Electronic Supplementary Material (ESI) for Digital Discovery.

Supporting Information: Towards the Automated Extraction of Structural Information from X-ray Absorption Spectra

Tudur David [1], Nik Khadijah Nik Aznan [2], Kathryn Garside[2], and Thomas J. Penfold [1]

List of Figures

[1]Chemistry - School of Natural and Environmental Sciences, Newcastle University, Newcastle upon Tyne, NE1 7RU, UK, tom.penfold@ncl.ac.uk

[2]Research Software Engineer Group, Catalyst Building, Newcastle University, Newcastle upon Tyne, NE1 7RU, UK

List of Tables

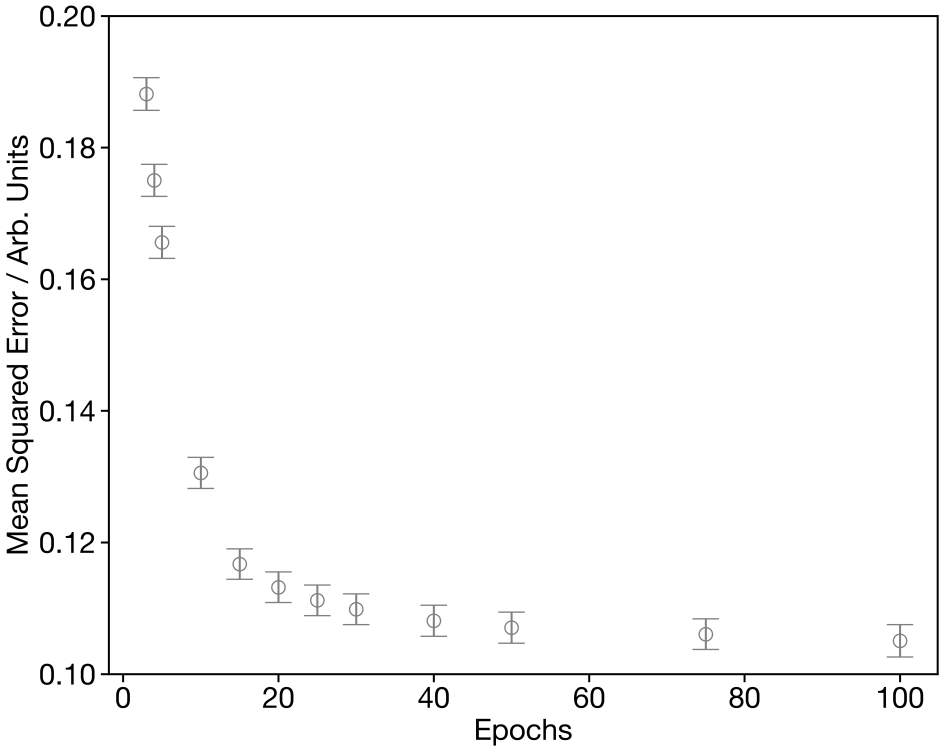

Figure S1: Evolution of the MSE as a function of the number of epochs. Data points are obtained from five-times-repeated five-fold cross-validation simulations on the full training set.

