



**HAL**  
open science

## Model protein excited states

Valérie Brenner, Thibaut Véry, Michael Schmidt, Mark Gordon, Sophie Hoyau, Nadia Ben Amor

► **To cite this version:**

Valérie Brenner, Thibaut Véry, Michael Schmidt, Mark Gordon, Sophie Hoyau, et al.. Model protein excited states: MRCI calculations with large active spaces vs CC2 method. *The Journal of Chemical Physics*, 2021, 154 (21), pp.214105. 10.1063/5.0048146 . hal-03307376

**HAL Id: hal-03307376**

**<https://hal.science/hal-03307376>**

Submitted on 25 Aug 2022

**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.

# Model protein excited states: MRCI calculations with large active spaces vs CC2 method

Cite as: J. Chem. Phys. **154**, 214105 (2021); <https://doi.org/10.1063/5.0048146>

Submitted: 19 February 2021 . Accepted: 02 May 2021 . Published Online: 02 June 2021

 Valérie Brenner, Thibaut Véry, Michael W. Schmidt,  Mark S. Gordon,  Sophie Hoyau, and  Nadia Ben Amor

## COLLECTIONS

Paper published as part of the special topic on [Special Collection in Honor of Women in Chemical Physics and Physical Chemistry](#)



View Online



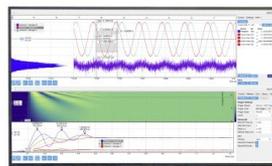
Export Citation



CrossMark

Challenge us.

What are your needs for  
periodic signal detection?



Zurich  
Instruments

# Model protein excited states: MRCI calculations with large active spaces vs CC2 method

Cite as: *J. Chem. Phys.* **154**, 214105 (2021); doi: [10.1063/5.0048146](https://doi.org/10.1063/5.0048146)

Submitted: 19 February 2021 • Accepted: 2 May 2021 •

Published Online: 2 June 2021



View Online



Export Citation



CrossMark

Valérie Brenner,<sup>1</sup>  Thibaut Véry,<sup>1,a)</sup>  Michael W. Schmidt,<sup>2</sup>  Mark S. Gordon,<sup>2</sup>  Sophie Hoyau,<sup>3</sup>   
and Nadia Ben Amor<sup>3,4,b)</sup> 

## AFFILIATIONS

<sup>1</sup>LIDYL, CEA, CNRS, Université Paris-Saclay, 91191 Gif-sur-Yvette, France

<sup>2</sup>Department of Chemistry, Iowa State University, Ames, Iowa 5001, USA

<sup>3</sup>Université de Toulouse, UPS, LCPQ (Laboratoire de Chimie et Physique Quantiques), IRSAMC, 118, rte de Narbonne, F-31062 Toulouse Cedex, France

<sup>4</sup>CNRS, UPS, LCPQ (Laboratoire de Chimie et Physique Quantiques), IRSAMC, 118, rte de Narbonne, F-31062 Toulouse Cedex, France

**Note:** This paper is part of the JCP Special Collection in Honor of Women in Chemical Physics and Physical Chemistry.

**a) Present address:** IDRIS–CNRS, Orsay Cedex F-91403, France.

**b) Author to whom correspondence should be addressed:** [nadia.benamor@irsamc.ups-tlse.fr](mailto:nadia.benamor@irsamc.ups-tlse.fr)

## ABSTRACT

Benchmarking calculations on excited states of models of phenylalanine protein chains are presented to assess the ability of alternative methods to the standard and most commonly used multiconfigurational wave function-based method, the complete active space self-consistent field (CASSCF), in recovering the non-dynamical correlation for systems that become not affordable by the CASSCF. The exploration of larger active spaces beyond the CASSCF limit is benchmarked through three strategies based on the reduction in the number of determinants: the restricted active space self-consistent field, the generalized active space self-consistent field (GASSCF), and the occupation-restricted multiple active space (ORMAS) schemes. The remaining dynamic correlation effects are then added by the complete active space second-order perturbation theory and by the multireference difference dedicated configuration interaction methods. In parallel, the approximate second-order coupled cluster (CC2), already proven to be successful for small building blocks of model proteins in one of our previous works [Ben Amor *et al.*, *J. Chem. Phys.* **148**, 184105 (2018)], is investigated to assess its performances for larger systems. Among the different alternative strategies to CASSCF, our results highlight the greatest efficiency of the GASSCF and ORMAS schemes in the systematic reduction of the configuration interaction expansion without loss of accuracy in both nature and excitation energies of both singlet  $\pi\pi^*$  and  $n\pi^*$  excited states with respect to the equivalent CASSCF calculations. Guidelines for an optimum applicability of this scheme to systems requiring active spaces beyond the complete active space limit are then proposed. Finally, the extension of the CC2 method to such large systems without loss of accuracy is demonstrated, highlighting the great potential of this method to treat accurately excited states, mainly single reference, of very large systems.

Published under license by AIP Publishing. <https://doi.org/10.1063/5.0048146>

## I. INTRODUCTION

Quantum chemistry now provides a large panel of tools to tackle the excited state calculation of molecular systems.<sup>1–4</sup> However, still today, highly accurate methods, because of their high computational demands (consuming time as well as memory and disk resources), remain limited to small systems and even very small systems if their dynamics is addressed. One of the challenges

is then to benchmark approximate methods against highly accurate ones, the objective being to define computational protocols using more efficient but less reliable methods. Once such benchmarks have been carried out and once such a protocol has been defined and validated, excited state potential energy surfaces can be investigated, and then, their photoinduced chemical dynamics can be explored. In this spirit, we recently developed a protocol<sup>5–8</sup> combining three levels of theory to study both the excited state

spectroscopy and dynamics of models of phenylalanine protein chains. First, the time-dependent density functional theory (TDDFT)<sup>9</sup> is used in non-adiabatic dynamics simulations in order to qualitatively investigate the deactivation mechanisms, and then, two higher levels of theory, the standard approximate coupled cluster singles and doubles method (CC2)<sup>10–14</sup> and a multireference configuration interaction (MRCI) method,<sup>15–17</sup> are used in order to address them quantitatively.

One key point in the development of this computational protocol was to demonstrate the validity of the CC2 method for models of phenylalanine protein chains by comparison with MRCI calculations, the method that allows one to access the properties of excited states of such systems, the first and the second derivatives of the energy being affordable. The comparison between CC2 and MRCI calculations was already done in our previous work on a building block of proteins, a capped peptide containing one residue.<sup>5</sup> We now evaluate if these reliable performances of the CC2 method can be extrapolated to larger systems such as capped peptides containing more residues and/or containing more extended side chains. However, the active space size that can be affordable in a complete active space self-consistent field (CASSCF)<sup>18</sup> calculation is reached with the capped peptide containing one residue, and alternatives to the CASSCF must, therefore, be found to tackle larger systems.

Two theoretical challenges arise for the multireference (MR) methods when expanding the size of the systems: the ability to obtain accurate zeroth-order reference wave functions (WFs) with the optimization of the orbitals and configuration interaction (CI) coefficients, on one hand, in order to obtain the major part of the non-dynamical or static correlation and, on the other hand, the introduction of the dynamical electronic correlation. Multiconfiguration self-consistent field (MCSCF)<sup>19–21</sup> or CASSCF are usually the methods used in the first step where, at least, all the orbitals for which the occupation numbers vary significantly are defined as active. However, these methods are currently limited to 18 electrons in 18 orbitals as the number of determinants becomes too large in the configuration interaction part. To overcome this limit, a solution is to restrict the number of determinants, for example, by partitioning the active space into groups of orbitals as has been proposed in different methods such as the configuration interaction spaces with restrictions on the orbital occupancies,<sup>22,23</sup> the macroconfiguration approach,<sup>24</sup> the Restricted Active Space (RAS) SCF (RASSCF)<sup>25,26</sup> method, the Generalized Active Space (GAS) SCF (GASSCF)<sup>27–30</sup> method, or the Occupation-Restricted Multiple Active Space (ORMAS)<sup>31,32</sup> method. The RASSCF scheme splits the active space into three subspaces, while the restriction acts on the degree of excitation of the determinants. In the GASSCF and in the ORMAS methods, the number of subspaces is not limited. The difference between them consists essentially in how the electron excitation between subspaces is managed. Starting from these zeroth-order wave functions, the dynamical correlation can be taken into account by multireference second-order Perturbation Theory (PT): multireference Møller–Plesset (MRMP) perturbation theory,<sup>33,34</sup> multiconfigurational quasi-degenerate perturbation theory (MCQDPT),<sup>35,36</sup> Complete Active Space with Perturbation at the Second Order (CASPT2),<sup>37</sup> or second-order N-electron valence state perturbation theory (NEVPT2)<sup>38,39</sup> methods on the CASSCF wave functions, RASPT2<sup>40</sup> on the RASSCF ones, and ORMAS-PT<sup>41</sup> on the

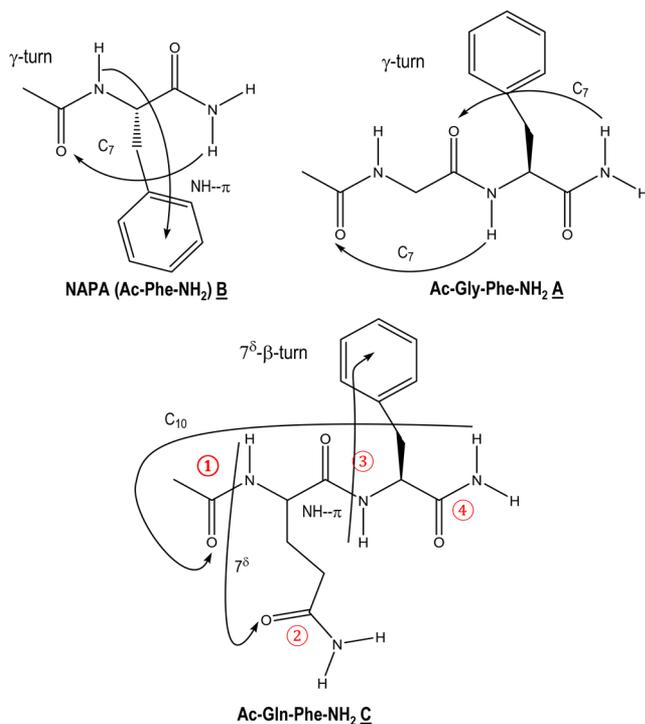
ORMAS ones. The GASPT2<sup>42</sup> method has also been developed but is not yet available in the standard version of MOLCAS.<sup>43–46</sup> Multireference Configuration Interaction (MRCI) can also be used to introduce dynamical correlation on top of these zeroth-order wave functions. However, the computational cost of these methods is much greater than that of perturbative ones, and without further approximations, the MRCI methods are restricted to rather small systems. The reduction in the CI-matrix size can be obtained by selecting only the most important determinants or configurations. The most evident way is to restrict the excitation degree, for example, from the full CI to the single and double CI or Difference Dedicated CI (DDCI),<sup>47,48</sup> where all two hole–two particle excitations external to the active space are excluded, which is suitable for the calculation of vertical excitation energies. As dynamic electron correlation is a local phenomenon, long-range interactions can also be neglected, and linear-scaling CASPT2,<sup>49</sup> NEVPT2,<sup>50</sup> or MRCI<sup>15–17,51–54</sup> methods have been developed.

