

Design of new DNA-interactive agents by molecular docking and QSPR approach

Antonino Lauria, Marco Tutone, and Anna Maria Almerico*

*Dipartimento Farmacochimico, Tossicologico e Biologico
Università di Palermo, Via Archirafi 32, 90123 Palermo, Italy
E-mail: almerico@unipa.it*

DOI: <http://dx.doi.org/10.3998/ark.5550190.0011.b02>

Abstract

The design of new series of pyrrolo-pyrimidine derivatives, further annelated with a third heterocycle of different size, which also present several chain shape moieties of variable length and with different physico-chemical character, is reported. In this contribution we showed that the combination of docking-based and QSPR-based methods could lead to good models for ligand-DNA interaction prediction. By means of these computational approaches on 360 proposed inhibitors, we were able to select the most promising candidates as DNA-interactive drugs potentially endowed with antitumor activity.

Keywords: DNA-interactive agents, pyrrolo-pyrimidine, molecular docking, QSPR

Introduction

DNA represents a traditional target for chemotherapeutic intervention in human cancers, especially for those where high proliferation rates of some tumor cell types have resulted in sensitivity to drugs, which block replication and transcription of their DNA.¹

Molecular recognition of DNA by small molecules is a fundamental problem in drug design. Polycyclic heterocycles having a planar structure can be effective pharmacophore moieties for DNA-interactive drugs because they can insert between the stacked base paired oligonucleotides. Moreover, if they bear suitable side chains, further interactions of these ligands with the other important architectural feature of DNA, its minor groove, can be envisaged.

For many years we have been interested in the design and synthesis of planar polyaromatic heterocycles, containing pyrrole or indole moiety, as good DNA intercalating candidates, endowed with anticancer activity.²⁻⁵ These classes of compounds can be considered bioisosters of the antitumor drugs of the actinomycins family, the most representative of which, Dactinomycin (Actinomycin D), has been extensively investigated.⁶ Its biological activity

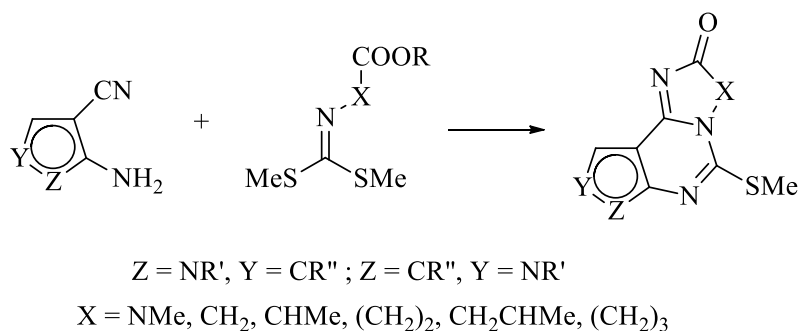
appears to depend on the very slow rate of dissociation of the complex between DNA and the drug, which reflects the intermolecular hydrogen bonds, the planar interaction between the purine rings and the chromophore, the phenoxazine moiety, and the several van der Waals interactions between the side chain and the DNA.

In connection with these studies, we have recently reported the design, synthesis and biological activity of a series of 4-substituted indolo[3,2-*e*][1,2,3]triazolo[1,5-*a*]pyrimidine derivatives.⁷ Compounds, selected on the basis of the docking studies and suitably synthesized, showed antiproliferative activity against each type of tumor cell lines investigated, generally in the low micromolar range. The molecular modeling results correlated well with the experimental activity data and the choice of suitable chains was theoretically supported: a more negative ΔG_{bind} value was mirrored by a higher inhibitory activity. In that context we have demonstrated how the introduction of the side chains was determinant for the increase of the biological activity.

Now we intend to extend this approach to different polycyclic scaffolds to explore how different structural features into the planar portion can affect the DNA binding capability. Therefore we report herein molecular modeling studies on pyrrolo-pyrimidine derivatives (both [2',3':4,5] and [3',4':4,5] condensed), further annelated with a third heterocycle of different size, which also present several chain shape moieties of variable length and with different physico-chemical character. Moreover QSPR models were also developed to further support the proposed mechanism of action of these series of inhibitors.

Results and Discussion

The structures of the designed compounds, reported in Figure 1, show that either the size of the third ring (5, 6, 7 atoms) was varied as well as the pyrrole-pyrimidine junction (3,4-*e* or 2,3-*e*).



Scheme 1. Synthetic approach to the tricycles.

The polycondensed tricycles were chosen having also in mind the easiness of synthetic methodology. In fact all these compounds could be accessible from known (and already studied

by us⁸⁻¹⁰) 2- and 3-aminopyrroles, ortho-cyano substituted, and the versatile BMMA (N-(BisMethylthio)Methylenamino Acids) according to the procedure already successfully employed in previous work^{11,12} (Scheme 1).

