Available online at www.sciencedirect.com
SCIENCE DIRECT®
European Journal of Medicinal Chemistry 38 (2003) 141–155
EUROPEAN JOURNAL OF MEDICINAL CHEMISTRY
www.elsevier.com/locate/ejmech

Original article

QSAR and molecular modelling studies on B-DNA recognition of minor groove binders

André Mauricio de Oliveira, Flávia Beatriz Custódio, Cláudio Luis Donnici, Carlos Alberto Montanari *

Núcleo de Estudos em Química Medicinal (NEQUIM), Departamento de Química, Universidade Federal de Minas Gerais, Campus da Pampulha Av. Pres. Antonio Carlos 6627, Belo Horizonte, MG 31270-901, Brazil

Received 27 March 2002; received in revised form 9 September 2002; accepted 26 September 2002

Abstract

Aromatic bisamidines have been proved to be efficient compounds against *Leishmania* spp. and *Pneumocystis carinii*. Although the mode of action is still not known, these molecules are supposed to be DNA minor groove binders (MGBs). This paper describes a molecular modelling study for a set of MGBs in order to rank them through their complementarity to the Dickerson Drew Dodecamer (DDD) according to their interaction energies with B-DNA. A comparative molecular field analysis (CoMFA) has shown the importance of relatively bulky positively charged groups attached to the MGB aromatic rings, and small and negatively charged substituents into the middle chain. Models were obtained for DNA denaturation related to H-bonding processes of binding modes. Validation of the model demonstrated the robustness of CoMFA in terms of independent test set of similar MGBs. GRID results allotted bioisosteric substitution of –O– by –NH– in furan ring of furamidine and related compounds as being capable to enhance the binding to DDD.

Keywords: *Pneumocystis carinii* pneumonia; Leishmaniasis; CoMFA; Comparative QSAR

1. Introduction

The antimicrobial activity of aromatic bisamidines against a widespread range of microorganisms, like *Acanthamoeba*, *Babesia canis*, *Crithidia fasciculata*, *Cryptosporium parvum*, *Giardia lambia*, *Leishmania* spp., *Plasmodium* spp., *Toxoplasma gondiie* and *Trypanosoma* spp. is well known [1–3,43]. In case of visceral leishmaniasis, pentamidine isethionate (Pentacarinat®-Rhodia), which belongs to this class of compounds, has been used clinically to treat patients affected by *Pneumocystis carinii* pneumonia (PCP), occurring in ca. 80% of the HIV-infected individuals, as well as antimony-resistant leishmaniasis patients [4]. Its use has been partially restricted due to its high toxicity [5], and more potent and less toxic analogues have been sought, based on the rigid analogue approach [6]. These drugs usually accumulate in the kidneys (62.9 ± 53.2 μg g^{-1}), lungs (66.8 ± 14.3 μg g^{-1}), spleen/pancreas (54.3 ± 25.5 μg g^{-1}), heart (45.7 ± 24.6 μg g^{-1}), muscle (32.0 ± 5.0 μg g^{-1}) and stomach (28.6 ± 4.4 μg g^{-1}), remaining below 20% of the drug injected to be eliminated.

Side effects are expected along the treatment, like abscesses at the site of intramuscularly injection, severe hypotension, leucopoenia, thrombocytopenia, hypocalcaemia, hepatic, renal and pancreatic complications [3,43]. In case of leishmaniasis, if the treatment is kept for too long, more serious adverse effects may occur, such as hypoglycaemia and acute pancreatitis, leading to diabetes [7]. Hence, pentamidine is second-line in the treatment of PCP, although it is as effective as the combination of trimethoprim and sulfamethoxazole (TMP-SMZ) [5].

Leishmaniasis and HIV virus co-infection has drawn a new profile for this disease in the last years [8]. A

Abbreviations: AT_m, denaturation temperature variation; DDD, Dickerson Drew Dodecamer; PLS, partial least square; MM, molecular mechanics; QSAR, quantitative structure–activity relationships; PDB, Brookhaven protein database.

* Corresponding author.
E-mail address: montana@dedalus.lee.ufmg.br (C.A. Montanari).

0223-5234/03/$ - see front matter
doi:10.1016/S0223-5234(02)01419-8

synergistic relationship occurs between both microorganisms when promastigotes triggers tumour necrosis factor-alpha (TNF-α) synthesis, which is responsible to regulate HIV replication [9].

It has also been suggested [10] that aromatic bisamidines bind specifically through a non-covalent and non-intercalative mode to AT-rich regions of B-DNA. The ligand adopts an 'isohelical' conformation to fit the minor-groove, in which both the aromatic rings stand in a 'twisted' position, following the helix curvature [11,12], as illustrated in Fig. 1.

Some alternatives have arisen to explain the mode of action of these molecules. It has been proposed that they can inhibit the synthesis of phospholipids, polyamides and proteins [13–16], polymerases [17], the complex dihydrofolate reductase–tymidilate synthetase [18], and some topoisomerases (some exhibit an IC_{50} over 50 μM in the presence of ATP-independent topoisomerase [19,20]).

Previous quantitative structure-activity relationship (QSAR) and molecular modelling studies carried out by our group [12] have demonstrated the importance of the bisamidine group in the *para*-position of the aromatic ring to enhance the potency, and the novel structures were built obeying this criterion.

The structure-based ligand design emphasises the discovery of new lead compounds using the 3D structure of the receptor, towards a continuous increase in the ligand–receptor complementarity [21]. In this paper, a comparison based on molecular mechanics (MM) energy calculations is made among classical minor-groove binders (MGBs) and some rigid analogues, in order to rank them in their self-complementarity and affinity to $d(CGCGAATTCGCG)_2$ DNA sequence (also known as Dickerson Drew Dodecamer, (DDD) [22]). Further analysis is reported by the calculation of physico–chemical parameters and QSAR. Some rational bases for the proposal of new chemical substances that bind to DNA are proposed.

2. Experimental protocols

2.1. Molecular mechanics

A 45-compound set of MGBs (some uncommon structures are assigned at Fig. 2B) was built using the SYBYL program, version 6.5, [23], on a Silicon Graphics (O2 type) workstation. Three dimensional structures were obtained using Corina® [24], and their atom-centred ESP charges were derived from MNDO wave functions, available in Mopac 93.0 [25]. MM calculations were achieved using MacroModel v. 5.5 [26], applying modified AMBER force field [27].

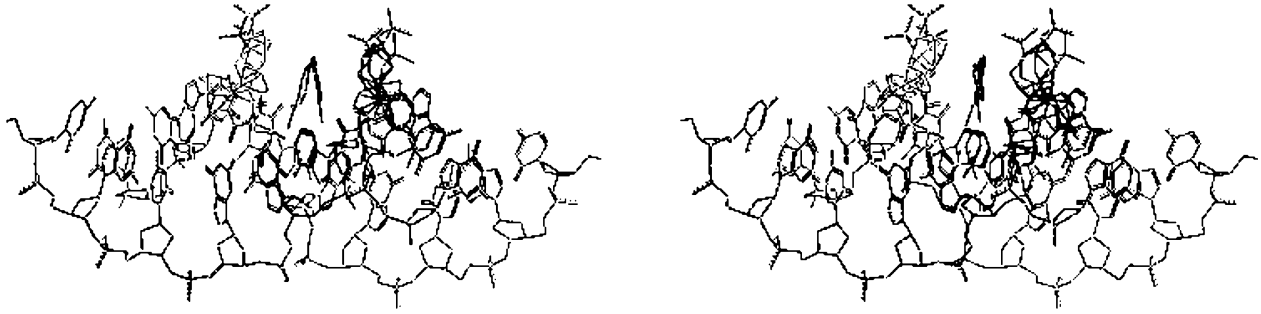
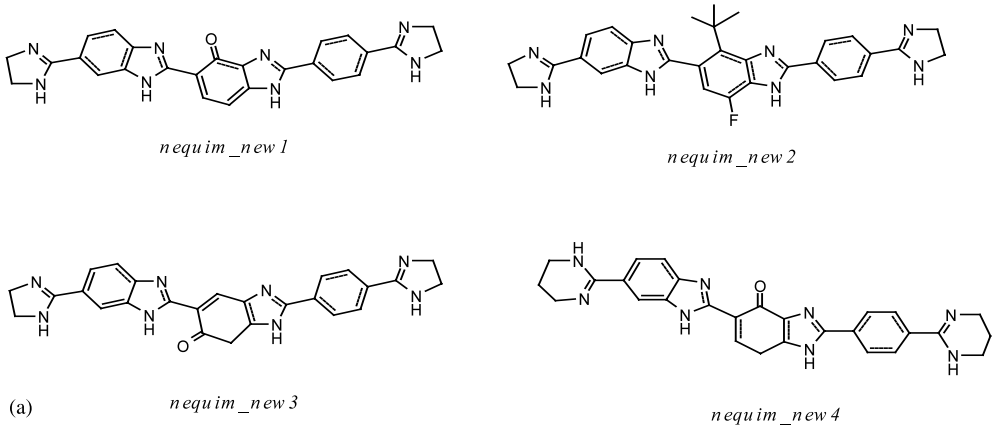
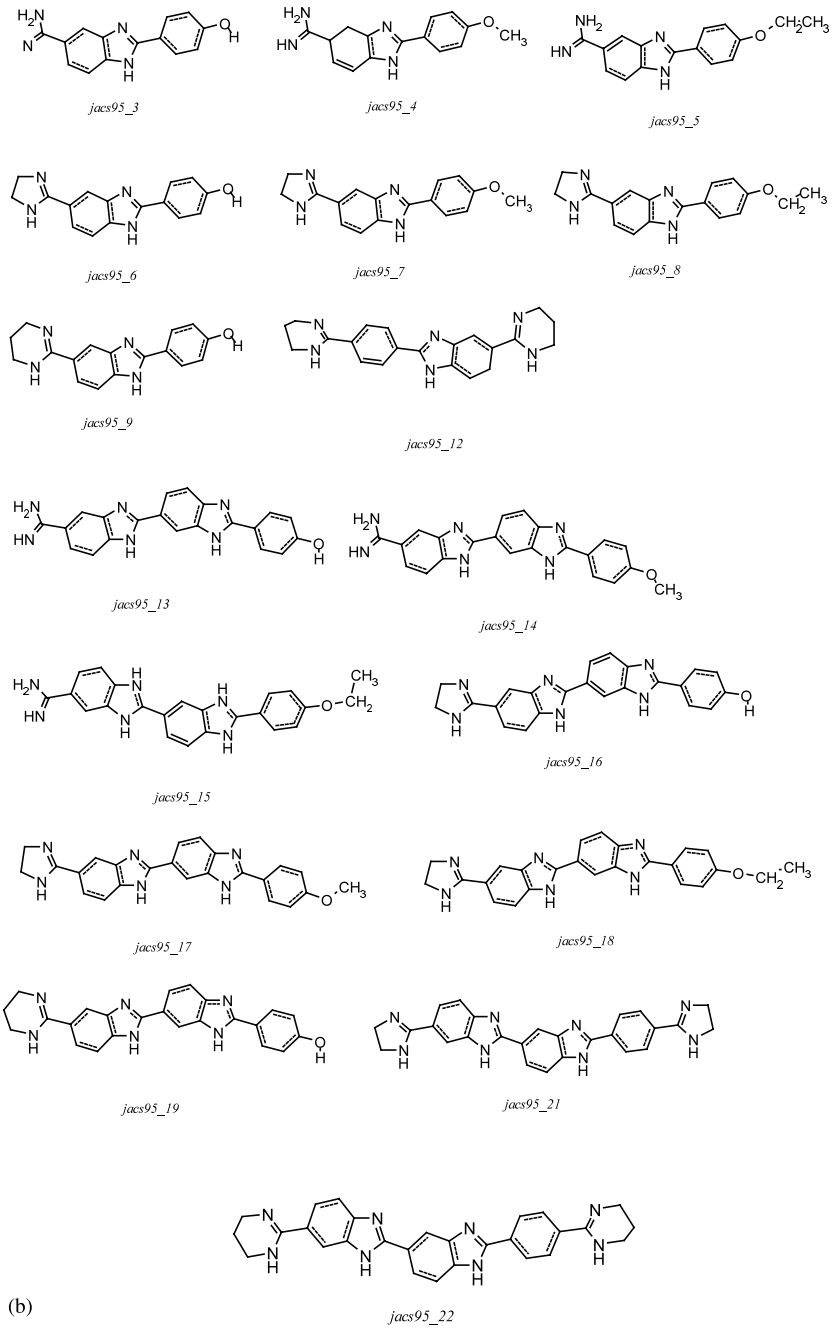


