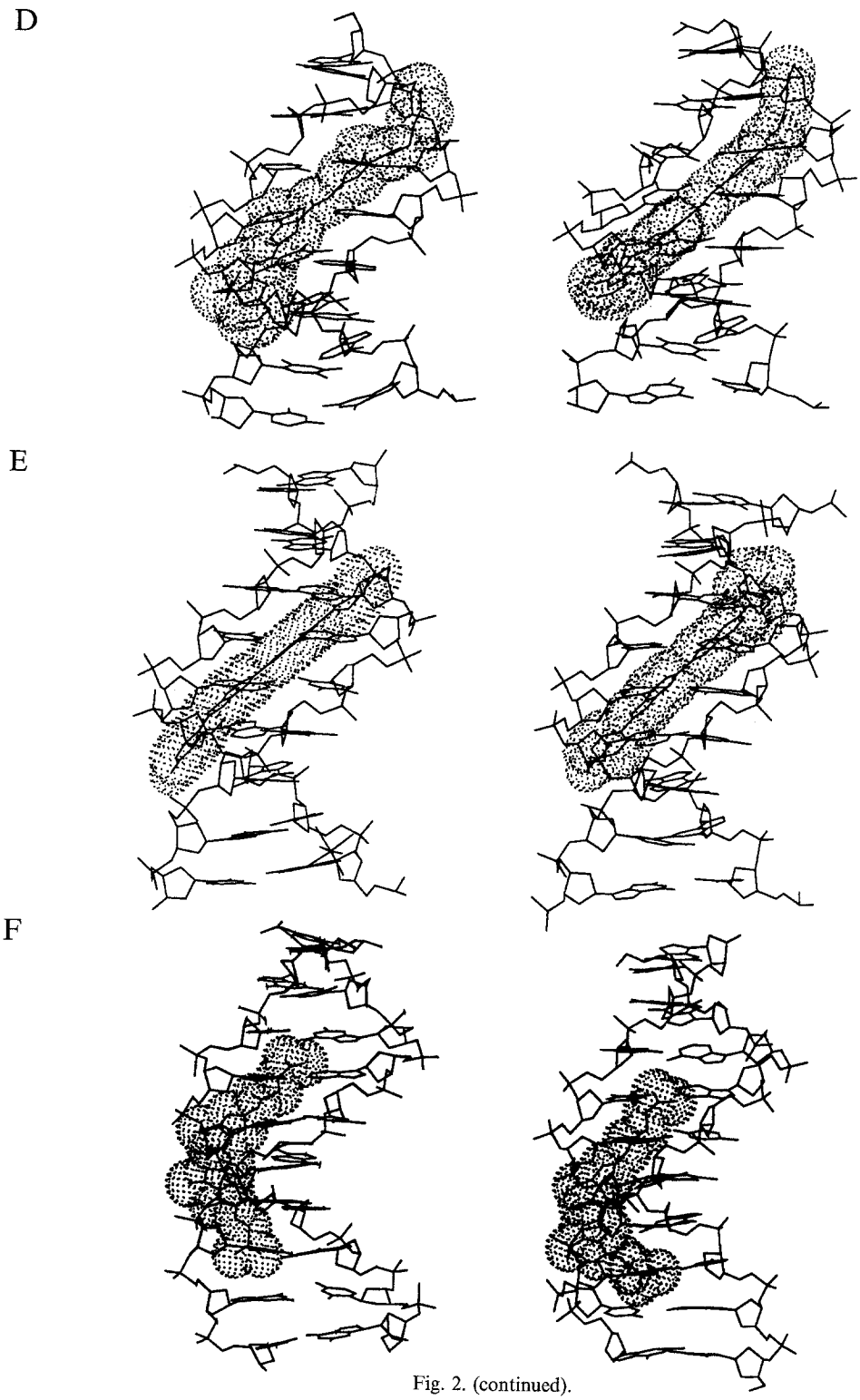


Fig. 2. (continued).

G

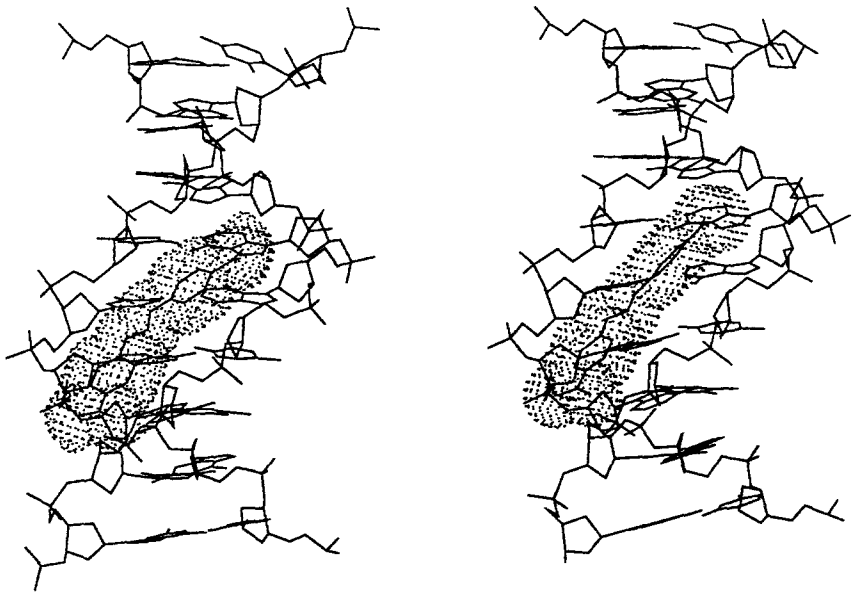


Fig. 2. (continued).

Minor and major groove binding

Docking the minor groove of d(CGCGAATTCGCG)$_2$

The DOCK program was run in the 'search' mode, in which molecules from a database are systematically positioned into the DNA site. The run was targeted at the minor groove of DNA for the duplex of the experimentally determined X-ray structure of the d(CGCGAATTCGCG)$_2$• netropsin complex. A graphical representation of the DOCK-generated spheres is given in Fig. 1A. For comparison, idealized A- and B-type DNA with the equivalent base-pair sequence and Z-type DNA with the d(CGCGCGCGCGCG)$_2$ sequence were also targeted. The 20 compounds with the highest scores for fitting in the minor groove of the crystal structure are collected in Table 2.

