

Docking of indolo- and pyrrolo-pyrimidines to DNA. New DNA-interactive polycycles from amino-indoles/pyrroles and BMMA

Antonino Lauria, Patrizia Diana, Paola Barraja, Alessandra Montalbano,
Gaetano Dattolo, Girolamo Cirrincione, and Anna Maria Almerico*

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

**Dedicated to Professor Vincenzo Tortorella in the occasion of his “Fuori Ruolo” status
(received 05 Jan 04; accepted 29 Mar 04; published on the web 01 Apr 04)**

Abstract

New indolo- and pyrrolo-pyrimidines of type **1-4** were studied for their ability to form stable complexes with DNA fragments. The calculated free energies of binding were found in the range $-8.39 \div -16.72$ Kcal/mol. The docking studies revealed a common binding mode with the chromophore intercalated between GC base pairs whereas the side chain lies along the minor groove.

Keywords: Docking, DNA interaction, indolopyrimidine, pyrrolopyrimidine, aminoindoles, aminopyrroles, BMMA

Introduction

DNA represents one of the most important molecular cellular targets of several chemotherapeutic drugs. Molecular recognition of DNA by small molecules and proteins is a fundamental problem in drug design. Polycyclic heterocycles having a planar structure can be effective pharmacophore moieties of 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. Relatively little is understood at present about the mode of action at the molecular level of the majority of minor groove-interacting drugs, although there is increasing evidence that may act by directly blocking or inhibiting protein–DNA recognition.¹

Among the different classes of antitumor drugs which interact with DNA the actinomycins, a family of chromopeptide antibiotics, represent a peculiar one that combines the two above mentioned features. The most representative derivative, Dactinomycin (**Actinomycin D**), has

been extensively investigated: it binds to double-stranded DNA and the phenoxazine moiety intercalates between bases whereas the peptide substituents lie in the minor groove (Figure 1).² The biological activity appears to depend from a 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, and the numerous van der Waals interactions between the polypeptide side chain and the DNA.

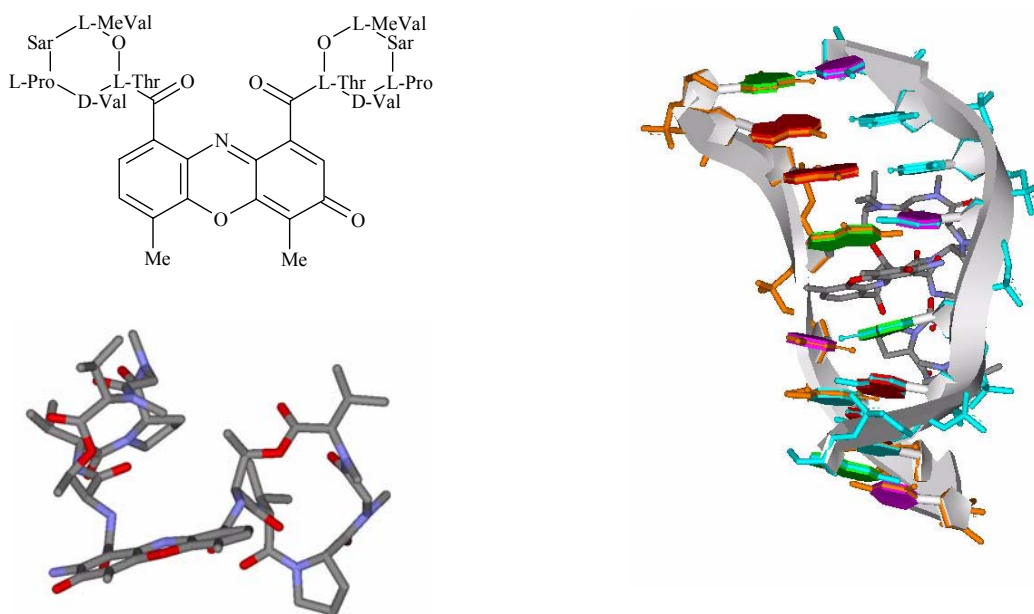
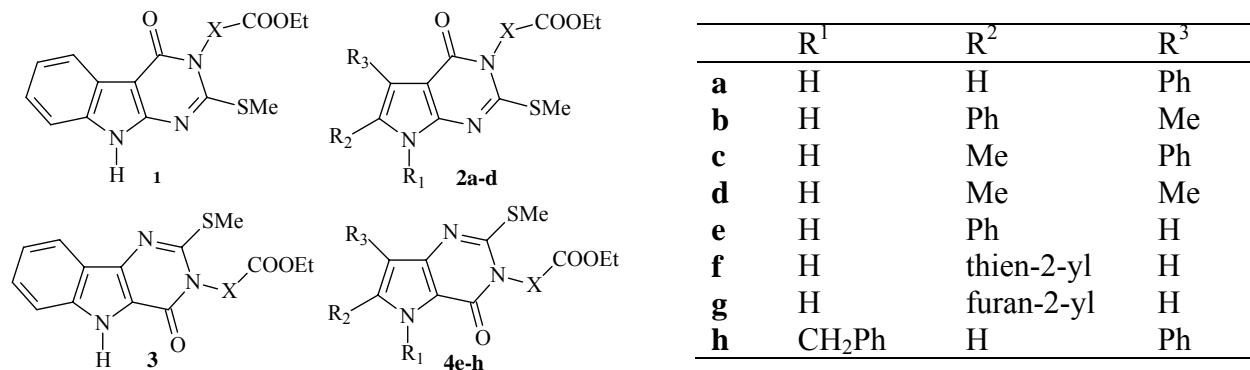


