PCCP

ARTICLE TYPE

Cite this: DOI: 00.0000/xxxxxxxxx

Supplementary Information: A systematic study of the valence electronic structure of cyclo(Gly-Phe), cyclo(Trp-Tyr) and cyclo(Trp-Trp) dipeptides in gas phase^{\dagger}

Elena Molteni,^{*a,b} Giuseppe Mattioli,^a Paola Alippi,^a Lorenzo Avaldi,^a Paola Bolognesi,^a Laura Carlini,^a Federico Vismarra,^{c1,c2} Yingxuan Wu,^{c1,c2} Rocio Borrego Varillas,^{c2} Mauro Nisoli,^{c1,c2} Manjot Singh,^d Mohammadhassan Valadan,^{d,e} Carlo Altucci,^{d,e} Robert Richter^f and Davide Sangalli^{a,b}

1 Unbound states in the continuum: reciprocal space vs localized basis-set

As shown in Figure S1 for c-GlyPhe conformer 1, delocalized unbound continuum states can be captured when using plane waves basis sets, i.e in our case in the QuantumESPRESSO (QE) Hartree-Fock (yellow) and B3LYP (magenta) datasets, while approaches using localized basis sets, such as our CCSD dataset starting from a ORCA HF calculation can only capture "resonances", i.e. unbound states which are (partially) localized on the molecule. Indeed our two QuantumESPRESSO datasets show a continuum-like distribution of states with positive energy, yielding a huge increase in the density of states with respect to the negative energy (bound states) region, while in the CCSD dataset the unbound states, although more dense than the bound ones, are not able to mimick a continuum distribution, and therefore the DOS of unbound states does not show a relevant increase with respect to that of bound states. The GW dataset cannot be taken into account from this point of view, since we have performed this calculation for few QuantumESPRESSO B3LYP empty levels only.

In Figure S2 we show an example of "resonance": the state (red tick) at $E \approx 2.3$ eV, obtained within Hartree-Fock using the localized basis ORCA code, in the highlighted region in the plot of unbound states for c-GlyPhe conformer 1, lies at approximately the same energy as two unbound states (yellow ticks) obtained with HF using the plane wave QE code. In the top and bottom left panels we show the spatial localization of these latter two states, which indeed are rather delocalized, but they also show some density on the molecule, at a difference with e.g. the QE HF LUMO (bottom right panel) - lying at en energy value (not shown here) with no ORCA HF states - which is definitely delocalized, without any density on the molecule.

^aIstituto di Struttura della Materia-CNR (ISM-CNR), Area della Ricerca di Roma 1, Via Salaria km 29,300, CP 10, Monterotondo Scalo, Roma, Italy; E-mail (EM): elena.molteni@mlib.ism.cnr.it

^bDipartimento di Fisica, Università degli Studi di Milano, via Celoria 16, I-20133 Milano, Italy

^{c1}Dipartimento di Fisica, Politecnico di Milano, Piazza Leonardo da Vinci, 32, Milano, Italy

^{c2}CNR-Istituto di Fotonica e Nanotecnologie, Piazza Leonardo da Vinci, 32, Milano, Italy

^d Dipartimento di Scienze Biomediche Avanzate, Università degli Studi di Napoli Federico II, via Pansini 5, I-80131 Napoli, Italy

^eIstituto Nazionale Fisica Nucleare (INFN), Sezione di Napoli, Napoli, Italy

^fSincrotrone Trieste, Area Science Park, Basovizza, Trieste, Italy



Figure S1 c-GlyPhe conformer 1. QuantumESPRESSO Hartree-Fock (yellow) and B3LYP (magenta) DOS for the occupied and empty electronic levels, compared with the GW (green), and CCSD (blue) PE spectral function and with experimental PE spectrum (black). GW corrections are reported for the first eleven QuantumESPRESSO B3LYP empty levels only. The vacuum level is set at 0 energy. The EOM-CCSD calculation has been performed starting from the HF calculation with the localized basis ORCA code. In the HF, B3LYP and GW cases occupied levels are shifted by +2 eV, -2.5 eV, -1.47 eV respectively, with empty levels unshifted.



