Title: Ionizable networks mediate pH-dependent allostery in SH2 signaling proteins

Papa Kobina Van Dyck[1,2], Luke Piszkin[1], Elijah A. Gorski[1,2], Eduarda Tartarella Nascimento[1,2,3], Joshua A. Abebe[1,2], Logan M. Hoffmann[1,2], Jeffrey W. Peng[1], Katharine A. White[1,2,*]

1 Department of Chemistry and Biochemistry, University of Notre Dame
251 Nieuwland Science Hall Notre Dame, IN 46556 USA

2 Harper Cancer Research Institute, University of Notre Dame
1234 N. Notre Dame Avenue South Bend, IN 46617 USA

3 Saint Mary's College, Notre Dame, IN 46556 USA

*Corresponding Author: kwhite6@nd.edu

Introduction

Transient intracellular pH dynamics[1] regulate mammalian proliferation[2,3], migration[4], and differentiation[5]. However, for many pH-dependent cell processes, the molecular mediators are unknown[6]. Prior work identified histidine residues as molecular switches in pH-sensitive proteins, but how other ionizable residues contribute to pH-dependent protein allostery is understudied. Here, we develop an *in silico* computational pipeline to identify putative pH-sensitive proteins and their molecular mechanisms. We first apply this pipeline to SHP2, a known pH-sensitive signaling protein with an uncharacterized molecular mechanism. We show wild-type SHP2 phosphatase activity is pH-sensitive *in vitro* and in cells, and mutation of identified H116 and E252 to non-titratable alanine residues abolishes pH-sensitive function. We also show that c-Src is a previously unrecognized pH-dependent kinase, and mutation of the identified ionizable network again abolishes pH-sensitive activity. Constant pH molecular dynamics simulations support a conserved allosteric mechanism of pH-dependent binding of inhibitory SH2 domains to the functional catalytic domains of SHP2 and c-Src. We apply our computational pipeline across SH2 domain-containing signaling proteins and identify evolutionarily conserved putative pH-sensing networks. Our results reveal that pH is an allosteric regulator of SH2 domain-containing signaling proteins providing insight into normal pH-dependent cell biology and diseases where pHi is dysregulated, such as cancer.

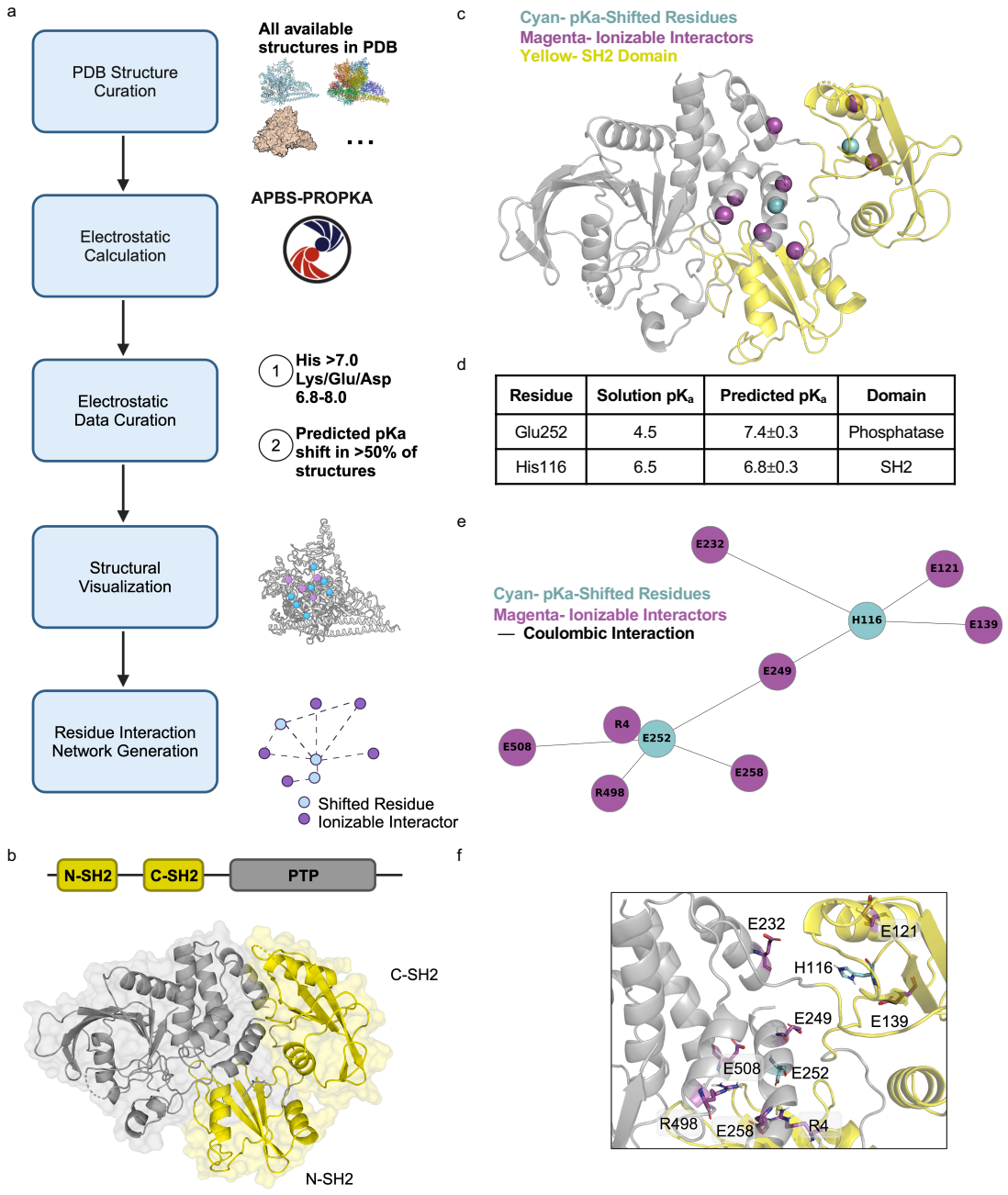
Main

All proteins have ionizable residues that can bind or release H^+, thereby changing their charged state. For an ionizable residue to have pH-dependent biological function, the H^+ binding affinity (pK_a) of one or more residues needs to be in the physiological pH range (6.8-8.0). It is not surprising that most pH-sensitive proteins (pH sensors) with characterized molecular mechanisms function through titratable histidine residues (solution pK_a of 6.5). However, the pK_as of ionizable residues are sensitive to electrostatic microenvironment[7]. Thus, local protein microenvironment can up- or down-shift pK_as of glutamates, aspartates, and lysines into the physiological pH range[8–10]. Further complicating this pH sensing landscape, ionizable residues can cooperatively regulate pH-dependent membrane proteins[11,12] and can be sequestered within cytoplasmic protein cores or buried at the interface of protein-protein complexes[13]. However, these complex networks of ionizable residues are difficult to identify *a priori* based on structure-gazing and thus many cytoplasmic proteins known to have pH-sensitive functions have unknown or incompletely characterized molecular mechanisms.

Pipeline to identify putative pH-sensitive networks

To more quickly identify pH-sensitive proteins and characterize unknown pH sensing mechanisms, we developed an *in silico* prediction pipeline to identify putative pH sensing nodes based on 3D protein structure (**Fig. 1a**, see methods for details). Briefly, we use PROPKA[14] to predict the pK_as of ionizable groups based on available structure(s) of a protein of interest. Next, we filter these results for buried ionizable residues with physiological pK_as and generate 2D and 3D maps showing the identified residue networks and other ionizable coulombic interactors. We can perform this pipeline on multiple full-length structures of the protein of interest, increasing

likelihood of identifying functional networks. Unlike similar approaches arising from work on membrane proteins, our approach is agnostic to protein localization and does not require a path for structural waters from bulk solvent to the identified buried ionizable networks[11,15].



Residue	Solution pK_a	Predicted pK_a	Domain
Glu252	4.5	7.4±0.3	Phosphatase
His116	6.5	6.8±0.3	SH2

Figure 1. New pipeline to identify pH-sensitive networks predicts SHP2 His116 and Glu252 have upshifted pK_as. (a) Schematic for *in silico* pK_a prediction method for proteins with solved structures (see text and methods for details). Briefly, all available structures in the protein database are curated, electrostatic properties are calculated using PROPKA, results are filtered

for ionizable residues with physiologically relevant predicted pK_a values, and data are visualized in a 3D structure or a 2D residue interaction network. (b) Crystal structure of SHP2 shown in cartoon and surface format (PDB ID:2SHP). Protein tyrosine phosphatase (PTP) domain colored in grey, SH2 domains colored in yellow. (c) Structure of SHP2 (PDB ID:2SHP) with protein tyrosine phosphatase (PTP) domain in grey and SH2 domains in yellow. Residues identified through *in silico* ionizable network prediction pipeline shown in spheres. Residues with predicted pK_a shifts (cyan) cluster with ionizable interactors (magenta) across the phosphatase-SH2 domain interaction interface of SHP2. (d) Table of predicted pK_as for cyan residues identified using *in silico* ionizable network prediction pipeline on 47 SHP2 structures (mean ± SD). (e) Residue interaction network of residues with predicted pK_a shifts (cyan) and their ionizable interactors (magenta). Length of edges reflect the strength of the coulombic interaction, with stronger coulombic interactions having shorter edge lengths (f) Zoom of SHP2 structure at the PTP-SH2 interaction interface. Networked residues from a and b are shown in stick. Residues with predicted pK_a shifts in cyan and ionizing interactors in magenta.

Analysis pipeline applied to SHP2 identifies His116 and Glu252 with predicted upshifted pK_as

We first applied our analysis pipeline to the SH2 domain-containing signaling protein SHP2. SHP2 is a regulator of multiple pH-sensitive processes including cell proliferation, migration, differentiation, growth, and survival[16–18]. SHP2 has also been shown to have pH-dependent phosphatase activity *in vitro*, with optimal activity at pH 7.0 and decreasing activity both below and above 7.0[19], suggestive of multi-group titration. While SHP2 activity is pH sensitive *in vitro*, the pH-dependent molecular mechanism has not been identified. While prior work showed that SHP2 activity positively regulates endothelial cell migration and angiogenesis and vascular smooth muscle cell function[17], the role for pH-sensitive SHP2 function in cells is less clear.

SHP2 consists of three globular domains: a protein tyrosine phosphatase (PTP) domain responsible for catalyzing the dephosphorylation of tyrosine-phosphorylated proteins, and two Src Homology 2 (SH2) domains that serve as phosphotyrosine (pTyr)-recognition domains[20] (**Fig. 1b**). In the absence of a tyrosine-phosphorylated binding partner, the N-terminal SH2 domain binds directly to the phosphatase domain, blocking the active site[21]. Conversely, upon ligand-induced receptor activation or signaling events, the SH2-PTP interaction is disrupted

when the SH2 domain is displaced by a tyrosine phosphorylated binding partner[21]. Thus, the SH2 domains control SHP2 activity by allosterically inhibiting the PTP domain in the bound state and influencing SHP2 localization via direct interactions with phosphorylated binding partners when signaling active[21].

