Supporting Information: Partial Density of States Representation for Accurate Deep Neural Network Predictions of X-ray Spectra

Clelia Middleton[1], Basile F. E. Curchod[2], Thomas J. Penfold[1*]

[1]Chemistry, School of Natural and Environmental Sciences, Newcastle University, Great North Road, Newcastle upon Tyne, NE1 7RU, UK.
[2]Centre for Computational Chemistry, School of Chemistry, Cantock's Close, University of Bristol, Bristol, BS8 1TS, UK.

*Corresponding author(s). E-mail(s): tom.penfold@newcastle.ac.uk;

List of Figures

S1 Supplementary Results: pDOS Featurisation

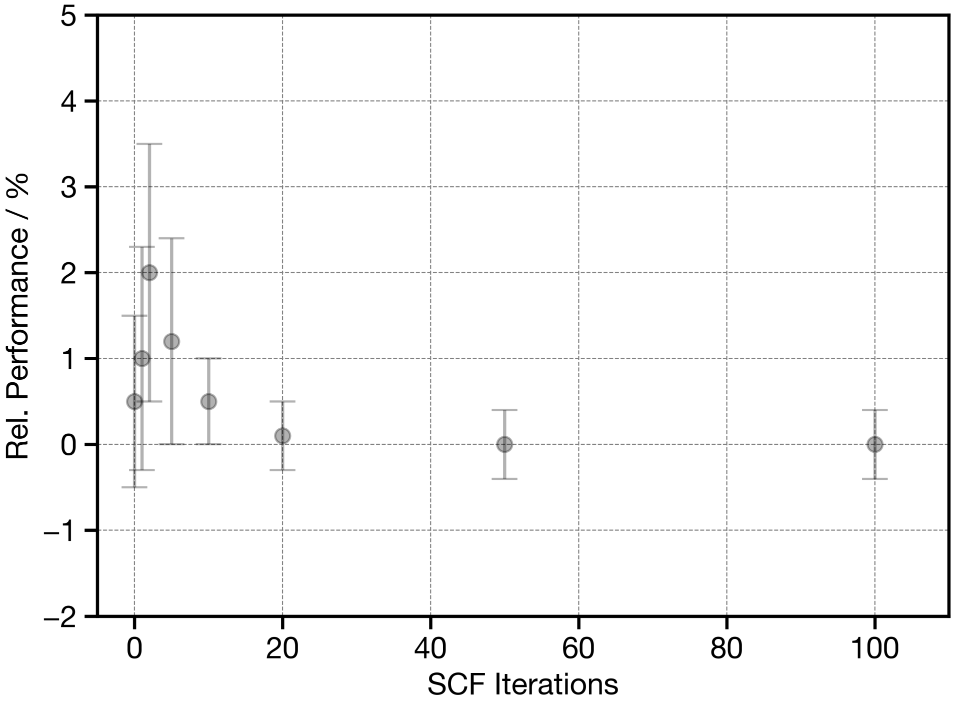

Fig. S1 Performance as a function of the number of SCF steps. Performance is plot relative (in %) to the best performance in the panel. Validation results; five-times-repeated five fold cross-validation.

