molecules

MDPI

Article

An Expedited Route to Optical and Electronic Properties at Finite Temperature via Unsupervised Learning

Fulvio Perrella [1,†], Federico Coppola [1,†], Nadia Rega [1,2,3] and Alessio Petrone [1,2,3,*]

[1] Scuola Superiore Meridionale, Largo San Marcellino 10, I-80138 Napoli, Italy; fulvio.perrella-ssm@unina.it (F.P.); federico.coppola@unina.it (F.C.); nadia.rega@unina.it (N.R.)
[2] Department of Chemical Sciences, University of Napoli Federico II, Complesso Universitario di M.S. Angelo, via Cintia 21, I-80126 Napoli, Italy
[3] Istituto Nazionale di Fisica Nucleare, Sezione di Napoli, Complesso Universitario di M.S. Angelo ed. 6, via Cintia 21, I-80126 Napoli, Italy
* Correspondence: alessio.petrone@unina.it
† These authors contributed equally to this work.

Abstract: Electronic properties and absorption spectra are the grounds to investigate molecular electronic states and their interactions with the environment. Modeling and computations are required for the molecular understanding and design strategies of photo-active materials and sensors. However, the interpretation of such properties demands expensive computations and dealing with the interplay of electronic excited states with the conformational freedom of the chromophores in complex matrices (i.e., solvents, biomolecules, crystals) at finite temperature. Computational protocols combining time dependent density functional theory and ab initio molecular dynamics (MD) have become very powerful in this field, although they require still a large number of computations for a detailed reproduction of electronic properties, such as band shapes. Besides the ongoing research in more traditional computational chemistry fields, data analysis and machine learning methods have been increasingly employed as complementary approaches for efficient data exploration, prediction and model development, starting from the data resulting from MD simulations and electronic structure calculations. In this work, dataset reduction capabilities by unsupervised clustering techniques applied to MD trajectories are proposed and tested for the ab initio modeling of electronic absorption spectra of two challenging case studies: a non-covalent charge-transfer dimer and a ruthenium complex in solution at room temperature. The K-medoids clustering technique is applied and is proven to be able to reduce by ∼100 times the total cost of excited state calculations on an MD sampling with no loss in the accuracy and it also provides an easier understanding of the representative structures (medoids) to be analyzed on the molecular scale.

Keywords: density functional theory; machine learning; computations of optical spectra; molecular dynamics; clustering techniques

Citation: Perrella, F.; Coppola, F.; Rega, N.; Petrone, A. An Expedited Route to Optical and Electronic Properties at Finite Temperature via Unsupervised Learning. *Molecules* **2023**, *28*, 3411. https://doi.org/10.3390/molecules28083411

Academic Editor: Maofa Ge

Received: 26 March 2023
Revised: 6 April 2023
Accepted: 7 April 2023
Published: 12 April 2023

1. Introduction

Photo-induced phenomena and optical properties are the grounds to investigate electronic states and their interactions with the environment [1–22]. Experimental spectra can be interpreted via computational approaches at the molecular scale, understanding the microscopic characteristics that determine the position, width and shape of absorption bands [23–34]. However, a number of challenges remain open and mainly concern the modeling of either floppy molecules or non-covalent complexes in solution. Ideal approaches to deal with the complexity of the conformational freedom have to ensure an adequate sampling of the phase space of the potential energy surface (PES) at a given temperature, since such systems cannot be easily described by minimum energy structures as starting points for subsequent more computationally expensive calculations required to compute electronic transitions and excited state properties [34–36]. Molecular dynamics (MD) is

the perfect technique for this goal since it can simultaneously describe the conformational freedom and the complexity of the environment (i.e., explicit solvent models) and can guarantee a satisfactory sampling of the phase space of the PES for selecting the initial states of the electronic transitions. On these bases, it is possible to reproduce the thermal fluctuations in a classical manner and simulate the shape of the electronic spectrum by classically considering the spreading of the vertical transitions of a representative sample of snapshots of the MD trajectory [37–39].

An accurate description of the electronic layout of a system is very important for the excited state properties and optical absorption. The electronic state separation, and the resulting UV-Vis absorption, strongly depend on the reference structure(s), indeed. A very detailed description of the PES ruling the system dynamics is demanded when standard force fields cannot by easily used, i.e., with non-covalent charge-transfer complexes [40–48], metal compounds or usually when the electronic density reorganization is involved during the time, even in the ground state. This is usually quite common also when an environment reorganization is involved as well. Since parameterized force fields cannot account for explicit electronic effects, an explicit treatment of electronic degrees of freedom is mandatory via ab initio methods. However, when reasonable large systems ($\geq$1000 atoms) are studied, accurate wavefunction-based methods cannot be employed due to their high computational cost (above all, for excited state properties), although some progress has been recently achieved using graphical processing units [49] and localization procedures [50,51]. Thus, density functional theory (DFT) and time dependent (TD-) DFT, the latter required for excited state quantities, are usually chosen as a good compromise between accuracy and computational costs [10,52–62].

Besides canonical computational chemistry fields, data analysis and machine learning (ML) methods have been increasingly employed as complementary approaches for an efficient data exploration, prediction and model development, starting from experimental data (structure, properties, reactivity) or from MD simulations and electronic structure calculations [63–76]. In particular, MD simulations often produce very big datasets (i.e., the collected trajectory in the phase-space), especially for long simulation times and extended systems, which can be difficult to manually inspect. Automated ML data analysis techniques thus can offer a valuable and efficient option to extract the significant and "physical" information from MD trajectories. In particular, unsupervised ML methods, such as clustering analysis, are able to partition a dataset according to similarities in some features space, employing only the *input* values and not requiring any *output* ones supplied by the user. Clustering proved a valuable tool for MD simulation analysis, allowing one to reduce the high number of sampled structures into a few representative ones, approximating conformational energy minima [77–92].

The simulation of electronic band-shapes at finite temperature through an MD sampling potentially requires hundreds of excited state calculations. Therefore, alternative routes such as the selection of a small number of representative frames could both reduce the computational cost of spectra calculations and simplify their interpretation [78,82]. In this work, dataset reduction capabilities via clustering techniques applied to MD trajectories in specifically tailored feature spaces were tested in the simulation of electronic absorption spectra of two model compounds. In particular, spectra computed only from the clusters' representative frames showed a remarkable reproduction of the main spectral features if compared to spectra from a uniform sampling of frames of the trajectory (a subset of $\sim$500 structures). This approach also allowed an easier interpretation of the calculated bands, which could result from many states close in energy but differing for their spatial properties.

Dataset reduction capabilities via unsupervised clustering techniques applied to MD trajectories are proposed and tested for the ab initio modeling of electronic absorption spectra of two challenging case studies. The first investigated model system is a prototypical π-stacked non-covalent dimer in dichloromethane (DCM) solvent (see Figure 1, left panel), consisting of an electron donor (1-chloronaphthalene, 1ClN) and an electron acceptor

(tetracyanoethylene, TCNE). This represents a challenging case study with respect to evaluating the performance of ML clustering techniques in reproducing its electronic/optical properties, considering that the potential energy surface of weakly bound systems, ruled by dispersive intermolecular forces, is quite flat and numerous isoenergetic orientational isomers can be present in the solution. The TCNE:π:1ClN dimer has been thoroughly investigated in recent years by means of Femtosecond Stimulated Raman Spectroscopy [93] and through electronic structure methods for the detailed characterization of the ground state properties [35] and to unveil the nuclear relaxation upon photoexcitation downhill from the Franck–Condon region of the first charge-transfer state [34,94]. The second system is instead a Ru(II) complex, $[Ru(dcbpy)_2(NCS)_2]^{4-}$ (dcbpy = 4,4′-dicarboxy-2,2′-bipyridine) in water solution, also called "$N3^{4-}$" (Figure 1, right panel), which is a popular example of Ru-based dye sensitizers for solar cells and light-harvesting applications [95–97]. Light absorption by $N3^{4-}$ in the visible region induced excitation to a dense manifold of metal-to-ligand charge-transfer (^{1}MLCT) states. $N3^{4-}$ photo-physical behavior is characterized by an ultrafast relaxation pathway among the singlet and triplet manifolds, induced by a complex interplay between closely spaced coupled electronic states, nuclear motion and solvent rearrangement [98–105], potentially influencing the dynamics and efficiency of the electron injection into a semiconductor substrate [106–110]. The $N3^{4-}$ complex, both for its dense ^{1}MLCT manifold and its conformational dynamics in the solution [111], represents therefore another ideal model system for testing an efficient MD/ML clustering approach for the simulation of electronic spectra including finite-temperature effects.

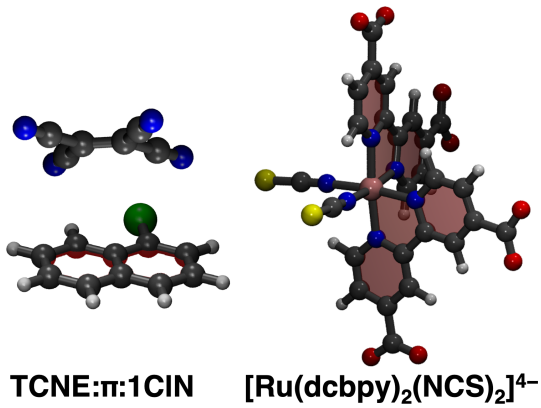

Figure 1. Case studies investigated in the present work. The TCNE:π:1ClN non-covalent dimer and Ru(II) complex ($[Ru(dcbpy)_2(NCS)_2]^{4-}$ or "$N3^{4-}$", dcbpy = 4,4′-dicarboxy-2,2′-bipyridine) are presented from left to right, respectively (Carbon is in gray, Hydrogen in white, Chlorine in green, Sulphur in yellow, Oxygen in red, Nitrogen in blue, Ruthenium in pink).

2. Results and Discussion

2.1. The TCNE:π:1ClN Case Study

The massive amount of data acquired during an MD simulation requires analyses that are capable of going beyond the visual inspection of snapshots and average structures. In this regard, the statistical analysis of the trajectories through the calculation of the distribution functions represents an advantageous choice with respect to taking into account conformational dynamics, solute–solvent interactions and so on. In Figure 2, a three-dimensional spatial distribution function (SDF) [112,113] is presented, computed along the 10 ps-long AIMD trajectory by considering the 1ClN as reference molecule (for which a local three points coordinate system was defined) and the center of the mass of the TCNE unit. From the SDF, it is observed that the TCNE remains on the same side of the ClN and slips on both rings during the exploration of the ground state PES, proving the presence of different mutual configurations and distances due to the weak Coulombic interactions that rule the π-stacked arrangement.

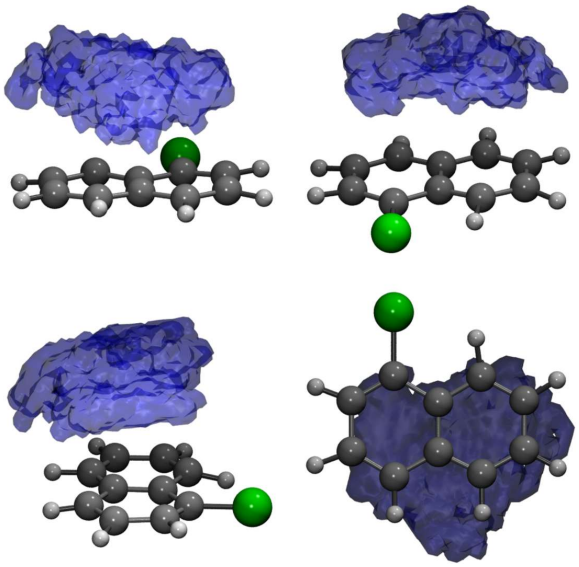

Figure 2. Side, front and top views of the spatial distribution function of the center-of-mass of the TCNE acceptor monomer around the 1ClN subunit.