Because SHP2 has no known molecular mechanism of pH-sensing, our goal was to identify the putative pH sensing nodes driving SHP2 pH-dependent phosphatase activity. We processed 47 structures of SHP2 available in the PDB through the computational pipeline. Importantly, there was no filtration of the structural input data: SHP2 structures included active conformations, inhibited conformations, and SHP2 proteins with gain- or loss-of-function mutations. Our analysis revealed a conserved network of resides with predicted physiological pK_as clustering at the interface between the SH2 and PTP domains of SHP2 (**Fig. 1c**). The two residues predicted to have upshifted pK_as were His116 and Glu252, with predicted pK_as of 6.8±0.3 and 7.4±0.3, respectively (**Fig. 1d**). His116 is in the C-SH2 domain and directly interacts with a cluster of glutamate residues (E121, E139, E232, and E249). Similar glutamate "cages" have been identified in other wild-type pH sensors including talin[22] and RasGRP1[23]. Glu252 is in the phosphatase domain at the interaction interface with the N-SH2 domain and interacts with R4, E249, E258, R498, and E508 (**Fig. 1e,f**). We next determined the structural conservation of the identified residues using Consurf[24]. Analyses of structural conservation frequently can reveal functionally important regions, such as catalytic and ligand-binding sites[24]. We found that the C-SH2 and N-SH2 are both highly conserved at the interaction interface with the PTP domain (**Extended Data Fig. 1a**). We also determined that His116 and Glu252 are sequence-conserved across SHP2 from lower-order organisms (**Extended Data Fig. 1a-c**). Taken together, these data

suggest that the SH2 domain binding interface in SHP2 has an evolutionarily conserved network of ionizable residues with predicted physiological pK_as.

SHP2 pH sensitive function requires both His116 and Glu252

We hypothesized that the previously uncharacterized pH-dependent molecular mechanism of SHP2 requires the identified ionizable network. We first assessed the *in vitro* phosphatase activity of wild-type SHP2 under different buffer pH (pH 6.1-8.0) using a generic para-nitrophenyl phosphate (PNPP) substrate. We found that wild-type SHP2 phosphatase activity is pH dependent, with maximal activity at pH 6.7 and reduced activity both above and below this value (**Fig. 2a, c, Extended Data Fig. 2a**). We found that k_{cat} increased from pH 6.1 through 6.7 and decreased from 7.0 through 8.0 forming a biphasic activity curve **(Fig. 2c)**, while K_m for the PNPP was pH-insensitive (**Extended Data Fig. 2a**). This result confirmed previous *in vitro* characterization of SHP2 activity[19], showing similar apparent pK_as of ~6.7 and ~7.1 in both the k_{cat} (**Fig. 2a, c, Extended Data Fig. 2b**) and catalytic efficiency (k_{cat}/K_m) pH profiles (**Extended Data Fig. 2c**). Reactions with bovine serum albumin (BSA) addition showed no pH-dependence, confirming that buffer pH is not independently affecting non-enzymatic PNPP hydrolysis (**Extended Data Fig. 2d**). To test whether the identified network is responsible for SHP2 pH-dependent activity, we generated a SHP2 double mutant where both H116 and E252 were mutated to non-titratable alanine residues. We repeated the phosphatase assay and found that the activity of SHP2-H116A/E252A was pH-independent (**Fig. 2b**). While k_{cat} of SHP2-H116A/E252A was reduced compared to WT SHP2, there was no pH-dependence of k_{cat} or K_m (**Fig. 2b, c, Extended Data Fig. 2a)**. This demonstrates that H116 and E252 are required for pH-dependent activity of SHP2.

Figure 2

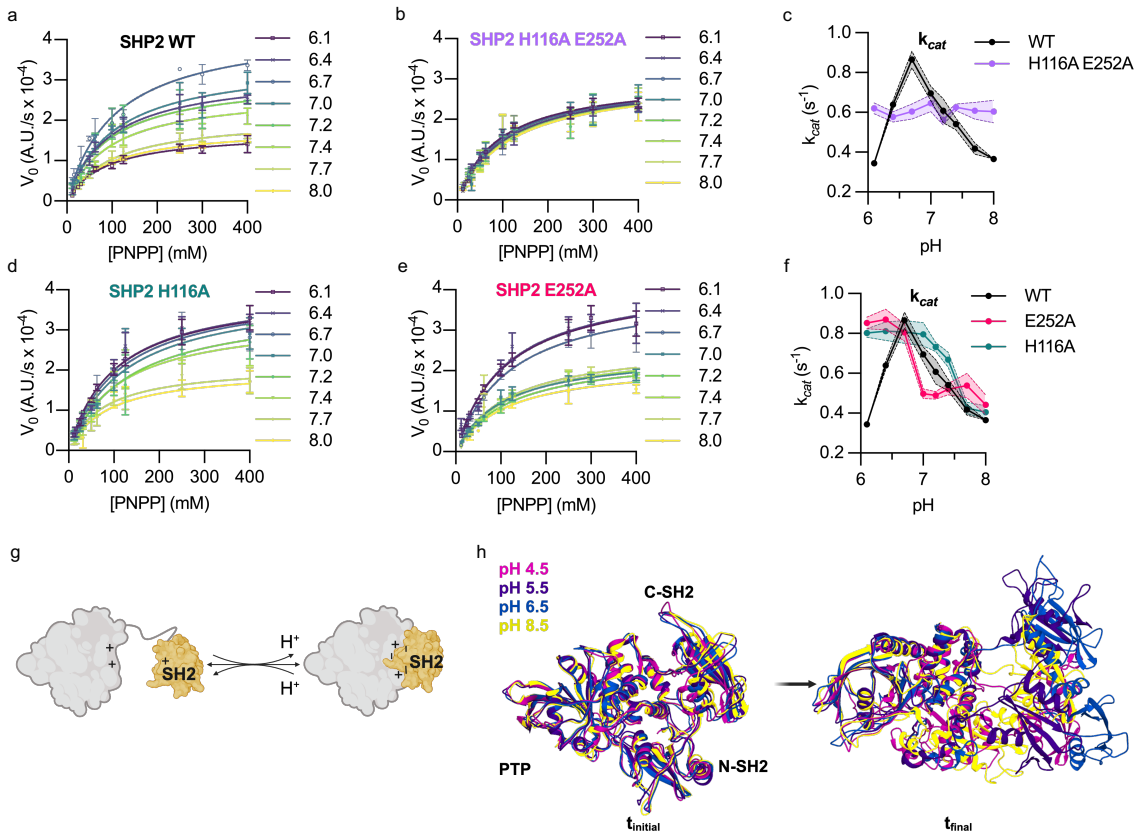


Figure 2. Molecular mechanism of SHP2 pH sensing requires both His116 and Glu252. (a) Wild-type (WT) SHP2 *in vitro* phosphatase activity curves with increasing concentrations of generic substrate p-Nitrophenyl Phosphate (PNPP) at buffer pH ranging from 6.1 to 8.0. (mean ± SEM; N=3 from ≥2 different protein preparations) (b) Double mutant (H116A/E252A) SHP2 *in vitro* phosphatase activity, assays performed as in A. (mean ± SEM; N=3 from ≥2 different protein preparations) (c) Plot of k_{cat} vs. pH for WT and double mutant (H116A/E252A) SHP2 activity. Calculated from activity curves in a and b. (mean ± SEM) (d) Single-mutant H116A-SHP2 *in vitro* phosphatase activity, assays performed as in a. (mean ± SEM; N=3 from ≥2 different protein preparations) (e) Single-mutant E252A-SHP2 *in vitro* phosphatase activity, assays performed as in a. (mean ± SEM; N=3, from ≥2 different protein preparations) (f) Plot of K_{cat} vs. pH for WT and double mutant (H116A/E252A) SHP2 activity. Calculated from activity curves in a, d and e. (mean ± SEM) (g) Proposed pH-sensing mechanism where SH2 domain (yellow) binding to catalytic domain (grey) is titratable by pH. (h) CpHMD (see methods for details) was performed on SHP2 at pH values from 4.0-10.0 (see supplemental videos). Shown are overlapping views of SHP2 structures at the start of CpHMD simulation (t = 0 ns) and end of simulation (t = 8 ns) for pH values 4.5 (pink), 5.5 (purple), 6.5 (blue) and 8.5 (yellow).

Our *in silico* pK_a prediction analysis suggests that H116 and E252 function in a network of conserved ionizable residues. To probe this interaction network in more detail, we next generated the single point mutants (SHP2-H116A and SHP2-E252A) and found that both mutants had pH-dependent phosphatase activity, each with a single apparent pK_a (**Figure 2d-f**). For both mutants, k_{cat} was maximal at low buffer pH and reduced with increasing pH (**Extended Data Figure 2a, c**). We predicted that the pH-dependent activity observed in the single point mutants reflects the pK_a of the remaining titratable residue. We compared apparent pK_as of the mutants and found that SHP2-H116A mutant had an apparent pK_a of ***7.5*** (**Fig. 2f, Extended Data Fig. 2a, e**)**,** which correlates with the higher apparent (pK_{a2}) of wild-type SHP2 (~7.1) and the *in silico* predicted pK_a of E252 (~7.4) (**Fig. 1d**). The E252A mutant has a lower apparent pK_a of 6.8 (**Fig. 2f, Extended Data Fig. 2a, e**), which matches the lower apparent (pK_{a1}) of the wild-type SHP2 (~6.7), and the *in silico* predicted pK_a of His116 (6.8) (**Fig. 1d).** This suggests that in the H116A point mutant, E252 drives the observed pH-dependent activity while in the E252A point mutant, H116 drives the observed pH-dependent activity. Compared to the wild-type SHP2 activity profiles, this suggests that H116 titration regulates activity between pH 6.1 to 7.0 while E252 titration regulates activity between pH 7.0 to 8.0. Thus, Glu252 and His116 function cooperatively to confer wild-type SHP2 pH-dependent activity.

Because E249 interacts with both Glu252 and His116, we next explored the role this bridging glutamate plays in the pH sensing mechanism of SHP2. Interestingly, we observed similar pH-dependent activity profiles of SHP2-E249A compared to WT (**Extended Data Fig. 2a, f**). The E249A mutant was still pH sensitive with biphasic k_{cat} and catalytic efficiency pH profiles (**Extended Data Fig. 2c, g**). However, the apparent pK_{a1} and pK_{a2} of the E249A mutant were both shifted higher (6.8 and 7.7) when compared to wild-type SHP2 (6.7 and 7.1)

(**Extended Data Fig. 2a, c**). This suggests that E249 is not acting as a direct proton conduit or relay between the identified pH sensing residues (H116 and E252) but instead tunes the pK_a values of these residues.

CpHMD reveals pH-dependent residue conformations and SH2 domain release in SHP2

Due to the clustering of pH-sensitive residues at the SH2 and PTP domain interface, we hypothesized that pH-dependent binding of the inhibitory SH2 domains drives pH-dependent SHP2 activity (**Fig. 2g**). We used continuous constant pH molecular dynamics (CpHMD) to probe pH-dependent structural conformatoins of wild-type SHP2. In traditional MD simulations, solvent pH is considered implicitly by fixing protonation states based on empirical or predicted residue pK_a values. However, protein conformational dynamics may depend explicitly on residue protonation states, so conventional MD methods cannot capture how dynamic changes in residue protonation affect structural dynamics. CpHMD addresses this gap by sampling protonation and conformation states simultaneously[25] and lends insight into proton-coupled mechanisms that are not captured in conventional MD methods.

