PigmentHunter: A point-and-click application for automated chlorophyll-protein simulations

S. Ahad,[1] C. Lin,[1] and M. Reppert[1, a)]

Department of Chemistry, Purdue University, West Lafayette, Indiana 47907, United States of America

(*Electronic mail: reppertm@purdue.edu)

Chlorophyll proteins (CPs) are the work-horses of biological photosynthesis, working together to absorb solar energy, transfer it to chemically active reaction centers, and control the charge-separation process that drives its storage as chemical energy. Yet predicting CP optical and electronic properties remains a serious challenge, driven by the computational difficulty of treating large, electronically coupled molecular pigments embedded in a dynamically structured protein environment. To address this challenge, we here introduce an analysis tool called PigmentHunter, which automates the process of preparing CP structures for molecular dynamics (MD), running short MD simulations on the nanoHUB.org science gateway, and then using electrostatic and steric analysis routines to predict optical absorption, fluorescence, and circular dichroism spectra within a Frenkel exciton model. Inter-pigment couplings are evaluated using point-dipole or transition-charge coupling models, while site energies can be estimated using both electrostatic and ring-deformation approaches. The package is built in a Jupyter Notebook environment, with a point-and-click interface that can be used either to manually prepare individual structures or to batch-process many structures at once. We illustrate PigmentHunter's capabilities with example simulations on spectral line shapes in the light harvesting 2 (LH2) complex, site energies in the Fenna-Matthews-Olson protein, and ring deformation in photosystems I and II.

[a)] https://mreppert.github.io/

I. INTRODUCTION

Biological photosynthesis, the process of converting light into usable chemical energy, is driven by pigment-protein complexes (PPCs) that capture sunlight and use it to induce chemical charge separation. PPCs provide a fascinating model for understanding light conversion at the molecular level and for designing synthetic light-harvesting systems.[1,2] However, the recognition that natural systems are optimized for competitive fitness rather than efficiency has also inspired a growing field of research into the use of targeted mutagenesis to "reoptimize" biological photosynthesis for agrofuel production.[3–12]

Unfortunately, efforts to both characterize and modify native photosynthetic function are limited by gaps in our understanding of the "structure-spectrum-function" relationship in photosynthetic PPCs, i.e., the connection between PPC structures, their optical spectra, and their corresponding biological function. Although the last several decades have seen immense progress in our ability to characterize both the atomistic structures and spectroscopic dynamics of light-harvesting PPCs, drawing detailed connections *between* structure and spectrum – e.g., assigning excitonic states to individual pigment clusters – remains both difficult and controversial.[3,13–16] While the atomistic detail provided by X-ray and cryogenic electron microscopy (CryoEM) structures provides invaluable insight into PPC organization and function, the structure alone gives little information on the expected optical properties or specific functional role played by most pigments.

To fill this gap, a variety of computational methods have been developed to "translate" PPC structural models into optical spectra.[16,19–22] The most complete of these treatments is offered by mixed quantum mechanics/molecular mechanics (QM/MM) methods, which describe the electronic properties of individual pigments using quantum-mechanical tools such as density functional theory (DFT) and treat the surrounding protein classically, e.g., using molecular dynamics (MD).[14,23–32] While these methods have shown considerable promise in calculating PPC optical spectra, they remain computationally expensive, and quantitative accuracy has only been shown in a handful of cases.[14,23,24,33] To lower computational cost, a variety of simpler methods have been devised that use electrostatic interaction energies or pigment deformation metrics to predict electronic frequency shifts.[20,21] Electrostatic methods in particular have shown promise for accurate frequency-shift predictions, but even here quantitative benchmarking against experiment is limited.[34]

Perhaps the most practical obstacle to the widespread application of these methods is the tech-

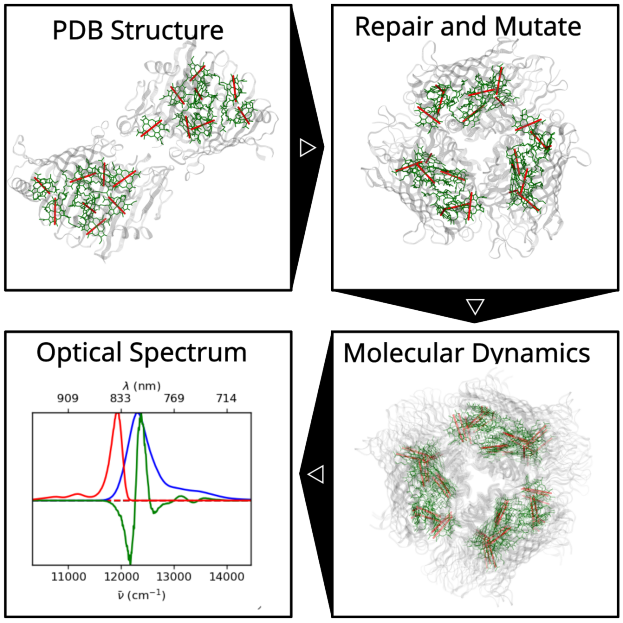

FIG. 1. PigmentHunter simplifies the process of predicting optical spectra from the Chl-protein structures. This is demonstrated schematically here for the 3ENI[17,18] PDB structure.

nical challenge of simply preparing the protein structure for calculation. While an electrostatic calculation itself may take only seconds, it can easily take hours or days to go from a raw protein data bank (PDB) structure to a chemically sensible model which is ready for simulation. Due to the constraints of finite experimental resolution, PDB structures usually do not include hydrogen atom locations and very often contain "gaps" in the form of missing atoms or omitted residues. If left unfixed (or, worse, unnoticed), such omissions can derail a calculation or produce dramatic inaccuracies in its output. Moreover, even if the input structure is chemically complete, running even a simple electrostatic calculation requires that pigments first be correctly identified and that accurate atomic charges be assigned to each atom. DFT and MD calculations require even more detailed preparation.[14] Each of these tasks is made more difficult by the fact that non-protein ligands like Chlorophyll (Chl) are not parameterized in standard MD force fields and are often represented under different naming conventions in different PDB structures.

To bridge this gap, we describe here the development of the PigmentHunter application, a 'point-and-click' simulation and analysis tool that automates the process of preparing PPC structures for MD simulation, running simple MD trajectories, analyzing electrostatic and ring-deformation statistics, and performing structure-based PPC excitonic simulations (see Figure 1).[35] PigmentHunter uses a Jupyter Notebook interface hosted on nanoHUB.org, a science gateway that allows for the creation of graphical interfaces and simulation tools that run remotely rather than

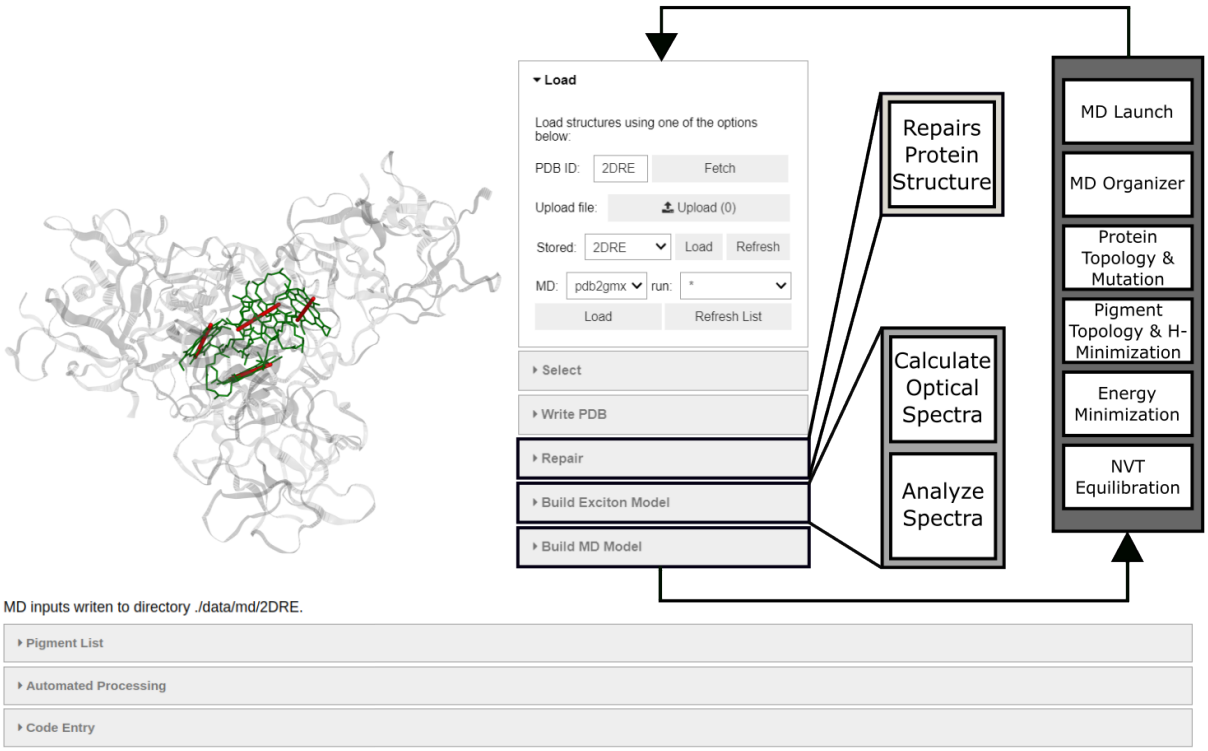

FIG. 2. Schematic overview of the PigmentHunter interface on nanoHUB.org[35]. The main PigmentHunter page can be used to calculate inter-pigment couplings and site energies and to prepare and analyze MD simulations. Call-out boxes indicate separate windows that can be launched for specific functions

locally.[36]

This paper is organized as follows: Section II describes the basic PigmentHunter workflow and usage, serving as an introductory "how-to" guide for PPC MD, electrostatic, and excitonic calculations. Section III describes in greater detail the underlying theoretical and computational methods upon which PigmentHunter depends. Section IV illustrates the application of PigmentHunter to several PPC structures, and Section V concludes with a discussion of PigmentHunter's current capabilities and an outlook on future developments.

II. IMPLEMENTATION

A. Overview

The PigmentHunter package is published and hosted as an online tool at nanoHUB.org, an online science gateway that allows users to execute scientific simulations without any need for local download, installation, or execution of code. (Source code for installing PigmentHunter locally

is also available on github at `https://github.com/mreppert/PigmentHunter`.) As a public science gateway (based on the HUBZero platform[37]), nanoHUB is free to use and requires no institutional affiliation to create an account. Once a nanoHUB account is created, PigmentHunter can be accessed at the url: `https://nanohub.org/tools/pigmenthunter`. nanoHUB provides all computational infrastructure necessary to run PigmentHunter calculations, including access to a high-performance compute cluster for short MD simulations.[36]

Broadly speaking, PigmentHunter handles four primary tasks:

1. Identify, classify, and visualize Chl-like pigments in PPC structures.

2. Repair errors (e.g., missing atoms) in pigment structures and build appropriate biosymmetric units.

3. Prepare and run PPC energy-minimization and MD simulations.

4. Calculate optical spectra from PPC structures and MD trajectories using the Frenkel exciton model.

Although some overlap exists between these functions, broadly speaking, each of these four tasks is accomplished through a separate window (or sequence of windows) launched from the PigmentHunter Main page. Figure 2 gives a graphical "map" of the PigmentHunter application. All simulations start on the Main page, which hosts an interactive structure-viewing window (powered by NGLView[38]) and a series of IPyWidgets[39] dialog menus, which launch separate windows for structure repair and biosymmetrization (the Repair window), MD simulations (the MD Launch sequence), or spectroscopic calculations (the Spectrum window). Details on each of these interfaces and the functions they perform are provided in the following sections.

B. Main Window

Perhaps the core technical challenge addressed by PigmentHunter (as opposed to other structure analysis tools) is the automation of the process of correctly recognizing chlorophyll-like pigments, fixing any structural anomalies, and assigning accurate molecular topologies. The entry point for this process is PigmentHunter's main window, where PPC structures can be either uploaded from the user's local computer or fetched directly from the Protein Databank (`https://www.rcsb.org/`). By default, Protein Databank structures are downloaded in *.cif format; users can upload

files in *.cif, *.pdb, or the GROMACS[40] *.gro format. MD simulations that have previously been run within the PigmentHunter app can be loaded from stored *.gro files using a dropdown menu that lists stored simulations. On loading a protein structure, PigmentHunter stores it internally as a ParMed structure[41] and searches it to identify potential Chl-like pigments.

Pigment	Code	μ_{eff}	PDC	TrEsp	NSD	MD	CDC
Chl *a*	CLA	4.3 D	Y	Y	Y	Y	Y
Pheo *a*	PHA	3.5 D	Y	Y	Y*	N	N‡
Chl *b*	CLB	3.6 D	Y	Y	Y	N	N‡
BChl *a*	BCA	5.45 D	Y	Y	Y†	Y	Y

TABLE I. Implementation information for four common pigments recognized by the PigmentHunter application. The "Code" entry indicates the unique three-letter code used by PigmentHunter to denote specific pigment types. The μ_{eff} column indicates the effective dipole moment magnitude used in electronic coupling calculations. The remaining columns indicate whether various computational methods are implemented ("Y") or not ("N") for each pigment. Notes: The $*$ marks that PHA NSD calculations use parameters originally developed for CLA. The † denotes that only NSD coefficients (not NSD site energies) can be calculated for the BCA pigment. The ‡ for PHA and CLB indicates that CDC parameters are installed but that these simulations cannot currently be run since MD parameters (for generating partial charges) are not yet implemented for these pigments.

