HF-STEX and RASSCF calculations on nitrogen K-shell X-ray absorption of purine base and its derivative

Yuji Mochizuki,[a]* Hiroshi Koide,[a] Toshiyuki Imamura[a] and Hiroshi Takemiya[a]

[a]*CCSE, Japan Atomic Energy Research Institute, Nakameguro 2-2-54, Meguro-ku 153-0061, Tokyo, Japan. E-mail: mochi@koma.jaeri.go.jp*

The nitrogen K-shell X-ray absorption spectra of the purine bases present in nucleic acids, adenine and guanine, were analyzed by using ab initio Hartree-Fock static exchange and restricted-active-space self-consistent-field calculations. A variety of derivative molecules were calculated to investigate the energetic shifts due to environmental effects on the nitrogen atoms. Shake-up excitations were also addressed.

Keywords: ab initio calculation; **purine base**; **nitrogen K-shell**; **environmental effect**; **shake-up.**

1. Introduction

Nucleic acids (DNA and RNA) are organic molecular systems crucial for proteins and living matter. There has been a sizable amout of studies about the electronic aspect of nucleic acids, most of which having concerned the valence electrons. The issue of thymine dimerization involving skin cancer induced by ultraviolet light can be mentioned as a well-known example in this respect. In order to study the electronic environment of nitrogen atoms in nucleic acids, Kirtley *et al.* (1992) performed near-edge X-ray absorption spectroscopoy (XANES) at the nitrogen K-edge, where a series of base molecules and related derivatives were used as samples. The chemical shifts of the nitrogen 1s core electron excitation energies were then focused on. The lowest or edge peak was attributed to 1s $\rightarrow \pi*$ excitation. For the K-edge position of adenine and guanine (or purine bases), Kirtley *et al.* investigated the shifts by bromine and/or glycosidic substitutions on the five-membered ring. They took interest also in the differences between adenine and guanine spectra due to oxygen subsitution on the six-membered ring. Furthermore, hydrogen-bondings were also addressed by considering the double-helix in natural forms of nucleic acids. Kirtley *et al.* assigned 1s $\rightarrow \sigma*$ excitations to higher lying broad bands using a couple of model compounds. To the author's knowledge, no theoretical study has been reported on these nitrogen K-shell spectra of nucleic bases.

In this paper, the nitrogen K-edge spectra of purine bases and related derivatives are investigated by ab initio Hartree-Fock static exchange (HF-STEX) (Ågren *et al.*, 1997) calculations. Environmental effects both by substitutions and by water-solvations are of main interest. Solvation was modelled explicitly by inclusion of water molecules, as was done in a previous investigation which illuminated the solvation effect on the sulphur K-edge of cysteine (Mochizuki *et al.*, 1999). Core-hole type restricted-active-space self-consistent-field (RASSCF) calculations (Ågren & Jensen, 1993) are additionally performed to check the screening of the $\pi*$ orbital. Shake-up excitations are treated also by RASSCF.

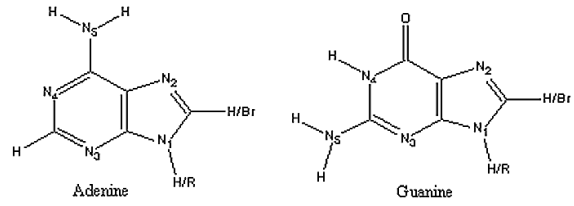

Figure 1
The numbering of five nitrogen atoms in adenine and guanine. Substitution points are also shown (refer to texts).

2. Calculation methods

Geometries of all molecules (the names of which are listed in the next section) were fully optimized by the standard HF/6-31G procedure provided by the Gaussian94 program package (Gaussian Inc., 1995). HF-STEX calculations with the so-called "dual basis sets technique" (Ågren *et al.*, 1997) were extensively carried out by using a local version of the DALTON 1.01 program (Department of Chemistry, University of Oslo, 1997). The calculation scheme is essentially the same as was done in the previous study on cysteine (Mochizuki *et al.*, 1999), where five nitrogen atoms in the base were treated independently. Figure 1 illustarates the numbering of nitrogen atoms. The basis set quality in the core-hole state optimizations was "triple-zeta plus polarization" for the nitrogen atoms, "double-zeta plus polarization" for the hydrogen atoms, and "double-zeta" for the other atoms. To describe excited orbitals, a set of diffuse 24s24p24d functions was augmented on the target nitrogen atom. Although a manifold of excited states are obtained by diagonalizing the STEX hamiltonian, the lowest energies of five nitrogen atoms having chemically different environments will mainly be discussed in this paper because they appear with the strongest intensity in the spectra.

