Unambiguous Determination of Protein Arginine Ionization States in Solution by NMR Spectroscopy

Yuichi Yoshimura, Nur Alia Oktaviani, Kento Yonezawa, Hironari Kamikubo, and Frans A. A. Mulder*

Dedicated to Dr. R. M. Scheek on the occasion of his 65th birthday

Abstract: Because arginine residues in proteins are expected to be in their protonated form almost without exception, reports demonstrating that a protein arginine residue is charge-neutral are rare and potentially controversial. Herein, we present a $^{13}$C-detected NMR experiment for probing individual arginine residues in proteins notwithstanding the presence of chemical and conformational exchange effects. In the experiment, the $^{15}$N and $^{13}$N chemical shifts of an arginine head group are correlated with that of the directly attached $^{13}$C. In the resulting spectrum, the number of protons in the arginine head group can be obtained directly from the $^{15}$N–$^1$H scalar coupling splitting pattern. We applied this method to unambiguously determine the ionization state of the R52 side chain in the photoactive yellow protein from Halorhodospira halophila. Although only three $^1$H atoms were previously identified by neutron crystallography, we show that R52 is predominantly protonated in solution.

The basic amino acid arginine plays key roles in biological systems. Arginine on the surface of proteins ensures their solubility and thermostability[1–3] and because of the unique potential of the arginine head group to form multiple hydrogen bonds,[4] arginine is a key component of protein–protein interfaces and interactions between protein and nucleic acid molecules.[5,6] Arginine also functions as a charge carrier in photoreceptor proteins and ion channels.[7–9] The intrinsic pK$_a$ value of arginine in an aqueous environment is close to 14,[10] and owing to its strongly basic character, arginine in proteins rarely exists in its deprotonated state.[11]

The charge state of an arginine in photoactive yellow protein (PYP), a photoreceptor protein from Halorhodospira halophila, was examined by a combined analysis of X-ray and neutron diffraction data.[12] PYP contains two arginine residues, R52 and R124, and knowledge of the ionization state of R52, which functions as a gateway in the reaction center of PYP,[13,14] is important for understanding the energetics of the photo cycle. The chromophore of PYP, p-coumaric acid (pCA), forms two very short hydrogen bonds to the side chains of Y42 and E46.[15] The crystallographic structure indicated that the E46 and pCA oxygen atoms share the hydrogen (Figure 1a) and that the electron density is spread over a large region of the active site.[11] This arrangement would suggest that the proton affinities of pCA and the coordinating groups are closely matched, allowing the formation of a low-barrier hydrogen bond (LBHB). The nuclear density also indicated that R52 was deprotonated, and this electrically neutral state would be stabilizing the delocalization of the isolated electron within the LBHB-conjugated system buried in the protein interior.[12] However, subsequent electrostatics and quantum mechanical/molecular mechanical (QM/MM) calculations have questioned the existence of the LBHB, and also proposed that R52 is protonated.[16] This controversy therefore calls for unambiguous determination of the charge state of R52 through experiments.

The NMR chemical shifts of the arginine head-group $^{15}$N, $^{13}$N, and $^{13}$C nuclei are sensitive to the charge state,[10] and a dispersion of the chemical shift in pH titration experiments can be used to determine the side chain pK$_a$ value. The $^{15}$N shifts...
and ${}^{13}$C chemical shifts can be obtained in 2D HD(CD)$_2$NE,$^{[17]}$ HD(CDNE)CZ,$^{[17]}$ and $^{13}$N–$^{13}$C correlation spectra.$^{[18]}$ For PYP we attempted pH titration experiments and followed the $^{13}$C and $^{15}$N resonances of R52. As can be seen in Figures S1 and S2 in the Supporting Information, these chemical shifts appeared constant with pH, and, owing to the absence of a (de)protonation event, it cannot be concluded whether R52 is charge-neutral or protonated.

