THE JOURNAL OF PHYSICAL CHEMISTRY A
A JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

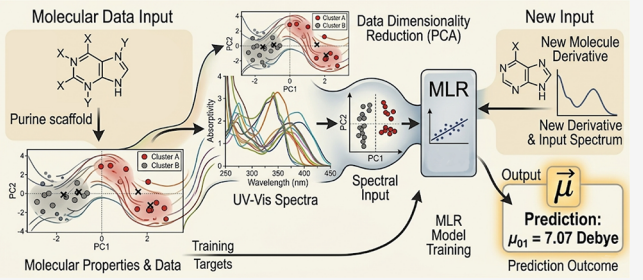
pubs.acs.org/JPCA
Article

Multivariate Analysis of Linear and Nonlinear Optical Properties in Purine Derivatives: A Predictive Framework from One-Photon Absorption Spectra

Ian R. Andrade and Leandro H. Zucolotto Cocca*

Cite This: J. Phys. Chem. A 2026, 130, 5205−5214
Read Online

ABSTRACT: The rational design of fluorescent organic molecules is central to the development of advanced linear and nonlinear photonic materials. Purine-based compounds have emerged as promise candidates for several photonics applications due to their structural similarity to biological nucleobases synthetic versatility and favorable photophysical properties. However, their optical characterization typically generates large and complex data sets that are difficult to interpret, particularly when multiple compounds are analyzed simultaneously. Here, we apply principal component analysis (PCA) to a series of purine derivatives to systematically investigate the relationships between molecular descriptors and photophysical performance. The PCA model applied in the optical properties of the set captures 76.8% of the total variance within the first two principal components, enabling clear clustering of molecules according to their electronic structure. Importantly, by applying PCA directly to one- and two-photon absorption spectra, we achieve effective spectral deconvolution with 91.87% and 94.51%, respectively, isolating contributions associated with intensity, spectral shifts, and bandwidth. The robustness of this approach is validated through accurate spectral reconstruction. To extend the analysis toward predictive modeling, multiple linear regression (MLR) was employed to correlate PCA-derived features from one-photon absorption data with the transition dipole moment (μ_{01}). The proposed PCA-MLR framework effectively captures the intrinsic relationships within the spectra of the studied group, minimizing the need for extensive experimental trials. The resulting model exhibits excellent predictive performance (R^2 = 0.9728) and accurately estimating the μ_{01} = 7.07D of an external validation molecule with a deviation of approximately 2.5%. Overall, this PCA-MLR framework provides a powerful and efficient strategy for interpreting complex photophysical data sets and accelerating the design and optimization of organic molecules for linear and nonlinear photonic applications.



1. INTRODUCTION

In recent years, the development of advanced photonic materials has been driven by the rational design of fluorescent organic molecules with tailored optical properties.[1−3] These materials play a central role in applications ranging from fluorescence microscopy,[4] and optical data storage[5] to photodynamic therapy[6−8] and nonlinear optics.[9,10] In this context, heterocyclic fluorophores have attracted particular attention due to their structural versatility and tunable electronic properties. Among them, purine-based compounds[11,12] stands out as highly promising systems owing their structural similarity to biological nucleobases,[13] ease of functionalization,[11] and their favorable photophysical properties.[14] These properties include large Stokes shifts,[15,16] high fluorescence quantum yields,[17] and the potential for multiphoton absorption processes.[17−19] Such features make purine derivatives attractive candidates for developing nucleobase analogs and fluorescent probes for biomolecular imaging and spectroscopy.[19−21]

Nitrogen-containing heterocyclic compounds have emerged as highly promising scaffolds by the incorporation of these heteroaromatics into conjugated π-systems, significantly enhancing molecular polarizability and intramolecular charge transfer dynamics.[22−24] A widely adopted strategy to enhance the optical response of organic chromophores includes the design of push−pull systems[17,25,26] or increase the conjugation π-conjugation.[27] The design of push−pull systems can be achieved through introducing electron-donating and electron-accepting substituents at strategic positions[28] within the molecular scaffold to enhance intramolecular charge transfer.[29,30] Functionalizing the purine core with appropriate donor−acceptor moieties offers a versatile way to tailor

Received: March 30, 2026
Revised: June 13, 2026
Accepted: June 15, 2026
Published: June 24, 2026

https://doi.org/10.1021/acs.jpca.6c02028
J. Phys. Chem. A 2026, 130, 5205−5214

excited-state properties and improve optical responses across a broad spectral range.[31] In these systems, such modifications significantly influence both linear and nonlinear optical responses. Moreover, these modifications govern critical nonlinear optical parameters, such as the nonlinear refractive index (n_2) and the nonlinear absorption coefficient (β), which directly impact the two-photon absorption (2PA) cross-section.[17,22−24,32−34]

2PA is a third-order nonlinear optical process in which a material is excited from the ground state to an excited electronic state[33] by the simultaneous absorption of two photons. The probability of 2PA is proportional to the square of the incident light intensity.[35] This relationship results in high spatial selectivity and minimizes photodamage to surrounding regions.[36] These features have driven the development of organic molecules with enhanced 2PA responses, particularly through systematic molecular design involving π-extension and donor/acceptor substitutions.[37,38] For example, Mckee et al.[39] investigated the impact of laser power, scanning speed, hatch distance, and layer height on the structural resolution and fabrication efficiency of microneedles produced via two-photon polymerization. Moreover, Liu and Qian[40] in their characterization and evaluation of novel BODIPY photosensitizers for singlet oxygen generation via two-photon absorption reported significant 2PA cross-sections evaluated at a single excitation wavelength of 1050 nm, resulting in 661.8 and 715.6 GM for Id-BDPI and Cz-BDPI, respectively, which highlight their potential for two-photon photodynamic therapy applications.

Driven by these promising applications, the linear and nonlinear optical characterization of novel molecules results in a complex and high-dimensional data set, including linear absorption spectra, steady-state and time-resolved fluorescence, solvatochromic behavior, fluorescence anisotropy, and nonlinear absorption responses, usually obtained via techniques like open-aperture Z-scan.[41,42] While detailed photophysical characterization provides critical insights into the structure−property relationships, the sheer volume and complexity of the data make interpretation challenging. This is especially true when multiple compounds are evaluated simultaneously, especially in the presence of subtle structural variations that lead to correlated optical responses.

In this context, multivariate statistical methods, particularly principal component analysis (PCA),[43] offer a powerful framework for extracting meaningful information from complex data sets. PCA reduces the dimensionality of correlated variables while preserving the most significant variance in the data set.[44] By transforming the original data into a set of orthogonal principal components, PCA enables identification of patterns, clustering behavior, and key descriptors governing optical properties. This approach is particularly useful for nonlinear optical studies, where structure−property relationships are influenced by interdependent electronic, geometric, and environmental parameters.[45,46] It is important to mention that previous studies have successfully applied PCA to analyze the 2PA response of imidazo[4,5-*b*]pyridine derivatives (purine isosteres[47]) and donor−acceptor systems.[44] This allowed for the identification of critical structural descriptors responsible for enhanced cross-section values. Although the photophysics of purine analogues, particularly 2-aminopurine (2AP), one of the most extensively studied representatives, has been thoroughly characterized experimentally,[16,17,48] the application of PCA to systematically map the linear and nonlinear optical behavior of purine-based chromophores remains largely underexplored.

Here, we bridge this gap by applying PCA to a series of push−pull purine derivatives[16,17] and already known purines derivatives such as 2AP.[48] These compounds have been previously characterized in depth using a range of linear and nonlinear spectroscopic techniques.[17] It should be mentioned that, to the best of our knowledge, the set of compounds investigated in this study encompasses all purine derivatives that have the photophysical properties of interest already reported in the literature. Our main objective is to systematically investigate how molecular and electronic descriptors (such as absorption maxima, emission properties, fluorescence lifetimes, anisotropy, solvatochromism, and 2PA cross-sections) correlate and contribute to the overall photonic performance of these compounds. By integrating experimental data with multivariate statistical analysis, we aim to (i) apply PCA to the experimentally determined photophysical parameters, evaluating the inherent clustering and group differentiation of the purine derivatives based on their optical responses, (ii) perform direct dimensionality reduction on the continuous 1PA and 2PA spectra, effectively decoupling overlapping photophysical features and validating the model through high-fidelity spectral reconstruction, (iii) establish a robust predictive framework, coupling PCA with MLR, capable of accurately forecasting complex optical properties directly from routine linear absorption data, and (iv) demonstrate the usefulness of PCA as an auxiliary tool to support the analysis and interpretation of structure−property relationships in these purine-based fluorophores.

