



**HAL**  
open science

## Low-lying excited states of model proteins: Performances of the CC2 method versus multireference methods

Nadia Ben Amor, Sophie Hoyau, Daniel Maynau, Valerie Brenner

► **To cite this version:**

Nadia Ben Amor, Sophie Hoyau, Daniel Maynau, Valerie Brenner. Low-lying excited states of model proteins: Performances of the CC2 method versus multireference methods. *Journal of Chemical Physics*, 2018, 148 (18), pp.184105. 10.1063/1.5025942 . hal-01802401

**HAL Id: hal-01802401**

**<https://hal.science/hal-01802401>**

Submitted on 12 Nov 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Low-lying excited states of model proteins: Performances of the CC2 method versus multireference methods

Nadia Ben Amor, Sophie Hoyau, Daniel Maynau, and Valérie Brenner

Citation: *The Journal of Chemical Physics* **148**, 184105 (2018); doi: 10.1063/1.5025942

View online: <https://doi.org/10.1063/1.5025942>

View Table of Contents: <http://aip.scitation.org/toc/jcp/148/18>

Published by the [American Institute of Physics](#)

---

---

**PHYSICS TODAY**

WHITEPAPERS

## ADVANCED LIGHT CURE ADHESIVES

Take a closer look at what these environmentally friendly adhesive systems can do

READ NOW

PRESENTED BY  
 **MASTERBOND**  
ADHESIVES | SEALANTS | COATINGS

## Low-lying excited states of model proteins: Performances of the CC2 method versus multireference methods

Nadia Ben Amor,<sup>1,2,a)</sup> Sophie Hoyau,<sup>2</sup> Daniel Maynaud,<sup>1,2</sup> and Valérie Brenner<sup>3</sup>

<sup>1</sup>CNRS, UPS, LCPQ (Laboratoire de Chimie et Physique Quantiques), IRSAMC, 118, Rte de Narbonne, F-31062 Toulouse Cedex, France

<sup>2</sup>UPS, LCPQ (Laboratoire de Chimie et Physique Quantiques), IRSAMC, Université de Toulouse, 118, Rte de Narbonne, F-31062 Toulouse Cedex, France

<sup>3</sup>Laboratoire Interactions, Dynamiques et Lasers, LIDYL, CEA, CNRS, Université Paris-Saclay, 91191 Gif-sur-Yvette Cedex, France

(Received 14 February 2018; accepted 10 April 2018; published online 9 May 2018)

A benchmark set of relevant geometries of a model protein, the *N*-acetylphenylalanylamine, is presented to assess the validity of the approximate second-order coupled cluster (CC2) method in studying low-lying excited states of such bio-relevant systems. The studies comprise investigations of basis-set dependence as well as comparison with two multireference methods, the multistate complete active space 2nd order perturbation theory (MS-CASPT2) and the multireference difference dedicated configuration interaction (DDCI) methods. First of all, the applicability and the accuracy of the quasi-linear multireference difference dedicated configuration interaction method have been demonstrated on bio-relevant systems by comparison with the results obtained by the standard MS-CASPT2. Second, both the nature and excitation energy of the first low-lying excited state obtained at the CC2 level are very close to the Davidson corrected CAS+DDCI ones, the mean absolute deviation on the excitation energy being equal to 0.1 eV with a maximum of less than 0.2 eV. Finally, for the following low-lying excited states, if the nature is always well reproduced at the CC2 level, the differences on excitation energies become more important and can depend on the geometry. *Published by AIP Publishing.* <https://doi.org/10.1063/1.5025942>

### I. INTRODUCTION

Many complex molecular systems absorbing light in the near UV spectral range, including those of paramount biological importance like DNA bases or proteins, are endowed with mechanisms of excited-state deactivation following UV absorption.<sup>1</sup> These processes are of major importance for the photochemical stability of these species since they provide a rapid and efficient way of dissipating the electronic energy in excess into vibration, thus avoiding photochemical processes to take place and then structural damages which can affect the biological function.<sup>2,3</sup> They are controlled by the energy and the nature of the electronic excited states of the chromophores, by their couplings and the resulting electron dynamics: UV light absorption populates excited states, which dissipate the electronic energy, either through a relatively slow radiative deactivation process, i.e., photon emission, or, more often and more efficiently, by a radiationless transition, e.g., internal conversion or intersystem crossing. The latter, the nonadiabatic (NA) transfers, often involve ultrafast energy transfers through regions of the potential energy surfaces (PES) corresponding to avoided or surface crossings, of conical nature or not.<sup>4–9</sup>

In order to investigate conformer-selective dynamics of biologically relevant molecular systems and, in particular, the building block of proteins such as capped peptides, we have developed an original innovative computational strategy. Our

main goal is to document the basic physical phenomena controlling the lifetime of excited states, highlighting the link between electronic dynamics and structure.<sup>10,11</sup>

The multi-step multi-level computational strategy allows us to both characterize the low-lying excited states of bio-relevant systems and to model efficiently their PES using, first, nonadiabatic dynamic simulations based on time-dependent density functional theory (NA-TDDFT) to provide hints about the critical motions that drive the deactivation. Two better levels of theory are then used to refine these simulations: (i) the standard approximate coupled cluster singles and doubles method (CC2)<sup>12–16</sup> and (ii) a multireference configuration interaction (MRCI) method.<sup>17–19</sup> The NA-TDDFT simulations and refinement of the energy profiles at the CC2 level were previously applied on small capped peptides, the *N*-acetylphenylalanylamine (NAPA) and its *N*-methylated derivative (NAPMA). It highlighted, for the first time in such systems, the quenching properties of the primary amide group (through its  $\pi^*_{CO}$  excited state) along with the effect of vibrational energy that facilitated access to the conical intersection (CI) area.<sup>10,11</sup> This paper is fully in line with these studies and focuses on the third step of the computational strategy that has, to our knowledge, never been addressed: assessment for such systems of the accuracy of the CC2<sup>12–16</sup> method by comparison with a MRCI method.<sup>17–19</sup>

The challenge of such calculations is multiple. First, one of our objectives is not only to assess the accuracy of the CC2 method on the equilibrium geometry of the initially excited state accessible from the Frank-Condon region (generally a

<sup>a)</sup>Author to whom correspondence should be addressed: benamor@irsamc.ups-tlse.fr

singlet  $\pi\pi^*$  excited state on the aromatic ring) but also to assess them on the equilibrium geometries of all close-lying excited singlet states of these systems as well as on relevant geometries of the energy profile of the deactivation mechanisms such as the conical intersections (CI) between excited states. Second, these systems present specificities that are decisive in the choice of methods, i.e., the level of theory: (i) their size (medium-size systems where the smallest one, a capped peptide with one residue, already contains at least thirty atoms), (ii) their lack of symmetry, (iii) their great flexibility due to the non-covalent interactions that govern their structure, and (iv) their multiple close-lying excited states featuring very different nature [the locally excited state on the peptide bonds ( $LE_{\text{pep}}$ ) or on the aromatic ring ( $LE_{\pi}$ ) and even the charge transfer (CT) state]. Finally, our ultimate goal is to apply this computational strategy to capped peptides which contain more than one residue but this latter point will be investigated and discussed in a future paper. Even if the validity of the CC2 method for equilibrium structures and energy profile of both ground and excited states of small peptides has been already established,<sup>20–22</sup> there is no benchmark of the CC2 method for application to the different close-lying excited states of such systems. Two of the most recent and extensive benchmarks of the CC2 method which use as reference data experimental values previously compiled from high-resolution gas-phase experiments,<sup>23,24</sup> and references therein, concern the adiabatic transitions of the lowest excited state of a set of different medium-size molecules containing either aromatic organic molecules including different conformers (66 molecules: Mean Absolute Error (MAE) of 0.08 eV and Mean Signed Error (MSE) of 0.04 eV)<sup>23</sup> or covering both organic (polyenes, carbonyl compounds, aromatic hydrocarbons, and heterocycle aromatic compounds) and inorganic (main-group and transition metal compounds) systems (79 molecules: MAE of 0.19 eV and MSE of 0.11 eV).<sup>24</sup> Concerning the theoretical benchmarks, only a few studies using reference values obtained at a very high level of theory, such as MRCI, focus on medium-size molecular systems such as the retinal-chromophore model<sup>25</sup> or 9H-adenine.<sup>26</sup> Moreover, these studies deal with energy profiles of mechanisms involving PES of only two states, i.e., the ground state and one excited state.

In this work, we have gone further in investigating the accuracy of the CC2 method for applications to low-lying excited states of capped peptides by performing calculations with the MR-Difference Dedicated Configuration Interaction (DDCI) method.<sup>27,28</sup> The performances of the CC2 method for the description of both the ground and excited states of a conformer of NAPA, i.e., NAPA B, are first evaluated at the ground state equilibrium geometry. Second, relevant geometries along the energy profile of deactivation mechanisms such as equilibrium geometries of the low-lying excited states or conical intersection (CI) geometry are considered. In this benchmark, the multireference method used as reference is the DDCI method.<sup>27,28</sup> This method was previously used for magnetic systems, organic molecules, and carbon nanotubes.<sup>17–19,29–32</sup> We report here on the conditions of the quasi-linear version of this multireference method to bio-relevant systems. The appropriate parameters used in this quasi-linear CI for the capped

peptides are defined and validated by comparison with calculations performed with a more standard multireference method, the multistate complete active space 2nd order perturbation theory (MS-CASPT2).<sup>33</sup>

## II. COMPUTATIONAL DETAILS

### A. Geometries

Let us define some notations used in the following: *pept1* or (1) refers to the first peptide bond in interaction with the phenyl ring (NH- $\pi$  interaction), while the other peptide bond is denoted (2) or *pept2* (see Fig. 1). In order to evaluate the performances of the CC2 method, different geometries of the NAPA B conformer (see Fig. 1 and Appendix S1 in the [supplementary material](#)) have been investigated with both CC2 and MR methods. First, calculations were performed at the B97-D2/TZVPP (noted DFT-D hereafter) equilibrium geometry of the ground state.<sup>34</sup> Second, three CC2/cc-pVDZ pertinent geometries of the energy profile of the deactivation mechanism identified recently and involving the second peptide bond (MII in Refs. 10 and 11) were chosen: (i) the equilibrium geometry of the first  $\pi\pi^*$  excited state ( $M_{\pi\pi^*}$ ), (ii) the equilibrium geometry of the  $n\pi^*_{\text{CO}}$  excited state localized on the second peptide bond [ $M_{n\pi^*_{\text{CO}}}(2)$ ], and (iii) the CI geometry connecting the equilibrium geometries of these two excited states ( $CI_{\pi\pi^*/n\pi^*_{\text{CO}}}(2)$ ). The last considered geometry was the partially optimized geometry [NH(1) bond length constrained to 1.216 Å] of the CT excited state ( $M'_{\text{CT}}$ ) recently identified in a deactivation mechanism (mechanism I in Refs. 10 and 11). In the following, NH(1) corresponds to the NH bond of the first peptide bond pointing to the phenyl ring.

### B. CC2 calculations

The CC2<sup>12–15</sup> calculations were carried out with the TURBOMOLE package.<sup>35,36</sup> All the calculations were performed by using the resolution-of-identity (RI) approximation for the electron repulsion integrals used in the correlation treatment and the description of the excitation processes. cc-pVXZ (X = D, T, and Q) Dunning's correlation consistent basis sets<sup>37</sup> were employed in connection with optimized auxiliary basis sets for the RI approximation.<sup>38</sup> Few calculations were also performed with the addition of diffuse functions. Frozen core for the 1s electrons was employed, and all calculations were carried out in the  $C_1$  point-group symmetry. Twenty singlet states were considered and  $D_1$  and  $D_2$  diagnostics and  $\% \langle \overline{E}_1 | E_1 \rangle$  biorthogonal norm were calculated in order to evaluate the capability of the CC2 method to properly describe the excited states of such systems.<sup>15,39,40</sup> Indeed, the  $D_1$  and  $D_2$  diagnostics, computed from the single and double substitution amplitudes in the CC2 wave function, have been found to be reliable indicators when static or dynamic correlation effects are not adequately treated at the CC2 level: their magnitudes are correlated with the performance of the CC2 method. In addition, the biorthogonal norm  $\langle \overline{E}_1 | E_1 \rangle$  gives a measure of the weight of the single excitation contributions to an excited state. Indeed, in order to be well described at the CC2 level, an excited state must be dominated by single excitations out of the ground state wave function.

