A Strategy for Interaction Site Prediction between Phospho-binding Modules and their Partners Identified from Proteomic Data*

Willy Aucher‡∥ ∥∥, Emmanuelle Becker§¶**∥∥, Emilie Ma‡, Simona Miron‡‡, Arnaud Martel§§, Françoise Ochsenbein§¶, Marie-Claude Marsolier-Kergoat¶¶‡, and Raphaël Guerois¶¶§¶

Small and large scale proteomic technologies are providing a wealth of potential interactions between proteins bearing phospho-recognition modules and their substrates. Resulting interaction maps reveal such a dense network of interactions that the functional dissection and understanding of these networks often require to break specific interactions while keeping the rest intact. Here, we developed a computational strategy, called STRIP, to predict the precise interaction site involved in an interaction with a phospho-recognition module. The method was validated by a two-hybrid screen carried out using the ForkHead Associated (FHA)1 domain of Rad53, a key protein of *Saccharomyces cerevisiae* DNA checkpoint, as a bait. In this screen we detected 11 partners, including Cdc7 and Cdc45, essential components of the DNA replication machinery. FHA domains are phospho-threonine binding modules and the threonines involved in both interactions could be predicted using the STRIP strategy. The threonines T484 and T189 in Cdc7 and Cdc45, respectively, were mutated and loss of binding could be monitored experimentally with the full-length proteins. The method was further tested for the analysis of 63 known Rad53 binding partners and provided several key insights regarding the threonines likely involved in these interactions. The STRIP method relies on a combination of conservation, phosphorylation likelihood, and binding specificity criteria and can be accessed via a web interface at http://biodev.extra.cea.fr/strip/. *Molecular & Cellular Proteomics* 9:2745–2759, 2010.

Cell processes are tightly coordinated through signal transduction pathways that heavily depend on reversible post-translational modifications, including the phosphorylation of serine, threonine, and tyrosine residues (1, 2). Reflecting the multiplicity of residues being phosphorylated at a given time in a cell, several modules are able to mediate the specific recognition of phosphorylated partners. Typically such modules are the 14–3-3, BRCT, C2, FHA, MH2, PBD, PTB, SH2, WD-40, and WW domains (3) (see a description in dedicated databases (4, 5)). These modules achieve their binding specificity primarily through the recognition of a region, usually a short sequence motif ($\sim$ 6–15 residues) containing the phosphorylated residues (6).

Small- and large-scale proteomic technologies such as the two-hybrid technique or affinity purification are providing a wealth of potential interactions between the proteins bearing these recognition modules and their substrates. The current protein-protein interaction maps reveal a dense network of interactions with a high degree of interconnections between nodes. Entire gene deletions bring about major perturbations that complicate the functional interpretation of a specific interaction. The functional dissection of an interaction network and the understanding of its molecular logic require a more local perturbation that breaks a specific interaction while keeping the rest of the network intact. In that scope, the precise identification of the phosphorylated residue(s) responsible for an interaction often turns out to be a laborious task. Here, we propose a strategy, called STRategy for Interacting site Prediction (STRIP),[1] to accelerate the faithful identification of these binding residues for a given phospho-binding module by coupling together several types of information : (i) the probability of a residue to be phosphorylated, (ii) the respect of the motif(s) bound with the highest affinity by the module around the modified residue, and (iii) the strict conservation of this motif in closely related species. The STRIP strategy can easily be used on the internet via a web server we designed for that purpose (http://biodev.extra.cea.fr/strip/).

To test this strategy we focused on one family of recognition modules, the ForkHead Associated (FHA) domains. We particularly addressed the case of the first FHA domain of

[1] The abbreviations used are: STRIP, strategy for interacting site prediction; GST, glutathione S transferase; FHA, forkhead associated.

Rad53, a *Saccharomyces cerevisiae* kinase involved in response pathways to genotoxic stresses whose catalytic domain is flanked by two FHA domains, named FHA1 and FHA2. The FHA (ForkHead Associated) domain was discovered by Hofmann and Bucher, who recognized a protein motif in a subset of forkhead-type transcription factors (7). This domain has since then been found in hundreds of proteins from eukaryotic, eubacterial, and archeal species. Biochemical studies of specific FHA domains (including Rad53 domains) have demonstrated that FHA domains bind specifically phosphothreonines and have little affinity for phosphoserines or phosphotyrosines or for unphosphorylated threonines *in vitro* (8–10). Moreover, powerful *in vitro* screening strategies using combinatorial phosphopeptide libraries showed that the amino acids surrounding the phosphothreonine (pT) contribute to the FHA domains binding specificities. The highest discrimination was usually found for the amino acids in positions either (pT+3) (9, 11–13) or (pT-3) (14) with a few notable exceptions (15).

The phosphopeptide binding function of the FHA domains was first demonstrated by studying Rad53 FHA1 and FHA2 (8–10). Rad53 FHA1 domain is probably the FHA module whose biochemical and physiological characteristics have been the most thoroughly analyzed (8, 16–23), which makes it an attractive target for a new predictive approach. In particular, two groups of investigators have found that FHA1 specifically binds phosphothreonines inside pTXXD motifs *in vitro* (8, 9).

Rad53 is part of the DNA checkpoints, response pathways that detect DNA lesions or replication blocks and coordinate various responses such as cell cycle arrests and transcriptional or post-translational modifications. Rad53 interacts with many different partners, and more than 30 FHA1 binding proteins have been described (18, 20, 24, 25), although it is not always clear whether the interaction is direct or not. Abolishing FHA1 phosphopeptide binding function by mutating conserved residues such as R70 and N107 leads to a slightly increased sensitivity to DNA damage generated by UV irradiation or by treatment with methyl methane sulfonate, but to a severe hypersensitivity to hydroxyurea (an inhibitor of ribonucleotide reductase that induces replication fork stalling through depletion of dNTP pools), which suggests that FHA1 has a specialized function related to replicational stress (19). However, because mutating FHA1 disrupts the interactions with all its partners, the interactions involved in resistance to replicational stress remain undetermined.

In this article, we set up the STRIP strategy designed to identify the ligands bound by phosphobinding modules and we first sought to investigate FHA1 ligands, and more precisely the FHA1 ligands involved in replicational stress. We performed two-hybrid screens using Rad53 FHA1 domain as a bait to identify new partners of FHA1 or to qualify previously described interactants of Rad53 as FHA1 ligands and we isolated two essential proteins involved in DNA replication,

Cdc7 and Cdc45. Using the STRIP strategy, we predicted the FHA1-bound phosphothreonines and we confirmed experimentally these predictions *in vivo* and *in vitro*. Mutating the FHA1-bound threonines of Cdc7 and Cdc45 led to no obvious phenotype. However, the STRIP strategy was also able to identify in Ptc2, a negative regulator of Rad53, the threonine T376 that had previously been characterized experimentally as an FHA1 ligand and whose mutation leads to defects in Rad53 inactivation. Finally, we applied the STRIP analysis to all Rad53 ligands.

EXPERIMENTAL AND COMPUTATIONAL PROCEDURES

Plasmids—The sequence encoding Rad53 residues 1 to 164 [Rad53(1-164)] was amplified by PCR and cloned between the EcoRI and the BamHI sites of pGBT9 (Clontech), so as to create pGBT9/FHA1. The pETM-30/FHA1(1–164) plasmid containing the sequence encoding Rad53 residues 1 to 164, which was used for the production of the FHA1 domain and of the glutathione S transferase (GST)-FHA1 fusion is described in (26). The *CDC7* and the *CDC45* genes (*i.e.* sequences comprising the entire coding sequences plus 500 bp upstream of the start codons and 500 bp downstream of the stop codons) were amplified by PCR and cloned between the XbaI and the EcoRI sites of pRS316-LYS2, a *LYS2*-marked derivative of the pRS316 plasmid (27), and between the BamHI and NotI sites of pRS314 (27), respectively, to give the plasmids pRS316-LYS2/*CDC7* and pRS314/*CDC45*. Point mutations in the sequences encoding Rad53 FHA1, Cdc7 and Cdc45 were realized using the QuickChange site-directed mutagenesis system (Stratagene, La Jolla, CA). All constructs were verified by sequencing. The wild-type and mutated *CDC7* and *CDC45* genes harbored by the pRS316-LYS2 and pRS314 plasmids were subsequently TAP-tagged using standard homologous recombination techniques. The plasmid pJA98 harboring a sequence encoding a HA-tagged version of Rad53 under the *GAL* promoter (28) was modified by replacing the *URA3* selection marker by *ADE2* using standard homologous recombination techniques. The resulting plasmid was named pJA98Ade. Further information is available upon request.

Two-Hybrid Screening—The yeast strains Y187 and Y190 were used for two-hybrid screening using the mating strategy as described in (29). We performed two-hybrid screenings using Rad53 FHA1 domain as a bait encoded by the pGBT9/FHA1 plasmid and the FRYL library of yeast genomic fragments cloned into the pACTII vector, a kind gift of Michèle Fromont-Racine and Pierre Legrain described in (29). Two screenings were performed, either in the absence of genotoxic stress or in the presence of camptothecin (5 µg/ml). About 40 × 10⁶ interactions were tested in each screening. Following selection for growth on plates lacking histidine complemented with 100 mM 3-amino-triazole and for X-Gal 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside staining, the pACTII-derived plasmids of the FRYL library were recov-

ered and reintroduced into the testing strain in order to validate the interaction in the absence of genotoxic stress. The *CDC7* fragment was isolated in the screen performed in the presence of camptothecin but the FHA1/Cdc7 interaction was also observed in the absence of camptothecin.

Affinity Measures—Phosphorylated and unphosphorylated peptides from Ptc2 (DDIpTDADTDAE), Cdc7 (DGESpTDEDD-VVS), and Cdc45 (DDEApTDADEVTD) spanning the sequence of the identified FHA1 binding sites were obtained by chemical synthesis. The FHA1 domain was produced using the pETM-30/FHA1(1–164) plasmid and purified as previously described in (26). A VP-ITC Microcal was used to measure the affinities using sample concentrations (FHA1 domain (20 μM), peptides (200 μM)) at 30 °C, Tris 50 mM, pH 8.

STRategy for Interacting Site Prediction (STRIP)—Phosphorylation likelihood was predicted as significant if one of the two scores obtained with the NetPhos2.0 (30) and DisPhos1.3 (31) dedicated programs raised above the 0.5 threshold. The binding motif recognized with the highest affinity by the FHA1 domain of Rad53 was identified as the pTxxD motif by two independent peptide library screening studies (8, 9). The conservation was analyzed from a multiple sequence alignment built with ClustalW (32) of the orthologous sequences from species closely related to *S. cerevisiae* (33, 34). For every putative phosphorylated residue (threonine for the FHA) and for each residue in the position specifically recognized with respect to the phosphoresidue (+3 in the case of the Rad53 FHA1), the percentage of sequences for which the residue was strictly conserved was determined. The output of five servers predicting the putative kinases likely phosphorylating the residues, namely NetPhosK (35), PredPhospho (36), PPSP (37), ScanSite (38), and KinasePhos (39), were combined to propose a consensus of the most probable kinase on the STRIP web server.