This PCA-guided approach not only complements traditional phenomenological and quantum chemical analyses but also offers a data-driven framework for accelerating the discovery of high-performance fluorophores for applications in fluorescence spectroscopy, nonlinear imaging, and nucleic acid sensing.

2. MATERIALS AND METHODS

2.1. Compounds

Purine derivatives are a class of heterocyclic compounds with high thermodynamic stability and notable optical and electronic properties, making them highly relevant for studies in materials physics and biophysics.[49] The molecular structure, which consists of a fused pyrimidine and an imidazole ring,[50] imparts a rich π-electron aromatic nature to these compounds. This electronic resonance is responsible for a strong absorption in the ultraviolet (UV−vis) range, with characteristic spectra that are dependent on pH and the local environment.[51] Furthermore, the presence of functional groups such as amines and carbonyls enables the formation of directional hydrogen bonds, which are crucial for self-assembly into supramolecular structures like the DNA and RNA double helices.[52] In materials science, purines can act as building blocks for semiconducting polymers and hybrid materials.[53] Their charge transport properties, influenced by π-orbital overlap, are of particular interest for applications in bioelectronics and optoelectronics.[54] The interaction of purines with surfaces and nanoparticles is also explored for the development of sensors and detection devices.[55] A representation of the molecular structures of the 19 molecules studied here is shown in Figure 1.

The multivariate analysis performed herein is based on the comprehensive data set of optical properties previously established by Cocca and co-workers[16,17] and by Mikhaylov et al.[48] described in Table 1.

https://doi.org/10.1021/acs.jpca.6c02028
J. Phys. Chem. A 2026, 130, 5205−5214

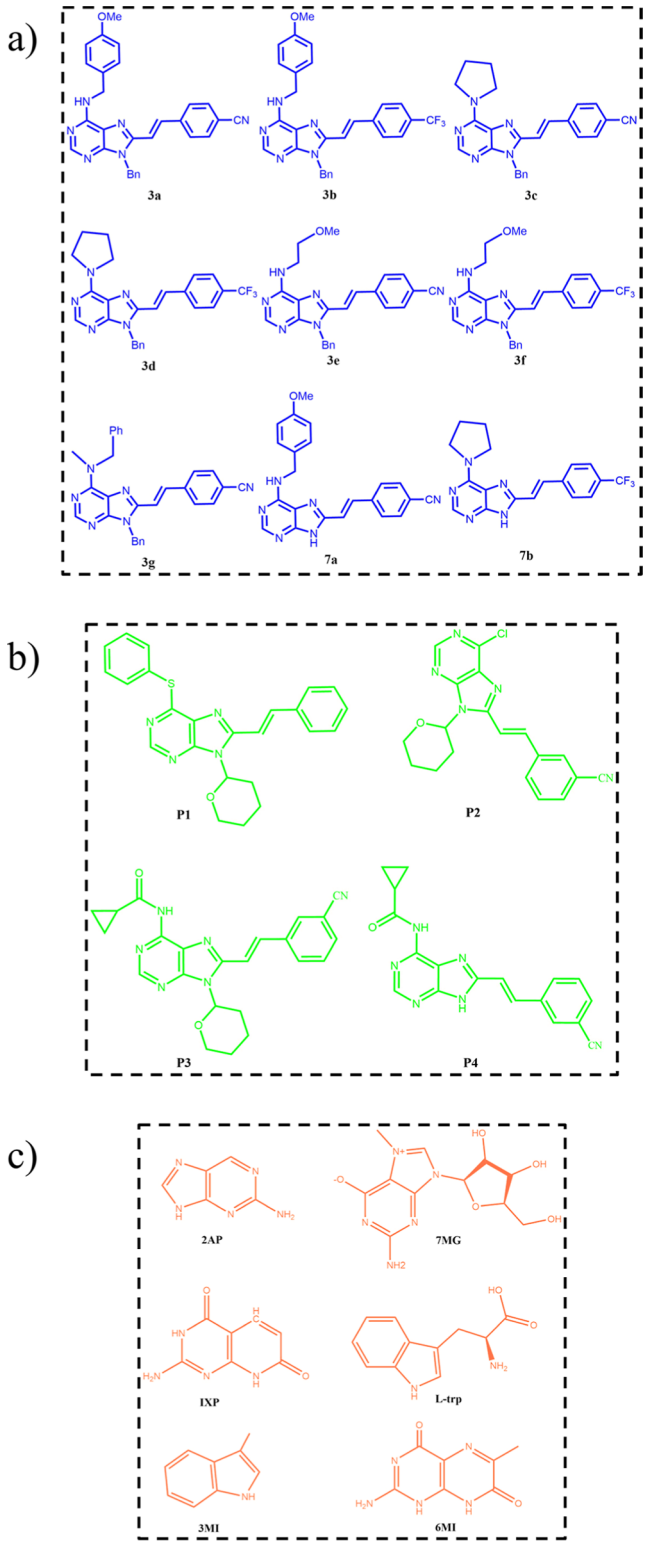


Figure 1. Representation of chemical structures of the investigated purine derivatives. Panels (a) and (b) depict push−pull systems from refs 16 and 17, respectively, while panel (c) shows additional purine derivatives structures from ref 48.

2.2. Principal Components Analysis

To systematically analyze the multidimensional data set of spectroscopic parameters, PCA was performed utilizing the Scikit-learn module in Python (v. 3.11). Given that the optical parameters evaluated inherently possess disparate magnitudes and units (e.g., transition dipole moments, spectral bandwidths, and cross-sections), direct application of PCA would mathematically bias the model toward variables with larger numerical scales. To prevent this artifact, each feature was standardized to unit variance according to

$$z = \frac{x - \mu}{\sigma} \quad (1)$$

Following standardization in eq 1, where x is the raw parameter value, μ is the mean of the feature across all samples, and σ is its standard deviation, PCA was executed via singular value decomposition (SVD) of the covariance matrix to orthogonally transform the correlated variables into a new set of uncorrelated variables, termed principal components (PCs). The optimal number of retained PCs was determined based on the cumulative explained variance, ensuring that the fundamental photophysical information on the data set was captured while effectively filtering out multidimensional noise. The outputs of the PCA were evaluated through two primary projections. Score plots were constructed to project the purine derivatives into the newly defined multidimensional space, facilitating the visual identification of inherent sample clustering, similarities, and potential outliers. Concurrently, loading vectors were extracted to quantify the statistical weight and correlation of each original spectroscopic parameter with the respective principal components, allowing for a physical interpretation of the mathematical axes.

Finally, to rigorously validate the visual groupings observed in the score plots, unsupervised machine learning algorithms were complementarily employed. Both *k*-means clustering and hierarchical clustering were applied directly to the principal component scores, providing a quantitative and objective framework to confirm the structure−property classifications within the data set.

