Highly Conserved Histidine Plays a Dual Catalytic Role in Protein Splicing: a pK_a Shift Mechanism

Zhenming Du[†], Philip T. Shemella[‡], Yangzhong Liu[¥], Scott A. McCallum[†], Brian Pereira[£§], Saroj K. Nayak[‡], Georges Belfort[§], Marlene Belfort[£] and Chunyu Wang[†]*

[†]*Biology Department, Center for Biotechnology and Interdisciplinary Studies, Rensselaer Polytechnic Institute, Troy, NY 12180;*

[‡]*Department of Physics, Applied Physics, and Astronomy, Rensselaer Polytechnic Institute, Troy, NY 12180;*

[¥]*Department of Chemistry, University of Science and Technology of China, Hefei, Anhui, P.R.China 230026;*

[£]*Wadsworth Center, New York State Department of Health, Center for Medical Sciences, 150 New Scotland Avenue, Albany, NY 12208;*

[§]*Department of Chemical and Biological Engineering, Center for Biotechnology and Interdisciplinary Studies, Rensselaer Polytechnic Institute, Troy, NY 12180*

Supporting information

Contents:

1. Table S1. Various mutants used to study the role of conserved histidines.

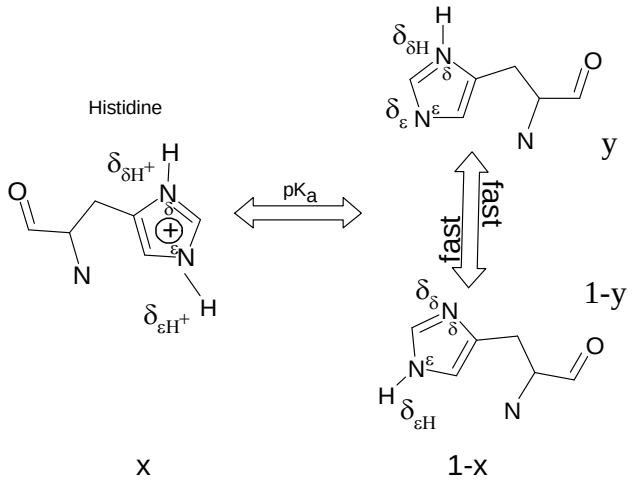
Name	N-extein	Changes relative to ΔΔIhh-SM in intein	C-extein	Usage in the study	Relevance to the steps in splicing
ΔΔIhh-SM (splicing mutant)	None	None	None	NMR pK_a determination in intein product	All four steps
ΔΔIhh-CM (cleavage mutant)		D121G			C-terminal cleavage
NI	17 AAs	C1A		NMR pK_a determination in a precursor with N-extein	N-S acyl shift
MIC	Maltose binding protein (MBP)	None	C-terminal domain of I-TEV1	*In vivo* splicing and cleavage assay with SDS-PAGE and Western blot	All four steps

2. Histidine pK_a estimation based on average chemical shift of ring nitrogens

The average chemical shift of histidine ring nitrogens is a good indication of the ionization state of the histidine ring[1-3]. When obtaining the full titration curve is not practical, the pK_a of histidines in NI was determined by the following equation:

$$pK_a = pH + \log\frac{(\tilde{A}_H - \tilde{A}_{obs})}{(\tilde{A}_{obs} - \tilde{A}_{H^+})} \qquad (1),$$

where $\tilde{A}_{obs}$ =½($\delta_{\delta(obs)}$+ $\delta_{\varepsilon(obs)}$), $\delta_{\delta(obs)}$ and $\delta_{\varepsilon(obs)}$ corresponds to the ^{15}N chemical shifts of histidine $N^{\delta1}$ and $N^{\varepsilon2}$ atoms measured by HMQC NMR spectra at individual pHs; $\tilde{A}_H$ =½(δ_δ+ $\delta_{\varepsilon H}$) =½(δ_ε + $\delta_{\delta H}$), δ_δ, δ_ε, $\delta_{\delta H}$, and $\delta_{\varepsilon H}$ being the ^{15}N chemical shifts of histidine (with either δ or ε position protonated) when the imidazole ring is neutral (Fig. below); $\tilde{A}_{H^+}$ =½($\delta_{\delta H}{}^+$+ $\delta_{\varepsilon H}{}^+$), $\delta_{\delta H}{}^+$ and $\delta_{\varepsilon H}{}^+$ is the ^{15}N chemical shifts of histidine with both δ and ε position protonated in a positively charged the imidazole ring.