Since there are many molecules studied each associated with many chemically shifted nitrogens, HF-STEX calculations were processed in a parallelized integral-driven way. To use the computational resources efficiently, the authors modified the message-passing interface (MPI) parts in the original DALTON codes by introducing the Stampi library (Japan Atomic Energy Research Institute, 1998) by which heterogeneous computers can be concertedly used via messages for the single HF-STEX calculation. Some portion of the calculations were carried out by the combination of NEC SX vector-parallel machine and IBM SP scalar-parallel machine, where NEC SX was used as the master which governs a job including diagonalizations suited for the vector processing.

In the HF-STEX approximation the excited electron is determined in the potential of the remaining N-1 electron core hole ion. If the excited orbital has a sizable amplitude in the valence regions a screening can take place, leading to an energetic lowering from the static exchange value. With the RASSCF technique, Yang *et al.* (1997) checked the screening contribution for the $\pi*$ levels at the carbon K-edge of substituted benzene molecules. There exists resemblance between benzene and purine as the conjugated molecules, and for the purpose of studying this shift for adenine and guanine molecules, RASSCF calculations were performed in the present work, imposing planar forms to identify σ and π symmetries in these molecules. Basis set types were almost the same as the valence sets for HF-STEX, but with Rydberg-p functions added on all the nitrogen and carbon atoms. Corresponding HF-STEX

calculations were done with the augmentation of 24s24p24d sets. Shake-up states were solved also by RASSCF under spin-doublet conditions. The numbers of electrons (and active orbitals) are two for core excitation and three for shake-up, respectively, which are the minimum possible RAS schemes. Electron correlation, which can be considerable for conjugated systems, were to some extent taken into account including ten electrons (four of σ type and six of π type) in the active space. The DALTON program was again used in the conventional RASSCF runs.

3. Results and discussion

Prior to discussing environmental effects evaluated by HF-STEX calculations, comparison with RASSCF results should be made to check the influence both from screening and correlation. Table 1 compiles the HF-STEX energies for the planar adenine and guanine molecules, with the results of two RASSCF schemes. These K-edge excitation transitions are confirmed here to be of 1s $\rightarrow$ $\pi*$ type, just as discussed by Kirtley *et al.* (1992). For both molecules, 2×2 RASSCF calculations show lowerings of 0.5-1.5 eV as the screening energy; the N2 and N3 cases obtain large lowerings relative to the remaining cases. Correlation through 12×12 RASSCF calcuations provide additional contributions, however, their directions are not uniform. These RASSCF calculations certainly provide energetic differences from the HF-STEX results but keep the trends that energies of guanine are somewhat blue-shifted relative to adenine and that the N2 or N3 atoms provide the lowest energies. The HF-STEX approximation is thus acceptable for the present purpose to investigate a variety of environmental effects. Checking by similar RASSCF calculations were also done for 1s $\rightarrow$ $\sigma*$ exicitations of adenine, and the influence was found to be at most 0.2 eV, confirming that the $\sigma*$ orbital is of Rydberg type.

Table 1
Nitrogen 1s $\rightarrow$ $\pi*$ excitation energies (in eV) by HF-STEX and RASSCF calculations for the planar adenine and guanine molecules (refer to texts and Figure 1). Experimental value by Kirtley *et al.* (1992) is included in square bracket.

	N1	N2	N3	N4	N5
Adenine [401.0]					
STEX	404.1	401.4	401.2	401.4	403.5
2×2-RAS	403.3	400.1	400.2	400.5	402.8
12×12-RAS	403.4	400.4	399.7	399.9	402.7
Guanine [401.2]					
STEX	404.3	401.5	402.3	404.3	404.8
2×2-RAS	403.8	400.1	400.8	403.4	403.6
12×12-RAS	404.0	401.2	400.7	403.7	404.3

Table 2
The lowest HF-STEX energies (in eV) for the adenine group. Symbols "**2W**" and "**3W**" indicate the double and triple water-solvations, respectively. Relative shifts are given for derivatives. Notation "ns" means that the shift is less than 0.1 eV. Experimental value by Kirtley *et al.* (1992) is included in square bracket.