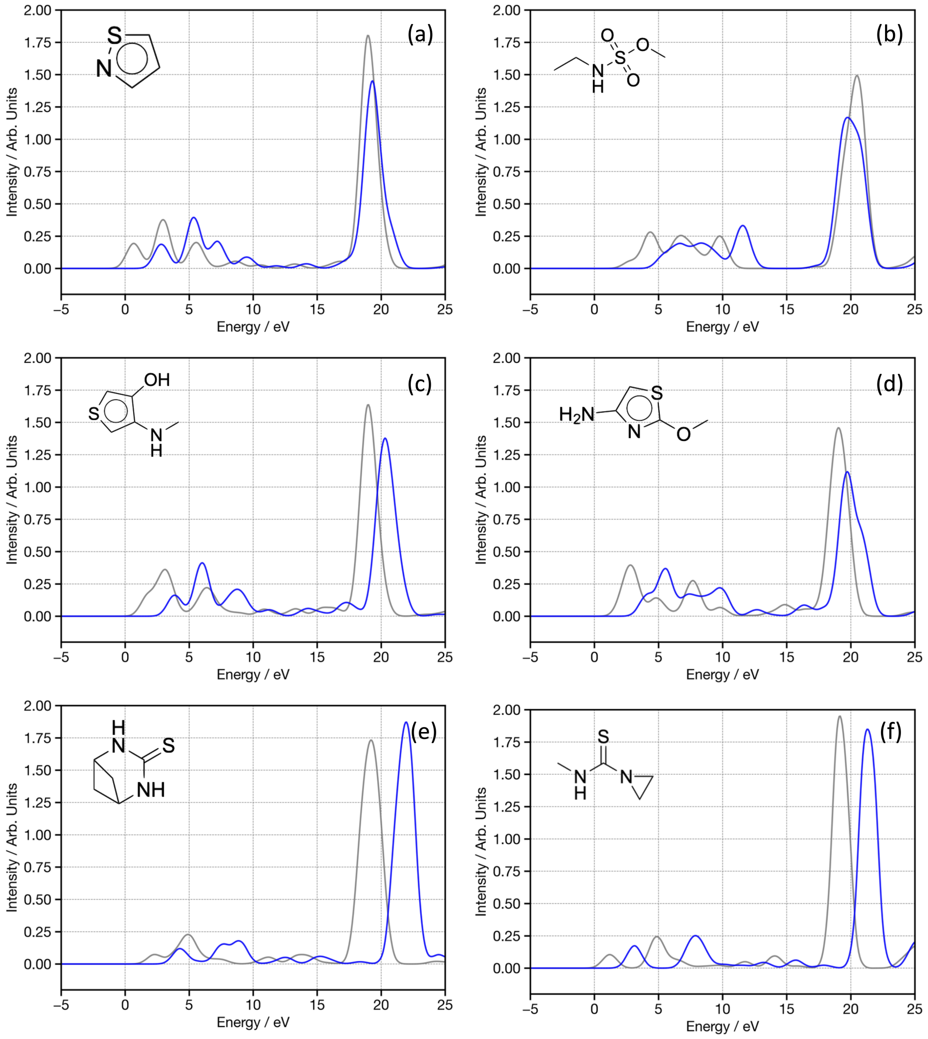

Fig. S2 Example p-DOS descriptors for 6 molecules (structures shown inset) without (grey) and with (blue) SCF optimisation.

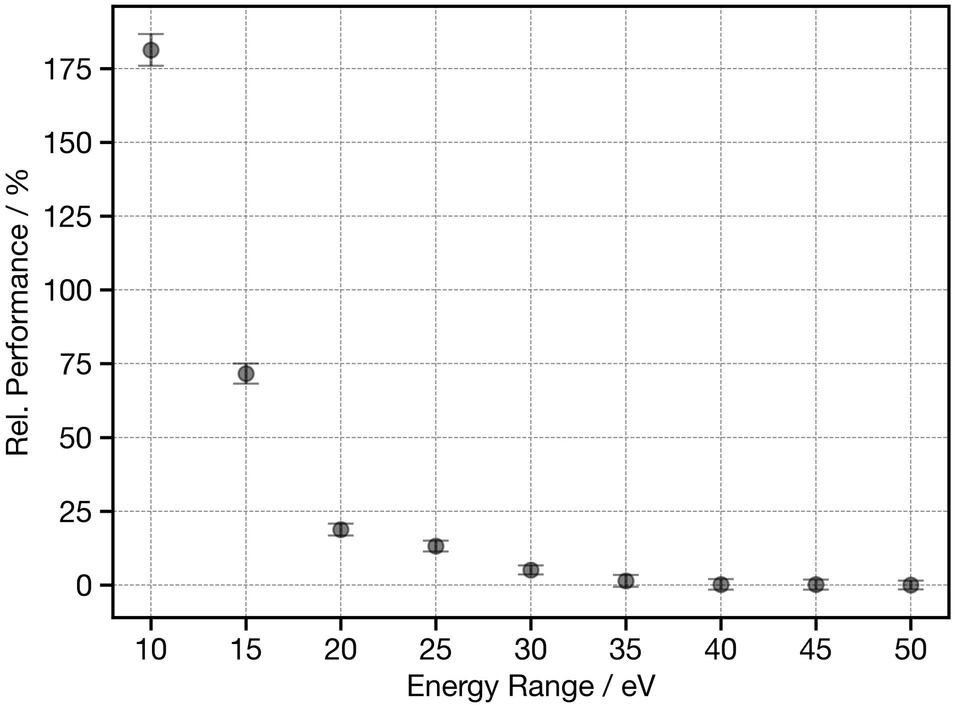

Fig. S3 Performance against energy range of pDOS descriptor. Performance is plot relative (in %) to the best performance in the panel. Validation results; five-times-repeated five fold cross-validation.