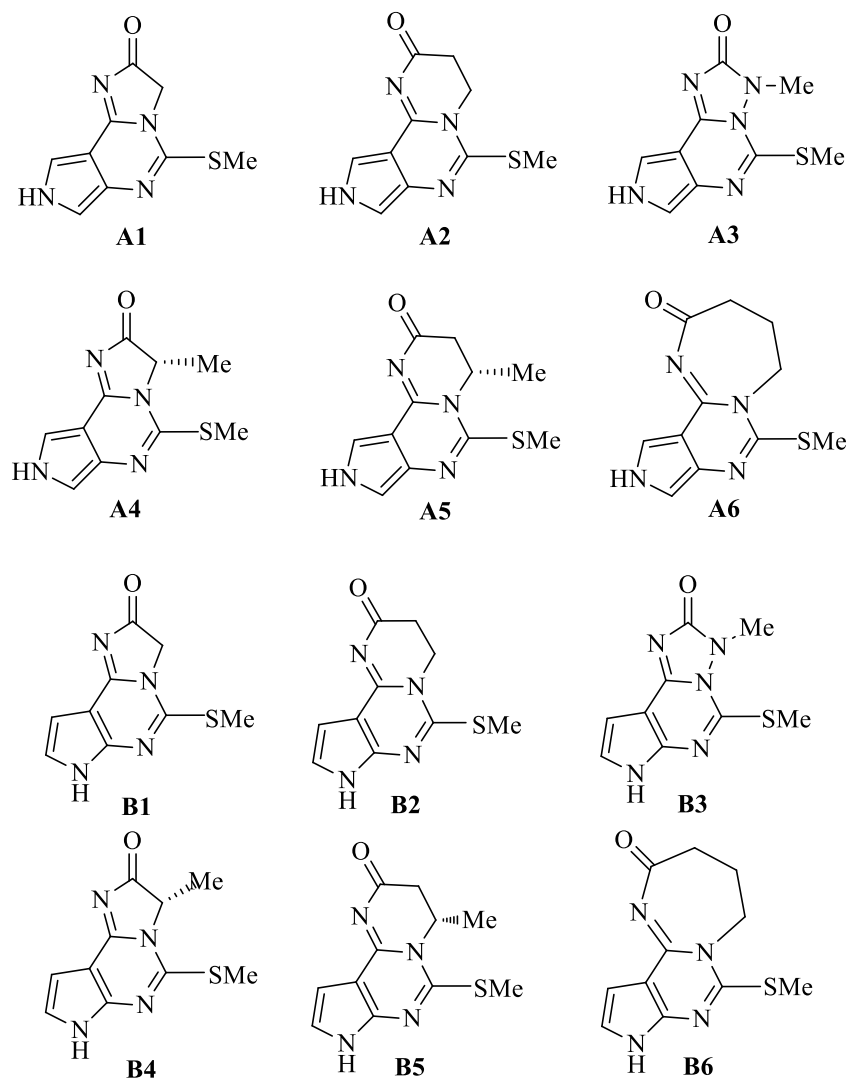
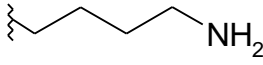
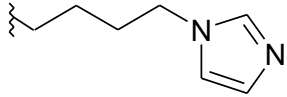
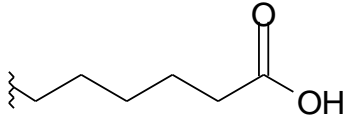
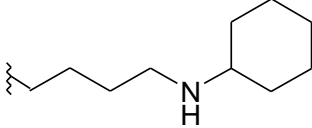
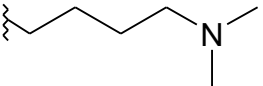
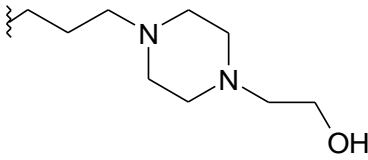
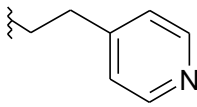
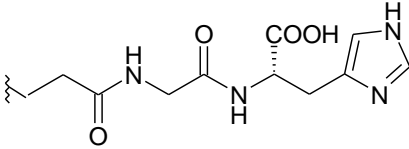
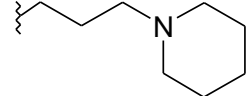
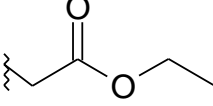
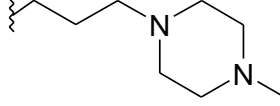
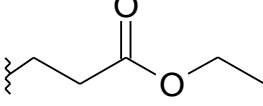
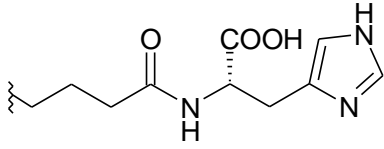
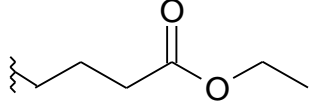
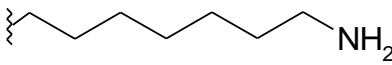
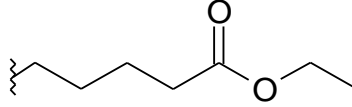
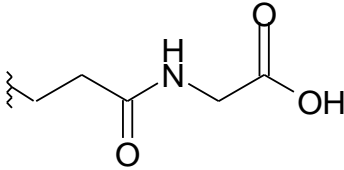
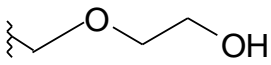
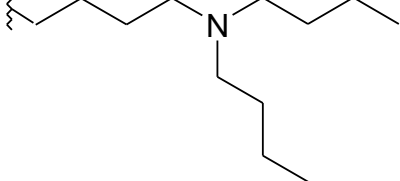
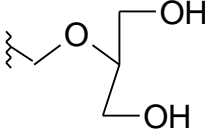


Figure 1. Structures of the selected scaffolds.

The twenty chains **a-t**, reported in Table 1, were selected and combined with the ring skeletons to generate the various ligands, presenting the substituents linked to the pyrrole NHs. Moreover a set of compounds of type **C** was also investigated to explore how simple substitutions on the base nuclei (such as the introduction of methyl groups) can influence the binding capability (Figure 2).

