Supporting Information

Convolutional Neural Network Analysis of Two-Dimensional Hyperfine Sublevel Correlation Electron Paramagnetic Resonance Spectra

Alexander T. Taguchi,† Ethan D. Evans,† Sergei A. Dikanov,‡ and Robert G. Griffin*,†*

†Department of Chemistry, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139, United States

‡Department of Veterinary Clinical Medicine, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, United States

Corresponding Authors
*A.T.T.: email, taguchi@mit.edu; R.G.G.: email, rgg@mit.edu

Simulation of HYSCORE Training Dataset. Training sets for machine learning were created with the EasySpin v5.2.13[1] saffron module in Matlab R2018a. 100,000 powder-type HYSCORE simulations were generated with the pulse sequence $\pi/2$-τ-$\pi/2$-t_1-π-t_2-$\pi/2$-τ-echo[2] assuming a two-spin system of a $S = ½$ electron spin coupled to a ^{14}N nuclear spin ($I = 1$). Ideal pulse widths were employed, and the echo intensity was recorded as a function of t_1 and t_2 in steps of 20 ns along a 500×500 grid of points.

The EPR spectrum was modeled by an anisotropic g-tensor with g_{YY} set to the center of the excitation pulse (determined by the microwave frequency fixed at 9 GHz and a randomized magnetic field in the interval 300-400 mT), and the g_{XX} and g_{ZZ} components randomized but constrained to have a maximum span ($g_{ZZ} - g_{XX}$) of 0.1. The EPR linewidths at the individual g_{XX}, g_{YY}, and g_{ZZ} orientations were randomized between 0-100 MHz using EasySpin's HStrain parameter. Strong hyperfine interactions, electron-electron dipolar interactions, or other magnetic interactions that would result in resolvable splittings in the EPR spectrum were not accounted for in the simulations as this is currently not a supported feature in EasySpin's saffron module. The excitation bandwidth of the pulse was randomly varied within the interval 10-100 MHz.

The coupling constants for the simulations were randomly sampled from the following ranges: isotropic hyperfine interaction (a) = ±0-8 MHz, anisotropic hyperfine interaction (T) = 0-1 MHz, hyperfine rhombicity (δ) = 0-1, nuclear quadrupole coupling constant (K) = 0-5 MHz, and nuclear quadrupole asymmetry parameter (η) = 0-1. The limits for a, T, and K were chosen based on the typical range of coupling constants observed by ^{14}N HYSCORE investigations in the literature. δ and η are naturally limited to the range 0-1. The relative orientations of the hyperfine and nuclear quadrupole tensors were randomized with EasySpin's sphrand function. Isotropic hyperfine coupling strain, or "a-strain", was also accounted for in the simulations. a-strain

describes the case where an EPR spectrum is not characterized by one single value for a, but rather a distribution of values for a. This occurs when the EPR sample is not perfectly monomorphic, but instead contains a heterogeneous mixture of molecular structures that gives rise to a distribution of isotropic couplings. a-strain was modeled with a weighted sum of HYSCORE spectra simulated with different values of a, in order to approximate a Gaussian distribution of isotropic couplings.[3] The shape of this Gaussian distribution was randomly varied between a standard deviation of 0 MHz (perfectly monomorphic case) and 0.3 MHz (large variance of a within the sample). Strain effects for T, δ, K, and η in HYSCORE spectra have not been reported in the literature to the best of our knowledge, and were therefore not accounted for in the simulations.

The experimental settings for the HYSCORE simulations were chosen in the following manner. The time delays between pulses were sampled from the following ranges: τ = 50-300 ns and t_1 and t_2 initial values = 0-100 ns. The microwave frequency was set to 9 GHz (X-band) and the magnetic field was randomized within the range 300-400 mT. The flip angle of the π-pulse was randomly set to a value between that of a $\pi/2$-pulse (90°) and a perfect π-pulse (180°) to mimic the appearance of diagonal peak artifacts in the experimental spectra. Powder averaging was performed with a 2° spacing between sampled molecular orientations.

Preprocessing of Time-domain Patterns. Simulated and experimental HYSCORE time-domain patterns were preprocessed before being passed through the neural network. A 3rd order polynomial baseline correction was applied, followed by fast Fourier transformation along both dimensions without any zero-filling.

For experimental spectra exhibiting multiple interacting ^{14}N nuclei, the signals were separated manually to obtain HYSCORE spectra corresponding to only a single ^{14}N (Figure S1).

Regions that were removed from the spectrum during this treatment were replaced with noise. A cubic interpolation, using Matlab's interp2 function, was used to recalculate the spectra to a frequency step size of 0.1 MHz as used in the simulations.

Simulated and experimental spectra were symmetrized by averaging the regions on one side of the diagonal/antidiagonal with their nearly symmetric counterparts on the other side of the diagonal/antidiagonal. After symmetrization, the redundant areas of the spectrum were removed by reorganizing the triangular halves into a compact square matrix representation. The spectrum was then truncated along both dimensions to the frequency range 0-12.8 MHz, resulting in a square matrix of 128×128 points due to the 0.1 MHz step size. Finally, all spectra were normalized to a maximum intensity of 1.

Neural Network Architecture. The neural network consists of convolutional layers for peak shape and pattern recognition, followed by fully connected layers that convert the spectral patterns into the desired parameters of interest. The 128×128 HYSCORE spectrum is concatenated with the experimental settings (magnetic field in units of T and pulse delays in units of μs), resulting in a 128×128×N tensor, where the size of the third dimension depends on the number of experimental parameters (N-1). This tensor is passed through three convolutional layers with 16, 32, and 64 channels, respectively. The kernel size is 3×3 and the stride is set to 2 in all cases. The output is then flattened along the image dimensions. After passage through one fully connected layer of size 256, the network branches into five parallel fully connected layers each of size 64 for predicting the values of a, T, δ, K, and η. ReLU activations[4] are applied to all layer outputs except for the five final classification layers, where the SoftPlus[5] activation $\ln(1 + e^x)$ is used instead. The SoftPlus

activation prevents the neural network from outputting negative numbers, constraining its form to that of a probability distribution for each parameter of interest.

Neural Network Training. Neural networks were created and optimized with TensorFlow v1.6.0. Training was performed on 95,000 of the preprocessed HYSCORE simulations in batch sizes of 100, with the remaining 5,000 set aside to validate there is no overfitting to the training data. Prior to each training pass, a random amount of noise between 0-15% was added to the training batch to improve the algorithm's performance on noisy data. A Gaussian broadening filter was also applied to the spectra with a random standard deviation between 0-0.2 MHz to make the neural network robust to different degrees of spectral broadening and resolution that occur in real experimental data. An Adam optimizer[6] set at a learning rate of 10^{-3} was used to minimize the cross-entropy loss function. Training was performed for 100,000 steps with 50% drop out[7] on a 14-core PC with 64 GB of RAM.

Neural Network Design Considerations. In this work, a neural network is trained on simulations to make predictions about experimental data. As such, significant thought was required in designing the network architecture to ensure that the machine learning algorithm had the tools necessary to model the complex HYSCORE spectra, but without allowing it to overfit to features in the simulations that do not appear in the experiment. Here, the major considerations that went into the neural network implementation are discussed.

Initial efforts made use of fully connected neural networks to model the ^{14}N HYSCORE data. For spin systems of low hyperfine anisotropy ($T \approx 0$) where peak shape is essentially circular and symmetric, these models performed well for estimating a and K. However, these simple

models failed to scale well to systems with $T > 0$ where peaks exhibit complex, anisotropic shapes. Convolutional networks were found to be better at handling hyperfine anisotropy, and demonstrated that an estimate of T could be made directly from the ^{14}N spectra. The number of convolutional layers and their channels were optimized while holding all other aspects of the network constant. A total of 3 convolutional layers was found to be optimal. Less than this number resulted in underfitting, and more than this number led to overfitting to features in the simulated HYSCORE spectra not present in the experimental spectra. In principle, if the quality of the simulations were improved (matching the characteristics of real spectra better), more convolutional layers could be added to improve the neural network's performance further. A stride of 2 was favored over max pooling as striding is computationally faster and showed no difference in final prediction accuracy relative to max pooling.