We were pleasantly surprised that the highest scoring molecule appeared to be CC-1065 (**1**), a potent antitumor agent known to interact with DNA [2b,20]. Other high-scoring molecules include (pre)vitamins, steroids, aromatic compounds, etc. Some of the molecules selected by DOCK are found in Scheme 2. Interestingly, eight compounds actually scored higher than netropsin itself. Follow-up molecular mechanics analyses of the energy-minimized complexes of CC-1065, netropsin and **15** with the DNA dodecamer found that the intermolecular van der Waals interaction energies in the complex amount to −63.7, −60.1 and −55.3 kcal/mol, respectively, which is in qualitative agreement with the trend found in the DOCK scores. In Fig. 3 a stereoview of the DNA dodecamer•CC-1065 complex is depicted in its conformation generated by DOCK.

It is clear that the ligand adopts a favorable orientation in the AT-rich part of the minor groove, consistent with results from experimental studies that indicate that CC-1065 prefers AT base pairs to CG base pairs near the adenine with which it forms a covalent adduct [2b]. Recent studies indicate that CC-1065 forms an adduct with DNA in which the ligand is covalently bound to the adenine N3 [20]. A detailed inspection of Fig. 3 suggests that CC-1065 adopts an orientation in the minor groove with its cyclopropane ring relatively close (4–5 Å) to an adenine N3.

The sequence to which it binds (d(GCTTA)) is very close, but different from the selective sequences it is known to react with (PuNTTA) [20d]. It would not be expected that DOCK could predict this selectivity, as studies have shown that the sequence selectivity is based on the sequence-dependent reactivity of adenine, rather than on the shape fit [20d]. However, the more complementary shapes have been shown to increase the rate of formation of the covalent adduct [20d], showing their importance. We suggest that the structure shown can be interpreted as a model for the supramolecular complex that precedes the formation of a covalently bound adduct.

Sensitivity for conformational changes of the minor groove

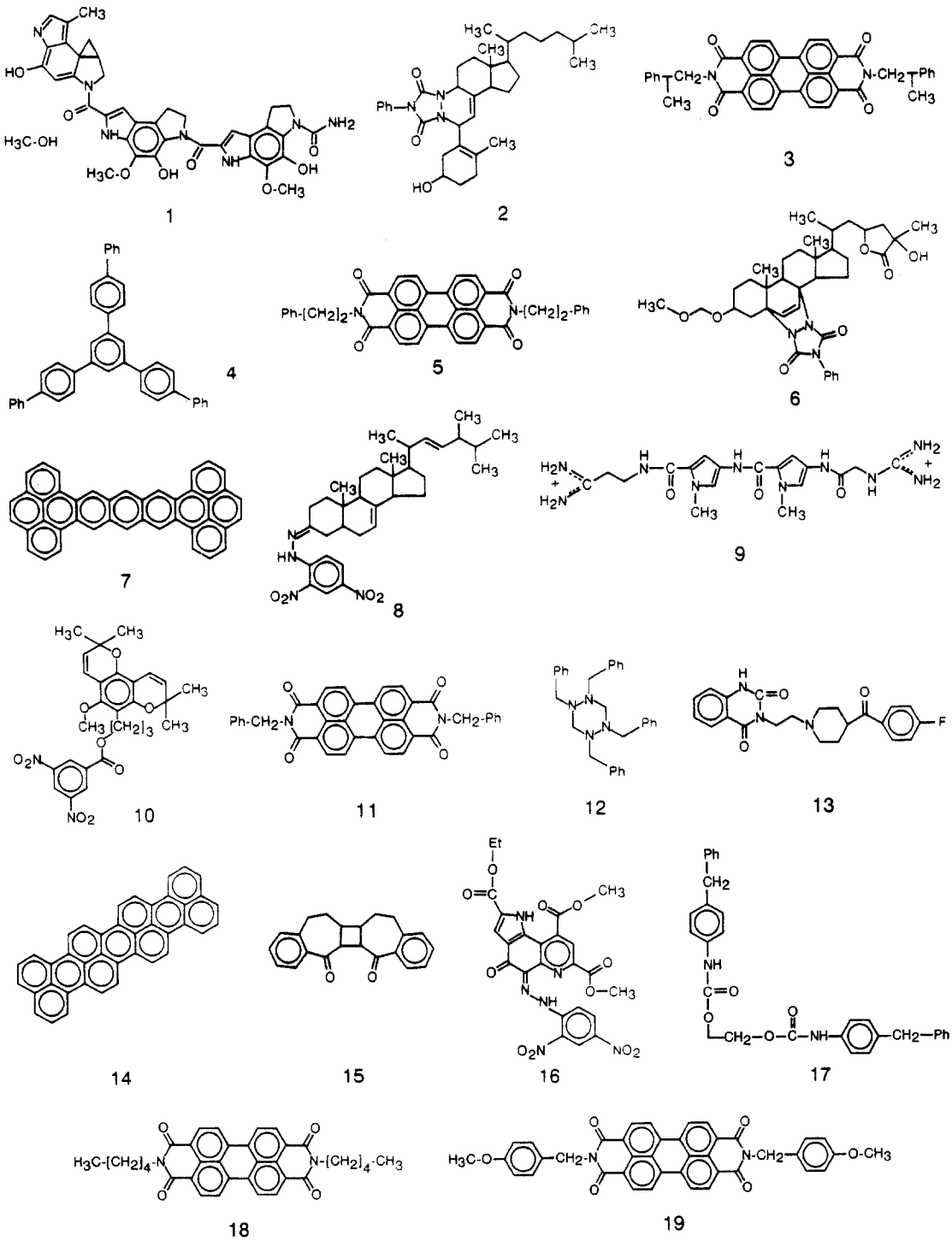
In order to study the sensitivity of the DOCK algorithm to *minor* conformational changes in the receptor site, we also calculated scores for the same top 20 molecules using an idealized B-DNA d(CGCGAATTCGCG)$_2$ structure. In Table 2 the ratios of the scores between the idealized receptor and the X-ray structure are found. In general, the scores are somewhat higher for the X-ray-determined receptor, but in most cases the relative scores are similar. Three molecules (**4**, **6** and **15**) are predicted to display more shape complementarity to the X-ray structure than the idealized B-DNA. The trend that the DOCK scores are somewhat higher in the case of the experimentally determined receptor might be caused by subtle local conformational changes due to the presence of netropsin in the minor groove of the crystalline complex. However, these results indicate that in the absence of experimentally determined structures of particular B-DNA sequences, model-

TABLE 2
(RELATIVE) DOCK SCORES OF THE 20 HIGHEST SCORING COMPOUNDS FITTING THE MINOR GROOVE OF d(CGCGAATTCGCG)$_2$ IN THE CONFORMATION OBSERVED IN THE CRYSTALLINE COMPLEX WITH NETROPSIN