Let x be the fractional population of the charged histidine ring and y be the fractional population with protonated Nδ in tautomer exchange. Based on the scheme in the above figure, we have:

$$\delta_{\delta}(obs) = \delta_{\delta H^+} x + (1-x) y \delta_{\delta H} + (1-x)(1-y)\delta_{\delta} \tag{2}$$

$$\delta_{\varepsilon}(obs) = \delta_{\varepsilon H^+} x + (1-x) y \delta_{\varepsilon} + (1-x)(1-y)\delta_{\varepsilon H} \tag{3}$$

Assuming $\delta_{\varepsilon} = \delta_{\delta}$, $\delta_{\delta H} = \delta_{\varepsilon H}$, and $\delta_{\delta H^+} = \delta_{\varepsilon H^+} = \tilde{A}_{H^+}$, (2)+(3) divided by 2 yields,

$$\begin{aligned} &\tfrac{1}{2}\left[\delta_{\delta}(obs) + \delta_{\varepsilon}(obs)\right] \\ &= \tilde{A}_{obs} = \delta_{\delta H^+} x + \tfrac{1}{2}(\delta_{\delta} + \delta_{\varepsilon H})(1-x) = \tfrac{1}{2}(\delta_{\delta} + \delta_{\varepsilon H}) - \left[\tfrac{1}{2}(\delta_{\delta} + \delta_{\varepsilon H}) - \delta_{\delta H^+}\right] x \end{aligned} \tag{4}$$

Or

$$x = \frac{\tfrac{1}{2}(\delta_{\delta} + \delta_{\varepsilon H}) - \tilde{A}_{obs}}{\tfrac{1}{2}(\delta_{\delta} + \delta_{\varepsilon H}) - \delta_{\delta H^+}} = \frac{\tilde{A}_H - \tilde{A}_{obs}}{\tilde{A}_H - \tilde{A}_{H^+}} \tag{5}$$

And

$$1 - x = \frac{\tilde{A}_{obs} - \tilde{A}_{H^+}}{\tilde{A}_H - \tilde{A}_{H^+}} \tag{6}$$

By definition,

$$pK_a = pH + \log\frac{x}{1-x} \tag{7}$$

Thus, equation (1) can be derived by substituting (5) and (6) into (7).

The canonical values of δ_{δ} or δ_{ε}, $\delta_{\delta H}$ or $\delta_{\varepsilon H}$, and $\delta_{\delta H^+}$ or $\delta_{\varepsilon H^+}$ are 249, 168, and 176 ppm respectively.[1] Incorporating those numbers can simplify equation 1 to,

$$pK_a = pH + \log\frac{(208.5 - \tilde{A}_{obs})}{(\tilde{A}_{obs} - 176)} \tag{8}$$

For His73 in the intein spliced product $\Delta\Delta_{hh}$I-SM, we have observed δ_{δ} =243.1 ppm and $\delta_{\varepsilon H}$ =165.5 ppm, resulting in $\tilde{A}$ = 204.3. So a slightly modified version of equation 7 gives

$$pK_a = pH + \log\frac{(204.3 - \tilde{A}_{obs})}{(\tilde{A}_{obs} - 176)} \qquad (8')$$

To eliminate the potentially big error in calculation when either the protonated form or the neutral form dominate, only data where pH is within one unit of pK_a, where $\frac{(208.5 - \tilde{A}_{obs})}{(\tilde{A}_{obs} - 176)}$ falls between 0.1 and 10, and 179.0 ppm < $\tilde{A}_{obs}$<205.5 ppm should be used.

In table S2, we validate the method by comparing pK_a values determined from equation (8), (8') and from curve-fitting when the full range of pH titration data are available. As seen from the table, pK_a values determined from average chemical shifts of ring nitrogens are consistently within 0.6 of the value determined from full curve-fitting, as long as only chemical shift values at pH within 1 unit of pK_a are used. Equation 8 and 8' yield very similar results, indicating that small errors in the estimation of $\tilde{A}_H$ and $\tilde{A}_{H+}$ won't significantly affect the calculated pK_a values. To ensure that $\tilde{A}_H$ and $\tilde{A}_{H+}$ do not have large variations, we performed statistical analysis based on histidine chemical shifts we obtained and those published in the literature.[4,5] The average and standard deviation of $\tilde{A}_H$ in 48 neutral histidine rings are 208.2 ± 3.4 ppm. The average and standard deviation of $\tilde{A}_{H+}$ from 18 imidazolium rings are 178 ± 2 ppm.

