Supplementary Information
Machine Learning Exciton Dynamics

Florian Häse,[1,2] Stéphanie Valleau,[1] Edward Pyzer-Knapp,[1] and Alán Aspuru-Guzik[1]

[1]*Department of Chemistry and Chemical Biology,*

Harvard University, Cambridge, Massachusetts 02138, USA

[2]*Physik-Department T38, Technische Universität München, Garching, 85748, Germany*

S.1. SUPPLEMENTARY INFORMATION

S.1.1. Coulomb matrix space cluster analysis: choosing the best grid-search BChl molecule

The neural networks were trained on Coulomb matrices representing BChl conformations generated during the classical MD simulation in a supervised training-scheme. TDDFT excited state energies were used as training target. To design an optimal neural network architecture with a minimal deviation between predictions and targets, a grid search on several neural network hyperparameters was performed. Learning rate and number of neurons in the first and second hidden layer were changed step-wise, as reported in Sec. S.1.2.

We found that BChls in the FMO complex show a high Coulomb matrix space overlap throughout the entire MD trajectory (see Fig. S.1). By determining the site with the highest Coulomb matrix space overlap with all other sites we were therefore able to limit the number of grid searches for optimal neural network hyperparameters to one. Optimal neural network hyperparameters obtained for the most representative site with the highest Coulomb matrix space overlap were used for all other sites.

To identify the most representative site we first started a cluster analysis on all Coulomb matrices representing site 1 based on the gromos method. Coulomb matrix distances were measured with the Frobenius norm (see main text, Sec. 2.2.3). The clustering cut-off was chosen to be $90\,\mathrm{e}^2/\mathrm{\AA}$ to clearly distinguish Coulomb matrices. Then, for all other sites, distances of all Coulomb matrices of all frames in the MD trajectory to all Coulomb matrices representing on particular cluster were calculated. If the distance was below the cut-off, the Coulomb matrix was attributed to the cluster. The pool of remaining Coulomb matrices was again clustered with the same method as site 1.

Coulomb matrix space overlap of one site with all other sites was then estimated by counting the number of Coulomb matrices of the considered site and of Coulomb matrices of all other sites in every individual cluster. Whenever two sites i and j contributed to the same cluster different numbers of Coulomb matrices n_i and n_j, the smaller number of Coulomb matrices $\min(n_i, n_j)$ was added to the Coulomb matrix space overlap estimation of both sites. This summation was carried out over all sites and all clusters. Quantitative results are presented in Fig. S.1.

We observed that of all BChls in the FMO complex, site 3 has the most shared Coulomb matrix space with all other sites. However, the difference of shared Coulomb matrix space volumes is

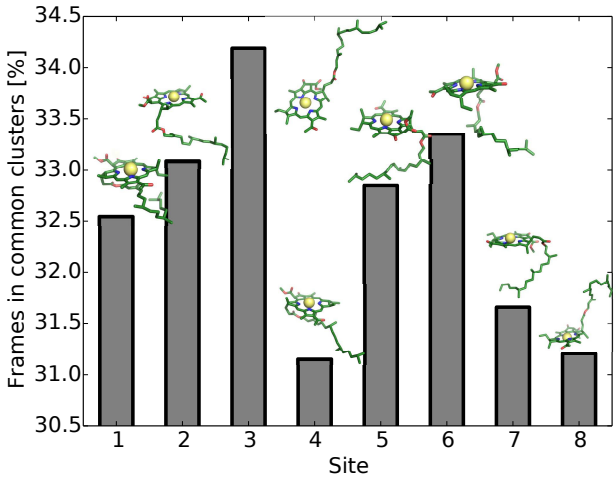

Figure S.1: Common Coulomb matrix space regions of bacteriochlorophylls observed during the $40\,\mathrm{ps}$ production run. We report the number of frames (in $\%$) of one site contributing to Coulomb matrix space clusters shared with at least one other site (cluster cut-off: $90\,\mathrm{e}^2/\mathrm{\AA}$). Stick representations of corresponding representative geometries of each individual site are shown for comparison.

small. The highest observed value ($34.19\,\%$ for site 3) is only about $10\,\%$ greater than the smallest observed value ($31.15\,\%$ for site 4) and for every site about one third of all Coulomb matrices lies in shared Coulomb matrix space regions. Hence, we expect neural network architectures optimized on one site to also work well for all other sites. This assumption was confirmed by the similarity in prediction accuracy of neural networks trained on different sites (see Sec.3.1).

S.1.2. Neural network parameter grid search: finding the best network architecture

The prediction accuracy of a neural network is highly influenced by its architecture. Introducing hidden layers to the neural network architecture allows for the distinction of data which is not linearly separable. In this study we used multi-layer perceptrons (i.e. fully connected neural networks with at least one hidden layer) with two hidden layers and logistic activation functions. Several neural networks with different learning rates and numbers of neurons in the first and second hidden layer were designed and trained on Coulomb matrices and corresponding excited state energies. We used the back-propagation algorithm and a supervised training scheme with excited state energies as target. Overfitting was avoided with the early stopping method (see main text, Sec. 2.2). Thus we determined an optimal set of hyperparameters to identify a neural network

architecture suitable for accurate excited state energy predictions for BChls in the FMO complex from classical MD simulations.