We report here first an evaluation of the performances of different alternatives to CASSCF on a series of capped peptides of increasing size, the NAPA *B* conformer of the *N*-acetylphenylalaninylamide and conformers of larger systems, the capped Ac–Gly–Phe–NH<sub>2</sub> *A* and Ac–Gln–Phe–NH<sub>2</sub> *C* conformers, two dipeptides with different side chains. The model proteins, the CC2 method as well as the multireference approaches used, and, in particular, the principle and parameters of three different alternatives to CASSCF among the most recent ones are described in the Sec. II. In the Sec. III, the CC2 and multireference results are presented and discussed on the series of systems, the criteria of the selection of the most efficient alternative to CASSCF being detailed. Finally, the validity of the CC2 methods for larger systems is demonstrated in Sec. IV by comparison with the multireference calculations.

## II. METHODS AND COMPUTATIONAL DETAILS

### A. Model proteins

The model proteins first consisted of a conformer of our reference system, a capped peptide with one residue, the *N*-acetylphenylalaninylamide (NAPA *B*), and second consisted of conformers of two larger systems, the *A* conformer of the capped Ac–Gly–Phe–NH<sub>2</sub> dipeptide, which contained one glycine (Gly) and one phenylalanine (Phe), and the *C* conformer of the capped Ac–Gln–Phe–NH<sub>2</sub> dipeptide, which contained one phenylalanine (Phe) and one glutamine (Gln) residue, a residue which also bears an amide group in the side chain. Low-lying excited states, i.e., the lowest  $\pi\pi^*$  excited state localized on the phenyl ring and the lowest  $n\pi^*$  excited states localized on the peptide bonds (one state per amide group) were investigated at the CC2/cc-pVDZ optimized geometry of the lowest  $\pi\pi^*$  excited state.<sup>6</sup> All these excited state conformers adopted prototypical secondary structural features of proteins, these global structures being preserved from their ground state (Fig. 1 and Ref. 6 for details). The NAPA *B* conformer corresponded to a  $\gamma$ -turn folded conformation stabilized by a C<sub>7</sub> H-bond and an NH $\cdots$  $\pi$  bond. The *A* conformer of Ac–Gly–Phe–NH<sub>2</sub> corresponded to the 2<sub>7</sub> ribbon extended conformation stabilized by two successive C<sub>7</sub> H-bonds (double  $\gamma$ -turn). The *C* conformer of Ac–Gln–Phe–NH<sub>2</sub> corresponded to a type I  $\beta$ -turn backbone, stabilized by a C<sub>10</sub> H-bond combined to a side chain/main chain C<sub>7</sub> H-bond bridging the NH site of the first peptide bond to the oxygen



**FIG. 1.** NAPA (Ac-Phe-NH<sub>2</sub>) **B**, Ac-Gly-Phe-NH<sub>2</sub> **A**, and Ac-Gly-Phe-NH<sub>2</sub> **C** conformers. The numbering of the peptide bonds starts from the CH<sub>3</sub>CO group and continues along the backbone, taking into account the amide groups of the side chains.

atom of the Gln residue side chain CO-NH<sub>2</sub> group labeled 7<sup>δ</sup>. In addition, it contained a NH ··· π bond, which implied the hydrogen atom of the second peptide bond.

## B. CC2 method

CC2<sup>10–14</sup> calculations were carried out with the TURBOMOLE package.<sup>55,56</sup> All the CC2 calculations were performed by using the resolution-of-identity (RI)<sup>57</sup> approximation for the electron repulsion integrals in the correlation treatment and the description of the excitation processes. The Dunning cc-pVDZ correlation consistent basis sets<sup>58</sup> were employed in connection with optimized auxiliary basis sets<sup>57</sup> for the RI approximation. Frozen core for the 1s electrons was employed, and all calculations were carried out in the C<sub>1</sub> point-group symmetry. Ten singlet states were considered, and D<sub>1</sub>, D<sub>2</sub> diagnostics and %  $\langle \bar{E}_1 | E_1 \rangle$  biorthogonal norm were calculated in order to evaluate the capability of the CC2 method to properly describe the ground and excited states of such systems.<sup>5,6,14,59,60</sup> The convergence criterion used in single point energy calculations was 10<sup>-8</sup> on the density for the HF calculation, 10<sup>-9</sup> for the RI-CC2 ground state energy in the iterative coupled-cluster methods, and 10<sup>-6</sup> for the convergence threshold for the norm of residual vectors in eigenvalue problems for the RI-CC2 excited state calculations. In the geometry optimization, the convergence criterion used corresponds to a norm of the Cartesian gradient lower than 10<sup>-4</sup> a.u. The harmonic frequencies were calculated by numerical differentiation

of the analytic gradients using central differences and a step length of 0.02 a.u. This also allowed verifying that the optimized geometries correspond to true minima.

Orbital-relaxed first-order properties were determined; in particular, the density and then CC2 differences between the density of excited states and that of the ground state were performed. In addition, a post-processing tool interfaced to TURBOMOLE, Nancy\_EX-2.0,<sup>61</sup> was used in order to analyze the density and character of the excited states and obtain, at the CC2 level, the so-called natural transition orbitals (NTOs)<sup>62,63</sup> of each excited state. Instead of describing one excitation with multiple canonical spin orbital couples, all the physical information on the nature of the electronic transition was gathered in one (sometimes two) couple(s) (from one occupied to one virtual) of NTOs allowing an unambiguous characterization of the nature of excited states. Moreover, the contributions of the NTOs to the wave function were more directly and accurately comparable to the MRCI weights of the determinants in the total wave function than the contributions of the canonical occupied–unoccupied HF orbitals.

## C. Multireference approaches

Multireference (MR) wave function (WF) approaches were used to investigate the electronic transitions in a two step procedure, the first one including the non-dynamical correlation led to zeroth-order reference wave functions and was followed by the introduction of dynamical correlation. Taking into account non-dynamical correlation was done—as a reference calculation—for the NAPA **B** and Ac-Gly-Phe-NH<sub>2</sub> **A** complexes by means of the CASSCF<sup>18</sup> method using the MOLCAS package. In order to reduce the bottleneck of the active space, three different methods were also used: RASSCF,<sup>25,26</sup> GASSCF,<sup>27–29,64</sup> and the quasi-equivalent ORMAS.<sup>31</sup> The RASSCF and GASSCF methods are implemented in the MOLCAS package, while ORMAS is part of the GAMESS (General Atomic and Molecular Electronic Structure System) package.<sup>65–67</sup> As the orbitals were assigned to different subspaces, localized orbitals were necessary. The SCF orbitals were localized with the DoLo code<sup>68,69</sup> as the starting point of the MOLCAS calculations, or using Pipek–Mezey localization<sup>70</sup> in the GAMESS package, for the ORMAS ones. All the calculations were based on the Dunning correlation consistent basis sets cc-pVDZ.<sup>58</sup> For NAPA **B**, some CASSCF/CASPT2, Q+DDCI, and CC2<sup>5</sup> calculations were also performed with the enlarged cc-pVTZ basis sets, but as the effects were weak with a very important computational cost (see [supplementary material](#), Appendix S4), only the cc-pVDZ basis set has been kept for the other systems. The Cholesky decomposition technique<sup>71,72</sup> has been used (threshold of 10<sup>-8</sup> a.u.) in the MOLCAS calculations.

To provide a good description of all considered states, it was necessary to include the involved orbitals into the active space. In the studied complexes, the low-lying states corresponded to local excitations: (i)  $\pi \rightarrow \pi^*$  centered on the phenyl group or (ii)  $n \rightarrow \pi^*$  centered on peptide bonds [where *n* is N or O lone pairs (pure-p lone pair)]. To accurately describe these states, the active space should include all the  $\pi$  and  $\pi^*$  orbitals of the phenyl and of the carbonyl groups as well as lone pairs of the nitrogen and oxygen atoms (the highest in energy, the pure-p lone pair). This active

space corresponded to 18 electrons in 14 orbitals [CAS(18,14)] for the NAPA *B* complex. For the larger systems (Ac-Gly-Phe-NH<sub>2</sub> *A* and Ac-Gln-Phe-NH<sub>2</sub> *C*), with one or two more peptide bonds, the active space contained 24 electrons in 18 orbitals or 30 electrons in 22 orbitals, respectively. The former led to a cumbersome but still tractable active space, while the latter one was unfeasible and required the use of RASSCF, GASSCF, and ORMAS alternative methods.