Table S2

$\Delta\Delta_{hh}$I-SM	pH	$\tilde{A}_{obs}$ (ppm)	pK_a (Eqn. 8)	pK_a (Eqn. 8')	pK_a (Titration)
His 17	5.19	198.5	4.84	4.60	4.5±0.3
	4.67	191.4	4.71	4.59	
	4.02	182.5	4.62	4.54	
His 30	7.28	200.9	6.77	6.42	7.1±0.2
	6.83	193.9	6.74	6.60	
	6.37	184.1	6.85	6.77	
	5.79	179.9	6.66	6.59	
His 41	7.28	198.2	6.95	6.72	7.1±0.2
	6.83	193.1	6.78	6.65	
	6.37	184.9	6.79	6.71	
	5.79	179.9	6.65	6.58	
His 429	8.2	192.5	8.19	8.06	8.9±0.2
	7.28	181.6	7.96	7.89	

3. Supporting figures

Figure S1: Protein splicing and its side reactions: N-terminal and C-terminal cleavage. N-extein, intein, and C-extein are colored in green, red, and blue.

Figure S2. Assignment of histidine resonances in ΔΔIhh-SM. A), 600 MHz (HB)CB(CGCD)HD spectrum and B), 600 MHz HMQC spectrum for $\Delta\Delta I_{hh}$-SM at pH 7.0. The assignments of the histidines in HMQC were based on chemical shift of $H^{\delta 2}$. Histidine $H^{\delta 2}$ was assigned in (HB)CB(CGCD)HD spectrum based on C^{β} derived from backbone sequential assignments.

Figure S3. Assignment of histidine resonances in ΔΔIhh-CM. A), 600 MHz (HB)CB(CGCD)HD spectrum and B), 600 MHz HMQC spectrum for $\Delta\Delta I_{hh}$-CM at pH 7.0. The assignments of the histidines in HMQC were based on chemical shift of $H^{\delta 2}$. Histidine $H^{\delta 2}$ was assigned in (HB)CB(CGCD)HD spectrum based on C^{β} derived from backbone sequential assignments.

Figure S4. Assignment of histidine resonances in NI. A), 600 MHz (HB)CB(CGCD)HD spectrum and B), 600 MHz HMQC spectrum for NI at pH 7.0. The assignments of the histidines in HMQC were based on chemical shift of $H^{\delta 2}$. Histidine $H^{\delta 2}$ was assigned in (HB)CB(CGCD)HD spectrum based on C^{β} derived from backbone sequential assignments.

