

Core-Ionized and Core-Excited States of Macromolecules

PIERRE-FRANÇOIS LOOS, XAVIER ASSFELD

Equipe de Chimie et Biochimie Théoriques, UMR 7565 CNRS-UHP, Institut Jean Barriol (FR CNRS 2843), Faculté des Sciences et Techniques, Nancy-Université, B.P. 239, 54506 Vandœuvre-lès-Nancy Cedex, France

Received 30 January 2007; accepted 2 April 2007

Published online 7 May 2007 in Wiley InterScience (www.interscience.wiley.com).

DOI 10.1002/qua.21410

ABSTRACT: We report the calculation of core-ionized states of small organic compounds and macromolecular systems in the framework of a new method based on the local self-consistent field (LSCF). This new theoretical scheme avoids the variational collapse of the empty core orbital (CO) of the core-excited states and ensures the orthogonality between the ground state and the excited states. Compared to experimental data and other theoretical methods, accurate carbon 1s ionization energies using the Boys-Foster (BF) localization criterion for the determination of the CO and the PBE0/6-311++G**//B3LYP/6-311++G** level of theory was obtained to calculate both the ground and excited states. The macromolecular systems, a sequence of 15 alanine amino acids in both α -helix and β -sheet conformations, are computed using hybrid quantum mechanics/molecular mechanics (QM/MM) method within the LSCF/MM framework. The results show a weak impact of the MM surrounding in the alanine polypeptides cases compared with that of previous studies based on electrically charged residue, such as glutamate in the crambin protein.
© 2007 Wiley Periodicals, Inc. *Int J Quantum Chem* 107: 2243–2252, 2007

Key words: QM/MM; core-excited; core-ionized; core electron binding energy; orthogonality

Introduction

Core-excited or core-ionized states are useful tools to obtain information on the structure and on the nature of molecules [1]. Although most of the studied systems are inorganic, or small organic molecules, it has recently been shown that biomolecules can also be treated [2–5].

Correspondence to: X. Assfeld; e-mail: Xavier.Assfeld@cbt.uhp-nancy.fr

From the point of view of theoretical chemistry, the core-excited state of macromolecules presents several challenges:

- To describe the core-excited state one has to avoid the so-called variational collapse and to be sure that the excited state is orthogonal to the ground state. This area of research is very fecund and many different techniques exist [6–11].
- The modeling of macromolecules with pure quantum mechanics is still out of reach from present day theoretical chemistry. However,

several hybrid methods (QM/MM), mixing quantum mechanics (QM) and molecular mechanics (MM), are available [12–30].

If both problems are correctly handled separately by many methods, they are very rarely considered simultaneously, although some attempt to combine QM/MM methods with the time dependent-density functional theory (TD-DFT) appear recently [31–33]. One of the above QM/MM methods, called LSCF, has been previously applied to the core-ionized state of polyethylene, polytetrafluoroethylene, and of the crambin protein [34]. In the LSCF method, the junction between the QM and the MM parts is materialized by means of a doubly occupied strictly localized bond orbital (SLBO) that remains frozen during the wave function optimization. This enables the correct description of the hybrid system and takes accurately the dangling bond problem into account. The core hole is represented by an empty core orbital (CO) that remains also frozen during the self-consistent field (SCF) optimization to avoid the variational collapse and to enforce the orthogonality between the states.

In this article we present a slight modification of the LSCF method that treats differently the overlap between an occupied SLBO and an empty CO. The method is applied on the pentane molecule as a test and compared to results obtained with the old scheme. The convergence toward the basis set size is investigated on the methane molecule for which the experimental value is unambiguous. The validity of the method is checked with calculations on small molecules containing at least one carbon atom and small molecules that contain the peptide bond. It is further applied on the alanine molecule in different surroundings. This molecule is either isolated or engaged in a tripeptide or pentadecapeptide in both α -helix and β -sheet conformations. The last two calculations are performed within the QM/MM framework.

Theory

The LSCF method has already been detailed in the following references [28–34], and only the most important points will be described here. The method is applicable on the Hartree-Fock and on the Kohn-Sham equations in the restricted or unrestricted formalism. The starting point is the knowledge of some, L , predefined spinorbitals $\{|f_i\rangle\}_{i=1}^L$. They are defined by the user. They are expressed,

thanks to the expansion coefficients $a_{\mu i}$, on the basis functions $\{|\phi_\mu\rangle\}_{\mu=1}^K$.

$$|f_i\rangle = \sum_{\mu=1}^K a_{\mu i} \phi_\mu \quad (1)$$

These functions can be SLBO, CO, localized orbitals, canonical orbitals, or whatever the user wants them to be.

The LSCF algorithm will optimize the total wave function of the system under investigation knowing that the predefined orbitals are frozen, i.e., they must remain unchanged! Note that the frozen orbitals can be occupied or empty.