In the RASSCF, the active space was separated into three subspaces: (i) RAS1 concerned occupied orbitals in which a limited number of holes was allowed, (ii) RAS2 corresponded to the complete active space, and (iii) RAS3 was the subspace containing the virtual orbitals, and the number of particles that could be created in this subspace had to be defined. Although this method is not commonly used, compared to the CASSCF one, in the same way, the orbitals that have notable varying occupation numbers have to be included in the RAS2. The notation used to define the RASSCF calculation was RAS (number of active electrons, number of holes in RAS1, number of particles in RAS3; number of orbitals in RAS1, in RAS2, and in RAS3), the number of active electrons referring to the total number of electrons in the three subspaces. In the present case, the partition of the RAS subspaces can be done in a state specific way: the aromatic cycle orbitals in the RAS2 (or the orbitals of each peptide bond) plus the complementary orbitals in RAS1 and RAS3 to reach the complete set of active orbitals. The obtained results are presented and discussed in Appendix S5 of the [supplementary material](#). Indeed, this led to different sets of orbitals, while the philosophy of the DDCI method is based on a common set of state averaged orbitals. This is why we chose in the following a common set of subspaces to describe all the states at the RASSCF/RASPT2 levels. For this, the RASSCF code was used as a Configuration Interaction (CI) one, coupled with the optimization of the orbitals. Indeed, the RAS2 subspace was empty; all the occupied active orbitals in the main determinant were placed in the RAS1 subspace and the virtual active ones in the RAS3 subspace, and the number of holes and particles defined the level of the CI (SD, SDT, SDTQ, etc.). For NAPA *B*, this led to a RAS(18,n,m; 9,0,5), where *n* was the number of holes, identical to the number *m* of particles, and took the values 3, 4, or 5. This partition is illustrated in Fig. 2. For the larger systems, we used for Ac-Gly-Phe-NH<sub>2</sub> *A* a RAS(24,n,n; 12,0,6), *n* = 3, 4, 5, and for Ac-Gln-Phe-NH<sub>2</sub> *C* a RAS(30,n,n; 15,0,7), *n* = 3. The goal was to determine, for the NAPA *B* and Ac-Gly-Phe-NH<sub>2</sub> *A* systems, how many holes and particles are necessary to reach the CASSCF results.

In the ORMAS/GASSCF methods, the total active space was divided into several subspaces. One subspace was dedicated to the  $\pi$

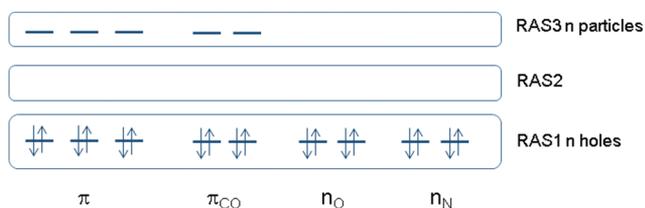


FIG. 2. NAPA *B*. Definition of the three active subspaces for the RASSCF.

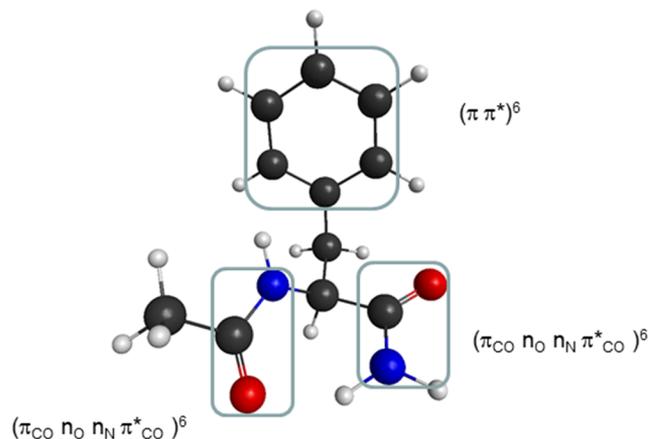


FIG. 3. Definition of the multiple active spaces for the ORMAS and GASSCF for NAPA *B*.

orbitals of the cycle (phenyl group), while the other ones concerned each peptide bond (Fig. 3). In each of these subspaces, all determinants were generated. Electron excitation between subspaces could be allowed by defining the minimum and maximum number of electrons in each subspace for ORMAS, while it was a cumulative number of electrons that was used in the GASSCF. In some cases, the two methods are not equivalent<sup>29</sup> even though in the present study they are. Indeed, as the targeted states were supposed to be local states,  $\pi\pi^*$  centered on the phenyl group or  $n\pi^*_{CO}$  on each peptide bond, electron excitation between subspaces was not allowed. However, this partition was not equivalent to multiple CASSCF calculations on each active space. In the GASSCF or ORMAS, all the states could be calculated in the same time and state averaged orbitals are obtained, necessary for DDCI calculations. Furthermore, simultaneous excitations inside the different sub-systems were allowed. Then, if  $n\pi^*_{CO}$  states involved excitations on different peptide bonds, they should be properly described. The generated determinants corresponded to all possible combinations of alpha and beta strings that kept the defined numbers of electrons in each subspace, leading, for example, to combinations of two triplets on the peptide subspaces on one side and a quintet on the aromatic cycle on the other side. There were many possibilities, and the maximal total excitation degree of a determinant could be large, 10 for NAPA *B*, corresponding to a di-excited determinant on each peptide subspace and a hexa-excitation on the aromatic cycle. This maximal total excitation degree was increased to 12 for Ac-Gly-Phe-NH<sub>2</sub> *A* (3 peptide bonds) and 14 for Ac-Gln-Phe-NH<sub>2</sub> *C* (3 peptide bonds and an amide group). However, the ORMAS/GASSCF allowed a drastic elimination of the less important determinants, compared to the CASSCF. One can note that ORMAS and GASSCF can also be used to perform RASSCF calculations. The orbitals of the singlet states are averaged in all cases.

CASPT2<sup>37</sup> calculations were performed on RASSCF/CASSCF reference wave functions to introduce dynamical correlation. GASPT2 is not yet available in the MOLCAS package. A level shift<sup>73,74</sup> of 0.2 a.u. was used as well as the standard Ionization Potential Electron Affinity (IPEA) shift<sup>75</sup> in the MOLCAS

calculations. The quasi-degenerate second-order perturbation theory for the ORMAS, ORMAS-PT method, implemented in the GAMESS package, was also applied. In that case, level shifting was not necessary.

The DDCI method developed by Caballol *et al.*,<sup>47,48</sup> a configuration interaction method, was also used to determine the correlated energies and wave functions, starting from the optimized and then localized RASSCF, CASSCF, or GASSCF orbitals. This method was used in order to reduce the number of determinants by neglecting those coming from two hole–two particle excitations external to the active space and was relevant to calculate the correlation energy that contributed to the energy difference between ground and excited states for which average starting orbitals were mandatory. In the DDCI calculations, the complete active space was reduced according to the targeted states, and several separate CI calculations were performed. The active spaces were the same as the subspaces defined in the GASSCF: (i)  $\pi$  and  $\pi^*$  orbitals for the local excitations centered on the phenyl group ( $\pi \rightarrow \pi^*$ ), corresponding to six electrons in six orbitals, and (ii) for each peptide group, 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 ( $n \rightarrow \pi^*_{\text{CO}}$ ), which led to six electrons in four orbitals. These small active spaces were denoted as CAS<sub>i</sub> with  $i = 1$  to the number of peptide bonds. For NAPA B, a larger active space was used with all the orbitals of the peptide bonds, i.e., eight electrons in 12 orbitals. This active space allowed to check the relevance of the reduced (6,4) active space. However, it could not be used for larger systems due to increasing number of peptide bonds. Size-extensivity errors were accounted by the *a posteriori* Davidson correction,<sup>76</sup> and the corrected excitation energies were denoted as Q+DDCI. In our previous study on NAPA B and in the present work, the DDCI method using localized orbitals was used in the quasi-linear-scaling version<sup>15–17</sup> performed with the EXSCI program<sup>77</sup> interfaced with MOLCAS. To neglect long-range interactions, thresholds on the exchange integrals were applied. This quasi-linear-scaling CI has been validated by comparison with calculations performed with CASPT2 in our previous work on NAPA B.<sup>5</sup> The same division of the molecular orbitals (MOs) in four zones was used in the present work. The first zone, zone 0, contained the active space of the targeted states completed by the remaining active orbitals of the full CAS. Zone 1 contained the remaining  $\sigma$  orbitals of the whole molecule. Zone 2 was defined by the non-valence virtual orbitals of the atoms involved in the active space, while zone 3 contained the rest of these non-valence orbitals. The same parameters as in our previous study on NAPA B were applied, i.e., a very small 0.0001 a.u. threshold on the zone 0 and largest ones on the other ones as defined in the [supplementary material](#) (Table S2) and detailed in our previous work.<sup>5</sup>

### III. RESULTS

#### A. CC2 results

The ground state calculations of the conformers exhibited  $D_1/D_2$  values in the 0.083–0.087/0.17–0.27 ranges, respectively, while the excited state calculations exhibited a  $D_2$  value equal to 0.25–0.27 with a biorthogonal norm  $\% \langle \bar{E}_1 | E_1 \rangle \geq 89\%$ . The  $D_1$  and  $D_2$  diagnostics computed from the single and double substitution amplitudes in the CC2 wave function were reliable indicators when static or dynamic correlation effects are not adequately treated at the

CC2 level: their magnitude is correlated with the performances of the CC2 method. The obtained values confirmed the reliability of the CC2 calculations on these systems even if some of them corresponded to the upper limit of the recommended values ( $D_1/D_2$  up to 0.15/0.25).<sup>14</sup> Moreover, in the excited states, the contributions of the canonical occupied–unoccupied HF orbitals to the total wave function change were larger or equal to 96% for the  $\pi\pi^*$  states and between 72% and 94% for the other states. The CC2/cc-pVDZ excitation energies of the first low-lying excited states (*a priori* one  $\pi\pi^*$  and one  $n\pi^*_{\text{CO}}$  per amide group) of NAPA B (three excited states), Ac–Gly–Phe–NH<sub>2</sub> A (four states), and Ac–Gln–Phe–NH<sub>2</sub> C (five states) conformers in their  $\pi\pi^*$  CC2/cc-pVDZ optimized geometry are reported in Table I. In addition, Table I contains the couple(s) of NTOs (occupied and virtual) of each state for which we obtained a contribution to the wave function greater than 10%. Moreover, the contours of the difference between the CC2 density of the different low-lying excited states and that of the ground state are shown in Fig. S1-1-3. Whatever the conformers, the first excited state is a locally  $\pi\pi^*$  excited state centered on the phenyl ring. For NAPA B, the second and third excited states are locally  $n\pi^*_{\text{CO}}$  states, each one centered on a peptide bond, the second peptide bond, that of the C-terminal side, for the second state and the first peptide bond, that of the N-terminal side, for the third one. For Ac–Gly–Phe–NH<sub>2</sub> A, the next three states are locally  $n\pi^*_{\text{CO}}$  excited states centered on two peptide bonds or one peptide bond. The second and the third excited states are localized on the two same peptide bonds, the second and the third, but with opposite weights ( $\sim 80\%$  and  $\sim 20\%$ ), whereas the fourth excited state is localized on the first peptide bond. For Ac–Gln–Phe–NH<sub>2</sub> C, the third, fourth, and fifth states are locally  $n\pi^*_{\text{CO}}$  excited states centered on one peptide bond (3, 1, and 2, respectively). The second state results from a combination of excitations of different nature,  $\pi\pi^*$  and  $n\pi^*$  (CT), resulting in an  $n[\pi^*_{\text{CO}(4)}\pi^*]$  couple of NTOs involving the same lone pair, the oxygen pure-p lone pair of the fourth peptide bond, and a virtual NTO showing  $\pi$  localized contributions both on the peptide bond 4 [ $\pi^*_{\text{CO}(4)}$ ] and on the phenyl group ( $\pi^*$ ).

