Electronic Supplementary Material (ESI) for Physical Chemistry Chemical Physics.

Supplemental Information for

Determining Interchromophore Effects for Energy Transport in Molecular Networks Using Machine-Learning Algorithms

Brian S. Rolczynski,[1,*] Sebastián A. Díaz,[2] Youngchan Kim,[3] Divita Mathur,[4] William P. Klein,[2] Igor L. Medintz,[2] and Joseph S. Melinger[1,*]

[1]Electronics Science and Technology Division, Code 6800, U.S. Naval Research Laboratory, Washington DC 20375, USA.
[2]Center for Bio/Molecular Science and Engineering, Code 6900, U.S. Naval Research Laboratory, Washington DC 20375, USA.
[3]Materials Science and Technology Division, Code 6300, U.S. Naval Research Laboratory, Washington DC 20375, USA.
[4] Department of Chemistry, Case Western Reserve University, Cleveland, OH 44106, USA.

*corresponding author. email: brian.rolczynski@nrl.navy.mil, joseph.melinger@nrl.navy.mil

Sample configurations

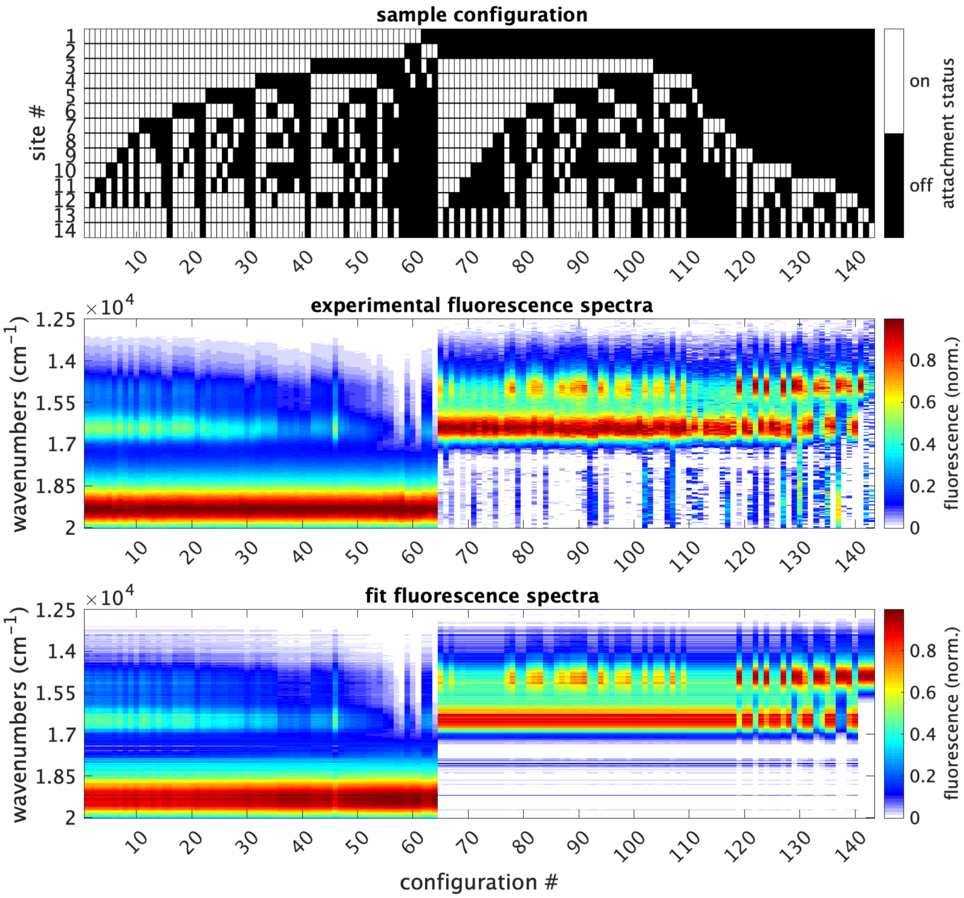

Figure S1. Sample configurations (top), corresponding measured fluorescence spectra at 21,459 cm^{-1} excitation (middle), and three-component fits (bottom) are shown. Each column corresponds to one of the 143 unique sample configurations used in this study. In the top figure, the rows correspond to Sites 1-14. A white square indicates that the corresponding Site is occupied, while a black space indicates that it is absent. The three-component fits were obtained by fitting the sample's fluorescence spectrum to a linear combination of the fluorescence spectra from each of the monomer types attached to DNA. The weight parameters obtained by these fits were used in the subsequent FRET analysis.