Hence, we have to consider two sets of molecular orbitals, the frozen ones (FOs) and the self-consistent ones that we call variational orbitals (VOs). We ask for the VOs to be orthogonal to the FOs to simplify the calculation. The very simple idea behind the LSCF method is that the VOs should be expanded over a set of functions already orthogonal to the FOs. To build this set of functions, we define a three-step algorithm:

1. To ease the computation, the set of FOs is orthonormalized.
2. Each initial basis function ϕ_μ is projected out of the subspace span by the orthonormal FOs. The resulting functions, $\tilde{\phi}_\mu$, are then orthogonal to any FO.

$$|\tilde{\phi}_\mu\rangle = \left[1 - \sum_i^L R_{\mu i}^2 \right]^{-1/2} \left[|\phi_\mu\rangle - \sum_i^L |f_i\rangle \langle f_i | \phi_\mu \rangle \right] \quad (2)$$

where $R_{\mu i}$ stands for the overlap integral between the FO f_i and the atomic orbital ϕ_μ .

3. Since L FOs are predefined, there are L linear dependencies in the set of functions obtained at Step 2. These linear dependencies are removed, thanks to the canonical orthogonalization procedure [35]. We then get a set of $(K-L)$ functions, orthogonal to the FOs and mutually orthogonal.

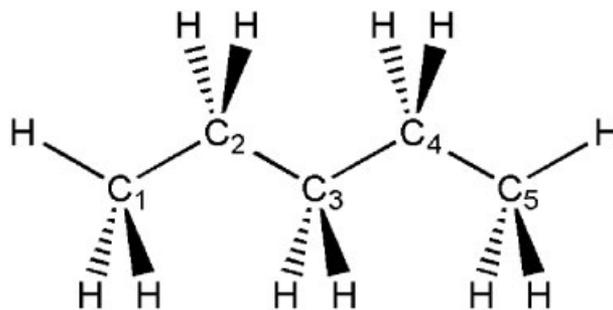
Steps 2 and 3 are combined together in a unique matrix transformation named B . This matrix plays the same role in the LSCF method as the Löwdin matrix ($S^{-1/2}$) plays in the standard SCF method.

To study core-excited or core-ionized state, one has to run two separate calculations, irrespective of the choice of the CO. The first one is performed on the ground state and the second one with the CO emptied and frozen, to avoid the variational collapse. The excitation or ionization energy is computed as the energy difference between the two calculations. There are two ways to define the CO that we call internal and external. The internal way consists of localizing the CO by using any kind of a posteriori localization techniques [36, 37] applied to the optimized wave-function of the ground state. The external way is when the CO is provided by the user. It can come from a prior atomic calculation, a priori localization scheme [38], or a user guess. In that case, the ground state calculation must be carried out with the LSCF method where the occupied CO must be frozen. This is the only way to enforce the orthogonality between the ground and the excited state [34].

When the FOs are all occupied, the orthogonalization method of Step 1 does not matter since the difference, between two orthogonalization procedures, corresponds to a rotation in the occupied space. Hence, generally, the Löwdin symmetric method [39, 40] is employed. However, for core-excited or core-ionized state, one has to pay attention. The Löwdin transformation will mix the occupied FOs with the empty FOs. The resulting functions will no longer represent the empty core function that was wanted. This situation can occur if the CO is too close from a SLBO for example. One way to avoid this slight error is to perform first a Löwdin transformation of the empty FOs alone, followed by a Gram-Schmidt [41] orthogonalization of the occupied FOs with respect to the empty ones. The results presented in this study are obtained with this new orthogonalization scheme. They are compared with those obtained with the old scheme [34] (Löwdin orthogonalization of all FOs simultaneously) in the next section.

Test Calculations

To check the magnitude of the error induced by the global Löwdin orthonormalization that was applied in a previous study [34], we carried out calculations on the pentane molecule with the 6-311++G** basis set. All calculations are run with the Gaussian 03 [42] package modified to perform LSCF calculations. The geometries have been optimized at the B3LYP/6-311++G** level of theory



SCHEME 1. Pentane molecule and carbon atom numbering.

[43–45]. Two types of internal CO (Carbon 1s) are obtained with the Pipek-Mezey (PM) or with the Boys-Foster (BF) localization criterion. In addition, the exchange-correlation functional effect is investigated by using the B3LYP and the PBE0 functional [46]. All five carbon 1s core orbitals are studied. The ionization energies are obtained without and with a SLBO between carbon atoms 1 and 2 (see Scheme 1). They are presented Table I.

Different features can be seen from these results. First of all, one can note that the differences between the old and the new orthogonalization scheme are not large (less than 0.07 eV). This is due to the fact that the overlap (last column of Table I) between the SLBO and the CO is small since the SLBO is mainly developed on the valence AOs. Hence, our previous reported results [34] are still valid. The other important point is that the presence of the SLBO disturbs the total wave function quite locally around it, i.e., the errors induced on the ionization energy of the 1s electron of carbon atom no. 3 is less than one tenth of an eV. On the other hand, computing the ionization energy of the carbon atoms bearing the SLBO gives error as much 1.5 eV, that are not acceptable. One has to keep that point in mind when treating macromolecules with QM/MM method. These conclusions are valid whatever the functional is (B3LYP or PBE0) and whatever the localization criterion is (PM or BF).

