SUPPLEMENTARY INFORMATION

Sudhakaran Prabakaran*, Robert A. Everley, Isabelle Landrieu,
Jean-Michel Wieruszeski, Guy Lippens, Hanno Steen, Jeremy Gunawardena*†

*Department of Systems Biology, Harvard Medical School
200 Longwood Avenue, Boston, MA 02115, USA

†To whom correspondence should be addressed; e-mail: jeremy@hms.harvard.edu
For full author affiliations, please see the main paper.

Abstract

This file provides supplementary information for the paper *"Comparative analysis of Erk phosphorylation suggests a mixed strategy for measuring phospho-form distributions"*. It should be read in conjunction with the main paper.

Contents

1 SUPPLEMENTARY METHODS

1.1 Additional NMR resonances

The TEY motif in Erk is part of a longer TEYVAT sequence. The (A)T residue within this sequence can be assigned in the 3D HNCACB spectrum of the $^{13}C/^{15}N$-labelled **Erk-pTpY** sample through the characteristic C_α and C_β resonances of both the T and A residues, [1], (data not shown). The position of this resonance depends on the Y phosphorylation state because the peak is absent from both the **Erk-TY** and **Erk-pTY** samples. The **Erk-TpY** spectrum has a related peak but in a different position, suggesting that the peak in the **Erk-pTpY** spectrum represents the *pTpY* phospho-form. However, the peak integral is 10% larger than the C-terminal (G)Y peak, suggesting that it is a superposition of two or more unidentified peaks. At 950MHz, a minor peak partially detaches from the (A)T resonance (Figure 10 below) and integration of the remaining peak leads to a normalised value of 90% for the *pTpY* phospho-form in the **Erk-pTpY** sample and 0% in the **Erk-TpY** sample (Paper Figure 6A). The (A)T resonance is marked as (pTEpYVA)T in Paper Figure 5A.

We attempted to get closer to the pY residue by identifying the resonance corresponding to the (V)A residue. With prior knowledge of the C_α and C_β chemical shifts from the 3D HNCACB spectrum (data not shown), we found several potential candidates. The one marked as (pTEpYV)A in Paper Figure 5A is an alanine residue that also has a valine residue immediately N terminal. However, the Erk sequence has four (V)A dipeptide motifs (Paper Figure 1A), so to confirm its identification we first assigned all the residues of a *pTpY* synthetic phospho-peptide. Upon superposing the spectra of the **Erk-pTpY** sample with that of the *pTpY* synthetic phospho-peptide, the (V)A residue in the latter resonates at almost the identical position as the resonance mentioned above. The two preparations are not at the same pH and the shift in positions is consistent with the change in pH. This confirms the (V)A resonance as characteristic of the *pTpY* phospho-form. Integration of this peak relative to the C-terminal (G)Y resonance leads to a value of 88% in the **Erk-pTpY** sample and 6% in the **Erk-pTY** sample (Paper Figure 6A).

1.2 Calculating the four-site phospho-form distribution

The NMR and proMS analyses reveal two phosphorylation sites on **Erk-pTpY** in addition to the canonical phosphorylations on TEY. These additional sites were localised to S9 and S15 of His_6-tagged **Erk-pTpY** by multi-stage MS, as explained in the paper. With four phosphorylations, there are sixteen potential phospho-forms. The combination of proMS and pepMS, together with the effect of phosphotyrosine dephosphorylation of **Erk-pTpY** to yield **Erk-pTY**, provide considerable information about the four-site phospho-form distribution.

Figure 6B of the main paper lists the sixteen phospho-forms, with their proportions labelled $a_1, \cdots, a_{16}$. With this notation, the proMS data for **Erk-pTpY** in Paper Figure 4C, implies the following equations, where proportions are given as percentages.

$$
\begin{aligned}
a_{16} &= 10 && (1)\\
a_{12} + a_{13} + a_{14} + a_{15} &= 30 && (2)\\
a_6 + a_7 + a_8 + a_9 + a_{10} + a_{11} &= 60 && (3)\\
a_2 + a_3 + a_4 + a_5 &= 0 && (4)\\
a_1 &= 0 && (5)
\end{aligned}
$$

We infer from equations (5) and (4) that $a_1 = a_2 = a_3 = a_4 = a_5 = 0$ and from equation (1) that $a_{16} = 10$.