Pigment identification is accomplished by searching for specific atom names according to the standard naming convention for Chl-like pigments in the PDB. Specifically, a residue is recognized as a Chl-like pigment if it contains *all* the atoms NX, C1X, C2X, C3X, C4X, C5X, and CHX, where X takes on each of the values A, B, C, and D. These 24 atoms define the four primary rings (traditionally labeled as rings A, B, C, and D) of the Chl macrocycle and are regarded in PigmentHunter as the minimal requirements for a residue to be recognized as "chlorophyll-like". Implementation information for the four most common pigments recognized by PigmentHunter is summarized in Table I.

The specific identify of the pigment (e.g., Chl *a* versus Chl *b* or BChl *a*) is assigned based on the other atoms contained in the residue. Note that in many cases pigment identities cannot be unambiguously assigned based on a PDB file alone, since many PDB structures are missing atoms due to the limited resolution of the experimental data. E.g., Chl *a* and Chl *b* differ by only the presence of a single heavy atom, the "OMC" atom of the Chl *b* formyl group. When such ambiguities arise, PigmentHunter makes a "best guess" as to the identity, but the user is free to modify this assignment using a dropdown menu on the "Pigment List" tab. However, PigmentHunter will *not* allow users to assign pigment identities that are inconsistent with the atoms actually present in the PDB structure. For example, a residue that contains an "OMC" atom

can never be assigned as Chl *a* since Chl *a* contains no such atom.

Once pigment parsing is complete, the structure is displayed in an embedded NGLView[38] window, with protein represented in a translucent gray "ribbons" format, and pigments represented as green stick structures.[38] The approximate orientation of the Q_y transition dipole of each pigment (along the NB-to-ND atom axis) is indicated by a red line. (See Figure 2.)

Once a structure is loaded, the user can proceed to run Repair, MD, or Exciton calculations, using the "Select" tab to determine which chains should be included in subsequent calculations. By default all chains are included.

C. Structure Repair Tool

Once pigments are correctly identifies, additional "tweaks" to both pigment and protein structure are often necessary to produce chemically correct inputs for analysis. Where possible, PigmentHunter leverages existing open-source tools (such as the GROMACS[40] pdb2gmx utility) to accomplish these tasks. But certain pigment-specific operations (e.g., adding missing side groups) must be handled separately. PigmentHunter's Structure Repair window accomplishes three such functions often (but not always) required before running MD or structure-based excitonic simulations:

- It can be used to delete specific pigments not wanted in subsequent calculations. This is accomplished by deselecting the unwanted pigments from the Pigment List.

- It can apply the biosymmetry operations sometimes specified in PDB files under "REMARK 350" entries.

- It can add certain atoms that might not be resolved in experimental structures, e.g., the first few atoms of the phytol tail or some side groups.

At present, PigmentHunter is able to read symmetry operations only from *.pdb format files; for this reason, the Pigment Repair tool fetches structures from the Protein Databank in *.pdb rather than *.cif format. At present, biosymmetry operations for *.cif files are not supported. Structures generated by the Repair tool can either be downloaded directly or loaded as a stored PDB structure in the main window.

D. Molecular Dynamics Sequence

Once a structure is loaded and the chains of interest are selected, simple MD simulations can be prepared and run using a sequence of MD windows. The open-source GROMACS[40] engine is used as a back-end for all MD simulations, while PigmentHunter (along with the GROMACS[40] pdb2gmx utility) automates the process of preparing MD-ready structure and topology files.

From the 'Build MD Model' menu in the Main window, the user first clicks "Write Inputs", which creates a new simulation folder for the current structure and copies into it all necessary pigment and protein structural and topology data. Note that as of this writing, PigmentHunter supports full MD simulations only for Chl *a* and BChl *a* pigments. Work is underway to include a broader selection of pigments.

After inputs are written, clicking the "Simulate" button opens an MD Launch menu, where the user can choose to work on any previously written MD simulation folder. Once a folder is selected and the MD process launched, the user is guided through a series of windows, each of which completes a separate step in the MD process, using GROMACS as the MD back-end:

- Construction of the protein topology file, including the incorporation of any desired mutations.
- Assigning pigment topologies and adding pigment hydrogens. This includes an energy-minimization for pigment H atoms (with restrained heavy atoms).
- Energy minimization of the pigment-protein complex (without restraints).
- A short (100 ps) heavy-atom-restrained NVT simulation. The first 25 ps of this trajectory are discarded during post-processing to avoid sensitivity to the initial coordinates.

Details on each of these steps can be found in the Methods section below. Here we comment only on key features that all users should consider.[40]

First, although PigmentHunter automates all steps of this process, it is critical (as in any MD simulation) that the user review outputs to ensure that they are physically sensible. For example, at the end of each energy-minimization process, the user should view the potential energy curve that pops up in the simulation window. A "good" potential energy curve should decay smoothly with cycle time; a curve that shows sharp spikes or oscillates without converging could indicate a problem with system stability.

Second, special attention should be paid to the first step in the MD process, the construction of the protein topology. This is the process by which force field parameters – partial charges, force constants, dihedral angle potentials – are defined for all non-pigment residues in the input structure. (Recognized pigments are handled separately.) Although PigmentHunter handles this process automatically (iteratively cycling inputs through the GROMACS[40] pdb2gmx function and attempting to fix recognized errors), it is critical that the user monitor the process to ensure that the output topology is appropriate for their purposes. PigmentHunter handles pdb2gmx errors in four ways:

- Eliminating unrecognized residues: Any residue that is not present in the pdb2gmx database is simply deleted from the input file.

- Repairing residue fragments: If a protein atom is recognized by the pdb2gmx database but is missing non-hydrogen atoms, PigmentHunter will attempt to repair the residue by "mutating" the fragment to the correct (full) residue using an external PyMol script.[42] If the repair is for any reason unsuccessful, the residue is simply mutated to glycine.

- Inserting chain breaks: Many PDB structures contain gaps in the protein backbone, where residues could not be resolved experimentally. Such gaps trigger a "long bond" warning in the GROMACS[40] pdb2gmx function, since the default pdb2gmx behavior results in an unphysically long amide bond between the two residues on either side of the gap. PigmentHunter eliminates such unphysical bonds by simply inserting a chain break marker into the protein *.pdb file, resulting in two separate protein chains.

The physical validity of these repairs will vary considerably from system to system. For a well-resolved structure with few non-pigment heteroatoms (e.g., the 2DRE structure[43,44] for the water soluble Chl protein of *Lepidium virginicum*), this process produces a reasonable MD model, appropriate even for long time-scale simulations. For complex structures such as the Photosystem II core complex[13] (which contains many non-pigment heteroatoms), the resulting MD model will be of much more limited use, appropriate for estimating electrostatic *changes* due to single-point mutations but ill-advised for calculating *absolute* electrostatic parameters or for long time-scale simulations.[34]

Ultimately, it is the user's decision whether the final structure is appropriate for their intended purposes. To simplify this analysis, PigmentHunter's pdb2gmx interface logs all structural changes

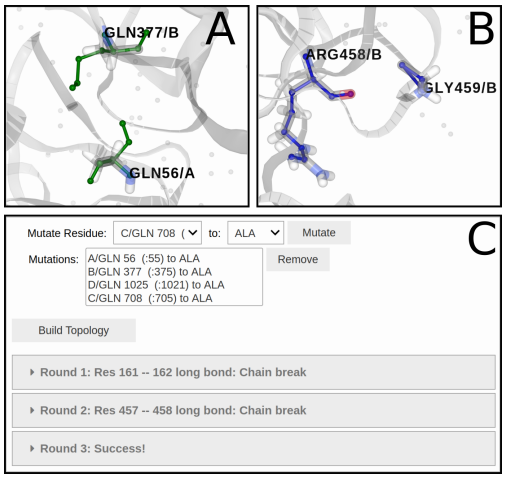

FIG. 3. Illustrative images from PigmentHunter's pdb2gmx interface using the 2DRE[43,44] PDB structure and a glutamine 57 site mutation to alanine. Frame A illustrates that mutations are highlighted in green in the final processed structure, while Frame B depicts the highlighting of chain breaks/repairs in blue. The mutation control panel is shown in Frame C.

both graphically (via an NGLView[38] representation) and in text both by displaying a simplified summary on the screen and by writing the full output from each GROMACS[40] pdb2gmx cycle to a downloadable log file.

During initial protein topology building, the user may also add any desired point mutations to the protein sequence. As illustrated in Figure 3, mutations are introduced by selecting the appropriate residue from a dropdown menu that lists all protein residues in the structure. Each residue is identified by its chain, residue name, residue sequence identifier from the original *.pdb or *.cif file, and its index in the ParmEd structure stored internally by PigmentHunter.[41] (ParmEd residue indices start at zero and increase sequentially through the structure.) Thus in Figure 3, PigmentHunter is to mutate to alanine (ALA) the residues identified as GLN 57 in all four chains (A, B, C, and D) in the PDB file, i.e., the residues at indices 55, 375, 705, and 1021 in the ParmEd structure. Mutations are accomplished using a PyMol script which selects from the PyMol rotamers library the orientation with fewest clashes with the surrounding environment.[42]

E. Exciton Models

Finally, once the final structural model is ready, exciton-based spectral simulations can be launched from the "Build Exciton Model" tab of the Main window. These simulations can be performed either on raw PDB structures or on structural trajectories loaded from MD simulations.

However, features that require the assignment of partial charges to the protein environment (most importantly the charge-density coupling model described below) can be run only after at least preparing the structure for MD by building topology files for both pigments and protein, i.e., the first two steps of the MD sequence just described. For such electrostatic-based calculations, it is strongly recommended to complete the entire MD sequence (through the initial NVT equilibration) to ensure reasonable sampling of structural fluctuations in the protein environment.

From the Exciton Model menu, the user can choose between point-dipole coupling[45] (PDC) and transition charge from electrostatic potential[22] (TrEsp) methods for calculating the electronic coupling constants that characterize interactions between pigments. When possible, the TrEsp model is usually preferred, since it provides a more accurate treatment of electrostatic interactions than the simpler PDC method.[22] (See Methods section for a more detailed description of both methods.) However, TrEsp is currently implemented for only certain pigments (see Table I), and it requires that all side group atoms be present in each Chl structure as well as the first few residues of the phytol tail. In many cases the necessary atoms can be "sprouted" using the Pigment Repair module already described. Otherwise, the simpler PDC model (which requires only the positions of the NB and ND atoms to run) can be used instead. Note that neither of these models handles the charge transfer interactions that are likely to be significant in very closely spaced pigments such as those found in photosynthetic reaction centers.

The Exciton Model menu also allows the user the choose between three options for handling site energies, i.e., the electronic transition frequencies of individual pigments in the absence of inter-pigment coupling. The simplest option is to simply assign all pigments an identical frequency under the "Uniform" dropdown setting. This option makes no attempt to gauge the impact of the protein structural environment on pigment site energies, but it can provide useful qualitative insight for highly symmetric and/or strongly excitonically coupled systems such as the B800 and B850 rings of the light harvesting 2 complex (LH2) of purple bacteria. (See Figure 7 below.)

Alternatively, the normal mode structural decomposition (NSD) model of Zucchelli[19] *et al.* (using the NSD algorithm of Jentzen et al.[46]) can be used to estimate how out-of-plane ring deformation affects site energies. Note, however, that this model has not been tested quantitatively against experimental data and that many crystal structures (even down to 2 Å resolution) are not sufficiently precise to accurately calculate the necessary deformation coefficients.[47] For this reason, NSD predictions should (for the present) be regarded as only suggestive.

