

# SCIENTIFIC REPORTS



OPEN

## Determination of the $pK_a$ of the N-terminal amino group of ubiquitin by NMR

Received: 14 November 2016

Accepted: 26 January 2017

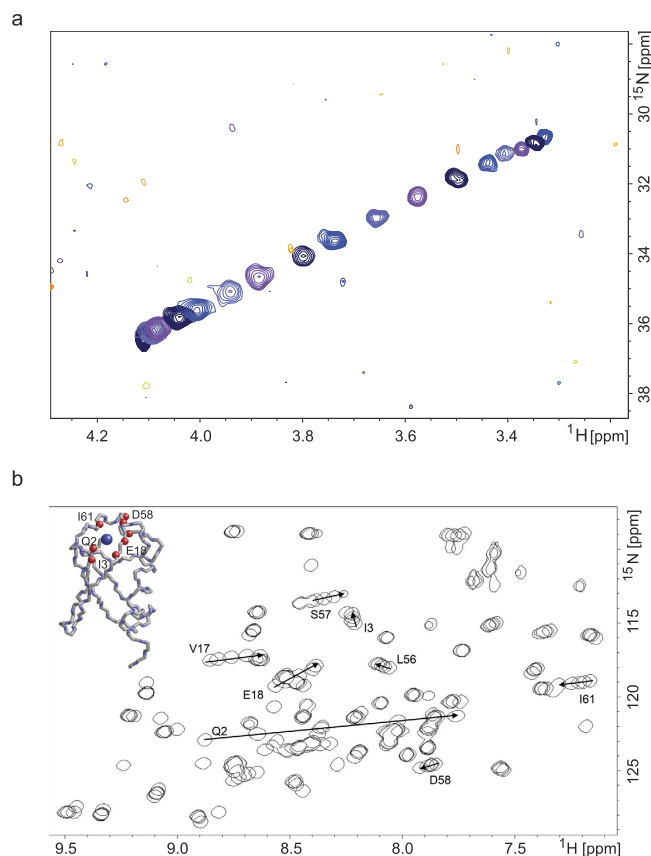
Published: 02 March 2017

Alain Oregioni<sup>1,\*</sup>, Benjamin Stieglitz<sup>2,†,\*</sup>, Geoffrey Kelly<sup>1</sup>, Katrin Rittinger<sup>2</sup> & Tom Frenkiel<sup>1</sup>

Ubiquitination regulates nearly every aspect of cellular life. It is catalysed by a cascade of three enzymes and results in the attachment of the C-terminal carboxylate of ubiquitin to a lysine side chain in the protein substrate. Chain extension occurs via addition of subsequent ubiquitin molecules to either one of the seven lysine residues of ubiquitin, or via its N-terminal  $\alpha$ -amino group to build linear ubiquitin chains. The  $pK_a$  of lysine side chains is around 10.5 and hence E3 ligases require a mechanism to deprotonate the amino group at physiological pH to produce an effective nucleophile. In contrast, the  $pK_a$  of N-terminal  $\alpha$ -amino groups of proteins can vary significantly, with reported values between 6.8 and 9.1, raising the possibility that linear chain synthesis may not require a general base. In this study we use NMR spectroscopy to determine the  $pK_a$  for the N-terminal  $\alpha$ -amino group of methionine1 of ubiquitin for the first time. We show that it is 9.14, one of the highest  $pK_a$  values ever reported for this amino group, providing a rationale for the observed need for a general base in the E3 ligase HOIP, which synthesizes linear ubiquitin chains.

The modification of proteins with polyubiquitin chains has profound consequences for their behaviour and can target proteins to the proteasome for degradation, activate signalling cascades or regulate the DNA damage response amongst many other functions. The fate of the modified protein depends on the linkage type between ubiquitin molecules within the chain: the C-terminal carboxylate of one ubiquitin molecule can form an isopeptide bond with one of the seven lysine residues within ubiquitin thereby creating seven distinct monotypic chains or, alternatively, form a peptide bond with the N-terminal  $\alpha$ -amino group of methionine1 to produce so-called linear or Met1-linked polyubiquitin chains<sup>1,2</sup>. These linear chains play important roles in the regulation of immune and inflammatory signalling pathways and contribute to the regulation of apoptotic signalling processes<sup>3,4</sup>. They are synthesized by the multi-component E3 ligase LUBAC that consists of three subunits termed HOIP, HOIL-1L and SHARPIN<sup>5-7</sup>. HOIP provides LUBAC with the ability to produce linear polyubiquitin chains in a highly specific manner using its C-terminally located RBR (RING between RING) domain<sup>8-10</sup>. RBR domain-containing E3 ubiquitin ligases form a subfamily of E3s that adopt a hybrid mechanism integrating the properties of RING and HECT-type ligases: a canonical RING domain ("RING1") initially recognizes the E2~ubiquitin conjugate and subsequently the ubiquitin is transferred onto a conserved cysteine residue located in the RING2 domain of the RBR to form a thioester intermediate before the final transfer of ubiquitin onto a substrate<sup>11-13</sup>. This mechanism is similar to that of HECT-type ligases, which also form a thioester intermediate with ubiquitin, whereas RING-type E3s play a more indirect role and act as a platform to bring the E2~Ub and substrate into close proximity and stabilise a conformation of the E2~Ub conjugate that is primed for ubiquitin transfer<sup>14</sup>. Regardless of the type of E3 ligase involved, for the transfer of ubiquitin onto the substrate to proceed efficiently the incoming nucleophile, which is either the  $\epsilon$ -amino group of a lysine side chain or the N-terminal  $\alpha$ -amino group of Met1, needs to be in its deprotonated form. The  $pK_a$  of lysine side chains is around 10.5, meaning they are protonated at physiological pH and hence to be effective nucleophiles a mechanism is required to depress the  $pK_a$ . Acidic residues in E2s, such as Asp127 in the SUMO-specific E2 Ubc9 and Asp117 in UbcH5, have been suggested to contribute to  $pK_a$  depression during ubiquitin transfer by RING ligases<sup>15-18</sup>. Similarly, an Asp residue in the active site of the HECT ligase Rsp5 has been reported to play a role in deprotonating the substrate lysine<sup>19</sup>.