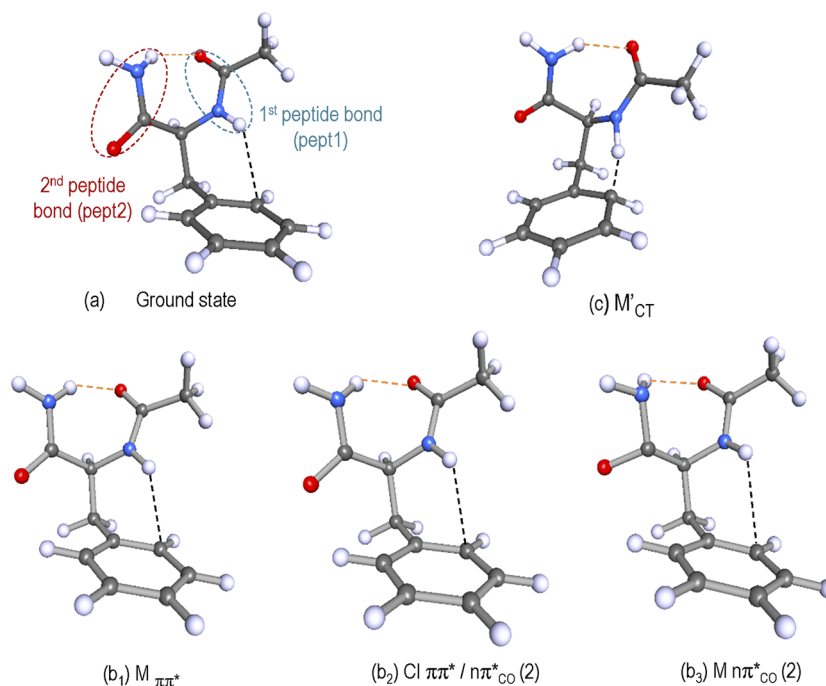


FIG. 1. (a) DFT-D optimized geometry of the ground state of the conformer B of NAPA exhibiting a C<sub>7</sub> H-bond (red dashes,  $d_{\text{NH}(2)-\text{O}} = 2.02 \text{ \AA}$ ) and an NH- $\pi$  interaction (black dashes,  $d_{\text{NH}(1)-\text{C}\pi} = 2.57 \text{ \AA}$ ); (b) CC2/cc-pVDZ geometries of the energy profile of the M<sub>II</sub> deactivation mechanism of NAPA B: (b<sub>1</sub>)  $M_{\pi\pi^*}$ , the optimized geometry of the  $\pi\pi^*$  excited state (red dashes,  $d_{\text{NH}(2)-\text{O}} = 1.95 \text{ \AA}$  and black dashes,  $d_{\text{NH}(1)-\text{C}\pi} = 2.35 \text{ \AA}$ ), (b<sub>2</sub>)  $\text{CI } \pi\pi^*/n\pi^*\text{CO}(2)$ , the geometry of the CI (red dashes,  $d_{\text{NH}(2)-\text{O}} = 2.00 \text{ \AA}$  and black dashes,  $d_{\text{NH}(1)-\text{C}\pi} = 2.42 \text{ \AA}$ ), and (b<sub>3</sub>)  $M_{n\pi^*\text{CO}(2)}$ , the optimized geometry of the excited state  $n\pi^*\text{CO}$  localized on the second peptide bond (red dashes,  $d_{\text{NH}(2)-\text{O}} = 2.28 \text{ \AA}$  and black dashes,  $d_{\text{NH}(1)-\text{C}\pi} = 2.66 \text{ \AA}$ ); (c)  $M'_{\text{CT}}$ , the CC2/cc-pVDZ partially optimized geometry with the NH(1) bond length constrained to  $1.216 \text{ \AA}$  of the intramolecular CT excited state  $n\pi^*\text{cycle}$  (red dashes,  $d_{\text{NH}(2)-\text{O}} = 1.87 \text{ \AA}$  and black dashes,  $d_{\text{NH}(1)-\text{C}\pi} = 1.35 \text{ \AA}$ ). All the coordinates corresponding to these geometries are reported in the [supplementary material](#).

### C. MR calculations

The complete active space self-consistent field (CASSCF)<sup>41</sup>/MS-CASPT2<sup>33</sup> calculations were carried out with the 7.8 MOLCAS package.<sup>42</sup> The linear scaling MRCI calculations<sup>17–19</sup> require the use of local orbitals. The CASSCF orbitals were then localized with the DOLO code<sup>43,44</sup> and the linear scaling MRCI calculations were performed with the EXSCI program.<sup>44</sup> The Davidson correction was introduced to correct the size-consistency error inherent in the MRCI methods. These multireference calculations were performed with different basis sets. First, the Atomic Natural Orbitals (ANO-L)<sup>45</sup> basis sets were used with the following contraction scheme: for *C*, *N*, and *O*, a (14s9p4d3f) set contracted to [3s2p1d] and for *H* a (8s4p3d) set contracted to [2s1p]. Dunning's correlation consistent basis sets cc-pVDZ and cc-pVTZ<sup>37</sup> were also chosen for comparison with the CC2 results. Finally, as for the CC2 calculations, few computations were performed with the addition of diffuse functions.

A well balanced active space has to be defined to provide a good description of all considered singlet states. These states may correspond to local excitations (LE): (i) centered on the phenyl group ( $\text{LE}_\pi$ ) and (ii) centered on each peptide bond [ $n \rightarrow \pi^*\text{CO}$  where *n* is a N or O lone pair (pure-p lone pair),  $\text{LE}_{\text{pep}}$ , and  $\pi\text{CO} \rightarrow \pi^*\text{CO}$ ] or to excitations corresponding to electronic charge transfer (CT) from the backbone to the phenyl group. Therefore, 18 electrons in 14 orbitals were included in the active space, corresponding to all the  $\pi$  and  $\pi^*$  orbitals of the phenyl and carbonyl groups and to lone

pairs on the nitrogen and oxygen atoms. However, for both  $n\pi^*\text{CO}$  and CT excited state geometries, this active space had to be enlarged in order to include one additional orbital: (i) an *n* orbital (pept2) for the equilibrium  $n\pi^*\text{CO}$  excited state geometry, i.e., the second lone pair (sigma lone pair) on the second peptide bond oxygen atom; (ii) a  $\sigma\text{NH}(1)$  (pept1) orbital pointing toward the phenyl group for the partially optimized CT excited state geometry, with the NH(1) bond length constrained to  $1.216 \text{ \AA}$ . In these two cases, 20 electrons in 15 orbitals were included in the active space.

MS-CASPT2 calculations were performed on CASSCF reference wave functions. Average molecular orbitals (MO) are generally a good compromise to obtain the whole set of states involved in the electronic spectrum. State-average orbitals on 14 and 20 singlet states were then optimized in order to obtain the charge transfer states (see discussion in Sec. III). In the MS-CASPT2 method, the use of a level shift allows us to avoid weak intruder states by the addition of a shift parameter to the zeroth-order Hamiltonian. Several trials were performed, with various level shift values: 0.0, 0.1, 0.2, 0.3, 0.4, and 0.5 a.u. (Figs. S2-1 and S2-2 of the [supplementary material](#)). The final choice was made on the basis of the deviation of the reference weight of the zeroth-order wave functions of the excited states compared to the ground state one, as well as on the basis of the stability of the excitation energies. A maximal deviation of 3% for the reference weight was obtained with a level shift of 0.5 a.u. and this value was then used for all the CASPT2 calculations. Another modification of the zeroth-order Hamiltonian, called the Ionization

Potential Electronic Affinity shift, was introduced in order to reduce the systematic error on the relative energies comparing closed shell and open shell states. The 0.25 default value was used for the IPEA shift. The core electrons were frozen. Finally, the Cholesky decomposition technique<sup>46–48</sup> was used, with a  $10^{-8}$  a.u. threshold.

The dimension of MRCI calculations is too large to be worth considering. Indeed, mono- and di-excitations on the large (18,14) or (20,15) active spaces would give rise to several billions of determinants. Therefore, different complementary computational strategies were introduced: (i) a reduction of the active space and two different selections of the CI space, (ii) the use of the MR-Difference Dedicated Configuration Interaction (DDCI) method<sup>27,28,30</sup> instead of MR-Single and Double CI (MR-SDCI), and (iii) the linear scaling MRCI method<sup>17–19</sup> using localized orbitals.<sup>43</sup> These three points are developed below.

### 1. Reduction of the active space

The large 18 electrons in 14 orbital active space were reduced, guided by the description of the different states, in the three different following active spaces: (i)  $\pi_{\text{cycle}}$  and  $\pi^*_{\text{cycle}}$  orbitals for the local excitations centered on the phenyl group ( $\text{LE}_\pi$ ), corresponding to 6 electrons in 6 orbitals; (ii) the nitrogen and oxygen pure-p lone pairs ( $n$ ),  $\pi_{\text{CO}}$  and  $\pi^*_{\text{CO}}$  orbitals for the local excitations on each peptide bond ( $\text{LE}_{\text{pep}}$  and  $\pi_{\text{CO}} \rightarrow \pi^*_{\text{CO}}$ ) leading to 12 electrons in 8 orbitals; (iii) for the charge transfer states (CT), the nitrogen lone pair pointing toward the phenyl group ( $n_{\text{N}(1)}$ ) and the lone pair ( $n$ ) of each oxygen atom as well as the anti-bonding  $\pi^*_{\text{cycle}}$  orbitals of the phenyl group were included in the last 6 electrons in 6 orbital active space. However, one active orbital was added to properly describe the CT states of the  $M'_{\text{CT}}$  and  $M_{n\pi^*\text{CO}(2)}$  excited states geometries. In the case of  $M'_{\text{CT}}$  geometry, the  $\sigma_{\text{NH}(1)}$  orbital of the first peptide bond was added as the amino hydrogen is pointing towards the phenyl group, leading to a CAS(8,7). For the  $M_{n\pi^*\text{CO}(2)}$  excited state geometry, the  $\pi^*_{\text{CO}}$  (pept2) orbital was added to describe the CT states as a small MR character with a non-negligible weight on the

$n\pi^*_{\text{CO}}$  (pept2) excitation was found. A CAS(6,7) was then considered.

### 2. DDCI method

This method was used to reduce the number of determinants by neglecting those involving only external orbitals, i.e., two inactive occupied and two virtual orbitals. As the most time consuming part of an SDCI calculation consists in the processing of the doubly excited determinants, the computational cost is considerably reduced by neglecting these two holes-two particles determinants. Including only the correlation energy that contributes to the energy difference in a variational CI allows us to obtain accurate vertical and adiabatic excitations, singlet-triplet gaps, and exchange magnetic coupling constants.<sup>27,28,49</sup> Beyond the reduction of the determinant basis, the DDCI method also reduces the inherent size-consistency error of the CI method. However, the Davidson correction was applied in order to get the most reliable excitation energies. The corrected excitation energies are noted CAS+DDCI+Q thereafter.