Finally, the electrostatic CDC model estimates pigment site energies based on Coulomb inter-

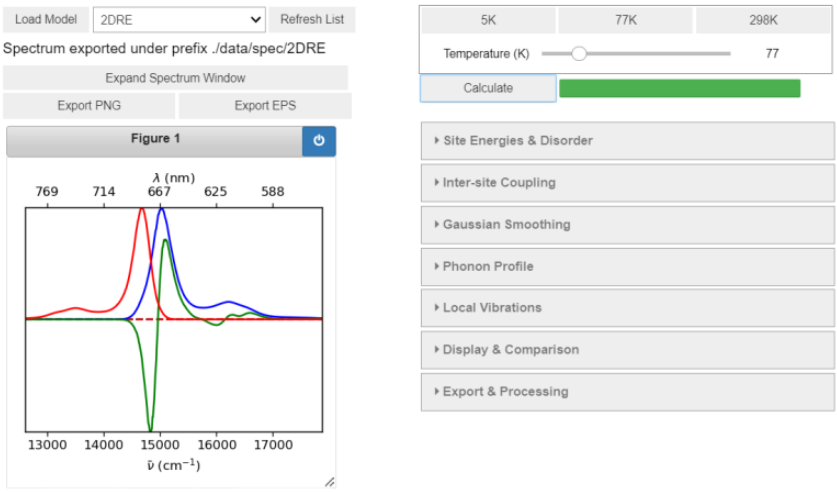

FIG. 4. The spectrum calculation page is shown here using parameters from the 2DRE[43,44] PDB structure. The fluorescence is shown in red; absorbance in blue; and circular dichroism in green. This calculator gives the user graphical control over simulation parameters including temperature, site energies, Gaussian disorder and smoothing, inter-site coupling, phonon profile and local vibrational spectral densities.

actions between the pigment and its environment.[45] This method has seen the most rigorous testing against experiment but even then only in predicting relative frequency shifts for point mutants that strongly perturb the local charge environment.[34] Its accuracy in predicting absolute frequencies for individual pigments remains to be rigorously tested, and it does not account for ligation or steric effects such as ring deformation. Note that the CDC method is available only when an MD trajectory (or pdb2gmx-generated structure) including atomic partial charges is loaded, or when the user supplies an external charge file through the Exciton Model window. External charge files should be in *.txt format and must contain only numeric data, with the n^{th} line in the file corresponding to the partial charge of the n^{th} atom in the structure.

F. Spectrum Calculations

After an exciton model has been constructed, spectroscopic simulations can be launched by clicking the "Calculate" button on the Exciton Model menu. This opens a separate Spectrum Calculation window (see Fig. 4), where the user can choose from a dropdown list of stored exciton models. Alternatively, the "Download" button opens a file tree where the user can download text files containing exciton model parameters. (See the Supplementary Material for a complete list of exciton model files.) With minor modifications to format, these parameters can be imported to several different third-party packages for both linear and nonlinear spectroscopic simulations, in-

cluding the NISE_2017 package[48] [https://github.com/GHlacour/NISE_2017], SPECTRON [https://mukamel.ps.uci.edu/software.html], and the the GPU-HEOM tool[49] hosted on nanoHUB itself https://nanohub.org/resources/gpuheompop.

In the Spectrum Calculation window, the user can choose a variety of options for the spectroscopic simulation method. The "Temperature" panel allows the user to set the simulation temperature using either the three pre-set buttons (5 K, 77 K, and 298 K) or a slide bar running from 1 K to 500 K. The temperature enters spectrum calculations in two ways: first, it sets the Boltzmann population for the contribution of each exciton state to fluorescence spectra; second, it alters the single-site line shape when phonon or vibrational convolutions are included. (See section III D for details.)

The "Site Energies & Disorder" panel allows the user to add Gaussian disorder to the pigment site energies (local transition frequencies), with a slidebar controlling how many times the Gaussian disorder is randomly sampled. By default, each pigment is assigned a Gaussian with a full-width at half-maximum (FWHM) of 400 cm^{-1} and the center frequency dictated by the exciton model.

The "Gaussian Smoothing" panel allows the user to convolve the final calculated spectrum with a Gaussian of variable width. This feature is turned on by default to eliminate sampling noise, though users may wish to turn it off (and increase the number of random samples) for publication-quality simulations.

The "Phonon Profile" menu allows users to convolve the spectrum with a temperature-dependent sideband structure that models the effect of coupling to low-frequency protein vibrational (phonon) modes. Using the dropdown menus, distinct phonon sideband shapes (extracted from low-temperature experimental data) and Huang-Rhys factors (S_{tot}) can be selected for Chl *a*, Chl *b*, and BChl *a*. A single shape and S_{tot} value is assigned to all "Other" pigment classes. By default, this option is enabled, with phonon sidebands selected based on the pigment types assigned by PigmentHunter's Main window and stored in the "names.txt" file stored in the exciton model folder.

The "Local Vibrations" menu performs a similar function but for higher-frequency local pigment vibrations rather than for broad low-frequency phonon profiles.

The "Display & Comparison" menu allows the user to control which spectra (Absorption, Emission, CD) are displayed in the simulation window. The user can also upload experimental spectra to be shown as dashed lines for comparison to calculations. The "Export & Processing"

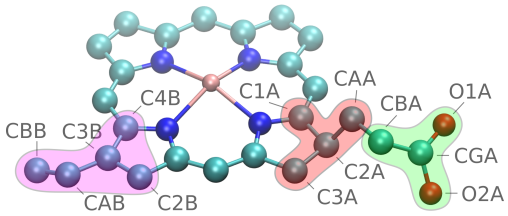

FIG. 5. Illustration of chemical side groups (colored regions) that can be "sprouted" using PigmentHunter's structure repair tool.

panel allows the user to export simulated spectra or to launch a separate "Processing" window for further comparison and manipulation (e.g., normalization, spectrum subtraction, etc.).

III. COMPUTATIONAL METHODS

The last section offered an overview of the type of simulations that can be performed using PigmentHunter; in this section, we provide a more technical description of the underlying methods.

A. Pigment Repair

PigmentHunter's missing-atom algorithm takes inputs from a set of tab-delimited template files for each implemented pigment (currently Chl *a*, Chl *b*, and BChl *a*). Each line in the text file indicates an atom that *should* be present in the structure; if absent, PigmentHunter will try to add it. For example, the template file for Chl *a* (misc/repair/templates/CLA.txt) contains as its first line the entry

```
CAA    sp3 C2A C1A C3A.
```

This indicates that an atom named "CAA" should be present in the structure; if *not* present, it should be "sprouted" from its nearest bonding partner, the atom "C2A", which is sp3 hybridized. Atoms "C1A" and "C3A" are also bonded to C2A. Based on this information, PigmentHunter will position the CAA atom in a tetrahedral geometry, at a 109.5^{o} angle relative to both C1A and C3A. The corresponding atoms are illustrated in the orange-shaded area of Figure 5. Note that the ordering of the C1A and C3A atoms is significant for chirality: the new atom is positioned on the side of the C1A-C2A-C3A plane indicated by the right-hand-rule via the cross product of the vectors C2A$\rightarrow$C1A and C2A$\rightarrow$C3A.

In some cases, the geometry of the missing atom may not be completely determined by the

existing structure. For example, the second line of CLA.txt is the entry

```
CBA sp3 CAA C2A -
```

This again indicates that if the atom "CBA" is missing, it should be sprouted in a sp3 geometry from its neighbor CAA, which also bonds to atom C2A. The final entry in thise case, however, is a hyphen, which indicates that there is no third atom that can be used to determine the final positioning of the missing atom, since the sp3 bond between atoms C2A and CAA can rotate freely. In this case, PigmentHunter steps through the possible rotation geometries (with a 1 degree step size), and selects the orientation with the largest separation between the sprouted atom and its nearest non-bonded neighbor.

A similar strategy is used for atoms to be sprouted from sp2-hybridized sites. In the lines

```
CAB sp2 C3B C2B C4B
CBB sp2 CAB C3B -
O1A:O2A sp2 CGA CBA -
```

the first entry indicates the atom or atoms to be added; the second entry indicates hybridization; the third entry is the sprouting site; and the second and third entries indicate other atoms bonded to the sprouting site. In the first line (see purple-shaded area of Figure 5), the geometry of the added atom is completely determined, since the sp2-hybridized sprouting site already has two bonded neighbors. In the second case, free rotation around the CAB-C3B bond is allowed, so that the position of the atom must be chosen as above by minimizing clashes with non-bonded neighbors. And in the final case, (see green-shaded region in Figure 5) two atoms are to be be sprouted simultaneously, with free rotation allowed.

Note that, at present, no missing-atom repairs implented in PigmentHunter require distinguishing between *cis* and *trans* isomers, since the chemical side groups sprouted from the central Chl macrocycle are generally free of such isomerization sites. An important exception is the Chl phytyl tail, which does contain a *cis*-configuration double bond. However, given the limited intelligence of our atom-placement algorithm (which works one atom at a time), we intentionally avoid constructing long chains such as the phytyl tail.

B. Molecular Dynamics

All MD simulations in PigmentHunter use the GROMACS MD package[40] (version 2018.4) installed on nanoHUB.org. The hublib library (https://pypi.org/project/hublib/) is used to handle

job submission and connection to the GROMACS[40] environment.

Protein MD parameters are taken from the OPLS-AA force field, while force fields for Chl *a* and BChl *a* are taken from Refs.[50,51] Although any water molecules resolved in the input structure are preserved in the MD simulation (using the SPC/E model), the system is otherwise not solvated, but placed in an empty cubic box with a minimum distance of 2 nm between the protein and the nearest wall. Electrostatic forces are handled with a simple 1.5 nm Coulomb cutoff. Full MD parameters can be found in the files hmin.mdp (energy minimization with heavy-atom restraints), em.mdp (energy minimization without restraints), and nvt.mdp (100 ps NVT equilibration) in the folder misc/gmx/mdp in the main PigmentHunter working directory or at `https://github.com/mreppert/PigmentHunter/blob/master/misc/gmx/mdp/`.

Protein MD topology files are created using the GROMACS[40] pdb2gmx function, with PigmentHunter iteratively passing the input structure to pdb2gmx and then modifying the structure in response to pdb2gmx outputs. Because pigments like Chl *a* and BChl *a* are not included in standard force field databases, their topology files must be handled separately. Briefly, topology input files CLA.itp (for Chl *a*) and BCA.itp (BChl *a*) are copied as needed into the MD working directory and included in the full complex topology file (complex.top) via include statements (e.g., '#include "CLA.itp"'). A schematic representation of file handling during topology generation is provided in the Supplementary Material.

Since hydrogen atoms are typically missing from PDB structures, these are "sprouted" based on the corresponding pigment topology file. After matching all non-hydrogen atoms in the pigment structure with exactly one entry in the topology file, the topology "[bonds]" table is used to identify identify both the hybridization state of each non-hydrogen atom (SP2 or SP3) and how many H atoms should be sprouted at each location. H atoms are then added sequentially to each non-hydrogen atom in the structure, with a standardized bond length of 1.1 Å and bond angles of either 120^{o} for SP2-hybridized atoms or 109.5^{o} for SP3 hybridization. Note that any H atoms present in the input structure are stripped and replaced by this protocol.

Mathematically, the Frenkel exciton model (illustrated graphically in Figure 6) describes electronic excitations via a Hamiltonian of the form[16,55,56]

$$\hat{H} = \hat{H}_l + \hat{H}_{ss} \tag{1}$$

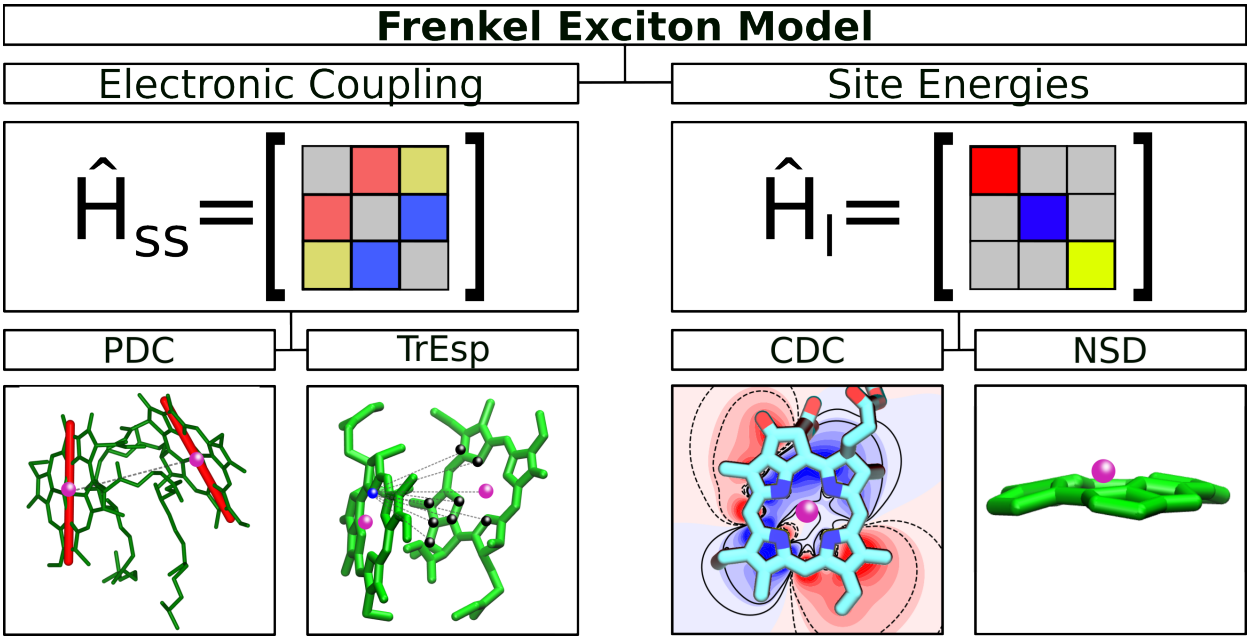

FIG. 6. The Frenkel (molecular) exciton model is the basis for PigmentHunter's excitonic calculations. Off-diagonal elements in the Frenkel Hamiltonian represent electronic couplings between pigments, while on-diagonal elements represent site energies (transition frequencies of individual pigments). PigmentHunter includes two different methods for predictiong electronic coupling values, the point-dipole coupling (PDC)[45] and Transition Electrostatic Potential (TrEsp)[22,52] methods. Similarly, two different site energy models are currently implemented: the electrostatic charge density coupling (CDC) method[22,53,54] and normal-coordinate structural decomposition (NSD) method, which monitors Chl ring deformation.[19,46].

where the "local" Hamiltonian

$$H_l = \sum_n \hbar\omega_n |n\rangle \langle n| \tag{2}$$

depends on the local excitation energies or "site energies" $\hbar\omega_n$ of each pigment, and

$$H_{ss} = \sum_{n,m} \hbar V_{nm} |n\rangle \langle m| \tag{3}$$

describes the site-to-site interaction energy or electronic coupling V_{nm}, between each pair of pigments.

