Published in final edited form as:
*Mol Biosyst*. 2008 June ; 4(6): 643–653. doi:10.1039/b801018h.

# A quantitative study of the recruitment potential of all intracellular tyrosine residues on EGFR, FGFR1 and IGF1R[†,‡]

**Alexis Kaushansky**[a,e], **Andrew Gordus**[b,e], **Bryan Chang**[c,e], **John Rush**[d], and **Gavin MacBeath**[e]

[a] Program in Molecular and Cellular Biology, Harvard University, 12 Oxford Street, Cambridge, MA 02138, USA

[b] Program in Biophysics, Harvard University, 12 Oxford Street, Cambridge, MA 02138, USA

[c] Program in Chemistry and Chemical Biology, Harvard University, 12 Oxford Street, Cambridge, MA 02138, USA

[d] Cell Signaling Technology, Inc., Danvers, MA 01923, USA

[e] Department of Chemistry and Chemical Biology, Harvard University, 12 Oxford Street, Cambridge, MA 02138, USA. macbeath@chemistry.harvard.edu

## Abstract

Receptor tyrosine kinases transmit and process extracellular cues by recruiting intracellular signaling proteins to sites of tyrosine phosphorylation. Using protein microarrays comprising virtually every human SH2 and PTB domain, we generated quantitative protein interaction maps for three well-studied receptors—EGFR, FGFR1 and IGF1R—using phosphopeptides derived from every intracellular tyrosine residue on each receptor, regardless of whether or not they are phosphorylated *in vivo*. We found that, in general, peptides derived from physiological sites of tyrosine phosphorylation bind to substantially more SH2 or PTB domains than do peptides derived from nonphysiological sites, supporting the idea that kinases and interaction domains co-evolve and suggesting that new sites arise predominantly through selection favoring advantageous interactions, rather than through selection disfavoring unwanted interactions. We also found substantial qualitative overlap in the recruitment profiles of these three receptors, suggesting that their different biological effects arise, at least in part, from quantitative differences in their affinities for the proteins they recruit.

## Introduction

Receptor tyrosine kinases (RTKs) are a large and diverse family of single-pass transmembrane proteins found only in metazoans. They mediate intercellular communication and regulate a wide variety of cellular processes, including growth, proliferation, migration, differentiation and apoptosis.[1–4] The basic mechanism by which RTKs transduce their signals is relatively well-understood. Ligand binding to the extracellular domain of an RTK induces either receptor oligomerization[5,6] or the reorientation of preclustered receptors.[7] This in turn promotes the intermolecular phosphorylation of one receptor chain by its neighbor, usually within the

[†]This article is part of a *Molecular BioSystems* 'Emerging Investigators' issue highlighting the work of outstanding young scientists at the chemical–and systems–biology interfaces.

[‡]Electronic supplementary information (ESI) available: Table S1: Equilibrium dissociation constants for interactions between SH2/PTB domains and phosphopeptides. See DOI: 10.1039/b801018h

Correspondence to: Gavin MacBeath.

activation loop of the tyrosine kinase domain. The activated kinase then phosphorylates other tyrosine residues on the receptor, which serve as docking sites for intracellular enzymes or adaptor proteins. These intracellular proteins process the signal by activating a variety of signaling pathways, which act in concert to elicit the appropriate physiological response. Misregulation of RTKs, or the pathways they control, results in developmental abnormalities, as well as a variety of human diseases, including cancer, diabetes, allergy and asthma, inflammation, osteoporosis, and immune deficiency.[1–4,8,9]

The sequence-specific recruitment of signaling proteins to sites of tyrosine phosphorylation is mediated by two types of protein interaction domains: Src Homology 2 (SH2) domains and Phosphotyrosine Binding (PTB) domains.[10,11] Over the past two decades, many interactions have been reported to occur between SH2- or PTB-containing proteins and sites of tyrosine phosphorylation on RTKs. Most of these interactions have been discovered or confirmed by immunoprecipitating activated receptors from cultured cells and detecting associated proteins by immunoblotting. This approach, though clearly effective, suffers from several limitations. First, most signaling processes are highly dynamic; the half-lives of many interactions that are regulated by tyrosine phosphorylation are on the order of seconds, severely limiting the ability of signaling proteins to co-purify with receptors.[12] Second, this strategy is neither systematic nor comprehensive. The researcher is restricted to studying those proteins that are expressed in their particular cell line at sufficient levels to be detected and for which good antibodies exist. Indeed, most review articles detailing early events in RTK signaling repeatedly describe the same five to ten molecules and neglect the other >100 proteins encoded in the human genome that contain either SH2 or PTB domains. The proteins PLC-γ, PI3K, Shc, Grb2, Src, Ras-GAP, Shp2, IRS1, FRS2 and Cbl are often described as being the primary proteins recruited to receptors as diverse as the epidermal growth factor receptor (EGFR), the fibroblast growth factor receptor 1 (FGFR1) and the insulin-like growth factor 1 receptor (IGF1R). That this is an oversimplification is highlighted by the observation that FGFR1 stimulates neurite outgrowth in PC12 cells, while EGFR stimulates mitogenesis.[13] In order to study phosphotyrosine-mediated interactions in an unbiased fashion and on a genome-wide scale, we have previously developed protein microarrays comprising virtually every SH2 and PTB domain encoded in the human genome.[14] By probing these arrays with fluorescently-labeled phosphopeptides representing sites of tyrosine phosphorylation on human RTKs, we gain an unbiased and quantitative view of SH2- and PTB domain-mediated recognition of RTKs.[14,15] Here, we report quantitative protein interaction maps for FGFR1 and IGF1R, as well as an updated map for EGFR,[14] which provide a wealth of new information that can be used to guide traditional investigations of cell signaling, as well as system-based efforts to model signal transduction.[16,17] In addition, since all three of these receptors have been studied extensively, and since comprehensive efforts have been reported in all three cases to map their sites of tyrosine phosphorylation,[18–22] we have the opportunity to compare the binding properties of phosphopeptides derived from known sites of tyrosine phosphorylation (‘physiological’ peptides) with the binding properties of phosphopeptides derived from intracellular tyrosines that have not been found to be phosphorylated (‘nonphysiological’ peptides).

Overall, the experiments we report here led to two system-level observations regarding RTKs. First, we found that, in general, physiological peptides bind to substantially more SH2 and PTB domains than do nonphysiological peptides, supporting the idea that kinases and interaction domains co-evolve [23,24] and suggesting that the recruitment profile of a pY site evolves primarily through selection to establish new, desirable interactions, rather than through selection to eliminate unwanted interactions. Second, in comparing the recruitment profiles of EGFR, FGFR1 and IGF1R, we found considerable qualitative overlap between these three receptors. We submit that differences in the biological outcome elicited by each receptor arise, at least in part, from quantitative differences in the affinities of the receptors for the proteins they recruit, rather than solely from the identities of these proteins. Purely descriptive signaling

diagrams are therefore unlikely to provide the necessary details to explain and predict signaling downstream of RTKs.

## Results and discussion

### Physiological and nonphysiological sites of tyrosine phosphorylation

The receptors EGFR, FGFR1 and IGF1R have 20, 15 and 15 intracellular tyrosine residues, respectively. We began by designating each tyrosine as 'physiological' or 'nonphysiological' based on previous efforts to uncover sites of tyrosine phosphorylation. In the case of EGFR, Lombardo *et al*. initially used tandem mass spectrometry (MS/MS) to identify eight sites of tyrosine phosphorylation (Y1016, Y1069, Y1092, Y1110, Y1125, Y1138, Y1172 and Y1197) on recombinant receptor tail (residues 1000–1210) that had been incubated with purified EGFR kinase or with c-Src.[19] Around the same time, Stover *et al*. found many of the same sites, as well as two additional sites (Y915 and Y944), by phosphopeptide mapping of endogenous receptor that had been immunoprecipitated from breast and colorectal tumor cell lines.[22] More recently, pY978 was found in a phosphoproteomics study of human HepG2 hepatocytes,[25] and pY998 was found in a phosphoproteomics study of human 184A1 mammary epithelial cells.[26] Finally, years after recruitment data had been reported concerning the activation loop tyrosine (Y869),[27] phosphorylation of this residue was observed by MS/MS in U87MG glioblastoma cells stably transfected with a naturally-occurring deletion mutant of EGFR.[28]

In the case of FGFR1, Mohammadi *et al*. initially found seven sites of tyrosine phosphorylation (Y463, Y583, Y585, Y653, Y654, Y730 and Y766) using an *in vitro* approach that combined mutagenesis with microsequencing.[20] Later, Hinsby *et al*. verified most of these sites and identified an additional site (Y605) through an MS/MS analysis of receptor that had been immunoprecipitated from transfected human embryonic kidney cells.[18] Although no direct evidence of phosphorylation of Y677 and Y701 has been reported, Foehr *et al*. showed through mutagenesis experiments that these sites are required for neurite outgrowth in PC12 cells and that changing these tyrosines to phenylalanine results in lower levels of phosphorylated receptor.[29] We therefore designated them as 'physiological'.