Figure S2 c-GlyPhe conformer 1. Top right panel: QuantumESPRESSO Hartree-Fock (yellow), Orca Hartree-Fock (red), QuantumESPRESSO B3LYP (magenta), Orca B3LYP (brown) unoccupied electronic energy levels compared with GW (green) and CCSD (blue) inverse PE energies. GW corrections are reported for the five bound QuantumESPRESSO B3LYP empty levels only. The vacuum level is set at 0 energy. Top left and bottom left panels: spatial localization of the two QuantumESPRESSO Hartree-Fock orbitals highlighted in the top right panel, with energies corresponding to a "resonance" i.e. an empty state partially localized on the molecule, obtained by a HF calculation with the Orca localized basis code. Bottom right panel: spatial localization of the QuantumESPRESSO Hartree-Fock LUMO level.

2 Energy ordering of different conformers

In Table S1 we show results on the dependence of the plane-wave QuantumESPRESSO DFT B3LYP total energy of the five lowest energy conformers of the c-TrpTrp dipeptide on cell size, using a simple cubic (sc) cell. Conformers are labeled (conf1, conf2 etc) according to their energy ordering found in the initial tight-binding conformational search. Upon increasing the cell side length from 40 a.u. (the one used for all three dipeptides throughout the present work) up to 70 a.u., the energy ordering switches from 3, 2, 1, 4, 5 to 3, 1, 2, 4, 5, and the energy difference among conformations 1, 2 and 3 decreases, so that the situation tends to approach that found in the initial tight-binding conformational search (where the ordering was 1, 2, 3, 4, 5). In all these cases, the five analyzed conformers can be separated, according to their total energies, into a first group (conformers 1, 2 and 3) always within a maximum energy range of 0.013 eV from each other, and a second group (conformers 4 and 5) differing in energy by more than 0.041 eV from the highest energy conformer of the first group.

	40 a.u.	50 a.u.	60 a.u.	70 a.u.
conf1	-417.0847318 Ry	-417.0848135 Ry	-417.0850968 Ry	-417.0851789 Ry
	+0.011 eV	+0.013 eV	+0.004 eV	+0.003 eV
conf2	-417.0850402 Ry	-417.0849780 Ry	-417.0849742 Ry	-417.0850295 Ry
	+0.007 eV	+0.011 eV	+0.006 eV	+0.005 eV
conf3	-417.0855624 Ry	-417.0857903 Ry	-417.0853841 Ry	-417.0853827 Ry
	0 eV	0 eV	0 eV	0 eV
conf4	-417.0812527 Ry	-417.0812544 Ry	-417.0813554 Ry	-417.0814165 Ry
	+0.059 eV	+0.062 eV	+0.055 eV	+0.054 eV
conf5	-417.0747257 Ry	-417.0753195 Ry	-417.0752128 Ry	-417.0752561 Ry
	+0.147 eV	+0.142 eV	+0.138 eV	+0.138 eV

Table S1 c-TrpTrp dipeptide. QuantumESPRESSO DFT B3LYP total energies and energy differences with respect to the lowest energy conformer, of the five lowest energy conformers from the tight binding conformational search: dependence on the cell size, using a simple cubic (sc) cell.

Still within QuantumESPRESSO DFT geometry optimizations, we also find the 3, 1, 2, 4, 5 energy ordering if we switch from the B3LYP exchange-correlation functional with NC Martins-Troullier pseudopotentials to the PBE0 one with ONCV pseudopotentials, both with the D3 dispersion correction (first and second column in Table S2). Increasing the kinetic energy cutoff for the exact exchange operator from 90 Ry to 135 Ry (first and third column) within B3LYP+D3 yields instead no remarkable difference, and the energy ordering remains 3, 2, 1, 4, 5. Upon using a stricter convergence criterion on forces in the geometry relaxation (fourth vs. first column) the energy ordering of conformers becomes 3, 1, 2, 4, 5, with a very small energy difference (0.002 eV) between conformers 1 and 3. By combining the use of a larger cell (70 a.u.) with the above-mentioned strict convergence criterion, we obtain the following energy ordering: 1, 3, 2, 4, 5, with 0.001 eV only between conformers 1 and 3 (fifth column). Also in all the calculations reported in this Table, the energy difference between the group of the three lowest energy conformers (1,2,3) and the following conformers (4, 5), is larger than those among conformers 1, 2 and 3.