We now turn to the proMS data for **Erk-pTY** in Paper Figure 4C. **Erk-pTY** was made from **Erk-pTpY** using a tyrosine phosphatase that appears to be 100% efficient, as discussed in the paper. The effect of the phosphatase on the four-site phospho-form distribution can be determined by calculation and equated to the data in Paper Figure 4C, yielding the following equations.

$$
\begin{aligned}
a_{12} + a_{16} &= 5 && (6)\\
a_6 + a_7 + a_9 + a_{13} + a_{14} + a_{15} &= 30 && (7)\\
a_2 + a_3 + a_4 + a_8 + a_{10} + a_{11} &= 65 && (8)\\
a_1 + a_5 &= 0 && (9)
\end{aligned}
$$

Since $a_{16} = 10$, equation (6) suggests that $a_{12} = 0$, to within the experimental error of around 5%. Combining equation (2) with equation (7), we find that $a_6 = a_7 = a_9 = 0$. It follows from equation (7) or from equation (2) that, therefore,

$$a_{13} + a_{14} + a_{15} = 30 \quad (10)$$

Substituting some of the zero values just found into equation (3), we also find that

$$a_8 + a_{10} + a_{11} = 60 \quad (11)$$

while using the previous estimates of $a_2 = a_3 = a_4 = 0$ in equation (8), we find that

$$a_8 + a_{10} + a_{11} = 65 \quad (12)$$

Equations (11) and (12) are consistent to within experimental error.

The pepMS data for **Erk-pTpY** in Paper Figure 2B is obtained by restricting to the TEY peptide, which carries only two of the four phosphorylation sites. The effect of this can again be determined by calculation from the four-site phospho-form distribution and equated to the data in Paper Figure 2B, yielding the equations

$$
\begin{aligned}
a_{11} + a_{14} + a_{15} + a_{16} &= 85 && (13)\\
a_5 + a_8 + a_{10} + a_{13} &= 10 && (14)
\end{aligned}
$$

The remaining equations are not informative in the light of what has already been learned above. Since $a_5 = 0$, it follows from equation (14) that

$$a_8 + a_{10} + a_{13} = 10 \tag{15}$$

According to equation (15), $a_8 + a_{10}$ can be no more than 10%, so that a_{11} would have to be at least 50% to satisfy equation (11). If, on the other hand, a_{13} achieved 10%, so that $a_8 + a_{10} = 0$, then a_{11} must be no more than 65% to satisfy equation (12). In other words,

$$50 < a_{11} < 65 \tag{16}$$

Finally, equation (15) says that a_{13} could be as low as 0% but no more than 10%, so that equation (10) tells us that

$$20 < a_{14} + a_{15} < 30 \tag{17}$$

These conclusions are summarised in Figure 6B of the main paper.

The calculations above rely on simple linear equations among $a_1, \cdots, a_{16}$ that arise from the proMS and pepMS data for both **Erk-pTpY** and **Erk-pTY**. They illustrate how such data can be used to constrain the sixteen unknown quantities. A more robust and systematic procedure could be developed with the use of interval arithmetic, in which the value of a_i is treated as lying in an interval, $u \leq a_i \leq v$, whose length reflects the experimental error. Linear algebra can then be undertaken using the rules of interval arithmetic, following a similar sequence of steps to that above.

2 SUPPLEMENTARY FIGURES

The figures appear in the order in which they are referred to in the paper.

References

[1] D. S. Wishart, C. G. Bigam, A. Holm, R. S. Hodges, and B. D. Sykes. ^{1}H, ^{13}C and ^{15}N random coil NMR chemical shifts of the common amino acids. I. Investigations of nearest neighbour effects. *J. Biomol. NMR*, 5:67–81, 1995.