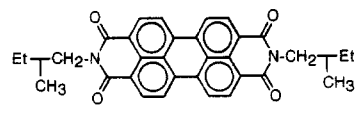
				Relative DOCK scores		
	CSD ref. code	No. of atoms	DOCK score	A-DNA[a]/X-ray	B-DNA[a]/X-ray	Z-DNA[b]/X-ray
1	CCATAG10	52	343.9	0.60	1.02	0.49
2	BINKAK	41	332.7	0.46	0.85	0.67
3	DICNEI	48	331.0	0.56	0.88	0.59
4	COSHUN	42	323.0	0.47	0.72	0.66
5	DICNIM	46	332.9	0.58	0.95	0.63
6	BOHHEL	47	322.3	0.58	0.76	0.67
7	TBZHCE	42	307.5	0.87	0.94	0.67
8	ERGDPH	42	306.4	0.55	1.00	0.62
9	Netropsin	31	300.4	0.52	0.99	0.66
10	ERSTND	38	296.8	0.55	0.85	0.61
11	DICPIO	44	295.8	0.70	0.89	0.71
12	TBZHTZ	34	293.3	0.51	0.90	0.56
13	BERGUA	29	291.9	0.51	0.81	0.79
14	QUATER10	40	288.9	0.73	1.01	0.73
15	DBZCTD	50	287.6	0.66	0.60	0.48
16	COWTAJ	41	286.1	0.51	0.90	0.61
17	CEVFOY	36	285.6	0.50	0.93	1.01
18	DICMUX	40	285.3	0.57	0.87	0.62
19	DICPEK	48	283.6	0.78	0.92	0.59
20	DICNOS	40	282.4	0.62	0.91	0.70

[a] Right-handed A- and B-type DNA of the duplex d(CGCGAATTCGCG)$_2$ with idealized geometries generated by NUCGEN.

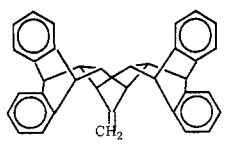
[b] Left-handed Z-type DNA of the duplex d(CGCGCGCGCGCG)$_2$ generated by NUCSEQZ.



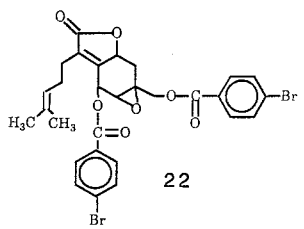
Scheme 1. Structures of molecules **1–37**.