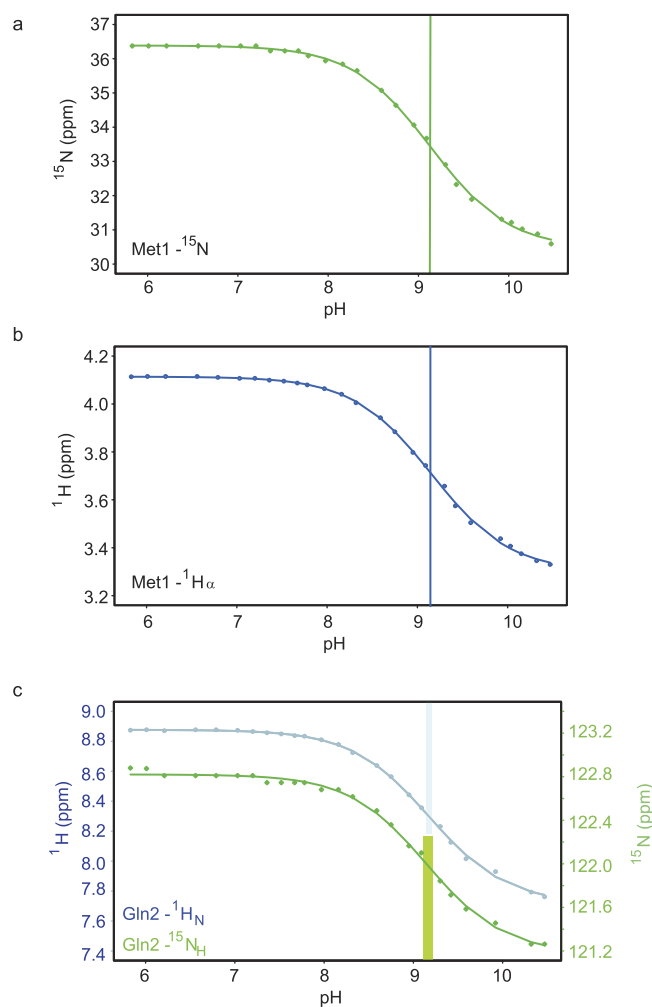
<sup>1</sup>MRC Biomedical NMR Centre, The Francis Crick Institute, 1 Midland Road, London NW1 1AT, UK. <sup>2</sup>Molecular Structure of Cell Signalling Laboratory, The Francis Crick Institute, 1 Midland Road, London NW1 1AT, UK. <sup>†</sup>Present address: School of Biological and Chemical Sciences, Department of Chemistry and Biochemistry, Queen Mary University of London, London, UK. \*These authors contributed equally to this work. Correspondence and requests for materials should be addressed to K.R. (email: [katrin.rittinger@crick.ac.uk](mailto:katrin.rittinger@crick.ac.uk)) or T.F. (email: [Tom.Frenkiel@crick.ac.uk](mailto:Tom.Frenkiel@crick.ac.uk))



**Figure 1. Overlap of ubiquitin spectra at different pH values.** (a) Overlay of successive 2D HA(CA)N spectra from the titration of human ubiquitin over the pH range 5.8 (lower left) to 10.5 (upper right). The peak arises from the correlation of the terminal  $^{15}\text{N}$  with the  $\text{H}_\alpha$  of the same residue; the pulse sequence has been adapted so that only this correlation has detectable intensity. In the interests of clarity only a subset of the recorded spectra have been shown on this plot. (b) Overlay of  $^1\text{H}$ - $^{15}\text{N}$  SOFAST-HMQC spectra at the end-points of the titration and three intermediate pH values. The peaks that show appreciable pH dependence in this range are annotated; the arrows show the movement of the peaks with increasing pH. In the inset structural diagram (1UBQ.pdb)<sup>50</sup> the backbone nitrogen atoms corresponding to these peaks are shown as red spheres, and the N-terminal amino group is shown in blue).

In contrast to lysine side chains, the  $\text{pK}_a$  values of the N-terminal amino group in proteins are less well characterised. Values between 6.8 and 9.1 have been reported, with an average of  $7.7 \pm 0.5$ <sup>20</sup>. These relatively low values raise the possibility that the synthesis of linear polyubiquitin chains may not require a general base to activate the nucleophile. However, our structural work on the catalytic core of the RBR ligase HOIP – which is the only E3 ligase capable of synthesizing linear ubiquitin chains<sup>8</sup> – provided a molecular explanation for the observed high chain linkage specificity of HOIP, and highlighted a histidine residue in the active site, His887, that was ideally positioned to carry out the role of a general base<sup>21</sup>. Indeed, substitution of this histidine residue with alanine severely suppresses catalytic activity, which however could be rescued at high pH, strongly indicating that His887 acts as a general base to deprotonate the  $\alpha$ -amino group of Met1. This apparent need for a general base prompted us to ask what the  $\text{pK}_a$  value of the N-terminal  $\alpha$ -amino group of Met1 might be. Although ubiquitin is one of the best studied proteins and the  $\text{pK}_a$  of most of its ionisable groups have been determined experimentally<sup>22</sup> we could not find any reports describing Met1  $\text{pK}_a$  determination. A number of computational approaches exist to calculate  $\text{pK}_a$  values based on the three-dimensional structure of proteins<sup>23,24</sup>. However, an overlap of different NMR and crystal structures of ubiquitin showed slight differences in the environment of Met1 and we therefore decided to use an NMR-based approach to determine the  $\text{pK}_a$  experimentally.