Table 1. Structures of the selected side chains

a		k	
b		l	
c		m	
d		n	
e		o	
f		p	
g		q	
h		r	
i		s	
j		t	

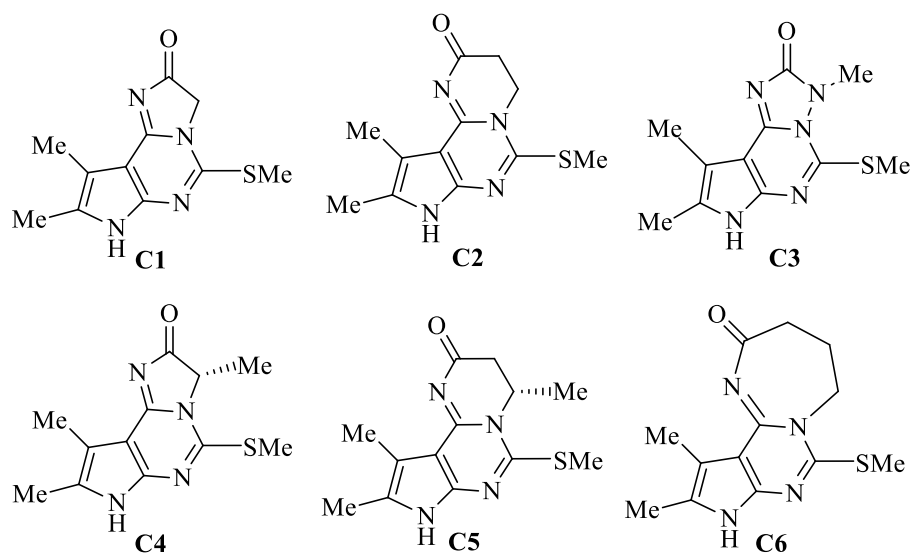


Figure 2. Structures of the selected dimethyl derivatives.

The PDB was searched for DNA fragments bound with intercalators and the structure 1DSC (an octamer complexed with Actinomycin D, already successfully used by us⁷ for this type of study) was selected. The original ligand was removed and the DNA sequence was utilized for the docking experiments. The 100 independent docking runs carried out for each ligand generally converged to a small number of different positions ("clusters" of results differing by less than 1.5 Å rmsd). Generally, in the standard Autodock output, the top clusters (i.e. those with the most favorable ΔG_{bind}) are also associated with the highest frequency of occurrence which suggests a good convergence behavior of the search algorithm (see materials and methods section for details). An example of the resulting clusters is depicted in Figure 3. These experiments allowed to estimate the interaction energies between DNA fragment and the different ligands. Thus the *in silico* screening of the selected 360 structures allowed us to predict the binding free energy of the lowest energy docked structure for each ligand considered in this study (values reported in Table 2) and to evidence the type of interaction with DNA (intercalating mode or groove binding mode).

The calculated free energy of binding was found in the range -3.88 to -19.68 Kcal/mol. It results immediately evident that generally all the designed compounds have shown high capability to bind DNA, with few exceptions for scaffolds belonging only to the **A3**, **A5**, **C3** series, especially when combined with chains of type **d**, **e**, or **f**. The data become more interesting if we consider that in the same experiment the value obtained for Actinomycin D was -10.37 Kcal/mol. All the classes of compounds are therefore able to bind DNA tighter than the reference drug, as also testified by the average values calculated either taking into account modulation of the chain on the selected skeletons (Table 2, last row) and the influence of the polycondensed system on each side chain (Table 3). In particular the compounds able to form

more stable complex with DNA have shown ΔG values from 1.5 to 1.9 times more negative than Actinomycin D, see for example **B1n** ($\Delta G = -19.68$ Kcal/mol), **C4i** ($\Delta G = -19.62$ Kcal/mol). Moreover these last two, together with **B6i** and **C2n**, resulted the best DNA binders, with calculated ΔG always more negative than -19.0 Kcal/mol.

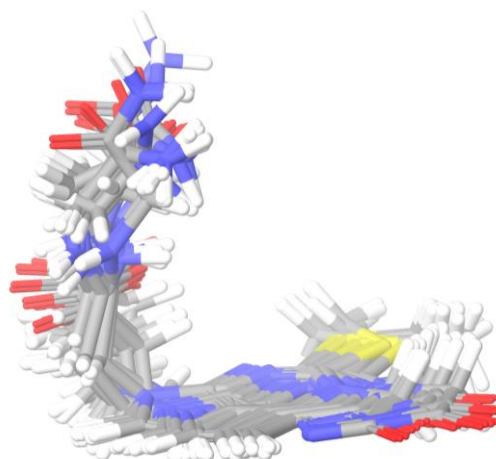


Figure 3. Cluster of best conformations.

An analysis of the chemical features of the different ring skeleton evidenced that always the shift of the NH from pyrrolo[3,4-e]pyrimidine (class **A**) to pyrrolo[3,2-e]pyrimidine (class **B**) lowered the DNA binding capability (compare column **A1** vs **B1**, **A2** vs **B2**, and so on).

