The ezSpectra Suite: An easy-to-use Toolkit for Spectroscopy Modeling

Samer Gozem\textsuperscript{a} and Anna I. Krylov\textsuperscript{b}

\textsuperscript{a} Department of Chemistry, Georgia State University, Atlanta, GA 30303, USA
\textsuperscript{b} Department of Chemistry, University of Southern California, Los Angeles, CA 90089, USA

March 1, 2021

Abstract

A molecule’s spectrum encodes information about its structure and electronic properties. It is a unique fingerprint that can serve as a molecular ID. Quantum chemistry calculations provide key ingredients for interpreting spectra, but modeling the spectra rarely ends there; it requires additional steps that entail combined treatments of electronic and nuclear degrees of freedom and account for specifics of the experimental setup (light energy, polarization, averaging over molecular orientations, temperature, etc.). This Software Focus article describes the ezSpectra suite, which currently comprises two stand-alone open-source codes: ezFCF and ezDyson. ezFCF calculates Franck-Condon factors, which yield vibrational progressions for polyatomic molecules within the double-harmonic approximation. ezDyson calculates absolute cross-sections for photodetachment/photoionization processes and photoelectron angular distributions using Dyson orbitals computed by a quantum chemistry program.
1 Introduction

“The spectrum is a message from the molecule” – Unknown

Spectroscopy is the most powerful tool for probing matter. Despite the mind-boggling sophistication of modern light sources and spectroscopic techniques, conceptually it remains just as simple as in the time of Newton’s experiments: we shine light at a sample and back comes the spectrum with information about the internal structure of the sample.

Spectroscopy allows us to communicate with molecules. Each spectrum is a message from the molecule, reporting on its structure. However, the message is coded. To translate an intricate pattern of spectral lines and intensities into what the nuclei and electrons are doing, we need a cipher. Theory provides the key to decipher the message (see Fig. 1). By using quantum chemical calculations, we can obtain essential parameters of the molecular structure. Using these parameters as an input, we can then simulate the spectra. By comparing the computed and experimentally measured spectra, we can confirm molecular identity, assign spectral features, and relate the spectra to such concepts as equilibrium bond lengths, vibrational frequencies, molecular orbitals, and so on.

Figure 1: Spectroscopic measurements (red panel) yield encrypted messages, which can be decoded using electronic structure calculations (blue panels) to reveal detailed information about molecular structure (black panel).

Different types of spectroscopies probe different types of transitions and report on different molecular properties. To appreciate the scope and variety of modern spectroscopy, consider the vast range of energies and intensities of light used,\(^1\) in addition to nonlinear,\(^2\) angle-resolved,\(^3,4\)
and time-resolved techniques. Experimental innovations constantly drive electronic-structure developments, pushing quantum chemistry into new domains. But modeling spectra rarely ends at the electronic structure calculations and often requires additional steps that entail combined treatments of electronic and nuclear degrees of freedom and account for the specifics of the experimental setup (light polarization, averaging over molecular orientations, temperature, etc).

This Software Focus article describes the ezSpectra suite, which bridges quantum-chemistry calculations and experimental observables (spectra). Our focus is on spectroscopies that probe the transitions between different electronic states, such as traditional one-photon absorption (UV/Vis, XAS), photoionization/photodetachment (UPS, XPS, ARPES, etc), as well as non-linear spectroscopies (2PA, RIXS, etc). Despite the vast differences between these techniques, the common thread is the need to combine electronic factors (relevant energy differences and transition moments) with vibrational structure. In what follows, we describe the theory of linear electronic absorption spectroscopy and photoionization/photodetachment spectroscopies, followed by a brief outline of the extensions to other types of spectra.

By treating light as a weak time-dependent perturbation of molecular eigenstates, we encounter singularities in the first-order perturbed amplitudes at frequencies resonant with energy differences between the (unperturbed) molecular states. The singularities mean that the molecule can undergo transitions between different states when the frequency matches the energy gap. Within the dipole approximation (i.e., assuming that the molecular size is smaller than the wavelength of the light), the probability of transition between an initial state \( i \) and a final state \( f \) is proportional to the square of the transition dipole matrix element:

\[
P_{if} \propto \left( \int \Psi_i(\mathbf{r}, \mathbf{R}) \hat{\mu} \Psi_f(\mathbf{r}, \mathbf{R}) \, d\mathbf{r} \, d\mathbf{R} \right)^2, \tag{1}
\]

where \( \hat{\mu} \) is the electronic dipole moment operator, \( \mathbf{R} \) and \( \mathbf{r} \) are the collective nuclear and electronic coordinates, respectively, and \( \Psi_i(\mathbf{r}, \mathbf{R}) \) and \( \Psi_f(\mathbf{r}, \mathbf{R}) \) are the wave functions of the two states. By using the separable ansatz (either within the Born-Oppenheimer approximation or within the exact treatment), the full wave functions can be represented as products of the electronic \( (\psi(\mathbf{r}; \mathbf{R})) \) and nuclear \( (\chi(\mathbf{R})) \) wave functions. Within the Born-Oppenheimer approximation, the electronic energies and wave-functions are determined by the electronic
Hamiltonian (full molecular Hamiltonian minus the nuclear kinetic energy operator), such that they depend only parametrically on the nuclear coordinates (as signified by ';' instead of ','):

\[ P_{i'f''} \propto \left( \int \psi_i(r; R) \chi_{i'}(R) \hat{\mu} \psi_f(r; R) \chi_{f''}(R) \, dr \, dR \right)^2. \]  

(2)

Here indices \( i' \) and \( f'' \) denote the vibrational states of the two electronic states (which, within the Born-Oppenheimer approximation, are determined by the nuclear problem with the potential defined by the electronic Schrödinger equation). Integration over the electronic coordinates yields the electronic transition dipole moment for the \( i \rightarrow f \) transition

\[ \mu_{if}(R) = \int \psi_i(r; R) \hat{\mu} \psi_f(r; R) \, dr, \]  

(3)

so that the probability of transition can be written as

\[ P_{i'f''} \propto \left( \int \chi_{i'}(R) \mu_{if}(R) \chi_{f''}(R) \, dR \right)^2. \]  

(4)

Eq. (4) is the starting point for understanding the spectrum. It contains electronic factor \( \mu_{if}(R) \), which encodes the information about the electronic and vibrational wave-functions of the two states. Within the Condon approximation,\(^{11}\) it is assumed that \( \mu_{if}(R) \) depends weakly on the nuclear coordinates and, therefore, can be evaluated at a fixed nuclear geometry (e.g., at the equilibrium geometry \( R_e \) of the initial state). In this case, the probability of the transition simply becomes:

\[ P_{i'f''} \propto \mu_{if}(R_e)^2 \left( \int \chi_{i'}(R) \chi_{f''}(R) \, dR \right)^2. \]  

(5)

The overlap integrals between the nuclear wave-functions of the two states are called the Franck-Condon factors (FCFs)\(^{11-13}\)—their squares are the intensities due to vibrational progressions (sometimes the intensities are called FCFs). The stick spectrum can then be assembled from the individual FCFs as follows:

\[ I(\omega) = \sum_{i',f''} P_{i'f''} \cdot \delta(\omega - E_{i'f''}), \]  

(6)

where \( \delta \) is the Kroneker \( \delta \)-function and \( E_{i'f''} \) is the energy difference between the two vibronic
states. The macroscopic observables, such as cross sections, Einstein coefficients, etc, can be computed from $I(\omega)$ (see, for example, Ref. 14).

