

Enhanced Prediction of Src Homology 2 (SH2) Domain Binding Potentials Using a Fluorescence Polarization-derived c-Met, c-Kit, ErbB, and Androgen Receptor Interactome*[§]

Kin K. Leung^{‡§}, Ronald J. Hause, Jr.^{§¶||}, John L. Barkinge^{**}, Mark F. Ciaccio^{‡§§}, Chih-Pin Chuu^{¶||}, and Richard B. Jones^{‡¶|||}

Many human diseases are associated with aberrant regulation of phosphoprotein signaling networks. Src homology 2 (SH2) domains represent the major class of protein domains in metazoans that interact with proteins phosphorylated on the amino acid residue tyrosine. Although current SH2 domain prediction algorithms perform well at predicting the sequences of phosphorylated peptides that are likely to result in the highest possible interaction affinity in the context of random peptide library screens, these algorithms do poorly at predicting the interaction potential of SH2 domains with physiologically derived protein sequences. We employed a high throughput interaction assay system to empirically determine the affinity between 93 human SH2 domains and phosphopeptides abstracted from several receptor tyrosine kinases and signaling proteins. The resulting interaction experiments revealed over 1000 novel peptide-protein interactions and provided a glimpse into the common and specific interaction potentials of c-Met, c-Kit, GAB1, and the human androgen receptor. We used these data to build a permutation-based logistic regression classifier that performed considerably better than existing algorithms for predicting the interaction potential of several SH2 domains. *Molecular & Cellular Proteomics* 13: 10.1074/mcp.M113.034876, 1705–1723, 2014.

Src homology 2 protein domains (SH2)¹ are modular self-folding entities of about 100 amino acids that bind to tyrosine-

From the [‡]Committee on Cancer Biology, [¶]Committee on Genetics, Genomics, and Systems Biology, and ^{‡‡}Committee on Cellular and Molecular Physiology, The Ben May Department for Cancer Research and the Institute for Genomics and Systems Biology, The Gwen and Jules Knapp Center for Biomedical Discovery, University of Chicago, Chicago, Illinois 60637

* Author's Choice—Final version full access.

Received October 7, 2013, and in revised form, March 16, 2014

Published, MCP Papers in Press, April 12, 2014 DOI 10.1074/mcp.M113.034876

Author contributions: K.K.L., R.J.H., J.L.B., M.F.C., C.C., and R.B.J. designed research; K.K.L., R.J.H., J.L.B., and C.C. performed research; K.K.L., R.J.H., M.F.C., and C.C. contributed new reagents or analytic tools; K.K.L., R.J.H., and M.F.C. analyzed data; K.K.L., R.J.H., J.L.B., and R.B.J. wrote the paper.

¹ The abbreviations used are: SH2, Src homology 2; ANN, artificial neural network; AR, androgen receptor; EGFR, epidermal growth

phosphorylated peptide sequences contained within target proteins. The SH2 domain (1–3) was originally described nearly 20 years ago as an N-terminal region of the FES protein kinase that was not required for kinase activity but was important for its regulation. More recent studies have demonstrated that SH2 domains exist in many signaling molecules, including PLC γ 1, Ras GAP, c-Src, and PI3KR. SH2 domains have been shown to enable the interaction of these signaling proteins with growth factor receptors such as FGFR1, EGFR, c-Met, and PDGFR in a phosphospecific manner (4–9). Subsequently, random peptide library screening approaches were used to define sequence motifs that resulted in the highest affinity interactions within particular SH2 domain classes (10, 11). For example, peptide sequences containing the pYEEI, pYXN, and pYMXM motifs were described to result in the highest affinity interactions with the SH2 domains from c-Src, Grb2, and the PI3KR SH2 domains, respectively. Data from such experiments have been used to generate predictions regarding the likelihood that any particular peptide sequence will interact with any particular SH2 domain (12–14).

Unfortunately, the predictive performance of these algorithms has not been thoroughly empirically tested or optimized for biologically derived peptide sequences. We and others reported the first comprehensive cloning, expression, and functional analysis of human genome-encoded SH2 domains using a protein microarray-based interaction analysis approach (15–17). Similarly, peptide arrays have been used to query the interaction potential of SH2 domains with biologically derived peptide sequences in a semi-quantitative manner (18). These studies demonstrated that most biologically derived peptide sequences contained within RTKs and signaling proteins do not represent best fit sequence motifs and

factor receptor; FGFR, fibroblast growth factor receptor; FP, fluorescence polarization; IGF1R, insulin-like growth factor receptor; PEBL, permutation-based logistic regression classifier; PDGFR, platelet-derived growth factor receptor; PM, protein microarray; PTB, phosphotyrosine binding; phosphotyrosine; RF, random forests; ROC, receiver operating characteristic; RTK, receptor tyrosine kinase; PLC, phospholipase C.

interact at a much lower affinity than with the optimal sequence motifs identified previously from random peptide libraries. Studies with biologically derived peptides indicated that context nonpermissive amino acids often contribute as much predictive information regarding interaction selectivity as positively contributing amino acids (19). Taken together, these results suggest that the collection of large quantitative protein interaction datasets between SH2 domains and biologically derived peptide sequences might be informative for building better algorithms that predict *bona fide* SH2 domain interaction sites within human protein sequences.

Although protein microarrays enabled the first systems-level glimpse at SH2 domain selectivity (15, 17), they had several limitations that resulted in reduced ability to identify low affinity interactions in comparison with solution phase methods (20). We therefore designed a high throughput fluorescence polarization approach that allowed for lower affinity interactions to be defined between SH2 domains and phosphopeptides of the ErbB family of receptor tyrosine kinases (RTKs) than was possible with protein microarrays (20).

RTKs are vital mediators of signal transduction in multicellular organisms. RTKs typically function as transmembrane receptors that contain a tyrosine kinase and other motifs that enable interaction with other intracellular proteins. Human cells often express many different RTK proteins from the set of 57 RTK genes encoded by the human genome (21). These RTKs may be activated in different combinations to transduce common and specific downstream signals (22). For a recent review of the complexity of RTK signaling networks, see Ref. 23. Following activation, RTKs are phosphorylated on several intracellular tyrosine residues that serve as recruitment sites for SH2 domains (15–18, 20). Activation of RTK signaling networks may cause changes in cellular motility, proliferation, survival, and cytoskeletal arrangement. Definition of their signaling capacity represents an important and unsolved problem in cell biology. Although most studies to date have focused on the role of singular RTKs in cancer progression, co-activation of RTKs derived from several unique RTK genes has recently emerged as an important driver of cancer progression (24–27). Co-activation of modules of RTKs may provide robustness against therapies designed to inhibit a single RTK (25).

Herein, we profiled the interaction potential of two RTKs and two signaling proteins and compared them with the recruitment potential of the ErbB family that we have previously profiled (28). The ErbB family, c-Met, and c-Kit RTKs have been shown to drive the progression of many cancer types, including breast, head and neck, lung (29), gastrointestinal, and stomach cancers (30). Downstream adaptor proteins often augment the signaling potential of RTKs by acting as scaffolds for recruitment of many additional proteins (31–33). Therefore, we also included peptides in our study derived from the Gab1 adaptor protein, which is critical for mediating

signaling networks downstream of c-Met and potentially other RTKs (34).

Finally, alternative oncogenic signaling networks may have points of cross-talk with tyrosine kinase signaling networks. Steroid hormone receptors such as the androgen receptor (AR) have been shown to associate with RTKs such as EGFR (35), to be substrates of tyrosine kinases (36, 37), and to drive the progression of prostate cancer (36). We therefore queried the interaction potential of phosphopeptides derived from AR with a set of 93 of the 120 SH2 domains encoded in the human genome. We subsequently used this interaction dataset to develop a permutation-based logistic regression classifier (PEBL) for predicting the interaction potential of SH2 domains and biologically derived phosphotyrosine-containing peptides.

MATERIALS AND METHODS

Reagents were produced and purified for use in automated high throughput fluorescence polarization assay as described previously (28).

SH2 and PTB Domain Proteins—The cloning of 109 SH2 and 44 PTB domains in the human genome is described previously (15). In this study, 93 SH2 and 2 PTB domain-containing constructs (supplemental Fig. S1 and supplemental Table S1) were selected that met each of the following criteria: 1) fraction of monomeric protein observed in a previous study following expression and purification $\geq 50\%$ by size exclusion chromatography; 2) previous evidence of functionality by PM as evidenced by interaction with one or more phosphopeptides with an apparent midpoint binding constant $K_D \leq 1 \mu\text{M}$. Where multiple SH2 domains were contained in a single gene, the tandem protein was included in our analysis with all internal amino acid residues linking the domains even if the percentage of monomeric tandem SH2 domains was less than 50%.

Peptide Synthesis and Purification—Peptides were synthesized and purified as described previously (28).

Fluorescence Polarization (FP) Saturation Binding Assay—The FP saturation binding assay was performed as described previously (see Fig. 1A) (28). Experimental values were output as millipolarization units and imported into MATLAB (The MathWorks, Inc., Natick, MA) in which Equation 1 was used to determine dissociation constants (K_D) for each protein/peptide pairing by least squares linear regression.

$$P_{\text{obs}} = \frac{P_{\text{max}} \cdot [\text{Protein}]}{K_D + [\text{Protein}]} \quad (\text{Eq. 1})$$

Protein and Gene Ontology Enrichment Analyses—The total number of phosphotyrosine (Tyr(P)) sites on each receptor or adaptor to which each SH2 domain-containing protein or gene ontology class bound was first determined. We then performed 10,000 permutations of the Tyr(P) sites to build a reference distribution of the null hypothesis of each receptor binding to a given protein or class of proteins at a random number of sites given the number of sites queried. We defined statistically significant enrichment and depletion of binding sites by identifying instances where the observed number of binding sites was unlikely to occur by chance given the number of sites bound across all receptors ($p < 0.05$).

Establishing Amino Acid Residue Location Importance in Predicting SH2 Domain Recruitment—The R package “randomForest” (65) was used to implement the random forest algorithm using 10,000 trees per run with two variables randomly sampled at each tree split. We examined the ability for all residue positions to predict binary binding

events for each SH2 domain. Variable selection was performed using the “varSelRF” package in R using 10,000 trees for the first forest, 300 trees for all additional forests, and excluding 20% of variables at each iteration.

PEBL Classifier—For each SH2 domain p , we randomly sampled q peptide sequences 100 times from the full set of 178 peptides used in our study, where q was the number of peptides that were determined to bind SH2 domain p by FP. From these permutations, for each amino acid residue at each site, we determined the relative statistical enrichment and depletion of that amino acid residue for each SH2 domain by comparing the observed amino acid frequencies to the permuted amino acid counts. We then \log_{10} -transformed these p values and inverted the depletion p values (such that -2 corresponds to depletion with $p = 0.01$) for enrichment heat maps.

We then built a PEBL to predict to which SH2 domains a given peptide would bind. For each peptide, we summed the transformed p values for each amino acid residue in the peptide for each SH2 domain (with depleted residues yielding negative values and enriched residues yielding positive values) to derive a prediction score for the likelihood that peptide would interact with an SH2 domain given its sequence. Accuracy, sensitivity, and specificity as a function of the PEBL score cutoff were calculated using the ROC package in R (72).

Evaluating PEBL on Consensus Motifs and an External Dataset—We first assembled a set of 62 consensus motifs from previous literature (13, 60, 61, 68, 69). SMALI and PEBL scores were calculated based on the highest scoring SMALI residues at each position relative to phosphotyrosine. We then acquired 1532 SPOT array measurements from 160 peptides from 14 different SH2 domains from Liu *et al.* (18). We then calculated SMALI and PEBL scores for each interaction based on the SH2 domain and the peptide queried. To calculate positive predictive values (the proportion of true positives over all positives called by the algorithm), we defined SMALI scores >1.0 and PEBL scores >4.5 as “interactions.”