NMR chemical shifts are exquisitely sensitive to local ionisation events and serve as excellent reporters in pH titrations. As a result, NMR titrations have been very widely used to measure  $\text{pK}_a$  values in a large range of molecules, including the biologically important ionization of the side-chains of aspartic and glutamic acid and histidine residues of proteins, for which many hundreds of measurements have been reported<sup>20</sup>. In contrast, the literature contains relatively few  $\text{pK}_a$  measurements of N-terminal amino groups. Direct detection of the  $^{15}\text{N}$  resonance has been used in a small number of studies, but suffers from poor sensitivity<sup>25,26</sup>. Higher sensitivity can in principle be achieved by exploiting indirect proton-detected methods such as HSQC and SOFAST, but these methods depend on scalar coupling between the  $^{15}\text{N}$  and  $^1\text{H}$  nuclei, and are therefore not applicable to free amino groups in aqueous solution where the amino protons exchange with solvent at a rate that is much faster than the scalar coupling.

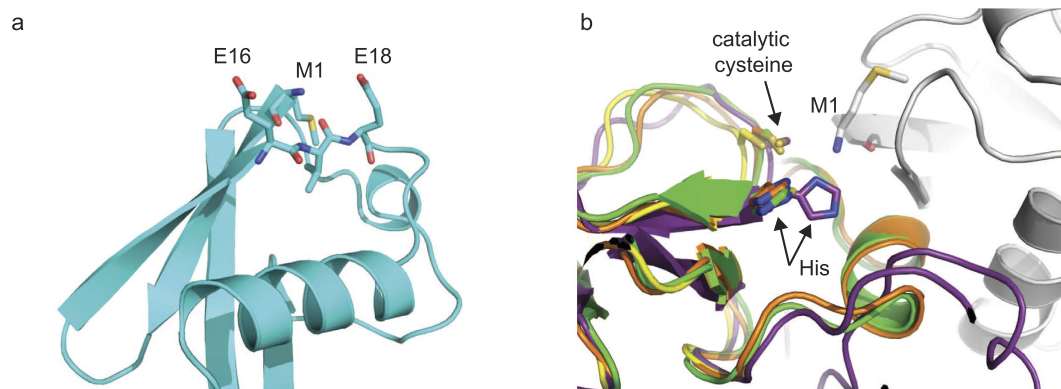


**Figure 2. Chemical shift dependence on pH.** The chemical shift dependence on pH of (a) Met1- $^{15}\text{N}$  and (b) Met1- $^1\text{H}_\alpha$  measured with the HA(CA)N experiment; (c) Gln2- $^{15}\text{N}$  and  $\text{H}_\text{N}$  measured with  $^1\text{H}$ - $^{15}\text{N}$  SOFAST-HMQC. The solid lines are the curves fitted to the Henderson-Hasselbalch equation. The widths of the vertical bars denote the 95.45% confidence intervals of the fits, as determined by the statistical bootstrap.

Measured nucleus	NMR experiment	$\text{pK}_a$	CL−	CL+	$\sigma_{\text{HA}}$	$\sigma_{\text{A}}$
Met1- $^{15}\text{N}$	HA(CA)N	9.14	9.105	9.163	36.386	30.444
Met1- $^1\text{H}_\alpha$	HA(CA)N	9.16	9.133	9.188	4.114	3.301
Gln2- $^1\text{H}_\text{N}$	$^1\text{H}$ - $^{15}\text{N}$ SOFAST-HMQC	9.18	9.143	9.209	8.877	7.718
Gln2- $^{15}\text{N}$	$^1\text{H}$ - $^{15}\text{N}$ SOFAST-HMQC	9.15	9.101	9.203	122.821	121.178

**Table 1.  $\text{pK}_a$  values derived from chemical shift measurements.**  $\text{pK}_a$  values derived from chemical shift measurements of four nuclei (Met1- $^{15}\text{N}$ , Met1- $^1\text{H}_\alpha$ , Gln2- $^1\text{H}_\text{N}$  and Gln2- $^{15}\text{N}$ ) from two NMR sequences (HA(CA)N and  $^1\text{H}$ - $^{15}\text{N}$  SOFAST-HMQC). CL− and CL+ are the lower and upper limits of the fitting confidence intervals calculated at the 95.45% confidence level as described in the text. The limiting chemical shifts (in ppm) obtained from the non-linear regression analysis are denoted by  $\sigma_{\text{HA}}$  and  $\sigma_{\text{A}}$  for the protonated and deprotonated forms respectively.

Alternative strategies involve detection of resonances from nuclei that are more remote from the ionization site, and whilst this approach provides opportunities for using more sensitive NMR methods it creates the risk that the chosen reporters will be affected by ionization events from multiple sites in the protein, resulting in chemical shifts that show complex pH dependence, with concomitant difficulties in interpretation of the data and the possibility of substantial errors in derived  $\text{pK}_a$  values<sup>27</sup>. It is generally recognised that the best reporter is the ionization site itself, where the frequency change associated with the ionization is likely to be largest, followed by the nucleus or nuclei that are covalently bound to the ionization site and other closely adjacent nuclei.



**Figure 3. Structural environment of Met1 of ubiquitin and active site of RBR ligases.** (a) Structure of ubiquitin highlighting the residues close to Met1 (1UBQ.pdb)<sup>50</sup>. (b) Overlap of the structures of HOIP in purple (4LJO.pdb)<sup>21</sup>, Parkin in green (5CAW.pdb)<sup>40</sup> and orange (5C23.pdb)<sup>38</sup> and HHARI in yellow (4KBL.pdb)<sup>36</sup> zoomed in onto the active site. The catalytic Cys885 and His887 of HOIP, plus Met1 of the acceptor ubiquitin (in gray) are shown.