The second half of the neural network consists of a single fully connected layer of size 256 followed by five fully connected layers of size 64 in parallel for predicting a, T, δ, K, and η. The role of the first fully connected layer is to interconnect all of the patterns recognized by the convolutional layers. As the magnetic coupling constants are all interrelated parameters, this information can be shared for predicting all five values, and allows for a reduction in both the number of neural network parameters, and consequently reduces overfitting to the training data. The numbers of layers and their sizes were optimized while holding all other aspects of the network constant. In total, the CNN comprises 3 convolutional layers and 1 fully connected layer, which then branches into 5 fully connected layers in parallel for predicting a, T, δ, K, and η (Figure 2).

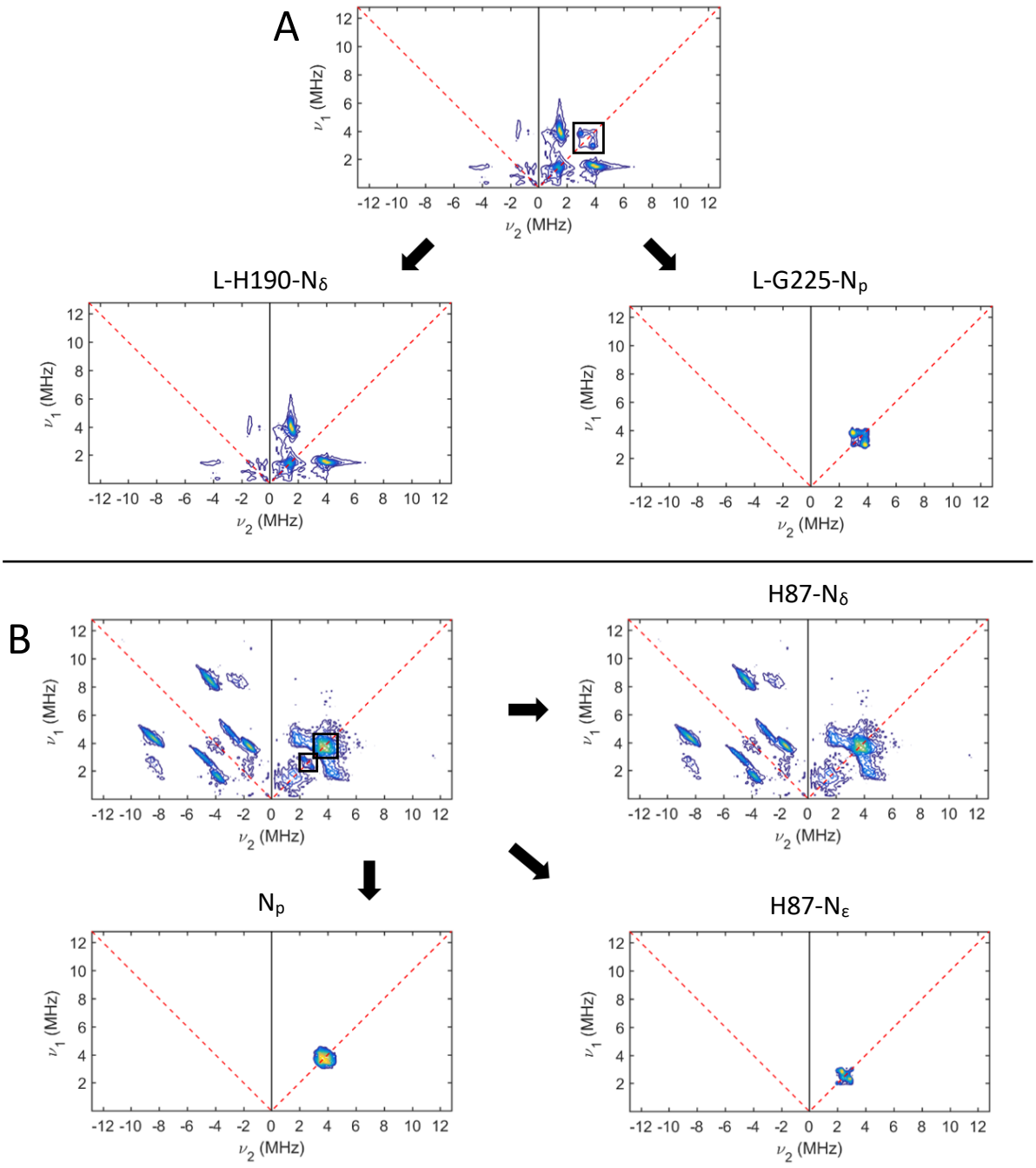
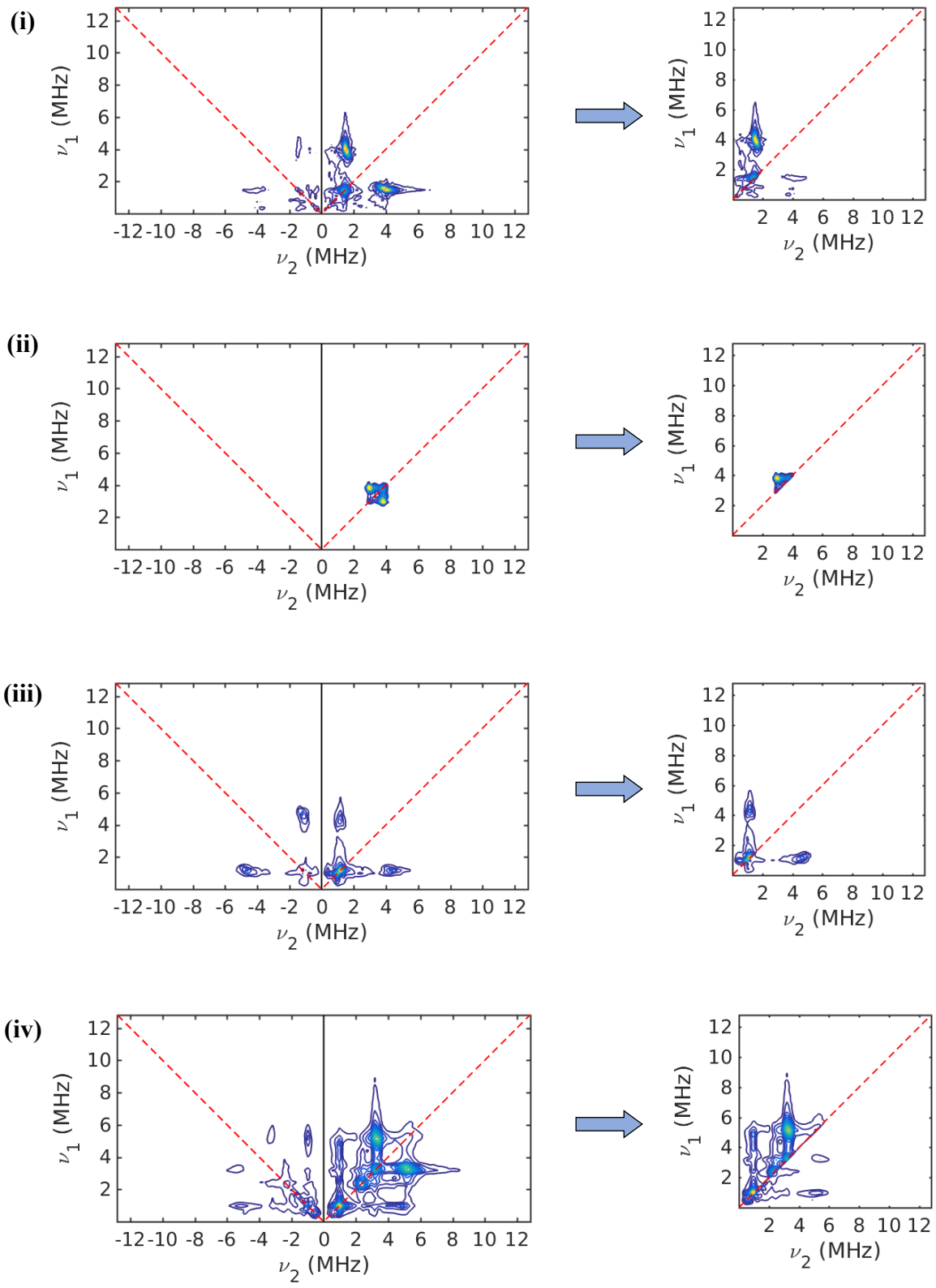
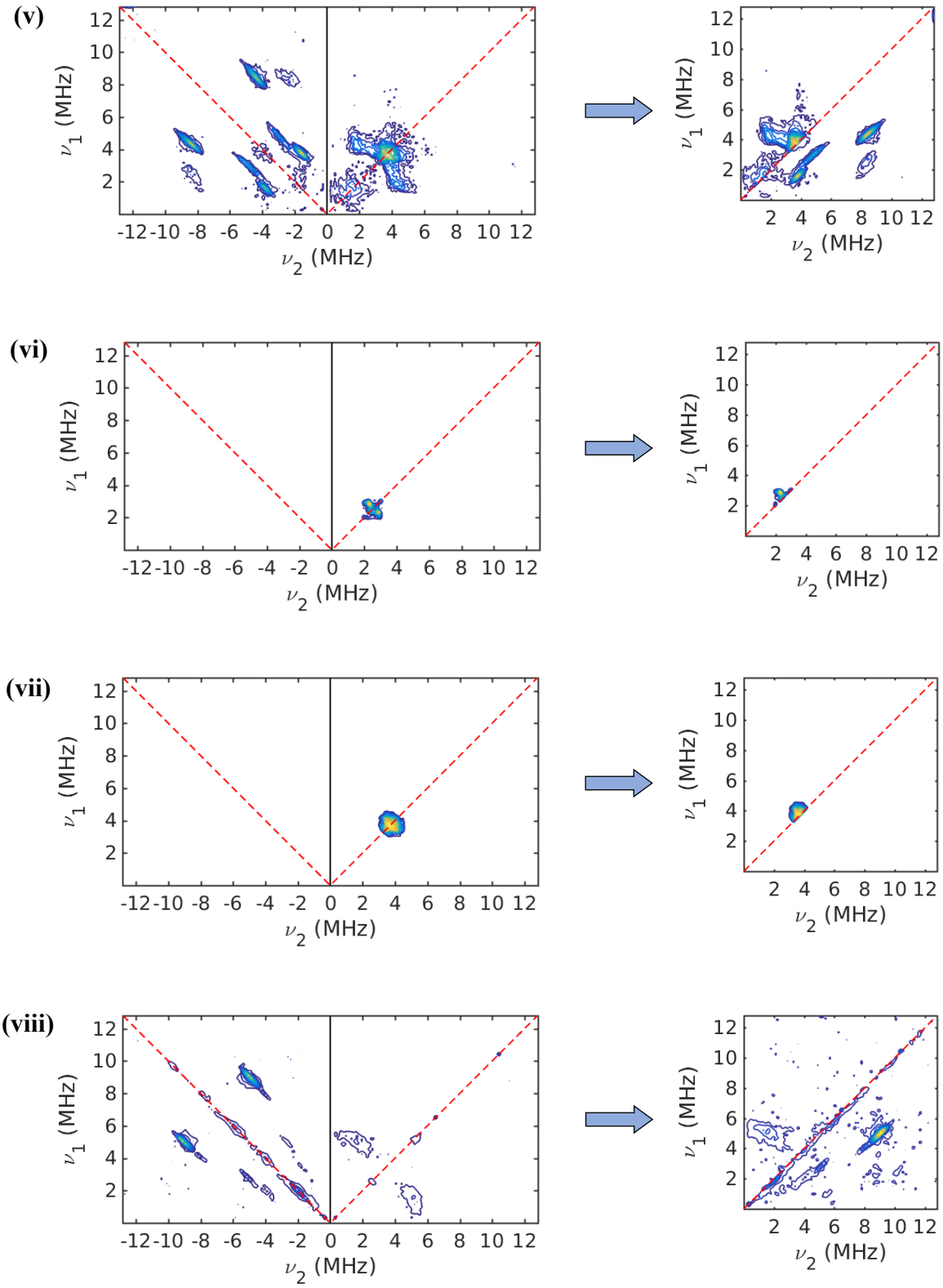