S2 Supplementary Results: Optimisation and Performance

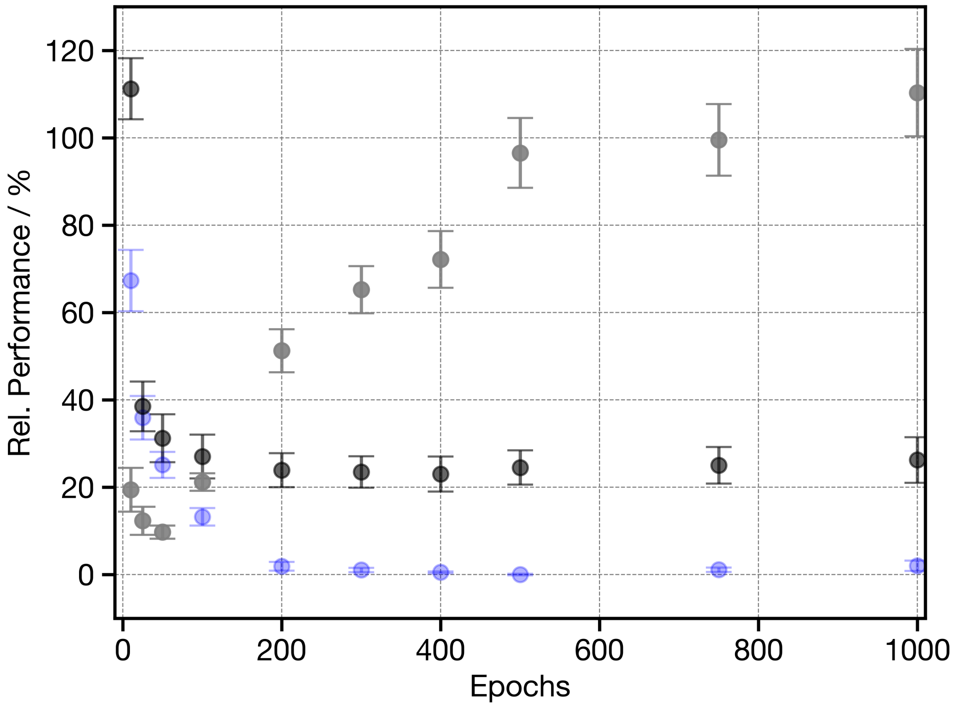

Fig. S4 Performance as a function of the number of epochs for the three descriptors considered in this work: wACSF: black, pDOS: grey, wACSF and pDOS: blue. Performance is plot relative (in %) to the best performance in the panel. Validation results; five-times-repeated five fold cross-validation.