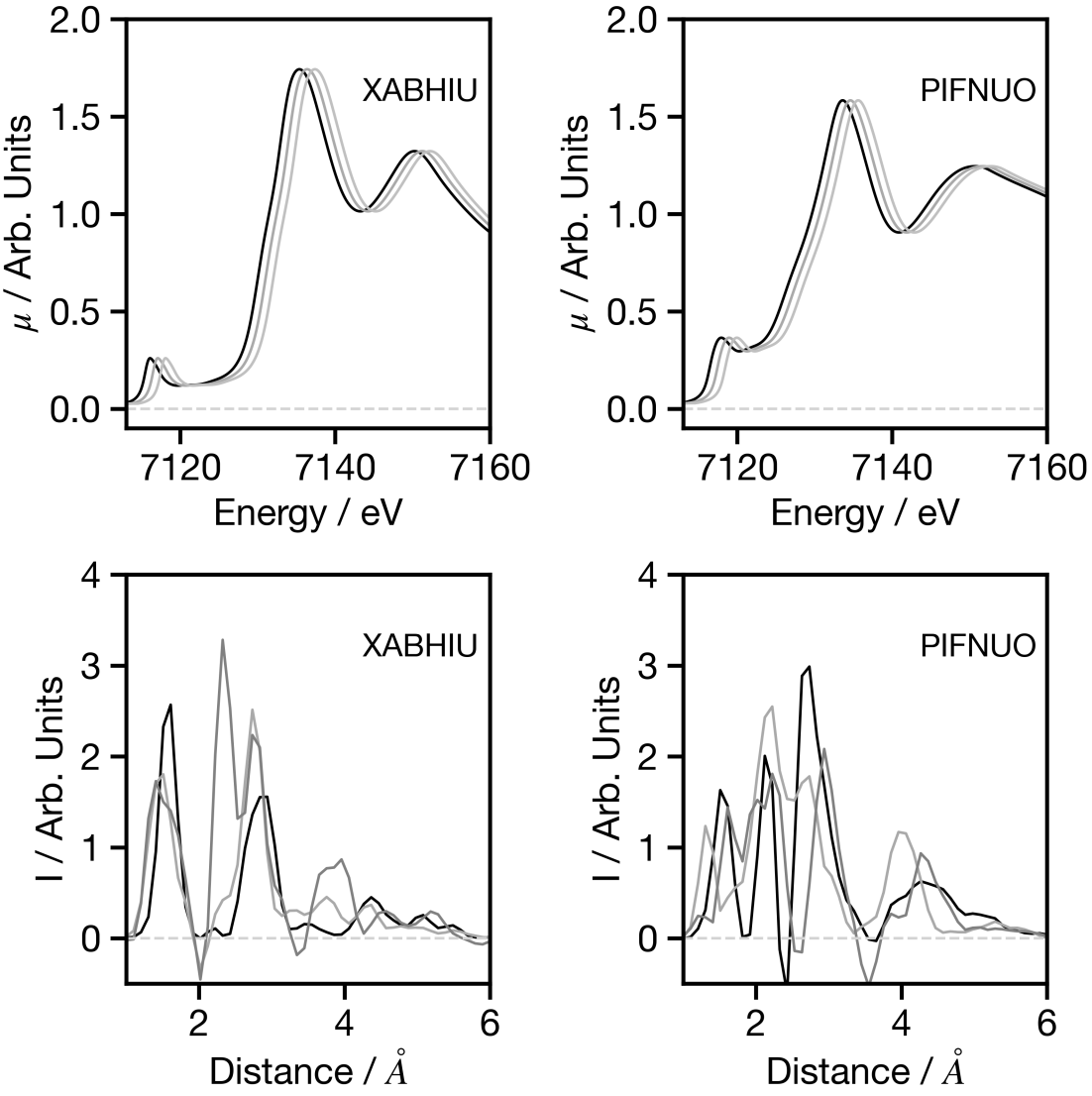

Figure S2: Spectra (*upper*) and G^2 wACSF (*lower*) as a function of spectral shift. Black: Original spectrum, Dark Grey: 1.0 eV shift, Grey: 2.0 eV shift. The six-letter codes are the Cambridge Structural Database identifiers upon which the original spectra are based.