Figure S1. Manual separation of the SQ_B[8] (**A**) and rat mitoNEET[9] (**B**) spectra into their individual ^{14}N components. The black boxes indicate the cropped regions of HYSCORE spectra. Deleted regions of the spectra were replaced with noise. In the case of rat mitoNEET, the feature at (4, 4) MHz marks a region of overlap between the H87-N_δ and N_p nitrogens and was left in in both cases.

Transformation of Experimental ^{14}N HYSCORE Spectra into Square Matrix Representation




(v)
(vi)
(vii)
(viii)
ν_1 (MHz)
ν_2 (MHz)
12
10
8
6
4
2
-12 -10 -8 -6 -4 -2 0 2 4 6 8 10 12

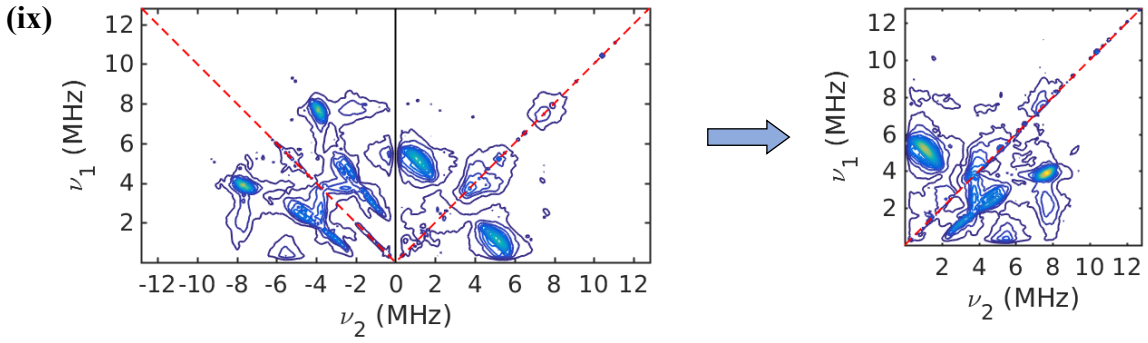


Figure S2. Transformation of experimental ^{14}N HYSCORE spectra into square matrix representation that were fed into the neural network to generate Figure 3 in the main text: (i) SQ_B RC (L-H190-N_δ)[8], (ii) SQ_B RC (L-G225-N_p)[8], (iii) SQ_H cyt *aa*$_3$ (H70-N_ε)[10], (iv) SQ_H cyt *bo*$_3$ (R71-N_ε)[11], (v) [2Fe-2S] rat mitoNEET (H87-N_δ)[9], (vi) [2Fe-2S] rat mitoNEET (H87-N_ε)[9], (vii) [2Fe-2S] rat mitoNEET (N_p)[9], (viii) VO^{2+} (Imidazole)[12], and (ix) VO^{2+} (Histidine-N_α)$_2$[12].

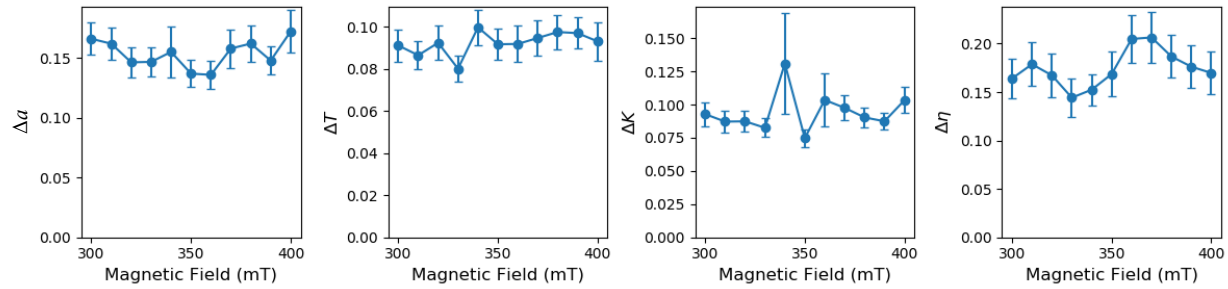


Figure S3. Performance of trained neural network HYnet-$B_0\tau$ in Table 1 as a function of magnetic field. The absolute difference between the real and machine learning predicted magnetic coupling parameters (Δa, ΔT, ΔK, and $\Delta \eta$) was calculated from an average of 100 simulated HYSCORE spectra for each data point. The magnetic field was incremented in steps of 10 mT between 300-400 mT.