We performed CpHMD simulations of SHP2 using the GBNeck2 implicit solvent model for pH values ranging from 4.0 to 10 (in increments of 0.5) (based on[25], see methods; trajectories in **Videos 1-13**). Each simulation was run for 8 nanoseconds, allowing for convergence of backbone RMSD and protonation states, for an aggregate 104 nanoseconds of total simulation time. In agreement with the *in vitro* results, the pK_a calculated from SHP2 CpHMD showed that H116 had a lower pK_a (6.0) compared to E252 (8.4) (**Extended Data Figure 3a, b**). In addition, we observed and quantified coordinated pH-dependent changes in the side chain conformations of H116 and E252. (**Extended Data Fig 3c-d**). At low (4.5) and high (8.5) pH, H116 and E252

are pointed toward each other, but at optimal pH for activity (6.5), E252 flips out and H116 swings in, disrupting its local helix (**Extended Data Fig. 3c-d**). From the CpHMD trajectories, we noted that the SH2 domain remains rigid and tightly bound to the PTP domain at pH values less than 5.5 and higher than 8.0 (**Fig. 2h**, **Video4 (2SHP_pH5.5) and Video9 (2SHP_pH8.0)**). However, at optimal activity pH (6.5) the SH2 domain is more dynamic and loosely bound to the PTP domain (**Fig 2h, Video6 (2SHP_pH6.5**)). To assess interdomain conformations, we quantified the center of mass distances between the N-SH2 or C-SH2 domain and the PTP domain during the final 3 nanoseconds of the trajectories. We found distinct pH-dependent dynamics between the N-SH2 and PTP domain, with maximal interdomain distances (65Å) at pH 6.5 and minimal distances (30Å and 35Å) at pH 4.5 and 8.5 respectively (**Extended Data Fig. 3f**). We observed that dynamics between the C-SH2 and PTP domain followed the same pH-dependent trend, with maximal distances (54Å) achieved at pH 6.5 **(Extended Data Fig. 3g).** While interdomain orientations were not included in this analysis, the distance trends alone show pH-dependent SH2 domain release and are in striking agreement with pH-dependent *in vitro* SHP2 activity experiments

Intracellular pH regulates SHP2 activity and downstream signaling in MCF10A cells

We next determined whether SHP2 activity is regulated by physiological pH changes in cells. We first experimentally manipulated intracellular pH (pHi) in normal breast epithelial cells (MCF10A), using inhibitors of the sodium proton exchanger (EIPA) and sodium bicarbonate transporter (S0859) to lower pHi from 7.45±0.12 to 7.10±0.01 and 30 mM ammonium chloride to raise pHi to 8.00±0.27 (**Fig. 3a**). After one hour of pHi manipulation, we assayed for endogenous SHP2 basal activity by immunoblotting for phosphorylation of SHP2 at Y542,

which is required for activity[26]. We found that endogenous SHP2 activity is pH sensitive, with increased pY542 levels at low pHi compared to control and high pHi (**Fig. 3b, c**). To investigate downstream signal transduction, we measured SHP2 binding to its signaling-active binding partner, Gab1[27] using a co-immunoprecipitation assay. SHP2 immune complexes contained significantly more Gab1 when pHi was low compared to control and high pHi (**Fig. 3b, d**). This demonstrates that endogenous SHP2 has pH-dependent activity in MCF10A cells, with higher activity and binding to the Gab1 signaling partner at low pHi compared to control and high pHi.

We next confirmed the effect of pHi dynamics on basal activity of endogenous SHP2 using a live, single-cell reporter of SHP2 activity (tagBFP-Grb2)[28]. We validated that the tagBFP-Grb2 reporter was robustly recruited to the cell membrane upon EGF stimulation (active SHP2) and that this recruitment was abolished by both a specific SHP2 inhibitor (PHPS)[29] and an upstream EGFR inhibitor (Erlotinib) (**Extended Data Fig. 4a**). We also validated inhibitor efficacy and specificity using western blots to assess EGFR and SHP2 phosphorylation (**Extended Data Fig. 4b**). Erlotinib treatment reduced EGF-induced pEGFR (pY1068; Grb2 phosphosite) and pSHP2 (pY542), whereas PHPS reduced pSHP2 while leaving pEGFR levels unaffected (**Extended Data Fig. 4b**). These data show that membrane recruitment of the biosensor is dependent on SHP2 activity.

Figure 3

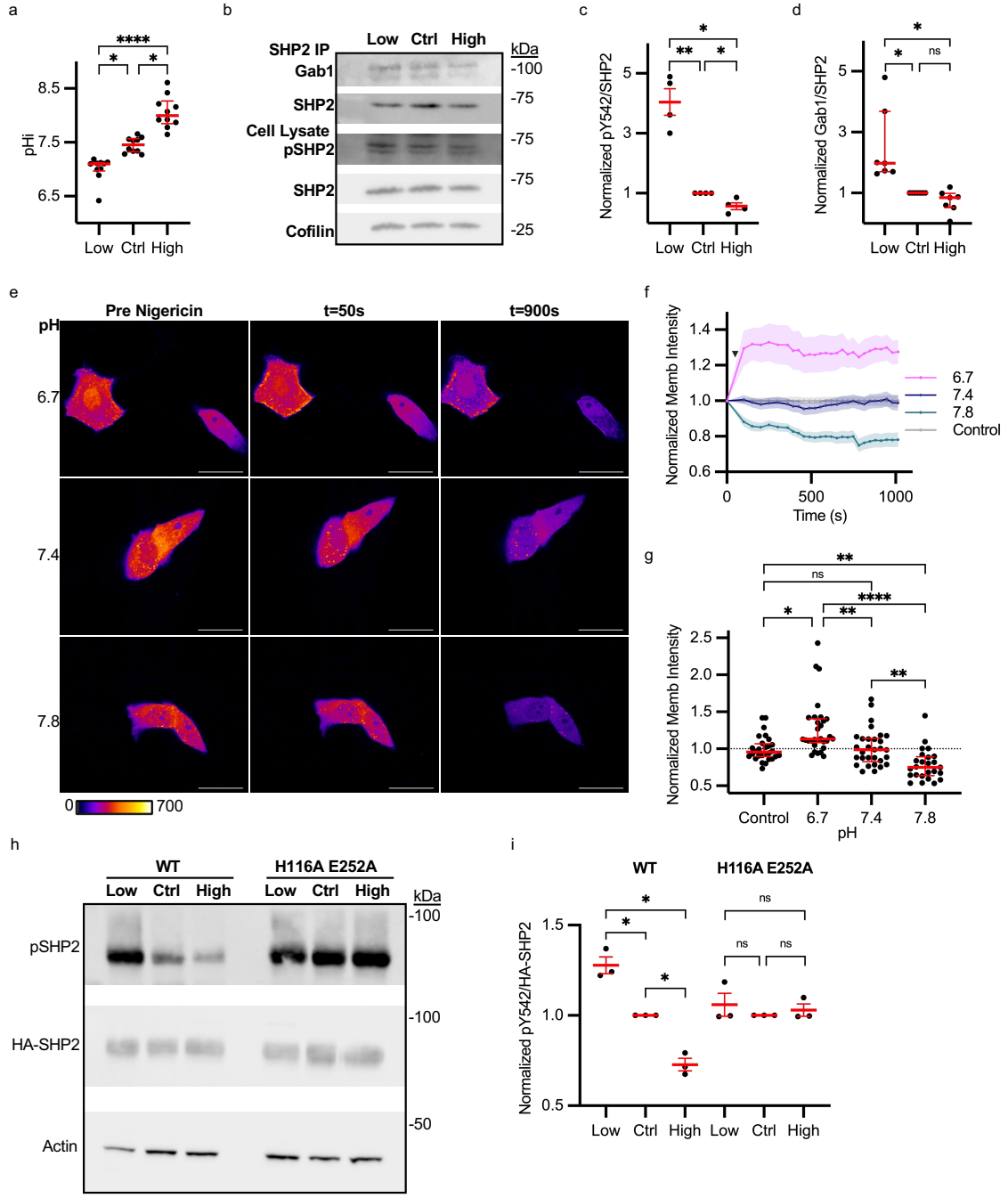


Figure 3. Intracellular pH regulates SHP2 activity and downstream signaling in MCF10A cells (a) pHi measurements of MCF10A cells. Cells were treated with 25 µM EIPA + 30 µM S0858 for 1 hour to lower pH to 7.10. To raise pHi, cells were treated with 30 mM ammonium chloride for 1 hour to raise pH to 8.00. Untreated cells had a pHi of 7.45. Scatter plot shows (median ± interquartile range, N =10) (b) Representative immunoblots of SHP2, Gab1, and phospho-SHP2 (pY542) from SHP2 immune complexes (SHP2 IP) or whole cell lysates (Cell

Lysate) isolated from MCF10A cells prepared as in A. (c) Quantification of replicate data collected as in B. Data was normalized to control in each biological replicate. Scatter plot shows (mean ± SEM, N=4). I (d) Quantification of Co-IP of SHP2 shown in b. Immunoblot intensities in the treatment conditions were normalized to control in each biological replicate. Scatter plot shows (mean ± SEM, N=7). (e) Representative images of MCF10A cells expressing the SHP2 activity reporter (Grb2 TagBFP) pseudocolored on an intensiometric scale. Images show cells prior to manipulating pHi with nigericin buffer (Pre Nigericin) (see methods for details), 50s - after manipulating pHi, and 900s after manipulating pHi. Scale bars: 25µm (f) Quantification of images as in E. Membrane intensity of SHP2 activity reporter was photobleach-corrected and then normalized to initial intensity over time. Line trace shows from single-cell data (mean ± SEM) (6.7 pH, n=30, 7.4 pH, n=30, 7.8 pH, n=25, control, n=28) collected across N=3 biological replicates. (g) Quantification of endpoint membrane intensities of single cells collected as described in f. Scatter plot shows (median ± interquartile range, N = 5) (h) Representative immunoblot of lysates prepared from MCF10A cells expressing either WT SHP2 or H116A/E252A SHP2 and treated as described in a. Immunoblots show total and pY542-SHP2 under low, control, and high pHi conditions. (i) Quantification of replicate data collected as in h. Scatter plots show (mean ± SEM, N=3). Intensities were normalized to the corresponding control condition. For a and g significance was determined using the Kruskal-Wallis test. For c, d, and i significance was determined using a ratio paired t-test to compare between treatment conditions and a one-sample t-test to compare to control. * $p<0.05$, ** $p<0.01$, *** $p<0.001$, **** $p<0.0001$.

After validating tagBFP-GRB2 as a specific SHP2 activity biosensor, we next measured biosensor recruitment to the MCF10A cell membrane under basal (unstimulated) conditions with and without pHi manipulation using the protonophore nigericin (see methods). When the pHi of the MCF10A cells was pinned to 6.7, we observed increased biosensor recruitment to the cell membrane, indicating increased SHP2 activity (**Fig. 3e-g).** Conversely, we observed no change in biosensor recruitment when MCF10A cell pH was pinned to basal pHi conditions (7.4) and reduced biosensor recruitment when the pHi of MCF10A cells was pinned at 7.8 (**Fig. 3f-g)**. Importantly, initial biosensor intensity at the membrane was not different prior to nigericin-based pHi manipulation (**Extended Data Fig. 4c**) and increased biosensor recruitment at low pHi was abolished with the addition of PHPS (**Extended Data Fig. 4f**). This shows that endogenous SHP2 basal activity is sensitive to pHi dynamics in single living cells.