Figure 2: Factors determining the spectral line shape for electronic transitions. The ground state (red), the first excited state (blue), and the second excited state (green) are represented by one-dimensional harmonic potentials along a normal mode $Q$. The change in the excited-state equilibrium geometry relative to the ground state is denoted by $\Delta Q'$ and $\Delta Q''$. The main progression is determined by the overlap of the initial state $\nu = 0$ wave function (red translucent area curve) and each of the excited-state vibrational wave functions (blue translucent area curves). If $T \neq 0$, hot bands may appear, such as peaks resulting from the $\nu = 1$ transitions indicated by dotted lines on the right. The total spectrum is an overlay of the main progression and hot bands. Instrumental broadening and interactions with the environment (e.g., inhomogeneous broadening in condensed phase) results in a broadened spectrum like the one shown by black dashed lines on the right. The definition of electronic vertical ($E_v'$) and adiabatic ($E_{ee}$) excitation energies are shown for the first excited state.

Fig. 2 illustrates how Eqs. (5) and (6) give rise to the spectral profile due to transitions from the initial state (red) to the two final states (blue and green). The $\mu_{if}$ for each state determines the magnitude of the electronic contribution to the cross-section, and the FCFs determine the probability of transition to different vibrational levels within each electronic state. As per Eq. (5), FCFs are simply the overlap integrals between the initial-state vibrational wave function (red translucent curve in Fig. 2) and the final-state vibrational wave functions (blue translucent curves). A large overlap leads to a large FCF, giving rise to a high probability of transition to that vibrational level. For two states represented by two identical one-dimensional harmonic potentials, only the diagonal transitions ($\Delta \nu = 0$) are allowed by virtue of the orthogonality of the Hermite polynomials, leading to a spectrum with no vibrational progression (e.g., for
a vibrationally cold molecule, only 00 transitions would be seen). Transitions to multiple vibrational states become allowed when the two potentials are offset with respect to each other (i.e., when \( \Delta Q \neq 0 \)) or when their curvatures are different (i.e., when the frequency changes upon the transition). The modes that are affected by the changes in the geometry or frequencies upon excitation are called Franck-Condon active modes. At finite temperatures, when higher vibrational states are populated, “hot bands” appear in the spectra, such as transitions from \( \nu = 1 \) shown in Fig. 2, in addition to the main progression from the lowest \( \nu = 0 \) vibrational state.

Fig. 2 also shows two important (and often misused) quantities: the vertical (\( E^v \)) and adiabatic (\( E^a_{ee} \)) electronic excitation energies. The vertical excitation energy is simply the electronic energy difference between the two potential energy surfaces at the equilibrium geometry of the initial (e.g., ground) state. The adiabatic electronic excitation energy is the energy difference between the two potential energy surfaces at their respective equilibrium geometries (in spectroscopic notations, it is denoted by \( E_{ee} \)); obviously, \( E^a_{ee} \) is always smaller than \( E^v \). Although \( E^v \) and \( E^a_{ee} \) are easy to define theoretically and are straightforward to compute by quantum-chemistry programs, they are not directly related to experimental observables, because they do not correspond to energy differences between vibronic states.

By taking into account energies of the vibrational levels, one can define the energy difference between the lowest vibrational states—this is the adiabatic excitation energy \( E^a_{00} (T_{00} \text{ in spectroscopic notations}) \). The adiabatic excitation energy includes the electronic energy difference (\( E^a_{ee} \)) and the zero-point energy correction (i.e., the difference in the zero-point vibrational energies of the two states). For a vibrationally cold molecule, \( E^a_{00} \) corresponds to the onset of the spectral band, provided that the corresponding FCF \( (\langle \chi'_0 | \chi''_0 \rangle) \) is sufficiently large. If the geometry relaxation in the target state is too large, the FCF for the 00 transition can become too small to be observed.

The vertical excitation energy is not an experimental observable. Using a semiclassical treatment of a one-dimensional harmonic model, one can argue that the largest FCF should correspond to the vertical transition, as shown for the \( \nu = 0 \) to \( \nu' = 2 \) transition in Fig. 2. Consequently, it is tempting to interpret the band maximum (\( \lambda_{max} \)) as the vertical excitation energy. While this is a reasonable assumption for many diatomics, calculations for polyatomic
molecules show that the maximum intensity often corresponds to the 00 transition, such that $\lambda_{\text{max}}$ represents the adiabatic excitation energy (more on this below). Hence, the most reliable way to compare theory and experiment is to complement quantum-chemistry calculations of the electronic energy differences by calculation of the FCFs.

The above treatment describes both the transitions between the electronically bound states (i.e., the ground and excited states of a molecule) and the transitions to the continuum, when the energy of light ($E_{h\nu}$) is sufficient to eject an electron. In the latter case, the spectra are measured as the yield of photo-electrons at different kinetic energy ($E_k$), as per the following energy balance:

$$E_k = E_{h\nu} - \Delta E_{i'f'}.$$  

(7)

In the experimental community, it is common to distinguish between photoionization (electrons ejected from a neutral molecule) and photodetachment (electrons ejected from an anion), but, for the sake of brevity, we will use the term photoionization to refer to both.

For the transitions between bound electronic states, Eq. (5) yields the total absorption cross section, and $\mu_{i f}$ is an electronic transition dipole moment element. For the transitions into the continuum, the probability of ejecting an electron is also determined by Eq. (5), but $\mu_{i f}$ is replaced by the photoelectron dipole matrix element $\mu_{i\infty}$:

$$\mu_{i\infty}(R) = \int \psi_f^n(r_1, r_2, \ldots, r_n; R) \hat{A} \left[ \psi_f^{n-1}(r_1, r_2, \ldots, r_{n-1}; R) \psi_{\text{el}}^k(r_n) \right] \, dr_1 \, dr_2 \cdots dr_n, \quad (8)$$

where $n$ is the number of electrons in the initial state, $\hat{A}$ is an anti-symmetrizing operator, $\psi_f^{n-1}$ is the wave-function of the ionized (detached) target, and $\psi_{\text{el}}^k$ is the wave-function of the ejected electron with wave vector $k$.

The ezSpectra suite currently comprises two stand-alone open-source codes (distributed under the GPL license): ezFCF and ezDyson. The former computes FCFs (the overlap integrals in Eq. (5)), and the latter computes properties related to the photoelectron dipole matrix element in Eq. (8). Both are written in C++ and use a built-in parser to read xml-formatted input files. The source code, precompiled executables, manuals, and examples of input and output files are available for download at http://iopenshell.usc.edu/downloads/. We also provide Python scripts for automatic preparation of the xml input files from raw outputs of quantum-
chemistry programs. The scripts currently support Q-Chem, Cfour/ACESII, Molpro, ORCA, GAMESS, and Gaussian outputs.

2 ezFCF

2.1 Software capabilities and features

Figure 3: ezFCF input parameters, options, and output. The grey box indicates selected features implemented in the code. The output contains information listed in yellow box.