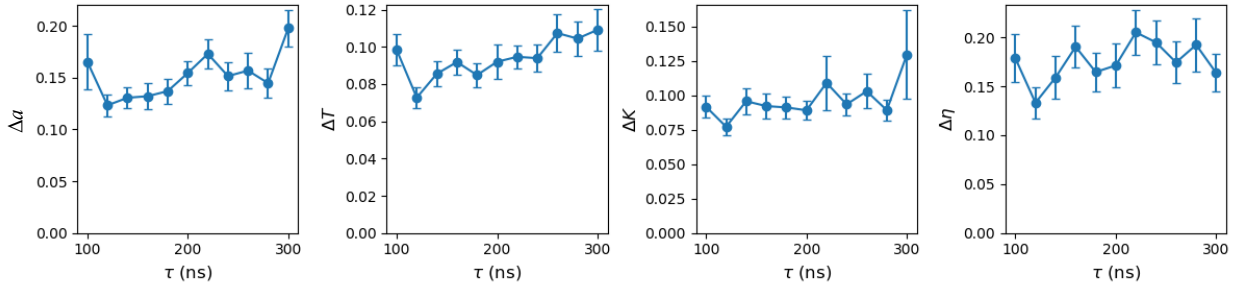


Figure S4. Performance of trained neural network HYnet-$B_0\tau$ in Table 1 as a function of τ-value. The absolute difference between the real and machine learning predicted magnetic coupling parameters (Δa, ΔT, ΔK, and $\Delta \eta$) was calculated from an average of 100 simulated HYSCORE spectra for each data point. The τ-value was incremented in steps of 20 ns between 100-300 ns.

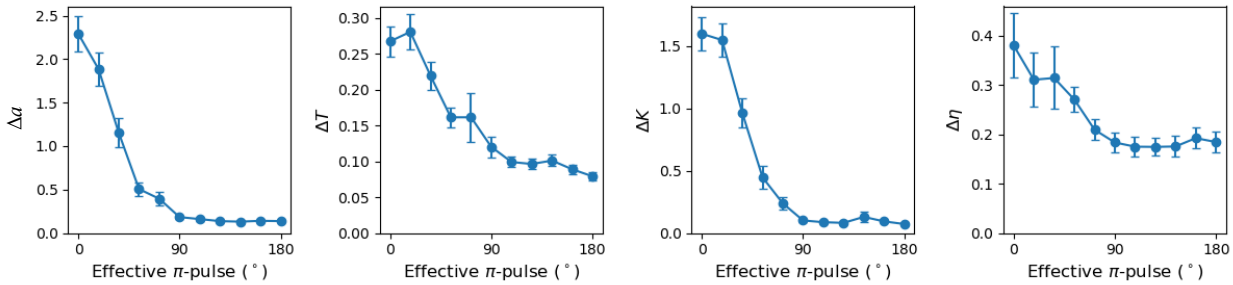


Figure S5. Performance of trained neural network HYnet-$B_0\tau$ in Table 1 as a function of the effective turning angle of the π-pulse. This parameter determines the ratio of the diagonal peak intensity to the cross-peak intensity, where turning angles <90° give very little spin inversion, resulting in almost no cross-peak intensity. The absolute difference between the real and machine learning predicted magnetic coupling parameters (Δa, ΔT, ΔK, and $\Delta \eta$) was calculated from an average of 100 simulated HYSCORE spectra for each data point. The π-pulse turning angle was incremented in steps of 18° between 0-180°.

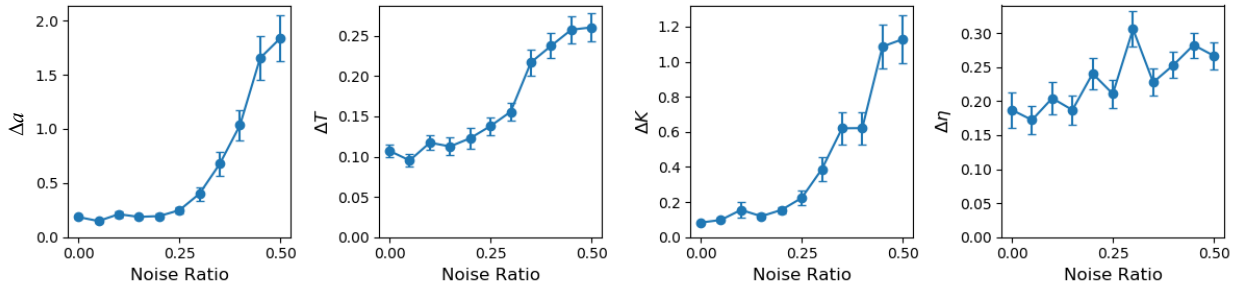


Figure S6. Performance of trained neural network HYnet-$B_0\tau$ in Table 1 as a function of the noise-to-signal ratio. Performance begins to degrade at a noise ratio of ~0.25 (signal-to-noise ratio of ~4). The absolute difference between the real and machine learning predicted magnetic coupling parameters (Δa, ΔT, ΔK, and $\Delta \eta$) was calculated from an average of 100 simulated HYSCORE spectra for each data point. The noise ratio was incremented in steps of 0.05 between 0-0.5.

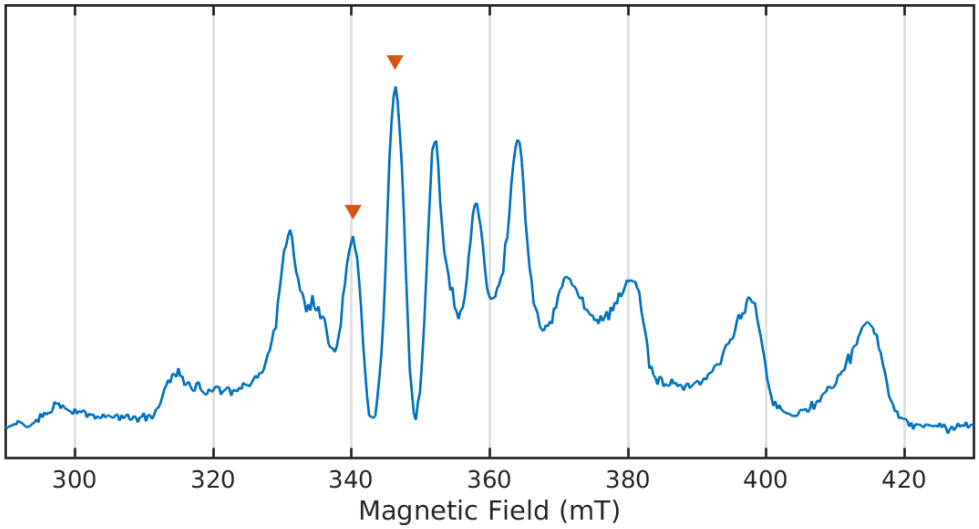


Figure S7. Two-pulse field-swept EPR spectrum of histidine-coordinated VO^{2+}.[12] HYSCORE spectra were acquired at 340 and 346 mT (red arrows). Experimental settings: Microwave frequency = 9.740 GHz, time between first and second pulses τ = 200 ns, temperature = 30 K.