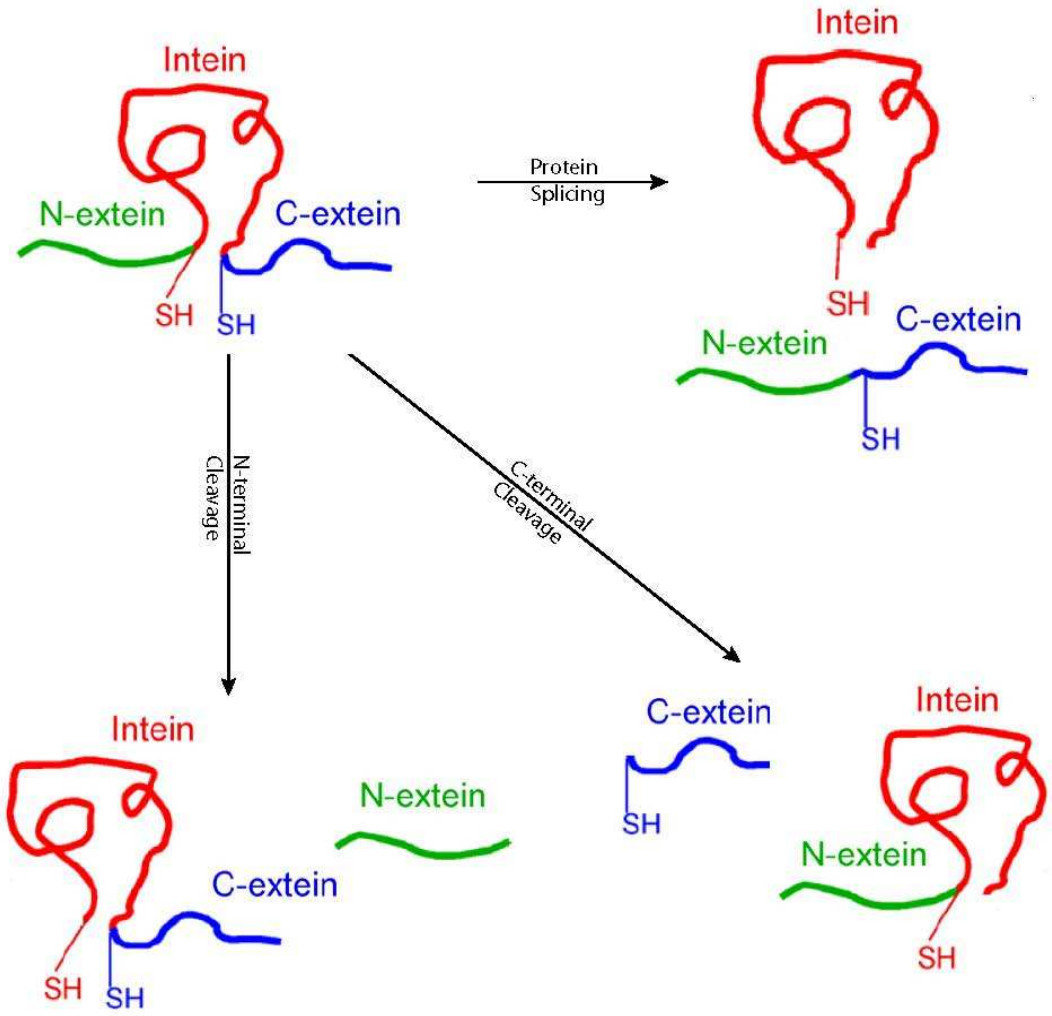

Intein
N-extein
C-extein
SH
SH
Protein
Splicing
Intein
SH
C-extein
N-extein
SH
N-terminal
Cleavage
C-terminal
Cleavage
Intein
N-extein
C-extein
SH
SH
C-extein
SH
Intein
N-extein
SH