**Overview:** ezFCF, formerly known as ezSpectrum, computes vibrational progressions (stick spectra) for electronic excitation or photoelectron spectra for polyatomic molecules within the double-harmonic approximation.

**Required input:** ezFCF requires equilibrium geometries, harmonic frequencies, and normal-mode vectors for each electronic state. These quantities can be computed using an ab initio electronic structure package (Fig. 3). Additionally, the user should specify the temperature (which depends on experimental conditions) and the adiabatic excitation or ionization energy (which can be either computed or taken from the experiment).

**What the code computes:** ezFCF computes FCFs, the overlaps between the multi-dimensional vibrational wave functions of the initial (usually, ground) state and the final (excited or ionized) states, as per Eq. (5). In the current version, both the initial and the final states are described
by harmonic potentials (hence, double-harmonic approximation).

For a one-dimensional case (diatomics or a molecule with only one Franck-Condon active mode), the FCFs are simply the overlaps between the eigenstates of the two shifted harmonic oscillators.\textsuperscript{18} For a multi-dimensional case, the vibrational wave functions are products of the eigenstates of one-dimensional harmonic oscillators

\begin{equation}
\chi'_{\nu'_{1},\nu'_{2},\ldots,\nu'_{N}}(Q') = \prod_{a=1}^{N} \chi'_{a,\nu'_{a}}(Q'_{a})
\end{equation}

\begin{equation}
\chi''_{\nu''_{1},\nu''_{2},\ldots,\nu''_{N}}(Q'') = \prod_{b=1}^{N} \chi''_{b,\nu''_{b}}(Q''_{b})
\end{equation}

here \(\chi'\) and \(\chi''\) denote the wave-functions for the initial and final states, respectively, each expressed in its own normal coordinates \(Q'\) and \(Q''\). Indices \(a\) and \(b\) run over all normal modes and strings \(\nu_{1},\nu_{2},\ldots,\nu_{N}\) give the number of quanta in each mode.

How can we evaluate the multi-dimensional overlap integral between \(\chi'(Q')\) and \(\chi''(Q'')\), which are expressed in different coordinate systems? One possibility is to transform one of the states into the normal coordinates of the other state; this can be achieved by using Duschinsky rotations\textsuperscript{19} (named after Duschinsky who first described the effect of normal-mode rotation in polyatomic molecules\textsuperscript{20}). Alternatively, the task can be significantly simplified by assuming that the normal coordinates of the initial and final states are the same, such that along each normal coordinate one state is offset (by \(\Delta Q\)) with respect to the other and the curvature of the potentials is different. This gives rise to the parallel-mode approximation in which the multi-dimensional integral becomes a product of one-dimensional integrals for each normal mode:

\begin{equation}
\langle\chi'_{\nu'_{1},\nu'_{2},\ldots,\nu'_{N}}(Q) | \chi''_{\nu''_{1},\nu''_{2},\ldots,\nu''_{N}}(Q + \Delta Q) \rangle = \prod_{a=1}^{N} \langle \chi'_{a,\nu'_{a}}(Q_{a}) | \chi''_{a,\nu''_{a}}(Q_{a} + \Delta Q_{a}) \rangle.
\end{equation}

In this expression, normal modes of the initial state are used, so that the final state is offset with respect to the initial state, but one can also use normal modes of the target state and treat the initial state as being offset. The one-dimensional integrals are evaluated analytically using the expressions from Ref. 18. The effect of normal modes rotations and parallel-mode approximation are illustrated in Fig. 4, which also explains the impact of choosing normal modes of the initial versus the final state.
The validity of the parallel-mode approximation (and the alignment of the normal modes) can be assessed by checking the normal-modes overlap matrix, which should be close to diagonal. The order of the normal modes can be flipped, which is manifested by near-unity off-diagonal values in the overlap matrix; this does not mean that the parallel mode approximation is invalid, as the diagonal form of the overlap matrix can be easily restored by the manual reordering of the normal modes to match. If the normal modes are rotated (i.e., when the values of the diagonal elements are significantly less than one, even after the modes are properly ordered), it is recommended to use the normal modes of the target state instead of the normal modes of the initial state. As explained in Fig. 4, the initial vibrational state ($\nu_1 = \nu_2 = \ldots = \nu_N = 0$), which gives rise to the main vibrational progressions, has a fully symmetric wave-function. Consequently, projecting such a state into the rotated coordinate system (normal modes of the target state $Q''$) does not affect its nodal structure and, therefore, introduces relatively small errors in the overlaps. In contrast, projecting the vibrationally excited state into another coordinate system can significantly affect the overlaps because of the incorrect representation of the nodal structure of the target state.

To further reduce the cost of the calculations, the maximum number of vibrational excitations in the target state are truncated at a user-defined threshold (this is justified because the overlaps for highly excited states eventually become small) and by excluding some modes (such as those that have small $\Delta Q$ and small change in frequencies) from the calculation. The
number of quanta in the initial state depends on the temperature: for $T=0$ K, $\nu'_{\text{max}}=0$, but for finite $T$ more states need to be included to describe hot bands. The populations of higher vibrational states are given by the Boltzmann factors.

Fig. 2 illustrates what is computed in ezFCF. If there are multiple final states (as in the figure), ezFCF reads the input parameters for each state and computes the overall spectrum by adding up the FCFs for all states, properly shifted by the respective excitation or ionization energies. Using the equilibrium geometries, harmonic frequencies, and normal mode vectors of the initial and final state(s) provided by ab initio calculations, ezFCF first aligns the equilibrium geometry of the initial and final states to maximize the overlap of the normal modes. The displacements between the initial state and each final state ($\Delta Q_s$ in Fig. 2) along each normal mode are then determined, using either the normal modes of the initial or the final state, as specified by the user (see Fig. 4). If $T=0$ K, only excitations from the $\nu = 0$ vibrational level of the initial state are considered. At finite temperatures, the thermal (Boltzmann) distribution is used to populate higher vibrational levels of the initial state (up to $\nu_{\text{max}}$) and hot bands are computed as well. ezFCF prints the energies and the FCFs (and their squares) for each vibronic transition, so that the result can be plotted as a stick spectrum, as per Eq. (6). The overall spectrum (printed in the ezFCF output) contains the superposition of the transitions from $\nu = 0$ and all hot bands (scaled by Boltzmann factors), as shown in Fig. 2, and can be directly compared with the vibrational progressions in experimental spectra.

The details of the algorithm, which uses a recurrence relation to evaluate the overlap integrals accounting for Duschinsky rotations, is described in the ezFCF manual and in more details in Refs. 22 and 23. Such calculations are more expensive than parallel-mode approximation and scale unfavorably with the system size. The cost can be controlled by limiting the number of Franck-Condon active modes and vibrational excitation levels in each mode. We note that double-harmonic approximation fails in systems with large-amplitude soft modes, such as molecular clusters; in such cases more sophisticated treatments are needed.

