Quantitative in-vivo characterization of intracellular and extracellular pH profiles in heterogeneous tumors: a novel method enabling multiparametric pH analysis

Norbert W. Lutz,[1] Yann Le Fur,[1] Johanna Chiche,[2] Jacques Pouysségur,[2,3] Patrick J. Cozzone[1]

SUPPLEMENTARY MATERIAL

[1] Centre de Résonance Magnétique Biologique et Médicale, UMR 7339, Faculté de Médecine de la Timone, CNRS, Aix-Marseille Université, F-13005 Marseille, France

[2] Institute of Research on Cancer and Aging, UMR 7284, Centre Antoine Lacassagne, Université de Nice, F-06189 Nice, France

[3] Centre Scientifique de Monaco, MC-98000 Monaco-ville, Monaco

(***J.C., present address:*** *Controle Métabolique des Morts Cellulaires, INSERM U 1065, Centre Méditerranéen de Médecine Moléculaire, F-06204 Nice, France*)

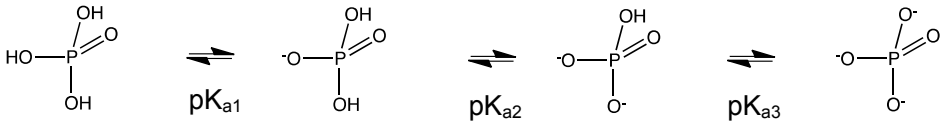
SUPPLEMENTARY MATERIALS AND METHODS

S1. Theory and algorithms

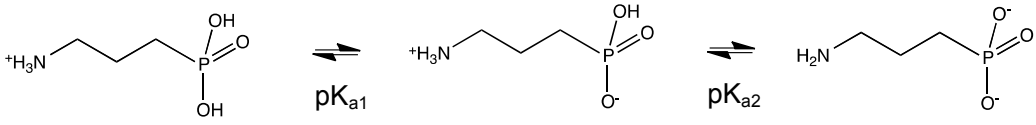
S1.1. Principles of pH measurement by ^{31}P MRS

The use of ^{31}P NMR spectroscopy for pH measurement is based on the pH dependence of the chemical shifts, δ, of particular ^{31}P NMR resonances. The chemical shift of a specific NMR resonance depends on the electron density at the nucleus generating this resonance, and this electron density is strongly influenced by nearby chemical bonds. Therefore, the chemical shift of a particular nucleus is pH-dependent in molecules in which these chemical bonds are affected by pH, *e.g.*, as a consequence of protonation.

A phosphate (phosphonate) moiety has one ^{31}P nucleus bonded to four (three) oxygens. P_i has three, and 3-APP has two sites of protonation: these sites are oxygens directly bonded to phosphorus, *i.e.* at a distance of only one bond from the observed nucleus. This bond is highly sensitive to oxygen protonation because electron density is readily redistributed between the oxygen atom and the oxygen-phosphorus bond as the oxygen-phosphorus bond changes from a double bond to a single bond upon protonation. This explains the high pH sensitivity of the P_i and 3-APP ^{31}P NMR resonances. The number of oxygens that are actually protonated depends on pH (protonation decreases as pH increases), as shown for P_i (eq. S1) and 3-APP (eq. S2) species:



eq. S1



eq. S2

The logarithm of the acid dissociation constant (pK_a) indicates the pH value at which two species at equilibrium are present at identical concentrations. The number of P_i or 3-APP species that exist over the entire pH range (zero to full protonation) equals the number of protonatable oxygens + 1, and the number of pK_a values for each compound equals the number of protonatable oxygens (pK_{a1} to pK_{a3} for P_i; pK_{a1} and pK_{a2} for 3-APP). For biomedical applications, only pK_a values close to the physiological range are of interest (pK_{a2} = 6.85 for P_i, and 6.91 for 3-APP; determined in model solutions for ionic strength and temperature close to physiological values). When pH is equal to pK_a, any variation in pH will cause maximum effects on protonation and, therefore, on ^{31}P chemical shift. Thus, it is desirable to use compounds with a pK_a close to *in vivo* pH for best sensitivity of pH measurements. The relationship between pH and the relevant pK_a is given by the Henderson-Hasselbalch equation (eqs. S3 for P_i, and S4 for 3-APP):

$$\mathrm{pH} = \mathrm{pK_{a2}} + \log\frac{[\mathrm{HPO_4^{2-}}]}{[\mathrm{H_2PO_4^-}]} = \mathrm{pH_i} \qquad \text{eq. S3}$$

$$\mathrm{pH} = \mathrm{pK_{a2}} + \log\frac{[\mathrm{3\text{-}AP\text{-}PO_3^{2-}}]}{[\mathrm{3\text{-}AP\text{-}HPO_3^-}]} = \mathrm{pH_e} \qquad \text{eq. S4}$$

In the chemical-exchange Henderson-Hasselbalch pH model a measured NMR-active species is taken to have a pH-dependent chemical shift resulting from a rapid equilibrium between the acid and conjugate base forms of this species (1). This is the equilibrium between the species in the numerator (conjugate base) and the denominator (acid) in the logarithmic expression of each of the equations S3 and S4. By substituting in these equations each molecular-species concentration with the difference between the measured chemical shift, δ, and the chemical shift of the acid or base-form species (relative to $\delta_{\alpha\text{-NTP}}$ = −10.05 ppm), the following equations are obtained for the corresponding pK_{a2} values (1, 2):

$$\mathrm{pH} = 6.85 + \log\frac{[\delta - 0.58]}{[3.14 - \delta]} = \mathrm{pH_i} \qquad \text{eq. S5}$$

$$\mathrm{pH} = 6.91 - \log\frac{[\delta - 21.10]}{[24.32 - \delta]} = \mathrm{pH_e}, \qquad \text{eq. S6}$$

where 0.58 and 3.14 are the ^{31}P chemical shifts of the $H_2PO_4^-$ and HPO_4^{2-} species, whereas 21.10 and 24.32 are the chemical shifts of the 3-AP-PO_3^{2-} and 3-AP-HPO_3^- species, respectively. Note the negative sign in front of the logarithm in eq. S6; in 3-APP solutions high pH results in low δ, whereas in P_i solutions high pH results in high δ.

Equations S5 and S6 are used to convert a ^{31}P NMR resonance into a pH curve by converting the scale of the abscissa from ppm to pH units. In addition, the intensity axis of the resonance must be divided by the derivative of the Henderson-Hasselbalch equation to correct for the nonlinearity between the δ chemical-shift (ppm) and pH scales (2, 3):

$$\mathrm{I_{Pi}^{corr}} = \mathrm{I_{Pi}}(\delta - 0.58)\left(\frac{[3.14 - \delta]}{[3.14 - 0.58]}\right) \text{ for } \mathrm{pH_i} \text{ calculation, and} \qquad \text{eq. S7}$$

$$\mathrm{I_{3-APP}^{corr}} = \mathrm{I_{3-APP}}(\delta - 21.1)\left(\frac{[24.32 - \delta]}{[24.32 - 21.1]}\right) \quad \text{for } \mathrm{pH_e} \text{ calculation,} \qquad \text{eq. S8}$$

where $\mathrm{I_{Pi}}$ and $\mathrm{I_{3-APP}}$ are the uncorrected intensities (I), and $\mathrm{I_{Pi}^{corr}}$ and $\mathrm{I_{3-APP}^{corr}}$ the corrected intensities ($\mathrm{I^{corr}}$). In these equations, the chemical shifts for the acid and base forms of the pH-sensitive species are used as introduced above for equations S5 and S6.

Without this intensity correction, the nonlinear relationship between chemical-shift and pH would result in a distorted pH curve because the intensity of each point of this curve would be intrinsically dependent on its distance from the pK_a in question. By contrast, applying equations S7 and S8 allows one to correctly translate the measured chemical-shift data points to pH data points. This improvement has first been suggested for pH_i

measurement based on the P_i resonance (3), and has subsequently been applied to pH_e measurement based on the 3-APP resonance (2).

