

Supporting Information

Absolute Minimal Sampling in High Dimensional NMR Spectroscopy

Alexandar L. Hansen and Rafael Brüschweiler

The Ohio State University, Columbus, OH 43210

Materials and Methods

The absolute minimal sampling (AMS) method was applied to the 8.5 kDa (76 amino-acids) protein ubiquitin and the 42 kDa enzyme arginine kinase (AK) [1-2]. Both proteins were uniformly ^{13}C , ^{15}N labeled in 90%/10% $\text{H}_2\text{O}/\text{D}_2\text{O}$ solution with final concentrations of ca. 1 mM protein. Standard Bruker pulse sequences were used for the acquisition of the 3D datasets. The ubiquitin data was recorded using the `trhncoetgp3d` sequence on a Bruker Avance III HD 600 MHz spectrometer equipped with a with a 5mm TXI Cryoprobe and the following parameters:

Table S1. Parameters used for Ubiquitin HNCO data acquisition

UBQ HNCO	^1H	^{13}C	^{15}N
Carrier	4.77 ppm	175.7 ppm	117.6 ppm
Sweep width	16.0 ppm	16.0 ppm	35.2 ppm
Dwell time	104.0 us	413.4 us	466.8 us
Complex points	1024	40	32

The arginine kinase data was recorded using the `trhncogp2h3d` (HNCO) and `trhncacogp2h3d` (HNCACO) sequences, with only minor modifications. Experiments were carried out on a Bruker Avance II 800 MHz spectrometer equipped with a with a 5mm TCI Cryoprobe and the following parameters:

Table S2. Parameters used for Arginine kinase HNCO and HNCACO data acquisition

AK HNCO/HNCACO	^1H	^{13}C	^{15}N
Carrier	4.77 ppm	175.7 ppm	118.1 ppm
Sweep width	20.0 ppm	14.0 ppm	40.0 ppm
Dwell time	104.0 us	308.0 us	354.0 us
Complex points	1024	44	55

The NMR raw data were first processed using the NMRPipe suite of programs [3] and SPARKY [4-5]. Conventional protocols were applied for the processing of the full 3D datasets. For AMS analysis, only the 2nd and 3rd dimensions were Fourier transformed (¹⁵N and ¹H) leaving the first (¹³C) dimension in the time-domain. In all datasets, the first complex point was acquired with zero delay in t_1 . A peak list was generated from the first 2D plane and used as input for nonlinear least-squares fitting of the remaining 2D planes to extract peak intensities. Peak fitting was accomplished using the NMRPipe functions `nlinLS` and `seriesTab` using the same minimally sampled HNCO and HN(CA)CO datasets used for subsequent application of the AMS approach based on the arctan (Eq. (1) of the main text) and the non-linear least squares fitting (Eq. (3) of main text).

¹³C resonance frequencies were determined from the minimally sampled cosine and sine modulated datasets along the t_1 domain, $C_{\text{exp}}(t_1)$ and $S_{\text{exp}}(t_1)$, with in-house Matlab programs either using the arctan relationship of Eq. (1) or by a non-linear least squares fit based on Eq. (3) that minimizes the function:

$$\chi^2 = \sum_{n=0}^{N-1} \sum_{k=1}^M (A_k \exp(-R_2 n \Delta t_1) \cos(\omega_{C,k} n \Delta t_1) - C_{\text{exp}}(n \Delta t_1))^2 + (A_k \exp(-R_2 n \Delta t_1) \sin(\omega_{C,k} n \Delta t_1) - S_{\text{exp}}(n \Delta t_1))^2 \quad (\text{S1})$$

where M (= 1, 2, 3, or 4) is the number of resonances with angular frequencies $\omega_{C,k}$ and amplitudes A_k and R_2 was set to a uniform value for all residues in the protein, which was either zero or $R_2 = 20 \text{ s}^{-1}$ for ubiquitin and $R_2 = 75 \text{ s}^{-1}$ for arginine kinase. The $2M$ fit parameters are the M frequencies $\omega_{C,k}$ and M amplitudes A_k . The angular frequencies $\omega_{C,k}$ were allowed to vary freely within the spectral range given by the spectral width, i.e. between $-\pi/\Delta t_1$ and $+\pi/\Delta t_1$, and the amplitudes were allowed to take any positive value or zero. Initial frequencies were randomly and uniformly chosen within the above frequency range defined by the spectral width. For each ¹⁵N,¹H frequency pair, which was defined by the center of the ¹⁵N,¹H HSQC cross-peak in the $\omega_2(^{15}\text{N})$ - $\omega_3(^1\text{H})$ plane, 25 independent optimizations were performed with randomly selected initial values for the $2M$ fit parameters and the one with the lowest χ^2 was selected as the best fit.

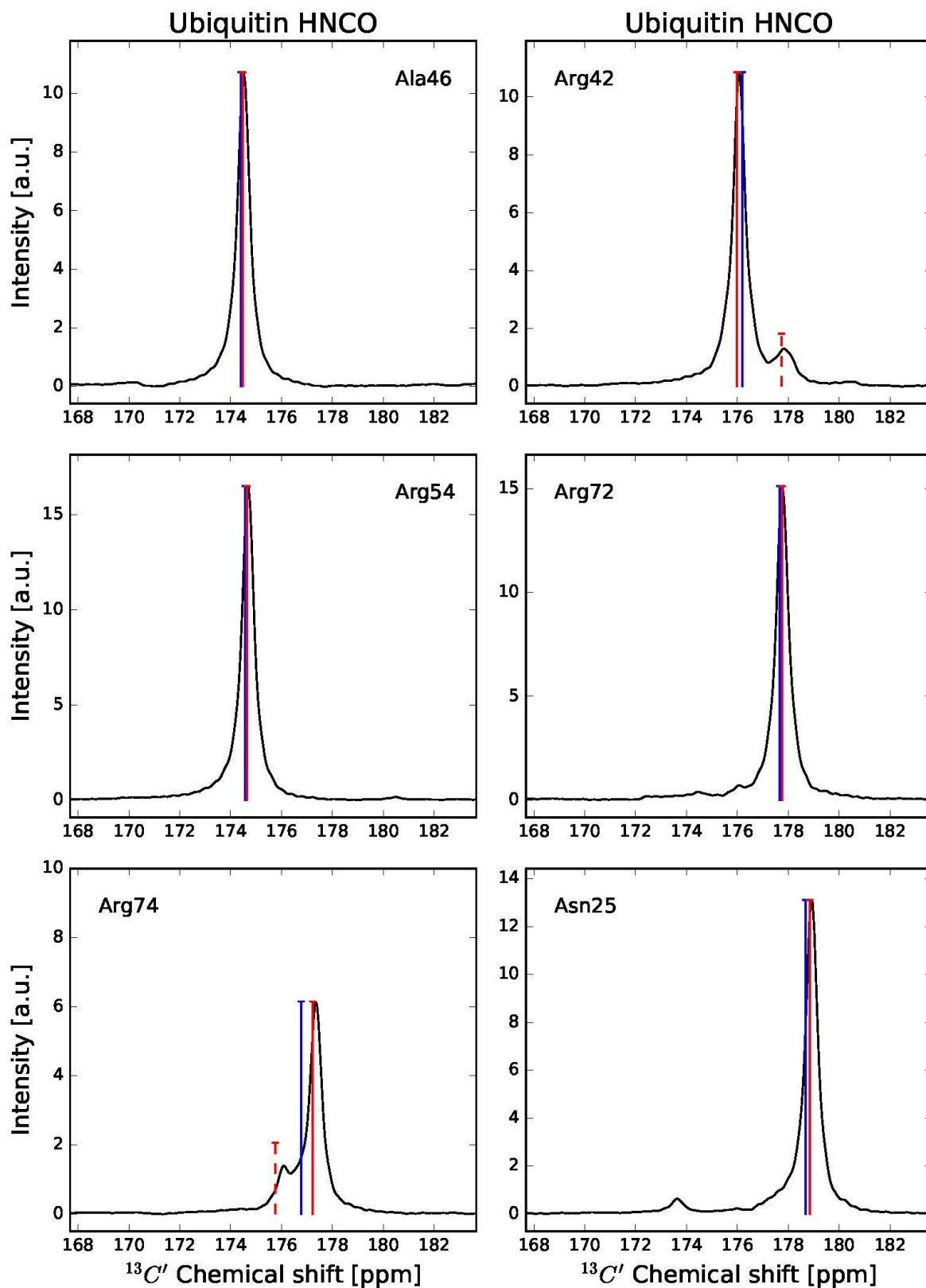
Residue-by-residue comparison of results from AMS with full-size 3D spectrum