According to the procedure presented in Section 3.2, the MD trajectory clustering thus yielded, for the TCNE:π:1ClN case study, five medoids (each one is representative of the corresponding cluster), characteristic of the accessible conformational space, in agreement with the SDF previously discussed. The clustering structural feature values shown by the five medoids, namely the rotation angle (θ_r) and the rotation axis ($\hat{n}_r$) of the two subunits molecular planes and the position vector between the two geometric centers ($\vec{r}_{N-E}$), are collected in Table 1 (see also Section 3.2 for features definitions). A detailed analysis of medoid structural features reveals that they do not significantly differ in the rotation angle θ_r between the two TCNE and 1ClN planes (values within a small range, ∼5 degrees), but mainly in the planes' relative orientation, as suggested by clearly different $\hat{n}_r$ axis components and secondarily in the TCNE-1ClN relative position. A closer inspection of the rotation axis $\hat{n}_r$ components for medoids 2 and 4 reveals that they differ along the x- and z-axes (see Table 1), while for medoids 3 and 5 only the component along the z-axis is reoriented for these latter. Additionally, analyzing the five medoids obtained from the clustering approach (side view presented in the right panel of Figure 3), considerable geometric deformations are present and the relative position of the molecular planes is also different. In order to acquire a visual representation of such clustering, the trajectory was projected onto the subspace of the features' first two principal components (PCs; please refer to Section 3.3 for technical details). A clear cluster separation was obtained, with each medoid representing a different portion of the conformational space (see Figure 4). The partial superposition of cluster 5 with cluster 3 in the principal components subspace is only an artifact (being instead separated in the full space) due to the reduced variance explained by the first two PCs (∼55.5% of the total variance).

Table 1. Clustering feature values of the five cluster medoids from TCNE:π:1ClN trajectory. θ_r: rotation angle (angle between versors normal to the two molecular planes, degrees), $\hat{n}_r$: rotation axis (versor normal to the former ones), $\vec{r}_{N-E}$: relative position vector (between 1ClN and TCNE geometric centers). Vector quantities are given as cartesian components (Å) in a fixed frame of reference.

Medoid	θ_r	$n_{r,x}$	$n_{r,y}$	$n_{r,z}$	$r_{N-E,x}$	$r_{N-E,y}$	$r_{N-E,z}$
1	16.68	−0.660	0.398	0.638	1.513	2.847	2.190
2	14.60	−0.816	0.141	0.561	2.604	1.673	1.852
3	19.36	0.723	−0.407	−0.558	1.543	2.444	2.276
4	15.19	0.714	0.021	−0.699	2.666	1.773	1.736
5	16.27	0.112	−0.736	0.668	1.833	2.401	2.436

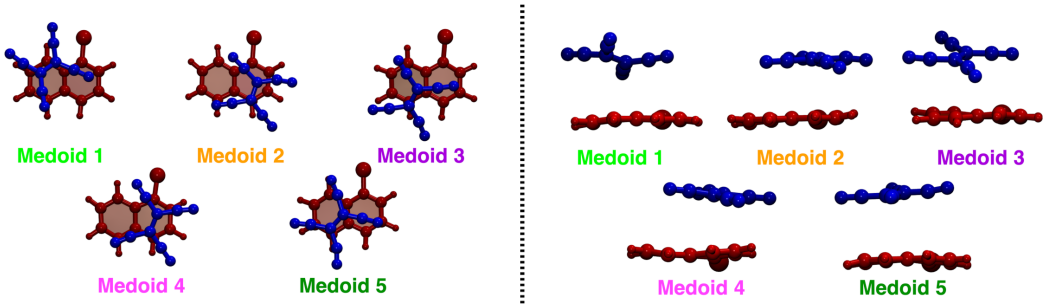

Figure 3. Structures of the five cluster medoids in top (**left** panel) and side (**right** panel) views. The TCNE and 1ClN are represented as ball and stick in blue and red, respectively. The color code is uniform with that of Figure 4.

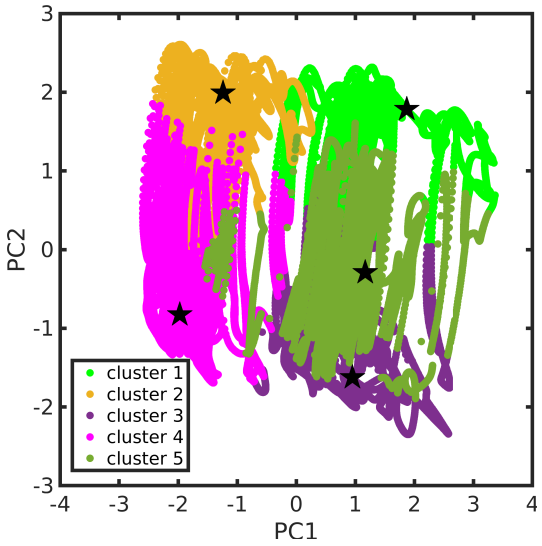

Figure 4. TCNE:π:1ClN trajectory in the features' first two principal components space. Cluster partition is represented through different colors. Cluster medoids are also highlighted (as star symbols). The color scheme adopted is kept fixed throughout this section.

From Figure 3, the found medoids overall show pairwise structural similarities (see 2–4 and 3–5) if observed in a top-down direction; see Table 1 for a more quantitative evaluation. The conformational flexibility of the TCNE:1ClN π-stacked complex, which is due to the weak dispersion forces, is thus fully captured with the trajectory clustering approach.

The UV-Vis spectrum comprising the first low-lying singlet states computed for each medoid within linear response TD-DFT formalism is reported in Figure 5. The estimation of the whole electronic spectrum (red curve) was obtained according to Equation (5) and the procedure explained in Section 3.4. A comprehensive analysis of the electronic spectrum has been recently provided by some of the authors in Ref. [35]. On the other hand, we report a detailed summary of the characterization of the transitions towards the S_1 and S_2 excited states in Table 2. We recall that the weaker electronic transitions below 4.00 eV have a charge transfer (CT) nature (for S_1 and S_2 see ω_{CT} charge transfer descriptor parameter in Table 2). Conversely, the very bright ones are characterized by electronic density reorganization occurring in the same molecular unit, hence they are of a local excitation (LE) character. For medoid 1, the TCNE is located on an edge of the 1ClN ring and it mainly contributes to the absorption bands above 2.50 eV (see light green curve in Figure 5), while for the S_0–S_1 electronic transition at 1.807 eV characterized by a strong CT nature ($\omega_{CT} = 0.968$) the probability is negligible, $f = 0.002$. For medoids 2 and 4, the TCNE lies on the ring bearing the chlorine atom and they share roughly the same electronic properties in terms of transition probability and energy range (see orange and magenta curves in Figure 5, respectively). Both show absorption bands in all regions

of the spectrum. In this case, the first two states S_1 and S_2 of both medoids contribute, respectively, to the bands at ∼2.00 and 2.80 eV. Also in these cases, the S_1 and S_2 states are characterized by a strong charge transfer nature as can be easily deduced from the values of the ω_{CT} descriptor close to unity, reported in Table 2. In medoids 3 and 5, the TCNE is placed on the unfunctionalized six-membered ring of the 1ClN and we observe that the electronic properties show considerable differences. The electronic features of medoid 3 (violet curve in Figure 5) cover the entire spectral range considered, 1.50–5.00 eV, while only high energy electronic transitions (>3.50 eV) are bright for medoid 5 (dark green curve in Figure 5).

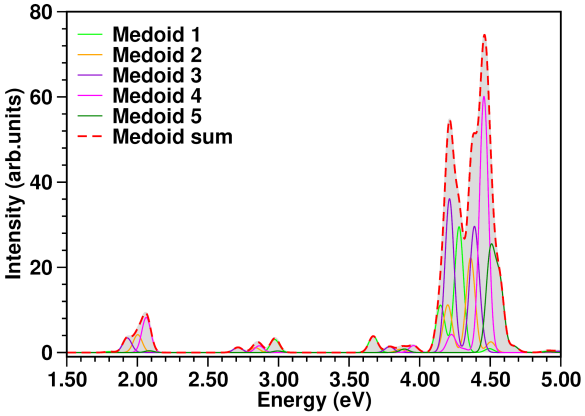

Figure 5. TCNE:π:1ClN absorption spectrum (in eV) calculated at TD-CAM-B3LYP/6-31G(d,p)/GD3/C-PCM(DCM) level of theory from each medoid and as the sum spectrum of the structures representative of the conformational equilibrium in the ground state. The color code is presented in the graph legend. The sum spectrum (red dashed curve) was obtained as the sum of individual medoid contributions (presented in the figure as well, see color legend), each one already multiplied by the k-th cluster population. See Equation (5) and the procedure explained in Section 3.4 for more details.

Comparing the spectrum from the five medoids to that from the complete MD sampling, an excellent agreement is observed (Figure 6, top panel). The experimental optical spectrum profile in solution (Figure 6, bottom panel) shows two distinct absorption bands with maxima centered at 408 nm (3.04 eV) and 537 nm (2.31 eV), as well as the calculated spectrum. In particular, the first calculated band at ∼2.00 eV has contributions from the S_1 states of representative frames 4, 2 and 3, each having, in turn, a clear 1ClN → TCNE charge-transfer character. Analogously, the second band at ∼2.80 eV appears constituted by the S_2 CT states of medoids 1, 4, 3 and 2.

Table 2. Characterization of S_1 and S_2 excited states of TCNE:π:1ClN cluster medoids. ν_i (eV): vertical excitation energy, f_i: oscillator strength (arb. units), Ω_{AB}: transition density population analysis for A (hole) and B (electron) fragments, ω_{CT}: charge transfer descriptor (please refer to Section 3.4 for definitions). Fragment labels: E: TCNE, N: 1ClN.

Medoid		ν_i	f_i	Ω_{EE}	Ω_{EN}	Ω_{NE}	Ω_{NN}	ω_{CT}
1	S_1	1.807	0.002	0.014	0.000	0.968	0.018	0.968
	S_2	2.973	0.035	0.016	0.000	0.965	0.018	0.966
2	S_1	2.003	0.052	0.031	0.001	0.943	0.025	0.944
	S_2	2.835	0.014	0.018	0.001	0.955	0.027	0.955
3	S_1	1.928	0.026	0.019	0.000	0.963	0.017	0.963
	S_2	2.713	0.009	0.018	0.000	0.963	0.018	0.964
4	S_1	2.063	0.067	0.031	0.001	0.938	0.030	0.939
	S_2	2.863	0.013	0.014	0.000	0.953	0.032	0.954
5	S_1	2.084	0.005	0.009	0.000	0.980	0.012	0.980
	S_2	2.994	0.005	0.017	0.000	0.971	0.012	0.971

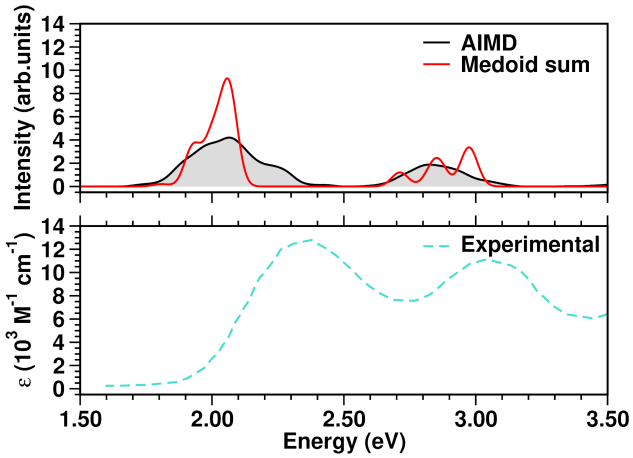

Figure 6. Top panel: comparison of TCNE:π:1ClN simulated absorption spectra in the 1.50–3.50 eV range. **Bottom** panel: experimental UV-Vis spectrum, retrieved from Ref. [93], of the TCNE:π:1ClN complex measured in DCM solution (molar absorptivity, ε). The color code is presented in the graph legend.

Such a case study proves the clustering technique to be an efficient way to estimate the electronic spectrum at finite temperature, avoiding excited state calculations on a large number of frames (for TCNE:π:1ClN model system, a 100-fold decrease in total computational cost). Moreover, the medoid excited state characterization (Table 2) allows one to perform a more accurate spectral assignment of the absorption bands. This further confirms that the cluster medoids, taken as representative frames, can efficiently resume the collected conformational dynamics.