In the case of IGF1R, six sites (Y973, Y980, Y1161, Y1165, Y1166 and Y1346) were initially identified by stably expressing human IGF1R in rat fibroblasts and using two-dimensional thin layer chromatography coupled with Edman degradation to observe phosphorylation.[21] Although no direct evidence has been reported for the phosphorylation of Y1280 and Y1281, several functional studies have demonstrated their importance in activating the mitogen-activated protein kinase (MAPK) signaling cascade and in suppressing apoptosis in lymphocytic cells.[30,31] We therefore designated these sites as 'physiological'. A summary of our designations is provided in Fig. 1.

### Quantitative protein interaction maps using protein microarrays

To study the interactions of SH2 and PTB domains with physiological and nonphysiological sites on each of the three receptors, we synthesized pY-containing peptides with sequences that correspond to the sequences surrounding every intracellular tyrosine. Structural studies have shown that recognition can occur as far upstream as the −7 position of a phosphopeptide for some PTB domains[32–34] and as far downstream as the +5 position of a peptide for some SH2 domains.[35–37] To ensure an accurate analysis of all domains, we synthesized peptides with sequences that include nine residues upstream and seven residues downstream of the pY (Table 1). To visualize peptides that are bound by SH2 and PTB domains, we labeled each peptide on its amino-terminus with 5-(and-6)-carboxytetramethylrhodamine [5(6)-TAMRA]. 5(6)-TAMRA serves as a chromophore for quantification, as well as a fluorophore for visualization on the microarrays. The labeled peptides were deprotected, cleaved from the

resin, and purified by reversed-phase high performance liquid chromatography (HPLC). Fractions containing the desired product were identified by matrix assisted laser desorption/ionization time-of-flight (MALDI–TOF) mass spectrometry and the purified peptides were recovered by lyophilization and quantified by absorption spectroscopy. Following this procedure, we obtained pure product for 46 of the 50 phosphopeptides. Of these, 43 were sufficiently soluble in aqueous buffer to use as probes for our experiments (Table 1).

We have previously reported cloning, expressing and purifying virtually every human SH2 and PTB domain and preparing microarrays of these domains on chemically-derivatized glass surfaces.[14,38] Briefly, the coding regions for each domain were cloned from human cDNA and the corresponding proteins were produced recombinantly in *Escherichia coli* using the T7 expression system. Each domain features an amino-terminal $His_6$-tag, as well as a thioredoxin tag to facilitate the high-level production of soluble protein. After purifying each domain from large-scale bacterial culture, we assessed its purity by SDS-polyacrylamide gel electrophoresis and its aggregation state by size-exclusion column chromatography. In the current version of our arrays, we eliminated domains that were impure or did not contain soluble, monomeric protein. Notably, SH2 domains derived from the STAT and SOCS families of proteins did not behave well. By cloning larger portions of STAT1 and STAT2 that included their entire SH2 domain-containing cores,[39] we obtained soluble, monomeric material for these two proteins. In addition, we cloned, expressed, and purified the N-terminal domain of Cbl,[40] which contains a noncanonical SH2 domain. In total, 133 domains representing 103 proteins were used in these studies.

To facilitate the rapid and automated processing of microarrays, we spotted our proteins in quadruplicate on aldehyde-displaying glass substrates, cut to a size that spans all the wells of a microtiter plate (Fig. 2). Ninety-six separate arrays were prepared on each glass substrate, and the glass was attached to the bottom of a bottomless microtiter plate using an intervening silicone gasket. Two 16 × 17 microarrays were required to accommodate all 133 domains as well as the appropriate controls ($His_6$-tagged thioredoxin and buffer). Proteins were spotted at high concentration (40–200 μM), and a low concentration of Cy5-labeled bovine serum albumin (200 nM) was included in each sample to facilitate image analysis. Scanning for Cy5 fluorescence enables us to identify the location of all the spots on the microarrays and scanning for 5(6)-TAMRA fluorescence enables us to visualize and quantify domain–peptide interactions (Fig. 2).

We have previously shown that probing a protein microarray with a single concentration of a labeled peptide can produce very misleading results.[15] We therefore probed our arrays with eight concentrations of each peptide and fit the resulting spot intensities, $F_{obs}$, to eqn (1):

$$F_{obs} = F_0 + \frac{F_{max}[\text{peptide}]}{K_D + [\text{peptide}]} \quad (1)$$

where $F_0$ is the background fluorescence, $F_{max}$ is the maximum fluorescence at saturation, [peptide] is the total peptide concentration, and $K_D$ is the equilibrium dissociation constant (see, for examples, Fig. 2(E)). For each peptide, we fit all 133 curves, one for each domain. Interactions were scored as positive if the data fit well to eqn (1) ($R^2 > 0.9$), with $K_D < 2$ μM and $F_{max}$ at least two-fold higher than the mean fluorescence of control spots ($His_6$-tagged thioredoxin). The interactions that met these criteria for IGF1R 1346 are shown in Fig. 2(E). Following this strategy, we performed the same quantitative analysis for all 43 phosphopeptides. One of the peptides (FGFR1 677) exhibited high levels of nonspecific binding to the glass surface and to the control spots. High quality data were obtained for the

other 42 peptides (ESI,‡ Table S1 and http://www.sysbio.harvard.edu/csb/macbeath/data/data.html).

Using the data derived from this large-scale analysis, we constructed two quantitative protein interaction maps for each receptor, one using the physiological peptides and the other using the nonphysiological peptides (Fig. 3). The maps derived from the physiological peptides provide an unbiased, genome-wide view of each receptor, showing biophysical interactions between signaling proteins and sites of tyrosine phosphorylation. Which proteins are actually recruited to each receptor in the context of a cell depends on many additional factors, including the relative concentrations of receptors and proteins. As such, these diagrams should be viewed as maps of the receptors, rather than as a depiction of protein recruitment in any particular cell type or cell state. Just as a city map provides all possible routes from one destination to another, but does not specify the route that a specific individual follows on a given day, so too these maps provide the sum total of possible pY-mediated interactions between receptors and their downstream proteins, but do not represent which interactions are occurring in a given cell. We anticipate that these maps will help guide future investigations into the biology of these receptors and will facilitate efforts to construct quantitative models of RTK signaling.[16,17]

What do we learn by comparing the maps based on physiological sites with those based on nonphysiological sites (Fig. 3)? From visual inspection, it is clear that the physiological peptides bind to substantially more SH2 and PTB domains than the nonphysiological peptides. On average, the physiological peptides bind to 16.5 SH2 or PTB domains with a $K_D < 2$ μM, while the nonphysiological peptides bind to 3.5 domains. In addition, many of the interactions with the nonphysiological peptides are with a single peptide: EGFR 813 (Fig. 3). Excluding this peptide, the average number of interactions with nonphysiological peptide drops to 1.5. It is possible that Y813 of EGFR is, in fact, phosphorylated *in vivo* and that this event has escaped detection to date. Given its location in a well-structured, α-helical segment of the kinase domain,[41] however, it is more likely that this tyrosine is not accessible, but is surrounded by residues that, by chance, bind to SH2 and PTB domains.

