## EDGE ARTICLE

CrossMark
click for updates

# Machine learning exciton dynamics†

Florian Häse,[ab] Stéphanie Valleau,[a] Edward Pyzer-Knapp[a] and Alán Aspuru-Guzik[*a]

Obtaining the exciton dynamics of large photosynthetic complexes by using mixed quantum mechanics/molecular mechanics (QM/MM) is computationally demanding. We propose a machine learning technique, multi-layer perceptrons, as a tool to reduce the time required to compute excited state energies. With this approach we predict time-dependent density functional theory (TDDFT) excited state energies of bacteriochlorophylls in the Fenna–Matthews–Olson (FMO) complex. Additionally we compute spectral densities and exciton populations from the predictions. Different methods to determine multi-layer perceptron training sets are introduced, leading to several initial data selections. In addition, we compute spectral densities and exciton populations. Once multi-layer perceptrons are trained, predicting excited state energies was found to be significantly faster than the corresponding QM/MM calculations. We showed that multi-layer perceptrons can successfully reproduce the energies of QM/MM calculations to a high degree of accuracy with prediction errors contained within 0.01 eV (0.5%). Spectral densities and exciton dynamics are also in agreement with the TDDFT results. The acceleration and accurate prediction of dynamics strongly encourage the combination of machine learning techniques with *ab initio* methods.

## 1 Introduction

Studying the exciton dynamics of large photosynthetic complexes, such as the Fenna–Matthews–Olson (FMO) and light-harvesting II (LHII) complexes, has been a topic of much theoretical and experimental interest in recent years.[1–14] The theoretical community has focused on employing and developing reduced models to understand and describe the dynamics of these complexes.[2–7,13–30] These models rely on the knowledge of a system Hamiltonian for the interacting chromophores as well as spectral densities to describe the coupling of chromophores to their environments (protein, solvent).[31] Computing a system Hamiltonian or spectral density is an arduous computational task due to the large number of degrees of freedom in these complexes. The most detailed approaches used to obtain these quantities have been mixed quantum mechanics/molecular mechanics (QM/MM) or semi-classical simulations.[32,33] In particular, one QM/MM approach which has become popular in recent years consists in propagating the nuclei in the electronic ground state of the photosynthetic complex.[9,10,34,35] This approximation ignores the change in electronic structure due to excitation of the chromophores. Subsequently, for a subset of time frames, excited state energies for the chromophores are computed using a quantum method

such as time-dependent density functional theory (TDDFT).[10] The energy trajectories are then employed to extract system-bath correlation functions and finally spectral densities to use in exciton dynamics.

The downside of this approach is the large computational cost. Long molecular dynamics (MD) equilibration times of several tens of nanoseconds are required.[36,37] The typical computational scaling of MD codes with the system size $N$ is $\mathcal{O}(N \log N)$.[38] In contrast, TDDFT calculations scale as $\mathcal{O}(N^2)$.[39] Very often calculations need to be repeated for identical chromophores in similar environments to account for the effect of small variations. For instance in the case of a single-point mutation typically one would need to rerun the entire set of simulations.

In this work we propose an alternative route: using multi-layer perceptrons, a special class of neural networks, to predict the excited state along a MD trajectory. Such approaches typically scale as $\mathcal{O}(N)$ and were found to perform significantly faster than TDDFT approaches. As a test system we consider the Fenna–Matthews–Olson (FMO) complex of *P. aestuarii*. Neural networks have previously been used to predict potential-energy surfaces.[40] We use multi-layer perceptrons as fully connected neural networks to predict the values of the first singlet excited state for the chromophores. We train the neural networks on the excited state energies obtained from QM/MM calculations. Several sampling methods are used to select the training data for the neural networks. In particular we tested a sampling method based on correlations of nuclear positions to improve on the spectral density predictions. Once trained, the neural networks are employed to make excited state energy

*[a]Department of Chemistry and Chemical Biology, Harvard University, Cambridge, 02138, USA. E-mail: aspuru@chemistry.harvard.edu; Tel: +1-617-384-8188*

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

predictions. Then one can build a Hamiltonian from the predictions and compute the exciton dynamics.

With optimal neural network training and 12 trained neural networks per BChl we predicted excited state energies with errors contained to 0.01 eV (0.5%) from the neural network ensemble average. Further, with neural networks trained on data based on correlation sampling we correctly predict the shape of the spectral density and observe an error which is squared with respect to the excited state prediction error. This demonstrates the power of machine learning in chemistry, as has also been found in recent work where machine learning was employed to extract other chemical properties.[41–44]

# 2 Methods and computational details

## 2.1 Ground state QM/MM simulations

A semi-classical description of the FMO complex was obtained by combining ground state MD simulations with TDDFT calculations of the first singlet excited state, known as the $Q_y$ state,[45] at given molecular conformations along the time-dependent trajectories. The MD runs were carried out using the NAMD software package[46] with the AMBER99 force field (ff99SB).[47] The BChl-a parameters employed are reported in ref. 48. The X-ray crystal structure of the FMO trimer in *P. aestuarii* (PDB 3EOJ,[49] see Fig. 1) was chosen as initial configuration. The trimer was solvated using a TIP3P water box.[50] The minimum distance between the crystal structure and edges of the solvent box was taken to be 15 Å.[37,51] The charge was neutralized by adding sodium ions. The total number of atoms in the system was 141624. The simulation was equilibrated for 4 ns and the production run was 40 ps long with a 2 fs time-step. Electrostatic interactions were calculated with the particle-mesh Ewald method. Shake constraints were used for all bonds containing hydrogen. Simulations were carried out at 300 K and 1 atm using a Langevin thermostat and a Langevin piston as implemented in NAMD.

The time-dependent $Q_y$ excited ground state gaps of the BChl-a molecules were obtained using TDDFT with the PBE0 functional,[52] within the Tamm–Dancoff Approximation (TDA),[53] using Q-Chem.[54] We employed the 3-21G basis set due to the high computational cost of these simulations. The $Q_y$ excited

state was taken to be the one with the largest oscillator strength and the orientation of the transition dipole was verified using the same methodology as in ref. 37 and 55. Excited state energy distributions are shown in the ESI (see Section S.1.4†) and trajectories can be downloaded from ref. 56. Excited state energies were computed at every 4 fs of the production run. The number of TDDFT calculations performed initially is based on the time scale and time step required to reproduce the dynamics accurately. Some values were excluded based on the oscillator strength/angle criterion/failed convergence. The excluded values were at most 2.15% of the full trajectory.