As seen from Figure 1 b, a count of the number of protons in the arginine head group would distinguish whether it is in the guanidinium (positive) or guanidine (neutral) form. Unfortunately, NMR experiments relying on $^1$H detection are compromised by exchange broadening effects, as indicated by a strong dependence of NMR signals on the magnetic field strength and temperature (see Figure S3 in the Supporting Information). First, chemical exchange with the solvent will render the labile (that is, $^1$H and $^3$H) protons invisible in spectra at high pH values. As a consequence, $^1$H-based experiments are never adequate for monitoring titrations of arginine residues in proteins, as these have pK$_a$ values well above neutral pH. Another factor that compromises $^1$H detection is millisecond timescale rotation around the C–N bond and C–N bond that exchanges the chemical and magnetic environments of the $^1$H protons, leading to strong exchange broadening.$^{[19,20]}$ Figure 2a shows the region of the 2D $^1$N–$^1$H heteronuclear single-quantum correlation (HSQC) spectrum containing the arginine head groups for PYP at 298 K. For R124, two broad signals are observed at the $^{15}$N frequency of 72.3 ppm. R124 is situated at the C-terminus of the protein, and is exposed on the surface,$^{[15]}$ such that C–N bond rotation is sufficiently rapid to make the two $^{15}$N frequencies coalesce. For the protons, two separate broadened signals are observed and can be explained by rotations about C–N that are fast enough to average the $^1$H chemical shifts within each amino group but slow rotation on the $^1$H chemical-shift scale about C–N.$^{[21]}$ For R52, on the other hand, only one signal is observed in the $^{15}$N–$^1$H HSQC spectrum. This observation could be interpreted to mean that the two $^{15}$N and the four $^1$H chemical shifts are all averaged, indicating that the partly buried R52 in the protein would rotate faster than R124. As we will show below, this conclusion is incorrect, and the correct conclusion can only be reached using data from a new NMR experiment.

In the following, we present a novel 2D NMR experiment for probing arginine head groups in proteins that assuages the chemical and conformational exchange effects that plague $^1$H detection. In this experiment, the head-group $^{15}$N and $^{15}$N chemical shifts are correlated with the directly attached $^{13}$C chemical shift using the pulse sequence shown in Figure S4 in the Supporting Information. Briefly, the magnetization transfer can be described as $^1$H $\leftrightarrow$ $^{15}$N$^0(t_1) \rightarrow$ $^1$C$^0(t_2)$. The signal loss during the initial $^1$H $\rightarrow$ $^{15}$N transfer due to solvent exchange is minimized by the use of cross-polarization (CP)$^{[22,23]}$ which makes efficient use of water magnetization that is being continuously pumped through $^1$H$^0$ to $^{15}$N$^0$. The CP scheme works efficiently for both NH and NH$_2$ spin systems at the same time. The simultaneous transfer of polarization from the $^{15}$N and $^{15}$N spins to $^{13}$C is also achieved by CP. The use of CP in this step is crucial to mitigate coherence losses from the chemical exchange processes.$^{[20,24]}$

Lastly, the experiment makes use of $^{13}$C detection, taking advantage of the favorable relaxation properties of the $^{13}$C nucleus because of the absence of directly attached protons.$^{[33]}$

Figure 2b shows the arginine $^{15}$N$^0$–$^{13}$C correlation spectrum for PYP. In contrast to the $^{15}$N–$^1$H HSQC spectrum, two $^{15}$N signals are detected for R52 at 74.0 and 70.6 ppm. The observation of separate $^{15}$N resonances indicates that the R52 side chain displays impeded mobility, in complete agreement with the engagement of the two NH$_2$ groups in hydrogen-bonds to the backbone carbonyl oxygen atoms of T50 and Y98 (Figure 1a) $^{[15]}$ R124, on the other hand, is not involved in hydrogen bonding and rotates faster, leading to a single averaged peak. It is clear that the single peak observed for R52 in the $^{15}$N–$^1$H HSQC spectrum (Figure 2a) corresponds to only one of the amino groups, while the other resonance is not detected. A possible explanation is that the protons for one of the amino groups are in fast exchange on the NMR timescale and therefore give rise to a sharp signal, whereas those in the other group are in intermediate exchange and broadened beyond detection.$^{[25]}$ Such a situation occurs if the proton chemical shift differences and/or rotation rates for the two amino groups differ. As indicated in the Supporting Information Figure S3c, two separate $^{15}$N frequencies are discernible at low temperature, which coalesce at higher temperature. The difference between R52 and R124 is that R52 has a higher activation barrier and it remains in slow exchange at all temperatures explored in this study.

Previous report showed that the $^{15}$N nucleus of the amino acid arginine exhibited a large chemical shift change upon deprotonation.$^{[19]}$ In aqueous solution, the $^{15}$N chemical shifts...
for the protonated and charge-neutral forms are circa 71 and circa 93 ppm, respectively. The two $^{15}$N chemical shifts of R52 obtained in this study are close to that for the guanidinium form, indicating that R52 is predominantly protonated. However, the partial burial of R52 in the low dielectric environment of the protein may affect the chemical shifts, potentially curbing the assignment of a unique protonation state.

