Multi-level Quantum Monte Carlo Wave Functions for Complex Reactions: The Decomposition of α-Hydroxy-Dimethylnitrosamine

Francesco Fracchia,[a]* Claudia Filippi,[b] and Claudio Amovilli[c]

We present here several novel features of our recently proposed Jastrow linear generalized valence bond (J-LGVB) wave functions, which allow a consistently accurate description of complex potential energy surfaces (PES) of medium-large systems within quantum Monte Carlo (QMC). In particular, we develop a multilevel scheme to treat different regions of the molecule at different levels of the theory. As prototypical study case, we investigate the decomposition of α-hydroxy-dimethylnitrosamine, a carcinogenic metabolite of dimethylnitrosamine (NDMA), through a two-step mechanism of isomerization followed by a retro-ene reaction. We compute a reliable reaction path with the quadratic configuration interaction method and employ QMC for the calculation of the electronic energies. We show that the use of multideterminantal wave functions is very important to correctly describe the critical points of this PES within QMC, and that our multilevel J-LGVB approach is an effective tool to significantly reduce the cost of QMC calculations without loss of accuracy. As regards the complex PES of α-hydroxy-dimethylnitrosamine, the accurate energies computed with our approach allows us to confirm the validity of the two-step reaction mechanism of decomposition originally proposed within density functional theory, but with some important differences in the barrier heights of the individual steps. © 2013 Wiley Periodicals, Inc.

DOI: 10.1002/jcc.23461

Introduction

The ability to compute accurate potential energy surfaces (PES) is one of the fundamental goals of quantum chemistry. The knowledge of the PES of a given chemical system allows one to perform dynamical simulations and acquire crucial knowledge on the chemical reactivity of the system. Particular attention must be paid to the differences in energy between the critical points of the PES, as these differences strongly impact the thermodynamics and kinetics of the system. To obtain these quantities, quantum Monte Carlo (QMC) represents a viable alternative to standard highly correlated quantum chemistry methods, which has recently been used in a variety of biochemical applications (see, e.g., Refs. [1–7]).

The computational cost of QMC is closely related to the size of the employed trial wave function and it is therefore important to work with wave functions characterized by a small expansion in Slater determinants. Recently, we have developed a new class of compact QMC wave functions [8,9] which are based on the use of localized orbitals of bonding and antibonding type, and inspired to the generalized valence bond (GVB) approach. The determinantal expansion in our wave functions has a size-extensive modular form and can include progressively different classes of electron excitations. We named these wave functions of the Jastrow linear GVB (J-LGVB) form.

Here, we illustrate a new potential capability of the J-LGVB wave functions which allows the QMC study of a complex PES with consistent accuracy at a reduced cost. The modular nature of the J-LGVB wave functions offers the possibility to treat the different parts of the molecule at a different level of the theory. In particular, we can treat at a lower level of theory the regions of the system that remain unchanged during the process under study, and concentrate the computational effort on the regions where the chemical changes occur. We believe that this study opens new perspectives for the future use of QMC in calculations on large molecular systems.

As prototypical study case, we investigate the mechanism of formation of a potent carcinogen, the methyldiazohydroxide (CH_3N_2OH) molecule, originated from dimethylnitrosamine (NDMA). We focus on a low-energy mechanism recently proposed in a theoretical study based on density functional theory (DFT) methods,[10] and evaluate the relevant barrier heights and reaction energies with QMC on a reliable reaction path we compute with the quadratic configuration interaction (QCISD) method. The application of the QMC method in

[a] *F. Fracchia*
Dipartimento di Scienze Chimiche e Farmaceutiche, Università di Ferrara, Via L. Borsari 46, 44121, Ferrara, Italy
Email: francesco.fracchia@unife.it

[b] *C. Filippi*
MESA+ Institute for Nanotechnology, University of Twente, P.O. Box 217, Enschede, 7500 AE, The Netherlands

[c] *C. Amovilli*
Dipartimento di Chimica e Chimica Industriale, Università di Pisa, Via Risorgimento 35, Pisa, 56126, Italy

Contract grant sponsor: European Commission Capacities Area - Research Infrastructures Initiative [HPC-EUROPA2 project (project number: 228398)]; Contract grant sponsor: Italian Ministry of Research and University (PRIN 2009) (national research project)

combination with our multilevel J-LGVB wave functions will allows us to obtain accurate thermodynamic and kinetic data of the process of formation of the methyldiazohydroxide carcinogen.

The outline of the article is as follows. In section General Theory, we illustrate the possible mechanisms that could be involved in the transformation of NDMA into the carcinogenic methyldiazohydroxide compound. In section Theoretical Method, we describe the J-LGVB approach, focusing on the important novel features of the method proposed here. In section Computational Details we describe the procedures used to perform the calculations. In section Results, we present all calculations with a detailed analysis of the results. Finally, in the last section, we draw the conclusions.

General Theory

Most of the known carcinogens are electrophilic species[11,12] and many of them are direct or indirect alkylating agents. The indirect alkylating agents become active as a result of metabolic transformations, whereas the direct alkylating agents react with the DNA-bases transferring an alkyl group. Organic nitrosamines are a particularly insidious class of indirect alkylating agents, as these molecules occur in foods as a consequence of the reaction between nitrites and amino groups.[13] Nitrosamines are also present as contaminants in water[14,15] and in tobacco smoke.[16] The NDMA is the most common nitrosamine in the environment and represents a serious danger for human health, particularly for liver and lung.[17] The enzymes involved for carcinogenic activation of NDMA are the P450 cytochromes. When P450 catalyzes an oxidation reaction, the active site of the enzyme can be formally represented by a $[FeO]^{3+}$ iron-oxo porphyrin complex. The iron atom is bound to the P450 protein via a thiolate ligand derived from a cysteine residue. If the substrate of the enzyme is the NDMA molecule, the reaction can proceed by extraction of a hydrogen from a methyl group. The result of this process is α-hydroxy-dimethylnitrosamine. The mechanism is shown in Figure 1 (top). The oxidation of NDMA via P450 can also proceed through an alternative path of denitrosation as shown in Figure 1 (bottom). The denitrosation process results in nitrogen monoxide and *N*-methyl-methanimine (CH_3NCH_2), which hydrolyzes to formaldehyde and methylamine. A recent computational study[18] has found that the two mechanisms have a common α-nitrosamino radical as key intermediate as also shown in Figure 1.

