



A new insight into the structure and stability of Hoogsteen hydrogen-bonded G-tetrad: an ab initio SCF study

Jiande Gu^{a,b}, Jerzy Leszczynski^{a,*}, Manju Bansal^c

^a *The Computational Center for Molecular Structure and Interactions, Department of Chemistry, Jackson State University, Jackson, MS 39217, USA*

^b *State Key Laboratory of Drug Research, Shanghai Institute of Materia Medica, Chinese Academy of Sciences, Shanghai 200031, PR China*

^c *Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560012, India*

Received 26 April 1999; in final form 29 June 1999

Abstract

Hoogsteen-bonded guanine tetrad (G-tetrad) has been studied at the HF/6-311G(d,p) and B3LYP/6-311G(d,p) levels of theory. The optimized structure of G-tetrad is significantly different from the traditionally suggested structure. Bifurcated hydrogen bonds are predicted to be responsible for the formation of internal G–G pairs in G-tetrad. The calculated stabilization energy of 62 kcal/mol at the HF/6-311G(d,p) level suggests that a quite stable G-tetrad is formed through four pairs of bifurcated H-bonds. Inclusion of electron correlation (the B3LYP/6-311G(d,p) approach) recovers only 4 kcal/mol of additional stabilization energy. The electrostatic potential map of G-tetrad shows significant concentration of negative charges in the central area of the G-tetrad. The neutralization of this charge by a cation placed in this central area is expected to bring an auxiliary stabilization of the G-tetrad. © 1999 Elsevier Science B.V. All rights reserved.

Guanine-rich oligonucleotides are known to be inhibitors for fibrinogen action in thrombin and HIV viral mediated cell fusion [1–5]. The formation of Hoogsteen-bonded guanine tetrads leads to novel four-stranded structures [6–10] in guanine-rich oligonucleotides such as d(T2G4) found in *Tetrahymena* telomeric DNA, d(T4G4) in *Oxytricha*, d(T2G3) in human, and d(T3AG3) in *Arabidopsis* [11–14]. A number of proteins have been identified to have specific binding to G-tetraplex structures [15,16]. G-tetraplexes have been postulated to be crucial for dimerization of HIV RNA [17].

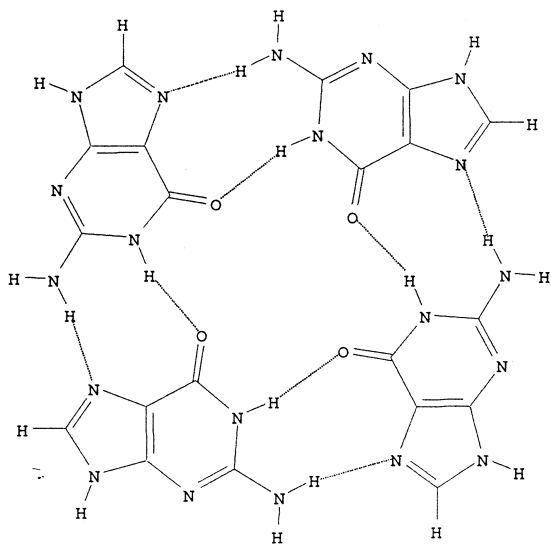
Previous theoretical studies on the structure of G-tetrad have been performed using the molecular mechanics and the molecular dynamics methods [18–25] in which classical empirical potentials depend on the atom–atom pair-wise additivity and cover only electrostatic, dispersion, and repulsion contributions. The ab initio studies on trimers of DNA bases have demonstrated the importance of the nonadditivity of interaction in the H-bonded triads of nucleobases [26]. Besides, the drastic and different approximations are introduced by different potential models, and consequently, the calculated properties of nucleobases are known to be force-field dependent [27,28].

It has been found that the presence of a metal ion is essential in the formation of the G-tetrad com-

* Corresponding author. Fax: +1-601-973-3647; e-mail: jerzy@tiger.jsums.edu

plexes in both parallel and folded-back quadruples structures [25,29–34]. Recent studies have also demonstrated the function of ammonium ion in stabilizing the G-tetraplex [35,36]. It seems that in order to stabilize this complex, charge compensation in the center of G-tetrad is necessary. However, based on the structure depicted in Scheme 1 in which there are eight H-bonds holding four guanines together, one could expect that the G-tetrad might be stable even without the presence of a metal ion or other cations. There is an increasing evidence that the four-stranded Hoogsteen-bonded G-tetrad structure is thermodynamically more stable than duplex DNA [36–39]. Although the structure and stability of G-tetraplex does not solely depend on the interaction of isolated Hoogsteen-bonded tetrads of the bases, the base pairing is one of the most important factors in the formation of tetraplexes, and the details of this phenomenon could only be revealed by accurate computational studies. Due to a relatively large size of such system, no theoretical study of the energy minimum structure of G-tetrad based on the reliable *ab initio* quantum chemistry methods has been reported so far.