One has to note however that the ionization energies are very sensitive to the choice of functional and of the localization criterion. The only way to decide which one is best is to perform a comparison of the results obtained with our method against experimental data or against other theoretical calculations. This is done in the section in which we compare the carbon 1s ionization energy of a series of small molecules containing at least one carbon atom.

TABLE I**Carbon 1s ionization energies (in eV) computed at the DFT/6-311++G** level of theory for the pentane molecule.**

SLBO Loc. DFT	No			New scheme			Old scheme			Overlap
	PM B3LYP	BF B3LYP	BF PBE0	PM B3LYP	BF B3LYP	BF PBE0	PM B3LYP	BF B3LYP	BF PBE0	
C ₁	289.54	291.02	290.41	290.85	292.41	291.94	290.90	292.48	291.99	0.0055
C ₂	290.26	290.98	290.41	291.57	292.26	291.81	291.60	292.31	291.83	0.0024
C ₃	290.32	290.86	290.29	290.38	290.92	290.36	290.38	290.92	290.37	-0.0064
C ₄	290.26	290.98	290.41	290.29	291.00	290.43	290.30	291.01	290.44	-0.0048
C ₅	289.54	291.02	290.41	289.56	291.04	290.42	289.56	291.04	290.42	-0.0012

The calculations are performed without or with a frozen SLBO between C₁ and C₂, and with the old or the new orthogonalization scheme. See Scheme 1 for atom numbering. Two different localization criteria are used, Pipek-Mezey (PM) and Boys-Foster (BF).

Basis Set Influence

To understand the important features a basis set must have for the computation of core holes with the LSCF method, we decide to study the convergence of the vertical carbon 1s ionization energy of the methane molecule. We choose this molecule since it contains only one carbon atom and thus the experimental value is unambiguous. All calculations have been carried out with the PBE0 functional. For each basis set, the geometry is completely optimized. The BF localization criterion has been chosen to locate the CO. Several characteristics can be extracted from the data gathered in Table II. One can see that a nice convergence is obtained for each basis set family. One can also note that augmenting the basis set with diffuse functions do not influence the results, as one could have been guessed since these functions expand quite far away from the nucleus. As long as the basis set is flexible enough, the polarization of the core orbitals is irrelevant. Pople's triple- ζ basis sets give quite constant values showing that for isolated species polarization and diffuse functions are not compulsory. The ionization energies obtained with the 6-311G series are very close to those obtained with the correlation consistent triple- ζ quality basis sets. From a practical point of view, the values obtained with the 6-311G type basis sets are closer to the experimental data (290.84 eV), certainly due to an adequate cancellation of error (basis set limitation versus relativistic corrections). As the polarization and diffuse functions could be important in some molecular situations, we decide to stick with the 6-311++G(d,p) basis set for the remaining calculations of the present study.

Small Molecules

The carbon 1s ionization energies of a series of small molecules are collected in Table III. The molecules are considered in two sets. The first one contains only hydrocarbon molecules while the second considers molecules possessing heteroatoms (oxygen or nitrogen). All ground state

TABLE II**Carbon 1s ionization energies (in eV) of the methane molecule at the PBE0 level of theory, using the Boys-Foster core orbital, with respect to various basis set.**

Basis set	IE(1s _C)
cc-pVDZ	292.23
cc-pVTZ	290.75
cc-pVQZ	290.56
cc-pV5Z	290.53
Aug-cc-pVDZ	292.05
Aug-cc-pVTZ	290.73
Aug-cc-pVQZ	290.55
Aug-cc-pV5Z	290.52
cc-pCVDZ	291.71
cc-pCVTZ	290.61
cc-pCVQZ	290.55
cc-pCV5Z	290.52
6-311G	290.78
6-311G(d)	290.78
6-311G(d,p)	290.77
6-311G(2d,p)	290.75
6-311G(2d,2p)	290.69
6-311+G(d,p)	290.78
6-311++G(d,p)	290.78

TABLE III

Carbon 1s ionization energy (in eV) of small molecules computed at various levels of theory and with various localization schemes.