Experimental studies indicate that only 14–20% of the NDMA undergoes denitrosation.[19,20] The cytotoxicity of the metabolites of the two processes has been investigated in several studies[19,20] and it was found that only the first metabolic pathway is responsible for mutagenic effects. The α-hydroxy-dimethylnitrosamine decomposes, through a nonenzymatic mechanism, to formaldehyde and methyldiazohydroxide:

$$(HOCH_2)(CH_3)NNO \rightarrow CH_2O + CH_3N_2OH\,. \qquad (1)$$

The reactant and products of this decompositions reaction are also shown in Figure 2. The methyldiazohydroxide is a precursor of the strong electrophile methyldiazonium cation $(CH_3N_2^+)$, which can methylate the DNA-bases and is ultimately responsible for the carcinogenicity of NDMA.

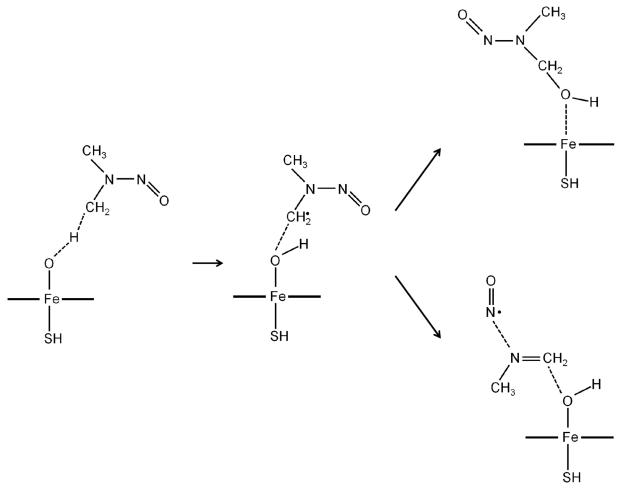

Figure 1. H-abstraction in the oxidative dealkylation (upper) and in the denitrosation (lower) of NDMA on the active site of P450 enzyme.

In this work, we will focus on the decomposition of α-hydroxy-dimethylnitrosamine in formaldehyde and methyldiazohydroxide. This process has already been the subject of previous theoretical work.[10,21] At the B3LYP/6-31G** level,[21] a reaction mechanism was identified passing through the intermediate monomethylnitrosamine, which isomerizes from the Z to the E form and finally tautomerizes to methyldiazohydroxide. In Figure 3, we show the sequence of reactions for the proposed mechanism. The first step of this mechanism involves a transition state with a four-atom ring and is characterized by a barrier height of 61.4 kcal/mol. The barrier heights for the following steps are lower, namely, 26.3 kcal/mol for the E-Z isomerization and 32.0 kcal/mol for the tautomerization.

The first barrier looks strikingly high for a process that takes place without enzyme. A more recent study[10] at the B3LYP/6-311+G(d,p) level has in fact identified a pathway with a lower activation barrier, which occurs only in two steps (see Fig. 4), namely, a E-Z isomerization with rotation around the N—N bond and a retro-ene reaction passing through a transition state with a six-atoms ring. According to these calculations, the rate-determining step is the E-Z isomerization with a barrier height of 24.11 kcal/mol.

Some caution should, however, be used in considering the proposed mechanisms of both B3LYP studies as the B3LYP

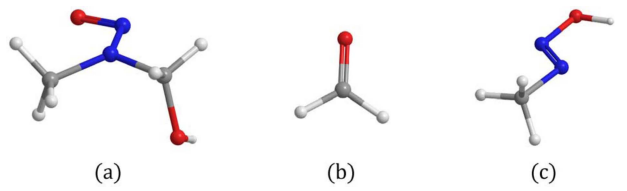

Figure 2. (a) The α-hydroxy-dimethylnitrosamine, (b) formaldehyde, and (c) methyldiazohydroxide molecules. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

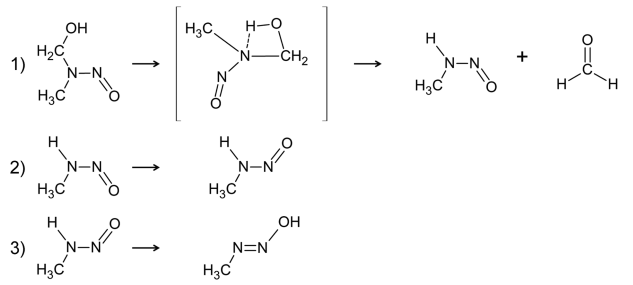

Figure 3. The decomposition of α-hydroxy-dimethylnitrosamine through the three steps identified in the theoretical work of Ref. [21] at the B3LYP/6-31G** level.

functional does generally not perform well in the prediction of barrier heights as documented in various systematic studies.[22,23] Here, we will investigate the α-hydroxy-dimethylnitrosamine decomposition at a higher level of theory both for the optimization of the geometries and for the single-point energy calculations, and focus on the two-step mechanism proposed in Ref. [10].

Theoretical Method

In the diffusion Monte Carlo (DMC) calculations, we use as trial functions the multideterminantal Jastrow–Slater J-LGVBn wave functions introduced in our previous work.[8] These wave functions exploit the chemical idea of electron pair and are constructed with localized orbitals (bonding, antibonding, lone pair, and diffuse lone pair with nodes). The J-LGVBn wave functions represent a class of wave functions of increasing complexity, which account for electron correlation through the progressive inclusion of configuration state functions (CSFs). The first-order wave functions, J-LGVB1, consider the correlation within electron pairs through double excitations from the bonding orbitals to the respective antibonding orbitals (E1). The second-order wave functions, J-LGVB2, include also the dispersion interaction between the adjacent electron pairs through double excitations given by the coupling of single bonding-antibonding excitations (E2). In our tests on the performance of J-LGVBn wave functions to describe bond dissociation energies and barrier heights of prototypical reactions, we found that there are generally no significant improvements beyond second order.