RESULTS

We previously assessed the comprehensive SH2 domain recruitment profile for phosphopeptides derived from ErbB1/EGFR, ErbB2, ErbB3, and ErbB4 RTKs using a high throughput fluorescence polarization (FP) interaction analysis assay (supplemental Fig. S1) (28). This approach has previously been shown to yield a false-positive rate of 18.4% and a false-negative rate of 4% based on validation of a subset of random interactions by surface plasmon resonance (20). We determined that peptides containing 4 residues N-terminal and 8 residues C-terminal to the phosphotyrosine residue generated maximal polarization changes upon SH2 domain binding while retaining maximal binding selectivity. Therefore, we synthesized 13-mer phosphopeptides corresponding to 85 of 89 cytosolic tyrosine sites from the c-Met and c-Kit RTKs, the Gab1 adaptor protein, and the AR protein. We then tested these peptides for interaction with 93 SH2 domains and 2 phosphotyrosine binding (PTB) domains (Fig. 1 and supplemental Table S1) using high throughput FP.

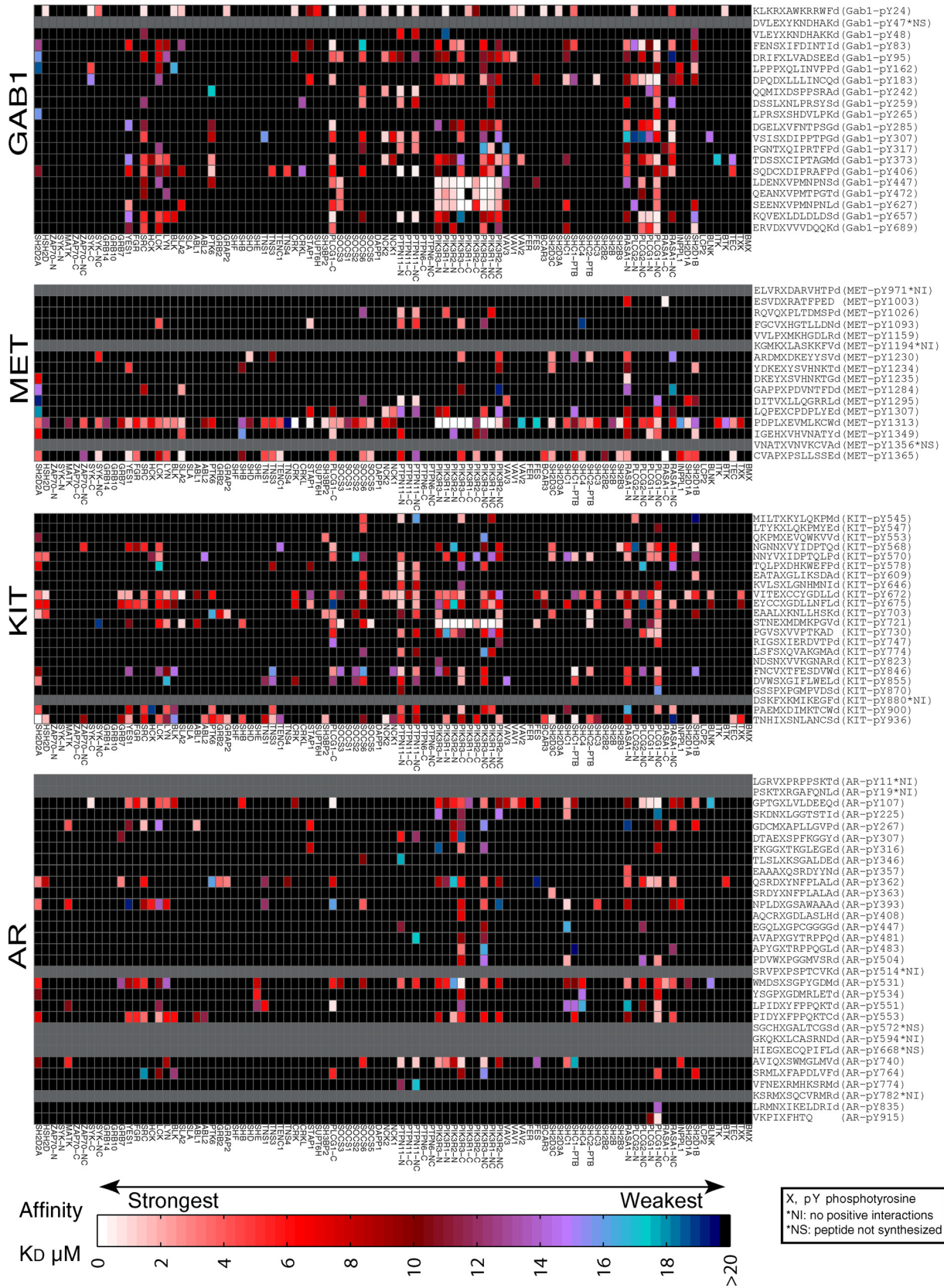
c-MET and Gab1—We synthesized phosphopeptides representing 15 of 16 intracellular tyrosine motifs on the oncogenic RTK c-Met and 19 of 20 tyrosines on the downstream adaptor protein GAB1 for interaction analysis via the FP assay (supplemental Table S2). The set of 15 c-Met phosphopeptides analyzed here represents a 2-fold increase in the num-

ber peptides queried for interaction *versus* previous PM-based interaction studies of c-Met recruitment with SH2 domains and thus offered the potential for novel insight (16). From 3115 unique queries, we identified 174 peptide-protein interaction pairs with c-Met and 310 interactions with GAB1 (Fig. 1).

We also compared our assay results to previously published PM data similarly to our prior ErbB study (Fig. 2) (16, 20). c-Met was the only available protein microarray interaction dataset orthogonal to the receptors examined in this study. Of the 176 FP-derived and 116 PM-derived interactions, 54 interactions were detected by both methodologies. As described previously, the FP method can identify a wider range of interaction affinities, whereas PMs are confined to only the strongest interactions (20). Based on previous estimates of the false-positive rate of PMs to be as high as 59% because of technical artifacts related to surface immobilization of proteins, disulfide bonding of SH2 domains, and incomplete peptide solubility (20), we conclude that many of the 62 missed interactions originally identified by PMs but missed by FP are likely false positives. Although peptides derived from c-Met Tyr-1313 and Tyr-1365 resulted in the majority of interactions, we detected 122 additional interactions that were either not previously queried or detected by protein microarrays (16). By contrast to the focused recruitment of SH2 domains to a small number of phosphosites within c-Met, SH2 domains were recruited to many phosphosites within GAB1.

Previous studies have demonstrated that c-Met Tyr(P)-1307/Tyr(P)-1313 can interact with SH2 domains from PIK3R proteins (38); we observed moderate recruitment of select PIK3R-derived SH2 domains by c-Met Tyr(P)-1307 but strong ($K_D < 1 \mu\text{M}$) recruitment of most PIK3R-derived SH2 domains at Tyr(P)-1313. The Tyr(P)-1349/Tyr(P)-1356 dual phosphorylation site has been previously shown to be important for recruitment of Grb2. Because we were unable to synthesize the peptide corresponding to Tyr(P)-1356, we were unable to empirically assess its recruitment potential. However, Tyr(P)-1356 contains the sequence pYVN and is therefore predicted based on previous studies and this study to recruit the GRB2 SH2 domain (10). We observed that Tyr(P)-1349 was able to recruit primarily SH2 domains from PLCG1 and SH2D1B with moderate affinity. c-Met Tyr(P)-1365 has been shown to be important for full enzymatic activity of the receptor and has been suggested to recruit downstream signaling mediators (40). We identified interactions with SH2 domains from several c-Src family kinases, tensin family, PLCG1, RASA1, SH2D1A, and SH2D1B proteins (Fig. 1), with this site. Although Tyr(P)-1313 had the ability to recruit most SH2 domains, it notably lacked the ability of Tyr(P)-1295, Tyr(P)-1307, and Tyr(P)-1365 to recruit the Shp2/PTPN11 SH2 domain.

GAB1 had six predicted recruitment sites for SH2 domains from CRK and CRKL based on SH2 domain interaction prediction algorithms: Tyr(P)-242, Tyr(P)-259, Tyr(P)-307, Tyr(P)-



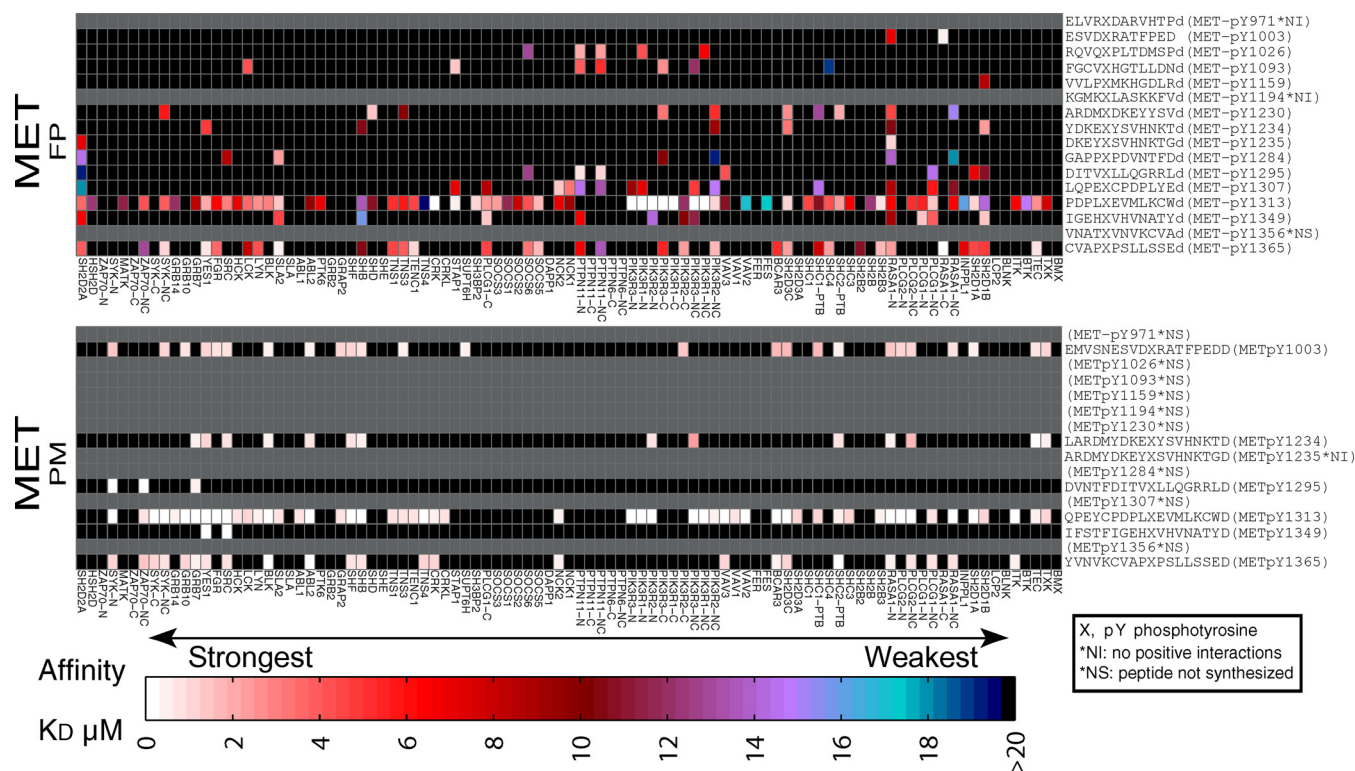


FIG. 2. Comparison of recruitment profiles for MET as determined by protein microarrays versus fluorescence polarization. Color-coded heat maps represent K_D values for FP interactions between SH2 and PTB domains and phosphopeptides representing all potential phosphotyrosine sites for which a peptide could be successfully synthesized in previously published protein microarray studies as well as this study. Black boxes indicate interactions that are too weak to be detected by the assay. Sequences of peptides used are indicated for each receptor site, where *d* denotes the pre-charged aspartic acid residue on the peptide synthesis resin and not a naturally occurring Asp. *NS* refers to peptides that were unable to be synthesized or, in the case of the protein microarray study, not queried at all. *NI* refers to synthesized peptides that produced no positive hits in the respective studies; therefore, we cannot confirm nor deny interactions at these sites with either assay. Rows of the heat maps for these hits peptides are *grayed out* to indicate that the protein microarray assay or our FP assay could neither confirm nor deny positive or negative interactions from these peptides.

