



HAL
open science

Atomic pseudopotentials for reproducing π -orbital electron behavior in sp² carbon atoms

Alexander Punter, Paola Nava, Yannick Carissan

► To cite this version:

Alexander Punter, Paola Nava, Yannick Carissan. Atomic pseudopotentials for reproducing π -orbital electron behavior in sp² carbon atoms. *International Journal of Quantum Chemistry*, 2019, pp.e25914. 10.1002/qua.25914 . hal-02048996

HAL Id: hal-02048996

<https://amu.hal.science/hal-02048996>

Submitted on 26 Feb 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.



Distributed under a Creative Commons Attribution 4.0 International License

Atomic pseudo-potentials for reproducing π -orbital electron behaviour in sp^2 carbon atoms

Alexander Punter, Paola Nava, Yannick Carissan *

January 14, 2019

Abstract

A pseudo-potential system for an sp^2 carbon atom is built and tested as a building block for various pseudo-hydrocarbon polyenes and polycyclic aromatic hydrocarbons. This pseudo-system has a central charge of $Z = 1$, it contains only one electron. It is employed in *ab-initio* calculations in which several physical characteristics including the orbital energies and first ionisation energy, as well as first excitation energy and UV spectra, are found to be well-reproduced by the pseudo-system. Remarkably, not only are the π excitation energies in good agreement with the reference calculations, but also transition densities and intensities, confirming that the virtual space obtained with the pseudo-potentials is of excellent quality. Finally, this approach is capable of reproducing the π electron systems of small or large, planar or near-planar hydrocarbons at low computational cost.

Keywords: Pseudo-potentials, *sp²carbon*, PAH, aromaticity, π system. ■

*Aix-Marseille Université, CNRS, Centrale Marseille, iSm2, Marseille, France

INTRODUCTION

Since Hellmann and Gombás' development of the first pseudo-potentials in the 1930s,^{1,2} pseudo-potential methods have been employed in a variety of theoretical chemistry problems. In recent years, pseudo-potentials have been used to provide enhancements or corrections to other methods, with examples including their use as link atoms in QM/MM calculations,³⁻²⁴ or the efforts to use them alongside DFT calculations in correcting for DFT errors in the effect of van der Waals forces.²⁵ Their primary use, however, has always been that of increasing computational efficiency as compared with all-electron solutions for the same systems. Most well-known among these methods is the Effective Core Potential (ECP),²⁶ where the replacement of core electrons in larger atoms with ECPs and RECPs (Relativistic Effective Core Potentials) permits calculations on molecules involving such atoms to be more efficient. These ECPs may themselves be separated into two further categories: the pseudo-potential, in which the radial nodal electronic structure is simplified, and the model potential, which aims to preserve the correct nodal structure of the valence orbitals.²⁷

This is not to say, however, that pseudo-potentials have been employed only in the treatment of heavy atoms. There have been efforts to create potentials for use on light atoms from the early work of Topiol *et al.* or Gresh and Pullman^{28,29} to the libraries spanning the whole periodic table (see the review by Dolg³⁰ and references therein) which are still under active development.³¹

In all of these methods, the technique has revolved around separating the electrons of an atom into core and valence electrons. We propose a different formulation. Our goal is to investigate the feasibility of pseudo-potentials reproducing electronic behaviour accurately, having been designed not only to replace the core electrons of light atoms, but also to replace specific valence electrons that do not take part in the electronic behaviour in which we are interested. Such potentials could find ready use in performing QM calculations on large hydrocarbons and polycyclic aromatic hydrocarbons.

It is a common idea that a chemical system can be thought of as comprised of (at least) two parts: an active one, where most of the chemistry takes place, and an inactive one, both of which must be taken into account in order to fulfil chemical requirements. Based on this

general statement, many successful theoretical approaches have been developed. Among them can be cited QM/QM' or QM/MM methods,²³ fragmentation methods,^{21,32,33} and frozen core and frozen density embedding methods.^{4,15,34,35} Of particular interest to us is the model potential method, which has seen several iterations as it has advanced, including ab-initio and relativistic formulations.³⁶⁻³⁸

In 1991 Huzinaga emphasised that the ‘dormant’ electrons of the inactive region had three main effects upon the active electrons, which any pseudo-potentials should maintain; these were the Coulomb and exchange forces, as well as a ‘no-collapse’ term.³⁹

Following his proposal, for an atom where active and dormant electrons have been separated, the Hamiltonian reads:

$$\hat{H} = \sum_{i=1}^n \hat{h}(i) + \sum_{i<j} \frac{1}{r_{ij}} \quad (1)$$

with $\frac{1}{r_{ij}}$ being the bi-electronic interaction between explicitly treated active electrons, and the mono-electronic operator:

$$\hat{h}(i) = -\frac{1}{2}\Delta_i - \frac{(Z - Z_c)}{r_i} + \hat{V} + \hat{\sigma} \quad (2)$$

where Δ_i is the Laplacian of the coordinates of electron i , and Z_C is the number of dormant electrons withdrawn from the reference system. The operator $\hat{\sigma}$ is the ‘no-collapse’ term that prevents active electrons collapsing into the dormant region; the operator \hat{V} reproduces the Coulomb and exchange interactions, for which Huzinaga proposed a spectral formulation:

$$\hat{V} = r^{-1} \left[\sum_I A_I \exp(-\alpha_I r^2) + \sum_J B_J r \exp(-\beta_J r^2) \right] \quad (3)$$

In this article, we perform the active/inactive split on a molecule with π electrons: the σ electrons comprising the inactive part and the π electrons the active part. The Schrödinger equation is solved for the latter part only in the field of the dormant part. The aim is to propose a method for extracting a potential for a hybridised carbon atom (sp^2), with a focus on reproducing basic properties of the electronic structure of the sp^2 carbon in a molecular framework (HOMO energy, ionisation potential and excitation energies). Our potentials can then be used as building blocks for constructing chemical systems, as several carbon pseudo-atoms, each containing one explicit nuclear charge and one electron, can be connected to

create extended systems. Moreover, the method should be usable ‘out of the box’, in any standard quantum chemistry software and no modification of the source code should be done.

This article is structured as follows. In the first part, the method is defined in detail. The second part shows the results obtained using the most refined version of the method; the extracted pseudo-potentials are then employed for extended systems, with results and discussion. A final perspective is presented in the conclusion.

METHODOLOGY

Choice of pseudo-potential criteria

Previous pseudo-potential methods have made use of different optimisation criteria, based on the pseudo-potentials’ intended function. The pseudo-bond method, for example, is intended to facilitate geometry-optimisation in QM/MM calculations, and uses a series of reference data, including bond lengths and angles, across a training set of different molecules.⁸ In our case, the aim is to simulate the electronic behaviour of π systems. Figure 1 helps to explain our choice of criteria for gauging the efficacy of our potentials. Since our reference system is the planar CH_3^\bullet radical (with the four atoms in the (xy) plane), the HOMO energy of CH_3^\bullet , $\epsilon_{\text{HOMO}(\text{CH}_3)}$, i.e. the p_z orbital energy, is an obvious choice, and should ensure the orbital of the pseudo-system is well-positioned to interact with other chemical moieties.

We would then take our CH_3^\bullet pseudo-potentials and create from them a pseudo-ethylene, the smallest π system. In addition to the HOMO energy of the pseudo-ethylene, we decided we would also compare the $\pi - \pi^*$ singlet-triplet energy gap, and the 1st ionisation energy. The reproduction of the Δ_{ST} ensures that the first virtual orbitals are well-positioned, which is an indicator of the quality of the system for the reproduction of spectroscopic and chemical interactions, whilst the reproduction of the ionisation potential shows the extent to which the pseudo-system can still behave like the reference system upon losing an electron. As discussed hereafter, this latter property is particularly critical, as all σ electrons are replaced by the pseudo-potentials, and therefore they are not explicitly treated.