In applying the Frenkel exciton model to real PPCs, the critical question is how to best estimate the site energies $\hbar\omega_n$ and electronic couplings V_{mn} that parameterize the model. In general terms, computational methods for estimating electronic couplings are well-established and believed to be quite accurate; for the most part, difficulties arise only for very closely-spaced pigments where

electrons may actually be shared between multiple pigments.[57,58] Estimating pigment site energies, by contrast, is quite difficult because (in contrast to couplings) these values are extremely sensitive to the local environment of the pigment. Both pigment-protein electrostatic interactions and steric deformation of the pigment ring are expected to play important roles in site energy determination, but quantifying these effects has proven to be a challenge.[19–21,34] The sections below describe the methods currently available in PigmentHunter for estimating both electronic couplings and site energies, along with guidelines for when each approach is appropriate.

1. Electronic Coupling Parameterization

Electronic couplings can be estimated in PigmentHunter using two different electrostatic methods: the point-dipole coupling (PDC) and Transition Electrostatic Potential (TrEsp) models. When available, the TrEsp model is generally more accurate since it explicitly sums over electrostatic interactions between all pairs of atoms in the two pigments.[22] The PDC method is much simpler, replacing the TrEsp's explicit sum over atomic interactions with a dipole-dipole approximation. For large separation distances, the two methods give similar results; for closely spaced pigments, TrEsp is generally more accurate. The chief advantage of the PDC method is that it only requires the positions of the four ring N atoms (NA, NB, NC, and ND) and thus can be applied to any chlorophyll-like pigment recognized by PigmentHunter. (Atoms NB and ND are used to assign the transition dipole orientation, while all four atoms are used to identify the "center" of the molecule.) The TrEsp method, since it explicitly sums over atom-atom interactions requires the locations of all (non-hydrogen) ring atoms and some side groups, which may not be available for some experimental structures.

2. PDC

PDC couplings are calculated in PigmentHunter as

$$\hbar V_{mn} = \frac{\mu_{\text{eff}}^2}{R_{mn}^3}[\boldsymbol{e}_m \cdot \boldsymbol{e}_n - 3(\boldsymbol{e}_m \cdot \boldsymbol{e}_{mn})(\boldsymbol{e}_n \cdot \boldsymbol{e}_{mn})] \tag{4}$$

where the transition dipole is oriented along the NB-to-ND axis and is represented by unit vector $\boldsymbol{e}_m$ for the mth pigment while $\boldsymbol{e}_{mn}$ represents the unit vector connecting the centers of the mth and

*n*th pigments[45].

The dipole strength for the Q_y transition is represented in Eq. (4) by μ_{eff}^2, which includes an empirically determined scaling factor reflecting the effect of the local environment on both the transition dipole strength and the dielectric screening of inter-pigment interactions. Although dipole strengths for Chl-like pigments in organic solvents have been carefully tabulated by Knox & Spring[59], there does not seem to be a single agreed-upon approach to determining μ_{eff} values for pigments embedded in a protein environment. The values used by PigmentHunter for μ_{eff} are reported in Table I along with the corresponding reference number. In the case of Chl *b* (CLB), the reported value is the square root of the cavity-extrapolated vacuum value reported by Knox and Spring scaled by the same coefficient (4.3/4.6) applied to CLA in Refs.[10,60] The value for BChl *a* (BCA) is likewise the square root of the Knox and Spring dipole strength scaled by the factor of 0.8 determined in Ref.[45], i.e., $5.45\ \text{D} = \sqrt{0.8 \cdot 37.1\ \text{D}^2}$.

3. TrEsp

TrEsp couplings are calculated in PigmentHunter via the Coulomb-interaction sum[22]

$$\hbar V_{mn} = \sum_{I=1}^{N} \sum_{J=1}^{N} \frac{q_I^{(m)}(1,0) q_J^{(n)}(0,1)}{|\boldsymbol{R}_I - \boldsymbol{R}_J|} \tag{5}$$

where the indices I and J sum over all atoms contained in the m^{th} and n^{th} pigments respectively, N is the number of parameterized atoms in each pigment, and $q_I^{(m)}$ and $q_J^{(n)}$ are the transition partial charges assigned to each atom.[22] $\boldsymbol{R}_I$ and $\boldsymbol{R}_J$, are the respective atomic positions. The atomic partial charges $q_J^{(n)}$ are determined by a fit to an electrostatic transition potential determined by *ab initio* time-dependent density functional theory (TD-DFT) using exchange-functional B3LYP.[22,52] PigmentHunter uses literature values for the transition charge parameters for Chl *a*[22], Pheo *a*[54], Chl *b*[53], and BChl *a*[22], but renormalizes each set of transition charges to reproduce the same transition dipole strength as in the PDC approach.

C. Site Energy Parameterization

Excitonic coupling is only one of the several factors that influence the optical properties of PPCs. Protein and environmental effects such as hydrogen bonding[21,61], charged amino acid

residues[34], and chlorophyll ring deformation[46] can all influence PPC optical spectra. PigmentHunter currently offers two models that can be used to estimate site-energy shifts based on either electrostatic interactions or ring deformation.

1. NSD

The simplest site-energy method in PigmentHunter is based on the normal-coordinate structural decomposition (NSD) method[46] developed by the Shelnutt group for characterizing ring-deformation in porphyrins, which was adapted to estimate Chl transition frequencies by Zucchelli et al.[19,46]

The NSD method compares the coordinates of target pigments and a copper(II) porphine reference. The deformation coefficients are the displacements of the target pigment from the reference along each normal mode. For the n^{th} mode we have

$$d_n = \hat{D}_n \cdot \vec{D}_{\text{obs}} \tag{6}$$

where $\hat{D}_n$ is normal mode displacement vector and D_{obs} is the observed displacement between the pigment and the reference.[46] In PigmentHunter, these values are stored in the output file "nsd.txt" following trajectory analysis using the NSD approach.

Under the site-energy model of Zucchelli et al.[19], pigment site-energy are determined by the three equations

$$\Delta E_{H\to L} = \left(E^0_{e_g(x)} - E^0_{a_{1u}}\right) + \left(\delta E_{e_g(x)} - \delta E_{a_{1u}}\right) \tag{7}$$

$$\Delta E_{H-1\to L+1} = \left(E^0_{e_g(y)} - E^0_{a_{2u}}\right) + \left(\delta E_{e_g(y)} - \delta E_{a_{2u}}\right) \tag{8}$$

$$E_{Q_y} = \frac{1}{2}\Big[\Delta E_{H\to L} + \Delta E_{H-1\to L+1} - \sqrt{\left(\Delta E_{H-1\to L+1} - \Delta E_{H\to L}\right)^2 + 4C^2}\Big] \tag{9}$$

Here $\Delta E_{H\to L}$ and $\Delta E_{H-1\to L+1}$ are the respective energies of the HOMO $\to$ LUMO and (HOMO-1) $\to$ (LUMO+1) transitions, while E_{Q_y} is the energy of the Q_y transition (i.e., the pigment site energy), to which both $\Delta E_{H\to L}$ and $\Delta E_{H-1\to L+1}$ contribute via the coupling constant C in Eq. (9). E^0_Γ is the equilibrium-geometry energy of the electronic state with symmetry Γ, while δE_Γ is the deformation-dependent first order energy change given by $\delta E_\Gamma = k_\Gamma d^2_\Gamma$, with k_Γ the force constant

for the relevant normal mode. Thus deformation along each normal mode is assumed to modulate only the energy of the electronic state with matching symmetry.[19]

While the NSD method is computationally fast and offers useful qualitative insight into ring deformation, it comes with two important caveats. First, monitoring ring deformation in PPCs is inherently approximate at the structural resolution currently available for most PPCs. The atomic displacements involved in PPC ring deformation are generally quite small, less than 1 Å even for strongly deformed pigments. The structural resolution needed to accurately monitor such effects experimentally is thus significantly higher than, for example, what is needed to characterize long-range electrostatic forces. In fact, in very low-resolution structures, NSD coefficients are determined largely by the position restraints used to process x-ray scattering data.[47] As experimental resolution increases, such external constraints become less important, but even down to 2 Å resolution, the refinement procedure still plays a significant role in determining the NSD wave and propeller coefficients.[47] Second, the particular site-energy method developed by Zucchelli et al.[19] and implemented in PigmentHunter has not been rigorously tested against experimental data at the single-site level. Although the method often provides a reasonable distribution of site energies when applied to an entire complex (such as the LHCII complex considered by Ref.[20]), this does not necessarily translate into accurate site energies for individual pigments. Thus, until further experimental validation is available, we recommend that NSD-based site energies be regarded as only suggestive.

2. CDC

The charge density coupling (CDC) method predicts site energies based on electrostatic interactions between the pigment and its local environment.[21] CDC calculations are necessarily somewhat more time-consuming than NSD estimates (since the entire protein environment must be considered), but the CDC approach comes with the benefit of having shown several definite successes in reproducing experimentally determined frequency-shift trends at the level of individual pigments.[21,34] Indeed, in several cases, electrostatic effects have been convincingly argued to be the dominant factor in determining Chl transition frequencies, with effects such as ring distortion playing only a minor role.[31,62,63] Even here, however, it is important to recognize that (at least in terms of head-to-head comparison with experiment) electrostatic methods have shown the most quantitative success in predicting *relative* site energies such as mutation-induced frequency

shifts.[31,34] Whether or not purely electrostatic methods like the CDC approach are capable of predicting *absolute* site energies requires further investigation.

As implemented in PigmentHunter, the CDC method uses atomic partial charges assigned by the MD force field (or uploaded by the user) to estimate pigment-environment electrostatic interaction energies for both the ground and excited states of the pigment. The overall site energy shift is then calculated as the difference between the interaction energies in the electronic excited and ground states; thus, for example, electrostatic interactions that raises the excited-state energy *more* than the ground-state energy result in a blue shift, while interactions that raise the ground-state energy more result in a red shift.

Explicitly, the transition-energy shift (relative to vacuum) is estimated as[21]

$$\Delta E_m = \frac{1}{\varepsilon_{\text{eff}}} \sum_I \sum_J \frac{\Delta q_I^{(m)} q_J^{(\text{bg})}}{\left| \boldsymbol{R}_I^{(m)} - \boldsymbol{R}_J^{(\text{bg})} \right|}, \tag{10}$$

where the index I runs over all atoms in pigment m, the index J runs over all atoms in the "background" (all atoms in the system other than pigment m), and $\boldsymbol{R}_I^{(m)}$and $\boldsymbol{R}_J^{(bg)}$ are the respective atomic positions. The effective dielectric constant ε_{eff} was originally set to 2.5 in Ref.[21] based on empirical fitting against experimental data. This value has subsequently been shown to provide remarkably good agreement against experimental mutation-induced frequency shifts without any further adjustment.[34] While PigmentHunter uses $\varepsilon_{\text{eff}} = 2.5$ for CDC calculations by default, the user can modify this parameter if desired using the ε_{CDC} widget on the Exciton Menu; this value affects only CDC calculations, not site-to-site electronic coupling constants calculated via the TrEsp or PDC methods.

Note that if pigment site energies were determined solely by electrostatic effects, it would be straightforward to convert these relative frequency shifts ΔE_m into absolute site energies

$$E_m = E_0 + \Delta E_m \tag{11}$$

by simple fitting against experiment to determine a zero-field transition frequency E_0 for each pigment type.[21] (E_0 would correspond to an overall frequency shift, while ε_{eff} determines the spectrum width.) The situation is complicated by the fact that non-electrostatic effects such as ring-deformation and ligation are likely also to play a role in determining pigment site energies, although the magnitude of such effects remains unclear. PigmentHunter currently uses zero-field

values of $\frac{hc}{E_0} = 670$ nm for Chl *a* and $\frac{hc}{E_0} = 800$ nm for BChl *a*, although for the present these values should be considered only as rough approximations.

Partial charges for the protein background $q_J^{(\mathrm{bg})}$ are taken from the MD force field, while the charge differences $\Delta q_I^{(m)}$ between excited- and ground-state potentials are taken from the gas-phase values of Refs.[22,53,54] Chl *b* values were not previously available in the literature but were kindly provided to us by Thomas Renger.

D. Spectrum

Once Hamiltonian model parameters are estimated, optical spectra can be calculated within the Frenkel exciton model, with phonon and vibrational coupling incorporated approximately via line shape functions. In this section we first briefly outline the approximations PigmentHunter uses to calculate optical spectra and then provide explicit expressions.