In Vitro GST Pull-Down Assays—The experiments were carried out using either strains in which the endogenous *CDC7* or *CDC45* genes had been TAP-tagged at their chromosomal loci or strains whose chromosomal copy of *CDC7* or *CDC45* had been deleted and that were complemented with TAP-tagged, either wild-type or mutated, alleles of *CDC7* or *CDC45* harbored by a centromeric plasmid. The strains with a TAP-tagged copy of *CDC7* or *CDC45* at the chromosomal locus (MCM974 and MCM973, respectively) were obtained from the EUROSCARF collection of TAP fusion strains (http://web.uni-frankfurt.de/fb15/mikro/euroscarf/cellzome.html) and are derived from the parental strain SC0000 (*MATa ade2 arg4 leu2–3,112 trp1–289 ura3–52*). The strains containing the plasmid-borne, TAP-tagged, mutant or wild-type, *CDC7* or *CDC45* alleles were constructed in the MCM185 background (*MAT a ura3–52 lys2–801amber ade2–101ochre trp1-Δ63 his3-Δ200 leu2-Δ1 bar1Δ::LEU2*), a derivative of YPH499 (27). The centromeric plasmids harboring the TAP-tagged *CDC7* or *CDC45* genes were first introduced into MCM185 and the *CDC7* and *CDC45* chromosomal alleles were subsequently
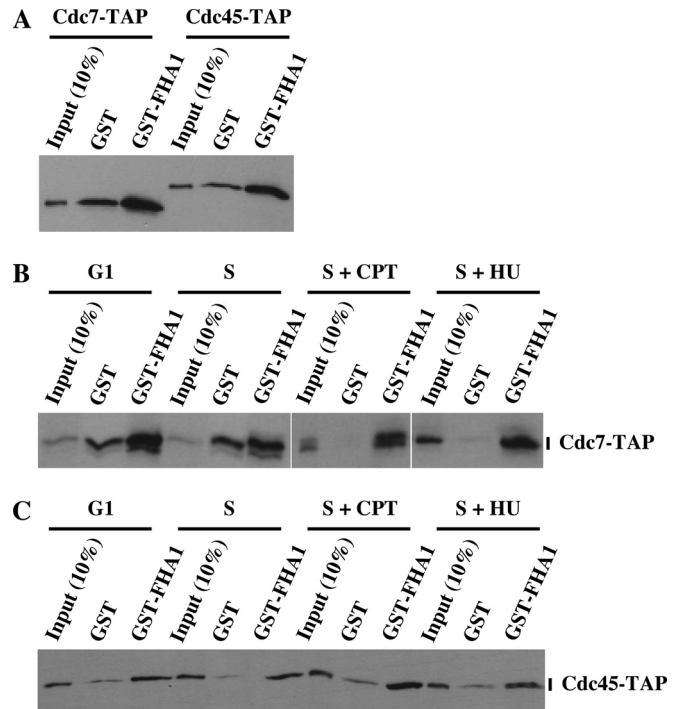
FIG. 1. **Analysis of Rad53 FHA1 interactions with Cdc7 and Cdc45 in the GST pull-down assay.** *A*, Total extracts (0.3 mg protein) from asynchronous cultures of yeast cells containing a TAP-tagged version of Cdc7 or Cdc45 (MCM974 and MCM973, respectively) were incubated with equal amounts (150 μg) of immobilized GST or GST-FHA1. Bound proteins were analyzed by Western blotting using anti-TAP antibody. *B* and *C*, Yeast cells containing a TAP-tagged version of Cdc7 or Cdc45 (MCM974 and MCM973, respectively) were synchronized in G1 with α-factor (G1), released into S phase and either left untreated for 30 min (S) or treated either with camptothecin (20 μg/ml) for 1 h (S + CPT), or with hydroxyurea (200 mM) for 2 h (S + HU). Total extracts were incubated with immobilized GST or GST-FHA1, and bound proteins were analyzed by Western blotting as above.

disrupted by PCR targeting using the *kanMX* cassette (40), giving rise to the strains MCM969 (as MCM185 *cdc7Δ::kanMX* + pRS316-LYS2/*CDC7-TAP*), MCM971 (as MCM185 *cdc7Δ::kanMX* + pRS316-LYS2/*cdc7T484A-TAP*), MCM965 (as MCM185 *cdc45Δ::kanMX* + pRS314/*CDC45-TAP*), and MCM967 (as MCM185 *cdc45Δ::kanMX* + pRS314/*cdc45T189A-TAP*).

The GST and GST-FHA1 proteins were purified as previously described in (26), except that they were incubated overnight with GSH-coated agarose beads at 4 °C and that the beads were ultimately washed in the TBS-N buffer described in (18). The assay was performed as described in (18) with the incubation of native extracts from yeast cells (0.3 mg of protein) with 150 μg of GST or GST-FHA1 proteins bound to GSH-coated beads.

Co-immunoprecipitation Assays—The pJA98Ade plasmid harboring *RAD53* under the control of a *GAL* promoter was introduced into the strains MCM969, MCM971, MCM965, and MCM967 containing the *CDC7-TAP*, *cdc7T484A-TAP*,

ORF name	Protein fragment	Hits $-$ CPT	Clones $-$ CPT	Hits $+$ CPT	Clones $+$ CPT	Description
YNL084C	End3(262–349)	16	7	4	4	Protein involved in endocytosis, actin cytoskeletal organization and cell wall morphogenesis
YLR103C	Cdc45(154–270)	5	3	11	4	DNA replication initiation factor
YDL017W	Cdc7(294–493)	0	0	4	4	Dbf4-dependent kinase catalytic subunit required for firing origins and replication fork progression
YFR022W	Rog3(383–604)	4	2	5	2	Protein that binds to Rsp5, a ubiquitin ligase; overexpression of *RSP5* rescues the *end3Δ* mutant
YGR013W	Snu71(173–359)	2	2	1	1	Component of U1 snRNP required for mRNA splicing
YJR031C	Gea1(1–292)	1	1	2	1	Guanine nucleotide exchange factor for ADP ribosylation factors (ARFs), involved in vesicular transport between the Golgi and endoplasmic reticulum, Golgi organization, and actin cytoskeleton organization
YPL120W	Vps30(104–190)	4	2	0	0	Protein that forms a membrane-associated complex essential for autophagy; involved in vacuolar protein sorting
YOL028C	Yap7(44–245)	1	1	1	1	bZIP transcription factor
YML016C	Ppz1(158–443)	1	1	1	1	S/T phosphatase
YHR202W	Yhr202w(498–602)	1	1	1	1	Putative protein of unknown function
YER089C	Ptc2(61–456)	1	1	1	1	PP2C phosphatase

CDC45-TAP, and *cdc45T189A-TAP* constructs, respectively, and into the control strain MCM185 devoid of any TAP-tagged protein. The MCM185 strain was supplemented with the empty vectors pRS314 (*TRP1*) and pRS316 (*URA3*) in order to have the same auxotrophies as the strains containing the TAP-tagged proteins Cdc7 and Cdc45.

Cells were grown overnight in a minimal medium containing 2% (w/v) raffinose. Expression of Rad53 from the pJA98Ade plasmid was induced by addition of 2% galactose at time zero. After 2 h, cells were harvested and native extracts were prepared as in (18) with the buffer (Tris-HCl, 20 mM pH 7.5; Nonidet P-40, 0.2%; Aprotinine, 6 μg/ml; Pefabloc (Roche), 1/1000; NaCl, 225 mM). Immunoprecipitations directed against the TAP tag were performed by incubating for 2 h at 4 °C 0.3 mg of protein and 75 μl of magnetic beads coated with mouse IgG Dynabeads (Invitrogen) under constant tilt rotation. Beads were washed three times with the above described buffer according to the manufacturer's instructions. SDS-PAGE samples were analyzed on 10% polyacrylamide gels and transferred onto nitrocellulose membrane. Bound proteins were revealed by Western blotting with a rabbit peroxidase anti-peroxidase antibody (Sigma) and with a goat polyclonal antibody raised against the C terminus sequence of Rad53p [RAD53 (yC-19); Santa Cruz Biotechnology]. The antibody directed against Rad53 cross-reacted with the TAP-tagged proteins.

RESULTS

Two-Hybrid Screening of Rad53 FHA1 Binding Proteins—We reasoned that some targets of FHA1 could bind it preferentially or exclusively in the presence of DNA damage and we performed two two-hybrid screenings using FHA1 as a bait, either in the absence of genotoxic stress or in the presence of camptothecin, an inhibitor of topoisomerase I religation reaction that induces the formation of double-strand breaks during DNA replication. Plasmid pGBT9/FHA1, encoding FHA1 (a fragment of Rad53 encompassing amino acids 1 to 164 [Rad53(1–164)]) fused to the DNA binding domain of Gal4 (Gal4BD) was used to screen a library of random genomic fragments fused to Gal4 activation domain sequence (Gal4AD). Following validation, 11 proteins were found to reproducibly interact with FHA1 in the two-hybrid system (Table I), in the absence as in the presence of camptothecin. Out of the 11 proteins, only one, Ptc2, had already been described as interacting with Rad53 FHA1 (18, 41) and one, Cdc7, had been shown to be an *in vitro* phosphorylation substrate of Rad53 (42). The small overlap in the results of different screenings for protein interactants is a common observation, which in this case can be partly attributed to the fact that we used the two-hybrid technique, which mostly detects direct interactions, in contrast with affinity purification, which was used in the study of Smolka and collaborators on FHA1 partners (18).