Pseudo-potential definition

The potentials were extracted on the basis of Hartree-Fock calculations, as it is unbiased in terms of electronic correlation. As in our previous work,⁴⁰ we take the planar CH_3^\bullet as our starting system to be reproduced. It is the smallest system containing one and only one sp^2 -hybridised carbon atom. This choice allows us to isolate the different electrostatic interactions to build a physically meaningful model. We make use of two kinds of Gaussian pseudo-potentials,⁴¹ of s and p shapes. As we want to avoid modifying the quantum chemistry software itself, the pseudo-potentials have a semi-local form and we choose to preserve the r^{-1} behaviour as in Huzinaga’s work. The mono-electronic operator is written as:

$$\hat{h}(i) = -\frac{1}{2}\Delta_i - \frac{(Z - Z_c)}{r_i} + \hat{W} \quad (4)$$

The multi-centered pseudo-potentials that we build read as follows:

$$\hat{W} = r^{-1} \left[\underbrace{A \exp(-\alpha r^2) \sum_m |Y_{1,m}\rangle \langle Y_{1,m}|}_{p \text{ projectors}} + \underbrace{\sum_J C_J r \exp(-\gamma_J (r - r_J^0)^2) |Y_{0,0}\rangle \langle Y_{0,0}|}_{s \text{ projectors}} \right] \quad (5)$$

with $Y_{0,0}$ the s spherical harmonic, $Y_{1,m}$ the p spherical harmonics and r_J^0 the relative fixed position of the J^{th} potential with respect to the origin of the pseudo-atom, which carries the pseudo-potentials. By analogy with Huzinaga model potentials defined in (3) we can say that the p projectors aim to mimic a coulombic interaction (see \hat{V} in Equation 3), while the s potentials have the task of recovering part of the bi-electronic interaction between the dormant and the active electrons, explaining their position *around* the pseudo-atom. Thus, this second term is not strictly equivalent to Huzinaga’s B_J terms as our s potentials aim at recovering the bi-electronic interactions *and* part of the no-collapse term. This later role was served by the $\hat{\sigma}$ operator in Huzinaga’s formulation.

Figure 2 displays the final pseudo-system: The CH_3^\bullet radical has been replaced by a hydrogen-like ‘pseudo-carbon’, with a nuclear charge of $Z_{\text{nucleus}} = 1$, and one electron occupying the p_z orbital. There are now no H atoms, and the system is surrounded by three potential sets at a planar distance of d , each consisting of two s -shaped potentials with a

distance above and below the xy plane of c . A further p -shaped potential is applied directly to the pseudo-carbon.

This gives multiple variables we can use to manipulate the properties of the system. We can alter the strength and diffuseness of the p and s potentials themselves, as well as vary the distances d and c by moving the s -potentials. In the effort to keep the effects of the s potentials as localised as possible, we chose $d = 0.5$ a.u. and $c = 0.25$ a.u., making the atom-atom distance approximately the shortest permitted in QM programs.

Table 1 displays the data resulting from the potential extraction process. Initially, we optimised potentials for the CH_3^\bullet system (initial pseudo). The next step was to take two pseudo- CH_3^\bullet systems to create a pseudo-ethylene. This meant a refinement of the exponents and coefficients (final pseudo). In the final version of the extraction, the pseudo-ethylene system was: (1) Optimised from several different starting guesses to give the correct Δ_{ST} value. (2) The pseudo-potentials from step (1) that also gave the best HOMO and 1st ionisation energies were chosen as the starting guess for a final optimisation, performed by hand, to give us the final set of pseudo-potentials used in the rest of this work. The final potential set gives an error of 5.1% for the HOMO energy of CH_3^\bullet , but it performs much better than the initial for the ethylene, giving us a 0.0% error on the Δ_{ST} , and errors of 2.9% and 7.8% for the HOMO and 1st ionisation energies respectively. As the σ electrons are not treated explicitly in our model, the effect of their reorganisation in the cation can not be recovered entirely and we expect an error in the ionisation energy that is inherent in the method. For comparison, the all-electron energy of the $(\text{C}_2\text{H}_4)^+$ cation was computed, by freezing all the orbitals but the SOMO, as they were obtained from the Hartree-Fock calculation on the neutral C_2H_4 (def2-SVP basis set). The deviation from the all-electron fully relaxed calculation is of 0.946 eV, which is comparable, and slightly higher, than the error obtained with our potentials (error of 0.715 eV).

The potentials were extracted on the basis of Hartree-Fock calculations, but they are transferable to other methods and they are employed within the DFT and TD-DFT frameworks in the application section.

COMPUTATIONAL DETAILS

All Hartree-Fock, DFT and time-dependent DFT (TD-DFT) energy calculations are performed with TURBOMOLE 7.1.⁴² The basis set used throughout is def-SV(P).⁴³ Wherever possible, planar (C_S) symmetry is used. The convergence energy is $10^{-7}E_h$ (`$scfconv = 7`) for SCF and $10^{-6}E_h$ for DFT. All optimisations used the Brent method in SciPy’s optimisation library, with a tolerance of $1.48 * 10^{-08}$, and used standard Hartree-Fock calculations.⁴⁴

Structures are optimised for the ground state at the all-electron level for each functional and for Hartree-Fock, and the pseudo-systems are built by preserving the optimised carbon-carbon distances. The structures obtained for the ground state are then used for the calculations on the triplet states and the cations.

All CASPT2 calculations and calculations with frozen orbitals were performed with the MOLCAS program package.⁴⁵ The PBE0 structures were employed for the CASPT2 calculations.

All-trans-polyenes and Polycyclic Aromatic Hydrocarbons. In addition to Hartree-Fock calculations, DFT is used with PBE0, PBE, TPSS and TPSSh functionals.⁴⁶⁻⁴⁹ The integration grid size is set at *m4*. Also used are TD-DFT calculations, where the Tamm-Dancoff approximation (CIS) is switched on to avoid triplet instability.⁵⁰

Topological analysis of the density. The topological analysis was performed using the Multiwfn program on a wfn file obtained from the TURBOMOLE program.⁵¹

RESULTS AND DISCUSSION

APPLICATIONS

All-trans-polyenes

Taking the final potentials, we test them against a series of all-trans-polyenes up to length $C_{100}H_{102}$. In order to check how the pseudo-potentials behave with different functionals, we performed a survey of four of them: one GGA (PBE), one meta-GGA (TPSS), and their hybrid counterparts PBE0 and TPSSh.⁴⁶⁻⁴⁹ Table 2 gives a breakdown of the percentage

errors for each method across all polyenes tested.

The non-hybrid PBE functional leads to acceptable results and the addition of exact exchange improves the overall agreement. Indeed, the best agreement is found with the PBE0 functional for which the quality of the results does not degrade with the size of the chain. TPSS and TPSSh display larger errors, up to 20% in the energy of the HOMO, which impacts the ionisation potential. However, TD-DFT results are very good.

Figures 3 and 4 show in detail the results obtained with the PBE0 functional for the small and large polyenes, respectively. For the small polyenes, in the pseudo-systems, the HOMO has systematically a slightly lower energy than in the reference, but absolute errors are small (between 0.18 and 0.3 eV) and the pattern of increasing HOMO energies with the size of the chain is well-replicated. As a consequence of this underestimation, ionisation energies are slightly higher than in reference calculations, even if tendencies are well-reproduced. We recall here that this quantity is particularly challenging for our method, as we do not account for the reorganisation of the σ electrons in the cation. For comparison, the mean ionisation errors obtained with our pseudo-potentials are between 7% (PBE0) and 11% (TPSS) and an average error of 10% is found in ROHF calculations for the small-polyenes where all the σ doubly-occupied orbitals are frozen in the cations, as they are obtained in the calculations for the neutral systems (see the Supplementary Information for details). More interestingly, a very good agreement between reference and pseudo-systems is obtained for TD-DFT calculations (for the $\pi - \pi^*$ excitation), which suggests that the virtual space is of good quality. We also compare TD-DFT results for this system and results of a previous work of some of the authors, which match to within 3% (see SI).⁴⁰ Let us recall here that in our previous work potentials were placed at the center of bonds, meaning that the pseudo-potentials generated, whilst effective, were time-consuming to set up for each molecule. For the smaller polyenes, the Δ_{ST} values were also computed and a good agreement is achieved (see SI).

Figure 4 refers to longer alkene chains (n up to 100). As for the smaller chains, the properties of large all-trans-polyenes are well-reproduced, which indicates that the results do not degrade with the size of the system. In the larger polyenes spin contamination is non-negligible, both in the reference and the pseudo-potential calculations, and the Δ_{ST} approach

is not viable as it is for small polyenes (see SI). Thus, the computation of the excited state has to be done through TD-DFT, which leads to an excellent agreement between all-electron and pseudo-potential calculations.

