PAPERS

A search for specificity in DNA-drug interactions

G. Cruciani and P.J. Goodford

Laboratory of Molecular Biophysics, Oxford, UK

The GRID force field and a principal component analysis have been used in order to predict the interactions of small chemical groups with all 64 different triplet sequences of B-DNA. Factors that favor binding to guanine-cytosine base pairs have been identified, and a dictionary of ligand groups and their locations is presented as a guide to the design of specific DNA ligands.

Keywords: COMFA, DNA, drug interactions, GRID, molecular design, principal components, selectivity

INTRODUCTION

The recognition of one molecule by another is fundamental to all living systems, and the replication of DNA and its transcription to RNA is perhaps one of the most remarkable and best known examples of molecular recognition. Small ligands also bind to double-stranded DNA by different modes that can be classified according to the nature of their DNA interactions. Intercalating compounds and nonintercalating groove binders make efficient physical associations with DNA, while covalently binding ligands make chemical bonds that are normally irreversible.

Many noncovalent DNA ligands are known and include both natural antibiotics and synthetic drugs, some of which were developed for cancer therapy. However, many of these molecules[1] tend to show a relatively low preference for the minor groove of guanine-cytosine (*GC*) sequences of DNA and it is not easy to design ligands capable of selectively binding to a predefined DNA sequence that has both *GC* and adenine-thymine (*AT*) pairs. Moreover, there is a particular interest in developing ligands with *GC* selectivity, because regions of high *GC* content are common in the genomes of mammals, including man,[2] and because certain oncogenes are particularly rich in *GC* base pairs.[3,4]

The aim of this research was to find the chemical groups that should be incorporated in a DNA ligand molecule and where they should be incorporated in order to confer a specificity on the ligand so that it will bind well to one selected triplet, but not as tightly to the others. First, the strategy was to use the (GRID) procedure[5,6] in order to describe the binding between certain known chemical groups and all 64 triplets, and then to use principal component analysis (PCA)[7] in order to condense and collate the predictive information from GRID. A general table of chemical groups and their locations was prepared (see Table 6) as a guide to the design of specific DNA ligands.

Address reprint requests to Dr. Cruciani at his permanent address: Dipartimento di Chimica, Universita' di Perugia, Via Elce di Sotto 10, 06123 Perugia, Italy.

Received 5 October 1993; accepted 9 December 1993; revised 30 November 1993

METHOD

The procedure can be divided into five major steps:

(1) modelling the 64 double-stranded DNA triplets in their beta conformation
(2) multivariate characterization of the triplets by their energies of interaction with small chemical groups using the GRID program
(3) building description matrices that contain the information collected from the GRID data
(4) using PCA to rationalize the structure of the data and to extract as much information as possible from the description matrices
(5) chemical interpretation of the statistical analyses

Modeling the DNA triplet structures

Because the B-conformation of DNA is the most commonly studied form, all 64 double-stranded DNA triplets have been modeled in that conformation. Each triplet was sandwiched between the same four *AT* end pairs in order to anneal the ends of the triplets themselves and to obtain a better definition of their chemical surroundings. The resulting 64 double-stranded heptaplets were modeled using standard Quanta geometry;[8] peripheral studies demonstrated that the choice of *AT* end pairs did not significantly influence the final statistical results.

To obtain strictly comparable coordinates for all of the target structures, the first two *AT* pairs and the last two *AT* pairs of each target were superimposed on the corresponding pairs of the TTAAATT double-stranded heptaplet. The sugar phosphate backbones of all of the molecules were thus aligned and their structural variations were only determined by the base variations of the middle three base pairs of each heptaplet.

In the absence of counterions, these molecules would have a large net anionic charge, so one potassium counterion

was added to each phosphate (except for the terminal phosphates, to which one magnesium counterion was added). Each cation was placed radially at 7.0 Å from the corresponding phosphate group so that it was not directly binding to the DNA. This resulted in an uncharged system for computational studies.

The GRID force field

It is well known that pure hydrogen bonding considerations are insufficient for the correct assessment of *GC* versus *AT* specificity of groove binding ligands.[9,10] Specificity can be accounted for only by taking into consideration the shape and overall electronic properties of the interacting species, and explicitly calculating all the relevant contributions to the energies of binding. The GRID program[5,6] was used in the present work in order to describe the interactions between small chemical groups (the probes) and each of the DNA heptaplets (the targets). GRID is a computational procedure for detecting energetically favorable binding sites on molecules of known three-dimensional (3D) structures such as DNA. The energies are calculated as the electrostatic, hydrogen bond and Lennard-Jones interactions of the probe group with the target structure, using a position-dependent dielectric function in order to modulate the strong electrostatic interactions between charged centers, since solvent molecules were not explicitly included with the DNA targets. The calculations were performed using 0.5 Å spacing between grid points in a rectangular box measuring 18 × 18 × 10 Å which included all the base atoms of the triplet. With a probe at the first grid point, the nonbonded interaction between probe and target was calculated as

$$E_{xyz} = \Sigma E_{LJ} + \Sigma E_{EL} + \Sigma E_{HB}$$

where E_{LJ} is the Lennard-Jones potential energy, E_{EL} is the electrostatic energy and E_{HB} the hydrogen-bonding energy.[5,6] The overall interaction energy, E_{xyz}, was then assigned to that grid point. The calculation was repeated with the probe at each successive grid point, until an energy had been assigned to each point. The resulting 3D matrix of energy values was then used for computation as described later.

The description matrices

For each heptaplet, the nonbonded interaction energies with 31 different chemical groups (the first 31 probes in Table 1) were calculated using the GRID program as described ear-

Table 1. Table of probes. The columns from left to right indicate the probe identification number, symbol, name, charge, number of hydrogen bonds donated, number accepted and the optimal angle between hydrogen bonds when more than one hydrogen bond can be formed. Full probe details are defined in the GRID program.[5,6]

1	C3	methyl group	0.0			
2	N:#	sp N with lone pair	0.0	0	1	
3	N:	sp3 N with lone pair	0.0	0	1	
4	N1	neutral flat NH	−0.08	1	0	
5	N1=	sp2 amine NH cation	0.66	1	0	
6	NH=	sp2 NH with lone pair	0.0	1	1	120
7	N2	neutral flat NH2	0.0	2	0	120
8	N2=	sp2 amine NH2 cation	0.66	2	0	120
9	N3+	NH3+ amine cation	0.66	3	0	109
10	O1	sp3 hydroxy group	−0.1	1	2	109
11	O−	sp2 phenolate oxygen	−0.5	0	2	120
12	O::	sp2 carboxy oxygen atom	−0.45	0	2	120
13	OES	sp3 ester oxygen atom	−0.2			
14	OS	O of sulphone/sulphoxide	0.0	0	2	109
15	O=	O of sulphate/sulphonamide	−0.4	0	2	109
16	PO4	phosphate dianion	−2.0	0	8	
17	S1	neutral SH group	0.0	1	0	
18	C1=	sp2 CH aromatic or vinyl	0.0			
19	N:=	sp2 N with lone pair	0.0	0	1	
20	N-:	anionic tetrazole nitrogen	−0.33	0	1	
21	N1+	sp3 amine NH cation	0.66	1	0	
22	N1:	sp3 NH with lone pair	0.0	1	1	109
23	N1#	sp NH with one hydrogen	0.4	1	0	
24	N2+	sp3 amine NH2 cation	0.66	2	0	109
25	N2:	sp3 NH2 with lone pair	0.0	2	1	109
26	OH	sp2 hydroxy group	−0.15	1	1	120
27	O	sp2 carbonyl oxygen	−0.25	0	2	120
28	OC2	ether or furan oxygen	−0.2	0	2	109
29	ON	oxygen of nitro group	−0.1	0	2	120
30	PO4H	phosphate anion	−1.0	1	7	
31	COO—	multi-atom carboxy anion	−1.0	0	4	
32	CONH2	aliphatic amide group	0.0	2	2	
33	C(NH2)2	aliphatic amidine cation	+1.0	4	0	

lier. These were selected from the list of available probes in 1991 in order to study a wide range of common chemical groups. Then the grid points that were in or near to the minor groove were selected for study, since this was the region of interest.

Each GRID calculation yielded 28,749 variables (the interaction energies calculated at each grid point) for one probe and one triplet. Nine thousand five hundred ten of these grid points were in or near the minor groove. These 3D data values were rearranged as a one-dimensional (1D) vector of variables so that the interaction between the probe and the target molecule was described simply by this 1D vector. The procedure was repeated for all the targets to obtain a target matrix, as shown in Figure 1a, containing the energies of interaction between all 64 triplets and the one specific interacting chemical group (the probe).