The J-LGVBn wave function depends on the parameters that define the orbitals, the coefficients of the CSFs, and the parameters of the Jastrow factor. All parameters are optimized by minimizing the variational Monte Carlo (VMC) energy. In our previous work,[8] we used as initial guess the localized orbitals from an multi-configurational self-consistent field (MCSCF) optimization of the LGVB1 expansion. These orbitals are the natural choice because they are the result of a variational optimization of a wave function very similar to the J-LGVBn ones. In this work, we want, however, to treat larger systems and an MCSCF calculation correlating all valence electrons would be an insurmountable obstacle. Therefore, we use here as initial guess the natural localized molecular orbitals (NLMO)[24] generated from the density matrices in DFT calculations with the M06-2X functional. We choose this particular functional because it proved to be one of the best among those employed in a systematic test on the barrier heights of chemical reactions.[22] The use of the NLMO procedure of localization is particularly convenient because, in addition to the bonding orbitals, it provides the antibonding ones, which are needed in the construction of the J-LGVBn wave functions. The orbitals are optimized together with the coefficients of the CSFs and the Jastrow parameters in the J-LGVB1 wave functions and then employed without reoptimization in the J-LGVB2 functions.

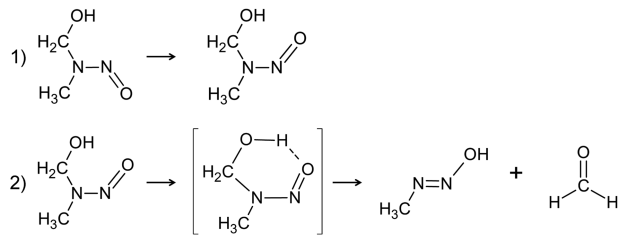

Figure 4. The decomposition of α-hydroxy-dimethylnitrosamine in the two steps identified in the theoretical work of Ref. [10] at the B3LYP/6-311+G(d,p) level.

In the present study of the α-hydroxy-dimethylnitrosamine decomposition, we employ the single-determinant, J-LGVB1, and J-LGVB2 wave functions. For the J-LGVB2 wave functions, we consider three different levels of approximation in the treatment of the dispersion interaction between the electron pairs. In total, we compare five types of wave functions:

1. J-LGVB0: single-determinant wave functions constructed with localized orbitals. For these wave functions, we did not optimize the orbitals simultaneously with the Jastrow factors but kept the NLMO.
2. J-LGVB1: these wave functions include the E1 excitations. Here, we optimize the orbitals simultaneously with the parameters of the Jastrow factor and the coefficients of the CSFs. The optimization of the orbitals is performed in a space that includes only the bonding and antibonding orbitals of each electron pair. Moreover, we divide the space into N_p subspaces, where N_p is the number of the electron pairs of the molecule: Each subspace comprises two orbitals, namely, the bonding and the antibonding orbital of a given electron pair. We therefore perform the optimization only allowing the mixing between the bonding and antibonding orbitals of the same electron pair. Although this choice considerably reduces the computational cost of the orbital optimization, it reduces the variational flexibility and renders the result dependent on the starting orbitals.
3. J-LGVB2-C: these are reduced forms of the J-LGVB2 wave functions, which include all the E1 and a part of the E2 excitations. In Figure 5(i), we illustrate which electron pairs are excluded (drawn in red) from the couplings in the α-hydroxy-dimethylnitrosamine molecule, namely, the C—H bonds and the C2—O1, C6—N3, and N3—N4 σ bonds. These bonds do not change during the chemical

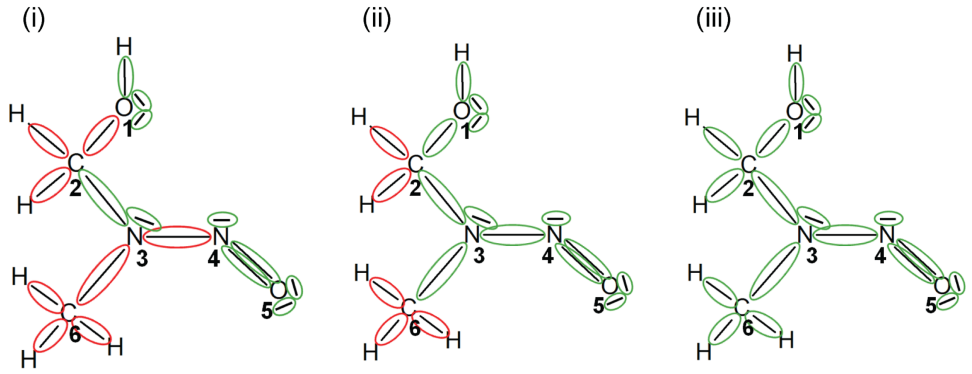

Figure 5. The different levels of coupling of the J-LGVB2 wave functions. The electron pairs labeled in red are not considered in the E2 excitations but only treated at E1 level. (i) Couplings considered in the J-LGVB2-C wave functions. From the E2 excitations, the C—H bond and the C2—O1, C6—N3 and N3—N4 σ bonds are excluded. (ii) In the J-LGVB2-B wave functions, only the C—H bonds are excluded from the E2 excitations. (iii) In the J-LGVB2-A wave functions, all the electron pairs are coupled.

process considered, and we can therefore maintain a balanced treatment of the reaction excluding all the E2 excitation involving these bonds.

4. J-LGVB2-B: they are a more accurate approximation of the J-LGVB2 wave functions than the J-LGVB2-C ones. As shown in Figure 5(ii), only the C—H bonds are excluded from the E2 couplings.
5. J-LGVB2-A: they are the full J-LGVB2 wave functions where all the E1 and E2 excitations are included.

