

Cite this: *Analyst*, 2025, **150**, 2872

# Programmable wide-range pH gradients for NMR titrations: application to antibody–drug conjugate linker group modifications†

Matthew Wallace,<sup>a</sup> James M. Sharpe,<sup>a</sup> Krzysztof Baj,<sup>b</sup> Michael Ngwube,<sup>a</sup> Jenny Thirlway,<sup>c</sup> Patrick L. Kerigan Higgs,<sup>c</sup> G. Richard Stephenson,<sup>a</sup> Jonathan A. Iggo,<sup>b</sup> Thomas E. Storr<sup>a</sup> and Christopher J. Richards<sup>\*a</sup>

We generate pH gradients spanning more than six units in standard NMR tubes and determine all the  $pK_a$  values of polyprotic compounds in single 20 minutes chemical shift imaging (CSI) NMR experiments. The modest demands of our method in terms of sample quantity and preparation time allow it be performed as part of the routine characterisation and optimisation of organic molecules during synthesis campaigns. As proof of concept, we measure the  $pK_a$  values of a family of vinylpyridines employed as antibody drug conjugate linkers. Our analysis reveals a strong correlation between the experimental aqueous  $pK_a$  and the rate of conjugate addition of thiol nucleophiles to the vinyl group, representing a powerful predictive method.

Received 8th April 2025,  
Accepted 25th May 2025  
DOI: 10.1039/d5an00406c  
rsc.li/analyst

## 1. Introduction

NMR titrations are a highly effective if laborious way to elucidate the pH-responsive behaviour of chemical systems including small molecule pharmaceuticals,<sup>1</sup> polymers,<sup>2</sup> proteins<sup>3</sup> and synthetic supramolecular systems.<sup>4</sup> As a faster alternative to the conventional titration, we demonstrated that pH gradients could be established in standard NMR sample tubes by placing a solution on top of crystals of an acid. <sup>1</sup>H NMR spectra were recorded along this gradient using chemical shift imaging (CSI) experiments, thus avoiding the preparation and analysis of separate samples at each pH investigated.<sup>5–7</sup> Here, we demonstrate how common pH buffer components can be used to generate tuneable, wide-range pH gradients spanning six or more units in standard NMR sample tubes. Adjustment of the buffer composition, guided by simple calculations, allows tuning of the pH gradient to span a wide range, or a narrow range covering acidic or basic pH as desired. Using this approach, we accurately determine all three  $pK_a$  values of histidine in one experiment, the side-chain and amino  $pK_a$  values

of tyrosine and lysine, and the side-chain and carboxyl  $pK_a$  values of aspartic acid. Polyprotic molecules are especially difficult to analyse using conventional NMR titrations due to the large number of pH values (points) required.<sup>8–10</sup> Finally, using substituted 4-vinylpyridines, we demonstrate how our method can be used to optimise the reactivity of families of organic molecules where protonation or deprotonation is a key mechanistic step. We compare our experimental  $pK_a$  values to site specific micro- $pK_a$  values predicted by machine learning (MolGp $K_a$ <sup>11</sup> and Epik 6.8<sup>12</sup>) and DFT (Jaguar  $pK_a$ <sup>13–15</sup>) methods. For monobasic pyridine derivatives, these computational approaches return predictions in reasonable agreement (<1 unit) with the NMR-derived values. However, significant differences are observed for pyridine derivatives with more than one basic site both in terms of the absolute  $pK_a$  values and relative basicity of the sites. These results highlight the necessity for the experimental determination of the  $pK_a$  for polyprotic compounds during synthesis campaigns. Our method can be applied using automatic NMR equipment and provides a useful addition to the organic chemist's analytical toolbox.

## 2. Materials and methods

### 2.1. Materials

All reagents were purchased from commercial suppliers and used as received. Milli-Q water (18.2 MΩ cm) was used throughout the study. Amino acids were analysed at a concen-

<sup>a</sup>School of Chemistry, Pharmacy and Pharmacology, University of East Anglia, Norwich Research Park, Norwich, NR4 7TJ, UK. E-mail: matthew.wallace@uea.ac.uk, chris.richards@uea.ac.uk

<sup>b</sup>Department of Chemistry, University of Liverpool, Crown Street, Liverpool, L69 7ZD, UK

<sup>c</sup>Iksuda Therapeutics Ltd., The Biosphere, Draymans Way, Newcastle Helix, Newcastle upon Tyne, NE4 5BX, UK

† Electronic supplementary information (ESI) available. See DOI: <https://doi.org/10.1039/d5an00406c>

tration of 2 mM. The 4-vinylpyridine derivatives **A–M** were synthesised in house as described in section S11 of the ESI.† Aqueous solutions of these pyridines for analysis were prepared by weighing sufficient compound into a 5 mL volumetric flask to give a theoretical concentration of 2 mM assuming complete dissolution upon stirring with a PTFE stirrer bar for at least one hour. Analysis was performed on the supernatant solution.