The entire grid search was performed on BChl 3, which was shown to represent the most of the Coulomb matrix space covered by all BChl molecules in the FMO complex during the $40\,\mathrm{ps}$ production run (see Sec. S.1.1). Since optimal neural network architectures were not known beforehand we chose particular initial hyperparameter values prior to the grid search. Unless specified otherwise neural networks were trained with a learning rate of 10^{-3} and 180 neurons in the first and second hidden layer. Neural networks were trained on 3000 frames to keep the training times during the grid search reasonably small.

S.1.2.1. Learning rate.

To determine an optimal learning rate we set up five different neural networks with learning rates of: 10^{-3}, $5 \cdot 10^{-4}$, 10^{-4}, $5 \cdot 10^{-4}$ and 10^{-5}. All of these neural networks were trained on $30\,\%$ of all trajectory frames representing BChl 3 by Coulomb matrices. The neural networks were designed with two hidden layers consisting of 180 respectively.

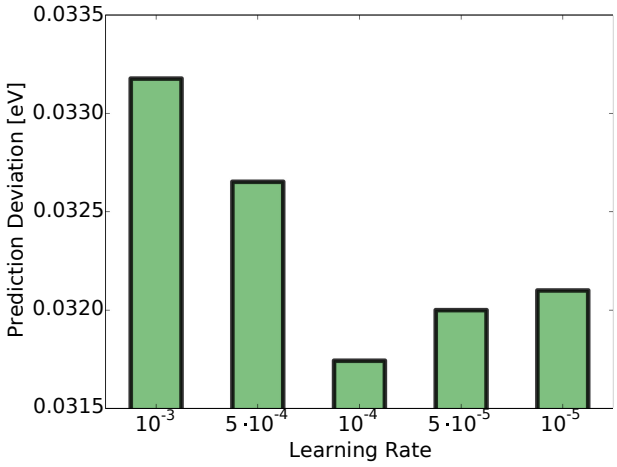

Figure S.2: Average absolute deviations of predicted excited state energies from TDDFT excited state energies for neural networks with different learning rates. Neural networks were trained on 3000 Coulomb matrices randomly drawn from the 10000 frame data set. A learning rate of 10^{-4} resulted in the lowest prediction error. However, prediction error for other learning rates are less than $10\,\%$ larger than the minimal value.

The average absolute deviations of predicted excited state energies from TDDFT results for

each neural network are depicted in Fig. S.2. We see that deviations of single predicted excited state energies from TDDFT excited state energies ranges from $31.8\,\mathrm{meV}$ for a learning rate of 10^{-4} to $33.2\,\mathrm{meV}$ for a learning rate of 10^{-3}. In any case, we found that the prediction error increases as the learning rate deviates from 10^{-4}. Hence, we consider a learning rate of 10^{-4} to be optimal for the case of excited state energies of BChls in the FMO complex. However, the small changes of the prediction accuracy with different applied learning rates indicates that the prediction accuracy of a neural network is not very sensitive to the learning rate for this application.

S.1.2.2. Choice of number of neurons.

The influence of the number of neurons in the first and second hidden layer on the prediction accuracy of a neural network was investigated with several neural networks with a learning rate of 10^{-3}. Neuron numbers in each of the two hidden layers were varied from 96 to 240 in steps of 12. Each of the constructed neural networks was trained on site 3 with 3000 Coulomb matrices randomly selected from the 10000 frame data set. Results are illustrated in Fig. S.3. The neural network with 204 neurons in the first hidden layer and 192 neurons in the second hidden layer showed the smallest average absolute deviation of $31\,\mathrm{meV}$ between single predicted excited state energies and TDDFT excited state energies.

S.1.2.3. Training set size.

Aside from neural network hyperparameters we also investigated the effect of the number of Coulomb matrices in the training set on the prediction accuracy of the neural network. In general, the more data is provided to the neural network during training, the more accurately it can predict the targets. However, with the training set size the computational time spent on neural network training also increases. So we looked for a balance between the amount of data provided to the neural network during training and the computational cost of the neural network training.