We summarize the size of the trial wave functions in Table 1.

Computational Details

All the geometries of the critical points are optimized using the QCISD/6-311++G** method. The trial functions for the QMC calculations are optimized by energy minimization within VMC using the linear method.[25] As already aforementioned, for the single-determinant wave functions (J-LGVB0), we use the occupied NLMO from M06-2X calculations and optimize only the parameters of the Jastrow factor. For the J-LGVB1 wave functions, we optimize the orbitals, the coefficients of the CSFs, and the Jastrow parameters. In the construction of the J-LGVB2 wave functions, we use the orbitals of the J-LGVB1 wave functions and only optimize the coefficients of the CSFs and the parameters of the Jastrow factor. The Jastrow factors contain electron-nuclear, electron-electron, and electron-electron-nuclear terms.[26] In the QMC calculations, we employ the Burkatzki-Filippi-Dolg (BFD) pseudopotentials[27] and the VTZ basis set specifically developed for these pseudopotentials. The pseudopotentials are treated beyond the locality approximation[28,33] and a time step of 0.05 a.u. is used in the fixed-node DMC calculations. To estimate the time-step error, we also compute the BH(Z-TS1) barrier height with the J-LGVB2-A wave function using a smaller time step of 0.025 a.u. and obtain a value which is statistically equivalent to the result obtained with a time step of 0.05 a.u. This finding is in line with what observed for the dissociation bond energies in our previous work.[8] For comparison, we also compute the energies with the meta-GGA hybrid functional M06-2X, the hybrid functional B3LYP, the double-hybrid functionals B2PLYP and B2PLYP-D, and the global-hybrid meta-GGA functional MPWB1K. Moreover, we assess the QCISD method using the aug-cc-pVTZ basis set. The zero-point energy corrections are calculated with PBE1PBE/6-311++G**. The QMC calculations are performed with the CHAMP program,[34] whereas the QCISD and DFT calculations are performed with Gaussian09.[29]

Table 1. Size of the trial functions used in the DMC calculations for the critical points considered. J-LGVB0 is a single-determinant wave function.

Critical point	J-LGVB1	J-LGVB2-C	J-LGVB2-B	J-LGVB2-A
Number of CSFs				
E	19	45	69	91
TSr1	19	45	69	91
Z	19	45	69	91
TS1	19	53	77	99
P1	19	49	69	93
CH_2O	7	13	19	29
CH_3N_2OH	13	33	47	59
Number of determinants				
E	19	97	169	235
TSr1	19	97	169	235
Z	19	97	169	235
TS1	19	121	193	259
P1	19	109	169	241
CH_2O	7	25	43	73
CH_3N_2OH	13	73	115	151

Results

We focus here on the pathway with a lower activation barrier for the decomposition of α-hydroxy-dimethylnitrosamine in formaldehyde and methyldiazohydroxide (see Fig. 4). The process occurs in two steps, that is, a E-Z isomerization and a retro-ene reaction passing through a transition state with a six-atoms ring. The molecular structures at the critical points of the PES are shown in Figures 6 and 7. In Figure 6, the molecules labeled as E and Z are the geometric isomers of the α-

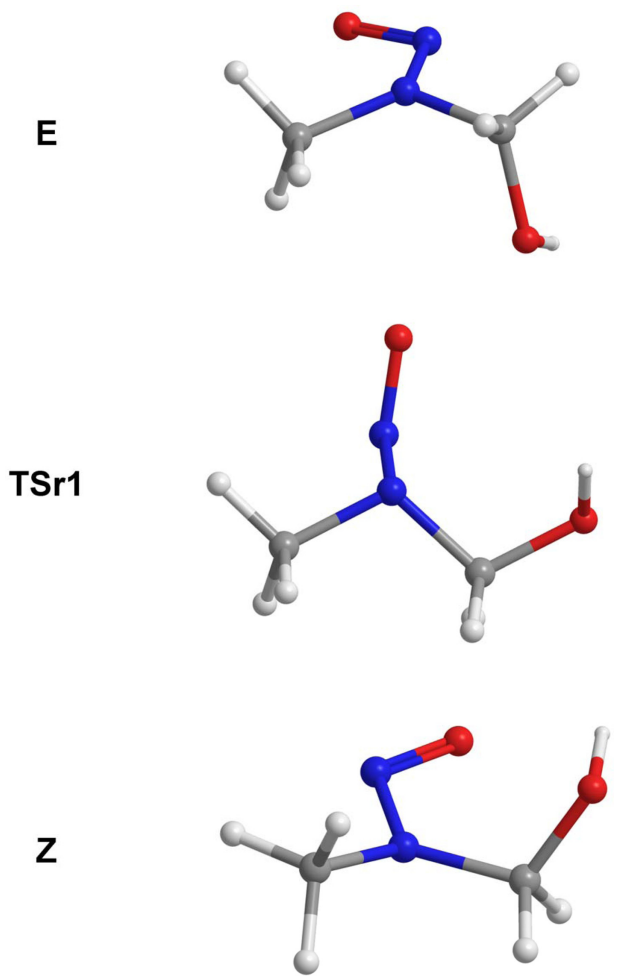

Figure 6. E and Z isomers of α-hydroxy-dimethylnitrosamine, and the transition state (TSr1) of the interconversion reaction between them.

