Supplementary Information (SI) for Digital Discovery.

SUPPLEMENTARY INFORMATION

A reproducible Python workflow for absorber–light-source spectral matching: overlap-calculator

Pinar Seyitdanlioglu*[a]

Department of Chemistry, Faculty of Science, Hacettepe University, Ankara, Turkiye

S1. Scope of the Supplementary Information

The main manuscript describes the scientific motivation, workflow concept, descriptor definitions, and demonstration results. This document records the technical details required to reproduce and audit the reported analysis:

- the exact input manifest used for the demonstration run;
- runtime parameters and default light-source selections;
- the treatment of theoretical and spreadsheet-based spectra;
- additional notes on broadening, absorptance conversion, and overlap metrics;
- output-file inventory and ranking-table interpretation;
- command-line, API, Docker, and test commands used to verify the workflow.

The supplementary material is intended to function as a practical reproducibility record rather than a replacement for the full user manual distributed with the software.

Table S1 Functional comparison of overlap-calculator with commonly used molecular post-processing or scripting approaches for spectral analysis.

Feature	Multiwfn[1] GUI/script workflow	ORCA[2]-related tools	ASE[3]/custom Python scripts	overlap-calculator
TD-DFT transition parsing	Available through GUI or scripted analysis	Native for ORCA-type outputs	Requires manual/custom parsing	Native batch parsing of Gaussian TD-DFT outputs
Light-source overlap analysis	Not built in	Not built in	Requires manual implementation	Built-in spectral-overlap descriptors
Indoor LED/fluorescent light-source library	Not built in	Not built in	Not built in	Built-in CIE LED, CIE fluorescent, and AM1.5G support
CSV/XLSX experimental UV-vis input	Not designed as a direct batch input branch	Not built in	Requires custom formatting and scripts	Built-in CSV/XLSX spreadsheet input support
Batch manifest-based execution	Not native	Limited or workflow-specific	Manual/custom setup	Native input.json manifest workflow
Reproducible command-line workflow	Partly possible through scripting	Tool-dependent	Manual scripting required	Built-in CLI workflow
HTTP API / service deployment	Not available	Not available	Not available unless custom-developed	Flask API and Docker-supported deployment
Structured output tables and plots	Requires user-defined export steps	Tool-dependent	Manual/custom export	Built-in CSV, JSON, XLSX, timing, skip-report, and TIFF plot outputs

S2. Demonstration Input Manifest

The manuscript demonstration used a mixed manifest containing five Gaussian TD-DFT output files and eight spreadsheet-based UV-vis spectra. The theoretical inputs were parsed from Gaussian output files.[4] The spreadsheet spectra were read from an Excel workbook in which each named molecule is treated as an independent spectral series.

Table S2. Input entries used in the manuscript demonstration.

Input family	Sample identifier	Source file	Sheet name	Series name
Theoretical	3e_td	files/slurm-2829207.out		
Theoretical	3f_td	files/slurm-2829212.out		
Theoretical	3g_td	files/slurm-2829256.out		
Theoretical	4c_td	files/slurm-2829260.out		
Theoretical	4g_td	files/slurm-2829265.out		
Spreadsheet	Coumarin_1	files/organic_uvvis_photochemcad_dataset.xlsx	program_input_absorbance	Coumarin_1
Spreadsheet	Coumarin_6	files/organic_uvvis_photochemcad_dataset.xlsx	program_input_absorbance	Coumarin_6
Spreadsheet	Coumarin_314	files/organic_uvvis_photochemcad_dataset.xlsx	program_input_absorbance	Coumarin_314
Spreadsheet	Fluorescein	files/organic_uvvis_photochemcad_dataset.xlsx	program_input_absorbance	Fluorescein
Spreadsheet	Rhodamine_B	files/organic_uvvis_photochemcad_dataset.xlsx	program_input_absorbance	Rhodamine_B
Spreadsheet	Rhodamine_6G	files/organic_uvvis_photochemcad_dataset.xlsx	program_input_absorbance	Rhodamine_6G
Spreadsheet	Eosin_Y	files/organic_uvvis_photochemcad_dataset.xlsx	program_input_absorbance	Eosin_Y

Spreadsheet	Pyrene	files/organic_uvvis_photochemcad_dataset.xlsx	program_input_absorbance	Pyrene

The corresponding JSON manifest is distributed as input/input.json.