## 2.2 Machine learning: neural networks training

**2.2.1 Input data representation.** Multi-layer perceptrons (neural networks) were used to predict quantum mechanical excited state energies of BChls in the FMO complex from the MD classical coordinate trajectory. Neural networks were trained in a supervised training scheme using the back propagation algorithm.[57] Excited state energies from the TDDFT calculations described previously were provided to the neural network as targets. BChl conformations were represented by Coulomb matrices,[43] which were successfully used to predict excited state energies of various molecules with one neural network.[58] Both, input and target feature distributions, were rescaled to a zero mean and a unitary standard deviation prior to neural network training.[59]

By using Coulomb matrices as input features, neural networks can be trained on a representation of BChls which is translation and rotation invariant. Coulomb matrices are particularly suitable to describe BChls in the FMO complex as these molecules do not undergo large conformational changes within time scales of several tens of picoseconds.[36,37,60] Coulomb matrices were adapted to account for external charges within and around the represented BChl. The electrostatic influence of particles in the environment $N$ was described by adding additional Coulomb potential terms to the corresponding Coulomb matrix entries (see eqn (1)):

$$M_{ij} = \begin{cases} Z_i^{2.4}/2 + \sum_{n \in N} \dfrac{Z_i Z_n}{|\vec{r}_i - \vec{r}_n|} & \text{for } i = j, \\[2ex] \dfrac{Z_i Z_j}{|\vec{r}_i - \vec{r}_j|} + \sum_{n \in N} \dfrac{Z_i Z_n}{|\vec{r}_i - \vec{r}_n|} + \sum_{n \in N} \dfrac{Z_j Z_n}{|\vec{r}_j - \vec{r}_n|} & \text{for } i \neq j. \end{cases} \quad (1)$$

Partial charges $Z_i$ of atoms in the system were taken from the system topology (Amber 99SB force field,[47] and ref. 48). Studies have shown that the tails of the BChls have little influence on the $Q_y$ excited state energies.[37,61] Thus, instead of representing the entire BChl in a Coulomb matrix, the phytyl tail was neglected and only the 90 atoms closest to the magnesium in the BChls were represented in Coulomb matrices to reduce their dimensionality. We included all external partial charges present in the system to generate Coulomb matrices.

**2.2.2 Neural network architecture, choice of BChl molecule and over-fitting.** We chose to use multi-layer perceptrons (neural networks) with logistic activation functions and two hidden layers. This set-up has been shown to perform
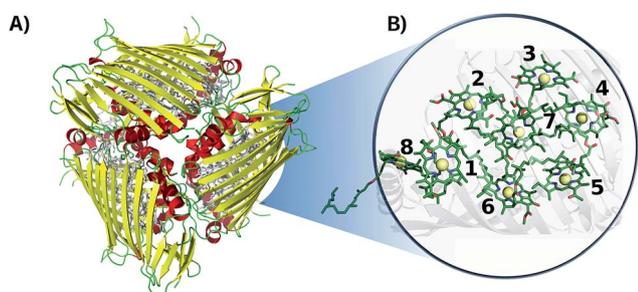


Fig. 1 Crystal structure of the FMO complex in *P. aestuarii* (PDB 3EOJ).[49] (A) 3EOJ trimer crystal structure. (B) BChls' geometric arrangement in monomer A (residues 360 to 367, corresponding to sites 1 to 8). Hydrogens are not shown in this representation.

particularly well for supervised regression problems.[62] Optimal neural network hyperparameters were identified from a grid search. Both the learning rate and the number of neurons in the first and second hidden layer were varied to find the lowest deviations between predictions and target data.

Instead of performing the grid search on each BChl in the FMO complex individually, only the most representative BChl was used to determine optimal neural network hyperparameters to reduce the computational cost. We identified this BChl in terms of shared Coulomb matrix space. Coulomb matrices of all eight BChls were clustered and compared. We found that site 3 shares the most Coulomb matrix space with all other sites (see ESI Section S.1.1†).

From the grid search, we found that a learning rate of $10^{-4}$ with 204 neurons in the first hidden layer and 192 neurons in the second hidden layer results in the smallest average absolute deviation of predicted and target excited state energies.

Target feature over-fitting was avoided by using *early stopping*.[63] For all training sessions a total of 4000 trajectory frames was assigned to the training set as a balance between information and computational cost. Neural networks were trained on Intel(R) Xeon(R) CPUs (X5650 @ 2.67 GHz). Training one neural network on four cores took about $(23.9 \pm 5.0)$ h. Training times for other investigated training set sizes ranging from 500 to 5000 frames are reported in the ESI (see Section S.1.2†).

**2.2.3 Reducing training set redundancies through clustering: taxicab, Frobenius and "correlation" clustering.** We employed different methods to select Coulomb matrices and corresponding excited state energies for neural network training. In a first approach training set Coulomb matrices were drawn randomly from the entire trajectory.[63] This led to excited state energy predictions with an average accuracy of 13 meV.

However, as BChl conformations in the data set are not uniformly distributed a training set consisting of randomly drawn Coulomb matrices likely contains redundant information. MD simulations were carried out in the NPT ensemble with constant temperature and pressure. Thus, the BChl conformations are sampled from a Boltzmann distribution.[64] To avoid selecting many similar conformations and thus similar Coulomb matrices we performed a cluster analysis on all Coulomb matrices of the entire trajectory to determine the most distinct Coulomb matrices. We implemented a Coulomb matrix cluster algorithm following the principles of the gromos method.[65]

Distances between Coulomb matrices were measured using $p$-norms. Two different metrics were applied: $p = 1$ (taxicab norm) and $p = 2$ (Frobenius norm). Both clustering approaches resulted in more accurate predictions of excited state energies with accuracies of 9 meV but the prediction of exciton dynamics remained quite inaccurate (see Section 3.2).

To improve the prediction of exciton dynamics, we developed a clustering method based on coordinate correlations in the classical MD trajectory. We will refer to this approach as the "*correlation*" clustering method. The $Q_y$ state in BChls is mostly distributed along one of the diagonals which connects opposite nitrogen atoms.[37] Training set frames were thus selected based on high correlations in the nitrogen root-mean square deviation (RMSD). In particular, for the $n$-th BChl we sampled from

$$|C_n^{\mathrm{RMSD}}(t)|^2 = |\langle D_n^{\mathrm{nitrogen}}(t)D_n^{\mathrm{nitrogen}}(0)\rangle|^2. \quad (2)$$

until 4000 frames with the largest RMSD correlation were selected. Here, $D_n^{\mathrm{nitrogen}}(t) = \sqrt{\frac{1}{4}\sum_{i=1}^{4}\|\vec{r}_{n,i}(t) - \vec{r}_{n,i}(0)\|^2}$ refers to the root-mean square difference in position of the four nitrogen atoms in the $n$-th BChl at a given time $t$ with respect to their position in the energy minimized crystal structure at time $t = 0$. This sampling led to a more accurate prediction of the spectral density (see Section 3.2).