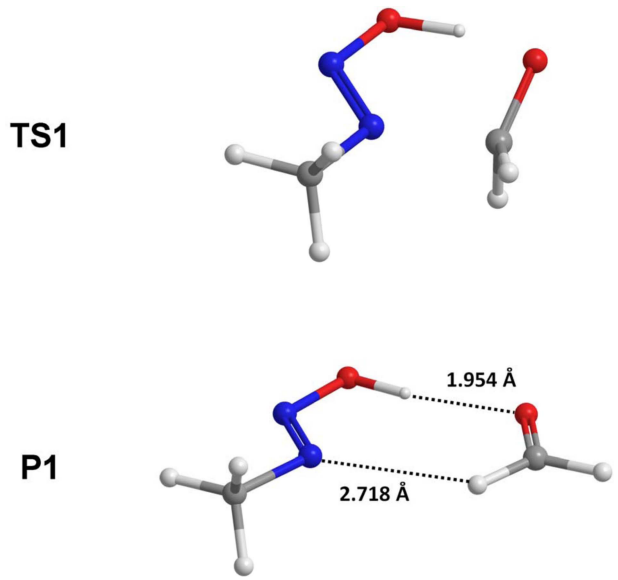

Figure 7. The transition state (TS1) between the Z isomer of α-hydroxy-dimethylnitrosamine and the intermolecular complex (P1) consisting of formaldehyde and methyldiazohydroxide. The geometric parameters of the intermolecular forces in the P1 complex are also given. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

hydroxy-dimethylnitrosamine, and TSr1 is the transition state of the interconversion reaction between these two isomers. In Figure 7, TS1 is the transition state leading to the dissociation of the Z isomer to formaldehyde and methyldiazohydroxide. P1 is the intermolecular complex between formaldehyde and methyldiazohydroxide that is formed after the dissociation. The complex forms a planar ring of seven atoms with the distance between the carbonylic oxygen atom and the hydroxylic hydrogen (1.954 Å) being compatible with a hydrogen bond of medium strength. The distance between the formyl hydrogen and the nitrogen of the diazo group is instead considerably larger (2.718 Å), so this second interaction has to be considered as due to dispersion forces. The schematic diagram of the potential energy profile of the α-hydroxy-dimethylnitrosamine decomposition computed within DMC with the J-LGVB2-A wave function is shown in Figure 8.

In the following, we denote the barrier heights for the forward and reverse E-Z isomerization reaction as BH(E-TSr1) and BH(Z-TSr1), and the forward and reverse barriers for the retro-ene reaction as BH(Z-TS1) and BH(P1-TS1), respectively. The relevant reaction energies are the isomerization reaction [ΔE(Z-E)], the retro-ene reaction [ΔE(P1-Z)], and the dissociation energy of the P1 complex into the formaldehyde and methyldiazohydroxide molecules [ΔE(F-P1)].

We collect the barrier heights calculated with DMC in Table 2. A monotonic lowering of the barrier heights is observed by increasing the quality of the wave function. In the single-determinant case, BH(E-TSr1) is 21.7 kcal/mol and is lowered by 1.7 kcal/mol in the J-LGVB1 case. The use of the second-order wave functions reduces the J-LGVB1 value by 2.2, 2.2, and 2.7 kcal/mol for J-LGVB2-C, J-LGVB2-B, and J-LGVB2-A, respectively. The same regular trend is observed for the reverse BH(Z-TSr1), where J-LGVB1 lowers the J-LGVB0 value by 2 kcal/mol and a further decrease of 0.3, 0.6, and 0.7 kcal/mol with respect to J-LGVB1 is obtained with J-LGVB2-C, J-LGVB2-B, and J-LGVB2-A, respectively. Also for the BH(Z-TS1) barrier height, the single-determinant wave function gives a higher result than the multideterminant ones but the difference is now of about only 1 kcal/mol. The multideterminantal wave functions yield the same result within statistical error

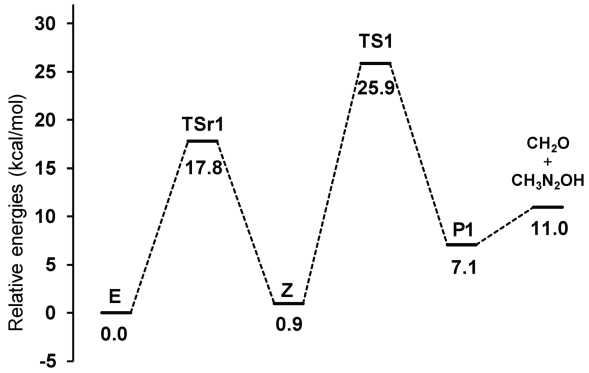

Figure 8. Schematic potential energy surface of α-hydroxy-dimethylnitrosamine decomposition calculated with DMC and the J-LGVB2-A trial wave function, corrected for ZPE (in kcal/mol).

Table 2. Barrier heights (kcal/mol) calculated with DMC and different J-LGVBn wave functions for the decomposition of the α-hydroxy-dimethylnitrosamine molecule.

BH	J-LGVB0	J-LGVB1	J-LGVB2-C	J-LGVB2-B	J-LGVB2-A
Barrier heights					
E-TSr1	21.7	20.0	19.5	19.5	19.0
Z-TSr1	21.0	19.0	18.7	18.4	18.3
Z-TS1	28.3	27.5	27.9	27.3	27.4
P1-TS1	15.1	17.5	17.9	17.6	18.0
Deviations with respect to DMC[J-LGVB2-A]					
E-TSr1	2.6	1.0	0.4	0.5	
Z-TSr1	−2.7	−0.7	−0.5	−0.1	
Z-TS1	0.9	0.1	0.5	−0.1	
P1-TS1	3.0	0.5	0.1	0.4	
MAD	1.8	0.5	0.3	0.2	

The statistical errors are 0.2 kcal/mol. We also report the differences and MAD (kcal/mol) of the DMC barrier heights with respect to the DMC values computed with J-LGVB2-A.

with the exception of the J-LGVB2-C wave function, which gives a slightly higher barrier by about 0.5 kcal/mol. For the barrier of the reverse reaction, BH(P1-TS1), there is a difference of more than 2 kcal/mol between the result of the single determinant and the other wave functions. In this case, the J-LGVB2 wave functions yield slightly higher barriers than J-LGVB1.

