



HAL
open science

Pseudopotential-fragment spectroscopy for organic molecules and carbon allotropes

Alexander Punter, Paola Nava, Yannick Carissan

► **To cite this version:**

Alexander Punter, Paola Nava, Yannick Carissan. Pseudopotential-fragment spectroscopy for organic molecules and carbon allotropes. *International Journal of Quantum Chemistry*, 2020, 120 (11), 10.1002/qua.26180 . hal-02980288

HAL Id: hal-02980288

<https://hal.science/hal-02980288>

Submitted on 27 Oct 2020

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.

Pseudopotential-Fragment Spectroscopy for Organic Molecules and Carbon Allotropes

Alexander Punter, Paola Nava, and Yannick Carissan Aix Marseille Univ, CNRS, Centrale Marseille, iSm2, Marseille, France

Following on from a previous work (Punter et al., IJQC 2019, 119, 23), pseudopotential sets are developed and tested for a variety of sp^2 and sp^3 carbon fragments. These fragments contain only one or two explicit protons and electrons, and make use of non-atom-centred potentials. They are tested with Density Functional Theory calculations in a selection of chemical environments in which several physical characteristics, including orbital and first ionisation energies, are found to be well-reproduced. They are then employed in the reproduction of molecular absorption spectra for large organic molecules and carbon allotropes, and are found to recreate both absorption and electronic circular dichroism spectra to a high accuracy. They are also found significantly to increase the computational efficiency of TDDFT calculations in which they are used.

I. INTRODUCTION

In most pseudopotential methods, the techniques have revolved around separating the electrons of an atom into core and valence electrons, and replacing the core. This remains true whether they are plane-wave or molecular pseudopotentials, whether they are shape or energy consistent, whether they are norm-conserving or ultrasoft, or indeed if they follow the alternate model core potential formulation of Huzinaga.²⁻⁸ There are several excellent reviews of these methods.⁹⁻¹¹

In our formulation, the goal is to reproduce electronic behaviour accurately, while replacing not only the core electrons of small molecular fragments, but also to replace specific valence electrons that do not take part in the electronic behaviour in which we are interested, as well as any unnecessary nuclei and protons. In a recent paper¹, we created a set of pseudopotentials designed to emulate an sp^2 -hybridised carbon fragment, eliminating all but one electron and one proton from the whole fragment.

In this paper we describe the most recent developments of the method. We begin in Section II by recalling briefly the defining features of our pseudopotential method, before setting out the new additions made since the previous work. In Section III we derive and test some new pseudopotentials across a series of simple but varied molecules. This involves creating new pseudofragment geometries for different hybridisations of carbon some of which, unlike the originals, have no σ - π separation. In Section IV, we derive and test a new pseudopotential with the aim of reproducing the absorption and electronic circular dichroism (ECD) spectra of π systems. Then, in Section V, we take the pseudopotentials created in previous sections and test them in several much larger and more complicated molecules, with a view to seeing how well the pseudopotentials are able to reproduce the spectroscopy of each. Some final remarks and thoughts are then presented in the conclusion.

II. METHODOLOGY

A. Pseudopotential Form

We restrict ourselves to an ‘out of the box’ approach, where no modification of quantum chemistry package source code should be necessary. This was done so as to keep the potentials independent of the quantum chemistry package used. The main quantum chemistry package we use is Turbomole.¹² Our potentials are of the standard form

$$\sum_k A_k (r^{n_k-2}) e^{-\alpha_k r^2} \sum_l \sum_m |Y_{l,m}\rangle_k \langle Y_{l,m}|_k \quad (1)$$

where n_k is an integer and where the total number of expressions k can run across multiple such expressions for each angular momentum l .¹³ From Equation 1 we see that we have a choice of three variables per potential function, A_k , n_k and α_k . Throughout this work, the value of n_k is fixed at 1. We limit ourselves to a maximum of one potential function per angular momentum l for each potential centre. This is done only for the sake of simplicity and the possibility of increasing this limit is discussed in Section VII. There is however a well-known precedent for the main group elements being sufficiently well-described by a single potential function per angular momentum.²

We stated above that we remove whole atoms from the molecular fragments for which we create potentials, as well as some further protons. We also make use of non-atom-centred potentials. In full, the multi-centred pseudopotential operator can be written as

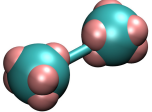
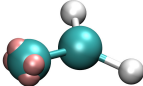
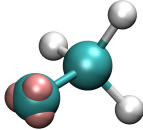
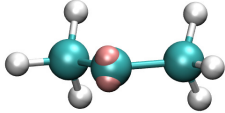
$$\hat{W} = \underbrace{\frac{A}{r} \exp(-\alpha r^2) \sum_m |Y_{1,m}\rangle \langle Y_{1,m}|}_{\text{atom-centred potentials}} + \underbrace{\sum_J \frac{C_J}{r - r_J^0} \exp(-\gamma_J (r - r_J^0)^2) |Y_{0,0}\rangle \langle Y_{0,0}|}_{\text{non-atom-centred potentials}} \quad (2)$$

with $Y_{0,0}$ the s spherical harmonic, $Y_{1,m}$ the p spherical harmonics (across all m) and r_J^0 the relative fixed position of the J^{th} potential with respect to the origin of the pseudoatom to which the potentials are assigned.

B. Pseudopotential Setups

In keeping with our philosophy described in Section II A of retaining only the chemically-relevant parts of the system, we designed three new potential set-ups that kept only the bonding atoms. These consisted of a further sp^2 carbon fragment, and two sp^3 carbon fragments, termed β , γ and δ potentials respectively. In the previous work, we designed a set of pseudopotentials for an sp^2 -hybridised carbon atom containing only one proton and one electron, to which we shall now refer as the α potential set-up, with the original potential parameters of the previous work referred to as the *set1* potentials. All four set-ups, α , β , γ and δ , are summarised in Table I.

Table I. A summary of different pseudopotential set-ups presented in this work. Non-atom-centred potentials are highlighted in red.

Diagram	Designation	Description
	α	sp^2 , 1 explicit electron
	β	sp^2 , 2 explicit electrons
	γ	sp^3 , 1 explicit electron
	δ	sp^3 , 2 explicit electrons

1. α pseudopotentials

It was shown in the previous work that α pseudopotentials could accurately reproduce absorption spectra for the excitations of the remaining electrons in the systems (as the pseudopotential calculations contain only π orbitals, excitations to or from σ orbitals cannot be reproduced). The pseudosystem spectra were however consistently shifted by a 30-40 nm as compared to the all-electron spectra.

Figure 1 displays a series of singlet excitations for a range of polycyclic aromatic hydrocarbon (PAH) systems

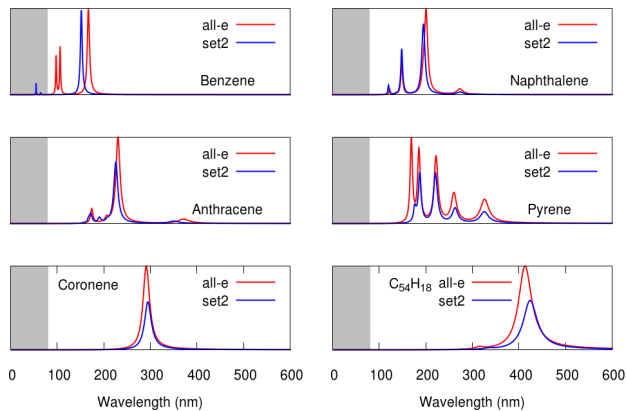


Figure 1. Comparison of the absorption spectra for the 20 first singlet excitations obtained with *set2* pseudopotentials and all-electron calculations (def2-SV(P)(s-less)/TD-PBE0) within the RPA framework. The grey zone covers energies for which no all-electron excitations have been calculated.

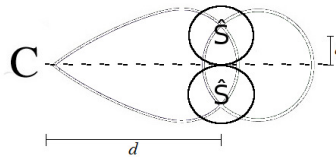


Figure 2. Definition of distances c and d for α and β pseudopotential set-ups

using a new set of α potentials, which we call *set2*, optimised specifically with the reproduction of UV spectra in mind. A detail should be noted in the benzene spectra. The all-electron spectrum contains two peaks at 100 nm , and the pseudomolecular spectrum contains two smaller peaks at 60 nm . The peaks in the all-electron spectrum represent excitations between σ orbitals, meaning that they cannot be reproduced by the pseudo-system. But since we instruct the chemistry package to calculate the same number of excitations in both systems, the pseudosystem produces two additional excitations at a higher energy. These excitations are not unphysical. This phenomenon is present in many of the pseudopotential spectra to follow. The grey zones are used to indicate the energy ranges for which no all-electron excitations are calculated.

