MODELLING OF DNA COMPLEXES WITH DISTAMYCIN ANALOGUES USING AN *ab initio* CONTINUUM SOLVENT MODEL

Petr BOUŘ[a1,b,*] and Vladimír KRÁL[a2]

[a] *Department of Analytical Chemistry, Prague Institute of Chemical Technology, Technická 5, 166 28 Prague 6, Czech Republic; e-mail: [1] bourp@vscht.cz, [2] kralv@vscht.cz*

[b] *Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic, Flemingovo nám. 2, 166 10 Prague 6, Czech Republic; e-mail: bour@uochb.cas.cz*

Received January 14, 2000
Accepted March 7, 2000

Dedicated to Professor Otakar Červinka on the occasion of his 75th birthday.

Model systems related to non-covalent minor groove DNA complexes with distamycin analogues were investigated using the Turbomole and Gaussian quantum chemical packages. The role of molecular shape, electrostatic field and conformer energies in the complex formation was discussed. The *ab initio* calculations included the COSMO solvent model. If compared to vacuum computations, polar solvent significantly destabilizes such complexes and increases conformational flexibility of distamycin. The DNA complex formation appears to be driven mainly by entropy lowering and complementarity of molecular shapes. The NH moiety of the amide group preferably points to the base pair according to the computations, in agreement with experimental data.

Key words: *Ab initio* calculations; DFT; DNA; COSMO; Distamycin; Amides; Solvation; Minor groove.

Molecular recognition of cellular components plays a central role in biological processes such as enzyme catalysis and inhibition, immunological response, transport, drug action or DNA replication. Cooperative interactions between DNA and binding proteins are critical for the regulation of the gene expression[1–6]. Hence understanding and good theoretical models of such systems appear important for a rational design of DNA-binding drugs.

Although the design of new compounds is primarily based on empirical knowledge, theoretical tools can substantially enhance the search. In spite of the limitation of current computational techniques, many molecular properties can be already obtained in advance. For example, the strength of the hydrogen binding can be determined and thus preferred ligands tested before actual synthesis. Previous computational studies suggest that impor-

tant properties of DNA ligands can be modelled using already rather low-level *ab initio* techniques[7]. Geometries of DNA–amide group complexes were studied behind the HF level[8]. For large molecules combination of empirical and quantum-chemical procedures can provide useful estimates of molecular reactivity[9]. Extensive post-HF studies based usually on the many body perturbation theory are routinely performed in vacuum for smaller DNA fragments[10–12]. Consideration of explicit water molecules, however, revealed a significant contribution of the solvent not only to the stability of weak complexes, but also to their geometry[13,14]. In this study, we compare usual *ab initio* computations of binding energies in vacuum with computations involving the implicit solvent COSMO model, as it became available in latest commercial software packages.

The COSMO model[15] has simpler boundary conditions for the electrostatic field than the more usual dielectric models[16]. Nevertheless, it was shown that the model could be adapted to describe solvents of various permittivity[17]. Also, the COSMO solvation energies and other molecular properties agree well with experimental values for small molecules[18]. Nevertheless, similar systematic computations for larger molecules are not known to us. Thus one of the main purposes of this study is to determine the implication of the more realistic *ab initio* modelling for larger systems. For such a case, it has been feared that current implementations do not provide energies with sufficient accuracy. However, as shown below, obtained solvent corrections are consistent and still reproducible.

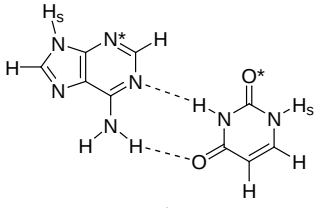
We calculate energies and geometries of model systems related to DNA protein complexes. Particularly interesting are proteins that recognize DNA sequences. The genetic information of DNA can be thus read *via* a binding to the major and minor groves. Regulatory proteins (promoters, repressors) are bound to the major groove, while the minor groove is reserved for polymerases and xenobiotics including many antibiotics. These proteins have often similar structure (see distamycin and netropsin in Fig. 1), and become a natural target for derivative synthesis[3].