In view of these considerations we have adopted an indirect-detection approach to determine the  $pK_a$  of the  $\alpha$ -amino group of Met1 by monitoring the  $^{15}\text{N}$  resonance frequency, detected indirectly via the non-labile alpha hydrogen of the terminal residue. This approach is based on that used by André *et al.* to determine the  $pK_a$  values of side-chain ionizable groups of lysine and arginine residues<sup>28</sup>, and by Lorieau *et al.* for the amino group of the influenza hemagglutinin fusion peptide<sup>29</sup>.

## Results and Discussion

**Data acquisition and analysis.** The terminal amino  $^{15}\text{N}$  data were collected using a two-dimensional version of the HACAN pulse sequence<sup>30</sup>, essentially in the form of Kanelis *et al.*<sup>31</sup>, modified to provide discrimination against the unwanted signals from backbone amide groups and side-chain amino groups of lysine residues. The modifications (described in Methods) reduced the unwanted signals to undetectable levels, and, as a consequence, the assignment of the signal from the N-terminal amino group was self-evident.

The pH titration was carried out by progressive transfer of aliquots between a pair of samples that were initially adjusted to pH 5.8 and 10.5 respectively, as described in Methods. At each stage of the titration the pH of the samples was measured via the  $^1\text{H}$  chemical shifts of a set of indicator molecules present as co-solutes<sup>32</sup>; this avoids the difficulties that arise with conventional electrode-based pH measurement of small-volume samples. The experimental methodology and the resulting  $pK_a$  values at 21.5 °C for the indicator molecules are described in Methods.

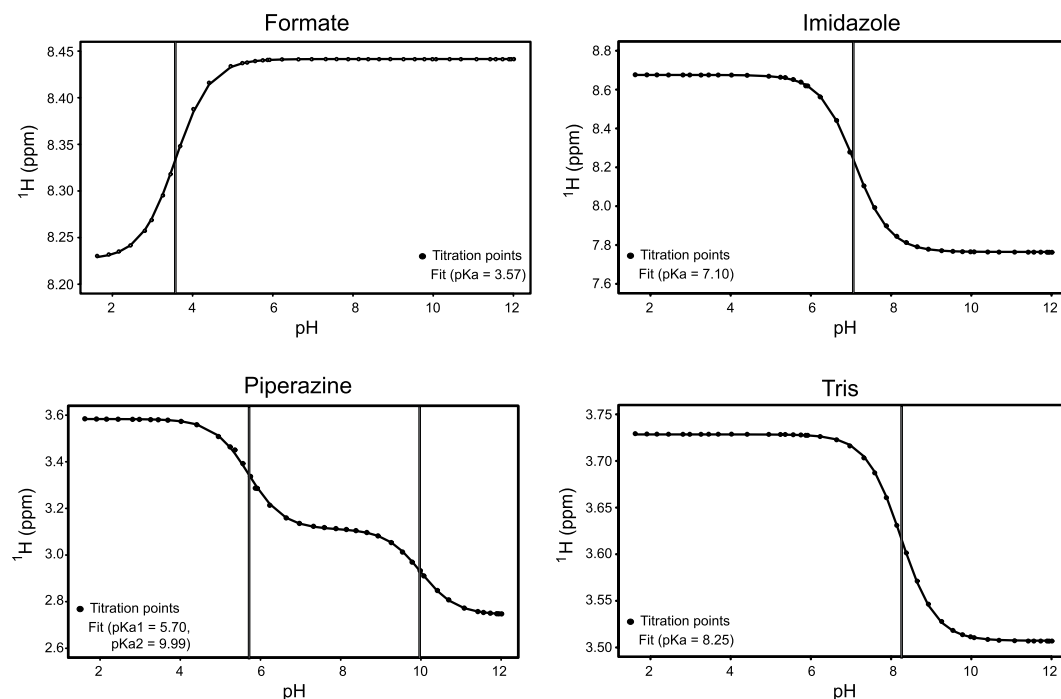
The results of the HA(CA)N experiment are shown in Fig. 1a, and although the primary purpose of this experiment was to detect the  $^{15}\text{N}$  chemical shift of the terminal nitrogen, the alpha hydrogen clearly also shows a pH dependence that is strongly correlated to that of the  $^{15}\text{N}$ . In addition to the HA(CA)N experiment, a two-dimensional  $^1\text{H}$ - $^{15}\text{N}$  SOFAST-HMQC data set<sup>33</sup> was recorded at each pH point to assess the integrity of the protein, which was found to remain folded up to the highest pH used in this work (Fig. 1b). In the SOFAST spectra a number of cross-peaks migrate as a function of pH, most notably the peak from the backbone amide of Gln2, and the coordinates of this peak were also analysed.

The peak positions for the ubiquitin titrations were fitted using non-linear regression to the Henderson-Hasselbalch equation, recast in terms of chemical shifts:

$$\sigma_{\text{peak}} = \sigma_{\text{HA}} + \frac{\Delta\sigma}{1 + 10^{(pK_a - pH)}} \quad (1)$$

where  $\sigma_{\text{peak}}$  is the measured NMR chemical shift of the peak of interest,  $\sigma_{\text{HA}}$  is the chemical shift of the protonated form and  $\Delta\sigma$  is the difference between  $\sigma_{\text{HA}}$  and the shift of the deprotonated form. All fitting and statistical analysis was carried out using the statistics package R, as described in Methods<sup>34</sup>. Good fits to eq. (1) were obtained for all four measured ubiquitin chemical shifts (Fig. 2).