Histograms showing the number of interactions per peptide (Fig. 4(A)) clearly highlight the difference between these two types of sequence. Since many of the nonphysiological peptides reside in the kinase domain while many of the physiological peptides reside in the solvent-accessible receptor tail, it is possible that the difference in their binding profiles arises simply from differences in their physicochemical properties. To test this hypothesis, we used three previously reported, quantitative descriptors of the physicochemical properties of amino acids (*z*-scales) to characterize each peptide sequence: $z_1$ is considered a descriptor of hydrophilicity, $z_2$ a descriptor of molecular weight and surface area, and $z_3$ a descriptor of polarity and charge.[42] Each peptide was expressed as an 18-dimensional vector, with three dimensions ($z_1$, $z_2$ and $z_3$) for each of the three amino acids upstream of the pY and three dimensions ($z_1$, $z_2$ and $z_3$) for each of the three amino acids downstream of the pY. The phosphopeptide vectors were then clustered using Euclidean distance as the similarity metric and a dendrogram was prepared using the centroid linkage method. As can be seen in Fig. 4(B), the physiological peptides are indistinguishable from the nonphysiological peptides. Similar results were obtained using other linkage methods (minimum, maximum, median) or using different numbers of residues up- and downstream of the pY (four or five; data not shown). This finding rules out the simplistic explanation that differences in the SH2 and PTB domain binding properties of physiological and nonphysiological peptides arise simply from differences in their physicochemical properties.

The observation that nonphysiological peptides (sequences that are not substrates for tyrosine kinases) bind to fewer domains than physiological peptides (sequences that are substrates) supports the hypothesis, first proposed by Cantley and co-workers,[23,24] that kinases and

interaction domains co-evolve. It also suggests, at least with respect to these three receptors, that selection favoring desirable interactions plays a larger role in establishing new recruitment sites than selection disfavoring deleterious interactions. Sequences that have not evolved to be phosphorylated tend, by default, not to bind to SH2 or PTB domains when they are phosphorylated artificially, even though the genome features over one hundred domains with diverse binding properties that recognize tyrosine-phosphorylated sequences. Even for a small binding site, there are many possible ligands and apparently only a small fraction of sequence space intersects with the binding preferences of natural SH2 and PTB domains. Incidental intersection cannot be ruled out, as is evident from the binding profile of EGFR 813, but, in general, selective pressure favoring new interactions must be applied in order to establish a physiological recruitment site within an RTK.

The striking preferential ability of the physiological peptides to bind to SH2 and PTB domains has another important implication: It strongly suggests that the sites of tyrosine phosphorylation from which they were derived play a biological role in recruiting SH2 or PTB domain-containing proteins in cells. Our finding is not a surprise for the well-characterized pY sites that have previously been shown to recruit signaling proteins *in vivo*. At least 35 proteins containing SH2 or PTB domains have been reported to interact with EGFR (ref. [43] and http://proteome.incyte.com) and our microarrays detect approximately two-thirds of them.[14] The situation is quite different, however, for FGFR1 and IGF1R. Substantially fewer direct interactions have been reported for these receptors and most of their pY sites have not yet been shown to act as recruitment sites. In the case of FGFR1, it is well-established that phospholipase C-γ (PLC-γ) is recruited to pY766 through its SH2 domain[44–46] and, as anticipated, we observe interactions with both PLC-γ1 and PLC-γ2 on our arrays. Very few other SH2- or PTB-containing proteins, however, have been shown to bind directly to sites of tyrosine phosphorylation on FGFR1. As a result, it is generally believed that recruitment to this receptor occurs primarily through interactions with the constitutively associated adaptor protein FRS2.[45,47] Our data suggest that there is substantially more direct recruitment to FGFR1 than is currently thought, and the same can be surmised for IGF1R. We submit that more interactions are occurring *in vivo* than have been described to date, but that these interactions are transient and so are not easily detected using standard biochemical approaches. This argues strongly for the development of new technologies that enable us to visualize dynamic, short-lived interactions in real time, in live cells.

From a pragmatic point of view, it is also instructive to know, for each receptor, if the nonphysiological peptides bind to domains that are not targeted by the physiological peptides. For the receptors studied here, we find that this is rare (Fig. 4(C)–(E)). In the case of EGFR, all but five of the domains that recognize nonphysiological peptides also recognize physiological peptides (Fig. 4(C)). With IGF1R, the number drops to three (Fig. 4(E)) and with FGFR1, every domain that binds a nonphysiological peptide also binds a physiological peptide (Fig. 4(D)). This has important implications for future studies of less well-characterized receptors. Even if the physiological sites of tyrosine phosphorylation are not known, quantitative interaction maps built using phosphopeptides derived from every intracellular tyrosine residue should approximate the ‘true’ maps built using only physiological sites of phosphorylation. Although such maps will contain extraneous, incorrect pY sites, many of these incorrect sites will feature no interactions and the others will feature interactions that largely overlap with correct interactions.

Based on the interaction maps of EGFR, FGFR1 and IGF1R, how do these receptors achieve specificity in signaling? It is likely that, in many cases, cellular context defines biological outcome. The identities and concentrations of signaling proteins vary from one cell type to the next, as do the levels of the transcription factors that carry out their instructions. Nevertheless, it has been shown that different RTKs induce different biological outcomes in the same cellular

background. For example, FGFR1 stimulates neurite outgrowth in PC12 cells, while EGFR stimulates mitogenesis.[13] When we compare the recruitment profiles of EGFR, FGFR1 and IGF1R at a qualitative level, we find that these receptors are remarkably similar (Fig. 5). Thirty-nine of the SH2- or PTB-containing proteins that recognize physiological sites of tyrosine phosphorylation are common to all three receptors; only 11, 7 and 3 proteins are unique to EGFR, FGFR1 and IGF1R, respectively. It is possible that specificity of signaling can be explained entirely by the proteins that are unique to each receptor. A closer examination of the proteins that are common to all three receptors, however, shows that they are very well-characterized proteins that activate well-studied signaling pathways (Fig. 5). Moreover, these proteins have been described to promote disparate and even opposing cellular responses: proliferation (*e.g.*, Shc and RasA1);[48] migration (*e.g.*, Abl2, CrkL and Tenc1);[49–51] differentiation (*e.g.*, Jak2 and Src);[52,53] survival (*e.g.*, PI3K and PLC-γ);[54,55] and apoptosis (*e.g.*, STAT1).[56,57] We hypothesize, based on these observations, that quantitative differences in the affinities of each receptor for the proteins they recruit play an important role in defining specificity. For example, although all three receptors feature pY sites that bind to the SH2 domains of the three isoforms of phosphoinositide-3 kinase (PI3K), these connections are strongest in FGFR1 and weakest in IGF1R. Similarly, although all three receptors features sites that bind to the PTB domain of Shc, there are five such connections in EGFR and only one each in FGFR1 and IGF1R. We are currently investigating the hypothesis that quantitative differences in the biophysical recruitment profile of RTKs are predictive of quantitative differences in the signaling pathways they induce.

## Experimental

### Peptide synthesis and purification

Peptides were synthesized at 2 μmol scale in a 96-well plate with an Intavis MultiPep synthesizer (Koeln, Germany) using standard Fmoc chemistry. Pre-loaded NovaSyn TGT resin was from Novabiochem (San Diego, CA). Fmoc-protected amino acids were activated *in situ* with HBTU/*N*-methylmorpholine and coupled at five-fold molar excess over peptide. Phosphotyrosine was incorporated using *N*-α-Fmoc-*O*-benzyl-L–phosphotyrosine. Each coupling cycle was followed by capping with acetic anhydride to avoid accumulation of one-residue deletion peptide byproducts. At the end of the synthesis, peptides were labeled on the resin with 5-(and-6)-carboxytetramethylrhodamine [5(6)-TAMRA] from Anaspec (San Jose, CA). 5(6)-TAMRA was coupled using HATU/diisopropylethylamine at a seven-fold molar excess over peptide. After labeling, the peptides were deprotected and cleaved from the resin with 95% trifluoroacetic acid, 2.5% triisopropylsilane, 2.5% dd$H_2O$ and were subsequently precipitated by addition of cold ether.