Fig. 1. X-ray structure (stereoview representation along the groove) of the complex between pentamidine and DDD (PDB entry 1D64 [18]).



(a)

Fig. 2. Structures of the compounds studied in the MM studies. (A) Novel compounds proposed by CoMFA studies. (B) Compounds by Czarny et al. (1995), used in QSAR studies. (C) Novel compounds proposed by previous studies made by our group.



(b)

Fig. 2 (*Continued*)

Parameters were set to dicationic bisamidine groups: Lennard–Jones potential as 10–12 and coulombic term was corrected to incorporate the continuous solvent model (dielectric constant $\varepsilon = 4r_{ij}$, where r_{ij} is the distance between atoms i and j [28,29]) as an attempt to prevent H-bond to phosphate groups in DNA backbone. This model follows Lavery and co-workers' proposal [30] of modelling the effect of the solvent by the use of a distance-dependent dielectric function. Although no explicit solvent was used in our calculations, a test employing crystallisation waters (in DDD and propamidine) present in X-ray structures showed a difference in ligand–receptor potential energy up to ca. 4% (structures with crystallisation waters have slightly more negative energies). This difference is quite negligible for our ranking purposes and within the inherent limitations of the method. Nonetheless, these results were also compared with the X-ray crystallographic structure of available Brookhaven protein database (PDB) complexes, which are in fair agreement with postulated 'isohelicity' to B-DNAs. Moreover, we have had pointed out earlier this to be the mode of action for

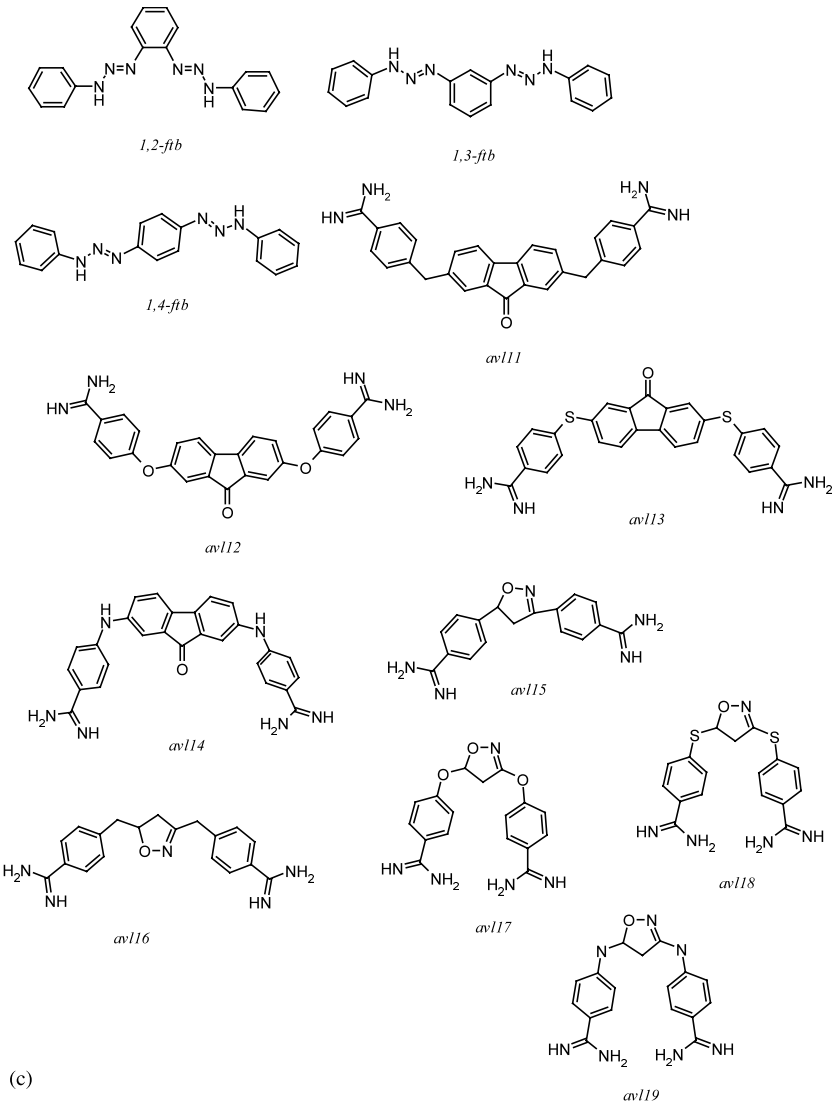


Fig. 2 (*Continued*)

a set of 37 pentamidine analogues [12] with antileishmaniasis activities. Yet, the method was also applied for a furamidine MGB and, once again, it conveyed the same trend that is, the method can be used for such interaction studies [29].

Ligand–DDD complexes were constructed keeping the following criteria: (1) 12-mer coordinates were picked up from those deposited in the Brookhaven Database, PDB [31,32]; (2) each compound was led to adopt the isohelical conformation likewise pentamidine, by the manual adjustment of the appropriated torsional angles; (3) structures were manoeuvred in order to fit into the minor groove by their superposition on pentamidine within the AT-rich region in the crystallographic modelled complex. The pentamidine molecule was then deleted and an additional adjustment was done in order to soothe the improper bump interactions with the walls of the groove. These intermolecular contacts were computationally minimised by using a 'belly'-type MM refinement of 500 cycles, steepest-descent type, to adjust the ligand alone, keeping the DNA strands rigid. (4) A MM minimisation of 4000 cycles was applied upon the entire complex, using the Polak–Ribiere procedure until reaching a RMS energy gradient of 0.05 kJ mol^{-1} at most. This task was pursued by the extension of non-bonded contacts to 8.0 and 20 A for charge–charge interactions. (5) Afterwards, the system was submitted to a molecular dynamics of 40 ps at 300 K, with an integration time step of 1.5 fs (among others of 100 and 1.0 fs, in a routine of MM-MD-MD-MD-MM). The procedure 'SHAKE' available in MacroModel package was used. The routine ends with a subsequent minimisation, leading to the final drug–DNA complex. The ranking of the 45 molecules was done by their complexation energy (E_C) calculated as from Ref. [33] (Eq. (1)).

$$E^{\mathrm{DNA-LIG}} = E^{\mathrm{CPLX}} - [E^{\mathrm{LBC}} + E^{\mathrm{DBC}}]$$

$$\Delta E^{\mathrm{LIGAND}} = E^{\mathrm{LBC}} - E^{\mathrm{LFC}}$$

$$\Delta E^{\mathrm{DNA}} = E^{\mathrm{DBC}} - E^{\mathrm{DFC}}$$

$$E_{\mathrm{C}} = \Delta E^{\mathrm{DNA}} + \Delta E^{\mathrm{LIGAND}} + E^{\mathrm{DNA-LIG}} \quad (1)$$

Here E^{CPLX} is the (absolute) potential energy of the final complex, E^{DBC} is the single-point energy of the DNA in the conformation adopted in the final complex (DNA bound conformation), and E^{LBC} is the energy of the ligand in its final conformation. $\Delta E^{\mathrm{LIGAND}}$ and ΔE^{DNA} are, respectively, the distortion energies involved in the conformational change of ligand and DNA from the free-minimum state to the pharmacophoric conformation. This quantity was chosen to describe the intensity of the affinity of the drug for the DNA in each case. The experimental binding affinity towards B-DNA was measured by log (ΔT_{m}), the logarithm of the variation of the denaturation temperature of the DNA when bound to the drug in relation to this temperature in its free state (without the drug).