The same procedure was then repeated using one single triplet and varying the type of probe. In this case, a different matrix was obtained (the probe matrix) containing different

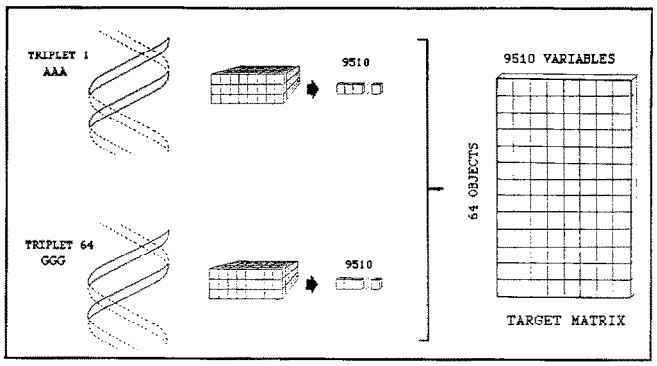


Figure 1a. A target matrix contains the interaction energies between one chemical group (the probe) and all 64 DNA triplet structures. The variables (9510 columns for 9510 grid points) are the interaction energies between the chosen probe and the triplet structures, while the objects (64 rows) are the triplet structures themselves.

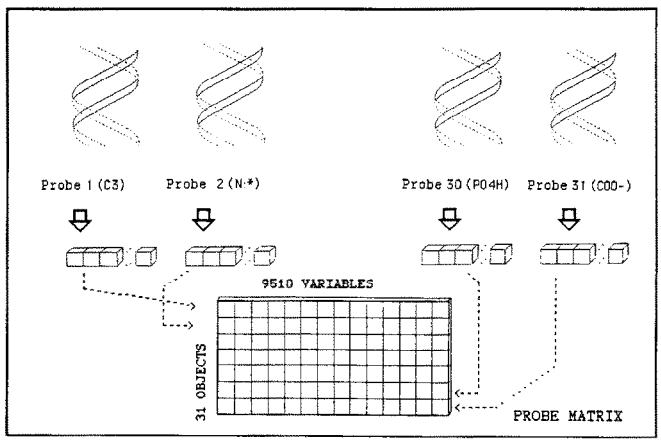


Figure 1b. A probe matrix contains the energies of interaction between the first 31 chemical groups (probes) in Table 1 and one double-stranded triplet. Each row contains the interaction energies between one probe and the chosen triplet, while the columns are related to the 9510 grid points as before.

description energies. As shown in Figure 1b, a probe matrix would contain the results for 31 different probes and one triplet.

The choice of target or probe matrices for detailed study depends on what one wants to know from the data and will be exemplified later in detail.

Statistical method

Principal component analysis[7] is a statistical multivariate technique for extracting and rationalizing the maximum common amount of information from a multivariate description of a biological system. The theoretical background of PCA has been presented in detail elsewhere.[7] Briefly, PCA can handle collinearities, thus permitting a matrix with more variables than objects to be studied. It does this by defining a smaller number of new variables, that retain much of the information in the observed data without the associated collinearities. PCA can also be used to build a model describing how a physical or chemical system behaves.

The principal component (PC) model depends on the transformation and scaling of the data matrix. A cut-off of +5 Kcal was used in order to make the data more symmetrically distributed about zero, because GRID energies can be very large and positive when the probe is close to the target. Then, each column variable was transformed by subtracting the column average. This, from a statistical point of view, corresponds to moving the multivariate system to the center of the data, which becomes the starting point of the mathematical analysis.[7] It should be noted that the autoscaling procedure[7] was not used, since all the variables were energies measured in the same units and autoscaling might introduce noise from variables with small variance, which were present in large numbers in these systems.

PCA results depend critically on the alignment criteria for the target molecules. However, the choice of the DNA heptaplets solved the alignment problem, because there was only one manner in which to perform the superimposition.

Counterion positions

Two additional target matrices were prepared and studied with a carboxy probe in order to assess the effect of varying the position of the potassium and the magnesium counterions. In one matrix, the cation-to-phosphate distance was increased by 35%, while in the other it was reduced by the same amount. The results showed that the alterations of counterion positions did somewhat affect the magnitude of the interaction energies between the probe and the DNA structures. However, the PC models computed from these two matrices showed indistinguishable behavior, demonstrating that the explicit location of the counterion does not have a significant influence on the mathematical model used to study the selectivity of the probe for the various triplet structures.

RESULTS AND DISCUSSION

PCA on the probe matrix

Initially, the binding of 31 different probes to two different targets (the TTGGGTT and TTAAATT double-stranded base pair sequences) was investigated. The aim of this study

was to describe the binding specificity of the 31 probes in order to make a simple and informative table as a guide to the design of specific DNA ligands. Nine thousand five hundred ten variables were obtained for each probe and each double-stranded DNA heptaplet, and the whole data array was represented by two probe matrices, each containing 31 objects (the probes) and 9510 variables (the interaction energies calculated at each grid point in the minor groove of one DNA triplet).

The two probe matrices were combined and two significant PCs were found from the combined matrix of 62 objects and 9510 variables according to the cross-validation technique as modified in the nonlinear iterative partial least squares (NIPALS) algorithm.[11] These components explained about 87% of the total variance in the combined matrix. The score plot for the PC model is shown in Figure 2. It suggests to a first approximation that the first PC distinguishes between the probes, while the second PC distinguishes between the targets. However, a more detailed interpretation of Figure 2 is required.

The first PC clearly shows that probes carrying a partial negative charge (such as number 16, phosphate dianion) are clustered on the left of Figure 2, while those with a partial positive charge (e.g., number 8, an NH2 cation) are on the right of the plot. In general, the cations (on the right of the plot) interact more favorably with DNA than the anions.

It may also be seen from the first PC, that the cations tend to interact more favorably with AAA than GGG, since the point representing the AAA interaction of each cation (i.e., its open diamond) is to the right of the GGG point (closed diamond) for the same probe. The vertical lines on the right of Figure 2 represent high-affinity cations. These lines slope forward, which demonstrates that cations, in general, have an even higher affinity for AAA than for GGG.

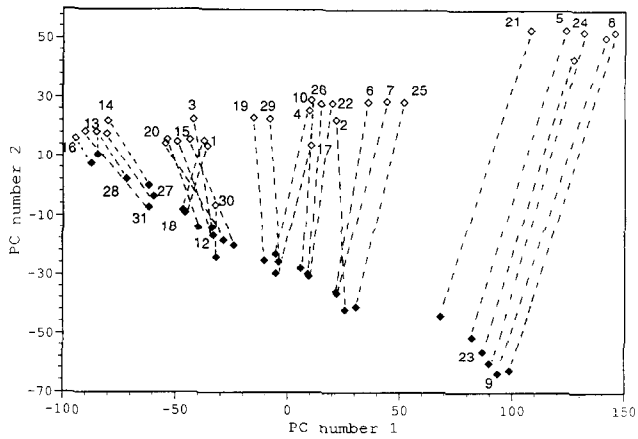


Figure 2. The first PC plotted against the second for the interactions between the first 31 probes of Table 1 with the AAA and GGG triplets. The filled points represent the interactions between the probes and the GGG triplet, while the open points represent the interactions with the AAA triplet. The dotted lines connect the same probe interacting with the two triplets and the slope of these lines is related to the selectivity of the probes. A negative slope indicates a GC*-selective probe, while a positive slope indicates* AT *specificity. For the identification of probe numbers and names, see Table 1.*

On the other hand, the first PC shows that the low-affinity anions (on the left of Figure 2) tend to have the opposite selectivity. Their open diamonds, representing AAA interaction, are to the left of the corresponding closed diamond for GGG. Their vertical lines slope backward, showing that anions tend to have a somewhat higher affinity for GGG than for AAA, although both affinities are low compared with the cations on the right of the plot.

This finding may account, to a considerable extent, for the general finding that ligands tend to bind more favorably to *AT*-rich sequences. Interest has been focused on high-affinity ligands that tend to have high-affinity groups. These groups tend to be cations and cations show an *AT* preference as demonstrated in Figure 2.

Turning now to the second PC, it may be seen that the AAA points (open diamonds) are invariably above the GGG points (closed diamonds) in Figure 2. Thus, as far as the second PC is concerned, AAA sequences are favored by all the probes. Moreover, this discrimination is greatest for the cationic probes, because their open and closed diamonds are always joined by a longer vertical line than the diamonds for anions.

In conclusion, both PCs demonstrate that the AAA sequence has the highest affinity for cations and that cations discriminate selectively against GGG in the minor groove. However, this general picture does not exclude the possibility that some regions of the minor groove of GGG can interact strongly with certain particular probes. This is in fact the case, as the detailed analysis in the following sections will show.