Crude peptides were purified by reversed-phase HPLC using a Kromasil 100-5-C18 semipreparative column (Peeke Scientific, Sunnyvale, CA). Fractions containing the correct product were identified by MALDI-TOF mass spectrometry using a Voyager DE Pro (Applied Biosystems, Foster City, CA). Solvent was removed by lyophilization and the purified peptides stored at −80 °C.

### Production and purification of recombinant proteins

Coding regions for STAT1, STAT2 and Cbl were cloned as previously described[14] and transferred into the expression vector pET-32-DEST.[14] The coding region of STAT1 included residues 132–683; the coding region for STAT2 included residues 135–680; and the coding region for Cbl included residues 24–351. Expression vectors were introduced into BL21(DE3) pLysS *Escherichia coli* and cells from a single ampicillin-resistant colony were grown at 37 °C in 500 mL LB medium supplemented with 100 μg $mL^{-1}$ of ampicillin to an $OD_{600}$ of 0.6. Isopropyl-1-thio-β-D-galactopyranoside (IPTG) was added to a final concentration of 1 mM

and the cultures shaken at 20 °C for 15 h. Cells were recovered by centrifugation and resuspended in 10 mL lysis buffer (50 mM $Na_2HPO_4$, 150 mM NaCl, 10 mM imidazole, 0.1% NP40, 1 mM benzamidine, 1 mM phenylmethylsulfonyl fluoride, pH 8.0). After adding 125 U of lysozyme and 150 kU of benzonase, the cells were lysed by sonication and insoluble material was removed by centrifugation at 30 000 *g* for 30 min, 4 °C. $His_6$-tagged proteins were purified by immobilized metal affinity chromatography using 500 μL Ni-NTA agarose beads (Qiagen, Valencia, CA) and eluted using 50 mM $Na_2HPO_4$, 300 mM NaCl, 250 mM imidazole, pH 8.0. Purified proteins were exchanged into phosphate buffered saline (PBS), pH 8.0 and their concentrations were determined based on their absorbance at 280 nm. Since the Cbl domain requires $Ca^{2+}$ for stable folding, the Ni-NTA-immobilized protein was washed with 20 mM HEPES, 2.5 mM $CaCl_2$, 20 mM imidazole, pH 7.4 and eluted with 20 mM HEPES, 2.5 mM $CaCl_2$, 250 mM imidazole, pH 7.4. Purified protein was exchanged into 20 mM HEPES, 2.5 mM $CaCl_2$, pH 7.4 using a NAP-10 column (GE Healthcare, Piscataway, NJ). Glycerol was added to all samples to a final concentration of 20% (v/v) and the proteins were divided into aliquots and stored at −80 °C.

### Fabrication and processing of protein microarrays

Purified recombinant domains were spotted at a concentration of 40–200 μM onto 112.5 mm × 74.5 mm × 1 mm aldehyde silane-coated glass substrates (Erie Scientific Company, Portsmouth, NH) using a NanoPrint microarrayer (TeleChem International, Inc., Sunnyvale, CA). 96 microarrays (2 different sets of 48 identical arrays) were fabricated in a 12 × 8 pattern on the glass, with a pitch of 9 mm. Each array consisted of a 16 × 17 pattern of spots, with a 250 μm pitch. The proteins were all spotted in quadruplicate. Following a 1 h-incubation, the glass was attached to the bottom of a bottomless 96-well microtiter plate (Greiner Bio-one, Kremsmuenster, Austria) using an intervening silicone gasket (Grace Bio-Labs, Bend, OR). Following attachment, the arrays were frozen and stored at −80 °C. Immediately before use, the plates were quenched with buffer A (20 mM HEPES, 100 mM KCl, 0.1% Tween-20, pH 7.8) containing 1% BSA (w/v) for 30 min at room temperature, followed by several rinses with buffer A. Arrays were then probed with 8 different concentrations of 5(6)-TAMRA-labeled peptides, dissolved in buffer A: 5 μM, 3 μM, 2 μM, 1 μM, 500 nM, 200 nM, 100 nM and 10 nM. Following a 1-h incubation at room temperature, the peptide solution was removed and the arrays washed with 150 μL buffer A for 10 s. The arrays were rinsed briefly with $ddH_2O$ and spun upside down in a centrifuge for 1 min to remove residual water.

### Scanning and analysis of microarrays

Protein microarrays were scanned at 10 μm resolution using a Tecan LS400 microarray scanner (Tecan, Salzburg, Austria). Cy5 fluorescence was imaged using a 633 nm laser, while 5(6)-TAMRA fluorescence was imaged using a 543 nm laser. Images were analyzed using Array-Pro Analyzer 4.5 (Tecan). Spots were identified based on the Cy5 image and the mean fluorescence of each spot was calculated as the ratio of the 5(6)-TAMRA fluorescence to the Cy5 fluorescence. These data were imported into Mathematica 5.0 (Wolfram Research, Inc., Champaign, IL) and concentration-dependent measurements fit to eqn (1) for each domain–peptide pair. The resulting quantitative data were displayed graphically using the software package Cytoscape 2.1, which is freely available on the web at http://www.cytoscape.org/.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

We thank John E. Allen for synthesizing a soluble version of FGFR1 766; Jeffrey Knott, Susan Rogers and Colleen Hunter for synthesizing the other phosphopeptides; Jiunn R. Chen for suggestions and help with analyzing the physicochemical properties of the peptides; and the Faculty of Arts and Sciences Center for Systems Biology for support with instrumentation. This work was supported by awards from the W. M. Keck Foundation, the Arnold and Mabel Beckman Foundation, and the Camille and Henry Dreyfus Foundation and by a grant from the National Institutes of Health (1 R33 CA128726-01). A. K. and B. C. were supported in part by the NIH Molecular, Cellular, and Chemical Biology Training Grant (5 T32 GM07598-25) and A. G. is the recipient of an NSF Graduate Research Fellowship.

## References

1. Schlessinger J, Lemmon MA. Sci STKE 2003;2003:RE12. [PubMed: 12865499]
2. Porter AC, Vaillancourt RR. Oncogene 1998;17:1343–1352. [PubMed: 9779982]
3. Yarden Y, Sliwkowski MX. Nat Rev Mol Cell Biol 2001;2:127–137. [PubMed: 11252954]
4. Zwick E, Bange J, Ullrich A. Endocr Relat Cancer 2001;8:161–173. [PubMed: 11566607]
5. Heldin CH. Cell 1995;80:213–223. [PubMed: 7834741]
6. Plotnikov AN, Schlessinger J, Hubbard SR, Mohammadi M. Cell 1999;98:641–650. [PubMed: 10490103]
7. Remy I, Wilson IA, Michnick SW. Science 1999;283:990–993. [PubMed: 9974393]
8. Li E, Hristova K. Biochemistry 2006;45:6241–6251. [PubMed: 16700535]
9. Robertson SC, Tynan J, Donoghue DJ. Trends Genet 2000;16:368. [PubMed: 10904266]
10. Sadowski I, Stone JC, Pawson T. Mol Cell Biol 1986;6:4396–4408. [PubMed: 3025655]
11. Kavanaugh WM, Williams LT. Science 1994;266:1862–1865. [PubMed: 7527937]
12. Felder S, Zhou M, Hu P, Urena J, Ullrich A, Chaudhuri M, White M, Shoelson SE, Schlessinger J. Mol Cell Biol 1993;13:1449–1455. [PubMed: 7680095]
13. Hondermarck H, McLaughlin CS, Patterson SD, Bradshaw RA. Proc Natl Acad Sci USA 1994;91:9377–9381. [PubMed: 7937773]
14. Jones RB, Gordus A, Krall JA, MacBeath G. Nature 2006;439:168–174. [PubMed: 16273093]
15. Gordus A, MacBeath G. J Am Chem Soc 2006;128:13668–13669. [PubMed: 17044677]
16. Schoeberl B, Eichler-Jonsson C, Gilles ED, Muller G. Nat Biotechnol 2002;20:370–375. [PubMed: 11923843]
17. Wiley HS, Shvartsman SY, Lauffenburger DA. Trends Cell Biol 2003;13:43–50. [PubMed: 12480339]
18. Hinsby AM, Olsen JV, Bennett KL, Mann M. Mol Cell Proteomics 2003;2:29–36. [PubMed: 12601080]
19. Lombardo CR, Consler TG, Kassel DB. Biochemistry 1995;34:16456–16466. [PubMed: 8845374]
20. Mohammadi M, Dikic I, Sorokin A, Burgess WH, Jaye M, Schlessinger J. Mol Cell Biol 1996;16:977–989. [PubMed: 8622701]
21. Peterson JE, Kulik G, Jelinek T, Reuter CW, Shannon JA, Weber MJ. J Biol Chem 1996;271:31562–31571. [PubMed: 8940173]
22. Stover DR, Becker M, Liebetanz J, Lydon NB. J Biol Chem 1995;270:15591–15597. [PubMed: 7797556]
23. Songyang Z, Cantley LC. Trends Biochem Sci 1995;20:470–475. [PubMed: 8578591]
24. Songyang Z, Carraway KL 3rd, Eck MJ, Harrison SC, Feldman RA, Mohammadi M, Schlessinger J, Hubbard SR, Smith DP, Eng C, Lorenzo MJ, Ponder BAJ, Mayer BJ, Cantley LC. Nature 1995;373:536–539. [PubMed: 7845468]
25. Gevaert K, Staes A, Van Damme J, De Groot S, Hugelier K, Demol H, Martens L, Goethals M, Vandekerckhove J. Proteomics 2005;5:3589–3599. [PubMed: 16097034]
26. Zhang Y, Wolf-Yadlin A, Ross PL, Pappin DJ, Rush J, Lauffenburger DA, White FM. Mol Cell Proteomics 2005;4:1240–1250. [PubMed: 15951569]
27. Wu W, Graves LM, Gill GN, Parsons SJ, Samet JM. J Biol Chem 2002;277:24252–24257. [PubMed: 11983694]