Figure. S1

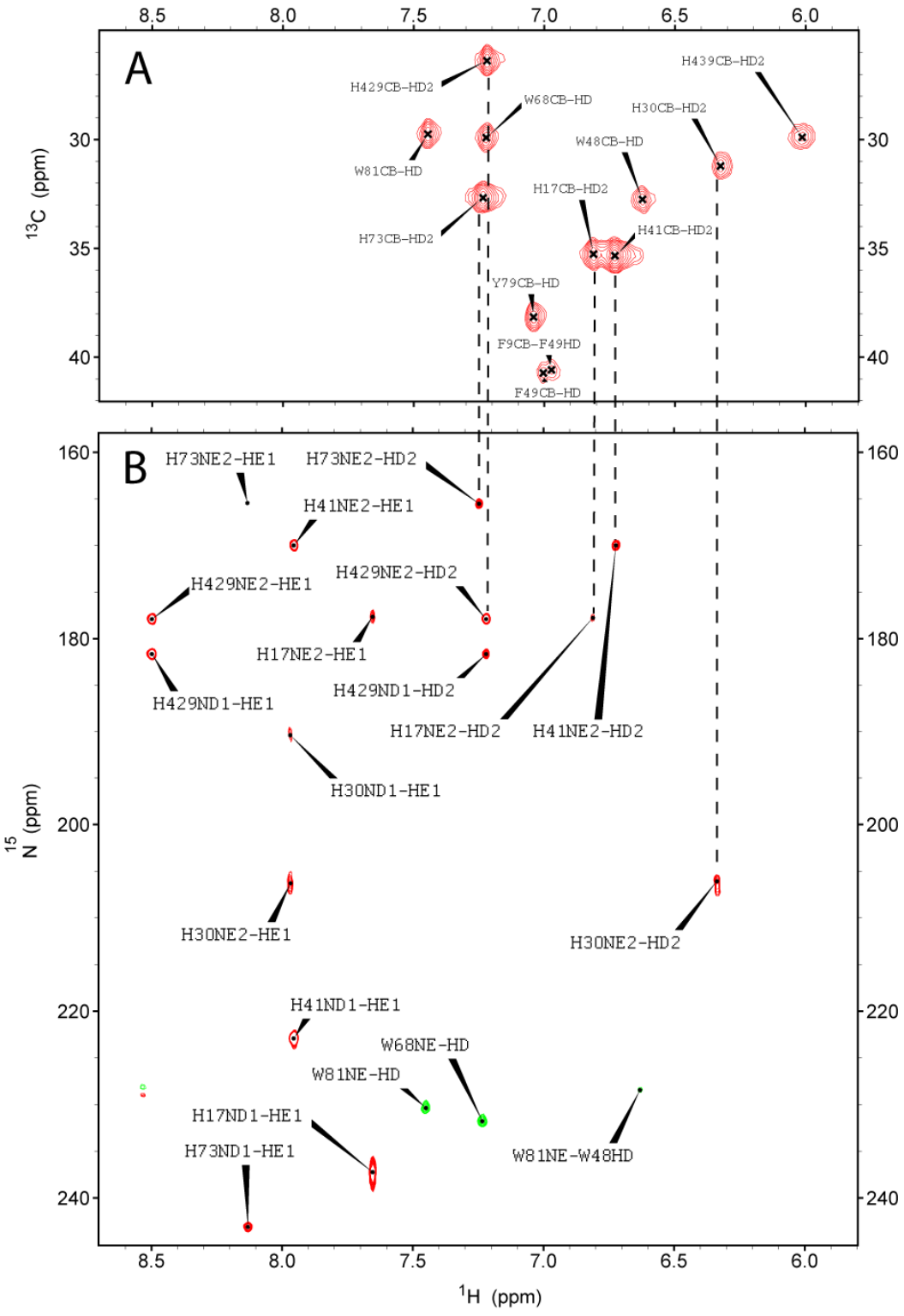

A
H429CB-HD2
W68CB-HD
W81CB-HD
H73CB-HD2
H439CB-HD2
H30CB-HD2
W48CB-HD
H17CB-HD2
H41CB-HD2
Y79CB-HD
F9CB-F49HD
F49CB-HD
B
H73NE2-HE1
H73NE2-HD2
H41NE2-HE1
H429NE2-HE1
H429NE2-HD2
H17NE2-HE1
H429ND1-HD2
H429ND1-HE1
H17NE2-HD2
H41NE2-HD2
H30ND1-HE1
H30NE2-HE1
H30NE2-HD2
H41ND1-HE1
W68NE-HD
W81NE-HD
H17ND1-HE1
H73ND1-HE1
W81NE-W48HD
^{13}C (ppm)
^{15}N (ppm)
^{1}H (ppm)