**$pK_a$  of the N-terminal amino group.** The  $pK_a$  values derived from the NMR measurements carried out on the four nuclei at or near the N-terminal residue are in close agreement (Table 1). The  $pK_a$  derived from the  $^{15}\text{N}$  chemical shift of the terminal amino group is 9.14; this nucleus is reasonably assumed to be the most faithful reporter. The values from the other three measured nuclei fall in the range 9.15 to 9.18, with fitting confidence intervals which overlap that of the Met1-N determination. This suggests that all of the observed chemical shift changes reflect the N-terminal amino group ionisation event alone. This is unsurprising as the measured value lies in a pH window devoid of other typical ionisations in proteins. The absence of other ionization events in this pH range is further evidenced by an examination of the full set of peaks in the SOFAST  $^1\text{H}$ - $^{15}\text{N}$ -correlation spectrum, where all of the  $^1\text{H}$ - $^{15}\text{N}$  correlations which display appreciable titration behaviour in this range arise from atoms that are in close proximity to the N-terminus in the three-dimensional structure (Fig. 1b).



**Figure 4. Titration curves of internal NMR reporters.** pH titration curves for the four internal NMR reporters: formate, piperazine, imidazole-H2 and Tris. The solid lines are the curves fitted to the Henderson-Hasselbalch equation as described in the text. The resulting  $pK_a$  values are marked with vertical lines.

	$pK_a$	$\sigma_{HA}$	$\sigma_A$	Lower pH limit	Higher pH limit
Formate	3.57	8.227	8.442	2.6	4.6
Piperazine $pK_{a1}$	5.70	3.584	3.111	4.7	6.7
Imidazole	7.10	8.675	7.765	6.1	8.1
Tris	8.25	3.729	3.507	7.2	9.2
Piperazine $pK_{a2}$	9.99	3.111	2.744	9.0	11.0

**Table 2. Measured  $pK_a$  values of four internal NMR reporters.** Measured  $pK_a$  of the four internal NMR reporters, where  $\sigma_{HA}$  and  $\sigma_A$  are the chemical shifts (in ppm) of the protonated and deprotonated state respectively. The second  $pK_a$  of piperazine has been included to extend the range of measurable pH. The suggested range of applicability (limits calculated as  $pK_a \pm 1$ ) of each reporter is also shown.

## Conclusions

The  $pK_a$  of 9.14 for the N-terminal amino group of ubiquitin, determined from the  $^{15}N$  chemical shift, is at the upper limit of experimental values for proteins reported in the literature, which span 6.8 to 9.1, and is also higher than that determined for model peptides<sup>20,35</sup>. The  $\alpha$ -amino group of Met1 is solvent exposed making it difficult to rationalize why its  $pK_a$  is higher than in most other cases and we speculate that the proximity of Glu16 and Glu18 (Fig. 3a) create a negatively charged environment that raises the  $pK_a$ . Interestingly, the conserved histidine residue in HOIP that we previously proposed to act as a general base to deprotonate the N-terminal amino group is conserved in a number of RBR ligases, all of which transfer ubiquitin onto a lysine sidechain<sup>21</sup>. At present the structure of the HOIP/ubiquitin complex is the only snapshot of an RBR ligase/substrate active site. The only other structures of RBR ligases available are of HOIP bound to E2~ubiquitin, and HHARI and Parkin in the autoinhibited state, or partially active forms of Parkin<sup>36–41</sup>. The histidine is in a different conformation in those structures (Fig. 3b) and points away from the ubiquitin which could indicate that substrate binding might either induce changes around the active site in HHARI or Parkin or that the substrate is presented to the active site in a different orientation in these RBRs. Further studies are required to fully understand how the incoming nucleophile is activated in other RBRs.

## Methods

**Internal pH indicator validation.** We used the set of internal indicator molecules proposed by Baryshnikova *et al.*<sup>32</sup> – Tris, formate, piperazine, and imidazole – but first re-determined the indicator  $pK_a$  and limiting chemical shift values at 21.5 °C, the temperature of interest in this study. We also re-assessed the range of applicability of each indicator for this temperature, and extended the applicability to a higher pH range by utilising the second ionisation of piperazine. The chemical shifts of the titrating signals of Tris, formate, and

pH	Met1- <sup>1</sup> H <sub>α</sub>	Met1- <sup>15</sup> N	Gln2- <sup>1</sup> H <sub>N</sub>	Gln2- <sup>15</sup> N
5.83	4.114	36.37	8.874	122.88
6.01	4.115	36.37	8.877	122.87
6.21	4.115	36.37	8.871	122.81
6.56	4.115	36.37	8.877	122.81
6.79	4.111	36.37	8.877	122.81
7.03	4.107	36.37	8.873	122.81
7.2	4.107	36.37	8.865	122.81
7.36	4.099	36.23	8.858	122.75
7.52	4.095	36.23	8.850	122.75
7.67	4.088	36.23	8.838	122.75
7.78	4.080	36.08	8.834	122.75
7.97	4.064	35.94	8.811	122.68
8.16	4.041	35.84	8.779	122.68
8.32	4.005	35.65	8.724	122.62
8.59	3.943	35.08	8.638	122.49
8.75	3.884	34.64	8.564	122.36
8.95	3.798	34.06	8.443	122.17
9.09	3.743	33.67	8.357	122.10
9.3	3.657	32.90	8.231	121.84
9.42	3.575	32.33	8.126	121.71
9.59	3.505	31.89	8.016	121.59
9.92	3.438	31.31	7.930	121.46
10.03	3.407	31.22	N.D.	N.D.
10.15	3.375	31.03	N.D.	N.D.
10.32	3.346	30.88	7.793	121.26
10.47	3.330	30.59	7.762	121.26

**Table 3. Ubiquitin titration data.** Chemical shifts (in ppm) for the ubiquitin signals used in the Met1 pKa determination, measured from HA(CA)N (Met1) and SOFAST-HMQC (Gln2) recorded at the indicated pH values. N.D. – not determined owing to peak overlap.

imidazole-H2 were analyzed using the Henderson-Hasselbalch equation (in the form shown above as eq. (1)), whereas the piperazine chemical shift was fitted to the extended form of the equation which accommodates two ionization events<sup>42</sup>:

$$\sigma_{peak} = \sigma_{HA} + \frac{\Delta\sigma_1}{1 + 10^{(pKa_1 - pH)}} + \frac{\Delta\sigma_2}{1 + 10^{(pKa_2 - pH)}} \quad (2)$$

where  $\Delta\sigma_1$  is the variation of chemical shift associated with the titration of the deprotonation governed by  $pK_{a1}$  and  $\Delta\sigma_2$  the added chemical shift variation governed by  $pK_{a2}$ .