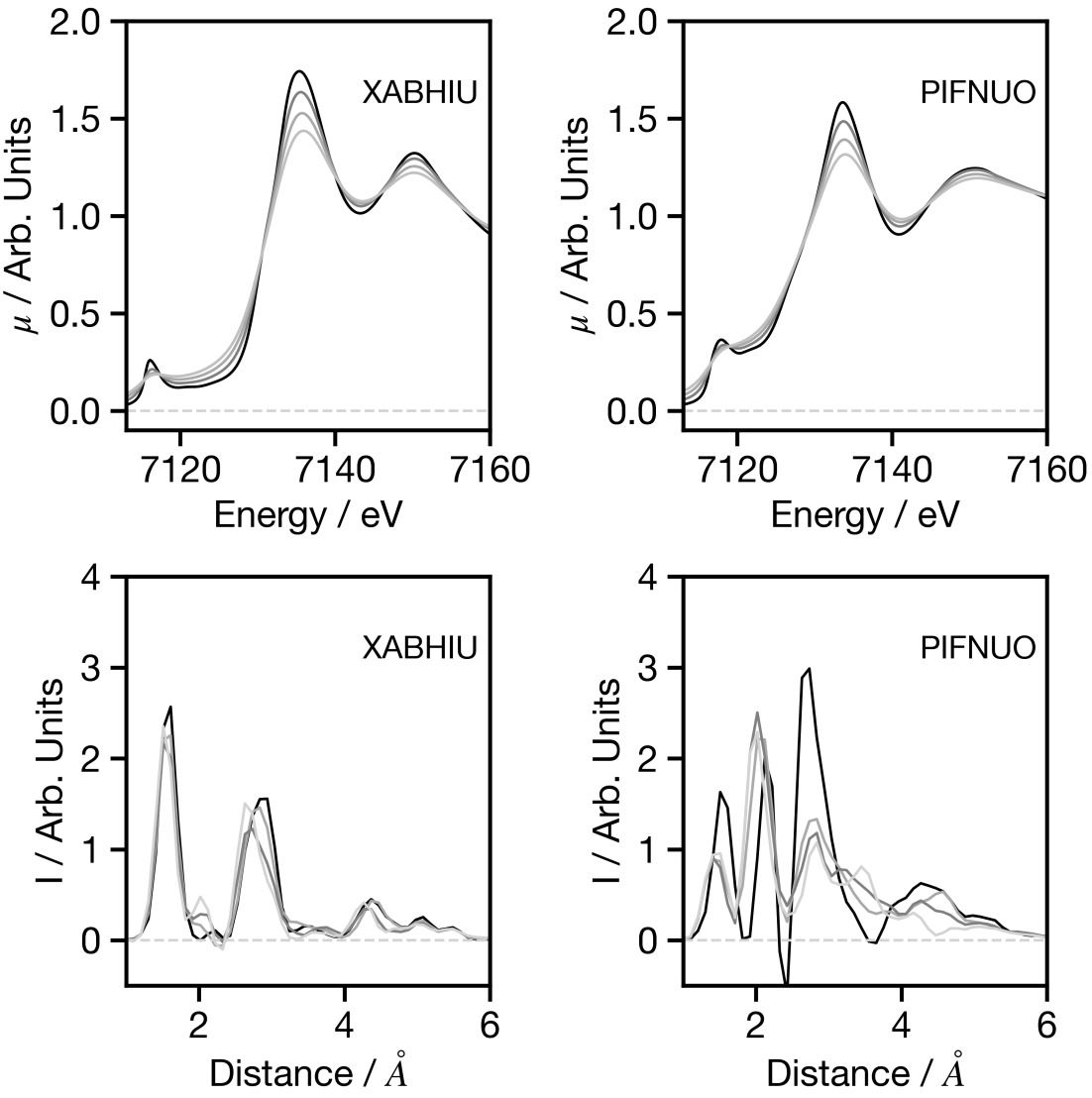

Figure S3: Spectra (*upper*) and G^2 wACSF (*lower*) as a function of spectral broadening. Black: Original spectrum, Dark Grey: 0.5 eV additional Gaussian Broadening, Grey: 1.0 eV additional Gaussian Broadening and Light Grey: 3.0 eV additional Gaussian Broadening. The six-letter codes are the Cambridge Structural Database identifiers upon which the original spectra are based.