2.3. Spectral PCA

To directly evaluate the continuous photophysical profiles, PCA was independently applied to the 1PA and 2PA spectral data. To ensure mathematical validity, the input data matrices were strictly confined to the common intersecting spectral windows shared across all evaluated purine derivatives. In contrast to the preprocessing of discrete

Table 1. Optical Parameters of the 22 Molecules Addressed in This Work

Sample	ϕ_f (%)	$\Delta\mu_{01}$ (D)	λ_{1PA} (nm)	ε (10^4 M^{-1} cm^{-1})	σ (GM)	λ_{2PA} (nm)
3a	75	10.2	370	1.9	20	745
3b	81	9.9	357	1.5	19	700
3c	62	12.4	391	1.9	24	782
3d	79	9.7	375	1.6	24	750
3e	63	12.5	370	1.55	20	740
3f	65	10.7	357	1.6	18	705
3g	63	6.1	383	2.4	30	760
7a	81	14.1	373	1.9	19	740
7b	99	11.4	367	1.6	28	720
P1	8	4	349	3.3	15.3	700
P2	14	2.7	334	3.1	11.9	690
P3	57	6	340	3.3	23.7	695
P4	64	8	340	3.2	22.3	660
L-trp	12	2.1	279	0.55	0.4	580
3MI	2.3	2	280	0.55	0.5	580
7MG	1.2	2	258	0.55	3.6	560
IXP	7	1	340	1.4	0.7	670
6MI		2.8	343	1.009	1.53	701
2AP-H_2O	58	1.1	305	0.556	0.2	612
2AP-Methanol	80	1	310	0.631	0.2	622
2AP-DMSO	70	1.2	313	0.593	0.4	626
2AP-Glycerol	22	1.1	312	0.625	2.4	625

https://doi.org/10.1021/acs.jpca.6c02028
J. Phys. Chem. A 2026, 130, 5205−5214

parameters, the spectral data were subjected solely to mean-centering, calculated as

$$X = x_i - \bar{x} \quad (2)$$

In eq 2, x_i is the original value and $\bar{x}$ is the mean value of the wavelength across all samples, and X is the centered spectra. This step is a methodological choice for continuous spectral data to prevent the artificial amplification of baseline noise in nonabsorbing regions.[56,57] For both the 1PA and 2PA spectral analyses, scree plots were computed to quantify the cumulative explained variance, specifically validating the high proportion of spectral information retained within the first two principal components. Subsequently, score plots were generated to map the spatial distribution of the molecules in the reduced dimensional space. This projection allowed for the rapid identification of molecular clusters exhibiting analogous overall spectral profiles.

2.4. PCA and Multiple Linear Regression (MLR)

To transition from exploratory analysis to predictive modeling, MLR was conducted.[58] In this step, the principal component scores extracted from the linear absorption spectra were utilized as independent variables to predict target optical parameters. This methodology offers a robust and reproducible framework for applying PCA to spectroscopic data sets, facilitating the understanding of the optical behavior and structure−property relationships across diverse material systems. In this framework, the previously extracted principal component scores functioned directly as the independent predictor variables, while the photophysical parameter of interest served as the dependent response variable. By utilizing these orthogonal scores rather than the highly correlated raw spectral data, the regression strictly avoids multicollinearity artifacts. This predictive relationship is defined by the generalized equation[58]

$$y = b_0 + \sum_{i=1}^{k} b_i x_i + \epsilon \quad (3)$$

In this equation, y is the estimated target parameter, b_0 represents the intercept, b_i denotes the regression coefficients for each i-th independent variable (x_i), that could be the principal component scores for PCA-MLR, and ϵ accounts for the residual error.

3. RESULTS AND DISCUSSION

3.1. PCA of Photophysical Parameters

For the PCA based on photophysical parameters, a data set containing 22 molecules was analyzed. The feature space was limited to the intersection of parameters available for all compounds, specifically, the linear optical characterization provided by Mikhaylov et al.,[48] which provides the smallest set of measured variables and defines the maximum number of parameters that could be consistently included in the analysis. Additionally, the fluorescence quantum yield value for the 6MI molecule was not reported in the same study as show in Table 1. To avoid excluding this compound from the data set, the missing value was estimated using the k-nearest neighbors (k-NN) method.[59] Prior to PCA, all variables were standardized to zero mean and unit variance in order to ensure a proper comparison between parameters with different scales.

For the set of optical parameters evaluated for the purine derivatives, the following variables were considered: fluorescence quantum yield (ϕ_f), permanent dipole moment difference between the first excited state and ground one ($\Delta\mu_{01}$), wavelength of 1PA (λ_{1PA}), molar absorptivity (ε), the two-photon absorption cross-section (σ), and wavelength of 2PA (λ_{2PA}), as summarized in Table 1. In addition, the solvent used for each molecule, dichloromethane (DCM), dimethyl sulfoxide (DMSO), or water, was included as the categorical variable. This was implemented using one-hot encoding,[60] where 1 indicates presence and 0 indicates absence, as shown in Table 2. For the molecules 7MG, IXP, 6MI, and 2AP-glycerol, no solvent descriptor was included since each compound required a distinct solvent. The inclusion of additional unique solvent categories would introduce variables with no shared variance across the data set, leading only to a redistribution of the explained variance among the principal components without providing meaningful correlations.

Table 2. Solvent Parameters for Each Molecule with One-Hot Encoding

Amostra	DMSO	DCM	H_2O
3a	0	1	0
3b	0	1	0
3c	0	1	0
3d	0	1	0
3e	0	1	0
3f	0	1	0
3g	0	1	0
7a	1	0	0
7b	1	0	0
P1	1	0	0
P2	1	0	0
P3	1	0	0
P4	1	0	0
l-trp	0	0	1
3MI	0	0	1
7MG	0	0	0
IXP	0	0	0
6MI	0	0	0
2AP-H_2O	0	0	1
2AP-Methanol	0	0	0
2AP-DMSO	1	0	0
2AP-Glycerol	0	0	0

Using the optical parameters listed in Table 1 and the solvent descriptors in Table 2, the first two principal components (PC1 and PC2) account for 76.82% of the total variance. The cumulative variance explained by the first nine principal components is presented in Figure 2. The high contributions of PC1 and PC2 indicate that the dimensionality of the problem can be significantly reduced without substantial loss of information, enabling a reliable interpretation of the correlations among the optical parameters.

The analysis of PCA scores for the optical parameters showed a clear separation and clustering in 2 different groups, as exhibited in Figure 3. The first cluster (highlighted in red) exclusively comprises the molecules featuring a push−pull architecture, characterized by a unified donor−π−acceptor (D−π−A) conjugated system. Conversely, the second cluster (colored gray) encompasses the well-established, commercially available purine derivatives. Unlike the first group, this subset lacks a common D−π−A structural motif, representing a broader class of standard purine architectures. Both clusters were closed by 90% confidence ellipses through the centroid of each set.

The PCA results clearly reveal the clustering behavior of the investigated molecules. PC1 is strongly influenced by the optical parameters, including ϕ_f, μ_{01}, λ_{1PA}, ε, σ, λ_{2PA}. The strong contribution of these variables allows a clear separation of the data points, reducing the likelihood of artificial clustering in the score plot.

https://doi.org/10.1021/acs.jpca.6c02028
J. Phys. Chem. A 2026, 130, 5205−5214

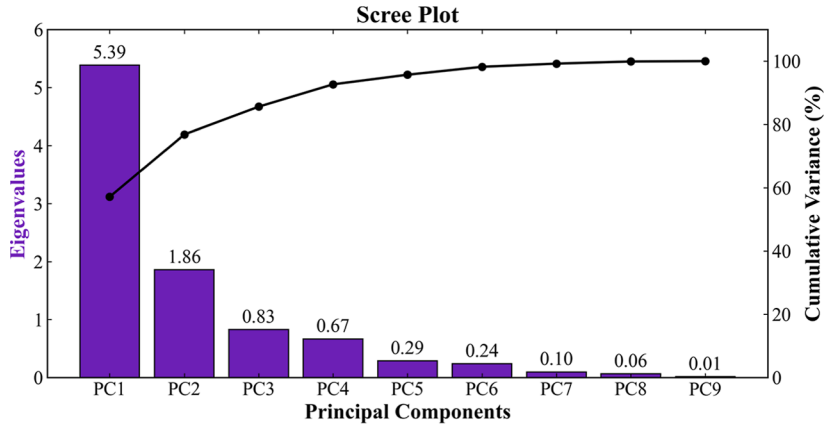


Figure 2. Scree plot showing the eigenvalues and cumulative explained variance from the PCA applied to the optical parameters of the purine derivatives. The first two PCs account for 76.82% of the total variance.