If we focus on the deviations from the best J-LGVB2-A values, we find that, for the J-LGVB2-C and J-LGVB2-B wave functions, these differences never exceed 0.5 kcal/mol and the mean absolute deviations (MAD) are, respectively, 0.3 and 0.2 kcal/mol. Therefore, by excluding the interaction between certain electron pairs according to chemical intuition, we obtain an accuracy comparable to the best J-LGVB2-A reference, where all electron pairs are included. The MAD for the J-LGVB1 wave function is instead slightly larger with a maximum difference of 1 kcal/mol. Finally, the performance of the J-LGVB0 wave function is significantly worse with a MAD of 1.8 kcal/mol and a maximum deviation of 3 kcal/mol.

The barrier heights computed with the DFT and QCISD methods are presented in Table 3, together with the corresponding deviations and MAD with respect to the DMC values obtained with the J-LGVB2-A wave functions. In general, the differences between the DFT and DMC results are very pronounced, while we obtain a good agreement between QCISD and DMC. For the BH(E-TSr1) barrier height, these functionals provide larger values than DMC, while the QCISD barrier is close to the multideterminantal DMC estimate. The behavior is the same for the barrier of the reverse reaction. For BH(Z-TS1), the agreement between the DFT and DMC results is better, except for the B3LYP functional that yields a too small value. In this case, the difference between the QCISD and J-LGVB2-A data is 1.3 kcal/mol. For the BH(P1-TS1) barrier height, all functionals give differences of 4 kcal/mol or more when compared to the J-LGVB2-A, while the deviation of QCISD is 1 kcal/mol. We note that the best functional is M06-2X, which gives a mean absolute deviation of 2.7 kcal/mol, while the worse is MPWB1K with a MAD of 3.8 kcal/mol. Also the B3LYP functional is generally unsatisfactory, its MAD being 3.5 kcal/mol.

The DMC reaction energies are shown in Table 4. The isomerization energy is very small (0.7–1.1 kcal/mol) and there are no significant variations between the different levels of the J-LGVBn calculations. For the dissociation energy of the Z isomer into the P1 complex, we observe instead an important lowering of the reaction energy in passing from the single determinant to the multideterminat wave functions. Furthermore,

Table 3. Barrier heights (kcal/mol) calculated with the DFT and QCISD methods and the aug-cc-pVTZ basis set.

BH	M06-2X	B3LYP	MPWB1K	B2PLYP	B2PLYP-D	QCISD
Barrier heights						
E-TSr1	22.6	24.3	24.5	23.7	23.5	19.7
Z-TSr1	22.3	23.7	24.3	23.2	23.1	19.3
Z-TS1	27.9	24.6	29.0	26.0	26.4	28.7
P1-TS1	12.4	14.1	12.0	13.9	13.1	17.1
Deviations with respect to DMC[J-LGVB2-A]						
E-TSr1	3.6	5.2	5.5	4.6	4.5	0.7
Z-TSr1	−4.0	−5.4	−6.1	−4.9	−4.8	−1
Z-TS1	0.5	−2.9	1.6	−1.4	−1.1	1.3
P1-TS1	5.6	3.9	6.0	4.1	4.9	1.0
MAD	2.7	3.5	3.8	3.0	3.1	0.8

The differences and MAD with respect to the reference DMC values obtained with the J-LGVB2-A wave functions are also given.

Table 4. Reaction energies (kcal/mol) calculated with the DMC method and different J-LGVBn wave functions for the Z-E isomerization, retro-ene reaction, and dissociation of the P1 intermolecular complex.

ΔE	J-LGVB0	J-LGVB1	J-LGVB2-C	J-LGVB2-B	J-LGVB2-A
Reaction energies					
Z-E	0.7	1.0	0.7	1.1	0.8
P1-Z	13.2	10.0	10.0	9.7	9.4
F-P1	5.7	5.6	5.4	4.9	5.5
Deviations with respect to DMC[J-LGVB2-A]					
Z-E	−0.1	0.3	0.0	0.4	
P1-Z	3.8	0.6	0.6	0.3	
F-P1	0.2	0.1	−0.2	−0.7	
MAD	1.4	0.3	0.3	0.4	

The statistical errors are 0.2 kcal/mol. We also report the differences and MAD (kcal/mol) of the DMC reaction energies from the DMC values obtained with the J-LGVB2-A wave functions.

when we increase the level of the multideterminantal J-LGVBn wave functions, ΔE(P1-Z) is lowered from 10.0 for J-LGVB1 to 9.4 kcal/mol for J-LGVB2-A. The calculated dissociation energy of the P1 complex into formaldehyde and methyldiazohydroxide are statistically equivalent for all wave functions with the exception of J-LGVB2-B, which underestimates the interaction energy of the P1 complex by 0.7 kcal/mol with respect to the J-LGVB2-A reference. We can possibly explain this deviation by noting that, in J-LGVB2-B, the dispersion force between the formyl hydrogen and the nitrogen atom of the diazo group in P1 is not properly described: The J-LGVB2-B wave function does not account for the interactions of the C—H bonds with other electron pairs and, therefore, does not include the CSFs responsible for the interaction between the C—H bond of formaldehyde and the lone pair on the nitrogen atom. These CSFs are also not present in the J-LGVB2-C wave function but, in this case, a favorable cancellation of the errors occurs due to the neglect of other couplings, namely, the C—O, C—N, and N—N σ bonds. This explanation is based on the assumption that the E2 CFSs give a minor contribution in the description of the dynamic correlation. The estimate of the intermolecular energy in P1 complex is the only case where the multilevel approach in the construction of the J-LGVB2 wave functions does not provide fully satisfactory results.

We collect the DFT and QCISD reaction energies and compare them to the DMC reference data obtained with the J-LGVB2-A wave functions in Table 5. The Z-E isomerization energy is slightly underestimated by all methods. The functionals that provide the best agreement with DMC are B3LYP and B2PLYP. For the dissociation energy of the α-hydroxydimethylnitrosamine Z isomer into the P1 complex, we observe a wide spread in the DFT results. With the exception of B3LYP, all functionals significantly overestimate the reaction energy. In particular, the M06-2X and MPWB1K functionals give errors in ΔE(P1-Z) as large as 6.1 and 7.6 kcal/mol, respectively. Finally, for the dissociation energy of the P1 complex, there is good agreement between the DFT and DMC data. The only functionals which give somewhat larger deviations are B3LYP and B2PLYP-D, the former underestimating the interaction by 1 kcal/mol and the latter overestimating it by 0.9 kcal/mol.