Analysis of Cdc7 and Cdc45 Interaction with Rad53 FHA1 Domain—Two two-hybrid hits, Cdc7 and Cdc45, appeared as plausible ligands for mediating FHA1 part in resistance to replication stress and were further analyzed. Cdc7 is the catalytic subunit of a kinase required for origin firing and replication fork progression (for review, see (43)). Cdc45 is a DNA replication initiation factor recruited to pre-replicative
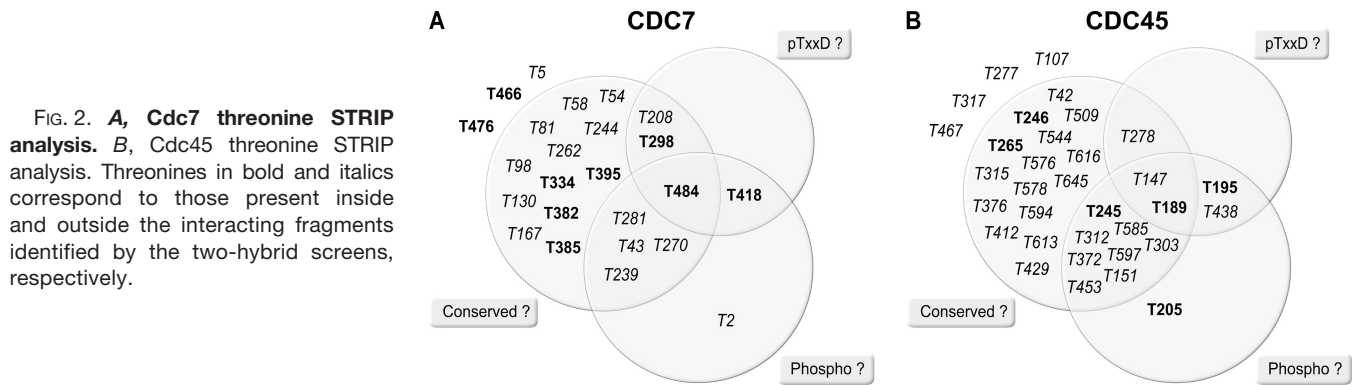
A CDC7

B CDC45

FIG. 2. *A, Cdc7 threonine STRIP analysis. B*, Cdc45 threonine STRIP analysis. Threonines in bold and italics correspond to those present inside and outside the interacting fragments identified by the two-hybrid screens, respectively.

complexes at replication origins and also required for replication elongation (44, 45). We first sought to confirm the interactions between Cdc7 and Cdc45, and Rad53 FHA1 domain, using the glutathione S-transferase (GST) pull-down assay. Extracts from exponentially growing cells containing a TAP-tagged version of Cdc7 or Cdc45 were incubated with immobilized GST or GST-FHA1, and bound proteins were analyzed by Western blotting. As shown in Fig. 1*A*, Cdc7-TAP and Cdc45-TAP fusion proteins bound specifically GST-FHA1. We then investigated the potential impact of the phase of the cell cycle and of the presence of genotoxic stresses on these interactions. Cells containing TAP-tagged versions of Cdc7 or of Cdc45 were synchronized in G1 with α-factor, released into S phase and either left untreated for 30 min or treated either with camptothecin or with hydroxyurea. We found that the interactions between Rad53 FHA1 and Cdc7-TAP and Cdc45-TAP were comparable in G1 and in S phase, with or without genotoxic stress (Fig. 1*B* and *C*).

Designing a Strategy for Predicting the Precise Interacting Sites of the FHA1 Domain of Rad53—The binding partners identified from two-hybrid screens or affinity-based experiments generally bear multiple residues likely to be recognized by a given phospho-recognition module. We explored whether a restricted set of residues could be isolated by screening the sequence of a binding partner for three conditions : (i) the probability of a residue to be phosphorylated, (ii) the respect of the motif recognized with the highest affinity by the phosphobinding module, and (iii) the strict or more relaxed conservation of the motif recognized with the highest affinity in closely related species. This strategy, called STRIP (for STRategy for Interaction site Prediction), is rationalized in more detail in the following.

(i) Phosphorylation likelihood was probed using a consensus information provided by two dedicated methods, Net-Phos2.0 (30) and DisPhos1.3 (31). Other efficient predictors such as GPS (46), KinasePhos (39), NetPhosK (35), PPSP (37), ScanSite (38) were not considered at this stage because they are more oriented toward the substrates of specific kinase classes such as CDK, CK2, PKA, or PKC. NetPhos2.0 utilizes a neural network trained on a database of short phosphotyrosine, phosphoserine, or phosphothreonine peptide frag-

ments likely to be phosphorylated *in vivo* (30) whereas DisPhos1.3 algorithm is based on a logistic regression approach whose training was enhanced by including a prediction of the structural disorder and of secondary structures along the substrate sequence (31). Both methods were estimated to reach accuracies ranging between 70% and 80% and are thus expected to provide complementary insights. (ii) The motif recognized with the highest affinity by the FHA1 domain of Rad53 has been characterized experimentally as pTxxD. (iii) Short linear binding motifs constitute interfaces that were shown to evolve faster than globular domain-domain complexes (47). Consequently, the conservation analysis of the phosphoresidue and of its neighboring positions was restricted to six fully sequenced genomes closely related to *S. cerevisiae*, namely *S. mikatae, S. paradoxus, S. bayanus, S. kluyveri, S. kudriavzevii,* and *S. castelii* (33, 34). For every putative phosphorylated residue and for each residue defined in the motif bound with the highest affinity (position +3 for the FHA1), the percentage of sequences for which the residue was strictly conserved was determined. If no putative phosphoresidue was detected as strictly conserved, the condition was relaxed allowing for residues conserved in more than half the set of sequences. The rationale behind this tolerance is the possible existence of alignment flaws in long disordered regions likely to be phosphorylated and to frequent truncations in the sequences of the six *Saccharomyces* genomes (see Discussion).

Analysis of the Threonine Targeted by FHA1 in Cdc7—It has to be noted that both Cdc7 and Cdc45 are essential proteins, which precludes the analysis of deletant strains and makes compulsory the design of point mutations. We applied to the analysis of Cdc7 and Cdc45 the STRIP strategy described above.

The fragment of Cdc7 identified from the two-hybrid experiment spans the segment 294–493 and bears 9 out of the 24 threonines present in Cdc7, with four TxxD motifs in the fragment (Fig. 2*A*). Only one threonine, T484, fulfilled all three criteria with high phosphorylation probability (88% and 55% according to the NetPhos and Disphos predictors, respectively), respect of the TxxD motif, and strict conservation of the Thr and Asp residues among closely related species. In the Cdc7(294–493) fragment, another threonine, T298, re-
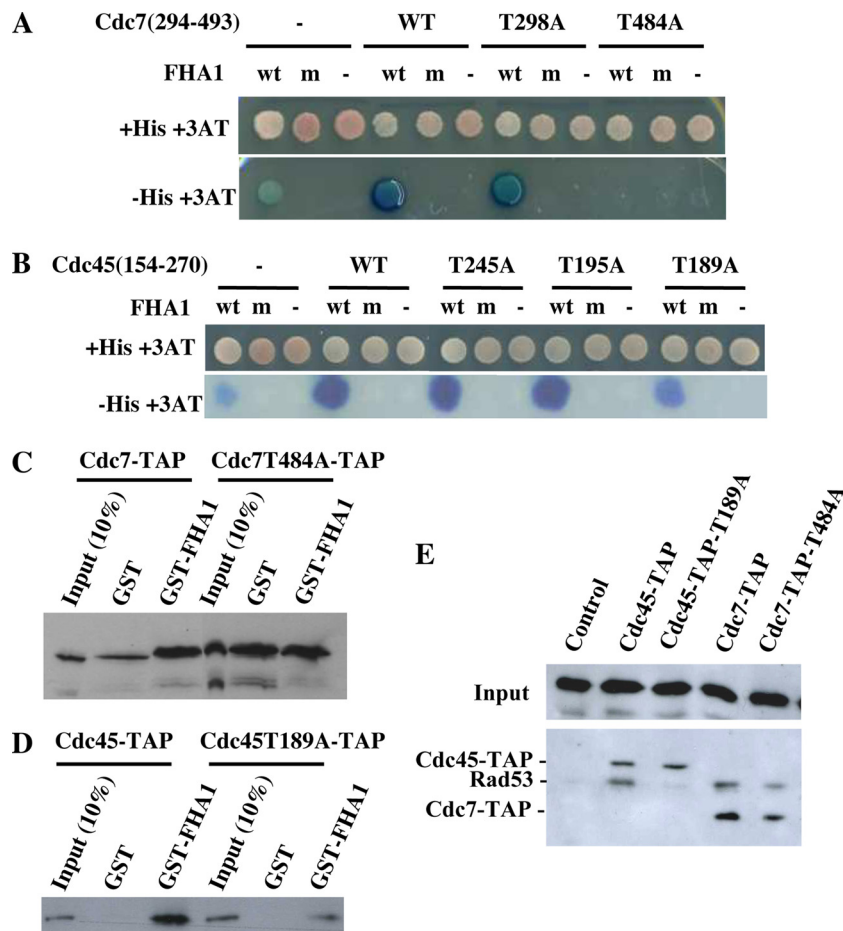
FIG. 3. *A, B,* **The interactions between FHA1 [Rad53(1–164)] and either Cdc7(294–493) (*A*) or Cdc45(154–270) (*B*) depend upon T484 and T189, respectively, in the two-hybrid assay.** pGBT9/FHA1 or pGBT9/FHA1R70A expressing the Gal4BD-Rad53(1–164) (wt) and Gal4BD-Rad53(1–164)R70A (m) fusion proteins, respectively, were introduced into the tester strain along with wild-type or mutated pACTII/Cdc7(294–493) or pACTII/Cdc45(154–270) vectors harboring the sequence encoding Gal4AD fused to the sequences encoding wild-type (WT) Cdc7(294–493) or Cdc45(154–270), or their mutated derivatives as indicated. - indicates that empty vectors (either pGBT9 or pACTII) were used as controls. The two-hybrid interactions were revealed by growth on plates lacking histidine (-His) complemented with 100 mM 3-amino-triazole (3AT) and by X-gal staining. Control plates contained 3AT but were complemented with histidine (+His). It has to be noted that the Gal4BD-Rad53(1–164) fusion protein can activate by itself the transcription of the reporter genes at a low level, hence the residual growth and the slight blue coloration visible on the first spot of the plates (for transformants containing pGBT9/FHA1 and the empty vector pACTII). *C* and *D,* Cdc7 and Cdc45 interactions with Rad53 FHA1 are dependent upon T484 and T189, respectively, in the GST pull-down assay. Total extracts (0.3 mg protein) from asynchronous cultures of yeast cells containing a TAP-tagged version of either the wild-type Cdc7 or Cdc45 (MCM969 and MCM965, respectively) or the mutated Cdc7T484A or Cdc45T189A (MCM971 and MCM967, respectively) proteins were incubated with immobilized GST or GST-FHA1. Bound proteins were analyzed by Western blotting as above. *E,* Co-immunoprecipitation analysis of Cdc7 and Cdc45 interactions with Rad53. The pJA98Ade plasmid harboring *RAD53* under the control of a *GAL* promoter was introduced into the control strain MCM185 (Control) and into the strains MCM969, MCM971, MCM965, MCM967 containing the *CDC7-TAP*, *cdc7T484A-TAP*, *CDC45-TAP* and *cdc45T189A-TAP* constructs, respectively. Cells were grown overnight in minimal media containing raffinose and expression of Rad53 from the pJA98Ade plasmid was induced by addition of galactose. The TAP-tagged proteins were immunoprecipitated using PAP antibodies. Bound proteins were analyzed by Western blotting using an antibody directed against Rad53 that also cross-reacted with the TAP-tagged proteins. 10% of the extract (input) was analyzed using antibodies directed against the C terminus sequence of Rad53p.