To confirm that the pH sensing network of H116 and E252 mediates pH-dependent SHP2 activity in cells, we overexpressed HA-tagged WT and H116A/E252A SHP2 in MCF10A cells and quantified SHP2-pY542 levels with and without pHi manipulation. We found that overexpressing WT SHP2 recapitulated the endogenous SHP2 results, with increased activity (SHP2-pY542) at low pHi and decreased activity at high pHi compared to control MCF10A (**Fig. 3h, i**). As expected, the double mutant SHP2-H116A/E252A had pH-independent activity (SHP2-pY524) in MCF10A (**Figure 3h, i**), confirming that His116 and E252 are essential mediators of pH-sensitive SHP2 activity. Through our cell-based analyses, we show that SHP2 activity is sensitive to physiological changes in pHi, with increased basal activity at low pHi, and that H116 and E252 mediate pH-dependent allostery in SHP2.

pH sensing nodes cluster at SH2 interface in other SH2 signaling proteins

SH2 domains are modular signaling domains that bind to phosphotyrosine residues, which have experimentally determined pK_as near the physiological range (5.7-6.1)[30,31] and are highly structurally and sequence-conserved across signaling proteins[20]. We hypothesized that networks of residues with physiological pK_as serve a functional role in regulating pH-dependent SH2 binding to phosphotyrosine residues across SH2 domain-containing proteins. To explore this, we applied our *in silico* computational pipeline to a collection of SH2 domain-containing proteins with multiple solved structures. We found that most SH2 domain-containing proteins tested had ionizable residue networks with pK_as predicted to be shifted into the physiological pH range (**Extended Data Fig. 5**). As we observed with SHP2, these ionizable residue networks clustered at the SH2 binding interfaces (**Extended Data Fig. 5**). Interestingly, we observed this pattern of clustering across functional protein classes—finding it in phosphatases (SHP1)

(**Extended Data Fig. 5a**), kinases (ZAP70, JAK1 JAK2, and ABL1) (**Extended Data Fig. 5b-d, h**), transcription factors (STAT6, STAT5A) (**Extended Data Fig. 5e-f**), and GTPase activating proteins (CHIO) (**Extended Data Fig. 5g**).

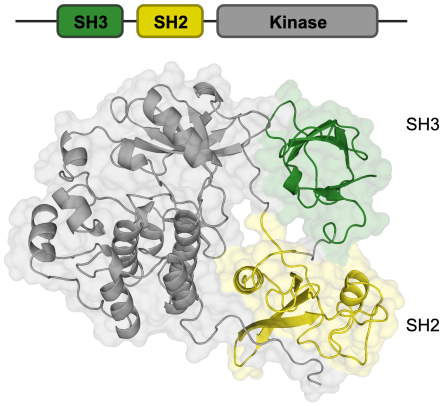
As a control, we also analyzed a collection of SH3 domain-containing signaling proteins. SH3 domains are modular signaling domains that are similarly highly evolutionarily and structurally conserved[32], but bind proline-kink regions[33] and electrostatics should not drive function. While we identified ionizable networks with physiological pK_as in some SH3 domain-containing proteins, the shifted residues did not cluster at the binding interfaces of SH3 domains (**Extended Data Fig. 5h, i-k**). This analysis suggests that ionizable networks with upshifted pK_as are not a general feature of all signaling domain interaction interfaces, but instead are enriched at functional SH2 domain interfaces.

Analysis pipeline applied to c-Src identifies four networked ionizable residues with predicted physiological pK_as

We next characterized c-Src to further validate the role of pH in the allosteric regulation of SH2 domain-containing signaling proteins. The signaling kinase c-Src has three globular domains: a kinase domain (SH1), an inhibitory SH2 domain that binds phosphotyrosines, and an inhibitory SH3 domain that binds proline-rich regions (**Fig. 4a**). The SH2 domain inhibits kinase activity when bound intramolecularly to the C-terminal phosphotyrosine (pY527)[33]. However, when the SH2 and SH3 domains are released, pY527 is dephosphorylated, the activation loop unfurls, and the SH1(kinase) domain autophosphorylates a conserved tyrosine (Y416) in the activation loop (A-loop). Activation loop unfolding and phosphorylation is required for full activation and facilitates c-Src binding to downstream targets[33].

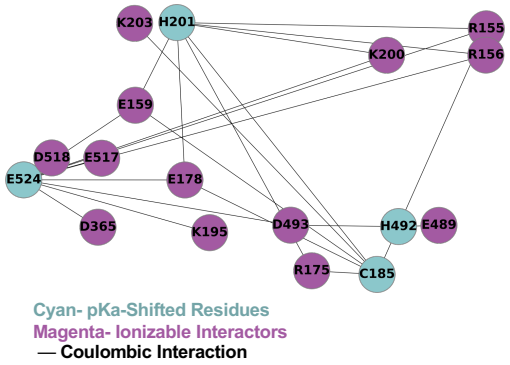
Figure 4

a

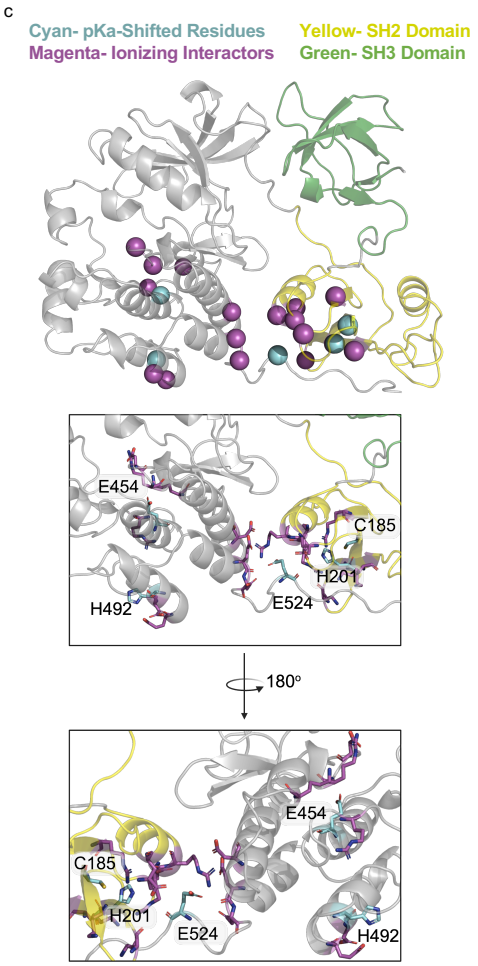


b

Residue	Solution pK_a	Predicted pK_a	Domain
Cys185	9	7.7 ± 0.0	SH2
His201	4.5	7.8 ± 0.1	SH2
Glu454	4.5	6.9 ± 0.4	Kinase
His492	6.5	7.0 ± 0.5	Kinase
Glu524	4.5	7.3 ± 0.4	Kinase



c



d

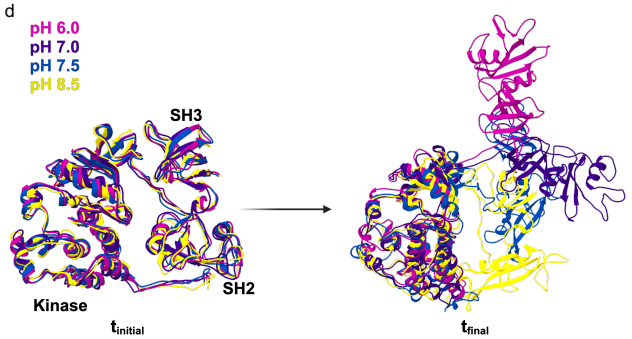


Figure 4: ***Residues with predicted pK_a shifts cluster at the binding interface of the SH2 and kinase domains in c-Src*** (a) Crystal structure of pilot protein c-Src shown in cartoon and surface format (PDB ID:2SRC). Protein Kinase domain colored in grey and SH2 domain colored in yellow, and SH3 colored in green (b) Table of predicted pK_as for shifted residues (mean ± SD).

Residue interaction network of residues with predicted pK_a shifts (cyan) and their coulombic interactors (magenta). Edge length reflects the strength of the interaction, with stronger coulombic interactions having shorter edge lengths. (c) Structure of c-Src (PDB ID:2SRC) showing the kinase domain in grey, SH2 domain in yellow, and SH3 domain in green. Residues identified through *in silico* ionizable network prediction pipeline shown in spheres. Residues with predicted pK_a shifts (cyan) cluster with ionizable interactors (magenta) across the kinase-SH2 domain interaction interface of c-Src. Zoom in shows c-Src structure at the kinase-SH2 domain interaction interface. Networked residues from are shown in stick: residues with predicted pK_a shifts in cyan, interactors in magenta. (d) CpHMD (see methods for details) was performed on Src at pH values from 4.0-10.0. Shown are overlapping views of Src structures at the start of CpHMD simulation (t = 0 ns) and end of simulation (t = 8 ns) for pH values 6.0 (pink), 7.0 (purple), 7.5 (blue) and 8.5 (yellow).

We processed seven structures of c-Src available in the PDB through the computational pipeline. Our analysis identified 5 residues predicted to have upshifted pK_as: C185 (pK_a of 7.7 ± 0.0), H201 (pK_a of 7.8 ± 0.1), E454 (pK_a of 6.9 ± 0.4), H492 (pK_a of 7.0 ± 0.5), and E524 (pK_a of 7.3 ± 0.4) (**Fig. 4b, c)**. Four residues (C185, H201, E524, and H492) are networked in a cluster at the interface between the SH2 domain and the kinase domain (**Fig. 4b, c**). E524 is in the C-terminal tail of c-Src and electrostatically interacts with H201 and C185 in the SH2 domain through bridging glutamate residues (E159, E178) and interacts with H492 in the kinase domain through a bridging aspartate residue (D493) (**Figure 4b, c**). While E454 has a predicted physiological pK_a, it is in the kinase catalytic site and distant from the SH2 regulatory domain interface (**Fig. 4b**). This analysis of c-Src again shows that ionizable residue networks with predicted physiological pK_as cluster at SH2 but not SH3 domain interfaces. We again determined the structural conservation of the identified residues in c-Src and found very high structural conservation of the ionizable network region and individual residues (**Extended Data Fig. 6a-c**). The COSMIC database[34] also showed recurrent charge-changing somatic mutations at each of these sites: C185, H201, H492, and E524 (**Extended data Fig. 6d**), suggesting functional importance of the identified network and potential for network disruption in cancer.