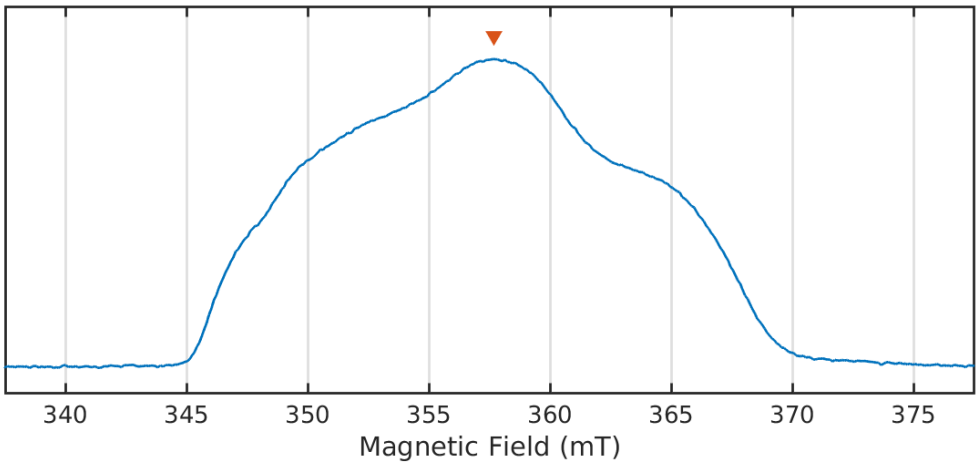
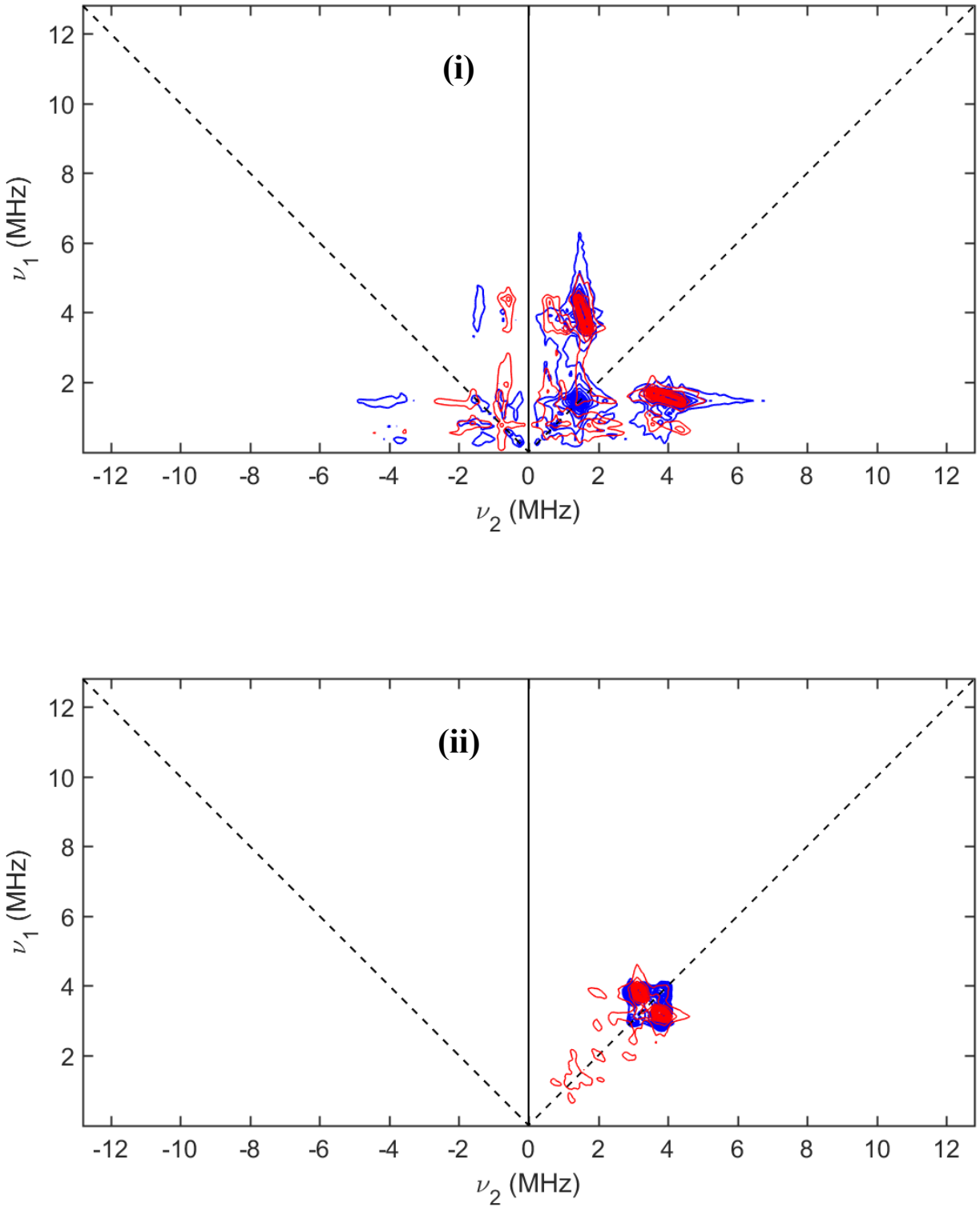
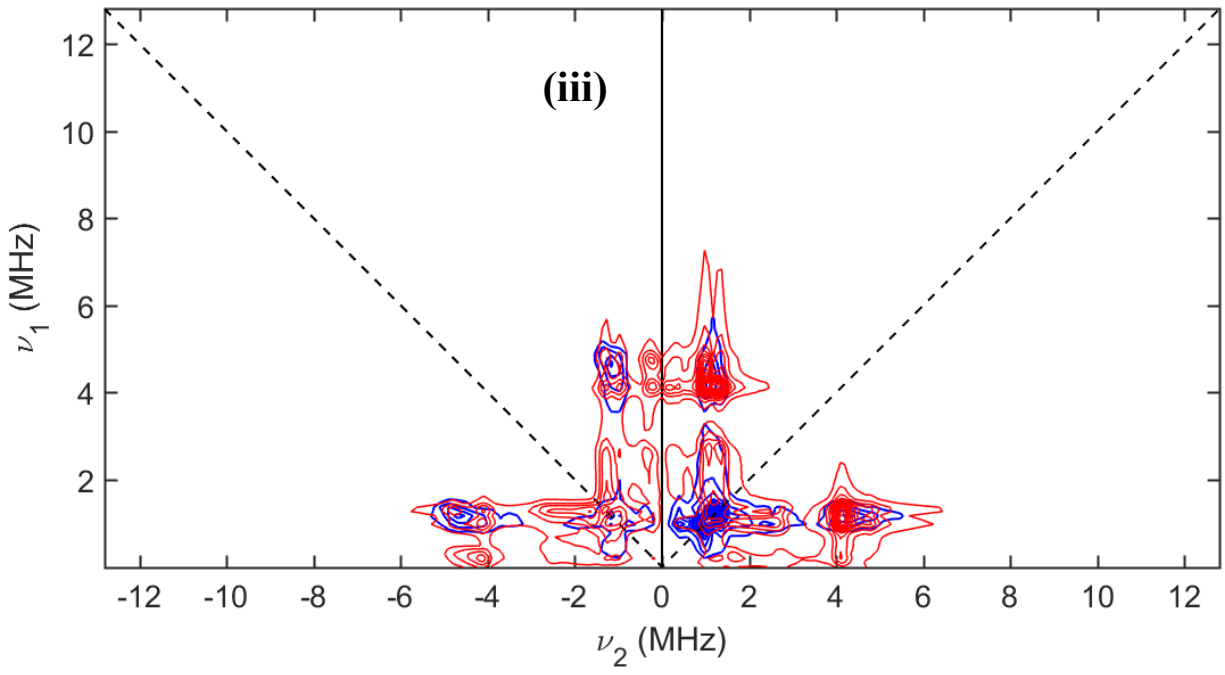
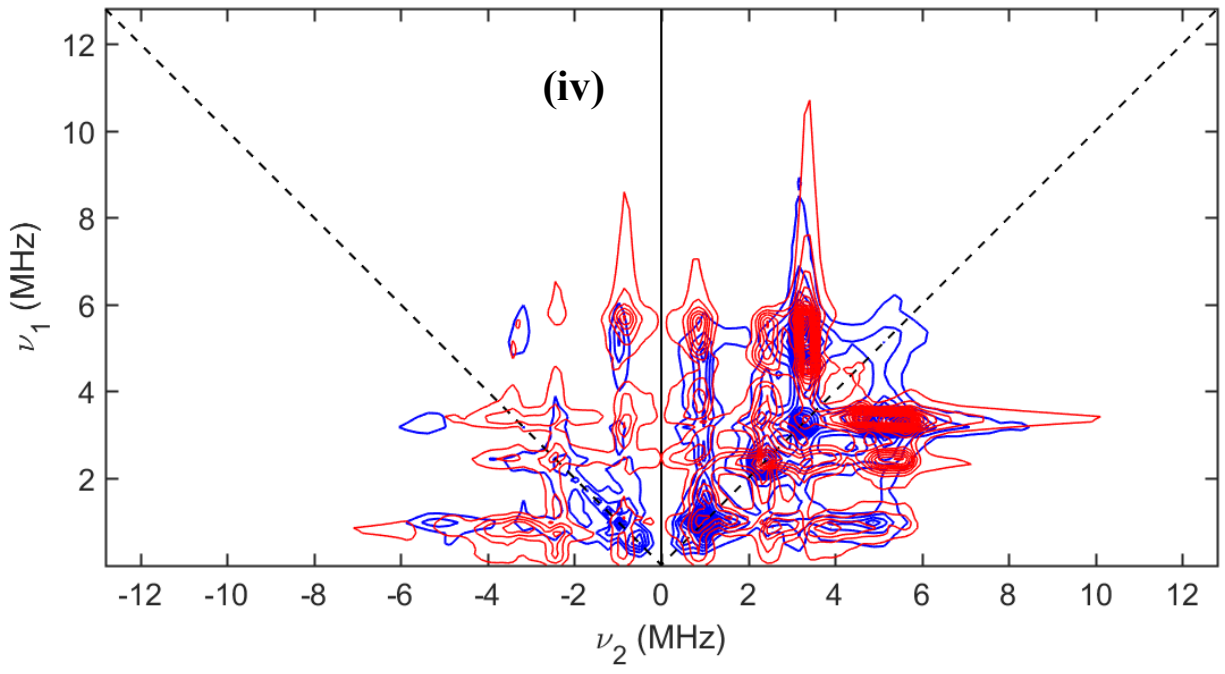


