Factorial knockouts diagnose missing physical structure in a calibrated PSI–LHCI transport surrogate

Hai Zhang Boyi Feng Huan Tan Yuqi Wang Haoming Luo

Abstract

Site-energy landscapes and pigment–pigment couplings jointly shape excitation trapping in photosystem I–light-harvesting complex I (PSI–LHCI), but randomization tests do not isolate their contributions. We applied a 2×2 factorial knockout to a calibrated, 155-pigment PSI–LHCI transport surrogate. Energy-KO flattened all site energies. Coupling-KO replaced every occupied edge by one root-mean-square coupling while preserving the 409-edge topology and total $\sum J_{ij}^2$. Double-KO applied both operations. All conditions shared one Full-model calibration and were not refitted. At the calibrated point, Full yielded 0.9650 with a conditional charge-separation mean first-passage time of 34.99 ps. Energy-KO, Coupling-KO, and Double-KO instead produced yields of 0.9735, 0.9702, and 0.9735 and first-passage times of 25.15, 29.24, and 25.15 ps, respectively. Strictly setting all couplings to zero abolished reaction-center access, whereas removing only the reaction-center energetic offset worsened trapping. Input and sensitivity audits resolved the apparent contradiction: all 23 lowest-energy sites were peripheral LHCI pigments, and matching the Full yield required a global coupling ceiling of 32825 cm^{-1}, far beyond ordinary chlorophyll coupling scales. At 120 cm^{-1}, Energy-KO also exposed a yield–speed tradeoff. We therefore do not infer that real PSI favors flat energies or uniform couplings. Instead, factorial knockouts reveal that this coarse energy map and extreme coupling calibration cannot support native-design claims. Controlled ablation is valuable here as a falsification-oriented model test: coupling is necessary, an RC energy offset is helpful, and the remaining counterintuitive responses identify missing physical structure.

Significance. Mechanistic models can reproduce an endpoint for the wrong reason. A useful model should also respond plausibly when a proposed design feature is removed. By separating energy heterogeneity from coupling-strength heterogeneity, we show that a calibrated PSI–LHCI surrogate fails this intervention test. The failure localizes what must be repaired—pigment-resolved energies, physically constrained couplings, and multi-observable calibration—before simulations can distinguish native organization from a numerical compensation.

1 Introduction

Photosystem I (PSI) couples broadband light harvesting to rapid charge separation with high quantum efficiency. Its antenna is neither spatially nor spectrally uniform. High-resolution structures resolve a pigment-rich core surrounded by light-harvesting complex I (LHCI), with specific interfacial chlorophylls linking the antenna belt to the core [1, 2]. Site energies and electronic couplings provide the two most immediate physical coordinates for explaining transport on this architecture: energy differences create thermal biases, while couplings set pairwise communication.

Low-energy chlorophyll states make the functional interpretation nontrivial. Far-red states extend absorption, but escape from sufficiently red pigments can become thermally activated and slow trapping [3, 4]. Their effect depends on location. Structure-based calculations suggest that some low-energy states near antenna–core interfaces can aid transfer [5], whereas spectroscopy assigns the deepest Lhca states partly to exciton–charge-transfer mixing rather than to isolated scalar site energies [6, 7]. Energy organization can therefore be useful, costly, or multifunctional depending on which pigments carry it and which observable defines performance.

Couplings are similarly multiscale. They are indispensable—without nonzero pair couplings, excitation cannot migrate—but that fact does not establish that the native placement of stronger and weaker couplings optimizes terminal trapping. Structure-based network models translate pigment coordinates into transport and have generated important hypotheses about routes and timescales [8, 9]. Yet a calibrated endpoint can remain insensitive to errors in the Hamiltonian or bath description. Systems-biology models frequently contain sloppy parameter combinations [10]; if discrepancy is absorbed into one global scale, a fit need not validate the mechanism [11].

This distinction motivates a direct intervention. Randomly permuting energies or couplings retains the full distribution and asks whether its placement is special. It does not ask how performance changes when the associated organization is removed. We instead construct a 2×2 factorial knockout that independently abolishes site-energy heterogeneity and coupling-strength heterogeneity, while retaining a separate $J = 0$ necessity control. We evaluate terminal charge-separation yield, reaction-center access, and conditional first-passage time without recalibrating the knockout conditions.

The result is initially counterintuitive: all organization knockouts outperform the Full reference at the calibrated point. Rather than interpreting this as a flat-energy or uniform-coupling design principle, we audit the energy input, global coupling scale, and observable dependence. The analysis shows that the knockout is a successful *model test* precisely because the current parametrization fails it.

2 Results

2.1 A factorial design separates existence from organization

The reference graph contained 155 pigments, 409 undirected edges, and two designated reaction-center (RC) sites. Four conditions formed the factorial experiment (fig. 1). Full used the supplied site energies and edge couplings. Energy-KO set every site to the network-mean energy while retaining reference couplings. Coupling-KO assigned one RMS magnitude to every occupied edge, preserving topology and $\sum_{(i,j)\in\mathcal{E}} J_{ij}^2$, while retaining reference energies. Double-KO combined both interventions. Temperature, initial population, RC identity, charge-separation rate, dephasing, and loss channels were identical. A single global coupling scale was fitted to Full and then frozen.