spects the TxxD motif and is strictly conserved, but has a low phosphorylation likelihood (about 17%). We tested our *in silico* prediction by mutating T298 and T484 into alanine. As shown in Fig. 3*A*, both the wild-type Cdc7(294–493) fragment and the Cdc7(294–493)T298A mutant interacted in the two-hybrid assay with the wild-type FHA1 domain but not with a mutant FHA1 affected in its phosphopeptide binding function

(FHA1R70A). In contrast, mutating T484 into alanine abolished Cdc7(294–493) interaction with FHA1. We verified that the wild-type and the mutant Gal4AD-Cdc7(294–493) fusions were expressed to the same levels (data not shown). These data were confirmed by the fact that in contrast to TAP-tagged Cdc7, the TAP-tagged Cdc7T484A protein showed the same background affinity for GST and for the GST-FHA1
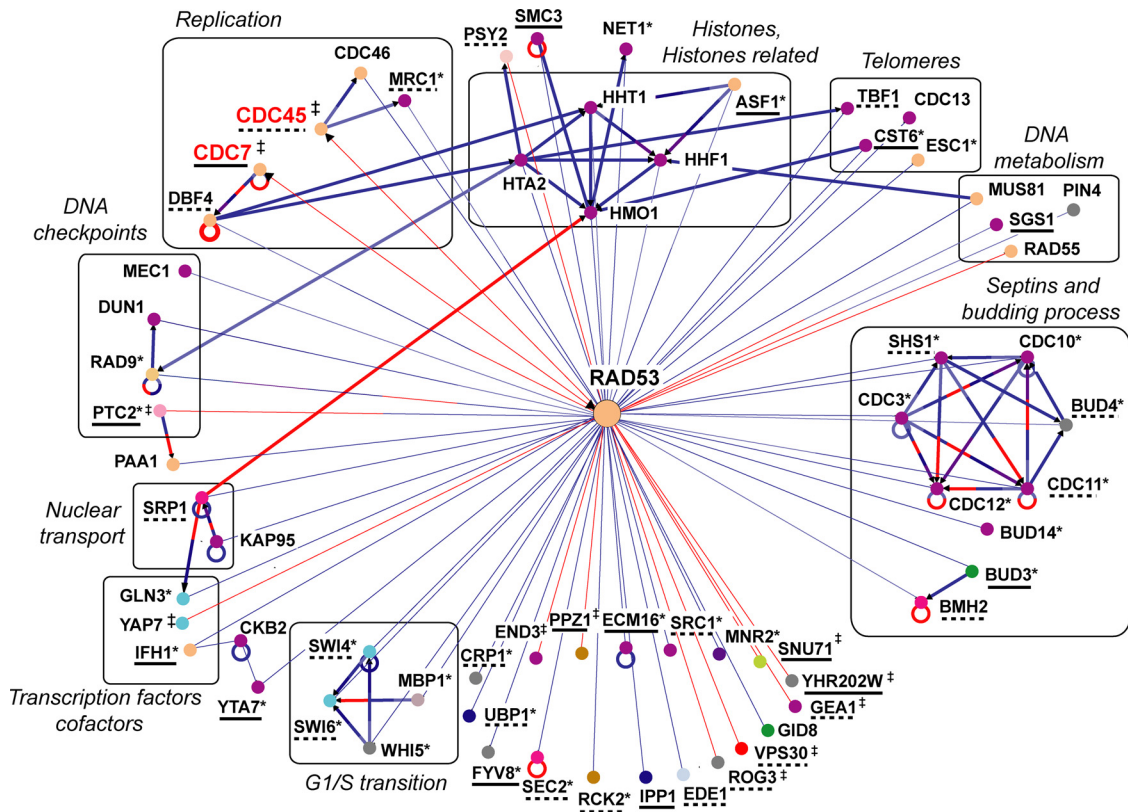
FIG. 4. **Graph summarizing the proteins found to physically interact with Rad53 in affinity-based (blue links) and two-hybrid (red links) experiments as collected in the Biogrid database.** Color codes refer to the Gene Ontology definitions used in the Osprey visualization program. Cdc7 and Cdc45 gene names are colored red and the other gene names that were studied in this work are labeled with an obelisk (‡). Proteins found to specifically interact with the isolated FHA1 domain of Rad53 through affinity-based experiments are labeled with an asterisk (*) (18). When possible, gene names were clustered with respect to their connectivity and to their biological function in black boxes. Plain or dashed underlines indicate proteins bearing a putative FHA1 binding phosphothreonine as detected using the stringent or the more permissive predictive criteria, respectively.

fusion in the GST pull-down assay (Fig. 3C). Interactions between full-length Rad53 and Cdc7 could also be observed by co-immunoprecipitation and immunoblot analysis (Fig. 3E). A slight decrease in Rad53 binding to the Cdc7 mutant compared with the wild-type was visible, consistent with the results obtained with the isolated FHA1 domain.

Our results thus indicate that Cdc7 threonine T484 should be the target of FHA1 and validate our prediction concerning the identity of FHA1 ligand. Interaction between the full-length Rad53 and Cdc7 proteins was also found to partially depend on T484 with likely contributions of alternative interaction sites. Interestingly, even considering the full-length Cdc7 rather than the Cdc7(294–493) fragment, we would have reached a similar conclusion because T484 was the highest scoring residue of all Cdc7 threonines.

*Analysis of the Threonine Targeted by FHA1 in Cdc45—*We identified the Cdc45(154–270) fragment as one of FHA1 interacting substrates in our screen. Cdc45 bears 33 threonines, six of which are located between amino acids 154 and 270 (Fig. 2B). None of the six threonines fulfilled the three stringent criteria altogether and stringency on the conserva-

tion was relaxed in a second step as stated in the description of the STRIP methodology. Then, only one threonine out of the six, T189, was found to fulfil the binding site criteria with a moderate conservation in three out of the five available sequences. Analysis of the multiple sequence alignment around T189 showed that it is located in a poly-acid stretch, likely disordered and difficult to align. A rapid inspection of the sequences lacking the TxxD conservation revealed that this motif could easily be identified in the neighborhood and realigned without disrupting the alignment consistency (Supplemental data Fig. 1). To validate our prediction three point mutants of Cdc45 were designed. Selected threonines corresponded either to T189 that fulfilled all three criteria or to T245 and T195 that fulfilled only two of them. As shown in Fig. 3B, the wild-type Cdc45(154–270) fragment and the Cdc45(154–270)T245A and Cdc45(154–270)T195A mutants interacted similarly in the two-hybrid assay with the wild-type FHA1 domain (and not with the mutant FHA1R70A domain). Conversely, the interaction between FHA1 and Cdc45(154–270)T189A was weaker and we verified that this was not because of a defective expression of the Gal4AD-Cdc45(154–

TABLE II

Thermodynamic parameters for FHA1 binding to Ptc2 (DDIpTDADTDAE), Cdc7 (DGESpTDEDDVVS), and Cdc45 (DDEApTDADEVTD) phosphopeptides measured by isothermal titration calorimetry

Peptide	Stoichiometry	K_d	ΔH	ΔG	$-T\Delta S$
		μM	kcal/mol	kcal/mol	kcal/mol
pT-Ptc2	1	0.7 ± 0.01	−21.5 ± 0.05	−8.5	−13.0
pT-Cdc7	1	1.69 ± 0.04	−20.6 ± 0.10	−8.0	−12.6
pT-Cdc45	0.97	0.42 ± 0.02	−20.7 ± 0.10	−8.8	−11.9

270)T189A fusion (data not shown). These results were also confirmed by the fact that the binding of the TAP-tagged Cdc45 protein to Rad53 FHA1 in the GST pull-down assay was strongly reduced by the T189A mutation (Fig. 3*D*). This observation was corroborated by a co-immunoprecipitation assay (Fig. 3*E*), which showed that the strength of the interaction between full-length Rad53 and TAP-tagged Cdc45 severely drops upon introduction of the T189A mutation. All in all, our results indicate that Cdc45 threonine T189 represents a ligand of FHA1 and validate our *in silico* predictions.

We would again have reached a similar conclusion considering the full-length Cdc45 rather than the Cdc45(154–270) fragment because T189 was one of the two highest scoring residues of all Cdc45 threonines. For the other residue, T147, close inspection of the alignment showed that the nonconservation of the TxxD motif in *S. kluyveri* could not be explained by alignment flaws or sequence truncations as for T189 (Supplemental data Fig. 1).

Rad53 FHA1 Binds In Vitro to Phosphopeptides Encompassing Cdc7 T484 and Cdc45 T189—In order to confirm the interactions between Rad53 FHA1 domain and Cdc7 and Cdc45, the direct binding of FHA1 to the phosphothreonine peptides [480]DGESpTDEDDVVS [pT(Cdc7)] and [185]DDEApTDADEVTD [pT(Cdc45)] derived from the Cdc7 and Cdc45 sequences, respectively, was probed using isothermal titration calorimetry. The dissociation constant, K_D, between Rad53 FHA1 and the pT(Cdc7) and the pT(Cdc45) reached 1.7 μM and 400 nM, respectively (Table II and Supplemental data Fig. 2). pT(Cdc45) is the peptide with the highest affinity described so far for an FHA1 substrate (the peptide corresponding to the near-optimal binding motif determined for Rad53 FHA1 by peptide library screening bound with a K_D of 780 nM). These data clearly indicate that Rad53 FHA1 interacts directly with peptides encompassing Cdc7 T484 and Cdc45 T189 *in vitro* and support our hypothesis that similar, direct interactions occur *in vivo* between FHA1 and Cdc7 and Cdc45.