2.2 Examples

To illustrate the theory, we consider several representative examples from the literature. Fig. 5 compares the experimental photoelectron spectrum of the phenolate anion with the stick
Figure 5: Photoelectron stick spectrum of the phenolate anion computed in double-harmonic parallel normal mode approximation at \( T = 300 \) K (red sticks). The experimental\(^{25}\) photoelectron spectrum is shown in black. The structure and displacement vectors indicate the primary Franck-Condon active mode responsible for the well-resolved vibrational peaks. Reproduced with permission from Ref. 26.

The computed spectrum reproduces the overall shape of the spectrum well, including the intensity pattern. At higher energies, the computed spacings between the peaks are slightly overestimated relative to the experimental ones, which is expected due to the harmonic approximation (the levels of the harmonic oscillator are equidistant whereas for the anharmonic Morse potential the separation between the levels decreases with energy). The most intense peak in the spectrum is the 0-0 transition (the smaller peak just under 2.2 eV is a hot band from the first vibrationally excited level of the main progression-forming normal mode). For this system, the difference between the vertical and adiabatic detachment energy is relatively small (0.05 eV), yet it is significant for quantitative comparisons of the theoretical results with the experimental values.

The same study\(^{26}\) also reported the photoelectron spectrum of the HBDI anion, a model system of the chromophore of the green fluorescent protein. For this larger molecule, the spectrum is broader and more congested due to multiple Franck-Condon active modes. The structural relaxation (the difference between \( E^v \) and \( E^a \)) is also larger in HBDI relative to phenolate—0.11 eV versus 0.05 eV; yet, the band maximum still corresponds to the 00 transition and not to the vertical detachment energy. A similar trend was observed\(^{27}\) for the excitation spectrum of the HBDI anion—here again the largest intensity corresponds to the 00 transition. This theoretical prediction was later confirmed by Weber and coworkers, who reported vibrationally resolved
action spectrum of the HBDI anion.\textsuperscript{28}

In the condensed phase, the spectra are broadened due to the interactions with the solvent (in addition to instrumental broadening); yet, they often retain discernible features of the underlying vibrational structure. An example, shown in Fig. 6, is flavin. The excitation and emission spectra of flavin were measured at high resolution by Dick and co-workers for lumiflavin in helium nanodroplets (Fig. 6A).\textsuperscript{29} The accompanying calculations reproduced reasonably well the Franck-Condon progression and ascribed the strongest progression-forming mode to an in-plane bending vibration.\textsuperscript{29} The UV/Vis spectrum in aqueous solution is broad and featureless, but in many proteins the spectra are better resolved and show three distinct bands in the $S_0 \rightarrow S_1$ transition (black spectrum in Fig. 6B). The computed spectrum (FCFs broadened by an empirically fit Gaussians, red spectra in Fig. 6B) reproduces these features and assigns them to the main progression-forming mode.\textsuperscript{30} The spectrum features two intense vibronic bands at ca. 2.6 eV and 2.8 eV, as well as a third, less-intense peak at ca. 2.95 eV. Neither of the two more intense bands is close to the vertical one; they are shifted by about 0.15 -0.40 eV relative to $E_v$.\textsuperscript{30} The magnitude of the zero-point energy correction, -0.065 eV, is also noticeable. The large difference between $E_v$ and a properly computed $\lambda_{max}$ has been pointed out in several studies on flavins using various levels of theory.\textsuperscript{30–33}

These examples illustrate that a common practice of interpreting the band maxima in photoelectron or UV/Vis absorption spectra as vertical ionization or excitation energies is not justified for polyatomic molecules, in which $\lambda_{max}$ often corresponds to the 00 transition, and is, therefore, much closer to the adiabatic ionization/excitation energies. In molecules discussed here, the adiabatic and vertical energy differ by as much as 0.1-0.4 eV, which is clearly important for quantitative comparisons between theory and experiment. The reason that the conclusions based on the one-dimensional model (often referred to as the Franck-Condon principle) do not hold up in polyatomic molecules is that small structural relaxation along many degrees of freedom adds up to a large energy value.\textsuperscript{15,26} Hence, despite a large energetic difference between $E_v$ and $E_a$, the actual displacements along Franck-Condon active modes are sufficiently small, so that the 00 transition remains the most intense one. We note that the smallness of structural changes can be attributed to the delocalized nature of molecular orbitals in polyatomic molecules.\textsuperscript{15}
Figure 6: A. Experimental and computed (stick spectra) Franck-Condon pattern of the excitation and emission spectra of flavin. B. Experimental (black, for flavin mononucleotide in iLOV) and computed (red) spectra for flavin in condensed phase. The FCFs were computed for a gas-phase lumiflavin model using ezFCF with B3LYP/cc-pVTZ (red stick spectra and solid red line broadened spectrum) and by Davari et al using a slightly different approach (dashed red line). The adiabatic and vertical excitation energies are indicated with vertical dashed lines. Reproduced with permission from Refs. 29 and 33.
Obviously, the best strategy for benchmark studies is to complement electronic structure calculations with calculations of FCFs, to unambiguously determine the position of the highest intensity relative to $E^v$ (or $E_{ee}^a$). If this is not possible, the next best approach is to compare experimental band maxima with the computed $E_{00}^a$ (or $E_{ee}^a$). These observations further highlight the importance of assessing the performance of electronic structure methods to describe potential energy surfaces. Even when comparing theory with theory, it is not sufficient to rely on vertical excitation energies alone. As was recently illustrated by Szalay and co-workers\textsuperscript{34} for the coupled-cluster family of methods (ranging from CC2 to CCSDT), already at the Franck-Condon geometries, the nuclear gradients differ considerably among methods, indicating that the shapes of the resulting potential energy surfaces and, consequently, FCFs would be very different. One consequence is that the ability of low-level methods (such as CC2 or ADC(2)) to produce accurate vertical excitation energies (as compared with the high-level theoretical reference) is an artifact due to a fortuitous cancellation of errors and that the errors for the adiabatic excitation energies are much larger.\textsuperscript{34} Fig. 7 further illustrates this point by comparing three excited-state potential energy surfaces that have the same $E^v$ but different shapes, giving rise to vastly different FCFs. Thus, calculation of FCFs should be an integral part of benchmarking excited-state methods.\textsuperscript{35,36}

$ezFCF$ can be used to generate requisite data to carry out more sophisticated calculations, such as inclusion of non-Condon effects\textsuperscript{27} and modeling spectral line-shapes in non-linear spectroscopies, such as two-photon absorption\textsuperscript{27} and resonant inelastic x-ray scattering.\textsuperscript{37}

3 ezDyson

3.1 Software capabilities and features

Overview: \textit{ezDyson} computes absolute photoionization cross sections, photoelectron angular distributions (PADs), and anisotropy parameters ($\beta$) using Dyson orbitals computed by a quantum-chemistry program.

Required input: \textit{ezDyson} requires molecular geometry and orbital data from a quantum-chemistry package, as well as several experimental details (Fig. 8). The user should specify the ionization energy and the range of electron kinetic energies for which the cross sections will
Figure 7: A schematic illustration of FCFs revealing the differences between excited-state methods that are not captured by vertical energy calculations. Different excited-state geometries (compare panels A and B) give rise to different FCF intensity pattern and band position. Different excited-state curvatures (compare panels A and C) give rise to different spacings between the vibrational peaks.