S3. Runtime Parameters

All calculations reported in the manuscript were generated using the default wavelength grid and Beer-Lambert reference settings unless otherwise stated.

Table S3. Runtime settings used for the reported demonstration output.

Parameter	Value	Role
Wavelength minimum	200 nm	Lower integration and plotting bound
Wavelength maximum	800 nm	Upper integration and plotting bound
Grid points	10000	Shared interpolation and integration grid
Broadening width	0.30 eV	Gaussian and Lorentzian broadening parameter for TD-DFT transitions
Reference concentration	1e-5 M	Beer-Lambert concentration for TD-DFT-derived spectra
Reference path length	1.0 cm	Beer-Lambert optical path length
Default light sources	AM15G,LEDB4,LEDB2,LEDB3,CIEFL10	Built-in solar, LED, and fluorescent spectra
Plot resolution	400 dpi	Default TIFF export resolution
Ranking outputs	enabled	Exports grouped ranking tables and plots
Prefactor mode	constant	integrated-intensity convention; frequency-resolved keeps the ν̃ factor inside the integrand (NIR).
Sigma mode	fixed	fixed uses --sigma-ev; marcus-hush sets σ=√(2λk_BT) from --reorganization-ev and --temperature-k.
Calibration TOML	(none)	optional --calibration applying E_cal=a*E+b, f_cal=alpha*f, per-band overrides; identity is bit-identical.

The principal command-line run is:

```
overlap-calculator analyze \
  --input input/input.json \
  --out output \
  --sigma-ev 0.30 \
  --wl-min 200 \
  --wl-max 800 \
  --num-points 10000 \
  --concentration-m 1e-5 \
  --path-cm 1.0 \
  --default-light-sources AM15G,LEDB2,LEDB3,LEDB4,CIEFL10
```

S4. Treatment of Theoretical and Spreadsheet Spectra

For theoretical inputs, Gaussian TD-DFT excited-state wavelengths and oscillator strengths are parsed from .out or .log files. Each transition is reconstructed on the common wavelength grid using both Gaussian and Lorentzian line-shape functions. The two broadened spectra are retained separately throughout descriptor calculation and export.

The five Gaussian TD-DFT output files were generated in a previous OPV screening study by the author at the B3LYP/6-31G level of theory and are reused here only as representative calculated-spectra inputs for validating the overlap-calculator workflow.[5] No new quantum-chemical calculations are claimed for these absorbers in the present work.

For spreadsheet inputs, the numeric spectral series is treated as an absorbance-like signal with a wavelength axis. Each numeric series column can be analysed as a distinct sample. The demonstration spreadsheet was prepared from public organic UV-vis molar absorptivity spectra and transformed into absorbance-like profiles before ingestion. These spreadsheet spectra are used to demonstrate the tabular-input branch of the workflow and are not used as an experimental validation set for the TD-DFT candidates.

Table S4. Input branch behaviour.

Input branch	**Accepted files**	**Internal spectral quantity before absorptance**	**Broadening outputs**
Theoretical	Gaussian .out / .log	Molar extinction coefficient reconstructed from transitions	Gaussian and Lorentzian
Spreadsheet	.csv, .xlsx, .xls	User-provided absorbance-like signal	Direct absorbance branch; exported in the common table structure