2.2. The N3^{4-} Case Study

The N3^{4-} dynamics at room temperature in water solution is characterized, on the one hand, by the rigidity of the dcbpy ligands, due to the chelation to the Ru center and, on the other hand, by the flexibility of the NCS^{-} ligands, exploring conical-shaped regions (please see Figure 1, right panel, to recall the system under investigation). The vibrational dynamics induce therefore instantaneous deviations from the ideal C_2 symmetry, which could improve the transition probability of otherwise dark excited states [111]. The clustering procedure applied to the collected N3^{4-} trajectory suggested a partition into seven distinct clusters. Projection into the two-dimensional principal component subspace (actually accounting for 56.9% of the total variance) shows indeed a quite clear separation between the clusters and the medoids representing them (Figure 7). Again, the observed partial superposition could be a spurious effect of data visualization through a low-dimensional PCA. According to the feature values shown by the cluster medoids, these representative structures (reported in Figure 8) actually seem to capture both the conformational (torsional) freedom of the coordinated NCS^{-} ligands (ϕ_1 and ϕ_2 torsional angles) and the different degrees of asymmetry sampled by the N3^{4-} dynamics (Table 3, please refer also to Section 3.2 for N3^{4-} features definitions). In particular, the values of continuous symmetry measure of deviation from C_2 symmetry (C_2-CSM, Section 3.2) most sampled by the MD trajectory (distribution maxima at 0.09, 0.17, 0.22, 0.31, Figure 9) are close to the values by the cluster medoids, further confirming the representation capabilities of the latter.

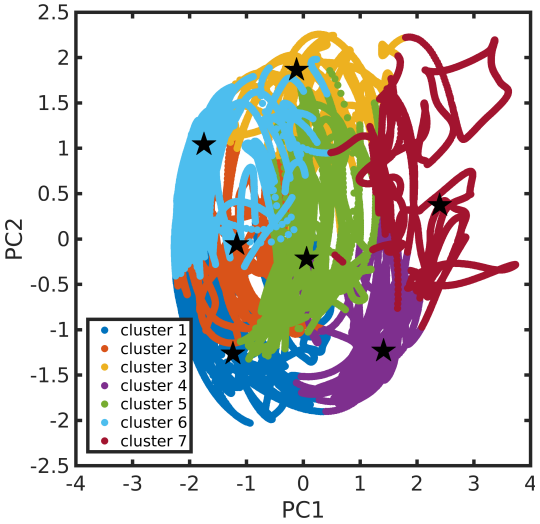

Figure 7. $N3^{4-}$ trajectory in the features' first two principal components subspace. Cluster partition is represented through different colors. Cluster medoids are also highlighted (as star symbols).

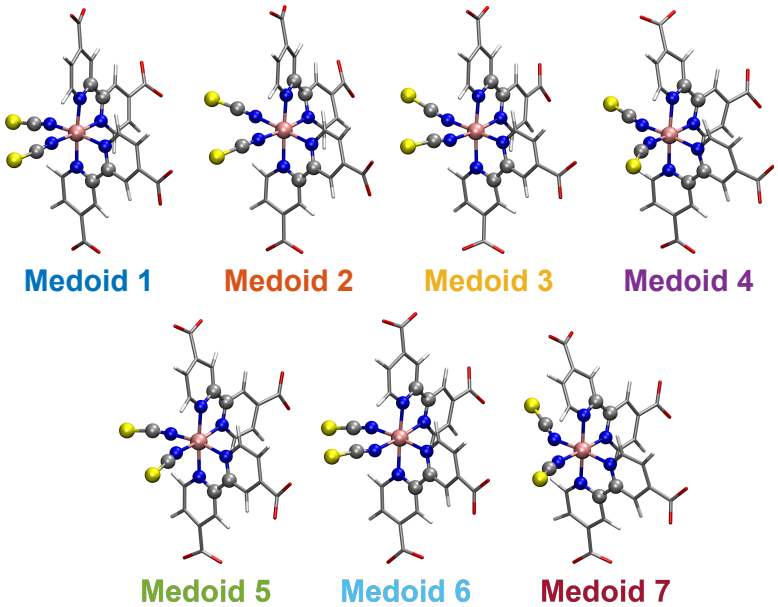

Figure 8. Structures of the $N3^{4-}$ seven cluster medoids. The atoms determining the features employed for clustering analysis are highlighted as ball and stick. The color code is uniform with that of Figure 7.

Table 3. Clustering feature values of the seven cluster medoids from $N3^{4-}$ trajectory. ϕ_1: C(NCS1)-N(NCS1)-Ru-N(dcbpy) dihedral angle (degrees), ϕ_2: C(NCS2)-N(NCS2)-Ru-N(dcbpy) dihedral angle (degrees), C_2-CSM: continuous symmetry measure for deviation from C_2 symmetry.

Medoid	ϕ_1	ϕ_2	C_2-CSM
1	−30.77	5.57	0.176
2	−54.10	107.01	0.172
3	−143.43	140.61	0.219
4	52.26	−131.41	0.174
5	69.09	83.05	0.215
6	−127.10	40.02	0.116
7	107.85	−137.07	0.352

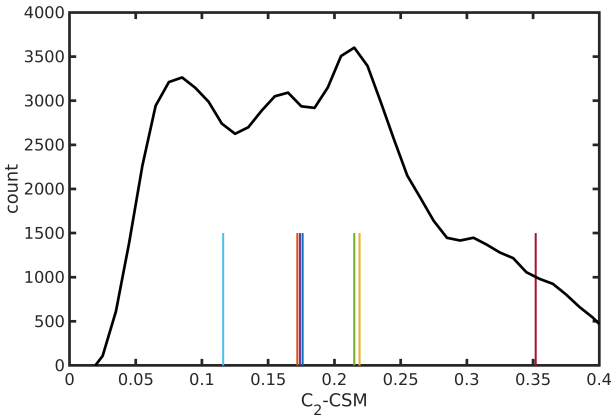

Figure 9. Distribution of C_2-CSM symmetry deviation parameter from $N3^{4-}$ trajectory in water solution. Values of the medoid structures from trajectory clustering analysis are also shown as vertical bars (with arbitrary heights). The color code is uniform with that of Figure 7.

Electronic absorption spectra of transition metal complexes are determined by several, closely spaced, excited states, differing in their spatial properties (i.e., metal and ligand-localized transitions, metal-to-ligand (ML) and ligand-to-metal (LM) charge-transfer (CT) transitions). From a practical point of view, this implies the computation (and characterization) of a high number of excited states with some level of theory to simulate the spectrum in a given energy range. Therefore, a clustering analysis performed on an MD trajectory (and so reducing the complete configuration dataset to a few, representative, structures) potentially appears even more convenient for the simulation and the interpretation of transition metal complex electronic spectra including finite-temperature effects.

The $N3^{4-}$ electronic spectrum was simulated up to ∼3.7 eV, comprising the two experimentally characterized bands at ∼2.50 and ∼3.36 eV [114]. The spectra calculated for each cluster medoid actually slightly differ in the absorption band positions (energies) and intensities, since the medoids represent different regions of the accessible conformational space (Figure 10). In particular, the spectrum obtained from the only seven representative frames can actually quite well reproduce that from the complete MD sampling at $T = 298\ K$ in water solution (Figure 11), although with an increased sub-structure, due to the lower number of frames involved in the spectrum calculation. The selection of representative frames through a clustering analysis allowed one therefore to achieve a remarkable ∼70-fold decrease in the total computational cost for $N3^{4-}$ electronic spectrum simulation.

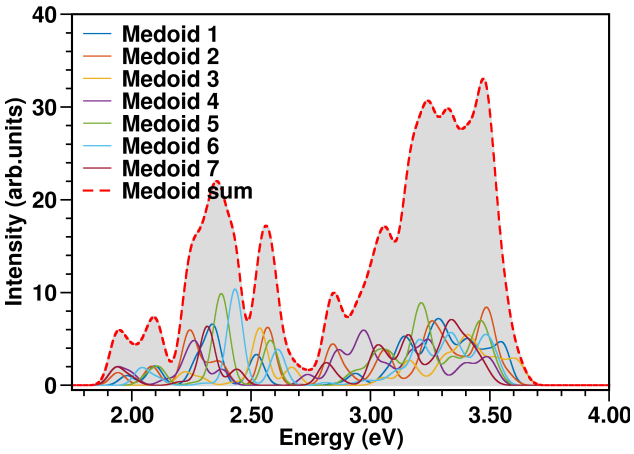

Figure 10. $N3^{4-}$ absorption spectra (in eV) calculated at TD-B3LYP/C-PCM/def2-SVP/SDD(Ru) level of theory from each medoid, weighted by the population of the corresponding cluster and the spectrum resulting from the sum over the medoids (red dashed curve). The color code is presented in the graph legend. See Equation (5) and the procedure explained in Section 3.4 for more details.

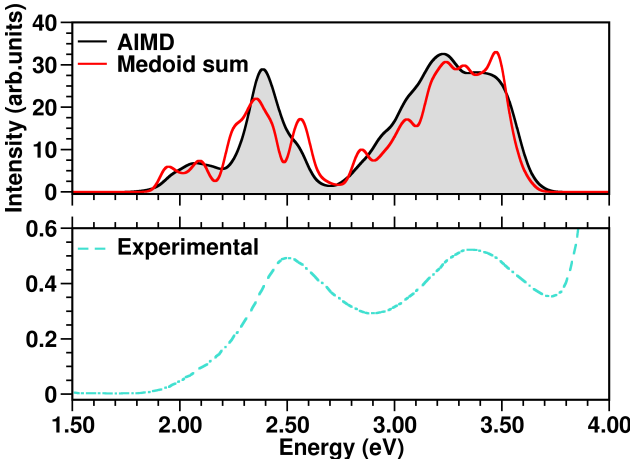

Figure 11. Top panel: comparison of $N3^{4-}$ simulated absorption spectra in the 1.50–4.00 eV range. **Bottom** panel: experimental $N3^{4-}$ UV-Vis spectrum, retrieved from Ref. [114], measured in water solution. The color code is presented in the graph legend.

Especially for transition metal complexes, the observed absorption bands can each be the result of many close transitions. Dataset reduction through a clustering analysis allowed an accurate $N3^{4-}$ spectral characterization, which could be otherwise difficult to perform. In particular, the calculated band at 2.07 eV (Figure 11) results from medoid 1 S_2, medoid 2 S_2, medoid 7 S_1, medoid 6 S_1 and medoid 5 S_2 states, which are mainly Ru→$(dcbpy)_2$ ($\Omega_{RP} \approx 0.55$, $\Omega_{SP} \approx 0.25$, Table 4) CT states. Analogously, medoid 6 S_5, medoid 7 S_5, medoid 2 S_6, medoid 1 S_5 and medoid 3 S_5, with similar metal-to-ligand charge-transfer (MLCT) spatial features, contribute to the more intense calculated band at 2.39 eV. The higher-energy bands are characterized instead by a less homogeneous set of excited states. In fact, the calculated band at 3.23 eV results from medoid 2 S_8, medoid 7 S_{13} Ru→$(dcbpy)_2$ states ($\Omega_{RP} \approx 0.55$, $\Omega_{SP} \approx 0.25$, Table 4), medoid 2 S_{18} Ru→$(dcbpy)_2$ state, but with an increased dcbpy localized-excitation character ($\Omega_{RP} \approx 0.40$, $\Omega_{PP} \approx 0.30$), medoid 4 S_{15} Ru→$(dcbpy)_2$ state, with increased $(NCS)_2$ donor contribution ($\Omega_{RP} \approx 0.50$, $\Omega_{SP} \approx 0.40$) and medoid 7 S_{19} state, which is mainly an $(NCS)_2$→$(dcbpy)_2$ CT state ($\Omega_{SP} \approx 0.40$, $\Omega_{RP} \approx 0.30$). The close calculated 3.38 eV band has instead a quite different average character. In fact, the contributing medoid 1 S_{40} state is mostly an $(NCS)_2$→$(dcbpy)_2$ CT state ($\Omega_{SP} \approx 0.60$), medoid 2 S_{37} and medoid 5 S_{33} states have an increased localized character ($\Omega_{SP} \approx 0.40$, $\Omega_{PP} \approx 0.30$), while medoid 3 S_{21} and medoid 6 S_{34} are localized excitations on dcbpy ligands ($\Omega_{PP} \approx 0.60$ and ≈ 0.50, respectively).