The denaturation temperature, T_{m}, also known as helix–coil transition temperature, has been shown to be an adequate parameter for evaluating experimentally the extra stability that the ligand conveys to the oligonucleotide once complexed [34]. T_{m} is defined as follows: the biopolymer (nucleic acid, protein, enzyme, etc.) is gradually heated in solution until its secondary structure is unravelled. The helix–coil transition is highly entropic-dependent, what explains its temperature dependence [35], and is detected by the abrupt increase in DNA UV absorption (ca. 20–30%). The higher this temperature is, the most stable is the secondary structure. A ligand bound to the biopolymer usually enhances its T_{m} the stronger is the interaction (within certain limits, as will be discussed ahead). Hence, T_{m} may satisfactorily reflect the binding intensity.

2.2. GRID analysis

We have also studied the interaction of the DDD dodecamer (NDB code bld038) with some common metallic cation probes (Na^+, K^+ and Mg^{2+}), some organic probes similar to substructures present in the compounds we have designed, water and a probe named DRY, which represents the hydrophobic interactions. A clarity radius around the target molecule of 5.00 A, an acceptable distance of 5.00 A and 54 planes in the grid were used in the analysis. The visualisation of the maps was pursued using SYBYL 6.5 [23].

Table 1
Intrinsic binding energies (kcal mol^{-1}) according to Andrews and coworkers' model

Group, X	Energy, E_{X}	Range
DOF	−0.7	−0.7 to −1.0
$C(sp^2)$	0.7	0.6–0.8
$C(sp^3)$	0.8	0.1–1.0
N^+	11.5	10.4–15.0
N	1.2	0.8–1.8
CO_2^-	8.2	7.3–10.3
OPO_3^{2-}	10.0	7.7–10.6
OH	2.5	2.5–4.0
C=O	3.4	3.2–4.0
O, S	1.1	0.7–2.0
Halogen	1.3	0.2–2.0

2.3. Average free energy calculation

An average Gibbs free energy of binding (Eq. (2)) was estimated for each compound in order to be compared with the complexation energy calculated by MM as described above. This quantity was obtained with the functional-group contribution approach proposed in 1986 by Andrews, Craik and Martin [36], as outlined in Eq. (2). Although this model (statistically predicted from 200 non-covalent drug–receptor interactions taken up from literature) is not intended to be predictive, it is useful to establish a ΔG_{bind} range ($\delta[\Delta G_{\mathrm{bind}}]$) within which the ligand must lie and circumscribe a more suitable profile for the thermodynamics of its interaction with DNA.

$$\Delta G_{\mathrm{bind}}{}^{\mathrm{av}} = T\Delta S_{\mathrm{r,t}} + n_{\mathrm{DOF}}E_{\mathrm{DOF}} + n_{\mathrm{X}}E_{\mathrm{X}} \quad (2)$$

Here $T\Delta S_{\mathrm{r,t}}$ represents the loss of entropy term associated with the rotational and translational restrictions imposed to the ligand fitted to the active site. It is estimated to be of about 14 kcal mol^{-1}; n_{DOF} is the total of the internal degrees of freedom (rotatable bonds) of the ligand; E_{DOF} is the average entropy loss associated to each bond due to the binding; n_{X} represents the number of times the functional Group X (the authors studied ten) appears in the ligand structure, and E_{X} is the binding (intrinsic) free energy contribution attributed to X. On behalf of clearness, Table 1 shows the intrinsic group energies of the ten corresponding functional groups originally proposed by Andrews.

A BASIC-based routine, developed in our group, was used to estimate the range of free energy of binding and the corresponding inhibition constant, K_{i}. The former quantity is defined as the free energy of binding according to Eq. (2) taking E_{DOF} and E_{X} as the lower limits of the intervals presented in Table 1 for the minimum ΔG_{bind}, and the upper limits to calculate the maximum ΔG_{bind}. The free energy range $\delta[\Delta G_{\mathrm{bind}}]$ has been proven to match the experimental values much

Table 2
Calculation of the complexation energy (E_C, see Eq. (1)) for molecules studied [a]

Ligand [b]	log (ΔT_m)	E_C [c]	E^{CPLX}	E^{DBC}	E^{LBC}	$E^{DNA\text{-}LIG}$	E^{LFC}	ΔE^{LIGAND}	ΔE^{DNA}
12-Ftb*	–	−21.96	−195.37	−158.17	11.53	−48.73	10.05	1.48	25.29
13-Ftb*	–	−13.09	−188.83	−159.41	8.12	−37.54	7.72	0.40	24.05
14-Ftb*	–	−16.05	−188.81	−155.1	10.94	−44.65	10.70	0.24	28.36
Avl11*	–	−28.71	−165.6	−152.25	66.21	−79.56	46.57	19.64	31.21
Avl12*	–	−42.35	−173.36	−158.24	62.94	−78.06	52.45	10.49	25.22
Avl13*	–	−34.51	−171.52	−151.4	57.87	−77.99	46.45	11.42	32.06
Avl14*	–	−46.53	−175.53	−155.92	57.47	−77.08	54.46	3.01	27.54
Avl15*	–	−3.35	−171.59	−160.84	24.66	−35.41	15.22	9.44	22.62
Avl16*	–	−34.81	−206.05	−155.41	24.49	−75.13	12.22	12.27	28.05
Avl17*	–	−37.01	−196.81	−154.56	18.2	−60.45	23.66	−5.46	28.9
Avl18*	–	−30.97	−200.18	−155.78	15.31	−59.71	14.25	1.06	27.68
Avl19*	–	−36.37	−203.77	−158.04	18.57	−64.30	16.06	2.51	25.42
Berenil	–	−41.62	−217.96	−158.81	8.33	−67.48	7.12	1.21	24.65
Dapi	1.233	−5.08	−168.85	−121.16	8.17	−55.86	9.53	−1.36	62.3
Dipi	1.193	−62.90	−212.55	−158.78	15.3	−69.07	33.81	−18.51	24.68
GOP	–	−40.14	−215.58	−158.95	13.02	−69.65	8.02	5.00	24.51
Guafur	–	−59.73	−212.5	−155.03	9.3	−66.77	30.69	−21.39	28.43
HD33258	1.516	−56.33	−220.31	−155.61	12.15	−76.85	19.48	−7.33	27.85
HOE58	1.253	−70.53	−208.53	−156.19	22.78	−75.12	45.46	−22.68	27.27
HOE42	–	−75.99	−219.44	−158.07	24.13	−85.50	40.01	−15.88	25.39
Imiben	–	−76.79	−208.52	−158.91	18.54	−68.15	51.73	−33.19	24.55
Isofur	–	−67.12	−218.11	−159.35	10.85	−69.61	32.47	−21.62	24.11
jacs95_12	1.233	−238.27	−214.17	−158.7	18.31	−73.78	207.56	−189.25	24.76
jacs95_13	1.362	−182.75	−189.36	−155.94	23.46	−56.88	176.85	−153.39	27.52
jacs95_14	1.362	−130.65	−134.67	−140.27	64.76	−59.16	179.44	−114.68	43.19
jacs95_15	1.362	−146.84	−153.45	−138.49	45.23	−60.19	176.85	−131.62	44.97
jacs95_16	1.378	−148.89	−151.85	−141.34	52.71	−63.22	180.50	−127.79	42.12
jacs95_17	1.387	−132.93	−133.67	−145.28	75.09	−63.48	182.72	−107.63	38.18
jacs95_18	1.412	−152.98	−152.9	−145.63	56.07	−63.34	183.54	−127.47	37.83
jacs95_19	1.405	−148.76	−149.68	−158.47	62.98	−54.19	182.54	−119.56	24.99
jacs95_21	1.479	−61.46	−151.72	−159.69	86.2	−78.23	93.20	−7.00	23.77
jacs95_22	1.487	−53.34	−171.12	−155.93	68.31	−83.50	65.68	2.63	27.53
jacs95_3	0.845	−44.74	−195.55	−159.74	12.26	−48.07	32.65	−20.39	23.72
jacs95_4	0.792	−64.66	−192.5	−156.72	13.58	−49.36	55.62	−42.04	26.74
jacs95_5	0.690	−65.60	−193.58	−156.73	13.91	−50.76	55.48	−41.57	26.73
jacs95_6	0.748	−65.52	−185.36	−156.18	21.34	−50.52	60.62	−39.28	27.28
jacs95_7	0.875	−59.96	−180.16	−157.32	23.82	−46.66	63.26	−39.44	26.14
jacs95_8	0.845	−58.90	−179.24	−151.44	24.62	−52.42	63.12	−38.50	32.02
Jacs95_9	0.845	−86.47	−191.46	−155.56	11.29	−47.19	78.47	−67.18	27.9
Mhoechst	–	−86.34	−210.78	−157.76	21.12	−74.14	59.02	−37.90	25.7
Pentamidine [d]	–	−43.18	−214.65	−157.65	12.06	−69.06	11.99	0.07	25.81
Propamidine	–	−38.42	−214.01	−157.04	9.89	−66.86	7.87	2.02	26.42
Sn6999	–	−65.92	−204.3	−157.35	31.31	−78.26	45.08	−13.77	26.11
Sn7167	–	−66.21	−204.97	−157.63	34.4	−81.74	44.70	−10.30	25.83
Tribiz	–	−126.49	−215.96	−157.11	27.62	−86.47	93.99	−66.37	26.35

Compounds marked with an asterisk are unpublished. GOP, gamma-oxa-pentamidine; HOE42, Hoescht 33342; HOE58, Hoescht 33258.

