Nonempirical Calculations of a Hydrated RNA Duplex

Jürg Hutter,[†] Paolo Carloni,[‡] and Michele Parrinello*,[†]

Contribution from the Max-Planck-Institut für Festkörperforschung, D-70569 Stuttgart, Germany, IBM Research Division, Zurich Research Laboratory, Säumerstrasse 4, CH-8803 Rüschlikon, Switzerland, and Department of Chemistry, University of Florence, Via G. Capponi, 7, I-50121 Florence, Italy

Received April 12, 1996. Revised Manuscript Received June 25, 1996[⊗]

Abstract: We have performed density functional theory based ab initio calculations on the crystal structure of sodium guanylyl-3′-5′-cytidine (GpC) nonahydrate. Our calculations are in good agreement with the experimentally determined X-ray structure. This is one of the first attempts to model ab initio nucleic acids in laboratory-realizable conditions. Comparison is also made with empirical force field based structure calculations.

I. Introduction

Molecular dynamics simulations based on effective potentials have been crucial in understanding the properties of a large variety of biological systems.[1−5] This approach derives its strength from its suitability to study very large systems and to follow their evolution on a relatively long time scale. However, this approach is not devoid of problems. For instance, effective potentials present some difficulties in describing the structure of nucleic acids,[6] so much so that very recently an ad hoc reparametrization of the effective potentials to fit nucleic acids properties explicitly has been attempted.[7] However, only very limited results are yet available on the overall performance of these new potentials. Furthermore, it is becoming increasingly clear that there is the need to transcend the effective potential approach if one wants to study biological processes that involve a change of the chemical bond such as enzymatic reactions.[8] These are better and more reliably described by ab initio quantum-chemical approaches.

Owing to the size of the biological molecule, the quantum-chemical calculations have been confined to the study of fragments in vacuum.[9−21] This, however, is far from being actually relevant because water and the environment are known to play a crucial role in determining the structure, dynamics, and function of proteins and nucleic acids.[1,5] Nevertheless, progress in ab initio molecular dynamics combined with the power of parallel computing has dramatically increased the size of systems currently accessible. Keeping future applications to biochemical processes in mind, it is important to investigate the accuracy of ab initio methods to describe biologically relevant processes in as realistic an environment as possible. To this end we have studied the structure of sodium guanylyl-3′-5′-cytidine (GpC) nonahydrate, which has been determined by single-crystal X-ray diffraction.[22,24]

This structure is favorable in many respects: it has a large but manageable number of atoms (368), and yet it contains all the basic ingredients concerning the stability of the nucleic acid helix. It is a small segment of right-handed, antiparallel double-helical RNA, with Watson−Crick base pairing (Figure 1a). Therefore, it contains both the base−base and the base−sugar backbone interactions. Furthermore, it is fully hydrated, thus allowing a study of the hydration process, and it contains the counterions (Figure 1b). The combination of all these elements has never been investigated by fully ab initio methods. This also provides a stringent test of the ability of ab initio methods to describe nucleic acids.

II. Computational Section

Our calculations are performed within the framework of density functional theory and use the generalized gradient approximation. The

[†] Max-Planck-Institut für Festkörperforschung.

[‡] University of Florence. Present address: IBM Research Division, Zurich Research Laboratory.

[⊗] Abstract published in *Advance ACS Abstracts,* August 15, 1996.

(1) McCammon, J. A.; Harvey, S. C. *Dynamics of Proteins and Nucleic Acids*; Cambridge University Press: Cambridge, 1987.

(2) Brooks, C. L., III; Karplus, M.; Petitt, B. M. In *Proteins: A Theoretical Perspective of Dynamics, Structure and Thermodynamics*; Prigodinen, I., Rice, S. Eds.; Wiley Series on Advances in Chemical Physics, Vol. 71; Wiley: New York, 1988.

(3) van Gunsteren, W. F.; Berendsen, H. J. C. *Angew. Chem., Int. Ed. Engl.* **1990**, *298*, 992−1023.

(4) Kollman, P. A.; Merz, K. M., Jr. *Acc. Chem. Res.* **1990**, *23*, 246−252.