373, and Tyr(P)-406 (41, 42). However, we identified no CRK SH2 domain interactions with these predicted recruitment sites (Fig. 1). We did observe CRK SH2 domain recruitment with peptides derived from Tyr(P)-24 and Tyr(P)-95 from GAB1. We also observed CRKL SH2 domain interactions with peptides derived from Tyr(P)-259 and Tyr(P)-406 but not with peptides derived from the other four previously suggested interaction sites. Based on prior interaction predictions, Tyr(P)-162 was suggested as an interaction site for the SHC1 SH2 domain (43), but we were unable to detect this interaction (Fig. 1). SHC1 SH2 domain interactions were identified with peptides derived from Tyr(P)-83, Tyr(P)-95, Tyr(P)-183, Tyr(P)-373, Tyr(P)-627, and Tyr(P)-689. SHC1 PTB domain interac-

tions were observed with peptides derived from Tyr(P)-24, Tyr(P)-83, and Tyr(P)-183, none of which represent the canonical NPXpY motif. Peptides derived from Tyr(P)-307 and Tyr(P)-317 were predicted to bind to SH2 domains from RASA1 (44) based on previous algorithms, and this prediction was confirmed by our analysis. FP analysis also confirmed the PLCG1 interactions that have been reported with Tyr(P)-307, Tyr(P)-373, and Tyr(P)-406 (45). Finally, Tyr(P)-447, Tyr(P)-472, and Tyr(P)-627 were previously demonstrated by mutational analysis to recruit PIK3R1, whereas Tyr(P)-657 and Tyr(P)-689 were shown to recruit PTPN11 (46). Our interaction analysis confirmed the three PIK3R1-binding sites and also found that these sites also recruited PIK3R2 and PIK3R3 with

Fig. 1. Recruitment profiles for GAB1, MET, KIT, and AR. Comprehensive SH2 domain recruitment potential of the adaptor protein GAB1, the MET and KIT receptor tyrosine kinases, and the human AR as determined by high throughput fluorescence polarization. Color-coded heat maps represent K_D values for FP interactions between SH2 and PTB domains and phosphopeptides representing all potential phosphotyrosine sites for which a peptide could be successfully synthesized. Black boxes indicate interactions that are too weak to be detected by the assay. Sequences of peptides used are indicated for each receptor site, where *d* denotes the pre-charged Asp residue on the peptide synthesis resin and not a naturally occurring Asp. *NS* refers to peptides that were unable to be synthesized. *NI* refers to synthesized peptides that produced no positive hits in the study; therefore, we cannot confirm nor deny interactions at these sites with our assay. Rows of the heat maps for these peptides are *grayed out* to indicate that our FP assay could neither confirm nor deny positive or negative interactions from these peptides.

high affinity. However, by contrast to previous reports, our analysis suggested that Tyr(P)-689 was unable to recruit PTPN11 but was able to recruit SH2 domains from PIK3R, PLCG1, SHC1, SH2D1B, and c-Src family tyrosine kinase family members.

We also identified previously unreported interaction sites between GAB1 phosphopeptides and SH2 domains derived from tensin family proteins, Vav family proteins, SH2D1B, and a subset of the SOCS proteins (47). Although SH2 domains from Vav and tensin family proteins interacted selectively with only a few phosphosites within GAB1, SH2 domains from SH2D1B, SOCS3, and SOCS6 displayed a recruitment pattern characterized by multiple redundant interaction sites, similarly as we observed between c-Met phosphopeptides and SH2 domains from PLCG1, PIK3R, and PTPN11.

c-Kit—c-Kit is an RTK with oncogenic potential in the PDGFR family (21, 30, 49). We successfully synthesized phosphopeptides representing all 22 potential intracellular tyrosine motifs (supplemental Table S2) of this receptor. From 2017 unique peptide-protein queries, we detected 307 interaction pairs (Fig. 1). We detected most literature-reported interactions between c-Kit and our SH2 domain set (49) and were able to infer the predicted respective binding sites because of c-Kit's homology with the PDGFR (50–52). Importantly, the peptide derived from Tyr(P)-721 (which represents the homologous PIK3R recruitment site shared by all PDGFR family members) was the only c-Kit peptide able to recruit all PIK3R SH2 domains with sub-micromolar midpoint dissociation constants. Peptides derived from Tyr(P)-703 and Tyr(P)-936 were the predicted GRB2/GRAP2-binding sites based on motif prediction software (10), and phosphopeptides derived from these sequences were the only two peptides that detectably recruited these domains. As expected from previous functional and interaction studies, peptides derived from Tyr(P)-568 and Tyr(P)-570 recruited several c-Src family kinase SH2 domains (49).

Although the peptide derived from Tyr(P)-721 displayed the highest affinity for PIK3R SH2 domains, we observed that many other c-Kit peptides not containing canonical PI3KR sequence motifs were also able to interact with these domains. These additional interaction sites in the PDGF receptor might allow for increased interaction avidity with PI3KR proteins because each PI3KR regulatory subunit has two SH2 domains. Similarly, we observed that domains from PTPN11, PLCG1, RASA1, and SOCS6 were recruited to many sites throughout c-Kit. Other protein families were recruited at a more limited number of c-Kit receptor phosphosites. For example, CRK was only recruited by peptides derived from Tyr(P)-672, Tyr(P)-675, and Tyr(P)-855. Domains from the tensin family proteins were recruited primarily by peptides derived from Tyr(P)-855 and Tyr(P)-936. In addition to interacting with peptides derived from Tyr(P)-568/Tyr(P)-570 sites, SH2 domains from the c-Src family tyrosine kinases were also recruited to peptides derived from Tyr(P)-672/Tyr(P)-675,

Tyr(P)-900, and Tyr(P)-936. As a hematopoietic RTK, it was noteworthy that several residues on c-Kit recruited the SH2 domain of SH2D1B, a signaling lymphocyte activation molecule member (53); peptides derived from Tyr(P)-568 and Tyr(P)-936 recruited SH2D1B with the highest affinity ($K_D \sim 1 \mu\text{M}$).

Androgen Receptor—The human AR is a type of nuclear receptor that is activated by the binding of testosterone or dihydrotestosterone (54). AR has 31 tyrosine residues, a subset of which has been shown by mass spectrometry studies to be phosphorylated by c-Src and other tyrosine kinases (36, 37, 55). AR has also been shown to associate with RTKs such as EGFR (35) and is an important therapeutic target in prostate cancer (56–58). AR activity may also be modulated by phosphorylation and other forms of post-translational modification. The Ack1 kinase has been shown to phosphorylate AR at Tyr-267 and Tyr-363 (36), whereas Tyr-534 has been shown to be a substrate for c-Src (37). The phosphorylation of AR is elevated in some forms of hormone refractory prostate cancer relative to hormone-sensitive cancer (37). The modulation of AR activity by tyrosine phosphorylation is thought to occur, in part, through conformational changes in protein structure. However, AR tyrosine phosphorylation may also modulate its ability to interact with the SH2 domains of cell signaling molecules.

To test the ability of AR to recruit downstream signaling proteins, we synthesized 13-mer phosphopeptides corresponding to 29 of the 31 AR tyrosine motifs (supplemental Table S2). From 2708 unique peptide-protein queries, we identified 215 unique interaction pairs (Fig. 1). The majority of the interactions had relatively low affinities ($K_D > 10 \mu\text{M}$) in comparison with receptor tyrosine kinase-mediated interactions. Our assay detected multiple interactions with the peptide derived from Tyr(P)-267, including a relatively high affinity interaction with the SRC SH2 domain ($K_D = 1.85 \mu\text{M}$) and weaker interactions with SH2 domains from other c-Src family kinase members, including YES1 and LCK. However, the peptide derived from Tyr(P)-363 interacted primarily with SH2D3C ($K_D = 2.28 \mu\text{M}$). The peptide derived from Tyr(P)-534, a known c-Src kinase substrate, did not recruit SH2 domains from c-Src family kinases but did recruit PIK3R3 (C-terminal domain) and PLCG1 (NC tandem domain). Tyr-362 has been shown to be phosphorylated independently and in tandem with Tyr-363 (37). The peptide derived from Tyr(P)-362 interacted with SH2 domains from SRC, RASA1, PLCG1, and the PI3KR phosphatidylinositol kinase regulatory subunits. Although the peptide derived from Tyr(P)-534 interacted with few SH2 domains, the peptide derived from Tyr(P)-531 recruited 29 domains, including several from the c-Src family kinases. AR recruited SH2 domains from PI3KR domains with relatively high affinity at several tyrosines, including peptides derived from Tyr(P)-107, Tyr(P)-362, Tyr(P)-531, Tyr(P)-553, and Tyr(P)-740. Similarly, recruitment sites for domains from the phospholipases PLCG1 and PLCG2 were distributed

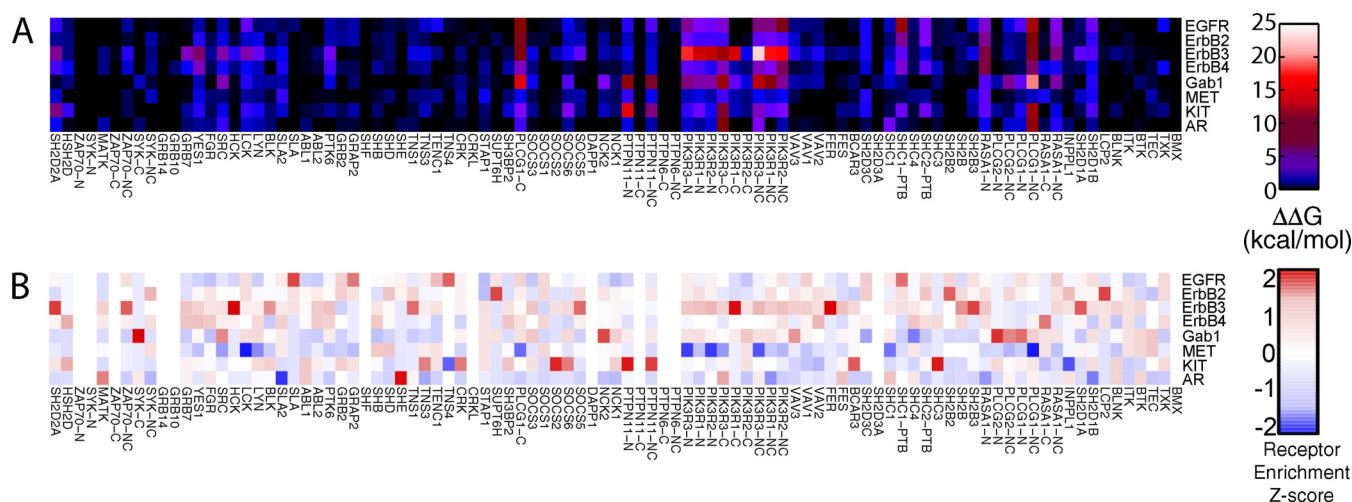


FIG. 3. **Relative binding energy and site enrichment for SH2 and PTB domains by signaling proteins.** *A*, relative binding free energies of interactions described by FP were summed across each receptor or adaptor protein. *B*, each receptor or adaptor was assessed for enrichment or depletion of binding sites for a given SH2 or PTB domain and is depicted by Z-score transformation of raw data.

throughout the length of the protein, including peptides derived from Tyr(P)-107, Tyr(P)-362, Tyr(P)-551, Tyr(P)-553, and Tyr(P)-915. The peptide derived from Tyr(P)-107 was the only peptide able to recruit VAV family SH2 domains with detectable affinities ($K_D \sim 2\text{--}8 \mu\text{M}$). Peptides derived from Tyr(P)-307, Tyr(P)-531, Tyr(P)-740, and Tyr(P)-774 recruited the PTPN11 protein-tyrosine phosphatase SH2 domain ($K_D \sim 2\text{--}8 \mu\text{M}$). The ability of AR-derived peptides to recruit PIK3R and PLCG1 domains is of particular interest because of the roles of these proteins in tumor processes such as cell survival, cell proliferation, and metastasis.