Table 4. Characterization of the excited states of $N3^{4-}$ cluster medoids most contributing to the calculated absorption bands. ν_i (eV): vertical excitation energy, f_i: oscillator strength, Ω_{AB}: transition density population analysis for A (hole) and B (electron) fragments, ω_{CT}: charge transfer descriptor (please refer to Section 3.4 for definitions). Fragment labels: S: $(NCS)_2$, R: Ru, P: $(dcbpy)_2$.

Medoid		ν_i	f_i	Ω_{SP}	Ω_{RP}	Ω_{PP}	ω_{CT}
1	S_2	2.097	0.029	0.269	0.566	0.113	0.858
	S_5	2.342	0.086	0.284	0.542	0.119	0.851
	S_{40}	3.556	0.056	0.578	0.235	0.127	0.851
2	S_2	2.083	0.026	0.235	0.580	0.111	0.846
	S_6	2.569	0.079	0.323	0.507	0.103	0.862
	S_8	2.840	0.055	0.275	0.536	0.114	0.843
	S_{18}	3.237	0.046	0.265	0.422	0.271	0.713
	S_{37}	3.498	0.063	0.401	0.219	0.310	0.662
3	S_5	2.536	0.118	0.294	0.545	0.108	0.862
	S_{21}	3.389	0.043	0.133	0.215	0.583	0.391

Table 4. *Cont.*

Medoid		ν_i	f_i	Ω_{SP}	Ω_{RP}	Ω_{PP}	ω_{CT}
4	S_{15}	2.983	0.056	0.366	0.511	0.087	0.894
5	S_2	2.114	0.022	0.276	0.572	0.086	0.876
	S_{33}	3.454	0.057	0.373	0.277	0.289	0.693
6	S_1	2.041	0.025	0.242	0.627	0.093	0.884
	S_5	2.433	0.141	0.267	0.570	0.109	0.859
	S_{34}	3.472	0.030	0.335	0.099	0.521	0.469
7	S_1	1.935	0.029	0.249	0.575	0.136	0.842
	S_5	2.316	0.101	0.267	0.551	0.121	0.844
	S_{13}	3.022	0.056	0.278	0.553	0.148	0.840
	S_{19}	3.165	0.059	0.427	0.312	0.212	0.773

3. Materials and Methods

3.1. Ab Initio Molecular Dynamics

The conformational flexibility of the TCNE:π:1ClN and $N3^{4-}$ model systems were sampled through ab initio molecular dynamics simulations. In particular, the Atom-centered Density Matrix Propagation (ADMP) method was employed: the density matrix in an orthonormalized atomic basis is included in an extended Lagrangian as an additional degree of freedom and propagated together with the nuclear degrees, avoiding a self-consistent procedure at each step [115–119].

The TCNE:π:1ClN ground state trajectory was collected for 10 ps with a 0.2 fs time step, at the B3LYP/6-31G(d,p) [120–122] level of theory [35,94]. Temperature was kept at 298 K, through a velocity rescaling every 1 ps. Dichloromethane solvent effects were included through the conductor-like polarizable continuum model (C-PCM) [123–128]. Moreover, due to the π-stacked, non-covalent nature of the TCNE:1ClN complex, dispersion forces had to be modeled, employing Grimme's correcting potential (GD3) [129–134].

The $N3^{4-}$ system was simulated instead for 8.6 ps with a 0.1 fs time step [111]. A velocity rescaling every 1 ps allowed to keep a 298 K temperature. Explicit water solvation was included in the $N3^{4-}$ ground state sampling, in order to better model the specific solute–solvent interactions at the several solvation sites. A 22 Å-radius spherical solvent box (∼1500 molecules) was extracted from a pre-equilibrated cubic one and placed around $N3^{4-}$. A hybrid quantum mechanics/molecular mechanics potential was employed: B3LYP/def2-SVP [135] for the QM portion (the $N3^{4-}$ molecule) with associated electronic core potential for the Ru atom [136] and the TIP3P water model [137] for the MM part (the water spherical box), re-parametrized to allow a bending motion [11]. The QM and MM potentials were combined through the ONIOM QM/MM scheme [138–140], including the MM charges into the QM hamiltonian (i.e., an "electronic embedding"). General AMBER Force Field [141] atom types (and so van der Waals non-bonding parameters) were assigned, moreover, to $N3^{4-}$ atoms. Non-periodic boundary conditions were introduced through a hybrid explicit/implicit solvent model. Long-range electrostatic effects and short-range dispersion–repulsion forces between the explicit and the bulk solvent were, respectively, modeled through C-PCM self-reaction field and an empirical confining potential, which has to be parametrized for the specific solvent model [10,124,142–144]. We refer the reader to previous works for more details about the ab initio molecular dynamics simulations of the model systems and the employed potentials [34,94,111].

3.2. Feature Selection and Clustering of Molecular Dynamics Trajectories

Due to its large dimensions, it is often useful to transform the original dataset of the collected, N-frames long, trajectory (the configurations, i.e., the positions of each of the N_{at} atoms in the system, at each time step) into a matrix $\boldsymbol{X} \in \mathbb{R}^{N \times d}$, representing the data in some d-dimensional ($d \ll 3N_{at}$) feature space, different from the coordinate space. The chosen features should adequately describe the properties of interest, without much loss of information [77]. Internal coordinates, such as bonds, angles and dihedrals or more

specifically tailored parameters, according to the problem under study, can be employed as features.

In particular, the TCNE:π:1ClN trajectory was transformed into a feature space able to describe the orientation of the two molecular planes and the relative position of the two molecules, comprising the angle ($[0, 180°]$) between the versors normal to the TCNE and ClN planes, the versor representing the axis of rotation of the two planes (i.e., the versor orthogonal to the former ones) and the relative position vector (i.e., the vector between the two geometric centers). For $N3^{4-}$, instead, a continuous symmetry measure of deviation from the ideal C_2 symmetry [145–147], calculated as the minimized root-mean-square deviation from the images generated through the C_2 symmetry operations, was considered. In particular, C_2-CSM was evaluated on the smallest subset of $N3^{4-}$ atoms showing a symmetry not higher than C_2, as the complete molecule. Since the non-linearity of the NCS^- coordination in the water solution (C(NCS)-N(NCS)-Ru angle less than 180°) and their torsional mobility were previously recognized [111], the C(NCS)-N(NCS)-Ru-N(dcbpy) dihedrals describing the NCS^- orientations were also included. In this regard, to avoid problems due to the periodicity around $\pm 180°$, each dihedral ϕ was included as a $(\cos(\phi), \sin(\phi))$ pair to keep a metric feature space [148]. The MD datasets in the feature space $\boldsymbol{X}$ were standardized (i.e., shifted to zero mean and scaled to unit variance) before following analyses.

Clustering machine learning techniques allow one to partition a dataset, grouping similar instances according to a similarity measure, such as a metric (for instance, Euclidean) in the feature space [149]. Instances within a cluster should be similar to each other and different from those belonging to the other clusters. In K-Means [150] and K-Medoids [151–153] approaches, for a given number K of clusters, K cluster centers are obtained. The feature space is partitioned (tessellated) by assigning each instance to the closest center. The latter are found by minimization of a loss function, defined as the sum of the squared distances between each instance and the cluster center to which it is assigned:

$$L(\boldsymbol{c}_k) = \sum_{i}^{N} \|\boldsymbol{x}_i - \boldsymbol{c}_k\|^2 \tag{1}$$

where $\boldsymbol{x}_i$ belongs to the cluster k, $\boldsymbol{c}_k$ is the corresponding center in the feature space and N is the number of "observations" (trajectory frames). While in the K-Means algorithm the cluster center is the mean of the cluster members and so does not have to correspond to any instance $\boldsymbol{x}_i$ of the dataset, in the K-Medoids approach it is forced to be some $\boldsymbol{x}_i$, such that the sum of the squared distances from the cluster members is the lowest (like a median).

MD trajectories of TCNE:π:1ClN and $N3^{4-}$ model systems in their respective feature spaces were clusterized with the K-Medoids algorithm [152], since the cluster medoids, which are representative of the corresponding clusters, are trajectory frames themselves and, compared to K-Means centroids, should be less sensitive to possible outliers [151].

The optimal number of clusters K was chosen searching for an "elbow" (i.e., a slope change) in the plot of the inertia parameter (i.e., the minimized value of Equation (1)) as a function of K and evaluating the Calinski–Harabasz index [154], which is the ratio of between-cluster and within-cluster dispersions, being higher for a better clusterization into compact and separated clusters.

3.3. Dimensionality Reduction for MD Data Visualization

Principal component analysis (PCA) [155] is a popular dimensionality reduction technique. For some centered (i.e., zero-mean) data matrix $\boldsymbol{X}$, its principal components $\boldsymbol{v}_j$ are the eigenvectors of its covariance matrix $\boldsymbol{C}$:

$$\boldsymbol{C} = \frac{1}{N}\boldsymbol{X}^T\boldsymbol{X} \qquad \boldsymbol{C}\boldsymbol{v}_j = \lambda_j \boldsymbol{v}_j \tag{2}$$

It can be shown that $\boldsymbol{v}_1$ (the eigenvector corresponding to the largest eigenvalue λ_1) is the direction along which the variance of the data is highest, $\boldsymbol{v}_2$ is the direction of highest variance in the subspace orthogonal to $\boldsymbol{v}_1$, etc., while the eigenvalues λ_j are the variance of the data along each $\boldsymbol{v}_j$. Projection of the data on the subspace of the first $d_r \leq d$ principal components can be performed via the following:

$$\boldsymbol{V} = \left(\boldsymbol{v}_1 \dots \boldsymbol{v}_{d_r}\right) \qquad \boldsymbol{X}_r = \boldsymbol{X}\boldsymbol{V} \tag{3}$$

where $\sum_j^{d_r} \lambda_j$ is the variance retained in the PC subspace.

PCA dimensionality reduction ($d_r = 2$) of TCNE:π:1ClN and $N3^{4-}$ trajectories was performed only for data visualization purposes on two-dimensional plots and not as a pre-processing step for clustering analysis. In fact, the dimensionalities of their respective feature spaces (Section 3.2) are actually quite small, likely not involving any "curse of dimensionality" issues.

3.4. Excited State Characterization and Spectra Simulations

TCNE:π:1ClN and $N3^{4-}$ excited states were computed with the linear-response TD-DFT approach at CAM-B3LYP/GD3/C-PCM(DCM)/6-31+G(d,p) and B3LYP/C-PCM(water)/def2-SVP/SDD(Ru) levels of theory, respectively. Electronic spectra in the solution at T = 298 K were simulated on 500 frame subsets of the collected MD samplings (i.e., every 20 fs and 17.2 fs, respectively). The first 8 and 40 singlet excited states were calculated for TCNE:π:1ClN and $N3^{4-}$, respectively. The complete spectra were obtained by summation of Gaussian-shaped contributions over each frame and each calculated excited state:

$$S_{li}(\nu) = f_{li}\, e^{-\frac{1}{2}\left(\frac{\nu-\nu_{li}}{\sigma}\right)^2} \qquad S(\nu) = \sum_{l}^{N_{\mathrm{fr}}} \sum_{i}^{N_{\mathrm{st}}} S_{li}(\nu) \tag{4}$$

where f_{li} and ν_{li} are the oscillator strength and excitation energy of the i-th state of l-th frame and σ is a width parameter, set at $\sigma^2 = 0.001$ eV^2. Spectra estimated from the only cluster medoids were similarly calculated:

$$S(\nu) = \sum_{k}^{K} p_k \sum_{i}^{N_{\mathrm{st}}} S_{ki}(\nu) \tag{5}$$

where K is the number of clusters and p_k is the k-th cluster population.