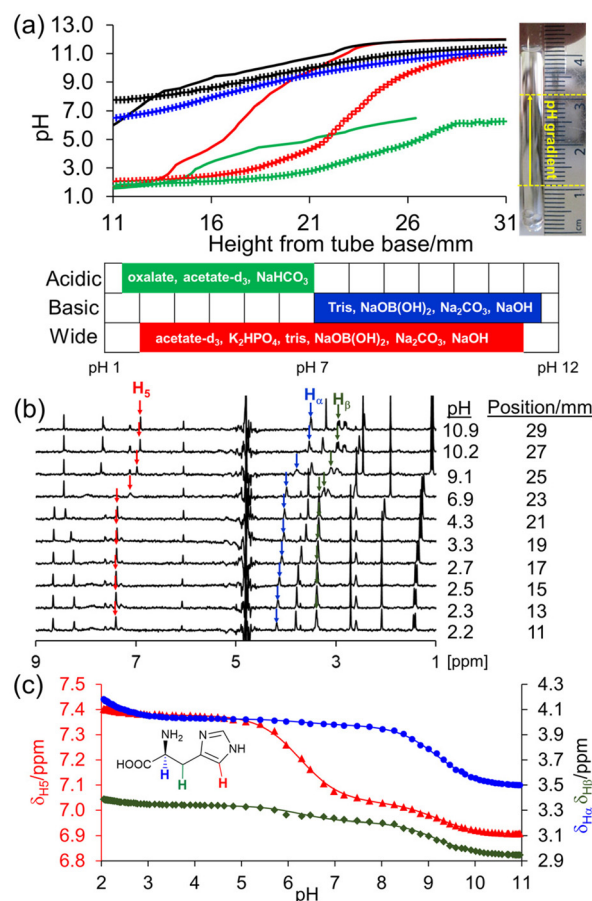
## 2.2. Creation of pH gradients

The following indicator compounds were included in the samples to allow determination of the pH between 1 and 12 *via* their  $^1\text{H}$  chemical shifts:<sup>5</sup> 2 mM sodium acetate, disodium methylphosphonate (MPA), sodium glycinate, methylammonium chloride, sodium dichloroacetate (DCA), 4 mM sodium formate, and 1 mM 2,6-lutidine. Sodium 3-(trimethylsilyl)-1-propanesulfonate (DSS, 0.2 mM) was used as a chemical shift reference for all spectra. The set of indicators used was adjusted depending on the pH range under assessment. To access more basic pH ranges (lysine and tyrosine, Fig. 1a), DCA, formate, acetate and 2,6-lutidine were excluded. Methylammonium chloride was excluded from experiments used to assess more acidic ranges (aspartic acid, Fig. 1a). For analysis of 4-vinylpyridine derivatives, *N*-hydroxysuccinimide (NHS, 2 mM) was substituted for 2,6-lutidine to avoid spectral overlap with the pyridine resonances.

To prepare a pH gradient, crystals of oxalic acid dihydrate were weighed directly into a Wilmad 528-PP 7" NMR tube. The nominal masses were within 20% of the stated values. Four 2 mm diameter glass beads were then placed in the tube. Finally, a solution of pH buffers, analyte and indicators was placed on top of the glass beads using a 9" NMR Pasteur pipette to a height of 40–45 mm above the absolute base of the NMR tube. The sample was then stood at 22 °C in the auto-sampler rack of the spectrometer prior to analysis.

## 2.3. NMR analysis and data processing

All experiments were performed on a Bruker 500 MHz Avance III spectrometer equipped with a Bruker Ascend superconducting magnet in 100%  $\text{H}_2\text{O}$ , without lock, at 298 K to enable direct comparison with published  $\text{pK}_\text{a}$  values. However, we note that CSI experiments on pH gradients can be performed with lock on a range of solvent systems.<sup>7</sup>  $^1\text{H}$  spectra of a sample recorded before and after a CSI experiment confirmed minimal drift (<1 Hz) in the frequency of  $\text{H}_2\text{O}$ .  $^1\text{H}$  CSI datasets were acquired using a gradient phase encoding sequence with water suppression (section S14.3†). The delay between successive pulses in the selective pulse train was set at 333  $\mu\text{s}$  corresponding to a 3000 Hz separation between the null points. The phase encoding gradient pulse was in the form of a smoothed square and was ramped according to the number of rows collected in the dataset. Amino acids, compounds **A–F**, **K–L**:  $-25$  to  $25\text{ G cm}^{-1}$  in 128 steps, 8 scans at each increment, giving a theoretical spatial resolution of 0.20 mm along the *z*-axis. Compounds **G–J**:  $-18.8$  to  $18.8\text{ G cm}^{-1}$  in 64 steps, 16 scans, 0.41 mm resolution. Compound **M**:  $-5.0$  to  $5.0\text{ G cm}^{-1}$  in 16



**Fig. 1** (a) Predicted (solid line) and experimental (cross) pH versus position from the base of the NMR tube for determination of the  $\text{pK}_\text{a}$  values of tyrosine ( $\text{OH}$ ,  $\text{NH}_2$ : blue), lysine ( $\text{NH}_2$ : black), histidine ( $\text{NH}_2$ , imidazole,  $\text{COOH}$ : red) and aspartic acid ( $\text{COOH}$ : green). Photograph of NMR sample with NMR-active region indicated as yellow dashed lines. Sketch of pH range accessed with each buffer system. (b) Example  $^1\text{H}$  spectra of histidine extracted from CSI image at pH value and vertical position indicated. (c) Plot of  $^1\text{H}$  chemical shift of  $\alpha$  (blue circle),  $\beta$  (green diamond) and  $\text{H}_5$  proton (red triangle) of histidine versus pH. The solid lines are fits to eqn (1).

steps, 64 scans, 1.6 mm resolution. The signal acquisition time and relaxation delay were 1 s and 0.26 s, respectively giving a total experimental time of 22 minutes 31 seconds for all experiments. A spoil gradient (1 ms,  $27\text{ G cm}^{-1}$ ) was employed at the end of the acquisition period to destroy any remaining transverse magnetization.

CSI datasets were processed in phase-sensitive mode with 32 K points and an exponential line broadening factor of 3 Hz. Each row was phased, baseline corrected, referenced to DSS (0 ppm) and the chemical shifts of indicators and analytes exported to the spreadsheet accompanying this work using automation scripts written in house (section S14†). Only rows 20–120 (128 point dataset) were used in the analysis to avoid off-coil effects. The pH was determined from the  $^1\text{H}$  chemical shifts of organic indicator molecules and ions (sections S1 and S3†). Fitting of the analyte chemical shifts to eqn (1) was

accomplished using the Solver module of Microsoft Excel optimising all variables, with the constraint for histidine that the fitted  $\delta_{\text{H3}}$  was less than 1 ppm higher than the lowest experimental chemical shift measured in the data series.