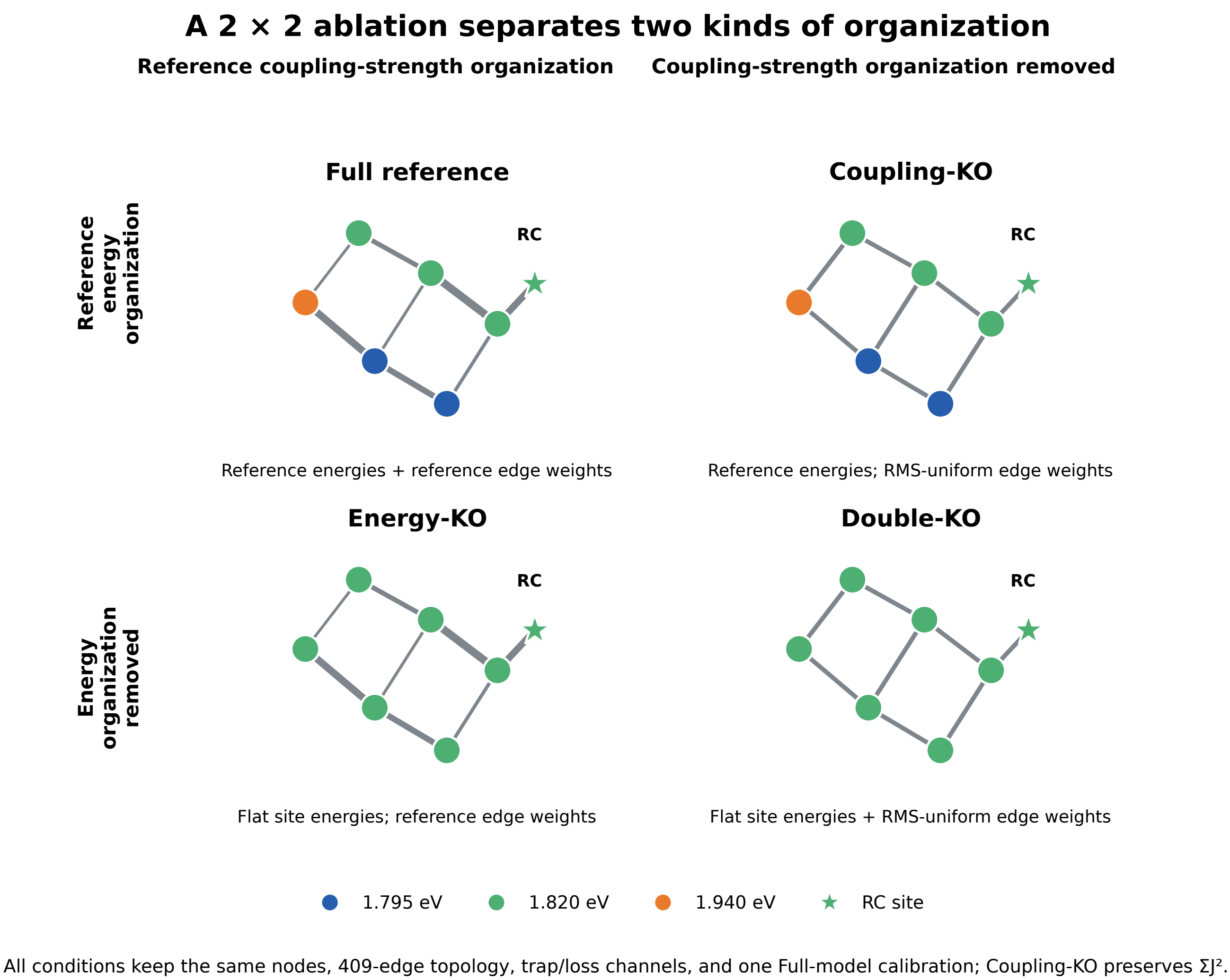


Figure 1: **The 2×2 organization knockout.** Energy-KO removes site-energy heterogeneity. Coupling-KO removes the allocation of strong and weak edge weights while preserving the graph and total coupling power. Double-KO removes both. The strict zero-coupling control is external to the factorial because it asks whether transport exists, not whether strengths are organized.

This distinction changes the estimand. Strict $J = 0$ asks whether inter-pigment transfer requires coupling; Coupling-KO asks whether the reference strength pattern adds value at fixed topology and aggregate scale. Energy-KO removes energy differences altogether; an energy permutation would retain them. No condition is claimed to be a realizable mutant.

2.2 The reference input is not an RC-directed energy funnel

The supplied energy map contained only three classes (fig. 2). All 23 pigments in the lowest class, 1.795 eV, were peripheral LHCI sites. The two RC sites were assigned to the middle 1.820 eV class, together with most core and interface pigments. Twelve high-energy LHCI sites occupied 1.940 eV. Thus, although the rate law penalized uphill motion and obeyed finite-temperature detailed balance, its lowest-energy bias pointed toward peripheral sites rather than exclusively toward the RC.