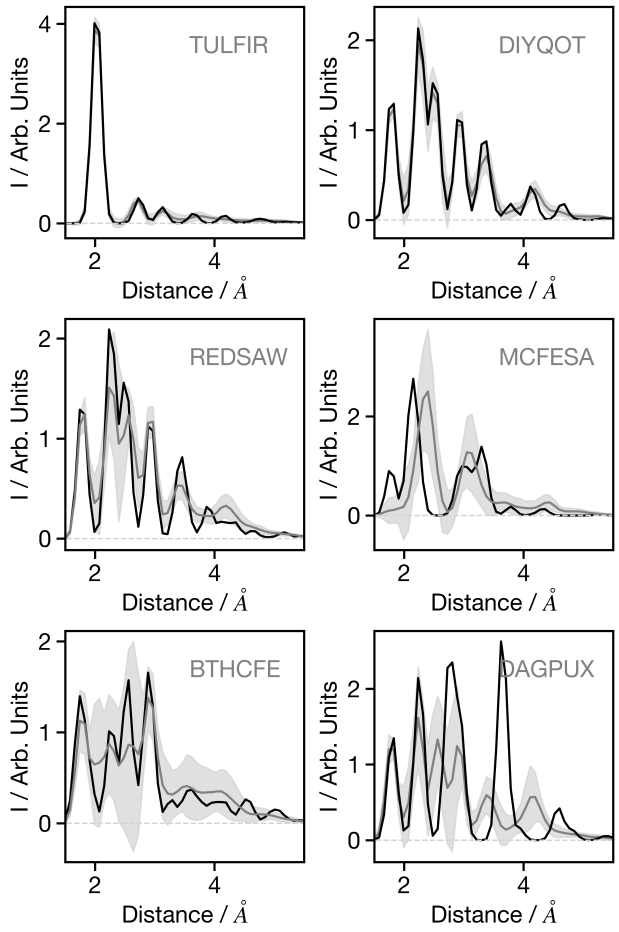

Figure S4: Example G^2 wACSF predicted from the *held-out* data set using the optimised network. The grey lines are the predicted structures with light grey regions showing $\pm 2\sigma$ calculated from the bootstrap resampling. The black traces show the true target G^2 wACSF. The upper two panels show predictions from the 0^{th}-10^{th} percentiles, *i.e.* the best performers when held-out set is ranked by MSE. The centre two panels show predictions from the 45^{th}-55^{th} percentiles, *i.e.* around the median. The bottom two panels show K-edge XANES spectra from the 90^{th}-100^{th} percentiles,*i.e.* the lowest performance. The six-character labels in the lower right of each panel are the Cambridge Structural Database (CSD) codes for the samples.