In Table 7, we present the final DFT, QCISD, and DMC energy barriers of the activation process of the α-hydroxydimethylnitrosamine corrected for the zero-point energies calculated at the PBE1PBE/6-311++G** level. According to a study by Alecu et al.[30] on a sample of 15 molecules, the root mean square deviation of the PBE1PBE functional in the ZPE calculations (when frequencies are scaled) is 0.15 kcal/mol. This value is comparable to the statistical error in the QMC calculations and has therefore a negligible effect on the estimate of the energy barriers. The ZPE values are separately reported in Table 6. Our highly correlated calculations confirm the mechanism proposed in the B3LYP investigation of Ref. [10] as the dominant barrier of the retro-ene reaction is not too dissimilar from the one reported in this previous study. On the other hand, we find that,

Table 5. Reaction energies (kcal/mol) calculated with the DFT and QCISD methods and the aug-cc-pVTZ basis set for the Z-E isomerization, retro-ene reaction, and dissociation of the P1 intermolecular complex.

ΔE	M06-2X	B3LYP	MPWB1K	B2PLYP	B2PLYP-D	QCISD
Reaction energies						
Z-E	0.4	0.6	0.2	0.5	0.4	0.5
P1-Z	15.5	10.4	17.0	12.1	13.3	11.7
F-P1	6.1	4.5	5.1	5.4	6.4	6.0
Deviations with respect to DMC[J-LGVB2-A]						
Z-E	−0.4	−0.2	−0.6	−0.2	−0.3	−0.3
P1-Z	6.1	1.0	7.6	2.7	3.9	2.2
F-P1	0.6	−1.0	−0.5	−0.1	0.9	0.5
MAD	2.4	0.7	2.9	1.0	1.7	1.0

The differences and MAD with respect to the reference DMC values obtained with the J-LGVB2-A wave functions are also given.

Table 6. Zero-point energy corrections (kcal/mol) calculated in the harmonic approximation with PBE1PBE/6-311++G**.

Molecule	ZPE
E	59.9
TSr1	58.7
Z	60.1
TS1	57.6
P1	56.8
CH_2O	16.7
CH_3N_2OH	38.5

in contrast to our and previous B3LYP results, the rate-determining step is not the isomerization but the retro-ene reaction. At the DMC and QCISD level, the difference between the barrier heights of these two steps is rather large. Within DMC with the J-LGVB2-A wave functions, the BH(Z-TS1) barrier is 7.1 kcal/mol higher than BH(E-TSr1). In the DFT calculations, the difference between the two barrier heights is less pronounced and the barrier to isomerization becomes slightly lower also in our B3LYP calculations employing the QCISD structures as in Ref. [10]. The geometries of the considered molecules, and the fixed-node DMC energies are given in the Supporting Informations.

Conclusions

In this work, we have presented several novel features of our recently proposed J-LGVBn wave functions, which allow a consistently accurate description of complex PES of medium-large systems within QMC. In particular, we have exploited the local nature of the J-LGVBn wave function to develop a multilevel scheme that describes electron correlation at different levels of accuracy in different regions of the molecule. As prototypical study case, we have investigated the carcinogenic activation of the NDMA molecule. This chemical process includes the E-Z isomerization reaction of α-hydroxy-dimethylnitrosamine, the retro-ene reaction of the Z isomer which dissociates to the intermolecular complex formed by formaldehyde and methyldiazohydroxide, and finally the dissociation of the complex in the two isolated molecules, of which methyldiazohydroxide is the actual carcinogen. We have applied the J-LGVBn trial functions with the fixed-node DMC method and optimized the molecular geometries of the critical points with the QCISD/6-311++G** method. In addition to the use of DMC, we have calculated the single-point energies with several DFT functionals and QCISD/aug-cc-pVTZ.

Compared to our previous work with the J-LGVBn wave functions, we introduced here some important innovations to extend the applicability of the method: (i) as initial guess, we used natural localized orbitals from localization of M06-2X orbitals instead of MCSCF orbitals; (ii) we used symmetry constraints in the orbital optimization, allowing the mixing only between the bonding and antibonding orbitals of the same electron pair; (iii) we have proposed a multilevel approach for the construction of the J-LGVB2 wave functions.

The results show that the use of multideterminant is crucial in the fixed-node DMC to accurately study the critical points of this complex PES. Furthermore, we find that our novel multilevel approach can be used to increase the compactness of the trial functions without a considerable loss of accuracy. Therefore, this technique opens, in perspective, the application of the method to considerably larger systems. It should, however, be noted that it is necessary to select carefully the parts of the molecule to be treated at different levels of the theory.

For the activation of the NDMA molecule, an important result obtained here is the confirmation of the validity of the two-step mechanism of α-hydroxy-dimethylnitrosamine decomposition proposed in a previous study as well as the identification of the correct rate-determining step of this process. At variance with the B3LYP results obtained here and in a previous study,[21] highly correlated methods yield a large difference between the barriers for the isomerization and for the retro-ene reaction, clearly indicating that the retro-ene reaction is the rate-determining step with a barrier height of 24.9 kcal/mol. We stress that the similarity between the global activation energies calculated here (with QMC on QCISD geometries) and in Ref. [21] (with the B3LYP functional on B3LYP geometries) must be viewed as accidental as B3LYP cannot generally achieve the accuracy of a highly correlated approach like QMC in the calculation of a complicate reaction path, as also demonstrated by its poor performance on the other critical points of this reaction.

Finally, we note that the barrier height of the process studied in this work could in principle be used as a quantum chemical descriptor[31] of NDMA in QSAR studies of carcinogenicity. Of course, similar calculations should be performed for a set of homologue molecules like those investigated in the study of Helguera et al.[32] to establish a relationship with carcinogenicity.