Investigated training set sizes ranged from 500 Coulomb matrices to 5000 in steps of 500. In every case, Coulomb matrices of site 3 were drawn randomly from the 10000 frame data set. For each training set size a total of 12 neural networks was trained. Neural networks had a learning rate of 10^{-4} with 204 neurons in the first hidden layer and 192 neurons in the second hidden layer. Recorded training times for different training set sizes are reported in Tab. S.1 as a 12 neural

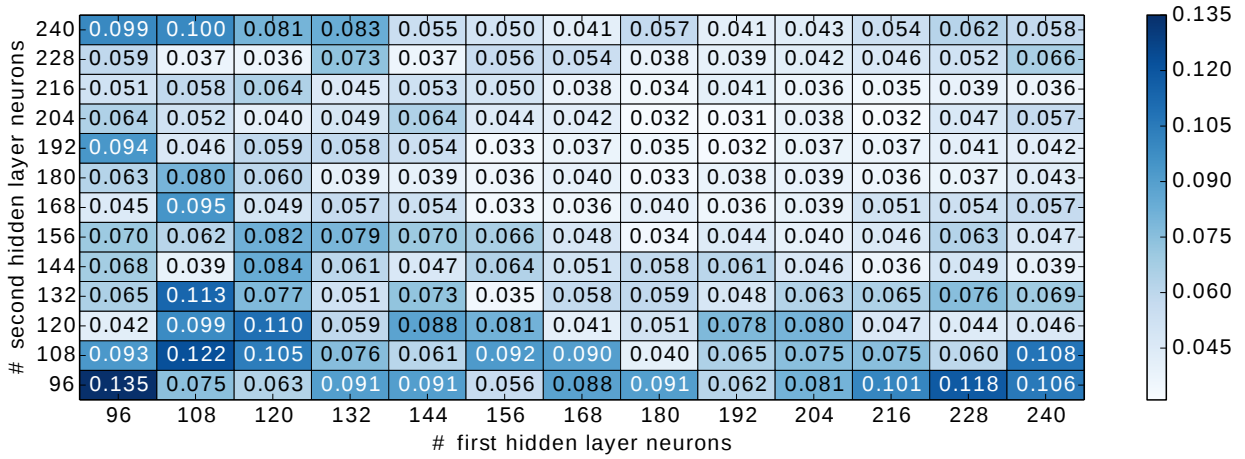

Figure S.3: Average absolute deviations of predicted excited state energies from TDDFT calculated excited state energies for different numbers of neurons in the first and second hidden layer. Deviations are reported in eV. Neural networks were trained on 3000 Coulomb matrices randomly drawn from the 10000 frame data set. A hidden layer combination of 204 neurons in the first hidden layer and 192 neurons in the second hidden layer resulted in the smallest deviation of $31\,\mathrm{meV}$.

network average. Prediction errors for all eight sites in the FMO complex and different training set sizes are illustrated in Fig. S.4. Four cores of Intel(R) Xeon(R) CPUs (X5650 @ $2.67\,\mathrm{GHz}$) with $4\,\mathrm{GB}$ of RAM were used to train one neural network.

We found that neural network training times significantly increase when including more than 4000 frames. Up to this training set size, neural network training takes about one day, which we considered a reasonable time for neural network training. Including 500 more frames to the training set increased the training time by about $6\,\mathrm{h}$ or 24 core hours. As a balance of the amount of input data and computational cost we therefore decided to use 4000 frames in the training set for neural network training.

Training set size [# frames]	Training time [h]
500	1.7 ± 0.6
1000	3.8 ± 1.4
1500	6.5 ± 2.8
2000	13.6 ± 6.5
2500	18.0 ± 7.9
3000	20.3 ± 5.5
3500	22.7 ± 7.3
4000	23.9 ± 5.0
4500	30.2 ± 1.5
5000	31.2 ± 4.8

Table S.1: Training times for neural network training on four cores. Each neural network was trained on Intel(R) Xeon(R) CPUs (X5650 @ 2.67 GHz) with 4 GB of RAM. Training set sizes are reported in number of frames randomly drawn from the 10000 frames of the trajectory of site 3. A total of 12 neural networks was trained on each training set size. Training times are reported with average and standard deviation of all neural network training sessions with the particular training set size.

