RZ 2830 (# 90133) 06/03/96
Chemistry 7 pages

Research Report

The Structure of a Drug-Nucleic Acid Complex: Density Functional Theory and Force-Field-Based Calculations

P. Carloni[1,2] and W. Andreoni[1]

[1]IBM Research Division
Zurich Research Laboratory
8803 Rüschlikon
Switzerland

[2]Department of Chemistry
University of Florence
Via G. Capponi, 7
50121 Florence
Italy

IBM Research Division
Almaden · T.J. Watson · Tokyo · Zurich

The Structure of a Drug-Nucleic Acid Complex: Density Functional Theory and Force-Field-Based Calculations

P. Carloni[1,2] and W. Andreoni[1]

[1] *IBM Research Division, Zurich Research Laboratory, 8803 Rüschlikon, Switzerland*

[2] *Department of Chemistry, University of Florence, Via G. Capponi, 7, 50121 Florence, Italy*

Abstract: The structural and bonding properties of an RNA–intercalator complex (2 DMPFL –2 5-iodo-cytidylyl(3'-5')guanosine • $16H_2O$) are investigated using both density functional theory (DFT) with gradient-corrected exchange and correlation functionals and force-field (FF) methods. Effects of temperature and hydration are explored using the empirical approach. A comparison is made between the DFT and FF structures and the one resulting from X-ray diffraction data. A consistent picture for the mode of action of the drug is derived that clarifies its crucial dependence on the water content.

INTRODUCTION

Understanding the action of nucleic-acid binding drugs at the microscopic level is very important for progress in modern pharmacology. Current knowledge relies on structural data obtained with X-ray diffraction[1] and NMR spectroscopy,[2] and on computer modeling based on the techniques of classical molecular mechanics (MM) and molecular dynamics (MD) calculations.[3]. The difficulties in developing accurate force-fields and in particular in properly describing the electrostatic interactions in the field of nucleic acid–drug interactions have recently been pointed out.[3]

Some of these difficulties could in principle be overcome by using *ab initio* calculations. The advantages are evident: from a fundamental point of view, the description of the interatomic interactions does not have to rely on a given *a priori* approximation, and from a practical point of view, there is no need to readjust the parameterization for each new specific drug of interest. However, it is also clear that before such computations become as efficient and as useful as classical approaches currently are, many steps have to be taken.

Ab initio approaches have so far been mainly hampered by the large size of the systems of interest, which involve costly time- and memory-intensive calculations both for the statics and the dynamics. Thanks to the advent of parallel computers and the progress of computational codes, we can now afford to treat systems that are still small compared to biologically relevant ones but cover some crystal structures of drug-nucleic acid complexes commonly synthesized in the laboratory. The accuracy of a nonempirical scheme, in fact, is not guaranteed *a priori*. It may still critically depend on the approximations used for the electron–electron interaction, and on the basis set adopted for the electron wave functions. Accordingly, specific tests are required on real systems for which experimental data are available. This is the advantage that computer-intensive calculations can now provide, especially compared to the *ab initio* studies available so far, which have been limited to oversimplified models of nucleic acid fragments.[4–8]

Here, we consider a crystal complex,[9] namely N,N,dimethyl proflavine (DMPFL) intercalated into an anti-parallel Watson–Crick base-paired duplex.[10] We make an energy minimization using local optimization procedures: a nonempirical one based on density-functional theory (DFT) and one based on classical force fields. We compare the resulting structure with experiment. This study allows us to investigate the accuracy of the nonempirical method, to get a feeling for the differences between the results obtained with it and those obtained with the force-field-based approach, and to find out whether there is a discrepancy between the two descriptions of the mode of action of the drug. In particular, both methods show the formation of an H-bond between the intercalating drug and one phosphate group of the nucleic acid duplex, in contrast with what X-ray data indicate.[9] Several reasons can cause the lack of agreement between theory and experiment. As the *ab initio* results concerning this feature validate the MM static picture, we use classical simulations to test our hypotheses.