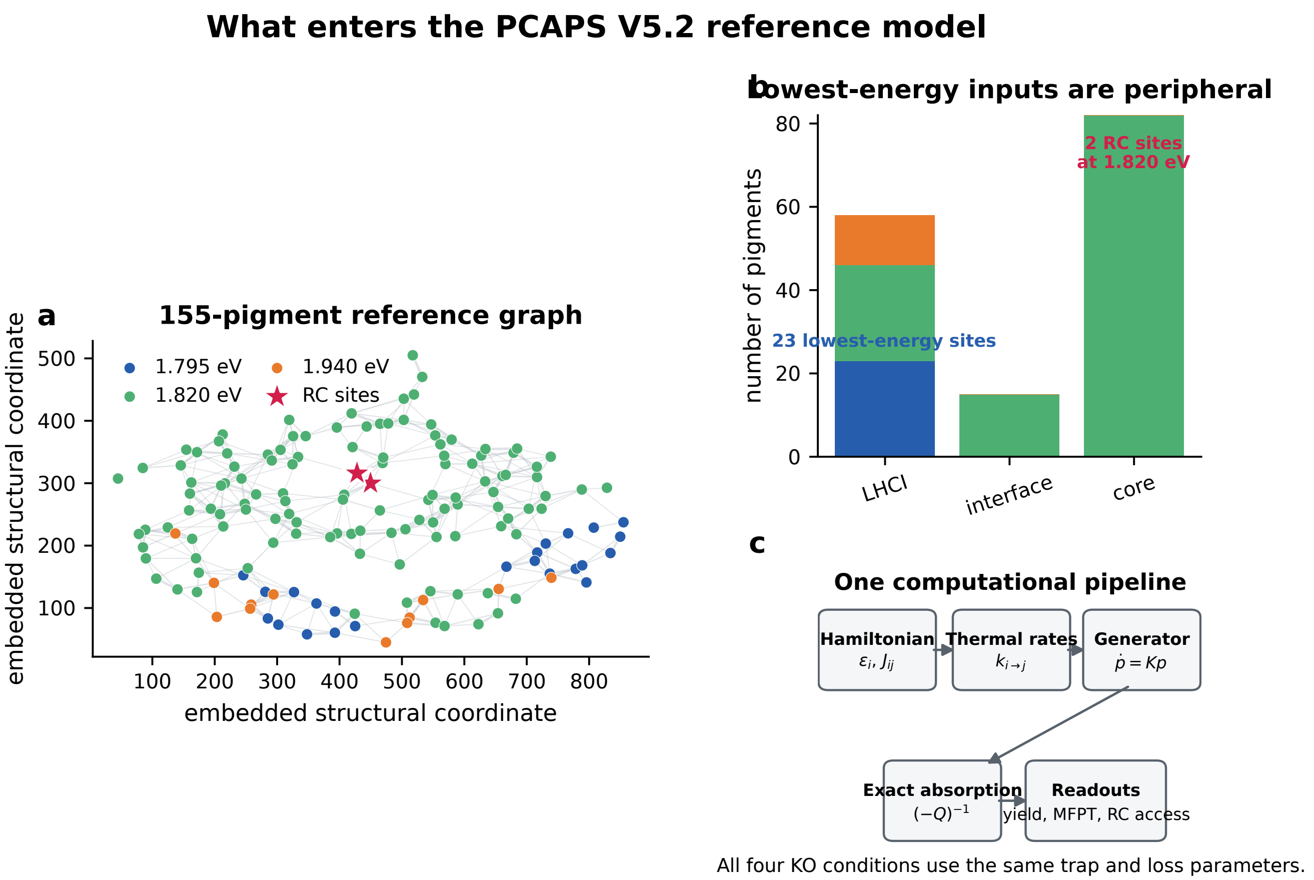


Figure 2: **Reference graph, energy classes, and computational pipeline.** **a**, embedded 155-pigment graph; stars mark RC sites. **b**, counts by energy and coarse region. All lowest-energy sites are peripheral LHCI, while both RC sites occupy the middle class. **c**, the same Hamiltonian-to-generator-to-absorption pipeline is applied to every condition.

This audit is essential for interpreting Energy-KO. Flattening does not remove a clean downhill funnel. It removes both a small RC offset relative to some antenna sites and a larger set of peripheral low-energy basins. These effects require targeted controls below.

2.3 Neither organization knockout lowers the calibrated performance measures

The Full yield was 0.965000 by calibration, with RC-access probability 0.988019 and conditional charge-separation mean first-passage time (CS-MFPT) 34.986 ps (table 1 and fig. 3). Energy-KO increased yield to 0.973472, raised RC access to 0.993460, and shortened CS-MFPT to 25.155 ps. Coupling-KO yielded 0.970224, RC access 0.993458, and CS-MFPT 29.242 ps. Double-KO closely matched Energy-KO.

The result does not mean that couplings are dispensable. Under strict $J = 0$, both RC access and charge-separation yield were zero. Coupling is therefore necessary in the implemented dynamics. The Coupling-KO result says only that the reference placement of edge strengths was not superior to an RMS-uniform placement under fixed topology and total J^2.