PCA on the target matrix

The target matrix with 64 objects and 9510 variables (Figure 1a) differs from the probe matrix for two reasons. First, it describes the interaction of only one probe with the target molecules and, second, it represents all 64 triplet targets instead of just the one triplet of a probe matrix. A similar target matrix can be obtained for each probe and can be analyzed by PCA. The probe matrix (Figure 2) actually demonstrates that the interaction of the GRID probes with DNA may be classified into several distinct groups. However, it is not appropriate to chose only one sample probe from each of the groups, because the probes behave in different ways giving different interactions with the GGG and AAA triplets (Figure 2). Accordingly, the more interesting target matrices will now be described in detail.

The multi-atom ionized carboxy probe, COO−

As shown in Table 2, the cross-validation technique[11] revealed three components, explaining about 69% of the total variance in the target matrix, which was obtained with the multi-atom ionized carboxy probe. The score plot for the first two components is shown in Figure 3. In this plot, each point represents one DNA triplet. When the points are near to each other, the corresponding triplets interact with the COO− probe in a similar way, but when the points are more separated, their interaction energies are different and the probe interacts in a more specific manner with each particular triplet.

In Figure 3, the *AT*-rich triplets are mostly clustered,

Table 2. PCA results for some data sets and the explained variance for each PC.

	PC1	PC2	PC3	PCtot
COO−	35.8%	24.2%	9.0%	69.0%
N3+	29.0%	28.1%	27.0%	83.1%
O=	29.2%	27.6%	19.9%	76.6%
O	30.5%	23.0%	21.1%	74.6%
OS	28.0%	25.0%	22.2%	75.2%
OH	26.2%	26.0%	25.1%	77.3%
O1	26.2%	26.0%	25.0%	77.2%
N:	28.1%	27.0%	24.0%	79.1%
N:=	27.7%	26.1%	24.5%	78.3%
CONH2	53.0%	13.9%	10.5%	77.4%

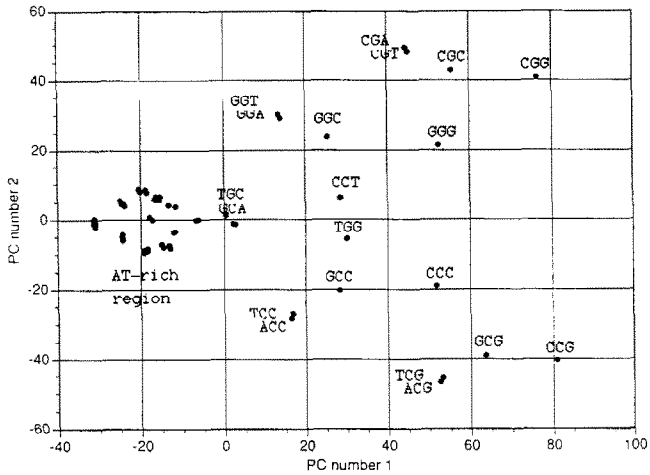


Figure 3. The first and second PCs for the interaction between the COO− probe with all 64 triplets. The diagram shows that this probe does not clearly distinguish between AT-*rich triplets, but can selectively recognize some* GC-*rich triplet structures. Note that the approximate symmetry of the second component is due to the symmetry of double-stranded DNA triplets.*

showing that a carboxy group does not differentiate them effectively from each other. However, the *GC*-rich triplets are well separated from the *AT* cluster and are usually separated from each other, showing that a ligand bearing a carboxy group could in principle be selective for *GC* as opposed to *AT* triplets and could also be selective between one *GC* triplet and another.

The PCs are related to the existence of different regions in the minor groove space. In order to evaluate the contributions of each of the regions in determining the triplet differences, we have constructed contour maps of the loading values for each PC (Figure 4). These maps illustrate the zones in which the interaction energies between the COO− probe and the triplets are most selective, favoring *GC*-rich affinity.

The region responsible for the separation in the first PC is located in the middle of the groove, opposite the central base pair of the triplets and near the amino group of a guanine. Some information about the interaction energies in this region is reported in the first column of Table 3. The best interaction energy is shown by the CGG (or CCG) triplet, but the GGC, CGC or GGG triplets can also interact very well with the carboxy probe. However, it must be emphasized that this region is not necessarily the place with the *best energy* of interaction between the probe and the triplets, but is the region where the probe can *best distinguish* between the triplets according to the score plot in Figure 3.

Table 3 shows that the first PC region is also influenced by the nature of the first and third base pair. The chemical

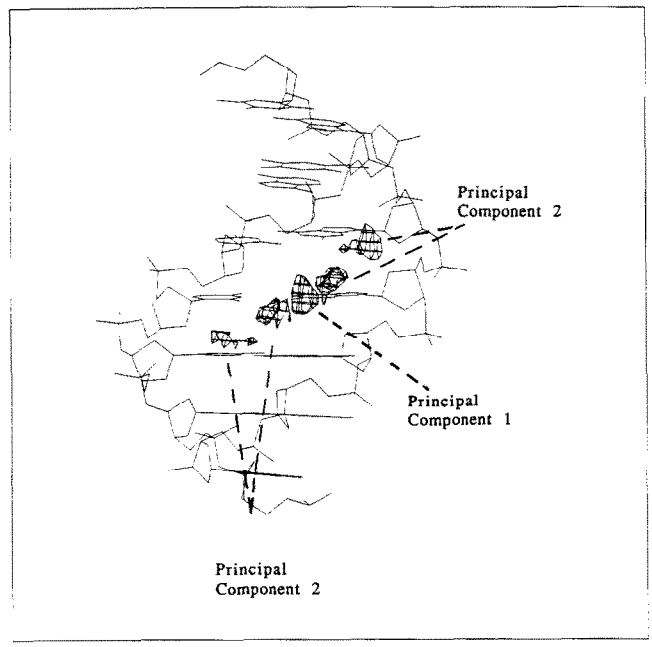


Figure 4. Contour maps (displayed using FRODO[12]) of the loading values for the first two PCs for the model of COO− interactions with all 64 triplets. The loadings show the position at which the carboxyl probe would interact most selectively in favor of the GGG triplet.

Table 3. Average interaction energies for selected targets in Kcal/mol in the regions indicated by the first and second components with the COO− probe (Figure 4). The values were calculated as the sum of the energies at each grid point included in the region, divided by the corresponding number of grid points.

Principal component number 1		Principal component number 2	
CGG	−6.2	CGX with X = T,A,C,G	−6.8
CCG	−6.2	XCG	−6.8
GGC	−5.7	GGX	−5.4
CGC	−5.6	XGG	−5.4
GGG	−5.5	CCX	−5.4
CGT	−5.3	XCC	−5.4
GGC	−3.9	GCX	−1.9
GGT	−3.2	XGC	−1.9
AAA	−0.4	AAA	−1.0

interpretation of this finding is that in this region the COO− probe cannot make any hydrogen bond with the N2 hydrogen of guanine, although *GC*-rich triplets do show favorable Lennard-Jones and electrostatic inteactions, which confer high affinity for the probe. However, the Lennard-Jones and electrostatic interactions are not so favorable with *AT*-rich triplets, which is why the first component of the carboxyl AAA interaction is only −0.4 Kcal/mol in Table 3.

The first PC also shows some small zones in which the COO− probe might interact near the C2 hydrogen of adenine. In these zones, which are not shown in Figure 4 for clarity, the probe had an average interaction energy of only −1.9 Kcal/mol with *AT* base pairs. This is not sufficient to select significantly for *AT*-rich triplets. However, the presence of the nearby *AT*-selective region demonstrates that the precise position of the interacting chemical group can be critically important for ligand selectivity.

The regions responsible for triplet separation by the second component are also shown in Figure 4. It may be noted that these regions are symmetrically distributed in relation to the symmetry of double-stranded DNA triplets. They are situated in different places, but exhibit the same chemical behavior. This is why the score plot in Figure 3 has an approximate symmetry axis more or less coincident with the x-axis of the plot.

The COO− probe makes a strong bond with the triplet CGX (or XCG) in these PC2 regions and is not appreciably influenced by the nature of the third X base. The corresponding energies of interaction are shown in column 2 of Table 3 and the strong interaction shown by GCX is due mainly to the geometry of the carboxy group. Graphical analysis (Figures 5–7) shows that the COO− probe in the region delineated by the second PC can interact with CGX triplets (Figure 5) by making a hydrogen bond with the guanine central base and another hydrogen bond with a guanine moiety in the adjacent base pair. This is made possible by the amino group positions of the CGX triplets isohelically located in the middle of the minor groove and therefore almost linearly aligned, as shown in Figure 5. Moreover, the interaction is further enhanced by favorable Lennard-Jones and electrostatic energies.