	B3LYP+D3	PBE0+D3	B3LYP+D3	B3LYP+D3	B3LYP+D3
	sc, 40 a.u.	sc, 40 a.u.	fock, sc, 40 a.u.	tight, sc, 40 a.u.	tight, sc, 70 a.u.
conf1	-417.0847318 Ry	-417.7950519 Ry	-417.0848732 Ry	-417.0865420 Ry	-417.08657453 Ry
	+0.011 eV	+0.017 eV	+0.012 eV	+0.002 eV	0 eV
conf2	-417.0850402 Ry	-417.7946484 Ry	-417.0851763 Ry	-417.0852177 Ry	-417.08530383 Ry
	+0.007 eV	+0.022 eV	+0.008 eV	+0.020 eV	+0.017 eV
conf3	-417.0855624 Ry	-417.7962707 Ry	-417.0857429 Ry	-417.0867170 Ry	-417.08646838 Ry
	0 eV	0 eV	0 eV	0 eV	+0.001 eV
conf4	-417.0812527 Ry	-417.7924806 Ry	-417.0813965 Ry	-417.0829814 Ry	-417.08292751 Ry
	+0.059 eV	+0.052 eV	+0.059 eV	+0.051 eV	+0.050 eV
conf5	-417.0747257 Ry	-417.7862192 Ry	-417.0748752 Ry	-417.0759140 Ry	-417.07570804 Ry
	+0.147 eV	+0.137 eV	+0.148 eV	+0.147 eV	+0.148 eV

Table S2 c-TrpTrp dipeptide. QuantumESPRESSO DFT total energies and energy differences with respect to the lowest energy conformer, of the five lowest energy conformers from the tight binding conformational search: dependence on: type of exchange-correlation functional (2nd column: PBE0 with ONCV pseudopotentials), kinetic energy cutoff for the exact exchange operator (3rd column: increased from 90 Ry to 135 Ry), convergence criterion on forces in the geometry relaxation (4th column: decreased from 5.0^{-4} Ha/bohr to 5.0^{-5} Ha/bohr), cell size (5th column). In all calculations we used a simple cubic (sc) cell.

In Table S3 we report results obtained with the ORCA code, using localized basis sets: calculations were performed either within DFT, varying the treatment of the vdW interactions and the exchange-correlation functional (columns 1 to 3), or with the configuration

interaction CCSD(T) method, known to be highly accurate (column 4). In all these calculations the lowest energy geometry is that of conformer 1, and the energy ordering of conformers is either 1, 2, 3, 4, 5 or 1, 3, 2, 4, 5. Interestingly, the ORCA DFT calculation with the M06-2X meta-GGA hybrid functional yields the same energy ordering - with similar total energy values - as the much more computationally expensive DPLNO-CCSD(T) method. For really small total energy differences such as those found in some of the calculations analyzed so far, differences in zero point energies (ZPE, neglected so far) may alter the energy ordering of the conformers. In column 5 we report the zero point energy and entropy corrections to the ORCA DFT B3LYP-D4 total energies, using the def2-TZVPP basis set: these corrections have been applied to the DPLNO-CCSD(T) energies, and the results with the energy difference with respect to the most stable conformer are listed as third and fourth item in each of the cells in the fourth column. The ZPE-corrected CCSD(T) energies yield a 3, 1, 2, 4, 5 conformer ordering, with however 0.0004 eV only between conformers 1 and 3.