S1.2. Characterization of pH heterogeneity in tissues

S1.2.1. Requirements for measuring heterogeneous tissue pH based on ^{31}P NMR spectra

To derive pH values from a P_i or a 3-APP ^{31}P NMR resonance (chemical-shift scale), the NMR resonance needs to be converted to a pH curve using the Henderson-Hasselbalch equation as described in the previous Section S1.1. The pH curve then represents the pH distribution within the measured tissue volume. As pointed out in the RESULTS section of the main body of this report, the resulting curve may also be influenced by other contributions to lineshape, linewidth and intensity: (i) by magnetic-field inhomogeneity and transverse relaxation (T_2) processes; (ii) potentially by residual unequal distribution of P_i or 3-APP across tissue regions of varying pH_i or pH_e, respectively; (iii) by phosphorus-proton J coupling (contributes somewhat to the spectral linewidth of 3-APP unless protons are decoupled from phosphorus during spectrum acquisition); and, of course, (iv) by the spectral processing parameters chosen, in particular filter parameters. In *in-vivo* spectroscopy, the T_2 effect (natural linewidth) is much smaller than the T_2^* effect (due to magnetic field inhomogeneity). In addition, natural linewidth is essentially determined by molecular interactions that cannot be manipulated under *in-vivo* conditions, e.g. by cationic (Ca^{2+}, Mg^{2+}) exchange of phosphates and phosphonates. Therefore it is appropriate to focus on optimizing the magnetic field homogeneity to minimize T_2^* effects on the linewidth, for instance by employing newest high-quality shim coil systems, and by using new generations of shim procedures that will replace the well-established FASTMAP technique (e.g. MAPSHIM; for a rigorous introduction into the development of new automated shimming methods see Koch *et al.* (4)). Note that besides linewidth effects, T_2 has no effects on spectra if they are acquired *via* free induction decays (FIDs) rather than echoes. Consequently, all spectra used in this work are based on FID acquisitions.

Corrected pH curves obtained with equations S5 to S8 accurately depict the underlying pH heterogeneity, within the limits described in the previous paragraph. Such a curve is, in fact, an envelope representing a mixture of individual tissue compartments covering a range of pH values. However, no significant effort has been made, according to published literature, to characterize and interpret these envelope curves. In the following we propose to analyze pH curves by multiple statistical methods to extract several parameters providing quantitative information on the nature of pH heterogeneity.

S1.2.2. Concept of describing pH heterogeneity by statistical parameters

The most basic objective of statistical analysis of pH in heterogeneous tissue is the determination of a pH value that is representative of the entire pH distribution. Simply using the highest point of the overall curve may be misleading in the case of asymmetric pH distribution because this choice would overrepresent tissue regions with a pH close to this maximum value, and neglect regions represented by the more extended flank of a given pH distribution. This problem has previously been addressed by calculating a "center of gravity" (5) or weighted-average (6) chemical shift, which was then converted to a pH value. However, since center of gravity and weighted average are based on the spectra themselves rather than on pH curves derived from spectra, they do not include the corrections described in Section S1.1, and may thus result in an inaccurate (shifted) pH maximum. Moreover, since this approach yields only one pH value, any details characterizing the pH distribution are lost in the process (compare Sections S1.2.4 to S1.2.9 below). We suggest a new strategy providing accurate weighted-average pH values ($\overline{pH}$, i.e. $\overline{pH_i}$ or $\overline{pH_e}$) and additional statistical parameters describing pH heterogeneity. This strategy is based on the concept that both the pH-sensitive spectrum of 3-APP or P_i, and the pH curve calculated from such a spectrum, are approximated as histograms. Such a histogram is formed by using the intensities of the digital points of a spectrum or a pH curve as the heights of the corresponding histogram bars, as described in the following paragraph. All algorithms used for calculating statistical pH heterogeneity parameters are based on established statistical equations; however, the

original equations have been transformed for use in pH curve point analysis instead of conventional histogram analysis.

S1.2.3. pH curves as histograms

A conventional histogram representing a discrete probability distribution is constructed from a frequency table based on a total of n independently measured parameter values (Fig. S1 A). Each observation falls into one of m adjacent categories forming the abscissa of the histogram, and the weight (or frequency) of each category k is W_k, from k = 1 to k = m. Obviously, W_k is always an integer. Weights are most commonly indicated by the heights (or areas) of vertical columns (sums of rectangles; see Fig. S1 A). Consequently, the total number of observations n equals the sum of all weights: $n = \sum_{k=1}^{m} W_k$ (eq. 26 in (7)). When the observed parameter is a continuous variable x, the total range of this variable is broken up into a number of equal intervals x_k ('bins' or 'buckets'; see Section S2.2 for a discussion of the number of bins), from k = 1 to k = m. The frequencies with which a measured parameter falls into these intervals x_k define the frequency distribution. In all modern NMR spectrometers, spectra are acquired and processed in digitized form. An NMR spectrum is made up of a sequence of equidistant digital points that can be considered as the representation of a statistical frequency distribution (ordinate) of chemical shifts (abscissa); however, the m intervals (bins) of conventional histograms (Fig. S1 A) are replaced with m discrete values of the measured variable δ. In addition, the heights of the digital points of a spectrum do not represent sums of individual observations, but are measured signal intensities; therefore, the weights obtained from NMR spectra are rational numbers rather than integers. After point-by-point conversion of ppm to pH values, the abscissa is made up of discrete pH values pH_k with weights W_k, from k =1 to k = m (Fig. S1 B).

In a ^{31}P NMR spectral line of 3-APP (Fig. S2 A) or P_i, the height (intensity) I_k of each of the equidistant digital spectral points k is a measure of its weight. Conversion of ppm to pH values according to eqs. S5 and S6 yields a pH distribution histogram (Fig. S2 B); however, this transformation has to be followed by I_k-to-I_k^{corr} conversion according to eqs. S7 and S8

to obtain accurate weights $W_k = I_k^{corr}$. The resulting pH distribution curve is approximated as a histogram made up of m digital points that show the frequency distribution of pH values (Fig. S2 C), akin to conventional histograms mentioned above in this paragraph (8). The weight W_k of a particular digital point k in a pH curve is a measure of the relative frequency with which the pH value represented by this point occurs in the measured sample. W_k can be considered as the sum of (real or hypothetical) individual contributions to the pH frequency at a given digital point of the pH curve. Furthermore, $n = \sum_{k=1}^{m} W_k$ is the sum of the weights of all digital points of the pH curve (7). By analogy to conventional histograms, n can be thought of as the (real or hypothetical) total number of individual contributions to the pH frequency distribution, and interpreted as being proportional to the total number of equal microscopic tissue volume elements contributing to the weights W_k of the digital points k of the pH curve. In Fig. S2 A to C, only the part of the data range that is essential for constructing the pH distribution curve is presented. Remarkably, the tails of the pH distribution curve (Fig. S2 C) are significantly lighter than the tails of the underlying spectral line (Fig. S2 A). While the W_k values generated for the pH distribution curve provide accurate pH profiles, they need to be modified as described in the following paragraphs if they are to be used for further statistical evaluation.

S1.2.4. Weighted-average pH

S1.2.4.1. The issue of non-equidistant digital points in pH curves

The digital pH curves introduced in the previous paragraph can be used to calculate weighted-average pH values and other statistical parameters. In these curves, the weight, W_k, of any given digital point is defined as its corrected intensity: $W_k = I_k^{corr}$, where k is the index counting the digital points to be used for calculating the weighted average from k=1 to k=m. Based on these curve points, a weighted-average pH value could be readily obtained by multiplying the pH value of each digital point by its weight, and by dividing the sum of these products by the sum of all weights. However, sequential ppm-to-pH conversion of the equidistant digital points of a spectrum results in non-equidistant points on the pH scale

because of the nonlinear relationship between chemical shift and pH discussed in Section S1.1. In fact, intervals between adjacent digital points in the pH domain increase with increasing distance from the relevant pK_a (Fig. S2 C). If these interval variations are not taken into account in $\overline{\mathrm{pH}}$ calculation, pH curve points distant from the pertinent pK_a in question are underrepresented *vs.* points close to this pK_a, as a function of their distance from this pK_a (Fig. S2 D, triangles; pK_{a2} = 6.91 for 3-APP). There are three possible solutions for this conundrum; these will be presented in the following paragraphs.