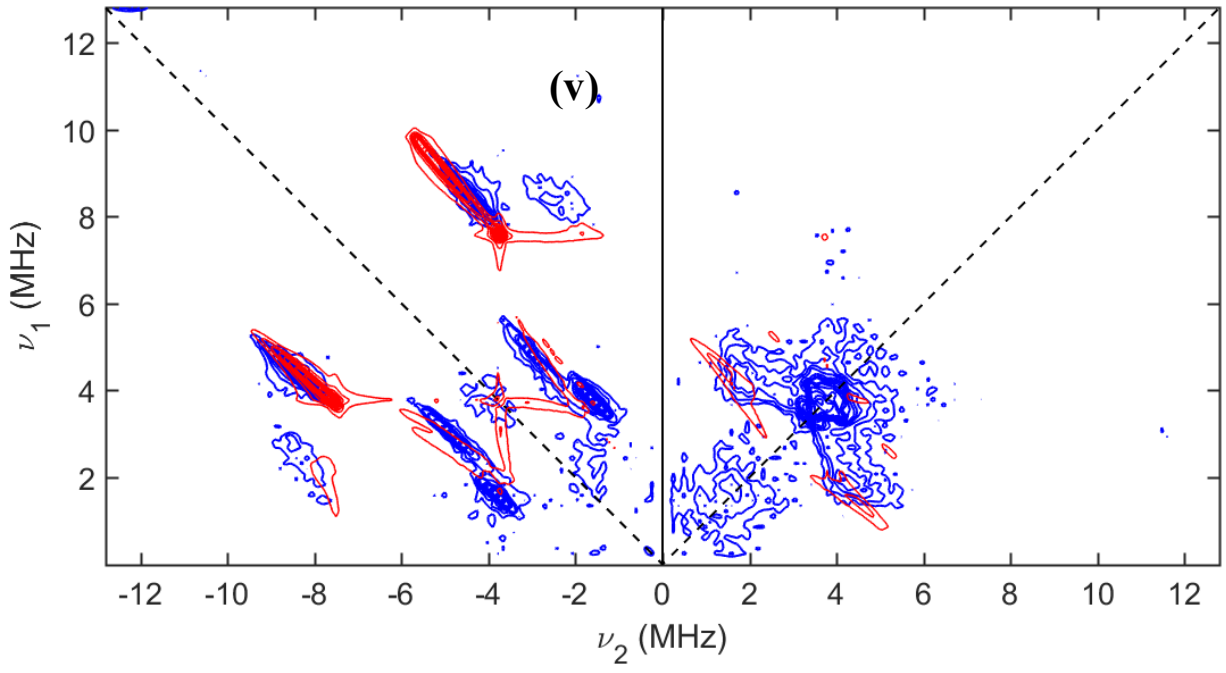
Figure S8. Two-pulse field-swept EPR spectrum of rat mitoNEET.[9] The HYSCORE spectrum was acquired at the g_{YY} orientation (red arrow). Experimental settings: Microwave frequency = 9.703 GHz, time between first and second pulses τ = 400 ns, temperature = 12 K.

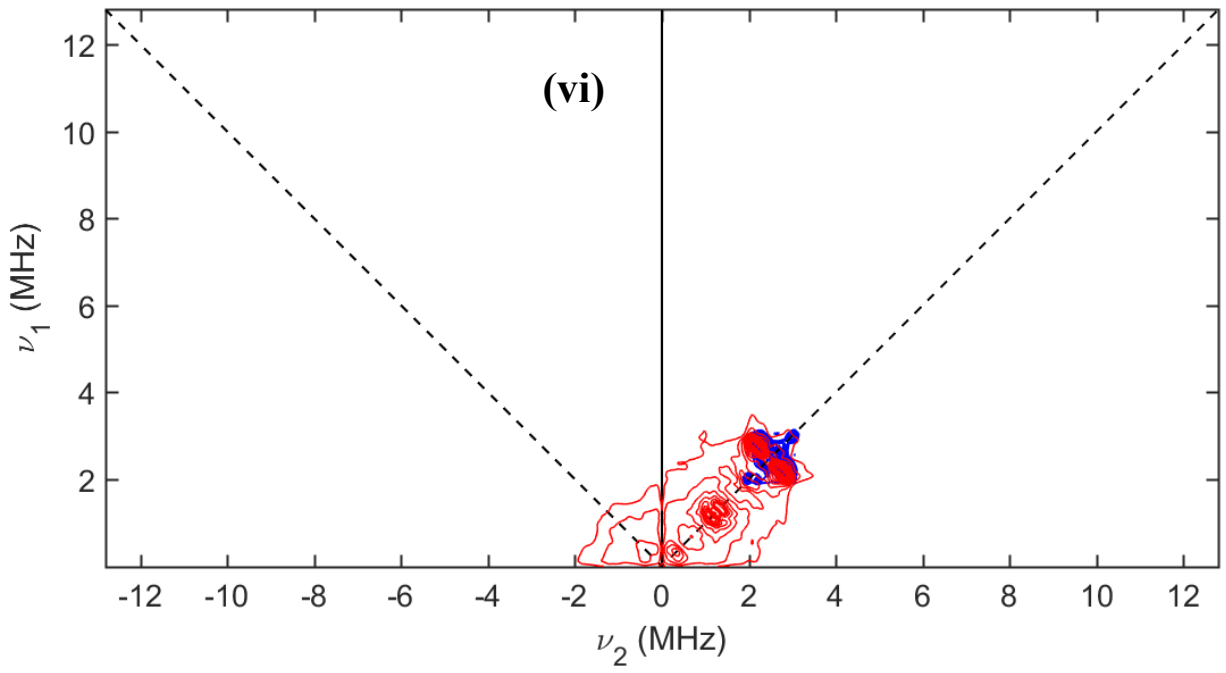
Comparison of Machine Learning Simulations with Experimental ^{14}N HYSCORE Spectra