		Ionization energy				Deviation from experiment		
		B3LYP, PM	B3LYP, BF	PBE0, BF	Exp	B3LYP, PM	B3LYP, BF	PBE0, BF
Methane	C	290.91	291.45	290.78	290.84	0.07	0.61	-0.07
Ethane	C	290.65	291.29	290.67	290.71	-0.06	0.58	-0.04
Ethylene	C	291.16	291.33	290.70	290.82	0.34	0.51	-0.12
Ethyne	C	290.67	291.82	291.19	291.25	-0.58	0.57	-0.06
Propane	CH ₃	290.23	291.11	290.49				
	CH ₂	290.73	291.20	290.63				
	Average	290.48	291.15	290.56	290.57	-0.09	0.58	-0.01
Propene	CH ₂	290.26	290.73	290.11				
	CH	291.28	291.21	290.64				
	CH ₃	290.51	291.44	290.82				
	Average	290.68	291.13	290.52	290.68	0.00	0.45	-0.16
Propyne	HC	290.48	290.82	290.21	290.37	0.11	0.45	-0.16
	C	290.94	291.42	290.84	290.93	0.01	0.49	-0.09
	CH ₃	289.98	292.32	291.70	291.76	-1.78	0.56	-0.06
Butane	CH ₃	289.74	291.05	290.44				
	CH ₂	290.52	291.03	290.46				
	Average	290.13	291.03	290.45	290.48	-0.35	0.56	-0.03
Pentane	C ₁	289.54	291.02	290.41				
	C ₂	290.26	290.98	290.41				
	C ₃	290.32	290.86	290.29				
	C ₄	290.26	290.98	290.41				
	C ₅	289.54	291.02	290.41				
	Average	289.99	290.97	290.39	290.42	-0.43	0.55	-0.03
RMSD 1						0.60	0.54	0.09
Formaldehyde	C	295.08	295.17	294.46	294.47	0.61	0.70	-0.01
Acetone	CO	294.34	294.28	293.70	293.71	0.63	0.57	-0.01
					293.88	0.46	0.40	-0.18
					291.23	-0.39	0.51	-0.12
Methanol	C	292.62	293.11	292.44	292.3	0.32	0.81	0.14
					292.8	-0.18	0.31	-0.36
					292.5	-0.22	0.41	-0.20
Ethanol	CH ₂	292.28	292.91	292.30	292.50	-0.22	0.41	-0.20
	CH ₃	290.56	291.32	290.69	291.10	-0.54	0.22	-0.41
Ethanoic acid	CH ₃	291.28	292.19	291.55	291.55	-0.27	0.64	0.00
					291.6	-0.32	0.59	-0.05
					295.38	0.54	0.58	-0.03
Dimethylether	CH ₃	292.25	292.83	292.17	292.17	0.08	0.66	0.00
					292.55	-0.30	0.28	-0.38
					295.90	0.46	0.52	-0.15
Formic Acid	C	296.36	296.42	295.75	292.80	0.56	0.62	-0.05
					295.90	0.46	0.52	-0.15
Methylamine	CH ₃	291.54	292.19	291.55	291.60	-0.06	0.59	-0.05
RMSD 2						0.40	0.54	0.19
RMSD Total						0.49	0.54	0.16

The column heads indicate the levels and the corresponding localizations. The first set contains only hydrocarbon while the second contains molecules involving heteroatoms (O or N). The experimental data are taken from references [47] and [48]. When several experimental data are available, the two extreme values are given.

TABLE IV

Carbon 1s ionization energy (in eV) of small biomimetic molecules (see Scheme 2 for carbon atom type and molecule conformation).

Carbon atom Molecules	C_{R1}		C_{carb}		C_{R2}		C_{R3}	
	Ref.		Ref.		Ref.		Ref.	
Formamide			294.16 (294.45)	294.41				
<i>Trans-N</i> -methylformamide			293.64	293.92			292.50	292.49
<i>Cis-N</i> -methylformamide			293.58	293.85	292.18	292.16		
<i>N,N</i> -dimethylformamide			293.25 (293.45)	293.52	291.92 (292.03)	291.86	292.27 (292.03)	292.24
Acetamide	291.20	291.19	293.87	294.17				
<i>Trans N</i> -methylacetamide	291.12	291.02	293.37	293.68	292.03	291.90		
<i>Cis-N</i> -methylacetamide		291.03		293.73				292.29
<i>N,N</i> -dimethylacetamide	290.90	290.89	292.99	293.34	291.69	291.61	292.11	292.03

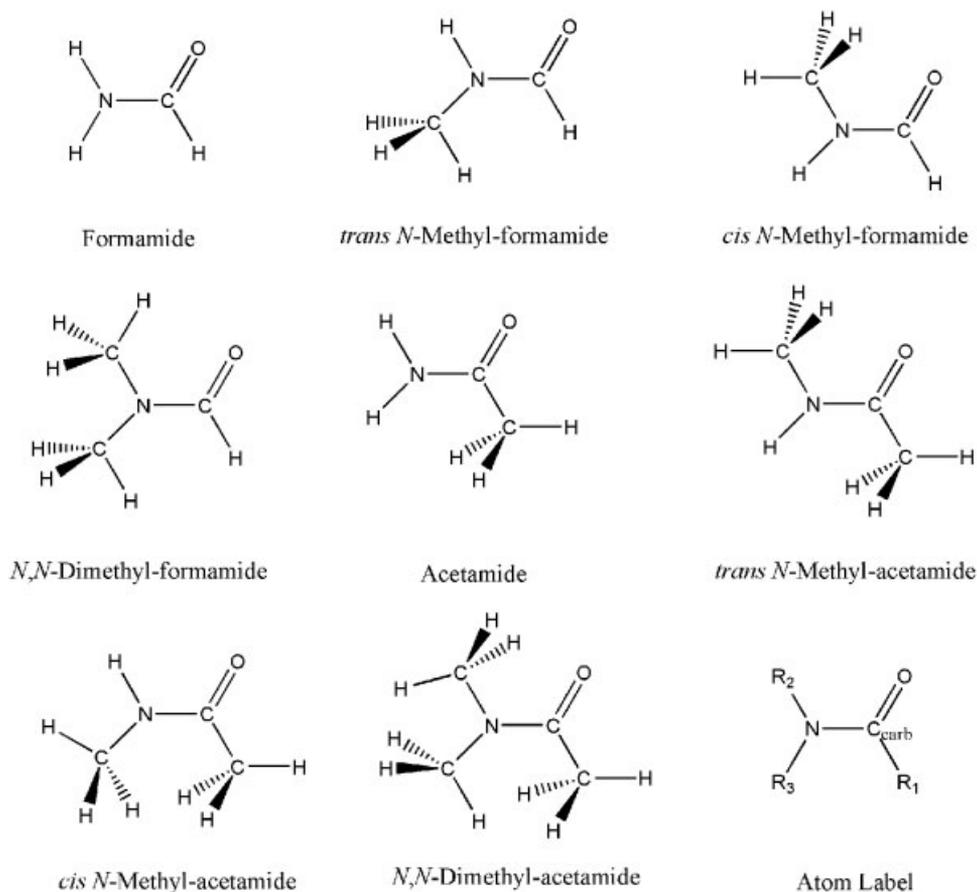
The reference data (ref. column) are extracted from the article of Chong et al. [49]. Experimental values are given in parenthesis [47].