- *Single-excitation Limit*: We assume that only one electronic excitation exists in the system at a time. This is generally a safe assumption for photosynthetic proteins in ambient light, but it means that PigmentHunter cannot (yet) describe nonlinear measurements such as pump-probe and two-dimensional spectroscopies.

- *Harmonic Environment*: In proteins, Chl electronic transitions interact with both local vibrations and environmental phonons, the so-called "bath" modes. The most important consequence of this interaction in the frequency domain is the modification of the electronic line shape by vibrational and phonon sidebands. (In the time domain, these interactions facilitate energy transfer and thermalization.) PigmentHunter makes several simplifying assumptions in treating these effects, the most fundamental of which is that each bath mode is taken to be a harmonic oscillator, whose potential energy surface is linearly displaced in the electronic excited state. (See, e.g., Ref.[64] for an explicit description of the model.)

- *Localized Lineshapes*: Even within this displaced harmonic oscillator model, exact spectroscopic simulations are numerically intensive. To speed up calculations, PigmentHunter neglects modifications to the electronic line shape due to excitonic delocalization. Instead, each excitonic state is assigned an absorption/fluorescence line shape function ($D_M^{(i)}(\bar{\nu})$ in the notation below) that is simply a weighted sum (by population) of the local absorption/fluorescence spectra ($L_p(\bar{\nu})$) of each pigment contributing to the excitonic state. This

treatment becomes exact in the limit of weak site-to-site interactions (i.e., the Förster limit). In the case of stronger interactions (e.g., the Redfield and intermediate regimes), excitonic states are still treated as delocalized, but effects like energy-transfer broadening and delocalization-induced changes to the reorganization energy and sideband line shapes are neglected.[64–67]

To implement this approach, the operator $\hat{H}$ of Eq. (1) is first represented as a real matrix $\boldsymbol{H}$ that can be diagonalized numerically. For a system with N_{sites} pigments, the site-basis Hamiltonian matrix $\boldsymbol{H}$ is an $N_{\text{sites}} \times N_{\text{sites}}$ matrix whose diagonal elements represent the site energies of each pigment in units of cm^{-1}, i.e.,

$$H_{nn} = \bar{\nu}_n \equiv \frac{\omega_n}{2\pi c} \tag{12}$$

where $\hbar\omega_n$ is the site energy of Eq. (2). (Here c is the speed of light, used to convert from angular frequency to wavenumbers.) The off-diagonal elements of $\boldsymbol{H}$ are likewise the coupling elements $\frac{V_{nm}}{2\pi c}$ of Eq. (3) in units of cm^{-1}.

If disorder is present in the simulation, the diagonal elements of $\boldsymbol{H}$ differ from one snapshot to the next since the site energies for each pigment are randomly sampled from a Gaussian distribution with parameters determined in the Spectrum window. (Note that when site energies are estimated using NSD and CDC methods, the Spectrum window loads only the average site energy for each pigment across the entire trajectory; PigmentHunter does not currently use frame-to-frame disorder from MD data to predict Gaussian distribution widths.) In what follows, a superscript "(i)" will be used to denote quantities that depend on the sample number i in simulations with disorder. For example, $\boldsymbol{H}^{(i)}$ represents the Hamiltonian matrix for the i^{th} snapshot, including randomly sampled site energies on the diagonal. The off-diagonal entries (coupling elements) of $\boldsymbol{H}$ are the same in all samples.

Once the Hamiltonian matrix is constructed, optical spectra are calculated in the exciton basis by numerically diagonalizing $\boldsymbol{H}$ to find the orthogonal matrix $\boldsymbol{u}$ such that $\boldsymbol{u}\boldsymbol{H}\boldsymbol{u}^{\text{T}}$ is diagonal. The diagonal elements

$$\tilde{\nu}_M \equiv \left[\boldsymbol{u}\boldsymbol{H}\boldsymbol{u}^{\text{T}}\right]_{MM} \tag{13}$$

of the resulting diagonal matrix represent the energies of each exciton state in the system in units

of cm^{-1}. Each row in $\boldsymbol{u}$ represents a different excitonic state, with columns representing the excitation amplitude for a given pigment, i.e., the element u_{Mm} gives the excitation amplitude for the contribution of pigment m to exciton M. The quantity $|u_{Mm}|^2$ is the probability of finding pigment m excited if the system is in exciton state M. Because $\boldsymbol{u}$ and $\tilde{\nu}_M$ both depend on $\boldsymbol{H}$, if disorder is present in the system, the quantities $\boldsymbol{u}^{(i)}$ and $\tilde{\nu}_M^{(i)}$ will also vary from one snapshot to the next.

Spec	$a_M^{(i)}(\bar{\nu})$
Abs	$\bar{\nu}\left\|\vec{\mu}_M^{(i)}\right\|^2$
CD	$-\bar{\nu}\sum_{m>n} u_{Mm}^{(i)} u_{Mn}^{(i)} \vec{R}_{mn}\cdot(\vec{\mu}_m\times\vec{\mu}_n)$
Flu	$\bar{\nu}^3 \sum_M \frac{\left\|\vec{\mu}_M^{(i)}\right\|^2 e^{-\frac{hc\bar{\nu}_M^{(i)}}{kt}}}{\sum_N e^{-\frac{hc\bar{\nu}_N^{(i)}}{kt}}}$

TABLE II. $a_M^{(i)}$ represents the signal amplitude coming from exciton M on i^{th} snapshot.

After diagonalizing $\boldsymbol{H}^{(i)}$, optical spectra are calculated in the exciton basis using the generic form[64]

$$\mathrm{Spec}(\bar{\nu}) = \sum_{i=1}^{N_{\mathrm{iter}}} \sum_{M=1}^{N_{\mathrm{exc}}} a_M^{(i)}(\bar{\nu}) D_M^{(i)}(\bar{\nu}). \tag{14}$$

Here "Spec($\bar{\nu}$)" represents the absorption (Abs), fluorescence (Flu), or circular dichroism (CD) spectrum, $a_M^{(i)}$ is the signal amplitude coming from exciton M on sample i, and $D_M^{(i)}(\bar{\nu})$ is the corresponding spectral line shape. Explicit expressions for $a_M^{(i)}$ are provided in Table II; for absorption and fluorescence, these amplitudes depend on the exciton transition dipole moments[64]

$$\vec{\mu}_M^{(i)} \equiv \sum_{n=1}^{N_{\mathrm{sites}}} u_{Mn}^{(i)} \vec{\mu}_n, \tag{15}$$

where $\vec{\mu}_n$ is the local (site basis) transition dipole moment for pigment n calculated by PigmentHunter during exciton model construction; although the local transition dipole vectors do not vary with sample number i, the exciton dipoles $\vec{\mu}_M$ vary with i due to their dependence on $\boldsymbol{u}^{(i)}$. The quantity $\|\vec{\mu}_M\|^2$ in Table II is the squared magnitude of the transition dipole vector.

The line shape $D_M^{(i)}(\bar{\nu})$ for each exciton is calculated in PigmentHunter as a weighted sum over

line shapes $L_p(\bar{\nu})$ for each type of pigment (Chl *a*, Chl *b*, etc.) contributing to the exciton:

$$D_M^{(i)}(\bar{\nu}) = \sum_{p=1}^{N_{\text{types}}} f_{Mp}^{(i)} L_p\left(\pm\left(\bar{\nu} - \bar{\nu}_M^{(i)}\right)\right) \tag{16}$$

where the quantity

$$f_{Mp}^{(i)} = \sum_{m}^{\substack{\text{pigments} \\ \text{of type } p}} \left|u_{Mm}^{(i)}\right|^2 \tag{17}$$

represents the fractional probability that the exciton is localized on a pigment of type p. The "+" sign applies to absorption and CD signals, while the "-" sign applies in fluorescence. The line shapes $L_p(\bar{\nu})$ represent the absorption spectrum of an isolated pigment of type p, calculated using the parameters (Huang-Rhys factors and vibrational and phonon spectral densities) selected from the drop-down menu on the Spectrum window. Explicit expressions for $L_p(\bar{\nu})$ are provided in the Supplementary Material. Note that these line shapes are *not* corrected for excitonic effects such as energy-transfer broadening and modified electronic-vibrational interactions.[64–67]

IV. APPLICATIONS

To give a sense for PigmentHunter's capabilities for spectroscopic simulations, we offer three example applications, focusing on the bacterial light-harvesting complex 2 (LH2) of *Rhodobacter sphaeroides*, the Fenna-Matthews-Olson protein of *Chlorobaculum Tepidum*, and the two core photosystems, Photosystem I (PSI) and Photosystem II (PSII), of oxygenic photosynthesis.

A. Excitons: LH2

For our first example, we take the highly symmetric LH2 complex to illustrate the role played by various simulation parameters – electronic couplings, site energies, and phonon and vibrational interactions – in determining absorption and fluorescence line shapes in PPCs. LH2 is a convenient system to explore these dependencies since its absorption spectrum features well-separated bands arising from two different pools of pigments with quite distinct excitonic properties. Structurally, LH2 consists of two spatially separated rings of pigments. The "B800" ring absorbs near 800 nm and contains 9 pigments (one pigment per monomer), which are weakly coupled due to the

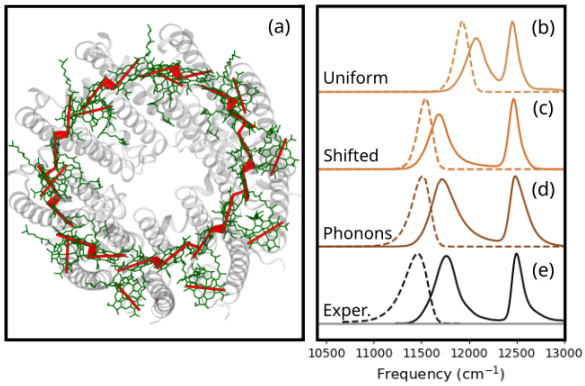

FIG. 7. The 7PBW[69,70] PDB structure (frame (a)) is used to demonstrate the effect of various parameters on simulated absorption (solid) and fluorescence (dashed) spectra. Frames (b) - (e) depict calculated spectra assuming uniform site energies ("Uniform"), shifted site energies for the B850 and B800 rings ("Shifted"; 12110 cm^{-1} and 12510 cm^{-1}, respectively), and after dressing with phonon sidebands ("Phonons"; $S = 0.8$). All simulations are at 5 K. Frame (e) depicts experimental 5 K spectra from Ref.[68]

relatively large separation from monomer to monomer. The second "B850" ring contains 18 pigments (two pigments per monomer) and due in part to the much stronger coupling of these closely spaced pigments absorbs near 850 nm. Due to the high efficiency of B800 $\rightarrow$ B850 energy transfer, fluorescence in LH2 comes exclusively from the B850 band, as reflected in the experimental 5K absorption and fluorescence spectra (curves (e), taken from Ref.[68]) in Figure 7.

Curves (a) to (d) in Figure 7 show simulated absorption (solid lines) and fluorescence (dashed lines) spectra for LH2 calculated in PigmentHunter with increasing refinement of simulation parameters against experiment. Curve (a) is calculated using TrEsp electronic couplings and uniform average site energies (12500 cm^{-1}) for all pigments, with no vibrational dressing. The higher-frequency peak here is due to the B800 pigment ring, while the lower-frequency peak is due to the more strongly coupled B850 pigments. Since site energies are the same for all pigments, the shift between the two bands comes entirely from the excitonic delocalization of the B850 ring. The only adjustable parameters in this case is the Gaussian disorder, which is set to a FWHM of 500 cm^{-1} for pigments in the B850 ring and 145 cm^{-1} for the B800 ring. (See Ref.[71] for a discussion of why the static disorder in the two rings varies so much.)

While this bare excitonic simulation correctly predicts that the B850 ring is red-shifted compared to B800, the magnitude of the splitting is much smaller than experiment. Within the Frenkel exciton model, there are two possible explanations for this: either the calculated coupling values are too small (so that the exciton-induced red-shift for B850 is too weak) or the site energies of the

B850 pigments are red-shifted relative to B800 (so that the excitonic shift builds on a pre-existing bias in transition energies). In fact, both factors likely contribute here: careful fitting against experimental data suggests for the B850 band a maximum inter-site coupling of around 315 cm^{-1} (compared to only 183 cm^{-1} estimated by PigmentHunter's TrEsp method) and a mean site energy of 12390 cm^{-1} (compared to PigmentHunter's default value of 12500 cm^{-1}).

Estimates for these parameters vary in the literature (e.g., the transition-density cube (TDC) method[72] gives a maximum coupling of 238 cm^{-1} for *R. acidophila* LH2, while an alternate fit to experiment[68] suggests 430 cm^{-1}), and we make no effort here to determine precise values. (Note that the difference between our TrEsp values and the TDC values of Ref.[72] are most likely due to differences in the assumed dielectric constant and effective oscillator strength; the TDC and TrEsp methods give essentially identical results given the same input parameters.[22]) Since our purpose is simply to illustrate how various parameters affect the simulated spectra, we opt to keep the inter-pigment couplings fixed at the calculated TrEsp values and adjust the relative site energies of the B800 and B850 sites to match experiment. (The converse option would be to uniformly scale the coupling values using the "Inter-site Coupling" box in PigmentHunter's Spectrum simulation window.) Curves (b) in Figure 7 show the result of shifting the site energies of the B850 pigments to 12110 cm^{-1} and the B800 sites to 12510 cm^{-1}. This single adjustment brings the simulated absorption bands into very near agreement with the experimental peaks.