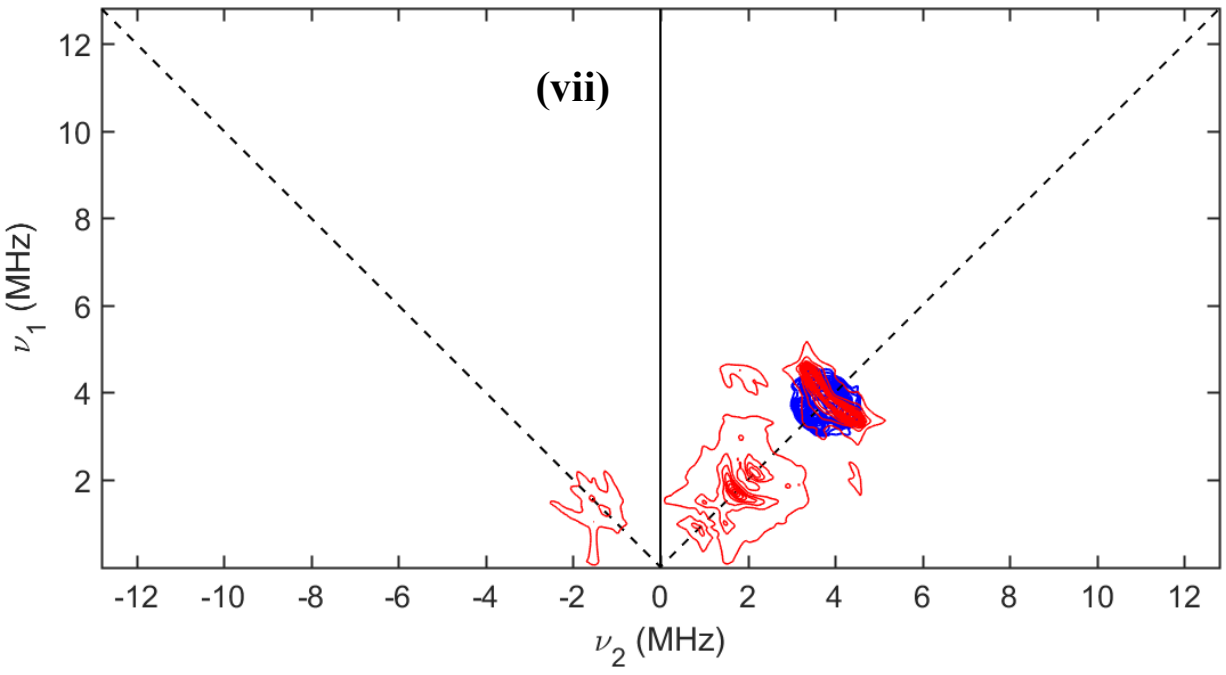


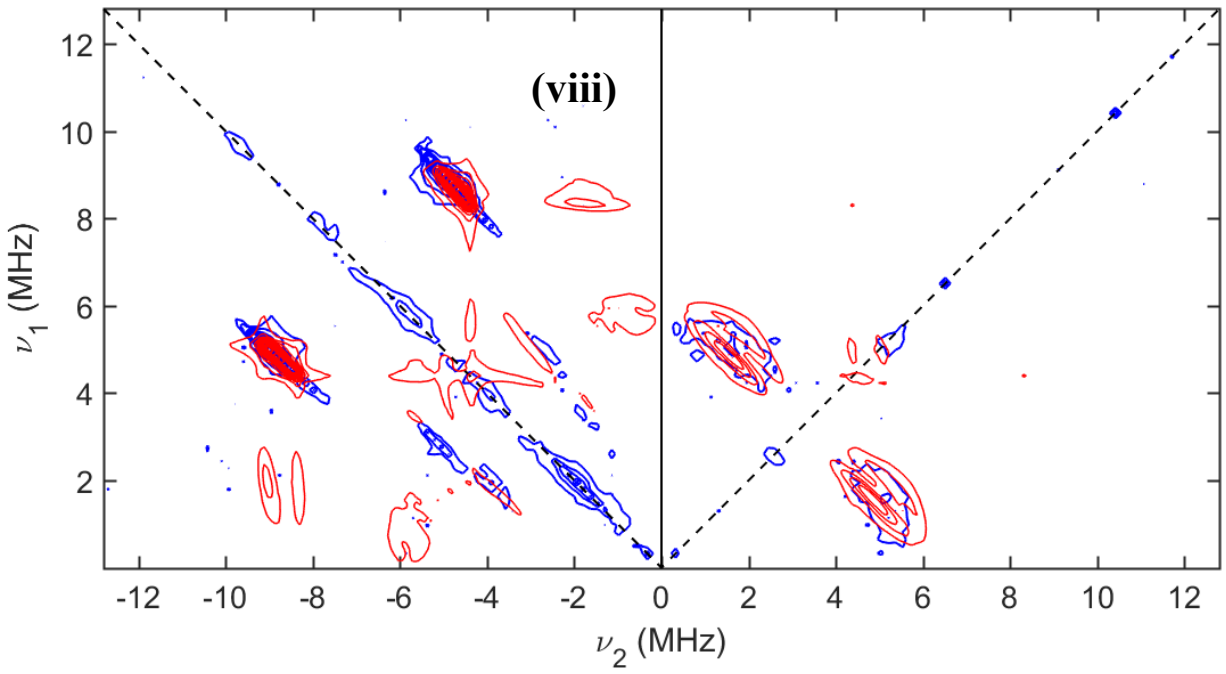

(iii)
12
10
8
6
4
2
ν_1 (MHz)
-12 -10 -8 -6 -4 -2 0 2 4 6 8 10 12
ν_2 (MHz)


(iv)
12
10
8
6
4
2
ν_1 (MHz)
-12 -10 -8 -6 -4 -2 0 2 4 6 8 10 12
ν_2 (MHz)


(v)
12
10
8
6
4
2
ν_1 (MHz)
-12
-10
-8
-6
-4
-2
0
2
4
6
8
10
12
ν_2 (MHz)


(vi)
12
10
8
6
4
2
ν_1 (MHz)
-12
-10
-8
-6
-4
-2
0
2
4
6
8
10
12
ν_2 (MHz)


(vii)
12
10
8
6
4
2
ν_1 (MHz)
-12 -10 -8 -6 -4 -2 0 2 4 6 8 10 12
ν_2 (MHz)


(viii)
12
10
8
6
4
2
ν_1 (MHz)
-12 -10 -8 -6 -4 -2 0 2 4 6 8 10 12
ν_2 (MHz)

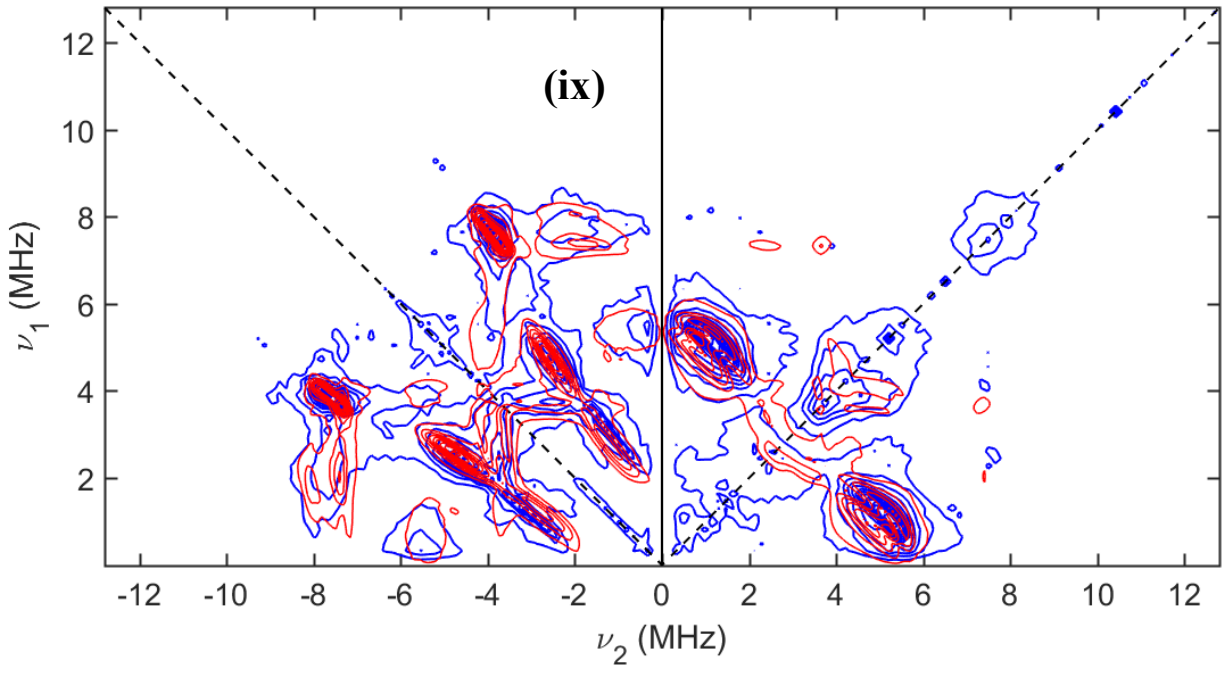


Figure S9. Comparison of simulations with machine learning predicted parameters with experimental ^{14}N HYSCORE spectra: (i) SQ_B RC (L-H190-N_δ)[8], (ii) SQ_B RC (L-G225-N_p)[8], (iii) SQ_H cyt *aa*$_3$ (H70-N_ε)[10], (iv) SQ_H cyt *bo*$_3$ (R71-N_ε)[11], (v) [2Fe-2S] rat mitoNEET (H87-N_δ)[9], (vi) [2Fe-2S] rat mitoNEET (H87-N_ε)[9], (vii) [2Fe-2S] rat mitoNEET (N_p)[9], (viii) VO^{2+} (Imidazole)[12], and (ix) VO^{2+} (Histidine-N_α)$_2$[12]. Simulations were performed in EasySpin[1] using the mean values highlighted in blue in Table S1 as the hyperfine and nuclear quadrupole parameters. The α and β Euler angles of the hyperfine tensor were both set arbitrarily to 45° to prevent any unusual resonance conditions due to perfect alignment of the hyperfine and nuclear quadrupole tensors. No orientation selection, hyperfine rhombicity (δ), or *a*-strain[3] were considered in these simulations.