[a] $E^{DFC} = -183.46$ kcal mol^{-1}, obtained by minimisation of crystallographic structure. ΔT_m data from Ref. [37].

[b] Compounds proposed from CoMFA studies (Fig. 2 (A)) were not included in this study.

[c] One may argue, by checking the results of the docking here outlined, that some unpublished compounds seem to have a worse complexation affinity than some classical compounds (e.g. avl13 vs. pentamidine). It should be considered, however, that this docking proceeding is more adequate to be applied in an overall statistical basis, by plotting the calculated complexation energies of the literature ligands with their experimental DNA affinities, and then interpolating the novel compounds to estimate their affinities, just like described in section of Section 2.1.

[d] Value for pentamidine as calculated by the method outlined herein bear a resemblance to that put forward by Samson et al. Anticancer Drug Des., 5 (1990) 243–8, E_C, −52 kcal mol^{-1}.

better (error below 10%) than the average binding energy Eq. (2). Notwithstanding this error seems to be too high when compared with classical free energy calculations based, for example, on free energy perturbation theory approaches. However, it is not time-consuming and is useful to draw out a general profile of the entire database binding range.

2.4. QSAR analysis

Further investigation of the system was carried out by the alignment of the compounds studied and calculation of their various physical–chemical descriptors (similarity, topological, hydrophobic, electronic and electrotopological), besides $E^{DNA-LIG}$, using Tsar™ [37].

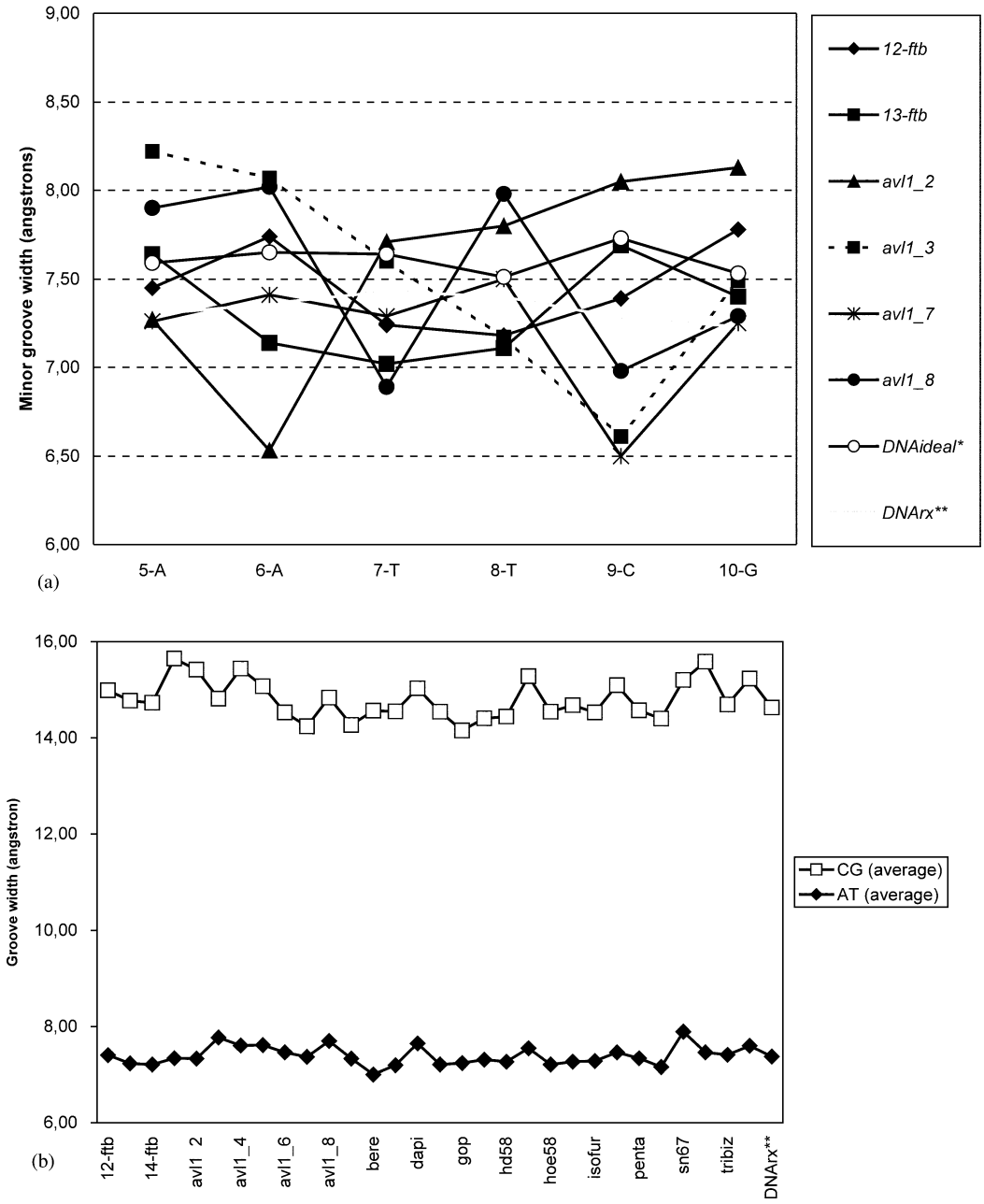


Fig. 3. (A) Average groove width in the AATT-rich region for some unpublished MGBs (on behalf of clearness, classical compounds data are not plotted); (B) Average groove width comparing AT and CG regions in the dodecamer d(CGCGAATTCGCG)$_2$.

In this QSAR analysis, a rigid alignment was performed by molecular weight extent, size weighting factor = 1 and charge weighting factor = 10. Conformations are the same as those bound to the DDD minor-groove as obtained by MM. Multiple regression analysis was used to obtain the relevant relationships between log (ΔT_m) and the physico–chemical parameters and $E^{DNA-LIG}$. To avoid combining strongly correlated descriptors in the same equation, a correlation matrix was built (through Principal Component Analysis) to eliminate such cases. The search for the best model was made taking the various combinations among each kind of descriptor (four descriptors at most for the 21-compound set).

2.5. Comparative molecular field analysis (CoMFA)

A 3D-QSAR type study (using comparative molecular field analysis (CoMFA) method [38]) was carried out using SYBYL 6.5 with the 21 analogues proposed by Czarny and co-workers [39], also present in our database (see structures in Fig. 2(B), plus DAPI, DIPI,

Table 3
GRID most negative energies (in decreasing order) for each probe employed with DDD

Probe	Most negative energy within the lattice (kcal mol^{-1})	Regions in the target molecule where the interaction is the strongest
Magnesium ion	−42.99	Phosphate groups > minor groove
Sodium ion	−31.84	Minor groove ≫ phosphate groups
Potassium ion	−31.29	Minor groove ≫ phosphate groups
Aliphatic amidine	−23.57	Minor groove, AT-rich region
Aromatic amidine	−23.07	Minor groove > major groove
Trimethylamino	−12.49	Minor groove and part of the GCG-rich region
Aromatic amide	−11.59	Low specificity (interacts with both grooves and the phosphate ones)
Water	−9.80	Hydrogen bonds with the nitrogen 3 and ribose oxygen (both at adenosine 16), beside phosphates
Furan ring oxygen	−4.73	Low specificity. Interacts with both grooves along the entire dodecamer
C sp^2 (aromatic and allyl)	−4.05	Interaction occurs only with the minor groove
Hydrophobic probe (DRY)	−2.41	Both ends of the double helix (where there are no phosphates), as well as some regions within the minor groove

HOECHST33258), with consequent proposal of some novel molecules potentially active (nequim_new1-4, Fig. 2A). These molecules were chosen once their ΔT_m values were given and have been measured in the same conditions. The structures, whose 3D structures were generated by CORINA™, were aligned according to