(5) Beveridge, D. L.; Swaminathan, S.; Ravishanker, G.; Withka, J. M.; Srinivasan, J.; Prevost, C.; Louise-May, S.; Langley, D. R.; DiCapua, F. M.; Bolton, P. H. In *Water and Biological Macromolecules*; Westhof, E., Ed.; MacMillan: London, 1993; p 165.

(6) McConnell, K. J.; Nirmala, R.; Young, M. A.; Ravishankar, G.; Beveridge, D. L. *J. Am. Chem. Soc.* **1994**, *116*, 4461−4462.

(7) MacKerell, A. D., Jr.; Wiórkiwicz-Kuczera, J.; Karplus, M. *J. Am. Chem. Soc.* **1995**, *117*, 11946−11975.

(8) Warshel, A. *Computer Modeling of Chemical Reactions in Enzymes and Solutions*; J. Wiley and Sons: New York, 1991.

(9) Clementi, E.; Mehl, J.; von Niessen, W. *J. Chem. Phys.* **1971**, *54*, 508.

(10) Del Bene, J. *J. Mol. Struct.*: *THEOCHEM* **1985**, *124*, 201.

(11) Hobza, P.; Sandorfy, C. *J. Am. Chem. Soc.* **1987**, *109*, 1302.

(12) Aida, M. *J. Comput. Chem.* **1988**, *9*, 362.

(13) Dive, G.; Dehareng, D.; Ghuysen, J. M. *Theor. Chim. Acta* **1993**, *85*, 409.

(14) Hrouda, V.; Florian, J.; Hobza, P. *J. Phys. Chem.* **1993**, *97*, 1542.

(15) Colson, A. O.; Besler, B.; Sevilla, M. D. *J. Phys. Chem.* **1993**, *97*, 13852.

(16) Gould, I. R.; Kollman, P. A. *J. Am. Chem. Soc.* **1994**, *116*, 2493.

(17) Nagata, C.; Aida, M. *J. Mol. Struct.: THEOCHEM* **1988**, *179*, 451.

(18) Aida, M. *J. Theor. Biol.* **1988**, *130*, 327.

(19) Aida, M.; Nagata, C. *Chem. Phys. Lett.* **1982**, *86*, 44.

(20) Hobza, P.; Sponer, J.; Polasek, M. *J. Am. Chem. Soc.* **1995**, *117*, 792−798.

(21) Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S.; Weiner, P. *J. Am. Chem. Soc.* **1984**, *106*, 765.

(22) Rosemberg, J. M.; Seeman, N. C.; Day, R. O.; Rich, A. *J. Mol. Biol.* **1976**, *104*, 145−167.

(23) Hydration patterns fairly consistent with experiment have been obtained also by Monte Carlo calculations on the RNA duplex in aqueous solution (Subramanian, P. S.; Pitchumani, S.; Beveridge, D. L.; Berman, H. *Biopolymers* **1990**, *29*, 771−783).

(24) X-ray structures have been determined also for other two salts of GpC: with ammonium (Aggarwal, A.; Islam, A. S.; Kuroda, R.; Sanderson, M. R.; Neidle, S. *Acta Crystallogr.* **1983**, *B39*, 98) and with calcium (Hingerty, B.; Subramanian, E.; Stellman, S. D.; Sato, T.; Broyde, S. B.; Langridge, R. *Acta Crystallogr.* **1976**, *B32*, 2998).