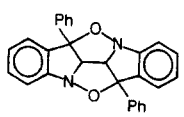
20



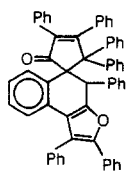
21



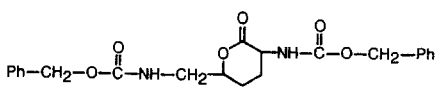
22



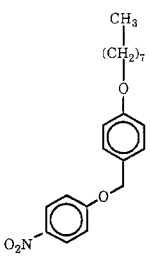
23



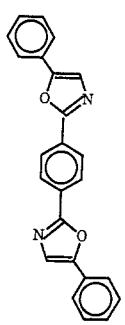
24



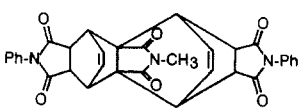
25



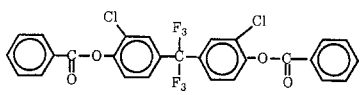
26



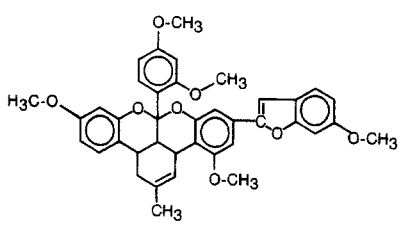
27



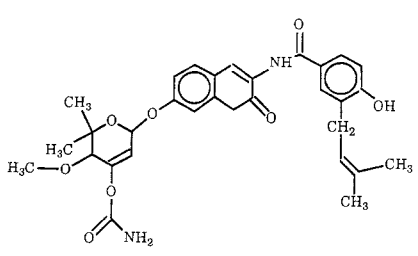
28



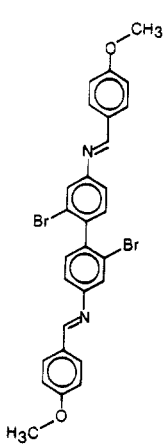
30



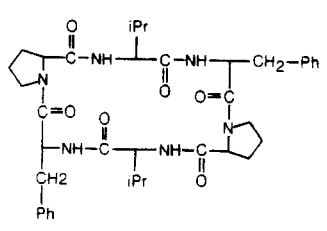
29



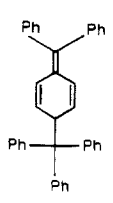
31



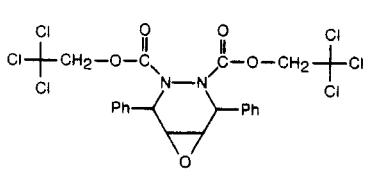
35



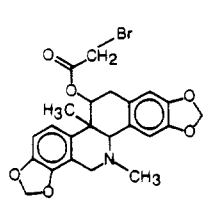
32



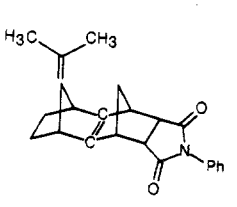
33



34



36



37

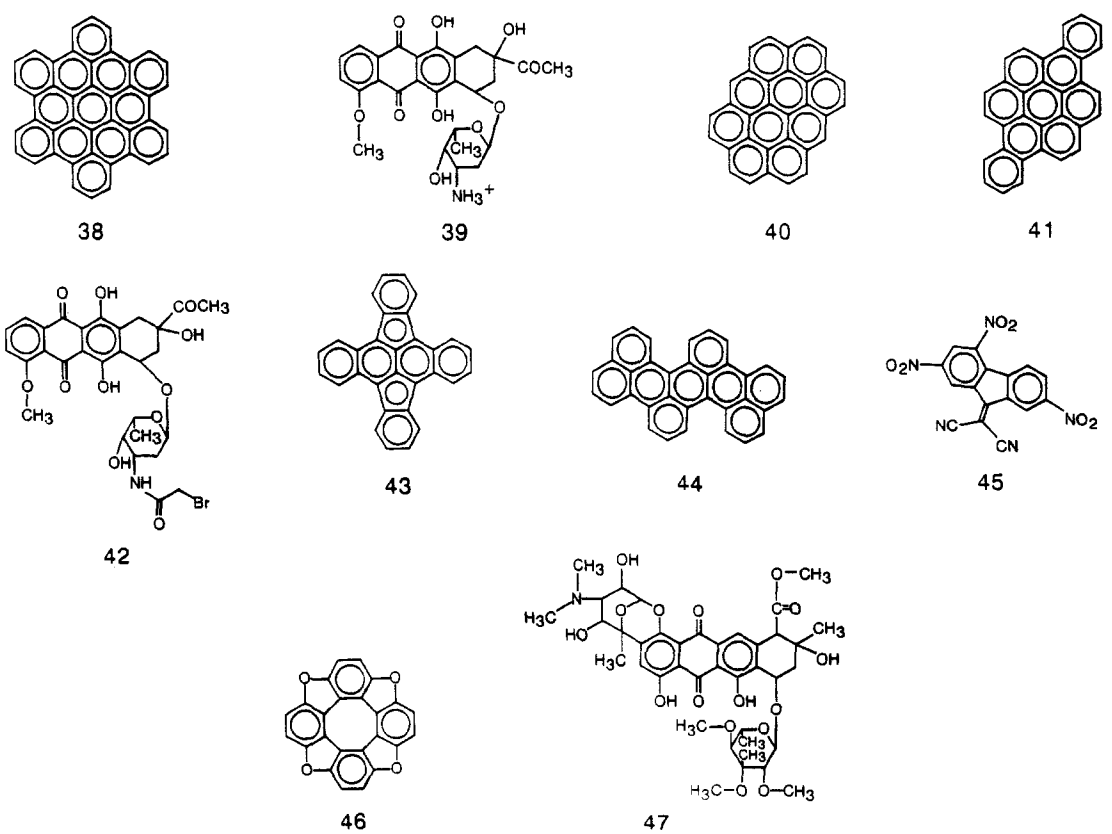
Scheme 1. (continued).

built structures of DNA can be used effectively to find ligands which might bind to the DNA.

The effect of *major* conformational changes on the DOCK scoring was studied by calculating the scores of the top 20 molecules for fitting the major groove of A-type $d(CGCGAATTCGCG)_2$ and the minor groove of Z-type $d(CGCGCGCGCGCG)_2$ double helices. The relative scores are summarized in Table 2. The average relative scores (0.59 ± 0.11 and 0.65 ± 0.11 for A-DNA and Z-DNA, respectively) clearly indicate that the fit of these molecules is poor compared to the fit with idealized B-DNA (average relative score 0.89 ± 0.10). This result is in agreement with the different shape of the grooves of A- and Z-DNA. It is known that the major groove of A-DNA is relatively deep and the minor groove of Z-DNA has been described as being very narrow [21]. On the basis of the DOCK scores one would predict compounds **2** and **4** to better fit the B- than the A-DNA minor groove. In contrast, the more elongated compound **48**, which was ranked 29th with a score of 276 on X-ray DNA, had a relative score of 0.29 for Z-DNA/X-ray. Therefore it is predicted to display specificity for B-DNA over Z-DNA.

The effect of altered base-pair sequences

The strong preference of netropsin to interact with AT-rich parts of DNA rather than with GC-rich parts has been explained in terms of base-dependent differences in the minor groove [10]. AT base sequences result in a narrower minor groove, which is more open to deep penetration



Scheme 2. Structures of molecules **38**–**47**.

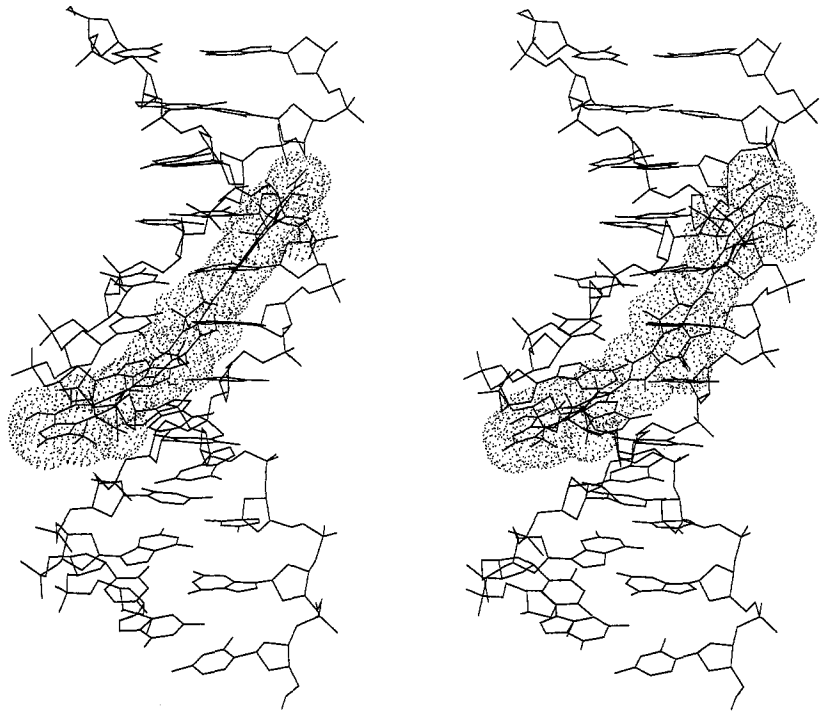


Fig. 3. Cross-eye stereoview of the DOCK-generated energy-minimized DNA dodecamer•CC-1065 complex.

by ligands than the wider but less penetrable minor groove in GC-rich areas. In order to find out if DOCK is able to reproduce these subtle shape differences, we decided to target DOCK at a $d(CGCGCGCGCGCG)_2$ duplex and calculate the corresponding scores for a d(CGCGAA-

TABLE 3
(RELATIVE) DOCK SCORES OF THE 12 HIGHEST SCORING COMPOUNDS FITTING THE MINOR GROOVE OF A B-TYPE $d(CGCGCGCGCGCG)_2$ DOUBLE HELIX AND THE CORRESPONDING SCORES FOR $d(CGCGAATTCGCG)_2$

			Base pairs 5–8		
Cmpd	CSD reference code	No. of atoms	CGCG	AATT	CGCG/AATT
21	BZEHPN	36	287.0	251.2	1.14
22	BEBPON10	38	283.6	246.3	1.15
23	EPPHIN10	32	276.4	215.1	1.28
24	CIJXUO	60	273.5	160.6	1.70
1	CCATAG10	52	269.9	352.5	0.77
25	LYSLAC	30	269.8	222.6	1.21
26	NPOXBZ10	27	265.6	237.1	1.12
27	PHOXBZ	28	264.6	253.3	1.02
28	MEAZML	42	262.5	163.3	1.61
29	CANREO	47	261.6	199.6	1.31
30	BOFPRO	41	260.6	225.0	1.16
31	NIVBIO	44	259.9	226.2	1.15