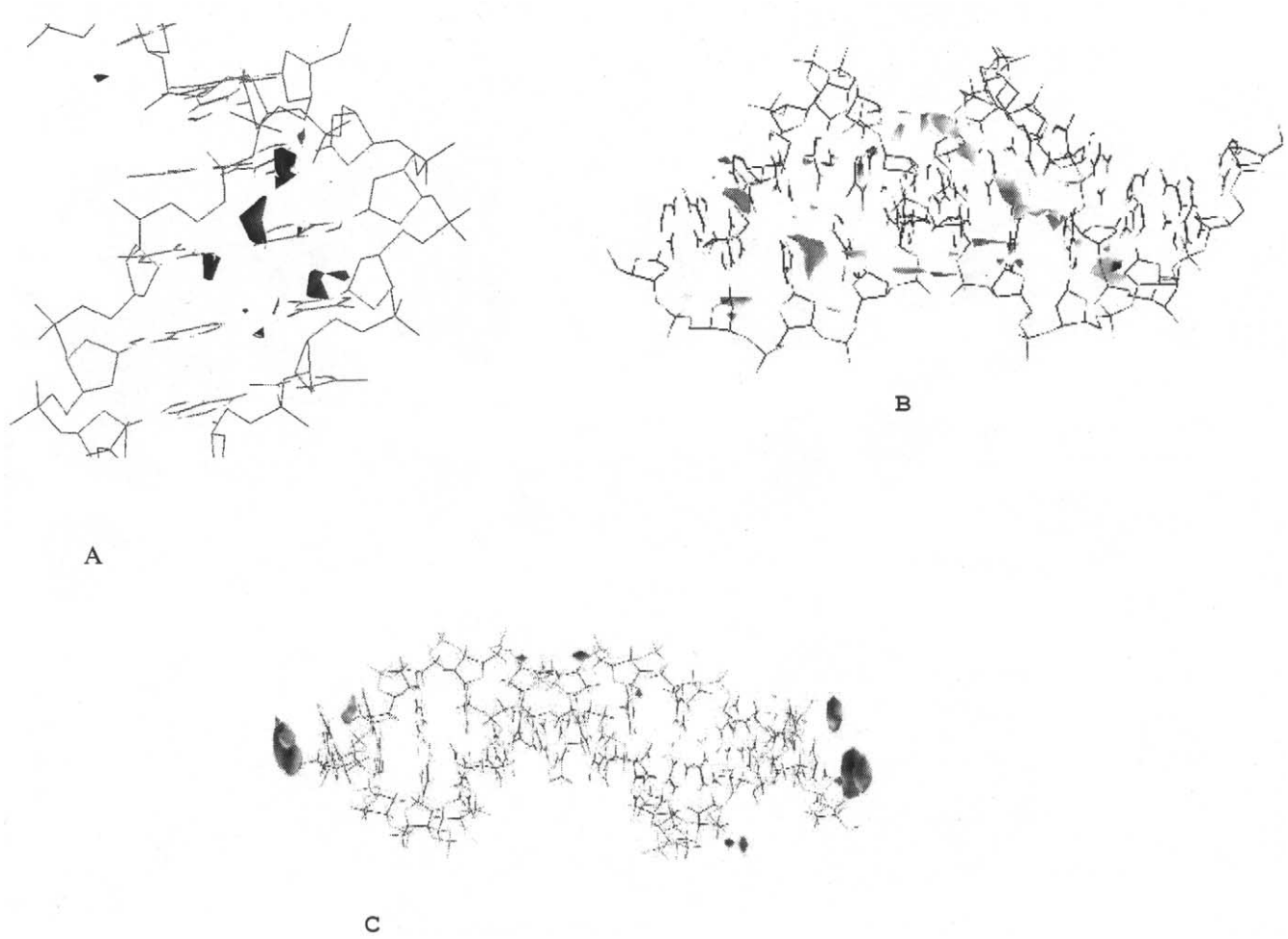


Fig. 4. GRID maps for metallic counterion probes sodium (A), for water (B), and for the hydrophobic one (C).

isohelical conformation of pentamidine (X-ray). The charge-based alignment was achieved using Gasteiger charges, although MNDO ESP charges were the values chosen for the MM calculations. Gasteiger charges have been shown to afford better statistical partial least square (PLS) results, despite the final qualitative results obtained from CoMFA are exactly the same as those obtained with MNDO ESP charges. ESP charges, by their turn, seem to be more adequate to be used in AMBER coulombic energy term in MM, since the values are closer to the formal charges.

Common benzimidazole pharmacophoric group was used for aligning, taking grid spacing = 2.0 A, and four different probes (lattice dimensions for each case is reported ahead): (1) C^+ (sp^3) (to draw a general profile of the structural implications of the complexation to DDD); (2) H^+ and O sp^3 (both to investigate the role of hydrogen bonds and oxygen to simulate interaction with water and DNA pentose); (3) Ca^{2+} (to check the partaking of divalent counter-ions in the stability of DDD).

PLS analysis was made in two steps: the first one was achieved by using the leave-one-out (LOO) method to evaluate cross-validation, and the second one omitted this procedure.

3. Results and discussion

3.1. Molecular mechanics

The interaction energies calculated for the series can be found in Table 2. Values seem to be in accord to the average experimental affinity order that can be inferred by ΔT_m determinations.

It can be seen that the majority of the novel structures (Fig. 2C) present a more negative $E^{DNA-LIG}$ than the classic molecules included. These results confirm the need of the charged bisamidine group to enhance the recognition power of the drugs. Hence, nequim1, nequim2 and nequim3 have low complementarity, but nequim12, nequim13 and nequim14, which differ of the former ones only for the presence of the amidine groups in their aromatic rings (*para*- to the lateral chain), have strong affinity energy to DNA.

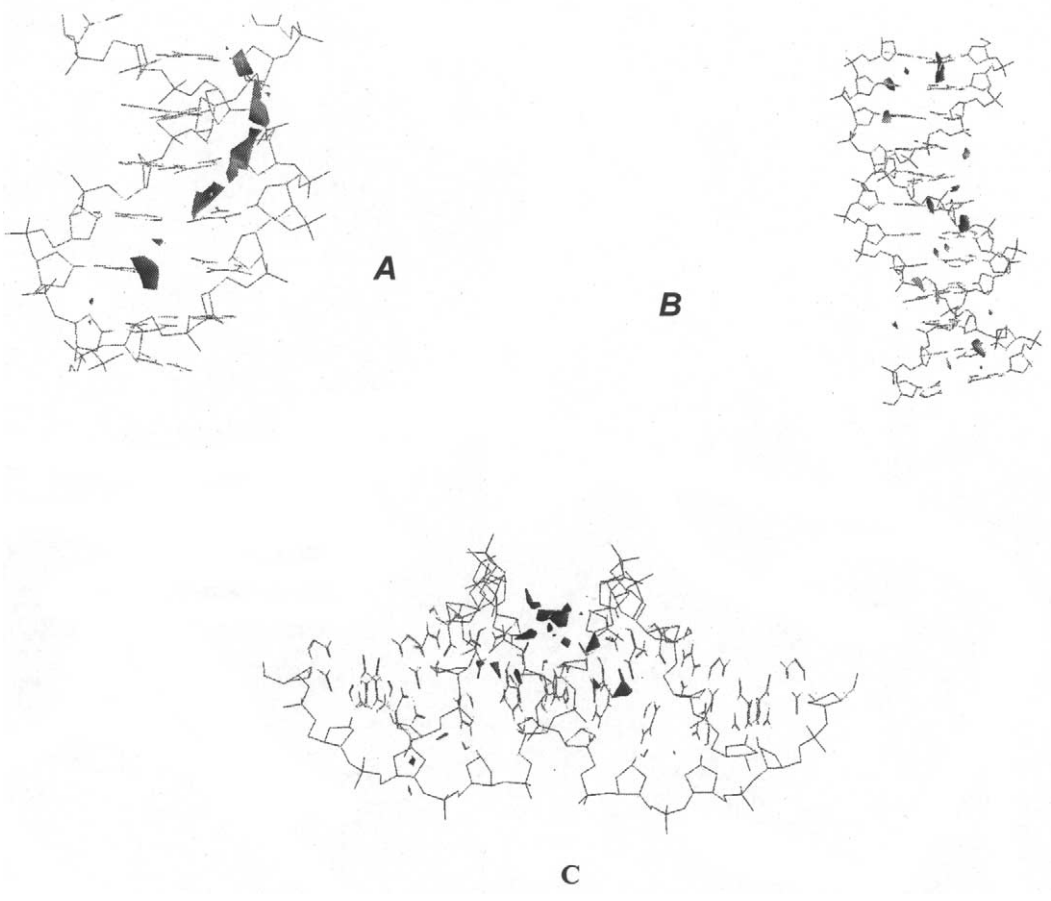


Fig. 5. GRID maps for organic probes: aromatic amidine (A), furan ring oxygen (B) and sp^2 aromatic and allylic carbon (C).

Moreover, we have checked some structural and energetic changes that come from the formation of the drug–DNA complex (mainly minor groove width). Fig. 3(A) illustrates the variation in this parameter in the AATT region for some of the ligands, and Fig. 3(B) shows the average width for AT and CG-rich sequences.

Starting from this point, an attempt to validate the method correlating the energetic quantities obtained from MM studies and ΔT_m determinations was made. Trials to achieve a linear relationship between complexation energy (including the distortion term) or interaction energy (omitting it) and denaturation temperature failed completely. Notwithstanding, an interesting result arose when a parabolic fitting was used to correlate log (ΔT_m) and the interaction energy (without the distortion energy), which is expressed in Eq. (3) (a similar but statistically worse equation can be obtained with E_C instead of $E^{DNA-LIG}$).

$$\log(\Delta T_m) = -9.38(\pm 7.0)10^{-4}[E^{DNA-LIG}]^2 - 0.138(\pm 0.090)[E^{DNA-LIG}] - 3.630(\pm 2.78)$$

$$E^{DNA-LIG}(\text{optimum}) = -73.5(-104/-68.4)$$

$$(n = 21;\ r^2 = 0.69;\ s = 0.166;\ F = 20.15;\ r_{cv}{}^2 = 0.578) \quad (3)$$

Eq. (3) reveals that the dodecamer stability (expressed by the denaturation temperature) goes better with an increasingly better fitted ligand, as expected. Nonetheless, some ligands, whose $E^{DNA-LIG}$ are more negative than ca. −73.5 kcal mol^{-1}, bind so tightly to the receptor that the complexation becomes thermodynamically undesirable (surrounding the limits of an irreversible binding). It means that this ligand diminishes the double helix overall stability, so that it would be better to the helix to denature than keep those ligands inside. This prevents lower ΔT_m values for high-interaction energy compounds.

Moreover, it was verified that the binding of MGBs to the minor groove provokes some geometric effects: (1) A smoothing narrowing occurs in the minor groove. This effect is probably due to the non-bonded interactions (mainly hydrogen bond and electrostatic interactions in the charged groups) with purine and pyrimidine heteroatoms and ribose oxygen (Fig. 3(A)). (2) The drug is continuously adjusted to the 'isohelical' conformation. (3) There is a mild preference for AT sequences instead of CG (Fig. 3(B)), as already reported [10]. The average width of the groove in AT region is about 5% smaller than this width in CG region.

Overall, the rigid isohelical conformers (e.g. dapi, berenil, Hoechst33258) exhibited higher complementarity to DNA than their more flexible partners (e.g. pentamidine, propamidine) did.

3.2. GRID analysis

Table 3 shows the most negative interaction energies for each probe and the most important characteristics of the interaction in each case. Some of the corresponding GRID maps are illustrated in Figs. 4 and 5.