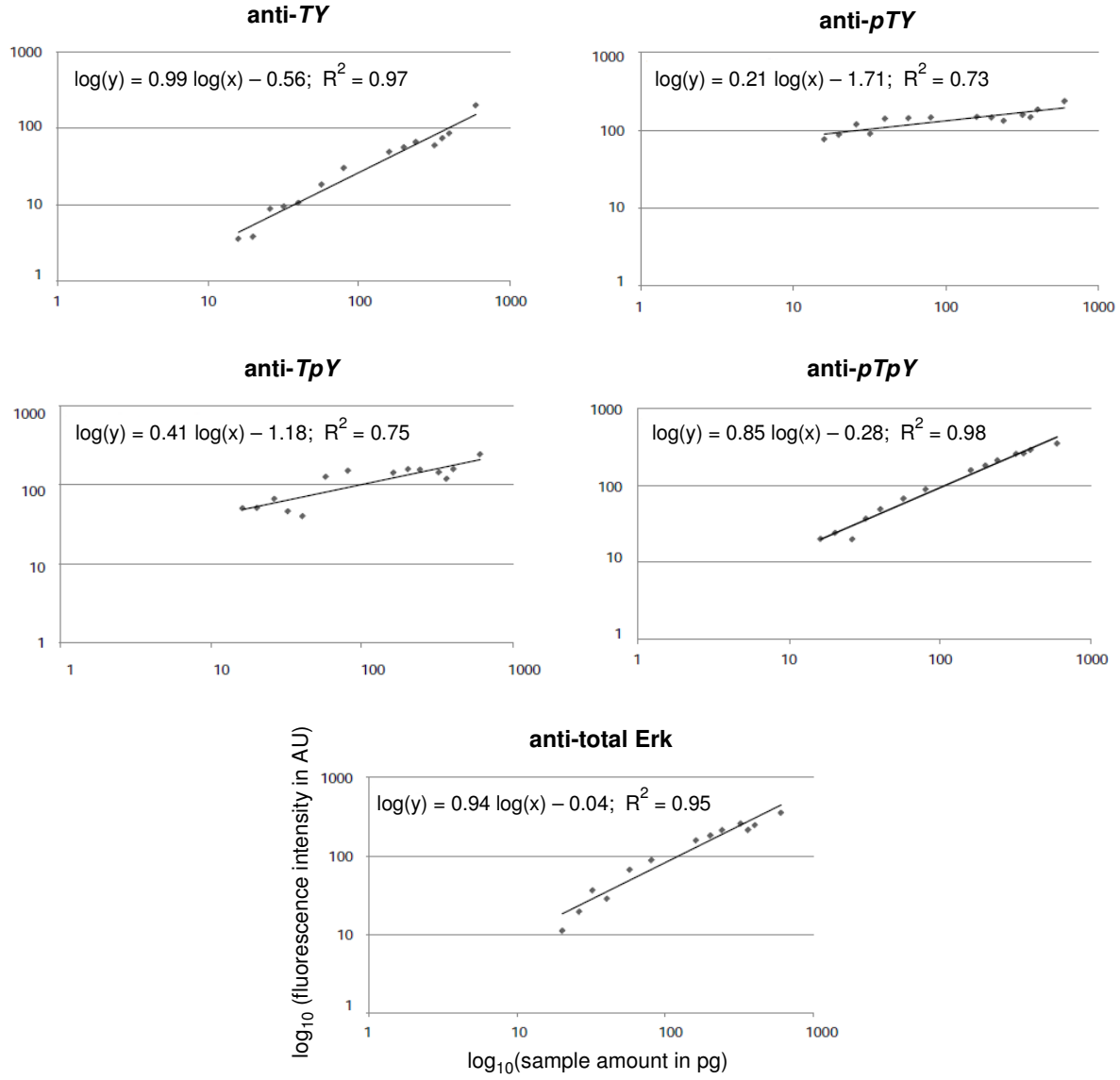

Figure 1: Antibody linearity for Paper Figure 1C. Monoclonal antibodies against each of the four phospho-forms of Erk and a polyclonal antibody against total Erk (ie: irrespective of any phosphorylation) were tested for linearity through a dilution series of their corresponding samples: anti-total and anti-*TY* against **Erk-TY**; anti-*pTY* against **Erk-pTY**; anti-*TpY* against **Erk-TpY**; anti-*pTpY* against **Erk-pTpY**. Log scales are used on all axes and the units for all series are the same, as shown for anti-total Erk. For each series, the equation of the line of best fit under least squares, in terms of sample amount, x, and fluorescence intensity, y, is shown, together with the Pearson correlation coefficient. The line corresponds to a power law, $y = Ax^{\alpha}$, where α is the coefficient of $\log(x)$ and $\log(A)$ is the constant term.