Figure 1. Actinomycin D (left) and X-Ray structure of Actinomycin D intercalated into a DNA fragment.

In connection with our researches on polycondensed nitrogen heterocycles as potential DNA-interactive drugs, we have already studied different classes of planar systems (either linear or angular) which incorporate the indole or pyrrole ring and that can be considered bioisosters of the pharmacophore of several classes of antitumor drugs including the actinomycins.³⁻⁹ Now our interest is focused on designing compounds that can form stable complexes with DNA and therefore we report docking studies on the new series of indolo- and pyrrolo-pyrimidines of type **1-4** which show a planar moiety joined to an iminoalkylcarboxylate chain of variable length, with the aim of performing a fast *in silico* screening of new potential DNA-interactive drugs.



X=CH₂, (CH₂)₂, (CH₂)₃, NMe, CHMe, CHMeCH₂

Fast computational tools are essential to most techniques for structure-based drug design. Reported procedures include docking simulation methods. These techniques can model the docking of a ligand to a target in greater detail: the ligand begins randomly outside the receptor, and explores translations, orientations, and conformations until an ideal site is found. These techniques are typically slower than other techniques, but they allow flexibility within the ligand to be modelled and can utilize more detailed molecular mechanics to calculate the energy of the ligand in the context of the putative active site. They allow medicinal chemists to investigate modifications of lead molecules suggested by the chemical intuition and expertise of organic synthetic chemists.

Results and Discussion

The starting point of all docking calculations is generally the crystal structure of a macromolecule (enzyme, DNA or RNA fragments and so on) usually obtained from a macromolecule-ligand complex. This complex may be taken from a database of compounds, such as the Cambridge Crystallographic Database¹⁰ or the Protein Data Bank (PDB).¹¹ The primary consideration in the design of a docking process has to be the development of methods which are both rapid and reasonably accurate.

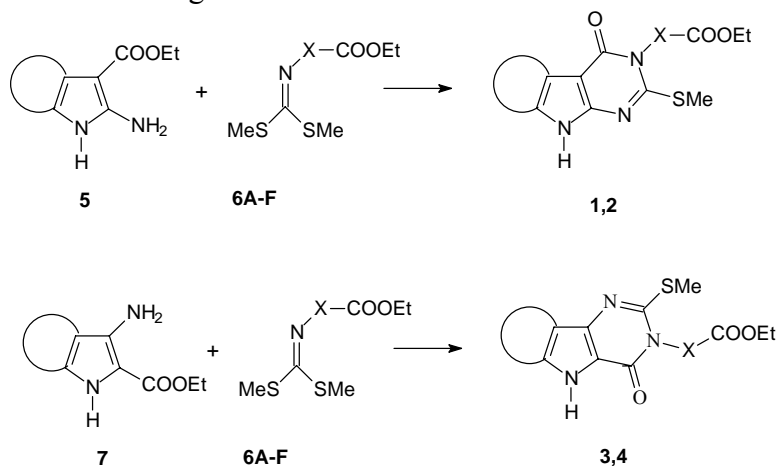
In any docking scheme two conflicting requirements must be balanced: the desire for a robust and accurate procedure, and the desire to keep the computational demands at a reasonable level. The ideal procedure would find the global minimum in the interaction energy between the substrate and the target receptor, exploring all available degrees of freedom (DOF) for the system. However, it must also run on a laboratory workstation within an amount of time comparable to other computations that a structural researcher may undertake, such as a crystallographic refinement. In order to meet these demands a number of docking techniques simplify the docking procedure. These programs can be separated functionally into roughly two parts, each somewhat independent of the other:

- Routines which determine the orientation and the position of a ligand relative to the receptor.
- Routines which evaluate (score) the stability of the complex.