The new pseudopotential parameters reported in Table II are not dissimilar to those of the original *set1* potentials. The coefficients and exponents are all within 0.5 of those of *set1*, and incorporating the distances d and c (see Figure 2) into the optimisation procedure has resulted in the non-atom-centred s potentials being moved by less than $0.1 a_0$ in both directions.

The significant difference between the potentials of the previous work and the *set2* potentials is in the optimisa-

Table II. List of optimisation criteria and results for the *set2* pseudopotential set. This is an α potential using ethylene as a reference. Column ‘Carbon l ’ shows the angular momenta of the atom-centred-potentials. The ‘Criteria’ column lists the reference values supplied to the program, as well as their weighting. The Mean Absolute Error (MAE) gives a normalised total error for the best result found by the minimisation procedure, and is the mean error of the values specified in the ‘Criteria’ column. c and d refer to distances in Figure 2.

Carbon l	Basis	Criteria	MAE (eV)	
p	def-SV(P)	HOMO; LUMO; LUMO+1	0.04037	
Potential	A ($H a_0$)	α (a_0^{-2})	d (a_0)	c (a_0)
p	-3.9020	0.6914	-	-
s	1.2266	0.5448	0.5821	0.2689

tion criteria. Rather than the HOMO, 1st ionisation and singlet-triplet gap energies chosen for the original potentials, the *set2* potentials are optimised on the HOMO energy, as well as the first two virtual orbitals. This includes the π^* orbital.

2. β , γ and δ pseudopotentials

In creating the β potentials, we wished to have explicit, all-electron systems interact successfully with neighbouring pseudosystems, meaning that the bonding electrons are required. The easiest way to do this from the original α potentials is to add one more electron and proton to the setup, allowing it to form both a σ and a π bond with a neighbouring atom. This means that one of the s potential sets can be removed, as these were intended to recapture the effect of σ electrons which, in the α potential setup, are no longer there. In the α and β set-ups, the planar distance to each potential set is denoted by d , and the distance above and below the xy plane by c . Figure 2 illustrates this.

The γ potential setup has a p -shaped potential on the central carbon, along with three s -shaped potentials replacing the hydrogen atoms at a distance from the central carbon of $0.5 a_0$.

In the δ pseudopotential setup there are two explicit electrons, thus it is able to act as a bridge between two other all-electron, bonding atoms. The central carbon not only has a p -shaped potential, but also its own s -shaped potential, and we have s potentials replacing two of the all-electron bonds at a distance from the pseudocarbon of $d=0.5 a_0$. In the γ and δ set-ups, the distance from central pseudocarbon to non-atom-centred potentials is denoted by d .

3. The optimisation code

We generalised the method of the previous work by creating a general minimisation program, the Multiple Orbital Optimiser (MOO), which minimises the error between all-electron and pseudosystems on selected criteria including molecular orbital energies, TDDFT excitation energies, UV spectra fitting and total energy differences, such as ionisation energies, singlet-triplet energy gaps, or even alternate geometries.

The program MOO can be used to alter the coefficients and exponents of the pseudopotentials, and the position of the non-atom-centred potentials. These optimisations are carried out at the Hartree-Fock level, but can be transferred to other levels of calculation. This code is publicly available¹⁴.

Table III. Optimisation criteria and parameters for the best β , γ and δ potential sets.

Optimisation Criteria				
Carbon l	Basis	Criteria	MAE (eV)	
β potentials				
p	def-SV(P)	HOMO($\times 3$); HOMO-1; HOMO-2	0.90670	
γ potentials				
p	def-SV(P)	HOMO($\times 2$); HOMO-1	0.4179	
δ potentials				
s, p	def-SV(P)	HOMO; HOMO-1; HOMO-2; HOMO-3; HOMO-4; HOMO-5	0.40067	
Optimised Parameters				
Potential	A ($H a_0$)	α (a_0^{-2})	d (a_0)	c (a_0)
β potentials				
p	-2.0031	0.4694	-	-
s	0.4376	0.4946	0.5	0.25
γ potentials				
p	-2.0949	0.3951	-	-
s	38.5052	14.8328	0.5	-
δ potentials				
p	-6.9569	3.4066	-	-
s	5.3934	6.4912	0.5	-
s (carbon)	0.6391	0.9382	-	-

Contained in Table III are the details of the β , γ and δ optimisations. Most of the various criteria on which the potentials are optimised can be written in terms of energy, which allows us to normalise the final MAE error by dividing it by the number of criteria (multiplied by any individual weighting applied to them). Various combinations of criteria were tested, and the potentials that produced the smallest overall errors were chosen.

For these β , γ and δ potentials, the aim was to reproduce bonding structure similarly to the α potentials of the previous paper. This meant reproducing the electronic density took priority over reproducing the virtual space excitations. These potentials are thus optimised on occupied orbitals of highest energies. The importance of the virtual orbitals in the reproduction of absorption spectra was not realised until later.

The β optimisation used the top three occupied orbitals in a closed-shell Hartree-Fock calculation on ethylene as a reference, with a $3\times$ weighting on the HOMO orbital, and the normalised total error is 0.9067 eV . Overall, this pseudopotential setup does not differ greatly from the α pseudopotentials, other than that these potentials are slightly weaker and more diffuse than the α potentials.

The γ optimisation was carried out on an eclipsed ethane molecule, using s and p central potentials, with the top two occupied orbitals as a reference (with the HOMO doubly-weighted). These s potentials are far more concentrated than those of either of the sp^2 potentials. In the relaxed, staggered geometry of ethane, the electrons at either end of the ethane molecule experience each others' influence less strongly. Therefore eclipsed ethane was chosen for this optimisation, in the hopes of creating a more accurate potential.

The δ potentials are optimised on propane and use both a p potential and an s potential on the central pseudocarbon, in addition to the non-atom-centred s potentials. The reference criteria used are the six highest-occupied molecular orbitals of an Unrestricted Hartree-Fock calculation.

In summary then, we have created and optimised four new pseudopotential setups, summarised in Table I, for different carbon fragments.

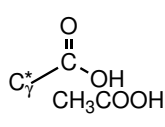
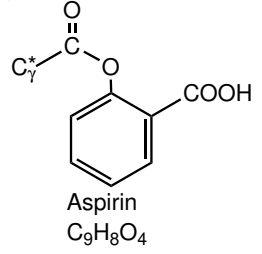
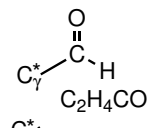
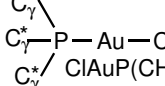
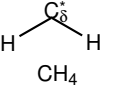
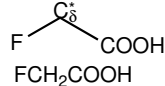
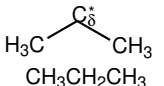
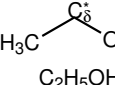
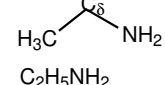
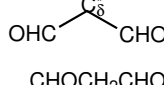
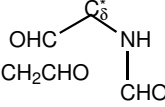
In each case, the optimisations were made on Hartree-Fock calculations, and this was done to keep the potentials correlation-independent. In the following sections, all calculations use DFT-PBE0/def-SV(P) unless otherwise specified. It was found possible to remove the s basis functions of the def-SV(P) *set2* pseudocarbon without altering any of the *set2* results. Hereafter in this work, the *set2* potentials use the def-SV(P) basis set without s functions.

III. RESULTS: SIMPLE SYSTEMS

The optimised potentials were tested across a sample of molecules chosen to expose the pseudopotentials to a range of different chemical environments. These molecules, along with the sites of the pseudopotentials, are shown in Table IV. In each case the properties measured are the HOMO, first ionisation, singlet-triplet gap and TDDFT first excitation energies.

The results are broken down by pseudofragment type in Figures 3, 4 and 5, with average differences in the various properties laid out in Table V. In most of these test

Table IV. Molecules used to test β , γ and δ pseudopotentials. The pseudofragments are denoted by C_β^* , C_γ^* and C_δ^* respectively.