	B3LYP+D3	B3LYP+D4	M062X+D3	DLPNO-CCSD(T)	ZPE-TS (B3LYP)
conf1	-1219.828587 Ha 0 eV	-1219.826531 Ha 0 eV	-1219.9846926 Ha 0 eV	-1218.138127763 Ha 0 eV -1217.797369793 Ha +0.0004 eV	0.34075797 Ha
conf2	-1219.827762 Ha +0.022 eV	-1219.825807 Ha +0.020 eV	-1219.9833174 Ha +0.037 eV	-1218.136843354 Ha +0.035 eV -1217.795910154 Ha +0.040 eV	0.34093320 Ha
conf3	-1219.827257 Ha +0.036 eV	-1219.824426 Ha +0.057 eV	-1219.9837206 Ha +0.026 eV	-1218.137427990 Ha +0.019 eV -1217.79738339 Ha 0 eV	0.34004460 Ha
conf4	-1219.825633 Ha +0.080 eV	-1219.822603 Ha +0.107 eV	-1219.9825641 Ha +0.058 eV	-1218.135709522 Ha +0.066 eV -1217.79544614 Ha +0.052 eV	0.34026338 Ha
conf5	-1219.822833 Ha +0.157 eV	-1219.820356 Ha +0.168 eV	-1219.9781133 Ha +0.179 eV	-1218.132405190 Ha +0.156 eV -1217.79223420 Ha +0.140 eV	0.34017099 Ha

Table S3 c-TrpTrp dipeptide. Total energies of the five lowest energy conformers from the tight binding conformational search, and energy differences with respect to the lowest energy conformer, obtained with the following methods: Columns 1-3: ORCA DFT with def2-QZVPP basis set: dependence on the type of vdW treatment (2nd column: D4 instead of D3) and on the type of exchange-correlation functional (3rd column: M06-2X meta-GGA hybrid functional); Column 4: ORCA DPLNO-CCSD(T) (without re-optimizing the geometries obtained with ORCA B3LYP-D4); Column 5: zero point energy and entropy corrections to the ORCA B3LYP-D4, def2-TZVPP basis set, energy values.

In Table S4 we report the conformer energy ordering found in the tight-binding conformational search using different tightbinding Hamiltonians and pairwise dispersion correction treatments, and the effects of including ZPE corrections. Although in the ZPE-uncorrected datasets (columns 1 and 2) conformer 1 is the lowest energy geometry, also here, as in the ORCA DPLNO-CCSD(T) dataset (fourth column of Table S3) the inclusion of the ZPE correction tends to lower the energy of conformer 3. In particular, adding the ZPE correction to the tight-binding GFN2(D4) dataset, conformer 3 becomes the lowest energy geometry of c-TrpTrp.

	GFN2(D4)	GFN1(D3)	GFN2(D4)+ZPE-TS	GFN1(D3)+ZPE-TS
conf1	-77.21030089 Ha	-79.54047362 Ha	-76.88384630 Ha	-79.21474531 Ha
60			+0.005 ev	
conf2	-77.20978280 Ha ± 0.014 eV	-79.53861641 Ha $\pm 0.051 \text{ eV}$	-76.88301986 Ha ±0.027 eV	-/9.21285896 Ha +0.051 eV
conf2	77 20000176 Ua	70 527/1507 Uo	76 99401700 Ha	70.001 CV
COIIIS	+0.033 eV	+0.083 eV	-70.88401709 Ha 0 eV	+0.036 eV
conf4	-77.20879438 Ha	-79.53792224 Ha	-76.88337462 Ha	-79.21344391 Ha
	+0.041 eV	+0.009 eV	+0.017 eV	+0.035 eV
conf5	-77.20777333 Ha	-79.53755072 Ha	-76.88199196 Ha	-79.21257408 Ha
	+0.069 eV	+0.080 eV	+0.055 eV	+0.059 eV

Table S4 c-TrpTrp dipeptide. Total energies of the five lowest energy conformers from a tight binding conformational search, and energy differences with respect to the lowest energy conformer, obtained using: column 1 (column 2) the GFN2-xTB (GFN1-xTB) tight-binding Hamiltonian with the D4 (D3) pairwise dispersion correction; columns 3 and 4: same two methods, adding the zero point energy and entropy corrections.