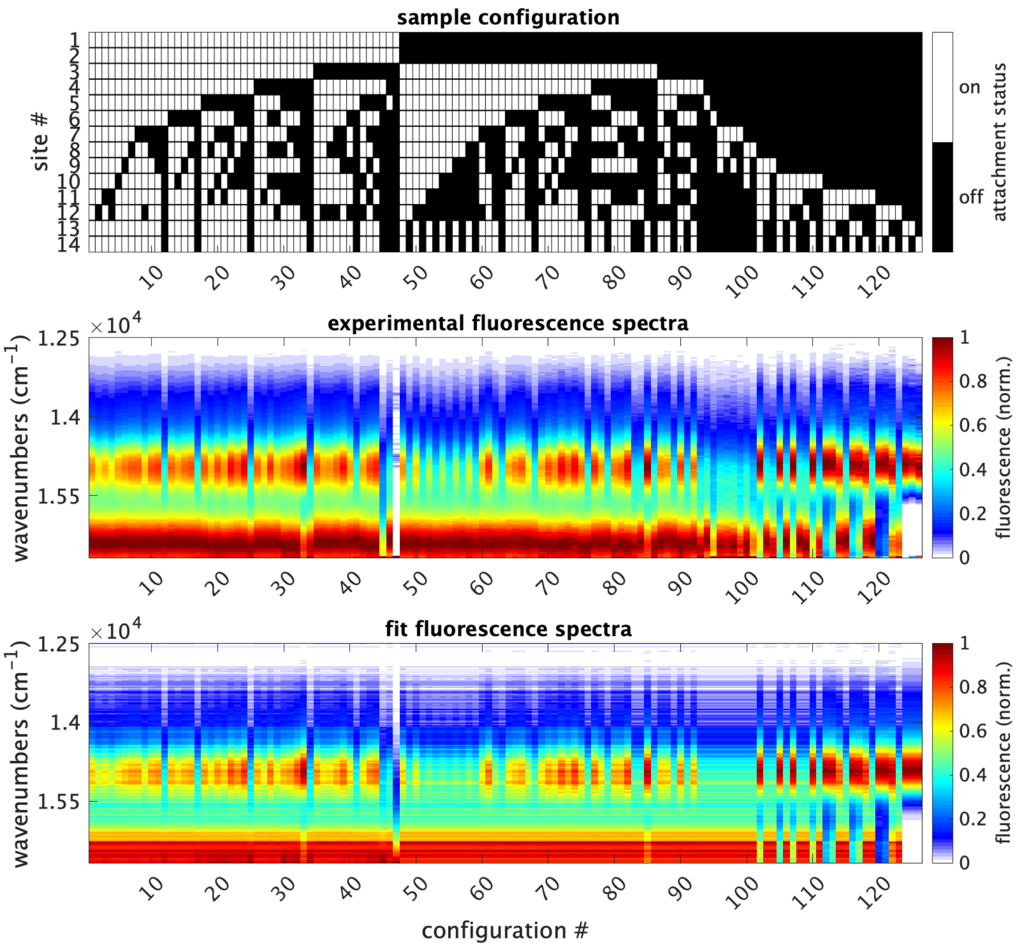

Figure S2. Sample configurations (top), corresponding measured fluorescence spectra at 17,094 cm^{-1} excitation (middle), and three-component fits (bottom) are shown. The information here is similar to Figure 2, except the excitation wavelength differs.