Our studies were carried out, according to the procedure reported in the experimental section, by using the software package AutoDock.¹² 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 receptor. While the macromolecule is required to be rigid, the program allows torsional flexibility in the ligand. The program searches for the best conformation and for the best place of binding of the ligand within the receptor structure. Several runs of space search were carried out to find the best ligand-receptor binding arrangement. Docking to DNA fragment was performed using the Lamarckian Genetic Algorithm (LGA)¹³ and an empirical binding free energy function.

The Protein Data Bank was searched for DNA fragments bound with intercalators and the structure **1DSC** (an octamer complexed with Actinomycin D) was selected. The original ligand was removed and the DNA sequence was utilized for the docking experiments. Ten different structures of planar bicyclic or tricyclic ring systems belonging to classes **1-4** were selected. Our choice was driven by the consideration that although pyrrolo[2,3-d]pyrimidine and pyrrolo[3,2-d]pyrimidine derivatives have been extensively studied as potential pharmaceuticals,¹⁴ being related to deazaguanine, only few reports can be found in literature on indolo[2,3-d]pyrimidines and indolo[3,2-d]pyrimidines.¹⁵ In any case however nothing has been reported on interaction with DNA of these classes of compounds and derivatives of type **1-4** are unknown.

All the ligands and related substituents on the basic structures **1-4**, utilized in this study, were chosen also in view of the easiness of synthetic procedure. In fact all these compounds could be accessible from known¹⁶⁻¹⁹ 2- and 3-aminoindoles/pyrroles, ortho-ethoxycarbonyl substituted, of type **5** and **7** and the versatile **BMMA**s (N-(BisMethylthio)Methylenamino Acids) of type **6** according to the procedure already successfully employed in our laboratory on 2- and 3-aminopyrroles ortho-cyano substituted²⁰ (Scheme 1). This method allows a one pot synthesis of the condensed pyrimidine bearing the desiderated side chain.



Scheme 1. Synthetic entry to derivatives **1-4**.

These docking experiments allowed to estimate the interaction energies between DNA fragment and the different ligands. The energies used, and reported by Autodock, can be distinguished in docked energies, which include the intermolecular and intramolecular interaction energies and are used during the dockings, and predicted free energies, which include the intermolecular energy and the torsional free energy, and are reported at the end of the docking. The predicted binding free energy of the lowest energy docked structure for each ligand considered in this study is reported in Table 1.

The calculated free energy change of binding was found in the range $-8.39 \div -16.72$ Kcal/mol and is compatible with the value obtained by us in this study for Actinomycin D (-10.37 Kcal/mol). Considering the different types of chains, generally a regular increase of the binding affinity in homologous series is observed (cf columns **A-C**). Only in the case of the two derivatives **4g** and **4h** the propionate (**B**, $X=CH_2-CH_2$) showed the higher value in the series. The introduction of a methyl in the side chain did not significantly effect the DNA binding capability (entries **A** and **E**); also generally the bioisosteric NMe derivatives (entry **D**) showed ΔG values higher than the corresponding C-substituted (entry **E**) or similar to the methylene analogs (entry **A**). However in isomeric series the branched side chains are less efficient in bringing about the interaction with the DNA minor groove (cf entries **B** and **E**, **C** and **F**).