Polycyclic aromatic hydrocarbons

The potentials derived above are also tested on polycyclic aromatic hydrocarbons (PAH) in order to check if they are able to reproduce the effect of aromaticity. This is particularly critical as the system used for the extraction, ethylene, has no such feature.

Figure 5 shows the excitation energies, the IE and HOMO energy values for several PAH. As with the polyenes, the general trend of the results is well-replicated by the pseudo-systems, and the percentage errors, displayed in Table 3, are similar. This is not surprising because aromaticity arises from the fact that the p orbitals overlap and are form a ring. Since we reproduce the p orbitals well, the aromatic character of the π ring appears.

Encouraged by the fact that low-lying first excited states as well as the ionisation potentials of a large range of PAH are well-reproduced by our pseudo-potentials, we decided to explore the virtual space further by computing the 20 first singlet π excited states, and to compare the results to all-electron calculations using the PBE0 functional. Absorption spectra with pseudo-potential and all-electron calculations for a selection of PAH are shown in Figure 6. As the pseudo-potential calculations contain only π orbitals, the excitations from σ to σ orbitals cannot be reproduced (these are denoted by a * above the peaks in the graphs). The overall agreement is very satisfactory. Excitation energies computed with the pseudo-potential are systematically too high, a red shift of 0.91 was used on the computed excited energies to match the reference spectra for the singlet. This is a very pleasing surprise owing to the fact that the pseudo-potentials used to obtain these results were extracted on the first excited state of ethylene. Furthermore, the analysis of the transition densities shows that the physics of the excitations is reproduced as well (see Figure 7 and Table 4). The same analysis done on the triplet excited states is available in the SI, which leads to the same excellent agreement. The most remarkable feature to us is the neatness of the agreement of the relative excitation intensities in both singlet and triplet cases. This certainly shows again that the pseudo-carbon we have extracted captures a lot of the physics of the reference

system as the intensities are related to the transition dipole moment.

Finally, our pseudo-potentials are tested within the CASPT2 framework, by computing the first singlet-singlet and singlet-triplet excitation energies for benzene, naphthalene and anthracene. For both reference and pseudo calculations, the active spaces are (6,6), (10,10), and (10,10), respectively for benzene, naphthalene and anthracene. Results are reported and compared to the TD-DFT data in Figure 8: trends are well-reproduced and errors with respect to the all-electron calculations are of a similar magnitude to those of the TD-DFT calculations.

Topological analysis of the density

A Bader analysis can be performed on the electronic density obtained with the pseudo-potentials.⁵² This gives a vision of the π -only electronic cloud. In the AIM framework, critical points are characterised by the sign of the eigenvalues of the Hessian of the electronic density and are classified as follows: as the maxima of the electronic density are located at the nuclear positions, atom critical points are characterised by three negative eigenvalues, noted (3, -3); a bond critical point is a saddle point, showing a minimum along the bond direction, thus noted (3, -1); cycle critical points are minima on the hyper-surface of the cycle and maxima along the direction perpendicular to the cycle, noted (3, +1). A last kind of critical point (cage critical point) is characterised by three positive eigenvalues (3, +3). The critical paths are obtained by connecting the critical points along a direction of zero gradient.

For benzene, we compare the critical points of the π cloud obtained in a reference calculation with pseudo-potentials at the same level of theory. The π system of the all-electrons calculation is obtained by populating the converged π molecular orbitals only. As the AIM procedure deals with the electronic density, no reoptimisation of the orbitals is necessary. In our case, as the critical point analyses were obtained with the π system only, the electronic density maxima are not located at the nuclei but rather at the maxima of the π density, *i.e.* above and below the nuclei. Accordingly, twelve bond critical points rather than six are expected, above and below the molecular plane. Moreover, due to the topology of the

density, two cycle critical points are found above and below the molecular plane on the C_6 symmetry axis. To fulfil the Poincaré-Hopf relationship, a cage critical point is required. However, this cage critical point at the center of the molecule could not be found, either in the reference or in the pseudo-potential calculations. After a careful inspection of the results, it turns out that the electronic density is below computer accuracy at this point (it is evaluated to be around 10^{-31}). Therefore we deduce, there is a cage critical point at the center of the molecule but it can not be detected.

Finally, Figure 9 shows that the description of the bond pattern by the pseudo-potential calculation corresponds to that by the reference, in terms of types of found critical points. The position of the critical points are not exactly the same and the pseudo-potential calculation leads to a slightly more diffuse π bonding pattern. However, as both images have the same scale, one can see that the topologies of the electronic densities are very similar. This confirms that the results obtained are not artefactual but that the pseudo-potentials lead to a physically relevant electronic density.

Helicene Spectra

We were interested to look at some still larger and more complex molecules. In particular, we were curious to see how the potentials would perform in a system that wasn't planar, as distortions in the plane would, in an all-electron system, alter the influence of the sp^2 -hybridised electrons.

In a comparison between CC2 theoretical calculations and experiment, Nakai *et al.* show that ECD spectra of helicenes can be accurately computed.⁵³ First of all let us show here that for small helicenes (6 to 9 rings), it is possible to reproduce with good accuracy the experimental spectra with DFT (PBE0/def2-SV(P)).

In this work, we are interested in carbon based helicenes made of ortho-fused-six-member-rings only. The agreement between experiment, CC2 and DFT calculations is very good according to Figure ?? (see SI). Relative intensities, sign and positions of the absorption peaks are remarkably well reproduced by the DFT calculations with no shift on the wavelengths. The ECD spectra of M helicenes feature a positive absorption at low energy followed by a

negative absorption about 100 nm further (the spectrum of ch[8] is plotted with a thicker line to emphasise this feature). This characteristic behavior (in the above defined framework) is indicative of the chirality M or P of the helicene. As a reminder, when the helix followed by a helicene rotates in the counter-clockwise direction, the helicene is labelled P (plus) and M (minus) otherwise.

Figure 10 displays the UV and ECD spectra for a series of [n]helicene molecules of $n = 6$ to $n = 9$ (structural diagrams are available in the SI), along with their pseudo-molecular counterparts. After shifting the pseudo-spectra by 32nm (a redshift), we can see the general shapes of the pseudo-system UV spectra are qualitatively very similar to the all-electron spectra, both in their spread and their intensities. We can also see that many of the peaks of the all-electron spectra appear to be clearly identifiable in the pseudo-spectra.

The ECD spectra are more complex. The general shapes of the pseudo-spectra are distorted compared to the all-electron results, with the distortion increasing from slight to more severe as the helicene becomes longer. We suggest that the reason for this is that with $n > 6$, the ends of the helicene overlap more and more, resulting in a much more complex electronic structure. However, one can still clearly identify the largest peaks for the smaller systems, and the direction of the dichroism remains clear throughout.

These spectra show that we can be confident that our potentials can retain much of the physics of more complex π -systems, even allowing for some distortion to the molecular plane. These calculations show that the ECD spectra of helicenes are mainly due to the π -like electrons and that our pseudo-potentials allow for the reproduction of properties which are much more difficult to reproduce than UV spectra as they are much more sensitive to the environment.

Coronene encapsulated in a carbon nanotube

In a recent work, Nakamura *et al.* have studied the photoexcitation spectrum of a coronene polymer encapsulated in a carbon nanotube.⁵⁴ In this work, they show that two absorption bands (at around 1.7 and 3.4 eV) lead to a charge transfer from the coronene polymer to the nanotube. This charge transfer process is backed up by DFT calculations in the plane wave

formalism using boundary conditions. These calculations do not allow for the computation of excited states. Instead, the nature of the transition is assumed from the calculation of band structure of a model system, which is a one (19,0) nanotube subset of the whole system. We tried to investigate this problem with our pseudo-potentials in the cluster approximation to see if we could find discrete electronic excitations, which would confirm a charge transfer between the coronene and the nanotube character.

An emphasis should be made of the fact that we are trying to use the pseudo-potentials out of their scope as they were extracted for planar molecules only. However, if the curvature radii are not too small, we are confident enough that our pseudo-potentials could still lead to physically relevant results. It is to be noted here that the curvature of the nanotube leads to the loss of strict σ and π definition of the electronic cloud. We shall refer to an orbital as π -like on a carbon atom when it exhibits one lobe above and one lobe below the plane made by the three nearest neighbours of the atom considered. We propose that the π -like electrons are mainly responsible for the photoelectronic properties of the system at low energy, otherwise, we would not be able to obtain any sensible result with our method.