Model validation

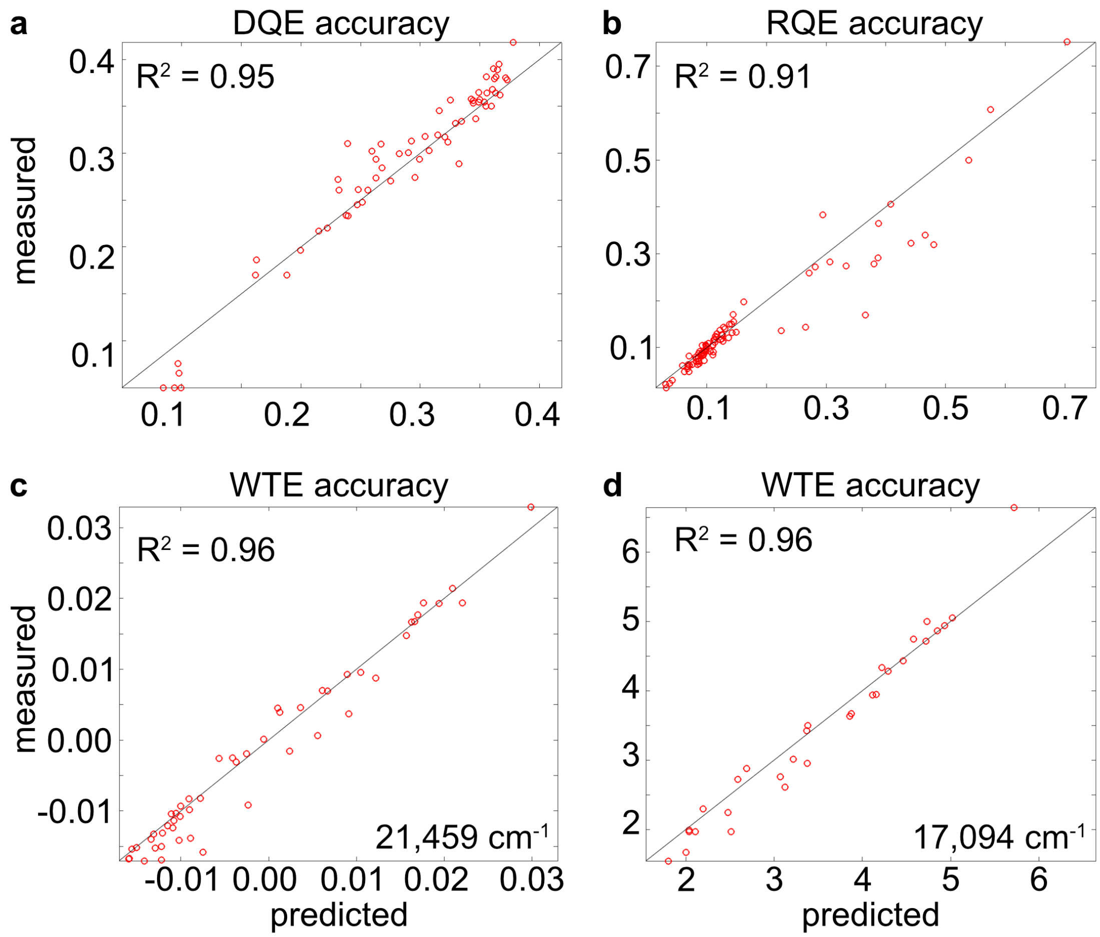

Figure S3. The predictions compared to measured outputs are shown for the DQE, RQE, and WTE outputs. Panels c and d specify the excitation wavenumber in their bottom-right corners. The red circles correspond to individual samples. The closer they are clustered to the diagonal line, the better the prediction. The correlation coefficients indicate good correspondence ($R^2 > 0.9$) for all outputs. Though these figures are produced using random variations and are not exactly identical, similar figures for panels a and c have been published previously.[1]