Figure. S2

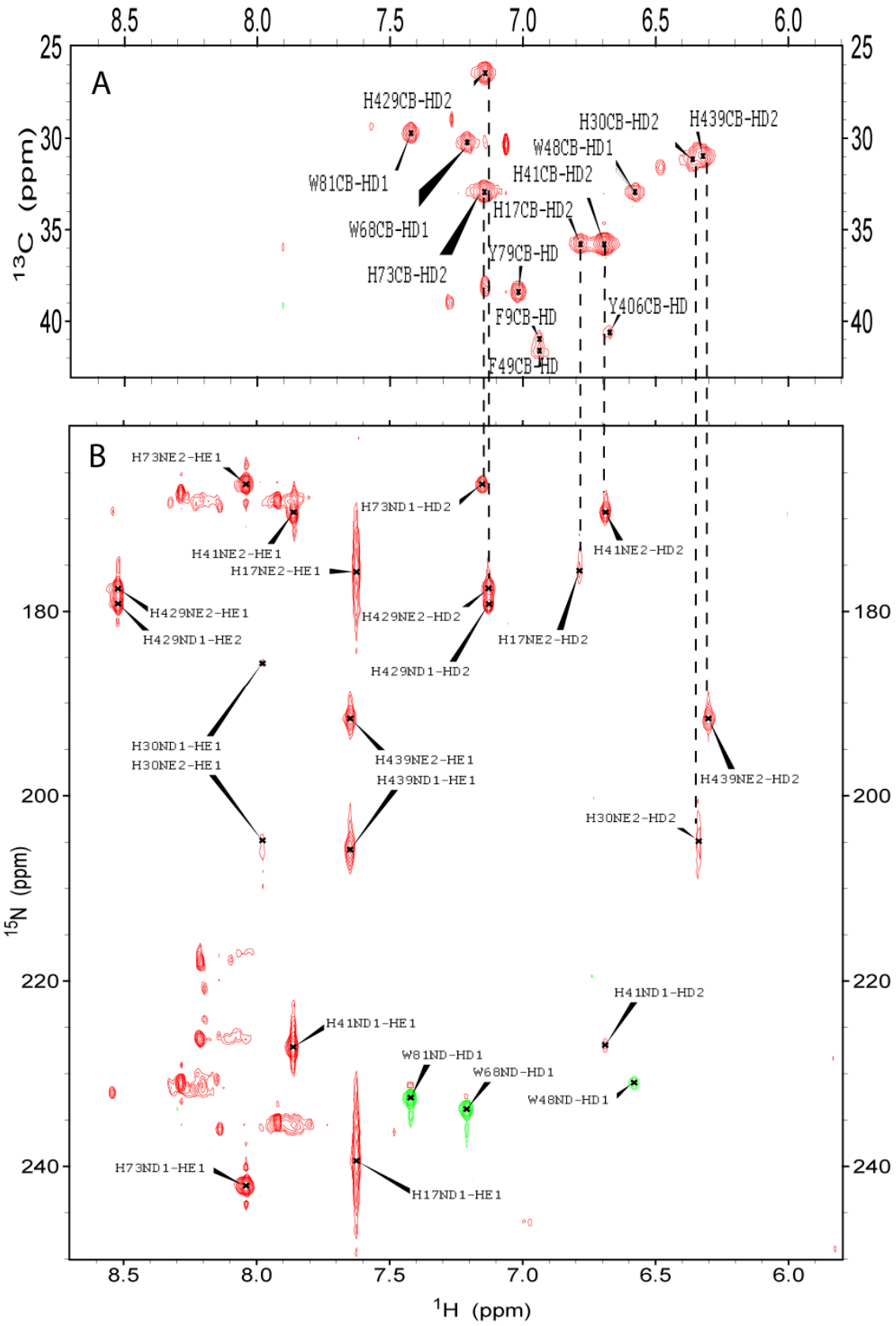

A
B
H429CB-HD2
W81CB-HD1
W68CB-HD1
H73CB-HD2
H17CB-HD2
H30CB-HD2
H439CB-HD2
W48CB-HD1
H41CB-HD2
Y79CB-HD
F9CB-HD
F49CB-HD
Y406CB-HD
H73NE2-HE1
H41NE2-HE1
H17NE2-HE1
H73ND1-HD2
H429NE2-HE1
H429ND1-HE2
H429NE2-HD2
H429ND1-HD2
H41NE2-HD2
H17NE2-HD2
H30ND1-HE1
H30NE2-HE1
H439NE2-HE1
H439ND1-HE1
H439NE2-HD2
H30NE2-HD2
H41ND1-HD2
H41ND1-HE1
W81ND-HD1
W68ND-HD1
W48ND-HD1
H73ND1-HE1
H17ND1-HE1
^{13}C (ppm)
^{15}N (ppm)
^{1}H (ppm)