Comparison of Overall SH2 Domain Recruitment Capacity of Signaling Proteins—We compared the overall SH2 domain recruitment potential of GAB1, c-Met, c-Kit, and AR with the ErbB interaction dataset that we previously described (supplemental Fig. S2) (28). We observed that every receptor had the potential to recruit most SH2 domains. However, they did so with different overall binding energies (Fig. 3A). We found that AR was enriched for SH2 domain-containing adaptor protein E-binding sites relative to the other receptors ($p = 0.0377$) but was depleted for sites that recruited most other domains (Fig. 3B). ErbB1, ErbB2, and ErbB3 were significantly enriched in phosphosites that recruited multiple domains and displayed no significant depletion for phosphosites that recruited any domain. ErbB4 had no significant enrichment for phosphosites recruiting any SH2 domains but was significantly depleted for phosphosites recruiting SOCS6, PTPN11, and PIK3R3 ($p < 0.05$). GAB1 was significantly enriched for PLCG1- and PIK3R1-binding sites but was depleted for ZAP70, SH2D3C, and the SHC2-PTB domain binding sites. c-Kit was significantly enriched for SOCS2-, SOCS6-, and PTPN11-binding sites. Notably, PTPN11 was one of the first SH2 domain-containing proteins found to be recruited to c-Kit (59). c-Met was significantly enriched for SHB-binding sites, but despite containing a few high affinity binding sites for

PIK3R1, PIK3R3, and PLCG1, it was depleted in total binding sites for these SH2 domains *versus* the other proteins that we examined. Also of note was that GAB1 was enriched in binding sites for which c-Met was depleted, underscoring the likely importance of GAB1 in complementing c-Met's recruitment ability.

Recruitment of Molecular Functions—We clustered the SH2/PTB domains used in our FP assay into groups based on functional ontologies (supplemental Table S3) such as phospholipase, phosphatidylinositol kinase, scaffolds, etc., and we compared the relative ability of each RTK and signaling protein to recruit them (Fig. 4A, supplemental Fig. S3, and supplemental Table S4) (28).

c-Met had the lowest binding free energy for recruitment of SH2 domains from most ontological classes. ErbB3 was the most efficient at recruitment of the phosphatidylinositol kinase ontology, followed by the GAB1 adaptor protein. GAB1 was nearly twice as efficient at recruitment of phospholipases compared with the other receptors. c-Met typically signals in tandem with GAB1, and the GAB1/c-Met module would display a binding free energy for phosphatidylinositol kinases similar to ErbB3. This GAB1/c-Met module would also display twice the binding free energy for SH2 domains from the adaptor ontology as every other RTK, and it would have similar binding free energies as the other RTKs for all other ontologies.

Surprisingly, AR also recruited SH2 domains from the phospholipase ontology at a similar overall binding free energy as ErbB1 and with a higher free energy than the other RTKs. Whereas AR lacked efficient recruitment potential for the adaptor ontology, c-Kit lacked recruitment potential for SH2 domains from the phosphatidylinositol phosphatase and Rho GEF ontologies.

We then tested for enrichment of binding sites that each receptor displayed for SH2 domains from each ontological

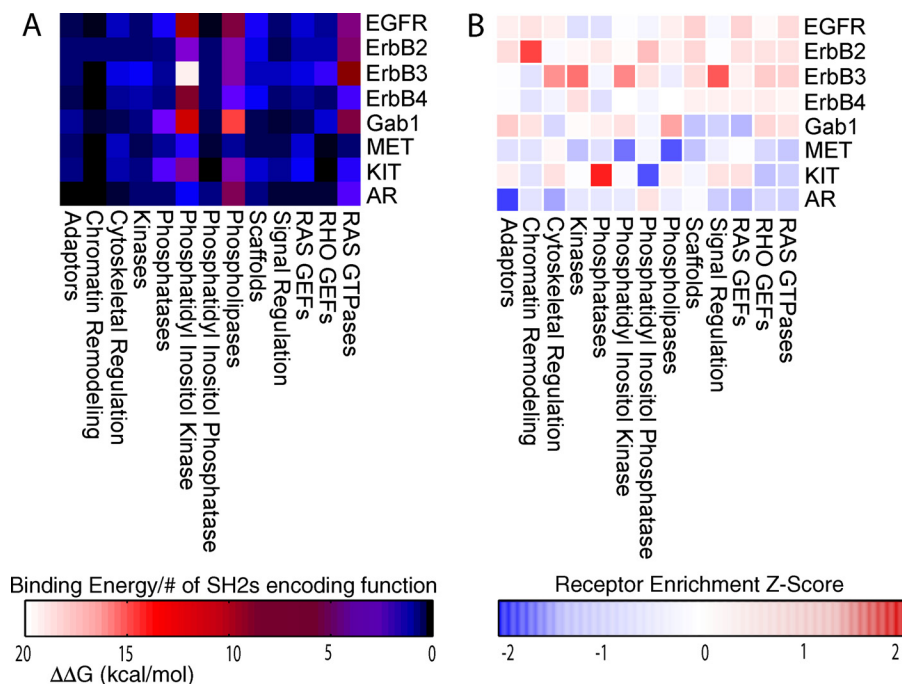
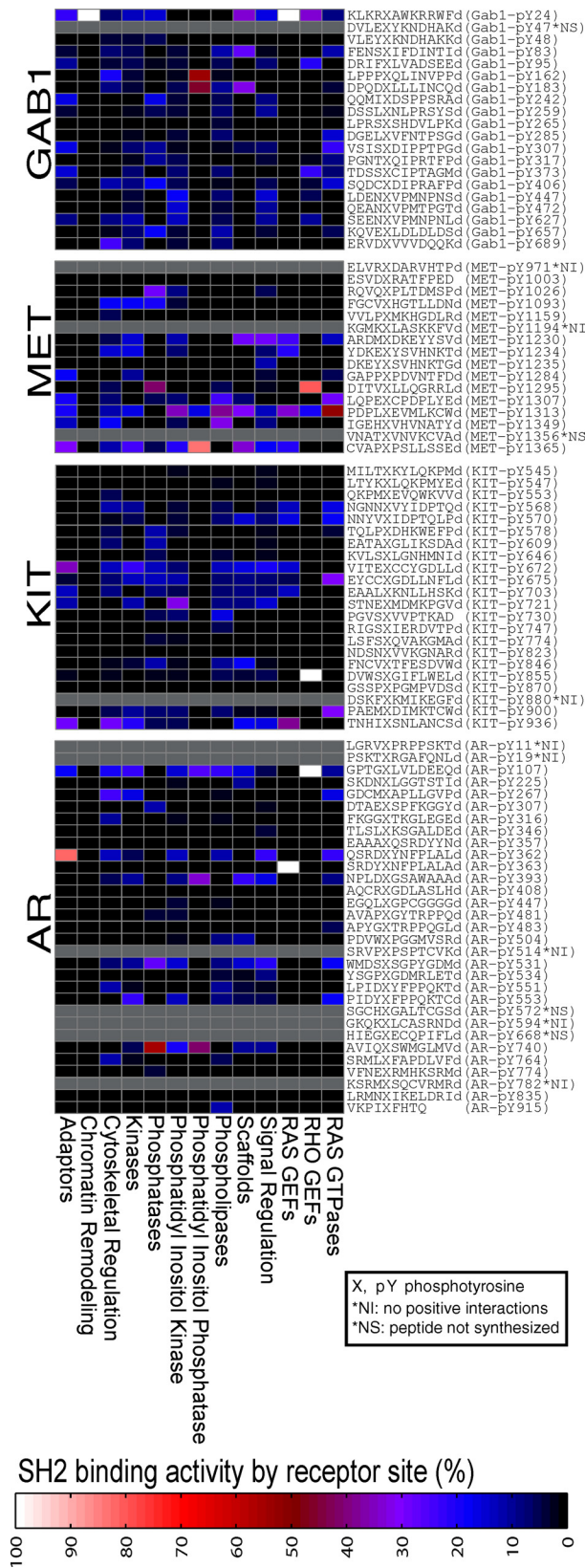


FIG. 4. Comparison of the recruitment of proteins representing different molecular function categories. *A*, relative binding free energies of interactions described by FP for the ErbB family, GAB1, MET, KIT, and AR were summed across all domains in each listed ontology and then divided by the number of domains in that ontology to determine an average recruitment potential for a particular molecular function group. *B*, each receptor or adaptor was assessed for enrichment or depletion of binding sites for a given ontology. Data are depicted by Z-score transforming the observed number of binding sites each receptor/adaptor had for a particular ontology relative to the average number of sites that bound the ontology across all receptors/adaptors assessed in our FP assay.

class (Fig. 4B and supplemental Table S5). AR was significantly depleted for binding sites for SH2 domains from most ontologies. ErbB1 and ErbB2 were enriched for binding sites for phospholipase, Ras GTPase, and scaffold ontologies. ErbB2 and ErbB3 were enriched for binding sites of SH2 domains from the kinase and phosphatidylinositol kinase ontologies. ErbB3 also displayed enrichment in binding sites for the cytoskeletal regulation and signal regulation ontologies. ErbB4 was depleted of binding sites for phosphatase and phosphatidylinositol kinase ontologies. c-Kit was enriched for binding sites to the phosphatase ontology but was depleted for binding sites to the Rho GEF ontology. GAB1 was enriched for binding sites for the phospholipase and phosphatidylinositol kinase ontology (Fig. 4B). Although c-Met was depleted in binding sites for SH2 domains of those ontologies, it was enriched for binding sites to SH2 domains from the cytoskeletal regulation ontology.

The FP-derived interaction matrix revealed that many SH2 domains were able to be recruited in a redundant manner by peptides derived from many phosphotyrosine sites within each protein. Therefore, we next asked how the overall ontological recruitment capacity of the RTKs, adaptor and AR protein, was distributed across each protein. For this purpose, we examined the ontological recruitment within each protein as a percentage of overall recruitment capacity for all phosphopeptides contained within each protein (Fig. 5 and

supplemental Table S6). Although domains from most ontological classes were recruited in a relatively even manner across each protein, a subset was recruited to a relatively small number of phosphosites. For example, our results suggested that a Tyr to Phe mutation of AR Tyr-363 would result in a substantial reduction in its ability to recruit the Ras GEF ontology without a substantial reduction in its ability to recruit domains from other ontologies. Similarly, a Tyr to Phe mutation of AR Tyr-107 would be predicted to result in a significant reduction in the ability of AR to recruit SH2 domains from the Rho GEF ontology but would be expected to have minimal effects on the recruitment of domains from other ontologies. The c-Kit receptor contained many redundant recruitment sites for a diverse set of SH2 domains, but a Tyr to Phe mutation of Tyr-855 would likely result in a complete loss in its ability to recruit SH2 domains from the Rho GEF ontology. c-Met contained no phosphosites that were exclusively responsible for the recruitment of domains from a particular ontology. However, a Tyr to Phe mutation of c-Met Tyr-1365 would be expected to result in a major reduction in its ability to recruit SH2 domains from the phosphatidylinositol phosphatase ontology, whereas mutation of c-Met Tyr-1295 would be expected to result in a major reduction in its ability to recruit SH2 domains from the Rho GEF ontology. GAB1 Tyr-24 was noteworthy in that its mutation would be expected to result in a nearly complete loss in recruitment of SH2



domains from proteins in the chromatin remodeling and Ras GEF ontologies.