Cluster medoid excited states were further characterized by fragment-based transition density Löwdin population analysis and related charge transfer descriptors, calculated with the TheoDORE package [156,157]:

$$\Omega_{AB} = \sum_{\mu \in A} \sum_{\nu \in B} \left(\boldsymbol{S}^{1/2}\boldsymbol{D}^{0i}\boldsymbol{S}^{1/2}\right)^2_{\mu\nu} \tag{6}$$

$$\omega_{\mathrm{CT}} = \frac{\sum_{A,B\neq A} \Omega_{AB}}{\sum_{A,B} \Omega_{AB}} \tag{7}$$

where A and B are two molecular fragments and $\boldsymbol{D}^{0i}$ is the transition density matrix for the $S_i \leftarrow S_0$ excitation.

Ab initio molecular dynamics simulations and excited state calculations were performed with the Gaussian16 software package [158].

4. Conclusions

Unsupervised clustering methods have been employed as complementary approaches for an efficient exploration of the data resulting from MD simulations and electronic structure calculations. In this work, MD dataset reduction capabilities via unsupervised clustering techniques were applied for the ab initio modeling of electronic absorption spectra of

the non-covalent charge-transfer TCNE:π:1ClN dimer and the $[Ru(dcbpy)_2(NCS)_2]^{4-}$ complex in solution at room temperature. Cluster medoids, taken as representative structures, were found and analyzed in terms of main structural parameters, principal component dynamics, electronic excitations and charge transfer indices, showing how such medoids can satisfactorily cover the system dynamics and optical properties with a very good agreement with experiments.

The simulation of electronic absorption spectra usually demands expensive computations and requires dealing with the interplay of electronic excited states with the conformational freedom of the chromophores in complex matrices (i.e., solvents, biomolecules, crystals) at finite temperature. This work highlights the power of the unsupervised K-medoid clustering technique combined with a tailored selection of the feature space in reducing by $\sim$100 times the total cost of electronic and optical property computations on an MD sampling with no loss of accuracy and in preserving the molecular interpretation via the cluster medoids. In this regard, it could be very interesting to study how the medoids and the several conformational minima are related and this is a subject for further spectroscopic and weighting scheme developments.

Author Contributions: F.P.: data curation, formal analysis, software, investigation, writing—original draft, writing—review & editing; F.C.: data curation, formal analysis, investigation, writing—original draft, writing—review & editing; N.R.: conceptualization, funding acquisition, methodology, writing—review & editing; A.P.: conceptualization, methodology, supervision, validation, writing—original draft, writing—review & editing. All authors have read and agreed to the published version of the manuscript.

Funding: Authors thank Gaussian Inc. and the Italian Ministry of University and Research (Projects: PRIN 2017YJMPZN001, PRIN 202082CE3T_002) for financial support.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

1ClN	1-chloronaphthalene
ADMP	Atom-centered density matrix propagation
C-PCM	Conductor-like polarizable continuum model
CT	Charge transfer
dcbpy	4,4′-dicarboxy-2,2′-bipyridine
DCM	dichloromethane
DFT	Density functional theory
LMCT	Ligand-to-metal charge-transfer
MD	Molecular dynamics
ML	Machine learning
MLCT	Metal-to-ligand charge-transfer
MM	Molecular mechanics
$N3^{4-}$	$[Ru(dcbpy)_2(NCS)_2]^{4-}$
PC	Principal component
PCA	Principal component analysis
PES	Potential energy surface
QM	Quantum mechanics
SDF	Spatial distribution function
TCNE	Tetracyanoethylene
TD-DFT	Time dependent density functional theory

References

1. Adamo, C.; Cossi, M.; Rega, N.; Barone, V. Chapter 12—New computational strategies for the quantum mechanical study of biological systems in condensed phases. In *Theoretical Biochemistry*; Eriksson, L.A., Ed.; Elsevier: Amsterdam, The Netherlands, 2001; Volume 9: Theoretical and Computational Chemistry; pp. 467–538. [CrossRef]
2. Barone, V.; Improta, R.; Rega, N. Quantum Mechanical Computations and Spectroscopy: From Small Rigid Molecules in the Gas Phase to Large Flexible Molecules in Solution. *Acc. Chem. Res.* **2008**, *41*, 605–616. [CrossRef] [PubMed]
3. Reichardt, C. Solvatochromic Dyes as Solvent Polarity Indicators. *Chem. Rev.* **1994**, *94*, 2319–2358. [CrossRef]
4. Barone, V.; Polimeno, A. Integrated computational strategies for UV/vis spectra of large molecules in solution. *Chem. Soc. Rev.* **2007**, *36*, 1724–1731. [CrossRef] [PubMed]
5. Krystkowiak, E.; Dobek, K.; Maciejewski, A. Origin of the strong effect of protic solvents on the emission spectra, quantum yield of fluorescence and fluorescence lifetime of 4-aminophthalimide: Role of hydrogen bonds in deactivation of S1-4-aminophthalimide. *J. Photochem. Photobiol.* **2006**, *184*, 250–264. [CrossRef]
6. Solntsev, K.M.; Huppert, D.; Agmon, N. Photochemistry of "Super"-Photoacids. Solvent Effects. *J. Phys. Chem. A* **1999**, *103*, 6984–6997. [CrossRef]
7. Solntsev, K.M.; Huppert, D.; Tolbert, L.M.; Agmon, N. Solvatochromic shifts of "super" photoacids. *J. Am. Chem. Soc.* **1998**, *120*, 7981–7982. [CrossRef]
8. Coppola, F.; Nucci, M.; Marazzi, M.; Rocca, D.; Pastore, M. Norbornadiene/Quadricyclane System in the Spotlight: The Role of Rydberg States and Dynamic Electronic Correlation in a Solar-Thermal Building Block. *ChemPhotoChem* **2023**, e202200214. [CrossRef]
9. Frank, H.A.; Bautista, J.A.; Josue, J.; Pendon, Z.; Hiller, R.G.; Sharples, F.P.; Gosztola, D.; Wasielewski, M.R. Effect of the Solvent Environment on the Spectroscopic Properties and Dynamics of the Lowest Excited States of Carotenoids. *J. Phys. Chem. B* **2000**, *104*, 4569–4577. [CrossRef]
10. Raucci, U.; Perrella, F.; Donati, G.; Zoppi, M.; Petrone, A.; Rega, N. Ab-initio molecular dynamics and hybrid explicit-implicit solvation model for aqueous and nonaqueous solvents: GFP chromophore in water and methanol solution as case study. *J. Comput. Chem.* **2020**, *41*, 2228–2239. [CrossRef]
11. Donati, G.; Petrone, A.; Rega, N. Multiresolution continuous wavelet transform for studying coupled solute–solvent vibrations via ab initio molecular dynamics. *Phys. Chem. Chem. Phys.* **2020**, *22*, 22645–22661. [CrossRef]
12. Coppola, F.; Perrella, F.; Petrone, A.; Donati, G.; Rega, N. A not obvious correlation between the structure of green fluorescent protein chromophore pocket and hydrogen bond dynamics: A choreography from ab initio molecular dynamics. *Front. Mol. Biosci.* **2020**, *7*, 569990. [CrossRef]
13. Raucci, U.; Savarese, M.; Adamo, C.; Ciofini, I.; Rega, N. Intrinsic and Dynamical Reaction Pathways of an Excited State Proton Transfer. *J. Phys. Chem. B* **2015**, *119*, 2650–2657. [CrossRef]
14. Petrone, A.; Caruso, P.; Tenuta, S.; Rega, N. On the optical absorption of the anionic GFP chromophore in vacuum, solution, and protein. *Phys. Chem. Chem. Phys.* **2013**, *15*, 20536–20544. [CrossRef]
15. Langella, E.; Rega, N.; Improta, R.; Crescenzi, O.; Barone, V. Conformational analysis of the tyrosine dipeptide analogue in the gas phase and in aqueous solution by a density functional/continuum solvent model. *J. Comput. Chem.* **2002**, *23*, 650–661. [CrossRef]
16. Cerezo, J.; Petrone, A.; Ferrer, F.J.A.; Donati, G.; Santoro, F.; Improta, R.; Rega, N. Electronic spectroscopy of a solvatochromic dye in water: Comparison of static cluster/implicit and dynamical/explicit solvent models on structures and energies. *Theor. Chem. Acc.* **2016**, *135*, 263. [CrossRef]
17. Cimino, P.; Raucci, U.; Donati, G.; Chiariello, M.G.; Schiazza, M.; Coppola, F.; Rega, N. On the different strength of photoacids. *Theor. Chem. Acc.* **2016**, *135*, 117. [CrossRef]
18. Kim, P.; Valentine, A.J.S.; Roy, S.; Mills, A.W.; Chakraborty, A.; Castellano, F.N.; Li, X.; Chen, L.X. Ultrafast Excited-State Dynamics of Photoluminescent Pt(II) Dimers Probed by a Coherent Vibrational Wavepacket. *J. Phys. Chem. Lett.* **2021**, *12*, 6794–6803. [CrossRef]
19. Lu, L.; Wildman, A.; Jenkins, A.J.; Young, L.; Clark, A.E.; Li, X. The "Hole" Story in Ionized Water from the Perspective of Ehrenfest Dynamics. *J. Phys. Chem. Lett.* **2020**, *11*, 9946–9951. [CrossRef]
20. Leger, J.D.; Friedfeld, M.R.; Beck, R.A.; Gaynor, J.D.; Petrone, A.; Li, X.; Cossairt, B.M.; Khalil, M. Carboxylate Anchors Act as Exciton Reporters in 1.3 nm Indium Phosphide Nanoclusters. *J. Phys. Chem. Lett.* **2019**, *10*, 1833–1839. [CrossRef]
21. Nascimento, D.R.; Zhang, Y.; Bergmann, U.; Govind, N. Near-Edge X-ray Absorption Fine Structure Spectroscopy of Heteroatomic Core-Hole States as a Probe for Nearly Indistinguishable Chemical Environments. *J. Phys. Chem. Lett.* **2020**, *11*, 556–561. [CrossRef]
22. Alberto, M.E.; De Simone, B.C.; Mazzone, G.; Quartarolo, A.D.; Russo, N. Theoretical Determination of Electronic Spectra and Intersystem Spin–Orbit Coupling: The Case of Isoindole-BODIPY Dyes. *J. Chem. Theory Comput.* **2014**, *10*, 4006–4013. [CrossRef] [PubMed]
23. Barone, V.; Alessandrini, S.; Biczysko, M.; Cheeseman, J.R.; Clary, D.C.; McCoy, A.B.; DiRisio, R.J.; Neese, F.; Melosso, M.; Puzzarini, C. Computational molecular spectroscopy. *Nat. Rev. Methods Prim.* **2021**, *1*, 38. [CrossRef]
24. Barone, V.; Bloino, J.; Biczysko, M.; Santoro, F. Fully Integrated Approach to Compute Vibrationally Resolved Optical Spectra: From Small Molecules to Macrosystems. *J. Chem. Theory Comput.* **2009**, *5*, 540–554. [CrossRef] [PubMed]
25. Santoro, F.; Lami, A.; Improta, R.; Barone, V. Effective method to compute vibrationally resolved optical spectra of large molecules at finite temperature in the gas phase and in solution. *J. Chem. Phys.* **2007**, *126*, 184102. [CrossRef]