β Test Molecules			
$C_\beta^* = \text{CH}_2$ C_2H_4	$C_\beta^* = \text{O}$ H_2CO	$C_\beta^* = \text{CHOH}$ CH_2CHOH	
$C_\beta^* = \text{NH}$ CH_2NH	$C_\beta^* = \text{CHNH}_2$ CH_2CHNH_2		
γ Test Molecules			
$C_\gamma^* - \text{H}$ CH_4	$C_\gamma^* - \text{CH}_3$ C_2H_6	$C_\gamma^* - \text{OH}$ CH_3OH	$C_\gamma^* - \text{NH}_2$ CH_3NH_2
eclipsed			
			
	$\text{C}_2\text{H}_4\text{CO}$	Aspirin $\text{C}_9\text{H}_8\text{O}_4$	
	Au-Cl $\text{ClAuP}(\text{CH}_3)_3$		
δ Test Molecules			
			
CH_4	FCH_2COOH	$\text{CH}_3\text{CH}_2\text{CH}_3$	
			
$\text{C}_2\text{H}_5\text{OH}$	$\text{C}_2\text{H}_5\text{NH}_2$	CHOCH_2CHO malonaldehyde	
			
	$\text{CHONH}_2\text{CH}_2\text{CHO}$		

molecules, the pseudocarbon is bound to an all-electron carbon, however among the test molecules are included several in which the pseudocarbon is bound to an all-electron heteroatom *i.e.* not a carbon. This was done to gauge how sensitive the potentials were to being placed in systems for which they were not optimised.

Looking across the results for sp^2 -hybridised carbon systems (*i.e.* the molecules containing β pseudofragments) shown in Figure 3 and Table V, we see that all molecules with homoatomic pseudofragment bonds have at most a difference of 5.9% between all-electron and pseudosystems. This is a good accuracy and is consistent with the results of the previous work. The differences between all-electron and pseudopotential systems for heteroatomic pseudofragment bonds however, are substantially larger, particularly in the case of a pseudofragment/oxygen bond. In the case of CH_2O , the molecular orbitals emerge in the wrong order energetically, includ-

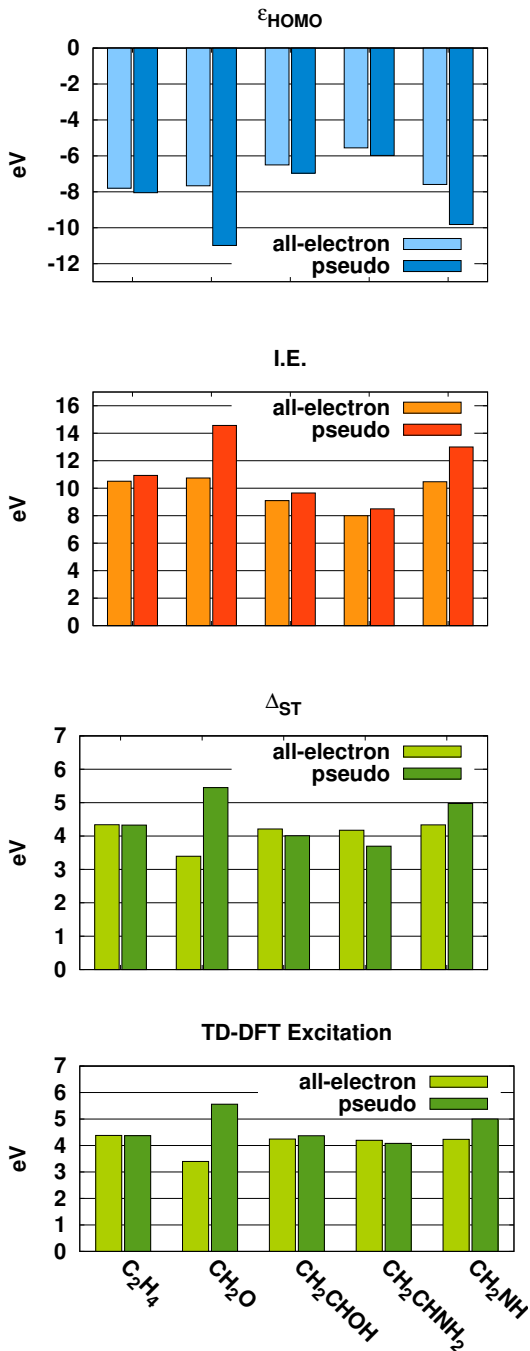


Figure 3. DFT-PBE0 and TDDFT-PBE0 comparison of all-electron and pseudosystem energies across a range of β potential systems.

ing the HOMO, explaining why its results are uniquely poor.

Turning to the sp^3 -hybrids (*i.e.* systems containing γ or δ pseudofragments) with their results displayed in Figures 4 and 5 and percentage differences shown again in Table V, it is seen that those systems with homoatomic

pseudofragment bonds have maximum differences between all-electron and pseudopotential systems of 5.3%, a similar result to the above. When considering the systems with heteroatomic pseudofragment bonds the differences are once again larger, particularly for pseudofragment/oxygen and pseudofragment/hydrogen bonds. δ pseudofragments allow for two bonds between all-electron atoms and the pseudofragment, and we see that the all-electron/pseudosystem differences in the one test molecule with two pseudofragment/heteroatom bonds, methane, are correspondingly large. In the particular case of the C^*_γ -P results, it is also worth noting that the pseudosystem is built with three separate sets of γ potentials, which would likely have compounded any error in the potentials themselves. One notes throughout though that the errors on sp^3 systems are much smaller than those of the sp^2 systems.

Table V. Average errors for molecules using β , γ and δ potentials, arranged by pseudocarbon-X bond type, for DFT-PBE0 and TDDFT-PBE0 calculations. ' Δ_{S-T} ' is the singlet-triplet gap, and '1st Ex' is the energy of the first TDDFT-PBE0 excitation.

Bond Type (C*-X)	Mean Error (%)			
	Δ_{S-T}	HOMO	1 st I.E.	1 st Ex
<i>sp</i> ² systems				
β pseudofragments				
All	18.4	18.1	15.2	17.5
C^*_β -C	5.5	5.9	5.5	2.0
C^*_β -O	60.6	43.3	35.6	63.7
C^*_β -N	14.7	29.4	24.1	18.1
<i>sp</i> ³ systems				
γ pseudofragments				
All	5.1	3.9	7.7	4.3
C^*_γ -C	5.3	2.8	1.7	4.6
C^*_γ -O	9.5	7.8	12.4	7.7
C^*_γ -N	3.1	3.8	4.9	2.9
C^*_γ -H	4.3	1.5	8.8	9.8
C^*_γ -P	9.0	2.3	1.3	13.7
δ pseudofragments				
All	8.9	3.1	2.3	10.5
H- C^*_δ -H	25.0	9.3	6.9	26.8
C- C^*_δ -C	5.3	0.2	0.9	2.8
C- C^*_δ -O	12.7	2.9	0.3	25.7
C- C^*_δ -N	5.5	2.5	1.1	5.9
C- C^*_δ -F	2.8	4.2	4.8	3.4

Overall, we deduce that the more polarised the bond between a pseudofragment and a heteroatom, the worse the result. This is to be expected, given that the pseudopotentials are optimised on carbon-carbon bonds, and given the different character of bonds between carbon atoms as compared to those between carbon and oxygen or nitrogen atoms. It is also found that the sp^2 β potentials are more sensitive to a change in the atom to which the pseudofragment is bound. Again this is not surprising, given that the β pseudofragment forms both a σ and a π bond to its all-electron neighbour. Meanwhile, carbon-bound pseudofragments performed consis-

tently well across all molecules tested and all properties measured (that is, the HOMO, first ionisation, singlet-triplet gap and TDDFT first excitation energies), with errors smaller than 6.0% in all cases. It can therefore be said that these pseudopotentials reproduce the basic properties of the test systems well, and to a degree consistent with the previous work. Particularly impressive to us is the accurate reproduction of the ionisation energies for so many of these systems, as most of the electron relaxation which takes place in all-electron systems after ionisation clearly cannot not happen in the pseudofragments.

IV. RESULTS: ABSORPTION SPECTRA

The *set2* pseudopotentials were tested for the same helicene molecules as in the previous work. The UV spectrum of the helicene molecules are improved, as the comparison for the [6]helicene reveals (Figure 6). One sees that the *set2* potentials do not require a shift in wavelength in order to line up the pseudosystem spectra with the all-electron spectra, which is an improvement on the previous work. Spectra for a range of pseudo[n]helicenes are included in the Supplementary Information (SI), along with an analysis of transition densities for excitations visible in the [10]helicene ECD spectrum.