Figure 8: ezDyson input parameters, options, and output. The grey box indicates options and features implemented in the code, which the user controls. The main information printed in the output are in yellow.
be computed. The user also should specify whether the calculation is for randomly oriented molecules or for a molecule aligned and fixed in space. In the latter case, the user should indicate the direction of the laser polarization in the molecular frame.

The quality of the calculations is sensitive to the treatment of the photoelectron wave function. For absolute cross sections, correct spin and orbital degeneracy factors also need to be provided by the user.

**What the code computes:** Absolute photoionization cross sections, PADs, and the $\beta$ anisotropy parameter, which are all functions of the photoelectron dipole matrix element from Eq. (8). In what follows we briefly introduce these quantities and explain how they are computed in *ezDyson*.

Eq. (8) defines the photoelectron dipole matrix element $\mu_{i\infty}$ as an integral over coordinates of all $n$ electrons (represented by $r$). Because electrons are indistinguishable and assuming strong orthogonality (i.e., that the photoelectron wave function is orthogonal to all orbitals of the molecule), the integration can be done in two steps, giving rise to

$$\mu_{i\infty} = \int \phi^d(r_1) \mu \psi^e_k(r_1) dr_1 \equiv \langle \phi^d | \mu | \psi^e_k \rangle,$$

(12)

where $\phi^d$ is obtained by integrating the initial $n$-electron and final $n-1$-electron wave functions:

$$\phi^d(r_1) = \sqrt{N} \int \psi^n_i(r_1, \ldots, r_n) \psi^{n-1}_f(r_2, \ldots, r_n) dr_2 \ldots dr_n.$$

(13)

For the sake of brevity, here we have dropped the implicit dependence on nuclear coordinates $R$.

$\phi^d(r)$, which is a function of only one electron, contains all the information about the two many-body wave functions $\psi^n_i$ and $\psi^{n-1}_f$ needed to evaluate $\mu_{i\infty}$; it is called a Dyson orbital or generalized overlap.\textsuperscript{38,39} Within the Koopmans’ framework, $\phi^d$ is simply the Hartree-Fock canonical orbital vacated by the electron upon ionization. When computed using correlated wave functions (such as EOM-CC,\textsuperscript{40–43} configuration interaction,\textsuperscript{44} Green’s functions and propagator methods,\textsuperscript{45,46} CASSCF,\textsuperscript{47} and CASPT2\textsuperscript{48}), Dyson orbitals account for electron correlation and orbital relaxation effects that are missing in Koopmans’ picture of ionization while retaining the conceptual simplicity of a one-electron picture. Among numerous correlated approaches,
EOM-CC is uniquely suitable for computing Dyson orbitals (and, consequently, photoionization and photodetachment cross sections) in a broad range of systems. The Fock-space formulation enables access to different electronic structure patterns including electronically excited and open-shell species as well as states with multi-configurational character whereas the single-reference formulation makes the derivations of working equations rather straightforward. Among recent extensions of EOM-CC, we note the implementation of Dyson orbitals for electronically metastable states (resonances) and core-level states. The versatility of EOM-CC has recently been exploited in extending the concept of Dyson orbitals to two-body Dyson orbitals suitable to describe Auger spectra in the framework of EOM-CC theory for core-level states.

Eq. (12) shows that the shape of the Dyson orbital determines the state of the ejected electron, by virtue of the dipole selection rules. For photoionization from an s orbital, for instance, the integral in Eq. (12) is only non-zero when the photoelectron wave function $\psi_{ek}$ has a p-like shape aligned with the direction of the light polarization vector. The photoelectron is said to be a p-wave. Likewise, when photoionization occurs from a p (or p-like) orbital, the photoelectron could be ejected as an s-wave or a d-wave. Because the direction in which the electron is ejected is related to the shape of the orbital from which it was ionized, it is informative to measure the PAD, which describes the angular distribution (or the anisotropy) of the ejected photoelectron with respect to the polarization of light. PAD is related to the differential cross section $(d\sigma/d\Omega_k)$ that describes the dependence of the cross section $(\sigma)$ on the solid angle of photoelectron emission $(\Omega_k)$.

Cooper and Zare showed that for an ensemble of randomly oriented molecules, this anisotropy persists and has the following simple form:

$$\frac{d\sigma}{d\Omega_k} = \frac{\sigma}{4\pi} \left[ 1 + \beta P_2 (\cos(\theta_k)) \right],$$

(14)

where $P_2$ is a second order Legendre polynomial and $\theta_k$ is the direction of the ejected photoelectron with respect to the polarization of the incident light. $\beta$ is the aforementioned anisotropy parameter that characterizes the angular distribution of the photoelectrons: it can vary between -1 and 2 (-1 is perfectly perpendicular, 0 is perfectly isotropic, +2 is perfectly parallel).
The plot of $\beta$ as a function of the energy of ionizing radiation (or photoelectron kinetic energy) encodes the information about the shape (i.e., angular momentum and the diffuseness) of the orbital from which the electron is ejected. Fig. 9 shows the $\beta$ values for H$^-$ and F$^-$ as a function of the photoelectron kinetic energy computed with $\text{ezDyson}$. While $\beta$ is constant for H$^-$ at all energies (consistent with a $p$-wave that is parallel to the polarization of light), F$^-$ has a more complex energy-dependent $\beta$ profile. At low energies, ionization of fluoride’s $p$-orbital results in a photoelectron wave function dominated by an $s$-wave (and, therefore, isotropic). In contrast, at high energies, the photoelectron wave function is dominated by a $d$-wave. At intermediate energies, cross-terms arising from the combination of $s$-waves and $d$-wave result in a perpendicular distribution of electrons, with $\beta = -1$ at some point.$^{63}$

![Figure 9: $\beta$ values for H$^-$ and F$^-$ as a function of energy of the photoelectron kinetic energy.](image)

The absolute photoionization cross section is a measure of the total probability of an electron being ejected in any direction. It is obtained by integrating the differential cross section over all solid angles ($\Omega_k$). The calculation of energy-dependent photoionization cross sections is useful for assigning the electronic origins of photoelectron spectra (e.g., to identify contributions from direct ionization and auto-ionizing metastable states). The ability to accurately compute absolute cross sections is also important in applications of photoelectron spectroscopy as an analytic tool for quantitative chemical analysis.$^{64}$

The differential cross section is computed from $\mu_{i\infty}$ using$^{16,65}$

$$\frac{d\sigma}{d\Omega_k} = \frac{4\pi^2 k E}{c} \langle \phi^d | \mu | \psi_{el}^k \rangle^2 = \frac{4\pi^2 k E}{c} |\mu_{i\infty}(\theta, \phi)|^2,$$

(15)
where \( k \) is the magnitude of the photoelectron wave vector \( \mathbf{k} \), \( E \) is the energy of the ionizing radiation, and \( c \) is the speed of light. \( \theta \) and \( \phi \) are the polar and azimuthal angles of the photoelectron relative to the polarization of light.

Eq. (15) is the starting point for computing total cross sections, PADs, and \( \beta \) anisotropy. \textit{ezDyson} evaluates the integral in Eq. (15) by numerical integration on a grid using Dyson orbitals computed by a quantum chemistry program (Dyson orbitals are specified as an expansion over the atom-centered Gaussian basis functions).