TABLE 4
(RELATIVE) DOCK SCORES OF THE COMPOUNDS SHOWING THE HIGHEST PREFERENCE FOR THE MAJOR GROOVE OF A-TYPE d(CGCGAATTCGCG)$_2$

	CSD reference code	No. of atoms	A-DNA[a]	A-DNA/X-ray	A-DNA/B-DNA[a]	A-DNA/Z-DNA[b]
32	PAPRVA	50	284.3	1.51	1.59	3.47
33	PMYCHD	38	234.4	1.70	1.69	1.92
34	DESJAM	35	221.2	1.58	1.48	1.78

[a] Right-handed A- and B-type DNA of the duplex d(CGCGAATTCGCG)$_2$ with idealized geometries generated by NUCGEN.
[b] Left-handed Z-type DNA of the duplex d(CGCGCGCGCGCG)$_2$ generated by NUCSEQZ.

TTCGCG)$_2$ duplex. The flanking sequences are identical in both dodecamers, so differences in the scores will be due to shape differences in the minor groove induced by the central four base pairs. The results for the 12 highest scoring molecules are collected in Table 3.

It is evident that in an *absolute* sense the scores calculated for the CG-rich duplex are generally lower than the scores found for the Dickerson dodecamer. This is in agreement with the observation that so far, few minor groove binders are known that expose a higher affinity for CG-rich DNA relative to AT-rich DNA. Remarkably, again CC-1065 (**1**) is among the highest scoring molecules, although the score is significantly smaller than that found with the Dickerson dodecamer. The same holds for netropsin which had a DOCK score of 237 (relative score 0.80), ranking it 47th for the d(CGCGCGCGCGCG)$_2$ duplex. These *relatively* lower scores are in qualitative agreement with the thermodynamic data, indicating a 5.6 kcal/mol difference in binding free energy of netropsin with poly dAT•poly dAT in favor of poly dGC•poly dGC [10]. Very interestingly, several molecules, i.e., **23**, **24**, **25**, **28** and **29**, display significantly (ratio of scores >1.2) higher scores. These compounds may serve as leads for designing molecules that prefer to bind CG-rich relative to AT-rich compounds.

Docking in the major groove of A-DNA

The DOCK algorithm was applied to the NUCGEN-generated, idealized A-type d(CGCGAATTCGCG)$_2$ duplex. For comparison the scores of the highest ranked molecules were also calculated for B- and Z-type DNA, the latter with the d(CGCGCGCGCGCG)$_2$ base sequence. Three molecules that turned out to display the highest *relative* scores (and therefore might have selectivity) for A-DNA are given in Table 4. In Fig. 4 the DOCK-generated orientations of the three molecules are displayed.

TABLE 5
(RELATIVE) DOCK SCORES OF THE COMPOUNDS SHOWING THE HIGHEST PREFERENCE FOR THE MINOR GROOVE OF Z-TYPE d(CGCGCGCGCGCG)$_2$

	CSD reference code	No. of atoms	Z-DNA[a]	Z-DNA/A-DNA[a]	Z-DNA/X-ray	Z-DNA/B-DNA[a]
35	BMXYAP	34	272.8	1.82	1.34	1.40
36	EPICOR10	31	260.3	1.95	1.47	1.43
37	DEPYEC	26	255.4	1.78	1.29	1.30

[a] Right-handed A- and B-type DNA of the duplex d(CGCGAATTCGCG)$_2$ with idealized geometries generated by NUCGEN.