### 2.3 Exciton dynamics and spectral densities

To further compare the predicted excited state energy trajectories with the TDDFT trajectories we computed the exciton dynamics in the FMO complex using two different methods. The first is a stochastic integration of the Schrödinger equation as used in previous work.[10] The second method is the Markov Redfield master equation.[66] Both of the methods are Markovian but the first relies only on the excited state energy trajectories while the latter also depends on the spectral density. Here we focus on the sensitivity of these methods to the changes related to using neural networks rather than to more subtle questions on dynamics such as those addressed by the comparison of Markovian with non Markovian methods (*e.g.* the hierarchy equation of motion approach).[2]

Finally the spectral density $j(\omega)$, as used in the Redfield equations of motion (see eqn (3)), is obtained by normalizing the Fourier transform of the two-time correlation function as we discussed in previous work.[31]

$$j(\omega) = \frac{J^{\mathrm{harm}}(\omega)}{\pi} \quad \text{with} \quad J^{\mathrm{harm}}(\omega) = \frac{\beta\hbar\omega}{2}\int_{-\infty}^{\infty} \mathrm{e}^{i\omega t}C^{\mathrm{cl}}(t)dt. \quad (3)$$

The superscript "harm" refers to the harmonic prefactor which is needed to connect the QM/MM results to the open quantum system approach. Following the notation of previous work,[31] $C^{\mathrm{cl}}(t)$ denotes the classical correlation function.

## 3 Results and discussion

### 3.1 Excited state energy prediction using neural networks

**3.1.1 Acceleration of excited state energy computations with neural networks.** Intel(R) Xeon(R) CPUs (X5650 @ 2.67 GHz) were used to train neural networks and predict excited state energies. A total of 12 neural networks was trained for each of the eight BChls in monomer A of the FMO complex. Predictions of each of the 12 neural networks per BChl were averaged in a neural network ensemble averaging approach to obtain a more accurate prediction for the excited state energy trajectory. The spread of predictions of individual neural networks is given in the ESI (see Section S.1.3†).

Training sets were generated with all four training set selection methods (see Section 2.2.3). Predicting $Q_y$ excited state

**Table 1** Time required to compute excited state energies (10000 frames) for all bacteriochlorophylls (BChls) for TDDFT (PBE0/3-21G) and neural network (NN) predictions from correlation clustered Coulomb matrices. Reported times include neural network training ($t_{train}$) on 4000 frames with input ($t_{Coul}^{input}$) and target feature ($t_E^{target}$) generation, excited state calculations/predictions ($t_{Calc}$) and the total time ($t_{tot}$). If trained neural networks are available, only Coulomb matrices need to be calculated for neural network predictions, reducing the required time to 48 h. Reported times correspond to training a total of 12 neural networks independently to obtain ensemble averaged excited state energies. All reported times refer to calculations on a single core of an Intel(R) Xeon(R) CPU (X5650 @ 2.67 GHz)

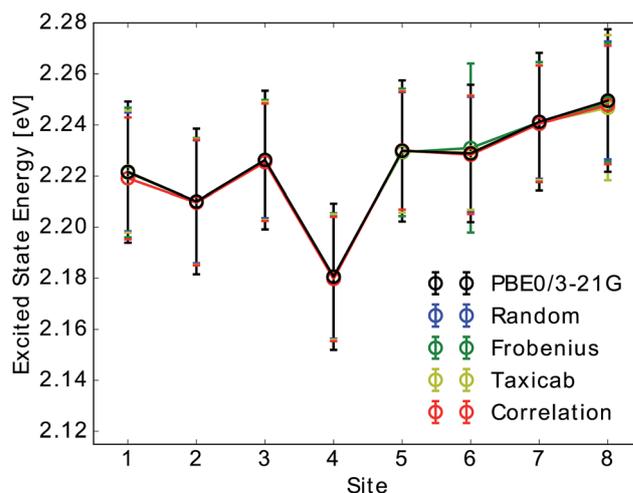| Method | Training [h] | | | Calculation [h] | Total [h] |
|---|---|---|---|---|---|
| | $t_{Coul}^{input}$ | $t_E^{target}$ | $t_{train}$ | $t_{Calc}$ | $t_{tot}$ |
| PBE0/3-21G | — | — | — | 480000 | 480000 |
| $NN_{Corr}$ | 48 | 192000 | 9178 | <0.1 | 201226 |



**Fig. 2** Mean and standard deviation of $Q_y$ excited state energy distributions for all eight sites obtained from TDDFT calculations (PBE0/3-21G) and compared to neural network predictions. Presented distributions were obtained from single neural network predictions. Neural networks were trained on Coulomb matrices selected from the classical MD trajectory by the indicated selection method. Error bars indicate the width of the excited state energy distribution.

energies for the entire trajectory ($10^4$ frames) for one BChl with one neural network took on average $(3.9 \pm 0.8)$ s on one core. In contrast, the quantum chemistry calculations using the TDDFT (PBE0/3-21G) model chemistry required approximately 60 000 h for the entire trajectory on one core. Target feature generation is only necessary if trained neural networks are not available and independent of the number of frames to be predicted, thus reducing the reported costs if longer trajectories are predicted. Required calculation times to compute excited state energy trajectories for each of the eight BChls in the FMO complex are reported in Table 1. While two cores of an Intel(R) Xeon(R) CPU (X5650 @ 2.67 GHz) were used for TDDFT calculations and neural networks were trained on four cores the recorded calculation times were multiplied by the number of used cores for comparison reasons.

If trained neural networks are available, excited state energies of given BChl conformations can be predicted directly from Coulomb matrices representing the BChl conformations. Thus, to obtain the excited state energies of an entire trajectory consisting of 10 000 frames, Coulomb matrices need to be calculated first. With a calculation time of 2.19 s per Coulomb matrix on one Intel(R) Xeon(R) CPU (X5650 @ 2.67 GHz) core, about 6 h are needed to compute all 10 000 Coulomb matrices. With excited state energy predictions requiring less than a minute, excited state energy trajectories can be obtained from neural networks about four orders of magnitude faster compared to TDDFT calculations.