### 3. Results and discussion

#### 3.1. Design of pH gradients and data analysis

pH gradients are created by placing a basic solution on top of solid crystals of oxalic acid dihydrate. This acid is particularly convenient owing to its absence of  $^1\text{H}$  NMR resonances, non-deliquescent crystals that enable direct weighing into the NMR sample tube, and high acid strength that enables access to a wide pH range (pH 2–12). The spreadsheet accompanying this work enables selection of the mass of oxalic acid, buffer composition and time to leave the sample following preparation,  $t_{\text{opt}}$ , so that the user can create pH gradients spanning any desired range. The theory and calculations underpinning the spreadsheet are described in section S2 of the ESI† The buffers comprise inorganic phosphate, carbonate and borate species along with the inexpensive deuterated organic buffers, tris(hydroxymethyl- $\text{d}_3$ )amino- $\text{d}_2$ -methane (tris) and acetate- $\text{d}_3$ . Combining these buffers in different proportions permits access to different pH ranges, based on the  $\text{pK}_{\text{a}}$  of the protonated buffer species,  $\text{pK}_{\text{a,H}}$  (Table 1). The pH of the sample at each position is determined from the  $^1\text{H}$  chemical shifts of organic indicator molecules as described in our previous work and section S3 of the ESI†<sup>5</sup>

This highly tuneable pH range allows us to determine the  $\text{pK}_{\text{a}}$  values of molecules that possess several ionisable groups. Here, we use amino acids bearing ionisable side chains as exemplar systems. To extract the three  $\text{pK}_{\text{a}}$  values of histidine in a single experiment, we require a smooth pH gradient spanning approximately pH 2 to pH 11. We choose a buffer system where the  $\text{pK}_{\text{a,H}}$  values of the buffers are separated by approximately two units or less: NaOH,  $\text{Na}_2\text{CO}_3$ ,  $\text{NaOB}(\text{OH})_2$ , tris,  $\text{K}_2\text{HPO}_4$  and acetate- $\text{d}_3$ . With 3.7 mg of oxalic acid placed at the bottom of the NMR tube and with 10 mM of each buffer component, the concentration of oxalic acid is predicted to vary from 110 mM to 2 mM from the bottom to the top of the NMR active region of the sample (Fig. 1a), 6.5 hours after

preparation, giving a pH gradient spanning the required range. We note that this calculated  $t_{\text{opt}}$  can vary considerably between samples due to differences in the identity and concentration of buffers selected, the pH range targeted, and the mass of acid weighed into the sample tube. However, there is a *ca.* 2-hour window either side of the calculated  $t_{\text{opt}}$  over which acceptable  $\text{pK}_{\text{a}}$  values are obtained (section S4†).

Resolution of the two  $\text{pK}_{\text{a}}$  values of aspartic acid associated with the COOH groups requires a pH gradient spanning pH 1 to pH 7. We choose buffer components with lower  $\text{pK}_{\text{a,H}}$  values (10 mM  $\text{NaHCO}_3$ , acetate- $\text{d}_3$  and 20 mM oxalate), 3.4 mg of oxalic acid and leave the sample seven hours before analysis. Similarly, resolution of the  $\text{pK}_{\text{a}}$  values of lysine ( $\text{NH}_2$ ) and tyrosine ( $\text{NH}_2$ , Ar-OH) requires pH gradients spanning pH 6 to pH 12. This alkaline pH range is accomplished using more basic buffer components (0.01 M NaOH, 20 mM  $\text{Na}_2\text{CO}_3$ , tris and  $\text{NaOB}(\text{OH})_2$ ), a smaller mass of oxalic acid (1.3 mg), and leaving the sample 8.5 hours before analysis.

When prepared and analysed by CSI, the pH exhibits a smooth variation over the desired pH ranges, as predicted by our spreadsheet (Fig. 1a). We attribute the discrepancy between the predicted and experimental pH values to the coupled diffusion of acid and base along the NMR tube and the  $\pm 0.3$  mg uncertainty in the mass of oxalic acid used in the experiments (section S6, ESI†).

The  $^1\text{H}$  NMR chemical shifts of the pH indicator molecules and amino acids were extracted from the CSI datasets using the routines and scripts provided in section S14 of the ESI† This data was then exported to the spreadsheets accompanying this work and fitted to extract the pH of each row. The  $\text{pK}_{\text{a}}$  values of the amino acids were obtained by fitting the observed chemical shift,  $\delta_{\text{obs}}$ , of each observed site on the molecule to eqn (1):

$$\delta_{\text{obs}} = \frac{\delta_{\text{L3}} + \delta_{\text{HL2}} 10^{\text{pK}_3 - \text{pH}} + \delta_{\text{H2L}} 10^{\text{pK}_3 + \text{pK}_2 - 2\text{pH}} + \delta_{\text{H3}} 10^{\text{pK}_3 + \text{pK}_2 + \text{pK}_1 - 3\text{pH}}}{1 + 10^{\text{pK}_3 - \text{pH}} + 10^{\text{pK}_3 + \text{pK}_2 - 2\text{pH}} + 10^{\text{pK}_3 + \text{pK}_2 + \text{pK}_1 - 3\text{pH}}} \quad (1)$$

where  $\delta_{\text{L3}}$ ,  $\delta_{\text{HL2}}$ ,  $\delta_{\text{H2L}}$  and  $\delta_{\text{H3}}$  are the fitted limiting chemical shifts of the fully deprotonated, monoprotonated ( $\text{pK}_3$ ), diprotonated ( $\text{pK}_2$ ) and fully protonated ( $\text{pK}_1$ ) states, respectively.  $\delta_{\text{L3}}$ ,  $\delta_{\text{HL2}}$ ,  $\delta_{\text{H2L}}$  and  $\delta_{\text{H3}}$  of each observed resonance and the  $\text{pK}_{\text{a}}$  values were treated as free variables in the fitting. Only two  $\text{pK}_{\text{a}}$  values and three limiting shifts were fitted for lysine, tyrosine and aspartic acid as high pH-resolution was required over the alkaline and acidic pH ranges, respectively (section S4 of the ESI†). The fitted  $\text{pK}_{\text{a}}$  values,  $\text{pK}_n$ , were converted to their thermodynamic values ( $I = 0$ ),  $\text{pK}_{n,0}$ , using eqn (2).<sup>5,16</sup>

$$\text{pK}_{n,0} = \text{pK}_n - \Delta z^2 \left[ 0.51 \frac{\sqrt{I}}{1 + \sqrt{I}} - 0.1I \right] \quad (2)$$

where  $\Delta z^2$  is the difference in the square of the overall charge of the molecule between the protonated and deprotonated sides of each protonation equilibrium ( $z_{\text{Hn}}^2 - z_{\text{H(n-1)}}^2$ ). The  $\text{pK}_{n,0}$  values obtained agree with literature values within experimental uncertainty (Table 2).