S1.2.4.2. Generation of equidistant digital points in pH curves by zero-filling

Theoretically, the problem of non-equidistant digital points in pH curves could be circumvented by employing the following two-step approach: first, process the NMR spectrum with multiple zero fillings to generate a large number of points per ppm; then, upon ppm-to-pH conversion and intensity correction, sample a subset of approximately equidistant digital points from the pH curve and obtain weighted-average pH values based on their weights $W_k = I_k^{corr}$. However, this approach would be rather complex, time consuming, and only approximately accurate.

S1.2.4.3. Rescaling weights of non-equidistant digital points by interval weighting

An alternative approach is based on using the non-equidistant digital points produced by ppm-to-pH conversion. To obtain accurate weighted-average pH values, this method requires that the weight of each point be adjusted for variations in intervals between points. The first variant of this method consists of determining, for each digital point, the intervals to its left and to its right, and multiplying its weight $W_k = I_k^{corr}$ by the mean of these intervals. Subsequently, the weighted-average pH is calculated based on all curve points rescaled in this manner (Fig. S2 D, circles).

S1.2.4.4. Rescaling weights of non-equidistant digital points by analytical weighting

The most accurate and elegant technique for nonlinear rescaling is replacing the intensity I_k^{corr} of each digital point of the pH curve with the uncorrected intensity I_k: $W_k = I_k$ rather than $W_k = I_k^{corr}$ (Fig. S2 D, diagonal crosses). This nonlinear scaling correction compensates exactly the effect on $\overline{pH}$ caused by the decrease in pH curve point density with increasing distance from the relevant pK_a, i.e., the decreasing number of points per pH unit. The weighted average for pH_i and pH_e is then calculated by way of the following equations (eqs. S9 and S10, respectively):

$$\overline{pH_i} = \frac{\sum_{k=1}^{m}(pH_{i,k} \times W_{i,k})}{\sum_{k=1}^{m} W_{i,k}} \qquad \text{eq. S9}$$

$$\overline{pH_e} = \frac{\sum_{k=1}^{m}(pH_{e,k} \times W_{e,k})}{\sum_{k=1}^{m} W_{e,k}} \qquad \text{eq. S10}$$

where $pH_{i,k}$ ($pH_{e,k}$) is the pH value for a given digital point k of the pH_i (pH_e) curve; m is the number of curve points used for $\overline{pH_i}$ or $\overline{pH_e}$ calculation; and $W_{i,k} = I_{i,k}$ ($W_{e,k} = I_{e,k}$) is the weight of curve point k as introduced above. Rescaling by analytical weighting and interval weighting yield almost identical results, notably for pH values close to the pK_a in question (Fig. S2 D). Major deviations exist for apparent pH values close to or greater than 8.0 where the intervals between adjacent curve points increase strongly and nonlinearly with increasing pH. Interval weighting is based on the arithmetic means of adjacent intervals (Section S1.2.4.3.), amounting to local linear interpolation. Therefore, this method becomes inadequate for curve points that change strongly and nonlinearly with pH.

Note that the curve point intensities based on rescaled weights (Fig. S2 D, circles and diagonal crosses) do not represent pH profiles; they are only a visualization of the weights to be assigned to the corresponding curve points for the purpose of obtaining accurate statistical parameters. The use of I_k^{corr} instead of I_k is still necessary for the generation of accurate pH curves where pH is represented on a linear scale (Fig. S2 D, triangles; see also Section S1.1). By contrast, I_k is used when weights of pH curve points need to be adjusted

for varying intervals. Besides $\overline{pH}$ calculation, I_k is used in the calculation of the weighted median $\widetilde{pH}$, but also in the determination of skewness, kurtosis, entropy and areas under curves (see Sections S1.2.5. to S1.2.7. and S1.2.9. below).

S1.2.4.5. Advantages of weighted-average *vs.* conventional tissue pH values

Major advantages of using a weighted-average pH value compared with the pH of a single (*i.e.* the highest) curve point are: (i) $\overline{pH}$ provides an unbiased mean pH that truly represents the entire pH range, (ii) $\overline{pH}$ can be obtained regardless of the shape of the pH distribution (broad or narrow; symmetric or asymmetric; unimodal, bimodal or multimodal), and (iii) ^{31}P NMR lineshape distortions caused by other factors than pH have little influence on the resulting $\overline{pH}$ value since lineshapes are affected independently of pH, unless magnetic-field inhomogeneity is very large and varies significantly between tissue regions of different pH. For extremely inhomogeneous magnetic fields, the lineshape of the pH-insensitive ^{31}P NMR resonance used for chemical-shift reference (usually NTP-α) would also be distorted, which would complicate pH analysis. For less distorting magnetic-field inhomogeneities, it is sufficient to calibrate the weighted-average chemical shift of the reference signal to its set value prior to ppm-to-pH conversion. Note that in an *in vivo* situation, the NTP-α peak predominantly represents ATP, with minor other nucleoside triphosphates contributing to the same signal. Since the NTP-α peak is rather broad, it strongly overlaps with the NDP-α peak that is slightly downfield from the NTP-α peak; the resulting line is an envelope comprising both NTP-α and NDP-α signals. The chemical-shift difference between NTP-α and NDP-α is on the order of 0.4 ppm, corresponding to 0.3 pH units. As NDP is usually much less concentrated than NTP, its effect on the precision of the NTP-α resonance is mostly negligible. However, in physiological situations where NTP is depleted and NDP levels are increased, the chemical shift of the NTP-α/NDP-α envelope increases due to a significant contribution of the NDP-α signal to the envelope. Since this envelope peak serves as a chemical-shift reference for all other ^{31}P NMR signals of the same

spectrum, nominal $\delta_{3\text{-APP}}$ and δ_{Pi} also increase as [NTP]/[NDP] decreases. Unless this effect is corrected for, the apparent pH_e will increase and the apparent pH_i decrease due to a reduced NTP content.

S1.2.5. Weighted median pH

The rationale for the use of weighted pH values for the calculation of a median pH is along the same lines as the rationale presented in Section S1.2.2. for weighted means: $\widetilde{pH}$ provides an unbiased median pH that truly represents the entire pH range; it can be obtained regardless of the shape of the pH distribution; and lineshape distortions caused by other factors than pH have little influence on the resulting $\widetilde{pH}$ value. In addition, a well-known advantage of medians *vs.* means is the robustness of the latter to outliers. Common spreadsheets such as EXCEL can be employed to determine weighted-median pH according to the following algorithm, which is equivalent to the general algorithm for median calculation from a frequency distribution (9). First, cumulative sums CSUM (corresponding to cumulative frequencies) are calculated for the rescaled weights of the individual pH curve points: $\mathrm{CSUM(k)} = \sum_{i=1}^{k} W_i$, for all k from k = 1 to k = m. Then, the cumulative sum of weights obtained for the last point of the pH curve, $\mathrm{CSUM(m)} = \sum_{i=1}^{m} W_i$, is divided by 2, yielding the half-sum $\mathrm{CSUM}(\tilde{k}) = \frac{\mathrm{CSUM(m)}}{2}$. Subsequently, the pH curve point k whose cumulative sum equals the obtained half-sum is identified, that is the curve point k for which $\mathrm{CSUM(k)} = \mathrm{CSUM}(\tilde{k})$. The pH value of this curve point, $k = \tilde{k}$, is defined as the weighted median pH = $\widetilde{pH}$. This relationship could be directly applied to $pH_k = pH_{i,k}$, or $pH_k = pH_{e,k}$, as defined above for equations S9 and S10.