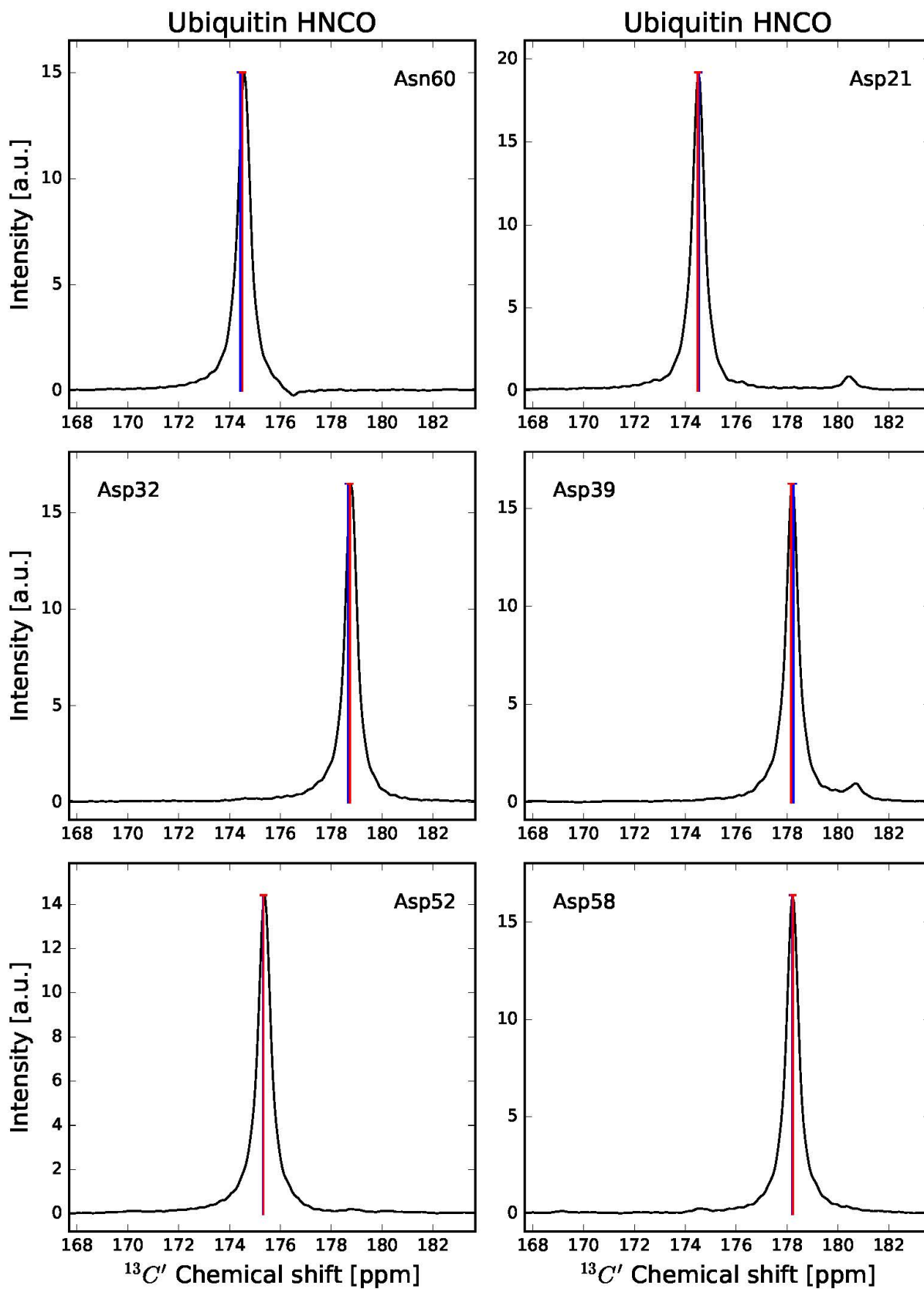
Traces along the carbonyl ^{13}C dimension of individual residues are plotted below in individual panels together with best fitting results. The blue lines refer to the ^{13}C chemical shifts determined using the inverse tangent of a single complex increment (Eq. (1)) and the red lines to the chemical shifts obtained by non-linear least squares fitting to 4 complex increments along the ^{13}C dimension using the expression of Eq. (3). The heights of the red lines reflect the fitted peak amplitudes A_k of Eq. (3). The dashed red lines correspond to minor peaks determined by AMS. Minor peaks with amplitudes smaller than 10% (ubiquitin) or 20% (arginine kinase) of the major peaks are not shown.