The comparison between experiment and such a cluster approach is risky as the system under study is a very crude approximation of the experimental system. Experimentally, Nakamura *et al.* make measurements over a variety of nanotubes (their diameter ranges from 1.0 to 1.5 nm), each encapsulating one coronene. We could study a model system of the experimental system using a coronene molecule encapsulated in a nanotube slice as can be seen in Figure 11. We chose a non-metallic chiral nanotube of an intermediate diameter as the ones used experimentally to be as close as possible to the experimental system. The calculation of the excited states of this model system is not routine at the TD-DFT level as it would require the calculation of a huge number of excited states, probably some thousands, as one must include both the σ and π excitations. Because of the reduction of the system when using our pseudo-potentials (only π -like electrons treated explicitly), we could do such a calculation in the π -like-only space. We performed the calculation of the photoabsorption spectrum with TD-DFT from the ground state to the first 200 excited states (see Supplementary Information). The computation of these 200 excitation energies leads to a maximum energy of roughly 3.6 eV. A peak appears in the region of interest of

the spectrum around 1.7 eV and two excitations (at 748 and 770 nm) have large oscillator strengths. The transition densities of these excitations are shown in Figure 12. The excitation with the higher oscillator strength at 748 nm shows electronic depletion and increase on both the coronene and the nanotube. This rearrangement is the reason for the large oscillator strength but cannot be assigned to a charge transfer process. On the other hand, the second excitation shows a charge transfer from the coronene to the nanotube (only blue π density on the coronene and the red π density is only on the tube). Such a result confirms the assignment made by Nakamura *et al.*

These calculations lead to informative results for the electronic structure of large system, provided that the properties of interest involve electrons with a strong π character. They allow for the use of quantum chemistry methods on systems for which such calculations, even if strictly not impossible, would be computationally intensive to say the least.

Timings

The model that we develop in this work allows significant computational gains. To illustrate this, we provide in Table 5 a small study done on the $C_{50}H_{52}$ alkene chain. Calculations were performed with two different basis sets: def2-SV(P) and QZVPPD. When using the pseudo-potential, the basis set was truncated by removing all basis functions not necessary to reproduce the π system. In other words, only p functions or those of higher angular momentum were kept. As can be seen from Table 5, the gain increases with the size of the basis set as expected. The gain ranges from 2.4 (def2-SV(P)) to 8 (QZVPPD).

Physical meaning of the p_z pseudo-potential

In this paragraph we propose a physical interpretation of our pseudo-potentials. This provides a conceptual advantage with respect to our previous work,⁴⁰ where the pseudo-potentials were simply relying on a shifting of the orbitals to map the reference ones.

The pseudo-system contains only one proton. However, the effective charge felt by the p electron in a carbon atom should be larger: for instance Slater’s rules suggest an effective charge of $Z_{eff} = 2.4$. Thus we must account for this difference. We shall see that this role

is filled by the p potential once optimised.

In order to evaluate the screening effect in the real CH_3^\bullet system, we computed the expectation value of the distance of the electron from the nucleus $\langle r \rangle$:

$$\langle r \rangle = \langle \psi_{p_z} | r | \psi_{p_z} \rangle = 1.80 \text{ a.u.} \quad (6)$$

where ψ_{p_z} is the molecular orbital extracted from a reference calculation of CH_3^\bullet .

Since the CH_3^\bullet pseudo system has only one electron (as in a hydrogen-like atom), the analytical form of the p_z orbital is:⁴¹

$$\phi_{210} = \frac{1}{\sqrt{\pi}} \frac{Z_{eff}^{\frac{5}{2}}}{2a_0} r e^{-\frac{Z_{eff}r}{2a_0}} \cos \theta \quad (7)$$

where Z_{eff} is the total nuclear attraction the electron would feel in the real CH_3^\bullet system, taking into account the screening effect of the core electrons that would be present in the all-electron system. This leads to the following expression for $\langle r \rangle(Z_{eff})$;

$$\langle \phi_{210} | r | \phi_{210} \rangle = \frac{5a_0}{Z_{eff}} \quad (8)$$

From Equations 6 and 8, we obtain the theoretical value of $Z_{eff} = 2.77$.

The ψ_{p_z} molecular orbital is influenced strongly by the p pseudo-potential, because of their overlap. In this paragraph we use a simplified definition \hat{p} of the p pseudo-potential, which allows for a quick evaluation of overlap effects with the basis set:

$$\chi = e^{-\frac{\alpha}{2}r^2} \sum_{m=-1}^{+1} |Y_{1,m}\rangle, \quad \hat{p} = -A|\chi\rangle\langle\chi| \quad (9)$$

We define

$$S = \langle \psi_{p_z} | \chi \rangle \quad (10)$$

the overlap between a molecular orbital ψ_{p_z} , and χ , leading us to evaluate the effect of the pseudo-potential on the ψ_{p_z} molecular orbital as:

$$\langle \psi_{p_z} | \hat{p} | \psi_{p_z} \rangle = \langle \psi_{p_z} | -A|\chi\rangle\langle\chi| \psi_{p_z} \rangle = -AS^2 \quad (11)$$

Finally, knowing that our hydrogen-like pseudo-system already contains a charge, $Z_{nucleus} = 1$, we expect that:

$$A = (Z_{eff} - Z_{nucleus})S^{-2} \quad (12)$$

Ultimately, we used this expression to make informed guesses for the starting coefficients of the p functions, from which we optimised the pseudo-potential. By using Equation 12, and the optimised exponent of $\alpha = 0.624$, an estimated value of $A = 3.614$ is obtained. The optimised coefficient $A = 3.909$ of our final potential set is very close to this estimation, supporting the idea that the p potential retrieves the incomplete screening effect of the missing electrons.

CONCLUSION AND PERSPECTIVE

In this paper pseudo-potentials have been developed for sp^2 -hybridised carbon units containing only one nuclear charge and one electron. Their main features are that: (1) They are fully atomic, meaning that they can be placed relative only to the carbon atom itself and the direction of its bonds. (2) We could show that not only were the occupied orbitals well-reproduced by the use of these new potentials, but also that the virtual space is of good quality for excited states calculation. They are able to reproduce physically reasonable absorption spectra and the electronic density leads to the same AIM view of the bonding pattern. Ionisation energies have also been tested and the expected trends are obtained, even if clearly our model can not recover the effects of the reorganisation of the σ electrons in the cation. However, deviations from the all-electron calculations decrease with the size of the system. (3) They can be employed with both Hartree-Fock and DFT methods including PBE0, PBE, TPSS and TPSSh functionals. CASPT2 calculations have also been positively tested. (4) They have been shown to function well whether their neighbouring atoms are other pseudo-carbons (as with fused carbon rings), hydrogen atoms (as in CH_3^\bullet), or a mixture of both (as in the polyenes or at the edge of fused ring systems). (5) They offer a significant gain in computational time as compared to all-electron calculations. (6) They were employed to compute UV and ECD spectra of extended, near-planar systems: helicenes and a coronene embedded in a nanotube. Despite the fact that our method does not reach chemical accuracy in ionisation energies, it can be useful to solve highly difficult assignments for large systems in a limited time.

Our work remains ongoing, and we envisage two areas for further development. (1) We are

at the early stages of creating potentials that can successfully interact with neighbouring all-electron atoms. This will mean producing potentials which are capable of bonding covalently with their all-electron neighbours, *i.e.* adding another electron to the pseudo-carbon for every bond. Our preliminary results (to be submitted in due course) show that optimisation of bond distances is possible. Such potentials could be useful as link atoms in QM/MM calculations, functioning similarly to Quantum Capping Potentials,⁹ though this is not our primary motivation.

(2) We would like further to test and to expand this work into a general method for increasing computational efficiency in large systems in which there are many sp^2 carbon units, both in SCF calculations and the generation of electronic spectra. This could be particularly useful in large molecules which are optically active to simulate ECD spectra. The challenge is to ensure the right occupation of the π -like orbitals despite the loss of the $\sigma - \pi$ separation.

ACKNOWLEDGMENTS

We are grateful to the French Research Ministry, Aix-Marseille University and CNRS for financial support.