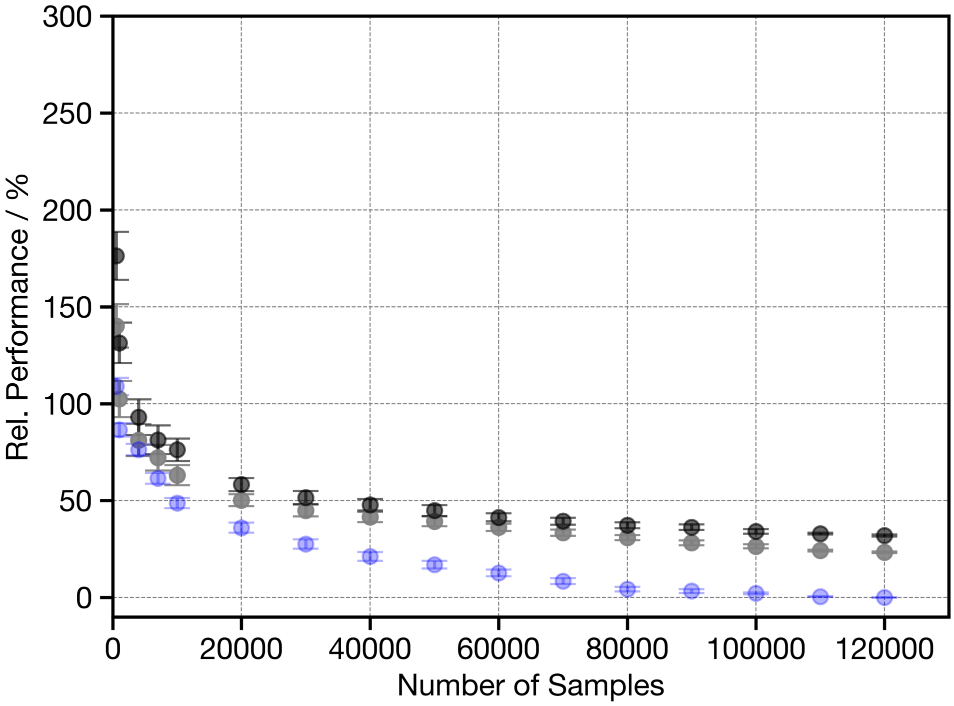

Fig. S5 Performance as a function of the number of training samples for the three descriptors considered in this work: wACSF: black, pDOS: grey, wACSF and pDOS: blue. Performance is plot relative (in %) to the best performance in the panel. Validation results; five-times-repeated five fold cross-validation

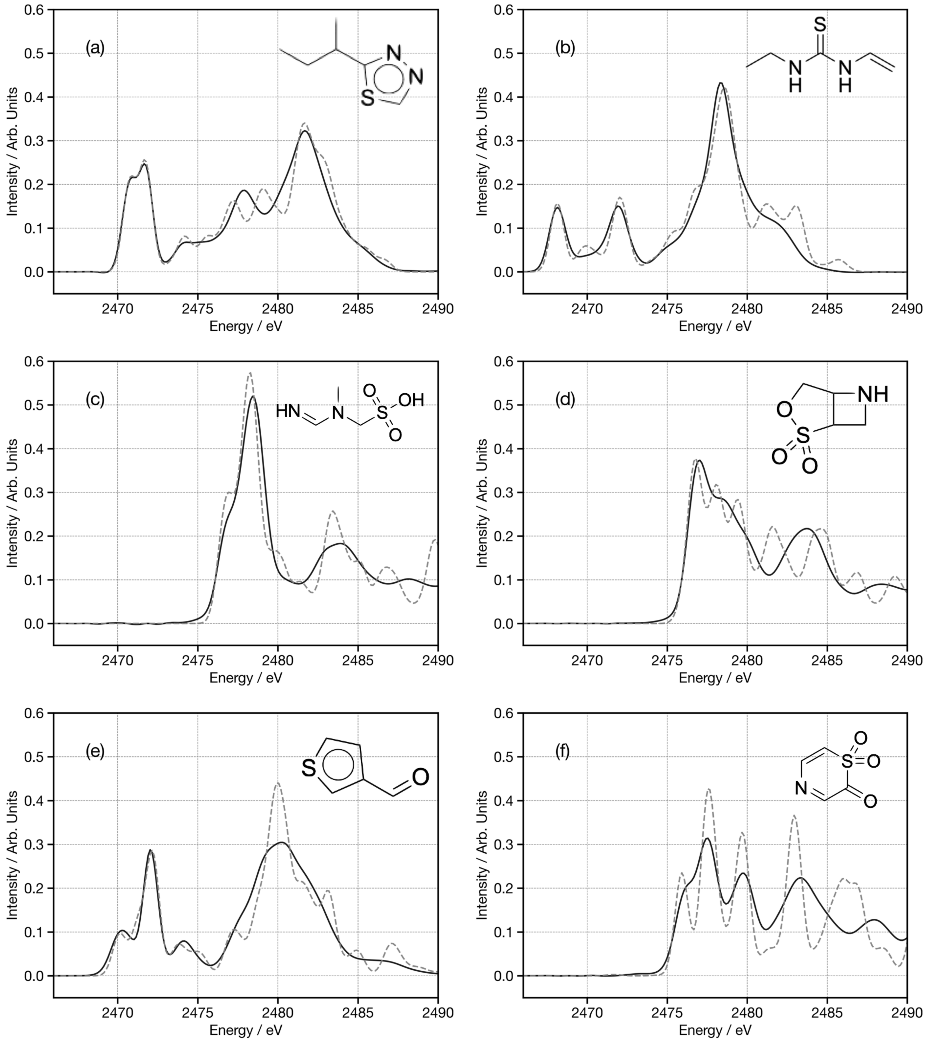

Fig. S6 Example sulphur K-edge XANES spectra. The upper two panels show K-edge XANES spectra from the 0^{th}-10^{th} percentiles, *i.e.* the best performers. The middle two panels show K-edge XANES spectra from the 45^{th}-55^{th} percentiles, *i.e.* around the median. The lower two panels show K-edge XANES spectra from the 90^{th}-100^{th} percentiles, *i.e.* the worse performers.