Summarizing, from this detailed analysis of the energy ordering of c-TrpTrp geometries (Tables S1 to S4), we can conclude that, as for the starting tight-binding conformational search, both the implementations we have tried within the CREST code, i.e. the one with the GFN1 tight-binding Hamiltonian and the D3 pairwise dispersion correction, and the one with GFN2 and D4 (expected to better describe electrostatic and vdW interactions), identify the same five lowest energy conformers.

When starting from the five lowest energy geometries found with the tight-binding conformational search, both DFT and CCSD calculations, with various choices/combinations of xc functionals, basis type, vdW treatment, yield a first "cluster" of lowest energy geometries (conformers 1, 2, and 3, with possibly varying mutual ordering) followed, at higher energies, by conformers 4 and 5. Often the total energies of conformers 1, 2 and 3 span a smaller energy range if compared to their energy separation from conformers 4 and 5.

In particular, within the first cluster of conformers, plane-wave QE DFT calculations (ORCA DFT calculations with localized basis sets) tend to favour conformer 3 (conformer 1). Within the very accurate - and computationally expensive - CCSD(T) method with the addition of zero point energy corrections, the most stable geometry is conformer 3, although at 0.0004 eV only from conformer 1.

These findings lead us to focus on conformers 1, 2 and 3 of c-TrpTrp as reliable candidates for lowest energy geometries of this dipeptide, on which we then calculated electronic properties.

3 Spatial localization of EOM-CCSD and of B3LYP orbitals



Figure S3 Hartree-Fock orbitals (left) contributing to the CCSD "IE-1" ionization energy of c-GlyPhe (center), and the B3LYP HOMO-2 (right).



Figure S4 Hartree-Fock orbitals (left) contributing to the CCSD "IE-2" ionization energy of c-GlyPhe (center), and the B3LYP HOMO-5 (right).

4 Calculated adiabatic ionization energies and electron affinities for the cyclo-dipeptides under study

eV	r ² -SCAN-3c	B3LYP-D3	M062X-D3	DLPNO-CCSD(T)		
	c-GlyPhe					
IP	8.15	8.35	8.93	9.11		
EA	0.37	0.48	0.50	0.74		
c-TrpTyr						
IP	6.94	7.01	7.36	-		
EA	0.06	0.17	0.15	-		
c-TrpTrp						
IP	6.82	6.94	7.40	-		
EA	0.01	0.15	0.20	-		

Table S5 Adiabatic ionization energies (IE) and electron affinities (EA) for the investigated cyclo-dipeptides, calculated with the meta-GGA exchangecorrelation functional r^2 -SCAN-3c and the mTZVPP basis set (column 1), the global hybrid B3LYP functional (20 % exact exchange (EXX)) (column 2) and the meta-GGA global hybrid M062X functional (54% EXX) (column 3), both using the D3 dispersion correction and the def2-QZVPP basis set, and finally with the DLPNO-CCSD(T) method (column 4). In each of the three cases reported in columns 1, 2 and 3 the geometry has been optimized with the chosen functional; in the DLPNO-CCSD(T) case, due to the large computational weight, IE and EA have been calculated on the B3LYP-D3 optimized geometry. In all cases the thermochemical contribution (ZPE-TS@298 $^{\circ}$ K) calculated with the r^2 -SCAN-3c functional has been added.

eV	B3LYP-D3 adiab	B3LYP-D3 vert	M062X-D3 adiab	M062X-D3 vert			
	c-GlyPhe						
IP	8.43	8.57	9.01	9.25			
EA	0.68	0.81	0.70	0.91			
	c-TrpTyr						
IP	7.05	7.29	7.40	7.76			
EA	0.38	0.63	0.36	0.71			
c-TrpTrp							
IP	6.97	7.08	7.42	7.70			
EA	0.37	0.47	0.42	0.64			

Table S6 Comparison between adiabatic and vertical values of IE and EA for the three investigated cyclo-dipeptides, all obtained using the def2-QZVPP basis set and without the thermochemical contribution ZPE-TS.