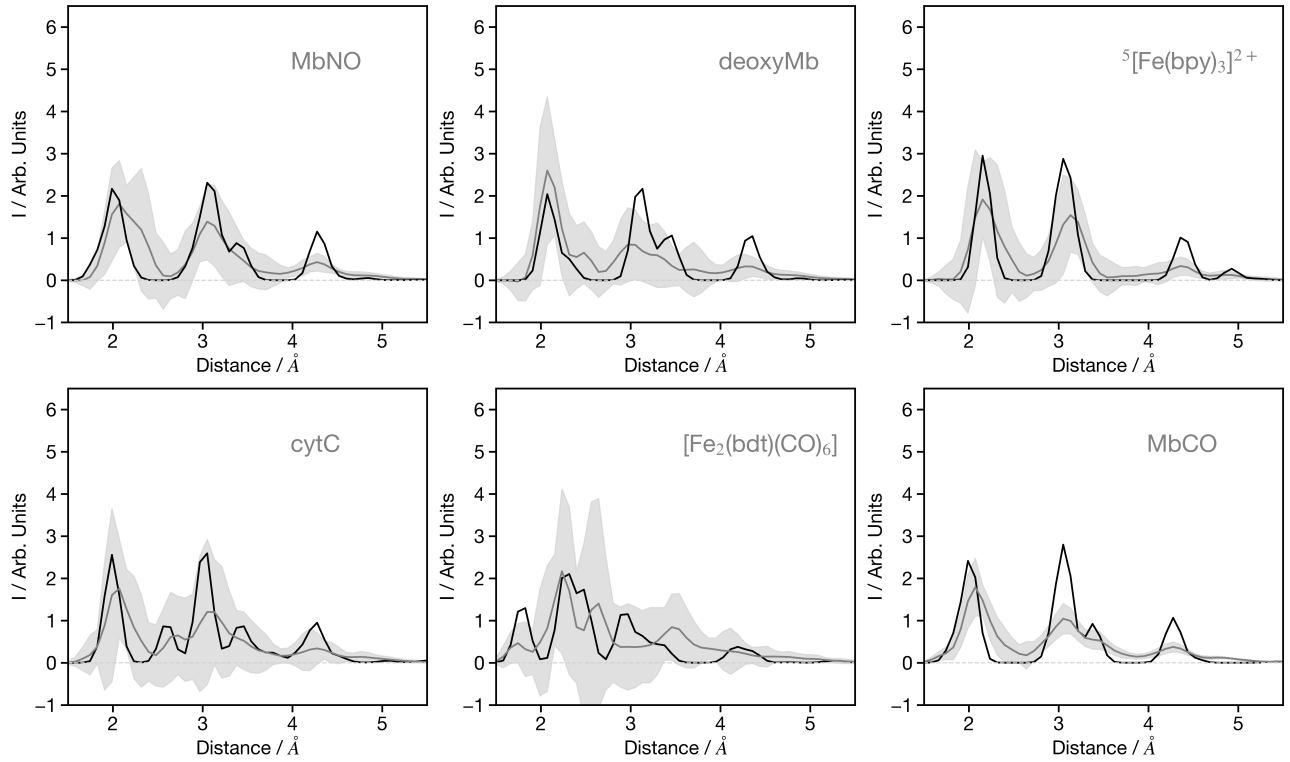

Figure S5: G^2 wACSF predicted from experimental spectra. The source of the experimental spectra is given in Table S1. The grey lines are the predicted structures with light grey regions showing $\pm 2\sigma$ calculated from the bootstrap resampling. The black traces show the expected G^2 wACSF from experimentally reported structures, as discussed in the main text