**Table 1** pH buffer components used in this work

| Buffer                     | $\text{pK}_{\text{a,H}}^a$ | pH range <sup>b</sup> |
|----------------------------|----------------------------|-----------------------|
| NaOH                       | —                          | >11                   |
| $\text{Na}_2\text{CO}_3$   | 10                         | 9.2–10.8              |
| $\text{NaOB}(\text{OH})_2$ | 9                          | 8.2–10.1              |
| Tris <sup>c</sup>          | 8                          | 7.2–9.0               |
| $\text{K}_2\text{HPO}_4$   | 7                          | 6.1–7.9               |
| $\text{NaHCO}_3$           | 6                          | 5.4–7.3               |
| Acetate                    | 4.7                        | 3.8–5.7               |
| Oxalate <sup>2-</sup>      | 4                          | 2–5 <sup>d</sup>      |

<sup>a</sup>  $\text{pK}_{\text{a}}$  of protonated buffer species. <sup>b</sup> pH range over which buffer is between 90% and 10% protonated. <sup>c</sup> Tris(hydroxymethyl- $\text{d}_3$ )amino- $\text{d}_2$ -methane. <sup>d</sup> pH range accessible using oxalic acid/oxalate alone.



**Table 2** Thermodynamic  $pK_a$  values of amino acids measured using CSI

| Amino acid                    | $pK_{1,0}$   | $pK_{2,0}$  | $pK_{3,0}$                                    | pH range <sup>a</sup> |
|-------------------------------|--|---|---|-----------------------|
| Histidine<br>$I = 0.12$ M     | $1.39 \pm 0.14$ (1.56)<br>$\Delta z^2 = 3$               | $6.12 \pm 0.12$ (6.00)<br>$\Delta z^2 = 1$                | $9.28 \pm 0.04$ (9.25)<br>$\Delta z^2 = -1$   | 2.0–11.1              |
| Lysine<br>$I = 0.13$ M        | —  | $9.12 \pm 0.04$ (9.09)<br>$\Delta z^2 = 1$                | $10.90 \pm 0.04$ (10.90)<br>$\Delta z^2 = -1$ | 7.7–11.4              |
| Tyrosine<br>$I = 0.13$ M      | —  | $9.22 \pm 0.08$ (9.09)<br>$\Delta z^2 = -1$               | $10.59 \pm 0.05$ (10.69)<br>$\Delta z^2 = -3$ | 6.5–11.2              |
| Aspartic acid<br>$I = 0.10$ M | $2.12 \pm 0.08$ (1.99 <sup>b</sup> )<br>$\Delta z^2 = 1$ | $3.98 \pm 0.03$ (3.89 <sup>b</sup> )<br>$\Delta z^2 = -1$ | —   | 1.8–6.3               |

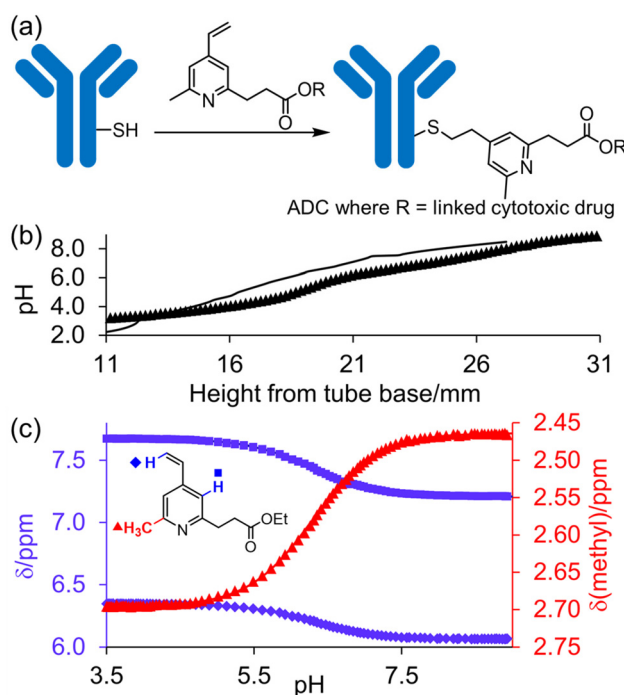
Literature values are provided in parentheses, taken from ref. 16 unless otherwise stated. <sup>a</sup> pH range accessed in CSI experiment (Fig. 1a). <sup>b</sup> Ref. 17. Uncertainties were calculated as described in section S5 of the ESI.†

Negligible variation in these  $pK_a$  values ( $\leq 0.05$  units) is observed upon repetition of the CSI experiments (section S6†) while acquisition of CSI images  $\pm 2$  hours around  $t_{\text{opt}}$  causes  $< 0.1$  units variation in the fitted  $pK_a$  values (section S4†). <sup>1</sup>H spectra are provided in section S7 of the ESI.† We note that the accuracy and reproducibility of CSI experiments are comparable to potentiometric titration and consume less compound (section S3.3).