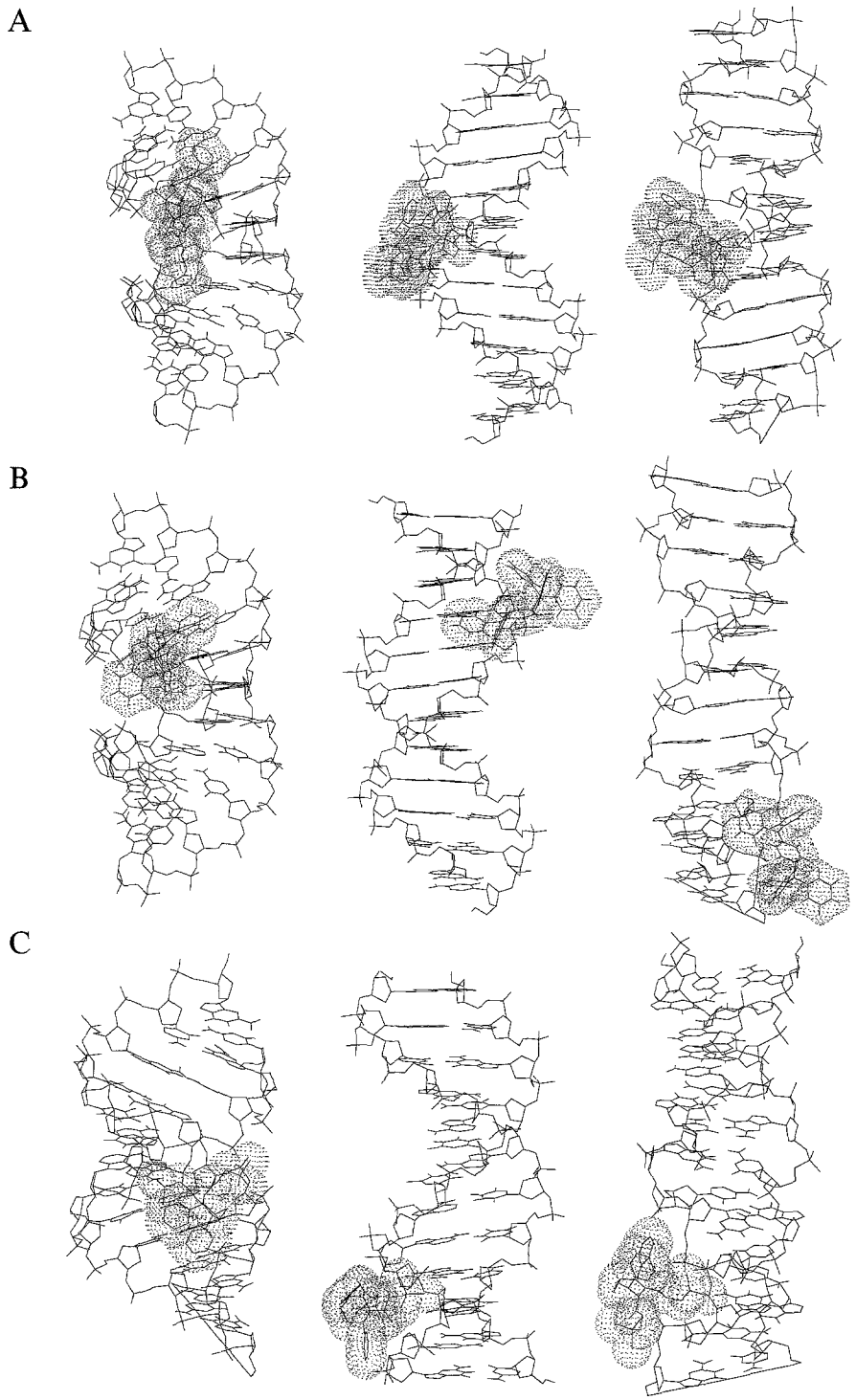


Fig. 4. DOCK-generated orientations of compounds (A) **32**; (B) **33**; and (C) **34** in the grooves of A-, B- and Z-type DNA dodecamers, respectively. See Table 4 for additional data on the used sequences.

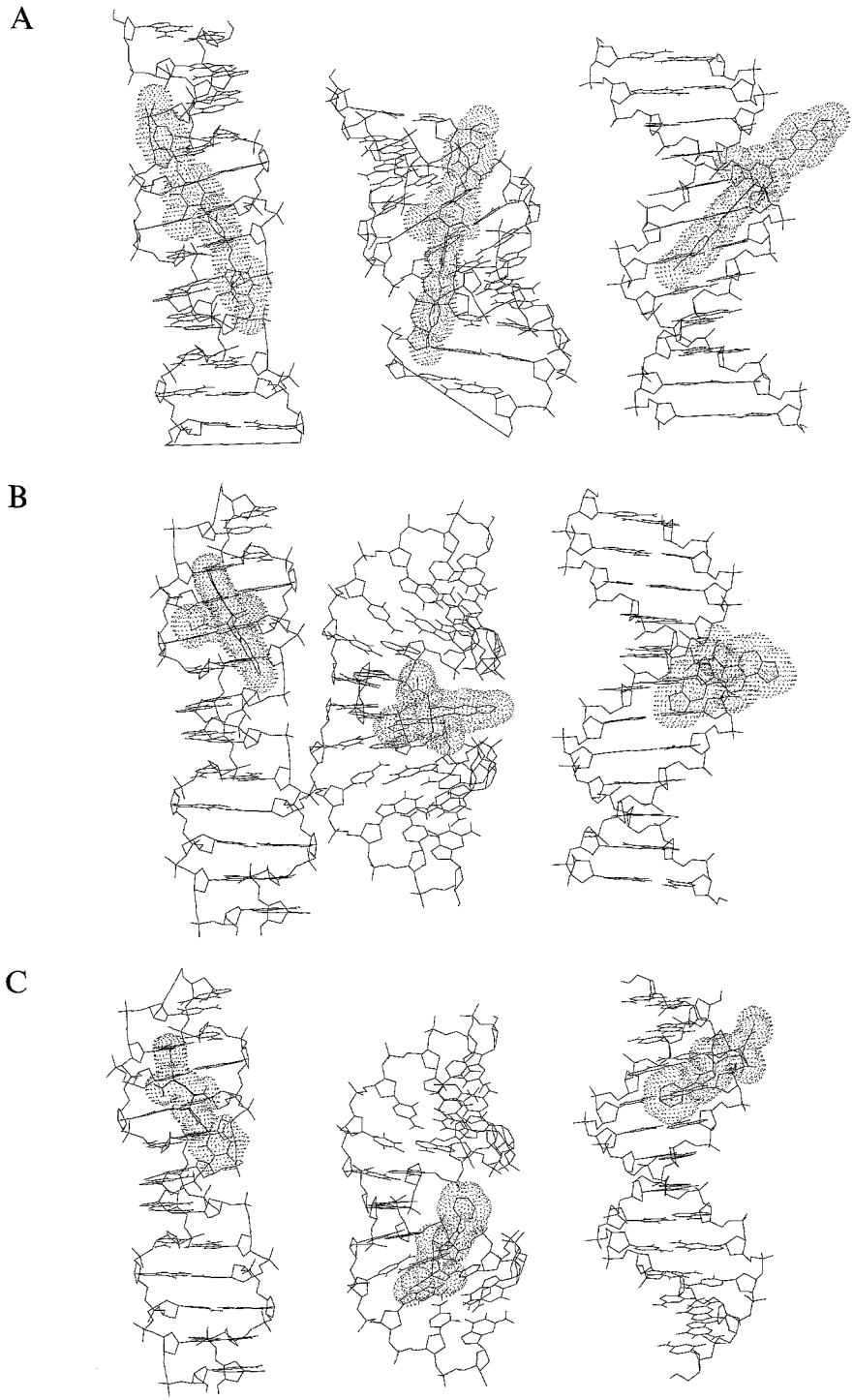


Fig. 5. DOCK-generated orientations of compounds (A) **35**; (B) **36**; and (C) **37** in the grooves of Z-, A- and B-type DNA dodecamers, respectively. See Table 5 for additional data on the used sequences.

Docking in the minor groove of Z-DNA

The three molecules exhibiting the highest relative scores with respect to preferring to be docked in the minor groove of Z-type d(CGCGCGCGCGCG)$_2$ are given in Table 5. Figure 5 depicts the DOCK-generated orientations of these compounds.

The scores for Z-DNA docking are in the same range as those found for the highest scoring B-type DNA minor groove binders and therefore significantly higher than the scores for the top A-DNA-selective molecules. This would indicate that the minor groove of Z-DNA seems more suited to act as a host for ligands than the major groove of A-DNA. Compounds **35**, **36** and **37** may serve as starting points for the design of compounds that would stabilize Z-type DNA.

Intercalation

Docking to a potential intercalation site

By removing daunomycin from its position between the first two base pairs of B-type d(CGTACG)$_2$, a potential intercalation site is formed that was targeted using the DOCK approach. It was checked that the overlapping spheres described both this intercalation site and the minor groove (see Fig. 1B). After that, DOCK was run in the 'search mode' and the highest scoring molecules are collected in Table 6 and Scheme 2.