The small yield differences also require scale awareness. Uniform fluorescence and heat loss sum to $k_{\text{loss}} =$

Table 1: **Primary knockout endpoints at the Full-model calibration point.**

Condition	CS yield	CS-MFPT (ps)	RC access	lifetime (ps)
Full	0.965000	34.986	0.988019	43.750
Energy-KO	0.973472	25.155	0.993460	33.160
Coupling-KO	0.970224	29.242	0.993458	37.221
Double-KO	0.973475	25.150	0.993464	33.156

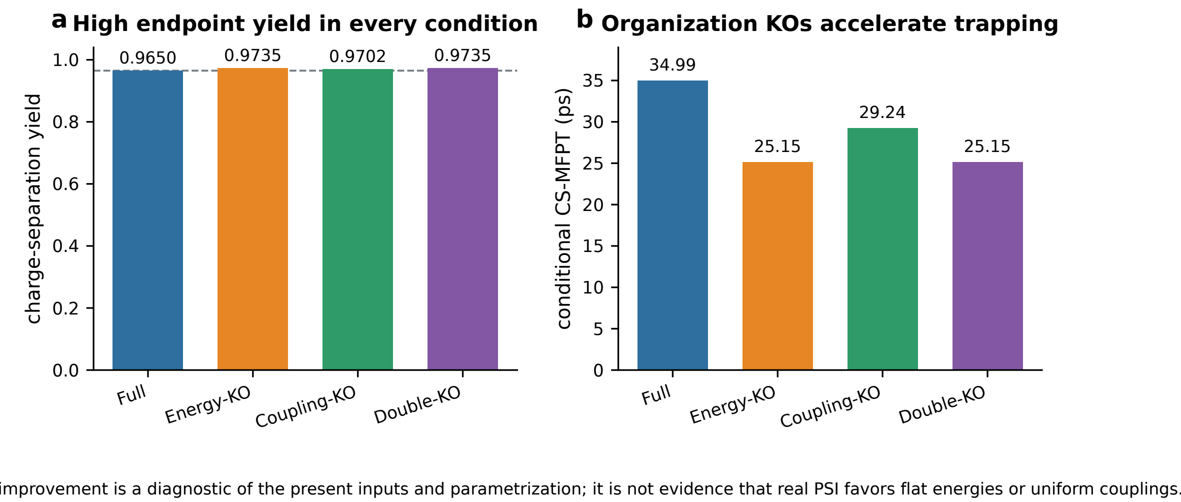


Figure 3: **Factorial knockout results.** **a**, charge-separation yield shown from zero. **b**, conditional CS-MFPT. **c**, probability of reaching an RC site. **d**, factorial interaction contrasts. Every knockout uses the Full calibration without refitting.

$0.0008\ \mathrm{ps}^{-1}$. For any condition in this model,

$$\Phi_{\mathrm{CS}} = 1 - k_{\mathrm{loss}}\langle\tau\rangle, \tag{1}$$

where $\langle\tau\rangle$ is the mean transient excitation lifetime. Fitting $\Phi_{\mathrm{CS}} = 0.965$ fixes $\langle\tau\rangle = 43.75$ ps. The calibrated yield and mean lifetime are therefore not independent validations, and the near-saturated yield compresses kinetic differences.

2.4 Targeted energy controls recover a beneficial RC offset

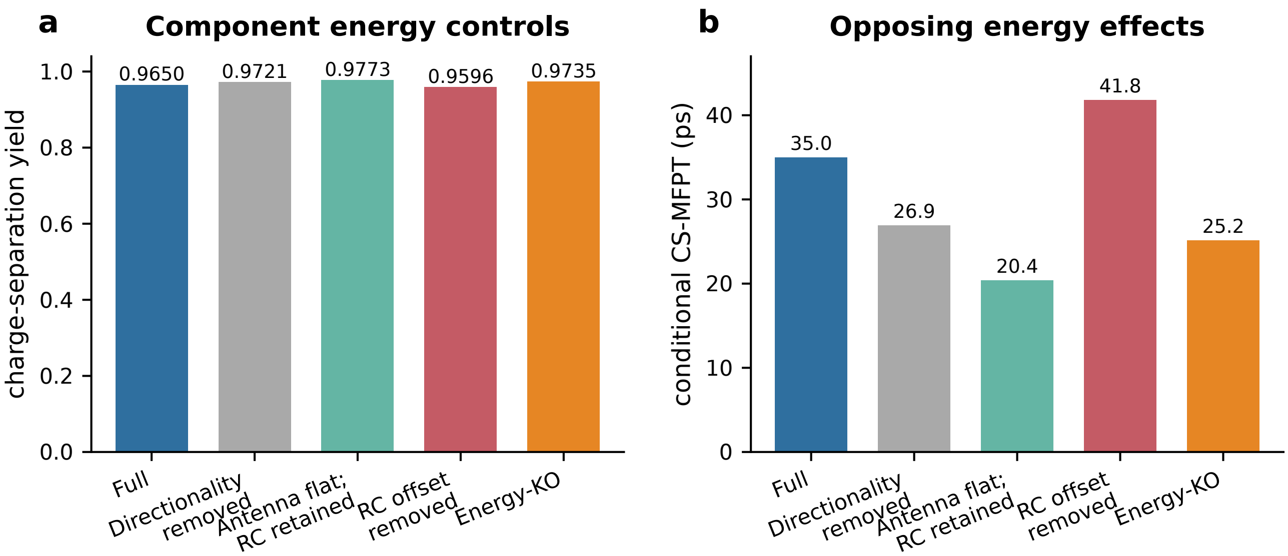
To separate the energetic effects, we performed three additional controls (fig. 4). Eliminating only thermal directionality while retaining energy-dependent detuning raised yield to 0.972055 and shortened CS-MFPT to 26.900 ps. Flattening antenna energies while keeping the RC energy raised yield to 0.977282 and shortened MFPT to 20.360 ps. In contrast, eliminating only the relative RC offset lowered yield to 0.959593 and lengthened MFPT to 41.802 ps.