DEFINITION OF THE SYSTEM AND COMPUTATIONAL METHODS

The crystal structure of 2 DMPFL–2 5-iodo-cytidylyl(3'-5')guanosine • $16H_2O$[9] was taken from the X-ray diffraction data of Bhandary *et al.*,[11] namely the lattice is triclinic (a = 11.78 Å, b = 14.55 Å, c = 15.50 Å, $\alpha = 89.20°$, $\beta = 86.20°$, $\gamma = 96.40°$), and the space group is P1. In particular, the unit cell contains two DMPFL molecules, one of which intercalates between the dinucleoside monophosphate duplex, and the other is stacked on the exterior, and sixteen water molecules. In total, the system is composed of 134 heavy atoms and 110 hydrogens. X-ray refinement obviously only provides information on the positions of the heavy atoms. For both

DFT and force-field-based energy minimizations, we used the X-ray structure supplemented by an educated guess for the hydrogens as a starting configuration. Within the DFT approach, we performed an additional minimization, starting from the result of the classical modeling, as an additional test.

For the classical calculations,[12] we adopted the "united atoms" CHARMM[13] force-field for the RNA duplex except for the 5-iodo-cytosine ring (IC), for which a simple parameterization scheme had to be constructed.[14] For the intercalator, our parameterization followed a procedure similar to that introduced by Boehncke *et al.*[15] in the case of a dystamycin/DNA complex.[16] For the water molecules, the TIP3P model[17] was assumed. Coulomb interactions were calculated with the minimum-image convention and a cutoff scheme for the nonbonded interaction, consistent with a "supercell" of eight unit cells. A cutoff value of 10 Å was chosen. Geometry optimization was carried out with the conjugate gradient method.[18]

DFT calculations were performed using periodic boundary conditions. For the exchange and correlation functionals, the local density approximation (LDA) was used, augmented by gradient corrections, which are necessary for an accurate description of hydrogen-bonded systems.[19] We have performed separate calculations both with the local density plus Becke–Perdew[20,21] (BP) approximation and with the Becke–Lee–Yang–Parr (BLYP) one,[20,22] and found that differences in the final structural parameters were negligible. *Ab initio* atomic angular-momentum-dependent soft-core pseudopotentials of the Martins–Trouiller type[23] describe the core–valence interaction. The electronic wave function basis set consists of plane waves up to an energy cutoff of 70 Ry. Electronic wave functions are optimized with the direct inversion in iterative subspace method in conjunction with a preconditioning scheme,[24,25] the geometrical parameters with a combination of it and the quasi-Newton procedure.[25]

RESULTS AND DISCUSSION

We have performed local structural minimization starting from the X-ray-refined atomic positions and by making an educated guess for the hydrogens, in both the empirical and the nonempirical framework. The nonempirical calculations are made both in the BP and BLYP approximations.[20–22]

The two schemes resulted in different geometries, with the classical one deviating more significantly from the original atomic positions; the MM structure turned out to be 1250 kcal/mol higher in energy than the DFT one. For these reasons, we attempted another DFT optimization starting from the MM structure. However, the system was trapped in a high energy minimum, at 100 kcal/mol. It is obvious in fact that single-point energy minimizations are unable to provide the global minimum and that the agreement with the X-ray refinement, if this is taken as the starting structure, may simply be due to the limited freedom left to the system to relax. This for instance occurs for sugar puckers, the confirmations of which are separated by interconversion energy barriers of a few kcal/mol.[26] However, these calculations are valid to test and compare the two frameworks, the minimization having been made in the same way. The question to ask at the moment is not how the two global minima compare but whether, near the experimental structure, there is a minimum of the theoretical potential energy surface. If there is not, the calculated forces will cause the system to move away from the starting geometry, and thus something important may be missing in the model.

A. RNA duplex

The root-mean-square deviations (rmsd) with respect to the X-ray structure are 0.2 and 1.3 Å for the DFT and MM structures, respectively. The experimental value for the distance between the two phosphorus atoms, which measures the width of the dinucleoside monophosphate duplex, is 16.6 Å. DFT calculations do not change it, and MM elongates it only by 0.3 Å. The backbone torsion angles, which are a crucial test of any modeling, are compared in Table I. Deviations between MM predictions and experiment are generally large. DFT values, however, do not exhibit significant deviations.