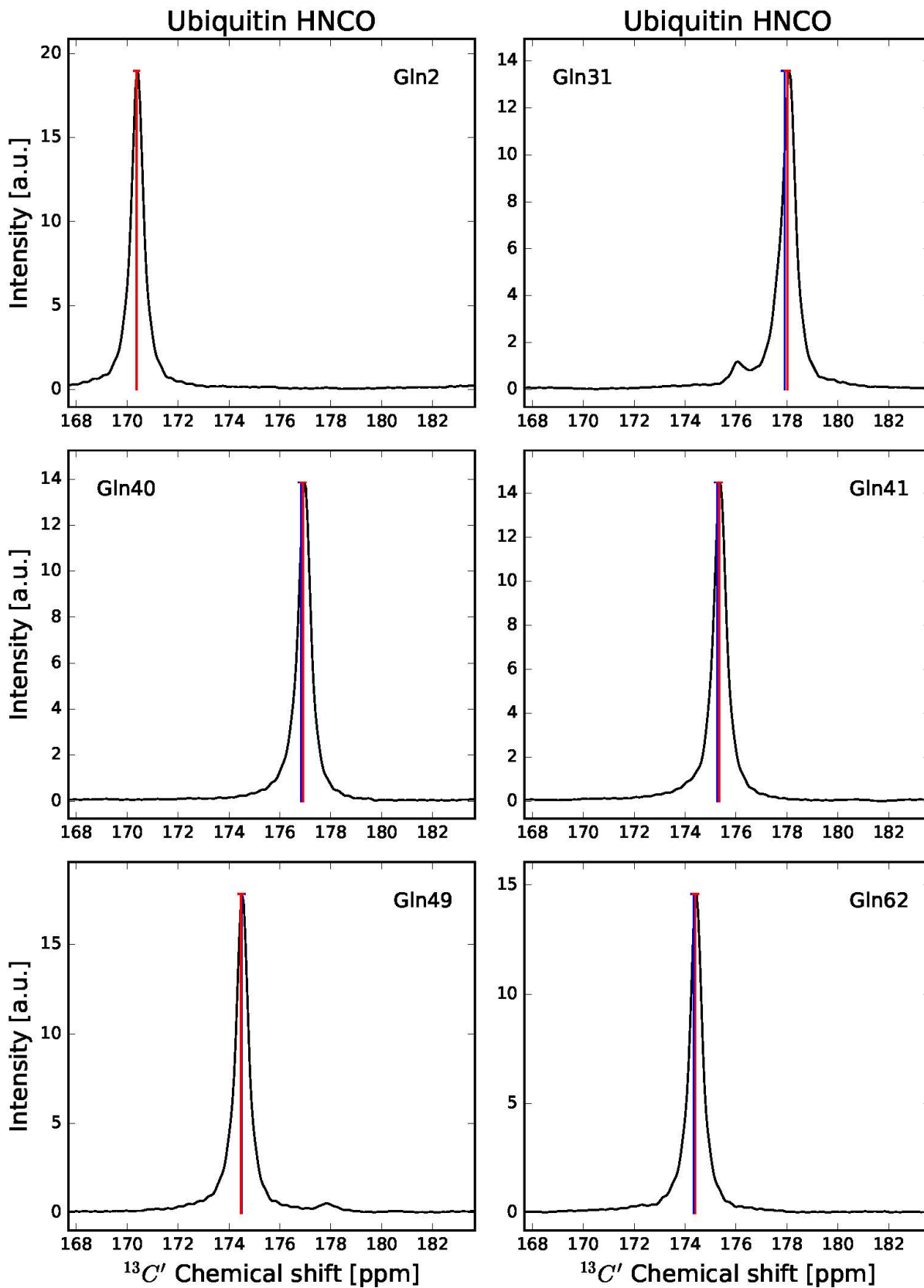
References

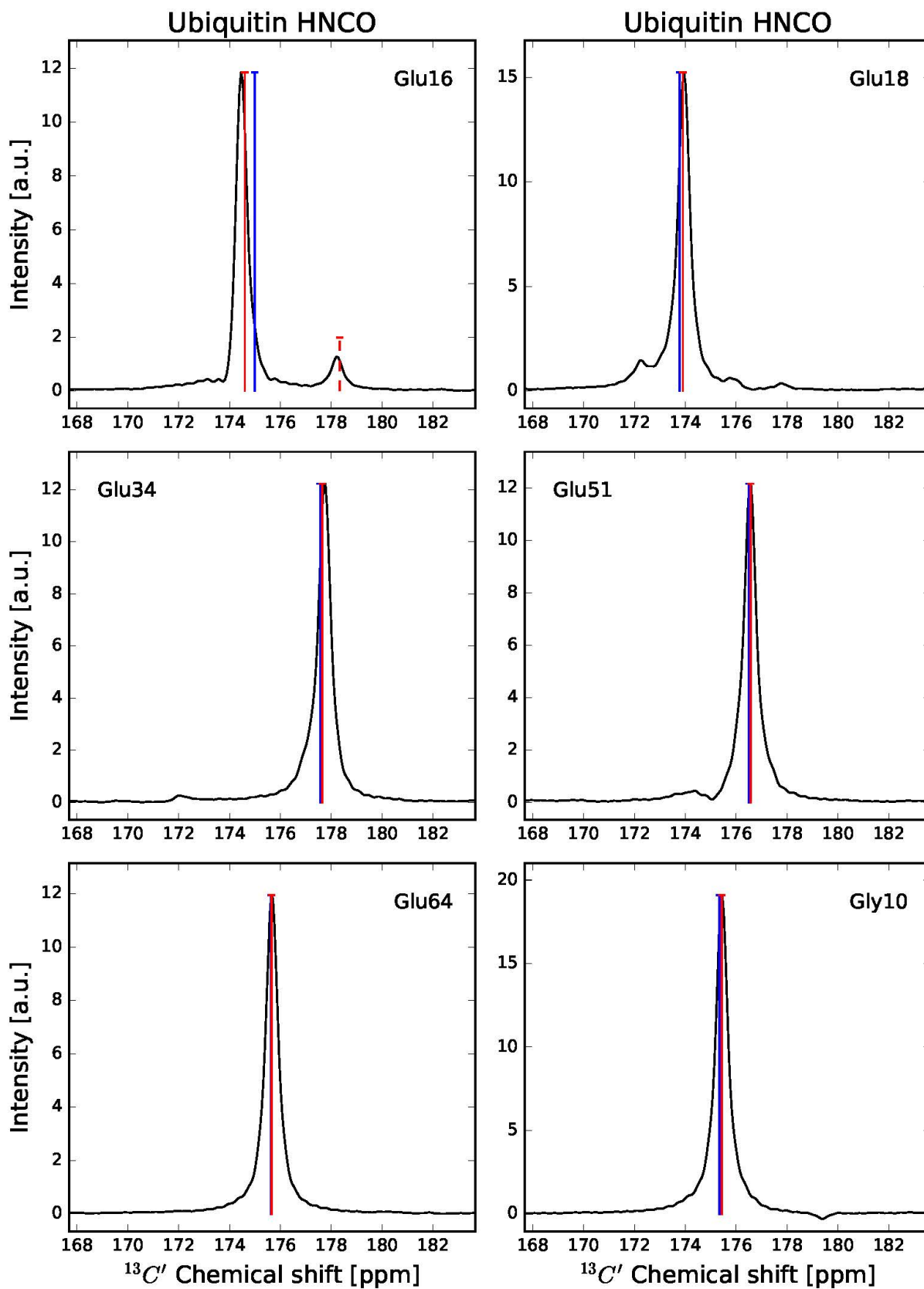
- [1] O. Davulcu, P. F. Flynn, M. S. Chapman, and J. J. Skalicky, *Structure* **2009**, *17*, 1356.
X. Niu, L. Bruschweiler-Li, O. Davulcu, J. J. Skalicky, R. Bruschweiler, and M. S. Chapman, *J. Mol. Biol.* **2011**, *405*, 479.
- [2] O. Davulcu, X. Niu, L. Bruschweiler-Li, R. Bruschweiler, J. J. Skalicky, and M. S. Chapman *Biomol. NMR Assign.* **2014**, *8*, 335.
- [3] F. Delaglio, S. Grzesiek, G. W. Vuister, G. Zhu, J. Pfeifer, and A. Bax, *J. Biomol. NMR* **1995**, *6*, 277.
- [4] T. D. Goddard and D. G. Kneller, SPARKY 3, University of California, San Francisco
- [5] W. Lee, M. Tonelli, and J. L. Markley, *Bioinformatics* **2015**, *31*, 1325.