CpHMD reveals pH-dependent residue conformations and SH2 domain release in c-Src

We next assessed pH-dependent structural dynamics in c-Src using CpHMD (see methods; trajectories in **Videos 14-26**). While we were not able to generate a pK_a for C185 using the CpHMD trajectories (**Extended Data Fig. 7a**), all other identified residues had upshifted and near-physiological pK_as calculated from the CpHMD trajectories: (H201, pK_a of 7.3; E454 pK_a of 8.2; H492 pK_a of 7.0; E524 pK_a of 6.8) (**Extended Data Fig. 7b-e**). Local side chain movement showed pH-dependent changes in the identified residue network, suggesting pH-dependent interaction pairs (**Extended Data Fig 7f,g**). Dihedral angle changes with pH for all identified residues, with coupled dynamics observed for connected residues (H201, C185, and H492) between pH 6.5 and 7.5 and E524 dihedral angles changing at pH above 7.0 **(Extended Data Fig 7f**). The CpHMD trajectories also show pH-dependent changes in interdomain interactions between c-Src kinase domain and the inhibitory SH2 and SH3 domains (**Fig. 4d**). In agreement with the SHP2 dynamics, the SH2 domain of c-Src was bound more tightly to the kinase domain at high pH (8.5) compared to low pHi (6.0) where both the SH3 and SH2 domain were unbound (**Fig. 4d, Video18 (2SRC_pH6.0), Video23 (2SRC_pH8.5)**). We also observed intermediate conformations in 7.0 and 7.5 trajectories, with c-Src adopting a conformation where the SH3 domain binds more tightly while the SH2 domain is more dynamic (**Fig 4d, Video20 (2SRC_pH7.0), Video 21 (2SRC_pH7.5**)). To assess interdomain conformations, we quantified interdomain distances between either the SH2 or SH3 domain and the central kinase domain of c-Src during the final 3ns of the trajectories. We found pH-dependent SH2-kinase interdomain distances with a maximum (least binding) at pH 6.0 and minimums at high (8.5) and low (4.5) pH (**Extended Data Fig 7h)**. We observed similar pH-dependent trends in pH-dependent SH3-kinase interdomain distances, with a maximum (least binding) distance at pH 5.5 and minimums

at high (8.5) and low (5.0) pH (**Extended Data Fig 7i**). To rule out pH-dependent general structural compaction in these CpHMD trajectories, we also assessed SH3-SH2 domain distances and found no pH dependence (**Extended Data Fig. 7j**). While these short trajectories are less likely to capture large protein domain movements, the trends across the SHP2 and c-Src CpHMD analyses are strikingly consistent: low physiological pH (6.0-7.0) leads to SH2 release from the catalytic domain compared to high physiological pH (7.0-8.0) where SH2-catalytic interdomain interactions are increased.

Intracellular pH regulates c-Src activity and signaling in MCF10A cells

We next determined whether basal activity of endogenous c-Src is sensitive to physiological pH changes in normal epithelial cells. We assessed c-Src activity by immunoblot quantification of the levels of pY527, a proxy for inhibited c-Src, and pY416, a proxy for active c-Src (**Fig. 5a**). We found that endogenous c-Src activity was pH sensitive, with increased activity at low pHi compared to control and high pHi (**Fig. 5b-d**). Importantly, the trends in pY527 and pY416 both support the hypothesis that low pHi leads to SH2 domain release, loss of the inhibited c-Src phosphosite (pY527) (**Fig. 5c**), and gain of the active c-Src phosphorylation site (pY416) (**Fig. 5d**). We also performed immunofluorescence labeling and single-cell analysis of c-Src activity in MCF10A cells and found that active c-Src (pY416) was increased at low pHi and decreased at high pHi compared to control cells (**Fig. 5e, f**). Additionally, c-Src-pY416 showed increased membrane accumulation at low pHi compared to both control and high pHi (**Fig. 5e**).

Figure 5

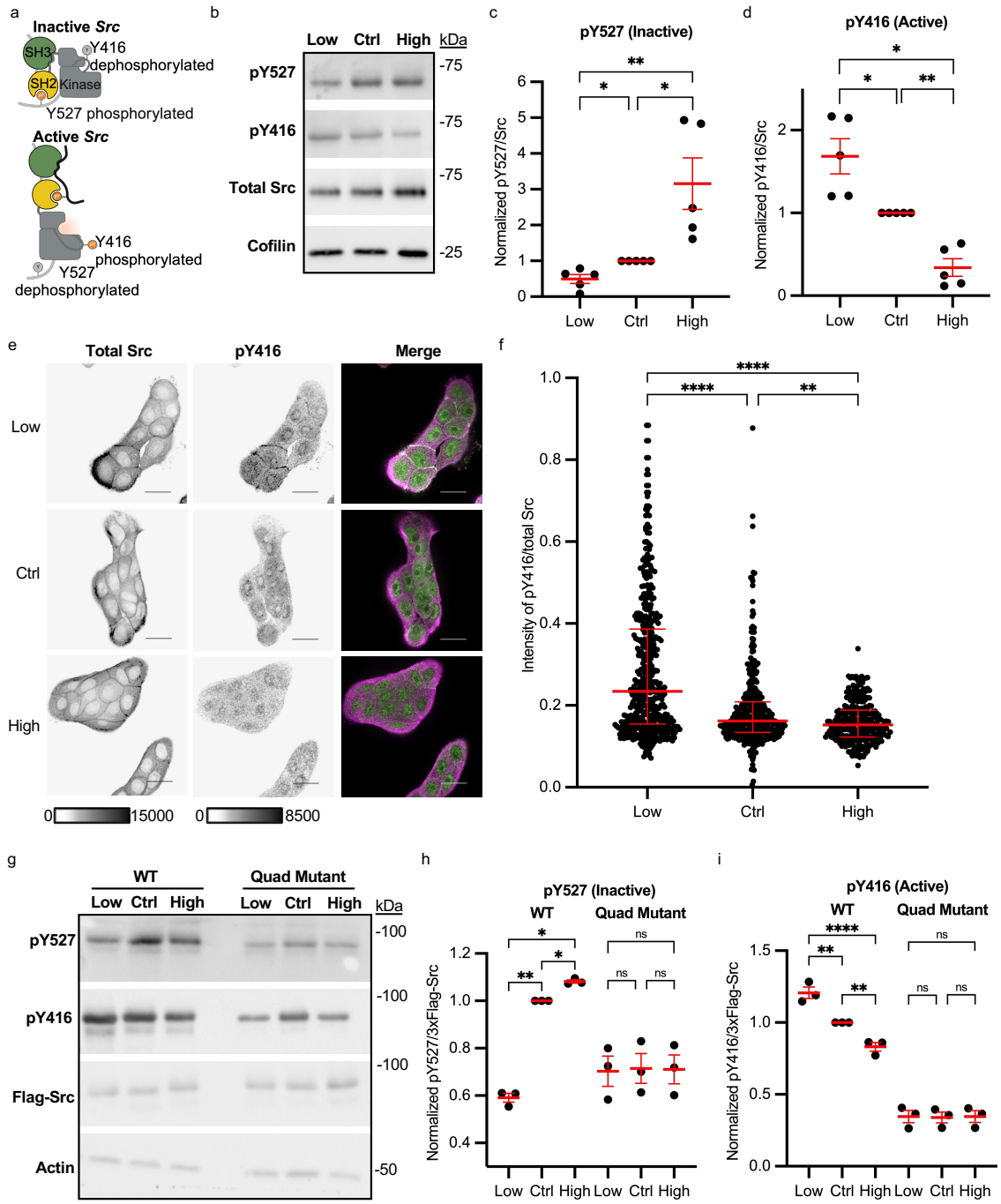


Figure 5. ***Intracellular pH regulates Src activity and localization in MCF10A cells and requires an intact ionizable residue network (C185 H201 H492 E524)*** (a) CARTOON OF c-SRC Activation (b) Representative immunoblot from MCF10A lysates prepared from cells maintained for 1 hour at low or high pHi or untreated (control). Immunoblots show total c-Src and indicated c-Src phosphorylation sites that reflect protein activation pY527 (inhibited) and pY416 (active). (c) Quantification of pY527 (inactive) from replicate data collected as in b. Data was normalized to control in each biological replicate. Scatter plots show (mean ± SEM, N=5)

(d) Quantification of pY416 (active) from replicate data collected as in B. Data was normalized to control for each biological replicate. Scatter plots show (mean ± SEM, N=5) (e) Representative confocal images of MCF10A cells fixed and stained for total c-Src and pY416 (active) c-Src. Total c-Src and pY416 are shown in inverse mono intensity display. In the merged overlay, total c-Src is shown in magenta and pSrc is shown in green, areas of high co-localization show white. Scale bar: 25µm. (f) Quantification of single-cell data collected as in F. Levels of pY416 c-Src are normalized to total c-Src in each single cell. Scatter plot shows individual cells and (median ± IQR, Low, n=486, Control, n=376, High, n=280) collected from 3 biological replicates. (g) Representative immunoblot from MCF10A lysates expressing either WT c-Src or the Quad Mutant (c-Src- C185A/H201A/H492A/E524A) and maintained for 1 hour at low or high pHi or untreated (control). Immunoblots show total c-Src and indicated c-Src phosphorylation sites that reflect protein activation pY527 (inhibited) and pY416 (active). (h) Quantification of pY527 (inactive) from replicate data collected as in g. Data was normalized to control in each biological replicate. Scatter plots show (mean ± SEM, N=5) (i) Quantification of pY416 (active) from replicate data collected as in g. Data was normalized to control for each biological replicate. Scatter plots show (mean ± SEM, N=5). For c, d, h, and i, significance was determined using a ratio paired t-test to compare between treatment conditions and a one-sample t-test to compare to control. For f, significance was determined using the Kruskal-Wallis test. * $p<0.05$, ** $p<0.01$, *** $p<0.001$, **** $p<0.0001$.

To determine whether the identified residue network in c-Src mediates pH-dependent activity, we overexpressed either FLAG-tagged WT-c-Src or the quadruple mutant c-Src-C185A/E454A/H492A/E524A in MCF10A cells. We then quantified pY527 and pY416 levels by immunoblot with and without pHi manipulation. We found that overexpressing WT c-Src recapitulated the pH-dependent endogenous c-Src activity results, with decreased inhibited c-Src (pY527) at low pHi compared to control and increased pY527 at high pHi compared to control MCF10A (**Fig. 5g, h**). As with endogenous c-Src, we also observed pHi-dependent phosphorylation of pY416, with increased active c-Src (pY416) at low pHi compared to control and high pHi (**Fig. 5g, i)**. Taken together these data demonstrate that pHi allosterically regulates c-Src inside cells, with low pHi promoting the active state (pY416, Y527) while high pHi promotes a more inhibited state (Y416, pY527). As expected, the quadruple c-Src mutant had pH-independent phosphorylation of both pY416 and pY527 while maintaining measurable catalytic activity (**Figure 5h, 5i**). Our cell-based analyses show that c-Src activity is sensitive to

physiological pHi changes, with increased activity at low pHi, and that the identified ionizable cluster is required for pH-dependent c-Src activity.