For comparison, the scores of these compounds were also calculated for binding in the minor groove of B-type d(CGCGAATTCGCG)$_2$. DOCK mainly found relatively flat, mostly aromatic molecules, that were oriented in the intercalation site. Nogalamycin, **47**, a known intercalator, was among the top 12 compounds, along with BADAUN, the *N*-bromoacetyl derivative of daunomycin. No molecules were selected that explore the ability of daunomycin to simultaneously bind in the intercalation site and in the minor groove. Daunomycin is among the highest scoring ligands.

DISCUSSION

DOCK's performance

In the present study, DOCK was applied to variously shaped grooves and intercalation sites occurring in DNA. The high-scoring molecules that were selected by DOCK originated from many different chemical classes. However, they all had in common a complementary shape of at least part of the molecular framework to the receptor shape under study.

How relevant are the high-scoring compounds to DNA binding? In general, it would be surprising to find a large fraction of the molecules selected by DOCK to actually bind DNA to an appreciable extent. A well-known DNA covalent minor groove binder was selected only in the searches where CC-1065 turned up. In the case of CC-1065, not only was the molecular shape complementary, but the electrostatic profile and H-bond donors and acceptors were in the proper positions as well. It is clear, therefore, that the molecules selected by DOCK with the scoring scheme in DOCK v. 1.01 may only serve as a starting point or template for further molecular design. However, there are other indications that the chemical classes selected by DOCK may have some relevance. We have studied a number of compounds that are known to bind to DNA to use as a control (Table 1). Although some of the more complex ligands are missed in this version of DOCK, and the smaller ones score less well because the scoring system favors larger ligands, many of the well-known intercalators and groove binders score well. This suggests the

TABLE 6
(RELATIVE) DOCK SCORES OF THE 12 COMPOUNDS SHOWING THE HIGHEST SCORES FOR DOCKING INTO THE INTERCALATION SITE OF $d(CGTACG)_2$ IN THE CONFORMATION AS OBSERVED IN ITS COMPLEX WITH DAUNOMYCIN

	CSD reference code	No. of atoms	X-ray	X-ray/B-DNA[a]
38	HBZCOR	42	525.4	2.75
39	Daunomycin	38	482.0	2.24
40	OVALENO1	32	465.0	2.12
41	DBZCOR	32	458.8	2.11
42	BADAUN	42	448.1	2.16
43	PHNAPH	30	447.3	2.13
44	NAPANT	34	441.0	2.01
45	CMNFLO10	30	440.8	2.37
16	COWTAJ	41	439.7	1.70
46	TPTFUR20	28	438.4	2.38
47	BIDROV10	56	432.1	2.34

[a] Right-handed B-type DNA of the duplex $d(CGCGAATTCGCG)_2$ with idealized geometry generated by NUCGEN.

promise of the method for suggesting other DNA binders. An example are the steroids that were selected as potential minor groove binders (**6**, **8**). It has been suggested that some steroids might indeed interact with DNA [22]. The flat aromatic compounds chosen by DOCK as potential intercalators also seem to be reasonable. Thus, we feel that the various receptor sites have been characterized rather well by DOCK and the new ligands suggested might indeed be relevant. Based on known DNA-binding ligands, one would obviously select ligands with both a high DOCK score and a positive charge. One of the most exciting results of this study is the suggestion of a number of compounds that could be developed which, by their shape properties, would be selective for A- or Z-DNA over B-DNA. The same holds for the compounds that are predicted to show preference for CG-rich DNA over AT-rich DNA (Table 2).

What are the major limitations to the version of DOCK used in this study? The first would be the absence of electrostatic and H-bonding interactions in the scoring function. The more recent version of DOCK (version 3.0) allows a more sophisticated 'force field'-based scoring method to be used. On the other hand, such force field scoring methods will likely exaggerate the electrostatic interactions in an unrealistic way, as there is no solvation correction in this scoring function, and the DNA backbone is highly charged. Therefore it is unlikely that more sophisticated scoring schemes would be a significant improvement over the one employed here until a more accurate solvation/desolvation free energy can be calculated.

A second limitation to DOCK is its dependence on the composition of the database. Of course the database composition will affect the choices made by DOCK, as will the conformation of the compounds in the database. We feel that the latter is an even more important limitation than the former. Since only one conformation of each compound in the database is docked and evaluated, many more interesting conformations of the compounds are not considered. Nevertheless, useful numbers of interesting leads were found by the currently used version of DOCK. In fact, the efficiency of DOCK to find and retrieve, for instance, an established minor groove binder from the database will of course depend on the conformation of the binder and its substituents.

Proper electrostatic complementarity is important in DNA binding. Still, obtaining the proper molecular shape is a prerequisite before this factor can play a role.

CONCLUSIONS

The DOCK algorithm has been used successfully in order to select both known DNA-binding molecules as well as molecules with significant shape complementarity to A-, B- and Z-DNA from a database. In addition, lead structures for designing DNA intercalators were generated. Apparently, the shape of the various grooves and intercalation sites can be well described by the DOCK approach. Interestingly, some of the molecules found by DOCK may well expose selectivity to bind a (minor or major) groove of A-, B- or Z-DNA. The effect of different base pairs on the minor groove shape of B-DNA was reproduced by DOCK, resulting in the prediction of potential lead structures that would serve as starting points for ligands that prefer to bind to CG-rich zones in B-DNA rather than to AT-rich zones. We feel that the approach as described in this paper has provided molecular frameworks that may act as a source of inspiration for further drug design.

ACKNOWLEDGEMENTS

P.A.K. and I.D.K. are pleased to acknowledge research support from Grants CA-25644 (P.A.K.), GM-31497 (I.D.K.) and GM-39552 (I.D.K. and P.A.K., George Kenyon, P.I.). The facilities of the UCSF Computer Graphics Laboratory, supported by Grant RR-1081, are gratefully acknowledged.