These spectra show that we can be confident that our potentials can retain much of the physics of complex π systems, even allowing for some distortion of the molecular plane. Using the definition of dihedral torsion shown in Figure 7, [6]helicene has a maximum dihedral torsion-per-ring of 14.5° . These calculations show that the ECD spectra of helicenes are mainly due to the π -like electrons and that our pseudopotentials 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. We can also say that the *set2* pseudopotentials prove to be more accurate than the *set1* potentials for the reproduction of helicene spectra. Overall then, it can be said that the use of virtual orbitals as reference criteria for optimisation improves the UV spectra produced.

V. RESULTS: COMPLEX SYSTEMS

Thus far, all potentials have been tested on systems related closely to those for which they have been optimised. In this section, potentials are tested in more complex systems, selected from recent literature, involving metal atoms, distorted π systems, and neighbouring π rings.

A. Twistacene

The phrase ‘twisted acene’, later ‘twistacene’, was introduced in 2004¹⁵, then more thoroughly described in 2006¹⁶. The electronic and optical properties of their parent acenes are altered by this twisting¹⁷, which introduces chirality to the molecule¹⁸. A study by Bedi *et al.*¹⁸ describes a system allowing the ‘helical-locking’ of an acene at a specific torsion angle, and it is to this publication that we turn for geometric data. Figure 8 shows the Ant-*cn* molecule (where $n=3-6$ is the length of the carbon bridge at the top of the molecule), which allows the helical locking of anthracene. The increasing length n of the carbon bridge at the top of the molecule allows the anthracene partially to relax back toward a planar alignment.

This molecule is of specific interest to us for similar reasons to the helicene described in Section IV, *i.e.* that the distorting of π rings starts to break the separation of σ and π orbitals, which will in turn make our α potentials less and less physically descriptive of the system. With a systematic test of increasingly distorted π rings therefore, we can gain some idea of the limits of the pseudopotentials used.

As our interest was solely in the part of the molecule we wished to model with pseudopotentials, *i.e.* the twisted anthracene itself, we first optimised the molecular geometries (at the DFT-PBE0 level, with def-SV(P) basis sets) before removing the atoms surrounding the central anthracene. This way, we could calculate the anthracene spectra alone.

Figure 9 displays the spectra for Ant-*cn*, for $n = 3 - 6$, using all-electron structures as well as the *set2* pseudopotential set. The all-electron UV spectra are very similar to one another and have peaks in approximately the 120 nm to 400 nm range, with most of the peaks clustered between roughly 120 nm and 280 nm, and the main peak at around 240 nm in each molecule.

The pseudopotentials reproduce in detail the shapes of the all-electron spectra. For the pseudopotentials, all of the peaks in the all-electron systems appear to be present in the pseudosystems. The relative intensities of the peaks also seem broadly correct. Turning to the ECD results we see a similar picture. The *set2* potential spectra peaks are spread across a similar range to those of the all-electron spectra. We see once again that all of the peaks are consistently identifiable in both all-electron and pseudosystem spectra, except for the positive peak at 160 nm in the all-electron spectra. The transition density of this peak shows that it is a σ - σ excitation, meaning that it is not possible for the pseudopotentials to reproduce. In the pseudosystems we instead see a negative peak at this wavelength. The intensity of the peaks is also broadly correct, albeit that main peak at around 260 nm is several times higher for the pseudosystem than for the all-electron system. We also see that the negative-to-positive shift indicative of chirality is found in this range, and remains clear even for the most dis-

torted system, Ant-c3.

An important and interesting feature of these spectra is that while the incorrect negative peak at 160 *nm* increases slightly in intensity as the system becomes more distorted, we do not see any additional unphysical peaks appear, nor do we see much further distortion in the peaks present. There is also no shifting of the pseudosystem spectra relative to the all-electron spectra, even for the most distorted systems. We can say then that errors in the pseudopotential spectra remain fairly consistent as the system becomes more distorted. We found this surprising, as we expected the pseudosystem spectra to degenerate as the π rings were distorted. An analysis of some transition densities for all-electron and pseudosystem excitations is presented in the SI.

These results are consistent with the helicene results of Section IV. Table VI shows the maximum and mean distortion angles per benzene unit across both helicene and twistacene molecules used in this work. The distortion per benzene unit is defined in Figure 7. Both maximum and average distortion rates per benzene unit are similar for [6]helicene and for the Ant-*cn*, with [6]helicene falling roughly in the middle. We see that for all the helicene and Ant-*cn* molecules, both UV and ECD spectra reproduced, albeit with flaws introduced by the distortion.

Table VI. Maximum and mean distortions per benzene unit for helicene and twistacene molecules used in this work. The maximum distortion per benzene unit is chosen to be the largest dihedral torsion angle in the molecule (see Figure 7).

System	Distortion ($^{\circ}$)	
	Mean	Maximum
[6]helicene	9.3	14.5
Ant-c3	14.1	17.9
Ant-c4	13.6	16.9
Ant-c5	10.4	13.3
Ant-c6	6.2	8.1

In conclusion then, we can say that the pseudosystems were successful at recreating the UV spectra of the distorted anthracene π systems, up to and including a distortion of around 14.0 $^{\circ}$ per benzene unit (Ant-c3), with the *set2* potentials proving the most successful. We can also say that while even at the comparatively low rate of twisting of around 6.2 $^{\circ}$ per benzene unit (Ant-c6), at least one spurious excitation is introduced into the ECD spectra of twistacene, the ECD spectra of the pseudosystems are also very accurate, even up to the maximum distortion of the systems.

B. Dodecaphenyltetracene

This molecule is taken from a synthesis performed by Xiao *et al.*¹⁹ Its complexity, along with the fact the team had reported a UV spectrum to which theoretical spectra

could be compared, made this an attractive challenge. The molecular structure is shown in Figure 10.

The reader will see that this molecule could encapsulate two different challenges. First are the phenyl groups, *i.e.* the presence of many π systems facing and overlapping one another at different angles. This is sure to involve the overlapping of many higher-energy orbitals, making for a complicated electronic structure. Second is the fact that the central tetracene is distorted by the phenyl groups. Since we already have both the twistacene molecule in Section V A and the helicene molecules in Section IV to investigate distorted π systems, we decided to ‘pseudopotentialise’ only the phenyl groups so as not to confuse the results. Figure 10 shows the pseudomolecular system. This setup uses both α potentials, along with β potentials connecting the phenyl rings to the central, all-electron tetracene.

The *set2* potentials show five clear peaks in near-agreement with the all-electron peaks, with the possible exception of the peak at around 300 *nm* (*set2*) as compared to around 330 *nm* (all-electron). The relative intensities of these peaks are also similar to those of the all-electron spectrum, although they are somewhat skewed toward the lower-energy excitations by comparison.

The next comparison to be made is that of these theoretical results with the experimental ones in the original synthesis. The experimental spectrum extends from around 650 *nm* to 250 *nm* and has a spread of shallow excitations across the 600-450 *nm* range, a clear and strong peak at 360 *nm*, and at least one further peak in the 250 *nm* region. The stronger peaks (below 450 *nm*) are reproduced in the theoretical spectra, both in all-electron and pseudosystem calculations. The range of peaks between 600 and 400 *nm* is not reproduced in any of the theoretical spectra. Comparison of theoretical and experimental spectra is complicated by the fact in the experimental results the molecules are in a CH₂Cl₂ solution.

An experimental spectrum for a simple tetracene molecule, without phenyl groups, has also been added to Figure 11.²⁰ We see that the characteristic shapes of the dodecaphenyltetracene spectrum appear to arise from the central tetracene itself, and that the main result of adding the phenyl groups appears to be that the largest excitations are now at wavelengths roughly 200 *nm* shorter than in the tetracene. We also see that this ‘shift’ is faithfully reproduced by the DFT calculations, both all-electron and pseudopotential.

Finally, we note that the HOMO energies of the all-electron and *set2* systems are -4.941 *eV* and -5.476 *eV* respectively, making for percentage errors that are not dissimilar to those of the higher polyenes modelled previously¹.

In conclusion, these results for dodecaphenyltetracene show that this pseudopotential technique is effective at recreating the UV spectra of overlapping and interacting π systems.