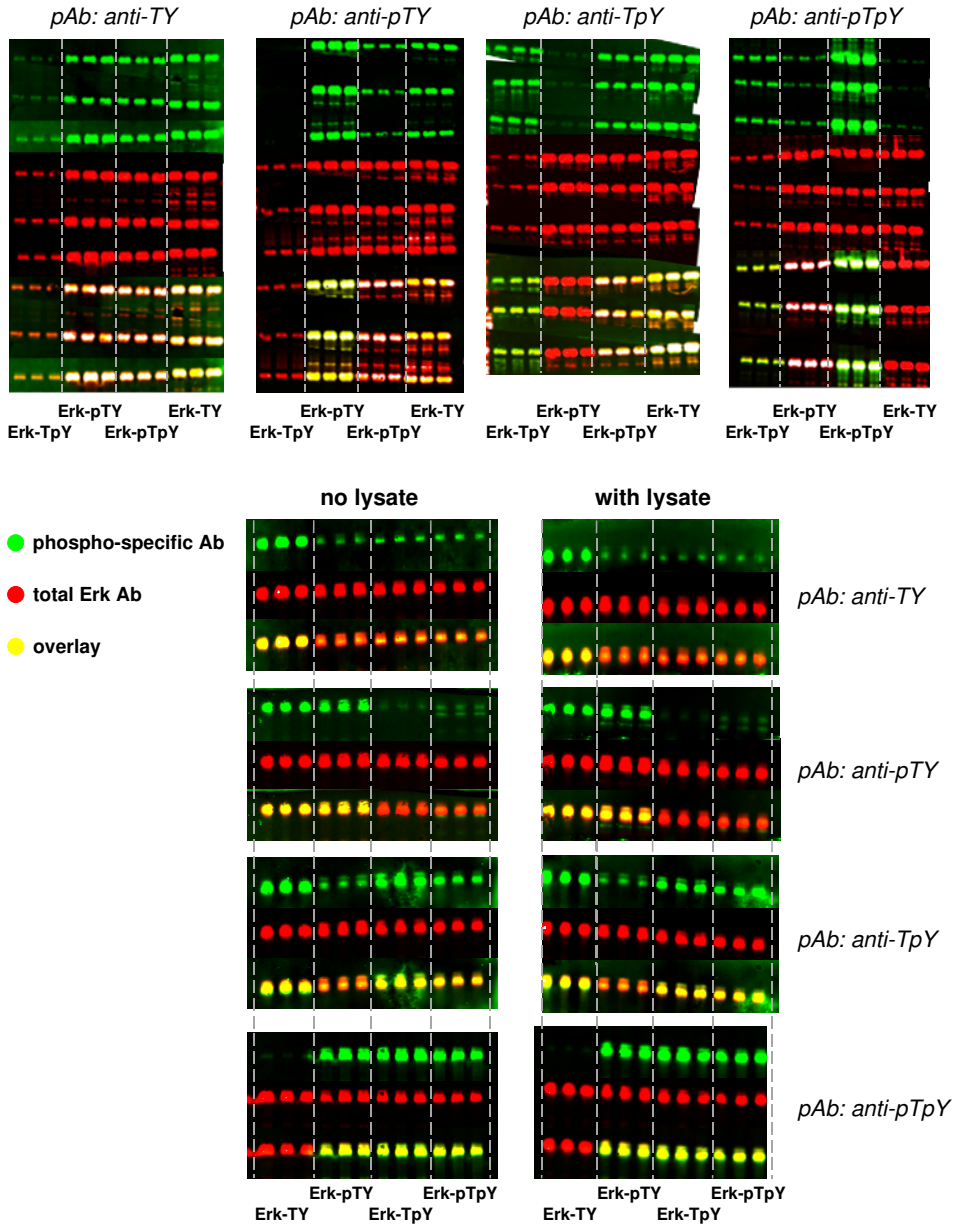

Figure 2: LI-COR Western blots for the first set of samples and Paper Figure 1D (top panel) and for the second set of samples and Paper Figures 3B and C (bottom panel). For the first set, three replicates lanes per gel and three replicate gels were obtained. Note that the **Erk-TY** sample was placed on the last column, not the first. For the second set, only one gel was obtained, with three replicate lanes per gel.

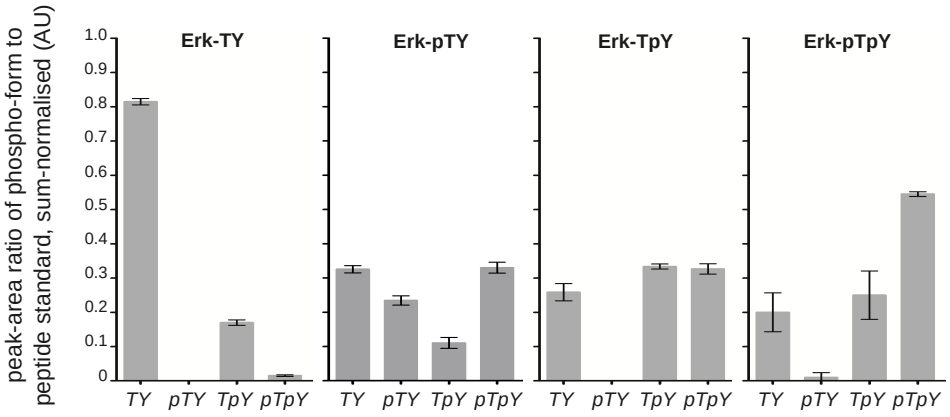

Figure 3: Phospho-form distributions of the second, non-isotopically labelled Erk samples. Peak-area ratios of the sample to the corresponding internal standard were sum-normalised as proportions of the total. The error bars given the mean $\pm$ standard deviation of the normalised values from three replicates. These data were rearranged to give Paper Figure 3A.