Detailed information about structure of many covalent DNA–oligopeptide complexes has been available from the X-ray diffraction analysis for a long time. For example, netropsin binds firmly to the minor groove of the dimer $[CGCGAATTCGCG]_2$, displacing the spine of hydration with the pyrrole rings approximately parallel to the walls of the minor groove[4]. Hydrogen bonds from amide NH groups bridge the strands to the exposed adenine N(3) and thymine O(2) on adjacent sites, thus mimicking the bridging provided by the displaced water. Rational design of DNA sequence-recognizing agents is largely based on information-reading oligo- peptides

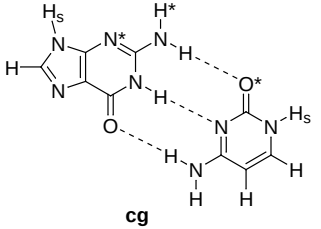
called lexitropsins, which are derived from the natural antibiotics[19,20]. Another important property of these compounds is their light-sensitivity[21] or the ability to intercalate between the base pairs. For all lexitropsin molecules the DNA sequence recognition is mediated *via* the hydrogen binding of amide groups.

MODEL SYSTEMS

The AT and CG pairs were modelled with complexes **at** and **cg**.



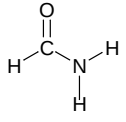
at



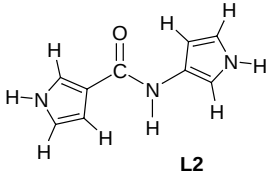
cg

Potential binding sites in minor grooves are denoted by asterisks (*). Hydrogens H_s are replaced by the sugar-phosphate backbone in natural DNA. With respect to the properties studied at this work the **at** system mimics both the AT and AU base pairs, since the methyl group of the thymin residue is far from the binding sites.

Two model ligands were considered:



L1



L2

Ligand **L1** is the simplest model of amide group, which is the most frequently encountered binding site in natural compounds. Ligand **L2** additionally mimics the aromatic residues in distamycin or netropsin. The substitution of the pyrrol ring attached to the carbonyl group resembles model compounds studied in our laboratory rather than those in the natural products. As trial computations suggest, however, such a change have a negligible influence on calculated energies at the level of approximation used in this study.

Geometries of these four molecules and their complexes were optimized by energy minimization using the Gaussian program[22]. Two levels of ap-

proximation were used: the HF/SCF method with standard 6-31G basis, and the density functional theory (DFT) with the Becke 88 exchange[23] and Perdew–Wang's 1991 correlation[24] functional (BPW91 key word in Gaussian). For the latter DFT model additional polarization functions were added to the binding atoms in base pairs (marked by asterisk) and ligands (H, O) corresponding to the 6-31G** basis. A full 6-31G** basis set was used for properties of the **cg** and **at** pairs. Planar (C_s) symmetry was imposed on geometries of the four molecules. Using a standard routine of Gaussian and a set of supplementary programs molecular electrostatic potentials were calculated. The COSMO solvent model was used as implemented both in Gaussian as well as in Turbomole[25]. While the Gaussian program package offers greater flexibility, we found that the Turbomole implementation of COSMO is more numerically stable. Both programs provide almost same energies and were used interchangeably in the computations.

In the complexes the planar symmetry of the base pairs and ligands was constrained. The ligand and base pair planes were mutually perpendicular. A tilt observed in real complexes[4] was neglected, since we do not assume that it has a substantial influence on binding energies. In DNA complexes the sugar-phosphate backbone sufficiently prevents a π-electron conjugation between the ligand and base. Five types of complexes were investigated for each ligand, in order to mimic the most probable interactions between the minor groove and the amide group.

(I)	**cg–L1**, **cg–L2**	(H*···O=C)
(II)	**cg–L1**, **cg–L2**	(O*···HN)
(III)	**cg–L1**, **cg–L2**	(N*···HN)
(IV)	**at–L1**, **at–L2**	(O*···HN)
(V)	**at–L1**, **at–L2**	(N*···HN)