The H-bond interaction of the two Watson–Crick C–G base-pairs [Fig. 1(a)] is characterized by H–O and H–N distances whose calculated values are in the range of 1.7–1.85 Å in DFT, as expected,[26] and lie between 1.7 and 1.9 Å in the MM structure. Experimentally, the conformation of the nucleobase amino groups, and in particular how it is affected by the presence of H-bond interactions, is not known. In general, this issue is of fundamental importance because the amino groups are involved in key interactions of information transfer processes (such as DNA duplication or RNA transcription)[26] and drug/DNA interactions.[27] DFT calculations show that H-bond interactions are dominant in this particular system and render the amino-group hydrogens not coplanar with the nucleobases. Deviations from planarity are up to 7°, except for the $-NH_2$ group of guanine of strand 2 (29°) that interacts with a water molecule (d(H–O) = 1.9 Å, d(N–O) = 2.9 Å). In the MM structure, instead, the maximum deviation is 4°; the guanine amino hydrogen deviates by 4°, and the H-bond with the water molecule is not formed (d(H–O) = 2.8 Å, N–O = 3.0 Å). Note that the experimental value for the N–O distance is 3.0 Å, which is well reproduced in both models. However, the water–guanine interaction is dramatically different. The missing H-bond in the MM geometry could be an intrinsic effect of the force field, modeled in such as way as to keep the amino groups coplanar with the nucleobases. We tried to see whether including temperature effects would alter this static picture. However, MD runs at room temperature resulted in a unphysical dynamical behavior for the water molecules. In fact, simulations suffered from their high dilution, so that the water molecules did not keep their crystallographic positions (see discussion below).

B. DMPFL molecules

There are two DMPFL molecules in the unit cell. One intercalates between the RNA duplex, and the other is located at the exterior. The rmsd deviations from the X-ray data for the intercalator are 0.2 Å (DFT structure) and 1.2 Å (MM structure), and 0.1 Å (DFT) and 1.8 Å (MM) for the other DMPFL molecule. We focus here on the intercalator.

Both DFT and MM calculations reproduce the experimental observation of H-bonds between the amino group and a water molecule (d(O–N) = 3.0 Å (Exp), 3.0 Å (DFT); 2.9 Å(MM)). Both theoretical structures, however, differ from experiment in the description of the amino–phosphate interaction, for which calculations predict the presence of one hydrogen bond (d(O–N) = 3.0 Å for both DFT and MM) not emerging from the X-ray refined structure (d(O–N) = 3.4 Å).

C. Drug–RNA interaction and the drug mode of action

The occurrence of an H-bond between the amino group and the nucleic acid phosphate backbone is important for characterizing the mode of action of the intercalating drug. The common knowledge, based on X-ray refined structures, is that in proflavine the free amino groups H-bond to phosphate oxygens ("complex intercalation"), whereas in acridine orange (N,N-N',N' tetramethyl proflavine) the methyl groups interact weakly with the hydrophobic pocket of the nucleic acid fragment ("simple intercalation")[9]. The system we consider here is indeed intermediate between these two, with one $-N(CH_3)_2$ and one $-NH_2$ group [Fig. 1(b)]. Following our discussion above, the X-ray refined parameters indicate that the $-NH_2$ group in DMPFL does not H-bind to the RNA duplex either, whereas our static calculations do. Why is there this discrepancy between theory and experiment?

One possibility is that the amino group fluctuates at room temperature, and/or that the 16 crystallographic water molecules reorient in such a way as to suppress the amino–phosphate H-bond. These effects are clearly neglected in our calculations. Another possibility is that other water molecules are responsible for this. They may have well eluded experimental observation owing to structural and/or dynamical disorder[28]. Indeed, this is consistent with the topology of the Connolly surface[29] of the X-ray structure that we have calculated. As shown in Fig. 2, the existence of hydration sites[30] located between the amino and the phosphate group is apparent. As both DFT and classical modeling predict the formation of this hydrogen bond, we have tested both hypotheses with classical MD.

Simulation of the system with the 16 crystallographic water molecules showed that at room temperature they diffuse in the crystal and the H-bond still forms (average N–O distance 2.9 Å). Instead, structural optimization performed at various water concentrations revealed the important role of hydration. In fact, only at high levels do we observe no formation of the H-bond (d(N–O) = 3.6 Å) and are the positions of the crystallographic water molecules maintained during the dynamics. The structure of the adduct in the hydrated system retained the features of the originally optimized geometry of the classical modeling. The correspondence of this structure to experiment was only slightly better (rmsd = 1.1 and 0.7 Å for the RNA duplex and the intercalator, respectively).