Examining the data more closely, however, one finds that the Stokes shift between the peak frequencies of B850 absorption and fluorescence are noticeably smaller in the simulations than in experiment. This deficiency is due to the neglect in these calculations of electron-phonon coupling, which increases the Stokes shift due to the fact that the phonon sideband appears in absorption spectra to *lower* frequency from the purely electronic transition but to *higher* frequency in fluorescence. Curves (c) in Figure 7 show simulated spectra after adding electron-phonon interactions with $S_{phonon} = 0.8$. Note that experiments show that the effective electron-phonon coupling in this system is in fact quite heterogeneous, so the use of a single static Huang-Rhys factor is a significant approximation.[68,73] While this simplistic treatment reasonably describes the non-line-narrowed spectra considered here, a more rigorous treatment is required to handle line-narrowed spectra such as hole burning and fluorescence line narrowing.

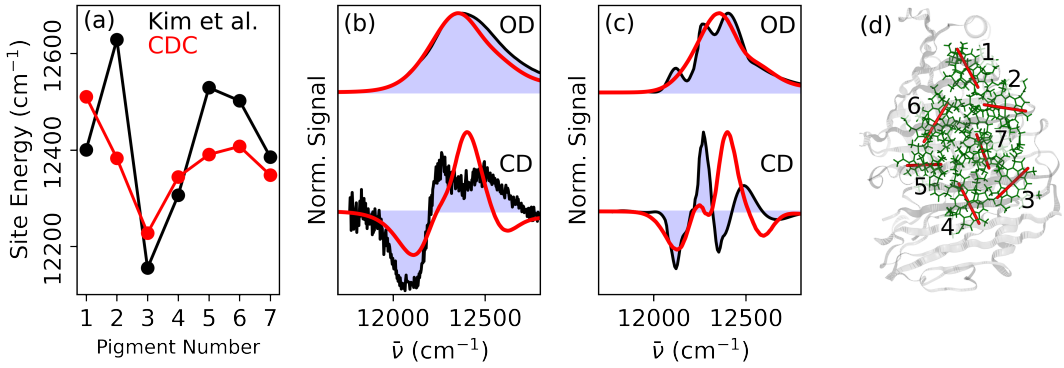

FIG. 8. The 3ENI[17,18] PDB structure of the FMO complex is used to demonstrate the effectiveness and limitations of the charge density coupling (CDC) model for predicting site energies. The CDC input FMO structure was collected after a 100 ps NVT trajectory. Frame (a) shows the comparison between CDC-calculated site energies in PigmentHunter (red) to values previously published (black) using QM/MM simulations.[14] Frames (b) and (c) show the comparison between PigmentHunter's CDC-calculated absorption (OD) and CD spectra (red lines) with the corresponding experimental spectra (blue-shaded curves) at 295 K (b) and 77 K (c). Frame (d) shows the numbering of sites in the 3ENI structure; the eighth pigment, which is situated between monomers, is omitted from all calculations.

B. Site energies: FMO

In the LH2 calculations just discussed, we adjusted the pigment site energies manually to achieve a reasonable description of the experimental absorption and fluorescence spectrum. This approach is fairly straightforward for a high-symmetry system like LH2 which contains only 3 distinct pigment-binding sites (repeated 9 times to produce a 27-pigment complex), so that only 3 site-energy parameters need to be adjusted.

Determining site energies is much more complex in low-symmetry systems such as the Fenna-Matthews-Olson (FMO) protein we consider next. This water-soluble antenna protein binds either 7 or 8 pigments depending on the preparation and *in vivo* plays a critical role in funneling energy from the peripheral antenna complexes to the bacterial reaction center.[74] Over the years, this protein has been extensively studied using a variety of both experimental and computational approaches, leading to a rough consensus on the site energies of each pigment in the complex.[14,21,23,75–77] For this reason, the FMO complex offers a useful reference system for benchmarking site-energy prediction methods; in this section, we use it to illustrate the utility and limitations of the CDC site-energy model implemented in PigmentHunter.

In PigmentHunter's Structure Repair tool, we selected a single FMO monomer from the

3ENI[17,18] crystal structure and eliminated the eighth pigment, which is sandwiched between two monomers and thus somewhat unstable in single-monomer MD simulations. We then ran a 100 ps NVT simulation using PigmentHunter's MD module, loaded the final NVT trajectory into the main window, and estimated average site energies across the trajectory using the CDC method.

The resulting site energy estimates are plotted in Figure 8A (red data points), in comparison to site energies calculated by Kim at al.[14] using more rigorous mixed Quantum/Classical methods (black data points), which agree rather well with the empirical site-energy estimates of Kell[76] and Brixner[75]. While the simpler CDC method misses several important features (e.g., the energetic ordering of sites 1 and 2), it does qualitatively capture capture several key features, most importantly the identity of the lowest-energy pigment at Site 3.

To give a sense for how the CDC-calculated site energies translate into experimental data, frames (b) and (c) compare experimental (blue-shaded curves) and CDC-calculated (red curves) absorption (OD) and circular dichroism (CD) spectra. Spectra in frame (b) are measured or calculated at 295 K, while spectra in frame (c) are at 77 K. Experimental data are taken from Ref.[14]. All calculated spectra include Gaussian disorder with a FWHM of 250 cm^{-1} (10^6 samples) and are convolved both with a 10 cm^{-1} Gaussian to eliminate sampling noise and with the experimentally determined vibrational and phonon profiles for the FMO protein. (These are PigmentHunter's default line shape parameters for BChl a.) A comparison of the four sets of curves in frames (b) and (c) highlights the utility of CD spectra, especially at low temperatures, for testing theoretical models: while the room-temperature absorption simulation is in nearly quantitative agreement with experiment, the CD spectrum reveals notable differences, especially in the structure of the positive lobe. Lowering the temperature to 77 K provides even greater sensitivity to the quality of the input parameters; although the calculated 77 K spectrum has qualitatively similarity to experiment, discrepancies are quite apparent.

C. Ring Distortion: PSI and PSII

While the CDC calculations in Figure 8 offer a useful look at how electrostatic forces affect (B)Chl spectra, the CDC method offers no information on other environmental forces, especially structural deformation of the pigment ring. While PigmentHunter does implement the NSD-based site-energy prediction method of Zucchelli et al., such calculations are quite sensitive to deformation along high-frequency normal modes that can be resolved experimentally only at very high

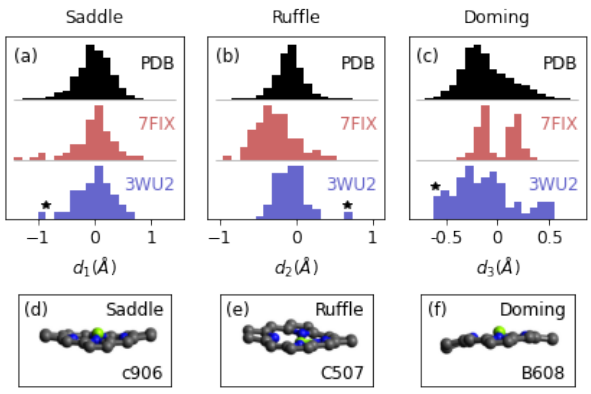

FIG. 9. Frames (a) - (c): NSD ring-deformation histograms for the saddle, ruffle and doming modes calculated for 94 Chl *a*-containing PDB structures at 2.5 Å resolution or better (black histograms), the 7FIX structure of PSI[78,79] (red), or the 3WU2 structure of PSII[80,81] (blue) Frames (d) - (f) illustrate selected PSII pigments that are strongly deformed along the three modes. Black stars in frames (a) - (c) indicate the deformation value for the pigments depicted in frames (d) - (f).

structural resolution. For this reason, we suggest that the NSD method should in most cases be used only to assess deformation along the lowest-frequency modes – the saddle, ruffle, and dome distortions – since these modes typically involve the largest real-space displacements.

Even without absolute site-energy predictions, however, NSD calculations offer a useful means of identifying strongly distorted pigments, especially when benchmarked against statistical data for many PPCs. PigmentHunter's automated processing tool is particularly helpful in generating such large data sets. To illustrate this capability, Figure 9 compares ring-deformation histograms for saddle, ruffle, and doming modes for the Photosystem I complex of *Thermosynechococcus vestitus BP-1* (PDB ID: 7FIX[78,79], 1.97 Å resolution, red histogram), the Photosystem II complex of Thermostichus vulcanus (PDB ID: 3WU2[80,81], 1.90 Å resolution, blue histogram) against all Chl *a* pigments contained in 94 distinct PDB entries returned by searching for the chemical name "Chlorophyll *a*" in the RCSB Advanced Search interface and restricting the results to only structures at 2.5 Å resolution or better. To illustrate how these numerical coefficients translate into real-space displacement, Frames (d) - (f) show the molecular structures of three pigments (c906, C507, and B608 in 3WU2[80,81] numbering) with strong displacements along one of the three modes; black stars in Frames (a) - (c) indicate the value of the corresponding deformation coefficient in the PSII histogram. Deformation coefficients for all 94 structures were calculated automatically using the Code Entry box in PigmentHunter's main window. (Explicit code is provided in the Supplementary Material.)

Comparison of the three sets of histograms reveals several interesting features. First, note that the overall shape of the histograms are very similar for the Saddle mode, slightly different for the Ruffle mode, and quite distinct for the Doming mode. The distinct bimodal distribution for doming in the 7FIX stands out in particular. Noteworthy also is the quite extreme ruffling distortion of site C507 (located in the CP43 subunit) of the 3WU2 PSII structure; comparison between the PSII and PDB histogram in frame (b) reveals that such strong ruffling is almost unique in PDB entries with resolution better than 2.5 Å . Given the small atomic displacements involved in such deformations (typically less than 1 Å), such features should be interpreted with caution. But given the rapid growth in available high-resolution structural data, however, statistical comparisons of this sort should be very useful in identifying sites with unusual deformations and which therefore merit further study in understanding the functional significance of steric strain in PPCs.

V. CONCLUSIONS AND OUTLOOK

In closing, we wish both to highlight PigmentHunter's existing capabilities as a practical tool for understanding PPCs and to note current limitations and potential areas for further development.

For immediate applications, PigmentHunter's strength is its ability to quickly process, parameterize, and compare spectroscopic predictions across multiple structures. For most PPCs, the TrEsp coupling method implemented in PigmentHunter provides a fast and realistic estimate of inter-pigment couplings and transition dipole moments, essential ingredients for understanding PPC excitonic structure and energy-transfer pathways. The automated NSD method offers a similarly efficient tool for quantifying structural deformation of individual pigments.

For spectroscopic simulations, PigmentHunter's Spectrum module provides a convenient means of visualizing how electronic couplings translate into experimental observables such as absorption, fluorescence, and CD spectra. The inclusion of phonon and vibrational dressing (using experimentally derived densities) is a rather unique feature of PigmentHunter's capabilities and is particularly useful for understanding the temperature dependence of measured spectra. To be sure, PigmentHunter's treatment of phonon and vibrational coupling is quite simplistic, convolving all electronic transitions with a single line shape. For many applications, however, even this simple treatment offers quite useful insight into the role phonons and local vibrations play in determining spectroscopic structure.

Perhaps PigmentHunter's most serious limitation is in the estimation of site energies, i.e., the

local electronic transition frequencies of individual pigments. For high-symmetry systems such as LH2 (see Figure 7), realistic site-energy estimates can often be inferred from experimental spectra. For low symmetry systems, however, (especially those with many pigments) such parameterization is much more difficult. The CDC method implemented in PigmentHunter has shown remarkable success in quantifying site-energy shifts in site-directed mutants, and (as illustrated in Figure 8 for the FMO protein) offers at least qualitative insight into the site energies of native systems. A distinct advantage of the CDC approach is its computational stability: whereas a poorly optimized electronic structure calculation can produce wildly inaccurate site energy predictions, the grounding of CDC calculations in a fixed set of atomic partial charges avoids much of this instability. However, agreement between CDC calculations and experiment is by no means perfect (see Figure 8(b) and (c)), and further testing is required to establish the limits within which it can be rigorously relied upon. Although PigmentHunter does include a NSD-based site energy model to capture steric effects (to which the CDC method is blind), this approach should for the present be considered only suggestive until it can be further validated in rigorous testing against experiment.

PigmentHunter's MD simulation module highlights both the application's power and limitations. On the one hand, the automation of structure repair, hydrogen addition, and pigment topology assignment dramatically simplifies the preparation of MD-ready GROMACS[40] inputs. On the other hand, PigmentHunter's in-application MD computational power is limited to only 100 ps of position-restrained, solvent-free NVT dynamics. While even such short MD runs are quite useful in assessing local electrostatic effects (as in the CDC site-energy method), their utility is obviously limited for assessing large-scale fluctuations away from the crystal structure (especially when mutations are inserted) or understanding solvation effects. Fortunately, these limitations should be straightforward to eliminate as computational power becomes increasingly cheap. Efforts to extend PigmentHunter's MD capacity to longer runs with explicit solvent are already underway, with substantial progress anticipated within the next several years.

