Supporting Information for

Helix-Capping Histidines: Diversity of N-H···N Hydrogen Bond Strength Revealed by $^{2h}J_{NN}$ Scalar Couplings

Matthew R. Preimesberger, Ananya Majumdar, Selena L. Rice, Lauren Que,

and Juliette T.J. Lecomte

Protein preparation

Protein preparation

Plasmid DNA encoding either *C. eugametos* CtrHb, *C. reinhardtii* THB1, *Synechococcus* sp. PCC 7002 GlbN or *Synechocystis* sp. PCC 6803 GlbN (and variants thereof) was used to transform *E. coli* BL21(DE3) cells, which were then grown at 37 °C for ~24 h on minimal medium plates. Transformant colonies were used to inoculate minimal medium (M9) starter cultures (~60–100 mL) and grown for an additional ~24 h at 37 °C. At this point, the starter culture was divided into fourths and used to seed 4 × 500 mL M9 expression cultures. To produce uniformly ^{15}N-labeled protein, $^{15}NH_4Cl$ (Sigma) was used as the sole nitrogen source. For *C. eugametos* CtrHb, *C. reinhardtii* THB1 and *Synechocystis* GlbN, expression cultures were allowed to continue growing until an O.D. (550 nm) ~ 0.8 – 1.0 was achieved. Overexpression was induced by addition of 0.5 mM (final concentration) IPTG (Santa-Cruz Biotechnology), and was allowed to continue for ~5–8 h prior to cell harvesting by centrifugation. In the case of *Synechococcus* GlbN, leaky overexpression occurs and the IPTG addition step is not required. Harvested cells were frozen at –20 °C until further use.

The recombinant hemoglobins expressed in this manner partitioned primarily into inclusion bodies. Cells were lysed by sonication; soluble and insoluble fractions were separated by centrifugation, and the lysate was discarded. Inclusion bodies were solubilized in ~8 M urea (50 mM Tris, 1 mM EDTA, pH 8) and passed over a G-50 size exclusion column to achieve apoprotein refolding and purification. Fractions deemed pure by SDS-PAGE were pooled and reconstituted with porcine hemin (> 98%, Sigma) to generate ferric holoprotein. Passage of the resulting solution over an anion exchange (DEAE) column removed any excess heme and resulted in > 95 % purity in the pooled protein samples. Buffer exchange (into ~1 mM phosphate, pH ~7) was achieved using an Amicon ultrafiltration unit and Millipore YM3 membrane (molecular weight cut off = 3 kDa). The pooled samples were concentrated to ~20 mL and protein concentration was assessed by optical spectroscopy as reported previously. Typical yields of recombinant hemoglobins were ~25–50 mg/L culture. Samples not used immediately for NMR spectroscopic studies were lyophilized and stored at –80 °C until needed.

A similar protocol was used for the expression and purification of ^{15}N-labeled rat microsomal cytochrome b_5. Following overexpression in *E. coli* BL21(DE3) cells, recombinant cytochrome b_5 primarily partitioned to the soluble fraction. An excess of hemin was added to the raw lysate in order to produce ferric holoprotein. Following the heme binding step, ferric cytochrome b_5 purification was achieved via DEAE anion exchange chromatography followed by G-50 size-exclusion chromatography. Apo cytochrome b_5 was prepared from ferric holoprotein using an acid-butanone heme extraction procedure (1:1 H_2O/butanone, pH 2.5, 4 °C) followed by extensive dialysis and buffer exchange to remove residual butanone. Optical absorbance measurements on the resulting apocytochrome b_5 sample confirmed that > 99% of the heme had been removed.

Table S1. List of heme protein sample conditions[a]

Protein	Distal ligand	Solvent
WT CtrHb[b]	CN	125 mM phosphate, 10 mM KCN, pH ~7.2, 10% D_2O
L75H CtrHb-B	CN	100 mM phosphate, 12 mM $KC^{15}N$, pH ~7.2, 10% D_2O
WT THB1	CN	90% H_2O, 3.3 mM KCN, pH 7.49, 10% D_2O
*Syn*7002 GlbN	His	100 mM phosphate, pH ~7.1, 10% D_2O
*Syn*7002 GlbN-A	His	100 mM phosphate, pH ~7.1, 10% D_2O
*Syn*6803 GlbN	CN	125 mM phosphate, pH 7.27, 10 mM KCN, 10% D_2O
*Syn*6803 GlbN-A	CN	125 mM phosphate, pH ~7.2, 10 mM KCN, 10% D_2O
*Syn*6803 GlbN	His	125 mM phosphate, pH ~7.2, 10% D_2O
*Syn*6803 GlbN-A	His	125 mM phosphate, pH ~7.2, 10% D_2O
*Syn*6803 GlbN-B	His	25 mM phosphate, pH ~7.1, 10% D_2O
*Syn*6803 GlbN-AB	His	125 mM phosphate, pH ~7.2, 10% D_2O
*Syn*6803 GlbN-B	CN	100 mM phosphate, pH ~7.1, 6 mM KCN, 10% D_2O
*Syn*6803 GlbN-AB	CN	125 mM phosphate, pH ~7.2, 6 mM KCN, 10% D_2O
Fe^{III} cytochrome b_5	His	5 mM phosphate, pH 7.02, 10% D_2O
Fe^{II} cytochrome b_5	His	50 mM phosphate, pH ~7.1, 10% D_2O
apo cytochrome b_5[c]	n/a[d]	5 mM phosphate, pH 6.97, 10% D_2O, 298 K

[a]All data acquired at 313 K unless otherwise noted. [b]Data were acquired at both 313 K and 283 K for this protein. [c]Data acquired at 298 K. [d]Not applicable.