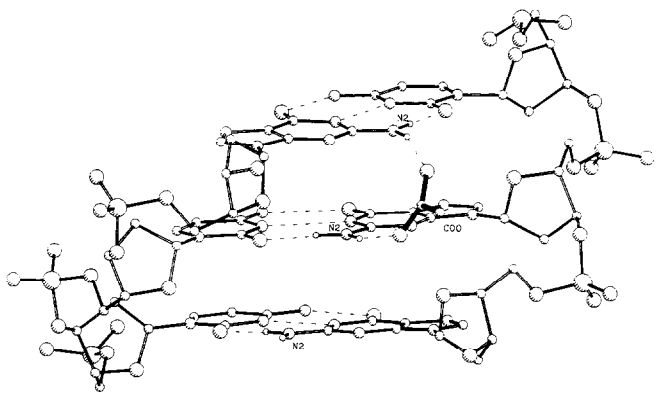


Figure 5. The minor groove of DNA (displayed using PLUTO[13]) showing the best predicted interaction between a CGG triplet and the COO− probe. The probe can make two strong hydrogen bonds, which are represented by dotted lines. Counterions are not shown.

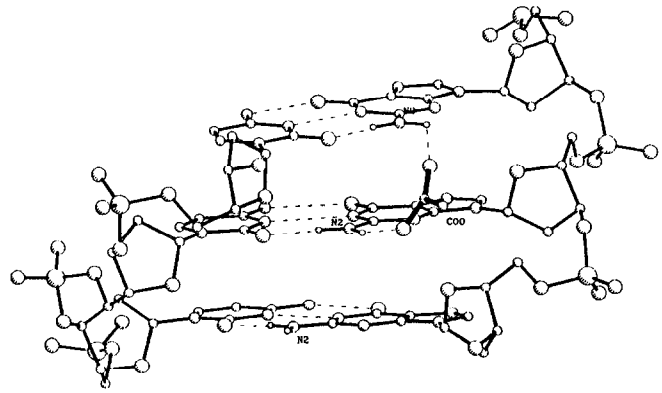


Figure 6. The predicted interaction between a COO− probe and the GGG triplet. The probe, positioned as defined by the second PC, is still able to make two hydrogen bonds with the N2 atoms of two adjacent guanine bases. However, these bonds are not as favorable as those shown in Figure 5.

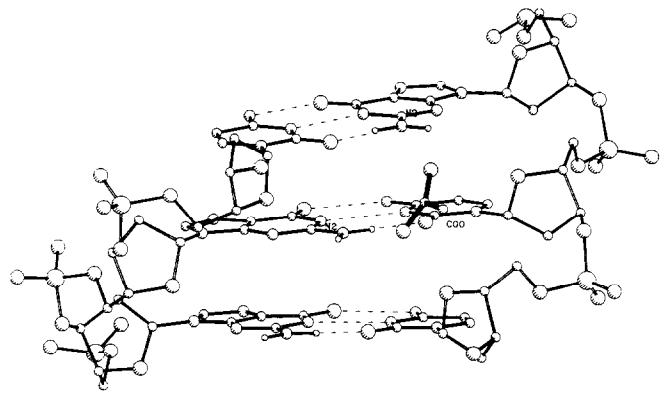


Figure 7. The predicted interaction between a COO− probe and the GCC triplet. This is a weak interaction because the position of the hydrogens does not permit significant hydrogen bonding to the probe.

The amino/carboxyl matching between target and probe is distorted with GGX triplets because the hydrogens of the amino groups are not so appropriately aligned (Figure 6). For example, the interaction energy of COO− with the GGC sequence in the second component region is less favorable by about 1.4 Kcal/mol than that with CGG. This was an unexpected result because the triplets look very similar, but the hydrogens of the amino groups in the GGC triplet are shifted with respect to those of the CGG sequence and this results in a weaker interaction with the probe. The interaction is even poorer with the GCX triplets, because the hydrogens of the amino groups are not well aligned (Figure 7) and are more widely separated than those of the CGX or GGX triplets, which is why the positive parts of the second component in Figure 3 differentiates these triplets into three distinct groups:

(1) CGX triplets with scores between 40 and 60
(2) GGX triplets with scores between 20 and 40
(3) the remaining triplets with scores near zero

The region identified by the third PC for the COO− probe, although statistically separate, is not chemically distinct, because it is composed of a mixture of the regions already discussed. However, it reinforces the preference of the COO− probe for *GC*-rich triplets (Figure 2). The overall analysis therefore suggests that the regions shown in Figure 4 may be appropriate places for the recognition of specific DNA base pair *GC* sequences by a ligand molecule bearing a suitably oriented carboxy group.

Amine cation probe, N3+

The results of the PCA for an N3+ probe are shown in Table 2 and Figure 8. This probe behaves toward DNA like most probes that carry a positive charge, such as those grouped on the right in Figure 2. However, Figure 8 shows that the triplets are grouped and the groups depend on the number of *AT* base pairs contained in a triplet. There is no significant relationship with the position of the *AT* base pairs inside the triplets themselves and, for this reason, the model may not distinguish well between particular *AT* triplets, although triplets that are rich in *AT* base pairs are well discriminated from those that are *GC* rich.

In this case, for each base pair there is one PC region located deep in the center of the DNA minor groove where a direct interaction occurs between the cationic probe and both the adenine and thymine bases. In this region, the N3+ probe strongly prefers to interact with *AT* base pairs, having an attractive interaction up to −8 Kcal/mol with *AT*, while it tends to be repelled by *GC*. This difference is mainly due to its interaction with the lone pairs of the O2 atom of thymine and with the lone pair of the N3 atom of adenine. It also seems that the probe can sometimes interact making a third hydrogen bond with the O2 or N3 atom of adjacent *AT* base pairs (see Figure 9). With *GC* containing triplets on the contrary, the presence of the amino group of guanine in

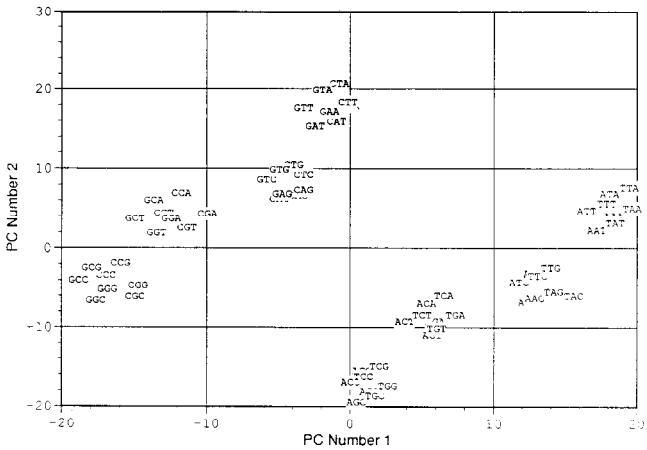


Figure 8. Score plot for the model describing the interactions of the N3+ probe. This probe does not clearly distinguish between triplets, but can differentiate between certain base pairs. The first PC separates TXX and AXX from GXX and CXX, while the second PC weakly distinguishes XTX and XAX from XGX and XCX. The third base pair of the triplet is also weakly differentiated by the third PC for this probe.

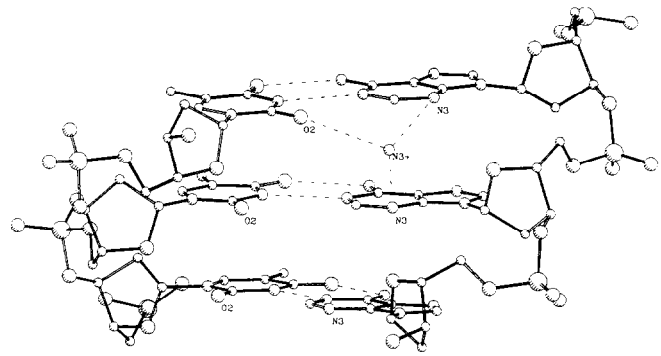


Figure 9. The predicted interaction between an N3+ probe and the AAA triplet. The probe is able to make three hydrogen bonds in PC region 1 with the atoms of the first two base pairs. A similar interaction could be made between the second and third pairs of the triplet, but this is not shown.

these regions tends to cause a steric and electrostatic repulsion of the N3+ probe, which is incapable of making a favorable hydrogen bond with the *GC* bases. These only have an average interaction energy of −0.9 Kcal/mole (Table 4) with N3+ compared with some −8.0 Kcal/mole for *AT*-containing triplets.

There is also a small zone in these regions in which the N3+ probe can interact making only one hydrogen bond each time with either the carbonyl or nitrogen of the base pairs. However, these places are not significant for specificity of binding.