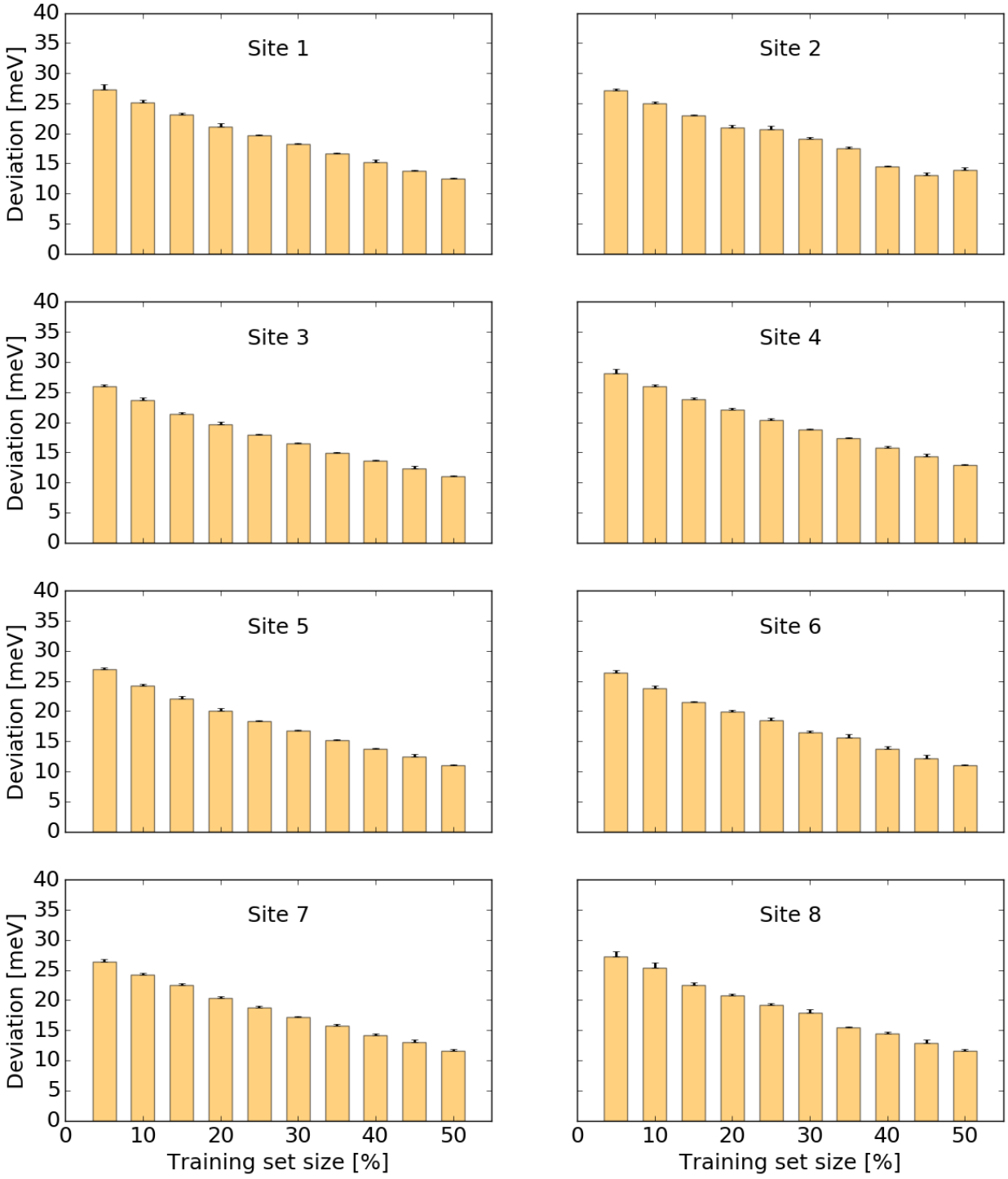

Figure S.4: Average absolute deviations of predicted excited state energies from TDDFT excited state energies for neural networks trained on different training set sizes. Neural networks were trained on the indicated fractions of the 10^4 frame data set. A total of 12 neural networks was trained on each investigated fraction of the data set. Neural networks were set up with a learning rate of 10^{-4} and 192 neurons in the first hidden layer and 204 neurons in the second hidden layer.

S.1.3. Spread of neural network predictions

We used a total of 12 independent neural networks to predict excited state energies for a particular BChl in the FMO complex. Predicted trajectories of all 12 neural networks were averaged to obtain a more accurate predicted excited state energy trajectory. To justify the usage of the average of predicted excited state energy trajectories we report the spread of individual neural network predictions in Fig. S.7 and illustrate predicted excited state energy trajectories of one neural network compared to an average of 12 neural networks in Fig. S.5.

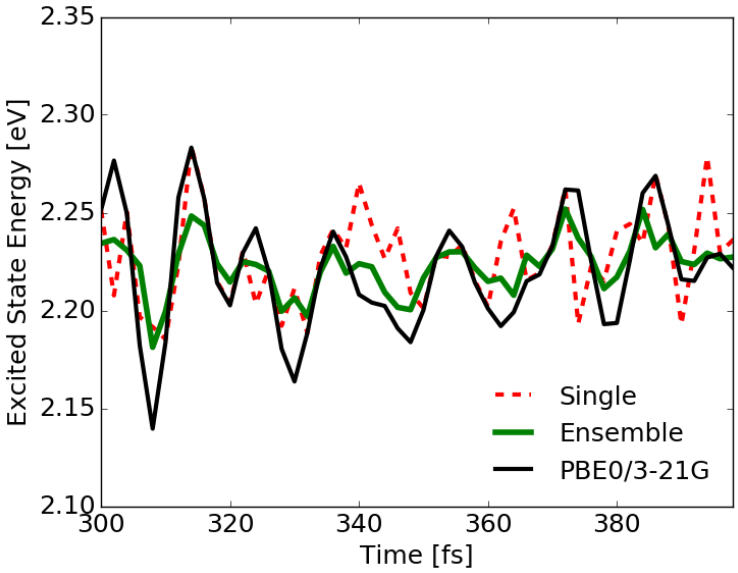

Figure S.5: Part of the entire excited state energy trajectory of site 1 calculated with TD-DFT (PBE0/3-21G, black), predicted from a single neural network (red, dashed) and obtained as a prediction average of 12 independently trained neural networks. Neural networks were trained on 4000 frames randomly drawn from the 10^4 data set.