To further probe the contributions of the c-Src pH-sensitive residue network, we generated single point mutants (c-Src-C185A, c-Src-E524A, c-Src-H201A, and c-Src-H496A) and measured their activity in MCF10A cells. While we were unable to robustly express the H496A mutant, we were able to express all other point mutants (**Extended Data Fig 8**). First, we found that the E524A mutant pushed c-Src into a more active, but still pH-dependent, conformation: Y527 was dephosphorylated compared to WT c-Src, while Y416 phosphorylation was increased compared to WT but still pH-dependent (**Extended Data Fig. 8b-d**). This suggests E524 is required for the observed pH-dependence of pY527 in wild-type c-Src: without E524, affinity for pY527 is reduced and c-Src C-terminal tail release from the SH2 domain is pH-insensitive. We can also conclude from the E524A data that the remaining residue network (C185, H201, and H495) is sufficient to preserve pH-dependent phosphorylation of active-loop residue (Y416) (**Extended Data Fig. 8b, d**). We found that C185A was still pH-sensitive with trends showing pH-dependent phosphorylation of Y527 and Y416 (**Extended Data Fig. 8e-g**), suggesting C185 is not essential for either pH-sensitive phosphorylation event on c-Src. Finally, the H201A mutant locked c-Src into an inhibited and pHi-independent conformation, with significantly reduced Y416 phosphorylation compared to WT (**Extended Data Fig. 8e-g**). This suggests H201 plays a predominant role in pH-dependent active-loop phosphorylation of Y416 observed in both WT and c-Src-E524A.

Our data show that our *in silico* analysis pipeline can be used to predict and mechanistically probe pH-dependent protein function. We applied this analysis to characterize molecular mechanisms of pH-dependent allostery for the known pH sensor SHP2 (**Fig. 6a**). We

also used it to identify novel pH regulation of c-Src activity and identify critical residues involved (**Fig 6b**). Finally, we demonstrated that ionizable residue networks with predicted physiological pK_as are conserved across many SH2 domain-containing signaling proteins including kinases, phosphatases, transcription factors, and GTPase activating proteins. Importantly, our network analysis can accurately identify previously validated pH-sensing residues based on existing structures including in mutant p53 (His273)[35], phosphofructokinase muscle isoform (His242, Asp309)[36], isocitrate dehydrogenase 1 (D273, E304, K217)[37], and the sodium proton exchanger (H544)[38] (**Extended Data Fig. 9)**.

Figure 6

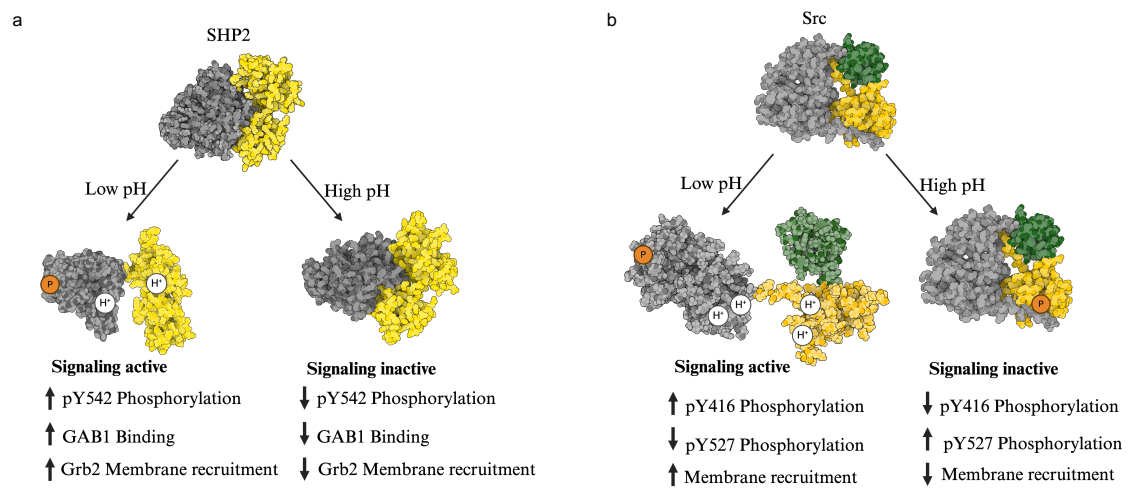


Figure 6. ***Molecular mechanism of pH-dependent allostery of SHP2 and Src*** (a) Schematic of pH-driven activation and inhibition of SHP2. At low pH, the SH2 domain is unbound and SHP2 becomes signaling active with increased phosphorylation of Y542, increased GAB1 binding, and increased Grb2 recruitment. (b) Schematic of pH-driven activation and inhibition of Src. At low pH, the SH2 domain is unbound and c-Src becomes signaling active with increased phosphorylation of Y416, decreased phosphorylation of Y527, and increased membrane recruitment.

Our pipeline is modular, enabling incorporation of improved pK_a prediction algorithms (e.g. recent ML pipelines with improved accuracy over PROPKA)[39]. Scientists can apply this approach to identify molecular drivers of pH-dependent cell behaviors that lack known pH-

sensitive molecular drivers including differentiation[40,41], cell cycle progression[2,3], and epithelial to mesenchymal transition[42,43]. In addition to revealing the pH-dependent structural proteome, this approach can be applied to molecularly characterize how dysregulated pHi drives pathophysiology in cancer (increased pHi)[6,44] and neurodegeneration (decreased pHi)[45].

Materials and Methods

pK_a Prediction Method

All full-length structures for proteins of interest were first identified based on the gene search and downloaded from the Protein Database (https://www.rcsb.org/). Next, the electrostatic properties for each structure were calculated using the APBS (Adaptive Poisson-Boltzmann Solver) webserver (https://server.poissonboltzmann.org/). The APBS calculation was performed using the following parameters: The pH for continuum electrostatic calculations was set to 7.4. The AMBER forcefield was used for the minimization step. We set the parameters in the APBS webserver to ensure that new atoms were not rebuilt too close to existing atoms and that hydrogen bonding networks were optimized during the structure repair and refinement step. The electrostatic data for each individual structure was combined into a single dataset and ionizable residues were then filtered for physiologically relevant pK_as. Histidine residues with pK_as above 7.0 were considered physiologically relevant while aspartates, glutamates, and lysines were filtered if pK_as were between 6.8.00. For ionizable residues that were classified using these criteria as "shifted" in ≥ 50% of structures, residue pK_as were averaged across all structures in the dataset. Next, the strength of coulombic interactions for the shifted residues were averaged for all structures. Structural visualization was done using Pymol and residue interaction networks were generated in Python (https://github.com/KWhiteLab/VanDyck_etal).

Structural Conservation

The ConSurf server was used as a bioinformatics tool to estimate evolutionary conservation in proteins of interest. To determine the evolutionary conservation in our proteins of interest, we followed the recommended guidelines outlined by Yariv *et al.*, 2023[24]. Briefly, the SHP2 (PDB: 2SHP) and c-Src (PDB:2SRC) protein structures were submitted to the ConSurf Web server (https://consurf.tau.ac.il/). ConSurf server utilizes an empirical Bayesian algorithm for conservation scores ranging from 1 to 9; where 1 is variable and 9 is conserved.

Purification of SHP2 proteins

Full-length SHP2 variants were cloned into the pGEX-4TI SHP2 WT plasmid[46], pGEX-4T1 SHP2 WT was a gift from Ben Neel (Addgene plasmid # 8322 ; http://n2t.net/addgene:8322 ; RRID:Addgene_8322)) and expressed in *E. Coli* for purification. Transformed BL21(DE3) cells were grown in Miller's Modified Luria Broth (RPI, L24080-2000) supplemented with 0.1 mg/mL ampicillin at 37°C with shaking (250 rpm). When cells reached an OD600 of 0.6, proteins were induced with Isopropyl β- d-1-thiogalactopyranoside (IPTG) (Thermo Scientific, R0392) (0.42 mM) and incubated at 18°C overnight with shaking (250 rpm). The induced cultures were centrifuged at 10,000 rpm at 4°C for 30 mins and resuspended in lysis buffer (B-PER™ Bacterial Protein Extraction Reagent (Thermo Scientific, 78243)) supplemented with EDTA-free Protease Inhibitor Cocktail (Roche, 04693132001)). Cell lysates were placed on a rocker at 4°C for 10 mins before centrifugating at 40,000 rpm at 4°C for 40 minutes. The supernatant was separated, and glutathione resin (GenScript, L00206) was added to bind the GST tagged proteins. The mixture was rocked for an hour at 4ºC and then purified in a

CrystalCruz chromatography column (Santa Cruz, sc-205554) at 4°C. Purified proteins were dialyzed into dialysis buffer (20mM Tris 50mM NaCl 1mM DTT 5% glycerol, pH 7.5) overnight at 4°C and stored at -20°C.

Phosphatase Activity Assays

SHP2 variants were quantified by measuring absorbance at 280 nm and comparison with BSA protein standards. Para-nitrophenyl phosphate (pNPP) (Thermo Scientific, 34045) was used as an artificial substrate for the protein tyrosine phosphatase (PTPase) activity of each SHP2 variant. Reactions were performed at 30°C in 100 µl volume of PTPase buffer (25 mM Tris–HCl, 50 mM NaCl, 2 mM EDTA and 10 mM dithiothreitol). Concentrations of pNPP ranged from 12-400 mM (12, 15.6, 25, 31.2, 50, 62.5, 100, 125, 250, 300, 400) and assay buffer pH ranged from 6.1 to 8.0. All assay conditions were performed in technical duplicates in 96-well plates. Reactions were initiated by the addition of 5µg of WT or mutant SHP2 to the reaction mixtures. Kinetic absorbance readings were collected at 405 nm over 30 mins on a spectrophotometer (Biotek Cytation 5). In all cases, after averaging technical replicates, the linear region of the reaction progress curve was determined by visual inspection and fit to a slope against the time of progression. These slopes were then plotted against the corresponding concentrations of substrate and fit to a the Michaelis-Menten equation using non-linear regression to determine catalytic parameters. Experiments were repeated at least three times across two batches of purified proteins, and the average and standard error of the mean for all biological replicates are reported.

Cell Culture

Normal breast epithelial MCF10A cells were grown in complete media: DMEM/50% F12 w/ GlutaMax (Invitrogen, 10565-018) supplemented with 5% Horse Serum (Invitrogen, 16050-122), 0.02 µg/mL EGF (Peprotech, AF-100-15), 5 µg/mL Hydrocortisone (Sigma, H-0888), 10 mg/mL Insulin (Sigma, I-1882), 0.1 µg/mL Cholera toxin (Sigma, C-8052), and 1% Penicillin-Streptomycin (Corning, 30-001-Cl). Cells were maintained in humidified incubators at 37°C with 5% CO_2. MCF10A cells (ATCC: CRL-10317) were split at sub-confluency. Cells were monitored for maintenance of epithelial morphology throughout experimental protocols.

pHi manipulation media:

Control: Complete media as defined above.