### 3. Linear scaling MRCI method using localized orbitals

Dealing with local orbitals offers the possibility to neglect long range interactions. Furthermore, it allows dividing the molecular system into regions of unequal importance. Indeed, in the most important region (where the phenomena occur), it is necessary to take into account interactions at a high level of accuracy. The interaction cutoff should then be set at a very small value, while in the rest of the system it can be taken rather large. This cutoff is obtained by means of a threshold on the exchange integral values between orbitals involved in each determinant or integral. Accurate results can then be obtained at a lower cost. Indeed, when thresholds are used on both determinants and integrals, the MRCI calculation becomes quasi-linear. The threshold used to eliminate the integrals was taken to be 0.0001 a.u. in all the study. Concerning the selection of the CI space determinants, the definition of the various zones (see Table I) was based on the nature of the localized orbitals:  $\sigma$ , lone pairs, or  $\pi$ . Zone 0 contained the targeted active space and the remaining orbitals of the CAS(18,14) or CAS(20,15) used at

TABLE I. Partition of the molecular orbitals in different zones. The corresponding thresholds applied on the exchange integrals in each zone to select the DDCI determinant space are given in a.u.

Nature of the states	Active space	Zone 0	Zone 1	Zone 2	Zone 3
$\pi_{\text{cycle}} \rightarrow \pi^*_{\text{cycle}}$	$\pi$ and $\pi^*$ cycle (6,6)			Non-valence virtual orbitals of the carbon atoms of the cycle	
$n \rightarrow \pi^*_{\text{CO}}$	$n_{\text{O}}, n_{\text{N}}, \pi_{\text{CO}}, \pi^*_{\text{CO}}$ (12,8)	Active space completed by the remaining orbitals of the CAS(18,14) or CAS(20,15) used at the CASSCF level + $\text{H}^c$	Remaining $\sigma$ orbitals	Non-valence atomic virtual orbitals of the peptide bonds	Remaining non-valence virtual orbitals
$n \rightarrow \pi^*_{\text{cycle}}$	$n_{\text{O}}, n_{\text{N}(1)}, \pi^*_{\text{cycle}}$ (6,6) + $\pi^*_{\text{CO}}$ (pept2) (6,7) <sup>a</sup> + $\sigma_{\text{NH}(1)}$ (pept1) (8,7) <sup>b</sup>			Non-valence virtual orbitals on the carbon atoms of the cycle	
Thresholds (a.u.)		0.0001	[0.1-0.006]	0.006	[0.1-0.006]

<sup>a</sup>For the  $n\pi^*_{\text{CO}}$  excited state geometry.

<sup>b</sup>For the CT excited state geometry.

<sup>c</sup>Orbitals implying the hydrogen atom of the amino group involved in the hydrogen bond (pept2) and the one which points in direction of the cycle (pept1).

the CASSCF level. All orbitals corresponding to the hydrogen atom of the amino group involved in the hydrogen bond (pept2) as well as the one pointing toward the cycle (pept1) were also included in this zone. The remaining  $\sigma$  orbitals were in zone 1. Two additional zones were defined by the non-valence virtual orbitals, those of the peptide bonds *or* those of the cycle (according to the considered states) for zone 2, and the corresponding remaining ones for zone 3. The thresholds were different among the various zones: 0.0001 a.u. was used for zone 0 while for the non-valence virtual orbitals (zone 2), a threshold of 0.006 a.u. was applied. The remaining non-valence orbitals (zone 3) as well as the  $\sigma$  zone (zone 1) were allowed to be more approximately described depending on the nature of the considered states. Several trials (0.1-0.006 a.u.) were done as presented in Sec. III, in order to get stable and accurate results on the excitation energies. When the exchange integral involved orbitals of different zones, a mean value between the two thresholds was applied. A more detailed presentation of the method can be found in previous studies.<sup>17-19</sup>

### III. RESULTS

Whatever the geometry and the basis set, there are mainly three types of excited states among the twenty first singlet excited states of NAPA B as illustrated in Fig. 2: locally ( $\pi_{\text{cycle}}, \pi_{\text{cycle}}^*$ ) excited states centered on the phenyl ring (noted hereafter  $\pi\pi^*$ ), locally ( $n, \pi_{\text{CO}}^*$ ) excited states centered on a peptide bond (noted hereafter  $n\pi_{\text{CO}}^*$ ), and intramolecular CT ( $n, \pi_{\text{cycle}}^*$ ) excited states involving an electronic charge transfer from the lone pairs of the backbone (one or two peptide bonds) to the phenyl ring (noted hereafter  $n\pi_{\text{cycle}}^*$ ).

#### A. Ground state geometry

##### 1. CC2 calculations

The relative CC2 excitation energies of the four first low-lying  $\pi\pi^*$ ,  $n\pi_{\text{CO}}^*$ , and  $n\pi_{\text{cycle}}^*$  excited states of NAPA B in its DFT-D ground state geometry [Fig. 1(a) and Table S1-1 of the [supplementary material](#)] and the contributions of the canonical occupied-unoccupied HF ( $\pi_{\text{cycle}}, n-\pi_{\text{cycle}}^*$ , and  $\pi_{\text{CO}}^*$ ) orbitals to the total wave function change are reported in Table II for a series of Dunning's basis sets. The NAPA B conformer presents a folded form exhibiting both a  $C_7$  hydrogen bond (2.02 Å) between one hydrogen of the amino group and the oxygen atom of the 1st peptide bond and a  $\text{NH}(1)-\text{C}\pi$

interaction [ $d_{\text{NH}(1)-\text{C}\pi} = 2.57$  Å; see Fig. 1(a)]. Whatever the basis set, the  $D_1/D_2$  values for the ground state were in the range 0.08/0.17 while the  $D_2$  value for the excited states was in the range 0.16-0.31 (0.19-0.27 for the first six excited states) with a biorthogonal norm  $\% \langle \bar{E}_1 | E_1 \rangle \geq 85\%$  ( $\% \langle \bar{E}_1 | E_1 \rangle \geq 87\%$  for the first six excited states). The initially recommended values for  $D_1/D_2$  for ground state minima were 0.04(0.05)/0.17(0.18) in the case of MP2(CCSD),<sup>33,34</sup> but Köhn and Hättig extended these  $D_1/D_2$  limit values up to 0.15/0.25 in particular from the evaluation of excited states of a set of small-size molecules computed with CC2. Furthermore, they proposed a new complementary diagnostics, i.e., the biorthogonal norms  $\% \langle \bar{E}_1 | E_1 \rangle$  which should be larger than 85%. More recently, for NA-ADC(2)/aug-cc-pVDZ simulations of the 9H-adenine, Barbatti *et al.* found a  $D_1$  (MP2) value in the range 0.04-0.06 for the ground state and up to values of 0.09 for the CI area while the  $D_2$  value [ADC(2)] of the first excited state was between 0.25 and 0.35, the upper region of this domain corresponding to the CI area.<sup>26</sup> Consequently, although our values remain close to the upper limit, they confirm the reliability of the CC2 calculations on these systems.

Whatever the basis set, the energy ordering as well as the main character of the first seven excited states remain unchanged: the first and fourth states correspond to  $\pi\pi^*$  states on the aromatic ring of the phenylalanine; the second, the third, and the seventh states correspond to  $n\pi_{\text{CO}}^*$  states localized on the peptide bonds of the backbone. The second state is localized on the second peptide bond whereas the third is localized on the first peptide bond. The fifth and sixth states correspond to intramolecular charge transfer  $n\pi_{\text{cycle}}^*$  states from the  $n$  lone pair orbitals of the backbone to the  $\pi^*$  aromatic ring orbitals of the phenylalanine. Moreover, from the double zeta to the quadruple zeta basis set, excitation energies do not vary very strongly; they decrease smoothly with a maximal decrease of around 0.2 eV for the second  $\pi\pi^*$  excited state, the first  $n\pi_{\text{cycle}}^*$ , and the third  $n\pi_{\text{CO}}^*$  states and a mean decrease of 0.17 eV. This leads to a small variation ( $<0.09$  eV and an absolute value of mean deviation of 0.06 eV) of their relative excitation energy with respect to the energy of first excited states. Finally, as this system presents CT states, the effect of diffuse function addition was also tested. In order to avoid redundancy problems that arise if diffuse functions are supplied on all atoms,<sup>50,51</sup> we added to the cc-pVXZ ( $X=D$  and  $T$ ) diffuse functions taken from the aug-cc-pVXZ ( $X=D$  and  $T$ ) basis set for each oxygen and nitrogen atom and for

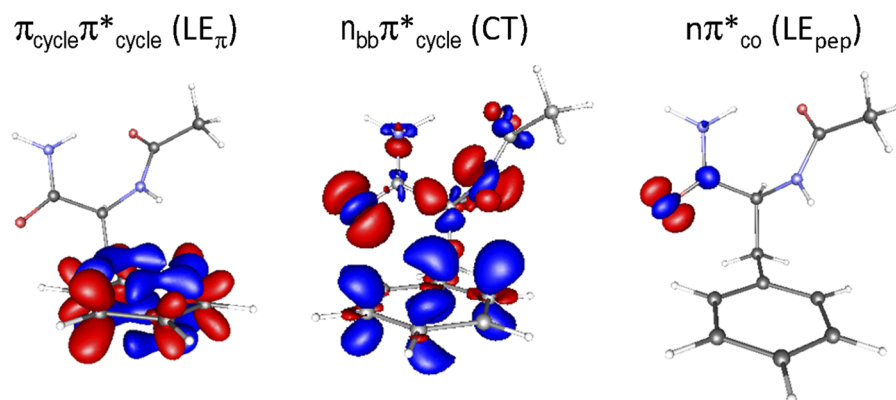


FIG. 2. Contours of the difference between the CC2 density of the different low-lying excited states ( $\pi\pi^*$  ( $\pm 0.0015$  a.u.),  $n\pi_{\text{CO}}^*$  ( $\pm 0.03$  a.u.), and  $n\pi_{\text{cycle}}^*$  ( $\pm 0.003$  a.u.)) and that of the  $S_0$  state of NAPA B calculated at the CC2/cc-pVDZ optimized geometry of the first  $\pi\pi^*$  excited state, the  $M_{\pi\pi^*}$  geometry. A density increase (decrease) is indicated in blue (red). Notation: bb for backbone.

TABLE II. CC2 excitation energy (eV) of the first four low-lying  $\pi\pi^*$ ,  $n\pi^*_{CO}$ , and  $n\pi^*_{cycle}$  excited states of the NAPA B conformer at its DFT-D optimized ground state geometry [Fig. 1(a) and Table S1-1 of the [supplementary material](#)]. The reference energy is the energy of the ground state and several basis sets have been used (cc-pVXZ, X = D, T, and Q). Values in brackets are the contributions of the canonical occupied ( $\pi_{cycle}$ , n)-unoccupied ( $\pi^*_{cycle}$ ,  $\pi^*_{CO}$ ) HF orbitals to the total wave function change. Only contributions superior or equal to 10% are indicated and only contributions superior or equal to 2% are taken into account in the summations for each type of contributions (LE $_{\pi}$  for  $\pi_{cycle} \rightarrow \pi^*_{cycle}$  excitation, LE $_{pep}$  for  $n \rightarrow \pi^*_{CO}$  excitation, CT for  $n \rightarrow \pi^*_{cycle}$  excitation, and CT $_{\pi-pep}$  for  $\pi_{cycle} \rightarrow \pi^*_{CO}$ ). The exponent numbers in brackets give the order of stability of each excited state among the twenty computed states.