Table 1. Calculated free energy of binding (ΔG , kcal/mol)

X	CH ₂	CH ₂ CH ₂	CH ₂ CH ₂ CH ₂	NMe	CHMe	CHMeCH ₂
Compound	A	B	C	D	E	F
1	-14.21	-16.07	-16.36	-13.20	-14.10	-15.28
2a	-9.71	-12.02	-12.44	-9.91	-10.26	-9.07
2b	-10.56	-11.18	-12.01	-8.98	-10.79	-10.74
2c	-9.76	-10.79	-11.62	-10.88	-10.06	-9.80
2d	-13.12	-13.19	-15.37	-12.38	-13.22	-13.06
3	-14.83	-15.63	-16.72	-14.31	-14.58	-15.56
4e	-8.39	-10.29	-10.31	-9.32	-9.73	-11.12
4f	-8.82	-10.16	-11.51	-10.25	-10.52	-10.27
4g	-10.31	-11.69	-11.24	-11.21	-10.61	-10.72
4h	-10.68	-10.97	-10.10	-11.93	-12.27	-11.20

The more interesting compounds resulted to belong to the indolo-pyrimidine series **1** and **3** in which a larger planar system is present. This is not surprising although the phenyl rings present in several of our pyrrolo-pyrimidine derivatives can mimic a wider conjugated system being nearly coplanar with the bicyclic system in most of the docked structures. However it is worthy to note that generally the N-benzyl series of pyrrolo[3,2-d]pyrimidine **4h** showed affinity higher than NH compounds. Of the pyrrolo-pyrimidines only compound **2d** showed a ΔG value, -15.37 Kcal/mol, comparable with those of the indolo-pyrimidine classes.

The ligand with the highest affinity was demonstrated to be derivative **3C** ($X=CH_2-CH_2-CH_2$), which had $\Delta G = -16.72$ Kcal/mol, and that therefore can be expected to form a complex

with DNA more stable than that formed by Actinomycin D ($\Delta G = -10.37$ Kcal/mol) with the same DNA fragment. In all the cases the indolo- and pyrrolo-pyrimidines were shown to assume a spatial arrangement in which the planar heterocyclic moiety intercalates in GC portion of the DNA sequence whereas the side chain lies close to the minor groove. Figure 2 (right) shows the DNA-binding mode for the most interesting derivative **3C**.

In order to rationalize structure-binding capability relationships, for all the compounds we calculated the LUMO and HOMO energies, considering that these variables are of importance when two molecules with π electron systems form charge-transfer complexes, especially in the case of DNA-interactive compounds. We utilized the structures of the ligands and of some known intercalators (Actinomycin D included), optimized *in vacuo* by SCF calculation with PM3 method.²¹ It was impossible to find a linear or simple correlation between the ΔG and energy values, however in the case of the more interesting series, the indolo-pyrimidines **1** and **3**, the range of HOMO energies was similar to that obtained for the reference drugs although the LUMO values resulted higher (Table 2). These findings further confirm the interest in the classes of compounds described herein as DNA-interactive polycycles.

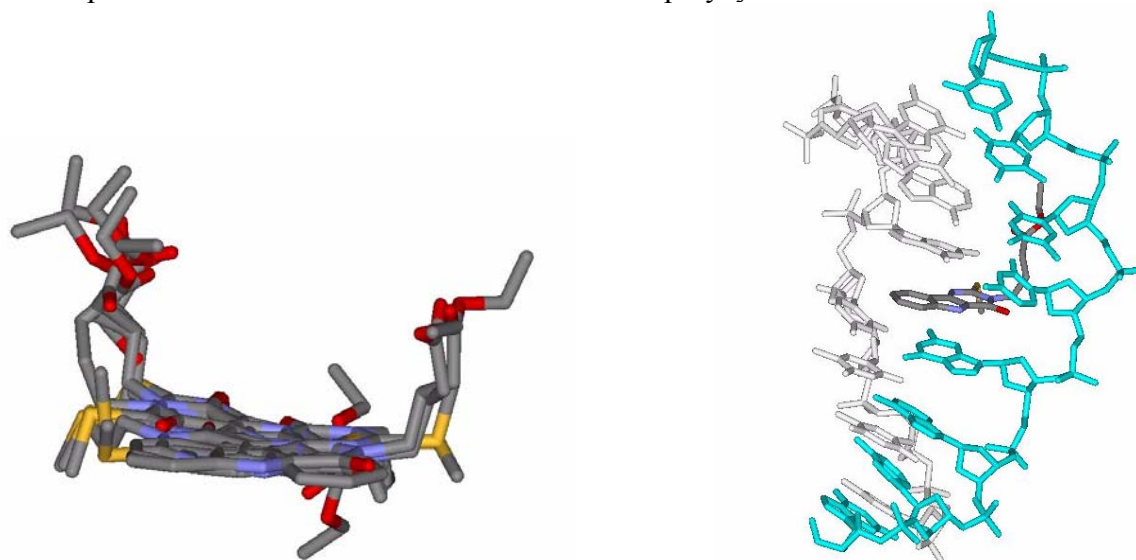


Figure 2. Stick view of ten superpositioned best scores (left) and best structure docked to DNA-fragment (right) for derivative **3C**.