The treatment of the photoelectron wave function \( \psi_{\mathbf{k}}(\mathbf{r}) \) is one of the key challenges for quantitative modeling of photoionization and photodetachment. Currently, \textit{ezDyson} uses two simple models—plane wave and Coulomb wave. The limitations of these models are well documented and can be addressed by more sophisticated approaches.\(^{66}\)

In the plane-wave representation, the photoelectron wave-function is

\[
\psi_{\mathbf{k}}(\mathbf{r}) = \frac{1}{(2\pi)^{3/2}} e^{i \mathbf{k} \cdot \mathbf{r}},
\]

where \((2\pi)^{-3/2}\) is the continuum normalization for plane waves.\(^{67}\) \textit{ezDyson} provides two implementations for computing the respective matrix elements: one uses Eq. (16) as-is and another uses a partial-wave expansion in which the plane wave is expressed as a sum of spherical waves:\(^{68}\)

\[
e^{i \mathbf{k} \cdot \mathbf{r}} = 4\pi \sum_{l=0}^{\infty} \sum_{m=-l}^{l} i^l R_l(kr) Y_{lm}(\mathbf{r}) Y_{lm}(\mathbf{k}).
\]

By using the partial-wave expansion, one can analyze the wave function of the ejected photoelectron in terms of angular momentum quantum numbers. Each spherical wave is characterized by its energy \((E = \frac{k^2}{2m})\) and angular momentum \((l, m)\), and is a product of a radial function \(R_l(kr)\) and spherical harmonic functions \(Y_{lm}(\mathbf{r})\). \(Y_{lm}(\mathbf{r})\) depends on the position vector \(\mathbf{r}\) and, therefore, determines the shape of the plane wave in position space, while \(Y_{lm}(\mathbf{k})\) determines the direction in which the plane wave travels.\(^{68}\) By providing the coefficients of each partial wave, \textit{ezDyson} enables quantitative application of the mixed \(s&c\) model of Sanov to polyatomic molecules.\(^{63,69}\)

In practice, the expansion in Eq. (17) is truncated at some \( l_{\text{max}} \) (specified by the user). In
atoms, the truncation is justified because strict selection rules preclude contributions from high angular momentum waves to the photoelectron wave function. In molecules the truncation may be justified when Dyson orbitals resemble atomic orbitals, which is a main rationale behind the mixed $s&p$ model.\textsuperscript{63,69} Moreover, the contributions from large angular momentum are small at low photoelectron kinetic energies.

Using a plane wave is most appropriate for photodetachment, where there is no electrostatic interaction between the ejected photoelectron and the remaining neutral core. In contrast, in photoionization the core is charged and the electrostatic interaction between the photoelectron and the core cannot be easily ignored. In this case, the Coulomb wave may be used instead of a plane wave. The Coulomb wave can also be expanded into a sum of Coulomb partial waves using the same expression as Eq. (17).\textsuperscript{6} The details of the implementation and the analytic expressions for computing the total cross section, PADs, and $\beta$ using plane or Coulomb waves can be found in the ezDyson manual.\textsuperscript{70}

When using sufficiently high energy, multiple ionized target states $f$ become accessible and one needs to account for their contributions to the total cross section. In addition, the calculation should account for electronic degeneracies. The contributions can be added up as follows:

$$\frac{d\sigma}{d\Omega_k} = \frac{4\pi^2E}{c} \cdot \sum_f k_f \left| \mu_{i\infty}^f(\theta, \phi) \right|^2.$$  \hspace{1cm} (18)

Because different target states may give rise to different photoelectron kinetic energies, $k_f$ is a function of the state $f$. Different states also give rise to different Dyson orbitals, according to Eq. (13). For non-degenerate target states, $k_f$ and $\mu_{i\infty}^f$ differ, and one can compute cross sections for the individual states separately and add them up. For systems with symmetry-imposed degeneracies, $k_f$ and $\mu_{i\infty}^f$ are identical for the degenerate states and, therefore, the cross sections may be computed for only one of the states and then multiplied by the number of the states in the manifold. Users also need to account for spin-degeneracy of the target states. For ionization of closed-shell singlet states, the cross section should be multiplied by 2.\textsuperscript{70} For example, the cross section for photodetachment from $F^-$ to the $F(^2P)$ channel is

$$\sigma(^2P) = 3 \times 2 \times \sigma(\phi^d \equiv p_z),$$  \hspace{1cm} (19)
where the factor of 3 is for orbital degeneracy and the factor of 2 is for spin degeneracy. Likewise, for ionization from \( \pi \) orbitals in CO or acetylene, the cross sections should be multiplied by 4 (2 for orbital degeneracy and 2 for spin degeneracy). These degeneracy factors need to be specified by the user in the \textit{ezDyson} input.

**Averaging over molecular orientations.** Experiments are most often performed on isotropic samples comprising randomly oriented molecules. \textit{ezDyson} code provides two averaging procedures: analytic and numeric. The analytic averaging algorithm largely follows the approach described in Ref. 71 and is documented in the \textit{ezDyson} manual. The analytic averaging is fast and employs the partial-wave expansion from Eq. (17). The numeric averaging approach does not rely on the partial-wave expansion and, therefore, does not entail the truncation at some \( l_{\text{max}} \); this is useful at high energies where the truncation of the partial-wave expansion becomes problematic because the contributions from high angular momentum are no longer negligible.

**Accounting for FCFs.** The shapes of photoionization spectra are also affected by FCFs, as per Eq. (5). The electronic \( \mu_\infty \) term is computed in \textit{ezDyson} with the appropriate averaging approach, while the second term is the FCF for a given vibronic transition computed by \textit{ezFCF}. The absolute total cross section can be computed for any vibronic transition using the following equation:

\[
\sigma_{\nu\nu'} = P_\nu \sigma_k \langle \chi_{i,\nu} | \chi_{f,\nu'} \rangle^2, \tag{20}
\]

where \( P_\nu \), the population of the initial state vibrational level \( \nu \), accounts for the thermal population of vibrational states at non-zero temperatures. The sum of all \( P_\nu \) must be normalized to 1 to preserve the meaning of the absolute cross sections. Note that \( \sigma_k \) depends on the energy of the ionizing radiation \( E \) (see Eq. (12)), which is equal to the energy of the vibronic transition plus the kinetic energy of the photoelectron \( (E_k) \). The latter can be computed from the electron 0-0 transition and the vibrational energy levels:

\[
E_k = \frac{h^2 \nu_{\nu'}}{2} = E - IE_{00}^{IF} - (E_{\nu}^{F}(\nu) - E_{\nu'}^{F}(\nu')) = E - IE_{00}^{IF} - \Delta E_{\nu}(\nu, \nu'), \tag{21}
\]

where \( IE_{00}^{IF} \) is the adiabatic ionization/detachment energy (including the zero-point energy correction) of the initial state, and \( E_{\nu}^{F}(\nu) \) and \( E_{\nu'}^{F}(\nu') \) are the excess vibrational energies of
the initial and final states above $\nu = 0$ and $\nu' = 0$, respectively. To compute the total cross section at a given energy, one needs to sum over all vibronic transitions accessible at that energy, in a fashion similar to Eq. (18). Because $\sigma_k$ is only computed for discrete values of $E_k$, ezDyson distribution includes a script ($\text{xsecFCF}$) to compute the result of Eq. (20) for discrete energies using the output of ezFCF (see ezDyson manual\textsuperscript{70} for details).