geometries have been optimized at the B3LYP/6-311++G** level of theory. Only the vertical ionization energy is considered, i.e. the geometry of the cation is the same as that the geometry of the neutral form. This approximation is valid since the ionization process occurs very rapidly and the geometry has not the time to relax. The ionization energy is computed with various levels of theory (B3LYP/6-311++G** and PBE0/6-311++G**) and with various localization criteria (PM or BF at the B3LYP/6-311++G** level). The accuracy of the calculations is checked against experimental data [47, 48]. Readily, one can see that the PBE0 functional performs a much better job than the B3LYP exchange-correlation functional. The maximum deviation from experiment is less than 0.2 eV for PBE0 and 0.6 eV for B3LYP on the subset of molecules containing only carbon and hydrogen atoms. The deviation increases to 0.4 eV for PBE0 and 0.8 eV for B3LYP for molecules containing heteroatoms. It is also noteworthy that the BF core orbital provides results as accurate as those obtained with the PM CO on average. If one looks case by case, one can see that the PM core orbital sometimes gives more precise results than the BF one (for example, the deviation for the methane molecule is 0.07 eV with the PM CO and 0.61 with the BF CO). Surprisingly, the PM localization scheme predicts a very poor ionization energy for the 1s CO of the methyl carbon atom of propyne. The deviation from experimental data of the values obtained with the PM CO varies between -1.78 and 0.63 eV. The BF criterion shows a much more constant behavior for all carbon atom types; all results differing from experimental values are in the range 0.22 – 0.81 eV with the

B3LYP functional and in the interval -0.41 to 0.14 eV with PBE0. This behavior seems to be systematic. Hence, in the rest of this study, the BF criterion is chosen in conjunction with the PBE0 functional.

Core-Ionized States of the Peptide Bond

Before studying large macromolecular systems, it is better to know the accuracy of our method on simple biomimetic molecules. We compute the carbon 1s ionization energy of various molecules containing the peptide bond. The methodology defined earlier is used, and the results presented in Table IV are compared with experimental data, when available [47], or with theoretical calculations [49]. We have chosen the same molecules studied by Chong et al. [49] with the *cis* isomer of *N*-methyl-acetamide added to the list. The molecules and the carbon atom type (C_{carb} , C_{R1} , C_{R2} and C_{R3}) are sketched on Scheme 2. From these computations one can see that our method gives results as accurate as the ΔE_{KS} method of Chong. The main discrepancy between the two series of theoretical values is 0.3 eV that is in an acceptable error range. Compared to the few experimental data, the largest deviation is 0.2 eV with our method and 0.3 eV with the ΔE_{KS} method. Again, both methods agree very well with the experimental values and are within a reasonable range of error since our calculations do not take the relativistic correction into account.



SCHEME 2. Biomimetic molecules containing the peptide bond.

The Alanine Molecule

Previously [34], we have reported hybrid QM/MM calculations on the carbon 1s ionization energy of the glutamate residue in crambin protein. Since the functional group is electrically charged, the influence of the surroundings point charges is of primary importance. We computed a chemical shift of 4 eV going from the isolated glutamate molecule to the protein. Here we would like to investigate the influence of a model protein surrounding on a neutral residue. We choose the alanine amino acid since experimental data [50] are available for this molecule in the gas phase. Hence, theoretical calculations are perfectly adapted. In a first step we compute the carbon 1s ionization energy of the isolated alanine molecule in the two most stable conformations (see Figure 1). In a second step, the direct surroundings influence is investigated by means of a tripeptide composed of three

alanine molecules in two conformations close to the α -helix and the β -sheet idealized geometries. Finally, in the third step, the alanine molecule is engaged in a poly-Ala-pentadecapeptide in the α -helix and β -sheet conformations. These last systems are computed with the hybrid LSCF/MM method (Scheme 3). All geometries have been fully optimized at their respective levels of theory, at the exception of the QM/MM calculations for which the backbone has been kept fixed. The results are given in Table V.