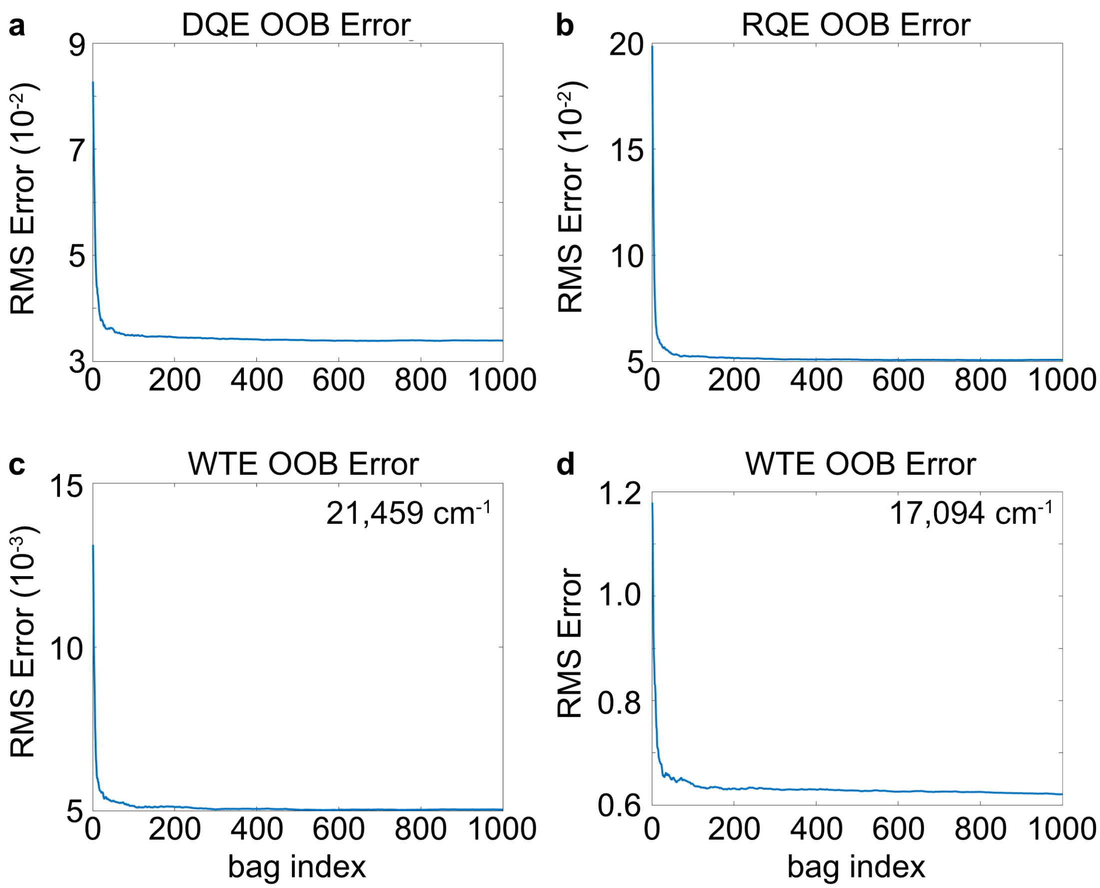

Figure S4. The out-of-bag validation scores are shown. The root-mean-square errors are shown as a function of number of predictors sampled (bag index) for the DQE (a), RQE (b), and WTE (c-d) outputs. The plots for the WTE outputs list the excitation wavenumber in the top-right corner. The biggest error contributions occur at <100 bags, and by 1000 bags all of the curves have reached a plateau.