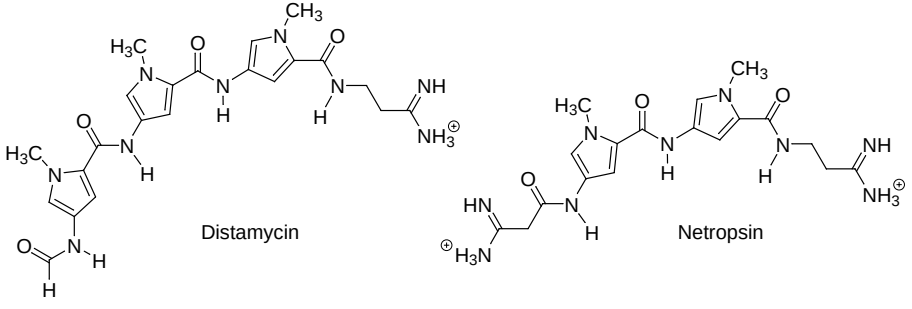


FIG. 1
Netropsin and distamycin as examples of natural DNA-binding drugs

Although many studies advocate the counterpoint correction (CP) of BSSE (refs[10,26]), only high-level benchmark computations[27] prove its convenience and even for such a small system as He_2 molecule problems have been encountered with combination of CP and many body perturbation theory[28]. Our results (small difference of energies when polarization functions were added) do not indicate a need for the BSSE for the present model. Thus, the basis set superposition error (BSSE) was not considered in the computations.

RESULTS AND DISCUSSION

Electrostatic Potential of Base Pairs

Many properties of molecules involved in the non-covalent complexes can be understood from their electrostatic potential. Its estimation requires relatively modest computational effort if compared to geometry optimization, for example. Supposedly, the longer-range electrostatic forces play an important role not only in the energy balance, but also during the initial dynamic phase of complex formation.

The electrostatic field of the DNA in the close vicinity of the grooves is determined by the base pairs. We calculated the field of the **at** and **cg** pairs (where hydrogens H_s were replaced by methyl groups) at the B3LYP/6-31G** level. According to trial computations, the electrostatic pattern is not too sensitive to computational model and reliable fields can be obtained already at lower levels of approximation (*e.g.* HF/6-31G). This is in agreement with other studies using molecular-dynamics and similar semiempirical modelling[9]. In Figs 2 and 3, the electrostatic potentials of **at** and **cg** base pairs are plotted, respectively, as cross-sections through the pair planes. The color scale captures the range of –1 to 1 V, the black isopotential curves are plotted each 0.1 V. The inner white area (following isodensity line) approximately corresponds to the van der Waals volume of the base pair. Apparently, negative potential indicated by the blue color prevails both in the minor and major grooves of each of the two pairs. Additionally, positive regions appear close to the center of the minor groove of CG as well as of the major groove of AT. Thus, for example, hydrogen as well as electron donation is “electrostatically allowed” in non-covalent binding involving the minor groove of CG. Because of this ambivalence, CG minor groove complexes are expected to be weaker than those of the AT pair, which was also observed experimentally[3,4,7]. Clearly, both the **cg**

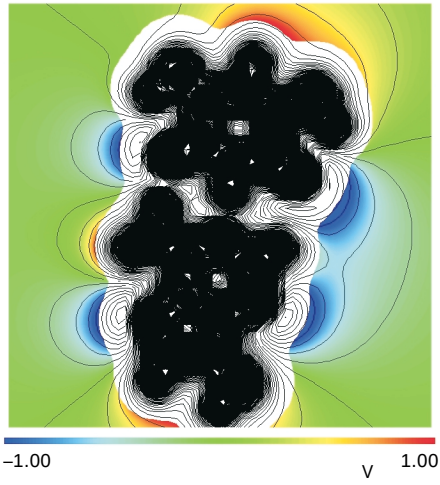


FIG. 2
Calculated (Becke3LYP/6-31G**) electrostatic field of the AT base pair, cross-section in the plane of the pair. Scale given in V, isopotential curves drawn each 0.1 V. White area is inside the van der Waals surface of the pair. In the vicinity of the minor groove (right) the potential is always negative

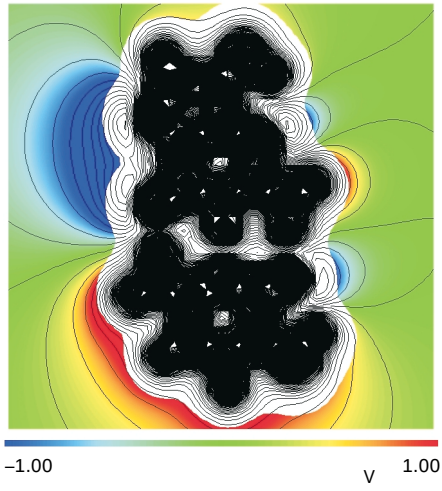


FIG. 3
Electrostatic field in the vicinity of the CG base pair, projection analogous to Fig. 2. In the minor grove (right) a positive potential between the two negative sites can be observed, presumably destabilizing the complexes

and **at** pairs are highly polar and prone to specific interactions with polar environment.