S0002-7863(96)01220-6 CCC: $12.00

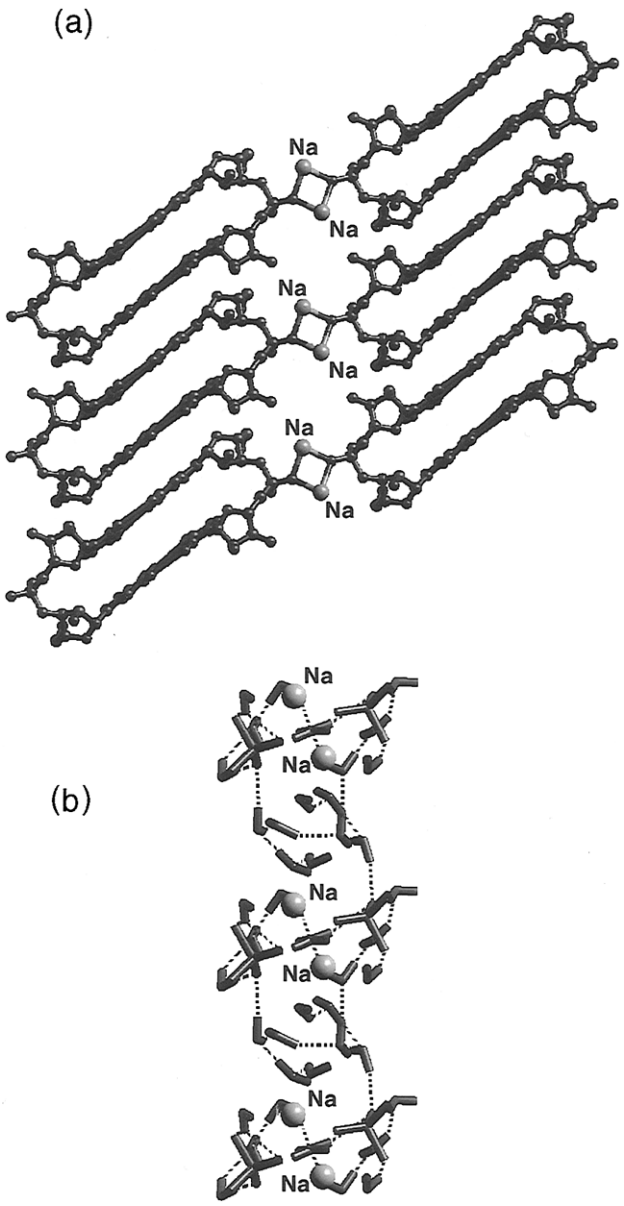

Figure 1. (a) Side view of the GpC crystal structure, which exhibits the stacking of the Watson−Crick base pairs. The water molecules and the hydrogen atoms are not shown. The two RNA fragments are held together by the sodium counterions shown in the picture. The interstice is filled with water. The structure of this water channel is shown in part b.

local functional for correlation in the Perdew−Zunger parameterization was used together with Becke's gradient-corrected exchange functional.[26] We treat explicitly only the 1160 valence electrons. The interaction between valence electrons and ionic cores is described by supersoft pseudopotentials of the Vanderbilt type.[27] The Kohn−Sham orbitals are expanded in plane waves up to an energy cutoff of 24 Ry, resulting in 45477 degrees of freedom per state. This scheme has been tested elsewhere.[30,31] We have used the CPMD[32] code, and optimized the structure using a combination of DIIS for electronic minimization and a Newton−Raphson method for ionic relaxation. This program uses periodic boundary conditions; the Coulomb interactions are evaluated by the Ewald sum method. No symmetry restriction was imposed on the calculation. We stopped the relaxation when the root-mean-square value of the force was less than 10^{-3} au. We estimate that the resulting uncertainty in the position is less than the experimental error.

Table 1. Root Mean Square Deviations (Å) (Heavy Atoms) with Respect of the X-ray Structure for the Elements Contained in the Unit Cell of GpC, Namely the Two RNA Duplexes, the 16 Water Oxygens, and the 4 Sodium Ions

	RNA	water	sodium
ab initio calculations	0.40	0.31	0.37
force field calculations	1.01	0.93	0.87

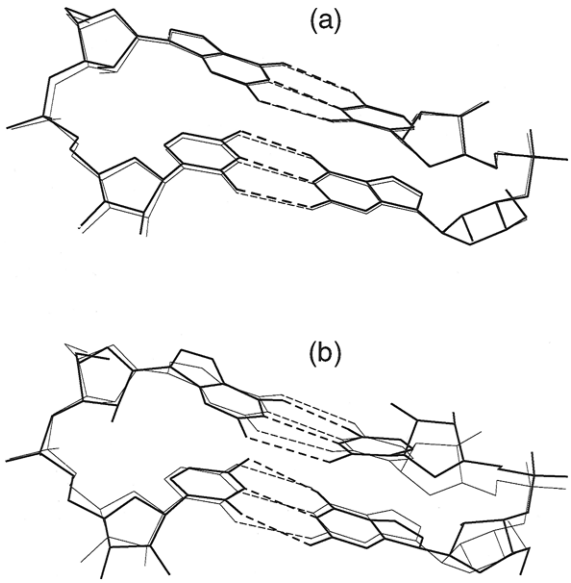

Figure 2. Comparison of the X-ray structure of the GpC duplex (thin lines) with (a) the ab initio structure (thick lines) and (b) the force field structure (thick lines).