Figure. S3

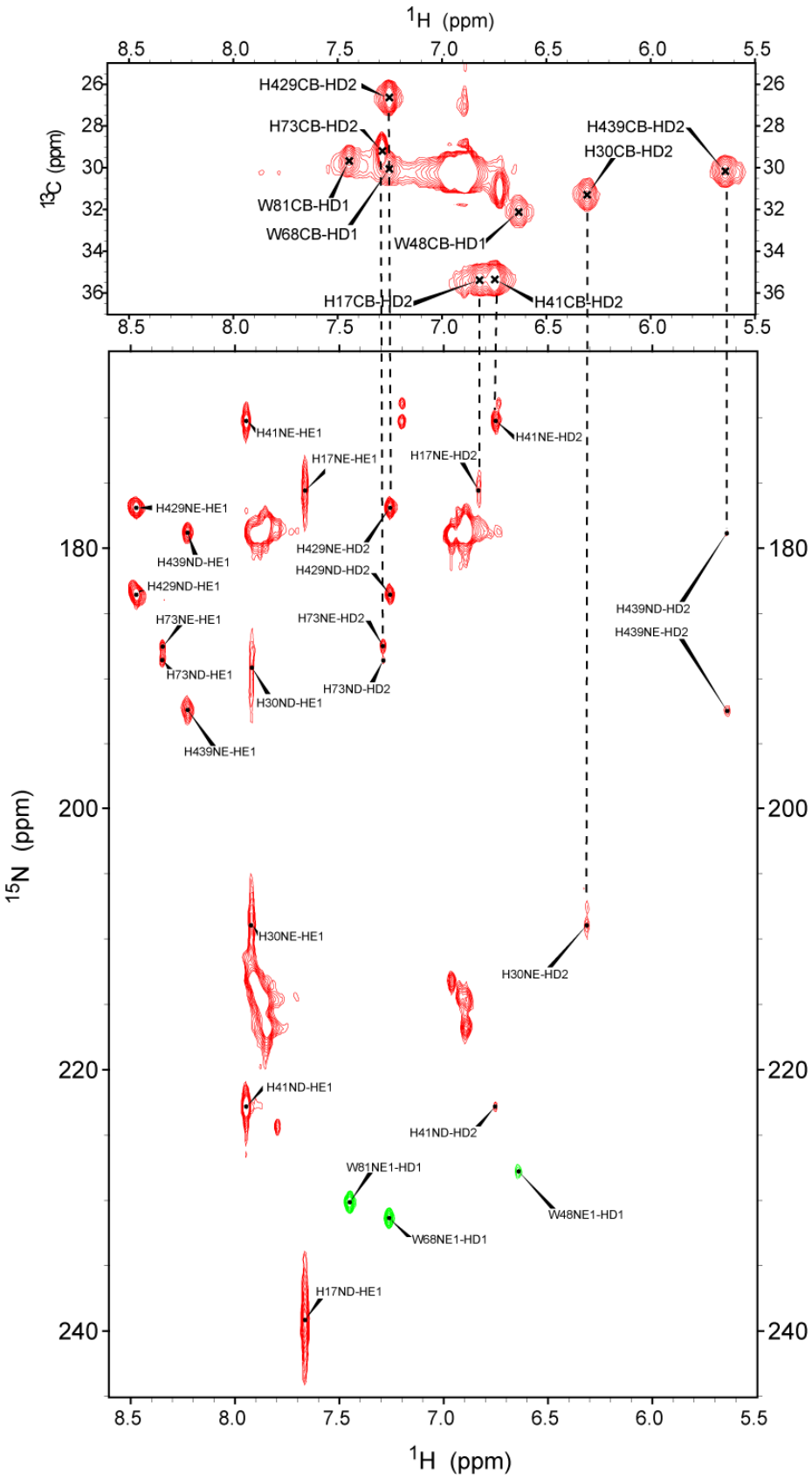

^{1}H (ppm)
^{13}C (ppm)
H429CB-HD2
H73CB-HD2
H439CB-HD2
H30CB-HD2
W81CB-HD1
W68CB-HD1
W48CB-HD1
H17CB-HD2
H41CB-HD2
H41NE-HE1
H41NE-HD2
H17NE-HE1
H17NE-HD2
H429NE-HE1
H429NE-HD2
H439ND-HE1
H429ND-HD2
H429ND-HE1
H439ND-HD2
H73NE-HE1
H73NE-HD2
H439NE-HD2
H73ND-HE1
H73ND-HD2
H30ND-HE1
H439NE-HE1
^{15}N (ppm)
H30NE-HE1
H30NE-HD2
H41ND-HE1
H41ND-HD2
W81NE1-HD1
W48NE1-HD1
W68NE1-HD1
H17ND-HE1
^{1}H (ppm)

Figure. S4

4. Complete Ref. 48

(48) Frisch, M. J. T., G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, Jr., J. A.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; and Pople, J. A. *Gaussian 03, Revision C.02, Wallingford, CT* 2004.

References:

(1) Bachovchin, W. W. *Magnetic Resonance In Chemistry* **2001**, *39*, S199-S213.

(2) Bachovchin, W. W.; Roberts, J. D. *J. Am. Chem. Soc.* **1978**, *100*, 8041-8047.

(3) Farr-Jones, S.; Wong, W. Y. L.; Gutheil, W. G.; Bachovchin, W. W. *J. Am. Chem. Soc.* **1993**, *115*, 6813-6819.

(4) Gabelli, S. B.; Azurmendi, H. F.; Bianchet, M. A.; Amzel, L. M.; Mildvan, A. S. *Biochemistry* **2006**, *45*, 11290-303.

(5) Schubert, M.; Poon, D. K.; Wicki, J.; Tarling, C. A.; Kwan, E. M.; Nielsen, J. E.; Withers, S. G.; McIntosh, L. P. *Biochemistry* **2007**, *46*, 7383-95.