Mutating Cdc7 T484 and Cdc45 T189 Induces No Obvious Replication Phenotype—Having identified Cdc7 T484 and Cdc45 T189 as probable ligands of FHA1, we sought to assess the part played by the FHA1/Cdc7 and FHA1/Cdc45 interactions by abrogating specifically these interactions via the mutation of Cdc7 T484 and Cdc45 T189 into alanine. We constructed yeast strains deleted for either *CDC7* or *CDC45* at their chromosomal loci and complemented with plasmids harboring either a wild-type or a mutated copy of the corresponding gene (*cdc7T484A* and *cdc45T189A*, respectively). The strains were tested for their growth on solid medium in the presence or in the absence of various genotoxic stresses including UV-irradiation, camptothecin, hydroxyurea, and 4-nitroquinoline 1-oxide (a reagent that produces bulky base damage of the type that is mainly repaired by the nucleotide excision repair system). No reproducible difference of viability or growth rate could be observed between the mutated *cdc7T484A* and *cdc45T189A* cells and the controls (data not shown). The double mutant *cdc7T484A cdc45T189A* also behaved as wild-type cells (data not shown). These results can be explained by the redundancy of interactions linking two proteins or even two complexes via different protein-protein interactions (48). Regarding Cdc45, the interaction between Rad53 FHA1 and the Cdc45T189A could be indirectly maintained in the prereplication complex through other Rad53 partners such as Mrc1 (18) and Cdc46 (Mcm5) (49). In the case of Cdc7, an interaction between FHA1 and Cdc7T484A could be indirectly maintained via other proteins such as Dbf4, the regulatory subunit of the Cdc7/Dbf4 kinase complex, also described as a Rad53 FHA1 binding partner (24).

Application of STRIP Strategy to a Complex and Large interactome—Cdc7 and Cdc45 represent two examples for which our strategy correctly determined the phosphothreonines targeted by FHA1 (as monitored by the complete or partial loss of interaction in several *in vivo* and *in vitro* assays). We had previously demonstrated experimentally that FHA1 binds the threonine T376 of the PP2C phosphatase Ptc2, which plays a part in Rad53 inactivation following double-strand breaks (26). We tested whether we could have predicted this site with the STRIP strategy and found that indeed T376 is the only Ptc2 threonine fulfilling the three criteria of our test. In this case, we had demonstrated that the T376A mutation not only abolishes the interaction between Ptc2 and Rad53 FHA1 but also induces a clear phenotype in terms of adaptation defects (26). Threonines bound by FHA1 are expected in many cases to be phosphorylated by CK2 because CK2 substrate consensus site (50) resembles the optimal binding motif of FHA1. In contrast to the interaction between Rad53 FHA1 and Ptc2 (26), the interactions between Rad53 FHA1 and Cdc7 and Cdc45 observed in the two-hybrid assay were not affected by the deletion of the genes encoding the regulatory subunits of CK2 (data not shown).

FIG. 5. **Screen capture of the STRIP web server with the query page on the left and the result page on the right.** A dark blue color indicates that the corresponding phosphoresidue satisfies either the phosphorylability, the consensus motif rule or the strict conservation condition. Light blue cells indicate that the conservation condition is only respected with the less stringent condition and in that case, further manual analysis may be performed easily. By clicking in the table cell of any phosphoresidue, a pop-up window allows for rapidly checking the conservation features between *S. cerevisiae* close homologs, detecting possible sequence truncations or identifying alignment flaws.

To further challenge the interest of the STRIP strategy in facilitating the dissection of large interactomes, we analyzed the whole set of physical interactions involving Rad53, derived either from the present work or from the literature. Several experimental works were devoted to unravel Rad53 binding partners using either affinity-based or two hybrid-based methods (18, 25, 41, 51). Several large scale yeast interactome analyses also provided a wealth of data connected to Rad53 (49, 52–54). The graph in Fig. 4 reports the 63 Rad53 binding partners extracted from this work and the Biogrid database (version of July 2008) (55) using the Osprey visualization tool (56). Blue and red linkages report for the affinity-based and the two-hybrid results, respectively. The 11 proteins identified from our FHA1 two-hybrid screen are labeled by an obelisk (‡) in Fig. 4 and in Table III. Among the affinity-based results, 30 binding proteins (labeled by an asterisk in Fig. 4 and in Table III) were identified in a proteomic survey that focused on the isolated FHA1 domain partners (18). These interactions were lost upon point mutation of the

phosphobinding site in the FHA1 domain, confirming that a phosphothreonine is mediating the interaction. For the remaining proteins, it is not known whether the FHA1, the FHA2, or another region of Rad53 is involved in the interaction. In the following, we limited our survey to the existence of putative threonines that may be recognized by FHA1 and explored how the STRIP strategy may restrict the numbers of putative binding sites.

All in all, there are 2922 threonines in the 63 binding proteins. Applying our protocol led to a unique candidate threonine for 25 out of 63 partners (Table III). For 11 additional cases, a limited set of two to three threonines could be proposed. For the remaining 27 proteins, no threonines fulfilled the set of constraints applied on the sequence of the binding partners. These partners could bind the FHA2 or another region of Rad53, or could bind Rad53 indirectly via the intermediate of other Rad53 bridging partners. Indirect interactions concern complexes bearing many cross interactions such as the septin complex (containing Cdc3/10/11/12,

TABLE III

STRIP-based prediction of the threonines possibly involved in the interaction with the FHA1 domain of Rad53. (‡) gene names of the proteins identified in the FHA1 two- hybrid screen performed in this work. Predictions were carried out using the full-length proteins. (#) predicted threonines that are comprised within the fragment identified in the two-hybrid screen performed in this work. (*) proteins that were identified in the affinity-based screen by Smolka and co-workers using the isolated FHA1 as a bait (18)

	Number of Threonines	Number of TxxD motifs	Threonines strict conditions	Threonines permissive conditions
End3‡	11	0		
Rog3‡	33	2		T453#
Snu71‡	29	3	T269#	
Gea1‡	76	4		T272#
Yap7‡	10	0		
Vps30‡	36	6	T434	T130#; T257
Ppz1‡	38	2	T171#	
Yhr202w‡	38	1	T504#	
Cdc7‡	24	5	T484	
Dbf4	68	5		T247; T253
Mrc1*	69	7		T242; T272; T977
Cdc45‡	33	5		T147; T189
Cdc46	49	1		
Rad9*	89	5		
Ptc2*‡	32	4	T376#	
Dun1	26	0		
Mec1	131	10		
Hhf1	6	1		
Hht1	9	0		
Hta2	5	0		
Hmo1	13	0		
Asf1*	9	1	T270	
Swi6*	41	2		T111
Mbp1*	63	4		
Swi4*	70	3		T64
Whi5*	39	0		
Kap95	44	1		
Srp1/Kap60	32	2		T273
Gln3*	49	2		
Ifh1*	55	7	T346	
Tbf1	42	1		T522
Cdc13	54	1		
Cst6*	48	2	T534	
Esc1*	98	2		
Sgs1	103	3	T423	
Mus81	81	2		
Rad55	22	1		
Crp1*	38	2		T170
Src1*	36	3		T399
Ecm16*	73	5	T869; T1169	
Net1*	82	2		
Yta7*	70	7	T247; T946; T1077	
Psy2	59	7		T12; T178; T524
Smc3	66	5	T18	
Cdc3*	18	1		
Cdc12*	27	0		
Shs1*	36	2		T539
Cdc11*	30	1		T62
Cdc10*	24	1		
Bud4*	88	8		T178; T237; T612

TABLE III—*continued*

	Number of Threonines	Number of TxxD motifs	Threonines strict conditions	Threonines permissive conditions
Bud3*	106	7	T255	
Bud14*	36	2		
Bmh2	15	1		T4
Sec2*	52	5		T423
Ubp1*	37	4		T592
Rck2*	39	4		T342; T539
Fyv8*	50	7	T602; T608	
Mnr2*	52	5		
Ipp1	19	2	T100; T247	
Ckb2	10	0		
Ede1	118	3		T948
Pin4	39	0		
Gid8	27	0		

Shs1, and Bud4), the G1/S transition complex (made of Swi4, Swi6, Mbp1, and Whi5), or the histone complex (Hht1, Hhf1, Hta2, and Hmo1), for which only 5 out of 14 partners contain a candidate threonine. We asked whether the STRIP strategy could help identify in these stable complexes the most likely direct partners.

For the septin complex (Cdc3/10/11/12, Shs1, and Bud4 network), which was found to interact with the isolated FHA1 domain (18), three proteins (Shs1, Bud4, and Cdc11, with dashed underline in Fig. 4) were detected as harboring five putative FHA1 binding threonines using the relaxed conservation criterion (none were found with the stringent one (see Methods)). As for Cdc45, a rapid inspection of the alignment around these five threonines (Supplemental data Fig. 3) revealed that the three threonines in Bud4 diverged much more extensively than the two threonines in Cdc11 and Shs1, leading to the proposition that FHA1 may rather interact with Cdc11 T62 and Shs1 T539. Shs1 T539 was experimentally found phosphorylated in three independent proteomic studies (57–59). Moreover, Shs1 is phosphorylated by Rad53 *in vitro* (18) and appears as the only member of the septin complex to undergo a Rad53-dependent phosphorylation following treatment with methyl methane sulfonate (59). However, the deletion of *SHS1* was not found to abrogate Cdc11 interaction with Rad53 FHA1 (18) leaving the possibility that both Cdc11 and Shs1 sites contribute to Rad53 binding in a redundant manner.

For the G1/S transition complex (the Swi4/Swi6/Mbp1/Whi5 network) only two threonines, T64 in Swi4 and T111 in Swi6, were detected as potential binding sites, again with the relaxed conservation criterion. We can notice that FHA1 optimal binding site is more conserved around Swi6 T111 than around Swi4 T64 (Supplemental data Fig. 4). We monitored the binding of a Mbp1-TAP containing complex to immobilized GST-FHA1 and found that Swi6 is required for FHA1 binding to Mbp1-TAP whereas Swi4 is not (Supplemental data Fig. 5). Furthermore, Swi6 was shown to be directly phosphorylated at residue S547 by Rad53, impacting on the delay of the G1/S transition following DNA damage (60). The STRIP analysis together with these experiments are thus compatible with a direct interaction between Swi6 and Rad53 mediated by Swi6 phosphothreonine T111.

Other interesting features of Fig. 4 coupled to the STRIP predictions are that the threonines T346 and T247, which are likely to be bound by FHA1 in Ifh1 and Yta7, respectively, are also predicted to be phosphorylated by the CK2 kinase according to the NetPhosK server (35). Interestingly, both proteins were found to associate with the Ckb2 CK2 regulatory subunit (53, 61).