### 3.2. Application to linker group modification

The  $pK_a$  value determines the charge and reactivity of a molecule. Nevertheless, the high costs of conventional  $pK_a$  measurements in terms of time, labour and sample quantity can discourage  $pK_a$  determination during molecular design campaigns. Important chemical insights can thus be missed, and time wasted on the synthesis of ineffective compounds. Antibody–drug conjugates (ADCs) are a molecular construct whereby a drug is appended to an antibody for targeted therapeutic treatments, notably oncology.<sup>18</sup> A reaction is required to attach a linker group to the antibody, with the soft nucleophilic thiol moiety of cysteine being the most targeted functional group due to its chemoselective reactivity with several electrophilic Michael acceptors.<sup>19</sup> Maleimide is one of the most frequently employed, and is utilised in several clinically approved ADCs, but suffers from deconjugation by retro-Michael reaction.<sup>20</sup> Selective alkylation of cysteine groups in proteins has been achieved with 2- and 4-vinylpyridines,<sup>21</sup> and 4-vinylpyridine based PermaLink® is an effective linker unit for ADC development which does not experience reversible addition with thiols (Fig. 2).<sup>22,23</sup> Related systems developed subsequently include vinylpyrimidines,<sup>24</sup> divinylpyrimidines (for disulfide rebridging)<sup>25</sup> and *N*-methylated 2-vinylpyridinium salts.<sup>26</sup>

We sought to synthesise several 4-vinylpyridine derivatives to which can be linked two cytotoxic payloads as a means to improve the drug-antibody ratio (DAR) of a resulting ADC, with this study providing also the opportunity to investigate the relationship between vinylpyridine  $pK_a$  and the rate of a model bioconjugation reaction. Synthesis yielded single-arm compounds A–E, dual-armed compounds F–K containing functionality linked additionally to position 6, and also two hetero-



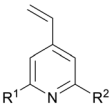
**Fig. 2** (a) Reaction linking a 4-vinylpyridine derivative to a thiol unmasked antibody. (b) Predicted (solid line) and experimental (cross) pH versus position from the base of the NMR tube for determination of  $pK_a$  value of A. (c) Plot of <sup>1</sup>H chemical shifts of A (2 mM) versus pH with fitted  $pK_{a,0}$  of  $6.18 \pm 0.09$ .

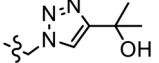
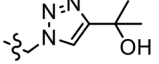
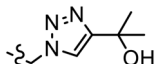
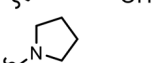
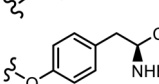
atom-substituted compounds L and M for additional reactivity comparison (Table 3).

Starting with compound A, we assumed a  $pK_a$  value of approximately 6. We chose a buffer system comprising 20 mM  $\text{NaHCO}_3$ , 10 mM tris,  $\text{K}_2\text{HPO}_4$  and acetate. 1.5–2 mg of oxalic acid was predicted to give a gradient between pH 2 and pH 9 with a  $t_{\text{opt}}$  of 8 hours. The  $pK_a$  of twelve other variously substituted 4-vinyl pyridines were then determined. For comparison, micro- $pK_a$  values were predicted using three computational packages: Jaguar  $pK_a$  (Schrödinger Release 2024-2: Schrödinger, LLC, New York, NY, 2024),<sup>13–15</sup> Epik version 6.8<sup>12</sup> and MolGp $K_a$ <sup>11</sup> (Table 3). The values returned by these calculations were used to guide the selection of the buffer systems





Table 3  $pK_a$  and rate analysis of substituted 4-vinylpyridines


| Entry | R <sup>1</sup>                                     | R <sup>2</sup>   | $k_{\text{obs}}/\text{M}^{-1} \text{min}^{-1}$ | $pK_{a,0} \text{ NMR}$   | $pK_{a,0} \text{ Epik}$ | $pK_a, \text{ MolGpK}_a$ | $pK_a, \text{ Jaguar}$  |
|-------|--|--|--|--------------------------|-------------------------|--------------------------|-------------------------|
| A     | Me   | (CH <sub>2</sub> ) <sub>2</sub> CO <sub>2</sub> Et                                 | 0.80   | 6.18 ± 0.09              | 6.62                    | 5.9                      | 6.40                    |
| B     |  | CH <sub>2</sub> CH(CH <sub>2</sub> CO <sub>2</sub> Et)CO <sub>2</sub> Et           | 0.17   | 5.69 ± 0.09              | 6.25                    | 5.5                      | 5.49                    |
| C     |  | CH <sub>2</sub> N(CH <sub>2</sub> CO <sub>2</sub> Et) <sub>2</sub>                 | 0.13   | 5.86 ± 0.07              | 5.50, 3.91 <sup>a</sup> | 4.7, 4.6 <sup>a</sup>    | 7.29, 0.95 <sup>a</sup> |
| D     |  | (CH <sub>2</sub> ) <sub>2</sub> N(CH <sub>2</sub> CO <sub>2</sub> Et) <sub>2</sub> | 0.39   | 6.69 ± 0.06, 1.4 ± 0.07  | 6.48, 3.95 <sup>a</sup> | 6.0, 5.1 <sup>a</sup>    | 7.38, 2.43 <sup>a</sup> |
| E     |  | (CH <sub>2</sub> ) <sub>3</sub> N(CH <sub>2</sub> CO <sub>2</sub> Et) <sub>2</sub> | 0.75   | 6.59 ± 0.04, 3.12 ± 0.37 | 6.87, 4.39 <sup>a</sup> | 6.3, 5.5 <sup>a</sup>    | 6.75, 3.58 <sup>a</sup> |
| F     | (CH <sub>2</sub> ) <sub>2</sub> CO <sub>2</sub> Et | (CH <sub>2</sub> ) <sub>2</sub> CO <sub>2</sub> Et                                 | 0.17   | 5.36 ± 0.04              | 5.39                    | 5.2                      | 5.89                    |
| G     |  |   | 0.0054   | 2.98 ± 0.05              | 3.37, 2.84 <sup>b</sup> | 4.5, 3.0 <sup>b</sup>    | 3.27                    |
| H     |  | (CH <sub>2</sub> )NHBoc  | 0.083  | 4.74 ± 0.03              | 4.57                    | 4.7                      | 4.47                    |
| I     | (CH <sub>2</sub> )NHBoc                            | (CH <sub>2</sub> )NHBoc  | 0.087  | 4.11 ± 0.02              | 2.55                    | 4.4                      | 5.39                    |
| J     |  |   | 0.0022   | 2.39 ± 0.02              | 1.97, 2.84 <sup>b</sup> | 4.2, 3.2 <sup>b</sup>    | 3.86                    |
| K     |  |   | n.d.   | 0.92 ± 0.15 <sup>c</sup> | 0.69, 2.84 <sup>b</sup> | 4.1, 3.1 <sup>b</sup>    | 1.22                    |
| L     | H  |   | 0.084 <sup>28</sup>                            | 7.60 ± 0.07              | 7.35, 7.27 <sup>d</sup> | 6.7, 5.4 <sup>d</sup>    | 7.60                    |
| M     |  |   | 0.0048   | 2.49 ± 0.49 <sup>e</sup> | 2.42                    | 3.1                      | 1.62                    |