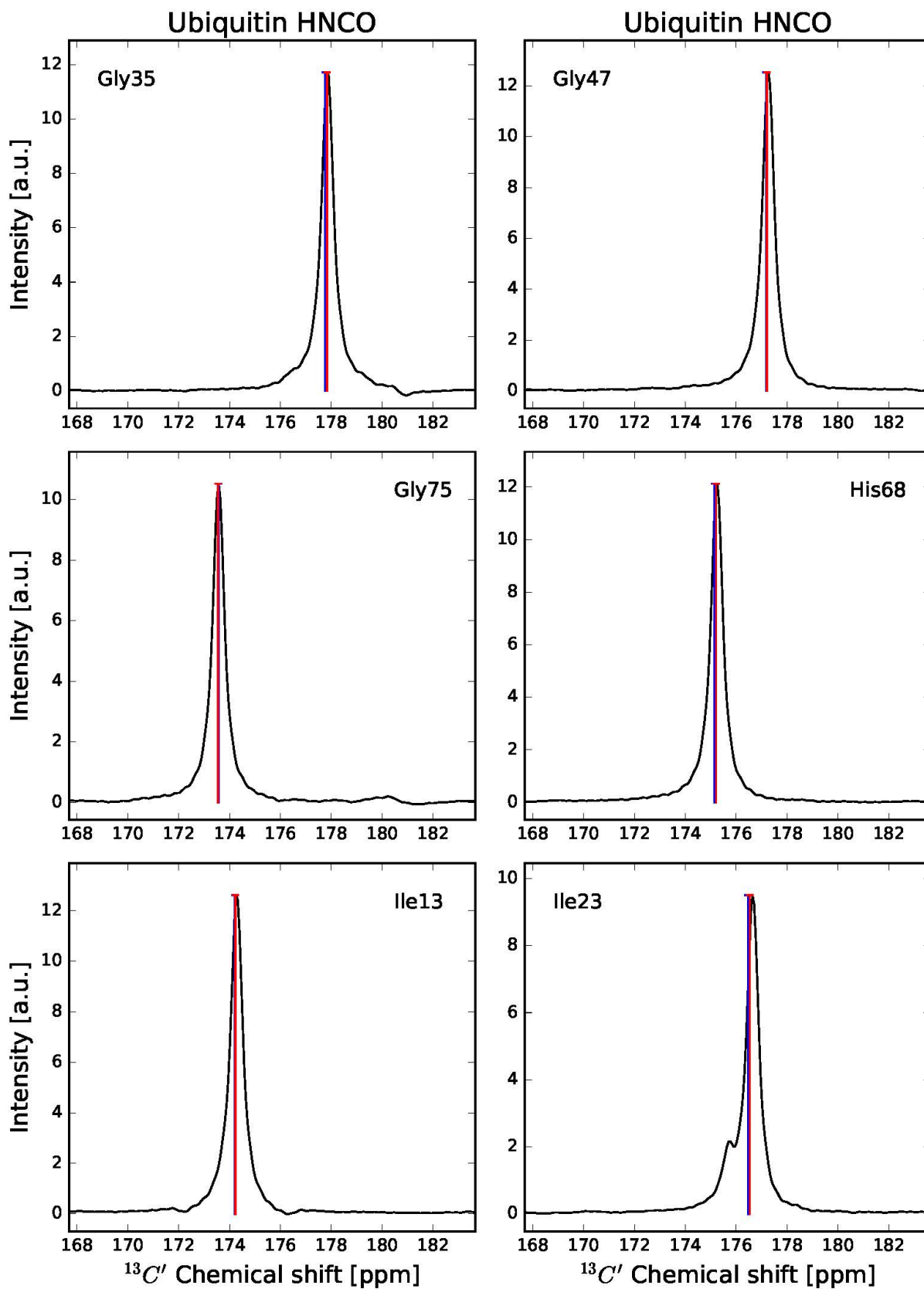
Figure S1. $^{13}\text{C}'$ traces for ubiquitin HNCO data. FT (black) was performed on 40 complex points, while AMS (red) used 4 points. In blue are the arctan results on the first complex point.