However, two adjustments are needed to obtain proper pH values for $\tilde{k}$ and, subsequently, accurate $\widetilde{pH}$ values. First, the algorithm for $\widetilde{pH}$ calculation described above has been developed for frequency distributions based on binned measurement of continuous parameters, as pointed out in Section S1.2.3. This implies that the cumulative sum of each bin k includes all observations from parameter interval 1 through parameter interval k. By

contrast, our pH histograms are not based on continuous-parameter measurements that fall into bins, as each point k represents a discrete pH value. In fact, a pH_k value can be thought of as a distinct data point (Fig. S1 B) located at the center of a conventional bin interval (Fig. S1 A). For this reason, the pH_k values used for the identification of $\widetilde{pH}$, as described in the previous paragraph, must be corrected by adding one half of the distance to the next curve point, $(pH_{k+1} - pH_k)$, such that $pH_{\tilde{k}} = pH_k + \frac{1}{2}(pH_{k+1} - pH_k)$. Second, in most practical cases none of the points k will possess a cumulative sum $CSUM(k)$ that is exactly equal to the half-sum. Instead, there will be two adjacent points, $\tilde{k}$ and $(\tilde{k} - 1)$, that possess cumulative sums such that $CSUM(\tilde{k} - 1) < \frac{CSUM(m)}{2} < CSUM(\tilde{k})$. The exact median pH will lie between these adjacent pH curve points and can be interpolated based on the differences between the half-sum $\frac{CSUM(m)}{2}$ on the one hand, and each of the cumulative sums, $CSUM(\tilde{k} - 1)$ and $CSUM(\tilde{k})$, on the other. Consequently, the equation to be used for the calculation of the weighted median pH is

$$\widetilde{pH} = pH_{\tilde{k}-1} + (pH_{\tilde{k}} - pH_{\tilde{k}-1}) \times f_{int} \,, \qquad \text{eq. S11}$$

where $pH_{\tilde{k}}$ is the pH value of curve point $\tilde{k}$ possessing the cumulative sum $CSUM(\tilde{k})$; $pH_{\tilde{k}-1}$ is the pH value of curve point $(\tilde{k} - 1)$ possessing the cumulative sum $CSUM(\tilde{k} - 1)$; and f_{int} is an interpolation factor defined as

$$f_{int} = \frac{\left(\frac{CSUM(m)}{2} - CSUM(\tilde{k}-1)\right)}{CSUM(\tilde{k}) - CSUM(\tilde{k}-1)} \qquad \text{eq. S12}$$

It is evident that $\widetilde{pH} = pH_{\tilde{k}}$ in the limit where $CSUM(\tilde{k}) = \frac{CSUM(m)}{2}$. Likewise, $\widetilde{pH} = pH_{\tilde{k}-1}$ in the limit where $CSUM(\tilde{k} - 1) = \frac{CSUM(m)}{2}$. This algorithm can be applied to $pH_{\tilde{k}} = pH_{i,\tilde{k}}$ or $pH_{\tilde{k}} = pH_{e,\tilde{k}}$, in analogy to pH_k above. A practical test of this algorithm is provided by way of

numerical simulation of a well-defined distribution function (Gaussian distribution resulting in identical mode, mean and median, see Table S2).

S1.2.6. Skewness and kurtosis

While the average value of a data set describes its location, skewness is a measure of the lack of symmetry of a given data set (9). A normal distribution is perfectly symmetric and, therefore, has the skewness = 0. A negative skew implies that the left tail of the distribution is heavier than the right tail, and that most of the data are clustered on the right. A positive skew means that the right tail of the distribution is heavier than the left tail, and that most of the data are clustered on the left. However, skewness = 0 does not necessarily imply a symmetric distribution. Kurtosis is a measure of whether the data are peaked or flat relative to a normal distribution (9); a leptokurtic distribution has a higher peak and heavier tails than a normal distribution, while platykurtic distributions are flatter and have lighter tails (10). Histograms constructed as described in Sections S1.2.2. and S1.2.3. can be analyzed for skewness and kurtosis in pH distribution (considering the limitations described in Section S1.2.1.), after rescaling according to Section S1.2.4.4. The equations for skewness and kurtosis, adopted from the statistics module of the commonly used EXCEL spreadsheet and adapted to pH distributions, are:

$$\text{skewness} = G1 = \frac{n}{(n-1)(n-2)} \sum_{l=1}^{n} \left(\frac{pH_l - \overline{pH}}{s}\right)^3 = \frac{n}{(n-1)(n-2)} \sum_{k=1}^{m} W_k \left(\frac{pH_k - \overline{pH}}{s}\right)^3 \quad \text{eq. S13}$$

$$\text{kurtosis} = G2 = \frac{n(n+1)}{(n-1)(n-2)(n-3)} \sum_{l=1}^{n} \left(\frac{pH_l - \overline{pH}}{s}\right)^4 - \frac{3(n-1)^2}{(n-2)(n-3)} = \frac{n(n+1)}{(n-1)(n-2)(n-3)} \sum_{k=1}^{m} W_k \left(\frac{pH_k - \overline{pH}}{s}\right)^4 - \frac{3(n-1)^2}{(n-2)(n-3)} \quad \text{eq. S14}$$

with $$s = \sqrt{\frac{\sum_{l=1}^{n}(pH_l - \overline{pH})^2}{(n-1)}}$$ eq. S15

$$= \sqrt{\frac{\sum_{k=1}^{m} W_k(pH_k - \overline{pH})^2}{(n-1)}}$$

and $$n = \sum_{k=1}^{m} W_k,$$ eq. S16

where s is the nominal standard deviation, a parameter analogous to the standard deviation of the mean based on individual observations; $pH_l = pH_{i,l}$, or $pH_l = pH_{e,l}$, is the pH value of the l-th (real or hypothetical) individual contribution to the measured pH_i or pH_e distribution; l is a parameter analogous to the index that counts individual observations in conventional skewness and kurtosis calculation, from l = 1 to l = n.

Both skewness and kurtosis characterize the shape of a frequency distribution. Hence, for the purpose of calculating skewness and kurtosis, the absolute value of n is of no importance (as long as it is not too small) since properties such as asymmetry, pointedness and the weight of the tails of a pH curve only depend on the relative weights of the individual pH values represented by digital points, and on their deviation from a normal distribution. Skewness and kurtosis asymptotically approach n-independent values for n greater than several times the number of digital points m (numerical examples are presented in Figure S3). Therefore, to obtain consistent results it is sufficient to scale the weights W_k such that n is much greater than m (W_k is typically on the order of several thousand or higher).

An alternative approach to calculating skewness and kurtosis would be to consider the pH curve as a representation of a continuous probability density function rather than a series of discrete (digital) points, and to derive skewness and kurtosis from this function. However, real-world pH distributions in tissues can rarely be described by well-defined mathematical functions because they are often highly irregular (11). Therefore, these pH distributions are

not parameterizable and, consequently, not amenable to analytical solutions (with the exception of some cases where mixed Gaussian/Lorentzian curve fitting is possible; see also S1.2.9.). A third alternative would be computer simulation of the underlying probability density function, but simulating complex distributions would be extremely problematic and demanding in computer time (12).

The interpretation of pH profiles in terms of skewness and kurtosis is most instructive when the profiles are inspected visually. For skewness > 0, the tail of the pH distribution extends to the right, indicating an asymmetric pH distribution with a higher prevalence of high-pH vs. low-pH tissue regions. This implies that the more acidic (or less alkaline) tissue fraction (< mean pH) consists of regions whose pH values vary little and are close to the mean, while there is a relatively large pH variability within the less acidic (or more alkaline) tissue fraction. For skewness < 0, the pH asymmetry consists in a higher prevalence of low-pH vs. high-pH tissue regions, which means that the pH variability is larger in the low-pH (< mean pH) than in the high-pH (> mean pH) tissue fraction. For kurtosis > 0, the pH distribution is more peaked than that of a normal distribution. In such cases a substantial tissue proportion has pH values that fall within a very narrow pH range around the center of the pH distribution, while the residual tissue has pH values that are spread over a large range beyond the center. By contrast, for kurtosis < 0 the bulk of the tissue mass has pH values distributed over a significant range around the mean (flat pH distribution), but very little tissue has extreme pH values. Accordingly, both skewness and kurtosis are parameters useful for characterization of tissue pH heterogeneity. Departure from normal pH distribution can be visually assessed by displaying pH distributions with their cumulative distributions plotted as normal quantile plots (9). However, since these plots provide little practical information on pH distributions apart from indicating their departure from normality, we do not present diagrams based on normal quantile pH distribution plots.