**3.1.2 Accuracy of neural network predictions.** As can be seen in Fig. 2 neural network predictions agree well with the TDDFT data. Predictions from the average over 12 networks deviate from TDDFT values by ~0.3 meV regardless of the site and the input selection method. Ensemble averaging was found to improve the prediction accuracy for all sites (see Table S.2†). When considering excited state energies predicted from a single network, we observed a deviation of $\sigma_{s,random}^{single} < 14$ meV for all sites s when the neural network was trained on randomly drawn Coulomb matrices. This prediction error can be decreased to $\sigma_{s,taxicab}^{single} < 9$ meV by selecting the training set based on the

Coulomb matrix space clustering taxicab method (see Section 2.2.3). The uncertainty of neural network predictions therefore corresponds to roughly 14% of the typical FWHM (64 meV) of the sites. Frobenius clustering showed similar deviations. In contrast, predictions from neural networks trained on correlation clustered Coulomb matrices show a slightly higher deviation of $\sigma_{s,correlation}^{single} < 15$ meV on average. The obtained prediction errors are about one order of magnitude smaller than those in ref. 58. This difference is due to the fact that we predict excited state energies of the same molecule in different time-dependent geometries while in ref. 58 the neural network was trained to predict the excited state energies of many different molecules.

**3.1.3 Cross-predictions: predicting excited state energies for other bacteriochlorophylls.** Since all BChl-a molecules in the FMO complex consist of the same atoms and show similar geometrical conformations, we also used neural networks trained on one BChl to predict excited state energies of other BChls in the same monomer. This enabled us to understand how well the trained network can adapt to changes in the environment from changes in the Coulomb matrices. We observed that for any clustering method, the prediction error is about two times larger when performing this type of cross-prediction, see Fig. 3. Nonetheless, as we see in panel A, the largest observed average absolute deviation is still below 1.14% (corresponding to 25 meV).

**3.1.4 Predicting excited state energies of bacteriochlorophylls in other monomers.** All neural networks were trained on BChl-a molecules in monomer A. These neural networks were then used to predict excited state energies of BChls in the other two FMO monomers. Each neural network predicted the excited state energies of the BChl corresponding to the one on which it was trained (i.e. a neural network trained on site 1 in monomer A predicted site 1 in monomer B and C).
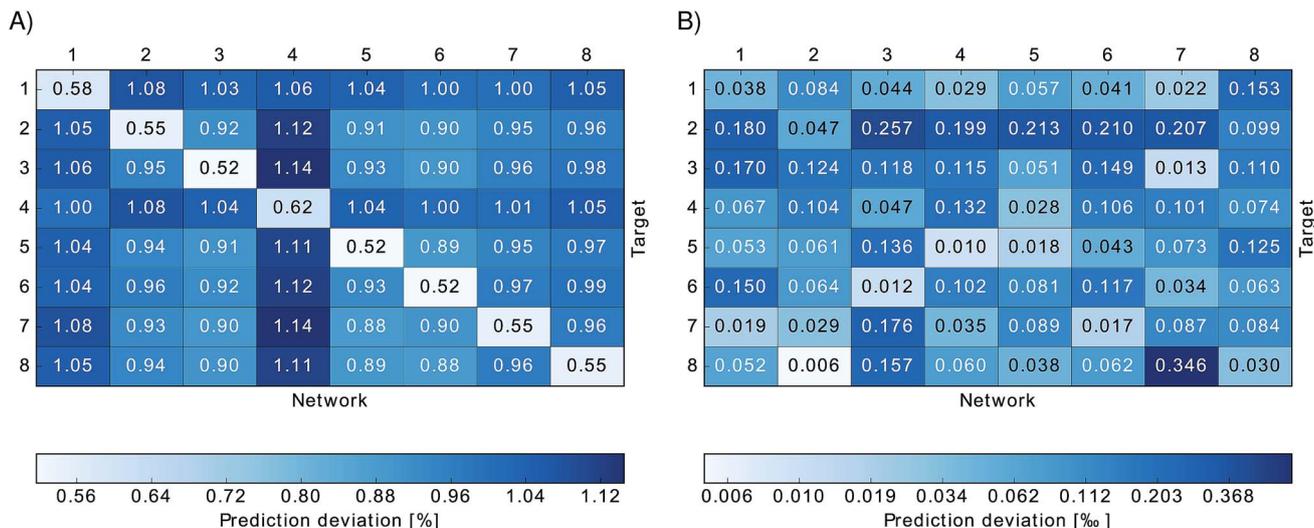
A)



B)



**Fig. 3** Relative absolute deviations of predicted excited state energies from TDDFT excited state energies. Neural networks trained on one particular site (indicated by "Network") were used to predict excited state energies of another site (indicated by "Target"). Panel (A) shows the relative absolute deviation $\bar{\sigma}_{BChl}^{rel}$ of predicted excited state energies from TDDFT excited state energies for each BChl, $\bar{\sigma}_{BChl}^{rel} = \sum_i |\varepsilon_{BChl}^{NN}(t_i) - \varepsilon_{BChl}^{TDDFT}(t_i)|/(N_{frames}\langle\varepsilon_{BChl}^{TDDFT}\rangle)$, in percent. Panel (B) shows the deviation $\sigma_{mean}^{mean}$ of the mean of the predicted excited state energies from the mean of the TDDFT calculated excited state energies, $\sigma_{mean}^{BChl} = |\langle\varepsilon_{BChl}^{NN}\rangle - \langle\varepsilon_{BChl}^{TDDFT}\rangle|/\langle\varepsilon_{BChl}^{TDDFT}\rangle$ in per-thousand.

Due to the fact that the FMO complex is a homo-trimer, similar BChl conformations should be sampled during the MD simulation in each monomer. Thus, excited state energy averages are expected to be identical for corresponding BChls of different monomers, provided that the same phase space regions were covered in the simulation time. The results are presented in Fig. 4. In that figure, we show the time-averaged excited state energy predicted from 12 independently trained

neural networks for each monomer as well as the time-averaged $Q_y$ energies obtained from TDDFT for monomer A. The bars represent the spread of the distribution and not an error. The predicted distributions are narrower than the TDDFT distributions. This is probably due to the fact that frames corresponding to energies at the tails of the distribution are less sampled. Regarding the error on the other hand, the largest deviation encountered between mean values of excited state energy distributions is 3 meV.

We have found that trained neural networks can accurately predict TDDFT excited state energies. This allows for a large reduction of computational time. It is possible to train on a single BChl and predict excited state energies of other BChls in the same monomer or in different monomers.

### 3.2 Spectral densities and exciton dynamics with neural networks

We then used the calculated excited state energy trajectories of all BChls to obtain information about interactions of the BChls with their environment by computing spectral densities. In addition, we built a Hamiltonian to extract the exciton dynamics in the system.