We compared our TDDFT results for excited state energies and the predictions of neural networks trained on these excited state energies to the values obtained in different studies conducted by Shim et al. in Ref. [1], Olbrich et al. in Ref. [2], and Jurinovich et al. in Ref. [3]. The results are presented in Fig. S.6.

Neural networks with optimal hyperparameters (see Sec. S.1.2) were trained on 4000 Coulomb matrices randomly drawn from the 10000 frame data set for every respective site. After training, neural networks were used to predict the entire excited state energy trajectory of the site on which they were trained. Trajectory averages of predicted excited state averages were calculated

Site	σ_{single} [meV]	σ_{ensemble} [meV]
1	14.0	11.0
2	13.7	12.1
3	12.6	11.6
4	14.3	13.5
5	12.8	11.5
6	13.4	11.5
7	13.4	12.3
8	13.8	12.3

Table S.2: Average absolute deviations σ of neural network predicted excited state energies and excited state energies calculated with TDDFT (PBE0/3-21G). Deviations were calculated for single neural network predictions (σ_{single}) and for ensemble averaged neural network predictions with 12 neural networks (σ_{ensemble}). Ensemble averaging improves the prediction accuracy by up to 3 meV for site 1.

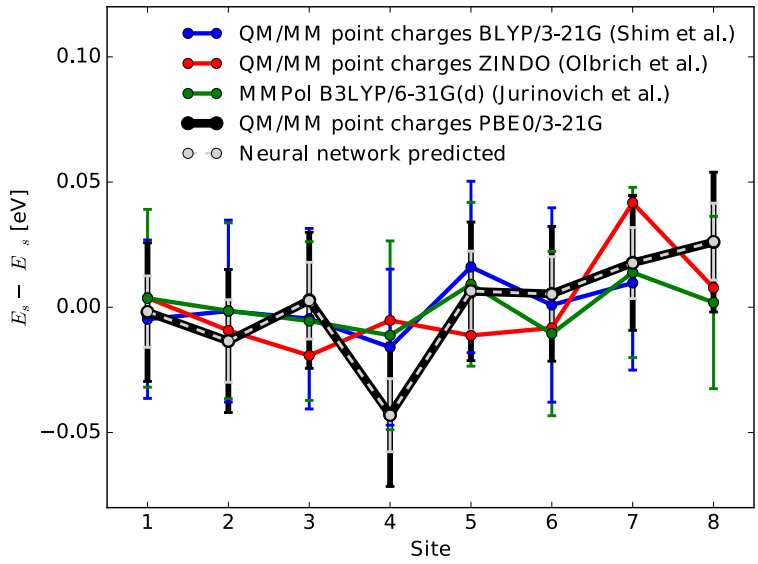

Figure S.6: Averages and standard deviations of excited state energy trajectories for all eight sites in the FMO complex calculated with different methods (solid lines) and predicted by neural networks (dashed lines) trained on TDDFT (PBE0/3-21G). To compare the energies, their values were re-scaled with respect to their mean for each method. Values were obtained from studies conducted by Shim,[1], Olbrich,[2] and Jurinovich.[3].

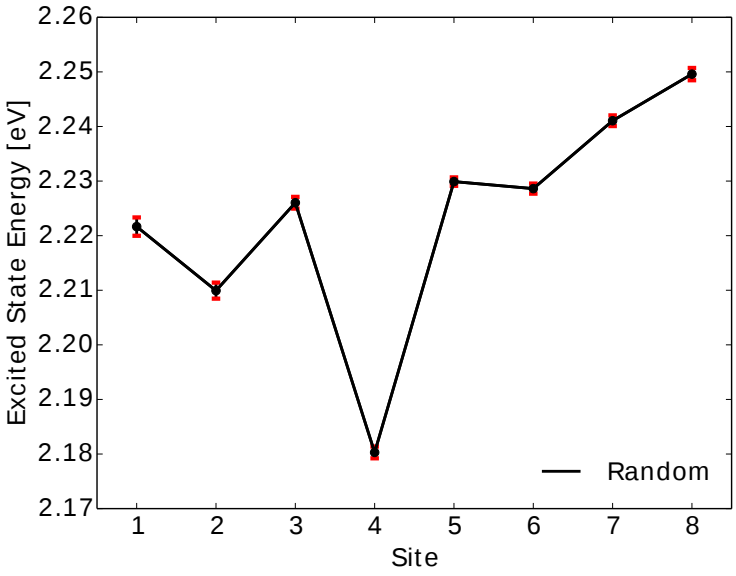

Figure S.7: Averages of neural network predicted excited state energy trajectory averages. Error bars indicate the standard deviations of the neural network predicted excited state energy trajectory distributions. A total of 12 neural networks was trained for each of the eight sites on 4000 Coulomb matrices randomly drawn from the 10000 data frames of the indicated site.

for every neural network prediction. All predicted excited state energy trajectory averages were then ensemble averaged over the 12 neural networks trained on the particular site. Results are reported in Fig. S.7 with the standard deviation of the neural network predicted excited state energy trajectory averages as error bars.