	N1	N2	N3	N4	N5
Adenine [401.0]	404.2	401.6	401.3	401.5	403.6
Bromoadenine [+0.1]	+0.4	+0.5	+0.1	ns	+0.1
Adenosine [ns]	+0.1	+0.1	+0.3	+0.2	ns
Bromoadenosine	+0.5	+0.5	+0.3	+0.2	ns
Adenylic acid	+0.1	+0.1	+0.3	+0.2	ns
3W-Adenine	-0.2	+0.2	+0.2	+0.2	-0.8
2W-Bromoadenine	+0.5	+0.6	+0.2	+0.2	-0.8
2W-Adenosine	+0.2	+0.3	+0.4	+0.3	-0.8
2W-Bromoadenosine	+0.6	+0.7	+0.4	+0.3	-0.8

Table 3
The lowest HF-STEX energies (in eV) for the guanine group. Composition of the table are the same as Table 2.

	N1	N2	N3	N4	N5
Guanine [401.2]	404.4	401.6	402.4	404.4	404.9
Bromoguanine [+0.1]	+0.3	+0.5	+0.1	ns	ns
Guanosine [+0.1]	ns	+0.1	ns	+0.1	-0.6
Bromoguanosine	+0.3	+0.5	+0.3	+0.1	-0.4
3W-Guanine	+0.2	+0.1	+0.3	-0.2	-0.7
2W-Bromoguanine	+0.5	+0.5	+0.4	-0.2	-0.7
2W-Guanosine	+0.1	+0.2	+0.3	-0.1	-0.7
2W-Bromoguanosine	+0.4	+0.6	+0.4	-0.1	-0.2

Table 2 summarizes the lowest HF-STEX energies for each nitrogen atom of the adenine base and relative energy shifts of its derivatives; all of these energies are due to 1s $\rightarrow$ $\pi*$ type excitations. The sites of bromine and/or glycosidic (symbolized as "R") substitutions can be found in Figure 1. Double water-solvation involves H-N5/N4 and H-N5/N2 pairs; one additional water molecule for the triple solvation is attached on H-N1/N3. Experimental energies (Kirtley *et al.*, 1992) are included also in Table 2. The present calculations imply that the excitations at the N3 atom could be responsible for the K-edge of spectra throughout the adenine group. Bromine substitution leads to larger shifts for the neighboring N1 and N2 atoms than the distant N3 atom. Such a situation is reverted for glycosidic substitution although its shifts are smaller than for bromine substitution. Water-solvation provides a sizable red-shift for the amino-group N5 atom, as expected by the position of the water molecules. For the other nitrogen (N1-N4) cases, the solvation gives small blue-shifts that are additive to the shifts by bromine and/or glycosidic substitutions. We believe that a similar situation may be the case for the actual double-helix of nucleic acids.

The results of guanine and its derivatives are summarized in Table 3; double water-solvation involves H-N5/H-N4 and O/N2 pairs, and one more water for triple solvation is put on H-N5/N3 pair. It is again found that oxygen substitution provides a blue-shift trend relative to the adenine group (compare with the entries in Table 2). This fact is in accord with experiment (Kirtley *et al.*, 1992). The bromine and glycosidic factors of the environment play a similar role with the adenine group. Water-solvation increases similarly the shifts for the N1, N2, and N3 cases. However, there is a difference that the N4 case shows small red-shifts as does the N5 case by water-solvation. This reflects the fact that the protonic hydrogen atom is bonding to the N4 atom in guanine. For the guanine group, excitations at the N2 atom could correspond to the K-edge (Kirtley *et al.*, 1992).

Kirtley *et al.* (1992) observed the edge peak at 401.0 eV and the second (and less intense) peak at 403 eV (see their original paper) for the adenine molecule. They also recorded a broad but sizable peak around 407-412 eV. They speculated that 1s $\rightarrow$ $\sigma*$ excitations correspond to this band. A Stieljes-imaging (SI) plot (Ågren & Carravetta, 1987) for adenine is given as Figure 2, where the solution sets of the STEX hamiltonian for each nitrogen atom have been merged. The simulated first peak consists of 1s $\rightarrow$ $\pi*$ excitations at the N2, N3, and N4 atoms. The second peak contains the contributions from the N1 and N5 atoms. Unfortunately, the higher band is not reproduced in this SI plot. The same situation has been found for the other molecules, which has motivated a discussion of

the possible contributions of multi-electron, "analgous shake-up" character (Ågren & Carravetta, 1992), in this region. Such contributions have certainly been verified for small species like first row diatomics, but no investigations have to our knowledge been made for larger, conjugated species of the size here studied. In order to find out where such shake-up transitions reside, we have computed higher excited state energies by the RASSCF method for planar adenine. Table 4 summarizes the ionization potentials (IPs) and the lowest shake-up energies. The 3×3 (or minimal) RASSCF results indicate that shake-up excitations can contribute to the higher band (Kirtley *et al.*, 1992). Note that the introduction of correlation (by the 11×12 scheme) provide some lowering in energy. Further calculations including intensities are evidently necessary to make a firmer assignment, however, such calculations are far from trivial.