26. Avila Ferrer, F.J.; Cerezo, J.; Stendardo, E.; Improta, R.; Santoro, F. Insights for an Accurate Comparison of Computational Data to Experimental Absorption and Emission Spectra: Beyond the Vertical Transition Approximation. *J. Chem. Theory Comput.* **2013**, *9*, 2072–2082. [CrossRef]
27. Dierksen, M.; Grimme, S. Density functional calculations of the vibronic structure of electronic absorption spectra. *J. Chem. Phys.* **2004**, *120*, 3544–3554. [CrossRef]
28. Isborn, C.M.; Gotz, A.W.; Clark, M.A.; Walker, R.C.; Martínez, T.J. Electronic absorption spectra from MM and ab initio QM/MM molecular dynamics: Environmental effects on the absorption spectrum of photoactive yellow protein. *J. Chem. Theory Comput.* **2012**, *8*, 5092–5106. [CrossRef]
29. Pagliai, M.; Mancini, G.; Carnimeo, I.; De Mitri, N.; Barone, V. Electronic absorption spectra of pyridine and nicotine in aqueous solution with a combined molecular dynamics and polarizable QM/MM approach. *J. Chem. Theory Comput.* **2017**, *38*, 319–335. [CrossRef]
30. Mendanha, K.; Prado, R.C.; Oliveira, L.B.; Colherinhas, G. TD-DFT absorption spectrum of (poly) threonine in water: A study combining molecular dynamics and quantum mechanics calculations. *Chem. Phys. Lett.* **2021**, *779*, 138876. [CrossRef]
31. Kasper, J.M.; Williams-Young, D.B.; Vecharynski, E.; Yang, C.; Li, X. A Well-Tempered Hybrid Method for Solving Challenging Time-Dependent Density Functional Theory (TDDFT) Systems. *J. Chem. Theory Comput.* **2018**, *14*, 2034–2041. [CrossRef]
32. Van Beeumen, R.; Williams-Young, D.B.; Kasper, J.M.; Yang, C.; Ng, E.G.; Li, X. Model Order Reduction Algorithm for Estimating the Absorption Spectrum. *J. Chem. Theory Comput.* **2017**, *13*, 4950–4961. [CrossRef]
33. Alberto, M.E.; Mazzone, G.; Quartarolo, A.D.; Sousa, F.F.R.; Sicilia, E.; Russo, N. Electronic spectra and intersystem spin-orbit coupling in 1,2- and 1,3-squaraines. *J. Comput. Chem.* **2014**, *35*, 2107–2113. [CrossRef]
34. Petrone, A.; Perrella, F.; Coppola, F.; Crisci, L.; Donati, G.; Cimino, P.; Rega, N. Ultrafast photo-induced processes in complex environments: The role of accuracy in excited-state energy potentials and initial conditions. *Chem. Phys. Rev.* **2022**, *3*, 021307. [CrossRef]
35. Coppola, F.; Cimino, P.; Perrella, F.; Crisci, L.; Petrone, A.; Rega, N. Electronic and Vibrational Manifold of Tetracyanoethylene–Chloronaphthalene Charge Transfer Complex in Solution: Insights from TD-DFT and Ab Initio Molecular Dynamics. *J. Phys. Chem. A* **2022**, *126*, 7179–7192. [CrossRef]
36. Segatta, F.; Nenov, A.; Nascimento, D.R.; Govind, N.; Mukamel, S.; Garavelli, M. iSPECTRON: A simulation interface for linear and nonlinear spectra with ab-initio quantum chemistry software. *J. Comput. Chem.* **2021**, *42*, 644–659. [CrossRef]
37. Petrenko, T.; Neese, F. Analysis and prediction of absorption band shapes, fluorescence band shapes, resonance Raman intensities and excitation profiles using the time-dependent theory of electronic spectroscopy. *J. Chem. Phys.* **2007**, *127*, 164319. [CrossRef]
38. Petrone, A.; Cerezo, J.; Ferrer, F.J.A.; Donati, G.; Improta, R.; Rega, N.; Santoro, F. Absorption and Emission Spectral Shapes of a Prototype Dye in Water by Combining Classical/Dynamical and Quantum/Static Approaches. *J. Phys. Chem. A* **2015**, *119*, 5426–5438. [CrossRef]
39. De Mitri, N.; Monti, S.; Prampolini, G.; Barone, V. Absorption and emission spectra of a flexible dye in solution: A computational time-dependent approach. *J. Chem. Theory Comput.* **2013**, *9*, 4507–4516. [CrossRef]
40. Hoffman, D.P.; Ellis, S.R.; Mathies, R.A. Characterization of a conical intersection in a charge-transfer dimer with two-dimensional time-resolved stimulated Raman spectroscopy. *J. Phys. Chem. A* **2014**, *118*, 4955–4965. [CrossRef]
41. Dubinets, N.; Safonov, A.; Bagaturyants, A. Structures and binding energies of the naphthalene dimer in its ground and excited states. *J. Phys. Chem. A* **2016**, *120*, 2779–2782. [CrossRef]
42. Hancock, A.C.; Goerigk, L. Noncovalently bound excited-state dimers: A perspective on current time-dependent density functional theory approaches applied to aromatic excimer models. *RSC Adv.* **2022**, *12*, 13014–13034. [CrossRef] [PubMed]
43. Cui, Z.h.; Lischka, H.; Mueller, T.; Plasser, F.; Kertesz, M. Study of the Diradicaloid Character in a Prototypical Pancake-Bonded Dimer: The Stacked Tetracyanoethylene (TCNE) Anion Dimer and the Neutral K2TCNE2 Complex. *ChemPhysChem* **2014**, *15*, 165–176. [CrossRef] [PubMed]
44. Valente, D.C.A.; Do Casal, M.T.; Barbatti, M.; Niehaus, T.A.; Aquino, A.J.; Lischka, H.; Cardozo, T.M. Excitonic and charge transfer interactions in tetracene stacked and T-shaped dimers. *J. Chem. Phys.* **2021**, *154*, 044306. [CrossRef] [PubMed]
45. Siddique, F.; Barbatti, M.; Cui, Z.; Lischka, H.; Aquino, A.J. Nonadiabatic dynamics of charge-transfer states using the anthracene–tetracyanoethylene complex as a prototype. *J. Phys. Chem. A* **2020**, *124*, 3347–3357. [CrossRef] [PubMed]
46. Mauck, C.M.; Bae, Y.J.; Chen, M.; Powers-Riggs, N.; Wu, Y.L.; Wasielewski, M.R. Charge-transfer character in a covalent diketopyrrolopyrrole dimer: Implications for singlet fission. *ChemPhotoChem* **2018**, *2*, 223–233. [CrossRef]
47. Cui, Z.h.; Aquino, A.J.; Sue, A.C.H.; Lischka, H. Analysis of charge transfer transitions in stacked π-electron donor–acceptor complexes. *Phys. Chem. Chem. Phys.* **2018**, *20*, 26957–26967. [CrossRef]
48. Müller-Dethlefs, K.; Hobza, P. Noncovalent interactions: A challenge for experiment and theory. *Chem. Rev.* **2000**, *100*, 143–168. [CrossRef]
49. Snyder, J.W.; Fales, B.S.; Hohenstein, E.G.; Levine, B.G.; Martínez, T.J. A direct-compatible formulation of the coupled perturbed complete active space self-consistent field equations on graphical processing units. *J. Chem. Phys.* **2017**, *146*, 174113. [CrossRef]
50. Demel, O.; Pittner, J.; Neese, F. A Local Pair Natural Orbital-Based Multireference Mukherjee's Coupled Cluster Method. *J. Chem. Theory Comput.* **2015**, *11*, 3104–3114. [CrossRef]
51. Riplinger, C.; Sandhoefer, B.; Hansen, A.; Neese, F. Natural triple excitations in local coupled cluster calculations with pair natural orbitals. *J. Chem. Phys.* **2013**, *139*, 134101. [CrossRef]

52. Chiariello, M.G.; Donati, G.; Raucci, U.; Perrella, F.; Rega, N. Structural Origin and Vibrational Fingerprints of the Ultrafast Excited State Proton Transfer of the Pyranine-Acetate Complex in Aqueous Solution. *J. Phys. Chem. B* **2021**, *125*, 10273–10281. [CrossRef]
53. Chiariello, M.G.; Raucci, U.; Donati, G.; Rega, N. Water-mediated excited state proton transfer of pyranine–acetate in aqueous solution: Vibrational fingerprints from ab initio molecular dynamics. *J. Phys. Chem. A* **2021**, *125*, 3569–3578. [CrossRef]
54. De Simone, B.C.; Alberto, M.E.; Marino, T.; Russo, N.; Toscano, M. The Contribution of Density Functional Theory to the Atomistic Knowledge of Electrochromic Processes. *Molecules* **2021**, *26*, 5793. [CrossRef]
55. Raucci, U.; Chiariello, M.G.; Rega, N. Modeling excited-state proton transfer to solvent: A dynamics study of a super photoacid with a hybrid implicit/explicit solvent model. *J. Chem. Theory Comput.* **2020**, *16*, 7033–7043. [CrossRef]
56. Chiariello, M.G.; Donati, G.; Rega, N. Time-resolved vibrational analysis of excited state ab initio molecular dynamics to understand photorelaxation: The case of the pyranine photoacid in aqueous solution. *J. Chem. Theory Comput.* **2020**, *16*, 6007–6013. [CrossRef]
57. Raucci, U.; Chiariello, M.G.; Coppola, F.; Perrella, F.; Savarese, M.; Ciofini, I.; Rega, N. An electron density based analysis to establish the electronic adiabaticity of proton coupled electron transfer reactions. *J. Comput. Chem.* **2020**, *41*, 1835–1841. [CrossRef]
58. Chiariello, M.G.; Raucci, U.; Coppola, F.; Rega, N. Unveiling anharmonic coupling by means of excited state ab initio dynamics: Application to diarylethene photoreactivity. *Phys. Chem. Chem. Phys.* **2019**, *21*, 3606–3614. [CrossRef]
59. Chiariello, M.G.; Rega, N. Exploring nuclear photorelaxation of pyranine in aqueous solution: An integrated ab-initio molecular dynamics and time resolved vibrational analysis approach. *J. Phys. Chem. A* **2018**, *122*, 2884–2893. [CrossRef]
60. Perrella, F.; Raucci, U.; Chiariello, M.G.; Chino, M.; Maglio, O.; Lombardi, A.; Rega, N. Unveiling the structure of a novel artificial heme-enzyme with peroxidase-like activity: A theoretical investigation. *Biopolymers* **2018**, *109*, e23225. [CrossRef]
61. Williams-Young, D.B.; Bagusetty, A.; de Jong, W.A.; Doerfler, D.; van Dam, H.J.; Vázquez-Mayagoitia, Á.; Windus, T.L.; Yang, C. Achieving performance portability in Gaussian basis set density functional theory on accelerator based architectures in NWChemEx. *Parallel Comput.* **2021**, *108*, 102829. [CrossRef]
62. Williams-Young, D.B.; Petrone, A.; Sun, S.; Stetina, T.F.; Lestrange, P.; Hoyer, C.E.; Nascimento, D.R.; Koulias, L.; Wildman, A.; Kasper, J.; et al. The Chronus Quantum software package. *WIREs Comput. Mol. Sci.* **2020**, *10*, e1436. [CrossRef]
63. Artrith, N.; Butler, K.T.; Coudert, F.X.; Han, S.; Isayev, O.; Jain, A.; Walsh, A. Best practices in machine learning for chemistry. *Nat. Chem.* **2021**, *13*, 505–508. [CrossRef] [PubMed]
64. Butler, K.T.; Davies, D.W.; Cartwright, H.; Isayev, O.; Walsh, A. Machine learning for molecular and materials science. *Nature* **2018**, *559*, 547–555. [CrossRef]
65. Pflüger, P.M.; Glorius, F. Molecular Machine Learning: The Future of Synthetic Chemistry? *Angew. Chem.* **2020**, *59*, 18860–18865. [CrossRef] [PubMed]
66. Stocker, S.; Csányi, G.; Reuter, K.; Margraf, J.T. Machine learning in chemical reaction space. *Nat. Chem.* **2020**, *11*, 5505. [CrossRef]
67. Sanchez-Lengeling, B.; Aspuru-Guzik, A. Inverse molecular design using machine learning: Generative models for matter engineering. *Science* **2018**, *361*, 360–365. [CrossRef]
68. Dral, P.O. Quantum Chemistry in the Age of Machine Learning. *J. Phys. Chem. Lett.* **2020**, *11*, 2336–2347. [CrossRef]
69. Keith, J.A.; Vassilev-Galindo, V.; Cheng, B.; Chmiela, S.; Gastegger, M.; Müller, K.R.; Tkatchenko, A. Combining Machine Learning and Computational Chemistry for Predictive Insights Into Chemical Systems. *Chem. Rev.* **2021**, *121*, 9816–9872. [CrossRef]
70. Ramakrishnan, R.; Dral, P.O.; Rupp, M.; von Lilienfeld, O.A. Big Data Meets Quantum Chemistry Approximations: The Δ-Machine Learning Approach. *J. Chem. Theory Comput.* **2015**, *11*, 2087–2096. [CrossRef]
71. von Lilienfeld, O.A.; Müller, K.R.; Tkatchenko, A. Exploring chemical compound space with quantum-based machine learning. *Nat. Rev. Chem.* **2020**, *4*, 347–358. [CrossRef]
72. Häse, F.; Roch, L.M.; Friederich, P.; Aspuru-Guzik, A. Designing and understanding light-harvesting devices with machine learning. *Nat. Chem.* **2020**, *11*, 4587. [CrossRef]
73. Rosen, A.S.; Iyer, S.M.; Ray, D.; Yao, Z.; Aspuru-Guzik, A.; Gagliardi, L.; Notestein, J.M.; Snurr, R.Q. Machine learning the quantum-chemical properties of metal–organic frameworks for accelerated materials discovery. *Matter* **2021**, *4*, 1578–1597. [CrossRef]
74. Häse, F.; Galván, I.F.; Aspuru-Guzik, A.; Lindh, R.; Vacher, M. How machine learning can assist the interpretation of ab initio molecular dynamics simulations and conceptual understanding of chemistry. *Chem. Sci.* **2019**, *10*, 2298–2307. [CrossRef]
75. Häse, F.; Valleau, S.; Pyzer-Knapp, E.; Aspuru-Guzik, A. Machine learning exciton dynamics. *Chem. Sci.* **2016**, *7*, 5139–5147. [CrossRef]
76. Schriber, J.B.; Nascimento, D.R.; Koutsoukas, A.; Spronk, S.A.; Cheney, D.L.; Sherrill, C.D. CLIFF: A component-based, machine-learned, intermolecular force field. *J. Chem. Phys.* **2021**, *154*, 184110. [CrossRef]
77. Glielmo, A.; Husic, B.E.; Rodriguez, A.; Clementi, C.; Noé, F.; Laio, A. Unsupervised Learning Methods for Molecular Simulation Data. *Chem. Rev.* **2021**, *121*, 9722–9758. [CrossRef]
78. Falbo, E.; Fusè, M.; Lazzari, F.; Mancini, G.; Barone, V. Integration of Quantum Chemistry, Statistical Mechanics and Artificial Intelligence for Computational Spectroscopy: The UV–Vis Spectrum of TEMPO Radical in Different Solvents. *J. Chem. Theory Comput.* **2022**, *18*, 6203–6216. [CrossRef]