In this Letter we report the first *ab initio* quantum chemistry study of the stability and structure of G-tetrad. The Hartree–Fock self-consistent-field (SCF) method in conjunction with the valence triple-zeta basis set augmented with d- and p-like



Scheme 1. Hoogsteen H-bonded G-tetrad.

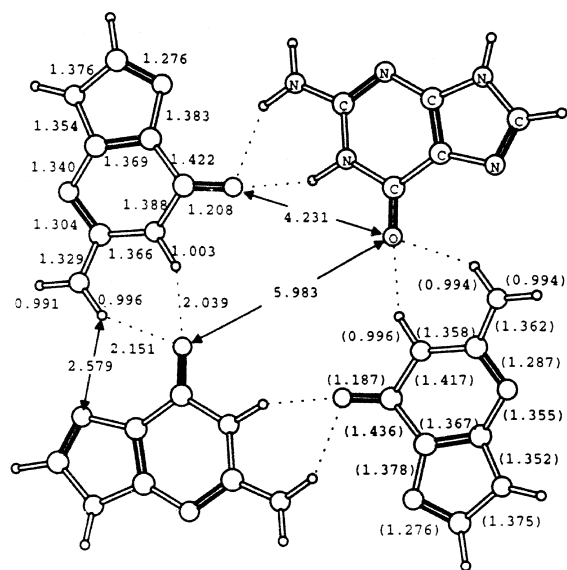


Fig. 1. The optimized structure of G-tetrad with C_4 symmetry. Values in parentheses are the optimized geometric parameters of guanine. Atomic distance in Å.

polarization functions [40] (6-311G(d,p)) was used to ensure a correct description of the H-bonded system. To cover the electronic correlation effects, the density functional theory (the B3LYP approach) [41–43] was used in the single point energy calculation. Our previous studies on H-bonded systems involving DNA bases have shown that the B3LYP approach predicts reliable interaction energies and is compatible to the MP2/6-31(d,p) method [44,45]. Gaussian-94 software [46] was used in the calculations.

The energy minimum form of the G-tetrad was located at the HF/6-311G(d,p) level. C_4 symmetry constraint was imposed during the optimization. Despite the assumed non-planar initial structure of the guanine units, the optimized geometry of the G-tetrad consists of almost planar monomers. The geometric parameters of the optimized G-tetrad are given in Fig. 1 together with the parameters of the guanine monomer which are also listed for comparison.

The structure of G-tetrad as obtained by this calculation is significantly different from the traditionally suggested structure in the presence of cations (see Scheme 1). In contrast to the traditionally suggested structure, the current study indicates that G-te-

trad is stabilized by bifurcated hydrogen bonds. In the optimized G-tetrad, bifurcated hydrogen bonds are predicted to form internal G–G pairs. The hydrogen of the amino group forms a relatively stronger H-bond with the O6 atom ($R_{O \dots H} = 2.15 \text{ \AA}$) rather than with the N7 atom ($R_{N \dots H} = 2.58 \text{ \AA}$). This is a result of the negative charge distribution on the O6 atoms of guanine. The oxygen atoms in the central part of G-tetrad repulse each other due to the electrostatic force. In order to minimize such interactions, guanine units in the tetrad orient themselves in such a way that the oxygen atom forms the bifurcated H-bonds with the hydrogen attached to the N1 and the hydrogen from the amino group. In this case, there are a number of factors stabilizing the G-tetrad: (1) the negative charge on the O6 atom is partly neutralized by two hydrogen atoms; (2) the oxygen–oxygen atomic distances increase; and (3) the neighboring oxygen–oxygen interaction is partially screened by the H atom at the N1 position. The significant non-planarity of the amino group of isolated guanine disappears in the G-tetrad due to the strong interaction between the hydrogen from the amino group with the O6 and the N7 atoms, of a neighbouring guanine.