28. Huang PH, Mukasa A, Bonavia R, Flynn RA, Brewer ZE, Cavenee WK, Furnari FB, White FM. Proc Natl Acad Sci USA 2007;104:12867–12872. [PubMed: 17646646]
29. Foehr ED, Raffioni S, Murray-Rust J, Bradshaw RA. J Biol Chem 2001;276:37529–37536. [PubMed: 11459840]
30. Leahy M, Lyons A, Krause D, O'Connor R. J Biol Chem 2004;279:18306–18313. [PubMed: 14963047]
31. O'Connor R, Kauffmann-Zeh A, Liu Y, Lehar S, Evan GI, Baserga R, Blattler WA. Mol Cell Biol 1997;17:427–435. [PubMed: 8972223]
32. Eck MJ, Dhe-Paganon S, Trub T, Nolte RT, Shoelson SE. Cell 1996;85:695–705. [PubMed: 8646778]
33. Zhou MM, Huang B, Olejniczak ET, Meadows RP, Shuker SB, Miyazaki M, Trub T, Shoelson SE, Fesik SW. Nat Struct Biol 1996;3:388–393. [PubMed: 8599766]
34. Zhou MM, Ravichandran KS, Olejniczak EF, Petros AM, Meadows RP, Sattler M, Harlan JE, Wade WS, Burakoff SJ, Fesik SW. Nature 1995;378:584–592. [PubMed: 8524391]
35. Case RD, Piccione E, Wolf G, Benett AM, Lechleider RJ, Neel BG, Shoelson SE. J Biol Chem 1994;269:10467–10474. [PubMed: 8144631]
36. Lee CH, Kominos D, Jacques S, Margolis B, Schlessinger J, Shoelson SE, Kuriyan J. Structure 1994;2:423–438. [PubMed: 7521735]
37. Pascal SM, Singer AU, Gish G, Yamazaki T, Shoelson SE, Pawson T, Kay LE, Forman-Kay JD. Cell 1994;77:461–472. [PubMed: 8181064]
38. MacBeath G, Schreiber SL. Science 2000;289:1760–1763. [PubMed: 10976071]
39. Mao X, Chen X. Acta Crystallogr, Sect F: Struct Biol Cryst Commun 2005;61:666–668.
40. Meng W, Sawasdikosol S, Burakoff SJ, Eck MJ. Nature 1999;398:84–90. [PubMed: 10078535]
41. Stamos J, Sliwkowski MX, Eigenbrot C. J Biol Chem 2002;277:46265–46272. [PubMed: 12196540]
42. Sandberg M, Eriksson L, Jonsson J, Sjostrom M, Wold S. J Med Chem 1998;41:2481–2491. [PubMed: 9651153]
43. Peri S, Navarro JD, Amanchy R, Kristiansen TZ, Jonnalagadda CK, Surendranath V, Niranjan V, Muthusamy B, Gandhi TK, Gronborg M, Ibarrola N, Deshpande N, Shanker K, Shivashankar HN, Rashmi BP, Ramya MA, Zhao Z, Chandrika KN, Padma N, Harsha HC, Yatish AJ, Kavitha MP, Menezes M, Choudhury DR, Suresh S, Ghosh N, Saravana R, Chandran S, Krishna S, Joy M, Anand SK, Madavan V, Joseph A, Wong GW, Schiemann WP, Constantinescu SN, Huang L, Khosravi-Far R, Steen H, Tewari M, Ghaffari S, Blobe GC, Dang CV, Garcia JG, Pevsner J, Jensen ON, Roepstorff P, Deshpande KS, Chinnaiyan AM, Hamosh A, Chakravarti A, Pandey A. Genome Res 2003;13:2363–2371. [PubMed: 14525934]
44. Burgess WH, Dionne CA, Kaplow J, Mudd R, Friesel R, Zilberstein A, Schlessinger J, Jaye M. Mol Cell Biol 1990;10:4770–4777. [PubMed: 2167438]
45. Eswarakumar VP, Lax I, Schlessinger J. Cytokine Growth Factor Rev 2005;16:139–149. [PubMed: 15863030]
46. Mohammadi M, Honegger AM, Rotin D, Fischer R, Bellot F, Li W, Dionne CA, Jaye M, Rubinstein M, Schlessinger J. Mol Cell Biol 1991;11:5068–5078. [PubMed: 1656221]
47. Zhang Y, McKeehan K, Lin Y, Zhang J, Wang F. Mol Endocrinol 2008;22:167–175. [PubMed: 17901128]
48. Katz M, Amit I, Yarden Y. Biochim Biophys Acta 2007;1773:1161–1176. [PubMed: 17306385]
49. Woodring PJ, Hunter T, Wang JY. J Cell Sci 2003;116:2613–2626. [PubMed: 12775773]
50. Feller SM. Oncogene 2001;20:6348–6371. [PubMed: 11607838]
51. Chen H, Duncan IC, Bozorgchami H, Lo SH. Proc Natl Acad Sci USA 2002;99:733–738. [PubMed: 11792844]
52. Horne WC, Sanjay A, Bruzzaniti A, Baron R. Immunol Rev 2005;208:106–125. [PubMed: 16313344]
53. Rane SG, Reddy EP. Oncogene 2000;19:5662–5679. [PubMed: 11114747]
54. Cantley LC. Science 2002;296:1655–1657. [PubMed: 12040186]
55. Patterson RL, van Rossum DB, Nikolaidis N, Gill DL, Snyder SH. Trends Biochem Sci 2005;30:688–697. [PubMed: 16260143]
56. Kim HS, Lee MS. Cell Signalling 2007;19:454–465. [PubMed: 17085014]