S.1.4. Excited state energy distributions

Excited state energies were calculated for all BChl in the FMO at every 4 fs of the 40 ps production run. Results were obtained from TDDFT calculations with the PBE0 functional and the 3-21G basis set using the Q-Chem quantum chemistry package. Distributions of the obtained excited state energy trajectories are depicted in Fig. S.8.

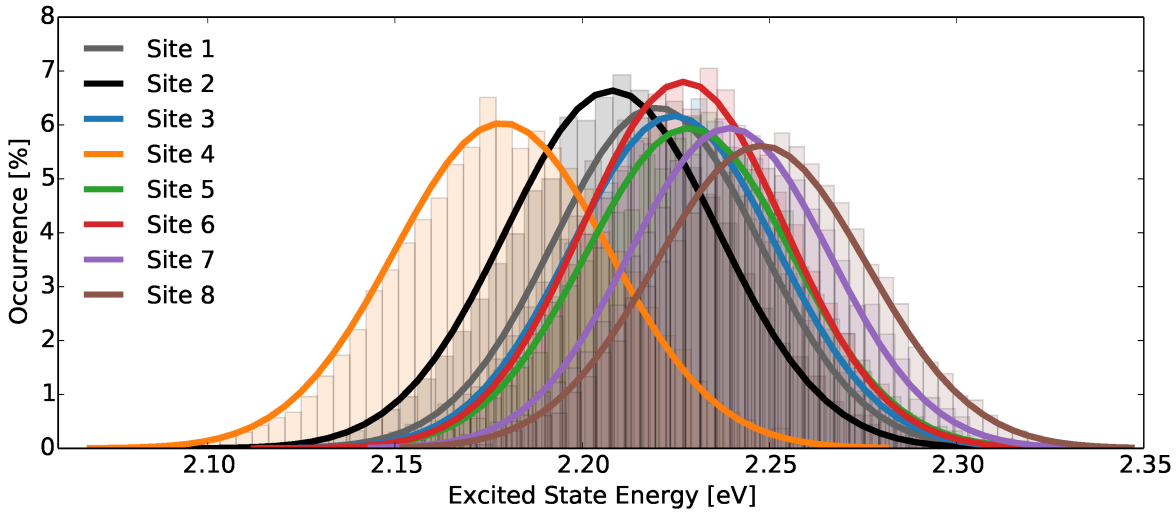

Figure S.8: Excited state energy distributions for all eight BChls in the FMO monomer A. Excited state energies were obtained from TDDFT calculations with the PBE0 functional and the 3-21G basis set. Distributions were calculated from a total of 10000 frames spanning 40 ps with a binning of 50.

S.1.5. Spectral Densities and Exciton Dynamics

S.1.5.1. Spectral Densities for individual sites.

Spectral densities for individual BChls in the FMO complex were calculated from TDDFT excited state energy trajectories and neural network predicted excited state energy trajectories. Neural networks were trained on Coulomb matrices selected with different selection methods from the 10000 frame trajectory. Harmonic spectral densities for individual BChls are shown in Fig. S.11.

For the average spectral density (see main text, Fig. 5 in Sec. 3.2) we observed that neural networks trained on correlation clustered Coulomb matrices predicted spectral densities significantly better than any other Coulomb matrix selection method. However, the advantage of correlation

clustering in terms of spectral density prediction accuracy is less obvious for spectral densities trained on individual BChls.

Nevertheless, for each of the introduced Coulomb matrix selection methods neural networks were able to predict the general shape of the spectral density, although the area below the curves is significantly smaller and correlation clustering identified more peaks than the other selection methods.