$\Delta E$ (eV)	CC2 cc-pVDZ	CC2 cc-pVTZ	CC2 cc-pVQZ
$\pi_{cycle} \rightarrow \pi^*_{cycle}$	(1)5.27 [0.97] (4)6.52 [0.76] + CT [0.15] (8)7.32 [0.79] (9)7.35 [0.78]	(1)5.17 [0.97] (4)6.35 [0.83] (8)7.08 [0.75] (9)7.13 [0.73]	(1)5.14 [0.94] (4)6.30 [0.87] (9)7.01 [0.44] + LE $_{pep}$ [0.16] + CT $_{\pi-pep}$ [0.10] (10)7.02 [0.37] + LE $_{pep}$ [0.21]
$n \rightarrow \pi^*_{CO}$	(2)5.68 [0.81] <sup>a</sup> (3)5.81 [0.79] <sup>b</sup> (7)7.00 [0.36] + CT [0.34] (13)7.85 [0.34] + CT [0.28]	(2)5.55 [0.61] <sup>a</sup> (3)5.74 [0.69] <sup>b</sup> (7)6.86 [0.50] + CT [0.26] (10)7.25 [0.57]	(2)5.50 [0.66] <sup>a</sup> (3)5.73 [0.69] <sup>b</sup> (7)6.79 [0.44] + LE $_{\pi}$ [0.32] (8)6.91 [0.39] + LE $_{\pi}$ [0.20] + CT $_{\pi-pep}$ [0.18]
$n \rightarrow \pi^*_{cycle}$	(5)6.70 [0.54] <sup>c</sup> + LE $_{\pi}$ [0.19] (6)6.86 [0.74] + LE $_{\pi}$ [0.14] (11)7.71 [0.76] + LE $_{pep}$ [0.10] (12)7.80 [0.42] + LE $_{pep}$ [0.36]	(5)6.55 [0.50] <sup>c</sup> + LE $_{pep}$ [0.13] (6)6.73 [0.73] + LE $_{\pi}$ [0.13] (13)7.57 [0.49] + LE $_{pep}$ [0.28] (15)7.68 [0.64]	(5)6.50 [0.50] <sup>c</sup> + LE $_{pep}$ [0.12] (6)6.68 [0.65] + LE $_{\pi}$ [0.11] (15)7.48 [0.58] + LE $_{pep}$ [0.12] (16)7.59 [0.38] + LE $_{pep}$ [0.14] + CT $_{\pi-pep}$ [0.11]

<sup>a</sup>Lone pair localized on the second peptide bond.

<sup>b</sup>Lone pair localized on the first peptide bond.

<sup>c</sup>Lone pair localized on the backbone, i.e., on the two peptide bonds.

only one in two carbon atoms of the aromatic ring. The results reported in Table S3 of the [supplementary material](#) highlight a relatively weak effect of these diffuse functions on the excitation energies of the low-lying first excited states. Indeed, the diffuse function addition leads to a decrease of around 0.10 eV (respectively, 0.05 eV) for both  $\pi\pi^*$  and  $n\pi^*_{CO}$  states and of around 0.20 eV (respectively 0.25 eV) for the CT states using the cc-pVDZ (respectively cc-pVTZ) basis set.

In view of these results, all the following calculations are performed within the cc-pVDZ basis set.

## 2. Multireference calculations

*a. MS-CASPT2.* First, as it is reported in Table III, two CASSCF/MS-CASPT2 calculations with the ANO-L Double Zeta plus Polarization (DZP) basis sets, differing by the number of roots, were performed. Indeed, in order to obtain the

TABLE III. MS-CASPT2 excitation energy (eV) of the NAPA B conformer at its DFT-D optimized ground state geometry. The ground state energy is taken as the reference in each method and the basis set used is the ANO-L DZP. 14 and 20 roots have been considered (respectively, noted 14R and 20R). Values in brackets are the sum of the weights of the corresponding determinants in the total wave function. The thresholds considered and the definition of the excitations are those adopted for CC2 calculations and are defined in the caption of Table II, except LE $_{\pi_{CO}\pi^*_{CO}}$  for the  $\pi_{CO} \rightarrow \pi^*_{CO}$  excitation. The MR states are reported by gray shading values, the values in bold correspond to the most important weight of the corresponding determinant in the total wave function of the MR states.

$\Delta E$ (eV)	MS-CASPT2 14R/ANO-L DZP	MS-CASPT2 20R/ANO-L DZP
$\pi_{cycle} \rightarrow \pi^*_{cycle}$	5.06 [0.73]	5.14 [0.67]
$n \rightarrow \pi^*_{CO}$	5.80 [0.63] <sup>a</sup> + LE $_{\pi_{CO}\pi^*_{CO}}$ [0.13]	5.78 [0.55] <sup>a</sup> + CT [0.17]
$n \rightarrow \pi^*_{CO}$	6.00 [0.74] <sup>b</sup>	6.09 [0.67] <sup>b</sup>
$\pi_{cycle} \rightarrow \pi^*_{cycle}$	6.38 [0.87]	6.45 [0.80]
MR	6.65 eV CT [0.20] + LE $_{\pi}$ [0.19] + LE $_{\pi_{CO}\pi^*_{CO}}$ [0.19] + LE $_{pep}$ [0.17]	6.70 CT [0.14] + LE $_{\pi}$ [0.12] + LE $_{\pi_{CO}\pi^*_{CO}}$ [0.20] + LE $_{pep}$ [0.13]
MR	6.88 eV LE $_{\pi}$ [0.34] + CT [0.22] + LE $_{pep}$ [0.10]	6.85 LE $_{\pi}$ [0.35] + LE $_{\pi_{CO}\pi^*_{CO}}$ [0.20]
MR		6.99 LE $_{\pi_{CO}\pi^*_{CO}}$ [0.29] + LE $_{\pi}$ [0.19] + $\pi_{CO} \rightarrow \pi^*_{cycle}$ [0.15]
$\pi_{cycle} \rightarrow \pi^*_{cycle}$	7.03 [0.54] + CT [0.24]	7.10 [0.75]
$\pi_{cycle} \rightarrow \pi^*_{cycle}$	7.08 [0.49] + CT [0.30]	
$n \rightarrow \pi^*_{cycle}$	7.24 [0.54]	7.14 [0.59] + LE $_{\pi}$ [0.11]
$n \rightarrow \pi^*_{cycle}$		7.32 [0.47] + LE $_{pep}$ [0.20]

<sup>a</sup>Lone pair localized on the second peptide bond.

<sup>b</sup>Lone pair localized on the first peptide bond.



second CT state, 14 roots were not sufficient and a 20-root calculation had to be performed to access this latter state.

At the CASSCF level, the four lowest excited states were similar using both root sets except the first  $n\pi^*_{CO}$  localized on the second peptide bond which was strongly modified. Indeed, it was found at 6.11 eV in the calculation with 14 roots (14R) and at 5.66 eV in the 20-root calculation (20R) (Table S4-1 of the [supplementary material](#)). In the MS-CASPT2/14R calculation, only one CT state was found (7.24 eV). On the contrary, in the MS-CASPT2/20R calculation, the two CT states were present among the 20 states and they followed one another in position 9 and 10 (7.14 and 7.32 eV). This illustrated the fact that for such specific states, CT states relatively high in energy, not only the choice of the active space but also the number of roots is crucial for MS-CASPT2 calculations. On the other hand, states which had a strong multireference character (MR states) and which were lower in energy than the first CT appeared in the two calculations: two for the 14R calculation, at 6.65 and 6.88 eV, and three for the 20R calculation, at 6.70, 6.85, and 6.99 eV. Concerning the lowest excited states, the two first  $\pi\pi^*$  and the two first  $n\pi^*_{CO}$ , the ordering of stability as well as the main character remained unchanged and the excitation energies were very close ( $<0.1$  eV) whatever the number of roots. For both 14R/20R calculations, the following excited states were 1 eV higher and corresponded mainly to di-excited states.

The basis set effects were also tested in order to quantify their order of magnitude. MS-CASPT2/20R calculations using both ANO and cc-pVXZ ( $X = D$  and  $T$ ) basis sets are presented in Table IV. Indeed, these latter were used for further comparison with the CC2 results (for which ANO auxiliary basis sets are not available) while ANO basis sets, well adapted to the localization step, were used to compare to CAS+DDCI+Q calculations. The results for the ANO-L DZP and cc-pVDZ basis sets were very similar: for the seven low no-MR excited states,

the maximal deviation was 0.11 eV and the mean absolute deviation was 0.05 eV. From the cc-pVDZ to the cc-pVTZ, the nature and the energy ordering of the seven low no-MR excited states remained unchanged. The corresponding excitation energies decreased slightly, the maximal deviation being equal to 0.23 eV with a mean value of 0.14 eV. For these two basis sets, two MR states lower in energy than the first CT excited state were also found. Finally, the addition of diffuse functions was tested for the cc-pVDZ basis set (the results are presented in Table S3 of the [supplementary material](#)). As for the CC2 method (cf. Sec. III A 1), the results highlighted a relatively weak effect on the excitation energies: a 0.05 eV average variation for the  $\pi\pi^*$  and the  $n\pi^*_{CO}$  states and a 0.12 eV for the CT states were observed.

Finally, the MS-CASPT2 calculations are then carried out in the following considering 20 roots within the ANO-L DZP basis set.

*b. CAS+DDCI.* First, the different thresholds implied in the CAS+DDCI calculations were adapted to these types of systems. The molecule was divided into four zones and different thresholds were applied according to the considered zone. The threshold used was always 0.0001 a.u. for zone 0. Similarly, for the non-valence virtual orbitals of the considered states (zone 2), the threshold was set to 0.006 a.u. These thresholds were low enough to give reliable results. Concerning zones 1 and 3, several thresholds were tested. The goal was to determine the highest thresholds giving accurate results for the lowest computational cost. The results are presented in Table V.

**Zone 1.** When decreasing the threshold used in  $\sigma$  zone from 0.1 to 0.01 a.u., the improvement was quite negligible on the  $n \rightarrow \pi^*_{CO}$  states (0.04 eV) whereas it was more important for the  $\pi_{cycle} \rightarrow \pi^*_{cycle}$  (at the most 0.26 eV) and huge for the  $n \rightarrow \pi^*_{cycle}$  ones (0.74 eV). Finally, the threshold of 0.01 a.u.

TABLE IV. MS-CASPT2/ANO-L DZP and cc-pVXZ ( $X = D$  and  $T$ ) excitation energy (eV) of NAPA B conformer at its DFT-D optimized ground state geometry. The reference energy is that of the ground state in each method and 20 roots have been considered in all calculations. For the definition of the values in bracket, see captions of Tables II and III. The MR states are reported by gray shading values, the values in bold correspond to the most important weight of the corresponding determinant in the total wave function of the MR states.

$\Delta E$ (eV)	MS-CASPT2 ANO-L DZP	MS-CASPT2 cc-pVDZ	MS-CASPT2 cc-pVTZ
$\pi_{cycle} \rightarrow \pi^*_{cycle}$	5.14 [0.67]	5.17 [0.73]	5.12 [0.74]
$n \rightarrow \pi^*_{CO}$	5.78 [0.55] + CT [0.17]	5.85 [0.51] + $LE\pi_{CO}\pi^*_{CO}$ [0.20]	5.77 [0.57] + $LE\pi_{CO}\pi^*_{CO}$ [0.18]
$n \rightarrow \pi^*_{CO}$	6.09 [0.67]	6.01 [0.51] + $LE\pi_{CO}\pi^*_{CO}$ [0.19]	5.98 [0.52] + $LE\pi_{CO}\pi^*_{CO}$ [0.19]
$\pi_{cycle} \rightarrow \pi^*_{cycle}$	6.45 [0.80]	6.54 [0.86]	6.36 [0.87]
MR	6.70 <b><math>LE\pi_{CO}\pi^*_{CO}</math> [0.20]</b> + CT [0.14] + $LE_{pep}$ [0.13] + $LE\pi$ [0.12]	6.80 <b>CT [0.26]</b> + $LE_{pep}$ [0.21] + $LE\pi$ [0.14]	6.62 <b><math>LE_{pep}</math> [0.27]</b> + $LE\pi_{CO}\pi^*_{CO}$ [0.17] + $LE\pi$ [0.15] + CT [0.10]
MR	6.85 <b><math>LE\pi</math> [0.35]</b> + $LE\pi_{CO}\pi^*_{CO}$ [0.20]	7.05 <b><math>LE\pi</math> [0.26]</b> + CT [0.21]	6.82 <b><math>LE\pi</math> [0.34]</b> + $LE\pi_{CO}\pi^*_{CO}$ [0.17] + CT [0.17]
MR	6.99 <b><math>LE\pi_{CO}\pi^*_{CO}</math> [0.29]</b> + $LE\pi$ [0.19] + $\pi_{CO} \rightarrow \pi^*_{cycle}$ [0.15]		
$\pi_{cycle} \rightarrow \pi^*_{cycle}$	7.10 [0.75]	7.11 [0.63] + CT [0.10]	6.88 [0.75]
$n \rightarrow \pi^*_{cycle}$	7.14 [0.59] + $LE\pi$ [0.11]	7.25 [0.40] + $LE\pi$ [0.39]	7.08 [0.42] + $LE\pi$ [0.30]
$n \rightarrow \pi^*_{cycle}$	7.32 [0.47] + $LE_{pep}$ [0.20]	7.35 [0.45] + $LE\pi$ [0.18]	7.14 [0.52] + $LE\pi_{CO}\pi^*_{CO}$ [0.12]
MR			7.28 <b><math>LE\pi_{CO}\pi^*_{CO}</math> [0.32]</b> + CT [0.28] + $LE_{pep}$ [0.10]
$\pi_{CO} \rightarrow \pi^*_{CO}$		7.51 [0.44] + $LE_{pep}$ [0.14]	

TABLE V. Effect of the thresholds used in the selected DDCI/ANO-L DZP on the excitation energy (in eV) of the different states and corresponding number of determinants (in millions). Starting orbitals are those of the CASSCF(18,14)/14R calculation. Active spaces are defined in Table I.