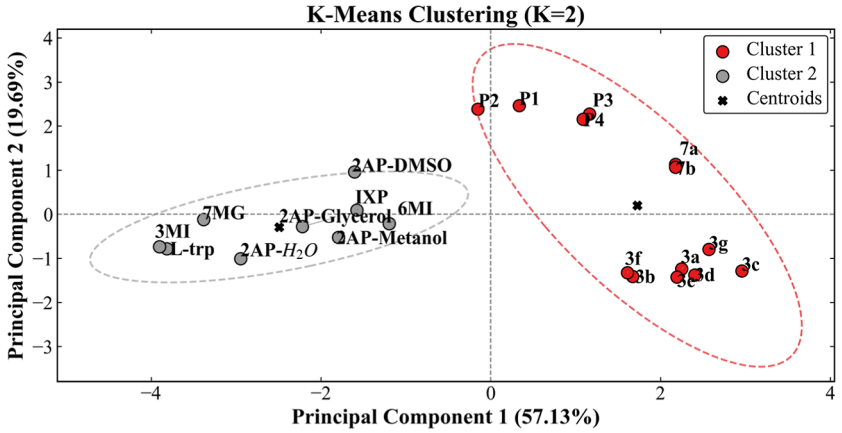


Figure 3. Score plots of the purine derivatives. Each data point represents an individual molecule, and the axes correspond to the principal components that capture the strongest and weakest correlations among their photophysical properties.

Figure 4 shows the loading plot, which is essential for identifying the variables that contribute most to the variance explained by the principal components. The projection of the variables in the PC1-PC2 space provides information about the direction and magnitude of each spectroscopic parameter. This allows for the identification of the optical properties that most strongly drive the data set distribution, particularly when evaluated in conjunction with the corresponding molecular positions in the score plot. In this representation, each vector starting from the origin corresponds to a variable, and its length is proportional to its contribution to the principal components. The angle between vectors reflects the correlation between variables: small angles indicate strong positive correlation, angles close to 180° indicate negative correlation, and angles near 90° indicate weak or negligible correlation.[61] The combined interpretation of the score plot (Figure 3) and the loading plot (Figure 4) makes it possible to identify molecular clusters associated with enhanced linear and nonlinear optical responses, according to the positive or negative correlations between the variables and the principal components.

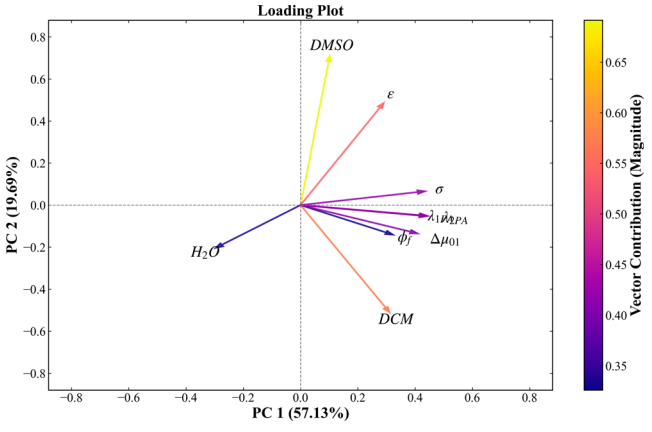


Figure 4. Loading plot of push−pull purines derivatives showing the direction and magnitude of each spectroscopic parameter, where the solvents have pronounced magnitude and the molar absorptivity.

For example, the vector associated with ϕ_f is predominantly oriented in the positive direction of PC1, indicating that PC1 is highly correlated with this optical property. This behavior is consistent with the score plot (Figure 3), where molecules located along the positive PC1 axis correspond to those exhibiting higher ϕ_f. Thus, the loading plot enables a direct correlation between the position of each molecule in the score space and its corresponding photophysical properties. This combined interpretation provides a quantitative framework for understanding the structure−property relationships governing the optical behavior of the purine derivatives. The observed correlations between molecular features and optical responses further highlight the effectiveness of PCA as a tool for linking

https://doi.org/10.1021/acs.jpca.6c02028
J. Phys. Chem. A 2026, 130, 5205−5214

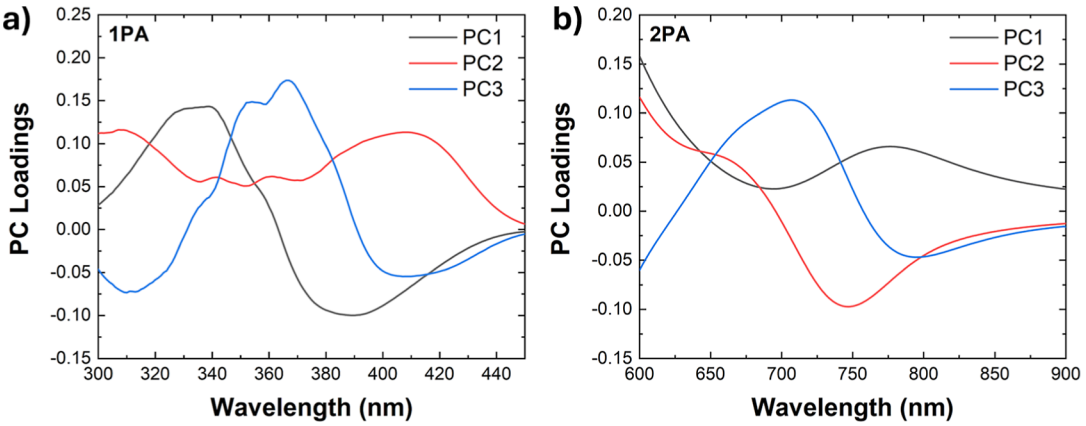


Figure 5. Loading plot showing the first 3PCs of (a) one-photon absorption and (b) two-photon absorption spectra of 13 purines.

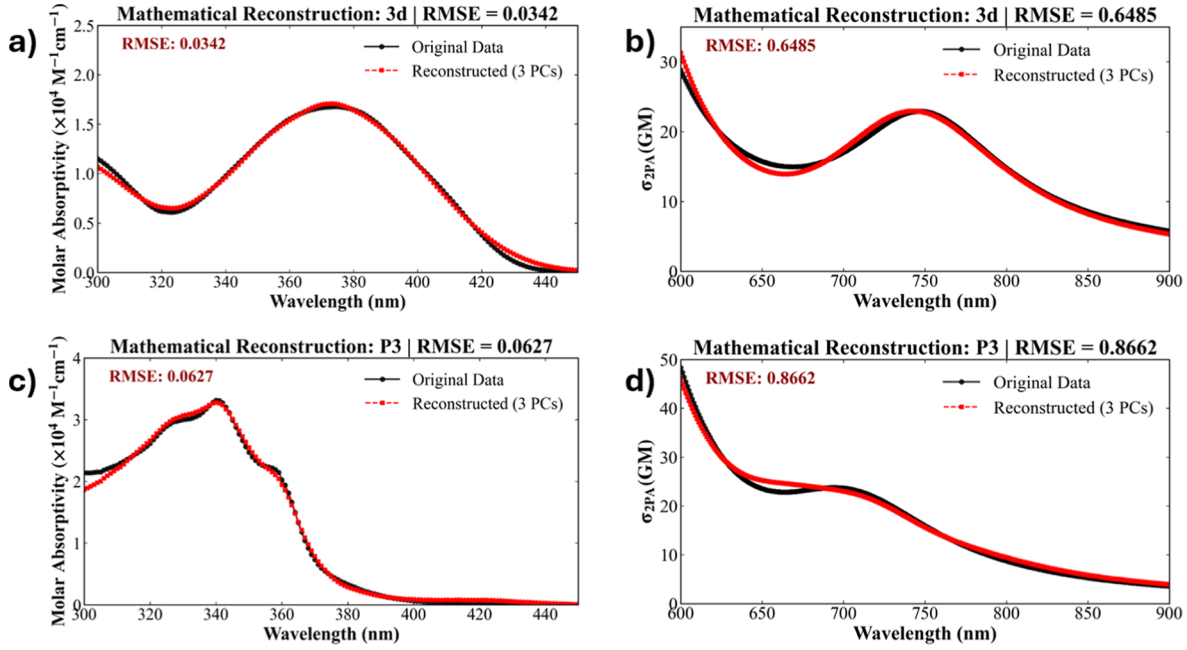


Figure 6. Spectral reconstruction of 1PA (a,c) and 2PA (b,d) profiles based on three principal components. The graphs (a,b) are related with molecule 3d from Figure 1a. The graphs (c,d) are related with molecule P3 from Figure 1b.

molecular architecture to optical properties in organic photonic materials.