Table S2. List of histidine pK_a values and $^{2h}J_{NN}$ couplings

Protein	Capping His	$^{2h}J_{NN}$ (observed) (Hz)	pK_a	$^{2h}J_{NN}$ (capped) (Hz)	pK_a Remark
CtrHb-CN, M	His79	5.0	< 3.5	5.0	acid unfolding, Figure S7
CtrHb-B-CN	His79	4.0	~ 5.0	4.2	Figure S7
*Syn*7002 GlbN	His83	4.7	< 4.0	4.7	acid unfolding, Figure S8
*Syn*6803 GlbN	His83	4.9	< 4.0	4.9	acid unfolding[1]
ferric cyt b_5	His80	5.0	< 5.5		[2]
apocyt b_5	His80	5.3[a]	< 5.5	5.5	closed[3]
apocyt b_5	His80	0[b]	~ 6.9	0[b]	open[3]
AR NRRC[a]	His47	4.0	< 4.0[c]	4.0	Figure S6
AR NRRC[a]	His80	3.9	< 4.0[c]	3.9	Figure S6
AR NRRC[a]	His113	2.1	~ 5.7	2.7	Figure S6

[a]Data were acquired at 298 K for this protein. [b]Assumed $^{2h}J_{NN}$ for non H-bonded state. [c]Estimated from absence of pH dependent chemical shift changes and comparison with previously published titration data acquired on the three-repeat consensus AR protein, NRR.

Table S3. Histidine tautomers, pK_a values, microstate populations, and $^{2h}J_{NN}$

	Microstate	pK_a	Population	$^{2h}J_{NN}$ (Hz)
wild-type CtrHb-CN[a]				
observed	Nε2H	< 3.5[b]	–	5.0
capped	Nε2H	3.00[c]	~1.0	5.0[d]
open	Nε2H	6.53[e]	~0	0
open	Nδ1H	6.92[e]	~0	0
protonated	–	–	~0	0
L75H CtrHb-B-CN[a]				
observed	Nε2H	< 5.5[b]	–	4.0
capped	Nε2H	5.02[c]	0.953	4.2[d]
open	Nε2H	6.53[e]	0.029	0
open	Nδ1H	6.92[e]	0.011	0
protonated	–	–	0.006	0
consensus AR protein[f]				
observed	Nε2H	~5.7[g]	–	2.1
capped	Nε2H	5.76[c]	0.772	2.7[d]
open	Nε2H	6.73[h]	0.083	0
open	Nδ1H	7.12[h]	0.034	0
protonated	–	–	0.112	0

[a]His79 populations evaluated at pH 7.2. [b]Estimated from LR-HMQC and HSQC spectra at 313 K (Figure S7). [c]Microscopic pK_a used in the calculations. [d]Calculated $^{2h}J_{NN}$ value for a hypothetical 100% capped state. [e]Open state imidazolium pK_a values for Nα-acetyl-histidine methylamide.[4,5] [f]His113 populations evaluated at pH 6.6. [g]Estimated from LR-HMQC and HSQC spectra at 298 K (Figure S6). [h]Values at 310 K[5] adjusted to 298 K with ΔH = 30 kJ/mol.[6]

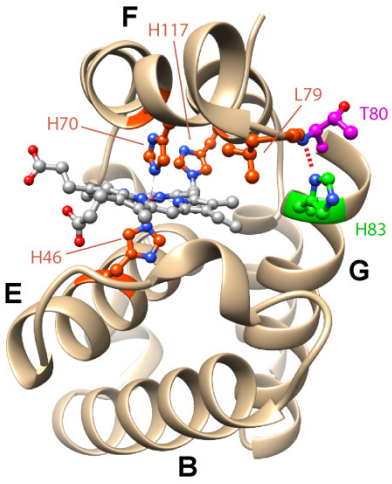

Figure S1A. Ribbon diagram of GlbN-A (4MAX, chain A) showing the location of the histidine N-cap in relation to the heme binding pocket. Helices B, E, F, and G are labeled according to canonical myoglobin nomenclature.

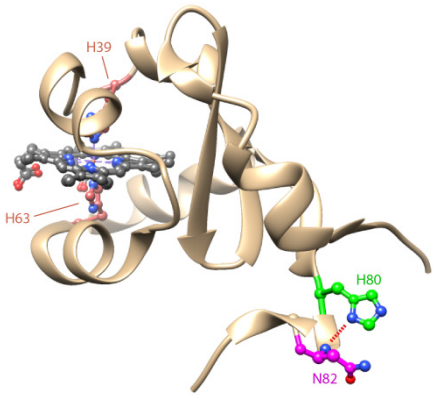

Figure S1B. Ribbon diagram of cytochrome b_5 (4HIL) showing the location of the histidine N-cap in relation to the heme binding pocket.