References

1. H. Hellmann, *J. Chem. Phys.* **3** (1935).
2. P. Gombás, *Z. Phys.* **94**, 473 (1935).
3. U. C. Singh and P. A. Kollman, *J. Comput. Chem.* **7**, 718 (1986).
4. X. Assfeld and J.-L. Rivail, *Chem. Phys. Lett.* **263**, 100 (1996).
5. S. Goedecker, M. Teter, and J. Hutter, *Phys. Rev. B* **54**, 1703 (1996).
6. J. Gao, P. Amara, C. Alhambra, and M. J. Field, *J. Phys. Chem. A* **102**, 4714 (1998).
7. C. Hartwigsen, S. Goedecker, and J. Hutter, *Phys. Rev. B* **58**, 3641 (1998).
8. Y. Zhang, T.-S. Lee, and W. Yang, *J. Chem. Phys.* **110**, 46 (1998).
9. G. A. DiLabio, M. M. Hurley, and P. A. Christiansen, *J. Chem. Phys.* **116**, 9578 (2002).
10. O. A. von Lilienfeld, I. Tavernelli, U. Rothlisberger, and D. Sebastiani, *J. Chem. Phys.* **122**, 014113 (2004).
11. O. A. von Lilienfeld, I. Tavernelli, U. Rothlisberger, and D. Sebastiani, *Phys. Rev. Lett.* **93**, 153004 (2004).
12. Y. Zhang, *J. Chem. Phys.* **122**, 024114 (2004).
13. O. A. von Lilienfeld, I. Tavernelli, U. Rothlisberger, and D. Sebastiani, *Phys. Rev. B* **71**, 195119 (2005).
14. G. A. DiLabio, R. A. Wolkow, and E. R. Johnson, *J. Chem. Phys.* **122**, 044708 (2005).
15. C. R. Jacob and L. Visscher, *J. Chem. Phys.* **125**, 194104 (2006).
16. J. M. Parks, H. Hu, A. J. Cohen, and W. Yang, *J. Chem. Phys.* **129**, 154106 (2008).
17. J. L. Lewin and C. J. Cramer, *J. Phys. Chem. A* **112**, 12754 (2008).
18. A. C. Ihrig, C. Schiffmann, and D. Sebastiani, *J. Chem. Phys.* **135**, 214107 (2011).

19. O. A. von Lilienfeld, *Mol. Phys.* **111**, 2147 (2013).
20. M. Hitzenberger and T. S. Hofer, *J. Comput. Chem.* **36**, 1929 (2015).
21. M. A. Collins and R. P. A. Bettens, *Chem. Rev.* **115**, 5607 (2015).
22. S. Pezeshki and H. Lin, *Mol. Sim.* **41**, 168 (2015).
23. L. W. Chung, W. M. C. Sameera, R. Ramozzi, A. J. Page, M. Hatanaka, G. P. Petrova, T. V. Harris, X. Li, Z. Ke, F. Liu, et al., *Chem. Rev.* **115**, 5678 (2015).
24. M. Hitzenberger, M. Ratanasak, V. Parasuk, and T. S. Hofer, *Theor. Chem. Acc.* **135**, 1 (2016).
25. G. A. DiLabio, *Chem. Phys. Lett.* **455**, 348 (2008).
26. M. Dolg, *NIC Series* **1**, 479 (2000).
27. M. Dolg and X. Cao, *WIREs Comput Mol Sci* **1**, 200–210 (2011).
28. S. Topiol, A. A. Frost, M. A. Ratner, and J. W. Moskowitz, *J. Chem. Phys.* **65**, 4467 (1976).
29. N. Gresh and A. Pullman, *Theoret. Chim. Acta.* **49**, 283 (1978).
30. M. Dolg and X. Cao, *Chem. Rev.* **112**, 403 (2012).
31. W. Zou, Z. Cai, J. Wang, and K. Xin, *J. Comput. Chem.* **39**, 1697 (2018).
32. M. S. Gordon, M. A. Freitag, P. Bandyopadhyay, J. H. Jensen, V. Kairys, and W. J. Stevens, *J. Phys. Chem. A* **105**, 293 (2001).
33. C. Steinmann, D. G. Fedorov, and J. H. Jensen, *PLoS ONE* **7** (2012).
34. A. N. Artemyev, A. D. Müller, D. Hochstuhl, and P. V. Demekhin, *J. Chem. Phys.* **142**, 244105 (2015).
35. T. A. Wesolowski, S. Shedge, and X. Zhou, *Chem. Rev.* **115**, 5891 (2015).

36. V. Bonifacic and S. Huzinaga, *J. Chem. Phys.* **86**, 2132 (1987).
37. S. Huzinaga, L. Seijo, Z. Barandiarán, and M. Klobukowski, *J. Chem. Phys.* **60**, 2779 (1974).
38. J. Seino, M. Tarumi, and H. Nakai, *Chem. Phys. Lett.* **592**, 341 (2014).
39. S. Huzinaga, *J. Mol. Struct. THEOCHEM* **234**, 51 (1991).
40. J. Drujon and Y. Carissan, *J. Comput. Chem.* **34**, 49 (2013).
41. T. Helgaker, P. Jørgensen, and J. Olsen, *Molecular Electronic-Structure Theory* (John Wiley and Sons Ltd, Baffins Lane, Chichester, West Sussex PO19 1UD, England, 2002).
42. *TURBOMOLE V7.1 2016* (a development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, 1989-2007, TURBOMOLE GmbH, since 2007; available from <http://www.turbomole.com>., 2016).
43. A. Schäfer, H. Horn, and R. Ahlrichs, *J. Chem. Phys.* **97** (1992).
44. E. Jones, T. Oliphant, and P. Peterson, *Scipy: Open source scientific tools for python* (2001), URL <http://www.scipy.org/>.
45. F. Aquilante, J. Autschbach, R. K. Carlson, L. F. Chibotaru, M. G. Delcey, L. De Vico, I. F. Galvan, N. Ferre, L. M. Frutos, L. Gagliardi, et al., *J. Comp. Chem.* **37**, 506 (2016).
46. J. Perdew, M. Ernzerhof, and K. Burke, *J. Chem. Phys.* **105** (1996).
47. J. Perdew, K. Burke, and M. Ernzerhof, *Phys. Rev. Lett.* **77** (1996).
48. J. Tao, J. Perdew, V. Staroverov, and G. Scuseria, *Phys. Rev. Lett.* **91** (2003).
49. V. Staroverov, G. Scuseria, J. Tao, and J. Perdew, *J. Chem. Phys.* **119** (2003).
50. M. J. G. Peach, M. J. Williamson, and D. J. Tozer, *J. Chem. Theory Comput.* **7** (2011).
51. T. Lu and F. Chen, *J. Comput. Chem.* **33**, 580 (2012).

52. R. F. W. Bader, *Atoms in Molecules: A Quantum Theory*. (Oxford University Press., 1994).
53. Y. Nakai, T. Mori, and Y. Inoue, *J. Phys. Chem. A* **116**, 7372 (2012), ISSN 1089-5639, URL <https://doi.org/10.1021/jp304576g>.
54. A. Nakamura, K.-i. Yamanaka, K. Miyaura, H. E. Lim, K. Matsuda, B. Thendie, Y. Miyata, T. Kochi, S. Okada, and H. Shinohara, *The Journal of Physical Chemistry C* **122**, 16940 (2018).

Figure 1: Definition of the criteria chosen to optimise the pseudo-potentials (IP: ionisation potential, Δ_{ST} : singlet-triplet excitation energy, ϵ_{HOMO} : energy of the HOMO orbital). IP allow the inclusion of the relaxation of the orbitals when the system is ionised, whereas ϵ_{HOMO} is more important in a chemical framework.

Figure 2: Diagrams of the planar CH_3^\bullet radical (left) and the pseudo- C^\bullet (right) used to mimic the p_z orbital. The mono occupied p_z orbital is represented in grey. The pseudo- C^\bullet diagram displays the s and p -potentials in blue, and the distances d and c .

Figure 3: DFT and TD-DFT (PBE0) comparison of all-electron and pseudo-system energies across a range of polyenes.

Figure 4: DFT and TD-DFT (PBE0) comparison of all-electron and pseudo-system energies across a range of long polyenes (C_{20} - C_{100}).

Figure 5: DFT and TD-DFT (PBE0) comparison of all-electron and pseudo-system energies across a range of Polycyclic Aromatic Hydrocarbons. $\text{C}_{54}\text{H}_{18}$ is shown in the bottom.