The spectral density $J^{harm}(\omega)$ (see eqn (3)) was computed for all eight sites in the FMO complex from the TDDFT excited state energies and neural network predicted excited state energies.[31] Spectral densities for each site were averaged over all BChls to obtain an averaged spectral density $J_{ave}(\omega)$. To minimize spurious effects in the Fourier transform, we multiplied the correlation function by a Gaussian of $\sigma^2 = 0.09t_{max}^2$ with $t_{max} = 1600$ fs.[31] The Gaussian is normalized to have unitary area in frequency domain so that in frequency domain this corresponds to a convolution with a Gaussian with a FWHM of 26 cm$^{-1}$.
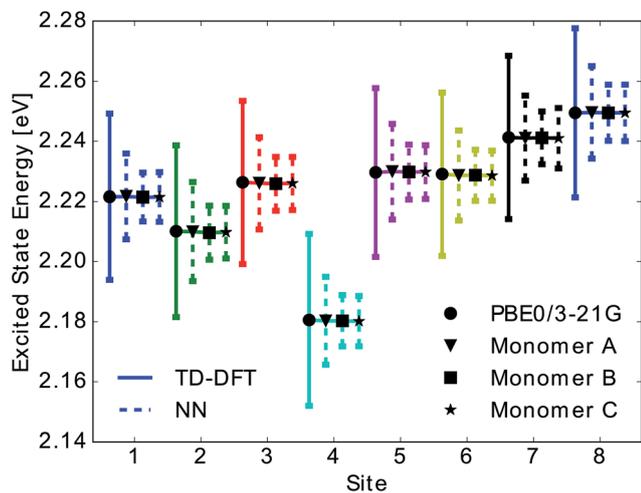


**Fig. 4** Mean and standard deviation of excited state energy distributions for all eight sites obtained using TDDFT PBE0/3-21G (solid line with circle), for monomer A, and using neural network prediction with training on 40% of all Coulomb matrices of monomer A for all three monomers (dashed lines and triangle, square and star symbol for monomers A, B, C). Presented distributions were obtained from the ensemble averaged predictions of 12 neural networks. Error bars indicate the standard deviation of the obtained excited state energy distributions.
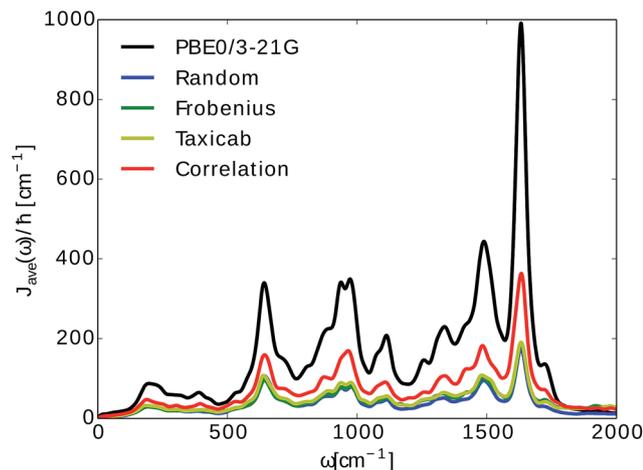
**Fig. 5** Spectral density averages $J_{ave}(\omega) = \sum\limits_{BChl} J_{BChl}^{harm}(\omega)/N_{BChl}$. The spectral densities were computed from excited state energy trajectories obtained from TDDFT calculations (PBE0/3-21G) and compared to spectral densities from neural network predicted excited state energy trajectories. Neural networks were trained on the bacteriochlorophyll they predicted with the indicated Coulomb matrix selection method. The correlation method (see Section 2.2.3) gives the best prediction.

In Fig. 5 we show the comparison to our neural network prediction with training and prediction on the same site and the various Coulomb matrix selection methods. We found that predicted spectral densities all have a shape which resembles the overall shape of the TDDFT spectral density. However, the height and accurate position of the peaks in the spectral density is most accurately predicted using correlation clustering.

Spectral densities for each BChl, calculated from TDDFT excited state energies and excited state energies predicted from neural networks trained on all input selection methods are plotted in the ESI (see Fig. S.11 in Section S.1.5†).

Average spectral densities were used to calculate the reorganization energy $\lambda = \int_0^\infty J_{ave}(\omega)/\omega\, d\omega$. Comparisons of reorganization energies are reported in Table 2. We observe that the smallest deviation between neural network predicted results and TDDFT results occurs for neural networks trained on correlation clustered Coulomb matrices.

We also computed the population dynamics in the FMO complex monomer with a stochastic integration method.[10] We averaged 4000 stochastic trajectories to obtain converged population dynamics. The initial excited site was chosen to be site 1 and the dynamics was propagated for all eight coupled sites at

**Table 2** We report the percentage deviation of neural network predicted reorganization energies $\lambda_{NN,\beta}$ from TDDFT (PBE0/3-21G) calculated reorganization energies $\lambda_{TDDFT}$ as $\sigma_\lambda = |\lambda_{TDDFT} - \lambda_{NN,\beta}|/\lambda_{TDDFT}$, where $\beta$ indicates the method used for Coulomb matrix selections. Results are given in percent (%)

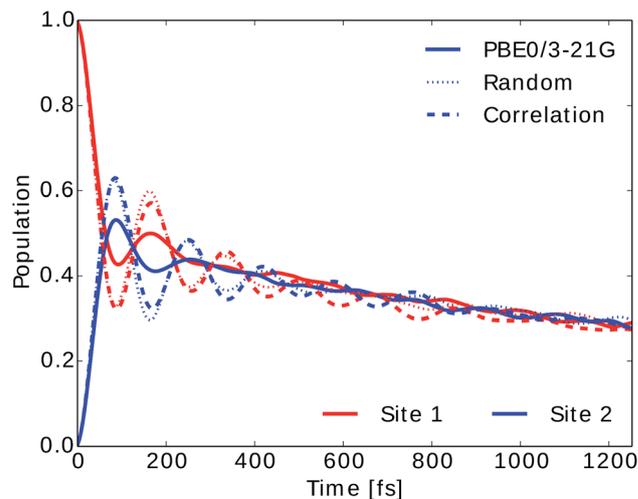| Method ($\beta$) | Random | Frobenius | Taxicab | Correlation |
|---|---|---|---|---|
| $\sigma_\lambda$ [%] | 70.0 | 54.9 | 53.4 | 46.6 |



**Fig. 6** Population dynamics of the FMO complex calculated at 300 K with initial state in site 1. Here only two site populations are shown but dynamics was carried out for the 8 BChls. Excited state energy trajectories were obtained from TDDFT calculations (PBE0/3-21G) as well as neural networks trained on randomly drawn and correlation clustered Coulomb matrices.

300 K. The couplings of the Hamiltonian were taken from ref. 55 and 67. Results are shown in Fig. 6. We see that neural network predictions from neural networks trained on randomly drawn and correlation clustered Coulomb matrices predict the exciton dynamics in agreement with TDDFT calculations.