S1.2.7. Entropy

While kurtosis can be used as a measure of the peakedness or flatness of a heterogeneous pH distribution, also the evenness (smoothness) of a pH curve can be determined by employing a statistical function. A convenient measure of the smoothness of a probability distribution is its statistical entropy (13). The basis of this concept is the discrete Shannon entropy (14, 15), originally developed in information theory:

$$H(P) = -\sum_{i=1}^{n} p_i \log_2[p_i] \qquad \text{eq. S17}$$

where H is the entropy, and P = p_1,.., p_n is the set of all probability distributions. When applied to the analysis of the rescaled weights W_k defined above for pH distributions, equation S17 becomes

$$H(W) = -\sum_{k=1}^{m} \left[\frac{W_k}{\sum_{k=1}^{m} W_k} \log_2 \left(\frac{W_k}{\sum_{k=1}^{m} W_k} \right) \right]. \qquad \text{eq. S18}$$

The base b of the logarithm (b= 2 in equations S17 and S18) is not important, as long as the same base is used consistently; change of base merely results in a rescaling of the entropy value. Although the notion of evenness, represented by entropy, may appear to be similar to the notion of flatness, represented by kurtosis, there is an important difference between entropy and kurtosis: the equation for kurtosis contains terms referring to pH values (average pH and pH of individual digital points of the pH curve, eq. S14) whereas the equation for entropy contains exclusively terms referring to the distribution of weights W_k of pH curve points (eq. S18). Consequently, pH entropy is a direct measure of how even a pH distribution is: low entropy indicates that there are significant tissue regions with particular pH values that occur at much higher frequencies than other regions within the measured volume. By contrast, kurtosis measures the deviation of the flatness/peakedness of a particular pH distribution from a normal distribution (about a mean pH value), and also includes information on the tails of the distribution (Section S1.2.6.).

S1.2.8. Identification of distinct pH modes and pH ranges in heterogeneous tissue

In statistics, the mode is the value that occurs most frequently in a data set or a probability distribution, and a multimodal distribution is a continuous probability distribution with two or more modes (9, 16). The distribution of pH values across a given volume may be multimodal in heterogeneous tissue, as described in the INTRODUCTION section of the main body of this report. This type of pH distribution manifests itself by two or more maxima (= modes) in a pH curve, although the underlying lines often overlap heavily. In some cases, weaker lines are only visible as shoulders on more intense neighboring lines. Where two or more lines can be distinguished, their modes (positions of the corresponding maxima) can be determined individually (e.g., pH_{e1}, pH_{e2} in (11)). This should be achieved either by inspecting the pH curves, or by parameter fitting of an appropriate function (Lorentzian, Gaussian or other) to the data. Obviously, the latter method is only applicable to pH curves whose shapes can be approximated reasonably well by available fit functions, which mostly excludes asymmetric distributions. Asymmetric fit functions, *e.g.* based on exponential functions (17), do exist but are not accessible through widely used software packages. At any rate, owing to the purely empirical character of these functions little would be gained by employing these. In cases where pH distributions are not symmetric, multiple characteristic pH ranges in a pH curve can be frequently identified visually. In summary, the pH modes and pH ranges-based methods presented here have the advantage of yielding more detailed pH information than the commonly used practice of determining one overall maximum only for an entire pH curve. Our approach amounts to an identification of multiple tissue subregions each of which is characterized by a specific pH distribution centered about a distinct, dominating pH value, or by a specific pH range.

In the presence of extreme pH values (*e.g.*, in cases of extreme acidosis (18, 19) or alkalosis (20)), more than two or three separate modes can potentially be detected. In the opposite case, *i.e.* when the pH difference between two phantom compartments or tissue regions is on the order of 0.4 units or smaller, it will be difficult to clearly separate modes in

pH curves derived from standard proton-coupled 3-APP spectra. Instead, pH "maxima" may appear as shoulders on broad lines as shown in Fig. 4 H (with pH_{e2} forming a shoulder on pH_{e3}), but can still be used to characterize pH heterogeneity. Situations in which tissue physiology or pathology results in relatively flat pH distributions that do not possess well-defined modes, cannot be mimicked easily by simple phantoms. Quantitative evaluation of such pH distributions will be described below.

S1.2.9. Quantification of multiple tissue regions with differing pH

In heterogeneous tissues, multiple regions may be distinguishable by their pH as outlined in the previous paragraph. In addition to this qualitative characterization of pH distributions, pH curves can also be used to calculate or estimate the physical sizes of these regions. Since the pH curves are based on ^{31}P NMR spectra, the area under a pH curve is, under appropriate measurement conditions, proportional to the amount of the compound measured. For this reason, we suggest to use ^{31}P NMR-derived pH curves to quantify tissue regions defined by specific pH values. Relative sizes of individual areas from a particular pH curve can be obtained as a result of (i) fitting of multiple analytical curves, or (ii) integration over defined pH ranges. The relative areas calculated in this way directly yield the relative volumes of the corresponding tissue regions, provided that P_i or 3-APP are equally concentrated over all regions studied (see Section S1.2.1).

To obtain accurate areas under curves, care must be taken to correctly scale the weights of the pH curve points as explained above for the calculation of statistical parameters. When the integration method is used, intensities of curve points are summed over chosen pH ranges. However, if this were to be completed based on a histogram of a pH curve (Fig. S3 A), pH regions distant from the pertinent pK_a would be underestimated relative to regions close to this pK_a because the histogram bars are not equidistant. In fact, the areas would be calculated as though the histogram had equal intervals (Fig. S3 B). To compensate the deemphasizing of pH regions distant from, and the overemphasizing of regions close to this pK_a, the rescaling procedure introduced in Section S1.2.4.4 needs to be applied: $W_k = I_k$

rather than $W_k = I_k^{corr}$, which yields the distribution curve of Fig. S3 C. Obviously, the resulting curve point intensities correspond exactly to the intensities present in the original ^{31}P NMR spectrum. For better precision, the data point range considered should include the extended tails found in the underlying spectral lines (Fig. S3 D), although these additional points do not contribute significantly to the pH distribution curve that possesses rather light tails (Fig. S3 A). In some favorable cases, the integration of separate chemical-shift regions can be replaced by fitting multiple analytical curves with Gaussian/Lorentzian lineshapes, a procedure yielding the areas under the fitted curves. Conventional NMR spectrometer software can be used to apply this "deconvolution" to pH data (see Section S2.2. below). However, when a pH curve (Fig. S3 A) is transferred to spectrum processing software, the intervals on the horizontal (pH) axis are usually equalized by default (Fig. S3 B). Therefore, the fitting procedure has to be performed on rescaled histograms where $W_k = I_k$ (Fig. S3 D) as described above for the integration method. Since the intensities I_k are equivalent to the corresponding intensities of the underlying ^{31}P NMR spectrum, deconvolution can be performed directly on the spectrum itself because ratios of areas under a curve do not depend on abscissa units, provided the abscissa is linearly scaled.

S1.2.10. Properties of bimodal and multimodal pH distributions

The global pH_e profile presented in Fig. 2 F has only a moderately positive skew which is reflected by a moderate G1 value (Table 1 F). All four pH_e profiles mentioned in this paragraph are platykurtic, *i.e.* have a flatter global distribution than a Gaussian. This is reflected by negative G2 values that are particularly pronounced for Fig. 2 B (particularly broad distribution of pH_e values without a dominating peak; large pH difference between modes; very few pH values near center of pH distribution; Table 1 B). The smoothness of the pH_e distribution was relatively low for the curve showing two equally high, well separated modes (Fig. 2 B), indicated by a relatively low H value. The bimodal distribution with the highest entropy is presented in Fig. 2 G (particularly even pH_e distribution, *i.e.* a broad curve without a deep minimum between modes). Among the pH_i distributions presented, both Fig. 3

A and C may be considered bimodal, although in the latter case the two modes are barely distinguishable; separate values for pH_{i1} and pH_{i2} can only be estimated (see also Table 2 A and C). Consequently, the difference between the pH_{i2}/pH_{i1} area ratios obtained from deconvolution and integration diverged for Fig. 3 C more than it did for Fig. 3 A. Both of these relatively even pH_i distributions presented lower G1 and G2 values and higher H values than did the two other pH_i distributions (Table 2 B and D). In summary, all quantitative parameters clearly represent the pH curve shapes of the underlying phantom and tissue samples, for both pH_e and pH_i.