SH2 Domain Binding Motif Analysis Based on FP-derived Data—Previously defined consensus binding motifs for SH2 domains are based on competitive assays between SH2 domains and oriented degenerate peptide arrays and/or random peptide libraries (10, 13, 60). Studies have also been performed to infer domain binding preferences using structural information (62). A recent study has built upon these analyses by using biologically derived peptide sequences from insulin receptor (IR), insulin-like growth factor receptor (IGF1R), and fibroblast growth factor receptor (FGFR) to generate biologically derived consensus binding motifs (18). We asked whether we could improve upon prior studies using the interaction data derived from our FP study, which is noncompetitive in nature and seeks to quantify interaction affinities rather than to identify a sequence representing a “perfect motif.”

We first determined the relative statistical enrichment and depletion of each residue at each position from -4 to $+7$ relative to phosphotyrosine for each SH2 domain. For each protein (representing one SH2 domain) p , we randomly sampled q peptide sequences 100 times from the full set of 178 peptides, where q was the number of peptides originally bound by each SH2 domain protein p . From these permutations, we compared the observed residue counts to the permuted residue counts for all peptides that bound to each SH2 domain protein p . For instance, if out of 100 permutations we did not observe an instance where the permuted proportion of peptides with prolines at site $+1$ exceeded the observed number of interacting amino acid residues at site $+1$ with an SH2 protein, we would conclude that a statistical enrichment existed for proline at site $+1$ at $p < 0.01$ (1/the total number of permutations). We then \log_{10} -transformed these p values and inverted the depletion p values (such that -2 corresponds to depletion with $p = 0.01$) (see results for all SH2 domains in supplemental Fig. S4). We next compared the motifs identified in our analysis with those previously defined for the subsets of SH2 domains previously examined by the SMALI and Scansite prediction algorithms.

Derived consensus motifs for only 10 of the 95 SH2 domains in our assay shared two or more residues of homology with consensus motifs previously described in the literature (supplemental Table S7). For example, the previous consensus motif for GRB2 and GRAP2 based on the SMALI predic-

Fig. 5. Phosphosite contribution on GAB1, MET, KIT, and AR for the recruitment of molecular function groups. Relative binding free energies were calculated for each phosphosite and individual protein domain and then summed according to the classification of the domain in each molecular functional ontology group. The binding energy for the ontology at each site was subsequently divided by the total binding energy summed across the entire receptor or adaptor protein and presented as a percentage of total binding activity of that receptor or adaptor.

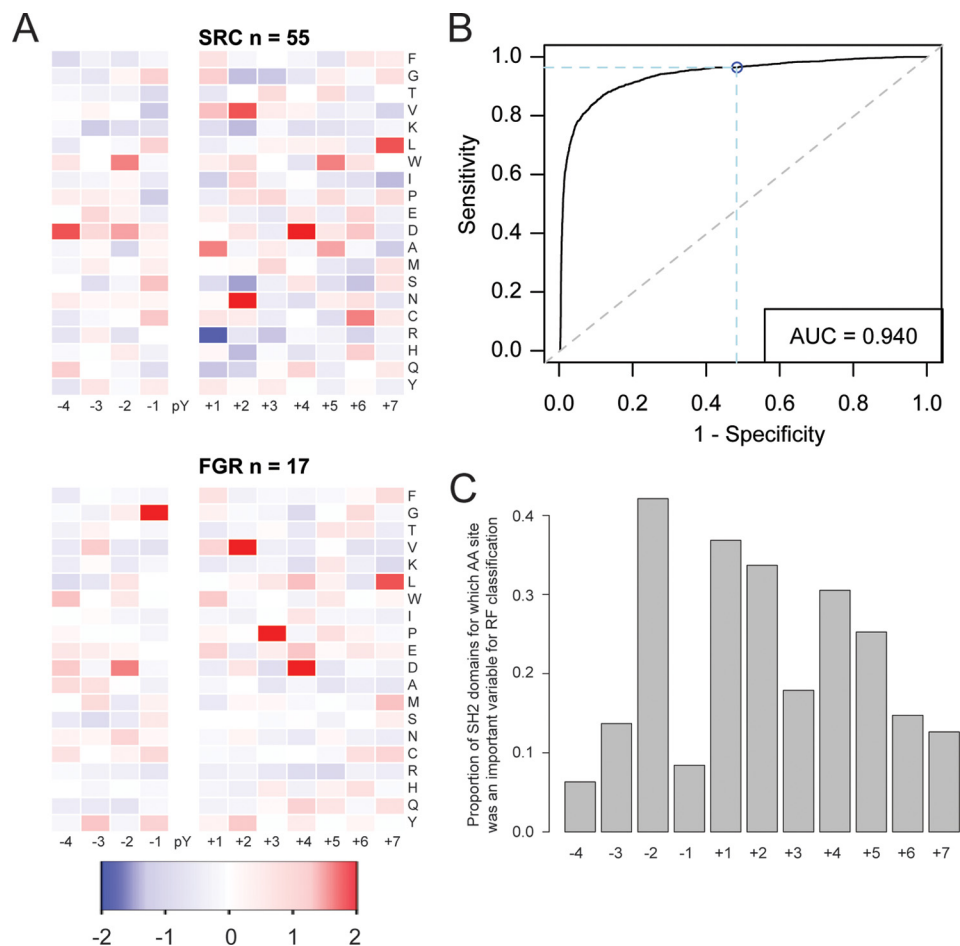


FIG. 6. Residue enrichment at positions relative to tyrosine that contribute to the recruitment of SH2 domains. *A*, representative plots depict \log_{10} transforms of the p values of each residue at each position relative to 100 permutations. The number of peptides that interacted with each domain is indicated at the top of each plot. *B*, ROC curve obtained by plotting the false-positive rate ($1 - \text{Specificity}$, x axis) by the true-positive rate (Sensitivity , y axis) for PEBL predictions of our FP interaction data. *C*, histogram depicting the number of SH2 domains for which each position relative to phosphotyrosine was selected as an important variable for binary interaction classification by random forests.

tion algorithm and random peptide library data were VXpYVNM and PPpYVNEL respectively, where pY represents the location of the phosphorylated tyrosine as a point of reference and X represents no enrichment or depletion for that amino acid residue. Similarly to the previous predictions, our analysis determined the following consensus motifs from our FP-derived data for GRB2 and GRAP2, respectively: PXpYXN and PXpYXNXXWT. However, the consensus motifs for most SH2 domains determined from our quantitative FP-derived interaction data were different from those previously determined from data derived from highest affinity interactions (Fig. 6A, supplemental Fig. S4, and supplemental Table S7). For example, in comparison with the previous consensus motif from SMALI for SRC SH2 of PIpYELID, our analysis suggested that the SRC SH2 domain was enriched for binding peptides with Asp at (-4, -2, and +4), Val and Asn at +2 but was depleted for binding peptides with Arg at +1. In comparison with the pYENL consensus ascribed to the SH2 domain of FGR using oriented peptide array libraries (13, 14), our

analysis confirmed an enrichment for Asp at +4 and also identified significant novel enrichments for Gly at -1, Val at +2, Pro at +3, and Leu at +7, implicating these amino acids as positive contributors to FGR SH2 affinity.

We subsequently built a PEBL regression classifier to predict the likelihood that a particular peptide sequence would result in an interaction with a particular SH2 domain. For each peptide, we summed the \log_{10} -transformed p values of enrichment (positive values) or depletion (negative values) for each residue in the peptide to derive an interaction prediction score for each SH2. To assess how well our logistic regression classifier categorized interactions from noninteractions in our experimental FP-determined data, we constructed a receiver operating characteristic (ROC) curve to examine the relationship between true-positive rate and false-positive rate as a function of the prediction score (Fig. 6B). Our classifier was highly accurate, with an ROC area under the curve (64) of 0.94. We achieved a maximum accuracy of 94% when defining interactions with a prediction score >4.50 as “interactions” and

<4.50 as “noninteractions” (supplemental Fig. S5). At this binary classifier threshold, although our model only achieved a sensitivity of 67.5%, it achieved a specificity of 97.8%.

To examine which residue positions relative to phosphotyrosine contributed the most in determining the probability of binding to each SH2 domain, we implemented the random forests (RF) algorithm to construct a predictive classifier and assess the proportion of variance explained by each site in the sequence for each SH2 domain (65). The random forests algorithm is a machine-learning technique that utilizes an ensemble of independent decision trees to perform classification or regression by building trees from sampling random subsets of all available variables (66). Our model had a median prediction accuracy of 91.0%, a median specificity of 97.8%, and a median sensitivity of 22.0% (supplemental Table S8) across all SH2 domains examined. Notably, many SH2 domains within our sample set contained few binders for which to inform our classifier. The RF model was better at predicting negative interactions than positive ones, except for peptides with higher prediction scores. However, the model had high sensitivity for many SH2 domains, such as PIK3R3.NC (100%), PLCG1.NC (98.9%), and RASA1.N (98.8%). Consistent with what has previously been shown, the -1 , $+1$, $+2$, $+4$, and $+5$ amino acid residues were the most informative variables for RF classification over all of the SH2 domains examined (Fig. 5C).

We compared the PEBL score domain interaction predictions to SMALI and Scansite score predictions for several SH2 domains (12, 13). We first identified peptide sequences representing the highest scoring residues at each position relative to phosphotyrosine (13, 14, 60, 68, 69) based on SMALI (supplemental Table S7). We then used this peptide set to calculate Scansite, SMALI, and PEBL scores for each SH2 domain. Among the 64 SH2 domains for which SMALI consensus motifs were available, 51 consensus sequences resulted in a positive PEBL score as compared with only 9 being predicted as a positive interaction by the Scansite algorithm. We observed a suggestive but not significant correlation between PEBL and SMALI scores (supplemental Fig. S6).

The PEBL analysis suggested that the probability of peptide interaction with RASA1-N would be increased if an aspartic or glutamic acid residue existed at the -1 position and if an isoleucine existed at the $+2$ position (Table I). The proline at the -1 position was well represented in our FP assay but was associated with peptides that were unable to bind RASA1-N. Although both Scansite and SMALI provide scoring metrics indicative of positive contributions of sequences to interaction probability, neither provides an assessment of the negative contributions of amino acids to probabilities of domain interaction. PEBL was able to circumvent this limitation by outputting scores that incorporated both positive and negative contributions by amino acid residues to determine whether a peptide would bind or not bind to a specific SH2 domain. The only significantly negative PEBL score obtained from peptide

sequences representing perfect SMALI consensus motifs was with the RASA1 N-terminal SH2 domain. The consensus motif suggested by SMALI (XPPpYTEM) had a PEBL score of -4.41 . A major source of this discrepancy stems from the observation that proline at the -1 , threonine at the $+1$, and glutamic acid at the $+2$ positions are suggested by PEBL to greatly reduce the probability of interaction with the RASA1-N SH2 domain (Table II).

We next sought to externally validate PEBL's performance and compare it with SMALI for predicting interaction events between SH2 domains and biologically derived peptides from an independent dataset that utilized a SPOT-array approach (supplemental Table S9) (18). In this dataset, 192 phosphopeptides were synthesized directly onto a support membrane and tested for interaction with 50 SH2 domains. The 11-mer peptides consisted of four N-terminal and six C-terminal residues to phosphotyrosine, which is within the parameters of PEBL based on the 13-mer peptides that we assayed via FP. The intensities of positive binders on the SPOT array were better correlated for PEBL scores than for SMALI scores (supplemental Fig. S7), although PEBL and SMALI scores were also significantly correlated ($\sigma = 0.17$, $p = 0.001$) in the predictions for this particular peptide by SH2 domain set. In total, the SPOT array dataset and our FP dataset had 46 overlapping SH2 domains. The 160 peptides that produced detectable SPOT intensities were derived from 13 proteins known to participate in insulin receptor, IGF1R, and FGFR signaling networks. 1275 interactions were detected, reflecting a signal intensity above the mean on the SPOT array (18). However, when we examined these interactions with SMALI, only 132 of the 1275 interactions (10.4%) had a score above the 1.00 SMALI cutoff. Therefore, 1143 of the SPOT array interactions were not predicted as positives from the SMALI analysis, indicating that SMALI had a sensitivity of only 10.4%.