C. Hemi-Cryptophane

The interest in cryptophanes and hemicryptophanes lies in their propensity to form Van der Waals complexes as hosts. Some of these complexes have interesting catalytic properties^{21–23}. Our own interest in such molecules is that they combine a range of different carbon environments with a complex electronic structure. A further consideration is that such ‘cage molecules’ are often characterised by their UV spectra, and so the ability to produce an accurate UV spectrum with a reduced pseudocage molecule would be useful practically.

The molecule we have adopted as a test subject is a Cu(II)hemicryptophane complex, shown in Figure 12²¹. In order to try to make the investigation as systematic as possible, we decided to group parts of the molecule together as shown in Figure 12, before building up the number of pseudopotentials used.

In light of the results from Section III, where we saw that the results were markedly less accurate when pseudofragments were permitted to bond with heteroatoms, it was decided to institute a ‘no-heterobonding’ rule, in order to see what the effects might be on the resulting spectra. An example of the ‘no-heterobonding’ rule for the ‘upper π ’ section of the cage is shown in Figure 13. Figure 13a shows the regular-all-electron upper π section of the cage. Figure 13b shows a version in which all carbon atoms have been replaced by pseudoatoms. Finally, Figure 13c shows the upper π section following the no-heterobonding rule, in which there must be at least one all-electron carbon atom in between pseudofragments and heteroatoms.

Table VII. Structures of pseudo-hemicryptophane complexes. Each column represents a part of the molecule (see Figure 12) that has been replaced with potentials. The ‘Type’ row shows the pseudopotential setups used for each part. The ‘rule’ column shows whether or not the no-heterobonding rule has been applied, *i.e.* whether pseudofragments have been placed next to non-carbon atoms in the complex. For those systems with the ‘Y (π)’ designation, heterobonding is forbidden for carbons which are members of π rings, but has been permitted elsewhere.

	upper π	upper bridges	methylys	mid-bridges	rule	peaks
Type	α, β	δ	γ	δ	-	-
A	Y	N	Y	N	N	2
B	Y	N	Y	N	Y (π)	1, 2
C	Y	Y	Y	N	Y (π)	1, 2
D	Y	Y	Y	Y	Y (π)	1, 2

Several schemes of pseudopotential positioning have been tested. Only some of them are discussed here, where no pseudopotentials are used below the mid-bridges (as the use of pseudopotentials on the lower π degrades the computed spectra). The full results are reported in the SI.

The reference spectra, shown in Figure 14, has two

major features, a sharp, intense peak at around 800 nm and a shallower and less intense peak centred at around 1900 nm. They are labelled as peak 1 and 2, respectively.

Let us start by comparing A and B. The methyl groups are replaced in both cases by the pseudopotentials. For the upper π , the schemes shown in Figures 13b and 13c are adopted for A and B (no-heterobonding rule), respectively. Peak 1 is not reproduced in A, but it is in B. Peak 2 is reproduced in both, even if the setup of B causes a reduced intensity. Two further setups are presented, with an increasing number of pseudopotentials: pseudocomplex C that builds on the setup of B; pseudocomplex D, where the pseudopotentials are applied from the upper bridges all the way down to the mid-bridges, while making sure to respect the no-heterobonding rule for the upper π systems. This makes for an overall reduction in the number of explicit electrons in the complex of 132, from 545 to 413. For both C and D, the UV spectrum is reproduced with an accuracy similar to that of B.

By summarising the results of the UV spectra we were able to conclude the following:

1. A heteroatom bond (*i.e.* a C-O bond) in the upper π region of the molecule is largely responsible for peak 1. In the original article, this peak was attributed only to copper transitions, independent of the cage. Here we see the cage is in fact necessary for peak 1 to be produced. The no-heterobonding rule is necessary to capture this transition correctly, as evident for A and confirmed by other setups in the SI.
2. Peak 2 has a strong upper π component. Its intensity is well-reproduced for A, but it is reduced in the other cases, suggesting that the setup in Figure 13b reproduces the upper π system slightly better than 13c (see the SI for comparison with other setups).

Figure 15 displays the all-electron transition densities for peaks 1 and 2. These provide further evidence that deductions above regarding the nature of the two peaks are correct. Peak 1 broadly shows an electron transfer between upper and lower parts of the molecules, and the density on the upper part is indeed focused on the upper π rings and the oxygen atoms directly below them. Peak 2 similarly contains a transfer of electron density between the top and bottom of the molecule, with a strong upper π component.

The chirality of this molecule arises in the top of the hemicryptophane itself, and so good pseudopotentials will be important for the accurate reproduction of the ECD spectrum. Looking at Figure 14, this is indeed the case. Pleasingly, with the exception of pseudocomplex A, all the ECD spectra retain a clear positive-to-negative shift, *i.e.* the signature of their chirality is preserved. The most distinctive features of the all-electron system are a positive peak at around 800 nm, leading into a large, shallow negative peak at around 1900 nm. Pseudocomplex spectra B, C and D broadly share these traits,

albeit with intensities differing by up to a factor of around three.

In conclusion then, this Cu(II)hemicryptophane complex is reproducible with simple carbon pseudopotentials, as we were able to recreate the key features of the complex’s spectrum. This is particularly impressive given the heavy delocalisation of electron density over the whole molecule. However, it should be noted that it was necessary to derive a new rule in order to be sure of retaining the necessary electronic complexity, which is that bonds between pseudocarbons and all-electron atoms should be restricted to carbon-carbon bonds only, and that bonding pseudocarbons to explicit heteroatoms should be avoided. Given the results seen in Section III, and given that all pseudocarbons in this work are optimised on carbon-carbon bonds, this seems reasonable.

VI. TIMINGS

Table VIII. Comparison of time and SCF iterations between all-electron and pseudosystems. The gain is defined as the ratio of the all-electron system calculation time to the pseudosystem calculation time. The final two columns give the percentage of electrons and basis functions remaining in the pseudosystem respectively.

SCF Iter. (all-e ⁻)	SCF Iter. (pseudo)	Gain	Gain / Iter.	TDDFT Gain	e ⁻ (%)	basis func. (%)
Dodecaphenyltetracene						
382	54	42.91	6.07	12.21	49	27
Ant-c3						
20	30	14.25	21.38	75.82	14	40
[10]helicene						
31	32	5.37	5.55	1.71	15	42
Hemicryptophane Cage (ps-complex D)						
110	434	0.62	2.39	3.15	76	83

Table VIII displays a comparison of calculation lengths between pseudopotential and all-electron systems, factoring in both the overall computational time and the number of SCF iterations. It can be seen that in every case, that the gain in time *per iteration* is larger than 1, varying from 2.39 to 21.38. The gain per iteration varies strongly, though not exclusively, according to the proportion of electrons remaining in the pseudosystem as compared to the all-electron system.

However, we also see that the number of SCF iterations need to converge pseudomolecule calculations varies a lot. For dodecaphenyltetracene many fewer iterations are needed; for the Ant-c3 and [10]helicene molecules the number is similar, and for the cage molecules many more are required, so much so that the overall gain is less than one, the SCF calculations on the pseudomolecules take *longer* to converge than those on all-electron systems. The reason for this is thought to be the quality of

the starting guesses for the molecular orbitals, the one-electron terms of the Hamiltonians. Finally, we see that across all systems there is a consistent and usually substantial gain in the time taken for TDDFT calculations to be completed.

VII. CONCLUSION

In this work we have developed and extended a previously-reported pseudopotential method¹.

1. Several new pseudopotential setups were created for different hybridised carbon environments, and found to reproduce electronic structural properties such as HOMO, singlet-triplet and 1st ionisation energies to a similar degree of accuracy to the previous work.
2. It was found possible to optimise pseudopotentials to improve the absorption spectra produced with pseudomolecules.
3. It was found that these pseudopotentials, used judiciously, were able to reproduce both the UV and ECD spectra of greatly more complex molecules than those on which they were optimised.

Rules derived for the application of these potentials are that in order to produced that best results they should (1) not be bonded to heteroatoms, i.e. that one all-electron carbon should be kept between pseudofragments and non-carbon all-electron atoms, and (2) that beyond a torsion angle of a few degrees α potentials become unreliable at reproducing ECD spectra, though the quality of UV spectra was maintained up to an average torsion-per-benzene unit of 14.1°. *A possible future improvement to the method would be the optimisation of custom basis sets for the pseudopotentials. Such basis sets could make them more accurate and may require fewer basis functions, increasing computational efficiency.*

VIII. FUNDING INFORMATION

The authors acknowledge the french Ministère de l’éducation nationale et de la recherche for providing the PhD grant of A. Punter.