Table S1. Neural network predictions for experimental spectra acquired at various experimental settings (results shown in Figure 3 are highlighted in blue, and a comparison with the previous literature values are in parentheses)

Spin System [a]	τ (ns)[b]	Δt (ns)[c]	a (MHz)	T (MHz)	K (MHz)	η
SQ_B RC (L-H190-N_δ)	136	20	1.33±0.53	0.47±0.23	1.50±0.39	0.79±0.30
	136	32	1.46±0.50	0.31±0.24	1.48±0.37	0.81±0.30
	200	20	1.29±0.50	0.33±0.24	1.47±0.38	0.82±0.27
	200	32	1.34±0.48	0.25±0.24	1.42±0.37	0.84±0.23
			(1.35)	(0.35)	(1.52)	(0.69)
SQ_B RC (L-G225-N_p)	136	20	0.44±0.40	0.16±0.26	2.93±0.31	0.77±0.26
	136	32	0.42±0.43	0.15±0.26	2.93±0.32	0.77±0.26
	200	20	0.41±0.41	0.13±0.25	3.10±0.35	0.64±0.24
	200	32	0.29±0.41	0.19±0.27	3.14±0.35	0.60±0.24
			(0.4)	(0.2)	(2.96)	(–)
SQ_H cyt aa_3 (H70-N_ε)	136	20	2.04±0.49	0.24±0.26	1.55±0.41	0.15±0.25
			(2.0±0.1)	(0.2±0.1)	(1.44±0.1)	(0.25±0.1)
SQ_H cyt bo_3 (R71-N_ε)	120	16	1.61±0.52	0.53±0.22	3.91±0.39	0.42±0.16
	152	16	1.55±0.52	0.49±0.22	3.80±0.34	0.45±0.15
	260	16	1.70±0.60	0.29±0.33	3.69±0.35	0.43±0.15
			(1.8)	(0.4)	(3.72)	(0.51)
[2Fe-2S] rat mitoNEET (H87-N_δ)[d]	136	32	5.73±0.53	0.80±0.29	2.69±0.41	0.87±0.20
			(6.1)	(0.8)	(2.40)	(–)
[2Fe-2S] rat mitoNEET (H87-N_ε)[d]	136	32	0.22±0.31	0.14±0.24	1.18±0.40	0.80±0.27
			(0.32)	(~0.2)	(1.32)	(–)
[2Fe-2S] rat mitoNEET (N_p)[d]	136	32	0.32±0.41	0.41±0.24	3.31±0.36	0.81±0.26
			(0.5)	(~0.2)	(3.04)	(–)
VO^{2+} (Imidazole)	256	24	6.23±0.53	0.28±0.24	2.35±0.44	0.76±0.31
			(6.3)	(0.45)	(–)	(–)
VO^{2+} (Histidine-N_α)$_2$[e]	104	24	4.86±0.46	0.17±0.25	2.82±0.39	0.88±0.21
	256	24	4.99±0.44	0.25±0.25	2.74±0.41	0.85±0.26
			(5.0)	(–)	(2.32)	(–)
VO^{2+} (Histidine-N_α)$_2$[f]	104	24	5.03±0.49	0.21±0.25	2.72±0.44	0.64±0.24
	256	24	5.10±0.53	0.28±0.22	2.64±0.45	0.81±0.26
			(5.0)	(–)	(2.32)	(–)

[a]Spin system descriptions – SQ_B RC: Semiquinone in the Q_B ubiquinone binding site of the bacterial reaction center from *Rhodobacter sphaeroides*. SQ_H cyt aa_3: The semiquinone in the Q_H high-affinity menaquinone binding site of the R70H mutant of the cytochrome aa_3-600 menaquinol oxidase from *Bacillus subtilis.* SQ_H cyt bo_3: The semiquinone in the Q_H high-affinity ubiquinone binding site of the cytochrome bo_3 ubiquinol oxidase from *Escherichia coli.* [2Fe-2S] rat mitoNEET: Reduced [2Fe-2S](Cys)$_3$(His)$_1$ cluster in rat mitoNEET.VO^{2+} (Imidazole): VO^{2+} coordinated to imidazole. VO^{2+} (Histidine-N_α)$_2$: VO^{2+} coordinated to histidine. [b]τ is the time between the first and second pulses of the HYSCORE sequence. [c]Δt is the step size that t_1 and t_2 are incremented in the HYSCORE time-domain pattern. [d]Spectrum acquired at the g_{YY} orientation (Figure S8). [e]Spectrum acquired at 340 mT (Figure S7). [f]Spectrum acquired at 346 mT (Figure S7).

References

1. Stoll, S.; Schweiger, A. EasySpin, a comprehensive software package for spectral simulation and analysis in EPR. *J. Magn. Reson.* 2006, *178*, 42-55.
2. Höfer, P.; Grupp, A.; Nebenführ, H.; Mehring, M. Hyperfine sublevel correlation (HYSCORE) spectrocopy: A 2D ESR investigation of the squaric acid radical. *Chem. Phys. Lett.* 1986, *132*, 279– 282.
3. Dikanov, S. A.; Taguchi, A. T. Two-Dimensional Pulsed EPR Resolves Hyperfine Coupling Strain in Nitrogen Hydrogen Bond Donors of Semiquinone Intermediates. *J. Phys. Chem. B* 2018, *122*, 5205-5211.
4. Nair, V.; Hinton, G. E. Rectified Linear Units Improve Restricted Boltzmann Machines. In *27th International Conference on Machine Learning*, 2010; pp 807-814.
5. Dugas, C.; Bengio, Y.; Bélisle, F.; Nadeau, C.; Garcia, R. Incorporating Second-Order Functional Knowledge for Better Option Pricing. *13th International Conference on Neural Information Processing Systems* 2000.
6. Kingma, D. P.; Ba, J. L. Adam: A Method for Stochastic Optimization. In *3rd International Conference on Learning Representations*, 2015.
7. Srivastava, N.; Hinton, G. E.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *J. Mach. Learn. Res.* 2014, *15*, 1929-1958.
8. Taguchi, A. T.; O'Malley, P. J.; Wraight, C. A.; Dikanov, S. A. Nuclear hyperfine and quadrupole tensor characterization of the nitrogen hydrogen bond donors to the semiquinone of the Q_B site in bacterial reaction centers: a combined X- and S-band ^{14,15}N ESEEM and DFT study. *J. Phys. Chem. B* 2014, *118*, 1501-1509.
9. Iwasaki, T.; Samoilova, R. I.; Kounosu, A.; Ohmori, D.; Dikanov, S. A. Continuous-wave and pulsed EPR characterization of the [2Fe-2S](Cys)$_3$(His)$_1$ cluster in rat MitoNEET. *J. Am. Chem. Soc.* 2009, *131*, 13659-13667.
10. Yi, S. M.; Taguchi, A. T.; Samoilova, R. I.; O'Malley, P. J.; Gennis, R. B.; Dikanov, S. A. Plasticity in the High Affinity Menaquinone Binding Site of the Cytochrome aa_3-600 Menaquinol Oxidase from *Bacillus subtilis*. *Biochemistry* 2015, *54*, 5030-5044.
11. Lin, M. T.; Baldansuren, A.; Hart, R.; Samoilova, R. I.; Narasimhulu, K. V.; Yap, L. L.; Choi, S. K.; O'Malley, P. J.; Gennis, R. B.; Dikanov, S. A. Interactions of intermediate semiquinone with surrounding protein residues at the Q_H site of wild-type and D75H mutant cytochrome bo_3 from *Escherichia coli*. *Biochemistry* 2012, *51*, 3827-3838.
12. Dikanov, S. A.; Samoilova, R. I.; Smieja, J. A.; Bowman, M. K. Two-Dimensional ESEEM Study of VO^{2+} Complexes with Imidazole and Histidine: Histidine Is a Polydentante Ligand. *J. Am. Chem. Soc.* 1995, *117*, 10579-10580.