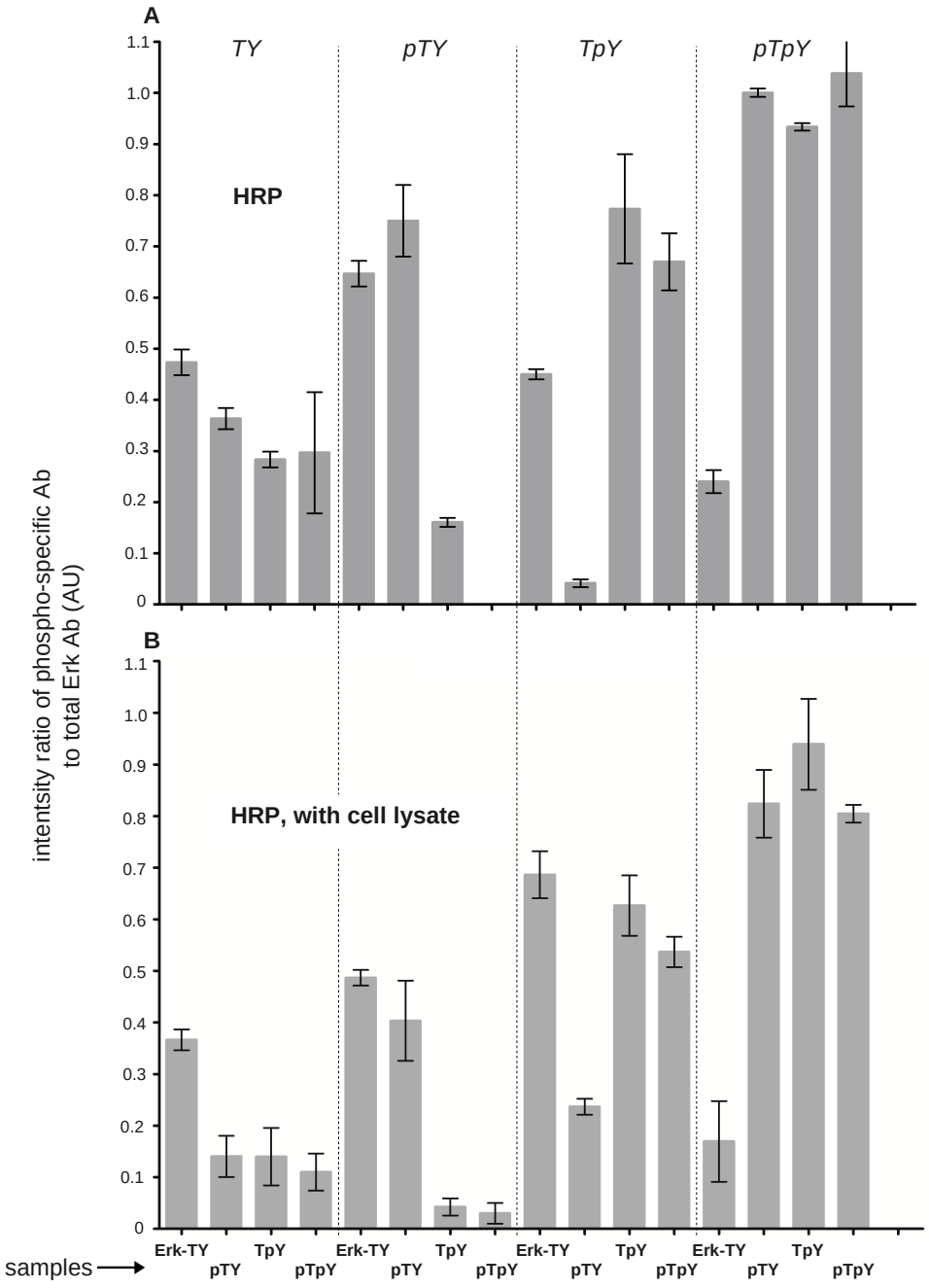

Figure 4: Western blots of the second, non-isotopically labelled Erk samples using HRP-conjugated secondary antibodies and CCD imaging. **A** without lysate. **B** with whole-cell lysate at 250pg of Erk to 10μg of lysate in each well. The with-lysate data shows semi-quantitative agreement with Paper Figures 3B and C.