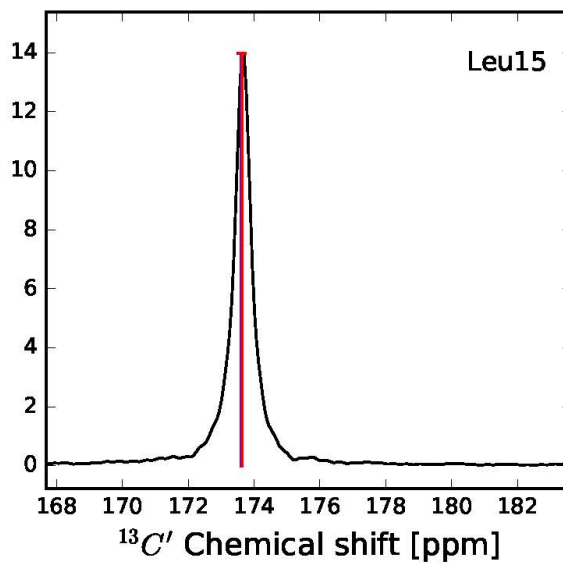
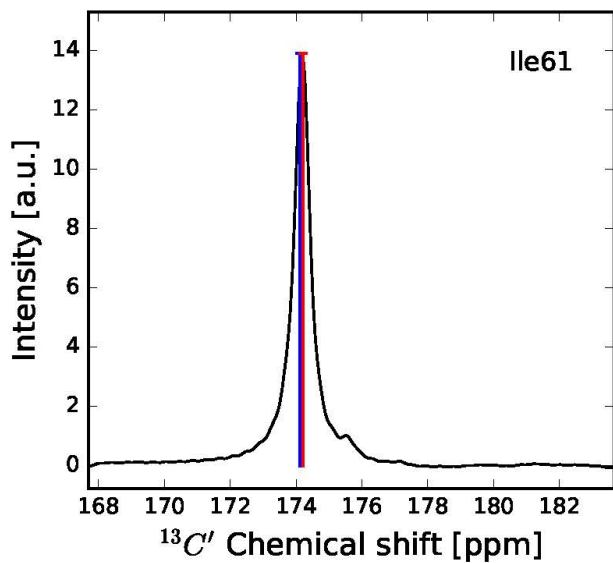
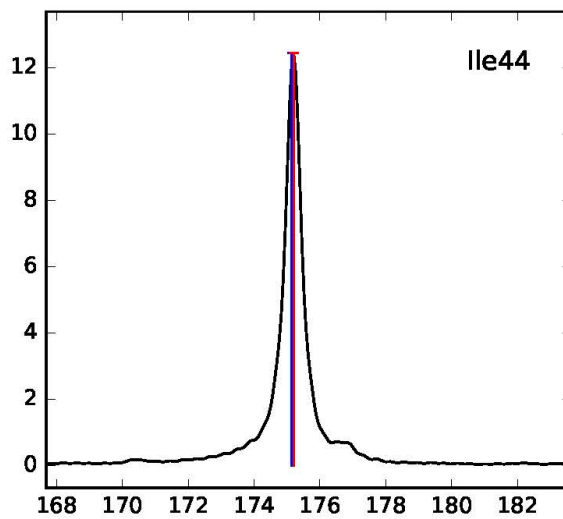
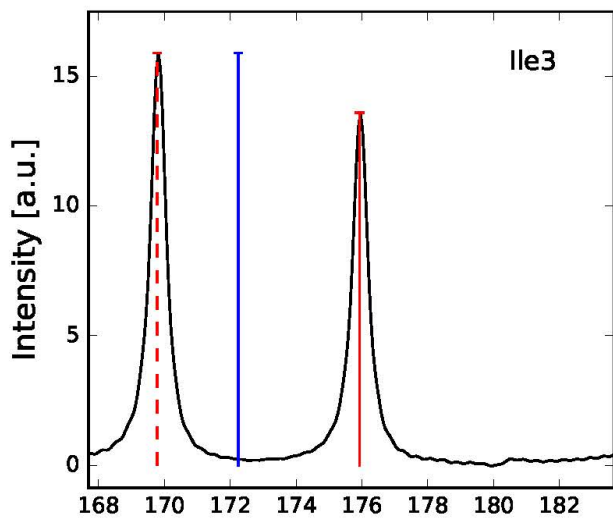
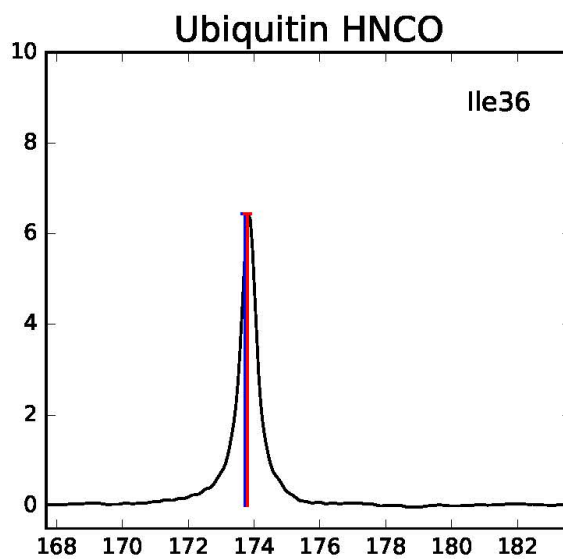
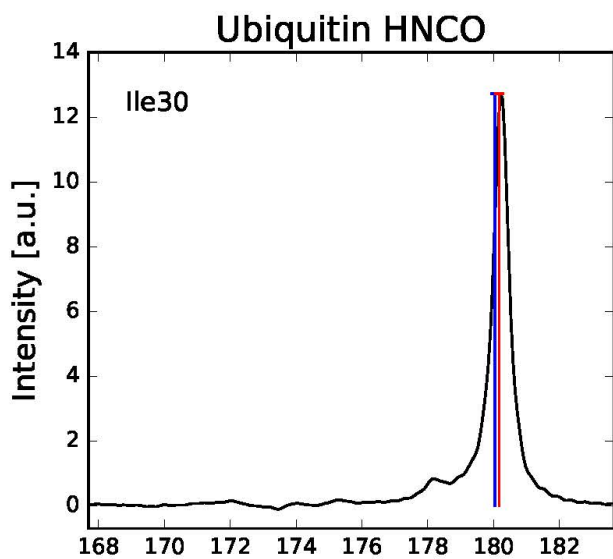