The crystal structure of sodium GpC nonahydrate, as determined in ref 22, contains four molecules per monoclinic unit cell. The space group is *C*2, with cell dimensions $a = 21.460$ Å, $b = 16.927$ Å, $c = 9.332$ Å, and $\beta = 90.54°$. The cell parameters were not optimized, and β was set to 90° for computational convenience. The coordinates were taken from the Cambridge Data Base.[33] The hydrogen atom positions, not resolved in the X-ray structure, were given in an arbitrary way, respecting only the constraint of standard bond angles and bond lengths. This yielded a rather unlikely initial configuration in which the water dipoles were pointing in the same direction. This has been meant to be a test of the ability of our scheme to generate spontaneously a hydrogen bond network (Figure 1b). As we shall demonstrate below, this test has been successful. In this respect the ab initio approach appears to be more robust than the effective potential, which at times has difficulties dealing with high-energy starting configurations.[34]

III. Results

The root-mean-square deviations from experiment are shown in Table 1 for the RNA moiety. The water oxygens and the sodium counterions are compared with the results of a standard force field model.[35]

(25) Perdew, J. P.; Zunger, A. *Phys. Rev. B* **1981**, *23*, 5048.
(26) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098.
(27) Vanderbilt, D. *Phys. Rev. B* **1990**, *41*, 7892.
(28) Sponer, J.; Florian, J.; Leszczynski, J.; Hobza, P. *J. Biomol. Struct. Dyn.* **1996**, *13*, 827.
(29) Florian, J.; Baumruk, V.; Strajbl, M.; Bednárová, S. J. *J. Phys. Chem.* **1996**, *100*, 1559.
(30) Laasonen, K.; Sprik, M.; Parrinello, M.; Car, R. *J. Chem. Phys.* **1993**, *99*, 9081.
(31) Tuckerman, M.; Laasonen, K.; Sprik, M.; Parrinello, M. *J. Chem. Phys.* **1995**, *103*, 150−161.
(32) CPMD version 2.5, written by Jürg Hutter, Max-Planck-Institut für Festkörperforschung, Stuttgart, 1995, with the help of the group for numerically intensive computations of IBM Research Laboratory Zurich and the Abteilung Parrinello of MPI Stuttgart.
(33) Allen, F. H.; Davies, J. E.; Galloy, J. J.; Johnson, O.; Kennard, O.; Macrae, C. F.; Mitchell, E. M.; Mitchell, G. F.; Smith, J. M.; Watson, D. G. *J. Chem. Inf. Comp. Sci.* **1991**, *31*, 187−204.
(34) The atomic coordinates of our optimized structure are available by anonymous ftp (address: parrix1.mpi-stuttgart.mpg.de, directory pub/outgoing).

Table 2. Backbone Torsion Angles for the Two RNA Duplexes

9	$\chi(5')$	$\xi(5')$	α	β	γ	δ	ϵ	$\chi(3')$	$\xi(3')$
X-ray	13	89	211	292	285	184	50	32	77
ab initio	14(2)	85(1)	210(5)	298(6)	281(4)	193(4)	49(1)	37(4)	81(4)
force field	15(20)	115(4)	213(10)	285(8)	238(17)	184(6)	98(28)	51(13)	105(5)

[a] For the ab initio and force field structures, the average values of the four GpC moieties are reported. Standard deviations are also reported in parentheses. For the definitions of the torsional angles, see ref 36.

We do not compare our results with the very recent parametrization in ref 7 because more extensive tests are necessary to establish its transferability.

We now examine the various structural building blocks of the crystal in more detail. First, and most important, we reproduce with good accuracy the Watson−Crick hydrogen bond distances (see Figure 2) and the conformation of the bases. A measure of the planarity of the nucleobase rings is given by the maximum deviation with respect to the ideal value of 180°. We find that this value is 5° for the ab initio and 10° for the force field-based structure.

These small but significant deviations between ab initio and force field structures are due to the well-known difficulty for force fields to ensure the planarity of the aromatic rings.[1] This is one effect of electronic origin, which in the force field method is mimicked by the so-called "improper" torsional forces.[1]

Another important feature of the conformation of two complementary nucleobases is their noncoplanarity.[36] The value of the "propeller twist" angle[28,36] of the X-ray structure of GpC is 9°. In the case of our optimized structure, we find values between 7° and 9°. For the force field structure, the values are larger, between 10° and 17°.