These controls provide a positive mechanistic statement within the surrogate: an RC energy offset helps trapping. They also prevent an overgeneralized interpretation of Energy-KO. The Full three-level map combines helpful and harmful contributions; the net effect reflects its particular assignment.

2.5 Coupling-scale sensitivity exposes a yield–speed tradeoff

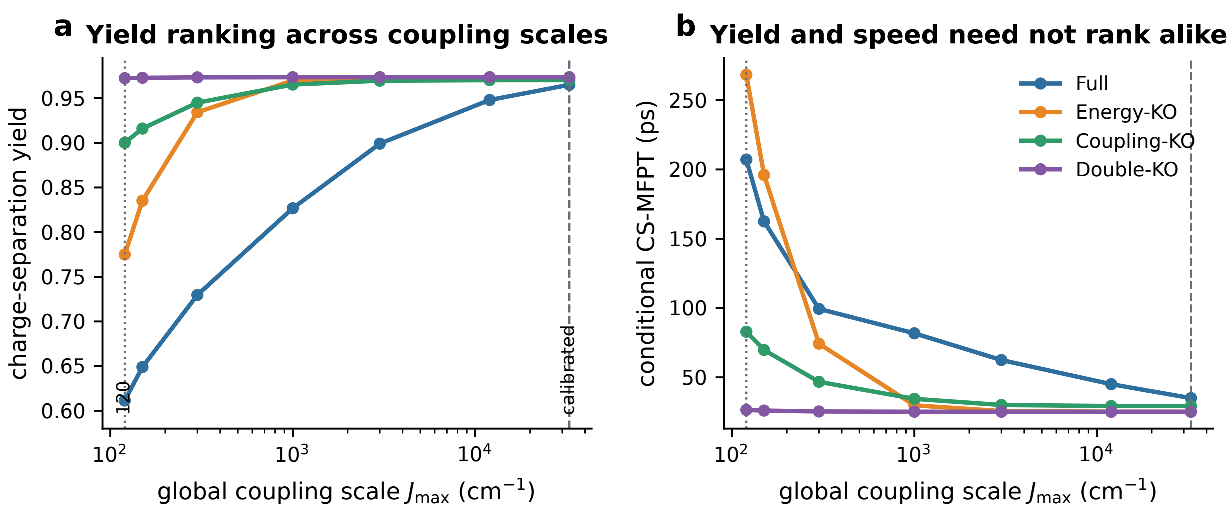
The Full yield calibration selected $J_{\mathrm{max}} = 32825.463836\ \mathrm{cm}^{-1}$. We therefore scanned the global coupling scale while keeping all other parameters fixed (fig. 5). The organization knockouts retained higher terminal yield throughout the sampled range, but their kinetic ranking was not invariant.

At $J_{\mathrm{max}} = 120\ \mathrm{cm}^{-1}$, Full yielded 0.611456 with CS-MFPT 207.060 ps, whereas Energy-KO yielded 0.774917 with a slower 268.269 ps conditional MFPT. Coupling-KO and Double-KO yielded 0.900080 and 0.972430 with MFPTs of 82.892 and 26.466 ps. Thus flattening energy can rescue more trajectories while delaying the successful subset. Terminal yield alone cannot rank transport organization.



Removing only the RC low-energy offset is detrimental; flattening the non-RC antenna is beneficial in this three-level input.

Figure 4: **Targeted controls decompose the energy knockout.** Removing the RC offset worsens both yield and conditional trapping time, showing that this local energetic organization is beneficial. The apparent Full-model penalty arises from the coarse peripheral heterogeneity and associated directionality, not from energy organization in general.



At $J_{\max} = 120$ cm^{-1}, Energy-KO increases yield but lengthens conditional MFPT (207 to 268 ps), exposing an endpoint-speed tradeoff.

Figure 5: **Endpoint and kinetic sensitivity to global coupling scale.** **a**, charge-separation yield. **b**, conditional CS-MFPT. At $J_{\max} = 120$ cm^{-1} (dashed line), Energy-KO increases terminal yield yet slows successful trapping, whereas the calibration point (dotted line) compresses all endpoint yields.

2.6 The calibrated coupling scale prevents a native-design interpretation

The calibrated $J_{\max}$ is far above ordinary chlorophyll coupling scales in structure-based PSI models, where even strongly interacting pairs are commonly on the order of 10^2 cm^{-1} [5, 9]. Under Coupling-KO, the common RMS edge magnitude is approximately 5122 cm^{-1}. The mean Hamiltonian participation ratio correspondingly rises from 2.97 in Full to 33.49 in Coupling-KO (fig. 6).