3.2. 1PA and 2PA Spectral PCA

While PCA based on scalar photophysical parameters provides valuable classifications regarding solvent effects and other molecular properties, it does not capture the full spectral response of each compound. To address this limitation, the analysis was extended to the linear and nonlinear absorption data by treating the entire spectra as data sets. Applying PCA directly to the UV−vis (1PA) and 2PA cross-section spectra of the selected compounds enables a more comprehensive description of their optical behavior.

Performing a direct PCA on these spectra requires a common wavelength range across all molecules. However, for the full set of 22 compounds (Figure 1), the absorption bands span significantly different regions, restricting the shared spectral overlap to a narrow window of approximately 60 nm (270−330 nm). Conducting PCA over such a limited range would yield insufficient variance representation, failing to capture the primary spectral features and the overall variability of the absorption bands. To overcome this, the data set was restricted exclusively to the push−pull purine derivatives (Figures 1a,b), ensuring a broad, common spectral window for all analyzed samples. This targeted selection enabled the inclusion of the lowest-energy absorption band for each compound, covering 300−450 nm for the 1PA spectra and 600−900 nm for the 2PA spectra. As all variables correspond to intensity values in identical units, the spectral data were mean-centered prior to analysis.

According to the preset established, the first two principal components accounted for approximately 91.87% of the total spectral variance in 1PA data set and 94.51% for 2PA data set (scree and score plots for the spectral analyses are provided in the Supporting Information).

The loading plot captures and identifies wavelengths that contribute most significantly to spectral variance. In the 2PA analysis, these wavelengths corresponded to the same transition as the 1PA;[17,62,63] consequently, the PCA applied to the 1PA and 2PA spectra is expected to exhibit similar spectral patterns. Notably, molecules with pronounced intra-

https://doi.org/10.1021/acs.jpca.6c02028
J. Phys. Chem. A 2026, 130, 5205−5214

molecular charge transfer (ICT) character show strong loadings at longer wavelengths,[64] confirming that PCA capture not only intensity-based differences but also spectral shape and position.

The 1PA PC1 loadings in Figure 5a reveal that the first principal component is primarily driven by the spectral position of the absorption band and intensity, alone accounting for 77.93% of the total data set variance. PC2 and PC3 capture spectral shifts and band broadening that result from the experimental response of molecules absorbing at different wavelengths and exhibiting different absorptivity intensities.[65] This grouping arises from the structural similarities and charge-transfer properties shared by the molecules, despite their distinct linked groups.

Likewise analogous features are observed in the 2PA loadings Figure 5b. In this case, PC1 comprises strictly positive loadings that reflect the overall absorption intensity and mean spectral profile. Furthermore, the subsequent components isolate specific photophysical alterations: PC2 accounts for bathochromic or hypsochromic spectral shifts,[66] whereas PC3 captures changes in bandwidth and spectral broadening.[65] The 2PA behavior is primarily governed by the π-conjugation of the central pyridine rings, which serve as the core of the molecular structure[67] and can impact in the PC2 directly.

The application of PCA to the 1PA and 2PA spectra provides robust evidence for the noncentrosymmetric push-pull architecture, a characteristic clearly reflected in both the experimental data sets and the individual PC loadings. The high degree of correlation between the experimental data and the theoretical model, as indicated by PCA, suggests a high level of confidence in the results. To validate the robustness of the PCA model, the original spectral data were reconstructed utilizing the derived principal component loadings and their corresponding scores, as shown in Figure 6. Notably, the root-mean-square error (RMSE) associated with each presented reconstruction is low, relative to the absolute scale of the spectral intensity. This minimal residual confirms the high fidelity of dimensionality reduction. The complete set of reconstructed spectra can be found in the Supporting Information.

3.3. PCA-MLR Prediction

Having established that the PCA-based reconstruction reproduces the experimental spectra with minimal residual error, the extracted principal components can be considered to retain the essential photophysical information on the data set. Building on this validated dimensionality reduction, the analysis was extended to the predictive modeling stage. In this way, here we investigate whether key spectroscopic parameters can be estimated using only 1PA spectral data as input variables. This approach is intended as a complementary tool for experimental research, enabling the prediction of optical properties from the standard linear measurements. Such a strategy allows rapid in silico screening of optical responses, reducing the need for exhaustive experimental characterization and providing a practical route to accelerating the rational design and optimization of organic materials for photonic applications.

To perform the predictive modeling, the μ_{01} was selected as the target physical parameter due to its fundamental role in governing the transition probability and its direct correlation with the optical absorption profile.[68] A multiple linear regression approach[69] was employed to stablish a relationship between spectral data and this photophysical property. To rigorously validate the predictive capability of the model, a target molecule (P3 shown in Figure 1b) was intentionally excluded from the data set of 13 push−pull purine derivatives. The PCA was then performed strictly on the remaining 12 molecules, ensuring that the multidimensional feature space was constructed without prior information from the P3 spectrum. Following the dimensionality reduction, MLR was conducted by regressing the experimental μ_{01} values of the 12 molecules training set against their respective principal component scores, yielding a predictive equation of the general form

$$\hat{\mu}_{01} = b_0 + \sum_{i=1}^{k} b_i PC_i \tag{4}$$

In this equation, $\hat{\mu}_{01}$ is the estimated transition dipole moment, b_0 is the intercept, b_i is the regression coefficient, and PC_i denotes the respective principal component scores. Finally, to test the model, the absorption spectrum of the withheld molecule (P3) was projected onto the established PCA space to extract its scores, which were then applied to the MLR equation to estimate its μ_{01} value.

Applying the coupled PCA-MLR methodology to the 12-purine training set yielded the following predictive equation based on the first three principal components

$$\hat{\mu}_{01} = 0.1013 \cdot PC_1 + 0.1927 \cdot PC_2 + 0.0082 \cdot PC_3 + 5.5500 \tag{5}$$

The regression model exhibited strong predictive linearity, this is evidenced by a coefficient of determination $R^2 = 0.9728$, indicating an excellent correlation between the reduced spectral features and the target physical property. To evaluate the predictive capability of the model, the experimental spectrum of the molecule P3 was projected onto the established PCA space. The model estimated a transition dipole moment of $\hat{\mu}_{01} = 7.07D$. When compared to the reference value of $\mu_{01} = 6.9$D previously reported by Cocca et al.,[16] this estimation yields a remarkably low relative error of approximately 2.46%. This high degree of accuracy underscores the robustness of the proposed framework, demonstrating its efficacy as a rapid, concise, and highly auxiliary tool for accelerating the extraction of complex photophysical parameters from standard optical analyses.

Overall, the obtained results confirm the reliability of the PCA-based framework for describing and predicting the optical behavior of purine derivatives, as discussed in the following conclusions.

4. CONCLUSIONS

In conclusion, this study successfully establishes a robust multivariate statistical framework to comprehensively evaluate and predict the photophysical behavior of purine derivatives. Initially, the application of PCA to established photophysical parameters demonstrated the method's capability to cluster molecules based on structure−property relationships, despite dimensionality constraints inherent to combined literature data sets. By refining the analytical scope to a consistent spectral window, we performed direct dimensionality reduction on the one- and two-photon absorption spectra of 13 purines set. The extracted principal components effectively decoupled critical, overlapping photophysical features, namely, the overall

https://doi.org/10.1021/acs.jpca.6c02028
J. Phys. Chem. A 2026, 130, 5205−5214

absorption intensity, spectral shifts, and band broadening. The high fidelity of this mathematical reduction was corroborated by the successful reconstruction of the experimental spectra with minimal residual variance.

Based on this validated model, we advanced the exploratory analysis into a highly accurate predictive tool. By coupling PCA with multiple linear regression, we demonstrated that optical parameters, such as the transition dipole moment, can be reliably forecasted relying exclusively on routine 1PA spectral scores. The external validation yielded an excellent predictive linearity of R^2 = 0.9728 and a remarkably low relative error ∼ 2.5%. Ultimately, this PCA-MLR methodology provides a powerful auxiliary tool for experimental results, in silico screening strategy that accelerates the determination of optical properties from standard linear measurements, optimizing the rational design of new materials for nonlinear photonics.