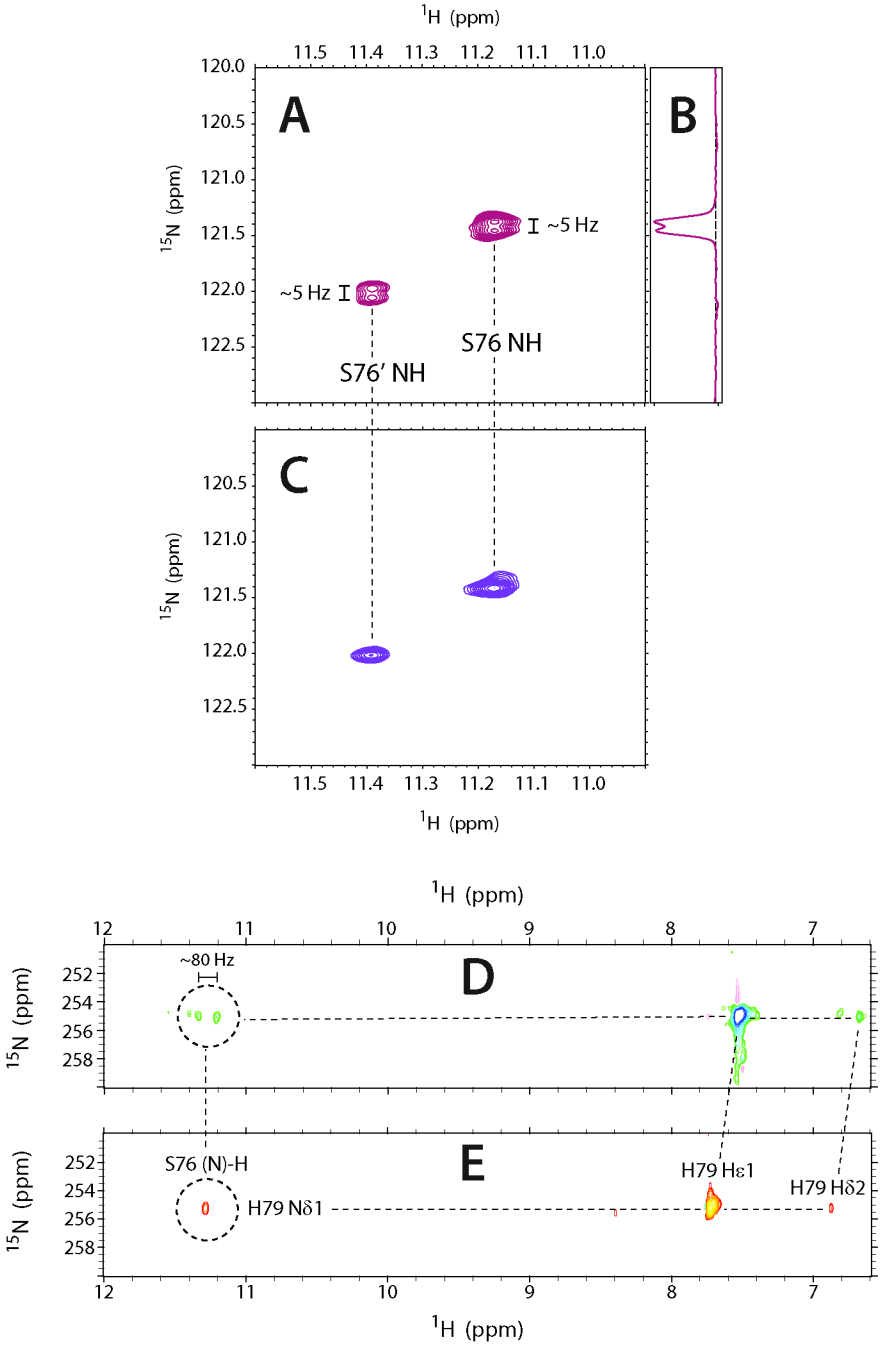

Figure S2. *C. eugametos* CtrHb-CN: initial experiments showing $^{1h}J_{HN}$ and $^{2h}J_{NN}$ couplings due to the Ser76 N-H···Nδ1 His79 H-bond

A. Initial evidence for homonuclear ^{15}N-^{15}N hydrogen bond scalar coupling. High-resolution ^{1}H-^{15}N HSQC (downfield ^{1}H region, ^{15}N evolution time = 300 ms) showing partially resolved Ser76 NH ^{15}N doublets. No other amide correlations in the spectrum display this splitting.

B. ^{15}N 1-D slice of the Ser76 major NH signal (^{1}H = 11.17 ppm).

C. High resolution ^{1}H-^{15}N HSQC (downfield ^{1}H region, ^{15}N evolution time as in a.) collected with a frequency-shifted ^{15}N decoupling pulse (histidine region, 255 ppm) during the ^{15}N amide evolution period. With this pulse sequence modification, the ^{15}N-^{15}N hydrogen bond scalar coupling is refocused and Ser76 NH signals collapse into singlets.

D. High-sensitivity ^{1}H-^{15}N LR HMQC (downfield ^{15}N, ^{1}H region), with ^{15}N decoupling (210 ppm, 298 K) during ^{1}H acquisition. The detection of the very weak doublet provided initial evidence for the identity of Ser76 N-H···Nδ1 His79 H-bonding partners. Here, the one-bond hydrogen bond scalar coupling, $^{1h}J_{HN}$, is responsible for coherence transfer. The directly detected Ser76 amide proton signal remains split owing to incomplete decoupling of $^{1}J_{NH}$.

E. As a test for the above hypothesis, a ^{1}H-^{15}N long-range HMQC spectrum (downfield ^{15}N, ^{1}H region) was acquired with frequency-shifted ^{15}N decoupling (122 ppm, 313 K) during ^{1}H acquisition. The $^{1h}J_{HN}$ –mediated correlation between Ser76 amide ^{1}H and His79 Nδ1 nuclei is observed as in d. However, in this instance, the $^{1}J_{NH}$ coupling is refocused and the Ser76 amide ^{1}H collapses into a singlet.