As already expected, aromatic and aliphatic amidines have played an important role in the recognition. Both exhibit similar interaction energies, around −23 kcal

Table 4
Average free energy (ΔG_{bind}), minimum ΔG_{bind}, maximum ΔG_{bind} and average free energy ranges ($\delta[\Delta G_{bind}]$) for the compounds of our database

Ligand	ΔG_{bind}	$\delta[\Delta G_{bind}]$	
		Minimum ΔG_{bind}	maximum ΔG_{bind}
Hoe58	52.20	40.10	69.40
GOP	52.00	39.90	69.80
Hoe42	52.00	39.90	69.80
Pentamidine	51.40	38.70	67.80
Avl16	37.60	25.20	53.50
Berenil	37.60	25.20	53.50
Jacs95_22	36.00	23.00	51.20
Jacs95_21	34.40	22.80	49.20
Avl11	31.20	22.40	45.20
Jacs95_12	29.40	18.20	42.70
Sn6999	26.40	16.30	40.40
Avl18	26.20	16.10	40.80
Avl15	26.20	16.10	40.80
Dapi	26.00	18.80	38.30
Mhoechst	25.40	16.70	38.20
Guafur	24.70	17.50	36.80
Hd33258	24.60	17.60	36.70
Dipi	23.90	10.90	37.00
Avl13	23.70	15.80	36.30
Sn7167	23.50	10.90	37.20
Jacs95_19	23.30	14.30	35.50
Avl14	23.10	10.90	37.40
Jacs95_18	22.70	12.40	35.10
Avl12	22.60	12.10	35.80
Jacs95_17	22.60	13.30	34.80
12-Ftb	22.50	14.20	34.50
Jacs95_16	22.50	14.20	34.50
Avl19	21.90	9.60	33.30
Imiben	21.90	9.60	33.30
14-Ftb	21.90	9.60	33.30
13-Ftb	21.90	9.60	33.30
Jacs95_15	21.10	12.20	33.10
Jacs95_14	21.00	13.10	32.80
Jacs95_13	20.90	14.00	32.50
Jacs95_9	16.70	9.50	27.00
Jacs95_6	15.90	9.40	26.00
Jacs95_3	14.30	9.20	24.00
Jacs95_8	13.60	5.10	22.60
Jacs95_7	13.50	6.00	22.30
Jacs95_5	12.00	4.90	20.60
Jacs95_4	11.90	5.80	20.30
Avl17	1.60	−4.40	7.00
Isofur	1.60	−4.40	7.00
Propamidine	1.60	−4.40	7.00

Values in kcal mol^{-1}.

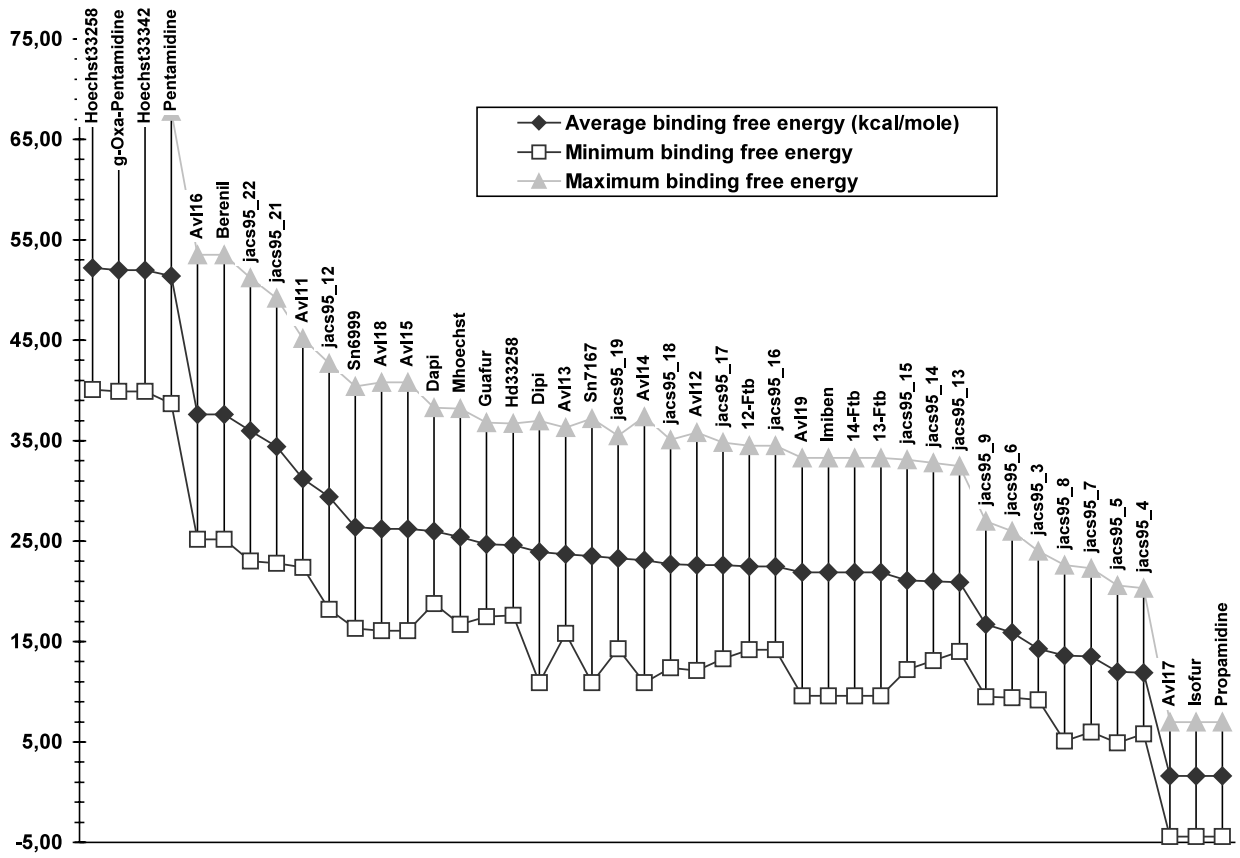


Fig. 6. Average free energy (ΔG_{bind}), minimum ΔG_{bind}, and maximum ΔG_{bind} for compounds of studied database. Values in kcal mol^{-1}.

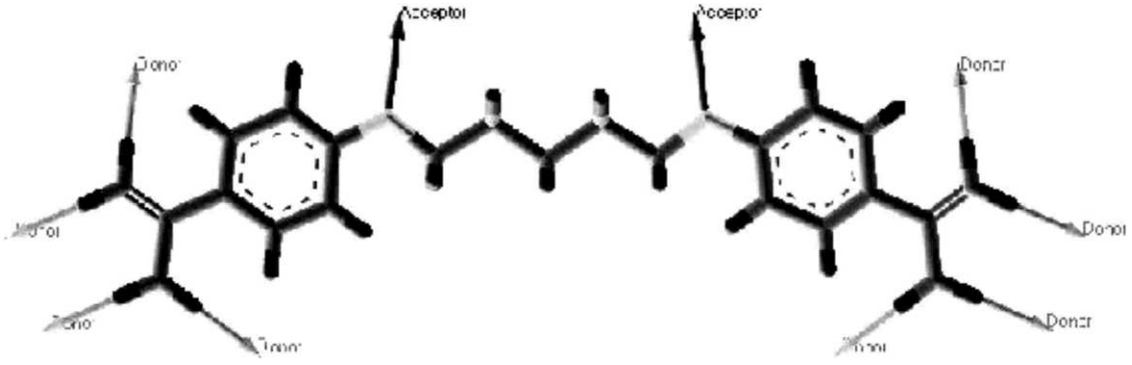


Fig. 7. Donor–acceptor profile of a typical MGB (pentamidine).

mol^{-1}. Moreover, the most positively charged the metallic ion is, the stronger is its interaction, including in the minor groove. Although furan groups are largely used in this kind of drugs (like furamidine), this study throws some doubt on its intrinsic AT-region specificity contribution. On the other hand, results have shown that the bioisosteric substitution of –O– by –NH– enhances the recognition. It corroborates the previous behaviour described by Montanari et al. [12]. The same observation can be made on the amide group, present in dystamicin and in netropsin, oligopeptide-type derivatives firstly reported in the literature. Trimethylamino may be a good choice as a group in this sort of structures, regardless the fact it differs AT and CG regions poorly.

Hydrophobic interactions once again have been proven to give their contribution to nucleic acids recognition. Dr. W.P. Jencks [40] have asserted that, despite their large dipole moments, nucleic acid bases present a slight hydrophobic profile once their solubility in water is often small, increasing gradually with the addition of hydrophobic compounds to the water. This kind of interaction seems to partake the recognition of specific regions within the DDD double helix, once the

Table 5
CoMFA statistical results for each one of the four probes used with Czarny's compounds

Probe	Charge	Lattice size (X, Y, Z steps)	r^2_{cv} (No. of components)	r^2	s	F	% Steric	% Electrostatic
C sp^3	+1	10, 9, 8	**0.891 (2)**	0.938	0.074	135.52	60.0	40.0
H	+1	24, 24, 24	**0.894 (2)**	0.941	0.073	142.98	71.2	28.8
O sp^3	0	24, 24, 24	**0.898 (2)**	0.942	0.074	147.11	100.0	0.0
Calcium	+2	24, 24, 24	**0.885 (2)**	0.943	0.071	147.85	74.4	25.6

interaction energies involved are small but not negligible (ca. −2 kcal mol^{-1}, see Table 3). It is probably due to the large amount of water molecules displaced from the inner part of the minor groove meanwhile the entrance of the ligand, which results in a substantial entropic contribution to the formation of the binary complex (ligand-DDD).

Neto [41] has demonstrated by the use of some molecular electrostatic potential calculations that these water molecules roughly assume a tetrahedral conformation within the AT-rich DDD segment (unlike CG segment, where this ordination is interrupted). Our results have also confirmed Dr. Neto's statement that the phosphate groups have high hydration ability.