<sup>a</sup>  $pK_{a,H}$  of tertiary amine. <sup>b</sup>  $pK_{a,H}$  of 1,2,3-triazole moiety. <sup>c</sup> Additional datapoint measured in homogeneous solution of 0.1 M HCl (Fig. S27†). <sup>d</sup> Reported  $pK_{a,H}$  of pyrrolidine moiety. <sup>e</sup> 16 point CSI experiment recorded due to low solubility.

for the NMR experiments. For example, the  $pK_a$  of compound **J** was predicted to be approximately 3 (Table 3). Hence a more acidic buffer system was used relative to **A**, comprising 20 mM oxalate, 10 mM acetate and NaHCO<sub>3</sub> and 3 mg of oxalic acid was used. The resulting pH gradient was predicted to span from pH 1.5 to 7 with a  $t_{\text{opt}}$  of 6 hours. Example predicted and experimental gradients for each compound are provided in section S8 of the ESI.†

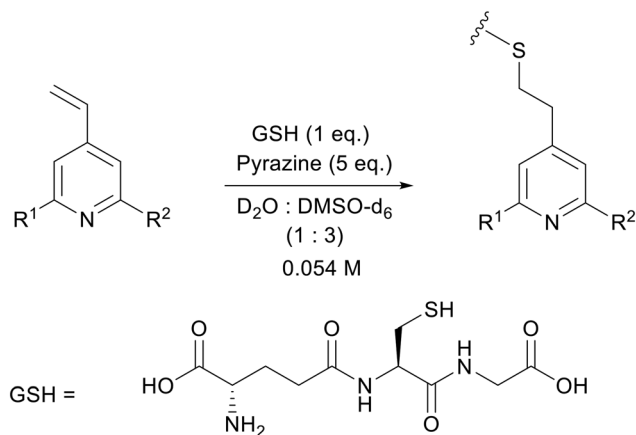
For compounds **D** and **E** additional  $pK_a$  values of 1.4 and 3.1 were obtained that can be associated with the tertiary amine functionality based on the pH-dependence of the chemical shifts of different parts of the molecule (section S9, ESI†). A  $pK_a$  value of 3.3 was determined for the analogous compound PhCH<sub>2</sub>N(CH<sub>2</sub>CO<sub>2</sub>Et)<sub>2</sub>, confirming the low basicity of the tertiary amine (Fig. S30, ESI†). DFT calculations confirmed a higher basicity of the pyridine in compounds **C–E** (section S10, ESI†). The  $pK_a$  values of the amine sites presented in Table 3 (Jaguar  $pK_a$ ) were calculated explicitly with the pyridine in the protonated state and are thus closer to the experimentally observed macroscopic values. However, the  $pK_a$  values returned by MolGpK<sub>a</sub><sup>11</sup> and Epik 6.8<sup>12</sup> consider the ionisation of each site in isolation (micro- $pK_a$ ) without regard for the effect of the ionisation of neighbouring groups. The  $pK_a$  values reported by these methods for the amine sites are thus higher than the second (lower) macroscopic  $pK_a$  observed. That no additional values were obtained for the triazole containing compounds (**G**, **J** and **K**) is consistent with the low basicity

city of this heterocycle ( $pK_a$  for 1-methyl-1,2,3-triazole 1.25).<sup>27</sup> However, the  $pK_a$  values of the triazole moiety calculated using Epik and MolGpK<sub>a</sub> are significantly higher, again highlighting the limitations of these predictors when applied to compounds that have more than one protonation site. The 2-pyrrolidinopyridine component of **L** may be regarded as a single functional group, giving this compound the highest basicity of the compounds studied.

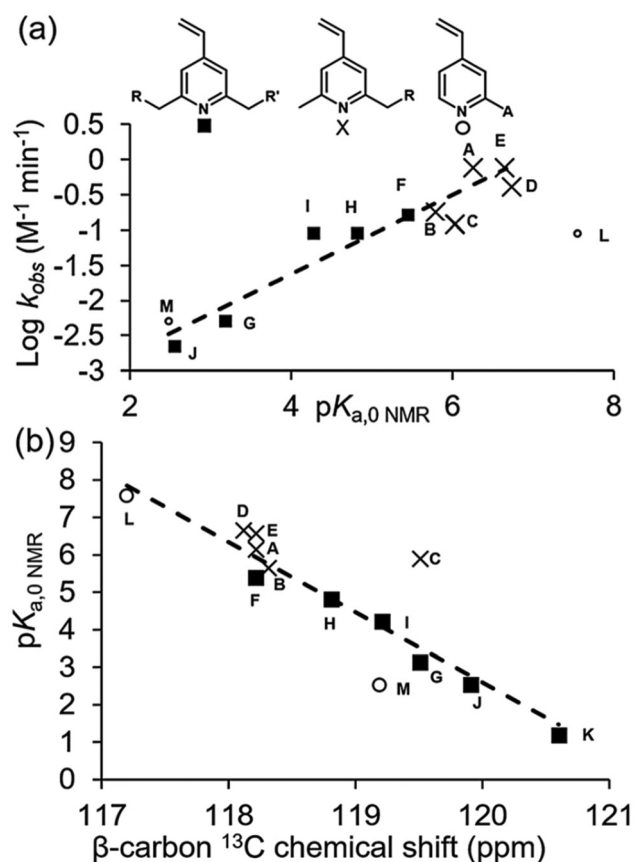
As a model bioconjugation reaction we employed a reaction of each vinylpyridine with glutathione (GSH, Scheme 1). This cysteine containing tripeptide is a non-volatile crystalline solid making it relatively easy to handle, and it has been used previously as a model for reactivity studies on electrophilic covalent inhibitors of cysteine-containing proteins.<sup>28–30</sup> One equivalent of glutathione was added as a solution in D<sub>2</sub>O (0.22 M) to the vinylpyridine in DMSO-*d*<sub>6</sub> (0.073 M), with pyrazine included as an internal standard for integration. The reaction was monitored using <sup>1</sup>H NMR spectroscopy and the rates of this second order addition<sup>25,26</sup> are given in Table 3. Synthetic schemes for the preparation of **A–M** and example kinetic plots are provided in sections S11 and S12 of the ESI,† respectively.