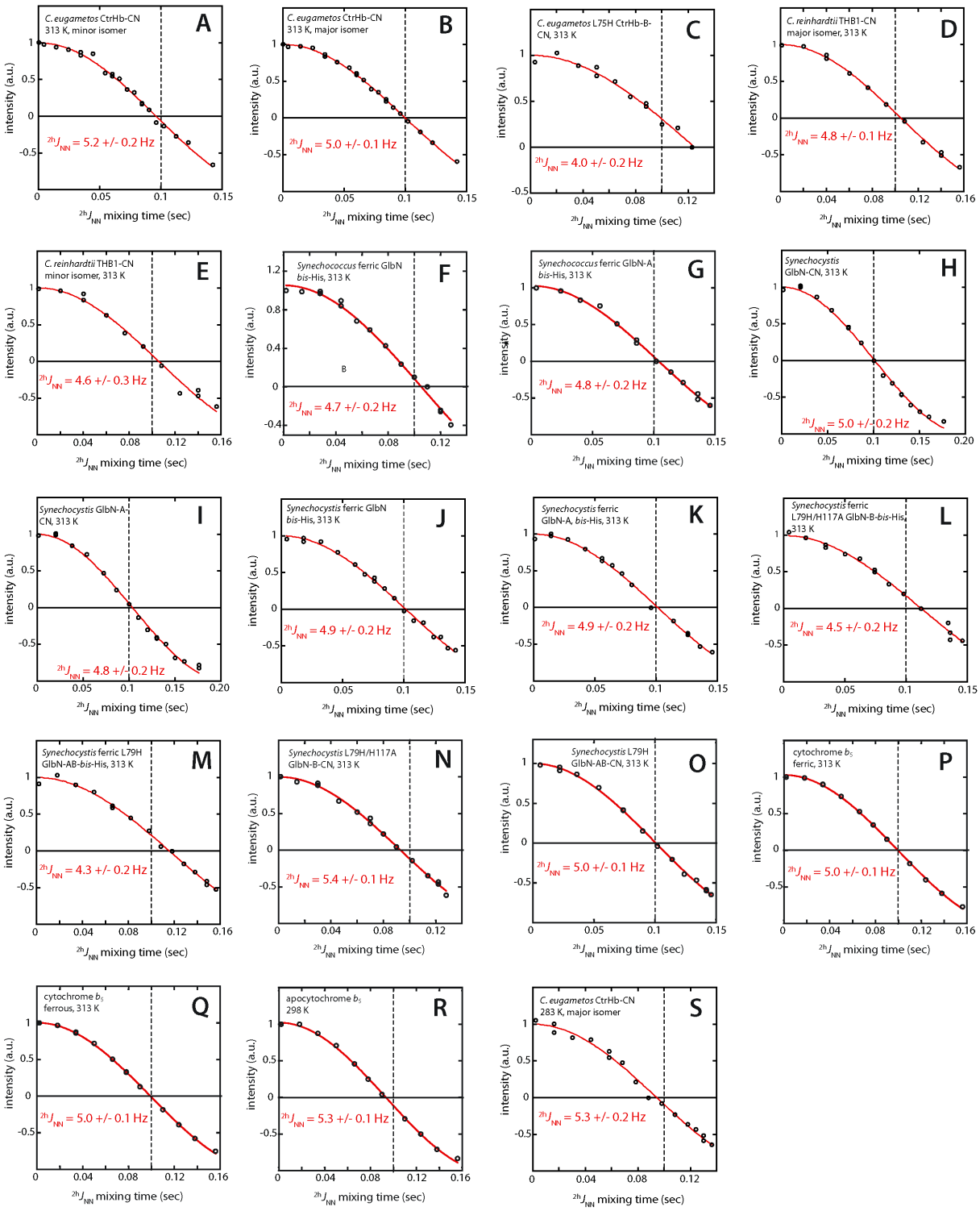

Figure S3. $^{2h}J_{NN}$ modulation curves and fits obtained from constant-time spin-echo difference HSQC experiments.

Data points from an individual experiment series were collected in random order, normalized according to the maximum signal intensity, and plotted (black circles). For each data set, the

non-linear least squares best fit curve is shown with a red line. In most cases, the fitting errors were < 1%, however the errors reported for each $^{2h}J_{NN}$ value combines an estimate for the uncertainty in peak intensity (~5 %) as obtained from duplicate data points collected at the beginning and end of each series. The black horizontal line marks zero intensity, and the vertical dashed line corresponds to the null time $^{2h}J_{NN}$ = 5 Hz (1/2J = 100 ms).