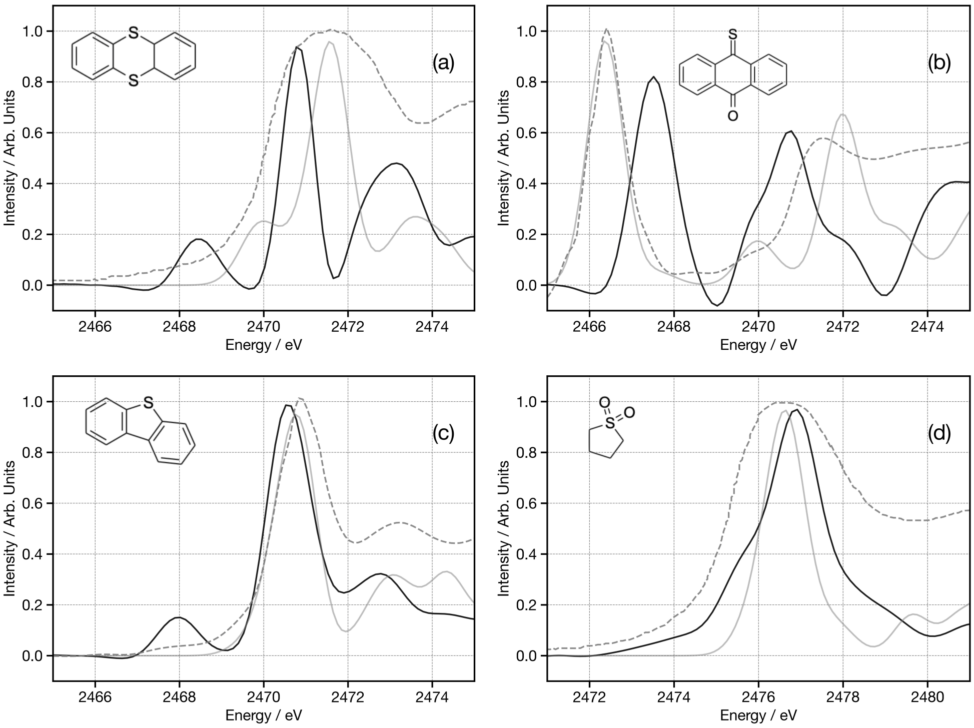

Fig. S7 Experimental (grey bashed), TDDFT(BLYP) calculated (grey solid) and DNN predictions (black) Sulphur K-edge spectra of (a) thianthrene, (b) thiohemianthraquinone, (c) dibenzothiophene and (d) tetramethylenesulfone. Experimental spectra have been digitised from Ref. [**?**]. The DNN predictions used a descriptor composed only a wACSF terms including 22 G^2 and 10 G^4 terms.