Thresholds (a.u.)							
Zone 1: remaining $\sigma$ orbitals	0.1	0.01	0.006	0.1	0.1	0.01	0.006
Zone 3: remaining non valence virtual orbitals	0.1	0.1	0.1	0.01	0.006	0.01	0.006
$\pi_{\text{cycle}} \rightarrow \pi^*_{\text{cycle}}$	5.07	5.03	5.03	5.05	5.05	5.04	5.03
	7.02	6.89	6.88	6.97	6.96	6.84	6.82
	8.20	7.97	7.96	8.13	8.12	7.87	7.85
	8.25	7.99	7.98	8.19	8.18	7.91	7.87
	8.29	8.19	8.19	8.22	8.22	8.18	8.17
No. of determinants (millions)	21	71	76	44	56	127	161
$n \rightarrow \pi^*_{\text{CO}}$	6.12	6.08	6.06	6.05	6.04	6.01	5.99
	6.30	6.26	6.24	6.22	6.20	6.15	6.12
No. of determinants (millions)	65	214	234	192	226	584	739
$n \rightarrow \pi^*_{\text{cycle}}$	7.82	7.08	7.05	7.78	7.76	7.14	7.10
	8.39	7.72	7.71	8.42	8.40	7.84	7.82
No. of determinants (millions)	26	92	97	70	85	205	255

gave accurate results for all the excited states since the lowest threshold value of 0.006 a.u. did not change the excitation energies by more than 0.03 eV.

**Zone 3.** As expected, the threshold of 0.1 a.u. for zone 3, i.e., the remaining virtual non-valence orbitals which do not involve the atoms of the local excitations, gave reasonable results ( $\leq 0.1$  eV) compared to the threshold of 0.01 a.u. for the three types of states. Concerning the excitation energy of the charge transfer states, these orbitals contributed in an equal amount to the differential energies and the excitation energies were stable whatever the threshold used. Decreasing the threshold from 0.01 to 0.006 a.u. improved the results only by 0.01 eV. To conclude, the threshold of 0.01 a.u. for zone 3 gave accurate excitation energies for all the states.

An additional trial was performed to validate these thresholds by decreasing the two thresholds to 0.006 a.u. However, this time consuming calculation gave very similar results to the one with thresholds of 0.01 a.u. Indeed, the largest difference was only of 0.04 eV. Consequently, in the following, the thresholds were set to 0.01 for zones 1 and 3.

Once the quality of the results has been validated, one can also take stock of the performance of the quasi-linear MRCI program. The number of determinants of the *non-selected* MR-DDCI calculations varied from 2 to  $9 \times 10^9$ , depending on the active space size. Using thresholds of 0.0001 a.u. for zone 0, 0.006 a.u. for zone 2, and 0.01 a.u. for zones 1 and 3 allowed a significant reduction of the CI size. Indeed, the largest CI space ( $n \rightarrow \pi^*_{\text{CO}}$  states) contained  $584 \times 10^6$  of determinants, i.e., 6% of the initial one. The corresponding computational time was 12 h per state and per iteration, and the whole calculation took 2 weeks for three roots on a bi-(4c) Intel Xeon E5-2637 v3 machine, using one processor and 30 GB of memory. For comparison, the 0.006 a.u. thresholds calculation took approximately twice as long.

Once the thresholds have been adapted, the effects of the number of roots (14R vs 20R) as well as those of the Davidson correction (noted +Q) were tested (Table VI). The effect of the number of roots on most of the excitation energies was small (at most 0.03/0.05 eV for no Q/+Q calculations). As the additional roots implied excitations from and to orbitals of the

TABLE VI. CAS+DDCI excitation energy (eV) of the NAPA B conformer at its DFT-D optimized ground state geometry. The reference energy is that of the ground state in each method and the basis set used is the ANO-L DZP. 14 and 20 roots have been considered and a Davidson correction (+Q) is applied or not.

$\Delta E$ (eV)	CAS+DDCI	CAS+DDCI+Q	CAS+DDCI	CAS+DDCI+Q
	ANO-L DZP	ANO-L DZP	ANO-L DZP	ANO-L DZP
	14R	14R	20R	20R
$\pi_{\text{cycle}} \rightarrow \pi^*_{\text{cycle}}$	5.04	5.09	5.04	5.09
$n \rightarrow \pi^*_{\text{CO}}$	6.01 <sup>a</sup>	5.89 <sup>a</sup>	5.81 <sup>a</sup>	5.74 <sup>a</sup>
$n \rightarrow \pi^*_{\text{CO}}$	6.15 <sup>b</sup>	6.10 <sup>b</sup>	6.18 <sup>b</sup>	6.13 <sup>b</sup>
$\pi_{\text{cycle}} \rightarrow \pi^*_{\text{cycle}}$	6.84	6.48	6.85	6.48
$n \rightarrow \pi^*_{\text{cycle}}$	7.14	7.13	7.11	7.08
$n \rightarrow \pi^*_{\text{cycle}}$	7.84	7.43	7.81	7.39
$\pi_{\text{cycle}} \rightarrow \pi^*_{\text{cycle}}$	7.87	7.64	7.88	7.64
$\pi_{\text{cycle}} \rightarrow \pi^*_{\text{cycle}}$	7.91	7.67	7.91	7.67

<sup>a</sup>Lone pair localized on the second peptide bond.

<sup>b</sup>Lone pair localized on the first peptide bond.

same nature, the average orbitals were not changed significantly. The only exception concerned the first  $n\pi^*_{CO}$  localized on the second peptide bond (deviation of  $-0.20/-0.15$  eV for no Q/+Q calculations) which was directly related to the change introduced by the CASSCF calculation (Table S4 of the [supplementary material](#)). On the other hand, the Davidson correction affected almost all excitation energies with a maximum effect, a decrease of around 0.40 eV, for the second  $\pi\pi^*$  and CT states. Not only the CAS+DDCI+Q results were in very good agreement with MS-CASPT2 ones, the nature and the energy ordering of the two first  $\pi\pi^*$ ,  $n\pi^*_{CO}$ , and  $n\pi^*_{cycle}$  excited states were the same but also the maximal deviation was weak, i.e., 0.10/0.07 eV for 14R/20R with a mean absolute value of 0.09/0.05 eV. Originally developed to calculate singlet-triplet gaps in biradicalar systems,<sup>52</sup> DDCI allows us to evaluate accurate vertical energy differences,<sup>28</sup> from a common set of MOs for all states. The CAS+DDCI+Q method led here to discrepancies within the expected error margins of the MS-CASPT2 method, i.e.,  $\pm 0.2$  eV for the excitation energies of the two first  $\pi\pi^*$ ,  $n\pi^*_{CO}$ , and  $n\pi^*_{cycle}$  excited states. On the other hand, the MR MS-CASPT2 states lower in energy than the first CT states could not be reproduced by the CAS+DDCI+Q method because of the restricted active space which does not allow the mixing between states of different nature.

Eventually, the CAS+DDCI calculations are carried out in the following considering 20 roots within the ANO-L DZP basis set and taking into account the Davidson correction (CAS+DDCI+Q).

### 3. CC2 versus MR method

Comparing to the MR method results (MS-CASPT2/20R and CAS+DDCI+Q starting orbitals coming from a 20R CASSCF calculation), it can be observed that the CC2/cc-pVDZ level was in good agreement for the four low-lying excited states while the discrepancy became larger for the states around and beyond 7 eV. The results are presented in Tables II, IV, and VI for comparison. Indeed, the first CT state  $n\pi^*_{cycle}$  was found to be 0.44/0.38 eV lower with the CC2 method compared to the MS-CASPT2/CAS+DDCI+Q whereas the two first low-lying  $\pi\pi^*$  states were found to be

0.14-0.07/0.18-0.04 eV higher and the two first  $n\pi^*_{CO}$  were found to be 0.10-0.28/0.06-0.32 eV lower. The difference between the CC2 excitation energies and the MRCI ones was then not larger than 0.3 eV for the lowest  $\pi\pi^*$  and  $n\pi^*_{CO}$  states and was of the same order of magnitude than the expected error of the MS-CASPT2 and CAS+DDCI+Q methods. On the contrary, a larger value was obtained for the excitation energies around and beyond 7 eV (around 0.40–0.60 eV for the CT states). As for CAS+DDCI+Q calculations, there were obviously no MR states in CC2 calculations due to the formalism of the method.

In view of these results, all calculations in the following are performed with these characteristics: MS-CASPT2/ANO-L DZP (20R), CAS+DDCI+Q/ANO-L DZP (20R), and CC2/cc-pVDZ.

### B. Energetic profile of the deactivation mechanisms of the conformer NAPA B involving the second peptide bond

The relative CC2, MS-CASPT2, and CAS+DDCI+Q excitation energies of the two first  $\pi\pi^*$ ,  $n\pi^*_{CO}$ , and  $n\pi^*_{cycle}$  low-lying excited states of NAPA B are reported in Tables VII–IX for the three pertinent geometries of the energy profile of the MII deactivation mechanism of the NAPA B conformer,<sup>10,11</sup> i.e., (i)  $M_{\pi\pi^*}$ , the CC2/cc-pVDZ equilibrium geometry of the  $\pi\pi^*$  [Fig. 1(b<sub>1</sub>) and Table S1-2 of the [supplementary material](#)], (ii)  $CI_{\pi\pi^*/n\pi^*_{CO}(2)}$ , the CC2/cc-pVDZ CI geometry connecting the equilibrium geometries of the two  $\pi\pi^*$  and  $n\pi^*_{CO}$  excited states [Fig. 1(b<sub>2</sub>) and Table S1-3 of the [supplementary material](#)], and (iii)  $M_{n\pi^*_{CO}(2)}$ , the CC2/cc-pVDZ equilibrium geometry of the  $n\pi^*_{CO}$  excited state localized on the second peptide bond [Fig. 1(b<sub>3</sub>) and Table S1-4 of the [supplementary material](#)].

Whatever the geometry, the  $D_1/D_2$  values obtained in the CC2 calculation for the ground state were in the range 0.08-0.13/0.20-0.27 while the  $D_2$  value for the excited states was in the range 0.17-0.33 with a biorthogonal norm  $\% \langle \overline{E}_1 | E_1 \rangle \geq 85\%$  ( $\% \langle \overline{E}_1 | E_1 \rangle \geq 87\%$  for the first six excited states). As previously mentioned, these values confirm the reliability of the CC2 calculations even if they remain close to the upper limit of acceptable values (see Sec. III A 1). As it has been already

TABLE VII. CC2, MS-CASPT2, and CAS+DDCI+Q excitation energy (eV) of the two first excited states of each nature of the NAPA B conformer. The geometry used is the  $M_{\pi\pi^*}$  CC2/cc-pVDZ optimized geometry [see Fig. 1(b<sub>1</sub>) and Table S1-2 of the [supplementary material](#)]. The energy values in bold correspond to the lowest excited state in each method. The exponent numbers in brackets give the order of stability of each excited state among the twenty ones. For the definition of the values in bracket, see captions of Tables II and III.