Two examples of trimodal pH_e distributions were analyzed. The three modes from our phantom could be quantified relatively easily (Fig. 2 D). However, *in vivo* situations amenable to precise quantification of three separate modes are rare; we present here an example (Fig. 2 H) in which one mode (pH_{e1}) appears as a shoulder on another mode (pH_{e3}). The pH_{e1} value could not be determined with great precision and had to be estimated. Consequently, a great deal of uncertainty is involved in area ratios based on curve fitting (deconvolution). In our example (Table 1 H), area ratios determined by integration and deconvolution differed by a factor of almost two. In fact, many real-world pH profiles are too irregularly shaped to hypothesize a specific number of modes with lineshapes amenable to curve fitting (see also Fig. 3 D for an irregular pH_i distribution). In these cases, it is more adequate to separately integrate readily recognizable sections of a pH distribution curve (for more details see Figure S5).

Global visual inspection of the trimodal distributions depicted in Figs. 2 D and H shows that they are positively skewed. This is confirmed by positive G1 values. G1 is larger for Fig. 2 H than for Fig. 2 D; in fact, the distribution of Fig. 2 H which has no significant negative tail has the highest absolute G1 value among all results presented in Table 1. Since in Fig. 2 D two low-intensity modes, pH_{e1} and pH_{e2}, are located almost symmetrically to the left and to the right of the dominating mode, pH_{e3}, the overall pH distribution is more centered (less tailed) than in a Gaussian distribution, resulting in increased peakedness (positive G2). The entropy of this pH distribution is relatively high (highest H value among all tested phantom samples

presented in Table 1) because the overall distribution is spread over a large pH range through overlapping modes, resulting in a rather even pH distribution. By contrast, in Fig. 2 H the intensities of the three strongly overlapping modes decrease steeply from left to right. This fact, in conjunction with the absence of a significant left tail on the pH_{e2} mode, results in an asymmetric but rather peaked global pH distribution (negative G2, Table 1 H) that is very narrow for a trimodal distribution. Consequently, the overall pH distribution is less smooth (relatively low H value) compared with Fig. 2 D. A similar picture arises from the asymmetric pH_i distribution of Fig. 3 D where strong asymmetry resulted in widely diverging pH_{i2}, $\overline{pH_i}$ and $\widetilde{pH_i}$ and values (Table 2 D, largest difference between principal mode and $\overline{pH_i}$). The strong asymmetry is also reflected by a high G1 value. Despite the peakedness of this distribution around its principal mode, pH_{i2}, G2 is slightly negative because of the heaviness of the tail, which also results in a large positive G1 value. Taken together, similar quantitative patterns were observed for irregular (tri- and higher modal) pH_e and pH_i distributions, and the statistical parameters suggested describe the underlying pH distributions adequately.

In a previously published study comprising mouse tumors with uni-, bi- and multimodal pH distributions we found that the median pH was more consistent within a given group of animals than was the mean pH (11). Since the median is generally more robust to outliers and extreme values than the mean (21), small foci of atypical pH values in tumors with otherwise "typical" pH distributions will influence the mean pH more than they will the median pH. Therefore, depending on the purpose of pH analysis either $\overline{pH_e}$ or $\widetilde{pH_e}$ or should be emphasized.

S1.2.11. Proton coupling in 3-APP ^{31}P NMR significantly affects measurable pH_e profiles

To test the potential and limitations for determining pH_e profiles by ^{31}P NMR spectroscopy of 3-APP, spectra were acquired for multiple combinations of 3-APP solutions at varying concentrations and pH values. First, the 3-APP fine structure of a non-decoupled

^{31}P NMR spectrum was obtained on a high-resolution spectrometer under optimal conditions. The ^{31}P-^{1}H coupling yielded a triple triplet (= triplet of a triplet, resulting in nine peaks, Fig. 4 A), as was to be expected based on the molecular structure of 3-APP (eq. S2). The two coupling constants in question were measured to be $^{2}J_{^{31}P\text{-}^{1}H}$ = 17.8 Hz and $^{3}J_{^{31}P\text{-}^{1}H}$ = 11.7 Hz. The highest peak of this multiplet comprises only 25% of the total signal intensity. Under the experimental conditions of *in vivo* spectroscopy this multiplet is not resolved and results in a single broad spectral line. Spectrometers/imagers used for *in vivo* work provide relatively poor magnetic-field homogeneity even when perfectly homogeneous solutions are measured. As a result, only five broadened 3-APP peaks could be distinguished at 4.7 T for a 3-APP solution contained in a spherical phantom (Fig. 4 B), although (i) localized shimming had been applied, and (ii) the acquisition parameters used allowed for a resolution of 3.2 Hz per point. In this spectrum, the three strongest peaks responsible for most of the 3-APP signal intensity cover a range of about 0.4 ppm at 4.7 T, corresponding to *ca.* 0.4 pH units. In fact, relevant pH_e variations within tissues may be of the same order of magnitude. Therefore, a significant improvement of the *in vivo* spectral linewidth of 3-APP would be desirable. This may be achieved by using proton-decoupled ^{31}P NMR spectroscopy provided that (i) the instrument used for measuring tissue pH heterogeneity is equipped accordingly, and (ii) the heat generated by proton decoupling can be kept at a low level lest there be a risk of interference with physiological processes. Narrower 3-APP spectral lines would not only result in better defined pH_e profiles, but also in a higher signal-to-noise ratio.

S1.2.12. pH regions derived from ^{31}P NMR-generated pH distributions *vs.* physical tissue regions

In the context of this study, it is essential to distinguish between the 'pH regions' that can be derived from pH curves on the one hand, and physical tissue regions on the other. The pH distributions obtained from ^{31}P NMR spectra represent pH profiles in the form of histograms. Therefore, they do not contain information on the spatial organization of the underlying physical tissue regions. In particular, no assumptions are made as to the physical

shapes, sizes or locations of the tissue volume elements responsible for these pH regions, beyond the trivial fact that all pH environments measured in one single-voxel ^{31}P NMR experiment are contained within the same macroscopic voxel. For instance, the appearance of two distinct pH regions in a particular pH profile may be generated by two macroscopic tissue regions, each one characterized by a distinct characteristic pH range. However, the same pH distribution pattern may be produced by a large number of microscopic tissue volumes forming a unstructured mixture of volume elements whose pH values fall in one of the distinct pH ranges represented by the pH profile. In either case, the ratios between tissue volumes characterized by different pH values can be estimated as described in Section S1.2.9.

Obviously, our spectroscopy-based technique indistinguishably takes into account macroscopic and microscopic pH heterogeneities, as opposed to heterogeneities based on pH maps ("pH imaging"). To compare the pH heterogeneities detectable by our proposed method and by pH maps, it is essential to distinguish between microscopic pH heterogeneity (microheterogeneity) and macroscopic pH heterogeneity (macroheterogeneity). Microheterogeneity refers to pH differences between tissue volume elements that are smaller than the voxel underlying a pixel in a pH map, but may be located inside the same voxel (microheterogeneity = intravoxel heterogeneity). Microheterogeneity is undetectable in pH maps, because only one pixel value per voxel is obtainable, and this pixel value will usually represent an average pH value for the entire voxel. Macroheterogeneity refers to pH heterogeneity that is detectable in pH maps because it is based on pH differences between (average) pH values of individual voxels underlying the pixel values in a pH map (macroheterogeneity = intervoxel heterogeneity). Therefore, pH imaging detects macroscopic pH heterogeneity only, while our ^{31}P NMR spectroscopy-based pH profiles represent macroscopic and microscopic pH heterogeneities combined. In the latter method, there is only one voxel that comprises the entire tissue volume of interest, unless ^{31}P NMR CSI is performed that produces one spectrum per voxel (not used in this report). In fact, the issue of macroheterogeneity *vs.* microheterogeneity is not specific to pH maps, but is common to all

voxel-based imaging. This has been highlighted in a recent paper discussing the relationship between tissue heterogeneity and anisotropy of apparent diffusion coefficients (ADC) determined by diffusion-weighted imaging (22).