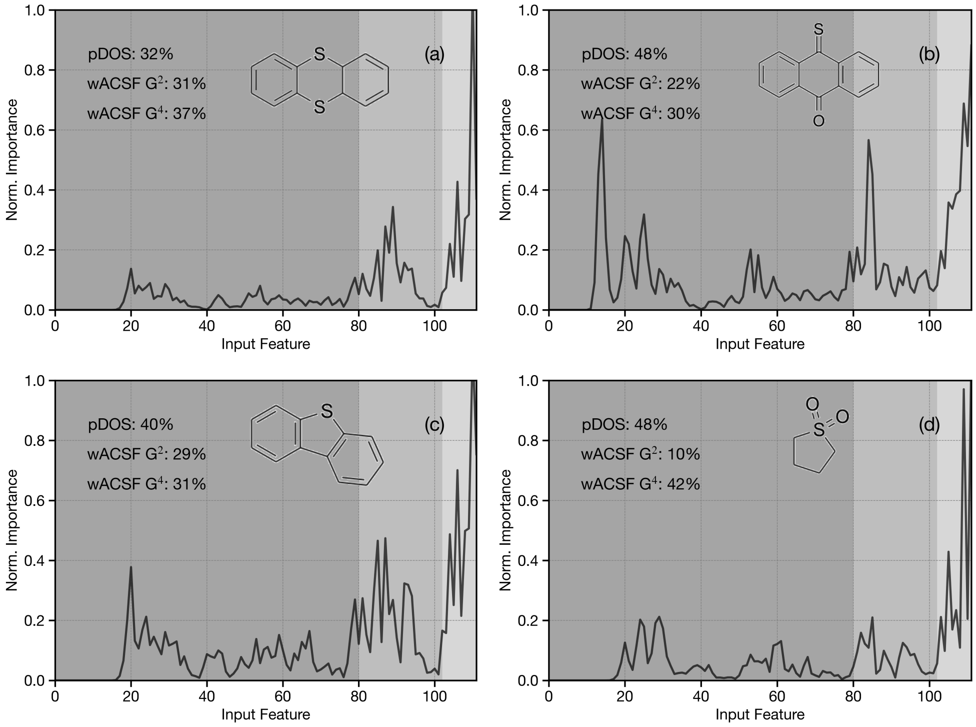

Fig. S8 The normalised feature importance arising from the Shapley analysis for the predictions of the sulphur K-edge spectra of (a) thianthrene, (b) thiohemianthraquinone, (c) dibenzothiophene and (d) tetramethylenesulfone.

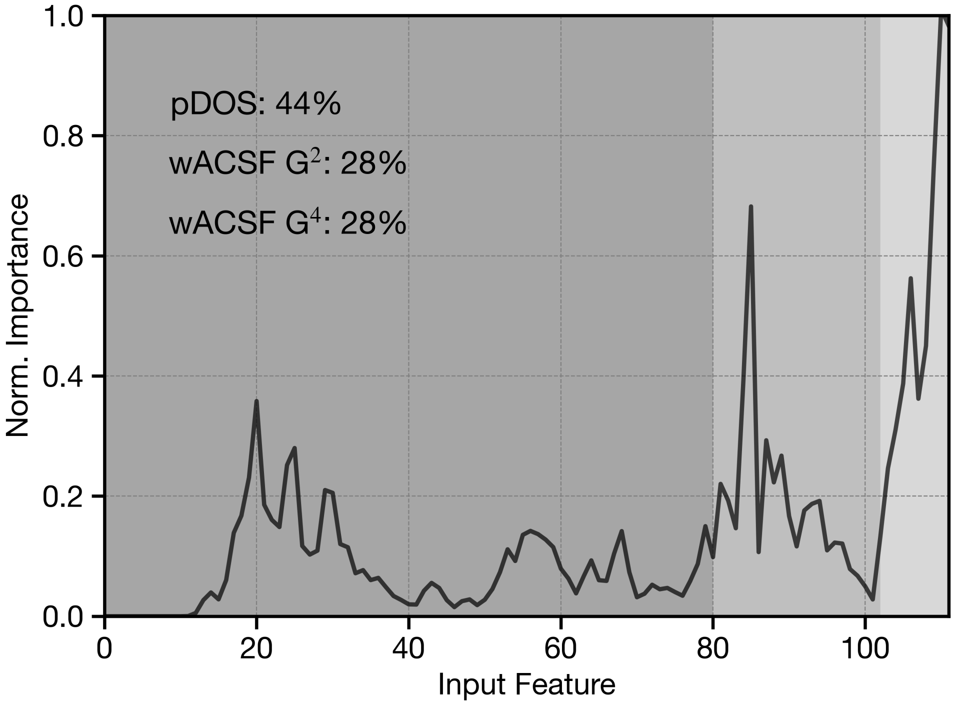

Fig. S9 The average normalised feature importance arising from the Shapley analysis from the 5000 held-out set. The features 0-80 corresponds to the p-DOS, 81-103 correspond to the G^2 wACSFS. terms and 103-113 correspond to the G^4 wACSFS. terms. Inset at the molecular structures and the percentage contribution of each term.