<center>DISCUSSION</center>

With the development of large scale phospho-proteomic experiments, several methods have been developed to analyze and predict the phosphorylation patterns in protein substrates. The widespread role of linear interaction motifs in protein-protein interactions and methodological breakthroughs in mass spectrometry methodology have prompted the development of several databases with special interest for phosphorylated sites (Phospho.ELM (6), PhosphoSite (62), Phosida (63), and PhosphoPep (64)). Other databases focus on the domains specialized in linear motif recognition (DOMINO (5) and ADAN (65)) and also combine both motif and binding domain information as in NetPhorest (66). The development of predictive algorithms to identify these sites from protein sequences has taken advantage of this mass of data. Several machine learning algorithms either dedicated to specific classes of kinases (NetPhosK (35), PredPhospho (36), PPSP (37), ScanSite (38), KinasePhos (39), and NetPhorest (66)) or with a broader scope (NetPhos2.0 (30), DisPhos1.3 (31)) have been proposed. Combining several of these approaches could lead to improved prediction rate (67). Many cellular factors such as localization, scaffolds, or expression also play important roles in determining the fate of kinase substrates. To cope with such level of complexity at the cell network level, NetworKIN (68) integrates these heterogeneous sets of data to help modeling phosphorylation networks and improve prediction specificity. Functionally important interaction sites can be identified through conservation analyses and small conserved motifs could be assigned as in the eMOTIF database (69) or through the PhosphoBlast program (70). However, high evolutionary rates for the regions containing these linear motifs may hamper their recognition across remotely related species (47, 71).

The STRIP approach that we devised was inspired by these works to provide a user-friendly interface helping the design of interaction mutants. We found that using conservation data from closely related species combined with the knowledge from the bait (phospho-binding modules with a specific consensus binding motif) and the preys (phosphorylation likelihood and conservation of the consensus motif) could greatly help to decipher the phosphorylated sites. The STRIP server in its user-friendly form can help to provide important clues to guide the dissection of interaction networks mediated by phosphobinding protein modules. Fig. 2 and the number of TxxD motifs in Table III illustrate that prediction of the interaction sites within FHA1 binding partners solely on the basis of the search for the motif bound with the highest affinity would lead to many more threonine candidates. In the case of the FHA1 domain, the binding specificity is known from targeted experiments but recent computational approaches, such as D-MIST, suggest that these specificities may shortly be inferred from prediction (72).

We validated the STRIP strategy by the identification of the threonines bound by Rad53 FHA1 domain. The analysis of FHA1 ligands was also motivated by the search of the FHA1 interactions involved in resistance to replicational stress. In that respect, the two replication proteins Cdc7 and Cdc45 appeared as possible candidates for Rad53 targets. Using the two-hybrid strategy, we found that Rad53 FHA1 interacts constitutively with Cdc7 and Cdc45 peptides in the presence as in the absence of genotoxic stress. These results were confirmed with GST pull-down assays, which showed that the full-length Cdc7 and Cdc45 proteins can interact with Rad53 FHA1 in G1 and in S phase, with or without treatment by genotoxic agents. Co-IP experiments with full-length Rad53 confirmed that the complexes can readily be formed *in vivo*. The decrease in interaction strength observed with the point mutated Cdc7 and Cdc45 further indicates that these interactions can be regulated by the phosphosites we predicted. These data suggest that Rad53 could interact constitutively with Cdc7 and Cdc45 irrespective of the phase of the cell cycle and of the presence of DNA damage or stalled replication forks. The Cdc7/Dbf4 kinase is considered to trigger the replication origins but not to remain associated with the elongating forks. In contrast, Cdc45 is required for both the firing of replication origins and replication elongation. In case of replicative stress, the constitutive Rad53/Cdc7 and Rad53/Cdc45 complexes could be recruited to replication origins, or Rad53/Cdc45 could replace Cdc45 in stalled replication forks because Rad53 has not been detected as a component of the replication machinery during a normal S phase (73).

To estimate the phosphorylation likelihood, the STRIP strategy relies on a meta-prediction approach combining two different algorithms NetPhos and DisPhos. However, the precision of these approaches may still be questionable and somehow the conservation of the phosphoresidue strengthens or decreases the reliability of the phosphorylation prediction. However, a characteristic feature of the phosphorylated regions is that they are often located in disordered regions that may turn out to be tricky to align properly. Moreover, the simple linear organization of the binding motifs may allow them to shift along the sequence during evolution without compromising the binding. The functional importance of these linear motifs recently prompted the development of specific alignment algorithms and dedicated benchmarks (74, 75). Cdc45 test case clearly illustrates how alignment pitfalls may hinder proper binding site prediction even with as closely related species as those of the *Saccharomyces* genus. Our large-scale analysis of Rad53 partners shows that inspection of the alignment in the vicinity of the phospho-residue may
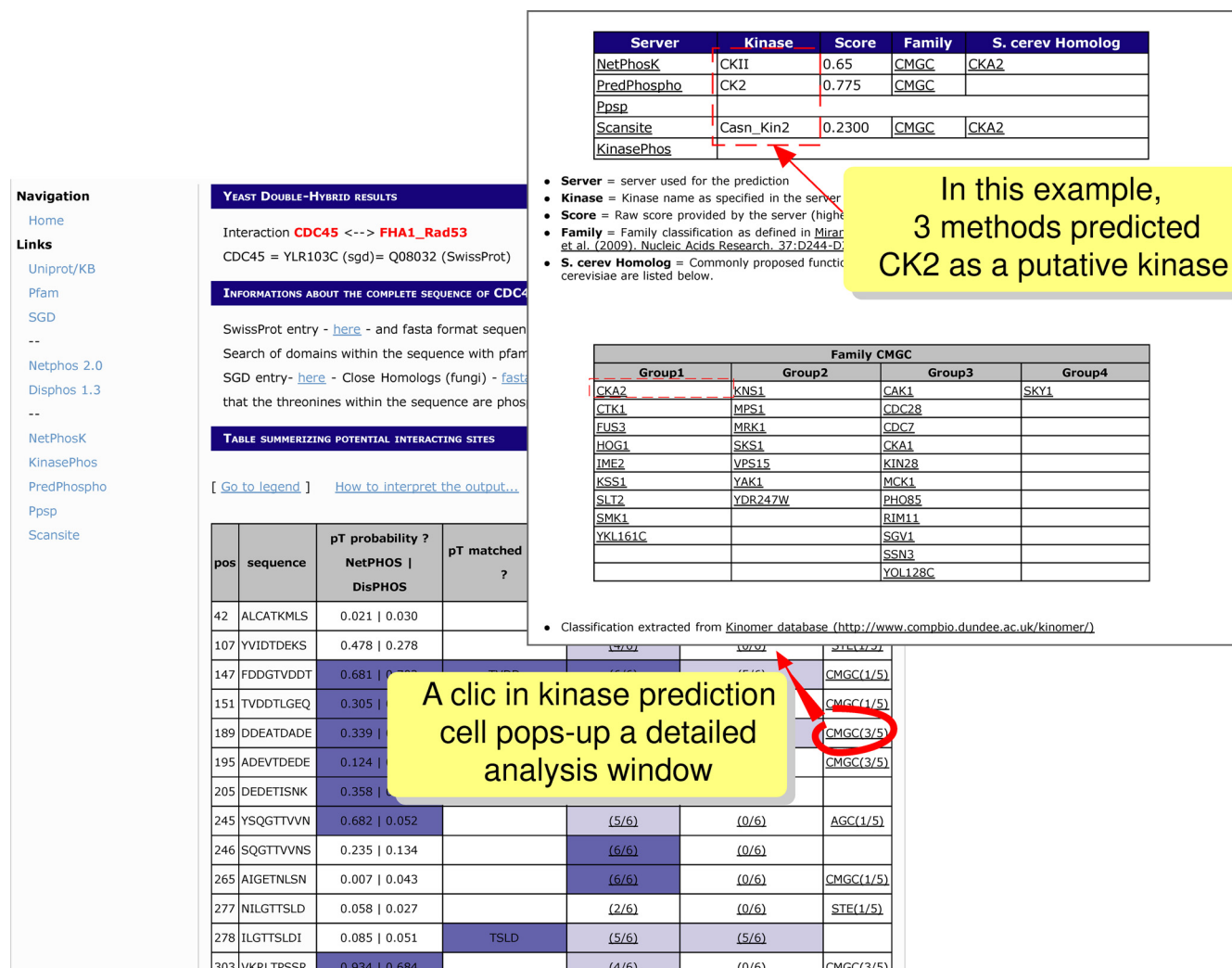
FIG. 6. **Screen capture of the STRIP web server with the prediction of likely kinases according to five different methods, namely NetPhosK (34), PredPhospho (35), PPSP (36), ScanSite (37) and KinasePhos (38).** In the main page the class name of the predicted kinase together with the number of servers that converged on the same prediction is provided. By clicking in the table cell of this consensus prediction, a pop-up window provides a detailed analysis of the predictions. In addition, all the *S. cerevisiae* kinases classified in the same class as the predicted kinase (82) are mentioned in the tables below.

provide crucial hints to rescore the binding sites. The STRIP web server facilitates such analysis by allowing the user to analyze around each putative phospho-residue a fragment of the multiple sequence alignment with different sequence highlights (Fig. 5 and 6). To date, the STRIP server is dealing with *Saccharomyces* datasets and will further progress by integrating data for plants and mammals.

One major question raised by the example of the FHA1 domain is whether the motif identified through peptide library screening as bound with the highest affinity is really useful to predict FHA1 binding motifs *in vivo*. In two well-studied interactions, Rad53 FHA1 was found to recognize its partners Rad9 and Pin4/Mdt1 through threonines within TxxV or TxxI motifs, respectively (20, 23, 76). Our analysis restores the reliability of the consensus motif analysis showing that it could guide efficiently the predictions for the Ptc2, Cdc7 and Cdc45 examples.

The presence of the pTxxD consensus motif significantly contributes to reach affinities in the range 0.5–1 μM, whereas the affinities of the pT motifs studied in Rad9 and Mdt1 were an order of magnitude lower. A specificity of Rad9 and Mdt1 is to be hyperphosphorylated upon genotoxic stress by the phosphatidyl inositol kinase-like kinases Tel1 and Mec1 on their SQ/TQ-rich clusters. In the case of Rad9, and probably Mdt1 also, these clusters are essential in mediating the interaction with Rad53 through its FHA domains. A possibility is that Rad53 association with these targets requires additional (direct or indirect) interactions separate from the phosphopeptide binding site or that polyvalent ligands such as the SQ/TQ-rich clusters have an increased affinity for FHA domains, which alleviates the stringency on the consensus motif (77, 78).

The molecular logic underlying phosphoproteome organization will surely benefit from the development of STRIP-like

strategies. Dissection of the intricate network of interactions between the components of cell signaling systems and/or cell machineries is all the more difficult that the redundancy of their contacts makes the role of each interaction difficult to analyze (79). Systematic prediction of the contacting sites is expected to help overcome these issues. Furthermore, competitive or synergistic interactions between interacting modules may be further predicted from the identification of the precise binding sites. A growing list of proteins involved in key signaling processes also demonstrate that alternative post-translational modifications such as acetylations or methylations may synergize with the phosphorylation of a particular site to implement complex regulatory signals (80, 81). These proteins are under specific focus but such level of complexity may be widespread and the development of predictive strategies to isolate a limited number of putative binding sites between proteins should have a major impact on the global understanding of cell components cross-talks.