Electrostatic Potential and Conformer Energies of Distamycin

A nice example of the electrostatic complementarity between the negative minor groove potential and a positive ligand field can be shown for distamycin (see Fig. 1 for its structure). A geometry of its simplified analogue was optimized at the HF/6-31G* level. The terminal $-(CH_2)_2-C(NH)NH_2$ group was replaced by $-CO-NH_2$ in order to avoid complication with charged and flexible residues, which are supposedly strongly solvated in real systems. The results indicate that the lowest-energy conformation corresponds to those found in a complex with DNA. Electrostatic potential on the van der Waals surface of the molecule (determined as an isodensity plane, $\rho = 0.0006$, where ρ is the electronic density) can be seen in Fig. 4 for four conformations. We have selected only those four "most interesting" conformations for demonstration, out of the total number of $64 = 2^6$. However, the lowest-energy conformer corresponds to the absolute minimum. Black numbers denote relative conformation energies in vacuum, blue values describe the COSMO solvent-corrected energies. Interestingly, the positively charged (red) edge of the lowest-energy conformation is not only complementary to the negative potential of the minor groove (*cf.* Figs 2, 3), but matches well also the shape of the groove in double- stranded DNA (*cf.* X-ray in ref.[4]). This coincidence of conformational energies, molecular shape, electrostatic potential and arrangement of N–H groups thus leads to the outstanding binding properties of the distamycin molecule. The upper conformation in Fig. 4, for example, could form a complex involving the NH_2 groups of the CG pair (*cf.* Fig. 3) if the energy barrier of 27.9 kcal/mol were overcome. This is, however, unlikely because of the small energy differences involved in the complex formation. Surprisingly, for the upper two conformations (in Fig. 4) a negligible difference of conformational energy (0.5–0.8 kcal/mol) is accompanied by a profound change of the electrostatic pattern. Presumably, such effects must be considered for development of new compounds interacting with DNA.

The solvent-corrected conformational energies are significantly lower than the vacuum values. Polar solvent thus contributes to the conformational flexibility of this type of compounds and, probably, to peptides generally. On the other hand, relative ordering of the conformers is preserved in this case, and thus the faster and easier vacuum calculation can be still used for approximate estimation of conformational energies.