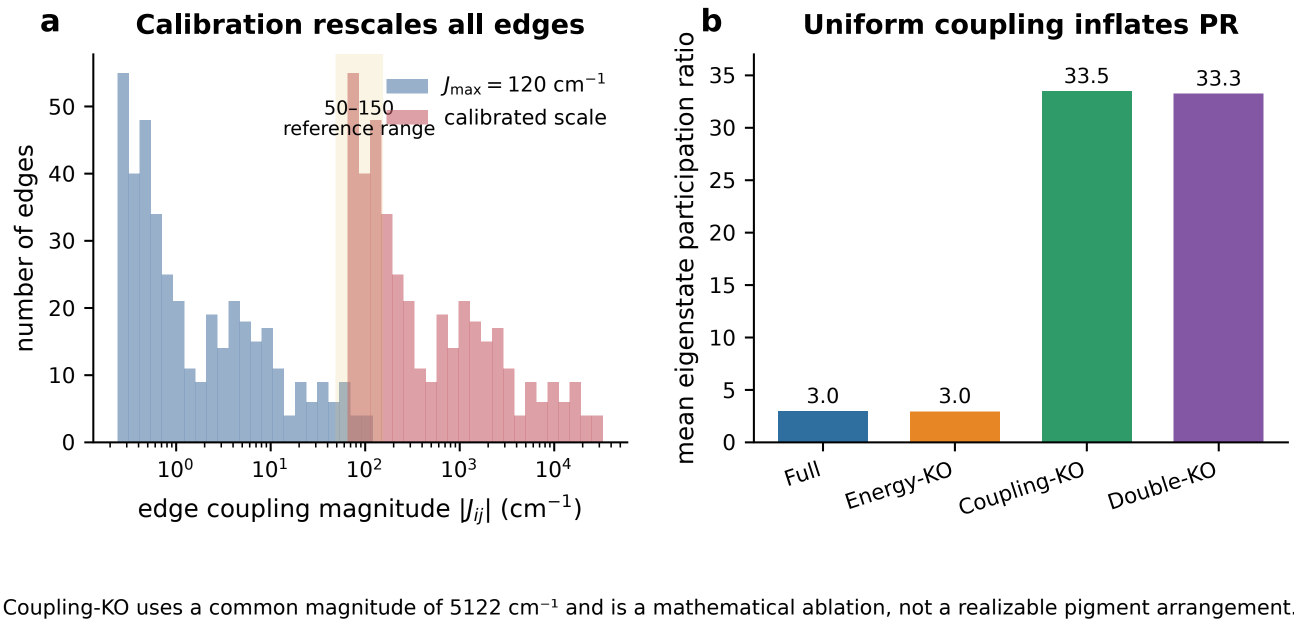


Figure 6: **Internal scale audit.** **a**, mean Hamiltonian participation ratio. **b**, the Full calibration and RMS-uniform KO edge strength compared with a representative 50–150 cm^{-1} chlorophyll-coupling band. Coupling-KO is a mathematical stress test, not a plausible alternative pigment arrangement.

The coupling KO therefore leaves the fixed graph but enters a qualitatively different eigenstate regime. Its improved performance cannot be used to propose uniform couplings as a biological design. Instead, the improvement reveals that an unconstrained global scale is compensating for missing physics or inconsistent inputs.

3 Discussion

3.1 What the knockout establishes

The factorial experiment answers three nested questions. First, does transfer require coupling? Yes: the zero-coupling control abolishes RC access. Second, does the reference allocation of strong and weak edge couplings outperform an aggregate-scale-matched control? Not under the present endpoints and parametrization. Third, does the supplied energy heterogeneity improve trapping? Not in aggregate, although a targeted control shows that the relative RC offset is beneficial.

This hierarchy matters. Coupling necessity and coupling-strength organization are not equivalent. Likewise, energy biases affecting rates and this three-level energy map being a favorable funnel are not equivalent. The factorial design replaces ambiguous importance language with explicit intervention targets.

3.2 Why the result is a model diagnosis, not a biological optimum

Real PSI–LHCI contains pigment-specific electrostatic environments, structured protein fluctuations, spectrally distinct red forms, and, in at least some Lhca clusters, mixed exciton–charge-transfer character [5–7]. The present model represents these features with three scalar energy classes and a globally rescaled coupling matrix. When that global scale is pushed more than two orders of magnitude above typical strong chlorophyll couplings to fit one endpoint, an improved uniform-coupling KO is evidence about the surrogate's compensation, not about evolutionary design.

The energy result is similarly localized. Experiments show that low-energy chlorophylls may slow transfer through thermally activated escape [3, 4], but red states also extend spectral coverage and can assist transfer when appropriately positioned [5, 12]. Our input places every lowest-energy site in peripheral LHCI. Energy-KO therefore tests that assignment, not all physically plausible PSI energy landscapes.

3.3 Calibration to one saturated endpoint is insufficient

Equation (1) makes the calibration limitation explicit. With uniform loss, matching terminal yield directly fixes mean lifetime. It does not independently validate RC-access paths, transient populations, spectra, or the Hamiltonian. Near saturation, a broad class of connected networks can achieve similar eventual trapping even when their first-passage distributions differ substantially.

This is a general systems-modeling problem. Many parameter combinations can fit one output while remaining mechanistically non-identifiable [10]. Model discrepancy should be represented and constrained through independent observables [11]. For PSI, those observables should include absorption and fluorescence spectra, wavelength-resolved transient kinetics, temperature dependence, minimal-RC charge separation, and held-out perturbations [13–16].