79. Mancini, G.; Fusè, M.; Lipparini, F.; Nottoli, M.; Scalmani, G.; Barone, V. Molecular Dynamics Simulations Enforcing Nonperiodic Boundary Conditions: New Developments and Application to the Solvent Shifts of Nitroxide Magnetic Parameters. *J. Chem. Theory Comput.* **2022**, *18*, 2479–2493. [CrossRef]
80. Mancini, G.; Fusè, M.; Lazzari, F.; Barone, V. Fast exploration of potential energy surfaces with a joint venture of quantum chemistry, evolutionary algorithms and unsupervised learning. *Digit. Discov.* **2022**, *1*, 790–805. [CrossRef]
81. Barone, V.; Puzzarini, C.; Mancini, G. Integration of theory, simulation, artificial intelligence and virtual reality: A four-pillar approach for reconciling accuracy and interpretability in computational spectroscopy. *Phys. Chem. Chem. Phys.* **2021**, *23*, 17079–17096. [CrossRef]
82. Mancini, G.; Del Galdo, S.; Chandramouli, B.; Pagliai, M.; Barone, V. Computational Spectroscopy in Solution by Integration of Variational and Perturbative Approaches on Top of Clusterized Molecular Dynamics. *J. Chem. Theory Comput.* **2020**, *16*, 5747–5761. [CrossRef] [PubMed]
83. Del Galdo, S.; Chandramouli, B.; Mancini, G.; Barone, V. Assessment of Multi-Scale Approaches for Computing UV–Vis Spectra in Condensed Phases: Toward an Effective yet Reliable Integration of Variational and Perturbative QM/MM Approaches. *J. Chem. Theory Comput.* **2019**, *15*, 3170–3184. [CrossRef] [PubMed]
84. Troyer, J.M.; Cohen, F.E. Protein conformational landscapes: Energy minimization and clustering of a long molecular dynamics trajectory. *Proteins* **1995**, *23*, 97–110. [CrossRef] [PubMed]
85. Wolf, A.; Kirschner, K.N. Principal component and clustering analysis on molecular dynamics data of the ribosomal L11·23S subdomain. *J. Mol. Model.* **2013**, *19*, 539–549. [CrossRef] [PubMed]
86. Papaleo, E.; Mereghetti, P.; Fantucci, P.; Grandori, R.; De Gioia, L. Free-energy landscape, principal component analysis and structural clustering to identify representative conformations from molecular dynamics simulations: The myoglobin case. *J. Mol. Graph. Model.* **2009**, *27*, 889–899. [CrossRef]
87. Shao, J.; Tanner, S.W.; Thompson, N.; Cheatham, T.E. Clustering Molecular Dynamics Trajectories: 1. Characterizing the Performance of Different Clustering Algorithms. *J. Chem. Theory Comput.* **2007**, *3*, 2312–2334. [CrossRef]
88. Torda, A.E.; van Gunsteren, W.F. Algorithms for clustering molecular dynamics configurations. *J. Comput. Chem.* **1994**, *15*, 1331–1340. [CrossRef]
89. Phillips, J.L.; Colvin, M.E.; Newsam, S. Validating clustering of molecular dynamics simulations using polymer models. *BMC Bioinform.* **2011**, *12*, 445. [CrossRef]
90. Karpen, M.E.; Tobias, D.J.; Brooks, C.L.I. Statistical clustering techniques for the analysis of long molecular dynamics trajectories: Analysis of 2.2-ns trajectories of YPGDV. *Biochemistry* **1993**, *32*, 412–420. [CrossRef]
91. Peng, J.H.; Wang, W.; Yu, Y.Q.; Gu, H.L.; Huang, X. Clustering algorithms to analyze molecular dynamics simulation trajectories for complex chemical and biological systems. *Chin. J. Chem. Phys.* **2018**, *31*, 404–420. [CrossRef]
92. González-Alemán, R.; Hernández-Castillo, D.; Rodríguez-Serradet, A.; Caballero, J.; Hernández-Rodríguez, E.W.; Montero-Cabrera, L. BitClust: Fast Geometrical Clustering of Long Molecular Dynamics Simulations. *J. Chem. Inf. Model.* **2020**, *60*, 444–448. [CrossRef]
93. Ellis, S.R.; Hoffman, D.P.; Park, M.; Mathies, R.A. Difference bands in time-resolved femtosecond stimulated Raman spectra of photoexcited intermolecular electron transfer from chloronaphthalene to tetracyanoethylene. *J. Phys. Chem. A* **2018**, *122*, 3594–3605. [CrossRef]
94. Coppola, F.; Cimino, P.; Raucci, U.; Chiariello, M.G.; Petrone, A.; Rega, N. Exploring the Franck–Condon region of a photoexcited charge transfer complex in solution to interpret femtosecond stimulated Raman spectroscopy: Excited state electronic structure methods to unveil non-radiative pathways. *Chem. Sci.* **2021**, *12*, 8058–8072. [CrossRef]
95. Hagfeldt, A.; Boschloo, G.; Sun, L.; Kloo, L.; Pettersson, H. Dye-Sensitized Solar Cells. *Chem. Rev.* **2010**, *110*, 6595–6663. [CrossRef]
96. Grätzel, M. Dye-sensitized solar cells. *J. Photochem. Photobiol. C* **2003**, *4*, 145–153. [CrossRef]
97. Grätzel, M. Solar Energy Conversion by Dye-Sensitized Photovoltaic Cells. *Inorg. Chem.* **2005**, *44*, 6841–6851. [CrossRef]
98. McCusker, J.K.; Vlcek, A., Jr. Ultrafast Excited-State Processes in Inorganic Systems. *Acc. Chem. Res.* **2015**, *48*, 1207–1208. [CrossRef]
99. Chergui, M. Ultrafast Photophysics of Transition Metal Complexes. *Acc. Chem. Res.* **2015**, *48*, 801–808. [CrossRef]
100. Pettersson Rimgard, B.; Föhlinger, J.; Petersson, J.; Lundberg, M.; Zietz, B.; Woys, A.M.; Miller, S.A.; Wasielewski, M.R.; Hammarström, L. Ultrafast interligand electron transfer in *cis*-[Ru(4,4′-dicarboxylate-2,2′-bipyridine)$_2$(NCS)$_2$]$^{4-}$ and implications for electron injection limitations in dye sensitized solar cells. *Chem. Sci.* **2018**, *9*, 7958–7967. [CrossRef]
101. Waterland, M.R.; Kelley, D.F. Photophysics and Relaxation Dynamics of Ru(4,4′-Dicarboxy-2,2′-bipyridine)$_2$cis(NCS)$_2$ in Solution. *J. Phys. Chem. A* **2001**, *105*, 4019–4028. [CrossRef]
102. Atkins, A.J.; González, L. Trajectory Surface-Hopping Dynamics Including Intersystem Crossing in [Ru(bpy)$_3$]$^{2+}$. *J. Phys. Chem. Lett.* **2017**, *8*, 3840–3845. [CrossRef] [PubMed]
103. Zobel, J.P.; González, L. The Quest to Simulate Excited-State Dynamics of Transition Metal Complexes. *JACS Au* **2021**, *1*, 1116–1140. [CrossRef] [PubMed]
104. Perrella, F.; Li, X.; Petrone, A.; Rega, N. Nature of the Ultrafast Interligands Electron Transfers in Dye-Sensitized Solar Cells. *JACS Au* **2023**, *3*, 70–79. [CrossRef] [PubMed]
105. Perrella, F.; Petrone, A.; Rega, N. Understanding Charge Dynamics in Dense Electronic Manifolds in Complex Environments. *J. Chem. Theory Comput.* **2023**, *19*, 626–639. [CrossRef] [PubMed]