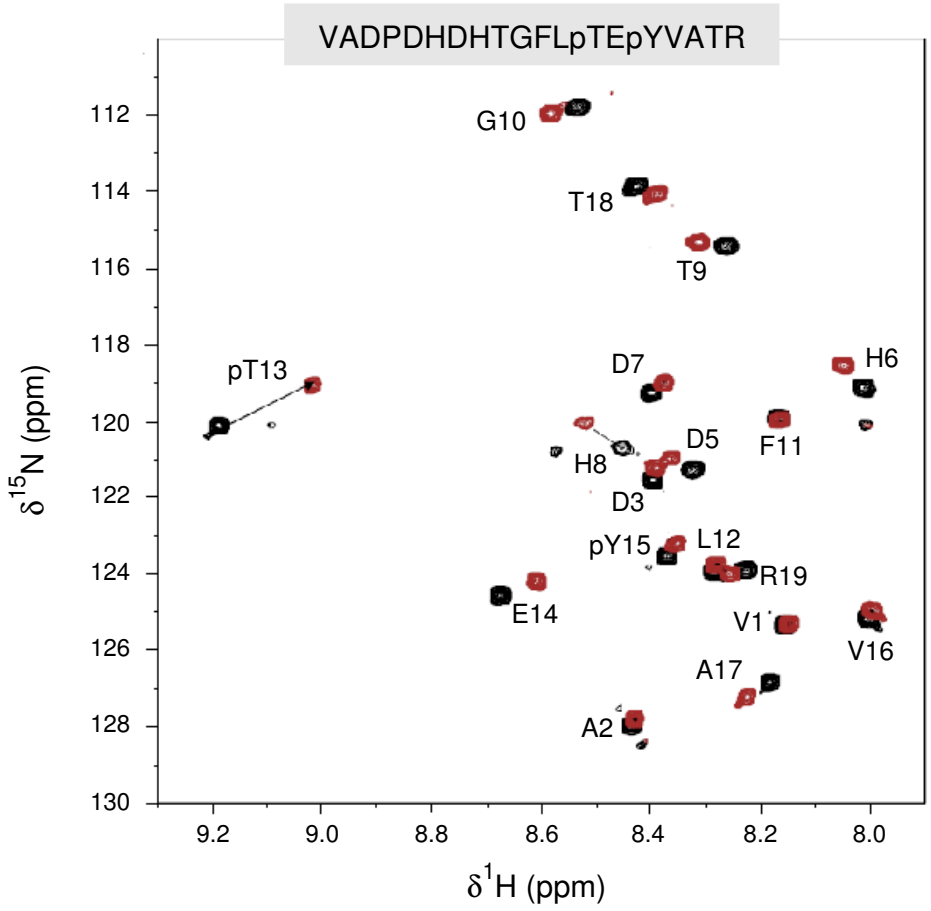

Figure 5: Shift of pT resonances with salt concentration. $^{1}H/^{15}N$ HSQC spectra of the *pTpY* synthetic phospho-peptide at low (black) and high (brown) salt concentrations. The spectra were recorded at 20degC with 5mM peptide in 300μl of aqueous buffer, 25mM Tris-D11 at pH 6.8, 2.5mM EDTA, 1mM DTT, 5% D20, 1mM TMSP, 25mM NaCl (black) and 425mM NaCl (brown). All resonances shift somewhat as a function of salt concentration or pH but the shift is most pronounced for pT.

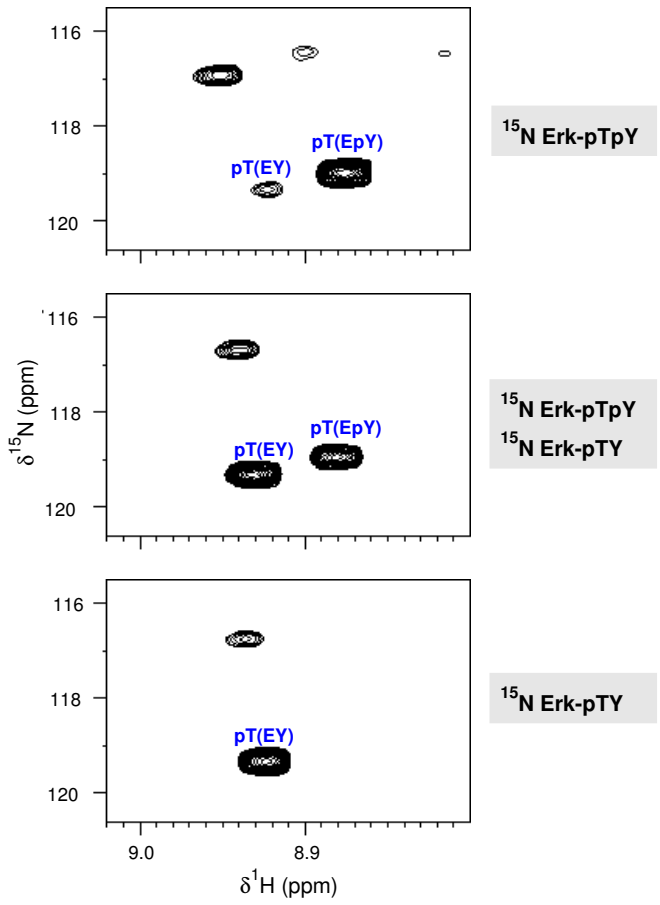

Figure 6: Identification of pT resonances for Paper Figure 5A. Three $^{1}H/^{15}N$ HSQC plots are shown covering the two resonances in the small box in Paper Figure 5A. The top plot is for **Erk-pTpY** at pH 6.70 while the bottom is for **Erk-pTY** at pH 6.77. To correct for pH variation, the C-terminal (G)Y peak, which was found not to change with pH, has been placed in the same position in both plots. Under these conditions the minor peak in the top plot occurs in the same position as the major peak in the bottom plot. The middle plot shows the spectrum with the two samples mixed in equal amounts. It has two distinct major peaks, one corresponding to both the major peak in the **Erk-pTY** sample and the minor peak in **Erk-pTpY** sample and the other to the major peak in the **Erk-pTpY** sample. The latter corresponds to pT in the *pTpY* phospho-form—that is, to pT(EpY)—while the former correspond to pT in the *pTY* phospho-form—that is, to pT(EY). This allows both the *pTpY* and *pTY* phospho-forms to be quantified.