S3 Supplementary Results: Ground State Interconversion of Highly Vibrationally Excited Photoproducts

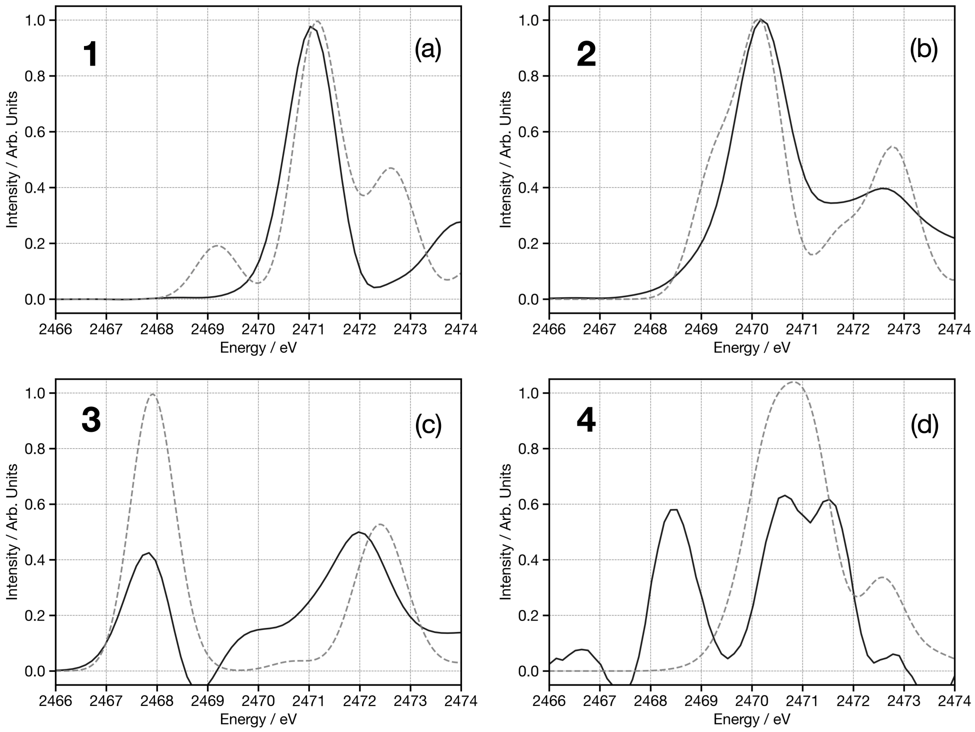

Fig. S10 TDDFT(BP86) calculated (dashed) and DNN predicted (solid) Sulphur K-edge spectra for (a) 2(5H)-thiophenone (**1**), (b) 2-(2-thiiranyl)ketene (**2**), (c) 2-thioxoethylkene (**3**) and (d) 2-(2-sulfanylethyl)kentene (**4**). All spectra have been shifted horizontally by 66 eV to account for the error in absolute transitions energies of TDDFT spectra. The DNN predictions used a descriptor composed only a wACSF terms including 22 G^2 and 10 G^4 terms.