106. Baldini, E.; Palmieri, T.; Rossi, T.; Oppermann, M.; Pomarico, E.; Auböck, G.; Chergui, M. Interfacial Electron Injection Probed by a Substrate-Specific Excitonic Signature. *J. Am. Chem. Soc.* **2017**, *139*, 11584–11589. [CrossRef]
107. Wei, H.; Luo, J.W.; Li, S.S.; Wang, L.W. Revealing the Origin of Fast Electron Transfer in TiO_2-Based Dye-Sensitized Solar Cells. *J. Am. Chem. Soc.* **2016**, *138*, 8165–8174. [CrossRef]
108. Tiwana, P.; Docampo, P.; Johnston, M.B.; Snaith, H.J.; Herz, L.M. Electron Mobility and Injection Dynamics in Mesoporous ZnO, SnO_2 and TiO_2 Films Used in Dye-Sensitized Solar Cells. *ACS Nano* **2011**, *5*, 5158–5166. [CrossRef]
109. Katoh, R.; Furube, A.; Yoshihara, T.; Hara, K.; Fujihashi, G.; Takano, S.; Murata, S.; Arakawa, H.; Tachiya, M. Efficiencies of Electron Injection from Excited N3 Dye into Nanocrystalline Semiconductor (ZrO_2, TiO_2, ZnO, Nb_2O_2, SnO_2, In_2O_2) Films. *J. Phys. Chem. B* **2004**, *108*, 4818–4822. [CrossRef]
110. Asbury, J.B.; Ellingson, R.J.; Ghosh, H.N.; Ferrere, S.; Nozik, A.J.; Lian, T. Femtosecond IR Study of Excited-State Relaxation and Electron-Injection Dynamics of $Ru(dcbpy)_2(NCS)_2$ in Solution and on Nanocrystalline TiO_2 and Al_2O_2 Thin Films. *J. Phys. Chem. B* **1999**, *103*, 3110–3119. [CrossRef]
111. Perrella, F.; Petrone, A.; Rega, N. Direct observation of the solvent organization and nuclear vibrations of $[Ru(dcbpy)_2(NCS)_2]^{4-}$, [dcbpy = (4,4′-dicarboxy-2,2′-bipyridine)], via ab initio molecular dynamics. *Phys. Chem. Chem. Phys.* **2021**, *23*, 22885–22896. [CrossRef]
112. Brehm, M.; Thomas, M.; Gehrke, S.; Kirchner, B. TRAVIS—A free analyzer for trajectories from molecular simulation. *J. Chem. Phys.* **2020**, *152*, 164105. [CrossRef]
113. Brehm, M.; Kirchner, B. TRAVIS—A free analyzer and visualizer for Monte Carlo and molecular dynamics trajectories. *J. Chem. Inf. Model.* **2011**, *51*, 8. [CrossRef]
114. De Angelis, F.; Fantacci, S.; Selloni, A.; Nazeeruddin, M.K. Time dependent density functional theory study of the absorption spectrum of the $[Ru(4,4'\text{-}COO^{-}\text{-}2,2'\text{-}bpy)_2(X)_2]^{4-}$ (X=NCS, Cl) dyes in water solution. *Chem. Phys. Lett.* **2005**, *415*, 115–120. [CrossRef]
115. Schlegel, H.B.; Millam, J.M.; Iyengar, S.S.; Voth, G.A.; Daniels, A.D.; Scuseria, G.E.; Frisch, M.J. Ab initio molecular dynamics: Propagating the density matrix with Gaussian orbitals. *J. Chem. Phys.* **2001**, *114*, 9758–9763. [CrossRef]
116. Iyengar, S.S.; Schlegel, H.B.; Millam, J.M.; Voth, G.A.; Scuseria, G.E.; Frisch, M.J. Ab initio molecular dynamics: Propagating the density matrix with Gaussian orbitals. II. Generalizations based on mass-weighting, idempotency, energy conservation and choice of initial conditions. *J. Chem. Phys.* **2001**, *115*, 10291–10302. [CrossRef]
117. Schlegel, H.B.; Iyengar, S.S.; Li, X.; Millam, J.M.; Voth, G.A.; Scuseria, G.E.; Frisch, M.J. Ab initio molecular dynamics: Propagating the density matrix with Gaussian orbitals. III. Comparison with Born–Oppenheimer dynamics. *J. Chem. Phys.* **2002**, *117*, 8694–8704. [CrossRef]
118. Iyengar, S.S.; Schlegel, H.B.; Voth, G.A.; Millam, J.M.; Scuseria, G.E.; Frisch, M.J. Ab initio molecular dynamics: Propagating the density matrix with Gaussian orbitals. IV. Formal analysis of the deviations from born-oppenheimer dynamics. *Isr. J. Chem.* **2002**, *42*, 191–202. [CrossRef]
119. Rega, N.; Iyengar, S.S.; Voth, G.A.; Schlegel, H.B.; Vreven, T.; Frisch, M.J. Hybrid Ab-Initio/Empirical Molecular Dynamics: Combining the ONIOM Scheme with the Atom-Centered Density Matrix Propagation (ADMP) Approach. *J. Phys. Chem. B* **2004**, *108*, 4210–4220. [CrossRef]
120. Becke, A.D. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **1993**, *98*, 5648–5652. [CrossRef]
121. Lee, C.; Yang, W.; Parr, R.G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B* **1988**, *37*, 785–789. [CrossRef]
122. Miehlich, B.; Savin, A.; Stoll, H.; Preuss, H. Results obtained with the correlation energy density functionals of becke and Lee, Yang and Parr. *Chem. Phys. Lett.* **1989**, *157*, 200–206. [CrossRef]
123. Tomasi, J.; Mennucci, B.; Cammi, R. Quantum Mechanical Continuum Solvation Models. *Chem. Rev.* **2005**, *105*, 2999–3094. [CrossRef] [PubMed]
124. Brancato, G.; Rega, N.; Barone, V. A hybrid explicit/implicit solvation method for first-principle molecular dynamics simulations. *J. Chem. Phys.* **2008**, *128*, 144501. [CrossRef] [PubMed]
125. Cossi, M.; Barone, V.; Cammi, R.; Tomasi, J. Ab initio study of solvated molecules: A new implementation of the polarizable continuum model. *Chem. Phys. Lett.* **1996**, *255*, 327–335. [CrossRef]
126. Cossi, M.; Scalmani, G.; Rega, N.; Barone, V. New developments in the polarizable continuum model for quantum mechanical and classical calculations on molecules in solution. *J. Chem. Phys.* **2002**, *117*, 43–54. [CrossRef]
127. Mennucci, B. Polarizable continuum model. *WIREs Comput. Mol. Sci.* **2012**, *2*, 386–404. [CrossRef]
128. Cossi, M.; Barone, V. Solvent effect on vertical electronic transitions by the polarizable continuum model. *J. Chem. Phys.* **2000**, *112*, 2427–2435. [CrossRef]
129. Grimme, S.; Ehrlich, S.; Goerigk, L. Effect of the damping function in dispersion corrected density functional theory. *J. Comput. Chem.* **2011**, *32*, 1456–1465. [CrossRef]
130. Grimme, S. Density functional theory with London dispersion corrections. *WIREs Comput. Mol. Sci.* **2011**, *1*, 211–228. [CrossRef]
131. Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J. Chem. Phys.* **2010**, *132*, 154104. [CrossRef]
132. Ehrlich, S.; Moellmann, J.; Grimme, S. Dispersion-Corrected Density Functional Theory for Aromatic Interactions in Complex Systems. *Acc. Chem. Res.* **2013**, *46*, 916–926. [CrossRef] [PubMed]

133. Risthaus, T.; Grimme, S. Benchmarking of London Dispersion-Accounting Density Functional Theory Methods on Very Large Molecular Complexes. *J. Chem. Theory Comput.* **2013**, *9*, 1580–1591. [CrossRef] [PubMed]
134. Grimme, S. Do Special Noncovalent π–π Stacking Interactions Really Exist? *Angew. Chem.* **2008**, *47*, 3430–3434. [CrossRef] [PubMed]
135. Weigend, F.; Ahlrichs, R. Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297–3305. [CrossRef] [PubMed]
136. Andrae, D.; Haeussermann, U.; Dolg, M.; Stoll, H.; Preuss, H. Energy-adjusted ab initio pseudopotentials for the second and third row transition elements. *Theor. Chem. Acc.* **1990**, *77*, 123–141. [CrossRef]
137. Jorgensen, W.L.; Chandrasekhar, J.; Madura, J.D.; Impey, R.W.; Klein, M.L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926–935. [CrossRef]
138. Svensson, M.; Humbel, S.; Froese, R.D.J.; Matsubara, T.; Sieber, S.; Morokuma, K. ONIOM: A Multilayered Integrated MO + MM Method for Geometry Optimizations and Single Point Energy Predictions. A Test for Diels-Alder Reactions and $Pt(P(t\text{-}Bu)_3)_2$ + H_2 Oxidative Addition. *J. Phys. Chem.* **1996**, *100*, 19357–19363. [CrossRef]
139. Vreven, T.; Byun, K.S.; Komáromi, I.; Dapprich, S.; Montgomery, J.A.J.; Morokuma, K.; Frisch, M.J. Combining Quantum Mechanics Methods with Molecular Mechanics Methods in ONIOM. *J. Chem. Theory Comput.* **2006**, *2*, 815–826. [CrossRef]
140. Chung, L.W.; Sameera, W.M.C.; Ramozzi, R.; Page, A.J.; Hatanaka, M.; Petrova, G.P.; Harris, T.V.; Li, X.; Ke, Z.; Liu, F.; et al. The ONIOM Method and Its Applications. *Chem. Rev.* **2015**, *115*, 5678–5796. [CrossRef]
141. Wang, J.; Wolf, R.M.; Caldwell, J.W.; Kollman, P.A.; Case, D.A. Development and testing of a general amber force field. *J. Comput. Chem.* **2004**, *25*, 1157–1174. [CrossRef]
142. Brancato, G.; Rega, N.; Barone, V. Molecular dynamics simulations in a NpT ensemble using non-periodic boundary conditions. *Chem. Phys. Lett.* **2009**, *483*, 177–181. [CrossRef]
143. Rega, N.; Brancato, G.; Barone, V. Non-periodic boundary conditions for ab initio molecular dynamics in condensed phase using localized basis functions. *Chem. Phys. Lett.* **2006**, *422*, 367–371. [CrossRef]
144. Brancato, G.; Barone, V.; Rega, N. Theoretical modeling of spectroscopic properties of molecules in solution: Toward an effective dynamical discrete/continuum approach. *Theor. Chem. Acc.* **2007**, *117*, 1001–1015. [CrossRef]
145. Zabrodsky, H.; Peleg, S.; Avnir, D. Continuous symmetry measures. *J. Am. Chem. Soc.* **1992**, *114*, 7843–7851. [CrossRef]
146. Pinsky, M.; Casanova, D.; Alemany, P.; Alvarez, S.; Avnir, D.; Dryzun, C.; Kizner, Z.; Sterkin, A. Symmetry operation measures. *J. Comput. Chem.* **2008**, *29*, 190–197. [CrossRef]
147. Pinsky, M.; Dryzun, C.; Casanova, D.; Alemany, P.; Avnir, D. Analytical methods for calculating Continuous Symmetry Measures and the Chirality Measure. *J. Comput. Chem.* **2008**, *29*, 2712–2721. [CrossRef]
148. Mu, Y.; Nguyen, P.H.; Stock, G. Energy landscape of a small peptide revealed by dihedral angle principal component analysis. *Proteins* **2005**, *58*, 45–52. [CrossRef]
149. Géron, A. *Hands-On Machine Learning with Scikit-Learn, Keras and TensorFlow: Concepts, Tools and Techniques to Build Intelligent Systems*; O'Reilly Media: Sebastopol, CA, USA, 2019.
150. Lloyd, S. Least squares quantization in PCM. *IEEE Trans. Inf. Theory* **1982**, *28*, 129–137. [CrossRef]
151. Kaufman, L.; Rousseeuw, P.J. *Finding Groups in Data: An Introduction to Cluster Analysis*; John Wiley & Sons: Hoboken, NJ, USA, 2009.
152. Schubert, E.; Lenssen, L. Fast k-medoids Clustering in Rust and Python. *J. Open Source Softw.* **2022**, *7*, 4183. [CrossRef]
153. Schubert, E.; Rousseeuw, P.J. Fast and eager k-medoids clustering: O(k) runtime improvement of the PAM, CLARA and CLARANS algorithms. *Inform. Syst.* **2021**, *101*, 101804. [CrossRef]
154. Caliński, T.; Harabasz, J. A dendrite method for cluster analysis. *Commun. Stat.* **1974**, *3*, 1–27. [CrossRef]
155. Jolliffe, I.T.; Cadima, J. Principal component analysis: A review and recent developments. *Phil. Trans. R. Soc. A* **2016**, *374*, 20150202. [CrossRef] [PubMed]
156. Plasser, F. TheoDORE: A toolbox for a detailed and automated analysis of electronic excited state computations. *J. Chem. Phys.* **2020**, *152*, 084108. [CrossRef] [PubMed]
157. Plasser, F.; Lischka, H. Analysis of Excitonic and Charge Transfer Interactions from Quantum Chemical Calculations. *J. Chem. Theory Comput.* **2012**, *8*, 2777–2789. [CrossRef] [PubMed]
158. Frisch, M.J.; Trucks, G.W.; Schlegel, H.B.; Scuseria, G.E.; Robb, M.A.; Cheeseman, J.R.; Scalmani, G.; Barone, V.; Petersson, G.A.; Nakatsuji, H.; et al. *Gaussian 16 Revision C.01*; Gaussian Inc.: Wallingford, CT, USA, 2016.