Table 2. Calculated ΔG_{bind} (Kcal/mol). In bold groove binding mode, yellow cells intercalating binding mode

	A1	A2	A3	B1	B2	B3	C1	C2	C3
a	-12.19	-13.46	-13.38	-11.06	-11.69	-12.25	-12.68	-14.42	-12.01
b	-14.25	-15.45	-16.02	-13.42	-14.99	-14.15	-15.60	-18.20	-15.33
c	-12.85	-13.76	-13.72	-12.22	-13.02	-11.78	-12.54	-14.33	-12.55
d	-12.91	-12.68	-3.88	-11.56	-11.69	-12.34	-11.89	-13.39	-8.44
e	-12.86	-13.38	-4.09	-12.19	-13.50	-11.88	-12.60	-14.59	-8.47
f	-14.01	-14.07	-5.02	-11.22	-13.89	-11.81	-9.22	-11.62	-8.73
g	-17.56	-18.58	-9.36	-13.90	-13.14	-12.05	-17.50	-13.30	-12.06
h	-13.30	-14.18	-15.08	-12.04	-13.42	-12.45	-13.20	-15.17	-12.24
i	-16.22	-17.07	-15.86	-16.83	-17.22	-17.06	-16.31	-18.66	-17.94
j	-13.20	-14.77	-14.46	-11.52	-11.39	-15.11	-12.12	-13.85	-11.38
k	-13.38	-11.35	-14.15	-12.09	-12.70	-12.62	-9.81	-16.06	-7.90

I	-13.82	-10.43	-14.63	-12.94	-15.71	-12.58	-8.14	-14.08	-7.59
---	--------	--------	--------	--------	--------	--------	-------	--------	-------

Table 2. Continued

	A1	A2	A3	B1	B2	B3	C1	C2	C3
m	-12.98	-10.92	-12.37	-10.59	-11.18	-11.04	-10.92	-13.45	-9.14
n	-14.06	-12.77	-17.47	-19.68	-16.76	-13.56	-18.82	-19.07	-13.02
o	-12.49	-13.02	-14.38	-13.50	-13.26	-13.66	-13.32	-13.71	-12.65
p	-13.85	-13.84	-14.05	-13.57	-13.81	-13.01	-13.41	-13.43	-13.77
q	-13.98	-15.12	-15.54	-13.87	-14.17	-13.96	-13.96	-15.32	-14.43
r	-13.66	-15.15	-15.46	-14.72	-13.59	-14.80	-14.73	-14.51	-15.63
s	-15.56	-16.34	-16.41	-14.10	-15.20	-14.96	-15.94	-15.45	-15.79
t	-16.19	-17.82	-17.20	-16.34	-17.55	-17.18	-16.22	-17.54	-15.87
average	-13.97	-14.21	-13.13	-13.37	-13.89	-13.41	-13.45	-13.64	-12.25
	A4	A5	A6	B4	B5	B6	C4	C5	C6
a	-14.08	-17.30	-13.48	-12.38	-12.23	-13.55	-13.04	-12.73	-14.63
b	-16.44	-16.64	-16.34	-12.14	-15.71	-18.04	-14.15	-15.77	-15.66
c	-14.10	-14.08	-13.99	-11.74	-14.33	-14.48	-12.13	-14.44	-14.04
d	-12.73	-4.91	-9.11	-12.59	-13.55	-13.74	-12.95	-13.24	-14.57
e	-13.13	-5.76	-9.77	-11.60	-13.14	-13.05	-12.59	-13.65	-13.76
f	-12.73	-5.19	-8.22	-11.94	-10.95	-12.86	-11.61	-12.63	-11.33
g	-18.14	-10.78	-12.59	-10.35	-6.86	-12.67	-12.53	-11.50	-11.93
h	-15.12	-15.38	-13.84	-12.45	-13.57	-14.40	-12.78	-14.53	-15.97
i	-18.22	-17.58	-17.78	-17.48	-16.25	-19.24	-19.62	-17.51	-17.88
j	-14.47	-7.55	-16.26	-10.76	-14.64	-13.89	-13.03	-13.67	-12.33
k	-14.34	-6.99	-9.13	-12.93	-13.97	-15.41	-13.76	-15.13	-15.73
l	-15.14	-6.57	-9.47	-13.64	-13.28	-14.62	-12.88	-14.12	-16.29
m	-14.24	-6.81	-8.24	-10.50	-13.83	-13.42	-11.17	-11.75	-12.05
n	-14.60	-10.34	-13.07	-14.12	-10.96	-12.56	-13.84	-10.74	-18.92
o	-13.84	-14.31	-14.57	-13.32	-14.38	-14.98	-14.29	-15.61	-16.13
p	-14.25	-14.43	-15.24	-13.07	-12.92	-14.61	-14.42	-15.43	-14.25
q	-14.99	-15.21	-16.57	-14.23	-15.05	-15.15	-14.24	-14.99	-16.55
r	-15.81	-16.29	-16.55	-14.29	-14.40	-15.02	-14.90	-15.06	-15.77
s	-15.52	-16.62	-16.98	-14.40	-16.16	-16.47	-16.71	-16.53	-17.13
t	-18.34	-18.08	-17.77	-17.26	-18.16	-18.73	-16.57	-16.39	-17.84
average	-15.01	-12.04	-13.45	-13.06	-13.72	-14.84	-13.86	-14.27	-15.14

However, by comparison of the behavior of the basic classes of compounds, it becomes evident that the introduction of methyl groups in the pyrrole moiety enhances the binding capability. In fact, with the sole exception of **C3** type derivatives, less tighter DNA binders (cfr **C3** vs **B3** column) and some example of **C6** series (less than a quarter of the set), generally the dimethyl scaffold furnishes ΔG values more negative than the corresponding **B** series. These data are also evident if the average value for all the chains is considered [see for example **B1** vs **C1** ($\Delta G = -13.37$ Kcal/mol vs $\Delta G = -13.45$ Kcal/mol) up to $\Delta G -14.84$ Kcal/mol vs $\Delta G = -15.14$ Kcal/mol as in the case of **B6** vs **C6**]. As a result, generally for derivatives of class **C** the stability of the complexes with DNA is of the same order of magnitude of that calculated in the case of class **A**.