When we analyzed the SPOT array peptides for interaction via PEBL analysis, we observed that 781 of the interactions identified by the SPOT arrays (61.3%) had a PEBL score greater than 0. These results were suggestive that PEBL was nearly six times more sensitive than SMALI in predicting interactions from an independent SPOT array dataset comprising biologically derived phosphopeptide sequences.

Validation of in Vivo Relevance of FP Interactions and PEBL Predictions by Phosphopeptide Pulldowns—To assess the ability of PEBL to accurately predict interactions identified in cells, we leveraged previously published data that examined the ability of synthetic phosphopeptides to pull down SH2 domain-containing proteins from HeLa cell extracts (supplemental Table S10) (70). From the 27 overlapping SH2 domain interactions identified in cells, 13 interactions (48%) were ranked highly by artificial neural network predictors (Z score >2) (70), 18 interactions (67%) had SMALI scores >1 , and 21 interactions had PEBL scores >0 (78%). Taken together, these results indicate that the PEBL algorithm trained with quantitative binding data outperformed the position weight

TABLE I
Amino acid residues that were statistically enriched at each position for each SH2 domain based on permutations ($p < 0.05$)

Protein	-4	-3	-2	-1	pTyr	+1	+2	+3	+4	+5	+6	+7
SH2D2A		V	W	P		G	N,H			K,W,D	E,C	L,W
HSH2D		N	P	E			N			N		
ZAP70.N												
SYK.N												
MATK	L		I	Q		E		M	W			
ZAP70.C		V	T	Q								
ZAP70.NC		V	W				V	T,M	V	W	E,C	L,W
SYK.C		P					L				W	Q
SYK.NC	F	G	T			K	W		L,W	I	P	
GRB14												
GRB10				K					I			
GRB7			P	E		E	N,Y	F		N		L,R
YES1	D	E,N	N	E				L	Q	D	C	
FGR			D	G			V	P	D			L
SRC	D		W,D			A	V,N		D	W,A	C	L
HCK						W,E	D			T	H	
LCK	D		D,N	V		A	V,I	L,P	W,D,Q		C	P
LYN	C			G			V	M	L,Q		C	
BLK	D	V	W			F,G		L	D	W	E,C	L
SLA2	I	E	N			W,E	W				P	
SLA	D,C	V	W	S		G	V	T	V	W	E	L
ABL1		T,I		G		A,M	P			T		
ABL2	P		H,Y			M	N					
PTK6	D	D,A,N	E			E	I,N	P	Q	T	P	
GRB2			P,R				N					
GRAP2			P				N				W	T
SHF												
SHB		N		E			V,N		Q	N,Q	C	
SHD	F		N								P	E
SHE	T,W	I	D	V		M	I	M	V		C	T,W
TNS1	C	V	L	D		Y			L,H			L
TNS3	C		L	M		D	I	E	W		C	W
TENC1	C	V	W			G	V,L	T	V,E	W	E,C	
TNS4	D	V	W	S		G	V	T	V	W	E	L

matrix and artificial neural network-based algorithms in predicting peptide-SH2 domain interactions from cells.

DISCUSSION

In this study, we used an automated high throughput fluorescence polarization assay to measure the interaction poten-

tials of 93 of the 120 SH2 domains in the human genome (71) with phosphopeptides derived from several RTKs, the GAB1 adaptor protein, and the androgen receptor. The FP-derived interaction data (28) uncovered interactions that have been previously identified in biological systems and a wealth of novel ones. The data allowed us to determine not just that

TABLE I—continued

CRK		L				G	V,C		L		W,E,H	L
CRKL				C					R			
BRDG1	F		C	D		W,A,C	W		L,I	L,N	F	P
SUPT6H	T	I	D	V		M	I	M	V	K	W,C	W
SH3BP2	A	N	P			E	N			H		
PLCG1.C	D			E,S		V,L	V	P,E	D	Q	E	L
SOCS3	S	E		G,N		E	P	M		P,D	G,N	W
SOCS1												
SOCS2			Y	V,L			D	L	L			W
SOCS6	L	V,H	T,L,W	S		G	L			W,M,S	E,D	L
SOCS5		E	N			E,M		M		K	C	W,E
DAPP1												
NCK2		N		E		C		A				
NCK1		D	I	E		C	T			D		
PTPN11.N	T	G,Q	T	V,Q		K	L,I	L	W			E
PTPN11.C		I				C	C	Y		D		
PTPN11.NC		V,I,Q	T,W	V,S,Q		G,K,D	G,I	T,L	K,W	W,S		L,E
PTPN6.C												
PTPN6.NC	T	I	D	V		M	I	M	V	K	C	W
PIK3R3.N		D,H	E,D	G,D,N		T,V,M,C	P	M	S	P,D		W,Q
PIK3R1.N		D	E	G		V,M,C	W,N	M		K,Q	C	
PIK3R2.N		E,D	E,D	G,E		V,M	P	M	I,S	K,P	G	W,Q
PIK3R3.C	F,A	V	W,D	G,S		G,T	I,P,D	L,M		W,Q	C	L,W
PIK3R1.C	P,S	E	E	G		E,M	P,Y	M	N	P,R	N	Q
PIK3R2.C	S	E		G,N		E	P	M	N	P	N	
PIK3R3.NC	P,D		E	G,D		V,E	V,I,P	M	Q		G,C	L,Q
PIK3R1.NC	P		E			V,E,M	V,I	P,M	V		C	L
PIK3R2.NC		N	P,D	E,D		V,M	N	L,M	Q	P,N,Q	C	
VAV3		E,D		G		V,E	I	P				
VAV1		H		G		L			D	T		Q
VAV2		A		G		L,A	T,V	P	Q	Q	G	F
FER											L	
FES				G		E	W	M			G	Q
BCAR3			H					L,Y			W	L
SH2D3C	D		W,D				N		V	W	E,C	
SH2D3A						A	W	K			W	
SHC1		P	W	S		G	V		W,I	W	E,H,Q	L
SHC1.PTB	D	V,N	T,W,P	V		W	W	E	I	W,D	P,E,C	L,W
SHC4	C	L				Y	C		L	D	D	N
SHC2.PTB	D	N	W,P			W	T	T	F	W	P	
SHC3			T,H			G			W,I	N	C	

TABLE I—continued

SH2B2	D	V	T,W	F,S		G,R	V	T,L	V	V,W	E	L,E
SH2B		E,D		E		E	D,Y	M				
LNK		V,H	W,N	G		T	V			W,D	E	L,W,Q
RASA1.N	D	D,A,N	W,E,D,N	E,D			I	L,P		W	C	
PLCG2.N	M			R					I,E		W	M
PLCG2.NC	D		W,D	S		G,V	F,V		I,D	W,D	E,C	L,W
PLCG1.N		P	W,P	S		V,L	V	T	I	W	E	Q
PLCG1.NC	D		W,P	S		V,L	V,I	P,E,M	D,Q	V,W	E	L
RASA1.C			D					P		T		
RASA1.NC	D		D	G			I	P	Q		F,D,C	W
INPPL1		V,P	W			G	V,L	T	V,I	W	E,C	W,Q
SH2D1A	G	V	T	V			G	I		I	S	L,E
SH2D1B	F,I		T,W	V		K	V,L	V,L,I	I	W,D,N	E	L,W
LCP2						W					P	
BLNK	W	Y	L				D		D		D	
ITK									E	L		M
BTK						Y	W	F	E		W	
TEC	T						I				C	
TXK	T	V,I	W,D	S		G,M	V	T,Y	V	K,W	E,C	L,W
BMX												

an interaction had occurred but also the strength of that interaction. We present these findings as a resource of interaction potentials that can be used to guide future biological inquiry.

We included the oncogenic receptor tyrosine kinases c-Met (73, 74) and c-Kit (30), the androgen receptor (57), and the adaptor protein Gab1, which functions downstream of many RTKs, including Egfr, c-Met, and c-Kit (34, 46, 49), in this interaction study to expand upon the recent analysis that we undertook regarding ErbB RTK interactions with SH2 domains (28). This dataset represents, to our knowledge, the largest existing interaction dataset comprising quantitative mid-point dissociation constants of SH2 domains with phosphopeptides. As reported previously for the ErbB receptors (28), many interaction affinities were relatively weak ($K_D > 2 \mu\text{M}$). These low affinity interactions may represent transient signaling events that would not have been easily observed using traditional in-cell interaction methodologies (75) and are likely of biological relevance (76). The protein interactions from this publication have been submitted to the IMEx consortium through IntAct (77) and assigned the identifier IM-22269.

By organizing our sets of domains into gene ontology groups, we assessed the relative degree that each signaling protein contributed to the recruitment of several cellular functions. As expected, ErbB3 and Gab1 were the most efficient at

recruitment of domains from phosphatidylinositol kinase regulatory subunits. Given the dependence of c-MET on Gab1 for signaling, it was not surprising to observe that c-Met displayed the least overall efficiency among the RTKs for recruitment of most ontologies. AR is commonly known as a transcriptional activator, but recent studies have identified sites of tyrosine phosphorylation that may influence its function (36, 37, 55). We expected few significant binding partners for AR based on its function as a steroid hormone receptor and were surprised that many domains were recruited to AR phosphotyrosine sites, including phospholipases and RASA1. These results are particularly notable given the reported association between AR and EGFR (35).

We observed that the signaling proteins recruited molecular functions at different efficiencies, both at the level of total binding affinities and numbers of recruitment sites. For example, although every receptor or adaptor recruited phosphatases, c-Kit and Gab1 recruited this ontology with twice the relative binding free energy as any other receptor, potentially resulting in more rapid dephosphorylation and down-regulation of tyrosine-phosphorylated signaling molecules. These analyses allowed us to discern the unique abilities of each receptor or adaptor to modulate signaling networks and, by extension, why different cell or tumor types may display disparate phenotypes despite employing the same core sets of cell-signaling proteins.

TABLE II

Amino acid residues that were statistically depleted at each position for each SH2 domain based on permutations ($p < 0.05$)

Protein	-4	-3	-2	-1	pTyr	+1	+2	+3	+4	+5	+6	+7
SH2D2A	K			G,N		V,L	K,E,M	G,P	D	P	G	
HSH2D												
ZAP70.N												
SYK.N												
MATK												
ZAP70.C												
ZAP70.NC												
SYK.C												
SYK.NC												
GRB14												
GRB10												
GRB7												
YES1	L	S						R	V			D
FGR												
SRC						R	S					
HCK						L						
LCK		S	K,R				E	R	H	R	S	A
LYN		G,S	P					R	V	R		
BLK												
SLA2												
SLA												
ABL1												
ABL2												
PTK6		P				R						V
GRB2												
GRAP2												
SHF												
SHB												
SHD												
SHE												
TNS1												
TNS3												
TENC1												
TNS4												
CRK												
CRKL												

TABLE II—continued

BRDG1								P				
SUPT6H												
SH3BP2												
PLCG1.C	T					K,S,R,H	K	G,R				
SOCS3												
SOCS1												
SOCS2												
SOCS6	G		P	P			N		N		G	
SOCS5												
DAPP1												
NCK2												
NCK1												
PTPN11.N		E,D	P,R	D					N			T
PTPN11.C												
PTPN11.NC		K,P,D	P,Q	D,N			K,A		T,N	V,R	G,T	T
PTPN6.C												
PTPN6.NC												
PIK3R3.N	G		K	P,R		P,D,S,H	S	I	K		K	A
PIK3R1.N				P		G,D,R	G,S	I			T	
PIK3R2.N	K		K	P,R		K,D	A,S	D	R	S	K	E,A
PIK3R3.C	E	S	K			R	E,S	I,R		R	I	K,A
PIK3R1.C												
PIK3R2.C						L						
PIK3R3.NC	K	S	K,R	I,P		K,S,R	G,K,S	R	A	I,S	K,S	I
PIK3R1.NC	G		K,R	P,R		S	K,S	I	P		T,K	I
PIK3R2.NC	K		K	P		T			K		V	
VAV3												
VAV1												
VAV2												
FER												
FES												
BCAR3												
SH2D3C		P							R	R		
SH2D3A												
SHC1		S					A		R			
SHC1.PTB	S		V,E	P		V		R	K,N	T,A		R
SHC4						L						
SHC2.PTB		P				V				T		
SHC3												
SH2B2												
SH2B												