#### B. Selected reference approaches: Non-dynamical correlation

In NAPA B, with all methods, the lowest  $n\pi^*_{\text{CO}}$  state is located on the second peptide bond (C-terminal side) and the second one on the first peptide bond (N-terminal side), following the numbering shown in Fig. 1. The RASSCF calculations were done for 3, 4, and 5 holes/particles as well as at the CASSCF and GASSCF/ORMAS levels. The results are presented in Table II. For  $n = 3$ , the difference to the CASSCF result for the  $\pi\pi^*$  state is +0.25 eV, reduced to +0.10 eV with  $n = 4$ , while for the  $n\pi^*_{\text{CO}}$ , the overestimation is 0.17 eV for  $n = 3$  and 0.25 eV for  $n = 4$ . With  $n = 5$ , the largest difference between the RASSCF and the CASSCF is again reduced to reach an overestimation of only 0.02 eV. The ORMAS and GASSCF reached almost the same accuracy (overestimation of 0.03 eV).

The number of determinants (Table V) is very different between each type of calculation as well as the number of iterations to obtain the convergence and the computational time of each iteration (Table II). The GASSCF number of determinants corresponds to 5.5% of the CASSCF one, while, for equivalent accuracy, the RASSCF  $n = 5$  generated 22.6% of the total number of the CASSCF

**TABLE I.** CC2/cc-pVDZ excitation energies of the first low-lying excited states (one  $\pi\pi^*$  and one  $n\pi^*_{CO}$  per amide group) of NAPA *B*, Ac-Gly-Phe-NH<sub>2</sub> *A*, and Ac-Gln-Phe-NH<sub>2</sub> *C* conformers and the couple(s) of NTOs with contribution to the wave function larger than 10%.

		$E_{\text{vert}}$ (eV)	NTOs (% , occupied $\rightarrow$ virtual)	
NAPA <i>B</i>	$\pi\pi^*$	4.82	55%	44%
	$n\pi^*_{CO(2)}$	5.63	98%	
	$n\pi^*_{CO(1)}$	5.78	99%	
Ac-Gly-Phe-NH <sub>2</sub> <i>A</i>	$\pi\pi^*$	4.77	58%	42%
	$n\pi^*_{CO(2,3)}$	5.60	77%	22%
	$n\pi^*_{CO(3,2)}$	5.64	78%	21%
	$n\pi^*_{CO(1)}$	5.69	99%	
Ac-Gln-Phe-NH <sub>2</sub> <i>C</i>	$\pi\pi^*$	4.76	56%	44%
	$n[\pi^*_{CO(4)}\pi^*]$	5.54	97%	
	$n\pi^*_{CO(3)}$	5.65	97%	
	$n\pi^*_{CO(1)}$	5.67	97%	
	$n\pi^*_{CO(2)}$	5.82	99%	

**TABLE II.** NAPA B. RASSCF, GASSCF, ORMAS, and CASSCF energies.

$\Delta E$ (eV)	RASSCF (3h3p) RAS(18,3,3; 9,0,5)	RASSCF (4h4p) RAS(18,4,4; 9,0,5)	RASSCF (5h5p) RAS(18,5,5; 9,0,5)	GASSCF (0h0p)	ORMAS (0h0p)	CASSCF
GS (a.u.)	-683.510 968	-683.522 861	-683.523 073	-683.523 403	-683.523 405	-683.523 796
$\pi\pi^*$	4.78	4.63	4.53	4.54	4.54	4.53
$n\pi^*_{CO}$	5.93 6.06	6.01 6.14	5.79 5.91	5.81 5.91	5.81 5.91	5.78 5.89
No. of iterations <sup>a</sup>	411		254	43	31	25
Time/iteration (min)	0.49	1.39	2.69	0.46		9.74
Total time	11 h 11 min		12 h	21 min	38 min	4 h 15 min

<sup>a</sup>The different calculations were performed on a bi-(4c) Intel Xeon E5-2637 v3 machine, using one processor and 19 GB of memory.

determinants. The convergence of the RASSCF was very slow. For convergence thresholds on the energy and on the rotation of the molecular orbitals of  $10^{-8}$  a.u. and  $10^{-4}$ , respectively, the RASSCF  $n = 3$  time per iteration is similar to that of the GASSCF but required many more iterations (411 vs 43) and led to a significant overestimation compared to the CASSCF values as for the RASSCF  $n = 4$ . One can note some difficulties in the convergence of the RASSCF  $n = 4$ , depending on the starting point. Indeed, starting—as in the other cases—from the SCF orbitals, the convergence was not only extremely slow but it also led to states whose nature was not totally in agreement with the CASSCF solutions. However, it was not the case when the starting orbitals were those of the CASSCF. Only the time per iteration was then relevant for a comparison with other methods (Table II). For the RASSCF  $n = 5$ , the number of iterations was reduced and the results gained accuracy. Finally, to reach an accuracy similar to that of the CASSCF, the RASSCF converged in 12 h ( $n = 5$ ), whereas the GASSCF converged in only 21 min, with a time per iteration and a number of iterations  $\sim 5$  times smaller. The CASSCF converged in fewer iterations than all the other methods, but the time per iteration was the greatest. In conclusion, the total time of the RASSCF calculations  $n = 3-5$  was much greater than that of the CASSCF, and only the GASSCF was more efficient, 21 min instead of 4 h 15 min, while allowing equivalent accuracy. The performances of ORMAS calculations (31 iterations in 38 min) were similar to those of the GASSCF (43 iterations in 31 min), slightly slowed down by the need to calculate all the spin multiplicities of the

intermediate states, for a total of 14 triplet and singlet states instead of the five singlets.

For Ac-Gly-Phe-NH<sub>2</sub> A (Table III), one can note that all the  $n\pi^*_{CO}$  states involve the same peptide bonds regardless of the method used; the first state is located on the third peptide bond, the second one on the first peptide bond, and the last state on the second peptide bond (Fig. 1). Concerning the energetics, overestimations of 0.2 eV ( $\pi\pi^*$  state) and 0.3 eV ( $n\pi^*_{CO}$  states) were obtained with the RASSCF  $n = 3$  reduced to 0.10 and 0.24 eV for  $n = 4$ . One more time, the RASSCF  $n = 5$  gave comparable results to the GASSCF/ORMAS ones, with an overestimation of 0.03 eV relative to the CASSCF results. As in the case of NAPA B, the convergence of the RASSCF was very slow: 197 iterations for  $n = 3$ , 86 for  $n = 5$ , while the calculation for  $n = 4$  did not converge after more than 400 iterations. The convergence thresholds of the RASSCF  $n = 4$  calculation were then raised to  $10^{-7}$  a.u. on the energy and  $10^{-3}$  on the MO rotation to converge in 153 iterations, but a larger error on the  $n \rightarrow \pi^*_{CO}$  excitation energies was observed. The CASSCF calculation,  $172 \times 10^6$  determinants (Table V), was at the limits of what could be done. The ratio of the number of determinants between the RASSCF  $n = 5$  and the GASSCF was still to the advantage of the latter (1.6% instead of 3.4% of the CASSCF space). However, it was to a lesser extent than for NAPA B (5.5% vs 22.6%).

The CASSCF for Ac-Gln-Phe-NH<sub>2</sub> C was intractable. Due to the near-degeneracies of the first and second  $n\pi^*_{CO}$  states as well as the third and fourth ones, the nature of the different states differed

**TABLE III.** Ac-Gly-Phe-NH<sub>2</sub> A. RASSCF, GASSCF, ORMAS, and CASSCF energies.

$\Delta E$ (eV)	RASSCF (3h3p) RAS(24,3,3; 12,0,6)	RASSCF <sup>a</sup> (4h4p) RAS(24,4,4; 12,0,6)	RASSCF (5h5p) RAS(24,5,5; 12,0,6)	GASSCF (0h0p)	ORMAS (0h0p)	CASSCF
GS (a.u.)	-890.354 353	-890.369 503	-890.372 827	-890.373 389	-890.373 389	-890.374 093
$\pi\pi$	4.82	4.68	4.54	4.53	4.53	4.52
$n\pi^*_{CO}^{(3)}$	6.02	6.06	5.86	5.86	5.86	5.83
$n\pi^*_{CO}^{(1)}$	6.12	6.16	5.95	5.95	5.95	5.92
$n\pi^*_{CO}^{(2)}$	6.13	6.17	5.96	5.97	5.97	5.93

<sup>a</sup>RASSCF  $n = 4$  convergence thresholds of  $10^{-7}$  a.u. on the energy and  $10^{-3}$  on the MO rotation instead of  $10^{-8}$  and  $10^{-4}$ , respectively.