S5. Spectral Reconstruction and Absorptance

For a TD-DFT transition i with excitation wavelength λ_i and oscillator strength f_i, the wavelength axis was first converted to a wavenumber scale according to

$$\tilde{\nu}(\lambda) = \frac{10^7}{\lambda}$$

where λ is given in nm and $\tilde{\nu}$ is obtained in cm⁻¹. The transition centre is therefore

$$\tilde{\nu}_i = \frac{10^7}{\lambda_i}.$$

Each transition was represented on the common wavelength grid using either a Gaussian or Lorentzian line-shape function. The Gaussian profile was defined as

$$g_i(\tilde{\nu}) = \frac{1}{\sigma\sqrt{2\pi}} exp\left[- \frac{(\tilde{\nu} - \tilde{\nu}_i)^2}{2\sigma^2} \right],$$

whereas the Lorentzian profile was defined as

$$l(\tilde{\nu}) = \frac{1}{\pi} \frac{\gamma}{(\tilde{\nu} - \tilde{\nu}_i)^2 + \gamma^2}.$$

$$\gamma = 2ln2\sigma \approx 1.1774\sigma$$

The reconstructed molar absorptivity spectrum was then obtained as the oscillator-strength-weighted sum of the broadened transitions:

$$\varepsilon(\lambda) = 2.315 \times 10^8 \sum_i f_i P_i(\tilde{\nu}(\lambda)),$$

where P_i denotes either G_i or L_i, depending on the selected broadening model. The resulting $\varepsilon(\lambda)$ values are expressed in M⁻¹ cm⁻¹.

For TD-DFT-derived spectra, absorbance was calculated using the Beer–Lambert relation:

$$A_{\text{TD - DFT}}(\lambda) = \varepsilon(\lambda)cL,$$

where c is the reference concentration and L is the optical path length. The corresponding wavelength-dependent absorptance was calculated as

$$\alpha_{\text{TD - DFT}}(\lambda) = 1 - 10^{-A_{\text{TD - DFT}}(\lambda)} = 1 - 10^{-\varepsilon(\lambda)cL}.$$

For spreadsheet-based spectra, the user-provided absorbance-like signal $A_{\text{user}}(\lambda)$ was inserted directly into the same absorptance definition:

$$\alpha_{\text{spreadsheet}}(\lambda) = 1 - 10^{-A_{\text{user}}(\lambda)}.$$

S6. Light Sources and Units

The demonstration uses one absolute solar spectrum and four relative indoor illuminant spectra.[6-10]

Table S5. Light-source definitions used in the demonstration.

Light source	Source type	Output unit label
AM15G	ASTM G-173 AM1.5 global reference spectrum	W m^-2 nm^-1
LEDB2	CIE LED illuminant data	relative
LEDB3	CIE LED illuminant data	relative
LEDB4	CIE LED illuminant data	relative
CIEFL10	CIE fluorescent illuminant data	relative

For AM1.5G, absorbed_flux is expressed as an integrated irradiance-like quantity over the selected wavelength grid. For relative CIE light-source spectra, the absolute scale is arbitrary, so absorbed_fraction and shape_overlap are generally more appropriate for ranking samples under the same light source.

S7. Overlap Descriptors

For each sample and light source, the workflow calculates:

```
absorbed_flux = integral alpha(lambda) I(lambda) d(lambda)
light_flux_total = integral I(lambda) d(lambda)
absorbed_fraction = absorbed_flux / light_flux_total
shape_overlap = integral min(alpha_hat(lambda), I_hat(lambda)) d(lambda)
                / integral I_hat(lambda) d(lambda)
```

Here I(lambda) is the light-source spectrum, and hats indicate max-normalised profiles. absorbed_fraction is intensity-weighted and bounded between 0 and 1 for non-negative inputs. shape_overlap is a dimensionless spectral matching descriptor that emphasises relative alignment between absorptance and the light-source emission shape. However, shape_overlap can approach unity when the normalised absorption envelope completely covers the normalised light-source envelope, that is, when $\hat{\alpha}(\lambda) \geq \hat{I}(\lambda)$ over the wavelength region where $\hat{I}(\lambda) > 0$. For very strongly absorbing molecules, this behaviour may compress the metric near unity. In such cases, absorbed_fraction can provide additional discrimination because it explicitly retains the Beer–Lambert absorption strength. For this reason, the two metrics are reported together as complementary descriptors rather than as interchangeable alternatives.