The sample used for these measurements consisted of a solution of 100 mM KCl, 2 mM Tris, 2 mM formate, 2 mM piperazine, 2 mM imidazole and 0.2 mM DSS in 95% H<sub>2</sub>O-5% D<sub>2</sub>O, matching the buffer used subsequently in the ubiquitin measurements. The pH was adjusted across the desired range and measured using a glass pH electrode.

The resulting limiting chemical shifts and  $pK_a$  values (Fig. 4 and Table 2) agree well with those reported by Baryshnikova *et al.* at 30 °C, demonstrating the validity of their observation that this approach is quite robust with respect to changes in temperature<sup>32</sup>.

**NMR sample preparation.** Recombinant wild-type human ubiquitin was expressed as untagged protein using pET15 and *E. coli* strain BL21(DE3), and purified by anion exchange using a Q Sepharose resin followed by gel filtration on a Sephadex G-100 column (GE Health-care). <sup>13</sup>C/<sup>15</sup>N labelled ubiquitin was obtained by expression in minimal medium containing glucose-<sup>13</sup>C<sub>6</sub> and <sup>15</sup>NH<sub>4</sub>Cl as sole carbon and nitrogen sources.

Two 0.5 mM ubiquitin samples for NMR were prepared by dialysis into a 95% H<sub>2</sub>O-5% D<sub>2</sub>O buffer (100 mM KCl, 2 mM Tris, 2 mM formate, 2 mM piperazine, 2 mM imidazole, 0.2 mM DSS). The two initial samples were adjusted to pH 5.8 or pH 10.5. Intermediate pH values were attained by transferring small aliquots between samples after each set of NMR measurements. The sample that was initially at pH 5.8 thereby increased with each transfer, whilst the pH of the other sample decreased. NMR measurements were made on both samples. Accurate sample pH was monitored after each transfer using the <sup>1</sup>H chemical shift of the appropriate buffer component as an internal indicator, making use of values for  $pK_a$ ,  $\sigma_{HA}$ ,  $\sigma_A$  and applicability regions re-derived experimentally for 21.5 °C as shown in Table 2.

**NMR measurements.** All NMR measurements were performed on a Bruker Avance 600 MHz equipped with a 5 mm TCI cryoprobe, at a sample temperature of 21.5 °C, as demonstrated by a prior calibration using the method of Findeisen *et al.*<sup>43</sup>.

Three NMR spectra were acquired at each pH value: a 1D <sup>1</sup>H spectrum using excitation sculpting<sup>44</sup>, a 2D <sup>1</sup>H-<sup>15</sup>N SOFAST-HMQC spectrum<sup>33</sup>, and a 2D HA(CA)N spectrum<sup>30,31</sup>. The latter was based on the sequence of Kanelis *et al.*<sup>31</sup>, adapted to confer selectivity and improve sensitivity for the desired N-terminal amino signal. Selectivity with respect to backbone amide signals was achieved by replacing the rectangular <sup>15</sup>N inversion pulses of the standard sequence with Q3-shaped selective pulses<sup>45</sup>, centred at a <sup>15</sup>N offset of 35 ppm, and having a duration of 1.3 ms. Discrimination against signals from the side-chain amino groups of lysine residues was obtained by setting the delay within the sequence that serves for evolution of proton-coupled <sup>13</sup>C magnetization to 3.4 ms, a value that was optimal for the methine C $\alpha$  group of N-terminal methionine and simultaneously minimized the signal from the methylene C $\epsilon$  groups of lysine side-chains. 2D spectra were processed using nmrPipe<sup>46</sup>. <sup>1</sup>H chemical shifts were referenced to internal DSS; <sup>13</sup>C and <sup>15</sup>N shifts were referenced indirectly using the gyromagnetic ratios of Wishart *et al.*<sup>47</sup>. Chemical shifts for the signals used in the ubiquitin pK<sub>a</sub> analysis are given in Table 3.

**Statistical methods.** Statistical analysis and Henderson-Hasselbalch equation fitting were done using R<sup>34</sup>. Both the NMR reporters and the ubiquitin titration curves were fitted using a non-linear least-squares method (Gauss-Newton). Because of the non-linear nature of the Henderson-Hasselbalch equation, the pK<sub>a</sub> values here are given with estimated confidence intervals rather than standard deviations<sup>48</sup>. Confidence intervals of 95.45% were calculated using a bootstrap procedure<sup>49</sup>; this interval corresponds to  $\pm 2\sigma$  for a normal distribution. The bootstrap calculation was run with N = 1000 and incorporated the bias correction and acceleration methods, which improve the accuracy by taking into account the non-normality of the bootstrap distribution<sup>48</sup>.