Table 2. Calculated HOMO and LUMO energies (eV)

Compounds	HOMO	LUMO
1	-8.43 ÷ -8.61	-0.71 ÷ -0.90
3	-8.20 ÷ -8.34	-0.76 ÷ -0.89
Reference drugs	-8.48 ÷ -9.08	-1.16 ÷ -1.57

Conclusions

In conclusion these classes of compounds combine two of the main features of DNA-interactive drugs and very tight DNA-binding capability has to be expected. The docking studies revealed a common binding mode with the chromophore intercalated between GC base pairs whereas the side chain lies along the minor groove. The more interesting new derivatives of the series will be synthesized and submitted for DNA-binding assays. The docking studies reported herein provided reliable information on the capability of new ligands to interact with DNA allowing us to quickly select the more interesting derivative in a series of congeners.

Experimental Section

Ligand setup. The 3D structures of the ligands were constructed, pre-optimized with implemented MM2 force field and further refined using Vega 1.5.²² The PM3 method was used for geometry optimization and charge calculation. To allow flexibility in the ligand, it is necessary to assign the rotatable bonds; however in our case all the rotamers were allowed.

DNA fragment setup. The PDB file of the selected structure 1DSC was downloaded from the database and the ligand was removed from the fragment. The resulting macromolecule was setup for docking as follows: polar hydrogens were added using the PROTONATE utility. Solvation parameters were added to the final protein 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 not only the active site but also significant portions of the surrounding surface. The points of the grids were thus $60 \times 60 \times 60$, with a grid spacing of 0.375 Å.

Docking. Docking of the ligands into receptor was carried out using AutoDock (version 3.0.5) set of programs.¹² It was carried out using the empirical free energy function and the Lamarckian genetic algorithm (LGA),¹³ 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 algorithm²³ 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. Ten independent docking runs were carried out for each ligand. At the end of each AutoDock execution, in which more than one run was performed, the program outputs a list of clusters and their energies. The clustering of docked conformations is determined by the tolerance specified in Å. In our cases the 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 favourable free energy of binding. The best representative from each cluster (the one with the lowest energy) is written out in PDBQ format at the end of the log file.²⁴ These structures can be read into any appropriate molecular modelling system. The table of ranked clusters shows the final docked energy for each conformation, and the RMS difference between the lowest energy member of the group and every other member. The RMS for the lowest member of the group is by definition zero. After this table, the structures are output in PDBQ format. Each conformation has a set of remark records, one of which describes the RMS difference between itself and the coordinates specified in the original input PDBQ file.

References and Notes

1. Smolina, I. V.; Demidov, V. V.; Frank-Kamenetskii, M. D. *J. Mol. Biol.* **2003**, *326*, 1113.
2. Lian, C. Y.; Robinson, H.; Wang, A. H. J. *J. Am. Chem. Soc.* **1996**, *118*, 8791.
3. Cirrincione, G.; Almerico, A. M.; Barraja, P.; Diana, P.; Lauria, A.; Passannanti, A.; Musiu, C.; Pani, A.; Murtas, P.; Minnei, C.; Marongiu, M. E.; La Colla, P. *J. Med. Chem.* **1999**, *42*, 2561.
4. Lauria, A.; Diana, P.; Barraja, P.; Almerico, A. M.; Cirrincione, G.; Dattolo, G. *J. Heterocycl. Chem.* **2000**, *37*, 747.
5. Lauria, A.; Diana, P.; Barraja, P.; Montalbano, A.; Cirrincione, G.; Dattolo, G.; Almerico, A. M. *Tetrahedron* **2002**, *58*, 9723.
6. Montalbano, A.; Diana, P.; Barraja, P.; Lauria, A.; Cirrincione, G.; Dattolo, G.; Almerico, A. M. *ARKIVOC* **2002**, (iii), 264.
7. Barraja, P.; Diana, P.; Lauria, A.; Montalbano, A.; Almerico, A. M.; Dattolo, G.; Cirrincione, G. *Heterocycles* **2003**, *60*, 2519.
8. Barraja, P.; Diana, P.; Lauria, A.; Montalbano, A.; Almerico, A. M.; Dattolo, G.; Cirrincione, G. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 2809.
9. Lauria, A.; Patella, C.; Diana, P.; Barraja, P.; Montalbano, A.; Cirrincione, G.; Dattolo, G.; Almerico, A. M. *Heterocycles* **2003**, *60*, 2669.
10. web address: <http://www.ccdc.cam.ac.uk>
11. web address: <http://www.rcsb.org/pdb>
12. web address: <http://www.scripps.edu/pub/olson-web/doc/autodock>
13. 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.
14. see for example: (a) Kataoka, K.; Kosugi, T.; Ishii, T.; Takeuchi, T.; Tutsumi, T.; Nakano, A.; Unoki, G.; Yamamoto, M.; Sakai, Y. PCT int. Appl. 2003; *Chem. Abstr.* **2003**, *139*, 214482. (b) Vidal Juan, B.; Esteve Trias, C.; Segarra Matamoros, V.; Ravina Rubira, E.; Fernandez Gonzalez, F.; Loza Garcia, M. I.; Sanz Carreras, F. PCT int. Appl. 2003; *Chem. Abstr.* **2003**, *138*, 73267. (c) Hayallah, A. M.; Sandoval-Ramirez, J.; Reith, U.; Schobert, U.;