The other important building block is the sugar−phosphate backbone, whose torsional angles are crucial in determining the RNA structure. As the corresponding torsional barriers are very small (typically of the order of 1 kcal/mol),[29] they are rather difficult to model with effective potentials. The results are shown in Table 2. We note that they are again in very good agreement with experiment and appear to give a better description than do molecular models.

Finally, the hydrogen bond network obtained in their relaxation process is identical to that postulated by the X-ray crystallographers.[22,23] The water molecules form hydrogen bonds with themselves as well as with the RNA moiety. This can be considered a real prediction because, as stressed above, protons are invisible to X-rays.

IN our calculations we find that some of the nucleobase amino groups are distorted: the hydrogens are out of the plane (maximum displacement: 17° for a guanine) to form hydrogen bonds with a neighboring water. Again, this effect is missed by the effective potential.

Our calculations also reproduce well the highly symmetric octahedral solvation structure of the sodium ions, which is more distorted in the effective potential calculations.

IV. Concluding Remarks

In conclusion, we have shown that it is now possible to perform ab initio simulations of molecules whose size approaches that relevant for biologically interesting systems. The quality of the results is very high even in the case of nucleic acids, which have traditionally been very difficult to model.[6] The ab initio modeling automatically includes all the physical and chemical effects that are so difficult to mimic in effective potential simulations, such as polarization effects, many-body forces, and the rigidity of aromatic rings. For instance, in our simulations the electronic structure of water is modified by the local environment, giving different properties to different water molecules.

Another important advantage to the ab initio method is that no painstaking parametrization is needed to extend the domain of applicability of the theory. We know it works for water,[30] water solutions,[31] and peptide bonds.[37] We have shown here its validity for the study of nucleic acids. It is expected that it will work with similar accuracy in a very large variety of biological environments, and have considerable predicting power. In particular, one can reliably model bioinorganic molecules, which contain transition metal ions, such as metalloproteins and metal-based drug−DNA adducts.[38]

Furthermore, and this is at odds with most effective potentials and ab initio codes, the Coulombic forces are calculated using the Ewald summation procedure. We believe that this is very important for periodic systems and that the use of cutoff in the Coulombic interactions introduces spurious effects.[39]

Of course, these kinds of simulations are several orders of magnitude more demanding than those based on effective potentials. However, progress in computer architecture and in the algorithms used gives us confidence that this gap can be narrowed in a short span of time and that a new dimension can be added to the simulation of biochemical processes.

Acknowledgment. We thank Pietro Ballone for many useful discussions. Marta Ferraroni is acknowledged for the use of the XPLOR program at EMBL, Hamburg, Germany. We also thank the Maui High Performance Computing Center (MHPCC) for a generous allocation of computing time on their IBM SP2 computer.

JA9612209

(35) For our classical simulations, we have used the XPLOR 3.1 package running on a Silicon Graphics Iris. The standard united atom CHARMM force field, available in the package, was used for the GpC and the sodium ions.[40] For water, we have used the TIP3P model.[41] The unit cell was duplicated along the *c* direction, in order to be able to use a larger nonbonding interaction cutoff. A cutoff of 8 Å was used. The crystal packing interactions are calculated in XPLOR with the periodic image convention. The geometry optimizations were performed with the conjugate gradient method up to an energy gradient of less than 1×10^{-3} kcal/mol for all the atoms.

(36) Saenger, W. *Principles of Nucleic Acid Structure*; Springer: New York, 1984.

(37) Parrinello, M.; Carloni, P.; Andreoni, W. Paper in preparation.

(38) Carloni, P.; Andreoni, W. Submitted for publication.

(39) Saito, M. *J. Chem. Phys.* **1994**, *101*, 4055−4061.

(40) Brooks, B.; Bruccoleri, R.; Olafson, B.; States, D.; Swaminthan, S.; Karplus, M. *J. Comp. Chem.* **1983**, *4*, 187−217.

(41) Jorgensen, W.; Chandrasekar, J.; Madura, J.; Impey, R.; Klein, M. *J. Chem. Phys.* **1983**, *79*, 926−935.