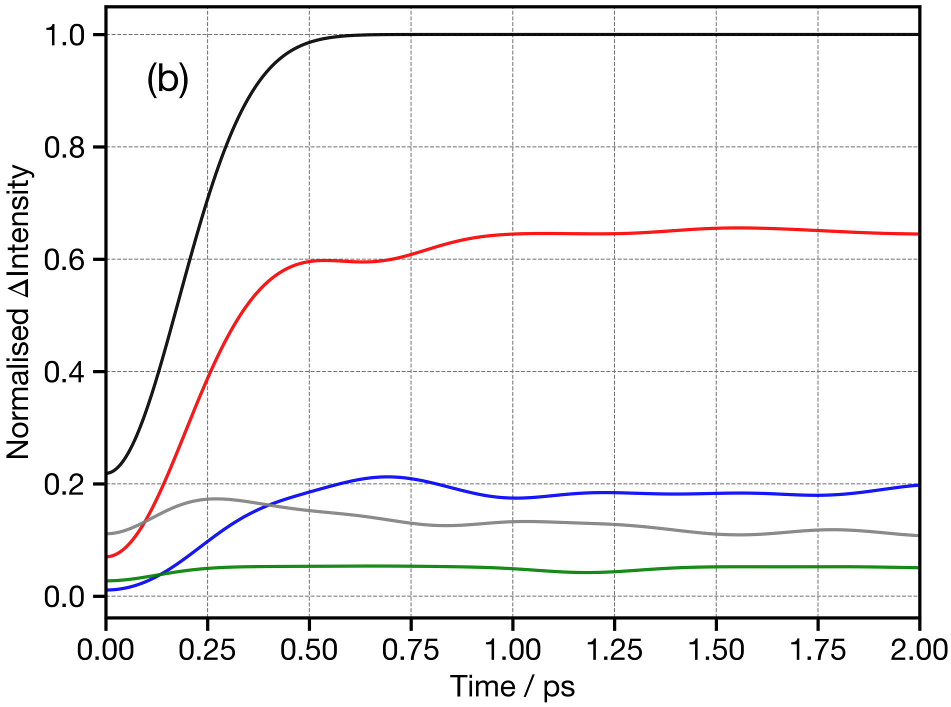

Fig. S11 Relative population of the ground state photoproducts formed after excitation of the second excited singlet state of 2(5H)-thiophenone. The overall ground state population is shown in black, while the different photoproducts include 2(5H)-thiophenone (blue), 2-(2- thiiranyl)ketene (red), 2-thioxoethylkene (green) and 2-(2-sulfanylethyl)kentene (grey). All population kinetic have been broadened with a Gaussian of full-width-at-half-maximum of 230 fs.

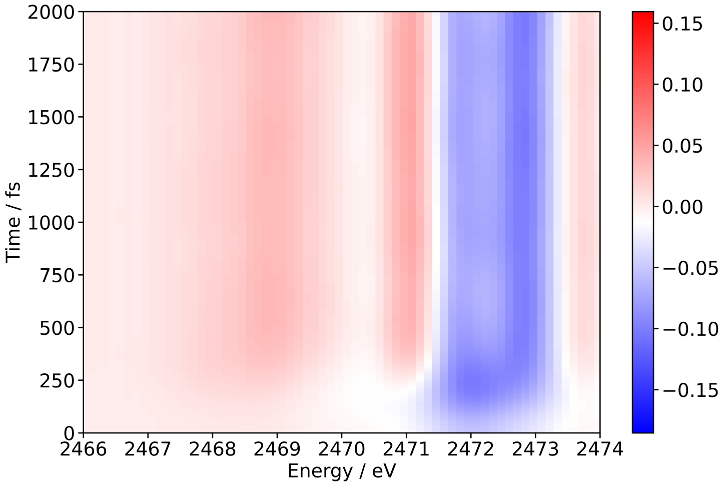

Fig. S12 DNN predicted transient ($\mu_{t\text{-}GS}$) S K-edge spectrum as a function of time calculated using 39 MD trajectories, starting from the point the photoexcited thiophenone repopulates the electronic ground states. The ground state spectrum used to generate the transient is predicted from cold ground state molecular dynamics, *i.e.* configurations prior to excitation. Each MD trajectory contains 20,000 steps meaning this spectrum has been generated using 800,000 spectral predictions from the DNN. The DNN predictions used a descriptor composed of only wACSF terms, comprising 22 G^2 and 10 G^4 terms.

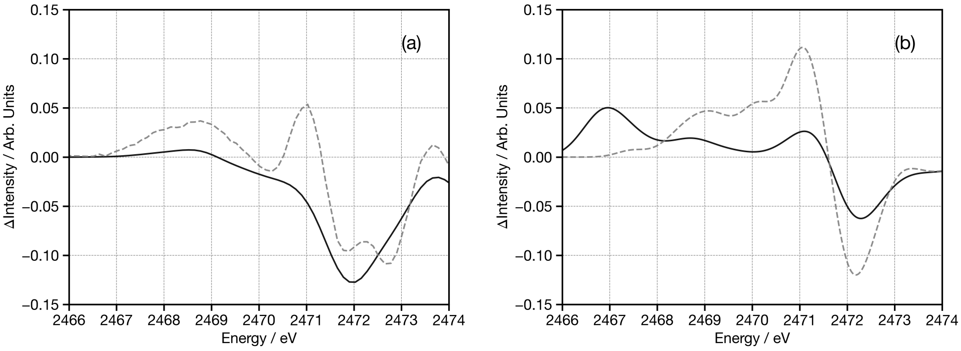

Fig. S13 The DNN predicted (a) and TDDFT(BP86) calculated (b) S K-edge spectra at 140 (solid) and 1000 (dashed) fs. The DNN predictions used a descriptor composed only a wACSF terms including 22 G^2 and 10 G^4 terms.