REFERENCES

1. Seet, B. T., Dikic, I., Zhou, M. M., and Pawson, T. (2006) Reading protein modifications with interaction domains. *Nat. Rev. Mol. Cell Biol.* **7,** 473–483

2. Bhattacharyya, R. P., Remenyi, A., Yeh, B. J., and Lim, W. A. (2006) Domains, Motifs, and Scaffolds: The Role of Modular Interactions in the Evolution and Wiring of Cell Signaling Circuits. *Annu. Rev. Biochem.*

3. Pawson, T., and Nash, P. (2003) Assembly of cell regulatory systems through protein interaction domains. *Science* **300,** 445–452

4. Gong, W., Zhou, D., Ren, Y., Wang, Y., Zuo, Z., Shen, Y., Xiao, F., Zhu, Q., Hong, A., Zhou, X., Gao, X., and Li, T. (2008) PepCyber:P~PEP: a database of human protein protein interactions mediated by phospho-protein-binding domains. *Nucleic Acids Res.* **36,** D679–683

5. Ceol, A., Chatr-aryamontri, A., Santonico, E., Sacco, R., Castagnoli, L., and Cesareni, G. (2007) DOMINO: a database of domain-peptide interactions. *Nucleic Acids Res.* **35,** D557–560

6. Diella, F., Gould, C. M., Chica, C., Via, A., and Gibson, T. J. (2008) Phospho.ELM: a database of phosphorylation sites–update 2008. *Nucleic Acids Res.* **36,** D240–244

7. Hofmann, K., and Bucher, P. (1995) The FHA domain: a putative nuclear signalling domain found in protein kinases and transcription factors. *Trends Biochem. Sci.* **20,** 347–349

8. Liao, H., Yuan, C., Su, M. I., Yongkiettrakul, S., Qin, D., Li, H., Byeon, I. J., Pei, D., and Tsai, M. D. (2000) Structure of the FHA1 domain of yeast Rad53 and identification of binding sites for both FHA1 and its target

9. protein Rad9. *J. Mol. Biol.* **304,** 941–951

9. Durocher, D., Taylor, I. A., Sarbassova, D., Haire, L. F., Westcott, S. L., Jackson, S. P., Smerdon, S. J., and Yaffe, M. B. (2000) The molecular basis of FHA domain:phosphopeptide binding specificity and implications for phospho-dependent signaling mechanisms. *Mol. Cell* **6,** 1169–1182

10. Durocher, D., Henckel, J., Fersht, A. R., and Jackson, S. P. (1999) The FHA domain is a modular phosphopeptide recognition motif. *Mol. Cell* **4,** 387–394

11. Li, J., Williams, B. L., Haire, L. F., Goldberg, M., Wilker, E., Durocher, D., Yaffe, M. B., Jackson, S. P., and Smerdon, S. J. (2002) Structural and functional versatility of the FHA domain in DNA-damage signaling by the tumor suppressor kinase Chk2. *Mol. Cell* **9,** 1045–1054

12. Huen, M. S., Grant, R., Manke, I., Minn, K., Yu, X., Yaffe, M. B., and Chen, J. (2007) RNF8 transduces the DNA-damage signal via histone ubiquitylation and checkpoint protein assembly. *Cell* **131,** 901–914

13. Lee, H., Yuan, C., Hammet, A., Mahajan, A., Chen, E. S., Wu, M. R., Su, M. I., Heierhorst, J., and Tsai, M. D. (2008) Diphosphothreonine-specific interaction between an SQ/TQ cluster and an FHA domain in the Rad53-Dun1 kinase cascade. *Mol. Cell* **30,** 767–778

14. Koch, C. A., Agyei, R., Galicia, S., Metalnikov, P., O'Donnell, P., Starostine, A., Weinfeld, M., and Durocher, D. (2004) Xrcc4 physically links DNA end processing by polynucleotide kinase to DNA ligation by DNA ligase IV. *EMBO J.* **23,** 3874–3885

15. Byeon, I. J., Li, H., Song, H., Gronenborn, A. M., and Tsai, M. D. (2005) Sequential phosphorylation and multisite interactions characterize specific target recognition by the FHA domain of Ki67. *Nat. Struct. Mol. Biol.* **12,** 987–993

16. Yuan, C., Yongkiettrakul, S., Byeon, I. J., Zhou, S., and Tsai, M. D. (2001) Solution structures of two FHA1-phosphothreonine peptide complexes provide insight into the structural basis of the ligand specificity of FHA1 from yeast Rad53. *J. Mol. Biol.* **314,** 563–575

17. Tam, A. T., Pike, B. L., and Heierhorst, J. (2008) Location-specific functions of the two forkhead-associated domains in Rad53 checkpoint kinase signaling. *Biochemistry* **47,** 3912–3916

18. Smolka, M. B., Chen, S. H., Maddox, P. S., Enserink, J. M., Albuquerque, C. P., Wei, X. X., Desai, A., Kolodner, R. D., and Zhou, H. (2006) An FHA domain-mediated protein interaction network of Rad53 reveals its role in polarized cell growth. *J. Cell Biol.* **175,** 743–753

19. Schwartz, M. F., Lee, S. J., Duong, J. K., Eminaga, S., and Stern, D. F. (2003) FHA domain-mediated DNA checkpoint regulation of Rad53. *Cell Cycle* **2,** 384–396

20. Pike, B. L., Yongkiettrakul, S., Tsai, M. D., and Heierhorst, J. (2004) Mdt1, a novel Rad53 FHA1 domain-interacting protein, modulates DNA damage tolerance and G(2)/M cell cycle progression in Saccharomyces cerevisiae. *Mol. Cell. Biol.* **24,** 2779–2788

21. Pike, B. L., Yongkiettrakul, S., Tsai, M. D., and Heierhorst, J. (2003) Diverse but overlapping functions of the two forkhead-associated (FHA) domains in Rad53 checkpoint kinase activation. *J. Biol. Chem.* **278,** 30421–30424

22. Pike, B. L., Hammet, A., and Heierhorst, J. (2001) Role of the N-terminal forkhead-associated domain in the cell cycle checkpoint function of the Rad53 kinase. *J. Biol. Chem.* **276,** 14019–14026

23. Mahajan, A., Yuan, C., Pike, B. L., Heierhorst, J., Chang, C. F., and Tsai, M. D. (2005) FHA domain-ligand interactions: importance of integrating chemical and biological approaches. *J. Am. Chem. Soc.* **127,** 14572–14573

24. Duncker, B. P., Shimada, K., Tsai-Pflugfelder, M., Pasero, P., and Gasser, S. M. (2002) An N-terminal domain of Dbf4p mediates interaction with both origin recognition complex (ORC) and Rad53p and can deregulate late origin firing. *Proc. Natl. Acad. Sci. U.S.A.* **99,** 16087–16092

25. Bjergbaek, L., Cobb, J. A., Tsai-Pflugfelder, M., and Gasser, S. M. (2005) Mechanistically distinct roles for Sgs1p in checkpoint activation and replication fork maintenance. *EMBO J.* **24,** 405–417

26. Guillemain, G., Ma, E., Mauger, S., Miron, S., Thai, R., Guérois, R., Ochsenbein, F., and Marsolier-Kergoat, M. C. (2007) Mechanisms of checkpoint kinase Rad53 inactivation after a double-strand break in Saccharomyces cerevisiae. *Mol. Cell. Biol.* **27,** 3378–3389

27. Sikorski, R. S., and Hieter, P. (1989) A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in Saccharomyces cerevisiae. *Genetics* **122,** 19–27

28. Allen, J. B., Zhou, Z., Siede, W., Friedberg, E. C., and Elledge, S. J. (1994)

The SAD1/RAD53 protein kinase controls multiple checkpoints and DNA damage-induced transcription in yeast. *Genes Dev.* **8,** 2401–2415

29. Fromont-Racine, M., Rain, J. C., and Legrain, P. (2002) Building protein-protein networks by two-hybrid mating strategy. *Methods Enzymol.* **350,** 513–524

30. Blom, N., Gammeltoft, S., and Brunak, S. (1999) Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. *J. Mol. Biol.* **294,** 1351–1362

31. Iakoucheva, L. M., Radivojac, P., Brown, C. J., O'Connor, T. R., Sikes, J. G., Obradovic, Z., and Dunker, A. K. (2004) The importance of intrinsic disorder for protein phosphorylation. *Nucleic Acids Res.* **32,** 1037–1049

32. Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22,** 4673–4680

33. Cliften, P., Sudarsanam, P., Desikan, A., Fulton, L., Fulton, B., Majors, J., Waterston, R., Cohen, B. A., and Johnston, M. (2003) Finding functional features in Saccharomyces genomes by phylogenetic footprinting. *Science* **301,** 71–76

34. Kellis, M., Patterson, N., Endrizzi, M., Birren, B., and Lander, E. S. (2003) Sequencing and comparison of yeast species to identify genes and regulatory elements. *Nature* **423,** 241–254

35. Blom, N., Sicheritz-Pontén, T., Gupta, R., Gammeltoft, S., and Brunak, S. (2004) Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence. *Proteomics* **4,** 1633–1649

36. Kim, J. H., Lee, J., Oh, B., Kimm, K., and Koh, I. (2004) Prediction of phosphorylation sites using SVMs. *Bioinformatics* **20,** 3179–3184

37. Xue, Y., Li, A., Wang, L., Feng, H., and Yao, X. (2006) PPSP: prediction of PK-specific phosphorylation site with Bayesian decision theory. *BMC Bioinformatics* **7,** 163

38. 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

39. Huang, H. D., Lee, T. Y., Tzeng, S. W., and Horng, J. T. (2005) KinasePhos: a web tool for identifying protein kinase-specific phosphorylation sites. *Nucleic Acids Res.* **33,** W226–229

40. Wach, A., Brachat, A., Pöhlmann, R., and Philippsen, P. (1994) New heterologous modules for classical or PCR-based gene disruptions in Saccharomyces cerevisiae. *Yeast* **10,** 1793–1808

41. Leroy, C., Lee, S. E., Vaze, M. B., Ochsenbien, F., Guerois, R., Haber, J. E., and Marsolier-Kergoat, M. C. (2003) PP2C phosphatases Ptc2 and Ptc3 are required for DNA checkpoint inactivation after a double-strand break. *Mol. Cell* **11,** 827–835