3.3. Average free energy calculation

Andrews' free energy calculation yielded the results assigned in Table 4 and plotted in Fig. 6. As expected, pentamidine, Hoechst33258 and berenil exhibit a high average binding energy, just like some of our novel structures (mainly avl16, avl11, avl18 and avl15). Taking into account the limitations of the method already commented, those compounds are potentially better ligands even than HD33258, one of the most profitable MGBs ever conceived.

3.4. QSAR analysis

The methodology earlier described resulted in the following relevant model:

$$\log(\Delta T_m) = 0.227(\pm 0.020)\mathrm{naH-bond} + 0.214(\pm 0.010)$$

$$n = 21;\ r^2 = 0.829;\ s = 0.119;\ F = 93.3;\ r_{cv}{}^2 = 0.805 \quad (4)$$

Here naH-bond is the number of H-bond acceptor atoms (mainly O, N). Donor and acceptor H-bond

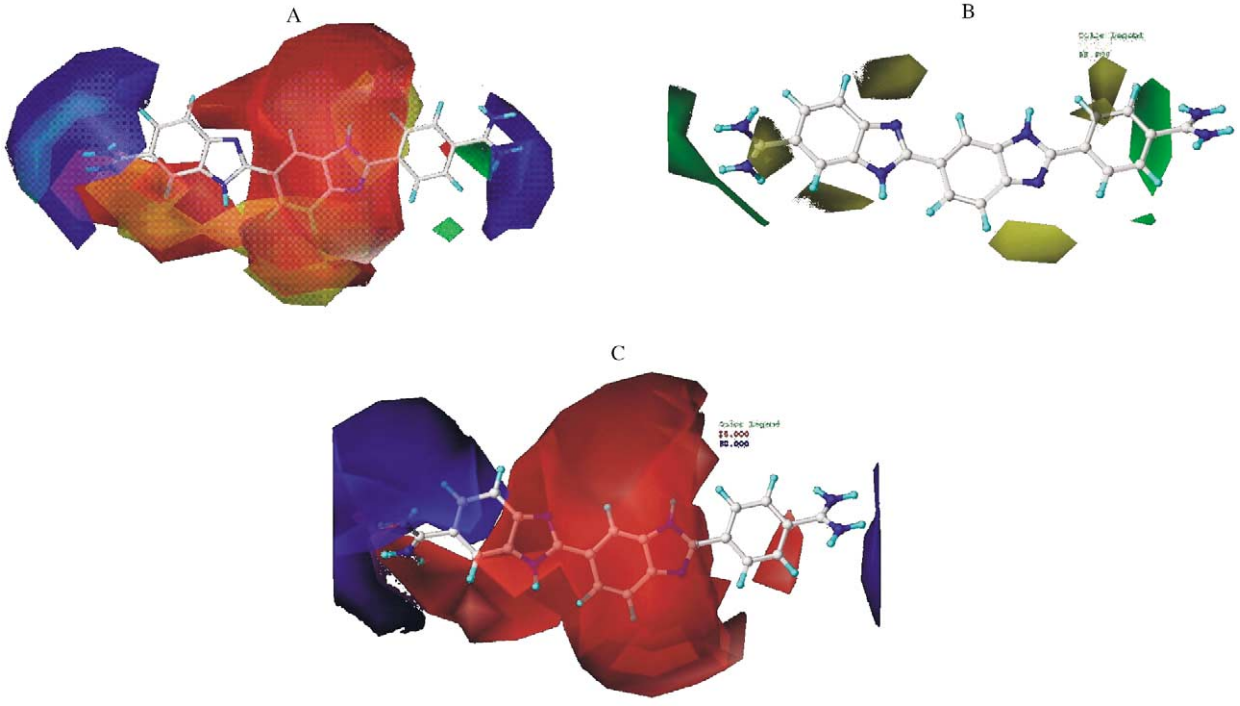


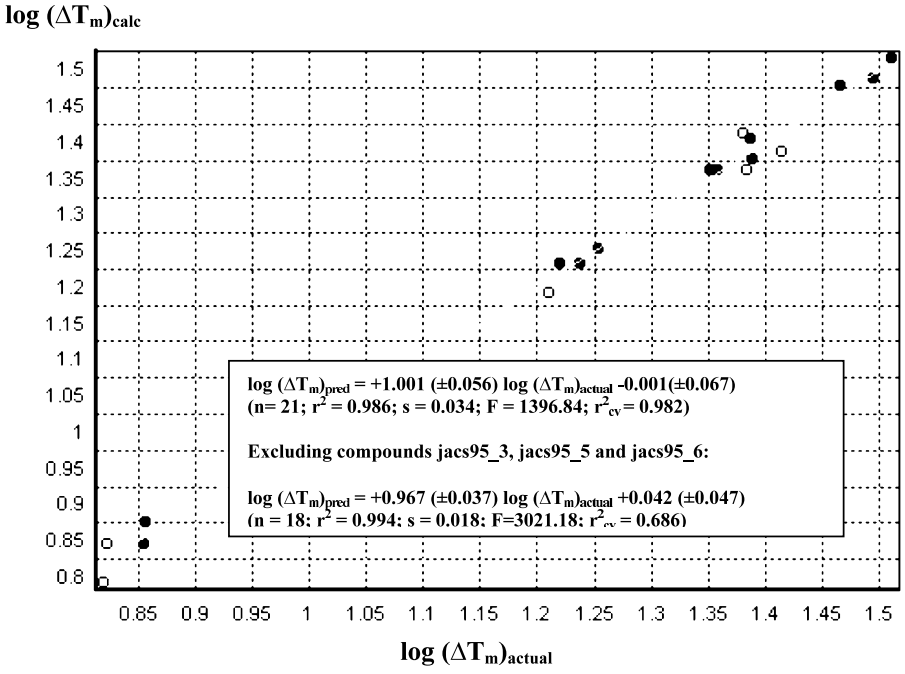
Fig. 8. CoMFA map showing HD33258. The colour patterns are: steric map is shown in yellow (less steric bulk favoured) and green (more bulk favoured), and electronic map in blue (more positive charge nearby is favoured) and red (less positive charge favoured). Probe: (A) C sp^{3+}, (B) H^+ steric and (C) H^+ electrostatic.

atoms are predominant in this class of molecules, as can be illustrated in Fig. 7.

A gainful discussion arises when we consider the latter result's closer implications. It has been proposed [33] that hydrogen bonds between ligand and DNA are not indispensable, although they enhance the binding when available. Unlike Czarny's compounds used in our QSAR analysis, in which H-bond was found to be important, some compounds described in the literature (like SN18071, which has no suitable H-bond acceptor atoms) exhibit good affinity to DNA, despite the predominance of apolar interactions.

3.5. Comparative molecular field analysis

Some suitable standard CoMFA models were obtained (see Table 5). Fig. 8 shows the contour map representation of the molecular fields corresponding to



Ligand	Actual log(ΔT_m)	Calculated log(ΔT_m)	Residue
DAPI	1.233	1.216	0.017
DIPI	1.193	1.206	-0.013
HD58	1.516	1.516	0.000
HOECHST33258	1.253	1.251	0.002
JACS95_12	1.233	1.234	-0.001
JACS95_13	1.362	1.359	0.003
JACS95_14	1.362	1.385	-0.023
JACS95_15	1.362	1.351	0.011
JACS95_16	1.378	1.389	-0.011
JACS95_17	1.387	1.416	-0.029
JACS95_18	1.412	1.381	0.031
JACS95_19	1.405	1.388	0.017
JACS95_21	1.479	1.469	0.010
JACS95_22	1.487	1.500	-0.013
JACS95_3	0.845	0.776	0.069
JACS95_4	0.792	0.802	-0.010
JACS95_5	0.690	0.767	-0.077
JACS95_6	0.748	0.814	-0.066
JACS95_7	0.875	0.841	0.034
JACS95_8	0.845	0.806	0.039
JACS95_9	0.845	0.838	0.007

Fig. 9. Log (ΔT_m) values calculated by CoMFA model vs experimental values. Compounds jacs95_3, jacs95_5 and jacs95_3 were excluded due to their high residues. Hollow points designate those compounds whose calculated value residues exceeded the SD.

the steric and electrostatic contribution to the model obtained. Some analogues were proposed based on the structure of the most complementary compound of the series, HD33258. Results are similar for all probes.

Fig. 9 illustrates the graph relating the actual and the CoMFA calculated log (ΔT_m) values, and demonstrates its robustness.

An independent additional validation procedure has been carried out in order to establish the predictive power of the 3D-QSAR model outlined. A test set containing some MGBs (furamidine derivatives) proposed by Francesconi et al. [42], was used to do this. Their structures, actual ΔT_m and predicted ΔT_m by our CoMFA model (with corresponding residues) are shown in Table 6. An appropriate PLS regression model between Francesconi's dataset and our predicted values for log (ΔT_m) was obtained ($n = 8$; $r^2 = 0.806$; $r^2_{cv} = 0.745$), which demonstrates the robustness of the designed CoMFA model.

The CoMFA analysis allows one to infer the requirements for the recognition of the minor groove, despite it does not take into account the 3D structure of the receptor. Fig. 8 shows that the presence of the positively charged amidine nearby the blue map confirms its fundamental role in the large potency of this molecules.