S2. Spectrum acquisition and processing

S2.1. Acquisition of ^{31}P NMR spectra and proton images

Phantoms consisting of 3-APP solutions at two different pH values and varying 3-APP concentrations were prepared by serial dilution to study effects on statistical parameters describing pH heterogeneity (for qualitative results see the RESULTS section in the main body of this report, and Table S1). Changes in pH due to dilution were limited to a few hundredths pH units. These variations were not corrected for because they were well below the pH differences between the phantoms (nearly 0.5 pH units), and were not significant in the context of this study.

The NMR tubes containing our solutions were inserted into the cap of a centrifuge tube through separate, tightly drilled holes near the center of the cap. This centrifuge cap, together with the inserted NMR tubes, was then placed into the bottom part of a shortened 50-mL centrifuge tube (height ca. 3.5 cm including bottom cone). The centrifuge tube volume surrounding the NMR tubes was filled with saline. The NMR tubes were secured in a stable, upright position by firmly attaching the centrifuge tube cap to the centrifuge tube, and by pushing the NMR tubes down to the bottom of the centrifuge tube. All chemicals used for the preparation of phantoms - including also HCl and NaOH for pH adjustment - were obtained from Sigma-Aldrich (Saint Quentin Fallavier, France).

All animal images and spectra shown in the current report were produced according to the protocol described in a previous study, using a BIOSPEC spectrometer/imager with a small-bore gradient/shim system (16 cm) (11). Paravision 3 (Bruker) software was used for acquisition and processing of phantom spectra and images. After three perpendicular ^{1}H (200.5 MHz) localizer images designed to verify the exact phantom position (TRIPILOT sequence), multi-slice/multi-echo images (MSME sequence) were acquired using the

following parameters : repetition time TR = 1 s; echo time TE = 14.45 ms; hermite pulses for excitation and refocusing (2700 Hz bandwidth); effective spectral bandwidth = 32 kHz; field of view between 4 and 8 cm, depending on slice orientation; 16-18 slices per orientation; slice thickness 1 mm, in-plane matrix 128x128, resulting in a resolution varying between ca. 0.3 and 0.6 mm/pixel. The purpose of these MSME images was to exactly determine the geometric parameters needed for localized ^{31}P NMR spectroscopy. Subsequently, the magnet was shimmed (FASTMAP) on a selected volume small enough to avoid phantom regions of large susceptibility differences (*e.g.*, interfaces between air and water), but large enough to fully contain the volume of interest used in subsequent ^{31}P NMR spectroscopy. The sequence employed for acquiring localized ^{31}P NMR spectra (81 MHz) consisted of a hermite excitation pulse, preceded by outer-volume saturation (OVS) based on the geometric information provided by the MSME sequence. The homemade ^{31}P MRS sequence was composed of modules from Bruker's Paravision software, and was based on the FROGS technique (23). The acquisition parameters used: repetition time TR = 8 s; number of transients NS = 32 to 512, depending on the concentration of the solutions; acquisition time AQ = 157 ms; 1024 points; spectral width SW = 6510 Hz (= 80 ppm); spectral resolution: 3.2 Hz/pt; number of OVS slices = 6 to 8; for each OVS slice (ca. 7 mm thickness), the sech saturation pulse was repeated 50 times. The acquisition parameters for the *in vivo* tumor spectra used for comparison were similar to the acquisition parameters for phantom spectra, and have been described elsewhere in great detail (11).

Significant deformations of *in vivo* 3-APP or P_i spectral lineshapes due to poor shim can be identified by inspecting other, pH-insensitive peaks from the same ^{31}P NMR spectrum. These peaks should be equally deformed if magnetic-field inhomogeneity rather than pH heterogeneity is the origin of the observed 3-APP and/or P_i peak deformation. For instance, the phosphorylcholine (PC) signal is rather pH-insensitive, and is usually quite strong in tumor ^{31}P NMR spectra. Appropriate algorithms could be used to correct, based on the PC lineshape, 3-APP and P_i peaks for magnetic-field inhomogeneity effects. However, this method would not adequately address the contribution of ^{31}P-^{1}H coupling to the linewidth of

3-APP since the ^{31}P nucleus of PC is also proton-coupled, albeit with a smaller coupling constant than the ^{31}P nucleus of 3-APP. Consequently, future experiments should be performed with proton decoupling where possible, as recommended in Section S1.2.10.

High-resolution ^{31}P NMR spectra of 3-APP in D_2O were acquired on a 9.4 T AVANCE 400 wide-bore Fourier transform NMR spectrometer from Bruker, using a quattro nucleus probe for 5-mm tubes (QNP tunable to ^{1}H, ^{31}P, ^{13}C and ^{19}F NMR resonance frequencies). During NMR measurements, samples were spun at a spin rate of 20 Hz. A standard Bruker Eurotherm BVT-3200 variable temperature unit was employed to maintain a sample temperature of 28°C. ^{31}P NMR spectra were acquired during AQ = 0.84 s (32k). One test spectrum was performed using inverse-gated WALTZ16 proton decoupling (power setting 17 dB for a pulse width of 100 µs). Further ^{31}P NMR acquisition parameters used: sweep width SW = 120 ppm, pulse repetition time TR = 15.85 s, pulse width (90°) PW = 6.0 µs. Generally, 128 transients were acquired, resulting in a total acquisition time of 34 minutes.

S2.2. Processing of ^{31}P NMR spectra and proton images

Lorentzian-Gaussian functions were applied to ^{31}P NMR FIDs acquired from phantoms. This method, *i.e.* resolution enhancement by Lorentzian-Gaussian lineshape transformation, proved to be well suited in typical *in vivo* situations (11). By contrast, apodization led to poor separation of 3-APP signals stemming from tissue regions with varying extracellular pH, due to the intrinsically broad bases of Lorentzian lines. The two adjustable parameters of the Lorentzian-Gaussian function (line broadening, LB, and Gaussian broadening, GB) were chosen to yield linewidths similar to those obtained *in vivo* (11). Suitable LB values ranged from -10 to -25 Hz, and GB values from 0.005 to 0.015. Moderate filtering of phantom spectra would produce unrealistically narrow spectral linewidths and, therefore, artificially well-defined modes in multimodal pH profiles (*e.g.*, LB = 15 Hz, Fig. 1 C). To convert 3-APP ^{31}P NMR signals to pH curves, the improved algorithms described in Sections S1.1 and S1.2 were used. For this purpose, the real parts of the processed ^{31}P NMR spectra were extracted from the complex spectra either using Bruker's TopSpin command, convbin2asc, or via

CSIAPO (24), and saved in ASCII format. These data were transferred to our EXCEL (Microsoft, Redmond, WA, USA) spreadsheet pH_param_template.xlsx to be multiplied with the appropriate mathematical functions for ppm-to-pH conversion (eq. S6) and appropriate weighting (intensity correction, eq. S8). The analogous functions for P_i ^{31}P NMR signals are equations S5 and S7, respectively. The resulting data were plotted to generate the pH profiles (pH_e and pH_i distribution curves, respectively). These and all subsequent calculations, with the exception of deconvolutions, can be performed with the same pH_param_template.xlsx spreadsheet (available as Supplementary Material; URL address for download is given in the MATERIALS AND METHODS section of the main body of this paper). Some pH_i profiles exhibit a broadened right flank, resulting from the PME resonance partially overlapping the base of the P_i resonance. However, it is possible to prevent this distortion of the pH_i profile by a curve fitting procedure; details are described in the Instructions file imbedded in the pH_param_template.xlsx spreadsheet.