A plot of the log of rate *versus* the  $pK_a$  value determined by NMR exhibits a strong positive correlation (Fig. 3a) consistent with a lowering of the activation energy barrier and an acceleration of reaction rate when the pyridine nitrogen bears a positive charge, as reported for the reaction of *N*-acetylcysteineamide with quaternized 2-vinylpyridine.<sup>26</sup>





**Scheme 1** Reaction of 4-vinylpyridines with glutathione (GSH) for the determination of rate as a function of  $R^1$  and  $R^2$ .



**Fig. 3** (a) Plot of  $\log k$  vs.  $pK_{a,0}$  NMR for 4-vinylpyridine derivatives exhibiting linear correlation. Compound L (Table 3) is plotted separately. (b) Plot of  $pK_{a,0}$  NMR versus  $^{13}\text{C}$  NMR chemical shift ( $\text{CDCl}_3$ ) of  $\beta$ -carbon of vinyl group.

Three different chain lengths were used to attach an iminodiacetate unit to the 2-position [ $(\text{CH}_2)_n$ , where  $n = 1-3$ , entries C–E] and an increase in chain length results in an increase in rate such that with a three-carbon spacer the electronegativity

of the nitrogen in **E** is essentially negated. The other electron-withdrawing functional groups are attached with either a two-carbon (**B**) or one-carbon spacer (**G** and **H**) which significantly diminishes the rate, especially when attached to both positions 2 and 6 (**I–K**). Similarly, replacement of the methyl group in **A** with an ester functionality reduces the reactivity (**F**). The conjugated 2-pyrroldine nitrogen in **L** significantly decreases the reactivity relative to **A**, despite **L** being the most basic of all the compounds studied as it is now too electron rich to react at a significant rate. The 2-oxygen substituent of **M** significantly reduces pyridine basicity, shutting down reactivity. Thus, a functionalised alkyl chain is the optimum substituent such that compounds **A**, **D** and **E** are in the ‘Goldilocks zone’ with respect to pyridine basicity and the enhanced electrophilicity of the protonated species.

To confirm that protonation of the pyridines took place under the conditions of the reaction, the  $^1\text{H}$  chemical shift of the methyl resonance of **A** was measured in  $\text{D}_2\text{O}/\text{DMSO}-d_6$  solutions containing 20 mM HCl (**A** fully protonated, 2.6074 ppm) and in the absence of acidic species (**A** non-protonated, 2.3548 ppm). Comparison to the chemical shift measured in the presence of GSH (2.3732 ppm) revealed  $7 \pm 2\%$  of **A** was protonated in the presence of GSH under the conditions of the reaction.<sup>31</sup> A full discussion of this calculation, including uncertainty, is provided in section S12 of the ESI.† We note that due to the acidic nature of GSH, an appreciable percentage protonation ( $>3\%$ ) is anticipated for vinylpyridines with  $pK_{a,0}$  NMR  $> 5.3$ , while the pH is expected to be weakly basic (Fig. S38†). An acceleration of the rate of reaction of *N*-acetylcysteineamide with 2-vinylpyridine has been reported under weakly acidic conditions in  $\text{H}_2\text{O}$  (pH 5.7) relative to neutral buffer (pH 7.6).<sup>26</sup> The acidity of GSH therefore aids the reaction under our experimental conditions, with a higher degree of protonation and thus a faster rate of reaction observed for the more basic vinylpyridines.

The correlation between the  $^{13}\text{C}$  NMR chemical shift of the  $\beta$ -carbon of *para*-substituted styrenes and the Hammett substituent constant ( $\sigma_p$ ) has previously been noted.<sup>32</sup> Accordingly, a correlation is observed between the  $pK_a$  values of **A–M** and the  $^{13}\text{C}$  NMR chemical shift of the electrophilic  $\beta$ -vinyl carbon (Fig. 3b). This correlation suggests that  $^{13}\text{C}$  NMR can be used to predict the  $pK_a$  values of substituted 4-vinylpyridines and thus estimate their reactivity towards cysteine bioconjugation.

## 4. Conclusions

In conclusion, a rapid and operationally simple method for the  $pK_a$  determination of organic molecules in buffered aqueous media has been developed for a wide range of pHs (*ca.* 2–11). As proof of principle, the thermodynamic  $pK_a$  values of four amino acids were ascertained using this method to obtain values closely matching literature values. As demonstration of application, the  $pK_a$  values of a range of substituted 4-vinylpyridine linker compounds were also determined which revealed a strong correlation between  $pK_a$  and reaction rate.



We are also able to rationalise other correlations, such as between observed  $^{13}\text{C}$  NMR chemical shifts and rates of reactions. Our approach makes experimental measurements of  $\text{pK}_\text{a}$  accessible and feasible to perform during the routine analysis of newly synthesised organic compounds by NMR.

## Author contributions

MW, TES and CJR conceived the project. MW and MN designed and performed the CSI experiments. JS prepared compounds **A–M** and performed the kinetic analysis, under the supervision of JT, PLKH, GRS, CJR and TES. KB performed the computational calculations of  $\text{pK}_\text{a}$  under the supervision of JAI. MW, TES and CJR wrote the manuscript, with suggestions from all other authors.

## Data availability

Additional data supporting this article (experimental procedures, synthetic details, stacked NMR plots) are available in the ESI.† A spreadsheet for prediction of pH gradients and processing of CSI data is also provided as ESI.† Raw CSI data files will be openly available at: <https://research-portal.uea.ac.uk/en/datasets/>.

## Conflicts of interest

There are no conflicts to declare.