REFERENCES

1 Hurley, L.H., J. Med. Chem., 32 (1989) 2027.
2 a. Dervan, P.B., Science, 232 (1986) 464.
b. Zimmer, C. and Wähnert, U., Prog. Biophys. Mol. Biol., 47 (1986) 31.
3 a. Neidle, S. and Abraham, Z., Crit. Rev. Biochem., 17 (1984) 73.
b. Pullman, B., In Sarma, R.H. and Sarma, M.H. (Eds.) Structure and Expression, Vol. 2: DNA and Its Drug Complexes, Adenine Press, Albany, NY, 1988, pp. 237–249.
c. Pullman, B., Adv. Drug. Res., 18 (1989) 1.
4 a. Cieplak, P., Rao, S.N., Grootenhuis, P.D.J. and Kollman, P.A., Biopolymers, 29 (1990) 717.
b. Frederick, C.A., Williams, L.D., Ughetto, G., Van der Marel, G.A., Van Boom, J.H., Rich, A. and Wang, A.H.J., Biochemistry, 29 (1990) 2538.
c. Liaw, Y.C., Gao, Y.G., Robinson, H., Van der Marel, G.A., Van Boom, J.H. and Wang, A.H.J., Biochemistry, 28 (1989) 9913.
5 Pyle, A.M., Rehmann, J.P., Meshoyrer, R., Kumar, C.V., Turro, N.J. and Barton, J.K., J. Am. Chem. Soc., 111 (1989) 3051.
6 a. Kuntz, I.D., Blaney, J.M., Oatley, S.J., Langridge, R. and Ferrin, T.E., J. Mol. Biol., 161 (1982) 269.
b. DesJarlais, R.L., Sheridan, R.P., Seibel, G.L., Dixon, J.S., Kuntz, I.D. and Venkataraghavan, R., J. Med. Chem., 31 (1988) 722.
c. DesJarlais, R.L., Seibel, G.L., Kuntz, I.D., Ortiz de Montellano, P., Furth, P.S., Alvarez, J.C., DeCamp, D.L., Babé, L.M. and Craik, C.S., Proc. Natl. Acad. Sci. USA, 87 (1990) 6644.
7 Part of this work appeared in preliminary form in: Grootenhuis, P.D.J., Kollman, P.A., Seibel, G.L., DesJarlais, R.L. and Kuntz, I.D., Anti-Cancer Drug Design, 5 (1990) 237.
8 a. Shoichet, B.K., Bodian, D.L. and Kuntz, I.D., J. Comput. Chem., 13 (1992) 380.
b. Meng, E.C., Shoichet, B.K. and Kuntz, I.D., J. Comput. Chem., 13 (1992) 505.
9 Kopka, M.L., Yoon, C., Goodsell, D., Pjura, P. and Dickerson, R.E., Proc. Natl. Acad. Sci. USA, 82 (1985) 1376.
10 Breslauer, K.J., Ferrante, R., Marky, L.A., Dervan, P.B. and Youngquist, R.S., In Sarma, R.H. and Sarma, M.H.

(Eds.) Structure and Expression, Vol. 2: DNA and Its Drug Complexes, Adenine Press, Albany, NY, 1988, pp. 273–290.
11 Arnott, S., Hukins, D.W.L., Dover, S.D., Fuller, W. and Hodgeson, A.R., J. Mol. Biol., 81 (1973) 7204.
12 Quigley, G.J., Wang, A.H.J., Ughetto, G., Van der Marel, G., Van Boom, J.H. and Rich, A., Proc. Natl. Acad. Sci. USA, 77 (1980) 7204.
13 Seibel, G.L., manuscript in preparation.
14 Connolly, M.L., Science, 221 (1983) 709.
15 Ferrin, T.E., Huang, C.C., Jarvis, L.E. and Langridge, R., J. Mol. Graphics, 6 (1988) 13.
16 DesJarlais, R.L., Ph.D. Thesis, University of California at San Francisco, San Francisco, CA, 1988.
17 a. Weiner, P.K. and Kollman, P.A., J. Comput. Chem., 2 (1981) 287.
b. Weiner, S.J., Kollman, P.A., Nguyen, D.T. and Case, D.A., J. Comput. Chem., 7 (1986) 230.
c. Weiner, S.J., Kollman, P.A., Case, D.A., Singh, U.C., Ghio, C., Alagona, G., Profeta Jr., S. and Weiner, P., J. Am. Chem. Soc., 106 (1984) 765.
18 Singh, U.C., Weiner, P.K., Caldwell, J. and Kollman, P.A., AMBER 3.0, University of California at San Francisco, San Francisco, CA, 1987.
19 Gao, X., Mirau, P. and Patel, D.J., J. Mol. Biol., 223 (1992) 259.
20 a. Chidester, C.G., Krueger, W.C., Mizsak, S.A., Duchamp, D.J. and Martin, D.G., J. Am. Chem. Soc., 103 (1981) 7629.
b. Hurley, L.H., Reynolds, V.L., Swenson, D.H., Petzold, G.L. and Scahill, T.A., Science, 226 (1984) 843.
c. Needham-VanDevanter, D.R., Hurley, L.H., Reynolds, V.L., Theriault, N.T., Krueger, W.C. and Wierenga, W., Nucleic Acids Res., 12 (1984) 6159.
d. Hurley, L.H., Lee, C.S., McGovren, J.P., Warpehoski, M.A., Mitchell, M.A., Kelly, R.C. and Aristoff, P.A., Biochemistry, 27 (1988) 3886.
e. Lee, M., Shea, R.G., Hartley, J.A., Kissinger, K., Pon, R.T., Vesnaver, G., Breslauer, K.J., Dabrowiak, J.C. and Lown, J.W., J. Am. Chem. Soc., 111 (1989) 345.
f. Scahill, T.A., Jensen, R.M., Swenson, D.H., Hatzenbuhler, N.T., Petzold, G., Wierenga, W. and Brahme, N.D., Biochemistry, 29 (1990) 2852.
21 a. Saenger, W., Principles of Nucleic Acid Structure, Springer, New York, NY, 1984.
b. Barton, J.K., Chem. Eng. News, 66 (1988) 30.
22 a. Strähle, U., Boshart, M., Klock, G., Stewart, F. and Schütz, G., Nature, 339 (1989) 629.
b. Arya, S.K. and Yang, J.T., Biochemistry, 14 (1975) 963.
c. Chandrasekhar, B., Mrigank and Kothekar, V., FEBS Lett., 225 (1987) 151.
d. Blackburn, G.M., Kellard, B., Rahid, A. and Thompson, M.H., Biochem. Pharmacol., 32 (1983) 2571.
e. Hendry, L.B., J. Steroid Biochem. Mol. Biol., 31 (1988) 493.
f. Spelsberg, T.C., Ruh, T., Ruh, M., Goldberger, A., Horton, M., Hora, J. and Singh, R., J. Steroid Biochem. Mol. Biol., 31 (1988) 579.
g. Hui, X., Gresh, N. and Pullman, B., Nucleic Acids Res., 17 (1989) 4177.
23 FCD-3D, version 89.2 is available from Molecular Design Ltd, San Leandro, CA.
24 a. Pearlman, R.S., CDA News, 2 (1987) 1.
b. Pearlman, R.S., Rusinko, A., Skell, J.M. and Balducci, R., CONCORD, Tripos Associates Inc., St. Louis, MO, 1987.