Beyond computing power, the other key limitation in PigmentHunter's MD capability is in the handling of non-protein ligands. Currently, MD parameters are available in PigmentHunter for only two pigments (Chl *a* and BChl *a*), so that structures containing even relatively common pigments such as Chl *b* and Pheo *a* cannot be processed through the MD module. More broadly, photosynthetic proteins often contain a wide variety of non-protein ligands such as carotenoids, quinones, hemes, lipids, and metal clusters; most of these molecules lack standard GROMACS[40]

topologies (or PigmentHunter add-ins) and so must be "stripped" from the structure before running MD.

The good news is that the "missing parameters" challenge should be straightforward to solve as force-field parameters for non-protein ligands become standardized. We are currently working to develop a standardized set of pigment parameters so that at minimum all common Chl-like pigments can be included in PigmentHunter MD simulations. (Where multiple force fields are available for a single pigment, users will be given a choice of parameters.) Non-pigment ligands will pose more of a challenge, since an exhaustive library of pre-tested MD parameters for all possible ligands is not feasible. Instead, the most plausible solution is likely to include pre-tested parameters for common ligands whenever possible and to allow users to upload custom parameters (e.g., from automated tools such as CGenFF) for uncommon ligands.

While these MD-oriented challenges are primarily technical, more fundamental questions surround the translation of MD data into spectroscopic signals. The accurate prediction of spectroscopic parameters (inter-pigment couplings, site energies, electronic-vibrational interaction coefficients) from structural data remains an open scientific challenge, as reflected by persistent debate in the literature regarding the composition and character of excitonic states in many PPCs. Our vision for PigmentHunter is not to "solve" all of these challenges directly but to provide a common, high-throughput, user-friendly interface through which theoretical and computational approaches can be tested and refined against real spectroscopic data. To this end, we intend to continue expanding PigmentHunter's repertoire of available theoretical and computational methods including

- Geometric "look-up" tables for electronic coupling parameters of tightly coupled pigments where charge-transfer states may be important.
- More rigorous line shape and dynamics methods grounded in, for example, the Redfield and Förster limits, the hierarchical equations of motion (HEOM),[82] or path integral methods.[83]
- Additional spectroscopic measurements such as linear dichroism, hole burning, fluorescence line narrowing, and pump-probe and 2D electronic spectroscopy.[12,75,84]

As spectroscopic models for other pigments (e.g., bilins and carotenoids) become available, we hope also to either implement these directly or to add an interface through which users can upload their own parameterized models.

While we will continue both to implement and develop such methods in-house, PigmentHunter's modular structure is also well-suited to the development by other research groups of

new modules that can be ported into the program and deployed online either as a plug-in to the main PigmentHunter application or as stand-alone tools on nanoHUB. (PigmentHunter is published as an open-source package, and developers are free to use, develop, and re-publish the code without restriction other than citing the original source.)

Most importantly, we hope that PigmentHunter's combination of a graphical point-and-click interface with its native recognition and automated processing of photosynthetic pigments and its pre-installed access to nanoHUB's high-performance-computing backend will help to demystify photosynthetic excitons for a new generation of researchers. Photosynthesis research is a richly interdisciplinary field, encompassing a diverse cohort of researchers in molecular and structural biology, plant and bacterial genetics, chemical theory and computation, and both time- and frequency-domain spectroscopies. Building a common vocabulary and conceptual framework for such a wide-ranging field is fundamentally challenging, particularly when it comes to specialized concepts such as exciton formation and vibrational-electronic mixing. Our hope is that the PigmentHunter interface will lower the barriers between photosynthetic theory, spectroscopy, structure, and genetics, facilitating fruitful "cross-pollination" in this exciting field.

VI. ACKNOWLEDGEMENTS

The authors thank Arvi Freiberg (University of Tartu) for kindly providing the experimental spectra shown in Figure 7. We thank Sergei Savikhin and Lyudmila Slipchenko (Purdue University) for providing the experimental spectra in Figure 8. And we thank Thomas Renger (Johannes Kepler University Linz) for ground and excited-state partial charges for Chl *b* used in the CDC calculations implemented in PigmentHunter. We thank Danilo Roccatano for providing GROMACS-format molecular topologies for Chl *a* and Lyudmila Slipchenko for GROMACS-format topologies for BChl *a*. We also thank nanoHUB.org and Steven Clark (San Diego Supercomputer Center) for technical support in implementing the PigmentHunter online. This research was supported by the U.S. Department of Energy, Office of Science, Basic Energy Sciences, under award #DE-SC0022884 and by startup funding from Purdue University.

VII. SUPPLEMENTARY MATERIAL

Online Supplementary Material includes instructions for local installation of the PigmentHunter tool, a description of PigmentHunter file outputs and storage locations, details on spectroscopic line shape calculations, and Python code for automated generation of the NSD data presented in Figure 9.

VIII. DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available in the article and its supplementary material. Source code for this version of the PigmentHunter app is openly available from nanoHUB.org at https://doi.org/10.21981/744F-DE56.

REFERENCES

[1]G. R. Fleming, G. S. Schlau-Cohen, K. Amarnath, and J. Zaks, "Design principles of photosynthetic light-harvesting," Faraday Discuss. **155**, 27–41 (2012).

[2]E. Romero, V. I. Novoderezhkin, and R. van Grondelle, "Quantum design of photosynthesis for bio-inspired solar-energy conversion," Nature **543**, 355–365 (2017).

[3]T. Reinot, J. Chen, A. Kell, M. Jassas, K. C. Robben, V. Zazubovich, and R. Jankowiak, "On the conflicting estimations of pigment site energies in photosynthetic complexes: A case study of the cp47 complex," Anal. Chem. Insights **11**, 35–48 (2016), aci-11-2016-035[PII].

[4]A. Melis, "Solar energy conversion efficiencies in photosynthesis: Minimizing the chlorophyll antennae to maximize efficiency," Plant Sci. **177**, 272–280 (2009).

[5]R. E. Blankenship, D. M. Tiede, J. Barber, G. W. Brudvig, G. Fleming, M. Ghirardi, M. R. Gunner, W. Junge, D. M. Kramer, A. Melis, T. A. Moore, C. C. Moser, D. G. Nocera, A. J. Nozik, D. R. Ort, W. W. Parson, R. C. Prince, and R. T. Sayre, "Comparing photosynthetic and photovoltaic efficiencies and recognizing the potential for improvement," Science **332**, 805–809 (2011), http://science.sciencemag.org/content/332/6031/805.full.pdf.

[6]L. Wobbe and C. Remacle, "Improving the sunlight-to-biomass conversion efficiency in microalgal biofactories," J. Biotech. **201**, 28–42 (2015), molecular Biotechnology: from enzymes and metabolically engineered microbes to superior and sustainable products and processes.

[7]N. Friedland, S. Negi, T. Vinogradova-Shah, G. Wu, L. Ma, S. Flynn, T. Kumssa, C.-H. Lee, and R. T. Sayre, "Fine-tuning the photosynthetic light harvesting apparatus for improved photosynthetic efficiency and biomass yield," Sci. Rep. **9**, 13028 (2019).

[8]R. Croce and H. van Amerongen, "Light harvesting in oxygenic photosynthesis: Structural biology meets spectroscopy," Science **369** (2020), 10.1126/science.aay2058, https://science.sciencemag.org/content/369/6506/eaay2058.full.pdf.

[9]M. Reppert, V. Naibo, and R. Jankowiak, "Analytical formulas for low-fluence non-line-narrowed hole-burned spectra in an excitonically coupled dimer," J. Chem. Phys. **131**, 234104 (2009), https://doi.org/10.1063/1.3270394.

[10]M. Reppert, V. Naibo, and R. Jankowiak, "Modeling study of non-line-narrowed hole-burned spectra in weakly coupled dimers and multi-chromophoric molecular assemblies," Chem. Phys. **367**, 27–35 (2010).

[11]M. Reppert, K. Acharya, B. Neupane, and R. Jankowiak, "Lowest electronic states of the cp47 antenna protein complex of photosystem ii: Simulation of optical spectra and revised structural assignments," J. Phys. Chem. B **114**, 11884–11898 (2010).

[12]R. Jankowiak, M. Reppert, V. Zazubovich, J. Pieper, and T. Reinot, "Site selective and single complex laser-based spectroscopies: A window on excited state electronic structure, excitation energy transfer, and electron-phonon coupling of selected photosynthetic complexes," Chem. Rev. **111**, 4546–4598 (2011).

[13]F. Müh and A. Zouni, "Structural basis of light-harvesting in the photosystem ii core complex," Protein Science **29**, 1090–1119 (2020), https://onlinelibrary.wiley.com/doi/pdf/10.1002/pro.3841.

[14]Y. Kim, D. Morozov, V. Stadnytskyi, S. Savikhin, and L. V. Slipchenko, "Predictive first-principles modeling of a photosynthetic antenna protein: The fenna–matthews–olson complex," The Journal of Physical Chemistry Letters **11**, 1636–1643 (2020).

[15]M. Reppert, "Bioexcitons by design: How do we get there?" The Journal of Physical Chemistry B **127**, 1872–1879 (2023).

[16]T. Renger and F. Müh, "Understanding photosynthetic light-harvesting: a bottom up theoretical approach," Phys. Chem. Chem. Phys. **15**, 3348–3371 (2013).

[17]D. Tronrud, A. Camara-Artigas, R. Blankenship, and J. Allen, "Crystal structure of the fenna-matthews-olson protein from chlorobaculum tepidum," (2009).

[18]D. E. Tronrud, J. Wen, L. Gay, and R. E. Blankenship, "The structural basis for the difference in absorbance spectra for the fmo antenna protein from various green sulfur bacteria," Photosynthesis Research **100**, 79–87 (2009).

[19]G. Zucchelli, D. Brogioli, A. P. Casazza, F. M. Garlaschi, and R. C. Jennings, "Chlorophyll ring deformation modulates qy electronic energy in chlorophyll-protein complexes and generates spectral forms," Biophysical Journal **93**, 2240–2254 (2007).

[20]G. Zucchelli, S. Santabarbara, and R. C. Jennings, "The qy absorption spectrum of the light-harvesting complex ii as determined by structure-based analysis of chlorophyll macrocycle deformations," Biochem. **51**, 2717–2736 (2012).

[21]J. Adolphs, F. Müh, M. E.-A. Madjet, and T. Renger, "Calculation of pigment transition energies in the fmo protein," Photosynthesis Research **95**, 197–209 (2008).

[22]M. E. Madjet, A. Abdurahman, and T. Renger, "Intermolecular coulomb couplings from ab initio electrostatic potentials: Application to optical transitions of strongly coupled pigments in photosynthetic antennae and reaction centers," The Journal of Physical Chemistry B **110**, 17268–17281 (2006).

[23]M. Higashi and S. Saito, "Quantitative evaluation of site energies and their fluctuations of pigments in the fenna-matthews-olson complex with an efficient method for generating a potential energy surface," J. Chem. Theor. Comput. **12**, 4128–4137 (2016).

[24]S. Saito, M. Higashi, and G. R. Fleming, "Site-dependent fluctuations optimize electronic energy transfer in the fenna-matthews-olson protein," J. Phys. Chem. B **123**, 9762–9772 (2019).

[25]C. Curutchet and B. Mennucci, "Quantum chemical studies of light harvesting," Chemical Reviews **117**, 294–343 (2017).

[26]X. Jia, Y. Mei, J. Z. Zhang, and Y. Mo, "Hybrid qm/mm study of fmo complex with polarized protein-specific charge," Scientific Reports **5**, 17096 (2015).

[27]S. Shim, P. Rebentrost, S. Valleau, and A. Aspuru-Guzik, "Atomistic study of the long-lived quantum coherences in the fenna-matthews-olson complex," Biophysical Journal **102**, 649–660 (2012).

[28]S. Chandrasekaran, M. Aghtar, S. Valleau, A. Aspuru-Guzik, and U. Kleinekathöfer, "Influence of force fields and quantum chemistry approach on spectral densities of bchl a in solution and in fmo proteins," The Journal of Physical Chemistry B **119**, 9995–10004 (2015).

[29]A. Sirohiwal, F. Neese, and D. A. Pantazis, "How can we predict accurate electrochromic shifts for biochromophores? a case study on the photosynthetic reaction cen-

ter," Journal of Chemical Theory and Computation **17**, 1858–1873 (2021), pMID: 33566610, https://doi.org/10.1021/acs.jctc.0c01152.

[30]A. M. Rosnik and C. Curutchet, "Theoretical characterization of the spectral density of the water-soluble chlorophyll-binding protein from combined quantum mechanics/molecular mechanics molecular dynamics simulations," Journal of Chemical Theory and Computation **11**, 5826–5837 (2015).