57. Schindler C. Trends Cell Biol 1998;8:97–98. [PubMed: 9695817]

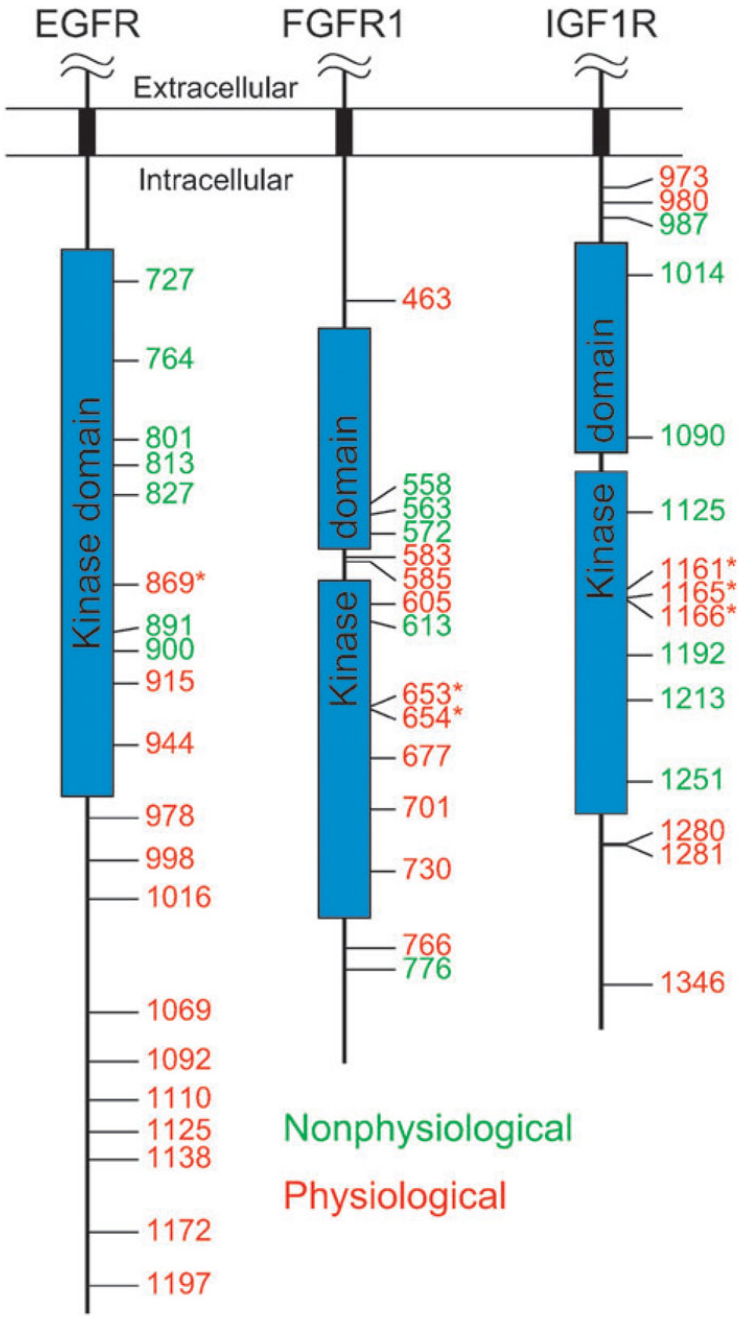


**Fig. 1.**

Intracellular domains of EGFR, FGFR1 and IGF1R. Tyrosine residues that have been shown to be physiological sites of phosphorylation are colored red; the other tyrosines have been designated ‘nonphysiological’ and are colored green. Asterisks denote tyrosine residues located in the activation loops of the kinase domains.

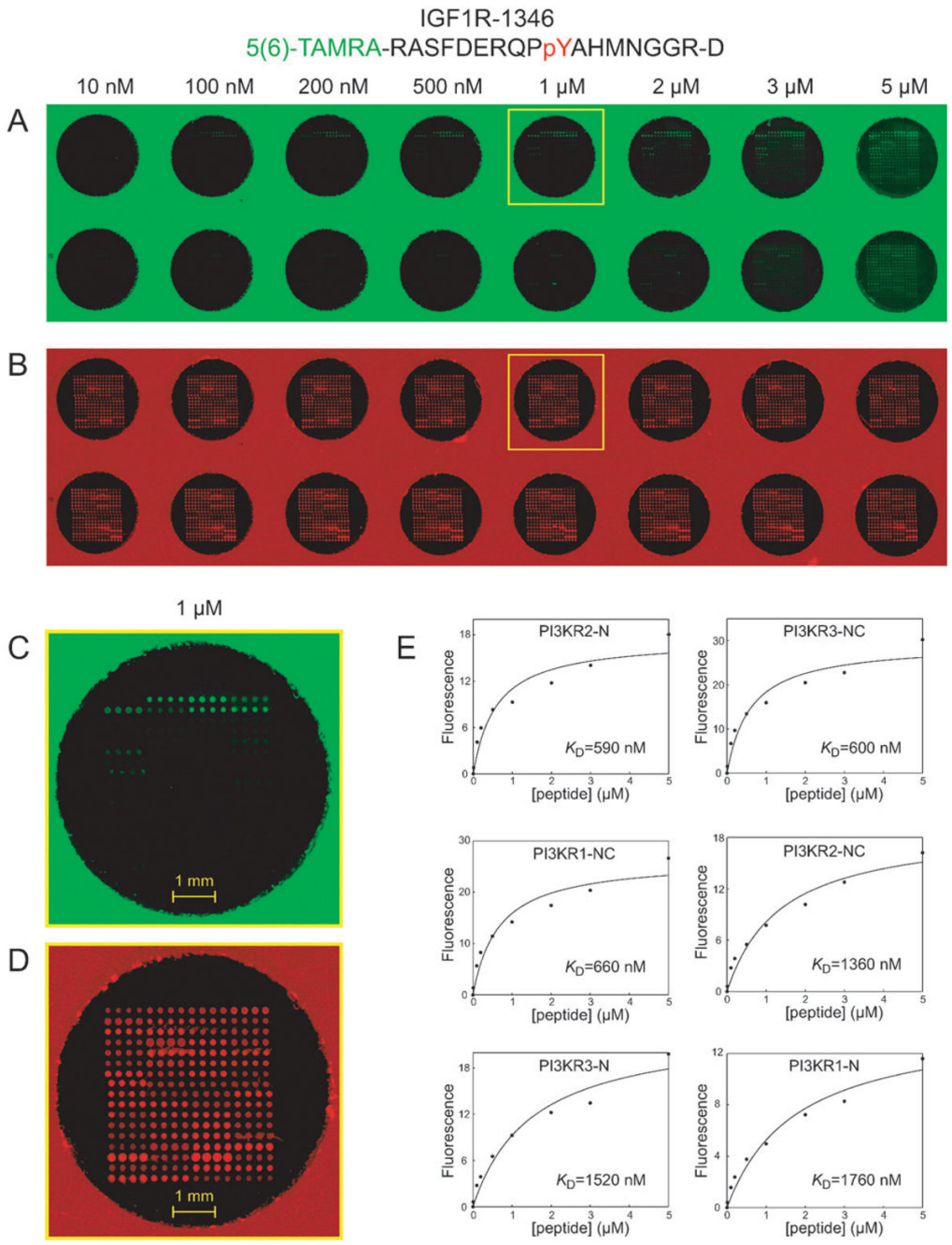


**Fig. 2.**

Measuring the binding affinity of SH2 and PTB domains for phosphopeptide IGF1R 1346 using protein microarrays. (A) Fluorescence images of SH2 and PTB domain microarrays in separate wells of a 96-well microtiter plate, obtained using a 633 nm laser. The fluorescence arises from a trace amount of Cy5-labeled BSA that was added to each protein before arraying. (B) Fluorescence images of the same SH2 and PTB domain microarrays, probed with eight concentrations of a 5(6)-TAMRA-labeled phosphopeptide derived from IGF1R (pY1346). The images were obtained using a 543 nm laser. (C, D) Close up images of a single well, obtained using 633 nm and 543 nm lasers, respectively. The well corresponds to the one highlighted by a yellow box in panels A and B. (E) Plots showing fluorescence as a function of peptide

concentration for six high-affinity interactions. The data were fit to eqn (1) to determine the $K_D$.