A. *C. eugametos* cyanomet CtrHb, Ser76 N-H···Nδ1 His79 (minor isomer, 313 K)
B. *C. eugametos* cyanomet CtrHb, Ser76 N-H···Nδ1 His79 (major isomer, 313 K)
C. *C. eugametos* Leu75His cyanomet CtrHb-B, Ser76 N-H···Nδ1 His79 (engineered His75-heme cross-link, 313 K)
D. *C. reinhardtii* cyanomet THB1, Asn87 N-H···Nδ1 His90 (major isomer, 313 K)
E. *C. reinhardtii* cyanomet THB1, Asn87 N-H···Nδ1 His90 (minor isomer, 313 K)
F. *Synechococcus bis*-histidine ferric GlbN, Thr80 N-H···Nδ1 His83 (non cross-linked, 313 K)
G. *Synechococcus bis*-histidine ferric GlbN-A, Thr80 N-H···Nδ1 His83 (native His117-heme cross-link, 313 K)
H. *Synechocystis* cyanomet GlbN, Asn80 N-H···Nδ1 His83 (non cross-linked, 313 K)
I. *Synechocystis* cyanomet GlbN-A, Asn80 N-H···Nδ1 His83 (native His117-heme cross-link, 313 K)
J. *Synechocystis bis*-histidine ferric GlbN, Asn80 N-H···Nδ1 His83 (non cross-linked, 313 K)
K. *Synechocystis bis*-histidine ferric GlbN-A, Asn80 N-H···Nδ1 His83 (native His117-heme cross-link, 313 K)
L. *Synechocystis* Leu79His/His117Ala *bis*-histidine ferric GlbN-B, Asn80 N-H···Nδ1 His83 (engineered His79-heme cross-link, 313 K)
M. *Synechocystis* Leu79His *bis*-histidine ferric GlbN-AB, Asn80 N-H···Nδ1 His83 (His117-heme, His79-heme double cross-link, 313 K)
N. *Synechocystis* Leu79His/His117Ala cyanomet GlbN-B, Asn80 N-H···Nδ1 His83 (engineered His79-heme cross-link, 313 K)
O. *Synechocystis* Leu79His cyanomet GlbN-AB, Asn80 N-H···Nδ1 His83 (His117-heme, His79-heme double cross-link, 313 K)
P. Rat microsomal cytochrome b_5, Asp82 N-H···Nδ1 His80 (ferric *bis*-histidine, 313 K)
Q. Rat microsomal cytochrome b_5, Asp82 N-H···Nδ1 His80 (ferrous *bis*-histidine, 313 K)
R. Rat microsomal apocytochrome b_5, Asp82 N-H···Nδ1 His80 (apoprotein, 298 K)
S. *C. eugametos* cyanomet CtrHb, Ser76 N-H ···Nδ1 His79 (major isomer, 283 K)

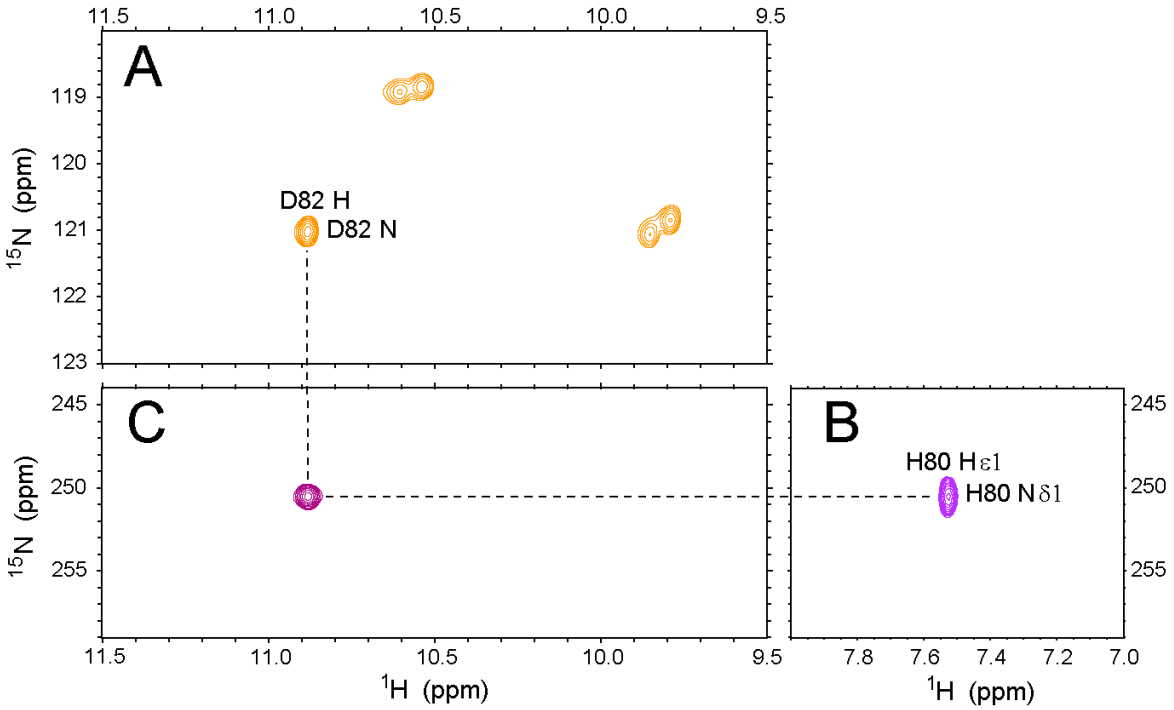

Figure S4. HSQC, long-range HMQC, and HNN-COSY correlation spectra used for Asp82 N-H···Nδ1 His80 H-bond assignment in ferric *bis*-histidine cytochrome b_5.

A. ^{1}H-^{15}N HSQC (downfield ^{1}H region) showing the Asp82 NH resonance. The major and minor isomer (6:4) amide signals are indistinguishable.

B. Histidine selective ^{1}H-^{15}N LR-HMQC (^{15}N downfield region) showing the Nδ1-Hε1 correlation expected from a neutral His80 Nε2-H tautomer.

C. HNN-COSY spectrum. A single cross-peak is observed, which correlates the Asp82 amide ^{1}H with His80 ^{15}Nδ1 via the two-bond HBC, $^{2h}J_{NN}$.