For the gas phase alanine molecule, we have considered two conformations because they are very close in energy (less than 1 kcal/mol apart). One can see that the ionization energies for the three carbon atoms for both conformations are in very good agreement with the experimental data [50]. It is thus not possible from our results to discriminate between the two geometries. However, we are quite confident with the accuracy of our results for this system.

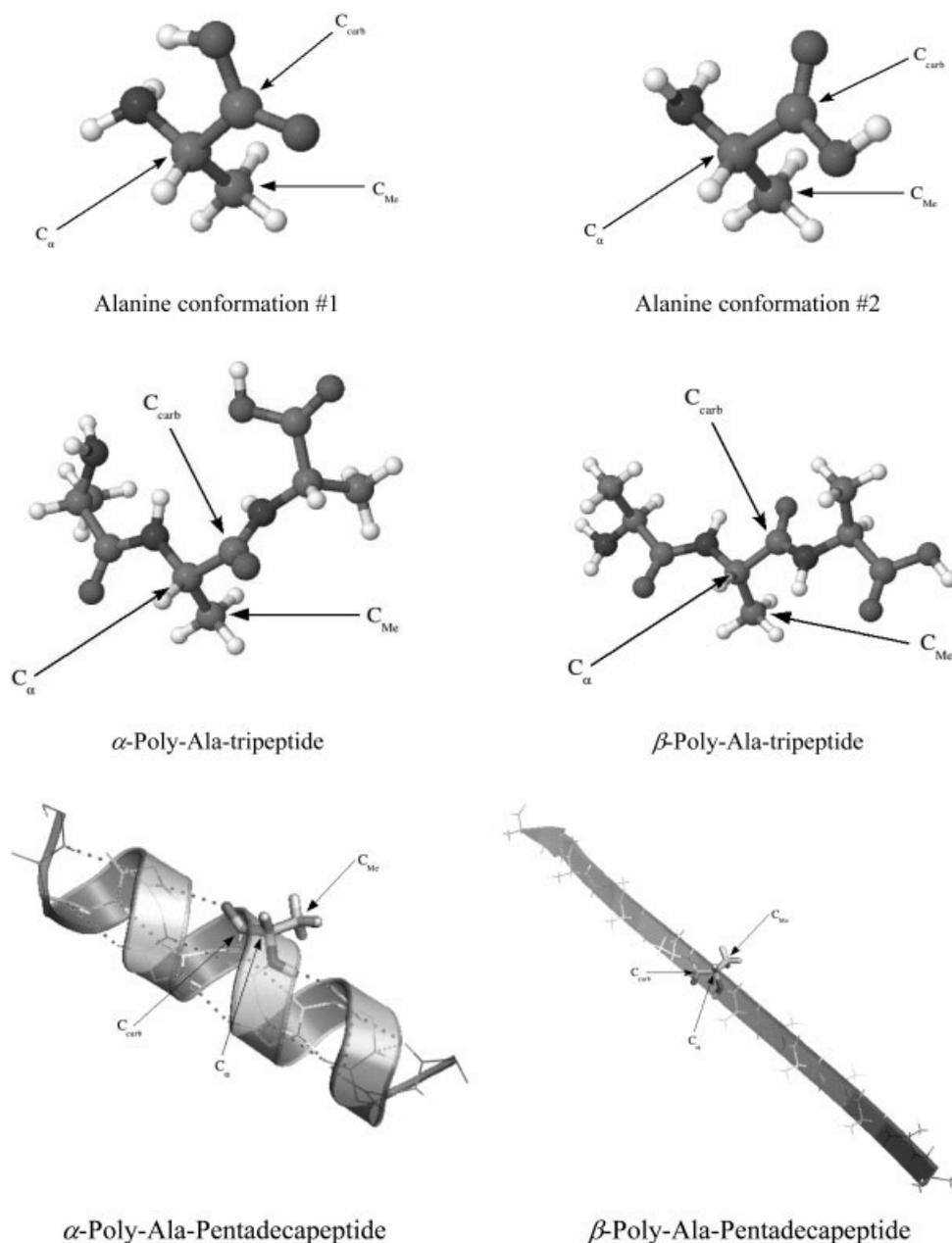
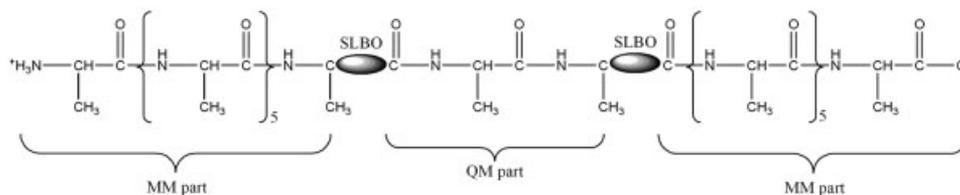


FIGURE 1. Geometries of the alanine molecule (two conformations), of the poly-Ala-tripeptide (α -helix and β -sheet conformations) and of the poly-Ala-pentadecapeptide (α -helix and β -sheet conformations).

For the tripeptide molecule, one can note that the ionization energies are very close for both conformations. The only significant change is for the carbon atom of the carbonyl group for which a shift about 1.5 eV is observed with respect to the gas phase data. This can be easily understood since this carbon atom changes from a carboxylic group to an amidic group. In fact, its ionization energy is very

close to the one computed for the peptide bond in the *N*-methylacetamide molecule (see Table IV).