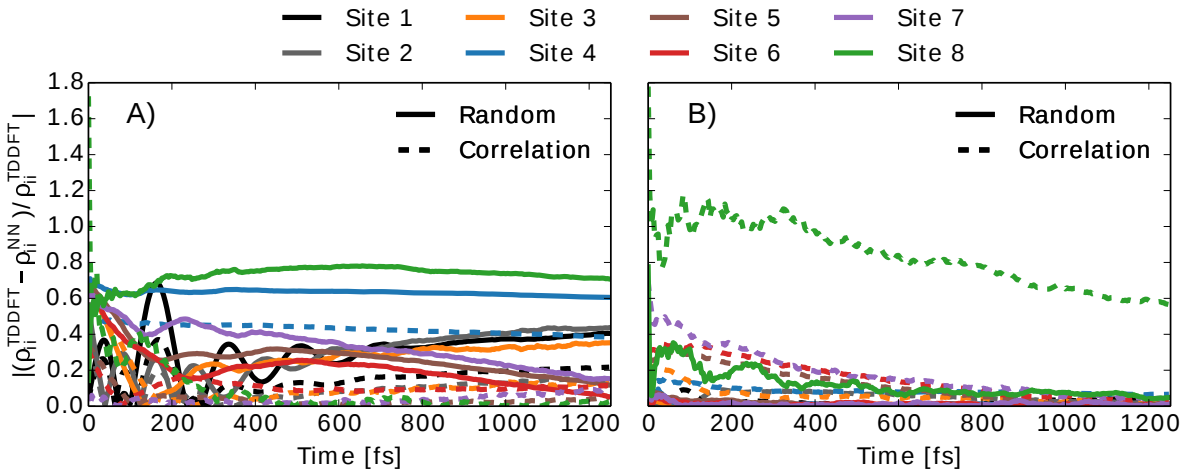

Figure S.9: Deviation of TDDFT calculated exciton dynamics and neural network predicted exciton dynamics over time. The deviation was calculated as $\sigma_\rho^i(t) = |\rho_{ii}^{\mathrm{TDDFT}}(t) - \rho_{ii}^{\mathrm{NN}}(t)|/\rho_{ii}^{\mathrm{TDDFT}}(t)$ with i indicating the BChl. Panel A) shows the deviation for exciton dynamics calculated with the Redfield method and neural network predicted harmonic average spectral densities. In Panel B), the TDDFT average spectral density was used for all calculations and neural networks only predicted excited state energy trajectories.

S.1.6. Error estimation of predicted reorganization energies.

Despite the fact that the trained neural networks predict excited state energies with small deviations, significant deviations are observed in the predicted reorganization energies. To investigate the propagation of error on the energies to the reorganization energies we computed reorganization energies for TDDFT trajectories with Gaussian noise. We choose Gaussian noise as it appears that the distribution of error of the neural networks predictions is approximately Gaussian. Results are depicted in Fig. S.10.

We observe that TDDFT trajectories with Gaussian noise of standard deviation $\sim$10 meV (this

value corresponds to the prediction errors observed for neural networks) lead to deviations in the reorganization energy of about 50 %. This is in agreement with the error we found was obtained by using the neural network predictions.

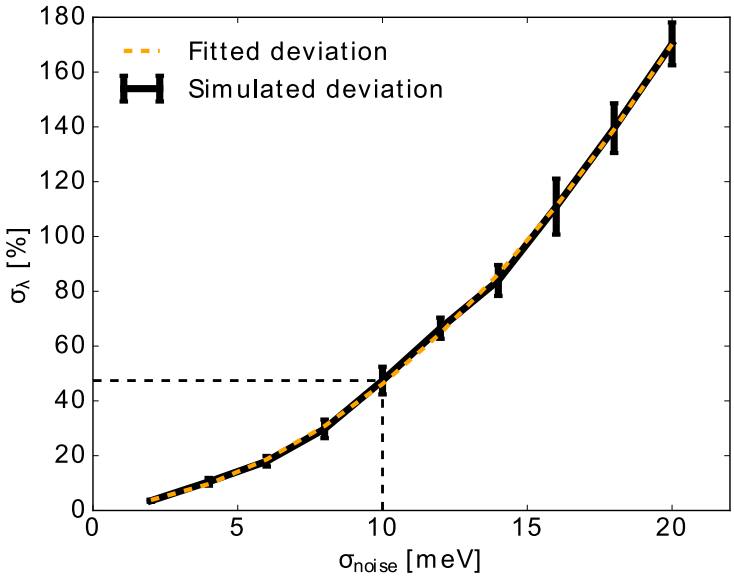

Figure S.10: Percentage deviation of reorganization energies σ_λ depending on the standard deviation of Gaussian noise σ_{noise} added to the TDDFT computed excited state energy trajectories. Gaussian noise with zero mean and different standard deviations was added to the TDDFT computed excited state energy trajectory of site 3. A total of 12 noisy trajectories per investigated standard deviation was generated to estimate the error on the reorganization energy. A quadratic polynomial (yellow dashed line) was found to be a good approximation for the deviation in reorganization energies of noisy trajectories to trajectories without noise. Black dashed lines indicate the error on the reorganization energy for a standard deviation of 10 %. This values corresponds to the prediction errors observed for neural networks.

S.1.6.1. Error estimation of predicted exciton dynamics.

The deviation of neural network predicted exciton dynamics and TDDFT calculated exciton dynamics of the i-th BChl was quantified by calculating $\sigma_\rho^i(t) = |\rho_{ii}^{\text{TDDFT}}(t) - \rho_{ii}^{\text{NN}}(t)|/\rho_{ii}^{\text{TDDFT}}(t)$. Deviations were calculated for the exciton dynamics obtained with the Redfield method as shown in Fig. S.9.

We observe that for exciton dynamics calculations with neural network predicted average spec-

tral densities (see panel A in Fig. S.9) the error of neural networks trained on randomly drawn Coulomb matrices is significantly higher than the error of neural networks trained on correlation clustered Coulomb matrices. This behavior can be observed for all sites for times up to $1\,\mathrm{ps}$. However, if we use the TDDFT calculated average spectral densities for all simulations instead of neural network predicted harmonic average spectral densities, we observe a much smaller error for neural networks trained on randomly drawn Coulomb matrices for all eight sites. This is due to the fact that the error on the energies is slightly smaller with the random sampling.