## Acknowledgements

MW thanks UKRI for a Future Leaders Fellowship (MR/T044020/1) and the Royal Commission for the Exhibition of 1851 for a Research Fellowship. Iksuda Therapeutics Ltd and the MRC are acknowledged for funding (JS). KB thanks AstraZeneca (Grant 10045297) and the University of Liverpool for financial support. We are grateful for use of the University of East Anglia (UEA) Faculty of Science NMR facility. We thank Schrödinger for a trial licence of their Small Molecule Drug Discovery Suite. We thank Mestrelab Research S.L. for technical advice and the gift of a software license for Mnova 14.2.

## References

- 1 J. Bezençon, M. B. Wittwer, B. Cutting, M. Smieško, B. Wagner, M. Kansy and B. Ernst, *J. Pharm. Biomed. Anal.*, 2014, **93**, 147–155.
- 2 D. R. Holycross and M. Chai, *Macromolecules*, 2013, **46**, 6891–6897.
- 3 D. Kohda, T. Sawada and F. Inagaki, *Biochemistry*, 1991, **30**, 4896–4900.
- 4 R. Liang, J. Samanta, B. Shao, M. Zhang, R. J. Staples, A. D. Chen, M. Tang, Y. Wu, I. Aprahamian and C. Ke, *Angew. Chem., Int. Ed.*, 2021, **60**, 23176–23181.
- 5 M. Wallace, D. J. Adams and J. A. Iggo, *Anal. Chem.*, 2018, **90**, 4160–4166.
- 6 G. Schenck, K. Baj, J. A. Iggo and M. Wallace, *Anal. Chem.*, 2022, **94**, 8115–8119.
- 7 M. Wallace, N. Abiama and M. Chipembere, *Anal. Chem.*, 2023, **95**, 15628–15635.
- 8 G. Hägele, *Molecules*, 2019, **24**, 3238.
- 9 A. H. Alkhzem, T. J. Woodman and I. S. Blagbrough, *ACS Omega*, 2021, **6**, 2824–2835.
- 10 R. Fraczekiewicz and M. Waldman, *J. Chem. Inf. Model.*, 2023, **63**, 3198–3208.
- 11 X. Pan, H. Wang, C. Li, J. Z. H. Zhang and C. Ji, *J. Chem. Inf. Model.*, 2021, **61**, 3159–3165.
- 12 J. C. Shelley, A. Cholleti, L. L. Frye, J. R. Greenwood, M. R. Timlin and M. Uchimaya, *J. Comput. Aided Mol. Des.*, 2007, **21**, 681–691.
- 13 J. J. Klicic, R. A. Friesner, S. Y. Liu and W. C. Guida, *J. Phys. Chem. A*, 2002, **106**, 1327–1335.
- 14 A. D. Bochevarov, M. A. Watson, J. R. Greenwood and D. M. Philipp, *J. Chem. Theory Comput.*, 2016, **12**, 6001–6019.
- 15 H. S. Yu, M. A. Watson and A. D. Bochevarov, *J. Chem. Inf. Model.*, 2018, **58**, 271–286.
- 16 H. Nagai, K. Kuwabara and G. Carta, *J. Chem. Eng. Data*, 2008, **53**, 619–627.
- 17 J. Partanen, P. Juusola, P. Minkinen, P. Nielsen, A. Claesson, J. Arnarp, L. Björk and R. Gawinecki, *Acta Chem. Scand.*, 1995, **49**, 163–171.
- 18 C. Dumontet, J. M. Reichert, P. D. Senter, J. M. Lambert and A. Beck, *Nat. Rev. Drug Discovery*, 2023, **22**, 641–661.
- 19 P. Ochtrop and C. P. R. Hackenberger, *Curr. Opin. Chem. Biol.*, 2020, **58**, 28–36.
- 20 S. C. Alley, D. R. Benjamin, S. C. Jeffrey, N. M. Okeley, D. L. Meyer, R. J. Sanderson and P. D. Senter, *Bioconjugate Chem.*, 2008, **19**, 759–765.
- 21 R. Sebastiano, A. Citterio, M. Lapadula and P. G. Righetti, *Rapid Commun. Mass Spectrom.*, 2003, **17**, 2380–2386.
- 22 A. G. Watts, T. Kantner and A. B. Mackenzie, WO2010/070300A2, 2010.
- 23 A. D. Baxter, C. M. Birchall, D. J. Mansell, J. H. Mysliwy and J. Thirlway, WO2016/067021, 2016.
- 24 H. Seki, S. J. Walsh, J. D. Bargh, J. S. Parker, J. Carroll and D. R. Spring, *Chem. Sci.*, 2021, **12**, 9060–9068.
- 25 S. J. Walsh, S. Omarjee, W. R. J. D. Galloway, T. T. L. Kwan, H. F. Sore, J. S. Parker, M. Hyvönen, J. S. Carroll and D. R. Spring, *Chem. Sci.*, 2019, **10**, 694–700.
- 26 M. J. Matos, C. D. Navo, T. Hakala, X. Ferhati, A. Guerreiro, D. Hartmann, B. Bernardim, K. L. Saar, I. Compañón, F. Corzana, T. P. J. Knowles, G. Jiménez-Osés and G. J. L. Bernardes, *Angew. Chem., Int. Ed.*, 2019, **58**, 6640–6644.
- 27 A. Albert, *Phys. Methods Heterocycl. Chem.*, 1963, **1**, 98.



- 28 J. Andrew, PhD Thesis, The University of East Anglia, 2022.
- 29 V. J. Cee, L. P. Volak, Y. Chen, M. D. Bartberger, C. Tegley, T. Arvedson, J. McCarter, A. S. Tasker and C. Fotsch, *J. Med. Chem.*, 2015, **58**, 9171–9178.
- 30 A. Birkholz, D. J. Kopecky, L. P. Volak, M. D. Bartberger, Y. Chen, C. M. Tegley, T. Arvedson, J. D. McCarter, C. Fotsch and V. J. Cee, *J. Med. Chem.*, 2020, **63**, 11602–11614.
- 31 M. Wallace, K. Lam, A. Kuraite and Y. Z. Khimyak, *Anal. Chem.*, 2020, **92**, 12789–12794.
- 32 C. E. Aun, T. J. Clarkson and D. A. R. Happer, *J. Chem. Soc., Perkin Trans. 2*, 1990, 645–649.