- Preiss, B.; Schumacher, B.; Daly, J. W.; Mueller, C. E. *J. Med. Chem.* **2002**, *45*, 1500. (d) Leonard, P.; Wiglenda, T.; Seela, F. *Nucleosides, Nucleotides & Nucleic Acids* **2001**, *20*, 1279.
15. (a) Romeo, G.; Materia, L.; Manetti, F.; Cagnotto, A.; Mennini, T.; Nicoletti, F.; Botta, M.; Russo, F.; Minneman, K. P. *J. Med. Chem.* **2003**, *46*, 2877. (b) Lowinger, T.; Shimazaki, M.; Sato, H.; Tanaka, K.; Tsuno, N.; Marx, K.; Yamamoto, M.; Urbahns, K.; Gantner, F.; Okigami, H.; Nakashima, K.; Takeshita, K.; Bacon, K.; Komura, H.; Yoshida, N. PCT Int. Appl. 2003; *Chem. Abstr.* **2003**, *138*, 368906. (c) Santagati, A.; Granata, G.; Santagati, M.; Cutuli, V.; Mangano, N. G.; Caruso, A. *Arzneim. Forsch.* **2002**, *52*, 448. (d) Koyama, K.; Nakayama, N.; Nakao, M. Jpn. Kokai Tokkyo Koho 1998; *Chem. Abstr.* **1998**, *129*, 131241.
 16. Cirrincione, G.; Almerico, A. M.; Aiello, E.; Dattolo, G. "Aminopyrroles" In *The Chemistry of Heterocyclic Compounds*, (Pyrroles. Part II), Jones, R. A., Ed.; Wiley, Inc.: Publisher, 1992; Vol. 48, p 299.
 17. Cirrincione, G.; Almerico, A. M.; Diana, P.; Barraja, P.; Mingoia, F.; Grimaudo, S.; Dattolo, G.; Aiello, E. *J. Heterocycl. Chem.* **1996**, *33*, 161.
 18. Diana, P.; Barraja, P.; Lauria, A.; Almerico, A. M.; Dattolo, G.; Cirrincione, G. *Tetrahedron* **2000**, *56*, 5177.
 19. Chen, N.; Lu, Y.; Gadamasetti, K.; Hurt, C. R.; Norman, M. H.; Fotsch, C. *J. Org. Chem.* **2000**, *6*, 2603.
 20. Almerico, A. M.; Lauria, A.; Diana, P.; Barraja, P.; Montalbano, A.; Cirrincione, G.; Dattolo, G. *18th International Congress of Heterocyclic Chemistry*, Yokohama, Aug 2001, abstr. p 308.
 21. Semi-empirical molecular orbital calculations were run on an Indigo-2 Silicon Graphics work station by using the Vamp (V 6.5) software, supplied by Oxford Molecular – Accelrys, according to Stewart, J. J. P. *J. Comput. Chem.* **1989**, *10*, 209; 221.
 22. web address: <http://www.unimi.it>
 23. Solis, F. J.; Wets, R. J.-B. *Math. Oper. Res.* **1981**, *6*, 19.
 24. The log file for each ligand is available as supplementary material.