TABLE II—continued

LNK												
RASA1.N	K		M,R	K,P,R		T	E	G,K,I,R		R		
PLCG2.N												
PLCG2.NC	S						S		R	R		
PLCG1.N	L			P		S		R	R			
PLCG1.NC		L	K	T,P		K,S,H	A,S	R	G,W	I,R		
RASA1.C												
RASA1.NC	V		A					G,R	G		K,L	
INPPL1				E				P				
SH2D1A												
SH2D1B	P					D		P	P,R		F	V
LCP2												
BLNK												
ITK												
BTK												
TEC												
TXK												
BMX												

The PEBL algorithm complements the “best fit” interaction predictions of Scansite and SMALI. We hypothesize that the incorporation of quantitative binding data from interactions across a wide spectrum of affinities represents a more realistic depiction of binding potentials than focusing on the sequence motifs of highest affinity binders. PEBL makes predictions based on permissive and nonpermissive amino acid residues and allowed us to improve upon the accuracy of existing algorithms. PEBL predictions outperformed SMALI in calling interactions from external SPOT array datasets and in cell phosphopeptide pulldown assays. We will continue to further develop and refine this model and will make it available as an on-line application for fellow researchers to assist with the rapid identification of novel and potentially biologically relevant interactions.

Further interaction experiments with larger and more diverse peptide libraries should enable more accurate interaction predictions for SH2 domains that resulted in too few interactions in this study. The current dataset provides many testable hypotheses regarding the interface of SH2 domains with RTK and AR signaling networks.

Acknowledgments—We thank Warren Bennion, Chia-Ying Chuang, and Vanessa Copeland for technical assistance; Elena Solomaha and staff at the Biophysics Core facility for SPR technical assistance and consultation; Sam Bettis, Nicolette Pavlovics, and staff at the IGSB cell screening facility for assistance in automating the high throughput FP assay; Bernard Liu for providing the raw interaction data for the SPOT arrays; and Geoffrey Greene, Carrie

Rinker-Schaeffer, Kenan Onel, and Ravi Salgia for their kind comments and advice on the manuscript.

* This work was supported, in whole or in part, by National Institutes of Health postdoctoral cardiovascular training grant (to J.B.), cancer biology postdoctoral training grant (to C.-P.C.), cancer biology predoctoral training grant (to K.K.L.), genetics and regulation predoctoral training grant (to R.J.H.), and systems biology of oxygen predoctoral training grant (to M.F.C.). This work was also supported by awards from the Cancer Research Foundation and the Illinois Department of Health.

§ This article contains [supplemental material](#).

§ Both authors contributed equally to this manuscript.

|| Present addresses: Dept. of Genome Sciences, University of Washington, Seattle, WA 98195. **Medsphere, Skokie, IL 60077. §§Dept. of Chemical and Biological Engineering, Northwestern University, Evanston, IL 60208. ¶¶Institute of Cellular and System Medicine, and Translational Center for Glandular Malignancies, National Health Research Institutes, Miaoli, Taiwan. ||||VMDR, Inc., Pullman, WA 99163.

|||| To whom correspondence should be addressed. Tel.: 773-702-2185; Fax: 773-834-3913; E-mail: richardbjones@gmail.com.

REFERENCES

- Weinmaster, G., Zoller, M. J., Smith, M., Hinze, E., and Pawson, T. (1984) Mutagenesis of Fujinami sarcoma virus: evidence that tyrosine phosphorylation of P130gag-fps modulates its biological activity. *Cell* **37**, 559–568
- Weinmaster, G., and Pawson, T. (1986) Protein kinase activity of FSV (Fujinami sarcoma virus) P130gag-fps shows a strict specificity for tyrosine residues. *J. Biol. Chem.* **261**, 328–333
- Sadowski, I., Stone, J. C., and Pawson, T. (1986) A noncatalytic domain conserved among cytoplasmic protein-tyrosine kinases modifies the kinase function and transforming activity of Fujinami sarcoma virus

- P130gag-fps. *Mol. Cell. Biol.* **6**, 4396–4408
4. Anderson, D., Koch, C. A., Grey, L., Ellis, C., Moran, M. F., and Pawson, T. (1990) Binding of SH2 domains of phospholipase C(γ)1, GAP, and Src to activated growth factor receptors. *Science* **250**, 979–982
 5. Moran, M. F., Koch, C. A., Anderson, D., Ellis, C., England, L., Martin, G. S., and Pawson, T. (1990) Src homology region 2 domains direct protein-protein interactions in signal transduction. *Proc. Natl. Acad. Sci. U.S.A.* **87**, 8622–8626
 6. Kazlauskas, A., Ellis, C., Pawson, T., and Cooper, J. A. (1990) Binding of GAP to activated PDGF receptors. *Science* **247**, 1578–1581
 7. Mohammadi, M., Honegger, A. M., Rotin, D., Fischer, R., Bellot, F., Li, W., Dionne, C. A., Jaye, M., Rubinstein, M., and Schlessinger, J. (1991) A tyrosine-phosphorylated carboxy-terminal peptide of the fibroblast growth factor receptor (Fg) is a binding site for the SH2 domain of phospholipase C- γ 1. *Mol. Cell. Biol.* **11**, 5068–5078
 8. Graziani, A., Gramaglia, D., Cantley, L. C., and Comoglio, P. M. (1991) The tyrosine-phosphorylated hepatocyte growth factor/scatter factor receptor associates with phosphatidylinositol 3-kinase. *J. Biol. Chem.* **266**, 22087–22090
 9. Cantley, L. C., Auger, K. R., Carpenter, C., Duckworth, B., Graziani, A., Kapeller, R., and Soltoff, S. (1991) Oncogenes and signal transduction. *Cell* **64**, 281–302
 10. Zhou, S., Shoelson, S. E., Chaudhuri, M., Gish, G., Pawson, T., Haser, W. G., King, F., Roberts, T., Ratnoffsky, S., Lechleider, R. J., Neel, B. G., Birge, R. B., Fajardo, J. E., Chou, M. M., Hanafusa, H., Schaffhausen, B., and Cantley, L. C. (1993) SH2 domains recognize specific phosphopeptide sequences. *Cell* **72**, 767–778
 11. Songyang, Z., and Cantley, L. C. (1995) Recognition and specificity in protein tyrosine kinase-mediated signalling. *Trends Biochem. Sci.* **20**, 470–475
 12. Obenauer, J. C., Cantley, L. C., and Yaffe, M. B. (2003) Scansite 2.0: Proteome-wide prediction of cell signaling interactions using short sequence motifs. *Nucleic Acids Res.* **31**, 3635–3641
 13. Li, L., Wu, C., Huang, H., Zhang, K., Gan, J., and Li, S. S. C. (2008) Prediction of phosphotyrosine signaling networks using a scoring matrix-assisted ligand identification approach. *Nucleic Acids Res.* **36**, 3263–3273
 14. Huang, H., Li, L., Wu, C., Schibli, D., Colwill, K., Ma, S., Li, C., Roy, P., Ho, K., Songyang, Z., Pawson, T., Gao, Y., and Li, S. S. C. (2008) Defining the specificity space of the human SRC homology 2 domain. *Mol. Cell. Proteomics* **7**, 768–784
 15. Jones, R. B., Gordus, A., Krall, J. A., and MacBeath, G. (2006) A quantitative protein interaction network for the ErbB receptors using protein microarrays. *Nature* **439**, 168–174
 16. Gordus, A., Krall, J. A., Beyer, E. M., Kaushansky, A., Wolf-Yadlin, A., Sevecka, M., Chang, B. H., Rush, J., and MacBeath, G. (2009) Linear combinations of docking affinities explain quantitative differences in RTK signaling. *Mol. Syst. Biol.* **5**, 235
 17. Kaushansky, A., Gordus, A., Chang, B., Rush, J., and MacBeath, G. (2008) A quantitative study of the recruitment potential of all intracellular tyrosine residues on EGFR, FGFR1 and IGF1R. *Mol. Biosyst.* **4**, 643–653
 18. Liu, B. A., Engemann, B. W., Jablonowski, K., Higginbotham, K., Stergachis, A. S., and Nash, P. D. (2012) SRC homology 2 domain binding sites in insulin, IGF-1, and FGF receptor mediated signaling networks reveal an extensive potential interactome. *Cell Commun. Signal.* **10**.1186/1478-811X-10-27
 19. Liu, B. A., Jablonowski, K., Shah, E. E., Engemann, B. W., Jones, R. B., and Nash, P. D. (2010) SH2 domains recognize contextual peptide sequence information to determine selectivity. *Mol. Cell. Proteomics* **9**, 2391–2404
 20. Hause, R. J., Leung, K. K., Barkinge, J. L., Ciaccio, M. F., Chuu, C., and Jones, R. B. (2012) Comprehensive binary interaction mapping of SH2 domains via fluorescence polarization reveals novel functional diversification of ErbB receptors. *PLoS ONE* **7**, e44471
 21. Robinson, D. R., Wu, Y. M., and Lin, S. F. (2000) The protein tyrosine kinase family of the human genome. *Oncogene* **19**, 5548–5557
 22. Ciaccio, M. F., Wagner, J. P., Chuu, C. P., Lauffenburger, D. A., and Jones, R. B. (2010) Systems analysis of EGF receptor signaling dynamics with microWestern arrays. *Nat. Meth.* **7**, 148–155
 23. Volinsky, N., and Kholodenko, B. N. (2013) Complexity of receptor tyrosine kinase signal processing. *Cold Spring Harb. Perspect. Biol.* **5**, a009043
 24. Guo, A., Villén, J., Kornhauser, J., Lee, K. A., Stokes, M. P., Rikova, K., Possemato, A., Nardone, J., Innocenti, G., Wetzel, R., Wang, Y., MacNeill, J., Mitchell, J., Gygi, S. P., Rush, J., Polakiewicz, R. D., and Comb, M. J. (2008) Signaling networks assembled by oncogenic EGFR and c-Met. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 692–697
 25. Engelman, J. A., Zejnullahu, K., Mitsudomi, T., Song, Y., Hyland, C., Park, J. O., Lindeman, N., Gale, C. M., Zhao, X., Christensen, J., Kosaka, T., Holmes, A. J., Rogers, A. M., Cappuzzo, F., Mok, T., Lee, C., Johnson, B. E., Cantley, L. C., and Jänne, P. A. (2007) MET amplification leads to gefitinib resistance in lung cancer by activating ERBB3 signaling. *Science* **316**, 1039–1043
 26. Xu, A. M., and Huang, P. H. (2010) Receptor tyrosine kinase coactivation networks in cancer. *Cancer Res.* **70**, 3857–3860
 27. Stommel, J. M., Kimmelman, A. C., Ying, H., Nabioullin, R., Ponugoti, A. H., Wiedemeyer, R., Stegh, A. H., Bradner, J. E., Ligon, K. L., Brennan, C., Chin, L., and DePinho, R. A. (2007) Coactivation of receptor tyrosine kinases affects the response of tumor cells to targeted therapies. *Science* **318**, 287–290
 28. Hause, R. J., Jr., Leung, K. K., Barkinge, J. L., Ciaccio, M. F., Chuu, C. P., and Jones, R. B. (2012) Comprehensive binary interaction mapping of SH2 domains via fluorescence polarization reveals novel functional diversification of ErbB receptors. *PLoS ONE* **7**, e44471
 29. Hynes, N. E., and Lane, H. A. (2005) ERBB receptors and cancer: the complexity of targeted inhibitors. *Nat. Rev. Cancer* **5**, 341–354
 30. Kitamura, Y., and Hirota, S. (2004) Kit as a human oncogenic tyrosine kinase. *Cell. Mol. Life Sci.* **61**, 2924–2931
 31. Bolanos-Garcia, V. M. (2005) MET meet adaptors: functional and structural implications in downstream signalling mediated by the Met receptor. *Mol. Cell. Biochem.* **276**, 149–157
 32. Holgado-Madruga, M., Emllet, D. R., Moscatello, D. K., Godwin, A. K., and Wong, A. J. (1996) A Grb2-associated docking protein in EGF- and insulin-receptor signalling. *Nature* **379**, 560–564
 33. Kallin, A., Demoulin, J. B., Nishida, K., Hirano, T., Rönstrand, L., and Heldin, C. H. (2004) Gab1 contributes to cytoskeletal reorganization and chemotaxis in response to platelet-derived growth factor. *J. Biol. Chem.* **279**, 17897–17904
 34. Sachs, M., Brohmann, H., Zechner, D., Müller, T., Hülsken, J., Walther, I., Schaeper, U., Birchmeier, C., and Birchmeier, W. (2000) Essential role of Gab1 for signaling by the c-Met receptor *in vivo*. *J. Cell Biol.* **150**, 1375–1384
 35. Bonaccorsi, L., Muratori, M., Carloni, V., Marchiani, S., Formigli, L., Forti, G., and Baldi, E. (2004) The androgen receptor associates with the epidermal growth factor receptor in androgen-sensitive prostate cancer cells. *Steroids* **69**, 549–552
 36. Mahajan, N. P., Liu, Y., Majumder, S., Warren, M. R., Parker, C. E., Mohler, J. L., Earp, H. S., and Whang, Y. E. (2007) Activated Cdc42-associated kinase Ack1 promotes prostate cancer progression via androgen receptor tyrosine phosphorylation. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 8438–8443
 37. Guo, Z., Dai, B., Jiang, T., Xu, K., Xie, Y., Kim, O., Nesheiwat, I., Kong, X., Melamed, J., Handratta, V. D., Njar, V. C., Brodie, A. M., Yu, L. R., Veenstra, T. D., Chen, H., and Qiu, Y. (2006) Regulation of androgen receptor activity by tyrosine phosphorylation. *Cancer Cell* **10**, 309–319
 38. Ponzetto, C., Bardelli, A., Maina, F., Longati, P., Panayotou, G., Dhand, R., Waterfield, M. D., and Comoglio, P. M. (1993) A novel recognition motif for phosphatidylinositol 3-kinase binding mediates its association with the hepatocyte growth factor/scatter factor receptor. *Mol. Cell. Biol.* **13**, 4600–4608
 39. Deleted in proof
 40. Gual, P., Giordano, S., Anguissola, S., and Comoglio, P.M. (2001) Differential requirement of the last C-terminal tail of Met receptor for cell transformation and invasiveness. *Oncogene* **20**, 5493–5502
 41. Sakkab, D., Lewitzky, M., Posern, G., Schaeper, U., Sachs, M., Birchmeier, W., and Feller, S. M. (2000) Signaling of hepatocyte growth factor/scatter factor (HGF) to the small GTPase Rap1 via the large docking protein Gab1 and the adapter protein CRKL. *J. Biol. Chem.* **275**, 10772–10778
 42. Lamorte, L., Kamikura, D., and Park, M. (2000) A switch from p130Cas/Crk to Gab1/Crk signaling correlates with anchorage independent growth and JNK activation in cells transformed by the Met receptor oncoprotein. *Oncogene* **19**, 5973–5981
 43. Bouscary, D., Lecoq-Lafon, C., Chretien, S., Zompi, S., Fichelson, S., Muller, O., Porteu, F., Dusanter-Fourt, I., Gisselbrecht, S., Mayeux, P.,