Low pHi: Complete media supplemented with 25 µM EIPA (Invitrogen, E3111) and 30 µM S0858 (Sigma-Aldrich, SML0638-5MG)

High pHi: Complete media supplemented with 30 mM ammonium chloride (NH_4Cl, Fisher Chemical, A661-500, Lot 185503)

All chemical concentrations are final concentrations (Cf).

pHi Manipulation and measurement

MCF10A cells were plated in 6 technical replicates for each treatment condition (6×100,000 cells/well; 24 well plate; 1 mL total volume). 24 hours after plating, media was exchanged with pHi manipulation media (see above). After 1 hour of incubation with pHi manipulation media, pHi was measured as previously described[47]. Briefly, 2′,7′-Bis-(2-Carboxyethyl)-5-(and-6)-Carboxyfluorescein, Acetoxymethyl (BCECF-AM) Ester (Biotium, 51011) was added at 2 µM (Cf) to cells for 30 min at 37°C with 5% CO_2. Cells were then washed 3x5 min with HEPES-

based wash buffer (30 mM HEPES, 145 mM NaCl, 5 mM KCl, 10 mM glucose, 1 mM KPO_4, 1 mM $MgSO_4$, and 2 mM $CaCl_2$) at 37°C. Wash buffers were supplemented with appropriate pH-altering compounds at the same concentrations as the media. After washes, kinetic fluorescence measurements were made at pH-sensitive wavelengths (490ex/535em) and the pH-insensitive wavelengths (440ex/535em) every 15sec for 5 minutes at Medium PMT sensitivity on a plate reader (Biotek Cytation 5). After the initial kinetic pHi read, wash buffers were removed and replaced with the isotonic standardization buffer at ~7.5 pH (25 mM HEPES, 80 mM KCl, 1 mM $MgCl_2$) supplemented with protonophore nigericin (Invitrogen, N1495) at 10 µM (Cf). Cells were incubated at 37°C for 10 min, and the plate was read again with the same parameters. The high pH standardization buffer was then replaced with a ~7.0 pH nigericin buffer standard, cells were incubated at 37°C for 10 min, and the plate was read again with the same parameters. The nigericin buffer was then replaced with a ~6.5 pH nigericin buffer standard, cells were incubated at 37°C for 10 min, and the plate was read again with the same parameters. Mean intensities from pHi measurement and the nigericin standards were exported and pHi was back calculated using a nigericin linear regression performed on each well with Nigericin standard pHi reported to the hundredths place.

Immunoblotting

After pHi manipulation (see above section) plates were removed from the incubator and placed on ice. The media was aspirated, and the plates were washed two times with ice-cold DPBS. 200µL of lysis buffer (500 mM NaCl, 10 mM EDTA, 500mM HEPES-free acid, 10 mM EGTA,

10% Triton X-100, protease inhibitor cocktail (Pierce™ Protease Inhibitor Tablets, A32965; 1 tablet/50 mL lysis buffer) and phosphatase inhibitor cocktail (PhosSTOP™, Sigma Aldrich, 4906845001), (pH 7.4) was added to each plate. After 15 minutes of incubation in the lysis buffer, Cells were removed using cell scrapers and transferred to pre-chilled microfuge tubes. Lysates were centrifuged at 14,000 x g for 15 minutes at 4°C. Clarified lysates were isolated to separate microfuge tubes on ice. The Bradford assay was used to quantify the total protein content of lysates (Bradford, 1976) using Coomassie Protein reagent (Thermo Scientific, cat: 1856209). Lysates were stored at -80°C if not immediately being used for SDS-PAGE sample preparation.

Protein samples were prepared in 6X Laemilli (reducing) buffer (Thermo Scientific, J61337-AC) and boiled for 5 minutes at 100°C. A total of 10 µg of MCF10A lysate was added to each lane for immunoblot analysis. Protein samples were loaded onto 12% SDS-PAGE gels, which were run at 175V and 0.08 Amps for 15 minutes to stack samples. Protein samples were separated by running the gel at 200 V and 0.08 Amps for an hour and a half. Polyvinylidene difluoride (PVDF) membranes were pre-soaked in 100% methanol for 5 minutes and rinsed in turbo transfer buffer. Turbo transfers were performed at 1.00 A and 25V for 30 mins. Membranes were blocked for 1 hour in 5% milk (or 5% BSA for phospho-proteins) in 1X Tris-buffered Saline + 0.1% Tween® 20 (TBS-T) at room temperature with shaking. Membranes were washed 3x5 min in 1X TBS-T before being cut and incubated with primary antibodies. The following primary antibodies were used: SHP2 (Cell Signaling Technology, 3752; 1:1000 dilution in 1% milk in 1X TBS-T), pSHP2(Tyr524) (Cell Signaling Technology, 3751T; 1:1000 dilution in 1% BSA in 1X TBS-T), Cofilin (Cell Signaling Technology, 5175; 1:1000 dilution in 1% milk 1X TBS-T),

HA(Cell Signaling Technology, 3724; 1:1000 dilution in 1% milk 1X TBS-T), Actin (Santa Cruz, sc58673; 1:1000 diluted in 1% milk in 1X TBS-T), Src (Cell Signaling Technology, 2109; 1:1000 dilution in 1% milk 1X TBS-T), pSrc(Tyr527) (Cell Signaling Technology, 2105; 1:1000 dilution in 1% BSA in 1X TBS-T), pSrc(Tyr416)(Cell Signaling Technology, 6943; 1:1000 dilution in 1% BSA in 1X TBS-T). Primary antibodies were incubated overnight at 4°C with shaking. Primary antibody solutions were removed, and membranes were washed 3x5 minutes in 1X TBS-T before incubating secondary antibodies (Goat anti-mouse/rabbit HRP-conjugated antibodies; 1:10,000 diluted in 1% milk/BSA in 1X TBS-T) for 1 hour at room temperature with shaking. Secondary antibody solutions were removed, and membranes were washed 3x5 minutes in 1X TBS-T before being developed using Pierce SuperSignal West Pico PLUS (Thermo Scientific, 34580) or Pierce SuperSignal West Femto PLUS (Thermo Scientific, 34095). Chemiluminescence and colorimetric images were acquired using a BioRad ChemiDoc imager.

Raw files were exported from ChemiDoc imager and opened in ImageJ. Band intensities were acquired by densitometry analysis (area under the curve of individual bands in respective lanes). For each replicate blot, target protein intensities were normalized to the intensity of loading controls before normalizing each condition ratio to that of the control within biological replicates.

Co-Immunoprecipitation Assays

After pHi manipulation (see above section) plates were washed 2x with ice-cold PBS. The cells were then lysed on ice (see above section). After Bradford measurements, 1µg of anti-SHP2 (Rabbit, Cell Signaling, 3752) was added to 100 µg of the lysate. The mixture was incubated on

the rocker overnight at 4°C. Next, 80 µl of protein A-dynabeads to each tube and the mixture was incubated on the rocker for 4 hours at 4°C. Next the beads were washed 3 times with 800 µl of wash buffer (TBS + 0.1% Tween, EDTA-free Protease Inhibitor Cocktail (Roche, 04693132001)). After the final wash, the beads were heated at 50°C in 50 µL of 2x Laemilli sample buffer (Biorad, 1610737) for 5 minutes to elute the proteins. Next the immunoprecipitation samples and total lysates were immunoblotted (see above section). The following antibodies were used for primary incubations: SHP2(abcam, ab76285; 1:1000 dilution in 1% milk in 1X TBS-T), GAB1 (Santa Cruz Biotechnology, sc-133191, 1:1000 dilution in 1% milk in 1X TBS-T).

Microscopy Experiments

GRB2 TagBFP Assay

The pHR Grb2-TagBFP plasmid, a gift from Jared Toettcher (Addgene plasmid # 188631 ; http://n2t.net/addgene:188631 ; RRID:Addgene_188631)[28], was transfected into MCF10A cells and grown for 24 hours before plating for experiments. The cells were plated on a 35mm imaging dish with a 14mm glass coverslip (Matsunami, D35-14-1.5-U) a day before imaging. Microscope objectives were preheated to 37°C, and the stage-top incubator was preheated to 37°C and kept at 5% CO2/95% air. Confocal images were collected on a Nikon Ti-2 spinning disk confocal with a 60× (CFI PLAN FLUOR NA 1.3) oil immersion objective. The microscope is equipped with a stage-top incubator (Tokai Hit), a Yokogawa spinning disk confocal head (CSU-X1), four laser lines (405 nm, 488 nm, 561 nm and 647 nm), a Ti2-S-SE motorized stage, multi-point perfect focus system and an Orca Flash 4.0 CMOS camera. The cells were imaged initially for starting intensities before nigericin stimulation. Next, the media was replaced with

nigericin buffer to pin the cells at different pH conditions to measure protein activity at the respective pH conditions. The cells were then imaged for 15 mins at 30 second intervals. The image quantification was performed in Nikon Elements software. Images were background-subtracted using a region of interest (ROI) drawn on a glass coverslip (determined by DIC). To only measure the intensity of the cell membrane, the ROI for membrane of each cell was outlined using the membrane marker and tracked over time. In case of improper tracking, manual tracking was used to redraw ROIs.

Immunofluorescence assays

MCF10A Cells were plated in 35mm imaging dish with a 14mm glass coverslip (Matsunami, D35-14-1.5-U) a day before treatment. The cells were then treated with pHi manipulation drugs for an hour. Cells were rinsed with DPBS and fixed in 3.7% formaldehyde (Alfa Aestar, 33314) at room temperature for 10 min. Cells were washed three times for 2 min each time with DPBS, then incubated in lysis buffer (0.1% Triton-X in DPBS) for 10 min at room temperature. Cells were washed three times for 2 min each time with DPBS, then incubated in blocking buffer (1.0% BSA in DPBS) for 1 hour at room temperature with rocking followed by three 2 min washes in DPBS. Cells were then incubated overnight at 4°C with anti-C-Src antibody (Proteintech, 60315-1-Ig) and anti-pY416-C-Src antibody (ThermoFisher Scientific, 44-660G). Cells were washed three times for 3 min each time in DPBS, then incubated with secondary antibody and phalloidin for 1 hour at room temperature. After three 3 min washes in DPBS, Hoechst 33342 dye (1:20,000) in antibody buffer was added for 10 min, then removed. Cells were imaged in DPBS on the spinning disk confocal microscope.

Statistical analysis

GraphPad Prism was used to prepare graphs and perform statistical analyses. Normality tests were performed on all data sets as well as outlier tests using the ROUT method (Q=1%). Where normally distributed data is shown with means and non-normally distributed data is shown with medians. For normally distributed data, significance was determined by ratio paired t-tests or appropriate ANOVA were used. For non-normally distributed data, significance was determined by Kruskal-Wallis test. "N" indicates the number of biological replicates performed and "n" represents the number of technical replicates or individual cell measurements collected.

Constant pH Molecular Dynamics

The 2SHP and 2SRC monomer chains were downloaded from Protein Data Bank (PDB) for initial structures. Constant pH MD (CpHMD) simulations were conducted following the protocol described by Henderson *et al*., 2022[25]. CpHMD simulations were performed with implicit solvent using the GBNeck2-CpHMD method implemented in AMBER 2022[48]. Preparation of input files was done using the cphmd_pred.sh script provided in the protocol[25], using the default titratable residue types (Asp, Glu, His, Cys, Lys). After energy minimization, structure relaxation, and protonation state initiation[25], independent production simulations were run over 4.0-10.0 pH in half-integer increments for 8 nanoseconds (ns) for each, resulting in 13 independent CpHMD simulations and 104 ns of total simulation time for each structure. Coordinate snapshots were saved every 20 picoseconds (ps), and protonation state snapshots

every 2 ps. The probability of protonated states became time-independent after 4 ns. The resulting coordinate and protonation state trajectories were analyzed using Python scripts provided in the protocol[48], the CPPTRAJ analysis program[49], and in-house Python scripts (https://github.com/KWhiteLab/VanDyck_etal).