## References

1. Yau, R. & Rape, M. The increasing complexity of the ubiquitin code. *Nat Cell Biol* **18**, 579–586, doi: 10.1038/ncb3358 (2016).
2. Komander, D. & Rape, M. The ubiquitin code. *Annual review of biochemistry* **81**, 203–229, doi: 10.1146/annurev-biochem-060310-170328 (2012).
3. Rieser, E., Cordier, S. M. & Walczak, H. Linear ubiquitination: a newly discovered regulator of cell signalling. *Trends Biochem Sci* **38**, 94–102, doi: 10.1016/j.tibs.2012.11.007 (2013).
4. Walczak, H., Iwai, K. & Dikic, I. Generation and physiological roles of linear ubiquitin chains. *BMC Biol* **10**, 23, doi: 10.1186/1741-7007-10-23 (2012).
5. Tokunaga, F. *et al.* SHARPIN is a component of the NF-kappaB-activating linear ubiquitin chain assembly complex. *Nature* **471**, 633–636, doi: 10.1038/nature09815 (2011).
6. Ikeda, F. *et al.* SHARPIN forms a linear ubiquitin ligase complex regulating NF-kappaB activity and apoptosis. *Nature* **471**, 637–641, doi: 10.1038/nature09814 (2011).
7. Gerlach, B. *et al.* Linear ubiquitination prevents inflammation and regulates immune signalling. *Nature* **471**, 591–596, doi: 10.1038/nature09816 (2011).
8. Kirisako, T. *et al.* A ubiquitin ligase complex assembles linear polyubiquitin chains. *Embo J* **25**, 4877–4887, doi: 10.1038/sj.emboj.7601360 (2006).
9. Stieglitz, B., Morris-Davies, A. C., Koliopoulos, M. G., Christodoulou, E. & Rittinger, K. LUBAC synthesizes linear ubiquitin chains via a thioester intermediate. *EMBO Rep* **13**, 840–846, doi: 10.1038/embo.2012.105 (2012).
10. Smit, J. J. *et al.* The E3 ligase HOIP specifies linear ubiquitin chain assembly through its RING-IBR-RING domain and the unique LDD extension. *Embo J* **31**, 3833–3844, doi: 10.1038/emboj.2012.217 (2012).
11. Spratt, D. E., Walden, H. & Shaw, G. S. RBR E3 ubiquitin ligases: new structures, new insights, new questions. *Biochem J* **458**, 421–437, doi: 10.1042/BJ20140006 (2014).
12. Smit, J. J. & Sixma, T. K. RBR E3-ligases at work. *EMBO Rep* **15**, 142–154, doi: 10.1002/embr.201338166 (2014).
13. Wenzel, D. M. & Klevit, R. E. Following Ariadne's thread: a new perspective on RBR ubiquitin ligases. *BMC Biol* **10**, 24, doi: 10.1186/1741-7007-10-24 (2012).
14. Berndsen, C. E. & Wolberger, C. New insights into ubiquitin E3 ligase mechanism. *Nat Struct Mol Biol* **21**, 301–307, doi: 10.1038/nsmb.2780 (2014).
15. Reverter, D. & Lima, C. D. Insights into E3 ligase activity revealed by a SUMO-RanGAP1-Ubc9-Nup358 complex. *Nature* **435**, 687–692, doi: 10.1038/nature03588 (2005).
16. Yunus, A. A. & Lima, C. D. Lysine activation and functional analysis of E2-mediated conjugation in the SUMO pathway. *Nat Struct Mol Biol* **13**, 491–499, doi: 10.1038/nsmb1104 (2006).
17. Plechanovova, A., Jaffray, E. G., Tatham, M. H., Naismith, J. H. & Hay, R. T. Structure of a RING E3 ligase and ubiquitin-loaded E2 primed for catalysis. *Nature* **489**, 115–120, doi: 10.1038/nature11376 (2012).
18. Dou, H., Buetow, L., Sibbet, G. J., Cameron, K. & Huang, D. T. BIRC7-E2 ubiquitin conjugate structure reveals the mechanism of ubiquitin transfer by a RING dimer. *Nat Struct Mol Biol* **19**, 876–883, doi: 10.1038/nsmb.2379 (2012).
19. Kamadurai, H. B. *et al.* Mechanism of ubiquitin ligation and lysine prioritization by a HECT E3. *Elife* **2**, e00828, doi: 10.7554/eLife.00828 (2013).
20. Grimsley, G. R., Scholtz, J. M. & Pace, C. N. A summary of the measured pK values of the ionizable groups in folded proteins. *Protein Sci* **18**, 247–251, doi: 10.1002/pro.19 (2009).
21. Stieglitz, B. *et al.* Structural basis for ligase-specific conjugation of linear ubiquitin chains by HOIP. *Nature* **503**, 422–426, doi: 10.1038/nature12638 (2013).
22. Sundd, M., Iverson, N., Ibarra-Molero, B., Sanchez-Ruiz, J. M. & Robertson, A. D. Electrostatic interactions in ubiquitin: stabilization of carboxylates by lysine amino groups. *Biochemistry* **41**, 7586–7596 (2002).
23. Alexov, E. *et al.* Progress in the prediction of pKa values in proteins. *Proteins* **79**, 3260–3275, doi: 10.1002/prot.23189 (2011).
24. Nielsen, J. E., Gunner, M. R. & Garcia-Moreno, B. E. The pKa Cooperative: a collaborative effort to advance structure-based calculations of pKa values and electrostatic effects in proteins. *Proteins* **79**, 3249–3259, doi: 10.1002/prot.23194 (2011).
25. Zhou, Z. *et al.* 15N NMR study of the ionization properties of the influenza virus fusion peptide in zwitterionic phospholipid dispersions. *Biophys J* **78**, 2418–2425, doi: 10.1016/S0006-3495(00)76785-3 (2000).
26. Zhu, L., Kemple, M. D., Yuan, P. & Prendergast, F. G. N-terminus and lysine side chain pKa values of melittin in aqueous solutions and micellar dispersions measured by 15N NMR. *Biochemistry* **34**, 13196–13202 (1995).
27. Webb, H. *et al.* Remeasuring HEWL pK(a) values by NMR spectroscopy: methods, analysis, accuracy, and implications for theoretical pK(a) calculations. *Proteins* **79**, 685–702, doi: 10.1002/prot.22886 (2011).