In conclusion, it may be seen that this probe does not show triplet-selective specificity, but only a localized, high affinity to single *AT* base pairs or adjacent pairs of *AT* base pairs. The N3+ probe does not interact favorably with *GC* base pairs and, moreover, it does not interact significantly with the backbone phosphate groups. This initially surprising finding is mainly due to the treatment of the aqueous environment by the GRID force field and is not in disagreement with other recent computations[14] nor with the crystal structure of a synthetic dodecamer DNA interacting with a positively charged ligand.[15]

Sulphate/sulphonamide probe, O=

This probe represents an oxygen atom bonded to sulphur as in a sulphate or sulphonamide group. Such an atom can accept up to two hydrogen bonds making a tetrahedral angle with each other. The PCA reveals some regions of noncovalent interaction at which this probe can play an important role in triplet differentiation. Three such PC regions were identified: the first located near the first and second base pairs (Figure 10), the second near the second and third base pairs, and the third distributed more deeply in the minor groove. The second and third regions are not shown in Figure 10 for clarity.

The first PC region is divided in two parts, the bigger of which lies on a base pair plane, while the smaller is fragmented and is shifted toward the adjoining base pair (Figure 10). Both parts show *GC* preference, but the probe is able to make only one strong hydrogen bond with the N2 atom of guanine in the larger part of the region, while it can make

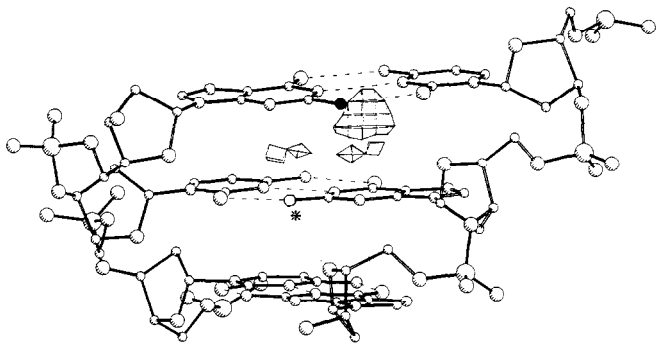


Figure 10. View of the region defined by the first PC for the interaction between an O= probe and the CGG triplet. In the larger part of this region (which lies in the base pair plane), the probe is able to make one strong hydrogen bond with the N2 atom of guanine (●), while in the smaller parts it can make one hydrogen bond to this nitrogen and another to the adjacent starred atom (*). A second region symmetrical to the first is not shown and neither is a third for clarity.

two bonds with the guanines of adjacent *GC* base pairs in the smaller part. However, the probe does not make good interactions in this second position and the energy (−2.7 Kcal/mol) from the two hydrogen bonds is less favorable than that from a single bond in the bigger part of the first PC region (−3.2 Kcal/mol). On the other hand, the smaller part seems to be more selective for *GC*-rich triplets (Tables 4 and 5).

There is another important difference between the two parts of the first PC region in Figure 10. In the larger part, the interaction energies are less dependent on the precise position of the probe, while the energy gradients in the smaller part are steeper. In other words, a slight alteration of the location of the probe in the smaller part leads to greater variations in the energy values. This suggests that the exact position of the probe might be critical in the smaller part of the first region.

The second PC region shows similar behavior, but in this case the parts of the triplets involved are those around the second and third base pairs, as required by the symmetry of the system. The properties of the third region are mainly due to steric hindrance between the probe and the N2 group of guanine. This third region does not influence the selectivity of binding.

Table 4. Average interaction energies in Kcal/mol (sum of the energies at each grid point included in the region, divided by the number of the points) of some different probes with regions identified by the PC models. The energy differences between the triplets provide a measure of the selectivities of the probes.

Probe	Triplet	Average energy	In PC regions
N3+	TTT	−8.0	1 or 2 or 3
N3+	TTG	−7.9	1 or 2
N3+	TCC	−7.6	1
N3+	CCC	−0.9	1 or 2 or 3
O=	CGG	−5.6	1
O=	GGG	−5.2	1
O=	CCA	−5.2	1
O=	GGT	−5.2	1
O=	GCG	−5.0	1
O=	TGA	−2.9	1
O=	AAA	−2.3	1
O	CGA	−4.2	1
O	CGT	−4.2	1
O	GGG	−3.8	1 or 2
O	GCA	−3.5	1
O	AAA	−1.2	1 or 2
OS	CGG	−3.4	1
OS	AAA	−0.7	1
O1/OH	GGG	−6.0	1 or 2 or 3
O1/OH	AAA	−4.0	1 or 2 or 3
N:=/N:	GGG	−4.1	1 or 2 or 3
N:=/N:	TTT	−1.9	1 or 2 or 3

Table 5. Details of the individual energy contributions (Kcal/mol) to the total interaction energy between an O= probe and two triplets in the position shown by the smaller part of the first PC.

		CGG TRIPLET			
Atoms	Base	Distance	EQ	ELJ	EHB
HN2	G4	2.35	—	—	−1.85
HN2	G5	2.73	—	—	−0.92
N2	G4	3.17	−0.61	−0.44	—
N3	G5	3.25	0.58	−0.38	—
O4*	A6	3.41	1.31	−0.22	—
N2	G5	3.45	−0.52	−0.45	—
N3	G4	3.65	0.48	−0.27	—
O4*	G5	3.66	1.16	−0.17	—
C2	G5	3.74	−0.21	−0.33	—
C2	G4	3.79	−0.20	−0.31	—

		AAA TRIPLET			
Atom	Base	Distance	EQ	ELJ	EHB
O2	T5	3.09	0.86	−0.29	—
O4*	A6	3.41	0.62	−0.22	—
O4*	A5	3.66	0.54	−0.17	—
N3	A4	3.68	0.24	−0.26	—
C4*	A6	3.82	−0.20	−0.23	—
C5*	A6	3.89	−0.43	−0.26	—
C2	A4	4.16	−0.26	−0.29	—
C2	T5	4.23	−0.30	−0.18	—
C4*	A5	4.23	−0.16	−0.15	—
O4*	T5	4.39	0.39	−0.08	—

Carbonyl oxygen probe, O

Three regions were extracted from the matrix obtained by GRID calculations with this probe, explaining 74.6% of the variance (Table 2). A numerical analysis of the selectivity regions (Table 4) shows that the probe interacts well with the *GC*-rich triplets, while it interacts rather weakly with *AT*-rich triplets.

Two PC regions shown in Figure 11 are responsible for the selectivity of binding, while a third is distributed more deeply in the minor groove space and is not important for selectivity. The first region has a wide boomerang shape in which the probe can interact, making one or two hydrogen bonds with two N2 atoms. In the second, smaller region, the probe makes only one hydrogen bond (Figure 11). This means that when the carbonyl probe interacts with CGX triplets, it can be positioned in any part of the wide region, since the interaction energies with the CGX triplets are very similar throughout.

The magnitude of the interaction energies and the size of the first region suggest that the carbonyl oxygen probe could be used to obtain significant selectivity for adjacent pairs of CGX base pairs. However, when the probe interacts in the same place with different *GC*-rich sequences such as GGX, it can make only one hydrogen bond at a time and the exact position of the probe becomes more critical. It must then be located at the top or bottom of the first PC region opposite the N2 atom of a guanine and, so, the interaction is less favorable than with CGX triplets.

The calculated positions of the carbonyl probe interactions with CGX sequences are compatible with experimental results showing the interactions of quinoxaline antibiotics with DNA.[16] This suggests that the carbonyl probe could be used as a CG recognition feature if properly positioned and aligned in the minor groove of DNA.

Sulphone/sulphoxide probe, OS

A GRID computation over the 64 triplet structures shows that the global interactions of this probe are similar with all triplets, giving overall interaction energies of about −4.0 Kcal/mol. However, three significant PCA components were obtained, accounting altogether for 75.2% of the data variance, and related to three distinct regions in which the probe is able to distinguish between selective molecular moieties.

As with the O= probe, the first region is located in the middle of the groove and extends from the first toward the second base pair. However, the region for an OS probe is a little smaller and nearer to the axis of the triplets than the corresponding O= region. Moreover, the magnitude of the interaction energies naturally differs according to the chemical characteristics of each individual probe.

As with the O= probe, there is a small zone in this first PC region between two adjacent base pairs in which the sulphone/sulphoxide OS probe can make two hydrogen bonds with *GC*-rich triplets, showing a preference for these triplets of about −2.7 Kcal/mol in comparison with AAA (Table 4). An analysis of the interaction energies indicates that the probe can make one strong hydrogen bond with one guanine and another weaker bond with an adjacent guanine base. However, the weaker second bond and the small size of this zone might prevent its effective exploitation for the design of triplet-selective ligands. The findings therefore suggest that the OS probe could be used only in order to obtain base pair differentiation.