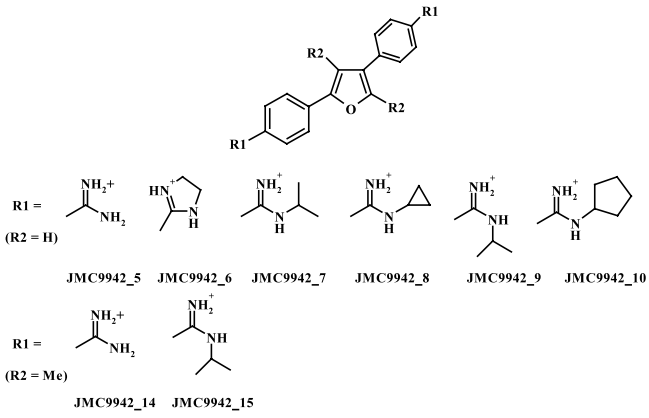
Another information that was obtained from CoMFA analysis is that the affinity increases when the middle chain is attached to small and electron-rich substituents. This fact seems to agree with the previous available model in which the drug is bound within the minor groove, requiring few bulk groups in the mid-chain, to avoid a hard steric clash against the groove's walls [10,11].

Besides, CoMFA shows that the steric contribution is always bigger than the one due to the electrostatic contribution. It corroborates our previous knowledge about the importance of the isohelical conformation for the improvement of the complementarity to DNA.

These results led us to design and propose some new structures obeying these criteria, and to predict the ΔT_m of the analogues already proposed. Table 7 exhibits the calculated log (ΔT_m) for the 12 new molecules studied by MM described above, as well as for the four new analogues designed after obtaining CoMFA model.

Table 7 permits to verify that some analogues from the list just proposed must have better affinities than other MBGs published in the literature.

Table 6
Predicted vs. actual log (ΔT_m) of the furamidine analogues proposed by Francesconi et al. [42]



Ligand	Actual log (ΔT_m)	Predicted log (ΔT_m)	Residual for log (ΔT_m)
JMC9942_5	0.959	0.868	0.090
JMC9942_6	0.857	0.870	−0.013
JMC9942_7	1.037	0.982	0.055
JMC9942_8	1.037	0.963	0.006
JMC9942_9	0.968	0.963	0.006
JMC9942_14	0.813	0.820	−0.007
JMC9942_15	0.732	0.827	−0.095
JMC9942_10	1.133	1.170	−0.037

Table 7
Predicted ΔT_m of the novel proposed analogues [a]

Compound	log (ΔT_m) predicted
HD33258 [b]	1.478
Berenil	1.275
12-Ftb	0.831
13-Ftb	1.056
14-Ftb	1.199
Avl11	0.857
Avl12	0.813
Avl13	0.760
Avl14	0.760
Avl15	0.802
Avl16	0.718
Avl17	1.384
Avl18	1.507
Avl19	1.443
Nequim_new1 [a]	1.635
Nequim_new2 [a]	1.447
Nequim_new3 [a]	1.555
Nequim_new4 [a]	1.270

[a] Structures proposed from the CoMFA model. Values in °C.
[b] The most complementary of the set, actual log (ΔT_m) = 1.516 (ΔT_m given in °C).

4. Conclusion

Models for the minor groove binders to recognise B-DNA have been demonstrated on the basis of molecular modelling and CoMFA studies, which provide further opportunity for structure-based design. The specificity observed in the bisamidine–DNA interaction seems to be closely related with the charge displacement all along the chain between both the bisamidine substructures. It can be understood by taking into account the strong

contribution of the straight directional hydrogen bonds established among the amidines and the available H-bond donor sites in the DNA (either in the pentose ring or in the nucleotides).

Overall, we demonstrate the robustness of the model by including an independent test set, which has validated its predictive ability.

Acknowledgements

The authors would like to thank the Conselho Nacional para o Desenvolvimento da Pesquisa (CNPq) and Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG) for financial support. One of us (A.M.O.) also wants to thank the research colleagues for their contributions to the technical aspects of this work, namely Vinícius Aniceto and Irwin Alencar Menezes for training helps in program GRID and Ramon K. Rocha for his valuable expertise in MM.

References

[1] M.M. Basselin, A.M. Badet Denisot, F. Lawrence, M. Robert Gero, Exp. Parasitol. 85 (1997) 274.
[2] M.M. Basselin, F. Lawrence, M. Robert Gero, Biochem. J. 315 (1996) 631.
[3] P.D. Walzer, Lung biology in health and disease, in: P.D. Walzer (Ed.), Pneumocystis carinii Pneumonia, 2nd ed (Revised and Expanded, v. 69), Marcel Dekker, New York, 1994.
[4] M. Kandpal, B.L. Tekwani, P.M.S. Chauhan, A.P. Bhaduri, Life Sci. 59 (1996) 75.
[5] A.G. Dirienzo, C. Van der Host, D.M. Finkelstein, P. Frame, S.A. Bozzette, K.T. Tashima, Aids Res. Human Retovir. 18 (2002) 89.
[6] R.B. Silverman, The Organic Chemistry of Drug Design and Drug Action, Academic Press, San Diego, 1992.
[7] A.J. Caldas, D.R. Silva, C.C. Pereira, P.M. Nunes, B.P. Silva, A.A. Silva, A. Barral, J. M. Costa, Rev. Soc. Bras. Med. Trop. 28 (1995) 405.
[8] H. Albrecht, AIDS 12 (1998) 2225.
[9] P.B. Carvalho, M.A.G. Arribas, E.I. Ferreira, Br. J. Pharm. Sci. 36 (2000) 69.
[10] D. Matesoi, I. Kittler, A. Bell, E. Unger, G. Lober, Biochem. Mol. Biol. Int. 36 (1996) 123.
[11] K.J. Edwards, T.C. Jenkins, S. Neidle, Biochemistry 31 (1992) 7104.
[12] C.A. Montanari, M.S. Tute, A.E. Beezer, J.C. Mitchell, J. Comp.-Aid. Mol. Des. 10 (1996) 67.
[13] R.S. Bornstein, J.W. Yarbro, J. Surg. Oncol. 2 (1970) 393.
[14] W.E. Gutteridge, J. Protozool. 16 (1969) 306.
[15] U. Bachrach, S. Brem, S.B. Wertman, L.F. Schnur, C.C. Greenblatt, Exp. Parasitol. 48 (1979) 464.
[16] A.J. Bitonti, J.A. Dumont, P.P. McCann, Biochem. J. 237 (1986) 685.
[17] M. Waring, J. Mol. Pharmacol. 1 (1965) 1.
[18] H.G. Kaplan, C.E. Myers, J. Pharmacol. Exp. Ther. 201 (1977) 554.
[19] C.C. Dykstra, R.R. Tidwell, J. Protozool. 6 (1991) 78S.
[20] T.A. Shapiro, P.T. England, Proc. Natl. Acad. Sci. USA 87 (1990) 950.
[21] N.C. Cohen (Ed.), Guidebook on Molecular Modeling in Drug Design, Academic Press, 1996.
[22] G. Minasov, V. Tereshko, M. Egli, J. Mol. Biol. 291 (1999) 83.
[23] SYBYL 6.5 (MIPS3—IRIX 6.2), Tripos Associates, St. Louis, MO, USA, 1998.
[24] CORINA v.1.7, Computer-Chemie-Centrum, Universität Erlangen-Nürnberg, 1996.
[25] MOPAC, v. 7.0 University of Indiana, Bloomington IN 47405.
[26] MacroModel v. 5.5, Department of Chemistry, Columbia University, New York, NY 10027.
[27] S.J. Weiner, P.A. Kollman, D.A. Case, U.C. Singh, C. Ghio, G. Alagona, S. Profeta, P. Weiner, J. Am. Chem. Soc. 106 (1984) 765.
[28] M. Orozco, C.A. Laughton, S. Neidle, C.H. Schwalbe, M.F.G. Stevens, Anti-Cancer Drug Des. 5 (1990) 243.
[29] C.A. Montanari, J.O. Trent, T. Jenkins, J. Braz. Chem. Soc. 9 (1998) 175.
[30] R. Lavery, K. Zakrzewska, B. Pullman, J. Biomol. Struct. Dyn. 3 (1986) 1155.
[31] Brookhaven Data Bank, http://rutgers.rcsb.org/pdb.
[32] F.C. Bernstein, T.F. Koetzel, G.J.B. Willians, E.F. Meyer, M.D. Brice, J.R. Rodgers, O. Kennard, K. Shimanouchi, M.J. Tasumi, J. Mol. Biol. 112 (1977) 535.
[33] B. Pullman, Molecular mechanics of specificity in DNA—antitumour drug interactions, in: Advances in Drug Research, vol. 18, Academic Press, London, 1989.
[34] S. Mazur, F.A. Tanious, D. Ding, A. Kumar, D.W. Boykin, I.J. Simpson, S. Neidle, W.D. Wilson, J. Mol. Biol. (2000) 321.
[35] G.M. Blackburn, M.J. Gait, R.I. Gumpart, Dynamics of nucleic acid structures, in: G.M. Blackburn, M.J. Gait (Eds.), Nucleic Acids in Chemistry and Biology, 2nd ed, Oxford University Press, 1996.
[36] P.R. Andrews, D.J. Craik, J.L. Martin, J. Med. Chem. 27 (1984) 1648.
[37] Tsar(TM) v. 3.0, Oxford Molecular Ltd., UK, 1997.
[38] R.D. Cramer, D.E. Patterson, J.D. Bunce, J. Am. Chem. Soc. 110 (1988) 5959.
[39] A. Czarny, D.W. Boykin, A.A. Wood, C.M. Nunn, S. Neidle, M. Zhao, D.W. Wilson, J. Am. Chem. Soc. 117 (1995) 4716.
[40] W.P. Jencks, Hydrophobic forces, in: Catalysis in Chemistry and Enzymology, Dover Publisher, New York, 1969.
[41] M.O. Neto, Quim. Nova 12 (1989) 248.
[42] I. Francesconi, W.D. Wilson, F.A. Tanious, J.E. Hall, B.C. Bender, R.R. Tidwell, D. McCurdy, D.W. Boykin, J. Med. Chem. 42 (1999) 12.
[43] P.D. Walzer, J. Foy, J. Runck, P. Steele, M. White, R.S. Klein, B.A. Otter, R.J. Sundberg, Ant. Agents Chemother. 38 (1994) 2572.