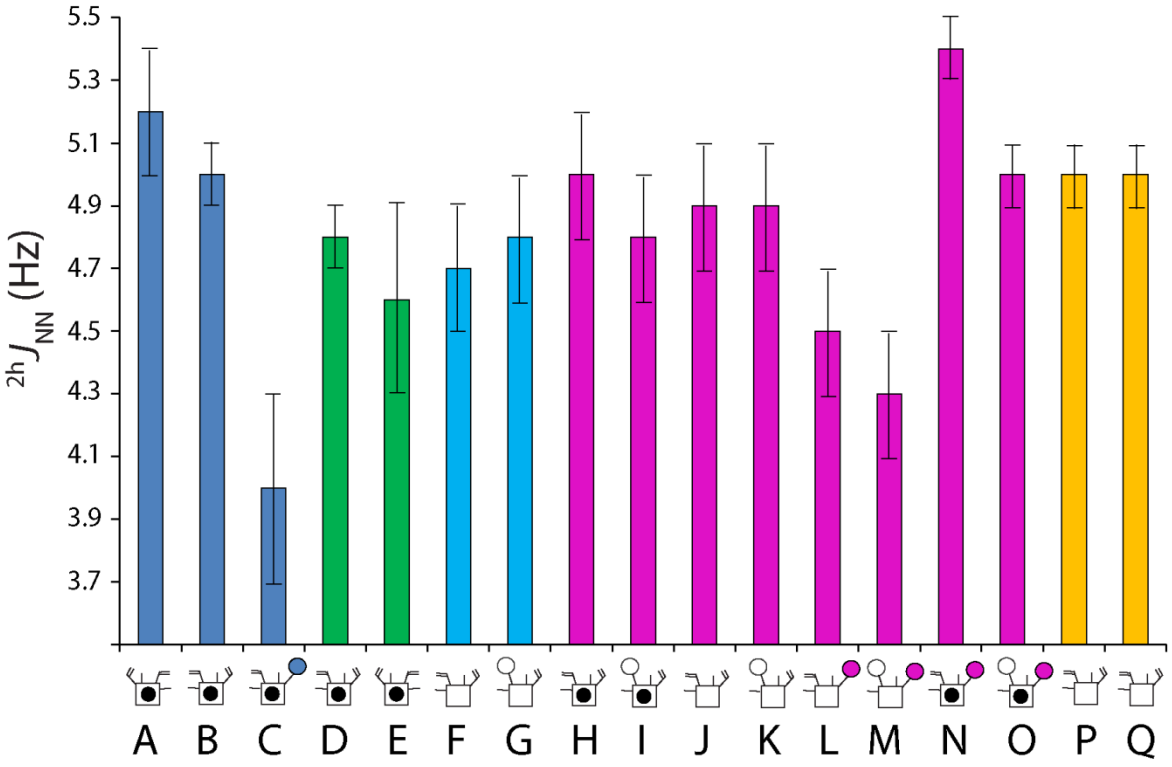

Figure S5. Bar graph summary of $^{2h}J_{NN}$ magnitudes and estimated uncertainties determined in the present study (313 K). The x-axis labeling is as defined in Table 1 and Figure 5 (see also Figure S3). Bars are colored according to host organism (blue, *Chlamydomonas eugametos* CtrHb; green, *Chlamydomonas reinhardtii* THB1; cyan, *Synechococcus* GlbN; magenta, *Synechocystis* GlbN; orange, rat microsomal cytochrome b_5). Heme pictographs are included to assist the reader in data interpretation: open squares correspond to *bis*-histidine complexes, squares containing black-filled circles correspond to cyanide-ligated proteins; open white circles depict the native crosslink, blue and magenta filled circles depict the engineered cross-link in CtrHb and GlbN, respectively.