$\Delta E$ (eV)	CC2	MS-CASPT2	CAS+DDCI+Q
$\pi_{cycle} \rightarrow \pi^*_{cycle}$	(1) <b>4.82</b> [0.97]	(1) <b>4.68</b> [0.70]	(1) <b>4.65</b>
	(4)6.18 [0.86]	(4)6.06 [0.79]	(4)6.22
$n \rightarrow \pi^*_{CO}$	(2)5.63 [0.67] <sup>a</sup>	(2)5.75 [0.64] <sup>a</sup>	(2)5.79 <sup>a</sup>
	(3)5.78 [0.87] <sup>b</sup>	(3)6.02 [0.66] <sup>b</sup>	(3)6.06 <sup>b</sup>
$n \rightarrow \pi^*_{cycle}$	(5)6.42 [0.67] <sup>c</sup> + $LE_{\pi}$ [0.26]	(5)6.94 [0.43] <sup>a</sup> + $LE_{pep}$ [0.13]	(5)6.76 <sup>a</sup>
	(6)6.61 [0.51] <sup>c</sup> + $LE_{\pi}$ [0.12]	(6)7.01 [0.66] <sup>a</sup>	(6)7.18 <sup>a</sup>

<sup>a</sup>Lone pair n localized on the second peptide bond.

<sup>b</sup>Lone pair n localized on the first peptide bond.

<sup>c</sup>Lone pair n localized on the backbone, i.e., on the two peptide bonds.

TABLE VIII. CC2, MS-CASPT2, and CAS+DDCI+Q excitation energy (eV) of the two first excited states of each nature of the NAPA B conformer. The geometry used is the  $CI_{\pi\pi^*/n\pi^*CO}$  CC2/cc-pVDZ geometry [see Fig. 1(b<sub>2</sub>) and Table S1-3 of the [supplementary material](#)]. The energy values in bold correspond to the lowest excited state in each method. The exponent numbers in brackets give the order of stability of each excited state among the twenty ones. For the definition of the values in bracket, see captions of Tables II and III.

$\Delta E$ (eV)	CC2	MS-CASPT2	CAS+DDCI+Q
$\pi_{\text{cycle}} \rightarrow \pi^*_{\text{cycle}}$	(2) <b>4.93</b> [0.52] + $LE_{\text{pep}}$ [0.20] (4)6.25 [0.82]	(1) <b>4.86</b> [0.59] (4)6.23 [0.54] + $LE_{\text{pep}}$ [0.19]	(1) <b>4.74</b> (4)6.27
$n \rightarrow \pi^*_{CO}$	(1) <b>4.91</b> [0.41] <sup>a</sup> + $LE_{\pi}$ [0.36] (3)5.77 [0.82] <sup>b</sup>	(2) <b>5.05</b> [0.63] <sup>a</sup> + CT [0.14] (3)6.21 [0.49] <sup>b</sup> + $LE_{\pi}$ [0.20]	(2) <b>5.00</b> <sup>a</sup> (3)6.12 <sup>b</sup>
$n \rightarrow \pi^*_{\text{cycle}}$	(5)6.47 [0.59] <sup>c</sup> + $LE_{\pi}$ [0.20] (6)6.62 [0.47] <sup>c</sup> + $LE_{\pi}$ [0.17] + $LE_{\text{pep}}$ [0.13]	(6)7.08 [0.63] <sup>a</sup> (7)7.20 [0.60] <sup>a</sup>	(5)6.77 <sup>a</sup> (6)7.17 <sup>a</sup>
$\pi_{CO} \rightarrow \pi^*_{CO}$		(5)6.36 [0.42] + $LE_{\text{pep}}$ [0.21]	

<sup>a</sup>Lone pair n localized on the second peptide bond.

<sup>b</sup>Lone pair n localized on the first peptide bond.

<sup>c</sup>Lone pair n localized on the backbone, i.e., on the two peptide bonds.

observed in such systems,<sup>22</sup> a  $\pi\pi^*$  excitation causes a significant change in the electron density distribution around the phenyl ring leading to a significant shortening of the NH(1)– $C_{\pi}$  distance in  $M_{\pi\pi^*}$  (0.22 Å) whereas parameters such as the covalent bonds, the valence angles, or torsional angles that do not involve the orientation of the backbone relative to the phenyl ring are only weakly changed upon  $\pi\pi^*$  excitation (Table S5-1 of the [supplementary material](#)). Indeed, upon excitation, the  $\pi$  system tends to extend farther from the ring  $C_6$  axis, with a density increase above the C atoms and on the ring edge and a decrease above the CC bonds as illustrated in Fig. 2. In  $M_{n\pi^*CO}(2)$ , the local excitation from the nitrogen and oxygen lone pairs of the backbone to the CO antibonding  $\pi^*$  orbital of the second peptide bond (see the density transfer illustrated in Fig. 2) leads to a characteristic deplanarization of the peptide bond with a pyramidalization of both the carbon and nitrogen atoms: the  $\omega_2$  angle is strongly reduced ( $\sim 80^\circ$ ) and both  $\Sigma\theta_{Ni}$

and  $\Sigma\theta_{Ci}$  differ from  $360^\circ$  by at least  $20^\circ$  (see Tables S5-1 and S5-2 and Fig. S5-3 of the [supplementary material](#)). This is accompanied by an elongation of 0.15 Å of the CO bond as well as an elongation of both the NH(2)–O the NH(1)– $C_{\pi}$  distances (+0.26 Å and +0.09 Å). Along the energy profile, the geometry of  $CI_{\pi\pi^*/n\pi^*CO(2)}$  is intermediate between the two previous geometries but much closer to the  $M_{\pi\pi^*}$  than to the  $M_{n\pi^*CO}(2)$  geometry (see Fig. 3): all the characteristic parameters, angles and distances, are modified by up to 30% of their total variation along the energy profile (Tables S5-1 and S5-2 and Fig. S5-3 of the [supplementary material](#)).

For  $M_{\pi\pi^*}$  and  $M_{n\pi^*CO}(2)$  equilibrium geometries, both the nature and the energetic of the first excited state obtained at the CC2 level are in good agreement with those obtained at the MR level, the energy difference being less than 0.2 eV (see Tables VII and IX). In the case of the  $CI_{\pi\pi^*/n\pi^*CO(2)}$  geometry, the geometry of the conical intersection identified with

TABLE IX. CC2, MS-CASPT2, and CAS+DDCI+Q excitation energy (eV) of the two first excited states of each nature of the NAPA B conformer. The geometry used is the  $M_{n\pi^*CO}(2)$  CC2/cc-pVDZ optimized geometry [see Fig. 1(b<sub>3</sub>) and Table S1-4 of the [supplementary material](#)]. The energy values in bold correspond to the lowest excited state in each method. The exponent numbers in brackets give the order of stability of each excited state among the twenty ones: (i) in the CC2 calculation, the fourth, fifth, and sixth states correspond to  $n\pi^*_{CO}$  at 5.23, 5.58, and 5.72 eV, respectively, and (ii) in the MS-CASPT2 calculation, the fourth and sixth state correspond to  $n\pi^*_{CO}$  at 5.84 and 6.03 eV. For the definition of the values in bracket, see captions of Tables II and III.  $CT_{\text{pep1-pep2}}$  correspond to the  $\pi_{CO}$  (peptide bond 2)  $\rightarrow \pi^*_{CO}$  (peptide bond 1). The orbitals have been optimized at the CASSCF(20,15) level and averaged on 20 roots; these orbitals are used in the MS-CASPT2 and the DDCI methods.

$\Delta E$ (eV)	CC2	MS-CASPT2	CAS+DDCI+Q
$\pi_{\text{cycle}} \rightarrow \pi^*_{\text{cycle}}$	(3)5.10 [0.63] + $CT_{\pi\text{-pep}}$ [0.31] (11)6.61 [0.86]	(3)5.00 [0.46] (7)6.16 [0.47]	(2)4.98 (4)6.30
$n \rightarrow \pi^*_{CO}$	(1) <b>2.14</b> [0.89] <sup>a</sup> (2)4.32 [0.67] <sup>a</sup> + $CT_{\text{pep2-pep1}}$ [0.19]	(1) <b>2.14</b> [0.74] <sup>a</sup> (2)4.79 [0.38] <sup>a</sup> + $\pi_{CO} \rightarrow \pi^*_{CO}$ [0.16] + $CT_{\pi\text{pep}}$ [0.14]	(1) <b>2.11</b> <sup>a</sup> (3)5.05 <sup>a</sup> + $CT_{\text{pep2-pep1}}$
$n \rightarrow \pi^*_{\text{cycle}}$	(8)5.96 [0.76] <sup>a</sup> (10)6.20 [0.85] <sup>a</sup>		(5)6.98 <sup>a,b</sup> (6)7.09 <sup>a,b</sup>
$\pi_{\text{cycle}} \rightarrow \pi^*_{CO}$	(7)5.89 [0.52] + $LE_{\text{pep}}$ [0.14] + $LE_{\pi}$ [0.14] (9)6.08 [0.82]		<sup>c</sup>
$D^+ \pi_{\text{cycle}} n \rightarrow \pi^*_{\text{cycle}} \pi^*_{CO}$		(5)5.92	

<sup>a</sup>Lone pair n localized on the second peptide bond.

<sup>b</sup>With CAS (6,7).

<sup>c</sup>These states do not appear by construction; the excitations leading to these states are not considered in the CAS+DDCI+Q calculation.

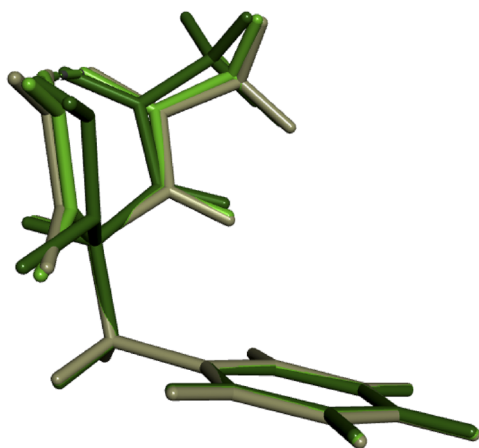


FIG. 3. Comparison between the CC2/cc-pVDZ geometries of  $M_{\pi\pi^*}$  (in gray),  $M_{n\pi^*CO(2)}$  (in dark green), and  $CI_{\pi\pi^*/n\pi^*CO(2)}$  (in light green). To illustrate the change from the second peptide bond and environment point of view, all the carbon atoms of phenyl rings as well as the carbon atom of the  $CH_2$  group and of the  $CH(CO-NH_2, NH-COCH_3)$  group are superimposed.

the CC2 method, the first  $\pi\pi^*$ , and  $n\pi^*_{CO}$  states are obviously quasi-degenerate at the CC2 level ( $\Delta E$  of  $-0.02$  eV) whereas at the MR level, the energy gap between these two states is more important (MS-CASPT2/CAS+DDCI+Q :  $+0.19/+0.26$  eV), the more stable one being the  $\pi\pi^*$  excited state. There is really no reason that the CC2 conical intersection geometry matches exactly the MR one. Moreover, the determination of this geometry at the CC2 level results from an interpolation between the geometries of the two minima and corresponds then to an estimation of the localization of the CI at the CC2 level.<sup>10</sup> On the other hand, considering the energy difference between the two excited states for this geometry at the MR level which corresponds to a weak lift of degeneracy ( $\sim 0.2$  eV), it is clear that the MR conical intersection geometry must be close to the CC2 one.