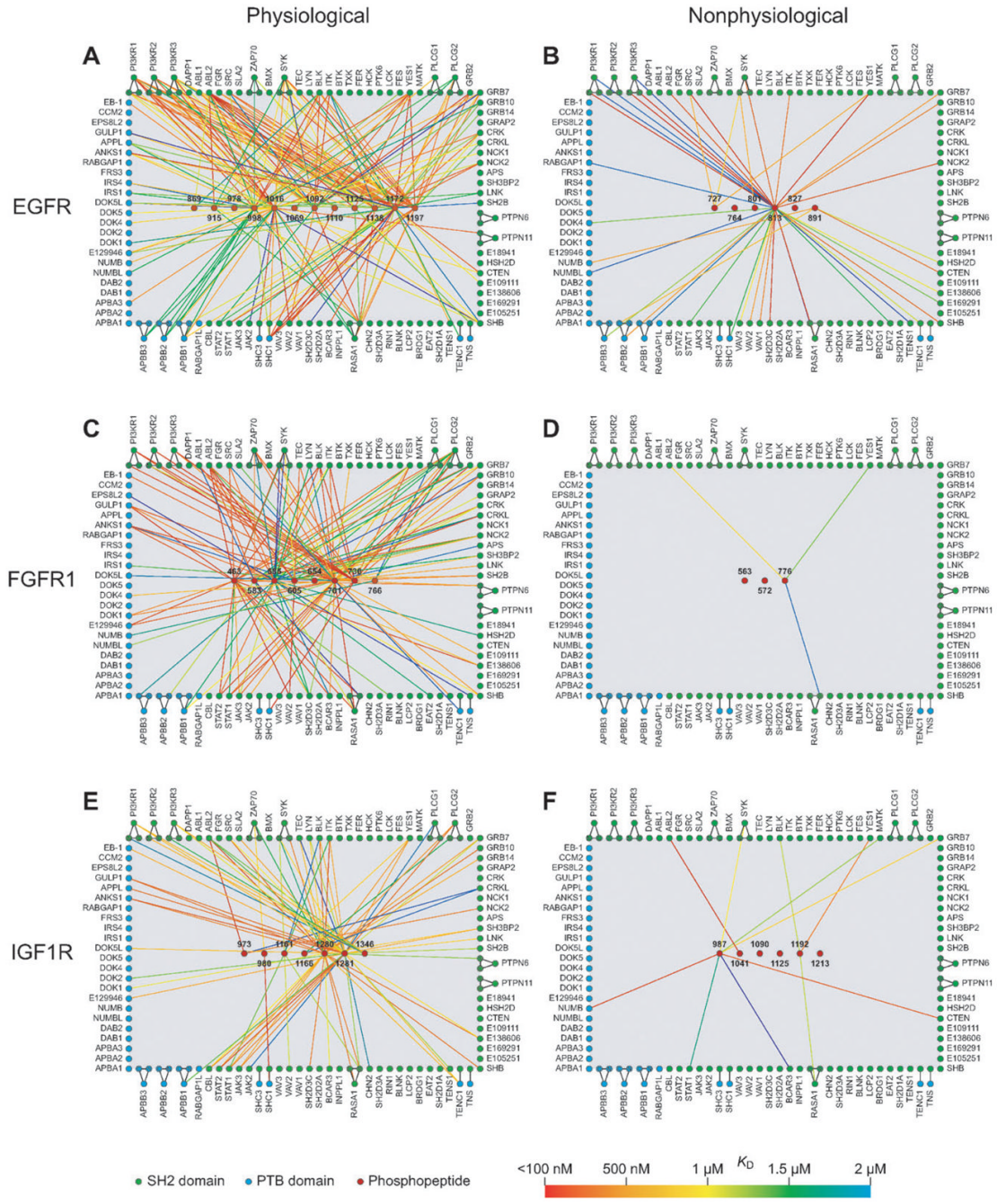


**Fig. 3.**

Quantitative protein interaction maps for EGFR, FGFR1 and IGF1R. Two maps are shown for each receptor. The maps on the left (A, C, E) are constructed using physiological peptides derived from EGFR, FGFR1 and IGF1R, respectively; the maps on the right (B, D, F) are constructed using nonphysiological peptides derived from EGFR, FGFR1 and IGF1R, respectively. Red circles represent phosphopeptides, green circles represent SH2 domains and blue circles represent PTB domains. Lines connecting peptides to domains indicate observed interactions, colored according to the affinity of the interaction (see legend). The circles that lie outside the rectangle of individual domains and are connected by black lines to two other circles, represent tandem domains. They comprise the two domains to which they are

connected. The circles that lie outside the rectangle but are connected to only one other circle represent proteins that contain both an SH2 domain and a PTB domain.

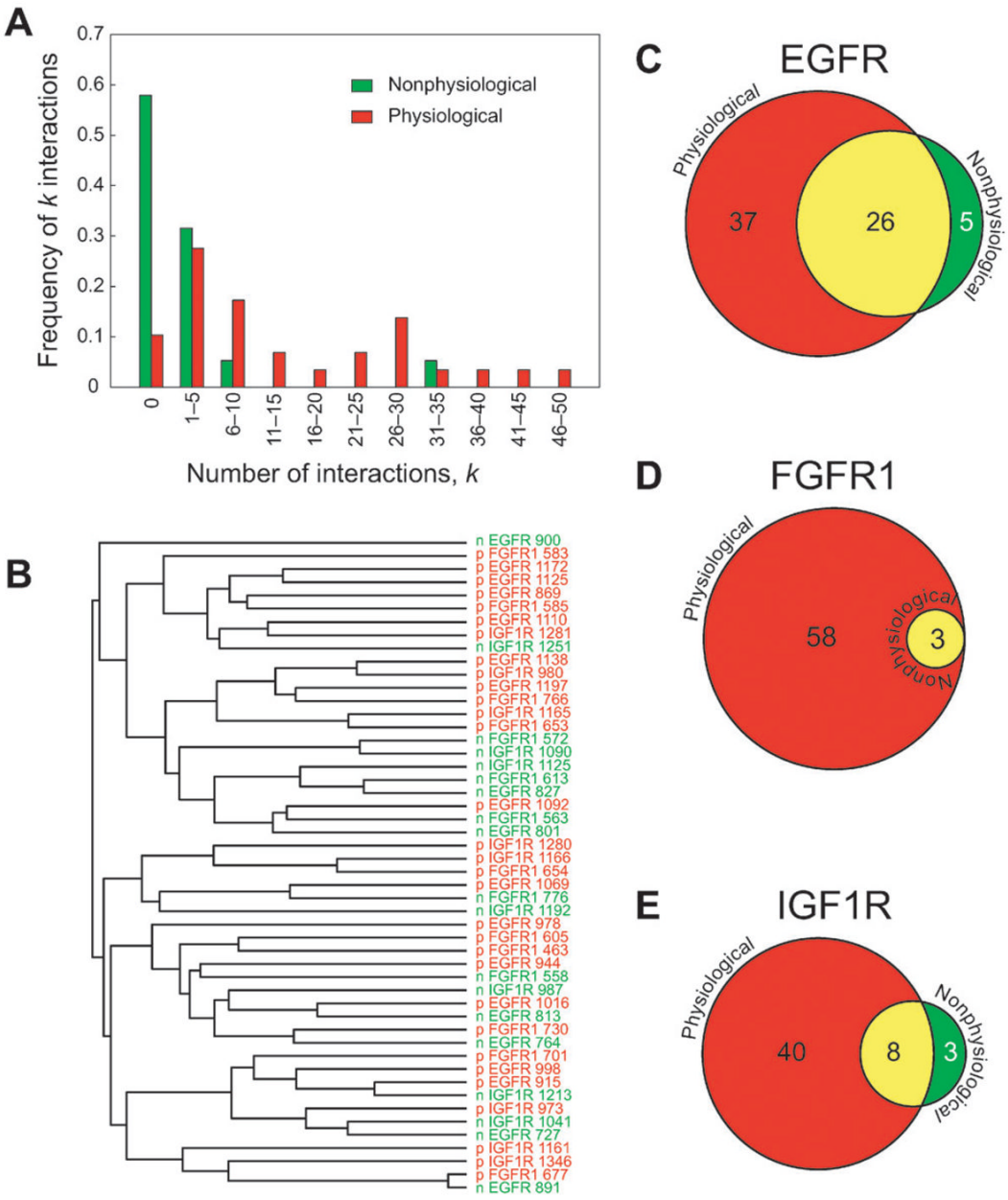


**Fig. 4.**

Comparison of the physiological and nonphysiological phosphopeptides. (A) Histograms showing the frequency of peptides with different numbers of interactions. Physiological peptides are shown in red; nonphysiological peptides are shown in green. (B) Dendrogram showing phosphopeptides clustered according to their physicochemical properties. Each peptide was described by an 18-dimensional vector comprising the first three *z*-scales of the three residues up- and downstream of the phosphotyrosine. The vectors were clustered using Euclidean distance as the similarity metric and the dendrogram was prepared using the centroid linkage method. Physiological peptides are designated by a ‘p’ and are colored red; nonphysiological peptides are designated by an ‘n’ and are colored green. (C–E) Venn diagrams showing proteins that interact with phosphopeptides derived from EGFR, FGFR1 and IGF1R, respectively. For each receptor, the circle on the left (red) represents all the proteins with at

least one SH2 or PTB domain that interacts with at least one physiological phosphopeptide and the circle on the right (green) represents all the proteins with at least one SH2 or PTB domain that interacts with at least one nonphysiological phosphopeptide. Overlap is colored yellow.