Figure 6: Comparison of the absorption spectra the first 20 excitations of PAH obtained with pseudo-potentials (blue) and all-electron (red) calculations (def2-SV(P)/TD-PBE0). Peaks marked with * cannot be reproduced by the pseudo-potentials as they are σ to σ excitations. Excitation energies obtained with the pseudo-potentials were corrected by a factor 0.91.

Figure 7: Transition densities of the first two (lowest-energy) pyrene absorption peaks, for singlet-singlet excitation spectra, as seen in Figure 6, with all-electron spectra (left) and pseudo-molecule spectra (right). In each transition, the electron density is decreased in the blue-coloured lobes and increased in the red.

Figure 8: TD-DFT (PBE0) and CASPT2 comparison of all-electron and pseudo-system first singlet-singlet and singlet-triplet transition energies for benzene, naphthalene and anthracene. The active spaces are (6,6), (10,10), and (10,10), respectively for benzene, naphthalene and anthracene.

Figure 9: Comparison of the bond critical points and bond critical paths obtained at the PBE0/def2-SV(P) level for the reference π system of benzene (left) and the π system obtained with the final pseudo-potentials (right). Colours indicate the nature of the critical points: cyan for (3, -3), blue for (3, 1) and red for (3, -1). The critical paths are drawn in yellow.

Figure 10: Comparison of the absorption and ECD spectra of the first 200 excitations of a series of helicene molecules obtained with pseudo-potentials (blue) and all-electron (red) calculations (def2-SV(P)/TD-PBE0). Excitation energies obtained with the pseudo-potentials were corrected by a shift of 32nm.

Figure 11: Model system of a coronene-polymer-encapsulating single wall carbon nanotube. It is made of a slice (1,14) armchair nanotube of length 1.10 nm and of diameter 1.25 nm encapsulating a coronene molecule. Hydrogen atoms were removed from the representation.

Figure 12: Transition densities for the excitation at 770 nm (left) and 747 nm (right) at the TD-DFT(PBE0) level with the pseudo-potentials developed in this work. Blue indicates a diminution of electronic density and red indicates an increase of electronic density.

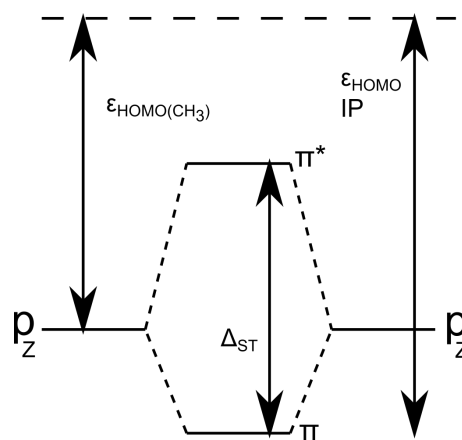


Figure 1
Alexander Punter, Paola Nava,
Yannick Carissan
Int. J. Quantum Chem.

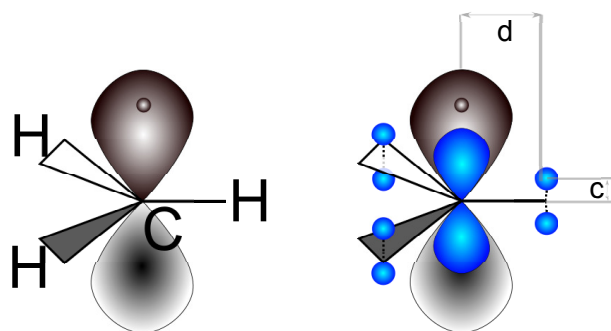


Figure 2
Alexander Punter, Paola Nava,
Yannick Carissan
Int. J. Quantum Chem.

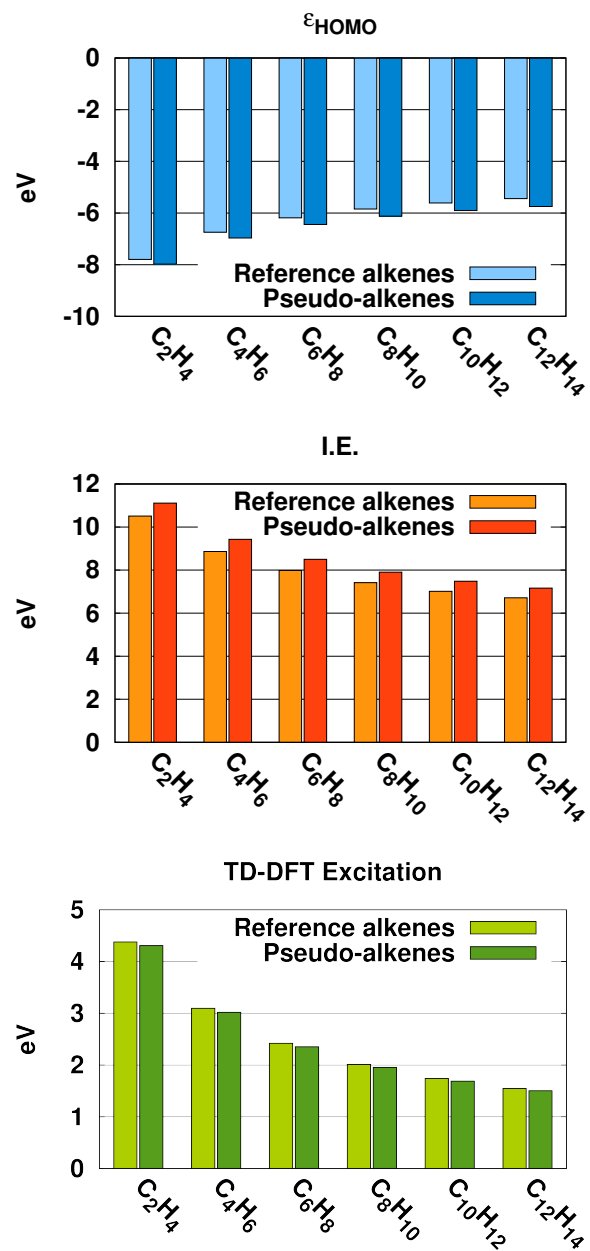


Figure 3
 Alexander Punter, Paola Nava,
 Yannick Carissan
 Int. J. Quantum Chem.

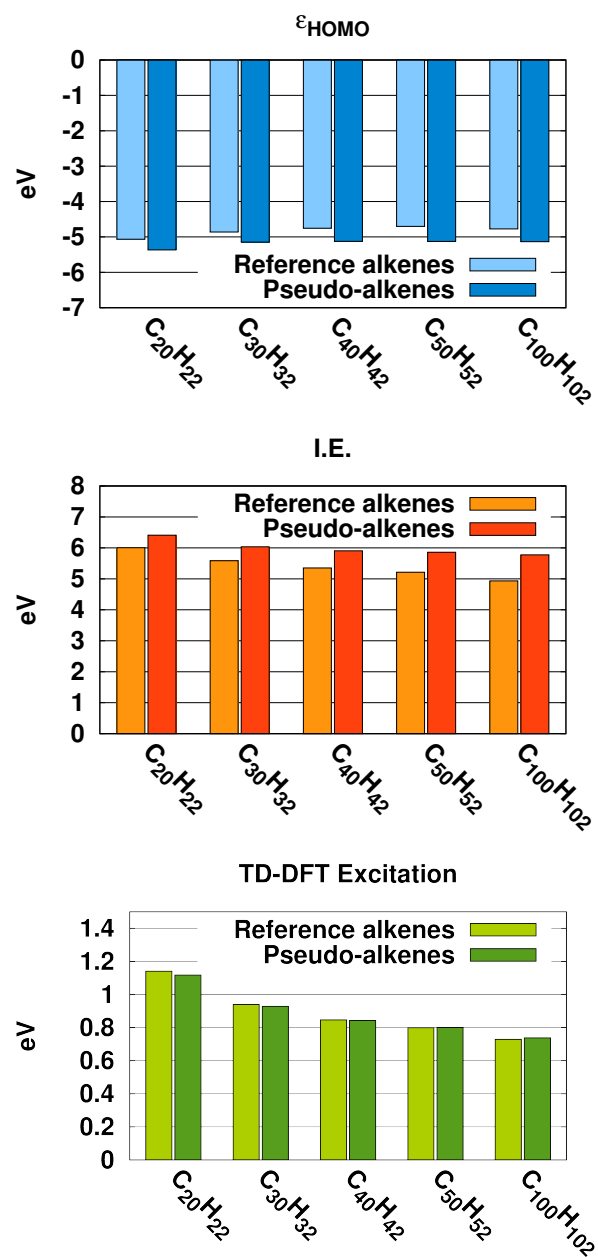


Figure 4
 Alexander Punter, Paola Nava,
 Yannick Carissan
 Int. J. Quantum Chem.