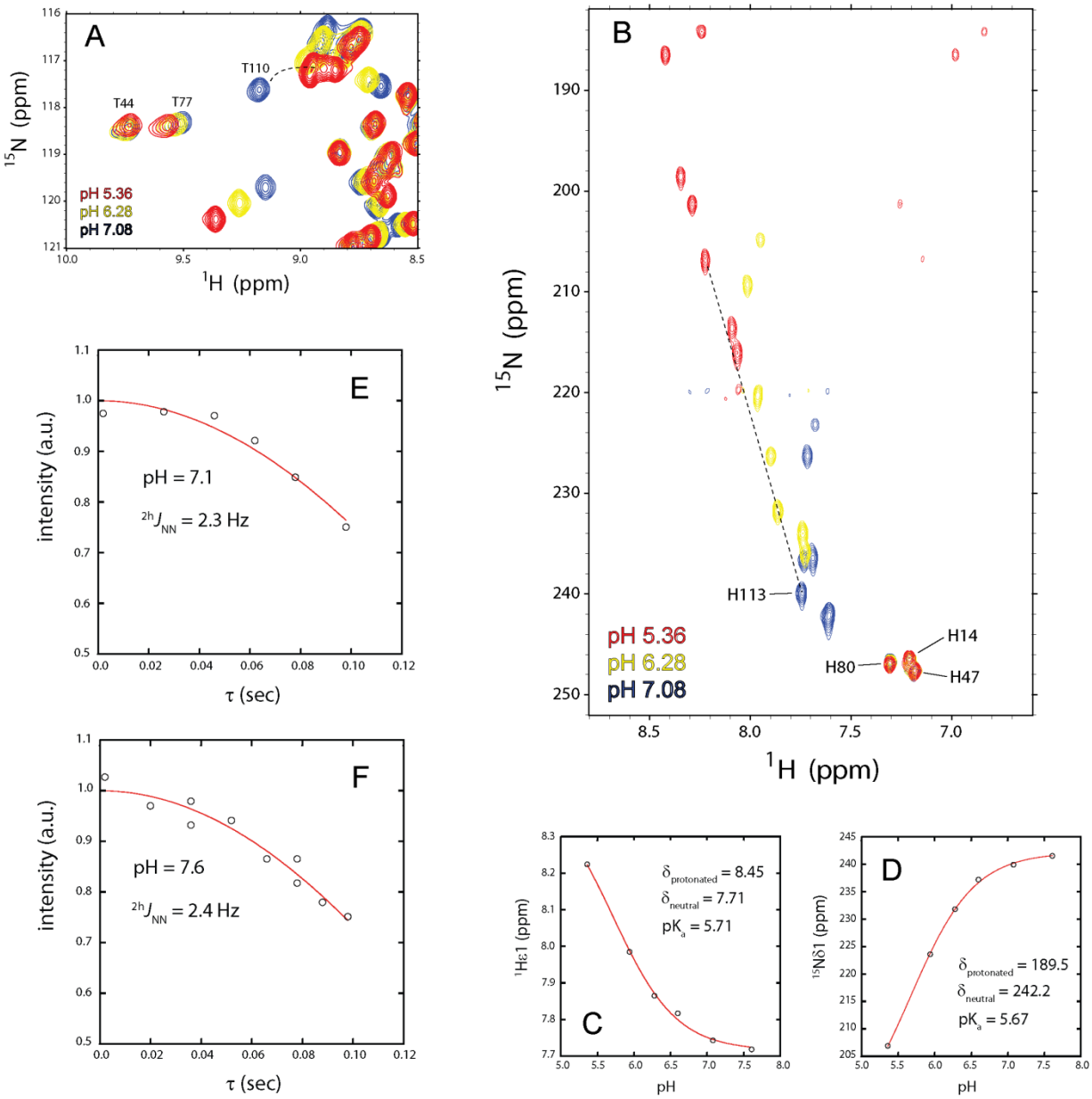

Figure S6. pH dependence of chemical shifts and $^{2h}J_{NN}$ couplings observed for the solvent-exposed T110-H113 H-bond within the four-repeat consensus AR protein, NRRC (298 K).

A. ^{15}N-^{1}H HSQC spectral overlay (downfield ^{1}H region) showing the chemical shift sensitivity of the T110 amide ^{1}H as the pH is lowered from 7.1 to 5.4. In contrast, T44 and T77 NH signals do not register major perturbation.

B Overlay of histidine-selective ^{15}N-^{1}H LR-HMQC spectra (downfield region). The ^{15}Nδ1 signals of the buried histidines (H14, H47, and H80) are insensitive to pH changes over the range 7.1–5.4 in an indication of low imidazole pK_a. Conversely, the H113 ^{15}Nδ1 undergoes a ~35 ppm upfield shift indicative of partial protonation.

C, D ^{1}Hε1 and ^{15}Nδ1 chemical shifts were fit individually using the Henderson-Hasselbalch equation

$$\delta_i = \delta_{i,\text{neutral}} + (\delta_{i,\text{protonated}} - \delta_{i,\text{neutral}}) \frac{10^{n_H(pK_a - pH)}}{1 + 10^{n_H(pK_a - pH)}}$$

with n_H set to 1. The estimated apparent pK_a of H113 is ~5.7, a value in good agreement with that obtained for the solvent exposed cap in a consensus three-repeat AR protein (pK_a = 5.7).[7]

E, F $^{2h}J_{NN}$-modulation curves measured for the T110–H113 N-H···Nδ1 H-bond as a function of pH. The HBC undergoes a nominal increase from ~2.1 Hz (pH 6.6) to ~2.4 Hz (pH 7.6).[7] This is consistent with expectations derived from the modeling illustrated in Figures 6 and 7. At pH values below 6.6, the T110 N-H signal is overlapped (see Figure S6A), and $^{2h}J_{NN}$ could not be reliably measured.