Looking at the results obtained with the LSCF/MM method, one can conclude that the effects of the surroundings are very weak and not comparable to the one obtained previously [34]. This was expected since the ground state is neutral and thus is only slightly polarized by the MM part.



SCHEME 3. QM/MM partition of the poly-Ala-pentadecapeptide.

One can persuade itself by inspection of the last column in Table V where are gathered the ionization energies obtained with the LSCF/MM method with the classical point charges turn off, i.e. without the electrostatic embedding. Both QM/MM calculations give values very close to one another. The main tendency one can extract is that the electrostatic embedding increases the ionization energy by 0.2–1.0 eV. It is noteworthy that the hydrogen bonds in the α -helix conformation are not polarizing enough to induce a significant shift on the 1s CO of the carbonyl carbon atom. One has to note that the results presented here do not take the polarization of the MM part into account and that this effect can slightly modify the above conclusions.

Conclusion

On the basis of the LSCF method, we have developed a new methodology to calculate accurate core-ionized and core-excited state of small organic compounds and of macromolecules. The new theoretical scheme allows handling correctly the orthogonality constraint between the occupied and virtual frozen orbitals, thanks to a modification of the orthogonalization procedure. The reliability of the method is demonstrated on the pentane mole-

cule by comparison with the old scheme. Concerning the description of the core hole, triple- ζ basis sets give results with an acceptable accuracy. Furthermore, we applied our method on a set of both small carbon containing compounds and small molecules containing a peptidic bond. The results have been compared with experimental and theoretical results. We have shown that the CO BF localization associated with the PBE0/6-311++G**//B3LYP/6-311++G** theoretical scheme gives very accurate ionization energy.

According to these conclusions and a preliminary study of the alanine mono-peptide and tripeptide, we have computed the alanine carbon 1s ionization energy of the pentadecapeptide in both α -helix and β -sheet conformation within the LSCF/MM framework. We conclude that the effects of the MM surroundings on the ionization energies are weak compared with those in a previous study [34]. However, the electrostatic embedding must be taken into account because of an increase of 0.2–1.0 eV of the ionization energies. In future work, the polarization of the MM part should be included.

As outlook, we are currently investigating and testing the reliability of this new theoretical scheme in the evaluation of the inner-shell absorption spectra of the most common proteinogenic α -amino

TABLE V Carbon 1s ionization energy of the alanine molecule, isolated, engaged in a tripeptide or in a pentadecapeptide in the α -helix or β -sheet conformation.

	Alanine			Tripeptide		Pentadecapeptide			
	PBE0		Exp.	PBE0		PBE0/AMBER		PBE0/AMBER*	
	conf 1	conf 2		α -helix	β -sheet	α -helix	β -sheet	α -helix	β -sheet
C_{α}	292.45	292.16	292.30	291.87	292.00	292.13	292.62	291.82	291.92
C_{carb}	294.60	294.92	295.30	293.69	293.87	293.78	294.63	293.06	293.64
C_{Me}	291.11	290.89	291.20	290.50	290.62	291.21	290.94	290.97	290.51

See Figure 1 for the geometries. The AMBER* [51,52] calculations are performed excluding the electrostatic polarization of the wave function by the classical point charges.

acids for which the experimental near edge X-ray absorption (NEXAFS) spectra are available [53, 54].