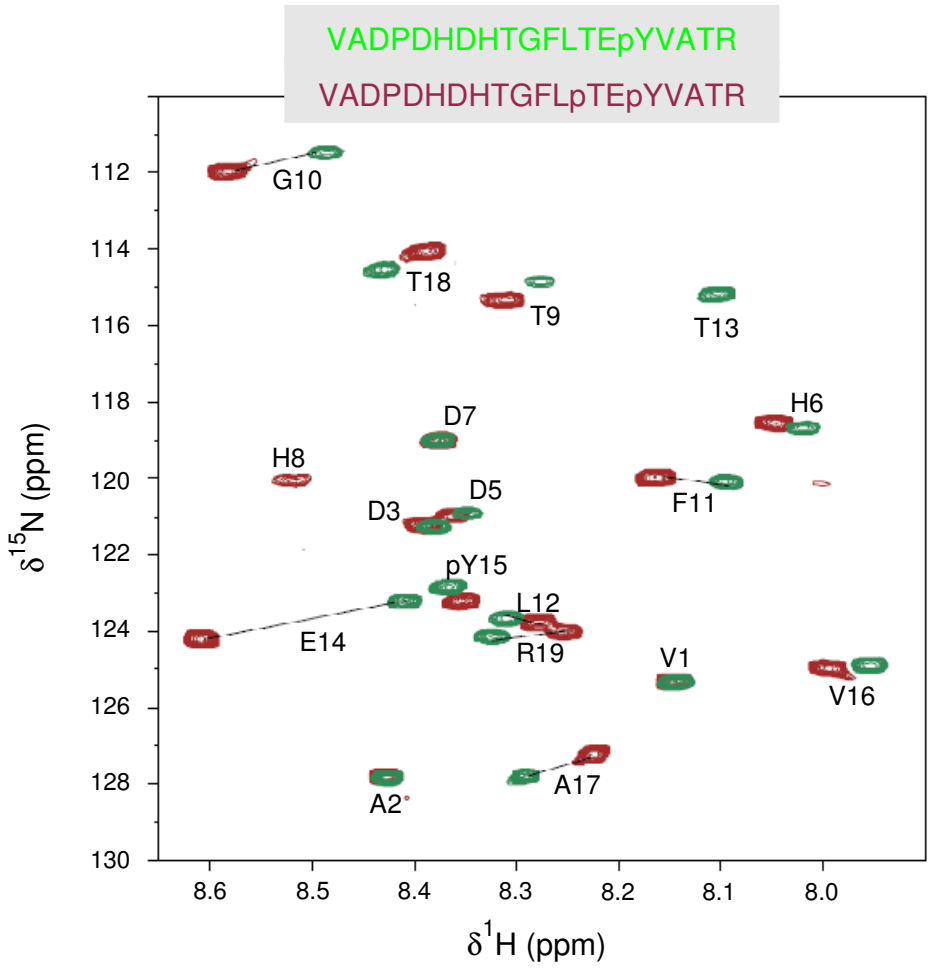

Figure 7: pY resonances are not identifiable. $^{1}H/^{15}N$ HSQC spectra of the *pTpY* and *TpY* synthetic phospho-peptides at high salt concentration, showing that the pY resonance does not shift out of the crowded region centered at 8.3ppm $\delta^{1}H$. The conditions were as in Figure 5 with high salt.

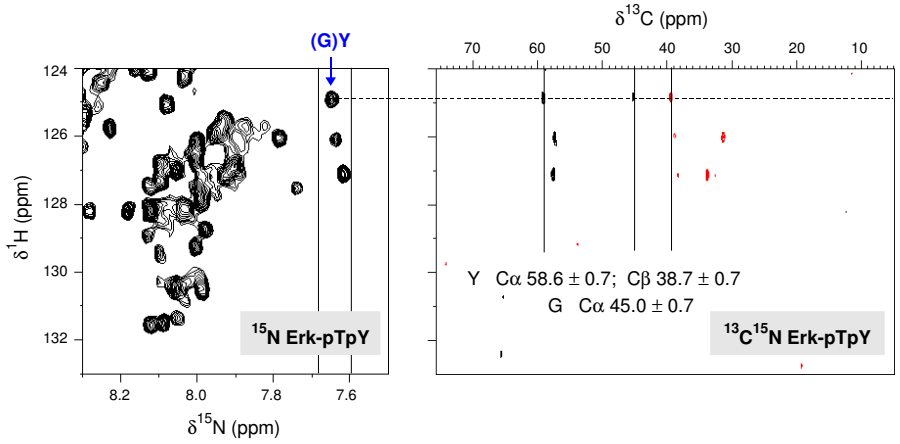

Figure 8: Identification of the (G)Y residue at the C terminal of Erk, used for NMR normalisation in Paper Figure 6A. The C-terminal residues of peptides are found in the region of the ^{1}H/^{15}N HSQC spectrum for **Erk-pTpY** on the left. The accompanying plot on the right is taken from the 3D HNCACB plot of ^{13}C/^{15}N-labelled **Erk-pTpY**. The plot shows a sum of ^{1}H/^{13}C-planes, with the sum being taken across the region between the vertical lines in the HSQC spectrum that covers the resonance marked by the blue arrow. This resonance is found in the spectra of all four samples. It has C_α and C_β peaks for a tyrosine residue at that position and the C_α peak for a glycine residue immediately N terminal [1]. There are two (G)Y dipeptides in Erk (Paper Figure 1A) but only the one at the C terminal of Erk also occurs at the C terminal of a tryptic peptide.