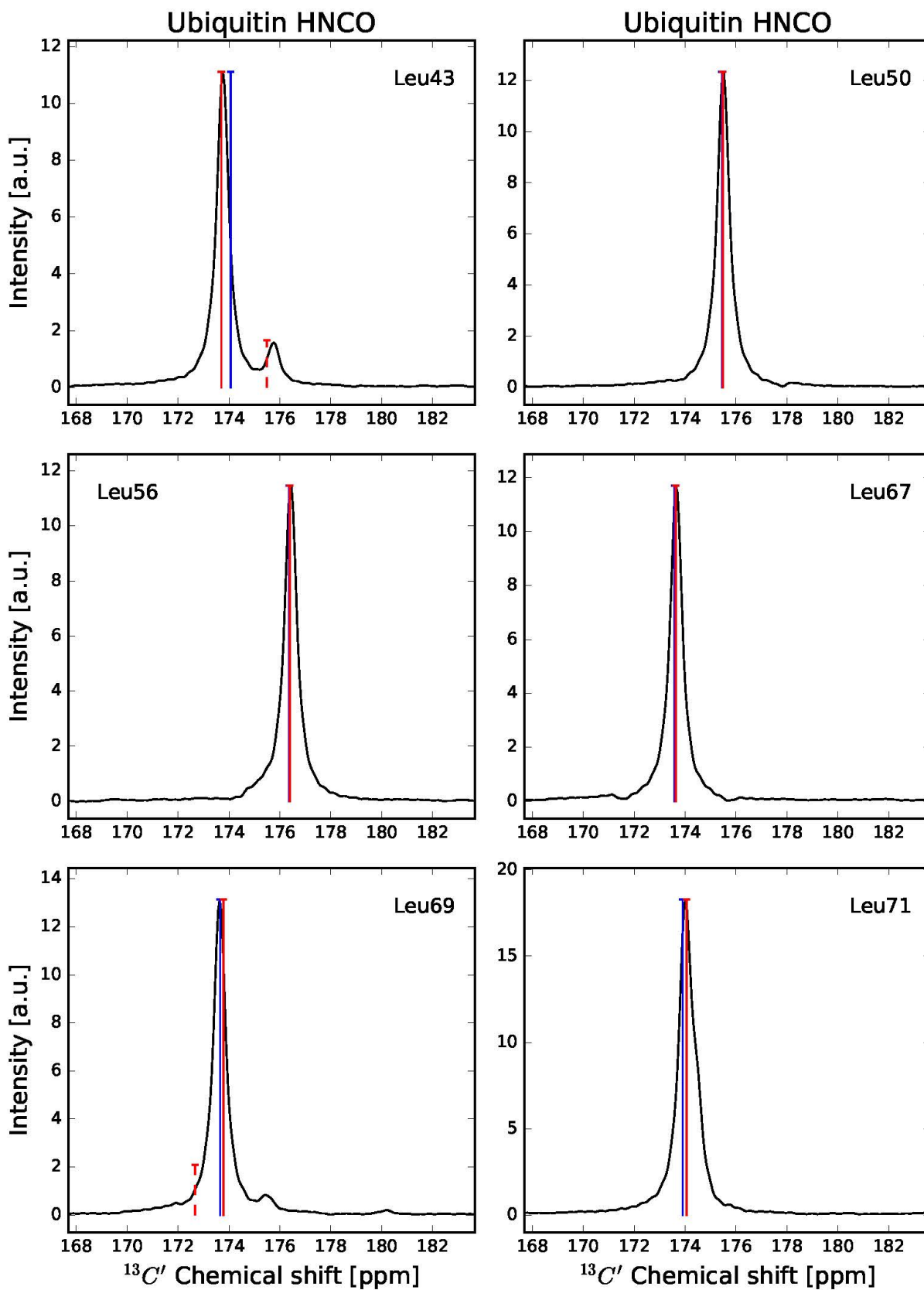


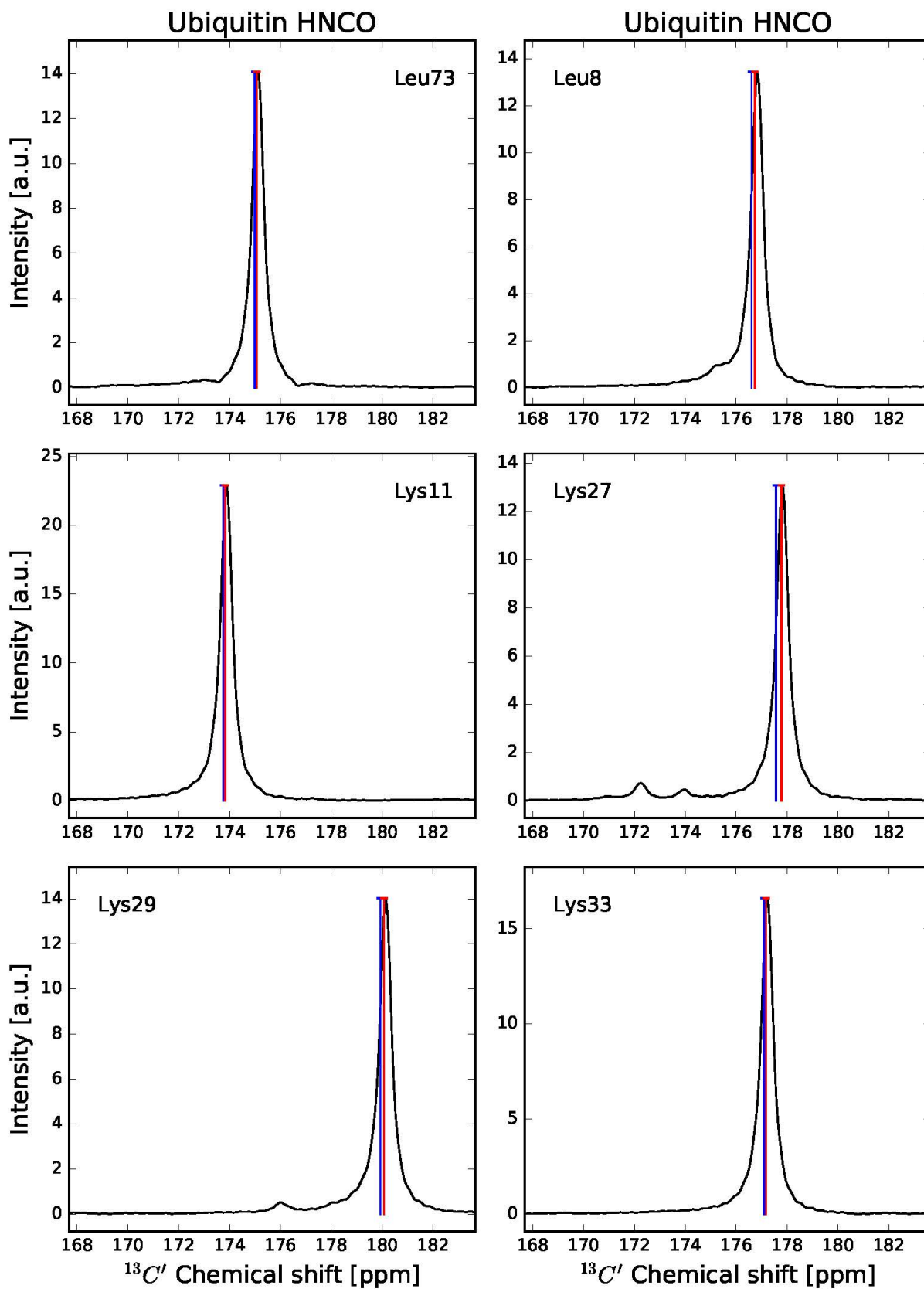


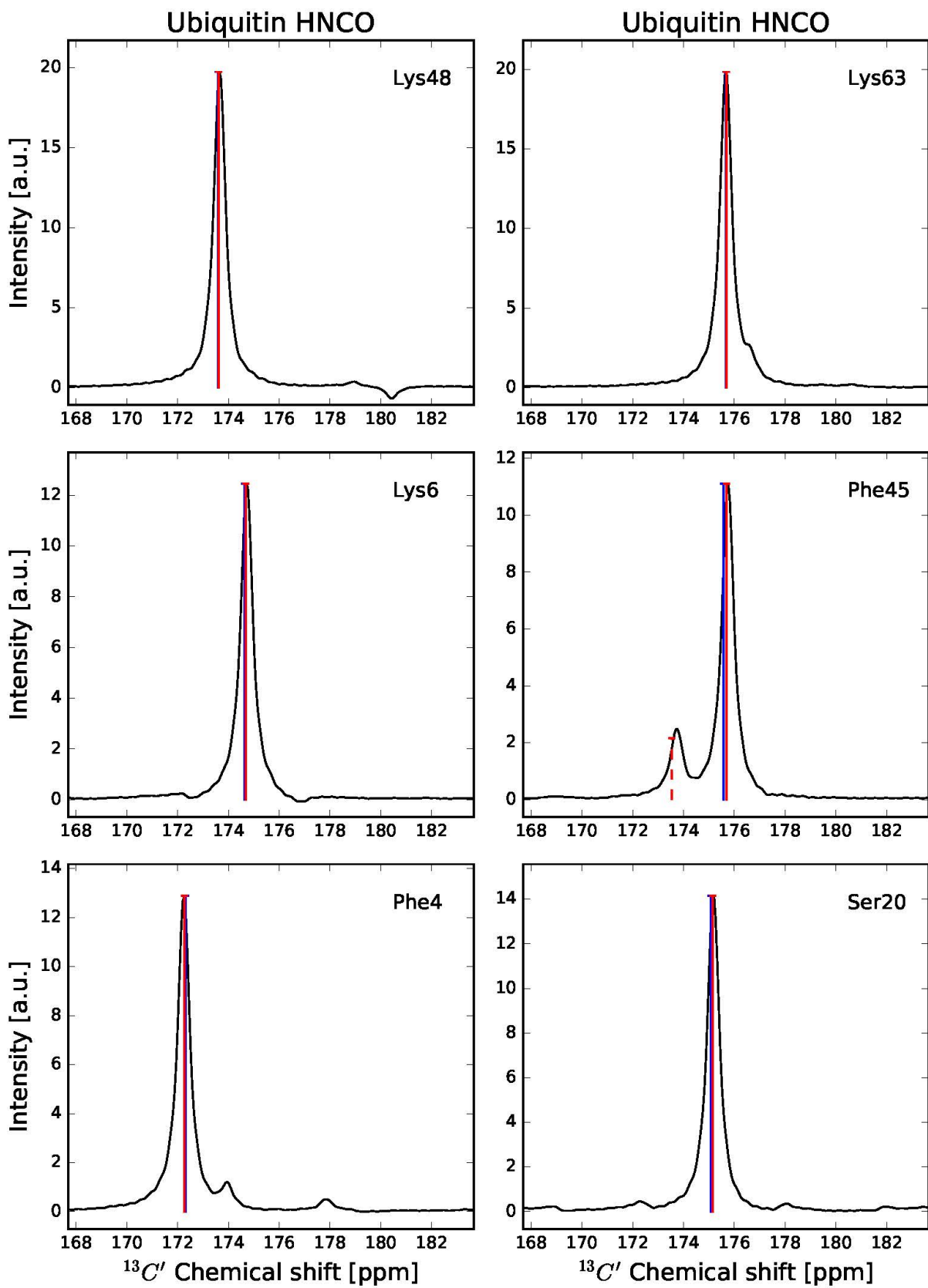


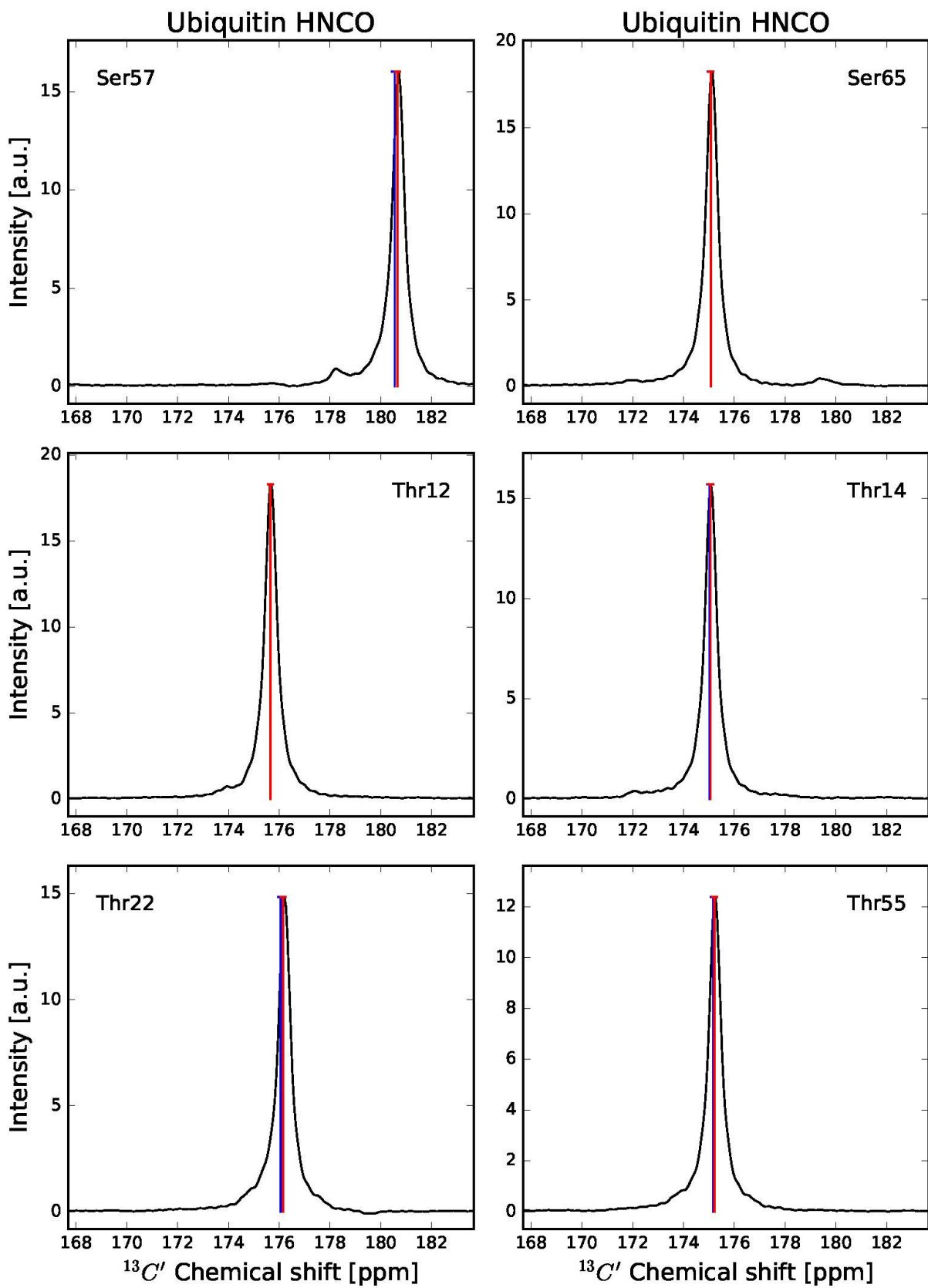