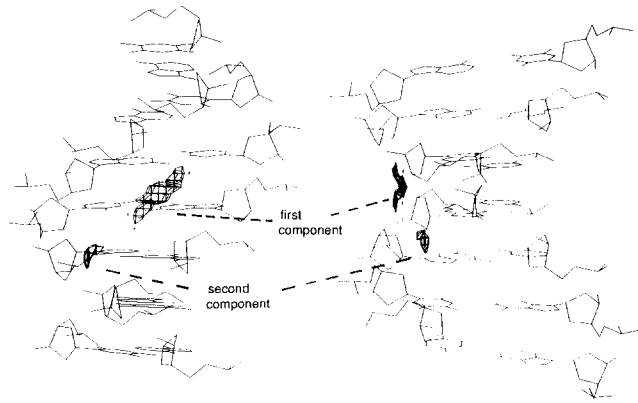


Figure 11. Two orthogonal views of the selectivity regions for the carbonyl oxygen probe. The wide boomerang-shaped region is composed of three parts, two of which are opposite two N2 atoms of adjacent guanine bases, while the third is between these two bases. With the CGX triplets, the probe makes two hydrogen bonds in this last part of the wide region, while it interacts with other triplets in the remaining parts.

The sp3 and sp2 hydroxy probes, O1 and OH

These probes represent a hydroxyl group bonded to an aliphatic (O1) or aromatic system (OH). Both groups can donate one hydrogen bond, but the O1 probe can accept two while the OH probe only accepts one. Moreover, the hydrogen bond pattern is different, principally because the electronic configuration of the OH probe interacts with the π system of the aromatic ring.

Two separate PC models were computed for the sp3 and sp2 hydroxy groups, but from Tables 2 and 4 it can be seen that the models are similar. In each case, the PCA extracts three components with the same amount of information and percentage of explained variance. For each probe, these PCs are related to the existence of three narrow regions located opposite the N2 atom of each guanine in a triplet. Although the probes show a good energy of binding with *GC* base pairs in these regions (Table 4), they also show a fair binding energy with *AT* pairs and the preference for *GC* versus *AT* is estimated as only −2 Kcal/mol. Furthermore, one of the two lone pairs of the sp3 hydroxy group cannot interact favorably with the N2 hydrogen of guanine and, so, this probe cannot accept two hydrogen bonds from two adjacent base pairs. It can only accept one hydrogen bond and does not differ significantly from the sp2 hydroxyl in this system.

One must conclude that the small differences in binding energies between the bases and the narrowness of the regions responsible for selectivity suggest that these probes could be used only with difficulty (or not at all) to obtain effective base pair or triplet differentiation.

Comparison between oxygen probes

Each GRID probe and each target atom is described by nine different properties that define the detailed geometry and energy of its interactions. It is these properties that confer specific characteristics on the various GRID probes.

In DNA systems, the polarizability and the hydrogen bond patterns play an important role in determining the predicted interactions. For example, the sulphate/sulphonamide probe (O=) has a high polarizability and interacts selectively with *GC* base pairs over a wide region. Its hydrogen bond pattern (Table 1 and Figure 10) is superimposed on this region and thus constrains the most favorable interaction to lie in the base pair plane.

The carbonyl oxygen probe (O) has a smaller polarizability than the sulphate/sulphonamide O= probe and has correspondingly smaller selectivity regions. On the other hand, its hydrogen bond interactions are stronger and its hydrogen bond pattern allows the carbonyl probe to bond in the middle of the base pair planes. For this reason, the probe is selective for adjacent pairs of base pairs.

The same chemical properties can also explain the behavior of the sulphone/sulphonamide probe (OS). This probe shows a hydrogen bond pattern similar to the sulphate/sulphonamide O= probe (Table 1), but it makes stronger hydrogen bonds and is much less polarizable. Therefore, it shows selectivity regions in the same places, but with a smaller size than the corresponding sulphate/sulphonamide regions.

The phenolic hydroxyl (OH) and aliphatic hydroxyl (O1) probes show an intermediate polarizability value between those of sulphate/sulphonamide O= and those of the sulphone/sulphoxide OS, and have quite a powerful hydrogen bond strength parameter that may account for the narrowness of their selectivity regions.

It therefore appears that the chemical properties of the target atoms and ligand groups can account for the different interaction selectivities predicted for oxygen groups that are, at first sight, rather similar. However, the particular selectivities of each probe will vary according to the characteristics of the target system. It would be unwise to generalize from the present DNA findings to protein or polysaccharide or lipid systems that might well interact differently.

NITROGEN PROBES WITH A LONE PAIR

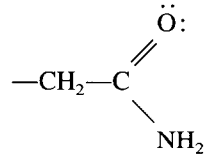
The N: and N:= probes (numbers 3 and 19) represent a nitrogen atom with one lone pair that can accept one hydrogen bond. N: is an aliphatic sp3, and N:= is an aromatic sp2 nitrogen. Neither of these probes can donate hydrogen bonds. They show similar behavior, interacting with *GC* base pairs as shown by the first PC in Figure 2.

GC affinity is due to the formation of one hydrogen bond, which accounts for an extra −2.5 Kcal/mol compared with the *AT* interaction. The selective region is slightly larger for the sp2 N:= probe, perhaps because it is somewhat more polarizable than sp3 N:. This suggests that an aromatic nitrogen that accepts hydrogen bonds might be a useful *GC*-selective group. However, it would not be triplet selective.

Experimental results showing the interactions of pyridine-2-carboxamide with DNA[17] are in agreement with this finding.

Aliphatic amide multi-atom probe

The aliphatic amide probe is planar, neutral and able to accept two hydrogen bonds at the carbonyl oxygen atom and donate two hydrogen bonds from the NH_2 group. In this probe, the amide carbon atom is explicitly included and a methylene group bonded to the sp2 amide carbon represents the start of an aliphatic side chain.



The probe's ability to accept and donate hydrogen bonds, at the same time, allows it to define a high number of interaction sites with all of the 64 DNA triplets. GRID shows that the probe can interact in the minor groove, making hydrogen bonds with the O2 oxygen of thymine and cytosine, N3 nitrogen of guanine and adenine, O4* oxygen of deoxyribose rings, OP oxygen of phosphate groups and N2 nitrogen of guanine. Moreover, many combinations of two, three or four hydrogen bonds are possible at these sites, according to the geometric and energetic characteristics of the triplets and the probe.

The aliphatic amide probe can therefore make multiple hydrogen bonds with each triplet, giving high energies of interaction. This interaction flexibility makes it a high-affinity ligand for all the DNA triplets. The PCA (Table 2) shows three different regions of selectivity in the minor groove space. However, the third region, although statistically and chemically significant, is very small and for this reason will be not discussed further. In fact, the size of a PC region is always an important factor to be considered, because it is not easy to exploit small regions for selective ligand design.

The score plot in Figure 12 shows that the triplets are clustered in two large groups and five subgroups by this

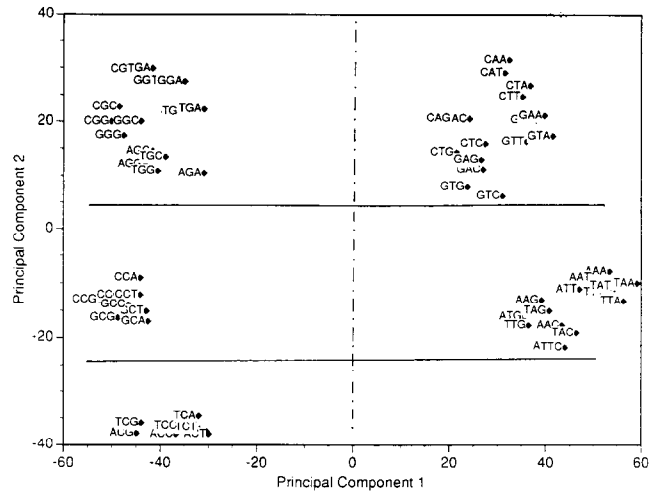


Figure 12. Score plot for the model describing the interaction of the CONH2 probe. The first PC can distinguish two main groups of triplets (indicated by dotted lines), while the second PC distinguishes three groups (indicated by continuous lines).

probe. The first PC is able to distinguish the two main groups, as indicated by the vertical dotted line in Figure 12. These groups correspond to the two regions shown in Figure 13. The probe can make multiple hydrogen bonds to triplets from the bigger region of Figure 13 (right of score plot), but can make fewer interactions from the smaller region (left of score plot). For example, the amide probe interacts with GTG (right of score plot) in the larger region, but with GCG (left of score plot) in the smaller region. As mentioned earlier, this smaller region is less favorable for ligand design simply because it is small. Moreover, the two regions are not spatially distant from each other and so it might not be easy to design a ligand that carries the amide group to the precisely required place in the precisely required orientation.

The situation is quite different with the second PC, which separates the three subgroups of triplets and is indicated by the continuous lines drawn across Figure 12. In this case, the selectivity regions are well separated from each other (as shown in Figure 14) and might well be used for ligand design. Interaction details are given in Table 6.