Sample Size discussion

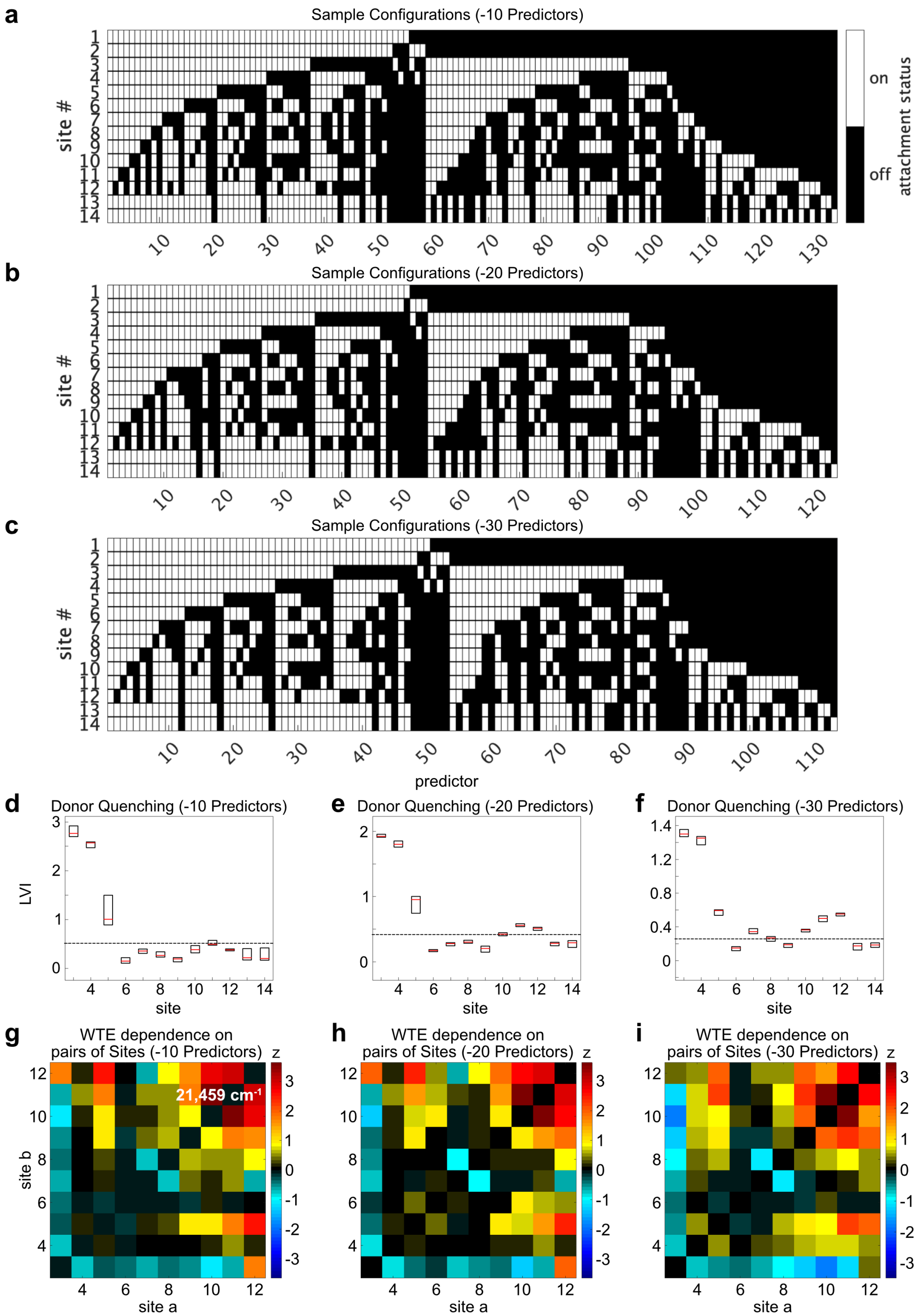

Figure S5. (a-c) Sample configurations are shown with 10-30 predictors randomly removed. (d-f) The corresponding Donor Quenching LVI box plots are shown. (g-i) The corresponding WTE NVI plots are shown. All of these panels use the data set with 21,459 cm^{-1} excitation.

Figure S5 addresses whether enough samples were obtained for the Random Forest model. The approach in this figure is to determine whether the major features of representative plots are greatly changed by the removal of random predictors. When 10, 20, or 30 predictors (selected at random) were removed, the major features in the Random Forest outputs were nonetheless not significantly changed. For example, inspecting the Donor Quenching LVIs in Figures S5d-f, the prominent features are Sites 3-5. These Sites retain their relative LVI scores in each of the figure panels. There is a small amount of variation in each of the Figures compared to each other, however this effect is expected due to the stochastic nature of the Random Forest model; and it does not disrupt the overall patterns. Likewise, the WTE NVI scores (Figures S5g-i) also have the same major features. These include the negative pattern between Site 3 and Sites 4-11, with the positive peak between Site 3 and Site 12; the positive signal between Site 5 and Sites 9-12; and the strong positive signals among Sites 9-12. Minor variations exist, but like before, these are expected due to the stochastic nature of the Random Forest model.

References

1. Klein, W. P.; Rolczynski, B. S.; Oliver, S. M.; Zadegan, R.; Buckhout-White, S.; Ancona, M. G.; Cunningham, P. D.; Melinger, J. S.; Vora, P. M.; Kuang, W.; Medintz, I. L.; Díaz, S. A., DNA Origami Chromophore Scaffold Exploiting HomoFRET Energy Transport to Create Molecular Photonic Wires. *ACS Appl. Nano Mat.* **2020,** *3* (4), 3323-3336.