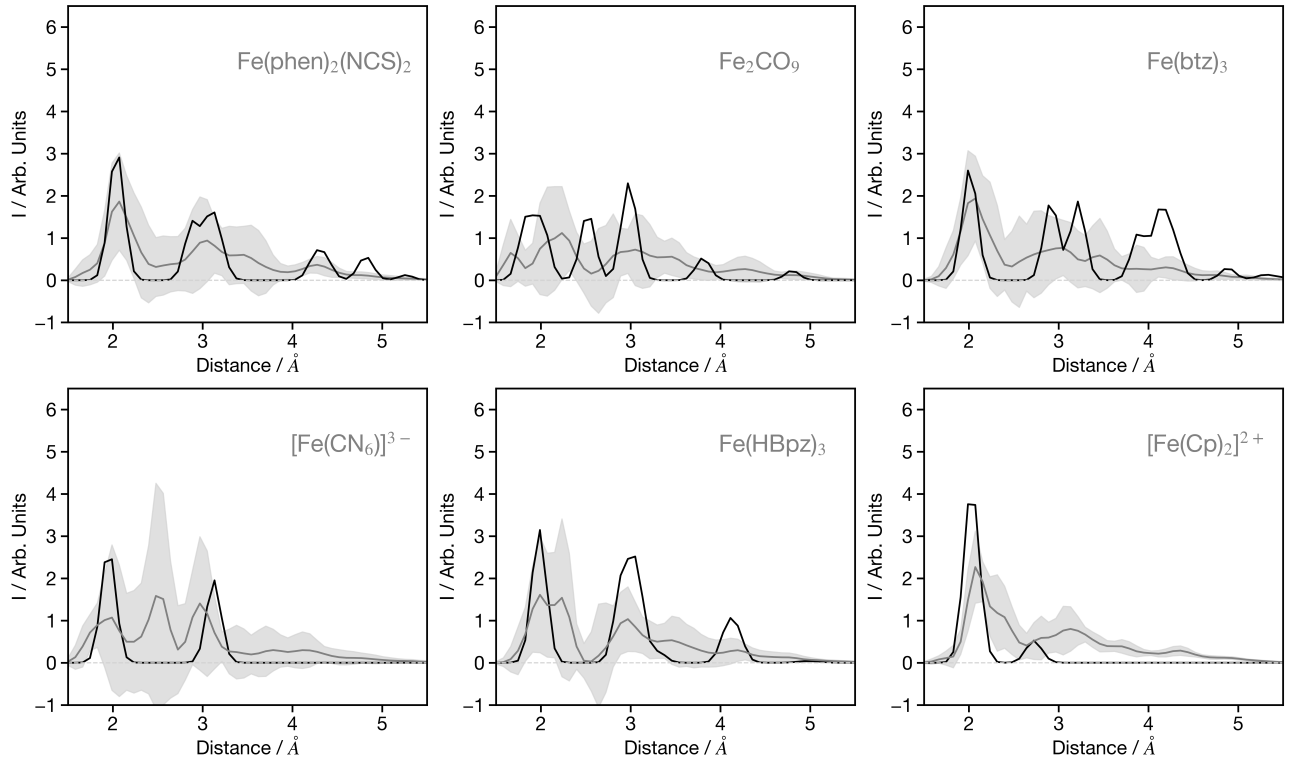

Figure S6: G^2 wACSF predicted from experimental spectra. The source of the experimental spectra is given in Table S1. The grey lines are the predicted structures with light grey regions showing $\pm 2\sigma$ calculated from the bootstrap resampling. The black traces show the expected G^2 wACSF from experimentally reported structures, as discussed in the main text.