[31]Y. Lahav, D. Noy, and I. Schapiro, "Spectral tuning of chlorophylls in proteins – electrostatics vs. ring deformation," Phys. Chem. Chem. Phys. **23**, 6544–6551 (2021).

[32]C. Curutchet, J. Kongsted, A. Muñoz-Losa, H. Hossein-Nejad, G. D. Scholes, and B. Mennucci, "Photosynthetic light-harvesting is tuned by the heterogeneous polarizable environment of the protein," Journal of the American Chemical Society **133**, 3078–3084 (2011).

[33]Y. Kim, Z. Mitchell, J. Lawrence, D. Morozov, S. Savikhin, and L. V. Slipchenko, "Predicting mutation-induced changes in the electronic properties of photosynthetic proteins from first principles: The fenna–matthews–olson complex example," The Journal of Physical Chemistry Letters **14**, 7038–7044 (2023).

[34]A. Srivastava, S. Ahad, J. H. Wat, and M. Reppert, "Accurate prediction of mutation-induced frequency shifts in chlorophyll proteins with a simple electrostatic model," The Journal of Chemical Physics **155**, 151102 (2021).

[35]S. Ahad, C. Lin, and M. E. Reppert, "Photosynthetic protein spectroscopy lab," (2020).

[36]K. Madhavan, L. Zentner, V. Farnsworth, S. Shivarajapura, M. Zentner, N. Denny, and G. Klimeck, "nanoHUB.org: cloud-based services for nanoscale modeling, simulation, and education," Nanotechnology Reviews **2**, 107–117 (2013).

[37]M. McLennan and R. Kennell, "Hubzero: A platform for dissemination and collaboration in computational science and engineering," Computing in Science & Engineering **12**, 48–53 (2010).

[38]H. Nguyen, D. A. Case, and A. S. Rose, "NGLview–interactive molecular graphics for jupyter notebooks," Bioinformatics **34**, 1241–1242 (2017).

[39]J. D. Team, *IPywidgets: Interactive HTML Widgets for Jupyter* (2018).

[40]J. Lemkul, "From proteins to perturbed hamiltonians: A suite of tutorials for the GROMACS-2018 molecular simulation package [article v1.0]," Living Journal of Computational Molecular Science **1** (2019), 10.33011/livecoms.1.1.5068.

[41]R. W. Amaro, C. O. H. Nakasone, M. M. Lerner, and A. E. Roitberg, "Parmed: Parameterization and equilibration of complex molecular force fields," Journal of Chemical Theory and Computation **12**, 3526–3530 (2016).

[42]W. L. DeLano, "Pymol: An open-source molecular graphics tool," CCP4 Newsletter on protein crystallography **40**, 82–92 (2002).

[43]D. Horigome, H. Satoh, N. Itoh, K. Mitsunaga, I. Oonishi, A. Nakagawa, and A. Uchida, "Crystal structure of water-soluble chlorophyll protein from lepidium virginicum at 2.00 angstrom resolution," (2006).

[44]D. Horigome, H. Satoh, N. Itoh, K. Mitsunaga, I. Oonishi, A. Nakagawa, and A. Uchida, "Structural mechanism and photoprotective function of water-soluble chlorophyll-binding protein," J. Biol. Chem. **282**, 6525–6531 (2007), http://www.jbc.org/content/282/9/6525.full.pdf+html.

[45]J. Adolphs and T. Renger, "How proteins trigger excitation energy transfer in the FMO complex of green sulfur bacteria," Biophysical Journal **91**, 2778–2797 (2006).

[46]W. Jentzen, X.-Z. Song, and J. A. Shelnutt, "Structural characterization of synthetic and protein-bound porphyrins in terms of the lowest-frequency normal coordinates of the macrocycle," The Journal of Physical Chemistry B **101**, 1684–1699 (1997).

[47]C. Lin, Y. Mazor, and M. Reppert, "Feeling the strain: Quantifying ligand deformation in photosynthesis," The Journal of Physical Chemistry B **0**, null (0), pMID: 38442033, https://doi.org/10.1021/acs.jpcb.3c06488.

[48]A. S. Sardjan, F. P. Westerman, J. P. Ogilvie, and T. L. C. Jansen, "Observation of ultrafast coherence transfer and degenerate states with polarization-controlled two-dimensional electronic spectroscopy," The Journal of Physical Chemistry B **124**, 9420–9427 (2020), pMID: 32990439, https://doi.org/10.1021/acs.jpcb.0c08126.

[49]C. Kreisbeck and T. Kramer, "Exciton dynamics lab for light-harvesting complexes (gpuheom)," (2013).

[50]M. Ceccarelli, P. Procacci, and M. Marchi, "An ab initio force field for the cofactors of bacterial photosynthesis," Journal of Computational Chemistry **24**, 129–142 (2003), https://onlinelibrary.wiley.com/doi/pdf/10.1002/jcc.10198.

[51]K. Karki and D. Roccatano, "Molecular dynamics simulation study of chlorophyll a in different organic solvents," Journal of Chemical Theory and Computation **7**, 1131–1140 (2011), pMID: 26606361, https://doi.org/10.1021/ct1004627.

[52]T. Renger and F. Müh, "Theory of excitonic couplings in dielectric media," Photosynthesis Research **111**, 47–52 (2011).

[53]T. Renger, I. Trostmann, C. Theiss, M. E. Madjet, M. Richter, H. Paulsen, H. J. Eichler, A. Knorr, and G. Renger, "Refinement of a structural model of a pigment-protein complex by accurate optical line shape theory and experiments," The Journal of Physical Chemistry B **111**, 10487–10501 (2007).

[54]G. Raszewski, B. A. Diner, E. Schlodder, and T. Renger, "Spectroscopic properties of reaction center pigments in photosystem II core complexes: Revision of the multimer model," Biophysical Journal **95**, 105–119 (2008).

[55]H. van Amerongen, R. van Grondelle, and L. Valkunas, *Photosynthetic Excitons* (WORLD SCIENTIFIC, 2000).

[56]V. May and O. Kühn, *Charge and Energy Transfer Dynamics in Molecular Systems* (Wiley, 2011).

[57]G. D. Scholes and K. P. Ghiggino, "Electronic interactions and interchromophore excitation transfer," The Journal of Physical Chemistry **98**, 4580–4590 (1994).

[58]C.-P. Hsu, "The electronic couplings in electron transfer and excitation energy transfer," Accounts of Chemical Research **42**, 509–518 (2009), https://doi.org/10.1021/ar800153f.

[59]R. S. Knox and B. Q. Spring, "Dipole strengths in the chlorophylls ¶†," Photochemistry and Photobiology **77**, 497–501 (2007).

[60]M. Reppert, V. Zazubovich, N. C. Dang, M. Seibert, and R. Jankowiak, "Low-energy chlorophyll states in the cp43 antenna protein complex: Simulation of various optical spectra. ii," J. Phys. Chem. B **112**, 9934–9947 (2008).

[61]S. Krawczyk, "The effects of hydrogen bonding and coordination interaction in visible absorption and vibrational spectra of chlorophyll a," Biochim. Biophys. Acta (BBA) - Bioenergetics **976**, 140–149 (1989).

[62]F. Müh, M. Plöckinger, and T. Renger, "Electrostatic asymmetry in the reaction center of photosystem ii," The Journal of Physical Chemistry Letters **8**, 850–858 (2017), pMID: 28151674, https://doi.org/10.1021/acs.jpclett.6b02823.

[63]A. Sirohiwal, F. Neese, and D. A. Pantazis, "Protein matrix control of reaction center excitation in photosystem ii," Journal of the American Chemical Society **142**, 18174–18190 (2020), pMID: 33034453, https://doi.org/10.1021/jacs.0c08526.

[64]T. Renger and R. A. Marcus, "On the relation of protein dynamics and exciton relaxation in pigment–protein complexes: An estimation of the spectral density and a theory for the calculation of optical spectra," The Journal of Chemical Physics **116**, 9997–10019 (2002).

[65]M. Reppert, "Modeling of resonant hole-burning spectra in excitonically coupled systems: The effects of energy-transfer broadening," J. Phys. Chem. Lett. **2**, 2716–2721 (2011).

[66]M. Reppert, "Delocalization effects in chlorophyll fluorescence: Nonperturbative line shape analysis of a vibronically coupled dimer," The Journal of Physical Chemistry B **124**, 10024–10033 (2020).

[67]C. Friedl, D. G. Fedorov, and T. Renger, "Towards a quantitative description of excitonic couplings in photosynthetic pigment–protein complexes: quantum chemistry driven multiscale approaches," Physical Chemistry Chemical Physics **24**, 5014–5038 (2022).

[68]A. Freiberg, M. Rätsep, K. Timpmann, G. Trinkunas, and N. W. Woodbury, "Self-trapped excitons in lh2 antenna complexes between 5 k and ambient temperature," The Journal of Physical Chemistry B **107**, 11510–11519 (2003).

[69]P. Qian, D. Swainsbury, T. Croll, P. Castro-Hartmann, K. Sader, G. Divitini, and C. Hunter, "Cryo-em structure of light harvesting complex 2 from rba. sphaeroides." (2021).

[70]P. Qian, D. J. K. Swainsbury, T. I. Croll, P. Castro-Hartmann, G. Divitini, K. Sader, and C. N. Hunter, "Cryo-em structure of the rhodobacter sphaeroides light-harvesting 2 complex at 2.1 Å," Biochemistry **60**, 3302–3314 (2021).

[71]L. Cupellini, S. Caprasecca, C. A. Guido, F. Müh, T. Renger, and B. Mennucci, "Coupling to charge transfer states is the key to modulate the optical bands for efficient light harvesting in purple bacteria," The Journal of Physical Chemistry Letters **9**, 6892–6899 (2018), https://doi.org/10.1021/acs.jpclett.8b03233.

[72]B. P. Krueger, G. D. Scholes, and G. R. Fleming, "Calculation of couplings and energy-transfer pathways between the pigments of lh2 by the ab initio transition density cube method," The Journal of Physical Chemistry B **102**, 5378–5386 (1998), https://doi.org/10.1021/jp9811171.

[73]K. Timpmann, M. Rätsep, C. N. Hunter, and A. Freiberg, "Emitting excitonic polaron states in core LH1 and peripheral LH2 bacterial light-harvesting complexes," J. Phys. Chem. B **108**, 10581–10588 (2004).

[74]D. E. Tronrud, J. Wen, L. Gay, and R. E. Blankenship, "The structural basis for the difference in absorbance spectra for the fmo antenna protein from various green sulfur bacteria," Photosynthesis Research **100**, 79–87 (2009).

[75]T. Brixner, J. Stenger, H. M. Vaswani, M. Cho, R. E. Blankenship, and G. R. Fleming, "Two-dimensional spectroscopy of electronic couplings in photosynthesis," Nature **434**, 625–628 (2005).

[76]A. Kell, R. E. Blankenship, and R. Jankowiak, "Effect of spectral density shapes on the excitonic structure and dynamics of the fenna–matthews–olson trimer from chlorobaculum tepidum," The Journal of Physical Chemistry A **120**, 6146–6154 (2016), pMID: 27438068, https://doi.org/10.1021/acs.jpca.6b03107.

[77]R. G. Saer, V. Stadnytskyi, N. C. Magdaong, C. Goodson, S. Savikhin, and R. E. Blankenship, "Probing the excitonic landscape of the chlorobaculum tepidum fenna-matthews-olson (fmo) complex: a mutagenesis approach," Biochim. Biophys. Acta (BBA) - Bioenergetics **1858**, 288–296 (2017).

[78](2021).

[79]J. Li, N. Hamaoka, F. Makino, A. Kawamoto, Y. Lin, M. Rögner, M. M. Nowaczyk, Y.-H. Lee, K. Namba, C. Gerle, and G. Kurisu, "Structure of cyanobacterial photosystem I complexed with ferredoxin at 1.97 å resolution," Commun. Biol. **5**, 951 (2022).

[80]Y. Umena, K. Kawakami, J.-R. Shen, and N. Kamiya, "Crystal structure of oxygen-evolving photosystem ii at a resolution of 1.9 å," Nature **473**, 55–60 (2011).

[81]Y. Umena, K. Kawakami, J. R. Shen, and N. Kamiya, "Crystal structure analysis of photosystem II complex," (2014).

[82]Y. Tanimura, "Numerically "exact" approach to open quantum dynamics: The hierarchical equations of motion (HEOM)," The Journal of Chemical Physics **153**, 020901 (2020), https://pubs.aip.org/aip/jcp/article-pdf/doi/10.1063/5.0011599/15575870/020901_1_online.pdf.

[83]A. Bose and N. Makri, "All-mode quantum–classical path integral simulation of bacteriochlorophyll dimer exciton-vibration dynamics," The Journal of Physical Chemistry B **124**, 5028–5038 (2020), pMID: 32496772, https://doi.org/10.1021/acs.jpcb.0c03032.

[84]J. Pieper, P. Artene, M. Rätsep, M. Pajusalu, and A. Freiberg, "Evaluation of electron–phonon coupling and spectral densities of pigment–protein complexes by line-narrowed optical spectroscopy," J. Phys. Chem. B **122**, 9289–9301 (2018), pMID: 30125502, https://doi.org/10.1021/acs.jpcb.8b05220.