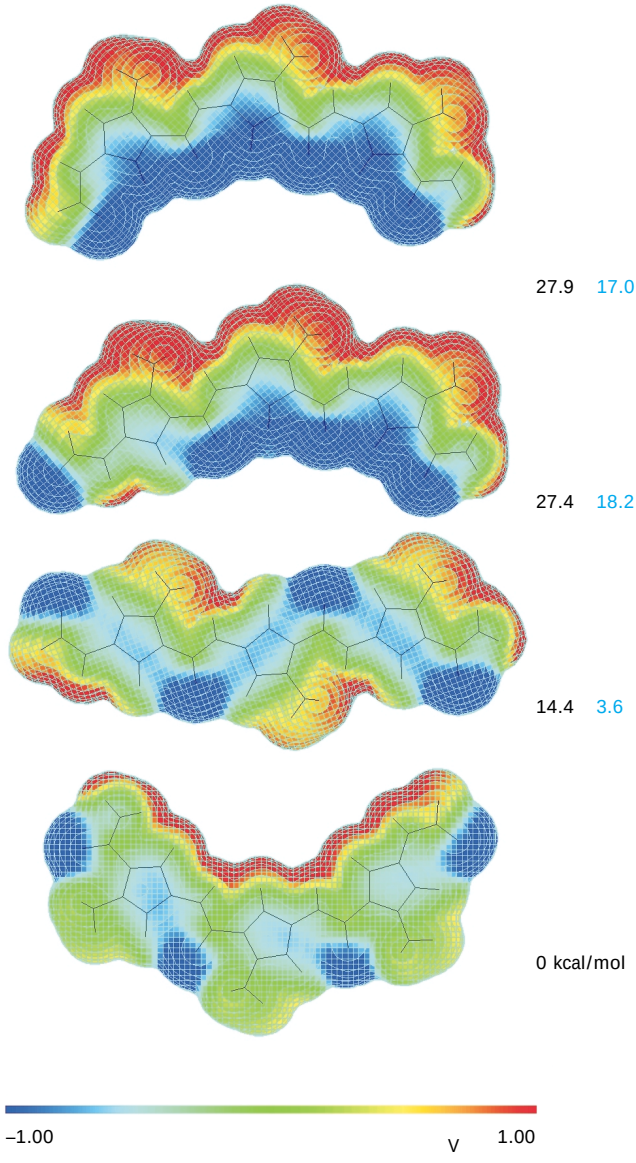


FIG. 4
Electrostatic potential (an HF/6-31G* calculation) on the van der Waals surface of the distamycin analogue, for the lowest-energy (bottom) and three randomly selected (top) conformations. Black numbers, vacuum calculation; blue numbers, the solvent-corrected values

AT and CG Formation Energies

The binding energies (**c** + **g** → **cg**, **a** + **t** → **at**) calculated for planar arrangement (C_s symmetry) at the Becke3LYP/6-31G** level are summarized in Table I. The **cg** binding energy is almost twice as high as that for **at**. This can be explained by the number of hydrogen bonds in the complex and by an additional mutual polarization and electron conjugation in the **cg** pair, as was shown in numerous studies previously[8,10]. When the COSMO model is switched on, binding energies are reduced to about 50% as can bee seen in the last two columns of Table I. This difference leads obviously to a dramatic decrease in the stability of the pairs and thus it can be concluded that the solvent correction is necessary for any modelling of such weakly-bound polar systems in water. The correction not only tends to smooth out differences between energies of individual conformers, but also between complex formation energies.

Two conclusions can be further drawn from the similarity of the binding energies computed for relaxed and rigid geometries (last two columns in Table I). (i) The solvent has a relatively minor effect on molecular structure and the major energy correction originates in the polarization of the solvent by the solute. (ii) The relaxation can be neglected in practical computations. This may be welcome since geometry optimizations can be performed much more easily in vacuum. Apart from a tremendous demand on the computer time, the solvent-corrected optimization algorithms are numerically unstable and often do not lead to the lowest-energy state at all[25].

TABLE I
Binding energies of DNA base pairs, calculated with and without solvent correction. Energies in kcal/mol, calculated at the Becke3LYP/6-31G** level

Pair	Vacuum	Water[a]	
		non-relaxed[b]	relaxed
at	16.4	8.2	7.9
cg	30.9	14.9	13.8

[a] COSMO model. [b] Vacuum geometry used without further optimization.

Energies of the Complexes

Calculated binding energies for the complexes **I–V** are listed in Table II. The DFT calculation predicts slightly lower values of binding energies than the HF approximation and thus no significant improvement can be expected by using the DFT at this level.

Formation energy of the complex **I** varries most with the model used. It is consistently the smallest for the vacuum calculations for both ligands. The relative energy rises for the ligand **L1** if the COSMO solvent model is applied (4.5 kcal/mol is the smallest energy in the first column, while the 3.2 kcal/mol is the highest value in the third column). However, since this rise is not observed for the **L2** ligand better related to real molecules, we may expect that orientation of the carbonyl group towards the DNA minor groove is not energetically convenient if another orientation is possible.

All the other complexes (**II–V**) where DNA donates electrons as a base thus appear more stable. The ligands **L1** and **L2** behave similarly in vacuum, while greater differences in the binding energies appear for water. The solvent also significantly lowers the energies, by up to 8.1 kcal/mol for the last complex. Thus, the solvent correction appears necessary for correct determination of molecular reactivity. However, calculated energy differences are small and on the brim of calculational error. Thus further work will be needed in order to incorporate the influence of the solvent-specific interactions as well as dynamic fluctuations of the geometry on binding energies.

TABLE II
Calculated energies (kcal/mol) for the model minor groove complexes

Complex	HF				DFT	
	vacuum		water		vacuum	
	L1	**L2**	**L1**[a]	**L2**	**L1**	**L2**
(**I**)	4.5	5.4	2.6/3.2	0.9	4.0	4.9
(**II**)	7.4	7.7	2.8/3.1	2.4	5.4	5.2
(**III**)	8.0	8.0	1.8/1.9	1.4	7.3	6.5
(**IV**)	7.7	8.5	0.8/1.2	1.8	7.2	6.9
(**V**)	7.9	9.1	0.2/0.2	0.6	6.8	7.4

[a] Non-relaxed/relaxed.