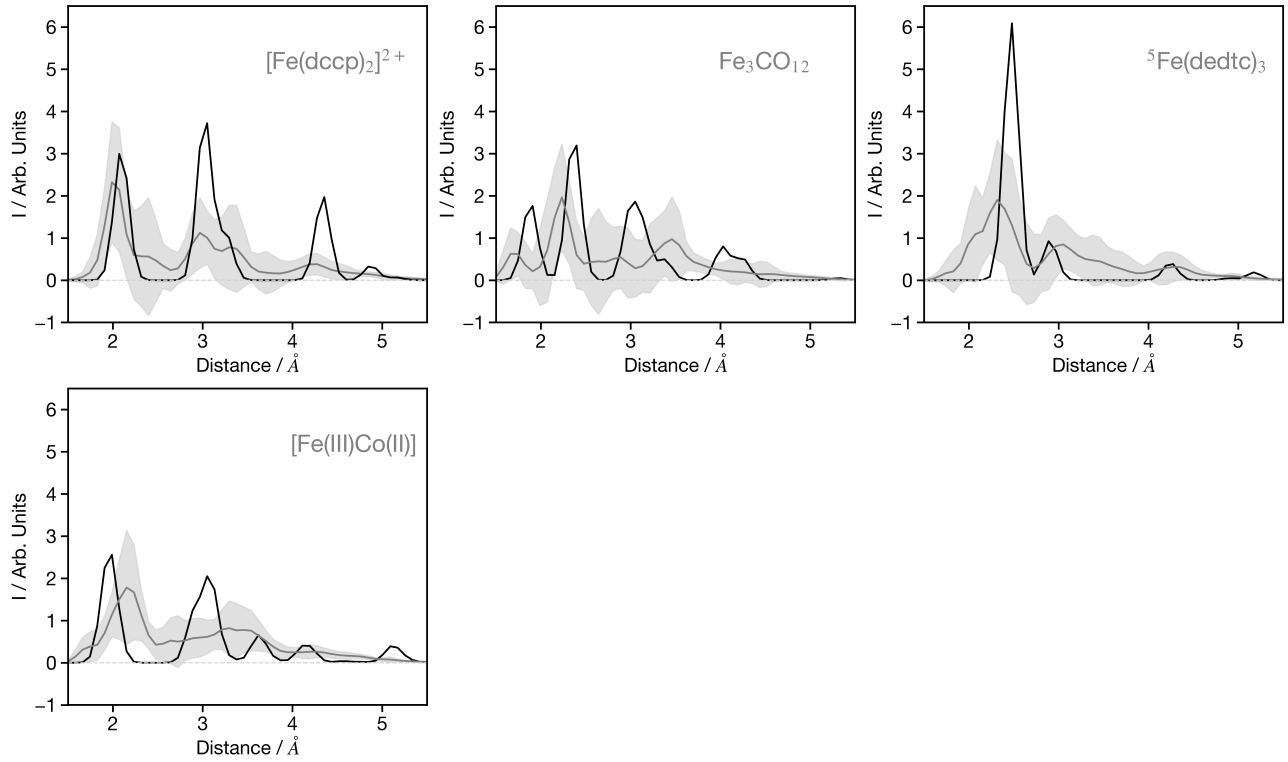

Figure S7: G^2 wACSF predicted from experimental spectra. The source of the experimental spectra is given in Table S1. The grey lines are the predicted structures with light grey regions showing $\pm 2\sigma$ calculated from the bootstrap resampling. The black traces show the expected G^2 wACSF from experimentally reported structures, as discussed in the main text

Compound	Notes	Ref.
MbNO	nitrosylmyoglobin	[2]
MbCO	carboxymyoglobin	[2]
MbO_2	oxymyoglobin	[2]
deoxyMb	unligated myoglobin	[2]
$^1[Fe(bpy)_3]^{2+}$	bpy = 2,2'-Bipyridine	[3]
$^5[Fe(bpy)_3]^{2+}$	bpy = 2,2'-Bipyridine	[3]
cytC	cytochrome C	[4]
$Fe(CO)_5$		[5]
$Fe_2(CO)_9$		[6]
$Fe_3(CO)_{12}$		[6]
$Fe_2(dbt)(CO)_6$	bdt = benzenedithiolene	[7]
$Fe(phen)_2(NCS)_2$	phen = 1,10-phenanthroline	[8]
$Fe(btz)_3$	btz = 3,3'-dimethyl-1,1'-bis(p-tolyl)-4,4'-bis(1,2,3-triazol-5-ylidene))	[9]
$[Fe(CN)_6]^{4-}$		[10]
$[Fe(CN)_6]^{3-}$		[10]
$Fe(acac)_3$		[11]
$Fe(HB(pz)_3$	pz = pyrazolylborate	[12]
$[Fe(Cp)_2]^{2+}$		[6]
$[Fe(dccp)_2]^{2+}$	dccp = 2,6-(dicarboxypyridyl)pyridine	[13]
$^1[Fe(dedtc)_3]$	dedtc = N,N'-diethyldithiocarbamate	[8]
$^5[Fe(dedtc)_3]$	dedtc = N,N'-diethyldithiocarbamate	[8]
[FeIICoIII]	[(Tp*)Fe(CN)3]2[Co(bpyMe)2]22+	[14]

Table S1: Name and reference associated with the experimental spectra used in this work. The data is available at [1]

References

[1] XANESNET Training Data, 2023, gitlab.com/team-xnet/training-sets.

[2] Frederico A Lima, Thomas J Penfold, Renske M Van Der Veen, Marco Reinhard, Rafael Abela, Ivano Tavernelli, Ursula Rothlisberger, Maurizio Benfatto, Christopher J Milne, and Majed Chergui. Probing the electronic and geometric structure of ferric and ferrous myoglobins in physiological solutions by fe k-edge absorption spectroscopy. *Physical Chemistry Chemical Physics*, 16(4):1617–1631, 2014.

[3] Ch Bressler, C Milne, V-T Pham, Amal ElNahhas, Renske M van der Veen, Wojciech Gawelda, S Johnson, Paul Beaud, Daniel Grolimund, Maik Kaiser, et al. Femtosecond xanes study of the light-induced spin crossover dynamics in an iron (ii) complex. *Science*, 323(5913):489–492, 2009.