Using the DFT approach, we have also investigated the nature of the drug–RNA bonding by means of the difference density $\Delta\rho(\vec{r}) = \rho_{\mathrm{complex}}(\vec{r}) - \rho_{[\mathrm{RNA}]^{2-}}(\vec{r}) - \rho_{2[\mathrm{DMPFL}]^{+}}(\vec{r})$, which describes the redistribution of the electron density $\rho(\vec{r})$ involved in the formation of the complex from the charged fragments. The dominant electrostatic character was apparent, thus confirming the commonly accepted picture of the intercalators.[26]

CONCLUSIONS

In summary, we have investigated structural and bonding properties of a drug-RNA duplex complex, using DFT and force-field-based calculations, and compared their outcome. Within the limits of single point energy minimizations, the two approximations resulted in a non-negligible difference concering the structural characterization but yielded a similar prediction for the drug-RNA binding. Comparison with experiment has forced us to consider further the effects of temperature and of hydration using classical simulations. In particular, the latter suggested that the hydration level plays a crucial role in determining the mode of action of the dye. It would be interesting to test this proposal with experiments made at low temperature (see, e.g., Ref. 28).

Acknowledgment. We are grateful to M.L. Klein, M. Boero and P. Guéret for useful discussions.

[1] Landolt–Börnstein *Numerical Data and Functional Relationships in Science and Technology*; Saenger, W., Ed., Springer: Berlin, 1989; Vol. VII/1b, Chapter 2.6, p. 247, and references therein.

[2] Searle, M. S. *Progr. NMR Spectros.* **1993**, *25*, 403, and references therein.

[3] 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, Chapter 6, and references therein.

[4] Prive, G. G.; Yanagi, K.; Dickerson, R.E. *J. Mol. Biol.* **1991**, *217*, 177.

[5] Clementi, E.; Mehl, J.; von Niessen, W. *J. Chem. Phys.* **1971**, *54*, 508.

[6] Del Bene, J. *J. Mol. Struct.: THEOCHEM* **1985**, *124*, 201.

[7] Aida, M. *J. Comput. Chem.* **1988**, *9*, 362, and references therein.

[8] Hobza, P.; Sponer, J.; Polasek, M. *J. Am. Chem. Soc.* **1995**, *117*, 792.

[9] Bhandary, K. K.; Sakore, T. D.; Sobell, H. M.; King, D.; Gabbay, E. J. *J. Biomol. Struct. Dyn.* **1984**, *1*, 1195.

[10] Lerman, L. S. *J. Mol. Biol.* **1961**, *3*, 18.

[11] Entry PFLCPC10 of the Cambridge Structural Database (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).

[12] We used the package X-PLOR 3.1, 1992, Brünger A. T., the Howard Hughes Medical Institute and Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT, USA.

[13] Brooks, B. R.; Bruccoleri, R. E.,; Olafson, B. D.; States, D.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.*, **1983**, *4*, 187.

[14] IC differs from the standard residue cytosine for the substitution of the H(5) with an iodine. For the latter, we have used the all-atom CHARRM force-field parameters of H(5)[13] except for the nonbonded interaction parameters. The Lennard–Jones parameters were those of bromine, and the charge was set to zero. The stability of the model with respect to changes in the value of the atomic charges was checked.

[15] Boehncke, K.; Nonella, M.; Schulten, K.; Wang, A. J. H. *Biochem.*, **1991**, *30*, 5465.

[16] The CHARMM[13] "united atoms" protein force-field was used. A planarity restraint of 10 kcal/mol was imposed on all atoms to penalize out-of-plane conformations. Following Kollman's scheme (Singh, U. C.; Kollman, P. A. *J. Comp. Chem.*, **1984**, *5*, 129), we obtained the partial atomic charges from the electrostatic potential generated by MNDO calculations. We ensured that the variation of charges derived from other semiempirical schemes, such as INDO and AM1, was negligible.

[17] Jorgensen, W. L.; Chandrasekhar, J.; Madura, J.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926.

[18] Classical molecular dynamics simulations were also performed at 300 K for 75 ps. A time step of 1.5 fs was chosen. All bond lengths were constrained with the SHAKE algorithm (Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327). The system was coupled to a thermal bath (Berendsen, H. J. C.; Postma, J. P. M.; van Gusteren, W. N.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684) of $T = 300$ K with a relaxation time of 0.2 ps.

[19] Laasonen, K.; Sprik, M.; Car, R.; Parrinello, M. *J. Chem. Phys.* **1993**, *99*, 9080.

[20] Becke, A. *Phys. Rev. A* **1988**, *38*, 3098

[21] Perdew, J. P. *Phys. Rev. B* **1986**, *33*, 8822; Erratum: *ibid.*, *34*, 7406.