28. Andre, I., Linse, S. & Mulder, F. A. Residue-specific pKa determination of lysine and arginine side chains by indirect 15N and 13C NMR spectroscopy: application to apo calmodulin. *J Am Chem Soc* **129**, 15805–15813, doi: 10.1021/ja0721824 (2007).
29. Lorieau, J. L., Louis, J. M. & Bax, A. Helical hairpin structure of influenza hemagglutinin fusion peptide stabilized by charge-dipole interactions between the N-terminal amino group and the second helix. *J Am Chem Soc* **133**, 2824–2827, doi: 10.1021/ja1099775 (2011).
30. Wang, A. C., Grzesiek, S., Tschudin, R., Lodi, P. J. & Bax, A. Sequential backbone assignment of isotopically enriched proteins in D<sub>2</sub>O by deuterium-decoupled HA(CA)N and HA(CACO)N. *J Biomol NMR* **5**, 376–382 (1995).
31. Kanelis, V. *et al.* Sequential assignment of proline-rich regions in proteins: application to modular binding domain complexes. *J Biomol NMR* **16**, 253–259 (2000).
32. Baryshnikova, O. K., Williams, T. C. & Sykes, B. D. Internal pH indicators for biomolecular NMR. *J Biomol NMR* **41**, 5–7, doi: 10.1007/s10858-008-9234-6 (2008).
33. Schanda, P. & Brutscher, B. Very fast two-dimensional NMR spectroscopy for real-time investigation of dynamic events in proteins on the time scale of seconds. *J Am Chem Soc* **127**, 8014–8015, doi: 10.1021/ja051306e (2005).
34. RCoreTeam. R.: *A language and environment for statistical computing*. (R Foundation for statistical computing: Vienna, Austria, 2016).
35. Thurlkill, R. L., Grimsley, G. R., Scholtz, J. M. & Pace, C. N. pK values of the ionizable groups of proteins. *Protein Sci* **15**, 1214–1218, doi: 10.1110/ps.051840806 (2006).
36. Duda, D. M. *et al.* Structure of HHARI, a RING-IBR-RING ubiquitin ligase: autoinhibition of an Ariadne-family E3 and insights into ligation mechanism. *Structure* **21**, 1030–1041, doi: 10.1016/j.str.2013.04.019 (2013).
37. Wauer, T. & Komander, D. Structure of the human Parkin ligase domain in an autoinhibited state. *Embo J* **32**, 2099–2112, doi: 10.1038/emboj.2013.125 (2013).
38. Kumar, A. *et al.* Disruption of the autoinhibited state primes the E3 ligase parkin for activation and catalysis. *Embo J* **34**, 2506–2521, doi: 10.15252/embj.201592337 (2015).
39. Riley, B. E. *et al.* Structure and function of Parkin E3 ubiquitin ligase reveals aspects of RING and HECT ligases. *Nat Commun* **4**, 1982, doi: 10.1038/ncomms2982 (2013).
40. Wauer, T., Simicek, M., Schubert, A. & Komander, D. Mechanism of phospho-ubiquitin-induced PARKIN activation. *Nature* **524**, 370–374, doi: 10.1038/nature14879 (2015).
41. Lechtenberg, B. C. *et al.* Structure of a HOIP/E2~ubiquitin complex reveals RBR E3 ligase mechanism and regulation. *Nature* **529**, 546–550, doi: 10.1038/nature16511 (2016).
42. Blomberg, F., Maurer, W. & Ruterjans, H. Nuclear magnetic resonance investigation of 15N-labeled histidine in aqueous solution. *J Am Chem Soc* **99**, 8149–8159 (1977).
43. Findeisen, M., Brand, T. & Berger, S. A 1H-NMR thermometer suitable for cryoprobes. *Magnetic resonance in chemistry: MRC* **45**, 175–178, doi: 10.1002/mrc.1941 (2007).
44. Stott, K., Stonehouse, J., Keeler, J., Hwang, T.-L. & Shaka, A. J. Excitation Sculpting in High-Resolution Nuclear Magnetic Resonance Spectroscopy: Application to Selective NOE Experiments. *J Am Chem Soc* **117**, 4199–4200, doi: 10.1021/ja00119a048 (1995).
45. Emsley, L. B. G. Optimization of Shaped Selective Pulses for NMR Using a Quaternion Description of Their Overall Propagators. *J Magnetic Resonance* **97**, 135–149, doi: 10.1016/0022-2364(92)90242-Y (1992).
46. Delaglio, F. *et al.* NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* **6**, 277–293 (1995).
47. Wishart, D. S. *et al.* 1H, 13C and 15N chemical shift referencing in biomolecular NMR. *Journal of Biomolecular NMR* **6**, 135–140 (1995).
48. DiCiccio, T. J. & Efron, B. Bootstrap confidence intervals. *Statistical Science* **11**, 189–228 (1996).
49. Efron, B. Bootstrap methods: another look at the jack-knife. *The Annals of Statistics* **7**, 1–26 (1979).
50. Vijay-Kumar, S., Bugg, C. E. & Cook, W. J. Structure of ubiquitin refined at 1.8 Å resolution. *J Mol Biol* **194**, 531–544 (1987).

## Acknowledgements

This work was supported by the Medical Research Council (grants U117533887 and grant U117565398 until March 2015) and by the Francis Crick Institute which receives its core funding from Cancer Research UK (FC001142, FC10029), the UK Medical Research Council (FC001142, FC10029), and the Wellcome Trust (FC001142, FC10029). We thank Serena Faggiano and John McCormick for preparing the <sup>13</sup>C, <sup>15</sup>N-ubiquitin used in this work.

## Author Contributions

A.O., B.S., G.K. carried out experiments; A.O., G.K. analysed the data, A.O., B.S., G.K., K.R., T.F. conceived the study; A.O., B.S., G.K., K.R., T.F. wrote the paper.

## Additional Information

**Competing Interests:** The authors declare no competing financial interests.

**How to cite this article:** Oregioni, A. *et al.* Determination of the pK<sub>a</sub> of the N-terminal amino group of ubiquitin by NMR. *Sci. Rep.* **7**, 43748; doi: 10.1038/srep43748 (2017).

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2017