Acknowledgements.
This research was supported in part by the Notre Dame Center for Research Computing; we specifically thank Dr. Dodi Heryadi and Dr. Scott Hampton for technical support. This work was supported by NSF CAREER AWARD (MCB-2238694) and an NIH New Innovator Award (DP2-CA26041601) to KAW; and an NIH R01 (R01GM123338) to JWP.

Author contributions.
PKVD contributed to project conceptualization, methodology development, data collection, analysis, figure generation, writing (original draft and review/editing). LP contributed to methodology development, data collection, analysis, figure generation, and writing (original draft and review/editing. EAG contributed to data collection, analysis, figure generation, writing (review/editing). ETN contributed to data collection, analysis, figure generation, writing (review/editing). JAA contributed to data collection, analysis, writing (review/editing). LMH contributed to contributed to data collection, analysis, writing (review/editing). JWP contributed to writing (original draft and review/editing), supervision, and funding acquisition. KAW contributed to project conceptualization, methodology development, analysis, figure generation, writing (original draft and review/editing), project administration, supervision, and funding acquisition.

Competing interests.

The authors declare no competing interests.

Materials & Correspondence.

For materials requests and correspondence contact Katharine White (kwhite6@nd.edu)

References:

1. Boron, W. F. Regulation of intracellular pH. *Adv. Physiol. Educ.* **28**, 160–179 (2004).
2. Spear, J. S. & White, K. A. Single-cell intracellular pH dynamics regulate the cell cycle by timing the G1 exit and G2 transition. *J. Cell Sci.* **136**, jcs260458 (2023).

3. Putney, L. K. & Barber, D. L. Na-H Exchange-dependent Increase in Intracellular pH Times G2/M Entry and Transition. *J. Biol. Chem.* **278**, 44645–44649 (2003).
4. Frantz, C., Karydis, A., Nalbant, P., Hahn, K. M. & Barber, D. L. Positive feedback between Cdc42 activity and H+ efflux by the Na-H exchanger NHE1 for polarity of migrating cells. *J. Cell Biol.* **179**, 403–410 (2007).
5. Boussouf, A. & Gaillard, S. Intracellular pH changes during oligodendrocyte differentiation in primary culture. *J. Neurosci. Res.* **59**, 731–739 (2000).
6. Czowski, B. J., Romero-Moreno, R., Trull, K. J. & White, K. A. Cancer and pH Dynamics: Transcriptional Regulation, Proteostasis, and the Need for New Molecular Tools. *Cancers* **12**, 2760 (2020).
7. Garcia-Moreno, B. Adaptations of proteins to cellular and subcellular pH. *J. Biol.* **8**, 98 (2009).
8. Harms, M. J. *et al.* A buried lysine that titrates with a normal pKa: role of conformational flexibility at the protein-water interface as a determinant of pKa values. *Protein Sci* **17**, 833–45 (2008).
9. Harms, M. J. *et al.* The pK(a) values of acidic and basic residues buried at the same internal location in a protein are governed by different factors. *J Mol Biol* **389**, 34–47 (2009).
10. Isom, D. G., Castaneda, C. A., Cannon, B. R. & Garcia-Moreno, B. Large shifts in pKa values of lysine residues buried inside a protein. *Proc Natl Acad Sci U A* **108**, 5260–5 (2011).
11. Isom, D. G. & Dohlman, H. G. Buried ionizable networks are an ancient hallmark of G protein-coupled receptor activation. *Proc. Natl. Acad. Sci.* **112**, 5702–5707 (2015).

12. Isom, D. G. *et al.* Protons as Second Messenger Regulators of G Protein Signaling. *Mol. Cell* **51**, 531–538 (2013).

13. Tews, I. *et al.* The Structure of a pH-Sensing Mycobacterial Adenylyl Cyclase Holoenzyme. *Science* **308**, 1020–1023 (2005).

14. Olsson, M. H. M., Søndergaard, C. R., Rostkowski, M. & Jensen, J. H. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical p K_a Predictions. *J. Chem. Theory Comput.* **7**, 525–537 (2011).

15. Bertalan, É., Rodrigues, M. J., Schertler, G. F. X. & Bondar, A. Graph-based algorithms to dissect long-distance water-mediated H-bond networks for conformational couplings in GPCRs. *Br. J. Pharmacol.* bph.16387 (2024) doi:10.1111/bph.16387.

16. Qu, C. K. The SHP-2 tyrosine phosphatase: Signaling mechanisms and biological functions. *Cell Res.* **10**, 279–288 (2000).

17. Mannell, H. & Krotz, F. SHP-2 Regulates Growth Factor Dependent Vascular Signalling and Function. *Mini-Rev. Med. Chem.* **14**, 471–483 (2014).

18. Lauriol, J., Jaffré, F. & Kontaridis, M. I. The role of the protein tyrosine phosphatase SHP2 in cardiac development and disease. *Semin. Cell Dev. Biol.* **37**, 73–81 (2015).

19. Merritt, R., Hayman, M. J. & Agazie, Y. M. Mutation of Thr466 in SHP2 abolishes its phosphatase activity, but provides a new substrate-trapping mutant. *Biochim. Biophys. Acta BBA - Mol. Cell Res.* **1763**, 45–56 (2006).

20. Schaffhausen, B. SH2 domain structure and function. *Biochim. Biophys. Acta BBA - Rev. Cancer* **1242**, 61–75 (1995).

21. Hof, P., Pluskey, S., Dhe-Paganon, S., Eck, M. J. & Shoelson, S. E. Crystal Structure of the Tyrosine Phosphatase SHP-2. *Cell* **92**, 441–450 (1998).

22. Srivastava, J. *et al.* Structural model and functional significance of pH-dependent talin-actin binding for focal adhesion remodeling. *Proc Natl Acad Sci U A* **105**, 14436–41 (2008).
23. Vercoulen, Y. *et al.* A histidine pH sensor regulates activation of the Ras-specific guanine nucleotide exchange factor RasGRP1. *Elife* **6**, e29002 (2017).
24. Yariv, B. *et al.* Using evolutionary data to make sense of macromolecules with a "face-lifted" ConSurf. *Protein Sci.* **32**, e4582 (2023).
25. Henderson, J. A. *et al.* A Guide to the Continuous Constant pH Molecular Dynamics Methods in Amber and CHARMM [Article v1.0]. *Living J. Comput. Mol. Sci.* **4**, (2022).
26. Sun, J. *et al.* Antagonism between binding site affinity and conformational dynamics tunes alternative cis-interactions within Shp2. *Nat. Commun.* **4**, 2037 (2013).
27. Cai, T., Nishida, K., Hirano, T. & Khavari, P. A. Gab1 and SHP-2 promote Ras/MAPK regulation of epidermal growth and differentiation. *J. Cell Biol.* **159**, 103–112 (2002).
28. Farahani, P. E. *et al.* pYtags enable spatiotemporal measurements of receptor tyrosine kinase signaling in living cells. *eLife* **12**, e82863 (2023).
29. Hellmuth, K. *et al.* Specific inhibitors of the protein tyrosine phosphatase Shp2 identified by high-throughput docking. *Proc. Natl. Acad. Sci.* **105**, 7275–7280 (2008).
30. Singer, A. U. & Forman-Kay, J. D. pH Titration studies of an SH2 domain-phosphopeptide complex: Unusual histidine and phosphate p K_a values. *Protein Sci.* **6**, 1910–1919 (1997).
31. Vogel, H. J. [13] Phosphorus-31 nuclear magnetic resonance of phosphoproteins. in *Methods in Enzymology* vol. 177 263–282 (Elsevier, 1989).
32. Kurochkina, N. & Guha, U. SH3 domains: modules of protein–protein interactions. *Biophys. Rev.* **5**, 29–39 (2013).

33. Xu, W., Doshi, A., Lei, M., Eck, M. J. & Harrison, S. C. Crystal Structures of c-Src Reveal Features of Its Autoinhibitory Mechanism. *Mol. Cell* **3**, 629–638 (1999).

34. Sondka, Z. *et al.* COSMIC: a curated database of somatic variants and clinical data for cancer. *Nucleic Acids Res.* **52**, D1210–D1217 (2024).

35. White, K. A. *et al.* Cancer-associated arginine-to-histidine mutations confer a gain in pH sensing to mutant proteins. *Sci. Signal.* **10**, eaam9931 (2017).

36. Kisor, K. pH-Sensors Regulating Transcription, Metabolism, and Cancer Cell Biology. (2023).

37. Luna, L. A. *et al.* An acidic residue buried in the dimer interface of isocitrate dehydrogenase 1 (IDH1) helps regulate catalysis and pH sensitivity. *Biochem. J.* **477**, 2999–3018 (2020).

38. Webb, B. A. *et al.* A Histidine Cluster in the Cytoplasmic Domain of the Na-H Exchanger NHE1 Confers pH-sensitive Phospholipid Binding and Regulates Transporter Activity. *J Biol Chem* **291**, 24096–24104 (2016).

39. Chen, A. Y., Lee, J., Damjanovic, A. & Brooks, B. R. Protein p K_a Prediction by Tree-Based Machine Learning. *J. Chem. Theory Comput.* **18**, 2673–2686 (2022).

40. Ulmschneider, B. *et al.* Increased intracellular pH is necessary for adult epithelial and embryonic stem cell differentiation. *J Cell Biol* **215**, 345–355 (2016).

41. Liu, Y. *et al.* Intracellular pH dynamics regulates intestinal stem cell lineage specification. *Nat. Commun.* **14**, 3745 (2023).

42. Amith, S. R., Wilkinson, J. M. & Fliegel, L. Na+/H+ exchanger NHE1 regulation modulates metastatic potential and epithelial-mesenchymal transition of triple-negative breast cancer cells. *Oncotarget* **7**, 21091–113 (2016).

43. Suzuki, A., Maeda, T., Baba, Y., Shimamura, K. & Kato, Y. Acidic extracellular pH promotes epithelial mesenchymal transition in Lewis lung carcinoma model. *Cancer Cell Int.* **14**, (2014).

44. White, K. A., Grillo-Hill, B. K. & Barber, D. L. Cancer cell behaviors mediated by dysregulated pH dynamics at a glance. *J. Cell Sci.* **130**, 663–669 (2017).

45. Harguindey, S., Reshkin, S. J., Orive, G., Arranz, J. L. & Anitua, E. Growth and trophic factors, pH and the Na+/H+ exchanger in Alzheimer's disease, other neurodegenerative diseases and cancer: new therapeutic possibilities and potential dangers. *Curr Alzheimer Res* **4**, 53–65 (2007).

46. O'Reilly, A. M., Pluskey, S., Shoelson, S. E. & Neel, B. G. Activated Mutants of SHP-2 Preferentially Induce Elongation of *Xenopus* Animal Caps. *Mol. Cell. Biol.* **20**, 299–311 (2000).

47. Choi, C.-H., Webb, B. A., Chimenti, M. S., Jacobson, M. P. & Barber, D. L. pH sensing by FAK-His58 regulates focal adhesion remodeling. *J. Cell Biol.* **202**, 849–859 (2013).

48. Case, D. A. *et al.* Amber 2022. Unpublished https://doi.org/10.13140/RG.2.2.31337.77924 (2022).

49. Roe, D. R. & Cheatham, T. E. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput.* **9**, 3084–3095 (2013).