[4] Camila Bacellar, Dominik Kinschel, Giulia F Mancini, Rebecca A Ingle, Jérémy Rouxel, Oliviero Cannelli, Claudio Cirelli, Gregor Knopp, Jakub Szlachetko, Frederico A Lima, et al. Spin cascade and doming in ferric hemes: Femtosecond x-ray absorption and x-ray emission studies. *Proceedings of the National Academy of Sciences*, 117(36):21914–21920, 2020.

[5] Wei-Ting Chen, Che-Wei Hsu, Jyh-Fu Lee, Chih-Wen Pao, and I-Jui Hsu. Theoretical analysis of fe k-edge xanes on iron pentacarbonyl. *ACS omega*, 5(10):4991–5000, 2020.

[6] Andrew J Atkins, Matthias Bauer, and Christoph R Jacob. High-resolution x-ray absorption spectroscopy of iron carbonyl complexes. *Physical Chemistry Chemical Physics*, 17(21):13937–13948, 2015.

[7] JPH Oudsen, B Venderbosch, DJ Martin, TJ Korstanje, JNH Reek, and M Tromp. Spectroscopic and theoretical investigation of the [fe 2 (bdt)(co) 6] hydrogenase mimic and some catalyst intermediates. *Physical Chemistry Chemical Physics*, 21(27):14638–14645, 2019.

[8] Stefan Mebs, Beatrice Braun, Ramona Kositzki, Christian Limberg, and Michael Haumann. Abrupt versus gradual spin-crossover in feii (phen) 2 (ncs) 2 and feiii (dedtc) 3 compared by x-ray absorption and emission spectroscopy and quantum-chemical calculations. *Inorganic Chemistry*, 54(24):11606–11624, 2015.

[9] Meiyuan Guo, Om Prakash, Hao Fan, Lisa HM de Groot, Valtỳr Freyr Hlynsson, Simon Kaufhold, Olga Gordivska, Nicolás Velásquez, Pavel Chabera, Pieter Glatzel, et al. Herfd-xanes probes of electronic structures of iron ii/iii carbene complexes. *Physical Chemistry Chemical Physics*, 22(16):9067–9073, 2020.

[10] Thomas James Penfold, Marco Reinhard, Mercedes Hannelore Rittmann-Frank, Ivano Tavernelli, Ursula Rothlisberger, Christopher J Milne, Pieter Glatzel, and Majed Chergui. X-ray spectroscopic study of solvent effects on the ferrous and ferric hexacyanide anions. *The Journal of Physical Chemistry A*, 118(40):9411–9418, 2014.

[11] Aniruddha Deb and Elton J Cairns. In situ x-ray absorption spectroscopyâa probe of cathode materials for li-ion cells. *Fluid phase equilibria*, 241(1-2):4–19, 2006.

[12] Valérie Briois, Ph Sainctavit, Gary J Long, and Fernande Grandjean. Importance of photoelectron multiple scattering in the iron k-edge x-ray absorption spectra of spin-crossover complexes: full multiple scattering calculations for several iron (ii) trispyrazolylborate and trispyrazolylmethane complexes. *Inorganic Chemistry*, 40(5):912–918, 2001.

[13] Alexander Britz, Wojciech Gawelda, Tadesse A Assefa, Lindsey L Jamula, Jonathan T Yarranton, Andreas Galler, Dmitry Khakhulin, Michael Diez, Manuel Harder, Gilles Doumy, et al. Using ultrafast x-ray spectroscopy to address questions in ligand-field theory: The excited state spin and structure of [fe (dcpp) 2] 2+. *Inorganic chemistry*, 58(14):9341–9350, 2019.

[14] Corine Mathonière, Dmitri Mitcov, Evangelia Koumousi, Daniel Amorin-Rosario, Pierre Dechambenoit, Sadaf Fatima Jafri, Philippe Sainctavit, Christophe Cartier dit Moulin, Loic Toupet, Elzbieta Trzop, et al. Metal-to-metal electron transfer in a cyanido-bridged $\{$Fe 2 Co 2$\}$ square complex followed by x-ray diffraction and absorption techniques. *Chemical Communications*, 58(86):12098–12101, 2022.