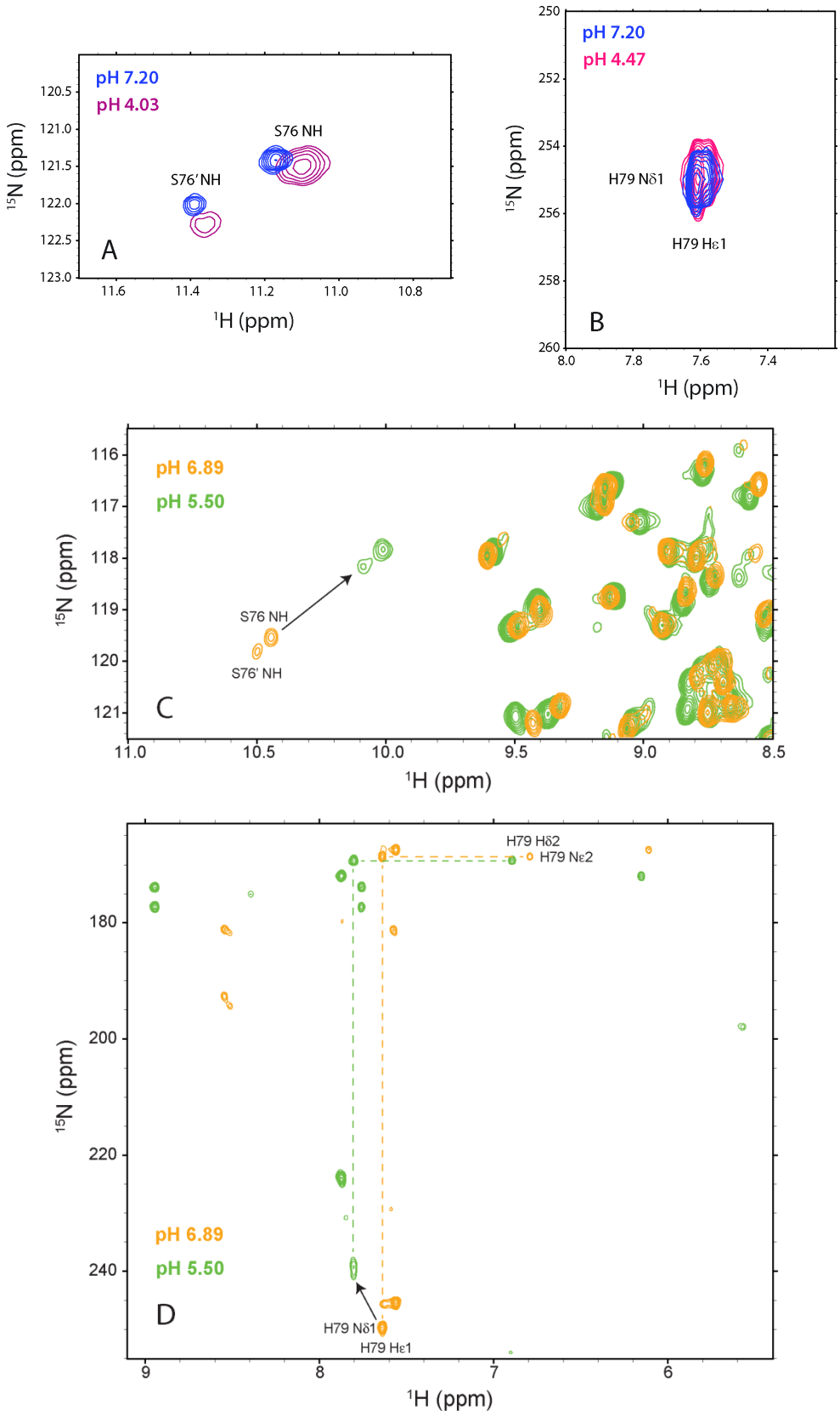

Figure S7. Estimate of His79 pK_a in WT CtrHb-CN and L75H CtrHb-B-CN (313 K).

A, B. ^{1}H-^{15}N HSQC (A) and ^{1}H-^{15}N LR-HMQC (B) spectra of WT CtrHb-CN showing the insensitivity of S76 amide N-H and H79 Nδ1 nuclei to sample pH. The absence of chemical shift changes over a pH range of 7.2–4.0 indicates a His79 $pK_a < 3.5$.

C, D. ^{1}H-^{15}N HSQC (C) and ^{1}H-^{15}N LR-HMQC (D) spectra of L75H CtrHb-B-CN showing the upfield shifts of S76 amide N-H and H79 Nδ1 nuclei as the sample pH is lowered from 6.9 to 5.5. Simulation of the chemical shift changes suggests a His79 $pK_a \sim 5.0$.

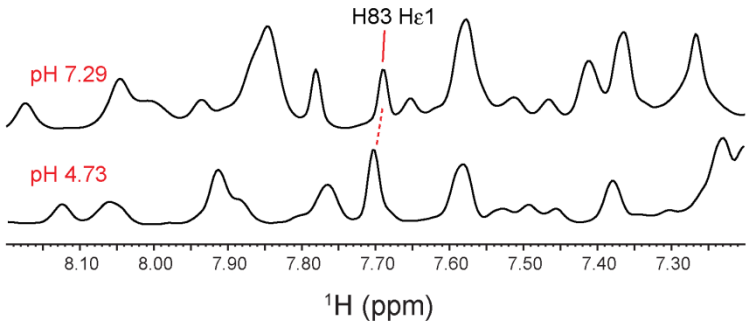

Figure S8. Estimate of His83 pK_a in WT *Synechococcus bis*-His GlbN (298 K). ^{1}H 1-D data collected as a function of pH^{*} (in D_2O) register a negligible Hε1 shift compared to a full titration (~1 ppm change) and demonstrate a $pK_a < 4.0$.

References

(1) Vu, B. C., Vuletich, D. A., Kuriakose, S. A., Falzone, C. J., and Lecomte, J. T. J. (2004) Characterization of the heme-histidine cross-link in cyanobacterial hemoglobins from *Synechocystis* sp. PCC 6803 and *Synechococcus* sp. PCC 7002. *J. Biol. Inorg. Chem. 9*, 183-194.

(2) Moore, C. D., al-Misky, O. N., and Lecomte, J. T. J. (1991) Similarities in structure between holocytochrome b_5 and apocytochrome b_5: NMR studies of the histidine residues. *Biochemistry 30*, 8357-8365.

(3) Lecomte, J. T. J., and Moore, C. D. (1991) Helix formation in apocytochrome b_5: the role of a neutral histidine at the N-cap position. *J. Am. Chem. Soc. 113*, 9663-9665.

(4) Reynolds, W. F., Peat, I. R., Freedman, M. H., and Lyerla, J. R., Jr. (1973) Determination of the tautomeric form of the imidazole ring of L-histidine in basic solution by carbon-13 magnetic resonance spectroscopy. *J. Am. Chem. Soc. 95*, 328-331.

(5) Tanokura, M. (1983) ^{1}H-NMR study on the tautomerism of the imidazole ring of histidine residues. I. Microscopic pK values and molar ratios of tautomers in histidine-containing peptides. *Biochim. Biophys. Acta 742*, 576-585.

(6) Boschcov, P., Seidel, W., Muradian, J., Tominaga, M., Paiva, A. C. M., and Juliano, L. (1983) Ionization constants and thermodynamic parameters of histidine and derivatives. *Bioorg. Chem. 12*, 34-44.

(7) Preimesberger, M. R., Majumdar, A., Aksel, T., Sforza, K., Lectka, T., Barrick, D., and Lecomte, J. T. (2015) Direct NMR detection of bifurcated hydrogen bonding in the α-helix N-caps of ankyrin repeat proteins. *J. Am. Chem. Soc. 137*, 1008-1011.