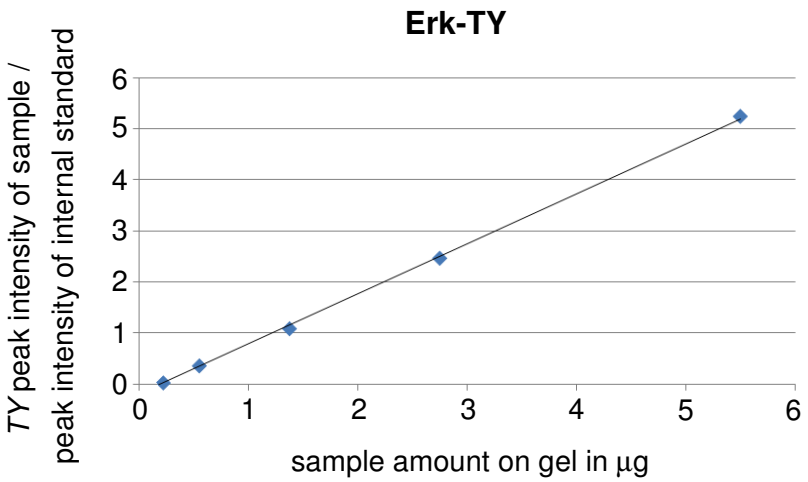

Figure 9: PepMS linearity and loading range, as discussed in the Methods section of the Paper. 500 fmol of the *TY* internal standard were added to various amounts of the **Erk-TY** sample, measured in μg, prior to gel digestion and LC/MS. The vertical axis gives the ratio of the intensity of the *TY* phospho-form in the sample to that of the internal standard, with the least-squares line of best fit through the data points. The linearity is excellent over this range, with a Pearson correlation coefficient of 1.00. A gel loading amount of 1μg was chosen for the pepMS measurements reported in Paper Figure 2.

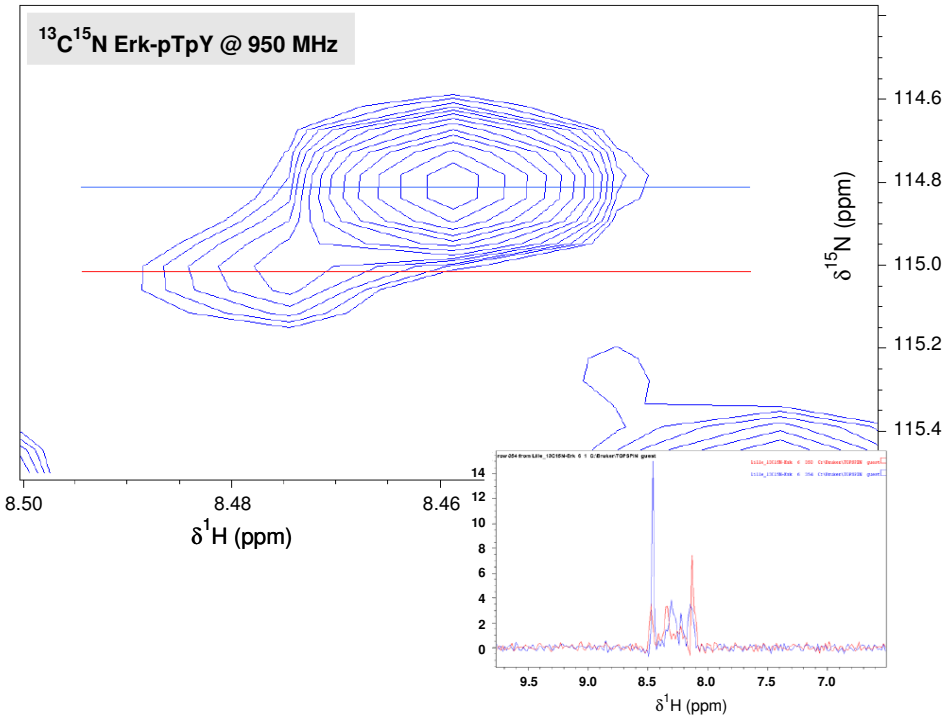

Figure 10: Determination of the (pTEpYVA)T resonance. A 950MHz $^{1}H/^{15}N$ HSQC spectrum of the $^{13}C/^{15}N$-labelled **Erk-pTpY** sample shows that the (A)T resonance splits into a minor and a major peak. The values used in Paper Figure 6A were determined from the major peak. The inset shows the intensity variation along the horizontal red and blue lines but note that it covers a much wider range of $\delta^{1}H$.