The order of stability as well as the nature of the five other states of  $M_{\pi\pi^*}$  given by the CC2 method (Table VII)

are equivalent to that given by the MR methods except for the  $n\pi^*_{cycle}$  states in which the lone pairs are localized on the backbone in the CC2 calculations and only on the second peptide bond in the MR methods. The difference in excitation energies between CC2 and MR methods are similar to that found for the DFT-D ground state geometry: the  $\pi\pi^*$  excited states are found to be around 0.1 eV higher, the  $n\pi^*_{CO}$  excited states are found to be around 0.2 eV lower, and the  $n\pi^*_{cycle}$  excited states are found to be around 0.45 eV lower. For the  $CI_{\pi\pi^*/n\pi^*CO(2)}$  geometry (Table VIII), the order of stability of the following excited states is equivalent for the three methods except that in the MS-CASPT2 calculations, a  $\pi_{CO} \rightarrow \pi^*_{CO}$  state is inserted in the fifth position. As for the  $M_{\pi\pi^*}$  geometry, the CC2 and MR nature of these states is equivalent except again for the lone pairs in  $n\pi^*_{cycle}$  states localized on the backbone in the CC2 calculations and only on the second peptide bond with the MR methods. Finally, the difference in excitation energies between CC2 and MR methods is similar to that found for the DFT-D ground state geometry. For  $M_{n\pi^*CO(2)}$  (Table IX), as no charge transfer state was found at the CASSCF/20R level, a 30-root CASSCF calculation was performed but, once more, no CT state was encountered. At the CAS+DDCI+Q level, the two CT states were obtained starting from 20R MOs, not optimized for these states, since these latter were not obtained at the CASSCF level. This can explain the quite large discrepancy (about 1 eV) between CC2 and CAS+DDCI+Q excitation energies of these states. The excitation energy of the other states does not present a so large discrepancy but this difference remains more important than those obtained for the other discussed geometries.

### C. Equilibrium geometry of the CT excited state of the NAPA B conformer identified in the mechanisms involving the first peptide bond

The relative CC2, MS-CASPT2, and CAS+DDCI+Q excitation energies of the two first low-lying  $\pi\pi^*$ ,  $n\pi^*_{CO}$ , and  $n\pi^*_{cycle}$  excited states of NAPA B for the partially optimized

TABLE X. CC2, MS-CASPT2, and CAS+DDCI+Q excitation energy of the two first excited states of each nature of the conformer NAPA B. The geometry is the CC2/cc-pVDZ partially optimized geometry with the NH(1) bond length constrained to 1.216 Å of the CT excited state [see Fig. 1(c) and Table S1-5 of the [supplementary material](#)]. The energy values in bold correspond to the lowest excited state in each method. The exponent numbers in brackets give the order of stability of each excited state among the twenty ones: (i) in the CC2 calculation, the fifth, sixth, and eighth states correspond to  $n\pi^*_{cycle}$  at 5.05, 5.36, and 5.63 eV, respectively, and (ii) in the MS-CASPT2 calculation, the fifth and sixth state correspond to  $n\pi^*_{CO}$  at 5.52 and 6.63 eV. For the definition of the values in bracket, see captions of Tables II and III. The orbitals have been optimized at the CASSCF(20,15) level and averaged on 20 roots; these orbitals are used in the CASPT2 and the DDCI methods.

$\Delta E$ (eV)	CC2	MS-CASPT2	CAS+DDCI+Q
$\pi_{cycle} \rightarrow \pi^*_{cycle}$	( <sup>2</sup> )4.28 [0.89] ( <sup>3</sup> )4.46 [0.72] + CT [0.22]	( <sup>2</sup> )4.07 [0.65] ( <sup>3</sup> )4.38 [0.66]	( <sup>2</sup> )4.11 ( <sup>3</sup> )4.86
$n \rightarrow \pi^*_{cycle}$	( <sup>1</sup> ) <b>3.23</b> [0.90] <sup>a</sup> ( <sup>4</sup> )4.98 [0.83] <sup>a</sup> + $LE_{\pi}$ [0.10]	( <sup>1</sup> ) <b>3.68</b> [0.69] <sup>b</sup> ( <sup>4</sup> )5.47 [0.58] <sup>b</sup>	( <sup>1</sup> ) <b>3.24</b> <sup>b,c</sup> ( <sup>4</sup> )5.80 <sup>b,c</sup>
$n \rightarrow \pi^*_{CO}$	( <sup>7</sup> )5.55 [0.59] <sup>a</sup> + CT [0.13] ( <sup>9</sup> )5.69 [0.62] <sup>a</sup> + $CT_{\pi-pep}$ [0.13]	( <sup>7</sup> )5.90 [0.38] <sup>a</sup> + $LE_{\pi}$ [0.16] ( <sup>8</sup> )5.94 [0.52] <sup>a</sup>	( <sup>5</sup> )5.85 <sup>b</sup> ( <sup>6</sup> )6.01 <sup>d</sup>

<sup>a</sup>Lone pair localized on the backbone, i.e., on the two peptide bonds.

<sup>b</sup>Lone pair localized on the first peptide bond.

<sup>c</sup>With CAS(8,7).

<sup>d</sup>Lone pair localized on the second peptide bond.

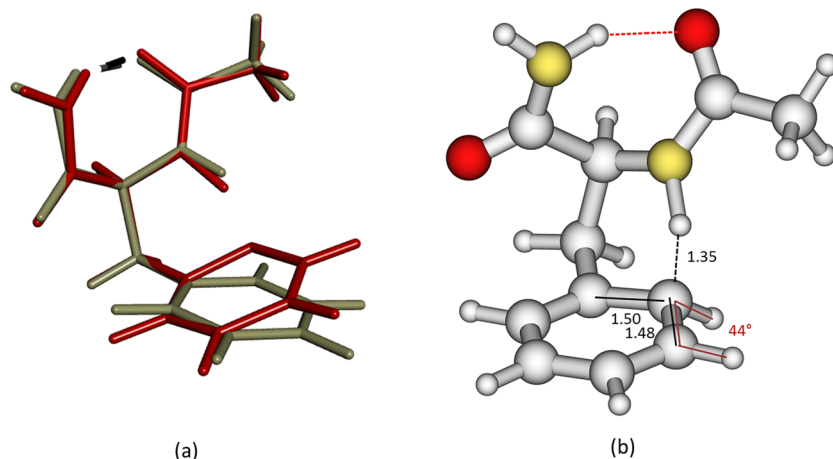


FIG. 4. (a) Comparison between the CC2/cc-pVDZ geometries of  $M_{\pi\pi^*}$  (in gray) and  $M'_{CT}$  (in red). To illustrate the change from the phenyl ring and environment point of view, backbones are superimposed. (b) For the sake of clarity, the main geometrical parameters around the ortho carbon atom of the phenyl ring are presented. Distances are given in Ångströms and the dihedral angle in degrees.

geometry, with the NH(1) bond length constrained to 1.216 Å, of the first CT excited state ( $M'_{CT}$ ) identified in the MI deactivation mechanism<sup>10</sup> of the NAPA B conformer [Fig. 1(c) and Table S1-5 of the [supplementary material](#)] are presented (Table X).

As previously discussed, the  $D_1/D_2$  values obtained in the CC2 calculation for the ground state, 0.11/0.23, as well as the  $D_2$  value obtained for the excited states which is in the range 0.19-0.31 with a biorthogonal norm  $\% \langle E_1 | E_1 \rangle \geq 85\%$ , confirm the reliability of the CC2 calculations even if they remain close to the upper limit of acceptable values (see Sec. III A 1). For  $M'_{CT}$  in which the NH(1) bond is elongated by around 0.2 Å compared to that of  $M_{\pi\pi^*}$ , the excitation corresponds to an electronic charge transfer from the  $\pi$  (carbonyl) orbitals and the  $n$  lone pairs of the backbone to the  $\pi^*$  orbitals of the phenyl ring as it is illustrated in Fig. 2. This leads to a strong shortening of 1.0 Å of the NH(1)– $C_\pi$  distance concomitant with a shortening of 0.08 Å of the NH(2)–O distance (see Table S5-1 of the [supplementary material](#)) as well as a strong distortion of the phenyl ring [Figs. 4(a) and 4(b)] resulting from the disruption of the  $\pi$  conjugation pattern of the phenyl ring in the neighborhood of the *ortho* carbon atom [an elongation of the two adjacent CC bonds to  $\sim 1.50$  Å, Fig. 4(b)].

In this geometry, the first excited state is a charge transfer state for both the CC2 and MR methods. If for the CC2 method, the involved lone pairs of the backbone are those of both peptide bonds, only the lone pairs of the first peptide bond are involved for the MR methods. Concerning the excitation energy, the CC2 and CAS+DDCI+Q values are very close ( $\Delta E = 0.01$  eV) whereas the MS-CASPT2 value is 0.45 eV higher. Concerning the following states, both the order of stability and the nature of the first three states are equivalent for the three methods except for the  $n\pi^*_{\text{cycle}}$  states in which the lone pairs are localized on the backbone in the CC2 calculation whereas only the lone pairs of the first peptide bond are involved in the MR calculations.

The treatment of the dynamic electronic correlation leads to a different correction in the two MR methods. In the DDCI method, the two holes-two particles determinants are not taken into account compared to the MR-SDCI method. The analysis of the CASPT2 results shows that these excitations contribute for the half to the dynamical correlation energy,

but their contribution is almost the same for all the ground and excited states and leads to a maximal differential value of 0.02 eV between the ground state and the excited states. However, we cannot directly conclude that the difference between DDCI and MS-CASPT2 methods does not come from the two holes-two particles determinants. The active space used in the DDCI method is indeed reduced compared to CASSCF/MS-CASPT2, even if the weights of the reference wave functions in the CI calculation are large enough ( $> 0.7$ ) to insure a reliable result. Finally, the mean absolute difference between the excitation energies obtained with the two MR methods (0.24 eV) is more than twice the value obtained in the other geometries (0.09 eV). This suggests that the difficulty encountered almost systematically at the CASSCF/MS-CASPT2 level to properly describe the CT states could be the reason of these discrepancies at the equilibrium geometry of the lowest charge transfer state.

#### IV. CONCLUSION

First of all, this work allows us to establish the validity of the quasi-linear CAS+DDCI+Q method to treat singlet excited states of bio-relevant systems. Indeed, once their parameters are adapted, this method leads—on a building block of proteins, the NAPA system—to a very good agreement with the results obtained with a more standard MR method, the MS-CASPT2. For a series of different geometries, both the excitation energies and the wave functions of the low-lying singlet excited states obtained at the CAS+DDCI+Q/20R/ANO-L-DZP level are equivalent to those obtained at the MS-CASPT2/20R/ANO-L-DZP independent of the nature of the excited state. The only non-negligible discrepancies which appear ( $\Delta E > 0.2$  eV) concern the excitation energies of CT excited states which either are high in energies (see  $n\pi^*_{\text{cycle}}$  states in Tables VIII and IX) or which correspond to a strongly distorted geometry (see Table X). Moreover, these calculations highlight various points to be taken into account in MR methods to properly deal with low-lying excited states of such systems. In the MS-CASPT2 calculations, for the low-lying singlet CT states which can be relatively high in energy, not only the choice of the active space is crucial but also the number of roots. For the CAS+DDCI+Q method, an enlargement

of the active space is required for the excitation energy of some low-lying excited states.

Second, the lowest singlet excited state of different CC2/cc-pVDZ geometries can be well described at the CC2/cc-pVDZ level whatever its nature, i.e., a  $\pi\pi^*$  excited state on the phenyl ring, an  $n\pi^*$  excited state localized on the peptide bond(s), or even an  $n\pi^*$  charge transfer (CT) excited state involving an electronic charge transfer from the lone pairs of the backbone to the phenyl ring. Compared to CC2, the only difference concerns this latter CT state which is localized on a particular peptide bond at the CAS+DDCI+Q level. Furthermore, the CC2/cc-pVDZ excitation energy of the lowest singlet excited state is close to that obtained at the CAS+DDCI+Q/20R/ANO-L-DZP level and the absolute deviation is inferior to 0.2 eV, a value close to the standard error of the MS-CASPT2 method ( $\pm 0.2$  eV). For the following low-lying excited states, the nature of these states is also well described by the CC2 method except, as previously, for the electronic charge transfer in the CT states, backbone vs localized lone pair. For the excitation energies, two different behaviors are observed. For the DFT-D ground state, the  $M_{\pi\pi^*}$  and the  $CI_{\pi\pi^*/n\pi^*CO}$  geometries, the difference between excitation energies is not greater than 0.3 eV for the lowest states—the  $\pi\pi^*$  and the  $n\pi^*$  excited states—and around 0.4 eV for the higher states—the CT states. For the two other geometries, the  $M_{n\pi^*CO}$  (2) and the  $M'_{CT}$ , the difference in excitation energies is more pronounced, around 0.4 eV on average with an absolute maximum deviation of 0.8 eV.