Table 3. Influence of the polycondensed system on each side chain (average rows ΔG_{bind} values)

Chain type	average	chain type	average
a	-13.14	k	-12.64
b	-15.46	l	-12.55
c	-13.34	m	-11.37
d	-11.45	n	-14.69
e	-11.67	o	-12.44
f	-10.95	p	-13.96
g	-13.04	q	-14.85
h	-13.84	r	-15.02
i	-17.49	s	-15.90
j	-13.02	t	-17.28

If the DNA-binding capability is considered from the point of view of the chain types, it is possible to evidence that the introduction of **i** and **t** moiety always gives rise to favorable interactions, regardless the position and the size of the tricyclic system. The compounds bearing those two chains are always found in the first three rankings and the calculated averaged ΔG are the more negative (-17.49 and -17.28 Kcal/mol, respectively). Also the presence of chains **b** and **s**, followed by **r** and **n** types, resulted of primary importance in modulating the binding capability at least for more than half of the designed structures, and in any case gave rise to the most effective binders.

From a structural point of view it seems that the presence in the side chain of a ring with basic nitrogen character, as in the case of **d**, **e**, **f**, or **m** types, is detrimental for the binding capability, whereas an aliphatic amino group is much more tolerated (see for example **a**, **c**, **h**, **j**

chains). On the contrary the presence of carboxylate functions and of amino acid moieties favors a more tighter binding, as shown in the case of chains **b**, **r**, **i**, **n**. All these results are in agreement with the literature data for DNA-interactive drugs, such as for example mitoxantrone, for which the specific molecular recognition also depends on the base/acid character of the chains laying along the grooves.¹³ Noteworthy as well is the behavior of derivatives presenting the chains **t** and **s** typical of the antiviral drugs acyclovir and ganciclovir.

By analyzing our data with respect to the ligand binding site, the docking studies revealed two different binding modes (Figure 4). The majority of structures were found to assume a position with the chromophore portion intercalated between GC base pairs whereas the side chain lies along the minor groove. A consistent number of derivatives (165) however were shown to bind essentially aligned along the minor groove.

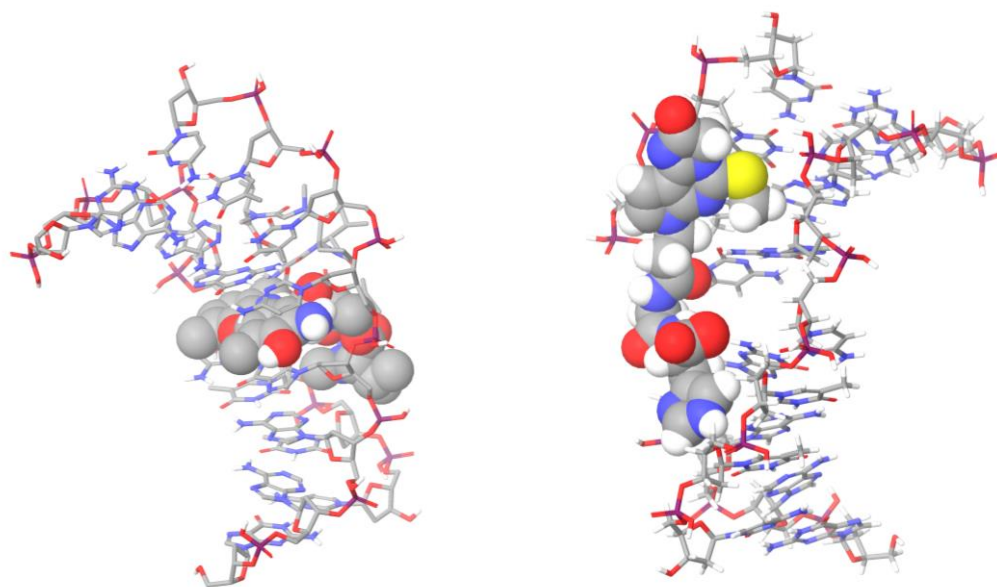


Figure 4. Left: intercalating binding mode. Right: minor groove binding mode.

Again if we consider the nature of the heterocyclic moiety as driving force for the chosen binding mode, it appears evident that scaffolds of type **A** strongly direct to the intercalation to DNA, especially when the ring size is 5/6 atoms. In fact only in 2-4 cases out of 20 the groove binding mode was evidenced. But the introduction of a methyl group into the pyrimidinone moiety, as in classes of type **5**, or ring enlargement to give the benzodiazepine, as in types **6**, favors both binding modes that therefore become driven by the character of the side chain. Moreover these systems probably present steric hindrance to the insertion between the DNA base pairs.

With respect to the nature of side chains, in the case of **n** and **d** types, the complex is forced to align along the DNA minor groove, and only 4 or 5 derivatives out of 18 show intercalating ability. On the contrary, in the case of **e**, **m** and **j** chain types is the intercalating mode preferred, since the alternative binding to the groove is limited to 2-4 examples out of the 18 considered.

The docking results allow us to evaluate the affinity of each ligand towards the DNA fragment, but they do not give us any information on the type of possible electronic interactions involved in such complexes. For this reason the predicted binding affinity values (ΔG_{bind}) of the 360 ligands, derived from the molecular docking, were taken as starting point for a QSPR analysis. Therefore, we submitted the results obtained for every single complex to structure-properties analysis by using CODESSA PRO software. The results of this approach consist in equations in which it is possible to appraise type and weight of the descriptors involved in the model built for each complex. In this case the obtained equations (*A1 – C6*), showing the dependency of ΔG as function of five descriptors, are listed in Table 4, whereas the molecular descriptors legend is reported in the Supplementary Materials, as Table S1.