■ ASSOCIATED CONTENT

Data Availability Statement

All data sets generated or analyzed during this study are included in this article.

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jpca.6c02028.

PCA applied to the one-photon absorption spectra, including the corresponding scree plot, score plot, and the complete set of reconstructed spectra with their respective root-mean-square error values; and analogous analysis for the two-photon absorption spectra, comprising the scree plot, score plot, and the full spectral reconstruction data set with RMSE values (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Leandro H. Zucolotto Cocca − *Photonics Group, Institute of Physics, Federal University of Goiás, 74690-900 Goiânia, GO, Brazil*; orcid.org/0000-0003-4587-4062; Email: leandro.zucolotto@ufg.br, ian_ribeiro@discente.ufg.br

Author

Ian R. Andrade − *Photonics Group, Institute of Physics, Federal University of Goiás, 74690-900 Goiânia, GO, Brazil*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jpca.6c02028

Author Contributions

Ian Ribeiro de Andrade: conceptualization, data curation, formal analysis, investigation, methodology, software, and writing—original draft. **Leandro H. Zucolotto Cocca**: conceptualization, formal analysis, methodology, validation, visualization, writing—review and editing, supervision, and funding acquisition.

Funding

The Article Processing Charge for the publication of this research was funded by the Coordenacao de Aperfeicoamento de Pessoal de Nivel Superior (CAPES), Brazil (ROR identifier: 00x0ma614).

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank the National Council for Scientific and Technological Development (CNPq, grants 303861/2024-0 L.H.Z.C.) and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES; Finance Code 001).

■ REFERENCES

(1) Zhang, C.; Zhao, Y. S. Flexible Photonic Materials and Devices: Synthetic Strategies, Sensing Properties, and Wearable Applications. *Adv. Mater.* **2025**, *37*, 2415856.

(2) Jeon, J.; Bukharina, D.; Kim, M.; Kang, S.; Kim, J.; Zhang, Y.; Tsukruk, V. Tunable and Responsive Photonic Bio-Inspired Materials and Their Applications. *Responsive Mater.* **2024**, *2* (1), No. e20230032.

(3) Yazhgur, P.; Muller, N.; Aeby, S.; Aguilar Uribe, A.; Scheffold, F. Printing Structurally Colored Photonic Materials. In *Proc. SPIE 12481, Bioinspiration, Biomimetics, and Bioreplication XIII* **2023**, *16*. DOI: .

(4) Van Der Velde, J. H. M.; Smit, J. H.; Hebisch, E.; Trauschke, V.; Punter, M.; Cordes, T. Self-Healing Dyes for Super-Resolution Fluorescence Microscopy. *J. Phys. D: Appl. Phys.* **2018**, *52* (3), 034001.

(5) Zhuang, Y.; Ren, X.; Che, X.; Liu, S.; Huang, W.; Zhao, Q. Organic Photoresponsive Materials for Information Storage: A Review. *Adv. Photonics* **2020**, *3*(01)..

(6) Tang, Y.; Li, Y.; Li, B.; Song, W.; Qi, G.; Tian, J.; Huang, W.; Fan, Q.; Liu, B. Oxygen-Independent Organic Photosensitizer with Ultralow-Power NIR Photoexcitation for Tumor-Specific Photodynamic Therapy. *Nat. Commun.* **2024**, *15* (1), 1−13.

(7) Joshi, P.; Soares, J. M.; Martins, G. M.; Zucolotto Cocca, L. H.; De Boni, L.; de Oliveira, K. T.; Bagnato, V. S.; Blanco, K. C. Enhancing the Efficacy of Antimicrobial Photodynamic Therapy through Curcumin Modifications. *Photochem. Photobiol.* **2025**, *101* (2), 359−372.

(8) Cocca, L. H. Z.; Oliveira, T. M. A.; Gotardo, F.; Teles, A. V.; Menegatti, R.; Siqueira, J. P.; Mendonça, C. R.; Bataus, L. A. M.; Ribeiro, A. O.; Souza, T. F. M.; Souza, G. R. L.; Gonçalves, P. J.; De Boni, L. Tetracarboxy-Phthalocyanines: From Excited State Dynamics to Photodynamic Inactivation against Bovine Herpesvirus Type 1. *J. Photochem. Photobiol., B* **2017**, *175*, 1−8.

(9) Zucolotto Cocca, L. H.; Valverde, J. V. P.; Leite, C. M.; Moreno, N. S.; Neto, A. L.; Macedo, A. G.; Pratavieira, S.; Silva, D. L.; Rodrigues, P. C.; Zucolotto, V.; Mendonça, C. R.; De Boni, L. Advancements in Organic Fluorescent Materials: Unveiling the Potential of Peripheral Group Modification in Dithienyl-Diketopyrrolopyrrole Derivatives for One- and Two-Photon Bioimaging. *J. Mater. Chem. B* **2025**, *13* (3), 1013−1023.

(10) Zucolotto Cocca, L. H.; Valverde, J. V. P.; le Bescont, J.; Breton-Patient, C.; Piguel, S.; Silva, D. L.; Mendonca, C. R.; De Boni, L. Photophysical Properties of 3-Arylthioimidazo[1,2-a]Pyridine Derivatives: The Role of Peripheral Electron-Donating and Electron-Withdrawing Groups in the Advance of Organic Materials Engineering. *J. Mol. Struct.* **2024**, *1300*, 137221.

(11) Shi, L.; Li, K.; Li, L. L.; Chen, S. Y.; Li, M. Y.; Zhou, Q.; Wang, N.; Yu, X. Q. Novel Easily Available Purine-Based AIEgens with Colour Tunability and Applications in Lipid Droplet Imaging. *Chem. Sci.* **2018**, *9* (48), 8969−8974.

(12) Chen, X.; Mao, Y.; Wang, A.; Lu, L.; Shao, Q.; Jiang, C.; Lu, H. Synthesis and Application of Purine-Based Fluorescence Probe for Continuous Recognition of Cu2+ and Glyphosate. *Spectrochim. Acta, Part A* **2024**, *304*, 123291.

(13) Guo, Y.; Zou, Y.; Chen, Y.; Deng, D.; Zhang, Z.; Liu, K.; Tang, M.; Yang, T.; Fu, S.; Zhang, C.; Si, W.; Ma, Z.; Zhang, S.; Peng, B.; Xu, D.; Chen, L. Design, Synthesis and Biological Evaluation of Purine-Based Derivatives as Novel JAK2/BRD4(BD2) Dual Target Inhibitors. *Bioorg. Chem.* **2023**, *132*, 106386.

(14) Martínez-Fernández, L.; Arslancan, S.; Ivashchenko, D.; Crespo-Hernández, C. E.; Corral, I. Tracking the Origin of