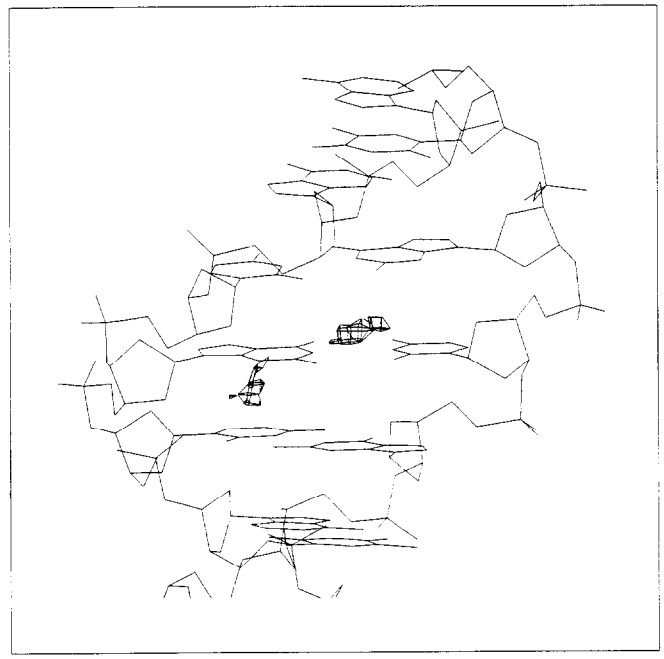


Figure 14. The minor groove of DNA showing the selectivity regions defined by the second PC of the amide-triplet interaction model.

Aliphatic amidine multi-atom probe

This probe carries a positive charge and is able to donate up to four hydrogen bonds. Like the amide probe, the amidine probe can interact with all the DNA triplets, making multiple

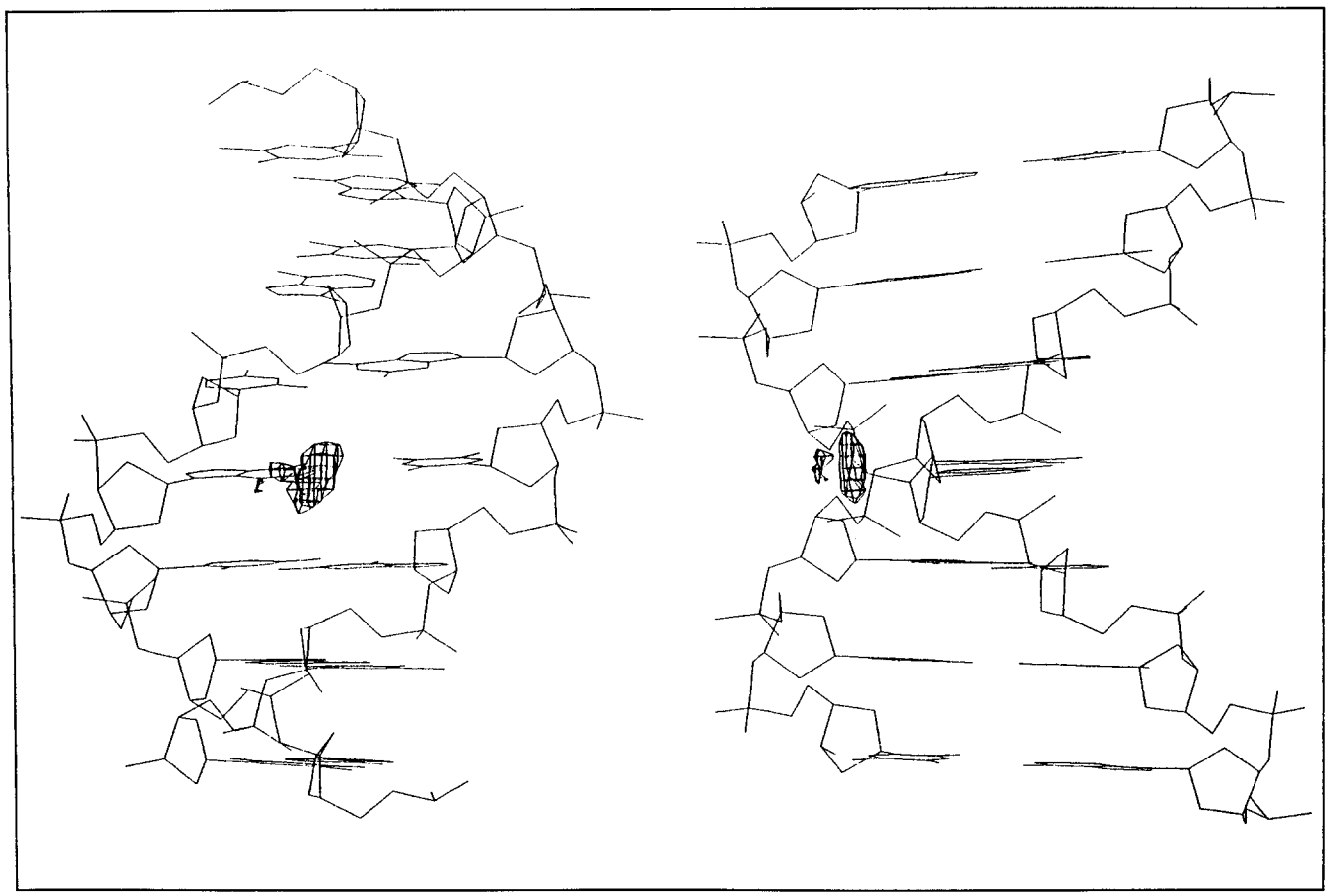


Figure 13. Two orthogonal views of the minor groove of DNA showing the selectivity regions defined by the first PC of the CONH2 interaction model.

Table 6a. Summary of the selectivity of certain probes that could be used as *GC* or *AT* base pair recognition features if properly aligned in the minor groove of DNA.

Probe name	Probe symbol	Specific for base pair	Interacting with	
amine cation	N3+	A..T	O2T	N3A
sulphate/sulphonamide	O=	G..C	N2G	
carbonyl	O	G..C	N2G	
sulphone/sulphoxide	OS	G..C	N2G	
sp2 nitrogen	N:=	G..C	N2G	

Table 6b. Probes showing selectivity for two adjacent pairs of base pairs.

Probe name	Probe symbol	Specific for adjacent pairs of base pairs		Interacting with	
sulphate/sulphonamide	O=	C..G G..C		N2G	N2G
carbonyl	O	C..G G..C	G..C G..C	N2G	N2G
carboxy	COO−	C..G G..C	G..C G..C	N2G	N2G

Table 6c. Probes showing selectivity for triplets.

Probe name	Probe symbol	Specific for triplets		Interacting with		
carboxy	COO−	C...G G...C G...C	C...G C...G G...C	Elect. and L.J. interactions		
carboxy	COO−	C...G G...C C...G	G...C C...G G...C	Elect. and L.J. interactions		
amide	CONH2	G...C* C...G G...C		N2G	N2G	O4*T
amide	CONH2	G...C* G...C G...C		N2G	O2C	
amide	CONH2	A...T* C...G A...T		N2G	N3A	O4*T

*cluster represented by this triplet (Figure 12)

hydrogen bonds at different sites in the minor groove space as well as bonds to the N3 nitrogen of guanine and adenine, O2 oxygen of thymine and cytosine, O4* oxygen of deoxyribose rings and the OP oxygen of the backbone phosphate groups. Moreover, the net positive charge of the probe makes it able to interact in a particularly strong way with these DNA bases.

However, the amidine probe does not show any exploitable preferences for particular base sequences because it can bind so well to all 64 DNA triplets. It was possible to find selectivity regions in the minor groove, but these regions were so close together that it would be virtually impossible to design a ligand bearing an amidine group in precisely the right place. Thus, once again, a probe with relatively high affinity results in no practical selectivity at all.

CONFORMATIONAL EFFECTS

The previous results were obtained with all 64 heptaplets superimposed in the B-conformation of DNA. In order to assess the influence of altering this conformation, four of the original triplet sequences (TTT, GCG, CGA, CGC) were also studied in conformations that have been reported by X-ray crystallography.[18,19] Each of these modified targets was embedded as usual between standard *AT* end pairs in order to make a heptaplet. These modified heptaplets were aligned with the main group of modeled target molecules. The calculations with the carboxy probe were then repeated with this extended set of 68 targets giving the results shown in Figure 15.

Simple inspection shows that Figures 3 and 15 are not dissimilar, and that the well-defined separation between *AT*-rich and *GC*-rich triplets is clearly conserved. AAA triplet in its observed conformation lies within the *AT*-rich region as before and the observed *GC*-rich triplets are still well distributed throughout the *GC* region. Their rankings in the second component are scarcely altered, because this component mainly represents hydrogen-bonding interactions between the carboxy probe and the targets. These hydrogen bonds can still be made to the *GC*-rich triplets in their experimentally observed conformations. On the other hand, the relative positions of the observed triplets in the first PC, which represents van der Waals and electrostatic interactions, are appreciably altered. The van der Waals interactions, in particular, are very sensitive to local alterations of the target structure, as would be expected.