Table 4. QSPR equations

<i>A1</i>	$\Delta G = -1006.08 + 9.81325 D_1 + 105.283 D_2 + 2.1532 D_3 + 0.535776 D_4 + 0.0152141 D_5$
<i>A2</i>	$\Delta G = 237.495 + 0.604662 D_6 + 21.3836 D_7 + 11.7714 D_8 - 35.6103 D_9 + 63.3637 D_{10}$
<i>A3</i>	$\Delta G = -1654.28 + 8.24508 D_{11} + 73.0018 D_{12} + 39.3151 D_{13} + 868.794 D_{14} - 33.6615 D_{15}$
<i>A4</i>	$\Delta G = 6.43958 + 0.0398856 D_{16} - 0.22963 D_{17} + 0.847178 D_{18} + 0.398599 D_{19} + 10.1329 D_{20}$
<i>A5</i>	$\Delta G = -2900.34 + 2.20996 D_{21} - 1.06686 D_{22} + 349.629 D_{23} - 6.87304 D_{24} + 28.2982 D_{25}$
<i>A6</i>	$\Delta G = -123717 + 5.60404 D_{11} - 0.506115 D_{25} + 734.538 D_{26} + 1191.42 D_{27} + 60.256 D_{28}$
<i>B1</i>	$\Delta G = 506.258 - 10.952 D_{29} + 0.642268 D_{30} + 73.0698 D_{31} - 35.1541 D_{32} - 0.936513 D_{33}$
<i>B2</i>	$\Delta G = -1836.71 + 1.2025 D_{33} + 79.8471 D_{34} - 28.6335 D_{35} + 291.592 D_{36} - 0.0529609 D_{37}$
<i>B3</i>	$\Delta G = 638.546 + 0.0461712 D_{16} - 0.0868716 D_{38} + 0.0161736 D_{39} + 0.0342804 D_{40} - 4.03483 D_{41}$
<i>B4</i>	$\Delta G = -1649.68 + 321.404 D_{42} - 87.9306 D_{43} + 60.2046 D_{44} + 278.056 D_{45} + 288.849 D_{46}$
<i>B5</i>	$\Delta G = -427.521 - 2.95153 D_{46} - 4.19156 D_{47} - 126.237 D_{48} + 662.138 D_{49} + 11.9784 D_{50}$
<i>B6</i>	$\Delta G = -378.78 + 45.6003 D_{50} + 0.0266912 D_{51} + 0.233501 D_{30} + 0.860957 D_{52} + 476.283 D_{14}$
<i>C1</i>	$\Delta G = 136.39 - 0.0671725 D_{53} - 0.172129 D_{54} - 119.016 D_{55} - 1.07812 D_{56} - 208.309 D_{57}$
<i>C2</i>	$\Delta G = 39.8479 - 0.085344 D_{58} - 4.15954 D_{59} + 92.8765 D_{10} + 0.123418 D_{60} - 0.821015 D_{61}$
<i>C3</i>	$\Delta G = -122.62 + 4.39165 D_{11} - 6.64417 D_{24} + 0.0764258 D_{62} - 424.707 D_{63} - 57.0088 D_{15}$
<i>C4</i>	$\Delta G = 16.5996 - 76.2123 D_{64} + 353.729 D_{65} + 0.0383017 D_{66} - 0.0667068 D_{67} - 29.3026 D_{68}$
<i>C5</i>	$\Delta G = 16.5042 + 1.44595 D_{69} - 16.4673 D_{70} + 142.455 D_{71} - 0.0730432 D_{37} - 12.093 D_{72}$
<i>C6</i>	$\Delta G = 292.037 - 1.28335 D_{73} + 6.19829 D_{74} - 65.5532 D_{55} - 58.9236 D_{75} + 11.8015 D_{76}$

The quality of the models, obtained according to the procedure described in the section material and methods, are summarized in Table 5. Very satisfactory correlations were obtained in each case ($R^2 \geq 0.91$; $Q^2 \geq 0.89$), with the sole exception of *B4* and *C2* equations. Only seventy six descriptors in total were included into the 18 equations. Among them, some contribute to a larger extent to the models and are present in more than one equation. In particular, according to the classification reported in Table S1, for these type of ligands the more important descriptors resulted:

- Among the constitutional descriptors: the total number of rings and the average information content;

- Among the geometrical descriptors: the inertia moment and the molecular area;
- Among the electrostatic descriptors: the partial charge (Zefirov) for all atom types;
- Among the quantum-chemical descriptors: the heat of formation of the molecule and the molecular orbital energies.

Table 5. Quality of the QSPR models

Equations	R ²	Q ²	F	s ²
A1	0.98	0.96	114	0.061
A2	0.98	0.91	122	0.129
A3	0.99	0.97	174	0.368
A4	0.97	0.94	84	0.123
A5	1.00	0.99	525	0.155
A6	0.99	0.99	402	0.095
B1	0.99	0.97	175	0.104
B2	0.94	0.89	40	0.302
B3	0.96	0.94	75	0.142
B4	0.93	0.66	35	0.367
B5	0.97	0.95	100	0.206
B6	0.95	0.89	49	0.272
C1	0.99	0.96	171	0.152
C2	0.91	0.71	25	0.532
C3	0.99	0.96	187	0.189
C4	0.98	0.97	185	0.078
C5	0.99	0.98	246	0.048
C6	0.97	0.95	103	0.162

R²: square of correlation coefficient; Q²: square of correlation coefficient obtained by performing leave-one-out (LOO) cross-validation; F: Fisher criteria; s²: square of the standard deviation.