REFERENCES

- ¹Punter, A.; Nava, P.; Carissan, Y. Atomic pseudopotentials for reproducing π -orbital electron behavior in sp² carbon atoms. *International Journal of Quantum Chemistry* **2019**, e25914.
- ²Igel-Mann, G.; Stoll, H.; Preuss, H. Pseudopotentials for main group elements (IIIa through VIIa). *Mol. Phys.* **1988**, *65*, 1321–1328.
- ³Hay, P. J.; Wadt, W. R. Ab initio effective core potentials for molecular calculations. Potentials for the transition metal atoms Sc to Hg. *The Journal of chemical physics* **1985**, *82*, 270–283.

- ⁴Hay, P. J.; Wadt, W. R. Ab initio effective core potentials for molecular calculations. Potentials for K to Au including the outermost core orbitals. *The Journal of Chemical Physics* **1985**, *82*, 299–310.
- ⁵Wadt, W. R.; Hay, P. J. Ab initio effective core potentials for molecular calculations. Potentials for main group elements Na to Bi. *The Journal of Chemical Physics* **1985**, *82*, 284–298.
- ⁶Hamann, D.; Schlüter, M.; Chiang, C. Norm-conserving pseudopotentials. *Physical Review Letters* **1979**, *43*, 1494.
- ⁷Vanderbilt, D. Soft self-consistent pseudopotentials in a generalized eigenvalue formalism. *Physical review B* **1990**, *41*, 7892.
- ⁸Bonifacic, V.; Huzinaga, S. Atomic and molecular calculations with the model potential method. I. *J. Chem. Phys.* **1987**, *86*, 2132.
- ⁹Dolg, M. Effective core potentials. *NIC Series* **2000**, *1*, 479–508.
- ¹⁰Dolg, M.; Cao, X. Pseudopotentials and Model Potentials. *WIREs Comput Mol Sci* **2011**, *1*, 200–210.
- ¹¹Klobukowski, M.; Huzinaga, S.; Sakai, Y. *Computational Chemistry: Reviews of Current Trends*; 1999; Chapter 2, pp 49–74.
- ¹²TURBOMOLE 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.
- ¹³McMurchie, L. E.; Davidson, E. R. Calculation of integrals over ab initio pseudopotentials. *Journal of Computational Physics* **1981**, *44*, 289–301.
- ¹⁴Punter, A. MOO. 2019; [github.org/cvzj57/chem_scripts](https://github.com/cvzj57/chem_scripts).
- ¹⁵Lu, J.; Ho, D. M.; Vogelaar, N. J.; Kraml, C. M.; Pascal, R. A. A Pentacene with a 144° Twist. *Journal of the American Chemical Society* **2004**, *126*, 11168–11169, PMID: 15355095.
- ¹⁶Pascal, R. A. Twisted Acenes. *Chemical Reviews* **2006**, *106*, 4809–4819, PMID: 17165675.
- ¹⁷Xiao, J.; Duong, H. M.; Liu, Y.; Shi, W.; Ji, L.; Li, G.; Li, S.; Liu, X.-W.; Ma, J.; Wudl, F., et al. Synthesis and structure characterization of a stable nonatwistacene. *Angewandte Chemie International Edition* **2012**, *51*, 6094–6098.
- ¹⁸Bedi, A.; Shimon, L. J. W.; Gidron, O. Helically Locked Tethered Twistacenes. *Journal of the American Chemical Society* **2018**, *140*, 8086–8090, PMID: 29905480.
- ¹⁹Xiao, Y.; Mague, J. T.; Schmehl, R. H.; Haque, F. M.; Pascal Jr., R. A. Dodecaphenyltetracene. *Angewandte Chemie International Edition* **2019**, *58*, 2831–2833.
- ²⁰Bryson, K.; Salama, F.; Ehrenfreund, P.; Ricco, A.; Peeters, Z.; Foing, B.; Jessberger, E.; Robert, F.; Mumma, M. Ground Control Monitoring for the Organics Experiment on the EXPOSE-R Facility on the International Space Station. 2010 NASA Laboratory Astrophysics Workshop. 2011; p C39.
- ²¹Ikbal, S. A.; Colomban, C.; Zhang, D.; Delecluse, M.; Brotin, T.; Dufaud, V.; Dutasta, J.-P.; Sorokin, A. B.; Martinez, A. Bioinspired Oxidation of Methane in the Confined Spaces of Molecular Cages. *Inorganic Chemistry* **2019**, *58*, 7220–7228.
- ²²Perraud, O.; Tommasino, J.-B.; Robert, V.; Albela, B.; Khrouz, L.; Bonneviot, L.; Dutasta, J.-P.; Martinez, A. Hemicryptophane-assisted electron transfer: a structural and electronic study. *Dalton Trans.* **2013**, *42*, 1530–1535.
- ²³Gosse, I.; Robeyns, K.; Bougault, C.; Martinez, A.; Tinant, B.; Dutasta, J.-P. Synthesis and Structural Studies of Gallium(III) and Iron(III) Hemicryptophane Complexes. *Inorganic Chemistry* **2016**, *55*, 1011–1013, PMID: 26771438.

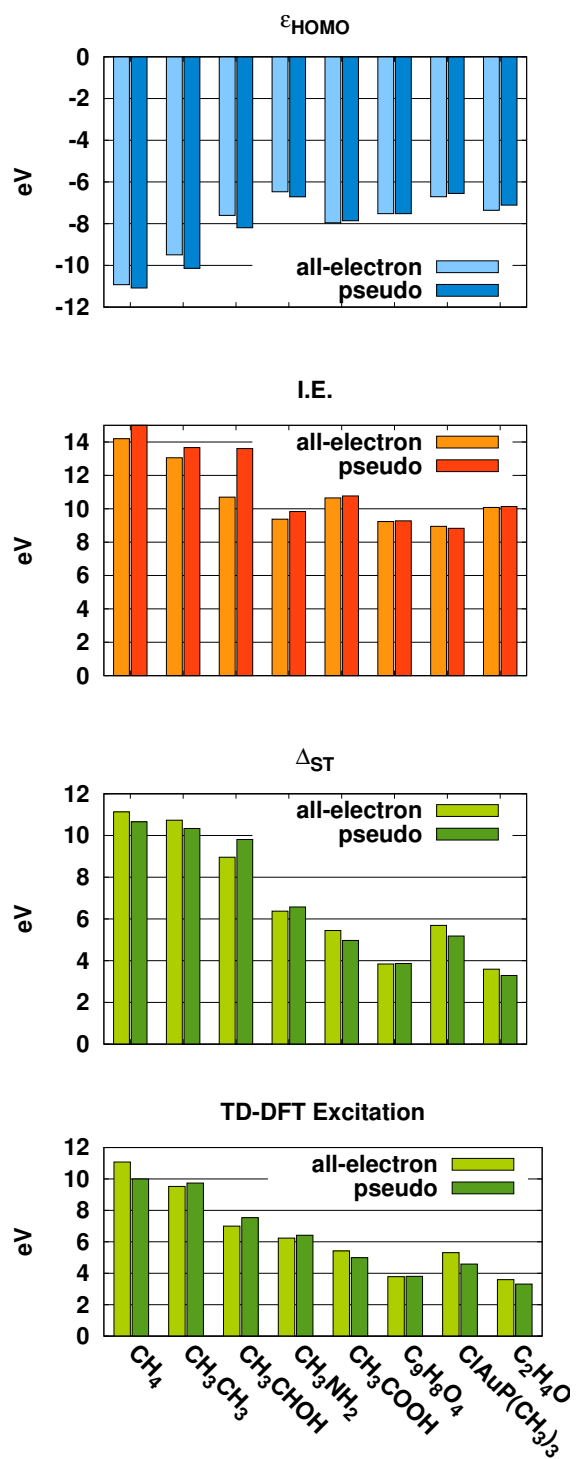


Figure 4. DFT-PBE0 and TDDFT-PBE0 comparison of all-electron and pseudosystem energies across a range of γ potential systems.