The calculated energies of the G-tetrad and guanine are summarized in Table 1. The stabilization energy of 62 kcal/mol at the HF/6-311G(d, p) level (BSSE [47] included) suggests significant stabilization of the G-tetrad formed through four pairs of bifurcated H-bonds. Inclusion of the electron correlation at the B3LYP/6-311G(d, p) level provides only 4 kcal/mol of additional stabilization energy. An additional study of the oxygen–oxygen repulsion contributions to stabilization of such structure has

Table 1

	E (hartree)	ΔE^a (kcal/mol)	BSSE (kcal/mol)	ΔE^{BSSE^b} (kcal/mol)
HF/6-311G(d, p):				
Guanine	-539.52751		-1.6	
G-tetrad	-2158.21951	-68.7		-62.1
B3LYP/6-311G(d, p)//HF/6-311G(d, p):				
Guanine	-542.69300		-2.4	
G-tetrad	-2170.89272	-75.8		-66.2

$$^a \Delta E = E(\text{G-tetrad}) - 4 \times E(\text{guanine}).$$

$$^b \Delta E^{BSSE} = \Delta E - 4 \times \text{BSSE}.$$

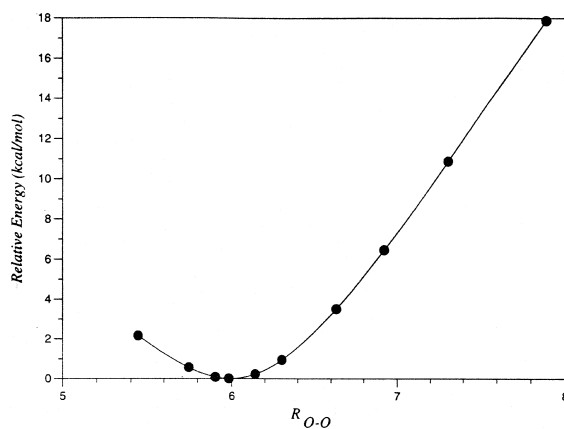


Fig. 2. The energy profile of G-tetrad along the energy minimum path of the diagonal O–O distance. Atomic distance in Å and energy in kcal/mol.

been carried out by optimizing the tetrad structure for different assumed diagonal O–O distances. The calculations reveal that an increase of the diagonal O–O distance leads to a larger energy increase than that caused by reduction of the diagonal O–O distance. Fig. 2 shows the energy profile along the energy minimum path of the diagonal O–O distance. At small O–O distances, it has been found that the

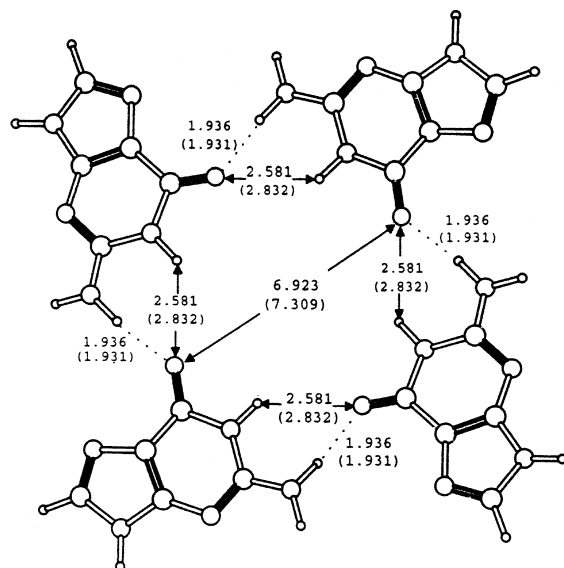


Fig. 3. Two optimized structures of G-tetrad with the constraints of diagonal O–O distances of 6.924 and 7.309 Å.