Moreover, from the analysis of the descriptors used to describe the binding on each single fragment and also in the whole, it is possible to get insights on the type of interactions involved in the formation of the complexes. The statistical distribution of all descriptors included in the models (Figure 5) showed a clean contribution of the quantum-chemical and electrostatic type. On the other hand constitutional and topological effects appear to be not determinant for the interactions DNA-ligand.

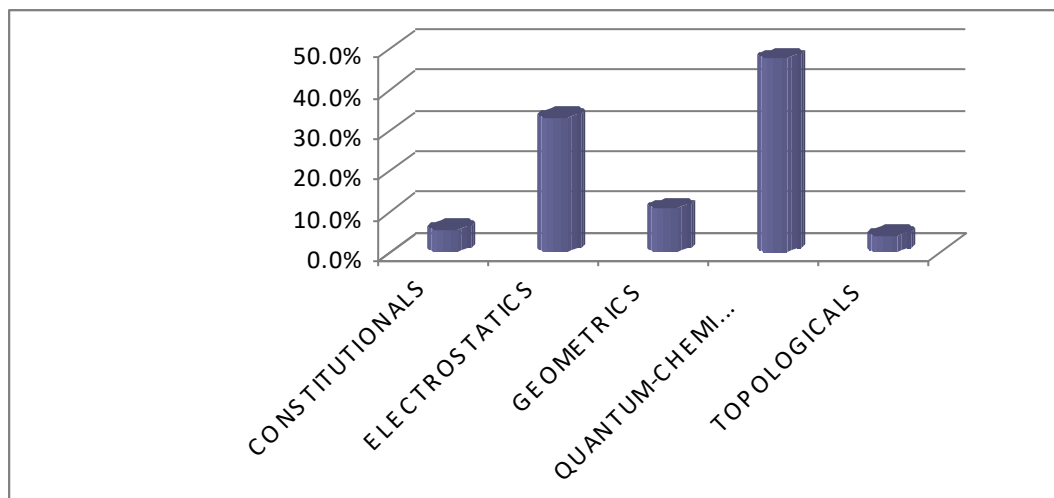


Figure 5. Statistical distributions of descriptors.

Conclusions

In this paper we described how docking experiments allowed to estimate the interaction energies between DNA fragment and the 360 different ligands, and to evaluate the binding mode.

From our results it was possible to recognize two different kinds of binding: an “intercalating mode” and a “groove mode”. The nature, size and physico-chemical properties of the designed ligands strongly influence the type of binding.

In this contribution we showed that the combination of docking-based and QSPR-based methods could lead to good models for ligand-macromolecule interaction prediction. In particular by docking procedure it was possible to calculate ΔG_{bind} values for a set of DNA-ligand complexes. A QSPR analysis allowed the characterization of the type of interactions in such complexes.

The more interesting derivatives of the new series will be synthesized and further investigated in DNA-binding assays. The studies reported herein provided reliable information on the capability of new ligands to interact with DNA allowing us to quickly select the more interesting derivatives in a series of congeners.

Experimental Section

Molecular docking

It was performed with AutoDock, version 3.0.5.¹⁴ It combines a rapid energy evaluation through precalculated grids of affinity potentials with a variety of search algorithms to find suitable binding positions for a ligand on a given macromolecule. While the protein is required to be

rigid, the program allows torsional flexibility in the ligand.¹⁵ The Protein Data Bank¹⁶ was searched for DNA fragments bound with intercalators and the structure 1DSC (an octamer complexed with Actinomycin D) was selected. Before performing docking experiment the ligand structure was removed, and Gasteiger-Marsili charges¹⁷ were assigned. Docking to DNA was carried out using the empirical free energy function and the Lamarckian genetic algorithm, applying a standard protocol, with an initial population of 100 randomly placed individuals, a maximum number of 1.5×10^6 energy evaluations, a mutation rate of 0.02, a crossover rate of 0.80, and an elitism value of 1. Proportional selection was used, where the average of the worst energy was calculated over a window of the previous 10 generations. For the local search, the so-called pseudo-Solis and Wets algorithm was applied using a maximum of 300 iterations per local search. The probability of performing local search on an individual in the population was 0.06, and the maximum number of consecutive successes or failures before doubling or halving the local search step size was 4. 100 independent docking runs were carried out for each ligand. Results differing by less than 1.0 Å in positional root-mean-square deviation (rmsd) were clustered together and represented by the result with the most favorable free energy of binding. The structures of the ligands (Figures 1, 2 and Table 1) were generated with Ligprep software¹⁸ and OPLS_2005 force field was used for the 3D optimization. The structures were setup for docking as follows: polar hydrogens were added using the PROTONATE utility (written by D.A. Case, K. Cross, and G.P. Gippert and distributed with AutoDock). Solvation parameters were added to the final DNA fragment file using the ADDSOL utility of AutoDock. The grid maps, representing the protein in the actual docking process, were calculated with AutoGrid. The grids (one for each atom type in the ligand, plus one for electrostatic interactions) were chosen to be sufficiently large to include significant portions of the minor groove. The dimensions of the grids were thus $60 \times 60 \times 60$, with a spacing of 0.375 Å between the grid points.

QSPR

Optimized atomic coordinates of ligands were used as an input data for the CODESSA PRO program,¹⁹ where calculation and selection of descriptors for QSPR were carried out.