Keywords: quantum Monte Carlo · dimethylnitrosamine · J-LGVB wave functions · potential energy surfaces

Table 7. Barrier heights and dissociation energy of the P1 intermolecular complex (kcal/mol) calculated with the DFT, QCISD, and DMC[J-LGVB2-A] methods corrected for the zero-point energies.

	M06-2X	B3LYP	MPWB1K	B2PLYP	B2PLYP-D	QCISD	DMC	Ref.[10]
BH(E-TSr1)	21.4	23.1	23.3	22.5	22.3	18.5	17.8	24.11
BH(Z-TSr1)	20.9	22.3	23.0	21.8	21.7	17.9	16.9	25.34
BH(Z-TS1)	25.4	22.1	26.5	23.5	23.9	26.2	24.9	23.17
BH(P1-TS1)	13.1	14.9	12.8	14.7	13.8	17.8	18.8	15.44
ΔE(F-P1)	4.5	2.9	3.4	3.8	4.8	4.4	3.9	

The B3LYP/6-311+G(d,p) data from Ref. [10] are also listed.

How to cite this article: F. Fracchia, C. Filippi, C. Amovilli. *J. Comput. Chem.* **2014**, *35*, 30–38. DOI: 10.1002/jcc.23461

Additional Supporting Information may be found in the online version of this article.

[1] A. Luchow, *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2011**, *1*, 388.
[2] M. Korth, A. Lüchow, S. Grimme, *J. Phys. Chem. A* **2008**, *112*, 2104.
[3] O. Valsson, C. Filippi, *J. Chem. Theory Comput.* **2010**, *6*, 1275.
[4] M. Dubecký, R. Derian, L. Mitas, I. Štich, *J. Chem. Phys.* **2010**, *133*, 244301.
[5] F. M. Floris, C. Filippi, C. Amovilli, *J. Chem. Phys.* **2012**, *137*, 075102.
[6] C. Filippi, F. Buda, L. Guidoni, A. Sinicropi, *J. Chem. Theory Comput.* **2012**, *8*, 112.
[7] O. Valsson, P. Campomanes, I. Tavernelli, U. Rothlisberger, C. Filippi, *J. Chem. Theory Comput.* **2013**, *9*, 2441.
[8] F. Fracchia, C. Filippi, C. Amovilli, *J. Chem. Theory Comput.* **2012**, *8*, 1943.
[9] F. Fracchia, C. Filippi, C. Amovilli, *J. Chem. Theory Comput.* **2013**, *9*, 3453.
[10] C. Lu, Y. Liu, R. Zhong, *J. Mol. Struct. (Theochem)* **2009**, *893*, 106.
[11] E. C. Miller, J. A. Miller, *Cancer* **1981**, *47*, 1055.
[12] E. C. Miller, J. A. Miller, *Cancer* **1981**, *47*, 2327.
[13] H. Bartsch, R. Montesano, *Carcinogenesis* **1984**, *11*, 1381.
[14] W. A. Mitch, J. O. Sharp, R. R. Trussell, R. L. Valentine, L. Alvarez-Cohen, D. L. Sedlak, *Environ. Eng. Sci.* **2003**, *20*, 389.
[15] S. S. Walse, W. A. Mitch, *Environ. Sci. Technol.* **2008**, *42*, 1032.
[16] S. S. Hecht, *Chem. Res. Toxicol.* **1998**, *11*, 559.
[17] S. S. Hecht, *J. Natl Cancer Inst.* **1999**, *91*, 1194.
[18] L. Ji, G. Schüürmann, *J. Phys. Chem. B* **2012**, *116*, 903.
[19] A. Streeter, R. Nims, P. Sheffels, Y. Heur, C. Yang, B. Mico, C. Gombar, L. Keefer, *Cancer Res.* **1990** *50*, 1144.
[20] V. Lee, L. Keefer, M. Archer, *Chem. Res. Toxicol.* **1996**, *9*, 1319.
[21] L. Li, A. Zhang, Z. Li, *J. Mol. Struct. (Theochem)* **2006**, *759*, 239.
[22] J. Zheng, Y. Zhao, D. Truhlar, *J. Chem. Theory Comput.* **2009**, *5*, 808.
[23] B. J. Lynch, D. G. Truhlar, *J. Phys. Chem. A* **2001**, *105*, 2936.
[24] A. Reed, F. Weinhold, *J. Chem. Phys.* **1985**, *83*, 1736.
[25] C. J. Umrigar, J. Toulouse, C. Filippi, S. Sorella, R. G. Hennig, *Phys. Rev. Lett.* **2007**, *98*, 110201.
[26] C. Filippi, C. J. Umrigar, *J. Chem. Phys.* **1996**, *105*, 213.
[27] M. Burkatzki, C. Filippi, M. Dolg, *J. Chem. Phys.* **2007**, *126*, 234105.
[28] M. Casula, *Phys. Rev. B* **2006**, *74*, 161102.
[29] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. J. A. Vreven, J. Montgomery, J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, O. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, D. J. Fox, Gaussian 09; Gaussian, Inc.: Wallingford, CT, **2009**.
[30] I. M. Alecu, J. Zheng, Y. Zhao, D. G. Truhlar, *J. Chem. Theory Comput.* **2010**, *6*, 2872.
[31] M. Karelson, V. S. Lobanov, A. R. Katritzky, *Chem. Rev.* **1996**, *96*, 1027.
[32] A. M. Helguera, M. N. D. S. Cordeiro, M. A. Cabrera Perez, R. D. Combes, M. P. Gonzales, *Toxicol. Appl. Pharmacol.* **2008**, *231*, 197.
[33] For the hydrogen atom, we use a more accurate BFD pseudopotential and basis set. M. Dolg and C. Filippi, private communication.
[34] CHAMP is a quantum Monte Carlo program package written by C. J. Umrigar, C. Filippi, and collaborators.

Received: 19 June 2013
Revised: 7 September 2013
Accepted: 16 September 2013
Published online on 21 October 2013