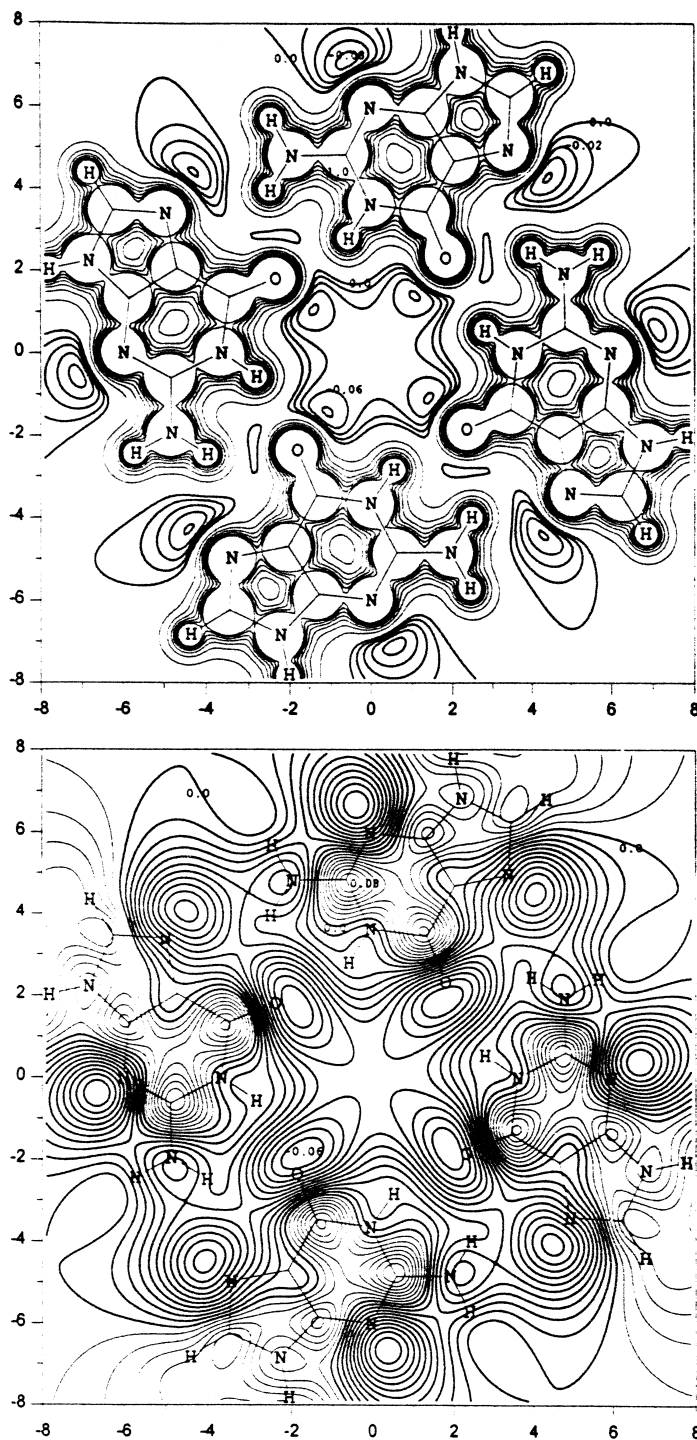


Fig. 4. The electrostatic potential map of G-tetrad. Top: G-tetrad plane. Bottom: 1.6 Å above G-tetrad plane. Thin line represents the positive part of electrostatic potential and thick line is the negative parts of electrostatic potential.

increase of the O–O repulsion is partly compensated by better H-bonding between the guanine pairs. At a large O–O atomic distance ($\sim 7 \text{ \AA}$) the bifurcated H-bonds becomes a single H-bond as can be seen in Fig. 3. The energy increases rapidly as the H-bonding changes from the bifurcated to the single form. From an analysis of Fig. 3 along with the energy profile, one can see that the stability of G-tetrad is governed by the formation of bifurcated H-bonding; the electrostatic effects have been reduced to a minimum through an adjustment of the relative position of the guanine units in the tetrad. It is interesting that the G-tetrads, in the absence of intercalated ions, achieve a similar bifurcated H-bonding arrangement during a molecular dynamics simulation of d(G)7 quadruples structure, carried out using the latest AMBER force field and PME method for treating the electrostatic interaction [48,49]. Hence the effect of including the sugar–phosphate backbone, and using a semi-empirical force field lead to model a larger system, leads to similar G-tetrads, with the O6 atom forming strong hydrogen bonds with both N1 and N2, while the N7 atom is moved away from the N2 amino group, leading to a considerably weaker hydrogen bond. The diagonal O6–O6 distance is $\sim 5.8 \text{ \AA}$, close to the value corresponding to the energy minima in Fig. 2, which reduces considerably if cations are located between the G-tetrads, leading to an increase in the intrinsic energy of the DNA quadruplex, which is compensated by ion–DNA interactions. Recent chemical probing data also indicates that the G-tetrads undergo cation dependent local variations and the N7 atom of guanine is exposed in some cases [50] in agreement with the G-tetrad structure revealed by theoretical studies.