We also employed the Markovian Redfield method to compute the exciton dynamics in the FMO complex (see Section 2.3).[66] In this case there is an explicit dependence of the exciton dynamics on the spectral density. The energies in the Hamiltonian were taken to be the averages from the TDDFT or neural network predicted excited state energy trajectories. The same couplings as for the stochastic integration method were used.

To investigate the importance of excited state energies and spectral densities on the Redfield exciton dynamics we calculated the exciton dynamics in two different ways. First, we computed the exciton dynamics with neural network predicted excited state energies and the average spectral density obtained from TDDFT calculations and then we used the neural network predicted energies as well as the neural network predicted spectral densities. Results are presented in Fig. 7 panel A, for predicted excited state energies and TDDFT spectral densities and panel B, for excited state energies and spectral densities predicted by neural networks. We initialized the dynamics with the excitation in BChl 1 and propagated the dynamics for all eight coupled sites at 300 K.

In panel A we see that given a constant spectral density the error on energies is small and does not strongly influence the exciton dynamics. The main role is played by the spectral density as can be seen in panel B. Here we see much larger differences depending on the neural network sampling method and we notice that the prediction of neural networks trained on correlation clustered Coulomb matrices agrees better with the TDDFT exciton dynamics. Deviations of neural network predicted exciton dynamics from TDDFT calculated exciton
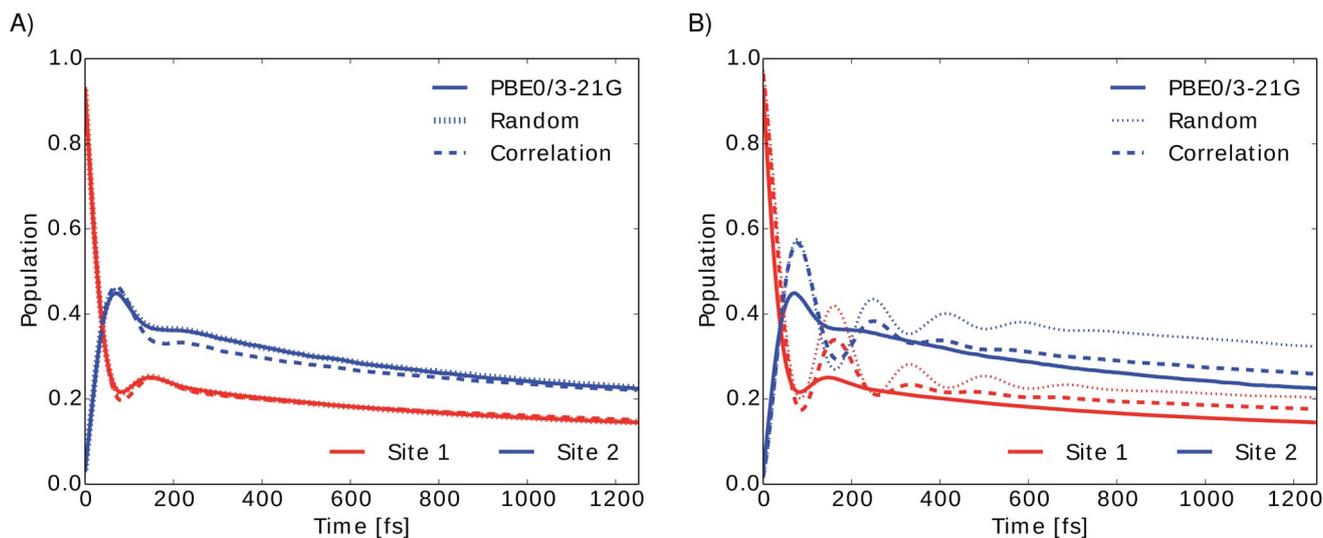
**Fig. 7** Time evolution of the exciton population for BChl 1 (red) and BChl 2 (blue) in the FMO complex calculated from excited state energy trajectories and average spectral densities using the Redfield method. The initial state is site 1 excited. Panel (A) shows the exciton dynamics for neural network predicted excited state energies using the same TDDFT calculated average spectral density in all cases. Panel (B) shows the exciton dynamics with both, excited state energy trajectories and harmonic average spectral densities predicted by neural networks trained with the indicated selection method.

dynamics are listed in detail in the ESI (see Section S.1.5†). Further methods should be investigated to improve the prediction of the spectral density, the error on the energy should be reduced further by an order of magnitude to predict spectral densities with the optimal reorganization energy.

## 4 Conclusion

The computational study presented in this article showed that multi-layer perceptrons (neural networks) can be used to successfully predict excited state energies of BChls in the FMO complex. Using different methods to select a training data set for neural network training from features in the classical MD trajectory alone we were able to generate neural networks which can predict TDDFT excited state energies with high accuracy (0.01 eV).

Furthermore, the neural networks can predict properties such as QM/MM derived spectral densities or exciton population dynamics. The prediction of excited state energies using the neural networks is about seven orders of magnitude faster than TDDFT. If we include training feature generations we still observe a speed-up of about four orders of magnitude. Even if neural networks need to be trained first, excited state energies are obtained in less than half the time needed for TDDFT. Based on the observations we made on the FMO complex we recommend the following procedure to apply machine learning to predict excited state properties of other systems:

(1) Represent all molecules of interest and their chemical environment with Coulomb matrices.

(2) Obtain optimal neural network architectures from hyperparameter grid searches. In particular, it is important identify the most representative molecule in terms of the space which contains the features used for neural network predictions.

(3) Determine an optimal training set. Training sets for neural networks with optimal network architecture can be generated by selecting Coulomb matrices based on properties which are related to the desired quantum mechanical properties. We observed that selecting Coulomb matrices which represent Coulomb matrix space clusters improves excited state energy predictions. However, these do not necessarily work well for dynamics. Investigate on the role of higher sampling of the tails of the energy distribution when selecting the training set.

(4) Make predictions. To predict spectral densities and exciton dynamics with high accuracy, Coulomb matrices should be selected for the training set if they reveal high excited state energy correlations. We found that nitrogen RMSD correlations in the BChls are a good indicator for excited state energy correlations. Thus, we selected Coulomb matrices based on high nitrogen RMSD correlations. Of course this might be complicated for some molecules with very delocalized excited states.

In conclusion, this approach provides a gigantic speedup to ground state QM/MM. From a neural network trained on a single BChl molecule, we can predict the excited states of 23 other molecules in the system at very low additional computational costs. This will be particularly useful to speed up or simply to enable the simulation of larger, more complex and challenging light-harvesting systems. Further, it will be helpful to study, for instance, the role of small changes in the environment on the exciton dynamics. Future questions we would like to address include the possibility of extending the prediction to other temperatures using for example multi-target machine learning.