**TABLE IV.** Ac-Gln-Phe-NH<sub>2</sub> C. RASSCF and GASSCF energies.

$\Delta E$ (eV)	RASSCF (3h3p) RAS(30,3,3; 15,0,7)	<<RASSCF>> (5h5p) RAS(30,5,5; 15,0,7)	GASSCF (0h0p)
GS (a.u.)	-1136.243 231	-1136.266 795	-1136.268 378
$\pi\pi^*$	4.89	4.57	4.56
$n\pi^*_{CO}$	6.22 <sup>(3)</sup>	6.07 <sup>(1,3)</sup>	6.04 <sup>(1)</sup>
	6.25 <sup>(1)</sup>	6.07 <sup>(3,1)</sup>	6.06 <sup>(3)</sup>
	6.32 <sup>(4)</sup>	6.15 <sup>(4)</sup>	6.15 <sup>(2)</sup>
	6.35 <sup>(2)</sup>	6.18 <sup>(2)</sup>	6.15 <sup>(4)</sup>

according to the method used. However, the first and the second  $n\pi^*_{CO}$  states involved the first and third peptide bonds, while the two last states involved the second and the fourth peptide bonds (Fig. 1). Even the RASSCF calculations were problematic due to the slow convergence, and only the RASSCF  $n = 3$  was obtained. The “RASSCF”  $n = 5$  results were obtained thanks to the GASSCF code, equivalent to the RASSCF one, with a better convergence. The RASSCF results are compared to the GASSCF ones in Table IV. As for NAPA B, the error is larger for the  $\pi\pi^*$  state (0.33 eV) than for  $n\pi^*_{CO}$  states (0.18–0.20 eV) for the RASSCF  $n = 3$ , while for  $n = 5$ , this discrepancy in accuracy according to the nature of the states disappeared, and the results were found to be similar to the GASSCF ones. One can note that the total number of determinants is now smaller for the RASSCF  $n = 5$  than for the GASSCF one (Table V), but the calculation time is largely in favor of the latter. The ORMAS method was not applied to this very large compound as the calculation required considerable computational time and huge memory resources and was not necessary for the subsequent CI calculations.

In conclusion, for the RASSCF strategies, it was necessary to go until an excitation degree of 5 to obtain results in good agreement with the CASSCF ones. The convergence was slow for all excitation degrees and even problematic for  $n = 4$ . The RASSCF code used as a configuration interaction coupled with the optimization of the

orbitals was not very efficient, and as mentioned previously, the partition of the RAS subspaces in a state specific way (Appendix S5 of the supplementary material) gave relevant results. However, the partition used was suitable for using the DDCI method, based on a common set of state-averaged orbitals. The number of determinants of the GASSCF grew more rapidly with the number of subspaces than that in the RASSCF case for which the number of subspaces was, in the present case, always limited to 2 whatever the number of peptide bonds. Indeed, for Ac-Gln-Phe-NH<sub>2</sub> C, the number of determinants became larger for the GASSCF compared to the RASSCF  $n = 5$ . Actually, the excitation degree of the determinants in the GASSCF could be larger than that in the RASSCF due to the multiplication of the subspaces. For example, in the case of Ac-Gln-Phe-NH<sub>2</sub> C, the combination of a mono-excitation on each peptide subspace coupled to a di-excitation on the cycle corresponded to hexa-excited determinants that were not present in the RASSCF  $n = 5$ . As mentioned in Sec. II C, the maximal total excitation degree is 14 for this system. Nevertheless, the GASSCF or ORMAS methods using small active subspaces seemed then to be the most relevant strategy to calculate non-dynamical energy for a very reasonable computational cost. The comparison with CASSCF results when possible also gave good agreement, with a maximal overestimation of 0.03 eV on the excitation energies whatever the nature of state.

### C. Selected reference approaches: Dynamical correlation

Dynamical correlation was taken into account using the CASPT2 method on RASSCF/CASSCF reference wave functions and also thanks to the Davidson corrected DDCI configuration interaction method. The optimized RASSCF, GASSCF, or CASSCF orbitals were used as the starting point of the DDCI. The results are presented in Tables VI–VIII. As presented in the computational details, the small active spaces used in the DDCI calculations were the subspaces defined in the GASSCF: (i) CAS(6,6) for the phenyl group, (ii) CAS(6,4) for each peptide bond denoted as CAS<sub>*i*</sub> with  $i = 1$  to the number of peptide bonds (Fig. 1), and (iii) for NAPA B, CAS(8,12) with all the orbitals of the peptide bonds.

**TABLE V.** Dimensions of the RASSCF, GASSCF, and CASSCF methods (in GAMESS). The percentages in brackets are the ratio of the number of determinants of the different methods compared to the number of determinants obtained in the CASSCF.

Subspaces	NAPA B	Ac-Gly-Phe-NH <sub>2</sub> A	Ac-Gln-Phe-NH <sub>2</sub> C
RASSCF	(18,n,n; 9,0,5)	(24,n,n; 12,0,6)	(30,n,n; 15,0,7)
Number of det ( $n = 3$ )	36 916 (0.9%)	158 669 (0.05%)	510 546 (0.002%)
Number of det ( $n = 4$ )	243 376 (6.1%)	1 787 219 (0.5%)	8 812 371 (0.03%)
Number of det ( $n = 5$ )	905 128 (22.6%)	11 577 923 (3.4%)	89 200 497 (0.3%)
Subspaces (number of e, number of Mos)			
GASSCF/ORMAS	(6,6)/(6,4)/(6,4)	(6,6)/(6,4)/(6,4)/(6,4)	(6,6)/(6,4)/(6,4)/(6,4)/(6,4)
Number of det	220 192 (5.5%)	5 510 848 (1.6%)	140 843 008 (0.5%)
CASSCF	(18,14)	(24,18)	(30,22)
Number of det	4 008 004 (100%)	344 622 096 (100%)	29 085 255 936 (100%)

**TABLE VI.** NAPA B. RASPT2, CASPT2, ORMAS-PT, and Q+DDCI corresponding to the different sets of optimized MOs.

Ref. WF	RASSCF n = 3 MOs		RASSCF n = 4 MOs		RASSCF n = 5 MOs		ORMAS MOs	GASSCF MOs	CASSCF MOs	
	RASPT2	Q+DDCI	RASPT2	Q+DDCI	RASPT2	Q+DDCI			ORMAS-PT	Q+DDCI
$\pi\pi^*$	4.64	CAS(6,6)	4.54	CAS(6,6)	4.54	CAS(6,6)	4.42	CAS(6,6)	4.57	CAS(6,6)
		4.73		4.72		4.71		4.72		
$n\pi^*_{CO}$	5.72 5.89	CAS(12,8)	5.77	CAS(12,8)	5.92	CAS(12,8)	5.67 5.85	CAS(12,8)	5.83	CAS(12,8)
		5.82		5.82		5.84		5.83		
		6.05		6.02		6.05		6.03		6.02
		CAS1/CAS2		CAS1/CAS2		CAS1/CAS2		CAS1/CAS2		CAS1/CAS2
		5.83		5.83		5.84		5.84		5.83
5.97	5.96	5.98	5.98	5.98						

**TABLE VII.** Ac-Gly-Phe-NH<sub>2</sub> A. RASPT2, CASPT2, and Q+DDCI corresponding to the different sets of optimized MOs. Superscripts (1), (2), and (3) refer to the peptide bond numbering (see Fig. 1).

$\Delta E$ (eV)	RASSCF n = 3 MOs		RASSCF n = 4 MOs		RASSCF n = 5 MOs		GASSCF MOs	CASSCF MOs		
	RASPT2	Q+DDCI	RASPT2	Q+DDCI	RASPT2	Q+DDCI		Q+DDCI	CASPT2	Q+DDCI
$\pi\pi^*$	4.62	CAS(6,6)	4.45	CAS(6,6)	4.46	CAS(6,6)	4.66	CAS(6,6)	4.57	CAS(6,6)
		4.66		4.64		4.65		4.66		4.65
$n\pi^*_{CO}$	5.63 <sup>(3)</sup> 5.71 <sup>(2)</sup> 5.76 <sup>(1)</sup>	CAS1/2/3	5.76 <sup>(3)</sup> 5.78 <sup>(2)</sup> 5.85 <sup>(1)</sup>	CAS1/2/3	5.62 <sup>(3)</sup> 5.67 <sup>(2)</sup> 5.73 <sup>(1)</sup>	CAS1/2/3	5.79 <sup>(3)</sup> 5.88 <sup>(1)</sup> 5.90 <sup>(2)</sup>	CAS1/2/3	5.65 <sup>(3)</sup> 5.66 <sup>(2)</sup> 5.71 <sup>(1)</sup>	CAS1/2/3
		5.79 <sup>(3)</sup>		5.80 <sup>(3)</sup>		5.78 <sup>(3)</sup>		5.79 <sup>(3)</sup>		5.77 <sup>(3)</sup>
		5.88 <sup>(2)</sup>		5.85 <sup>(1)</sup>		5.87 <sup>(2)</sup>		5.88 <sup>(1)</sup>		5.87 <sup>(1)</sup>
		5.89 <sup>(1)</sup>		5.87 <sup>(2)</sup>		5.87 <sup>(1)</sup>		5.90 <sup>(2)</sup>		5.88 <sup>(2)</sup>

For NAPA B, the analysis of the CASSCF wave function (expressed in local orbitals) of the  $n\pi^*_{CO}$  excited states shows a major contribution of the  $n_O \rightarrow \pi^*_{CO}$  mono-excitation and the  $n_N n_O \rightarrow (\pi^*_{CO})^2$  di-excitation in the same subspace counted for 5%. There is also a non-negligible contribution (8%–9%) of di-excitations, implying the second peptide bond ( $n_N \rightarrow \pi^*_{CO}$ ) coupled with  $n_O \rightarrow \pi^*_{CO}$  in the considered state. These simultaneous excitations are also present in the DDCI wave functions with a (12,8) active space and even in larger proportions: 8.5% of  $n_N n_O \rightarrow (\pi^*_{CO})^2$  intra-di-excitation and 16% ( $n_O \rightarrow \pi^*_{CO}$ )<sub>2</sub> ( $n_N \rightarrow \pi^*_{CO}$ )<sub>1</sub>

inter-di-excitation. For the second  $n_O \pi^*_{CO}$  state, they are almost equivalent, 9%–14%, respectively. Of course, the wave functions of the DDCI with reduced (6,4) active spaces, with a separate treatment of the peptide bonds, were different. Indeed, the weight of the ( $n_O \rightarrow \pi^*_{CO}$ )<sub>i</sub> ( $n_N \rightarrow \pi^*_{CO}$ )<sub>j</sub> intra-excitation was reduced (11%–12% vs 14%–16%) as it was only introduced by a mono-excitation on a reference determinant and then less correlated by the other ones. The weight of the main determinant is larger (61%) than that of the Q+DDCI on the CAS(12,8) active space (48%), while the weight of the inter-di-excitation is reduced (5% vs 9%).