Calculated energies follow the basic electrostatic rules (Figs 2–4), which can thus be used for a rough estimation of molecular reactivity. The importance of entropy rather than energy factors for the DNA complexes was also indicated by the latest molecular dynamics simulations[29].

The role of the aromatic residues attached to the amide group is ambiguous in our systems, since such a substitution can both increase as well as decrease binding energies. Its interaction with the sugar-phosphate DNA backbone would require more sophisticated computations not feasible with current means. The backbone was modelled by the geometry constraints where the ligand was perpendicular to the base pair plane. As a typical example of the optimized geometry the complex **V** is plotted in Fig. 5.

TABLE III
Calculated hydrogen bond lengths (Å) for the model minor groove complexes

Complex	HF			DFT	
	vacuum		water	vacuum	
	L1	**L2**	**L1**	**L1**	**L2**
(**I**)	2.015	1.988	1.897	1.933	1.988
(**II**)	1.956	2.098	1.907	2.116	1.941
(**III**)	2.049	2.257	2.084	2.164	1.969
(**IV**)	2.040	2.105	2.108	2.058	1.928
(**V**)	2.182	2.402	2.233	2.291	2.040

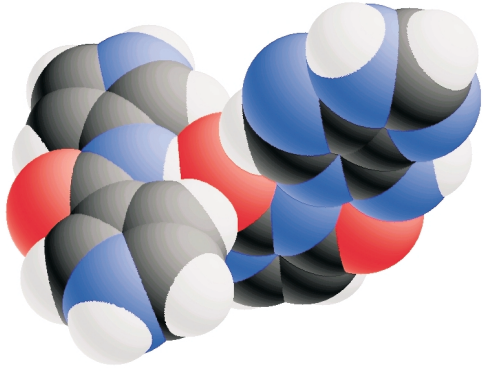


FIG. 5
Optimized geometry for complex **VI** of base pair **at** (left) and ligand **L2** (right)

In Table III hydrogen bond lengths are listed. Apparently, they may adopt a wide range of distances, which reflects the shallow interaction potential in the complexes. Moreover, their length depends on mutual orientation of the ligand and base pair. Surprisingly, the solvent correction leads both to lengthening (complexes **II**, **IV** and **V**) as well as shortening (complexes **I** and **II**) of the hydrogen bonds. Thus the lengths of the hydrogen bond cannot be used as an indication of its strength.

CONCLUSIONS

Stability of the non-covalent DNA complexes is determined by the molecular electrostatic and spatial complementarity, which can be roughly estimated on the basis of vacuum computations. Solvent corrections to conformational or binding energies, however, must be considered for more precise computations. The complex stabilization energies are more sensitive to the solvent model than the conformer energies.

Calculated properties of the model complexes agree with observed behavior of compounds covalently bonded to the minor groove of DNA. A minor groove hydrogen binding of an electron-donating ligand (*e.g.* C=O group) is not as convenient as that of an electron acceptor (*e.g.* the N–H group). Precise computations of molecular shapes, electrostatic fields and binding energies could enhance development and a rational design of DNA sequence-recognizing molecules in future.

The work was supported by the grants of the Grant Agency of the Czech Republic (No. 203/96/0740 and No. 203/97/1099 to V. K., No. 203/97/P002 to P. B.), the Ministry of Education of the Czech Republic (No. VS97135) and the Howard Hughes Medical Institute (No. 75195-541101).

REFERENCES

1. Szewczyk J. W., Baird E. E., Dervan P. B.: *J. Am. Chem. Soc.* **1996**, *118*, 6778.
2. Trauger J. W., Baird E. E., Dervan P. D.: *Angew. Chem., Int. Ed. Engl.* **1998**, *37*, 1421.
3. Neidle S., Waring M. J.: *Molecular Aspects of Anticancer Drug–DNA Interactions*. Macmillan Press, London 1993.
4. Neidle S.: *Biopolymers* **1998**, *44*, 105.
5. Fauvet M. P., Gresh N.: *J. Biomol. Struct. Dyn.* **1994**, *11*, 1203.
6. Swalley S. E., Baird E. E., Dervan P. B.: *Chem. Eur. J.* **1997**, *3*, 1600.
7. Sapse A. M., Jain D. C., Lown J. W.: *J. Biomol. Struct. Dyn.* **1997**, *14*, 475.
8. Šponer J., Burcl R., Hobza P.: *J. Biomol. Struct. Dyn.* **1994**, *11*, 1357.
9. Hobza P., Kabeláč M., Šponer J., Mejzlík P., Vondrášek J.: *J. Comput. Chem.* **1997**, *18*, 1136.
10. Hobza P., Šponer J.: *Chem. Rev. (Washington, D. C.)* **1999**, *99*, 3247.