42. Kihara, M., Nakai, W., Asano, S., Suzuki, A., Kitada, K., Kawasaki, Y., Johnston, L. H., and Sugino, A. (2000) Characterization of the yeast Cdc7p/Dbf4p complex purified from insect cells. Its protein kinase activity is regulated by Rad53p. *J. Biol. Chem.* **275,** 35051–35062

43. Masai, H., and Arai, K. (2002) Cdc7 kinase complex: a key regulator in the initiation of DNA replication. *J. Cell. Physiol.* **190,** 287–296

44. Zou, L., Mitchell, J., and Stillman, B. (1997) CDC45, a novel yeast gene that functions with the origin recognition complex and Mcm proteins in initiation of DNA replication. *Mol. Cell. Biol.* **17,** 553–563

45. Tercero, J. A., Labib, K., and Diffley, J. F. (2000) DNA synthesis at individual replication forks requires the essential initiation factor Cdc45p. *EMBO J.* **19,** 2082–2093

46. Xue, Y., Zhou, F., Zhu, M., Ahmed, K., Chen, G., and Yao, X. (2005) GPS: a comprehensive www server for phosphorylation sites prediction. *Nucleic Acids Res.* **33,** W184–187

47. Beltrao, P., and Serrano, L. (2007) Specificity and evolvability in eukaryotic protein interaction networks. *PLoS Comput. Biol.* **3,** e25

48. Chi, A., Huttenhower, C., Geer, L. Y., Coon, J. J., Syka, J. E., Bai, D. L., Shabanowitz, J., Burke, D. J., Troyanskaya, O. G., and Hunt, D. F. (2007) Analysis of phosphorylation sites on proteins from Saccharomyces cerevisiae by electron transfer dissociation (ETD) mass spectrometry. *Proc. Natl. Acad. Sci. U.S.A.* **104,** 2193–2198

49. Gavin, A. C., Aloy, P., Grandi, P., Krause, R., Boesche, M., Marzioch, M., Rau, C., Jensen, L. J., Bastuck, S., Dumpelfeld, B., Edelmann, A., Heurtier, M. A., *et al.* (2006) Proteome survey reveals modularity of the yeast cell machinery. *Nature* **440,** 631–636

50. Meggio, F., and Pinna, L. A. (2003) One-thousand-and-one substrates of protein kinase CK2? *Faseb J.* **17,** 349–368

51. Herzberg, K., Bashkirov, V. I., Rolfsmeier, M., Haghnazari, E., McDonald, W. H., Anderson, S., Bashkirova, E. V., Yates, J. R., 3rd, and Heyer, W. D. (2006) Phosphorylation of Rad55 on serines 2, 8, and 14 is required for efficient homologous recombination in the recovery of stalled replication forks. *Mol. Cell. Biol.* **26,** 8396–8409

52. Ho, Y., Gruhler, A., Heilbut, A., Bader, G. D., Moore, L., Adams, S. L., Millar, A., Taylor, P., Bennett, K., Boutilier, K., Yang, L., Wolting, C., *et al.* (2002) Systematic identification of protein complexes in Saccharomyces cerevisiae by mass spectrometry. *Nature* **415,** 180–183

53. Gavin, A. C., Bosche, M., Krause, R., Grandi, P., Marzioch, M., Bauer, A., Schultz, J., Rick, J. M., Michon, A. M., Cruciat, C. M., Remor, M., Hofert, C., *et al.* (2002) Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* **415,** 141–147

54. Krogan, N. J., Cagney, G., Yu, H., Zhong, G., Guo, X., Ignatchenko, A., Li, J., Pu, S., Datta, N., Tikuisis, A. P., Punna, T., Peregrin-Alvarez, J. M. *et al.* (2006) Global landscape of protein complexes in the yeast Saccharomyces cerevisiae. *Nature* **440,** 637–643

55. Stark, C., Breitkreutz, B. J., Reguly, T., Boucher, L., Breitkreutz, A., and Tyers, M. (2006) BioGRID: a general repository for interaction datasets. *Nucleic Acids Res.* **34,** D535–539

56. Breitkreutz, B. J., Stark, C., and Tyers, M. (2003) Osprey: a network visualization system. *Genome Biol.* **4,** R22

57. Albuquerque, C. P., Smolka, M. B., Payne, S. H., Bafna, V., Eng, J., and Zhou, H. (2008) A multidimensional chromatography technology for in-depth phosphoproteome analysis. *Mol. Cell Proteomics* **7,** 1389–1396

58. Gruhler, A., Olsen, J. V., Mohammed, S., Mortensen, P., Faergeman, N. J., Mann, M., and Jensen, O. N. (2005) Quantitative phosphoproteomics applied to the yeast pheromone signaling pathway. *Mol. Cell Proteomics* **4,** 310–327

59. Smolka, M. B., Albuquerque, C. P., Chen, S. H., and Zhou, H. (2007) Proteome-wide identification of in vivo targets of DNA damage checkpoint kinases. *Proc. Natl. Acad. Sci. U.S.A.* **104,** 10364–10369

60. Sidorova, J. M., and Breeden, L. L. (2003) Rad53 checkpoint kinase phosphorylation site preference identified in the Swi6 protein of Saccharomyces cerevisiae. *Mol. Cell. Biol.* **23,** 3405–3416

61. Rudra, D., Mallick, J., Zhao, Y., and Warner, J. R. (2007) Potential interface between ribosomal protein production and pre-rRNA processing. *Mol. Cell. Biol.* **27,** 4815–4824

62. Hornbeck, P. V., Chabra, I., Kornhauser, J. M., Skrzypek, E., and Zhang, B. (2004) PhosphoSite: A bioinformatics resource dedicated to physiological protein phosphorylation. *Proteomics* **4,** 1551–1561

63. Gnad, F., Ren, S., Cox, J., Olsen, J. V., Macek, B., Oroshi, M., and Mann, M. (2007) PHOSIDA (phosphorylation site database): management, structural and evolutionary investigation, and prediction of phosphosites. *Genome Biol.* **8,** R250

64. Bodenmiller, B., Campbell, D., Gerrits, B., Lam, H., Jovanovic, M., Picotti, P., Schlapbach, R., and Aebersold, R. (2008) PhosphoPep–a database of protein phosphorylation sites in model organisms. *Nat. Biotechnol.* **26,** 1339–1340

65. Encinar, J. A., Fernandez-Ballester, G., Sánchez, I. E., Hurtado-Gomez, E., Stricher, F., Beltrao, P., and Serrano, L. (2009) ADAN: a database for prediction of protein-protein interaction of modular domains mediated by linear motifs. *Bioinformatics* **25,** 2418–2424

66. Miller, M. L., Jensen, L. J., Diella, F., Jorgensen, C., Tinti, M., Li, L., Hsiung, M., Parker, S. A., Bordeaux, J., Sicheritz-Ponten, T., Olhovsky, M., Pasculescu, A., *et al.* (2008) Linear motif atlas for phosphorylation-dependent signaling. *Sci. Signal* 1, ra2

67. Wan, J., Kang, S., Tang, C., Yan, J., Ren, Y., Liu, J., Gao, X., Banerjee, A., Ellis, L. B., and Li, T. (2008) Meta-prediction of phosphorylation sites with weighted voting and restricted grid search parameter selection. *Nucleic Acids Res.* **36,** e22

68. Linding, R., Jensen, L. J., Pasculescu, A., Olhovsky, M., Colwill, K., Bork, P., Yaffe, M. B., and Pawson, T. (2008) NetworKIN: a resource for exploring cellular phosphorylation networks. *Nucleic Acids Res.* **36,** D695–699

69. Huang, J. Y., and Brutlag, D. L. (2001) The EMOTIF database. *Nucleic Acids Res.* **29,** 202–204

70. Wang, Y., and Klemke, R. L. (2008) PhosphoBlast, a computational tool for comparing phosphoprotein signatures among large datasets. *Mol. Cell Proteomics* **7,** 145–162

71. Beltrao, P., Trinidad, J. C., Fiedler, D., Roguev, A., Lim, W. A., Shokat, K. M., Burlingame, A. L., and Krogan, N. J. (2009) Evolution of phospho-regulation: comparison of phosphorylation patterns across yeast species. *PLoS Biol.* **7,** e1000134

72. Betel, D., Breitkreuz, K. E., Isserlin, R., Dewar-Darch, D., Tyers, M., and Hogue, C. W. (2007) Structure-templated predictions of novel protein interactions from sequence information. *PLoS Comput. Biol.* **3,** 1783–1789

73. Katou, Y., Kanoh, Y., Bando, M., Noguchi, H., Tanaka, H., Ashikari, T., Sugimoto, K., and Shirahige, K. (2003) S-phase checkpoint proteins Tof1 and Mrc1 form a stable replication-pausing complex. *Nature* **424,** 1078–1083

74. Chica, C., Labarga, A., Gould, C. M., López, R., and Gibson, T. J. (2008) A tree-based conservation scoring method for short linear motifs in multiple alignments of protein sequences. *BMC Bioinformatics* **9,** 229

75. Perrodou, E., Chica, C., Poch, O., Gibson, T. J., and Thompson, J. D. (2008) A new protein linear motif benchmark for multiple sequence alignment software. *BMC Bioinformatics* **9,** 213

76. Schwartz, M. F., Duong, J. K., Sun, Z., Morrow, J. S., Pradhan, D., and Stern, D. F. (2002) Rad9 phosphorylation sites couple Rad53 to the Saccharomyces cerevisiae DNA damage checkpoint. *Mol. Cell* **9,** 1055–1065

77. Mittag, T., Orlicky, S., Choy, W. Y., Tang, X., Lin, H., Sicheri, F., Kay, L. E., Tyers, M., and Forman-Kay, J. D. (2008) Dynamic equilibrium engagement of a polyvalent ligand with a single-site receptor. *Proc. Natl. Acad. Sci. U.S.A.* **105,** 17772–17777

78. Klein, P., Pawson, T., and Tyers, M. (2003) Mathematical modeling suggests cooperative interactions between a disordered polyvalent ligand and a single receptor site. *Curr. Biol.* **13,** 1669–1678

79. Palmbos, P. L., Daley, J. M., and Wilson, T. E. (2005) Mutations of the Yku80 C terminus and Xrs2 FHA domain specifically block yeast nonhomologous end joining. *Mol. Cell. Biol.* **25,** 10782–10790

80. Yang, X. J. (2005) Multisite protein modification and intramolecular signaling. *Oncogene* **24,** 1653–1662

81. Calnan, D. R., and Brunet, A. (2008) The FoxO code. *Oncogene* **27,** 2276–2288

82. Miranda-Saavedra, D., and Barton, G. J. (2007) Classification and functional annotation of eukaryotic protein kinases. *Proteins* **68,** 893–914