Photostability in Purine Nucleobases: The Photophysics of 2-Oxopurine. *Phys. Chem. Chem. Phys.* **2019**, *21* (25), 13467−13473.
(15) Alam, S.; Mao, Y.; Tao, X.; Zheng, S.; Jiang, C.; Chen, S. Y.; Lu, H. A Rapid and Highly Selective Purine-Based Fluorescent Probe for Hypochlorite Detection with Cellular Applications. *J. Mol. Struct.* **2025**, *1328*, 141345.
(16) Zucolotto Cocca, L. H.; Pelosi, A. G.; Piguel, S.; Mendonça, C. R.; Boni, L. De. Enhancement of Optical Properties of New Purine Nucleobases Containing Electron-Donating and -Withdrawing Peripheral Groups. *J. Photochem. Photobiol., B* **2022**, *234*, 112524.
(17) Zucolotto Cocca, L. H.; Abegaõ, L. M. G.; Sciuti, L. F.; Vabre, R.; De Paula Siqueira, J.; Kamada, K.; Mendonca, C. R.; Piguel, S.; De Boni, L. Two-Photon Emissive Dyes Based on Push-Pull Purines Derivatives: Toward the Development of New Photoluminescence Bioprobes. *J. Phys. Chem. C* **2020**, *124* (23), 12617−12627.
(18) Drobizhev, M.; Makarov, N. S.; Tillo, S. E.; Hughes, T. E.; Rebane, A. Two-Photon Absorption Properties of Fluorescent Proteins. *Nat. Methods* **2011**, *8* (5), 393−399.
(19) Bood, M.; Füchtbauer, A. F.; Wranne, M. S.; Ro, J. J.; Sarangamath, S.; El-Sagheer, A. H.; Rupert, D. L. M.; Fisher, R. S.; Magennis, S. W.; Jones, A. C.; Höök, F.; Brown, T.; Kim, B. H.; Dahlén, A.; Wilhelmsson, L. M.; Grøtli, M. Pentacyclic Adenine: A Versatile and Exceptionally Bright Fluorescent DNA Base Analogue. *Chem. Sci.* **2018**, *9* (14), 3494−3502.
(20) Yin, L.; Huang, J. B.; Yue, T. C.; Wang, L. L.; Wang, D. Z. Two 2D Metal-Organic Frameworks Based on Purine Carboxylic Acid Ligands for Photocatalytic Oxidation of Sulfides and CO2 Chemical Fixation. *Inorg. Chem.* **2024**, *63* (20), 9109−9118.
(21) Jones, A. C.; Neely, R. K. 2-Aminopurine as a Fluorescent Probe of DNA Conformation and the DNA−Enzyme Interface. *Q. Rev. Biophys.* **2015**, *48* (2), 244−279.
(22) Alam, Z.; Khan, S. A. Facile One-Pot Synthesis of Imidazole (MID): A Highly Sensitive Turn-Off Fluorescence Probe for Mercury (II) Ion Detection. *Jabirian J. Biointerface Res. Pharm. Appl. Chem.* **2026**, *3* (1), 1−7.
(23) Mujeeb, A.; Alam, M. Z.; Sultan, A.; Aleem Basha, H.; Khan, S. A.; Afzal, S. M. Synthesis, Physicochemical and Third Order Nonlinear Optical Properties of Bis-Chalcone (BBDP) as Donor-Pi Acceptor Chromophore in Organize Medium. *J. Fluoresc.* **2024**, *35* (3), 1393−1406.
(24) Mujeeb, M. A.; Sultan; Alam, M. Z.; Basha, H. A.; Afzal, S. M.; Khan, S. A. Synthesis, Spectroscopic Investigation of Imidazole Donor-π-Acceptor Chromophore: Opto-Electronics and Photonics Properties in Organized Medium. *J. Mol. Struct.* **2025**, *1341*, 142517.
(25) Lim, H. C.; Choi, M. S.; Chae, S.; Kim, H. J.; Kim, J. J.; Hong, J. I. Effect of a π-Linker of Push−Pull D−π−A Donor Molecules on the Performance of Organic Photodetectors. *J. Mater. Chem. C* **2020**, *8* (32), 11145−11152.
(26) Khadria, A.; De Coene, Y.; Gawel, P.; Roche, C.; Clays, K.; Anderson, H. L. Push−Pull Pyropheophorbides for Nonlinear Optical Imaging. *Org. Biomol. Chem.* **2017**, *15* (4), 947−956.
(27) Dou, D.; Zhou, X.; Wang, T.; Yang, Q.; Tan, X.; Ling, Z.; Manz, M.; Liu, X.; Wetzelaer, G. J. A. H.; Li, X.; Baumgarten, M.; Blom, P. W. M.; Li, Y. Intramolecular Through-Space Charge-Transfer Effect for Achieving Room-Temperature Phosphorescence in Amorphous Film. *Adv. Opt. Mater.* **2024**, *12* (34), 2400976.
(28) Huang, Y.; Zhang, M.; Ye, L.; Guo, X.; Han, C. C.; Li, Y.; Hou, J. Molecular Energy Level Modulation by Changing the Position of Electron-Donating Side Groups. *J. Mater. Chem.* **2012**, *22* (12), 5700−5705.
(29) Alyahya, F.; Wazzan, N. Systematic Molecular Engineering of π-Spacer in Indoline-Based Dyes with D-A'-π-A Framework to Enhance the Intramolecular Charge Transfer and Photovoltaic Properties in DSSC and NLO Applications: DFT Insight. *Chem. Phys.* **2025**, *594*, 112666.
(30) You, D. K.; Kim, M.; Kim, D.; Kim, N.; Lee, K. M. Improvement in Radiative Efficiency Via Intramolecular Charge Transfer in Ortho-Carboranyl Luminophores Modified with Functionalized Biphenyls. *Inorg. Chem.* **2023**, *62* (25), 10003−10013.
(31) Cappello, D.; Buguis, F. L.; Gilroy, J. B. Tuning the Properties of Donor-Acceptor and Acceptor-Donor-Acceptor Boron Difluoride Hydrazones via Extended π-Conjugation. *ACS Omega* **2022**, *7* (36), 32727−32739.
(32) Terenziani, F.; Katan, C.; Badaeva, E.; Tretiak, S.; Blanchard-Desce, M. Enhanced Two-Photon Absorption of Organic Chromophores: Theoretical and Experimental Assessments. *Adv. Mater.* **2008**, *20* (24), 4641−4678.
(33) Pawlicki, M.; Collins, H. A.; Denning, R. G.; Anderson, H. L. Two-Photon Absorption and the Design of Two-Photon Dyes. *Angew. Chem., Int. Ed.* **2009**, *48* (18), 3244−3266.
(34) Albota, M.; Beljonne, D.; Brédas, J. L.; Ehrlich, J. E.; Fu, J. Y.; Heikal, A. A.; Hess, S. E.; Kogej, T.; Levin, M. D.; Marder, S. R.; McCord-Maughon, D.; Perry, J. W.; Röckel, H.; Rumi, M.; Subramaniam, G.; Webb, W. W.; Wu, X. L.; Xu, C. Design of Organic Molecules with Large Two-Photon Absorption Cross Sections. *Science (1979)* **1998**, *281* (5383), 1653−1656.
(35) Rumi, M.; Perry, J. W. Two-Photon Absorption: An Overview of Measurements And. *Adv. Opt. Photonics* **2010**, *2* (4), 451−518.
(36) Brown, S. Photodynamic Therapy: Two Photons Are Better than One. *Nat. Photonics* **2008**, *2* (7), 394−395.
(37) He, G. S.; Tan, L. S.; Zheng, Q.; Prasad, P. N. Multiphoton Absorbing Materials: Molecular Designs, Characterizations, and Applications. *Chem. Rev.* **2008**, *108* (4), 1245−1330.
(38) Butler, R. S.; Myers, A. K.; Bellarmine, P.; Abboud, K. A.; Castellano, R. K. Highly Fluorescent Donor−Acceptor Purines. *J. Mater. Chem.* **2007**, *17* (19), 1863−1865.
(39) Mckee, S.; Lutey, A.; Sciancalepore, C.; Poli, F.; Selleri, S.; Cucinotta, A. Microfabrication of Polymer Microneedle Arrays Using Two-Photon Polymerization. *J. Photochem. Photobiol., B* **2022**, *229* (9), 112424.
(40) Liu, R.; Qian, Y. Near-Infrared BODIPY Photosensitizers for Two-Photon Excited Singlet Oxygen Generation and Tumor Cell Photodynamic Therapy. *Org. Biomol. Chem.* **2024**, *22* (27), 5569−5577.
(41) Van Stryland, E. W.; Said, A. A.; Sheik-Bahae, M. High-Sensitivity, Single-Beam n_2 Measurements. *Opt. Lett.* **1989**, *14* (17), 955−957.
(42) Sheik-Bahae, M.; Said, A. A.; Wei, T. H.; Hagan, D. J.; Van Stryland, E. W. Sensitive Measurement of Optical Nonlinearities Using a Single Beam. *IEEE J. Quantum Electron.* **1990**, *26* (4), 760−769.
(43) Ian, T. J. Principal Component Analysis; In *Springer Series in Statistics*; Springer-Verlag: New York, 2002; Vol. *2*, p 488. .
(44) Marbello, O. D.; De Boni, L.; Mendonça, C. R. Principal Component Analysis for Imidazo[4,5-b]Pyridine Linear and Nonlinear Optical Properties Description. *Opt. Mater. (Amst.)* **2025**, *159*, 116620.
(45) Yun, H.-S.; Lee, J.-E. The Principal Component Analysis Filtering Method for an Unbiased Spectral Survey of Complex Organic Molecules. *Astrophys. J.* **2023**, *958* (2), 113.
(46) Sherman, A. M.; Takanti, N.; Rong, J.; Simpson, G. J. Nonlinear Optical Characterization of Pharmaceutical Formulations. *TrAC, Trends Anal. Chem.* **2021**, *140* (4), 116241.
(47) Pelosi, A. G.; Zucolotto Cocca, L. H.; Piguel, S.; De Boni, L.; Mendonça, C. R. Effect of Peripherical Groups on the Two-Photon Brightness of (E)-3-Benzyl-6-Bromo-2-Styryl-3H-Imidazo[4,5-b]-Pyridine Derivatives. *J. Mol. Liq.* **2022**, *358*, 119186.
(48) Mikhaylov, A.; Lindquist, J. R.; Callis, P. R.; Kohler, B.; Pahapill, J.; de Reguardati, S.; Rammo, M.; Uudsemaa, M.; Trummal, A.; Rebane, A. Femtosecond Two-Photon Absorption Spectra and Permanent Electric Dipole Moment Change of Tryptophan, 2-Aminopurine and Related Intrinsic and Synthetic. Multiphoton Microscopy in the Biomedical Sciences XVII. *2017* **2017**, *10069*, 1006920.
(49) Canakdag, M.; Feizi-Dehnayebi, M.; Kundu, S.; Sahin, D.; İlhan, I. ·. O. ··; Alhag, S. K.; Al-Shuraym, L. A.; Akkoc, S. Comprehensive Evaluation of Purine Analogues: Cytotoxic and