S8. Output Inventory

The manuscript demonstration generated 65 result rows: 25 theoretical rows and 40 spreadsheet-input rows. Each of the 13 samples was evaluated under five light sources.

Table S6. Main output files generated under output/tables/.

Output file family	Formats	Description
results	CSV, JSON, XLSX	Main row-wise descriptor table for each sample-light pair
descriptor_summary	CSV, XLSX	Condensed ranked summary table
results_timings	CSV, JSON, XLSX	Runtime timing information for parsing, broadening, light-source handling, descriptor calculation, and plotting
ranking_by_light_source__<metric>	CSV, JSON, XLSX	Per-light-source sample ranking for each metric
ranking_by_sample__<metric>	CSV, JSON, XLSX	Per-sample light-source ranking for each metric
`run_manifest.json`	`JSON`	Per-run provenance manifest at the output root: software_version, git_commit, generated_at_utc (ISO-8601 UTC), the full resolved parameter set including the calibration block, the light-source

		set, one entry per (source, sample, series, sheet) with the SHA-256 of each input file, results_count, skipped_count.

Every run also writes a single run_manifest.json in the output root; result rows additionally carry prefactor_mode, sigma_mode, and software_version columns.

The ranking metric families are:

- gaussian_absorbed_fraction
- lorentzian_absorbed_fraction
- gaussian_shape_overlap
- lorentzian_shape_overlap

The run also generated 497 TIFF plots under output/plots/, including absorption plots, light-source plots, overlap plots, overlays, and grouped ranking bar charts.

S9. Additional Ranking Summary

The main text discusses source-dependent changes in candidate ranking. Table S7 gives the top three samples under each light source using Gaussian shape_overlap.

Table S7. Top-ranked samples by Gaussian shape overlap.

Light source	Rank 1	Rank 2	Rank 3
AM15G	4c_td (0.7952)	3e_td (0.7173)	3g_td (0.6437)
CIEFL10	4c_td (0.9852)	3g_td (0.9724)	3e_td (0.9664)
LEDB2	4c_td (0.9503)	3e_td (0.9439)	3f_td (0.9185)
LEDB3	3e_td (0.9514)	4c_td (0.9492)	3g_td (0.9333)
LEDB4	4c_td (0.9643)	3g_td (0.9536)	3e_td (0.9374)

These rankings illustrate why the manuscript avoids assigning a single universal optical ranking to all absorbers. Instead, the preferred candidate depends on the intended illumination environment and the selected descriptor family.

S10. Reproducibility Commands

The input manifest can be generated from the raw input directory:

```
overlap-calculator generate-input \
  --files-dir input/files \
  --out input/input.json
```

The full analysis can then be rerun from the manifest:

```
overlap-calculator analyze \
  --input input/input.json \
  --out output
```

To reproduce a table-only run without plot generation:

```
overlap-calculator analyze \
  --input input/input.json \
  --out output_table_only \
  --no-plot-outputs
```

To change only the plot resolution:

```
overlap-calculator analyze \
  --input input/input.json \
  --out output_600dpi \
  --plot-dpi 600
```

S11. HTTP API Reproduction

The same analysis logic is exposed through a Flask HTTP API. A minimal API health check is:

```
curl http://localhost:8000/health
```

A multipart analysis request can be sent as:

```
curl -X POST http://localhost:8000/analyze \
  -F "files=@input/files/slurm-2829207.out" \
  -F "files=@input/files/organic_uvvis_photochemcad_dataset.xlsx" \
  -F "sigma_ev=0.30" \
  -F "default_light_sources=AM15G,LEDB4" \
  --output analysis_outputs.zip
```

The response archive contains the same table and plot hierarchy as the command-line workflow.