11. Hobza P., Šponer J., Polášek M.: *J. Am. Chem. Soc.* **1995**, *117*, 792.
12. Florián J., Leszczynski J.: *J. Am. Chem. Soc.* **1996**, *118*, 3010.
13. Zhanpeisov N., Leszczynski J.: *J. Phys. Chem. B* **1998**, *102*, 9109.
14. Zhanpeisov N., Leszczynski J.: *J. Phys. Chem. A* **1998**, *102*, 6167.
15. Klamt A., Schürmann G.: *J. Chem. Soc., Perkin Trans. 2* **1993**, *2*, 799.
16. Miertus S., Scrocco E., Tomasi I.: *J. Chem. Phys.* **1981**, *55*, 117.
17. Klamt A.: *J. Phys. Chem.* **1995**, *99*, 2224.
18. Amovilli C., Barone V., Cammi R., Cances E., Cossi M., Mennucci B., Pomelli C. S., Tomasi J.: *Adv. Quantum Chem.* **1999**, *32*, 227.
19. Swalley S. E., Baird E. E., Dervan P. B.: *J. Am. Chem. Soc.* **1996**, *118*, 8198.
20. Lown J. W.: *Anti-Cancer Drug Des.* **1988**, *3*, 25.
21. Chatterjee S. R., Srivastava T. S., Kamat J. P., Devasagayam T. P.: *Chem.-Biol. Interact.* **1997**, *108*, 27.
22. Frisch M. J., Trucks G. W., Schlegel H. B., Gill P. M. W., Johnson B. G., Robb M. A., Cheeseman J. R., Keith T., Petersson G. A., Montgomery J. A., Raghavachari K., Al-Laham M. A., Zakrzewski V. G., Ortiz J. V., Foresman J. B., Cioslowski J., Stefanov B. B., Nanayakkara A., Challacombe M., Peng C. Y., Ayala P. Y., Chen W., Wong M. W., Andres J. L., Replogle E. S., Gomperts R., Martin R. L., Fox D. J., Binkley J. S., Defrees D. J., Baker J., Stewart J. P., Head-Gordon M., Gonzalez C., Pople J. A.: *Gaussian 98, Revision A.3*. Gaussian, Inc., Pittsburgh (PA), 1999.
23. Becke A. D.: *Phys. Rev. A: At., Mol., Opt. Phys.* **1988**, *38*, 3098.
24. Perdew J. P., Wang Y.: *Phys. Rev. B: Condens. Mater.* **1992**, *45*, 13244.
25. Klamnt A. in: *The Encyclopedia of Computatinal Chemistry* (P. v. R. Schleyer, N. L. Allinger, T. Clark, J. Gasteiger, P. A. Kollman, H. F. III. Schaefer and P. R. Schreiner, Eds), p. 604. John Wiley & Sons, Chichester 1998.
26. Chalasinski G., Szczesniak M. M.: *Chem. Rev. (Washington, D. C.)* **1994**, *94*, 1723.
27. Dunnnig T. H., Peterson T. A., Woon D. E. in: *The Encyclopedia of Computatinal Chemistry* (P. v. R. Schleyer, N. L. Allinger, T. Clark, J. Gasteiger, P. A. Kollman, H. F. III. Schaefer and P. R. Schreiner, Eds), p. 88. John Wiley & Sons, Chichester 1998.
28. Bukowski R., Jeziorski B., Szalewicz K.: *J. Chem. Phys.* **1996**, *104*, 3306.
29. Rodger A., Meistermann I., Snaders K., Hannon M.: *7th International Conference on CD, Mierki, Poland, 1999*; Book of Abstracts, p. 34. A. Mickiewicz University, Poznań 1999.