- and Lacombe, C. (2001) Role of Gab proteins in phosphatidylinositol 3-kinase activation by thrombopoietin (Tpo). *Oncogene* **20**, 2197–2204
44. Montagner, A., Yart, A., Dance, M., Perret, B., Salles, J. P., and Raynal, P. (2005) A novel role for Gab1 and SHP2 in epidermal growth factor-induced Ras activation. *J. Biol. Chem.* **280**, 5350–5360
 45. Gual, P., Giordano, S., Williams, T. A., Rocchi, S., Van Obberghen, E., and Comoglio, P. M. (2000) Sustained recruitment of phospholipase C- γ to Gab1 is required for HGF-induced branching tubulogenesis. *Oncogene* **19**, 1509–1518
 46. Yamasaki, S., Nishida, K., Yoshida, Y., Itoh, M., Hibi, M., and Hirano, T. (2003) Gab1 is required for EGF receptor signaling and the transformation by activated ErbB2. *Oncogene* **22**, 1546–1556
 47. Alexander, W. S. (2002) Suppressors of cytokine signalling (SOCS) in the immune system. *Nat. Rev. Immunol.* **2**, 410–416
 48. Deleted in proof
 49. Munugalavada, V., and Kapur, R. (2005) Role of c-Kit and erythropoietin receptor in erythropoiesis. *Crit. Rev. Oncol. Hematol.* **54**, 63–75
 50. Baxter, R. M., Secrist, J. P., Vaillancourt, R. R., and Kazlauskas, A. (1998) Full activation of the platelet-derived growth factor β -receptor kinase involves multiple events. *J. Biol. Chem.* **273**, 17050–17055
 51. Bernard, A., and Kazlauskas, A. (1999) Phosphospecific antibodies reveal temporal regulation of platelet-derived growth factor beta receptor signaling. *Exp. Cell Res.* **253**, 704–712
 52. Rosenkranz, S., DeMali, K. A., Gelderloos, J. A., Bazenet, C., and Kazlauskas, A. (1999) Identification of the receptor-associated signaling enzymes that are required for platelet-derived growth factor-AA-dependent chemotaxis and DNA synthesis. *J. Biol. Chem.* **274**, 28335–28343
 53. Engel, P., Eck, M. J., and Terhorst, C. (2003) The SAP and SLAM families in immune responses and X-linked lymphoproliferative disease. *Nat. Rev. Immunol.* **3**, 813–821
 54. Roy, A. K., Lavrovsky, Y., Song, C. S., Chen, S., Jung, M. H., Velu, N. K., Bi, B. Y., and Chatterjee, B. (1999) Regulation of androgen action. *Vitam. Horm.* **55**, 309–352
 55. Kraus, S., Gioeli, D., Vomastek, T., Gordon, V., and Weber, M. J. (2006) Receptor for activated C kinase 1 (RACK1) and Src regulate the tyrosine phosphorylation and function of the androgen receptor. *Cancer Res.* **66**, 11047–11054
 56. Huggins, C., Stevens, R., and Hodges, C. (1941) Studies on prostatic cancer: II. The effects of castration on advanced carcinoma of the prostate gland. *Arch. Surg.* **43**, 209–223
 57. Bonaccorsi, L., Muratori, M., Carloni, V., Zecchi, S., Formigli, L., Forti, G., and Baldi, E. (2003) Androgen receptor and prostate cancer invasion. *Int. J. Androl.* **26**, 21–25
 58. Chuu, C. P., Hiipakka, R. A., Fukuchi, J., Kokontis, J. M., and Liao, S. (2005) Androgen causes growth suppression and reversion of androgen-independent prostate cancer xenografts to an androgen-stimulated phenotype in athymic mice. *Cancer Res.* **65**, 2082–2084
 59. Tauchi, T., Feng, G. S., Marshall, M. S., Shen, R., Mantel, C., Pawson, T., and Broxmeyer, H. E. (1994) The ubiquitously expressed Syp phosphatase interacts with c-kit and Grb2 in hematopoietic cells. *J. Biol. Chem.* **269**, 25206–25211
 60. Rodriguez, M., Li, S. S., Harper, J. W., and Songyang, Z. (2004) An oriented peptide array library (OPAL) strategy to study protein-protein interactions. *J. Biol. Chem.* **279**, 8802–8807
 61. Deleted in proof
 62. Sánchez, I. E., Beltrao, P., Stricher, F., Schymkowitz, J., Ferkinghoff-Borg, J., Rousseau, F., and Serrano, L. (2008) Genome-wide prediction of SH2 domain targets using structural information and the FoldX algorithm. *PLoS Comput. Biol.* **4**, e1000052
 63. Deleted in proof
 64. Beck, J. R., and Shultz, E. K. (1986) The use of relative operating characteristic (ROC) curves in test performance evaluation. *Arch. Pathol. Lab. Med.* **110**, 13–20
 65. Liaw, A., and Wiener, M. (2002) Classification and regression by random forest. *R News* **2**, 18–22
 66. Breiman, L. (2001) Random Forests. *Mach. Learn.* **45**, 5–32
 67. Deleted in proof
 68. Beebe, K. D., Wang, P., Arabaci, G., and Pei, D. (2000) Determination of the binding specificity of the SH2 domains of protein tyrosine phosphatase SHP-1 through the screening of a combinatorial phosphotyrosyl peptide library. *Biochemistry* **39**, 13251–13260
 69. Sweeney, M. C., Wavreille, A. S., Park, J., Butchar, J. P., Tridandapani, S., and Pei, D. (2005) Decoding protein-protein interactions through combinatorial chemistry: sequence specificity of SHP-1, SHP-2, and SHIP SH2 domains. *Biochemistry* **44**, 14932–14947
 70. Tinti, M., Kiemer, L., Costa, S., Miller, M. L., Sacco, F., Olsen, J. V., Carducci, M., Paoluzi, S., Langone, F., Workman, C. T., Blom, N., Machida, K., Thompson, C. M., Schutkowski, M., Brunak, S., Mann, M., Mayer, B. J., Castagnoli, L., and Cesareni, G. (2013) The SH2 domain interaction landscape. *Cell Reports* **3**, 1293–1305
 71. Liu, B. A., Jablonowski, K., Raina, M., Arcé, M., Pawson, T., and Nash, P. D. (2006) The human and mouse complement of SH2 domain proteins—establishing the boundaries of phosphotyrosine signaling. *Mol. Cell* **22**, 851–868
 72. Sing, T., Sander, O., Beerenwinkel, N., and Lengauer, T. (2005) ROCr: visualizing classifier performance in R. *Bioinformatics* **21**, 3940–3941
 73. Birchmeier, C., Birchmeier, W., Gherardi, E., and Vande Woude, G. F. (2003) Met, metastasis, motility and more. *Nat. Rev. Mol. Cell Biol.* **4**, 915–925
 74. Gentile, A., Trusolino, L., and Comoglio, P. M. (2008) The Met tyrosine kinase receptor in development and cancer. *Cancer Metastasis Rev.* **27**, 85–94
 75. Jo, M., Stolz, D. B., Esplen, J. E., Dorko, K., Michalopoulos, G. K., and Strom, S. C. (2000) Cross-talk between epidermal growth factor receptor and c-Met signal pathways in transformed cells. *J. Biol. Chem.* **275**, 8806–8811
 76. Jones, R. B. (2013) Do low-affinity ErbB receptor protein interactions represent the base of a cell signaling iceberg? *Expert Rev. Proteomics* **10**, 115–118
 77. Kerrien, S., Aranda, B., Breuza, L., Bridge, A., Broackes-Carter, F., Chen, C., Duesbury, M., Dumousseau, M., Feuermann, M., Hinz, U., Jandrasits, C., Jimenez, R. C., Khadake, J., Mahadevan, U., Masson, P., Pedruzzi, I., Pfeifferberger, E., Porras, P., Raghunath, A., Roechert, B., Orchard, S., and Hermjakob, H. (2012) The IntAct molecular interaction database in 2012. *Nucleic Acids Res.* **40**, D841–D846