**TABLE VIII.** Ac-Gln-Phe-NH<sub>2</sub> C. RASPT2 (n = 3) and Q+DDCI with different sets of optimized MOs. Superscripts (1), (2), (3), and (4) refer to the peptide bond numbering (see Fig. 1).

$\Delta E$ (eV)	RASSCF n = 3 MOs		"RASSCF" <sup>a</sup> n = 5 MOs		GASSCF MOs	
	RASPT2	Q+DDCI CAS1/2/3/4				
$\pi\pi^*$	4.64	4.63	4.62	4.62	4.63	4.63
$n\pi^*_{CO}$	5.70 <sup>(3)</sup> 5.70 <sup>(1)</sup> 5.71 <sup>(4)</sup> 5.85 <sup>(2)</sup>	5.98 <sup>(3)</sup> 5.99 <sup>(4)</sup> 6.03 <sup>(2)</sup> 6.05 <sup>(1)</sup>	5.97 <sup>(3)</sup> 5.99 <sup>(4)</sup> 6.03 <sup>(2)</sup> 6.05 <sup>(1)</sup>	5.97 <sup>(3)</sup> 5.99 <sup>(4)</sup> 6.03 <sup>(2)</sup> 6.05 <sup>(1)</sup>	5.98 <sup>(3)</sup> 6.00 <sup>(4)</sup> 6.04 <sup>(2)</sup> 6.05 <sup>(1)</sup>	5.98 <sup>(3)</sup> 6.00 <sup>(4)</sup> 6.04 <sup>(2)</sup> 6.05 <sup>(1)</sup>

<sup>a</sup>"RASSCF" n = 5 was obtained thanks to the GASSCF.

**TABLE IX.** CC2/cc-pVDZ and GASSCF-Q+DDCI excitation energies of the first low-lying excited states (one  $\pi\pi^*$  and one  $n\pi^*_{CO}$  per amide group) of NAPA *B*, Ac-Gly-Phe-NH<sub>2</sub> *A*, and Ac-Gln-Phe-NH<sub>2</sub> *C* conformers.

$\Delta E$ (eV)	GASSCF-Q+DDCI		CC2		GASSCF-Q+DDCI		CC2	
	NAPA <i>B</i>		Ac-Gly-Phe-NH <sub>2</sub> <i>A</i>		Ac-Gln-Phe-NH <sub>2</sub> <i>C</i>			
$\pi\pi^*$	4.72	4.82	4.66	4.77	4.63		4.76	
$n\pi^*_{CO}$	5.84	5.63	5.79	5.64	5.98		5.65	
	5.98	5.78	5.88	5.69	<i>6.00<sup>a</sup></i>		<i>5.54<sup>a</sup></i>	
			5.90	5.60	6.04		5.82	
				6.05		5.67		

<sup>a</sup>Values in italics indicate that the nature of this state differs from CC2 to GASSCF-Q+DDCI (see the text).

The RASPT2 for  $n = 3$  presented a difference with the CASPT2 of +0.04 to +0.07 eV (Table VI). For  $n = 4$ , this difference was of -0.03 eV for the  $\pi\pi^*$  state and -0.06 eV for the  $n\pi^*_{CO}$  states. For  $n = 5$ , the deviation to the CASPT2 is -0.03 eV for the  $\pi\pi^*$  state and +0.08/+0.09 eV for  $n\pi^*_{CO}$  states. The ORMAS-PT results were found very similar to those of the CASPT2 ones for the  $n\pi^*_{CO}$  states, while the  $\pi\pi^*$  state presented a difference of 0.13 eV. The two perturbative treatments are not equivalent, and additionally, the use of the IPEA shift in the CASPT2 method corrects the underestimation of the excitation energies. Furthermore, a level shift of 0.2 a.u. was necessary to avoid intruder states in the CASPT2 contrary to the ORMAS-PT for which no level shift was necessary. The use of level shift tends to increase excitation energies.

Whatever the MO set used, i.e., RASSCF, GASSCF, or CASSCF MOs, the Q+DDCI results were not very sensitive and gave almost the same excitation energies. Compared to the CASPT2 results, the difference was of +0.15 eV for the  $\pi\pi^*$  state and +0.15/+0.19 eV for the  $n\pi^*_{CO}$  states. The excitation energies were then shifted by about the same amount.

By comparing the two active spaces used to describe the  $n\pi^*_{CO}$  states—CAS(12,8) or the two CAS<sub>i</sub>(6,4)—at the Q+DDCI level, one can note that there is a small difference: around 0.01 eV for the first excited state and 0.04 eV on the second one. To make possible the calculation of these excitation energies in the larger systems, this strategy could then be used, provided that the contribution of the other peptides in each  $n\pi^*_{CO}$  remained small.

For Ac-Gly-Phe-NH<sub>2</sub> *A*, the analysis of the wave functions of the  $n\pi^*_{CO}$  states between the different methods presents some differences. The lowest  $n\pi^*_{CO}$  state is located on the third peptide bond (see Fig. 1) in all cases. The second and third states calculated at the CASPT2 level are located on the second and first peptide bonds, respectively. At the RASPT2 levels, the nature of the different  $n\pi^*_{CO}$  states is the same. The Q+DDCI method found the second and the third  $n\pi^*_{CO}$  states almost degenerate, and inversion of these two states was encountered for the Q+DDCI calculations using the RASSCF  $n = 4$ , GASSCF, and CASSCF MOs. The first and second states were found close in energy at the CASPT2 level but not at the RASPT2 ones, except for the  $n = 4$  which was badly converged at the RASSCF level.

The analysis of the DDCI wave functions of the  $n\pi^*_{CO}$  states showed that the excitation  $n_O \rightarrow \pi^*_{CO}$  has a weight of 60% for the two first states and 58% for the last one, and the intra di-excitation  $n_N n_O \rightarrow (\pi^*_{CO})^2$  counts for 10%–12%, comparable to the equivalent weights in NAPA *B*. There were also contributions of the inter

di-excitations (8%), which were enhanced compared to NAPA *B* as there was one peptide bond more.

Concerning the energetics (Table VII), the difference between the CASPT2 and Q+DDCI excitation energies is smaller than that for NAPA *B*: +0.08 eV compared to +0.15 eV for the  $\pi\pi^*$  state and similar for the  $n\pi^*_{CO}$  states: +0.12 to 0.21 eV compared to +0.14 to 0.19 eV with the small CAS(6,4) active spaces. As for NAPA *B*, the set of molecular orbitals used did not have a notable effect on the Q+DDCI results (0.03 eV). The RASPT2 for  $n = 3$  presents an overestimation of +0.05 eV compared to the CASPT2 for the  $\pi\pi^*$  state, while it is an underestimation of -0.12 eV for  $n = 4$ . For  $n\pi^*_{CO}$  states, the RASPT2  $n = 3$  shows small differences with the CASPT2 (-0.02/+0.05 eV), while they are larger for  $n = 4$  (+0.11/+0.16 eV). For RASPT2  $n = 5$ , the agreement with the CASPT2 is similar to that obtained for  $n = 4$  for the  $\pi\pi^*$  state but strongly improved for the  $n\pi^*_{CO}$  states (-0.03/+0.02 eV).

For Ac-Gln-Phe-NH<sub>2</sub> *C* (Table VIII), in the DDCI wave functions of the  $n\pi^*_{CO}$  states, the weight of the intra-di-excitation  $n_N n_O \rightarrow (\pi^*_{CO})^2$  is similar for all the states (9.7%–10.2%) and comparable with that the smaller systems, with the exception of the second state for which it is 11.4%. The weight of the inter-di-excitations is 12.3%–12.4% for three of the states, the second one is again slightly different with a weight of 11.7%. The weight of the main reference is slightly lower as the number of peptides increases, between 57% and 58% instead of 58%–60% for Ac-Gly-Phe-NH<sub>2</sub> *A* and 61% for NAPA *B*. Only the RASSCF  $n = 3$  was affordable with MOLCAS in computing time and quality of convergence; then, the comparison was here between the RASPT2  $n = 3$  and the Q+DDCI on the same set of MOs and on the GASSCF one. The “RASSCF”  $n = 5$  was obtained thanks to the GASSCF code, and the perturbation treatment is not currently possible on these zeroth-order wave functions. For RASPT2  $n = 3$ , there is a difference of +0.01 eV for the  $\pi\pi^*$  state and -0.19/-0.35 eV for the  $n\pi^*_{CO}$  states compared to the Q+DDCI calculations. As for the smaller systems, the MOs has almost no influence on the Q+DDCI excitation energies. The three lowest  $n\pi^*_{CO}$  states are almost degenerate at the RASPT2 level, and the last one was found to be +0.14 eV higher. At the Q+DDCI level, the largest gap between these states is 0.07 eV. This small gap between the states explains the inversion of the  $n\pi^*_{CO}$  states between RASPT2  $n = 3$  and Q+DDCI calculations.

In conclusion, the Q+DDCI calculations are not very sensitive to the starting MOs, which suggests that they are all sufficiently satisfactory. The RASPT2 is much more sensitive to the zeroth-order

wave functions, and only the  $n = 5$  results are comparable to the CASPT2 ones.

The active spaces used in the DDCI calculations contain the same number of electrons and orbitals for the three systems; then, the dimensions of the quasi-linear-scaling CI calculations were almost the same: between  $119 \times 10^6$  and  $155 \times 10^6$  of determinants with that CAS(6,6) describing the  $\pi\pi^*$  state, while the CAS<sub>i</sub>(6,4) generate less than  $7 \times 10^6$  determinants for the  $n\pi^*_{CO}$  states. The large active space also used to describe these states in the case of NAPA B, CAS(12,8), generated about  $580 \times 10^6$  of determinants. It was then impossible to enlarge this active space with the additional peptide groups of the larger systems. All the dimensions of the CI calculations can be found in the [supplementary material](#) (S2).