## Acknowledgements

# References

1 J. Adolphs and T. Renger, *Biophys. J.*, 2006, **91**, 2778–2797.

2 A. Ishizaki and G. R. Fleming, *Proc. Natl. Acad. Sci. U. S. A.*, 2009, **106**, 17255–17260.

3 C. Kreisbeck, T. Kramer, M. Rodriguez and B. Hein, *J. Chem. Theory Comput.*, 2011, **7**, 2166–2174.

4 M. Sarovar, G. R. Fleming and K. B. Wahley, *Nat. Phys.*, 2010, **6**, 462–467.

5 J. Moix, J. Wu, P. Huo, D. Coker and J. Cao, *J. Phys. Chem. Lett.*, 2011, **2**, 3045–3052.

6 M. Mohseni, P. Rebentrost, S. Lloyd and A. Aspuru-Guzik, *J. Chem. Phys.*, 2008, **129**, 174106.

7 M. B. Plenio and S. F. Huelga, *New J. Phys.*, 2008, **10**, 113019.

8 B. P. Krueger, G. D. Scholes and G. R. Fleming, *J. Phys. Chem. B*, 1998, **102**, 5378–5386.

9 C. Olbrich and U. Kleinekathöfer, *J. Phys. Chem. B*, 2010, **114**, 12427–12437.

10 S. Shim, P. Rebentrost, S. Valleau and A. Aspuru-Guzik, *Biophys. J.*, 2012, **102**, 649–660.

11 G. D. Scholes, C. Curutchet, B. Mennucci, R. Cammi and J. Tomasi, *J. Phys. Chem. B*, 2007, **111**, 6978–6982.

12 P. Huo and D. F. Coker, *J. Chem. Phys.*, 2010, **133**, 184108.

13 J. Cao and R. J. Silbey, *J. Phys. Chem. A*, 2009, **113**, 13825–13838.

14 J. Wu, F. Liu, Y. Shen, J. Cao and R. J. Silbey, *New J. Phys.*, 2010, **12**, 105012.

15 S. Jang, M. D. Newton and R. J. Silbey, *J. Phys. Chem. B*, 2007, **111**, 6807–6814.

16 P. Rebentrost, M. Mohseni, I. Kassal, S. Lloyd and A. Aspuru-Guzik, *New J. Phys.*, 2009, **11**, 033003.

17 A. Ishizaki and G. R. Fleming, *J. Phys. Chem. B*, 2011, **115**, 6227–6233.

18 D. Abramavicius and S. Mukamel, *J. Chem. Phys.*, 2010, **133**, 064510.

19 N. Skochdopole and D. A. Mazziotti, *J. Phys. Chem. Lett.*, 2011, **2**, 2989–2993.

20 G. Ritschel, J. Roden, W. T. Strunz, A. Aspuru-Guzik and A. Eisfeld, *J. Phys. Chem. Lett.*, 2011, **2**, 2912–2917.

21 P. Rebentrost and A. Aspuru-Guzik, *J. Chem. Phys.*, 2011, **134**, 101103.

22 N. Singh and P. Brumer, *Faraday Discuss.*, 2011, **153**, 41–50.

23 N. Singh and P. Brumer, arXiv:1106.5911v1, 2011.

24 L. A. Pachón and P. Brumer, *Phys. Chem. Chem. Phys.*, 2012, **14**, 10094–10108.

25 S. M. Vlaming and R. J. Silbey, *J. Chem. Phys.*, 2012, **136**, 055102.

26 F. Caruso, S. K. Saikin, E. Solano, S. Huelga, A. Aspuru-Guzik and M. B. Plenio, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2012, **85**, 125424.

27 M. Mohseni, A. Shabani, S. Lloyd and H. Rabitz, arXiv:1104.4812v1, 2011.

28 J. Zhu, S. Kais, P. Rebentrost and A. Aspuru-Guzik, *J. Phys. Chem. B*, 2011, **115**, 1531–1537.

29 I. de Vega, *J. Phys. B: At., Mol. Opt. Phys.*, 2011, **44**, 245501.

30 J. Roden, A. Eisfeld, W. Wolff and W. T. Strunz, *Phys. Rev. Lett.*, 2009, **103**, 058301.

31 S. Valleau, A. Eisfeld and A. Aspuru-Guzik, *J. Chem. Phys.*, 2012, **137**, 224103.

32 I. P. Mercer, I. R. Gould and D. R. Klug, *J. Phys. Chem. B*, 1999, **103**, 7720–7727.

33 G. Tao and W. H. Miller, *J. Phys. Chem. Lett.*, 2010, **1**, 891–894.

34 C. Olbrich, T. L. C. Jansen, J. Liebers, M. Aghtar, J. Strümpfer, K. Schulten, J. Knoester and U. Kleinekathöfer, *J. Phys. Chem. B*, 2011, **115**, 8609–8621.

35 A. Damjanović, I. Kosztin, U. Kleinekathöfer and K. Schulten, *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.*, 2002, **65**, 031919.

36 C. Olbrich, J. Strümpfer, K. Schulten and U. Kleinekathöfer, *J. Phys. Chem. Lett.*, 2011, **2**, 1771–1776.

37 S. Jurinovich, C. Curutchet and B. Mennucci, *ChemPhysChem*, 2014, **15**, 3194–3204.

38 Y. Shan, J. L. Klepeis, M. P. Eastwood, R. O. Dror and D. E. Shaw, *J. Phys. Chem.*, 2005, **122**, 054101.

39 D. Rappoport and J. Hutter, Excited-State Properties and Dynamics, in *Fundamentals of Time-Dependent Density Functional Theory*, Lecture Notes in Physics 837, Springer Berlin/Heidelberg, 2012, pp. 317–336.

40 J. Behler and M. Parrinello, *Phys. Rev. Lett.*, 2007, **98**, 146401.

41 R. Ramakrishnan, M. Hartmann, E. Tapavicza and O. A. von Lilienfeld, *J. Chem. Phys.*, 2015, **143**, 084111.

42 K. Hansen, F. Biegler, R. Ramakrishnan, W. Pronobis, O. A. von Lilienfeld, K.-R. Müller and A. Tkatchenko, *J. Phys. Chem. Lett.*, 2015, **6**, 2326–2331.

43 M. Rupp, A. Tkatchenko, K.-R. Müller and O. A. von Lilienfeld, *Phys. Rev. Lett.*, 2012, **108**, 058301.

44 E. O. Pyzer-Knapp, K. Li and A. Aspuru-Guzik, *Adv. Funct. Mater.*, 2015, **25**, 6495–6502.

45 M. Olivucci, Density Functional Methods for Excited States: Equilibrium Structure and Electronic Spectra, in *Computational Photochemistry*, Elsevier, 2005, pp. 93–128.