**Single-center versus multi-center treatment of the photoelectron wave function.**

Because of the dipole moment operator in Eq. (8), the origin of the free-electron state $\psi_{el}^k$ matters. Consider, for example, an $s$-type Dyson orbital: if the partial-wave expansion has the same origin as the orbital, then only $p$-waves give rise to a non-zero matrix element, however, if the expansion center is offset, then many partial waves survive. Obviously, in atoms the expansion should be atom-centered, however, in polyatomic systems (and especially in clusters) the choice is no longer trivial.

One possibility is to use the centroid of the Dyson orbital as the origin:\textsuperscript{72} for atoms, this results in atom-centered expansions, whereas for relatively localized molecular orbitals this entails placing the origin where the density of the Dyson orbital is the largest. This choice also implicitly ensures the orthogonality of $\psi_{el}^k$ to the Dyson orbital.\textsuperscript{65,72} While working reasonably well for molecules, this choice becomes problematic in certain situations due to the limitations of our theoretical model.

To understand the issue, imagine two atoms at infinite separation; the total photoionization cross sections should simply be the sum of the atomic values computed using the atom-centered expansions. However, the Dyson orbital is delocalized by symmetry, such that its centroid falls in between the two atoms. As illustrated by numeric examples in Ref. 73, for two infinitely separated fragments the results of the two calculations (one using atom-centered expansion and one using the expansion at the midpoint) are vastly different, with atom-centered expansion yielding the correct description. However, at short inter-atomic distances, the correct result is obtained when the expansion is placed in between the atoms. This can be understood in the context of Sanov’s molecular interferometer,\textsuperscript{74} where the short-time evolution of time-resolved photoelectron images from dissociating I\textsuperscript{−}$_2$ was discussed in terms of the interference between the two waves coming from the individual atoms. This interference can be described as ionization of a single entangled state spanning both iodine atoms at short I\textsuperscript{−}$_2$ bond lengths. However,
at large inter-atomic separations the variation in $\beta$ ceases and it approaches the asymptotic value corresponding to the PAD of $I^-$. The distance at which the atoms become independent and the coherences die off is related to the energy of the photoelectrons. In principle, the exact treatment should correctly describe both regimes, but the single-centered model currently used in ezDyson would fail at large separations because it does not account for the interfering contributions of photoelectrons coming off the two degenerate orbitals, such as in-phase and out-of-phase combinations of $p$ orbitals in Sanov’s experiment.\textsuperscript{74}

This limitation provides a motivation for developing a multi-center model for photoionization. A simple variant of this model was used in Ref. 73 to describe photoionization of bulk water. In a multi-center model, the Dyson orbital is split into fragments (molecular or atomic) and the cross-section is computed as the sum of the fragments’ contributions. If the coherences between the centers are accounted for (as done, for example, in Refs. 75 and 76), the multi-center treatment should be able to smoothly cover both short-distance and long-distance cases (or, equivalently, low-energy and high energy regimes). In the current implementation in ezDyson, the coherences are neglected, so that the multi-center model is only appropriate for high-energy regime, when the de Broglie wave-length of photoelectrons is shorter than the distance between the centers. The inclusion of coherences is under development.

### 3.2 Examples

**Absolute Photoionization Cross Sections.**

Fig. 10 shows computed and experimental absolute cross sections for H$^-$ (photodetachment) and He (photoionization). For hydride, photodetachment results in a neutral atom (hydrogen) that does not exert electrostatic potential on the ejected electron. Therefore, the plane-wave treatment of the photoelectron is a reasonable approximation, as confirmed by the computed photoionization cross sections: the black curve in Fig. 10A falls within the experimental error (red dots and associated error bars). Fig. 10A also illustrates the impact of the quality of the Dyson orbital on the cross-section. The dashed line shows the cross section computed using the canonical Hartree-Fock orbital (Koopmans’ theorem)—it is rather different from the one computed using the correlated Dyson orbital and falls outside the experimental spectrum.

Helium becomes positively charged upon photoionization. Hence, ignoring the interaction
between the ejected electron and the positive He$^+$ core yields incorrect cross sections—compare the black curve with the red dots in Fig. 10B. Using a Coulomb wave with +1 charge yields much better agreement with the experimental spectrum (blue curve in Fig. 10B). Because ezDyson simulates direct photoionization, the computed cross section does not capture resonances, such as those observed in He around 60 eV. The contributions from resonances can be estimated following the procedure described in Ref. 77.

Figure 10: A. Absolute photodetachment cross section from H$^-$ using a plane wave description of the photoelectron and the Dyson orbital computed with EOM-IP-CCSD/aug-cc-pVTZ. The cross section computed using the Hartree-Fock canonical molecular orbital is shown with a dashed line. B. Absolute photoionization cross section from He using a plane wave (black line) and a Coulomb wave (blue line) descriptions of the photoelectron and the Dyson orbital computed with EOM-IP-CCSD/aug-cc-pVTZ. The plane-wave treatment neglects interactions between the photoelectron and ionized core, which is a reasonable approximation for photodetachment. The Coulomb wave accounts for an interaction with a +1 charge placed at the orbital centroid, which is a reasonable description for atomic photoionization. Reproduced with permission from Ref. 78.

Fig. 11 shows two examples of absolute photoionization cross section calculations in molecules. The simplest molecules, such as dihydrogen or small molecules with only one heavy atom,$^{78}$ show the same trend as helium: the plane-wave calculations are inaccurate, whereas the Coulomb-wave calculations agree quantitatively with the experimental photoionization spectra.

In larger molecules (e.g., formaldehyde in Fig. 11B), we often find that neither the plane wave nor the Coulomb wave with +1 charge yield quantitatively accurate photoionization cross sections. However, using a Coulomb wave with some partial effective charge ($Z_{eff}$) leads to a photoionization spectrum that matches both the shape and the absolute magnitude of the
experimental near-threshold photoionization efficiency spectrum. How to determine $Z_{\text{eff}}$ a priori is not clear. We have shown that the effective charge effectively tunes the shape of the photoelectron wave function to provide the correct overlap with the Dyson orbital in the interaction region and proposed a variational approach for determining optimal charges. Effective charges can be estimated using simple atomic models (e.g., Rydberg formula), giving rise to Belkić’s and Slater’s rules. The utility of these effective charges within the Coulomb-wave approximation have been recently illustrated in the context of modeling Auger spectra.

![Image](image_url)

**Figure 11:** Absolute photoionization cross section from molecular hydrogen (A) and formaldehyde (B) using a plane-wave (black line) and a Coulomb-wave (blue line) descriptions of the photoelectron and Dyson orbitals computed with EOM-IP-CCSD/aug-cc-pVTZ. In formaldehyde, neither the plane wave nor Coulomb wave reproduce the experimental total cross section, but a Coulomb wave with a partial effective charge of $Z_{\text{eff}} = 0.25$ (orange) does. The solid and dashed orange lines in panel B show the calculations with and without FCFs, respectively. By including the FCFs, the calculation reproduces the small rise in cross section around 11 eV. Reproduced with permission from Ref. 78.