[22] Lee, C.; Yang, W.; Parr, R. G.; *Phys. Rev. B* **1988**, *37*, 785.

[23] Troullier, N.; Martins, J. L. *Phys. Rev. B* **1992**, *46*, 1754.

[24] Pulay, P. *Chem. Phys. Lett.* **1980** *73*, 393.

[25] Hutter, J.; Lüthi, H. P.; Parrinello, M. *Comput. Mater. Sci.* **1994**, *2*, 244.

[26] Saenger, W. *Principles of Nucleic Acid Structure*, Springer: New York, 1984.

[27] Bailly, C.; Waring, M. J. *J. Am. Chem. Soc.* **1995**, *117*, 7311.

[28] Schneider, B.; Ginell, S. L.; Berman, H. M. *Biophys. J.* **1992**, *63*, 1572.
[29] Connolly, M. L. *Science* **1983**, *221*, 709.
[30] Clementi, E. in *Structure and Dynamics of Nucleic Acids and Proteins*; Clementi, E.; Sarma, R. H., Eds.; Adenine: New York, 1983, p. 321.

TABLE I. Torsion angles (in degrees) of the RNA duplex. Definition of the torsion angles as in Ref. 26.

	$\chi^{5'}$	γ	δ	ϵ	ζ	α	β	γ	δ	$\chi^{3'}$
X-ray structure										
strand 1	−164	51	84	−139	−71	−56	−131	52	142	−79
strand 2	−159	66	80	−156	−49	−76	−126	39	140	−85
DFT structure[a]										
strand 1	−160	57	75	−142	−72	−64	−139	67	147	−83
strand 2	−160	56	85	−156	−49	−90	−110	54	135	−81
MM structure										
strand 1	−146	77	104	−162	−134	−170	−143	−170	108	−136
strand 2	−164	59	97	−136	−61	−81	−121	40	120	−65

[a]Results of the BLYP calculation.

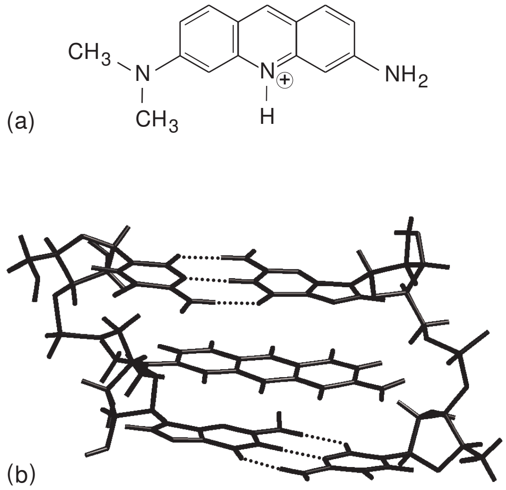

FIG. 1. (a) DFT Optimized structure of the intercalator-RNA complex. H-bonds are indicated by dashed lines. For the sake of clarity, the water molecules and the DMPFL located at the exterior are not shown. (b) Chemical structure of DMPFL.

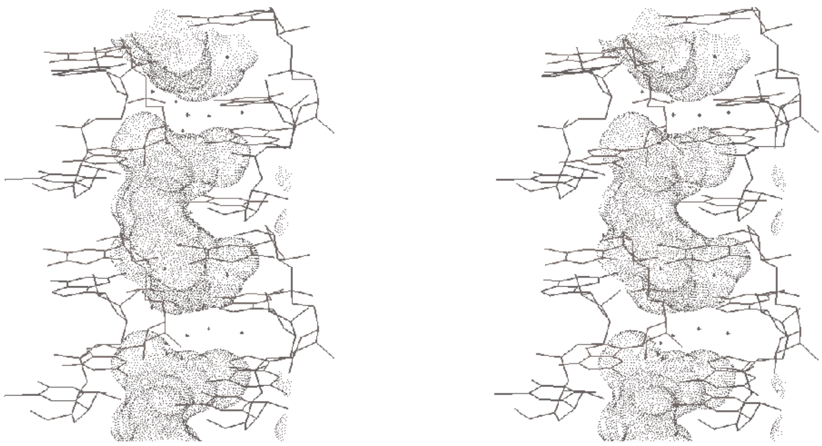

FIG. 2. Connolly surface calculated for the X-ray structure of the 2DMPFL–2ICpG complex. H atoms were added to the crystal structure, assuming standard bond lengths and bond angles.