Electrostatic potential map provides a simple way to predict how different geometries alter reactivities in intact DNA. The electrostatic potential map of G-tetrad shows significant concentration of the negative charge in the central area of G-tetrad (Fig. 4). The effects of cations in stabilizing G-tetrad is clearly through the neutralization of this charge. From the geometric parameters of the optimized structure in which the diagonal oxygen–oxygen atomic distance is $\sim 6 \text{ \AA}$, it is obvious that the position of a cation which interacts with G-tetraplex can be either between or within the planes of the H-bonded G-tetrad that is in agreement with the data obtained by X-ray

crystallographic technique for the sodium cations stabilized guanine-tetraplex [32,34]. It is unlikely that in such complex the oxygen atoms act as ‘ligands’ as suggested by Ross and Hardin [25]. The details of the interactions between G-tetrad and different cations are currently being investigated in our laboratory.

Acknowledgements

The authors would like to thank the Mississippi Center for Supercomputing Research for the computational facilities. This work was facilitated in part by the NSF Grant 9805465 and by the Office of Naval Research, Grant No. N00014-98-1-0592 and the Army High Performance Computing Research Center under the auspices of the Department of the Army, Army Research Laboratory cooperative agreement number DAAH04-95-2-0003/contract number DAAH04-95-C-0008, the content of which does not necessarily reflect the position or policy of the government, and no official endorsement should be inferred.

References

- [1] K.Y. Wang, S. McCardy, R.G. Shea, S. Swaminathanan, P.H. Bolton, *Biochemistry* 32 (1993) 1899.
- [2] R.F. Macaya, P. Schultze, F.W. Smith, J.A. Roe, J. Geigon, *Proc. Natl. Acad. Sci. USA* 90 (1993) 3745.
- [3] K. Padmanabhan, K.P. Padmanabhan, J.D. Ferrara, J.E. Sadler, A. Tulinsky, *J. Biol. Chem.* 268 (1993) 17651.
- [4] P. Schultze, R.F. Macaya, J. Feigon, *J. Mol. Biol.* 235 (1994) 1532.
- [5] J.R. Wyatt, T.A. Vickers, J.L. Robertson, W.J. Buckheit, T. Klimkait, E. DeBaets, P.W. Davis, B. Rayner, J.L. Imbach, D.J. Ecker, *Proc. Natl. Acad. Sci. USA* 91 (1994) 1356.
- [6] V. Sasisekharan, S.B. Zimmermann, D.R. Davies, *J. Mol. Biol.* 92 (1975) 171.
- [7] T.J. Pinnavaia, H.P. Miles, E.D. Becker, *J. Am. Chem. Soc.* 97 (1975) 7198.
- [8] Henderson, E., Hardin, C.C., Walk, S.K., Tinoco Jr., I., Blackburn, E.H., *Cell* 51 (1987) 899.
- [9] J.R. Williamson, M.K. Raghuraman, T.R. Cech, *Cell* 59 (1989) 871.
- [10] D. Sen, W. Gilbert, *Nature (London)* 334 (1988) 364.
- [11] E.H. Blackburn, J.W. Szostak, *Annu. Rev. Biochem.* 53 (1984) 163.
- [12] E.H. Blackburn, *Annu. Rev. Biochem.* 61 (1992) 113.
- [13] E.H. Blackburn, *Nature (London)* 350 (1991) 569.