#### IV. PERFORMANCES OF THE CC2 METHOD

Along the series of the capped peptides, both the nature and the energetics obtained at the CC2 level for the lowest  $\pi\pi^*$  excited states, the state for which the geometry has been optimized at this level, are in very good agreement with those obtained at the GASSCF-Q+DDCI level (Table IX). Indeed, the excitation energies were overestimated for all the capped peptides of around 0.11 eV, a systematic overestimation independent of the size and inferior to the standard error of the CASPT2 method ( $\pm 0.2$  eV). Moreover, this overestimation is of the same order as that observed between the MR methods themselves, i.e., between CASPT2 and GASSCF-Q+DDCI, for example. The nature of  $n\pi^*_{CO}$  states is also well described at the CC2 level, and only one exception difference was observed compared to the GASSCF-DDCI. This difference concerns the lowest  $n\pi^*_{CO}$  state of Ac-Gln-Phe-NH<sub>2</sub> C, whose CC2 wave function exhibits a non-negligible contribution of electronic charge transfer from the backbone to the phenyl ring, a CT contribution (see the NTO couple in Table I and the difference between the density of the ground state and this state on Fig. S1-3), whereas the wave function obtained at the GASSCF-DDCI level does not present such a contribution as electron excitation between spaces was not allowed. The CC2 excitation energies of the  $n\pi^*_{CO}$  states are well reproduced compared to the GASSCF-Q+DDCI level, but unlike the  $\pi\pi^*$  states, they are underestimated for all the capped peptides, and this underestimation was a little larger for the Ac-Gln-Phe-NH<sub>2</sub> C and larger than the overestimation observed for the  $\pi\pi^*$  states. Indeed, if a similar underestimation has been observed for NAPA B and Ac-Gly-Phe-NH<sub>2</sub> A (in average  $-0.21$  eV), this underestimation was in average equal to 0.35 eV for Ac-Gln-Phe-NH<sub>2</sub> with a maximum for the  $n\pi^*_{CO}$ , which exhibits a CT contribution ( $-0.46$  eV) and a minimum for the  $n\pi^*_{CO}$  state localized on the second peptide bond ( $-0.22$  eV).

In conclusion, the CC2 method reproduces well the nature of the excited states independent of the size of the systems. In addition, even if the discrepancies of the excitation energies can depend on the nature of the excited states [in average  $+0.11$  eV for the  $\pi\pi^*$  excited states and  $-0.21$  eV (NAPA B and Ac-Gly-Phe-NH<sub>2</sub> A) /  $-0.35$  eV (Ac-Gln-Phe-NH<sub>2</sub> C) for the  $n\pi^*_{CO}$  states], the CC2 excitation energies exhibit a systematic discrepancy compared to the GASSCF-Q+DDCI ones, an overestimation or an underestimation, for each type of excited state independent of the size of the systems. Moreover, these discrepancies are only a little higher than those obtained between the more sophisticated MRCI methods

investigated in this work and of the same order of the standard error obtained for these MRCI methods, i.e.,  $\pm 0.2$  eV.

#### V. CONCLUSIONS

All three alternatives to CASSCF, namely, RASSCF, ORMAS, and GASSCF, yield results in good agreement (wave function and energetics) with the CASSCF, where that can be performed. Of these, the ORMAS and the GASSCF are the most efficient, especially for larger systems such as capped peptides containing at least two residues. Indeed, the RASSCF requires going up to  $n = 5$  to obtain the CASSCF accuracy and quickly becomes not tractable such as in the case of Ac-Gln-Phe-NH<sub>2</sub> C.

When adding dynamical correlation, both the nature and the energetics of the excited states are also well reproduced at the GASSCF-Q+DDCI level compared with the CASPT2 level. Even if the CASPT2 and the CI methods are state-of-the-art methods, they also both present some drawbacks: (1) for the CI calculations, the use of the smallest active spaces and the size-consistency error, partially corrected by the Davidson procedure, and (2) for the CASPT2 calculations, the use of IPEA and level shifts or simply the fact that energies are not upper bounds of the real ones. The basis set also has a significant effect on the IPEA shift.<sup>78</sup> In the present study, the zeroth-order wave function plays an important role in the CASPT2 calculations, while the CI method seems not to be very sensitive to it. The CASPT2 method is currently used and gives very reasonable results on UV-visible spectroscopies, while the DDCI method is the most accurate method to study magnetic problems and hardly used for spectroscopic ones. The two methods can then be seen as complementary. In addition, the question of which method has to be used no longer arises for very large systems since the CASPT2 calculations become intractable.

In order to partition the full CAS into subspaces, the use of the RAS/GAS/ORMAS methods implies a good understanding of the electronic structure of the different states. Furthermore, in the present case, as the considered excitations are local, interspace excitations between different GAS spaces can be ignored. To go further, investigating the charge transfer states is a challenging task as electron excitation between subspaces will generate a large amount of determinants and as these states are also found higher in energy at the non-dynamical correlated level, with the necessity to average the orbitals on about 20 states.<sup>5</sup>

Finally, the discrepancies obtained by the CC2 method for the excitation energies exhibit a systematic overestimation or underestimation according to the nature of the excited states compared to those obtained by the MRCI methods, these discrepancies being independent of the size of the systems. The extension of the CC2 method to such large systems without loss of accuracy was then demonstrated, highlighting the great potential of this method to treat accurately excited states, mainly single reference, of very large systems with the crucial advantage of having access not only to the energy but also to both the gradient and the Hessian.

#### DEDICATION

The authors dedicate this article to Rosa Caballol and Ria Broer for their contributions to the development of theoretical methods

to accurately describe electron correlation effects. In particular, the difference dedicated configuration interaction method is an accurate and versatile strategy to calculate excitation energies as illustrated in this paper.

## SUPPLEMENTARY MATERIAL

See the [supplementary material](#) for Appendix S1: contours of the difference between the CC2 density of the different low-lying excited states [ $\pi\pi^*$  ( $\pm 0.0015$  a.u.) and others ( $\pm 0.03$  a.u.)] and that of the ground state of NAPA B, Ac-Gly-Phe-NH<sub>2</sub> A, and Ac-Gln-Phe-NH<sub>2</sub> C; Appendix S2: dimensions of the CI calculations and thresholds for the quasi-linear-scaling CI; Appendix S3: geometries; Appendix S4: NAPA B: effect of the basis set; and Appendix S5: NAPA B: different choices of RAS subspaces.

## 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. 2016-t2016087540, 2017-A0010807540, and 2020-A0070807540 awarded by GENCI (Grand Équipement National de Calcul Intensif) and to the CCRT High Performance Computing (HPC) facility at CEA under Grant No. CCRT2016/CCRT2017/CCRT2020-p606bren. N.B.A. thanks the computing facility CALMIP for the allocation of computing resources (Project No. P16009) at the University of Toulouse (UPS). M.S.G. and M.W.S. acknowledge support from a U.S. National Science Foundation Software Infrastructure (SI2) (Grant No. OCI-1047772).

## DATA AVAILABILITY

The data that support the findings of this study are available within this article and its [supplementary material](#).