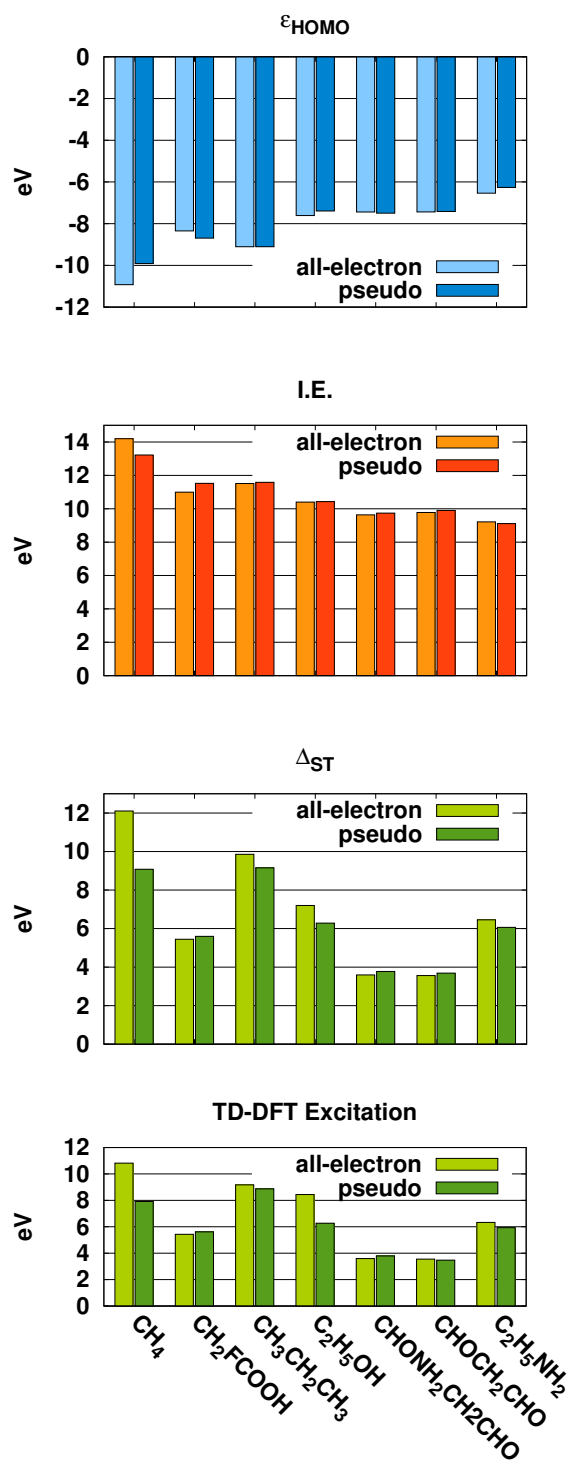


Figure 5. DFT-PBE0 and TDDFT (PBE0) comparison of all-electron and pseudosystem energies across a range of δ potential systems.

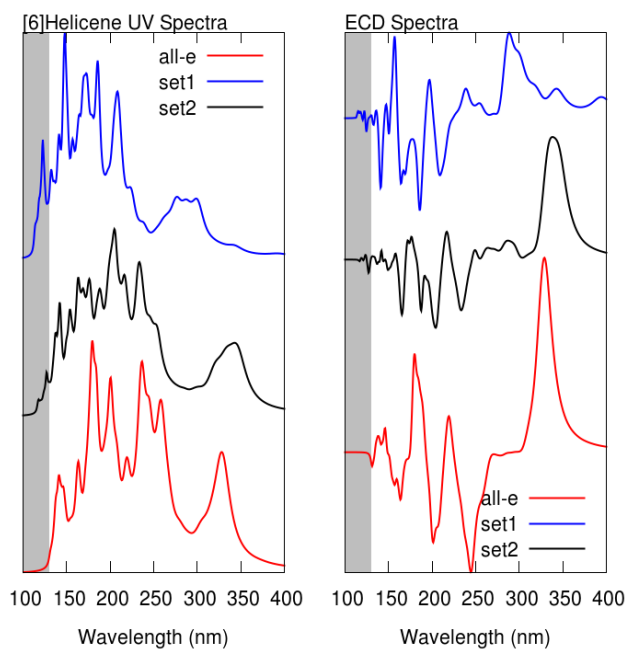


Figure 6. [6]Helicene UV and ECD spectra, for all-electron (red), *set1* (blue) and *set2* pseudopotential (black) systems. Calculations are performed at the TDDFT-PBE0 level. The grey zone covers energies for which no all-electron excitations have been calculated.

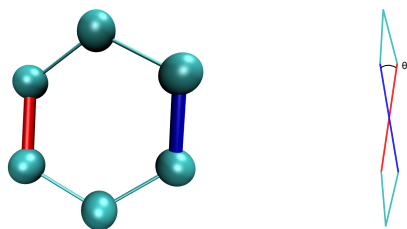


Figure 7. Dihedral torsion of a benzene ring, viewed from the front (left) and the side (right). In this work, the dihedral torsion of an individual benzene ring is defined as the maximum angle θ between any two opposing bonds in the ring.

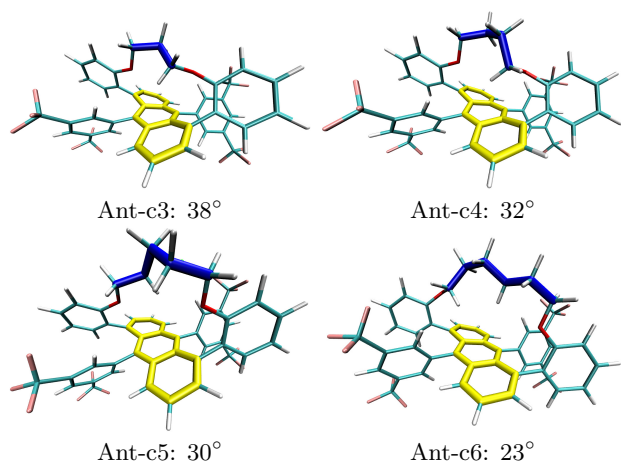


Figure 8. Structures of Ant- cn , where $n = 3 - 6$. From left to right and top to bottom: Ant-c3 (torsion angle 38°), Ant-c4 (torsion angle 32°), Ant-c5 (torsion angle 30°), Ant-c6 (torsion angle 23°). The torsion angle is defined as the angle between the ends of the central anthracene. In each case, the anthracene itself is highlighted in yellow, while the carbon bridge that determines the torsion is highlighted in blue.

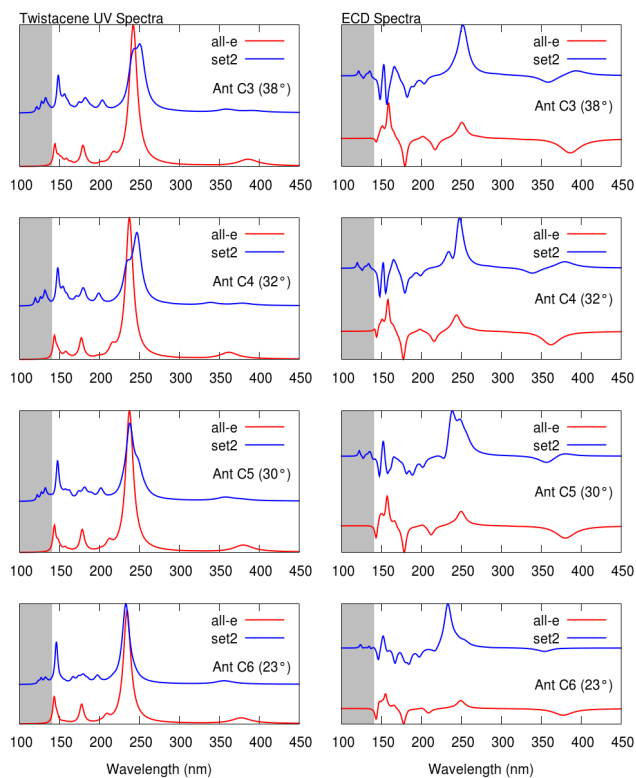


Figure 9. UV and ECD spectra for twisted anthracenes Ant- cn (for $n = 3 - 6$), calculated at the TDDFT-PBE0 level for the first 20 singlet excitations. These include the all-electron spectra (red) and the *set2* potential spectra (blue). The grey zone covers energies for which no all-electron excitations have been calculated.