- [14] E.R. Henderson, E.H. Blackburn, *Mol. Cell. Biol.* 9 (1989) 345.
- [15] K. Walsh, A. Fualberto, *J. Biol. Chem.* 267 (1992) 13714.
- [16] A.M. Pearson, A. Rich, M. Krieger, *J. Biol. Chem.* 268 (1993) 3546.
- [17] W. Sundquist, S. Heaphy, *Proc. Natl. Acad. Sci. USA* 90 (1993) 3393.
- [18] J.M. Darlow, D.R.F. Leach, *J. Mol. Biol.* 275 (1998) 3.
- [19] C.H. Lin, D. Patel, *J. Chem. Biol.* 4 (1997) 817.
- [20] I.G. Panyutin, O.I. Kovalsky, E.I. Budowsky, R.E. Dickerson, M.E. Rikhirev, A.A. Lipanov, *Proc. Natl. Acad. Sci. USA* 87 (1990) 867.
- [21] P. Balagurumoorthy, S.K. Brahmachari, D. Mohanty, M. Bansal, V. Sasisekharan, *Nucl. Acids Res.* 20 (1992) 4061.
- [22] D. Mohanty, M. Bansal, *Nucl. Acids Res.* 21 (1993) 1767.
- [23] D. Mohanty, M. Bansal, *Biopolymers* 34 (1994) 1187.
- [24] D. Mohanty, M. Bansal, *Biophys. J.* 69 (1995) 1046.
- [25] W.S. Ross, C.C. Hardin, *J. Am. Chem. Soc.* 116 (1994) 6070.
- [26] J. Sponer, J.V. Burda, P. Mejzlik, J. Leszczynski, P. Hobza, *J. Biomol. Struct. Dyn.* 14 (1997) 613.
- [27] J. Sponer, J. Kypr, *J. Biomol. Struct. Dyn.* 11 (1993) 277.
- [28] T.E. Cheatham III, P.A. Kollman, *J. Mol. Biol.* 259 (1996) 434.
- [29] J. Tohl, W. Eimer, *Biophys. Chem.* 67 (1997) 177.
- [30] W. Guschlbauer, J.F. Chantot, D. Thiele, *J. Biomol. Struct. Dyn.* 8 (1990) 491.
- [31] C. Kang, X. Zhang, R. Ratliff, R. Moyzis, A. Rich, *Nature (London)* 356 (1992) 126.
- [32] K. Phillips, Z. Dauter, A.I.H. Murchie, D.M.J. Lilley, B. Luisi, *J. Mol. Biol.* 273 (1997) 171.
- [33] Strahan, G.D., Keniry, M.A., Shafer, R.H., *Biophys. J.* (1998) 968.
- [34] G. Laughlan, A.I.H. Murchie, D.G. Norman, M.H. Moore, P.C.E. Moody, D.M.J. Lilley, B. Luisi, *Science* 265 (1994) 520.
- [35] N. Nagesh, D. Chatterji, *J. Biochem. Methods* 30 (1995) 1.
- [36] N.V. Hud, P. Schultze, J. Feigon, *J. Am. Chem. Soc.* 120 (1998) 6403.
- [37] C.C. Hardin, M.J. Corregan, D.V. Lieberman, B.A. Brown II, *Biochemistry* 36 (1997) 15428.
- [38] N.V. Hud, P. Schultze, V. Sklenar, J. Feigon, *J. Mol. Biol.* 285 (1999) 233.
- [39] C. Harrington, Y. Lan, S.A. Akman, *J. Biol. Chem.* 272 (1997) 24631.
- [40] W.J. Hehre, L. Radom, P.R. Schleyer, J.A. Pople, *Ab Initio Molecular Orbital Theory*, Wiley, New York, 1986.
- [41] A.D. Becke, *J. Chem. Phys.* 98 (1993) 5648.
- [42] C. Lee, W. Yang, R.G. Parr, *Phys. Rev. B* 37 (1988) 785.
- [43] B. Miehlisch, A. Savin, H. Stoll, H. Preuss, *Chem. Phys. Lett.* 157 (1989) 200.
- [44] J. Gu, J. Leszczynski, *J. Phys. Chem. A* 103 (1999) 2744.
- [45] J. Sponer, J. Leszczynski, P. Hobza, *J. Phys. Chem.* 100 (1996) 1965.
- [46] M.J. Frisch, G.W. Trucks, H.B. Schlegel, P.M.W. Gill, B.G. Johnson, M.A. Robb, J.R. Cheeseman, T. Keith, G.A. Petersson, J.A. Montgomery, K. Raghavachari, M.A. Al-Laham, V.G. Zakrzewski, J.V. Ortiz, J.B. Foresman, J. Cioslowski, B.B. Stefanov, A. Nanayakkara, M. Challacombe, C.Y. Peng, P.Y. Ayala, W. Chen, M.W. Wong, J.L. Andres, E.S. Replogle, R. Gomperts, R.L. Martin, D.J. Fox, J.S. Binkley, D.J. Defrees, J. Baker, J.P. Stewart, M. Head-Gordon, C. Gonzalez, J.A. Pople, *Gaussian 94, Revision D.3*, Gaussian, Inc., Pittsburgh, PA, 1995.
- [47] S.F. Boys, F. Bernardi, *Mol. Phys.* 19 (1970) 553.
- [48] M. Bansal, M. Ravikiran, S. Chowdhury, in: J. Leszczynski (Ed.), *Computational Molecular Biology*, Elsevier, Amsterdam, 1999, pp. 279–323.
- [49] W.D. Cornell, P. Cieplak, C.I. Bayly, I.R. Gould, K.M. Merz Jr., D.M. Gerguson, D.C. Spellmeyer, T. Fox, J.W. Caldwell, P. Kollman, *J. Am. Chem. Soc.* 117 (1995) 5179.
- [50] K. Poon, R.B. Macgregor, *Biopolymers* 45 (1998) 427.