## REFERENCES

- L. González, D. Escudero, and L. Serrano-Andrés, *ChemPhysChem* **13**, 28 (2012).
- M. E. Casida and M. Huix-Rotllant, *Annu. Rev. Phys. Chem.* **63**, 287 (2012).
- K. Sneskov and O. Christiansen, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2**, 566 (2012).
- D. Roca-Sanjuán, F. Aquilante, and R. Lindh, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2**, 585 (2012).
- N. Ben Amor, S. Hoyau, D. Maynau, and V. Brenner, *J. Chem. Phys.* **148**, 184105 (2018).
- M.-S. Dupuy, E. Gloaguen, B. Tardivel, M. Mons, and V. Brenner, *J. Chem. Theory Comput.* **16**, 601 (2020).
- M. Mališ, Y. Loquais, E. Gloaguen, H. S. Biswal, F. Piuze, B. Tardivel, V. Brenner, M. Broquier, C. Juvet, M. Mons, N. Došlić, and I. Ljubić, *J. Am. Chem. Soc.* **134**, 20340 (2012).
- M. Mališ, Y. Loquais, E. Gloaguen, C. Juvet, V. Brenner, M. Mons, I. Ljubić, and N. Došlić, *Phys. Chem. Chem. Phys.* **16**, 2285 (2014).
- E. Runge and E. K. U. Gross, *Phys. Rev. Lett.* **52**, 997 (1984).
- O. Christiansen, H. Koch, and P. Jørgensen, *Chem. Phys. Lett.* **243**, 409 (1995).
- C. Hättig and F. Weigend, *J. Chem. Phys.* **113**, 5154 (2000).
- C. Hättig and A. Köhn, *J. Chem. Phys.* **117**, 6939 (2002).
- C. Hättig, *J. Chem. Phys.* **118**, 7751 (2003).
- A. Köhn and C. Hättig, *J. Chem. Phys.* **119**, 5021 (2003).
- B. Bories, D. Maynau, and M.-L. Bonnet, *J. Comput. Chem.* **28**, 632 (2007).
- N. Ben Amor, F. Bessac, S. Hoyau, and D. Maynau, *J. Chem. Phys.* **135**, 014101 (2011).
- C. Chang, C. J. Calzado, N. B. Amor, J. S. Marin, and D. Maynau, *J. Chem. Phys.* **137**, 104102 (2012).
- B. O. Roos, P. R. Taylor, and P. E. M. Sigbahn, *Chem. Phys.* **48**, 157 (1980).
- G. Das and A. C. Wahl, *J. Chem. Phys.* **44**, 87 (1966).
- G. Das and A. C. Wahl, *J. Chem. Phys.* **47**, 2934 (1967).
- K. Ruedenberg and K. R. Sundberg, in *Quantum Science, Methods and Structure: A Tribute to Per-Olov Löwdin*, edited by J.-L. Calais, O. Goscinski, J. Linderberg, and Y. Öhrn (Springer US, Boston, MA, 1976), pp. 505–515.
- A. I. Panin and K. V. Simon, *Int. J. Quantum Chem.* **59**, 471 (1996).
- A. I. Panin and O. V. Sizova, *J. Comput. Chem.* **17**, 178 (1996).
- Y. G. Khaït, J. Song, and M. R. Hoffmann, *Int. J. Quantum Chem.* **99**, 210 (2004).
- P. Å. Malmqvist, A. Rendell, and B. O. Roos, *J. Phys. Chem.* **94**, 5477 (1990).
- J. Olsen, B. O. Roos, P. Jørgensen, and H. J. Aa. Jensen, *J. Chem. Phys.* **89**, 2185 (1988).
- T. Fleig, J. Olsen, and C. M. Marian, *J. Chem. Phys.* **114**, 4775 (2001).
- T. Fleig, J. Olsen, and L. Visscher, *J. Chem. Phys.* **119**, 2963 (2003).
- D. Ma, G. Li Manni, and L. Gagliardi, *J. Chem. Phys.* **135**, 044128 (2011).
- K. D. Vogiatzis, D. Ma, J. Olsen, L. Gagliardi, and W. A. de Jong, *J. Chem. Phys.* **147**, 184111 (2017).
- J. Ivanic, *J. Chem. Phys.* **119**, 9364 (2003).
- J. Ivanic and K. Ruedenberg, *J. Comput. Chem.* **24**, 1250 (2003).
- K. Hirao, *Chem. Phys. Lett.* **201**, 59 (1993).
- K. Hirao, *Chem. Phys. Lett.* **196**, 397 (1992).
- H. Nakano, *Chem. Phys. Lett.* **207**, 372 (1993).
- H. Nakano, *J. Chem. Phys.* **99**, 7983 (1993).
- K. Andersson, P. Å. Malmqvist, B. O. Roos, A. J. Sadlej, and K. Wolinski, *J. Phys. Chem.* **94**, 5483 (1990).
- C. Angeli, R. Cimraglia, S. Evangelisti, T. Leininger, and J.-P. Malrieu, *J. Chem. Phys.* **114**, 10252 (2001).
- C. Angeli, R. Cimraglia, and J.-P. Malrieu, *J. Chem. Phys.* **117**, 9138 (2002).
- P. Å. Malmqvist, K. Pierloot, A. R. M. Shahi, C. J. Cramer, and L. Gagliardi, *J. Chem. Phys.* **128**, 204109 (2008).
- L. Roskopf and M. S. Gordon, *J. Chem. Phys.* **135**, 044101 (2011).
- D. Ma, G. Li Manni, J. Olsen, and L. Gagliardi, *J. Chem. Theory Comput.* **12**, 3208 (2016).
- G. Karlström, R. Lindh, P.-Å. Malmqvist, B. O. Roos, U. Ryde, V. Veryazov, P.-O. Widmark, M. Cossi, B. Schimmelpfennig, P. Neogrady, and L. Seijo, *Comput. Mater. Sci.* **28**, 222 (2003).
- V. Veryazov, P.-O. Widmark, L. Serrano-Andrés, R. Lindh, and B. O. Roos, *Int. J. Quantum Chem.* **100**, 626 (2004).
- F. Aquilante, L. De Vico, N. Ferré, G. Ghigo, P.-Å. Malmqvist, P. Neogrady, 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).
- F. Aquilante, J. Autschbach, R. K. Carlson, L. F. Chibotaru, M. G. Delcey, L. De Vico, I. Fdez. Galván, N. Ferré, L. M. Frutos, L. Gagliardi, M. Garavelli, A. Giussani, C. E. Hoyer, G. Li Manni, H. Lischka, D. Ma, P. Å. Malmqvist, T. Müller, A. Nenov, M. Olivucci, T. B. Pedersen, D. Peng, F. Plasser, B. Pritchard, M. Reiher, I. Rivalta, I. Schapiro, J. Segarra-Martí, M. Stenrup, D. G. Truhlar, L. Ungur, A. Valentini, S. Vancocillie, V. Veryazov, V. P. Vysotskiy, O. Weingart, F. Zapata, and R. Lindh, *J. Comput. Chem.* **37**, 506 (2016).
- J. Miralles, J.-P. Daudey, and R. Caballol, *Chem. Phys. Lett.* **198**, 555 (1992).
- J. Miralles, O. Castell, R. Caballol, and J.-P. Malrieu, *Chem. Phys.* **172**, 33 (1993).
- D. Kats and H.-J. Werner, *J. Chem. Phys.* **150**, 214107 (2019).
- Y. Guo, K. Sivalingam, E. F. Valeev, and F. Neese, *J. Chem. Phys.* **144**, 094111 (2016).
- O. Demel, J. Pittner, and F. Neese, *J. Chem. Theory Comput.* **11**, 3104 (2015).
- T. S. Chwee, A. B. Szilva, R. Lindh, and E. A. Carter, *J. Chem. Phys.* **128**, 224106 (2008).

- <sup>53</sup>D. Walter, A. B. Szilva, K. Niedfeldt, and E. A. Carter, *J. Chem. Phys.* **117**, 1982 (2002).
- <sup>54</sup>T. S. Chwee and E. A. Carter, *J. Chem. Theory Comput.* **7**, 103 (2011).
- <sup>55</sup>TURBOMOLE, a development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, 2012.
- <sup>56</sup>F. Furche, R. Ahlrichs, C. Hättig, W. Klopper, M. Sierka, and F. Weigend, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **4**, 91 (2014).
- <sup>57</sup>F. Weigend, A. Köhn, and C. Hättig, *J. Chem. Phys.* **116**, 3175 (2002).
- <sup>58</sup>T. H. Dunning, *J. Chem. Phys.* **90**, 1007 (1989).
- <sup>59</sup>C. L. Janssen and I. M. B. Nielsen, *Chem. Phys. Lett.* **290**, 423 (1998).
- <sup>60</sup>I. M. B. Nielsen and C. L. Janssen, *Chem. Phys. Lett.* **310**, 568 (1999).
- <sup>61</sup>X. Assfeld, T. Etienne, A. Monari, and T. Véry, Nancy-EX, <https://sourceforge.net/projects/nancyex/>.
- <sup>62</sup>T. Etienne, X. Assfeld, and A. Monari, *J. Chem. Theory Comput.* **10**, 3896 (2014).
- <sup>63</sup>T. Etienne, X. Assfeld, and A. Monari, *J. Chem. Theory Comput.* **10**, 3906 (2014).
- <sup>64</sup>K. D. Vogiatzis, G. Li Manni, S. J. Stoneburner, D. Ma, and L. Gagliardi, *J. Chem. Theory Comput.* **11**, 3010 (2015).
- <sup>65</sup>M. S. Gordon and M. W. Schmidt, *Theory and Applications of Computational Chemistry* (Elsevier, 2005), pp. 1167–1189.
- <sup>66</sup>M. W. Schmidt, K. K. Baldridge, J. A. Boatz, S. T. Elbert, M. S. Gordon, J. H. Jensen, S. Koseki, N. Matsunaga, K. A. Nguyen, S. Su, T. L. Windus, M. Dupuis, and J. A. Montgomery, *J. Comput. Chem.* **14**, 1347 (1993).
- <sup>67</sup>G. M. J. Barca, C. Bertoni, L. Carrington, D. Datta, N. De Silva, J. E. Deustua, D. G. Fedorov, J. R. Gour, A. O. Gunina, E. Guidez, T. Harville, S. Irlé, J. Ivanic, K. Kowalski, S. S. Leang, H. Li, W. Li, J. J. Lutz, I. Magoulas, J. Mato, V. Mironov, H. Nakata, B. Q. Pham, P. Piecuch, D. Poole, S. R. Pruitt, A. P. Rendell, L. B. Roskop, K. Ruedenberg, T. Sattasathuchana, M. W. Schmidt, J. Shen, L. Slipchenko, M. Sosonkina, V. Sundriyal, A. Tiwari, J. L. Galvez Vallejo, B. Westheimer, M. Włoch, P. Xu, F. Zahariev, and M. S. Gordon, *J. Chem. Phys.* **152**, 154102 (2020).
- <sup>68</sup>D. Maynau, S. Evangelisti, N. Guihéry, C. J. Calzado, and J.-P. Malrieu, *J. Chem. Phys.* **116**, 10060 (2002).
- <sup>69</sup>D. Maynau, DoLo, a development of Laboratoire de Chimie et Physique Quantiques de Toulouse, [https://git.irsamc.ups-tlse.fr/LCPQ/Cost\\_package](https://git.irsamc.ups-tlse.fr/LCPQ/Cost_package).
- <sup>70</sup>J. Pipek and P. G. Mezey, *J. Chem. Phys.* **90**, 4916 (1989).
- <sup>71</sup>F. Aquilante, T. B. Pedersen, and R. Lindh, *J. Chem. Phys.* **126**, 194106 (2007).
- <sup>72</sup>F. Aquilante, P.-Å. Malmqvist, T. B. Pedersen, A. Ghosh, and B. O. Roos, *J. Chem. Theory Comput.* **4**, 694 (2008).
- <sup>73</sup>B. O. Roos, K. Andersson, M. P. Fülscher, L. Serrano-Andrés, K. Pierloot, M. Merchán, and V. Molina, *J. Mol. Struct.: THEOCHEM* **388**, 257 (1996).
- <sup>74</sup>N. Forsberg and P.-Å. Malmqvist, *Chem. Phys. Lett.* **274**, 196 (1997).
- <sup>75</sup>G. Ghigo, B. O. Roos, and P.-Å. Malmqvist, *Chem. Phys. Lett.* **396**, 142 (2004).
- <sup>76</sup>S. R. Langhoff and E. R. Davidson, *Int. J. Quantum Chem.* **8**, 61 (1974).
- <sup>77</sup>N. Ben Amor, B. Bories, S. Hoyau, and D. Maynau, EXSCI, [https://git.irsamc.ups-tlse.fr/LCPQ/Cost\\_package](https://git.irsamc.ups-tlse.fr/LCPQ/Cost_package).
- <sup>78</sup>J. P. Zobel, J. J. Nogueira, and L. González, *Chem. Sci.* **8**, 1482 (2017).