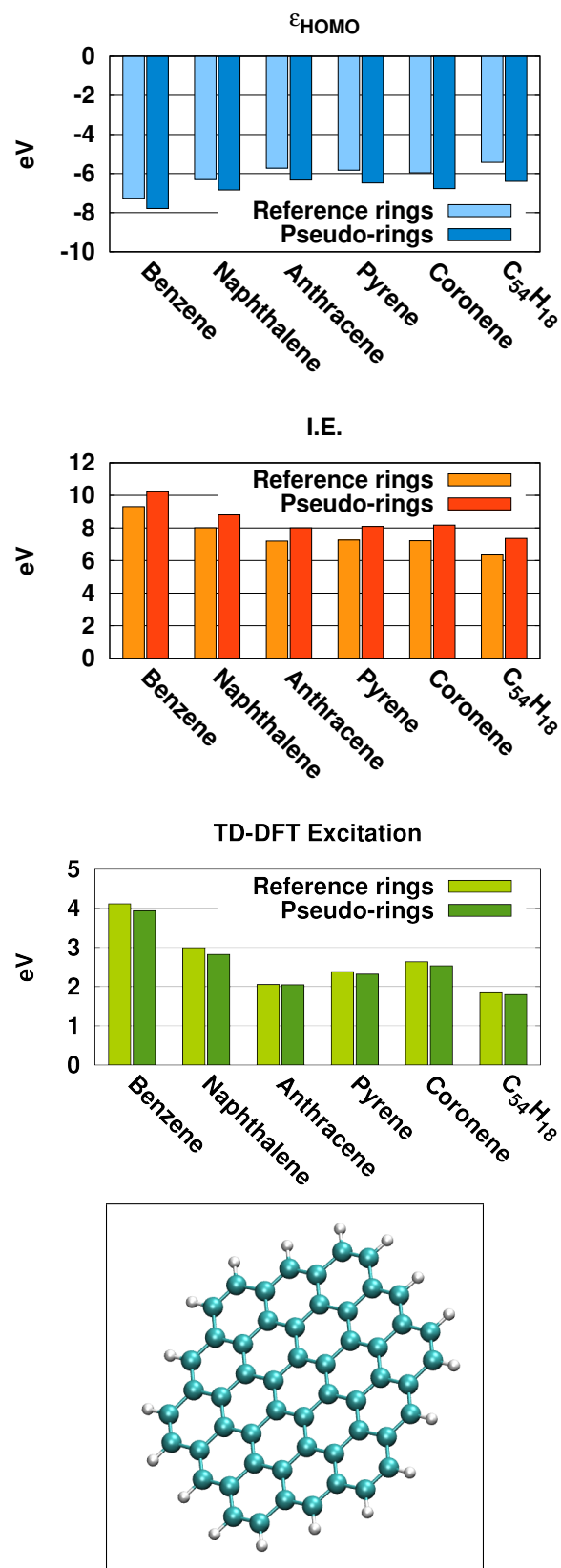


Figure 5
 Alexander Punter, Paola Nava, 28
 Yannick Carissan
 Int. J. Quantum Chem.

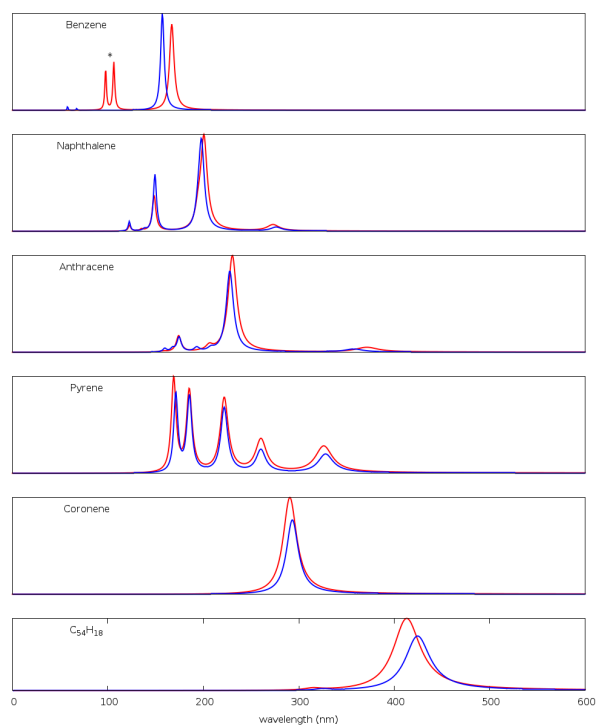


Figure 6
Alexander Punter, Paola Nava,
Yannick Carissan
Int. J. Quantum Chem.

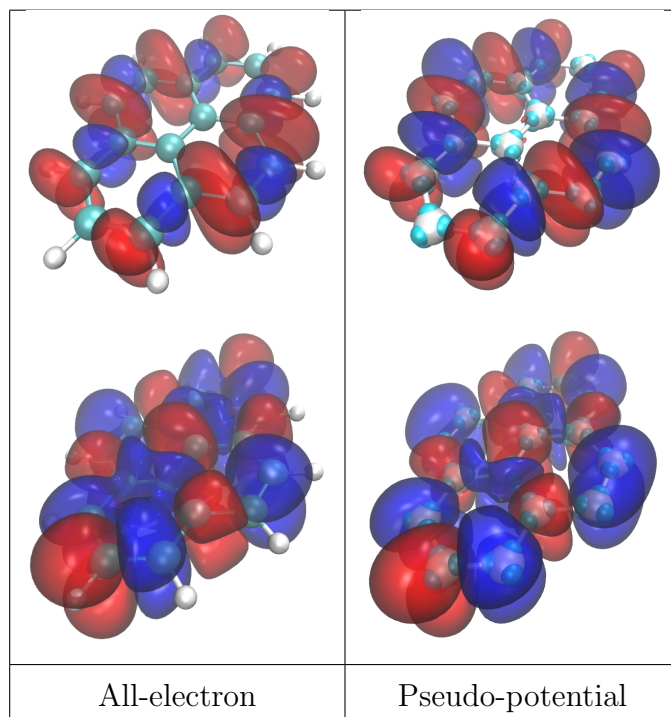


Figure 7
Alexander Punter, Paola Nava,
Yannick Carissan
Int. J. Quantum Chem.

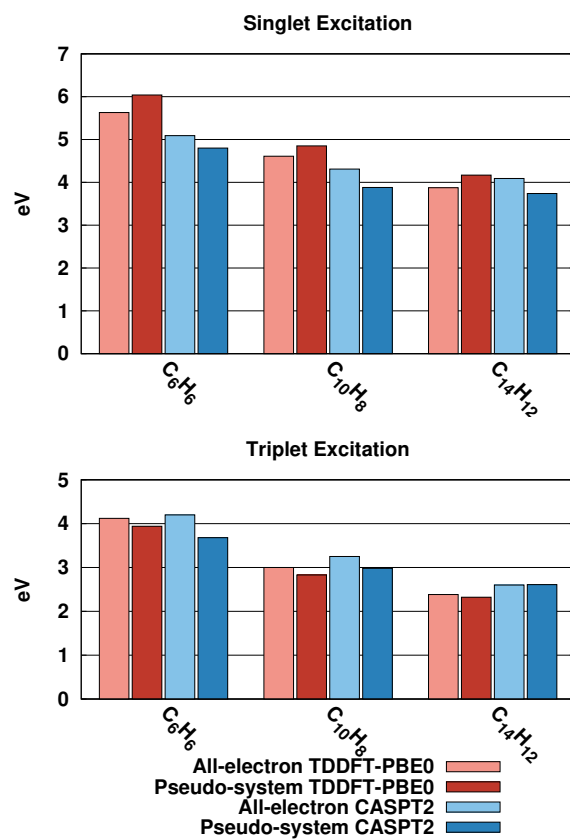


Figure 8
 Alexander Punter, Paola Nava,
 Yannick Carissan
 Int. J. Quantum Chem.