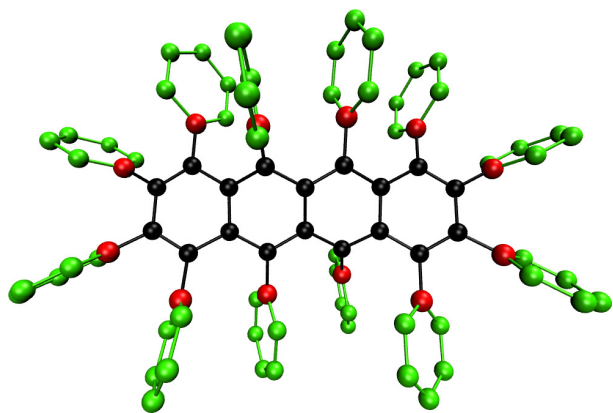


Figure 10. Pseudododecaphenyltetracene, with α pseudoatoms in green and β pseudoatoms in red. All-electron atoms remain in black.

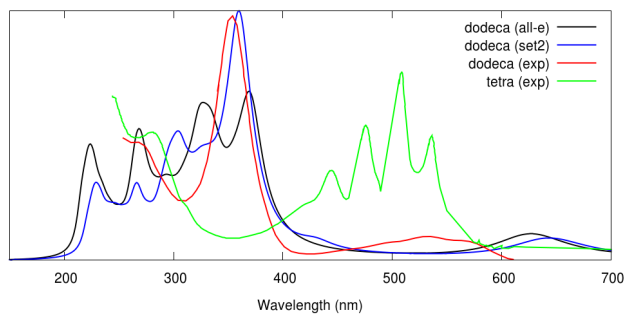


Figure 11. UV spectrum of dodecaphenyltetracene using all-electron (black) and pseudopotential (blue) systems, calculated at the DFT-PBE0 level. The original experimental data¹⁹ is also shown (green), along with an experimental spectrum for tetracene (green)²⁰.

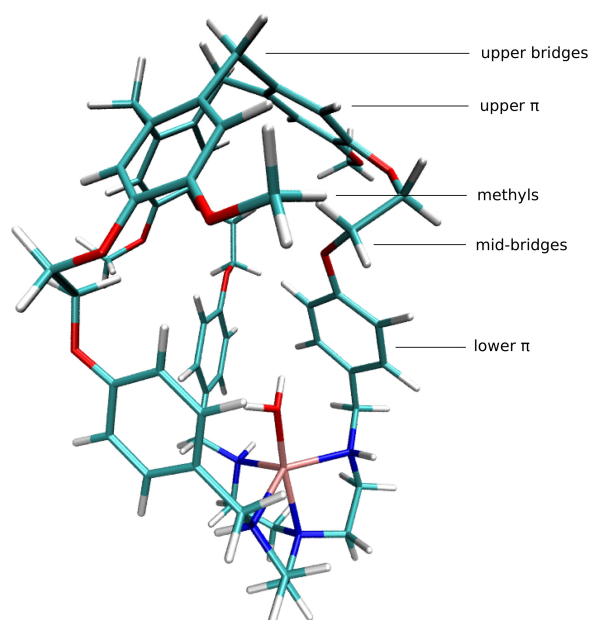


Figure 12. Labelling scheme for Cu(II)hemicryptophane complex, where oxygen is in red, nitrogen in blue, and copper is pink.

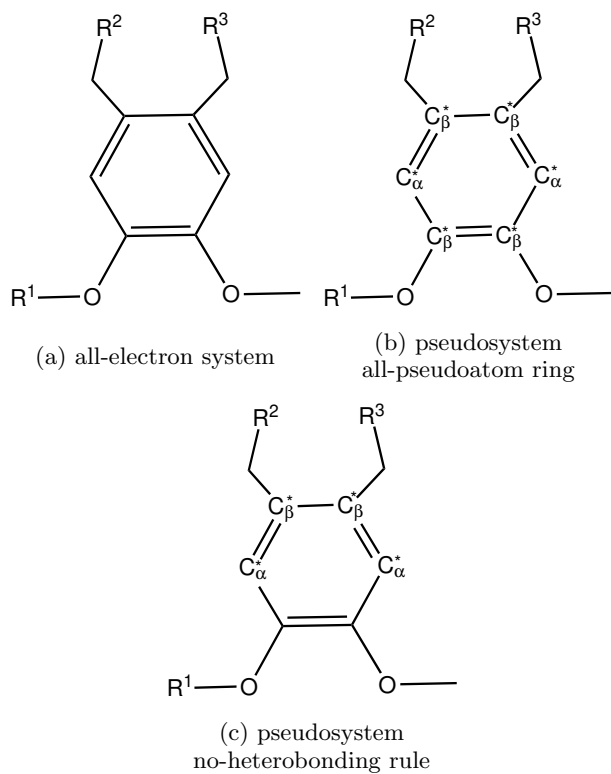


Figure 13. The 'no-heterobonding' rule for upper π ring pseudopotentials.

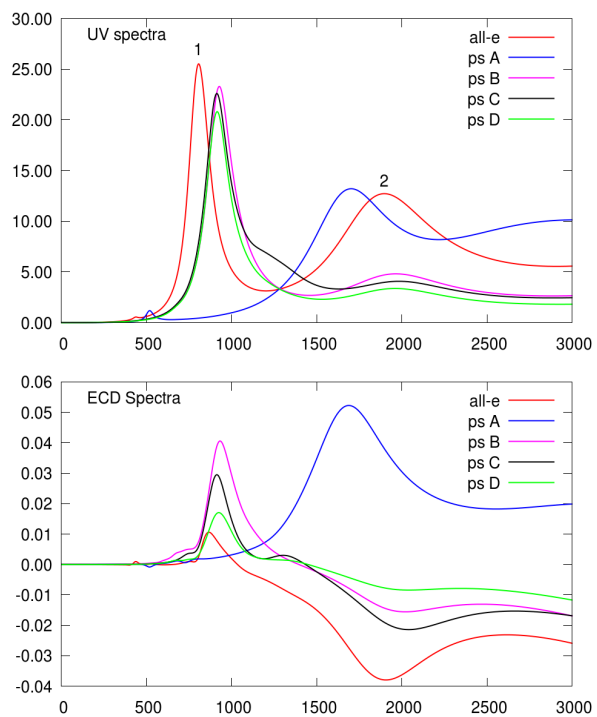


Figure 14. All-electron and pseudomolecular UV (above) and ECD (below) spectra for Cu(II)hemicryptophane. Peaks 1 and 2 are labelled. Calculations were carried out at the TDDFT-B3LYP level, with the first 20 singlet excitations.

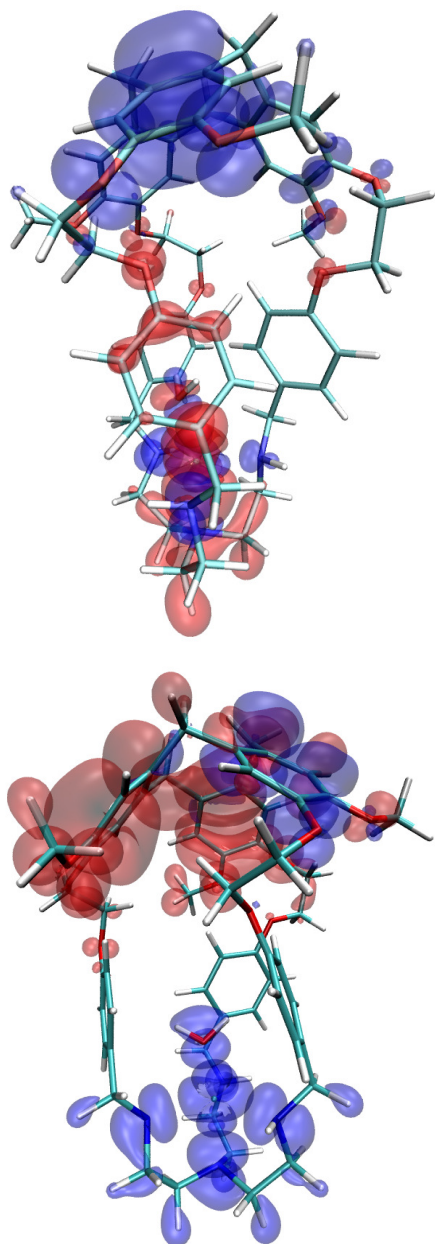


Figure 15. Transition densities based on all-electron TDDFT-B3LYP calculations for excitation peaks 1 (top) and 2 (bottom). In both cases, electron density is reduced in the blue zones and increased in the red.