Finally, in view of the good performances of the CC2 method on the NAPA system, a building block of proteins containing one residue, it will be very interesting to evaluate if these performances can be extrapolated to larger systems, i.e., systems containing more than one residue. However, even if this work demonstrates that the CAS+DDCI+Q method is appropriate to accurately describe the close-low-lying singlet excited states of a building block of proteins, the active space size which can be affordable in a CASSCF calculation is reached for the NAPA system. Therefore, it will be necessary to consider other solutions to optimize the orbitals, the first step of the CAS+DDCI+Q calculations. A promising solution that we are investigating on capped peptides containing at least two residues is given by the Generalized Active Space Self Consistent Field (GASSCF) method,<sup>53,54</sup> a method which allows us to expand active spaces beyond the CASSCF limit. This work will be reported in a future paper.

## SUPPLEMENTARY MATERIAL

See [supplementary material](#) for Appendix S1—Cartesian coordinates of all the geometries of the NAPA B conformer, Appendix S2—dependence of the excitation energy and reference weight of the zeroth-order wave function on the level shift in CASPT2 calculations, Appendix S3—diffuse orbital effects on the excitation energy in CC2 and MS-CASPT2 calculations, Appendix S4—first CASSCF excited states for the ground state geometry as a function of the number of roots, and Appendix S5—characteristic structural parameters of the different NAPA B geometries.

## ACKNOWLEDGMENTS

This work received financial support from the Agence Nationale de la Recherche (ANR), Grant No. ANR-14-CE06-0019-01-ESBODYR. This work was granted access to the HPC facility of [TGCC/CINES/IDRIS] under Grant Nos. 2015-t2015087412&t2015087372, 2016-t2016087540, and 2017-A0010807540, awarded by GENCI (Grand Equipement National de Calcul Intensif); to the CCRT High Performance Computing (HPC) facility at CEA under Grant No. CCRT2015/CCRT2016-p606bren; and to the HPC resources of CALMIP supercomputing center under Allocation No. 2016-[P16009].

- <sup>1</sup>W. Domcke and A. L. Sobolewski, *Phys. Chem. Chem. Phys.* **12**, 4897 (2010).
- <sup>2</sup>Y. Chen and M. D. Barkley, *Biochemistry* **37**, 9976 (1998).
- <sup>3</sup>P. R. Callis and T. Liu, *J. Phys. Chem. B* **108**, 4248 (2004).
- <sup>4</sup>M. A. Robb, M. Olivucci, and F. Bernardi, in *Encyclopedia of Computational Chemistry*, edited by P. von Ragué Schleyer, N. L. Allinger, T. Clark, J. Gasteiger, P. A. Kollman, H. F. Schaefer, and P. R. Schreiner (John Wiley & Sons, Ltd., Chichester, UK, 2002).
- <sup>5</sup>G. A. Worth and L. S. Cederbaum, *Annu. Rev. Phys. Chem.* **55**, 127 (2004).
- <sup>6</sup>B. G. Levine and T. J. Martínez, *Annu. Rev. Phys. Chem.* **58**, 613 (2007).
- <sup>7</sup>*Conical Intersections: Electronic Structure, Dynamics & Spectroscopy*, edited by W. Domcke, D. Yarkony, and H. Köppel (World Scientific, River Edge, NJ, 2004).
- <sup>8</sup>W. Domcke and D. R. Yarkony, *Annu. Rev. Phys. Chem.* **63**, 325 (2012).
- <sup>9</sup>D. R. Yarkony, *Chem. Rev.* **112**, 481 (2012).
- <sup>10</sup>M. Mališ, Y. Loquais, E. Gloaguen, H. S. Biswal, F. Piuze, B. Tardivel, V. Brenner, M. Broquier, C. Jouvét, M. Mons, N. Došlić, and I. Ljubić, *J. Am. Chem. Soc.* **134**, 20340 (2012).
- <sup>11</sup>M. Mališ, Y. Loquais, E. Gloaguen, C. Jouvét, V. Brenner, M. Mons, I. Ljubić, and N. Došlić, *Phys. Chem. Chem. Phys.* **16**, 2285 (2014).
- <sup>12</sup>C. Hättig and F. Weigend, *J. Chem. Phys.* **113**, 5154 (2000).
- <sup>13</sup>C. Hättig and A. Köhn, *J. Chem. Phys.* **117**, 6939 (2002).
- <sup>14</sup>C. Hättig, *J. Chem. Phys.* **118**, 7751 (2003).
- <sup>15</sup>A. Köhn and C. Hättig, *J. Chem. Phys.* **119**, 5021 (2003).
- <sup>16</sup>O. Christiansen, H. Koch, and P. Jørgensen, *Chem. Phys. Lett.* **243**, 409 (1995).
- <sup>17</sup>B. Bories, D. Maynau, and M.-L. Bonnet, *J. Comput. Chem.* **28**, 632 (2007).
- <sup>18</sup>N. Ben Amor, F. Bessac, S. Hoyau, and D. Maynau, *J. Chem. Phys.* **135**, 014101 (2011).
- <sup>19</sup>C. Chang, C. J. Calzado, N. Ben Amor, J. Sanchez Marin, and D. Maynau, *J. Chem. Phys.* **137**, 104102 (2012).
- <sup>20</sup>D. Shemesh, A. L. Sobolewski, and W. Domcke, *J. Am. Chem. Soc.* **131**, 1374 (2009).
- <sup>21</sup>N. Došlić, G. Kovačević, and I. Ljubić, *J. Phys. Chem. A* **111**, 8650 (2007).
- <sup>22</sup>W. Y. Sohn, V. Brenner, E. Gloaguen, and M. Mons, *Phys. Chem. Chem. Phys.* **18**, 29969 (2016).
- <sup>23</sup>N. O. C. Winter, N. K. Graf, S. Leutwyler, and C. Hättig, *Phys. Chem. Chem. Phys.* **15**, 6623 (2013).
- <sup>24</sup>C. Fang, B. Oruganti, and B. Durbeej, *J. Phys. Chem. A* **118**, 4157 (2014).
- <sup>25</sup>D. Tuna, D. Lefrançois, Ł. Wolański, S. Gozem, I. Schapiro, T. Andrzejowski, A. Dreuw, and M. Olivucci, *J. Chem. Theory Comput.* **11**, 5758 (2015).
- <sup>26</sup>F. Plasser, R. Crespo-Otero, M. Pederzoli, J. Pittner, H. Lischka, and M. Barbatti, *J. Chem. Theory Comput.* **10**, 1395 (2014).
- <sup>27</sup>J. Miralles, J.-P. Daudey, and R. Caballol, *Chem. Phys. Lett.* **198**, 555 (1992).
- <sup>28</sup>J. Miralles, O. Castell, R. Caballol, and J.-P. Malrieu, *Chem. Phys.* **172**, 33 (1993).
- <sup>29</sup>T. Kraus, N. Ben Amor, and V. Robert, *Phys. Chem. Chem. Phys.* **16**, 9509 (2014).
- <sup>30</sup>J. Zapata-Rivera, R. Caballol, and C. J. Calzado, *J. Comput. Chem.* **32**, 1144 (2011).
- <sup>31</sup>C. J. Calzado, N. Ben Amor, and D. Maynau, *Chem. - Eur. J.* **20**, 8979 (2014).
- <sup>32</sup>C. J. Calzado and D. Maynau, *J. Chem. Phys.* **135**, 194704 (2011).
- <sup>33</sup>J. Finley, P.-Å. Malmqvist, B. O. Roos, and L. Serrano-Andrés, *Chem. Phys. Lett.* **288**, 299 (1998).

- <sup>34</sup>W. Chin, M. Mons, J.-P. Dognon, R. Mirasol, G. Chass, I. Dimicoli, F. Piuze, P. Butz, B. Tardivel, I. Compagnon, G. von Helden, and G. Meijer, *J. Phys. Chem. A* **109**, 5281 (2005).
- <sup>35</sup>TURBOMOLE a Development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, 2012.
- <sup>36</sup>F. Furche, R. Ahlrichs, C. Hättig, W. Klopper, M. Sierka, and F. Weigend, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **4**, 91 (2014).
- <sup>37</sup>T. H. Dunning, *J. Chem. Phys.* **90**, 1007 (1989).
- <sup>38</sup>F. Weigend, A. Köhn, and C. Hättig, *J. Chem. Phys.* **116**, 3175 (2002).
- <sup>39</sup>I. M. B. Nielsen and C. L. Janssen, *Chem. Phys. Lett.* **310**, 568 (1999).
- <sup>40</sup>C. L. Janssen and I. M. B. Nielsen, *Chem. Phys. Lett.* **290**, 423 (1998).
- <sup>41</sup>B. O. Roos, P. R. Taylor, and P. E. M. Siegbahn, *Chem. Phys.* **48**, 157 (1980).
- <sup>42</sup>F. Aquilante, L. De Vico, N. Ferré, G. Ghigo, P. Å. Malmqvist, P. Neogrády, T. B. Pedersen, M. Pitoňák, M. Reiher, B. O. Roos, L. Serrano-Andrés, M. Urban, V. Veryazov, and R. Lindh, *J. Comput. Chem.* **31**, 224 (2010).
- <sup>43</sup>D. Maynau, S. Evangelisti, N. Guihéry, C. J. Calzado, and J.-P. Malrieu, *J. Chem. Phys.* **116**, 10060 (2002).
- <sup>44</sup>N. Ben Amor, B. Bories, S. Hoyau, and D. Maynau, DoLo and EXSCI codes are available on [https://github.com/LCPQ/Cost\\_package](https://github.com/LCPQ/Cost_package).
- <sup>45</sup>P.-O. Widmark, P.-Å. Malmqvist, and B. O. Roos, *Theor. Chim. Acta* **77**, 291 (1990).
- <sup>46</sup>F. Aquilante, T. B. Pedersen, and R. Lindh, *J. Chem. Phys.* **126**, 194106 (2007).
- <sup>47</sup>F. Aquilante, P.-Å. Malmqvist, T. B. Pedersen, A. Ghosh, and B. O. Roos, *J. Chem. Theory Comput.* **4**, 694 (2008).
- <sup>48</sup>H. Koch, A. Sánchez de Merás, and T. B. Pedersen, *J. Chem. Phys.* **118**, 9481 (2003).
- <sup>49</sup>E. Rodríguez and M. Reguero, *J. Phys. Chem. A* **106**, 504 (2002).
- <sup>50</sup>E. Papajak, J. Zheng, X. Xu, H. R. Leverentz, and D. G. Truhlar, *J. Chem. Theory Comput.* **7**, 3027 (2011).
- <sup>51</sup>G. Granucci, J. T. Hynes, P. Millié, and T.-H. Tran-Thi, *J. Am. Chem. Soc.* **122**, 12243 (2000).
- <sup>52</sup>P. De Loth, P. Cassoux, J. P. Daudey, and J. P. Malrieu, *J. Am. Chem. Soc.* **103**, 4007 (1981).
- <sup>53</sup>D. Ma, G. Li Manni, and L. Gagliardi, *J. Chem. Phys.* **135**, 044128 (2011).
- <sup>54</sup>K. D. Vogiatzis, G. Li Manni, S. J. Stoneburner, D. Ma, and L. Gagliardi, *J. Chem. Theory Comput.* **11**, 3010 (2015).