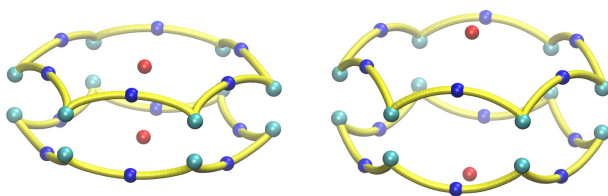


Figure 9
Alexander Punter, Paola Nava,
Yannick Carissan
Int. J. Quantum Chem.

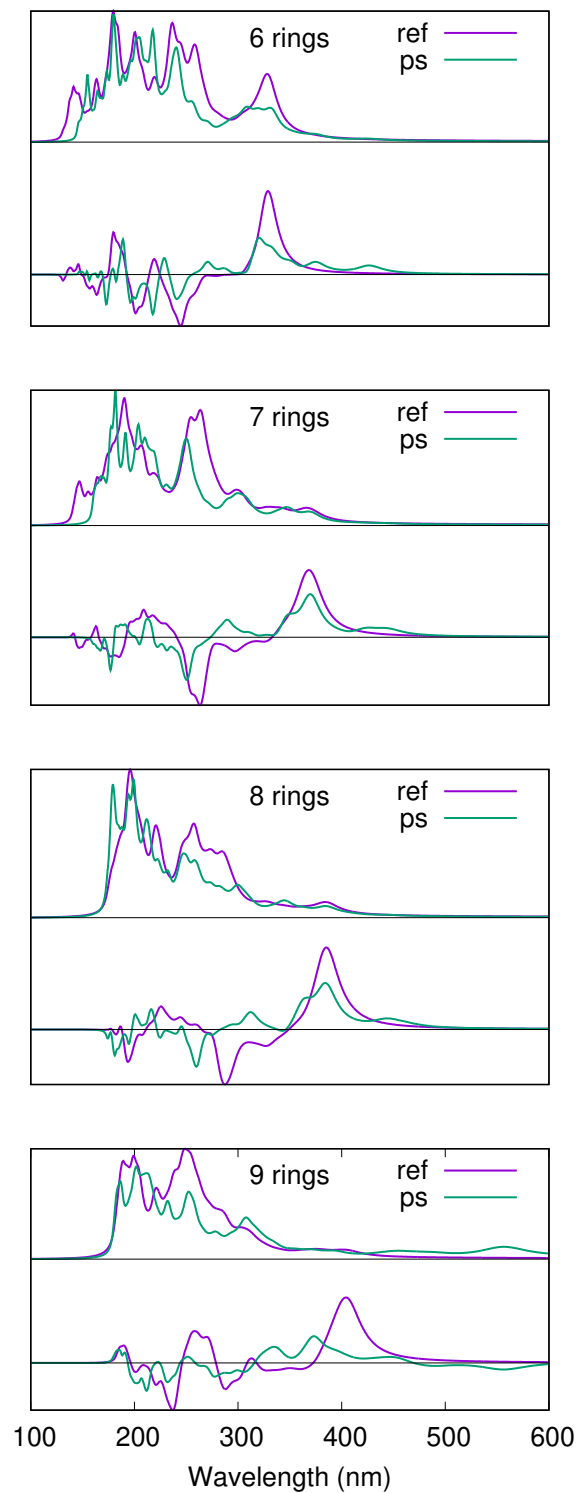


Figure 10
 Alexander Punter, Paola Nava,
 Yannick Carissan
 Int. J. Quantum Chem.

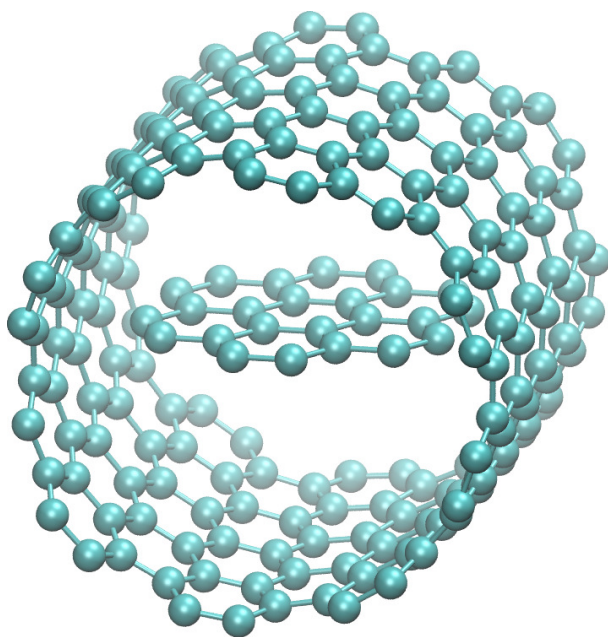


Figure 11
Alexander Punter, Paola Nava,
Yannick Carissan
Int. J. Quantum Chem.

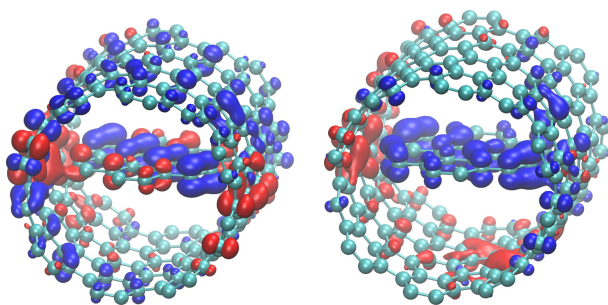


Figure 12
Alexander Punter, Paola Nava,
Yannick Carissan
Int. J. Quantum Chem.

Table 1: Comparison of molecular properties of pseudo-CH₃[•] and pseudo-C₂H₄ molecules, along with pseudo-potential parameters and all-electron values (Hartee-Fock level).

System	<i>p</i>		<i>s</i>		Energy (eV)		
	coeff.	exp	coeff.	exp.	HOMO	Δ_{ST}	1 st Ionisation
ref. CH ₃ [•]	-	-	-	-	-10.537	-	-
ref. C ₂ H ₄	-	-	-	-	-10.363	-3.522	9.091
initial pseudo-CH ₃ [•]	-3.268	0.295	10.381	10.000	-10.537	-	-
initial pseudo-C ₂ H ₄	-3.268	0.295	10.381	10.000	-14.061	-9.214	14.021
final pseudo-CH ₃ [•]	-3.910	0.624	1.500	0.500	-10.000	-	-
final pseudo-C ₂ H ₄	-3.910	0.624	1.500	0.500	-10.062	-3.533	9.806

Table 2: Errors across calculation types for all-trans-polyenes C_nH_{n+2}

Calculation Type	PBE0	PBE	TPSS	TPSSh
Short polyenes (n=2,4,6,8,10)				
Mean Ionisation error (%)	7.0	8.8	11.3	9.9
Mean Singlet HOMO error (%)	4.2	9.3	13.7	10.7
Mean TD-DFT Excitation error (%)	2.6	2.7	5.9	6.4
Long polyenes (n=20, 30, 40, 50, 100)				
Mean Ionisation error (%)	6.6	10.3	11.3	11.2
Mean Singlet HOMO error (%)	7.8	12.8	20.1	15.7
Mean TD-DFT Excitation error (%)	1.0	6.8	6.1	2.9

Table 3: Errors across calculation types for PAH

Calculation Type	PBE0	PBE	TPSS	TPSSh
Mean Ionisation error (%)	12.0	15.4	18.0	16.1
Mean Singlet HOMO error (%)	11.5	17.1	22.3	18.7
Mean TD-DFT Excitation error (%)	4.4	2.6	5.9	6.5

Table 4: Comparison of the weights (all-electron *vs.* pseudo-potentials) of the excitations obtained with TD-DFT to represent the triplet excited state from the closed shell singlet state. Example case of $C_{50}H_{52}$.

Excitation MO	Weight(%)	
	Ref.	Pseudo.
25 a" \rightarrow 26 a"	77.0	67.1
24 a" \rightarrow 27 a"	10.5	13.1
23 a" \rightarrow 28 a"	3.6	5.2

Table 5: Time (in seconds) per SCF iteration and relative gain (time ratio all-electron/pseudo-potential) for a standard and a large basis set.

Model	def2-SV(P)	QZVPPD
All electrons	67	68928
Pseudo-potential	28	8566
Gain (no unit)	2.4	8.0