3.4 A path toward a biologically interpretable model

Four upgrades follow directly from the audit. First, replace regional energy labels with pigment-resolved site-energy ensembles constrained by spectroscopy and electrostatic calculations. Second, compute couplings from transition charges or densities and preserve their physical scale rather than fitting one global ceiling. Third, benchmark the scalable rate model against a higher-fidelity open-system method on representative pigment clusters; Haken–Strobl-type models have known high-temperature limitations [17, 18]. Fourth, calibrate jointly to independent spectral and kinetic data and reserve subunit, temperature, or wavelength perturbations for validation.

Once these conditions are met, the same factorial logic becomes biologically informative. Targeted RC-offset, red-cluster, interface-coupling, and topology interventions can be combined with first-passage distributions, transition-path fluxes, and pathway entropy. Related first-passage work in photosystem II shows how path-resolved metrics can expose robustness hidden by terminal efficiency [19, 20]; this is a methodological precedent rather than direct evidence for PSI.

3.5 Conclusion

Factorial knockouts provide a stricter question than random shuffles: what remains when a defined kind of organization is removed? Applied to PCAPS V5.2, they show that coupling is necessary and that an RC energy offset is helpful, but they do not recover an advantage for the reference energy heterogeneity or reference coupling-strength allocation. Because the energy map is coarse and peripheral, the coupling calibration is extreme, and terminal yield is a saturated fitted quantity, the result cannot support flat-energy or uniform-coupling design. Its value is diagnostic. The knockout identifies the physical and inferential repairs required before a PSI–LHCI model can make credible claims about native organization.

4 Methods

4.1 Reference graph and Hamiltonian

The archived input contains 155 pigment nodes and 409 undirected edges from a PDB 9GBI preprocessing workflow. Two nodes are designated as RC sites. The Hamiltonian is

$$H_{ij} = \epsilon_i \delta_{ij} + J_{ij}(1 - \delta_{ij}), \tag{2}$$

with explicitly verified symmetry $J_{ij} = J_{ji}$. The site energies comprise 23 sites at 1.795 eV, 120 at 1.820 eV, and 12 at 1.940 eV.

4.2 Thermal population-transfer model

Directed rates use a symmetric base kernel and finite-temperature weighting,

$$k_{i \to j} = k_{ij}^{(0)} \exp\left[-\frac{\epsilon_j - \epsilon_i}{2k_{\mathrm{B}}T}\right], \tag{3}$$

where $k_{ij}^{(0)}$ depends on J_{ij}^2, detuning, and dephasing. This construction gives

$$\frac{k_{i\to j}}{k_{j\to i}} = \exp[-(\epsilon_j - \epsilon_i)/(k_\mathrm{B}T)] \tag{4}$$

for a symmetric coupling matrix. Temperature was 300 K, dephasing rate 1 ps^{-1}, RC charge-separation rate 3 ps^{-1}, fluorescence loss 5×10^{-4} ps^{-1}, and heat loss 3×10^{-4} ps^{-1}.

4.3 Knockout definitions

Energy-KO replaced each ϵ_i by the mean site energy. Coupling-KO preserved the edge set but assigned

$$J_\mathrm{RMS} = \sqrt{\frac{1}{|\mathcal{E}|}\sum_{(i,j)\in\mathcal{E}} J_{ij}^2} \tag{5}$$

to every occupied edge, thereby preserving $\sum J^2$. Double-KO combined these interventions. Strict coupling-null set all off-diagonal entries to zero. Directionality-KO retained detuning but removed Boltzmann weighting. Antenna-flat/RC-retained flattened non-RC energies while keeping the RC class. RC-offset-KO removed only the RC energetic offset.

4.4 Calibration and observables

The global coupling ceiling was adjusted once so that Full reached $\Phi_\mathrm{CS} = 0.965$, giving $J_\mathrm{max} = 32825.463836$ cm^{-1}. All knockout calculations reused this scale and all remaining parameters. For the transient generator Q, the fundamental matrix $N = (-Q)^{-1}$ yields absorption probabilities, mean lifetimes, and conditional moments. We report charge-separation yield, RC-access probability, and conditional CS-MFPT. Time courses from a detailed-balance eigendecomposition agreed with direct matrix exponentiation to within 1.5×10^{-8}; total-probability errors were of order 10^{-8}.

4.5 Sensitivity and reproducibility

The global coupling ceiling was scanned from lower, chlorophyll-scale values through the calibrated point without refitting. Automated tests verify Hamiltonian symmetry, detailed balance, KO constraints, probability conservation, absorption identities, and agreement between independent time-course calculations. Source code, input tables, raw result tables, and plotting scripts accompany the manuscript.

Data and code availability

The reproducibility bundle contains the exact node and edge tables, V5.2 knockout implementation, unit tests, raw primary and sensitivity results, and the script used to generate every figure. No values were manually transcribed into plots.

Author contributions

All listed authors are included as supplied. Specific contribution statements should be confirmed by the authors before submission.

Competing interests

The authors should confirm the competing-interest statement before submission.

Acknowledgement

This work is developed during the 2026 PEBBLE BioFusion Workshop at Westlake University in Hangzhou, China, from July 24th to August 4th.