---

[1] S. Shim, P. Rebentrost, S. Valleau and A. Aspuru-Guzik, *Biophys. J.*, 2012, **102**, 649 – 660. S.1.3, S.6

[2] C. Olbrich, T. L. C. Jansen, J. Liebers, M. Agthar, J. Strümpfer, K. Schulten, J. Knoester and U. Kleinekathöfer, *J. Phys. Chem. B*, 2011, **115** 8609 – 21. S.1.3, S.6

[3] S. Jurinovich, C. Curutchet and B. Mennucci, *ChemPhysChem*, 2014, **15**, 3194 – 3204. S.1.3, S.6

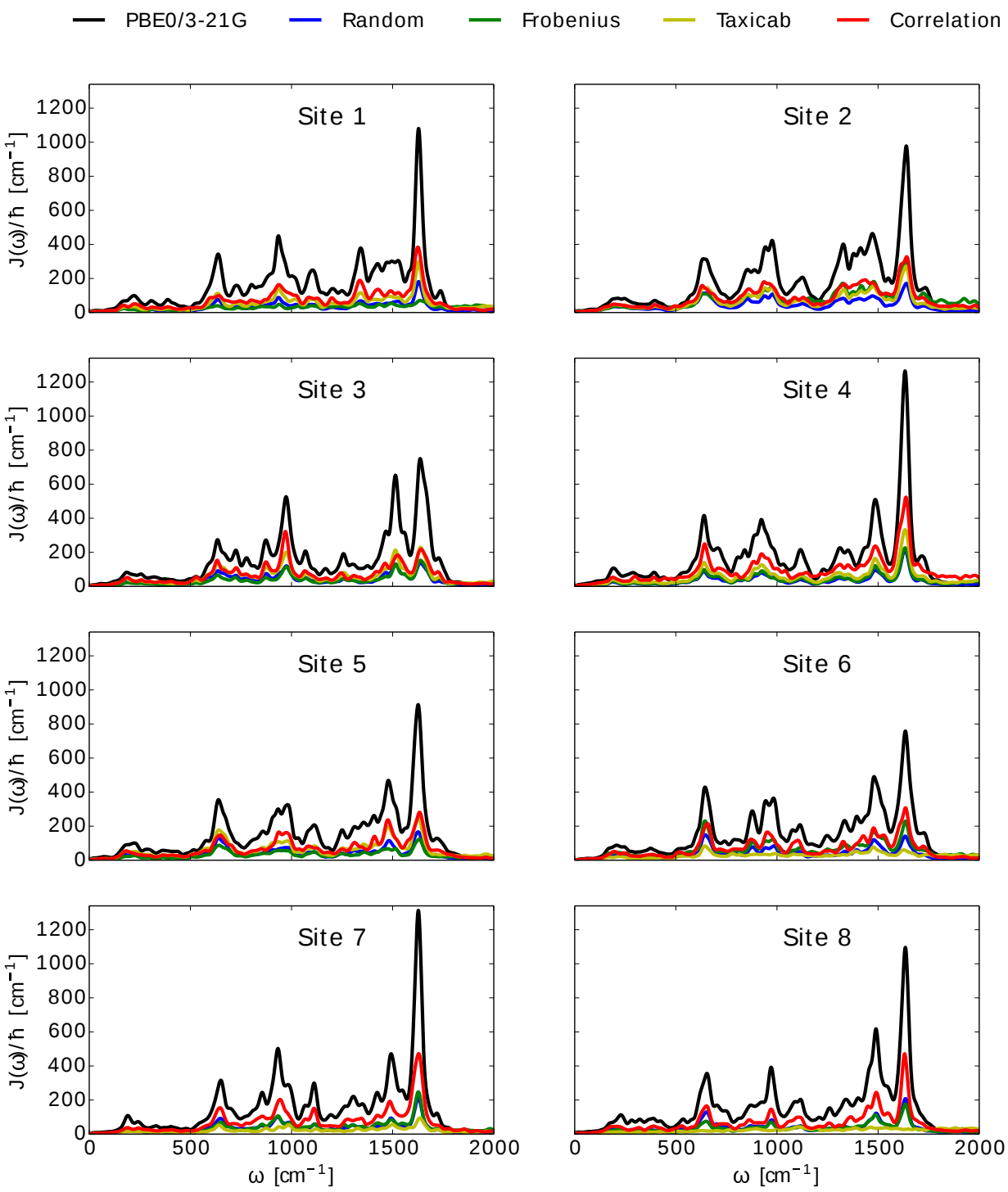

Figure S.11: Harmonic spectral densities for individual sites in the FMO complex. The spectral densities were calculated from excited state energy trajectories obtained from TDDFT calculations (PBE0/3-21G) and compared to spectral densities from neural network predicted excited state energy trajectories. Neural networks were trained on the BChl they predicted with the indicated Coulomb matrix selection method.