**Photoelectron Angular Distributions.** PADs and the associated $\beta$s are more sensitive to the treatment of the photoelectron wave function than absolute cross sections. Therefore, quantitative calculations of PADs are more challenging, especially in molecules. In the condensed phase, additional complications arise due to elastic scattering of photoelectrons. These challenges have been discussed in a recent study on photoionization of bulk water.

Since photoelectron anisotropies depend on orbital shapes, PADs report on electronic structure changes associated with chemical (reaction dynamics) or physical processes, such as solvation. Time-resolved photoelectron spectroscopy and imaging have been extensively used to monitor dynamics. Angular-resolved imaging can be used to understand intermolecular
interactions. For example, by comparing PADs in the gas and liquid phases, it is possible to ask interesting questions, for example, how inter-molecular interactions affect the shape of the molecular orbitals of solute. This was the focus of a recent study,\textsuperscript{73} which attempted to quantify how water’s $p$-like orbitals are altered by moving from the gas to the liquid phase by re-hybridization, loss of symmetry, scattering, and delocalization.

The PADs of water and the associated $\beta$ have been experimentally measured at different energies of the ionizing radiation both in the gas\textsuperscript{73,86,87} and liquid\textsuperscript{73,83,88} phases using the liquid microjet technique.\textsuperscript{89,90} Fig. 12 shows the experimental and computed gas-\textsuperscript{73,86,87} and liquid-phase\textsuperscript{73} $\beta$ values for the $3a_1$ orbital of water. Gas-phase calculations were performed using a Coulomb wave with $Z_{\text{eff}}=2.06$ (charge computed by Belkić’s rule\textsuperscript{79}). Whereas in the low-energy regime, where slow outgoing electrons interact with the perturbing molecular potential for a longer time, the PADs are very sensitive to the magnitude of the effective charge at large energies the differences between plane- and Coulomb-wave calculations with different charges are rather small, because fast photoelectrons are less sensitive to the shape of the perturbing cationic potential.

The computed energy-dependent $\beta$ values agree only qualitatively with the experiment, showing discrepancies in the low energy regime. Although the agreement for other valence orbitals of water ($1b_1$ and $1b_2$) is slightly worse than for $3a_1$, the gas-phase model is able to capture the variations of the angular distribution from different ionized states, meaning that it can be used to model changes in anisotropy as a result of perturbation due to solvent.

To quantify the effect of solvation of individual water molecules, $\beta$ values were computed using pentamer structures extracted from a molecular dynamics simulation of neat water. The corresponding Dyson orbitals of these snapshots (Fig. 12, bottom right) are typically delocalized. $\beta$ values computed using these delocalized orbitals and a single-center Coulomb wave placed at the centroid of the pentamer cluster orbital were almost isotropic ($\beta \approx 0$), suggesting a substantial loss of anisotropy (dotted line in Fig. 12). However, the multi-center approach yielded $\beta$ values almost identical to the ones computed in the gas phase. As explained above, in this regime the multi-center treatment is more appropriate, because X-ray ionization produces high-energy photoelectrons with de Broglie wavelengths smaller than the intermolecular distance between the water molecules. The experiment showed some loss of anisotropy of water
in the liquid phase relative to the gas phase (blue versus red dots in Fig. 12), which, through comparison with water $1s_O$ measurements,\textsuperscript{91} was attributed almost exclusively to electron scattering, rather than an intrinsic change in water’s orbital shape. Empirically correcting the computed localized liquid $\beta$ values for scattering using a scaling factor derived from the $1s_O$ experiment\textsuperscript{91} yields the dashed line in Fig. 12, which agrees well with the experiment. Surprisingly, this finding means that water’s molecular gas-phase orbitals are not sufficiently perturbed in the liquid to cause noticeable change in the $\beta$ anisotropy.

Figure 12: Experimental (symbols) and computed (solid lines) $\beta$ values for ionization $3a_1$ orbitals of gas-phase (red) and liquid-phase (blue) water. Gas-phase data are from Truesdale et al.\textsuperscript{86} (red triangles), Banna et al.\textsuperscript{87} (red rhombuses), and from Gozem, Seidel et al.\textsuperscript{73} (red circles), liquid-phase experimental data are from Gozem, Seidel et al.\textsuperscript{73} (blue circles). Computed anisotropies are shown as lines using a Coulomb wave with Belkić’s charges for the description of the photoelectron. Liquid $\beta$ calculations were performed on pentamer structures extracted from the molecular dynamics simulation of water using: (1) a multi-center model with expansion centers on individual water molecules (blue solid line); (2) the multi-center model including empirical scattering corrections (blue dashed line); and (3) a single-center model with the expansion center placed at the centroid of the pentamer cluster Dyson orbital (blue dotted line).

In sum, although quantitative computations of PADs require a careful treatment of the photoelectron wave function, phase shifts,\textsuperscript{72} and elastic scattering, and therefore remain elusive, simple models, such as those implemented in \textit{ezDyson}, deliver valuable qualitative (and often semi-quantitative) results. Such calculations can also provide insight into the evolution of the electronic structure in the course of dynamics. \textit{ezDyson} has been incorporated in the excited-state molecular dynamics program system \textit{Newton-X} to enable simulation of time-resolved photoelectron spectroscopies.\textsuperscript{92,93}
4 Conclusion and outlook

Absorption and photoionization spectra provide valuable information about the electronic and vibrational energy levels of the system. In addition, the intensity pattern, absolute values of cross sections, and anisotropies encode information about the electronic wave functions. The information contained in the spectra can be deciphered by using quantum-chemistry simulations. The ezSpectra suite provides a set of tools for computing spectra from electronic-structure calculations. Future developments include improving the theoretical models (e.g., multi-center treatments of photoionization), time-dependent modeling of FCFs for extensions to larger systems, inclusion of vibronic effects, and extensions to other spectroscopies and magnetic systems. We also plan to integrate these tools into Python workflows to create a convenient and flexible platform for spectroscopy modeling. We envision that ezSpectra tools can be integrated into other modeling softwares.

Acknowledgments

The ezSpectrum code was originally developed by Dr. Vadim Mozhayskiy and AIK. The original version of ezDyson was developed by Dr. Melania Oana and AIK. Dr. Evgeny Epifanovsky and Dr. Ilya Kaliman have helped to parallelize ezDyson. Dr. Kadir Diri, Dr. Anastasia Gunina, and Sahil Gulania contributed to the ezDyson manual. Dr. Takatoshi Ichino, Prof. John Stanton, Prof. Steven Bradforth, and Dr. David Osborn helped to validate the implementation for the calculation of photoionization total cross sections and angular distributions in ezDyson. Pawel Wojcik contributed to ezFCF. We thank all users who provided feedback and additional conversion scripts.

Funding Information

Development and support of ezFCF and ezDyson has been funded by the National Science Foundation and the Department of Energy (AIK). Some of the applications described were made possible thanks to funding from the American Chemical Society Petroleum Research Fund and the National Science Foundation (SG).
Conflicts of interest: A.I.K. is the President and a part-owner of Q-Chem, Inc.
References