References

[1] Xiaochun Qin, Michihiro Suga, Tingyun Kuang, and Jian-Ren Shen. Structural basis for energy transfer pathways in the plant PSI–LHCI supercomplex. *Science*, 348:989–995, 2015. doi: 10.1126/science.aab0214.

[2] Yuval Mazor, Anna Borovikova, Ido Caspy, and Nathan Nelson. Structure of the plant photosystem I supercomplex at 2.6 å resolution. *Nature Plants*, 3:17014, 2017. doi: 10.1038/nplants.2017.14.

[3] Robert C. Jennings, Giuseppe Zucchelli, Roberta Croce, and Francesca M. Garlaschi. The photochemical trapping rate from red spectral states in PSI–LHCI is determined by thermal activation of energy transfer to bulk chlorophylls. *Biochimica et Biophysica Acta*, 1557:91–98, 2003. doi: 10.1016/S0005-2728(02)00399-7.

[4] Michele Russo et al. Direct evidence for low-energy chlorophylls limiting excitation energy transfer in photosystem I. *Journal of Physical Chemistry B*, 125:3566–3573, 2021. doi: 10.1021/acs.jpcb.1c01498.

[5] Vladimir I. Novoderezhkin and Roberta Croce. Structure-based modeling of excitation energy transfer in photosystem I. *Photosynthesis Research*, 156:59–74, 2023. doi: 10.1007/s11120-022-00979-8.

[6] Elisabet Romero, Milena Mozzo, Ivo H. M. van Stokkum, Jan P. Dekker, Rienk van Grondelle, and Roberta Croce. The origin of the low-energy form of photosystem I light-harvesting complex Lhca4: mixing of the lowest exciton with a charge-transfer state. *Biophysical Journal*, 96:L35–L37, 2009. doi: 10.1016/j.bpj.2008.11.043.

[7] S. Capaldi et al. Structural and spectroscopic characterization of red-shifted chlorophylls in photosystem I from Chlamydomonas reinhardtii. *New Phytologist*, 248:2331–2346, 2025. doi: 10.1111/nph.70562.

[8] Xiaoqin Yang et al. Excitation energy transfer in the antenna system of photosystem I. *Biophysical Journal*, 85: 140–158, 2003. doi: 10.1016/S0006-3495(03)74461-0.

[9] Melih K. Sener, Craig Jolley, Adam Ben-Shem, Petra Fromme, Nathan Nelson, Roberta Croce, and Klaus Schulten. Comparison of the light-harvesting networks of plant and cyanobacterial photosystem I. *Biophysical Journal*, 89:1630–1642, 2005. doi: 10.1529/biophysj.105.066464.

[10] Ryan N. Gutenkunst et al. Universally sloppy parameter sensitivities in systems biology models. *PLoS Computational Biology*, 3:e189, 2007. doi: 10.1371/journal.pcbi.0030189.

[11] Chuan Lei et al. Quantifying and reducing model discrepancy in systems biology: parameter estimation, model selection and identifiability. *Philosophical Transactions of the Royal Society A*, 378:20190349, 2020. doi: 10.1098/rsta.2019.0349.

[12] Martijn Tros et al. Breaking the red limit: efficient trapping of long-wavelength excitations in chlorophyll antenna systems. *Physical Chemistry Chemical Physics*, 22:25720–25729, 2020. doi: 10.1039/D0CP03351K.

[13] J. A. Ihalainen et al. Kinetics of excitation trapping in intact photosystem I of Chlamydomonas reinhardtii and Arabidopsis thaliana. *Biochimica et Biophysica Acta*, 1706:267–275, 2005. doi: 10.1016/j.bbabio.2004.11.007.

[14] Emilie Wientjes, Ivo H. M. van Stokkum, Herbert van Amerongen, and Roberta Croce. The role of the individual Lhcas in photosystem I excitation energy trapping. *Biophysical Journal*, 101:745–754, 2011. doi: 10.1016/j.bpj.2011.06.045.

[15] P. Akhtar et al. Time-resolved fluorescence study of excitation energy transfer in photosystem I from Chlamydomonas reinhardtii. *Photosynthesis Research*, 135:239–250, 2018. doi: 10.1007/s11120-017-0427-2.

[16] P. Akhtar et al. Excitation energy transfer and charge separation in the minimal photosystem I reaction center complex. *Journal of the American Chemical Society*, 143:14601–14612, 2021. doi: 10.1021/jacs.1c05010.

[17] Hermann Haken and Gert Strobl. An exactly solvable model for coherent and incoherent exciton motion. *Zeitschrift für Physik A*, 262:135–148, 1973. doi: 10.1007/BF01399723.

[18] William Barford. Exciton dynamics in the Haken–Strobl–Reineker model. *Journal of Chemical Theory and Computation*, 20:6510–6517, 2024. doi: 10.1021/acs.jctc.4c00568.

[19] Meng Yang et al. Pathway entropy and first-passage analysis of photosynthetic energy transport. *Nature Communications*, 15:8763, 2024. doi: 10.1038/s41467-024-53138-z.

[20] J. Hall et al. Entropy and robustness in photosynthetic excitation-energy transport. *Proceedings of the National Academy of Sciences USA*, 122:e2426331122, 2025. doi: 10.1073/pnas.2426331122.