The obtained curve point intensities were rescaled as described in Section S1.2.4.4, and used as histograms for the determination of weighted pH medians (eq. S11), weighted-average pH means (eq. S10), skewness (eq. S13), kurtosis (eq. S14) and entropy (eq. 18). In addition, pH maxima were determined: single modes for unimodal pH distributions, or multiple modes for bimodal and multimodal pH distributions. These modes were compared to (i) weighted-average pH values, and (ii) pH values based on the dominating mode, corresponding to highest point of the entire pH curve investigated (classical method). For pH distributions permitting the distinction of two or more characteristic pH ranges, the areas under these individual pH ranges were integrated by summing the rescaled intensities of the digital data points in question. The relative sizes of these areas were determined and compared to results obtained by curve fitting using Bruker's TopSpin MDCON procedure (mixed Lorentzian and Gaussian (25) "deconvolution" of overlapping lines). Note that in this report pH modes are designated pH_{e1}, pH_{e2} and pH_{e3} (or pH_{i1} and pH_{i2}) for all multimodal pH distributions, regardless of whether they had been derived from animal, phantom or simulated ^{31}P NMR spectra. After validation through analysis of phantoms and numerical

simulations, these evaluation methods were applied to 3-APP and P_i signals of *in vivo* ^{31}P NMR spectra from mouse tumor xenografts (11). Since NMR spectra are available in digitized form, spectral data points were directly used for all of these procedures. Because the number of data points under a typical pH curve is moderately large (max. 100 to 200), no data reduction such as binning or bucketing was needed. Note that the selection of appropriate numbers of bins is still a matter of statistical debate; some theoreticians have attempted to determine criteria for an optimal number of bins, but these methods generally make strong assumptions about the shape of the distribution (7). For best quantitative comparison, the number n should always be greater than several times the number of digital points m present in histograms, m being the equivalent of the number of bins (for numerical examples see Figure S3, panels E to J; see also Section S1.2.6). Since intensity values produced by NMR spectrometers are usually on the order of 10^3, 10^4 or higher, it is not normally necessary to scale up the number n. Care should be taken not to overinterpret small differences in curve shape parameters between pH profiles, keeping in mind that skewness and kurtosis are also somewhat sensitive to the exact range of data points chosen, *i.e.* to the choice of the first and the last points k of the pH curve used for curve shape analysis. The determination of these points may not be trivial in cases where pH curves are derived from noisy ^{31}P NMR spectra, but can be chosen and visualized interactively using the "pH_param_template.xlsx" spreadsheet.

High-resolution ^{31}P NMR spectra of 3-APP solutions were processed using Bruker's TopSpin software. Modest apodization (LB = 0.2 Hz) was applied to the free induction decay before Fourier transform, phase and baseline corrections. Spectra were subjected to peak-picking and integration of peak areas.

Supplementary References

1. Ackerman, J. J., Soto, G. E., Spees, W. M., Zhu, Z., and Evelhoch, J. L. The NMR chemical shift pH measurement revisited: analysis of error and modeling of a pH dependent reference. Magn Reson Med, *36:* 674-683, 1996.
2. Raghunand, N. Tissue pH measurement by magnetic resonance spectroscopy and imaging. Methods Mol Med, *124:* 347-364, 2006.
3. Graham, R. A., Taylor, A. H., and Brown, T. R. A method for calculating the distribution of pH in tissues and a new source of pH error from the 31P-NMR spectrum. Am J Physiol, *266:* R638-645, 1994.
4. Koch, K. M., Rothman, D. L., and de Graaf, R. A. Optimization of static magnetic field homogeneity in the human and animal brain in vivo. Progress in nuclear magnetic resonance spectroscopy, *54:* 69-96, 2009.
5. Madden, A., Leach, M. O., Sharp, J. C., Collins, D. J., and Easton, D. A quantitative analysis of the accuracy of in vivo pH measurements with 31P NMR spectroscopy: assessment of pH measurement methodology. NMR Biomed, *4:* 1-11, 1991.
6. Rossiter, H. B., Ward, S. A., Howe, F. A., Kowalchuk, J. M., Griffiths, J. R., and Whipp, B. J. Dynamics of intramuscular 31P-MRS P(i) peak splitting and the slow components of PCr and O2 uptake during exercise. J Appl Physiol, *93:* 2059-2069, 2002.
7. Hargas, L., Koniar, D., and Stofan, S. Sophisticated biomedical tissue measurement using image analysis and virtual instrumentation. *In:* S. Folea (ed.), Practical applications and solutions using Labview™ software. Rijeka: InTech, 2011.
8. Kohler, U. and Kreuter, F. Data analysis using stata. College Station: Stata Press, 2005.
9. Sokal, R. R. and Rohlf, F. J. Biometry, 3rd edition. New York: Freeman, 1995.
10. DeCarlo, L. T. On the meaning and use of kurtosis. Psychol. Methods, *2:* 292-307, 1997.
11. Chiche, J., Fur, Y. L., Vilmen, C., Frassineti, F., Daniel, L., Halestrap, A. P., Cozzone, P. J., Pouyssegur, J., and Lutz, N. W. In vivo pH in metabolic-defective Ras-transformed fibroblast tumors: Key role of the monocarboxylate transporter, MCT4, for inducing an alkaline intracellular pH. Int J Cancer, *130:* 1511-1520, 2012.
12. Jensen, P. A. and Bard, J. F. Operations Research Models and Methods Vol. Supplementary material http://www.me.utexas.edu/~jensen/ORMM. Hoboken, NJ, USA: Wiley, 2003.
13. He, K. and Meeden, G. Selecting the number of bins in a histogram: A decision theoretic approach. J Stat Plan Inference, *61:* 49-59, 1997.
14. Frigg, R. and Werndl, C. Entropy: A Guide for the Perplexed. *In:* C. Beisbart and S. Hartmann (eds.), Probabilities in physics. Oxford: Oxford University Press, 2011.
15. Güçlü, B. Maximizing the entropy of histogram bar heights to explore neural activity: a simulation study on auditory and tactile fibers. Acta Neurobiol Exp, *65:* 399-407, 2005.
16. Butler, G. Mode. *In:* N. Salkind (ed.), Encyclopedia of Research Design, pp. 140-142. Thousand Oaks: SAGE, 2010.
17. OriginLab http://www.originlab.com. 2012.
18. Ahmad, S. and Beckett, M. Recovery from pH 6.38: lactic acidosis complicated by hypothermia. Emergency Med J, *19:* 169-171, 2002.
19. Langenbuch, M. and Portner, H. O. Changes in metabolic rate and N excretion in the marine invertebrate Sipunculus nudus under conditions of environmental hypercapnia: identifying effective acid-base variables. J Exp Biol, *205:* 1153-1160, 2002.
20. DeRubertis, F. R., Michelis, M. F., Beck, N., and Davis, B. B. Complications of diuretic therapy: severe alkalosis and syndrome resembling inappropriate secretion of antidiuretic hormone. Metabol Clin Exp, *19:* 709-719, 1970.
21. Ogunnaike, B. A. Random phenomena. Boca Raton: CRC, 2010.
22. Lope-Piedrafita, S., Garcia-Martin, M. L., Galons, J. P., Gillies, R. J., and Trouard, T. P. Longitudinal diffusion tensor imaging in a rat brain glioma model. NMR in biomedicine, *21:* 799-808, 2008.
23. Sauter, R., Mueller, S., and Weber, H. Localization in in vivo 31P NMR spectroscopy by combining surface coils and slice-selective saturation. J. Magn. Reson., *75:* 167-173, 1987.

24. Le Fur, Y., Nicoli, F., Guye, M., Confort-Gouny, S., Cozzone, P. J., and Kober, F. Grid-free interactive and automated data processing for MR chemical shift imaging data. Magma, *23:* 23-30, 2010.
25. Marshall, I., Higinbotham, J., Bruce, S., and Freise, A. Use of Voigt lineshape for quantification of in vivo 1H spectra. Magn Reson Med, *37:* 651-657, 1997.