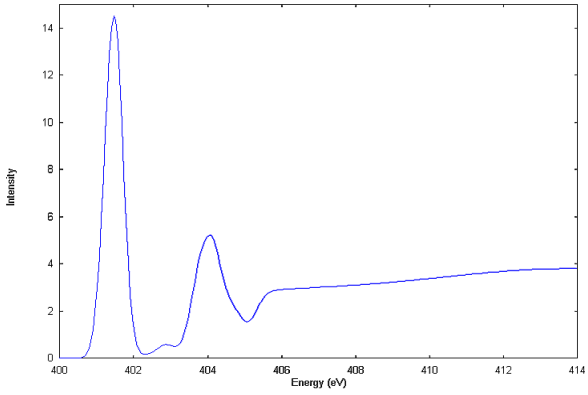

Figure 2
Stieltjes-imaging plot for the adenine molecule. Energy (horizontal axis) is given in eV and intensity (vertical axis) in au.

Table 4
Shake-up energies (eV) by RASSCF calculations for the planar adenine molecule (refer to texts for details). Ionization potentials (IPs) are also included. Wavefunctions of 11×12-scheme for $\sigma \rightarrow \pi*$ were unconverged.

	N1	N2	N3	N4	N5
3×3-RAS					
IP	406.8	404.9	404.4	404.2	405.7
$\pi \rightarrow \pi*$	411.2	407.7	408.0	407.7	409.6
$\pi \rightarrow \sigma*$	413.0	411.8	412.5	412.1	412.2
$\sigma \rightarrow \pi*$	411.2	408.6	408.3	408.5	410.1
11×12-RAS					
IP	406.1	404.2	403.8	403.6	405.9
$\pi \rightarrow \pi*$	410.1	408.0	407.4	407.8	409.5
$\pi \rightarrow \sigma*$	413.0	411.6	412.0	412.2	412.2

4. Summary

The nitrogen K-shell near-edge X-ray absorption spectra of adenine, guanine, and their derivatives have been investigated by extensive HF-STEX calculations. The K-edge band observed by Kirtley *et al.* (1992) was attributed to $1s \rightarrow \pi*$ excitations. The energetic shifts due to environmental differences of each nitrogen atom were evaluated and superposed on an overall feature that was in line with the experimental observation. The calculated energies of shake-up excitations indicated a role of such excitaions for features closely above the edge.

YM would thank Prof. Hans Ågren, Royal Institute of Technology, Sweden for comments. Parallelized HF-STEX calculations with the DALTON program were carried out on IBM SP and NEC SX systems equipped at the computer centers of JAERI and JST (Japan Science and Technology Corporation). Conventional RASSCF calculations and Gaussian94 geometry-opitimization were done on SGI PowerChallenge and Origin systems managed by JST and Nagoya City University. The present investigation was performed in a context of on-going projects of Earth Simulator and Stampi being promoted by Science and Technology Agency of Japanese government.

References

Ågren, H. & Carravetta, V. (1987). *J. Chem. Phys.* **87**, 370–380.

Ågren, H. & Carravetta, V. (1992). *Intern. J. Quant. Chem.* **42**, 685–718.

Ågren, H. & Jensen, H. J. A. (1993). *Chem. Phys.* **172**, 45–57.

Ågren, H., Carravetta, V., Vahtras, O. & Pettersson, L. G. M. (1997). *Theor. Chem. Acc.* **97**, 14–40.

Department of Chemistry, University of Oslo (1997). http:// www. kjemi. uio. no/ software/ dalton/ dalton.html

Gaussian Inc. (1995). http:// www. gaussian. com

Japan Atomic Energy Research Institute (1998). http:// guide. tokai. jaeri. go. jp/ program/ eng/ software/ list/

Kirtley, S. M., Mullins, O. C., Chen, J., van Elp, J., George, S. J., Chen, C. T., O'Halloranm T. & Cramerm S. P. (1992). *Biochim. Biophys. Acta* **1132**, 249–254.

Mochizuki, Y., Ågren, H., Pettersson, L. G. M. & Carravetta, V. (1999). *Chem. Phys. Lett.* **309**, 241–248.

Yang, L., Plachkevytch, O., Ågren, H. & Pettersson, L. G. M. (1997). *J. Phys. IV France* **7**, C2-227–228.

Received 31 July 2000 • Accepted 16 November 2000