References

- Schwarz, W. H. E. *Angew Chem Int Ed* 1974, 13, 454.
- Blomberg, E.; Claesson, P. M.; Froberg, J. C. *Biomaterials* 1998, 19, 371.
- Fitzpatrick, H.; Luckman, P. F.; Eriksen, S.; Hammond, K. J. *Colloid Interface Sci* 1992, 149, 1.
- Lebugle, A.; Subirade, M.; Gueguen, J. *Biochim Biophys Acta* 1995, 1248, 107.
- Gordon, M. L.; Tulumello, D.; Cooper, G.; Hitchcock, A. P.; Glatzel, P.; Mullins, O. C.; Cramer, S. P.; Bergmann, U. *J Phys Chem A* 2003, 107, 8512.
- Averbukh, V.; Cederbaum, L. S. *J Chem Phys* 2006, 125, 094107.
- Breidbach, J.; Cederbaum, L. S. *J Chem Phys* 2007, 126, 034101.
- Chong, D. P.; Takahata, Y. *Chem Phys Lett* 2006, 418, 286.
- Takahata, Y.; Okamoto, A. K.; Chong, D. P. *Int J Quant Chem* 2006, 106, 2581.
- Herbert, J. M.; Ortiz, J. V. *J Phys Chem A* 2000, 104, 11786.
- Plashkevych, O.; Privalov, T.; Ågren, H.; Carravetta, V.; Ruud, K. *Chem Phys Lett* 2000, 260, 11.
- Warshel, A.; Levitt, M. *J Mol Biol* 1976, 103, 227.
- Singh, U. C.; Kollman, P. A. *J Comput Chem* 1986, 7, 718.
- Field, M. J.; Bash, P. A.; Karplus, M. J. *Comput Chem* 1990, 11, 700.
- Ranganathan, S.; Gready, J. E. *J Phys Chem B* 1997, 101, 5614.
- Cummins, P. L.; Gready, J. E. *J Phys Chem B* 2000, 104, 4503.
- Das, D.; Eurenus, K. P.; Billings, E. M.; Sherwood, P.; Chatfield, D. C.; Hodosek, M.; Brooks, B. R. *J Chem Phys* 2002, 117, 10534.
- Maseras, F.; Morokuma, K. *J Comput Chem* 1995, 16, 1170.
- Zhang, Y.; Lee, T.-S.; Yang, W. *J Chem Phys* 1999, 110, 46.
- Antes, I.; Thiel, W. *J Phys Chem A* 1999, 103, 9290.
- DiLabio, G. A.; Hurley, M.; Christiansen, P. A. *J Chem Phys* 2002, 116, 9578.
- Gao, J.; Amara, P.; Alhambra, C.; Field, M. J. *J Phys Chem A* 1998, 102, 4714.
- Pu, J.; Gao, J.; Truhlar, D. G. *J Phys Chem A* 2004, 108, 632.
- Lin, H.; Truhlar, D. G. *J Phys Chem A* 2005, 109, 3991.
- Kairys, V.; Jensen, J. H. *J Phys Chem A* 2000, 104, 6656.
- Murphy, R. B.; Philipp, D. M.; Friesner, R. A. *J Comput Chem* 2000, 21, 1442.
- Lin, H.; Truhlar, D. G. *Theor Chem Acc* 2007, 117, 185.
- Assfeld, X.; Rivail, J.-L. *Chem Phys Lett* 1996, 263, 100.
- Ferré, N.; Assfeld, X.; Rivail, J.-L. *J Comput Chem* 2002, 23, 610.
- Moreau, Y.; Loos, P.-F.; Assfeld, X. *Theor Chem Acc* 2004, 112, 228.
- Wanko, M.; Hoffmann, M.; Strodel, P.; Koslowski, A.; Thiel, W.; Neese, F.; Frauenheim, T.; Elstner, M. *J Phys Chem B* 2005, 109, 3606.
- Tavernelli, I.; Roehrig, U. F.; Rothlisberger, U. *Mol Phys* 2005, 103, 963.
- Roehrig, U. F.; Frank, I.; Hutter, J.; Laio, A.; VandeVondele, J.; Rothlisberger, U. *ChemPhysChem* 2003, 4, 1177.
- Ferré, N.; Assfeld, X. *J Chem Phys* 2002, 117, 4119.
- Szabo, A.; Ostlund, N. S. *Modern Quantum Chemistry: Introduction to Advanced Structure Theory*; Dover: Mineola, NY, 1989.
- Foster, J. M.; Boys, S. F. *Rev Mod Phys* 1960, 32, 300.
- Pipek, J.; Mezey, P. *J Chem Phys* 1989, 90, 4916.
- Fornili, A.; Loos, P.-F.; Sironi, M.; Assfeld, X. *Chem Phys Lett* 2006, 427, 236.
- Lowdin, P. O. *J Chem Phys* 1950, 18, 365.
- Landshoff, R. *Z Phys* 1936, 102, 201.
- Gantmacher, F. C. *The Theory of Matrices, Vol. 1*; New York: Chelsea, 1960.
- Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03, Revision C. 03*; Gaussian, Inc.: Wallingford CT, 2004.
- Becke, A. D. *J Chem Phys* 1993, 98, 5648.
- Perdew, J. P. *Phys Rev B* 1986, 33, 8822.
- Perdew, J. P.; Burke, K.; Wang, Y. *Phys Rev B* 1996, 54, 16533.
- Adamo, C.; Barone, V. *J Chem Phys* 1999, 110, 6158.
- Jolly, W. L.; Bomben, K. D.; Eyermann, C. J. *Atom Data Nucl Data Tables* 1984, 31, 433.
- Saethre, L. J.; Berrah, N.; Bazek, J. D.; Borve, K. J.; Carroll, T. X.; Kukk, E.; Gard, G. L.; Winter, R.; Thomas, T. D. *J Am Chem Soc* 2001, 123, 10729.
- Chong, D. P.; Aplincourt, P.; Bureau, C. *J Phys Chem A* 2002, 106, 356.
- Powis, I.; Rennie, E. E.; Hergenbahn, U.; Kugeler, O.; Bussy-Socrate, R. *J Phys Chem A* 2003, 107, 25.
- Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J Am Chem Soc* 1995, 117, 5179.
- Wang, J.; Cieplak, P.; Kollman, P. A. *J Comput Chem* 2000, 21, 1049.
- Kaznatcheyev, K.; Osanna, A.; Jacobsen, C.; Plashkevych, O.; Vahtras, O.; Ågren, H.; Carravetta, V.; Hitchcock, A. P. *J Phys Chem A* 2002, 106, 3153.
- Zubavichus, Y.; Shaporenko, A.; Grunze, M.; Zharnikov, M. *J Phys Chem A* 2005, 109, 6998.