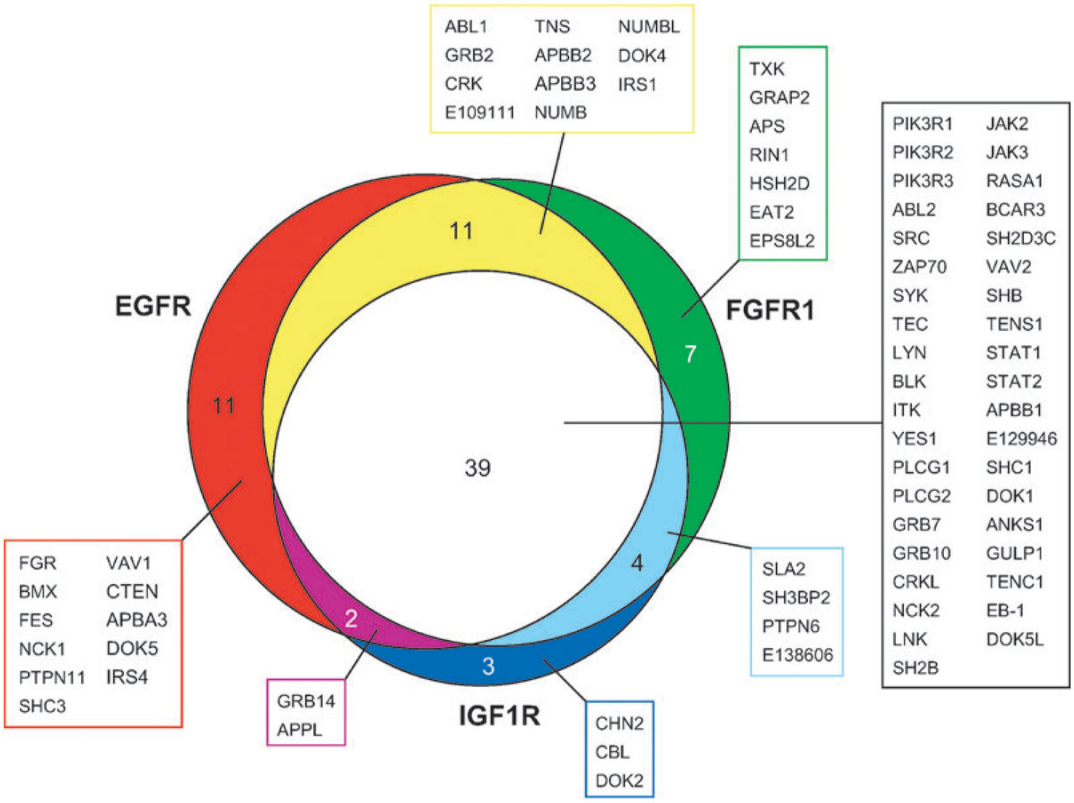


**Fig. 5.**

Venn diagram showing proteins that interact with phosphopeptides derived from EGFR, FGFR1 and IGF1R. The red, green and blue circles represent all the proteins with at least one SH2 or PTB domain that interacts with at least one physiological phosphopeptide derived from EGFR, FGFR1 and IGF1R, respectively. Overlap between EGFR and FGFR1 is colored yellow; overlap between EGFR and IGF1R is colored purple; overlap between FGFR1 and IGF1R is colored light blue; and overlap between all three receptors is colored white.

## Table 1

Phosphopeptides derived from intracellular tyrosines on EGFR, FGFR1 and IGF1R

| EGFR | | FGFR1 | | IGF1R | |
|---|---|---|---|---|---|
| Site | Peptide sequence | Site | | Site | Peptide sequence |
| | | **Physiological** | | | |
| 869 | KLLGAEEKEpYHAEGGKVD | 463 | TPMLAGVSEpYELPEDPRD | 973 | NSRLGNGVLpYASVNPEYD |
| 915 | ELMTFGSKPpYDGIPASED | 583 | QARRPPGLEpYCYNPSHND | 980 | VLYASVNPEpYFSAADVYD |
| 944[a] | QPPICTIDVpYMIMVKCWD | 585 | RRPPGLEYCpYNPSHNPED | 1161 | GDFGMTRDIpYETDYYRKD |
| 978 | SKMARDPQRpYLVIQGDED | 605 | SSKDLVSCApYQVARGMED | 1165[b] | MTRDIYETDpYYRKGGKGD |
| 998 | LPSPTDSNFpYRALMDEED | 653[b] | LARDIHHIDpYYKKTTNGD | 1166 | TRDIYETDYpYRKGGKGLD |
| 1016 | MDDVVDADEpYLIPQQGFD | 654 | ARDIHHIDYpYKKTTNGRD | 1280 | EPGFREVSFpYYSEENKLD |
| 1069 | IKEDSFLQRpYSSDPTGAD | 677[c] | APEALFDRIpYTHQSDVWD | 1281 | PGFREVSFYpYSEENKLPD |
| 1092 | DDTFLPVPEpYINQSVPKD | 701 | EIFTLGGSPpYPGVPVEED | 1346 | RASFDERQPpYAHMNGGRD |
| 1110 | PAGSVQNPVpYHNQPLNPD | 730 | KPSNCTNELpYMMMRDCWD | | |
| 1125 | NPAPSRDPHpYQDPHSTAD | 766[d] | ALTSNQEpYLDLSMD | | |
| 1138 | HSTAVGNPEpYLNTVQPTD | | | | |
| 1172 | HQISLDNPDpYQQDFFPKD | | | | |
| 1197 | KGSTAENAEpYLRVAPQSD | | | | |
| | | **NonPhysiological** | | | |
| 727 | LGSGAFGTVpYKGLWIPED | 558[a] | GACTQDGPLpYVIVEYASD | 987 | PEYFSAADVpYVPDEWEVD |
| 764 | ANKEILDEApYVMASVDND | 563 | DGPLYVIVEpYASKGNLRD | 1041 | LGQGSFGMVpYEGVAKGVD |
| 801 | LMPFGCLLDpYVREHKDND | 572 | YASKGNLREpYLQARRPPD | 1090 | LMTRGDLKSpYLRSLRPED |
| 813 | EHKDNIGSQpYLLNWCVQD | 613[a] | AYQVARGMEpYLASKKCID | 1125 | AGEIADGMApYLNANKFVD |
| 827 | CVQIAKGMNpYLEDRRLVD | 776 | LDLSMPLDQpYSPSFPDTD | 1192 | SLKDGVFTTpYSDVWSFGD |
| 891 | ALESILHRIpYTHQSDVWD | | | 1213 | EIATLAEQPpYQGLSNEQD |
| 900[a] | YTHQSDVWSpYGVTVWELD | | | 1251[b] | FELMRMCWQpYNPKMRPSD |

[a]No product of the correct molecular weight was obtained for these peptides after two attempts. [b]These peptides were not sufficiently soluble in aqueous buffer to permit microarray analysis. [c]This peptide exhibited high nonspecific binding on the microarrays, precluding data analysis. [d]This peptide was insoluble as a full-length peptide and so was truncated by two residues on either end to remove hydrophobic residues. The shortened peptide gave high-quality data.