These preliminary findings suggest that the present approach may be of value for interpreting the effects of conformational change, as well as for studying targets in one conserved conformation. For example, we had not originally anticipated that hydrogen-bonding interactions would be conformationally insensitive, although subsequent modeling studies show that a carboxy group can in fact make flexible hydrogen bonds to a range of *GC* sequences.

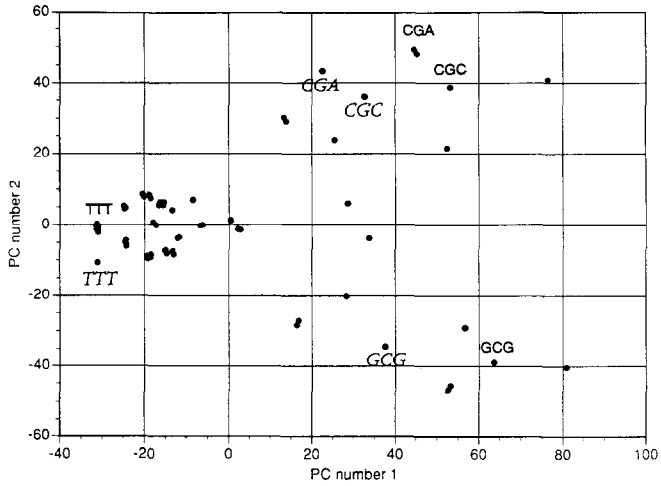


Figure 15. The first and second PCs for the interactions between the COO— probe and the extended set of 68 triplets. The four modified triplets are indicated in italics. (Compare with Figure 3.)

CONCLUSION

An important limitation of the present study is that flexibility of the DNA has not been taken into account for all the triplets. It would, of course, have been feasible to extend the investigation to include several conformations of each heptaplet, but this would have significantly increased the computational load and might not have yielded appreciably more reliable results (see conformational effects in the previous paragraph). Moreover, the experimental indications[20] are that the perturbation of the DNA structure due to the binding of common ligands is rather small and may not be particularly important for minor groove, noncovalently binding ligands.

Another limitation is that comparatively little experimental evidence is available to us, against which to check the present predictions. In some cases (e.g., the amine cation probe, the carbonyl oxygen probe or the sp2 nitrogen probe), the reported observations and computations are in good agreement. However, much more experimental work is needed. The present findings suggest a number of new approaches to the design of selective, high-affinity DNA ligands.

The results demonstrate that affinity and selectivity in the minor groove must be regarded as different concepts. Affinity can be defined by the binding constant for association with a particular DNA triplet, while selectivity may be measured by the ratio of two affinity constants for two different sequences. However, such a ratio is independent of the magnitude of the two constants. Selectivity can arise because one sequence binds a probe particularly strongly or because the other sequence hardly accepts it at all. The first of these two types is naturally to be preferred, but the second type of interaction could be used to introduce selectivity into a ligand that already binds strongly due to nonselective interactions.

Selectivity in the minor groove must be further subdivided into base pair selectivity and triplet selectivity, and there is an intermediate class in which a probe such as carbonyl oxygen selects for two adjoining base pairs. Moreover, selectivity depends on the type of probe, on the position and orientation of its interaction and on the magnitude of the selectivity regions. A large selectivity region with a small energy gradient is preferable to a smaller region, because the selectivity in the large region is less dependent on the precise position of the probe.

It should be noted that the chemical interactions between probe and target are not necessarily the same throughout one PC region. This is clearly demonstrated in Figure 10, which represents a single region with one large and one fragmented subregion. In the larger part, there is one hydrogen bond; in the smaller, there are two hydrogen bonds to two adjacent base pairs of the target. The statistics are unable to distinguish between these parts, because the computations are based on overall interaction energies without explicit mechanistic information. Clearly, it is essential to make a chemical evaluation of the statistical findings in order to exploit them for effective ligand design.

The probes that show high affinity for the minor groove, irrespective of sequence, are the cationic probes. This high overall affinity is due to the charge of the probe and to the

number of donating hydrogen bonds exhibited by the cationic probes.

The probes that show significant selectivity for individual base pairs are reported in Table 6a. Those that show significant selectivity for two adjoining base pairs are reported in Table 6b and those that show triplet selectivity are in Table 6c. These probes were chosen because their selectivity preferences are displayed over a sufficiently large region, so that their precise location in the minor groove should not be particularly critical for effective ligand design.

The results show that ligands of weak affinity may be selective and the existence of selectivity regions is closely related to the pattern of hydrogen bonds made by the probes. Anionic probes accepting two or more hydrogen bonds usually bind selectively to *GC* sequences. We were very surprised to find situations in which the selectivity changes dramatically from one position to a closely adjoining place. Such situations need to be identified at an early stage of ligand design, since they would almost certainly lead to capricious interactions.

The results also show that probes that can make one hydrogen bond prefer to interact in the base pair plane and can select only for a single base pair. On the other hand, probes that can make two or more hydrogen bonds often tend to interact in the middle of the planes generated by the base pairs and, in principle, can distinguish between triplets. Therefore, multiple hydrogen bonds are necessary to obtain triplet specificity. Because many GRID probes are small, they often exhibit only base pair or adjacent base pair selectivity. The new multi-atom GRID probes (e.g., carboxy, amide, amidine, etc.) are particularly informative, because they can make several hydrogen bonds over a wide region.

Two significant general problems with the PC and comparative molecular field analysis (COMFA) methods as practiced today are the alignment of the target molecules and the treatment of their flexibility. As the present work demonstrates, the unequivocal overall alignment of related biological macromolecules may be achieved in certain favorable cases. However, the mobility and flexibility of the targets are a serious problem that still needs to be addressed in PC and COMFA studies.

Finally, it should be pointed out that the present method is not restricted to DNA-ligand interactions. The DNA heptaplets were chosen for the present work, because it was easy to superimpose them unambiguously. The same approach could be applied to the major groove in order to assess specificity criteria for DNA-repressor interactions. Furthermore, it might be relatively easy to align related protein molecules such as mammalian renins and HIV proteases in order to apply the same approach to the design of selective protein ligands. This would be easiest when the proteins were very similar (i.e., when selective ligand design was most difficult).

ACKNOWLEDGMENT

G. Cruciani acknowledges financial support from the Consiglio Nazionale delle Richerche (CNR Italy).

REFERENCES

1 Zimmer, C. and Wahnert U. *Prog. Biophys. Molec. Biol.* 1986, **47,** 31
2 Mattes, W.B., et al. *Carcinogenesis* 1988, **9,** 2065
3 Hartley, J.A., et al. *Acta Oncologica* 1988, **27,** 503
4 Thurston, D.E. and Thompson, A.S. *Chemistry in Britain* 1990, **26,** 767
5 Goodford, P.J. *J. Med. Chem.* 1985, **28,** 849
6 Bobbyer, D.N.A., Goodford, P.J., McWhinnie, P.M. and Wade, R.C. *J. Med. Chem.* 1989, **32,** 1083
7 Wold, S., Esbensen, K. and Geladi, P. *Chem. & Intell. Lab. System* 1987, **2,** 37
8 Quanta Software. Molecular Simulations, Waltham, Massachusetts, USA,
9 Zakrzewska, K., Randrianarivelo, M. and Pullman, B. *J. Biomol. Struct. & Dyn.* 1988, **6,** 331
10 Zakrzewska, K. and Pullman, B. *J. Biomol. Struct. & Dyn.* 1988, **5,** 1043
11 Wold, S. *Technometrics* 1978, **20,** 397
12 Jones, A.T. *Methods Enzymol.* 1985, **115,** 157
13 Pluto Software. University Chemical Laboratory, Cambridge, England,
14 Zakrzewska, K. and Lavery, R. In *Computer-aided Molecular Design* (G. Richards, Ed.) Oxford University, Oxford (1989) 129
15 Kopka, M.L., Yoon, C., Goodsell, D., Pjura, P. and Dickerson, R.E. *J. Mol. Biol.* 1985, **183,** 553
16 Wang, J., Ughetto, G., Quigley, J.G. and Rich, A. *J. Biomol. Struct. & Dyn.* 1986, **4,** 319
17 Wade, W.S. and Dervan, P.B. *J. Am. Chem. Soc.* 1987, **109,** 1574
18 Dickerson, R.E. and Drew, H.R. *J. Mol. Biol.* 1981, **149,** 761
19 Gabriele, A.D., Sanderson, M.R. and Steitz, T.A. *Proc. Nat. Acad. Sci. USA* 1989, **86,** 1816
20 Zakrzewska, K., Lavery, R. and Pullman, B. *Nucleic Acid Research* 1983, **11,** 8825