46 J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kale and K. Schulten, *J. Comput. Chem.*, 2005, **26**, 1781–1802.

47 W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell and P. A. Kollman, *J. Am. Chem. Soc.*, 1995, **117**, 5179–5197.

48 M. Ceccarelli, P. Procacci and M. Marchi, *J. Comput. Chem.*, 2003, **24**, 129–142.

49 D. Tronrud, J. Wen, L. Gay and R. Blankenship, *Photosynth. Res.*, 2009, **100**, 79–87.

50 W. L. Jorgensen and J. D. Madura, *J. Am. Chem. Soc.*, 1983, 1407–1413.

51 C. Olbrich, J. Strümpfer, K. Schulten and U. Kleinekathöfer, *J. Phys. Chem. B*, 2011, **115**, 758–764.

52 J. P. Perdew, M. Ernzerhof and K. Burke, *J. Chem. Phys.*, 1996, **105**, 9982–9985.

53 J. C. Taylor, *Phys. Rev.*, 1954, **95**, 1313–1317.

54  Y. Shao, Z. Gan, E. Epifanovsky, A. T. B. Gilbert, M. Wormit, J. Kussmann, A. W. Lange, A. Behn, J. Deng, X. Feng, D. Ghosh, M. Goldey, P. R. Horn, L. D. Jacobson, I. Kaliman, R. Z. Khaliullin, T. Kús, A. Landau, J. Liu, E. I. Proynov, Y. M. Rhee, R. M. Richard, M. A. Rohrdanz, R. P. Steele, E. J. Sundstrom, H. L. Woodcock III, P. M. Zimmerman, D. Zuev, B. Albrecht, E. Alguire, B. Austin, G. J. O. Beran, Y. A. Bernard, E. Berquist, K. Brandhorst, K. B. Bravaya, S. T. Brown, D. Casanova, C.-M. Chang, Y. Chen, S. H. Chien, K. D. Closser, D. L. Crittenden, M. Diedenhofen, R. A. DiStasio Jr, H. Dop, A. D. Dutoi, R. G. Edgar, S. Fatehi, L. Fusti-Molnar, A. Ghysels, A. Golubeva-Zadorozhnaya, J. Gomes, M. W. D. Hanson-Heine, P. H. P. Harbach, A. W. Hauser, E. G. Hohenstein, Z. C. Holden, T.-C. Jagau, H. Ji, B. Kaduk, K. Khistyaev, J. Kim, J. Kim, R. A. King, P. Klunzinger, D. Kosenkov, T. Kowalczyk, C. M. Krauter, K. U. Lao, A. Laurent, K. V. Lawler, S. V. Levchenko, C. Y. Lin, F. Liu, E. Livshits, R. C. Lochan, A. Luenser, P. Manohar, S. F. Manzer, S.-P. Mao, N. Mardirossian, A. V. Marenich, S. A. Maurer, N. J. Mayhall, C. M. Oana, R. Olivares-Amaya, D. P. O'Neill, J. A. Parkhill, T. M. Perrine, R. Peverati, P. A. Pieniazek, A. Prociuk, D. R. Rehn, E. Rosta, N. J. Russ, N. Sergueev, S. M. Sharada, S. Sharmaa, D. W. Small, A. Sodt, T. Stein, D. Stück, Y.-C. Su, A. J. W. Thom, T. Tsuchimochi, L. Vogt, O. Vydrov, T. Wang, M. A. Watson, J. Wenzel, A. White, C. F. Williams, V. Vanovschi, S. Yeganeh, S. R. Yost, Z.-Q. You, I. Y. Zhang, X. Zhang, Y. Zhou, B. R. Brooks, G. K. L. Chan, D. M. Chipman, C. J. Cramer, W. A. Goddard III, M. S. Gordon, W. J. Hehre, A. Klamt, H. F. Schaefer III, M. W. Schmidt, C. D. Sherrill, D. G. Truhlar, A. Warshel, X. Xua, A. Aspuru-Guzik, R. Baer, A. T. Bell, N. A. Besley, J.-D. Chai, A. Dreuw, B. D. Dunietz, T. R. Furlani, S. R. Gwaltney, C.-P. Hsu, Y. Jung, J. Kong, D. S. Lambrecht, W. Liang, C. Ochsenfeld, V. A. Rassolov, L. V. Slipchenko, J. E. Subotnik, T. Van Voorhis, J. M. Herbert, A. I. Krylov, P. M. W. Gill and M. Head-Gordon, *Mol. Phys.*, 2015, **113**, 184–215.

55  S. Chandrasekaran, M. Aghtar, S. Valleau, A. Aspuru-Guzik and U. Kleinekathöfer, *J. Phys. Chem. B*, 2015, **119**, 9995–10004.

56  F. Häse, S. Valleau, E. Pyzer-Knapp and A. Aspuru-Guzik, *Machine learning for exciton dynamics: Qy trajectories*, figshare, 2015.

57  D. E. Rumelhart, G. E. Hinton and R. J. Williams, Learning Internal Representations by Error Propagation, *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, MIT Press, Cambridge, MA, USA, 1986, pp. 318–362.

58  G. Montavon, M. Rupp, V. Gobre, A. Vazquez-Mayagoitia, K. Hansen, A. Tkatchenko, K.-R. Müller and A. von Lilienfeld, *New J. Phys.*, 2013, **15**, 095003.

59  J. Heaton, *J. Mach. Learn. Res.*, 2015, **16**, 1243–1247.

60  O. A. von Lilienfeld, R. Ramakrishnan, M. Rupp and A. Knoll, *Int. J. Quantum Chem.*, 2015, **115**, 1084–1093.

61  N. H. List, C. Curutchet, S. Knecht, B. Mennucci and J. Kongsted, *J. Chem. Theory Comput.*, 2013, **9**, 4928–4938.

62  M. Riedmiller, *Comput. Stand. Interfac.*, 1994, **16**, 265–278.

63  L. Prechelt, *Neural Network*, 1998, **11**, 761–767.

64  D. Frenkel and B. Smit, Free Energies and Phase Equilibria, in *Understanding Molecular Simulation*, Academic Press, 2002, pp. 165–199.

65  X. Daura, K. Gademann, B. Jaun, D. Seebach, W. F. van Gunsteren and A. E. Mark, *Angew. Chem., Int. Ed.*, 1999, **111**, 249–253.

66  H.-P. Breuer and F. Petruccione, *The Theory of Open Quantum Systems*, Oxford University Press, New York, 2002.

67  M. Aghtar, J. Strümpfer, C. Olbrich, K. Schulten and U. Kleinekathöfer, *J. Phys. Chem. B*, 2013, **117**, 7157–7163.