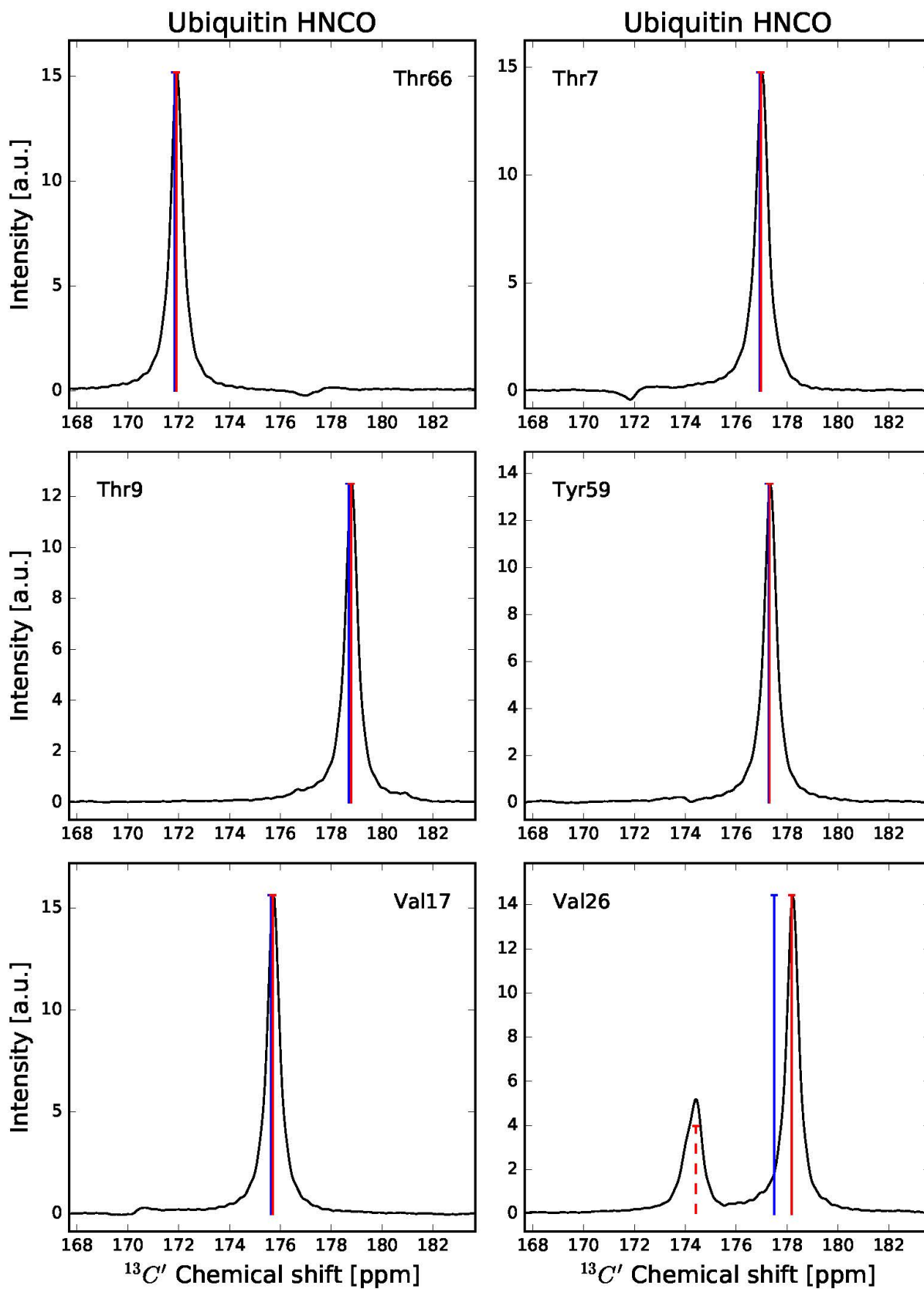












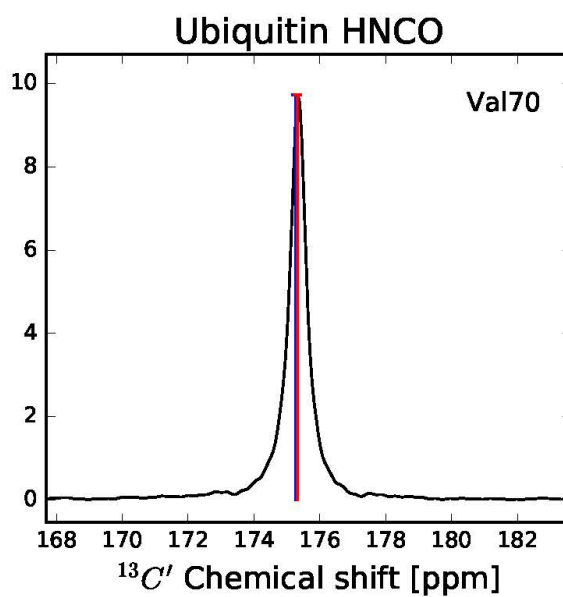
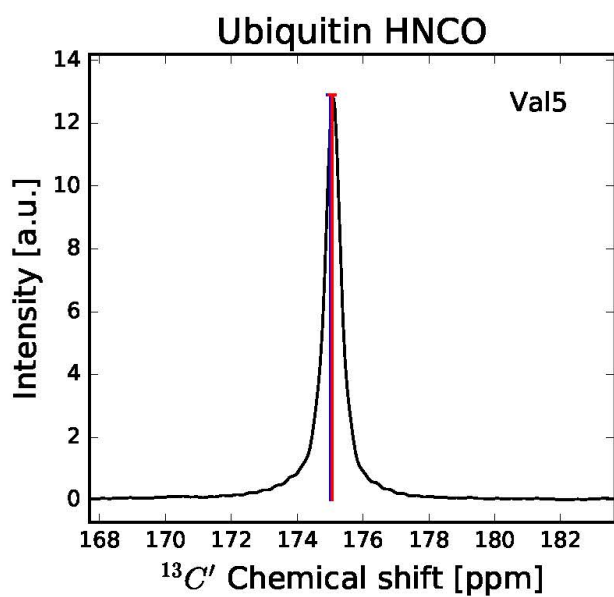


Figure S2. $^{13}\text{C}'$ traces for arginine kinase (AK) HNCO & HNCACO data. FT (black) was performed on 40 complex points, while AMS (red) used 4 (6) complex data points for the HNCO (HNCACO).