S12. Software Environment

The package is distributed as overlap-calculator version 1.1.1 and requires Python >=3.12,<3.13.The principal runtime dependencies are pinned in pyproject.toml.

Table S8. Main runtime dependencies.

Package	Version
numpy	2.1.1
pandas	2.2.2
pydantic	2.8.2
pydantic-settings	2.4.0
matplotlib	3.9.2
openpyxl	3.1.5
flask	3.0.3
typer	0.12.3
click	8.1.7

The recommended local installation is:

```
conda env create -f environment.yml
conda activate overlap-calculator
pip install -e .
```

The Docker build route is:

```
docker build -t overlap-calculator .
docker run --rm -p 8080:8000 overlap-calculator
```

S13. Test and Case-Study Coverage

The repository contains unit tests covering TD-DFT parsing, spectrum metrics, light-source loading, input generation, CLI behaviour, API helpers, export structure, and batch skip handling. Additional tests in `test_broadening_calibration_provenance` cover the revised broadening modes, calibration block, prefactor-mode handling, and run-manifest provenance fields. The test suite can be run with:

python -m pytestThe repository also includes worked case studies under case_studies/:

```
01_theoretical_only
02_experimental_only
03_mixed_theory_experiment
04_custom_light_source
05_default_light_source_selection
06_plot_dpi
07_table_only
08_broadening_sigma
09_beer_lambert_tuning
10_wavelength_grid
11_sheet_overrides

12_frequency_resolved_prefactor

13_marcus_hush_width

14_calibration_block

15_vibronic_tabular_branch
```

These examples are intended to document the expected behaviour of individual workflow branches and user-facing options.

S14. Notes on Interpretation

The reported descriptors are optical compatibility descriptors. They should not be interpreted as device efficiency, external quantum efficiency, short-circuit current density, power-conversion efficiency, or long-term stability. They also do not include exciton diffusion, charge generation, morphology, optical interference in multilayer stacks, electrode losses, or recombination.

The purpose of the workflow is to make a recurring spectral-overlap post-processing step explicit, reproducible, and comparable across molecule sets and light sources. The descriptors are therefore most useful for early-stage screening, prioritisation, and transparent comparison of optical compatibility under a defined illumination condition.

Supplementary References

1. T. Lu and F. Chen, *J. Comput. Chem.*, 2012, **33**, 580–592.

2. F. Neese, F. Wennmohs, U. Becker and C. Riplinger, *J. Chem. Phys.*, 2020, **152**, 224108.

3. A. H. Larsen, J. J. Mortensen, J. Blomqvist, I. E. Castelli, R. Christensen, M. Dułak *et al.*, *J. Phys.: Condens. Matter*, 2017, **29**, 273002.

4. M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman and D. J. Fox, Gaussian 16, Revision C.02, Gaussian, Inc., Wallingford CT, 2019.

5. P. Seyitdanlioglu, *J. Mol. Model.***,** 2026, **32**, 8.

6. M. Taniguchi and J. S. Lindsey, Database of absorption and fluorescence spectra of >300 common compounds for use in PhotochemCAD, Photochem. Photobiol., 2018, 94, 290-327.

7. International Commission on Illumination, Relative spectral power distributions of illuminants representing typical LED lamps, 1 nm spacing, CIE Central Bureau, Vienna, 2018, DOI: 10.25039/CIE.DS.dhcw57sd.

8. International Commission on Illumination, Relative spectral power distributions of illuminants representing typical fluorescent lamps, 1 nm wavelength steps, CIE Central Bureau, Vienna, 2018, DOI: 10.25039/CIE.DS.54hy6srn.

9. International Commission on Illumination, CIE 015:2018 Colorimetry, 4th edn, CIE Central Bureau, Vienna, 2018.

10. National Renewable Energy Laboratory, Reference Air Mass 1.5 Spectra (ASTM G-173-03), U.S. Department of Energy.