https://doi.org/10.1021/acs.jpca.6c02028
J. Phys. Chem. A 2026, 130, 5205−5214

Antioxidant Activities, Enzyme Inhibition, DFT Insights, and Molecular Docking Analysis. *J. Mol. Struct.* **2025**, *1323*, 140798.
(50) Leškovskis, K.; Zaķis, J. M.; Novosjolova, I.; Turks, M. Applications of Purine Ring Opening in the Synthesis of Imidazole, Pyrimidine, and New Purine Derivatives. *Eur. J. Org. Chem.* **2021**, *2021* (36), 5027−5052.
(51) Zhang, C.; Mo, S.; Liu, Z.; Chen, B.; Korshin, G.; Hertkorn, N.; Ni, J.; Yan, M. Interpreting PH-Dependent Differential UV/VIS Absorbance Spectra to Characterize Carboxylic and Phenolic Chromophores in Natural Organic Matter. *Water Res.* **2023**, *244*, 120522.
(52) Guo, X.; Jia, X.; He, Q.; Duan, W.; Zhang, Y.; Huang, Y.; Liu, L. Supramolecular Double-Helical Polymers: Supramolecular Chiral Induction and Asymmetric Catalysis. *Molecules* **2025**, *30* (7), 1517.
(53) Collier, G. S.; Brown, L. A.; Boone, E. S.; Kaushal, M.; Ericson, M. N.; Walter, M. G.; Long, B. K.; Kilbey, S. M. Linking Design and Properties of Purine-Based Donor−Acceptor Chromophores as Optoelectronic Materials. *J. Mater. Chem. C* **2017**, *5* (27), 6891−6898.
(54) Linking design and properties of purine-based donor−acceptor chromophores as optoelectronic materials − Journal of Materials Chemistry C, RSC Publishing. https://pubs.rsc.org/en/content/articlehtml/2017/tc/c7tc01835e (accessed Nov 08, 2025).
(55) Lai, H.; Niu, Y.; Ming, P.; Wang, J.; Wang, S.; Xu, Y.; Zhai, H. An Electrochemical Sensor for Purine Base Detection with ZIF-8-Derived Hollow N-Doped Carbon Dodecahedron and AuNPs as Electrocatalysts. *Anal. Bioanal.Chem.* **2023**, *415* (5), 855−865.
(56) Hanson, C.; Sieverts, M.; Vargis, E. Effect of Principal Component Analysis Centering and Scaling on Classification of Mycobacteria from Raman Spectra. *Appl. Spectrosc.* **2017**, *71* (6), 1249−1255.
(57) Beattie, J. R.; Esmonde-White, F. W. L. Exploration of Principal Component Analysis: Deriving Principal Component Analysis Visually Using Spectra. *Appl. Spectrosc.* **2021**, *75* (4), 361−375.
(58) Mohd Napi, N. N. L.; Noor Mohamed, M. S.; Abdullah, S.; Mansor, A. A.; Ahmed, A. N.; Ismail, M. Multiple Linear Regression (MLR) and Principal Component Regression (PCR) for Ozone (O3) Concentrations Prediction. *IOP Conf. Ser.: Earth Environ. Sci.* **2020**, *616* (1), 012004.
(59) Lubis, A. H.; Sihombing, P.; Nababan, E. B. Analysis of Accuracy Improvement in K-Nearest Neighbor Using Principal Component Analysis (PCA). *J. Phys. Conf. Ser.* **2020**, *1566* (1), 012062.
(60) Poslavskaya, E. Encoding Categorical Data: Is There yet Anything'hotter'than One-Hot Encoding?. arXiv:2312.16930, A. K. preprint. arxiv.orgE Poslavskaya, A KorolevarXiv preprint arXiv:2312.16930, arxiv.org **2023**.
(61) Abu-Rayash, A.; Dincer, I. Development and Application of an Integrated Smart City Model. *Heliyon* **2023**, *9* (4), No. e14347.
(62) Wang, K.; Tao, W.; Nagarajan, K.; Kushida, S.; Sandeep, K.; Genet, C.; Ebbesen, T. W. Study of the Selection Rules of Molecular Polaritonic Transitions by Two-Photon Absorption Spectroscopy. *J. Phys. Chem. C* **2024**, *128* (25), 10534−10541.
(63) Makarov, N. S.; Drobizhev, M.; Wicks, G.; Makarova, E. A.; Lukyanets, E. A.; Rebane, A. Alternative Selection Rules for One- and Two-Photon Transitions in Tribenzotetraazachlorin: Quasi-Centrosymmetrical π-Conjugation Pathway of Formally Non-Centrosymmetrical Molecule. *J. Chem. Phys.* **2013**, *138* (21), 214314.
(64) Li, Y.; Liu, T.; Liu, H.; Tian, M. Z.; Li, Y. Self-Assembly of Intramolecular Charge-Transfer Compounds into Functional Molecular Systems. *Acc. Chem. Res.* **2014**, *47* (4), 1186−1198.
(65) Shinn, M. Phantom Oscillations in Principal Component Analysis. *Proc. Natl. Acad. Sci. U.S.A.* **2023**, *120* (48), No. e2311420120.
(66) Yang, G.; Jin, J.; Mei, J.; Lin, Z.; Shi, J.; Lv, H. Insights into Anthocyanin Stability and Infusion Color in Purple Leaf Teas Driven by Co-Pigment Flavonoid Interactions. *Food Chem.* **2025**, *496* (17), 146711.
(67) Zhang, Q.; Tian, X.; Zhou, H.; Wu, J.; Tian, Y. Lighting the Way to See Inside Two-Photon Absorption Materials: Structure−Property Relationship and Biological Imaging. *Materials* **2017**, *10* (3), 223.
(68) Pandey, N.; Tewari, N.; Pant, S.; Mehata, M. S. Solvatochromism and Estimation of Ground and Excited State Dipole Moments of 6-Aminoquinoline. *Spectrochim. Acta, Part A* **2022**, *267* (15), 120498.
(69) Francis, J. O.; Michael, A. T.; Arawo, C. C. Multiple Linear Regression (MLR) Model: A Tool for Water Quality Interpretation. *Momona Ethiop. J. Sci.* **2020**, *12* (1), 123−134.

https://doi.org/10.1021/acs.jpca.6c02028
J. Phys. Chem. A 2026, 130, 5205−5214