CODESSA PRO (COMprehensive DEScriptors for Structural and Statistical Analysis) is a multi-purpose program providing various methods for statistical analysis of experimental data such as Partial Least Squares, (Multiple) Linear and Non-Linear Regressions, and Principal Components Regression. A set of 615 molecular descriptors derived from geometrical and quantum-chemical structures was computed in our study, as available in this program. All descriptors were divided into five groups: constitutional, topological, geometric, electrostatic, and quantum-chemical. Constitutional descriptors reflect only the molecular composition of the compounds as the number of atoms, the number of bonds and the molecular weight, and so on. Topological descriptors describe the atomic connectivity in a molecule. Geometric descriptors are calculated from the three-dimension atomic coordinates of the molecule, and comprise moments of inertia, shadow indices, molecular volume, and molecular surface area-like descriptors. Electrostatic descriptors reflect characteristics of the charge distribution in the molecule, calculated by the

Zefirov method. Quantum-chemical descriptors add important information to the conventional descriptors. They are divided into three subgroups: charge distribution-related, valence-related and quantum mechanical energy-related descriptors.

BMLR (Best MultiLinear Regression) analysis was applied for developing QSPR models. For selection of descriptors the heuristic method as available in CODESSA PRO was used. First, descriptors with missing or constant values for the set of structures are discarded from the original set. Further selection of descriptors is accomplished on the basis of the statistical parameters: R^2 , F-test and t-test for the one-parameter equation with the descriptors. The default values which were kept constant throughout the calculations were set as follows: the descriptor is eliminated if it does not meet the criteria for one-parameter equation (F-test value below 1.0; R^2_{\min} less than 0.001 and t-value below 0.01). Default value for the intercorrelation of descriptors was set to $R^2_{\max}=0.90$. After the preselection of descriptors, BMLR models are developed in a stepwise procedure. Thus, descriptors and correlations are ranked according to the values of the F-test and the correlation coefficient. Starting with the top descriptor from the list, two-parameter correlations are calculated. In the following steps new descriptors are added one-by one until the pre-selected number of descriptors in the model is achieved. The final result is a list of best models according to the values of the F-test and correlation coefficient.

References

1. Kumar, R.; Lown, J. W. *Eur. J. Med. Chem.* **2005**, *40*, 641.
2. Lauria, A.; Diana, P.; Barraja, P.; Almerico, A. M.; Cirrincione, G.; Dattolo, G. *J. Heterocycl. Chem.* **2000**, *37*, 747.
3. Lauria, A.; Diana, P.; Barraja, P.; Montalbano, A.; Cirrincione, G.; Dattolo, G.; Almerico, A. M. *Tetrahedron* **2002**, *58*, 9723.
4. Lauria, A.; Patella, C.; Diana, P.; Barraja, P.; Montalbano, A.; Cirrincione, G.; Dattolo, G.; Almerico, A. M. *Heterocycles* **2003**, *60*, 2669.
5. Lauria, A.; Patella, C.; Diana, P.; Barraja, P.; Montalbano, A.; Cirrincione, G.; Dattolo, G.; Almerico, A. M. *Tetrahedron Lett.* **2006**, *47*, 2187.
6. Bailey, S. A.; Graves, D. E.; Rill, R.; Marsch, G. *Biochemistry* **1993**, *32*, 5881.
7. Lauria, A.; Patella, C.; Dattolo, G.; Almerico, A. M. *J. Med. Chem.* **2008**, *51*, 2037.
8. Cirrincione, G.; Dattolo, G.; Almerico, A. M.; Aiello, E.; Jones, R. A.; Hinz, W. *Tetrahedron* **1987**, *43*, 5225.
9. Cirrincione, G.; Almerico, A. M.; Aiello, E.; Dattolo, G. Aminopyrroles. In: *The Chemistry of Heterocyclic Compounds Pyrroles Part II*, Jones, R. A. Ed., Wiley: New York, 1992; Vol. 48, pp 299.
10. Cirrincione, G.; Almerico, A. M.; Diana, P.; Barraja, P.; Mingoia, F.; Grimaudo, S.; Dattolo, G.; Aiello, E. *J. Heterocycl. Chem.* **1996**, *33*, 161.

11. Lauria, A.; Bruno, M.; Diana, P.; Barraja, P.; Montalbano, A.; Cirrincione, G.; Dattolo, G.; Almerico, A. M. *Bioorg. Med. Chem.* **2005**, *13*, 1545.
12. Lauria, A.; Patella, C.; Ippolito, M.; Almerico, A. M. *J. Mol. Struct. THEOCHEM* **2007**, *819*, 26.
13. Thomas, X.; Archimbaud, E. *Hematol. Cell Ther.* **1997**, *39*, 163.
14. (a) Goodsell, D. S.; Olson, A. J. *Proteins: Structure, Function and Genetics* **1990**, *8*, 195; (b) Available at Scripps. <http://www.scripps.edu/pub/olson-web/doc/autodock>.
15. Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J. *J. Comp. Chem.* **1998**, *19*, 1639.
16. Available via RCSB. <http://www.pdb.org>.
17. Gasteiger, J.; Marsili, M. *Tetrahedron* **1980**, *36*, 3219.
18. LigPrep, version 22, Schrodinger, LLC, New York, NY, 2005.
19. (a) Karelson, M.; Lobanov, V. S.; Katritzky, A. R. *Chem. Rev.* **1996**, *96*, 1027. (b) Katritzky, A. R.; Lobanov, V.S.; Karelson, M. 1994 CODESSA Reference Manual.