Although conformational and chemical exchange broadening precludes direct $^{1}H$ detection, a count of the number of protons can nevertheless be made if exchange with the solvent is slow enough to observe $^{15}$N–$^{1}H$ spin–spin couplings (Figure S5 in the Supporting Information). This was previously used to unambiguously determine the protonation state for lysine side chains in proteins.[27,28] In the same vein, Figure 2c shows a 2D arginine $^{15}$N–$^{13}$C correlation spectrum measured in the absence of $^{1}H$-decoupling during the $^{15}$N chemical shift encoding period. In the $^{1}H$-coupled spectrum, the $^{15}$N signals of R52 and R124 show a doublet pattern owing to the single protons attached, whereas all signals for $^{15}$N show triplets owing to two simultaneous $^{15}$N–$^{1}H$ couplings per nitrogen. Our experimental data demonstrate the presence of four $^{1}H$ protons for R52, indicating that the cationic form prevails in solution. These results are consistent with the observed $^{15}$N chemical shifts, which are much closer to that expected for a protonated form (ca. 71 ppm) than a deprotonated form (ca. 93 ppm). Although neutron crystallography identified only three $^{1}H$ atoms for R52,[12] a reconciliation may lie in intrinsic differences in the protein structure of the solution and solid states.[29] Future experiments by solid-state NMR of crystalline samples can help to resolve this difference.

Finally, the generality of our methodology is demonstrated by applications to human ubiquitin (2 mM) at pH 6.7 and phage T4 lysozyme (1 mM) at pH 5.5, which contain 4 and 13 arginine residues, respectively. As shown in Figure 3a,b, $^{15}$N–$^{1}H$ HSQC spectra are severely affected by signal broadening for these proteins as well. However, the corresponding $^{15}$N–$^{13}$C correlation spectra (Figure 3c,d) are sufficiently resolved, such that most of the head group signals can be assigned. As the $^{15}$N chemical shifts are between 70 and 75 ppm, it is expected that the arginine side chains in ubiquitin and T4 lysozyme are effectively protonated. This is confirmed by $^{1}H$-coupled $^{15}$N–$^{13}$C correlation spectra of ubiquitin and T4 lysozyme (Supporting Information, Figures S6 and S7).

In summary, we show that the charge states of individual arginine side chains in proteins are accessible by the $^{15}$C-detected experiments presented herein. The methodology demonstrated may be useful for large folded proteins as well because side-chain and backbone motions are often uncoupled.[18,30,31]

**Experimental Section**

NMR samples were prepared as described in the Supporting Information. Unless otherwise indicated, NMR experiments were performed at 25°C on a Bruker Avance III 950 MHz spectrometer with a cryogenic TCI probe. The experimental settings are provided in the Supporting Information, Table S3. The spectral assignment of R52 and R124 in PYP was taken from BMRB entry ID 18122[25] and further confirmed using 3D HNCA and $^{15}$N-edited NOESY-HSQC experiments.

**Acknowledgements**

This work was performed using the NMR spectrometers at the Danish Center for Ultra-High Field NMR Spectroscopy. Prof. Lewis E. Kay (University of Toronto) is acknowledged for providing the T4 lysozyme plasmid. We thank Dr. Mengjui Xue and Tania A. Nielsen (Aarhus University) for sample preparation, Camilla B. Andersen (Aarhus University) for SDS-PAGE and mass analysis, and Dr. Morten Bjerreing (Aarhus University) for assistance with the setup of the NMR experiments. We also thank Prof. Hiroshi Ishikita (University of Tokyo) for helpful discussions. This work was financially supported by Grant-in-Aid (25102003) for Scientific Research on Innovative Areas (to H.K.).

**Keywords**: arginine - cross polarization - $^{15}$N–$^{1}H$ spin–spin coupling - NMR spectroscopy - photoactive yellow protein

**Figure 3.** Probing arginine side chains of human ubiquitin and phage T4 lysozyme. $^{15}$N–$^{1}H$ HSQC spectra of a) ubiquitin and b) T4 lysozyme. The $^{15}$N–$^{1}H$ and $^{15}$N–$^{1}C$ correlations are indicated in red and blue, respectively. $^{15}$N–$^{13}$C correlation spectra of c) ubiquitin and d) T4 lysozyme. The $^{15}$N–$^{13}$C and $^{15}$N–$^{15}$C correlations are indicated in red and blue, respectively. The number of scans is as follows: a) 8, b) 16, c) 256, and d) 320. For further details, see the Supporting Information. The chemical shifts of assigned resonances for ubiquitin and T4 lysozyme are provided in the Supporting Information, Tables S1 and S2, respectively.
How to cite: Angew. Chem. Int. Ed. 2017, 56, 239–242

[21] Rotations about C=N that are fast enough to average the ’H chemical shifts between the two positions of the entire amino groups, coupled with slow motion on the chemical shift scale of the individual amino group due to slow rotation about C=N, would lead to the same observation.
[25] Solvent exchange that is too fast to achieve ’H-detection for one of the amino groups would lead to the same observation. However, this explanation is not valid in this case because it would lead to self-decoupling of the splitting in the ’H-coupled spectrum (see Figure 2c).
[33] The pulse sequence code is available from http://www.protein-nmr.org

Manuscript received: September 30, 2016
Final Article published: November 29, 2016