Structural and Functional Analysis of Phosphothreonine-Dependent FHA Domain Interactions

Structure
Article

Structural and Functional Analysis of Phosphothreonine-Dependent FHA Domain Interactions

Simon Pennell,[1,9,*] Sarah Westcott,[1,9] Miguel Ortiz-Lombardía,[2,8,9] Dony Patel,[1] Jiejin Li,[1] Timothy J. Nott,[1] Duaa Mohammed,[3] Roger S. Buxton,[4] Michael B. Yaffe,[3] Chandra Verma,[5,6,7] and Stephen J. Smerdon[1,*]
[1]Division of Molecular Structure, MRC National Institute for Medical Research, The Ridgeway, London NW7 1AA, UK
[2]Structural Biology and Biocomputing Program, Spanish National Cancer Research Centre, C. Melchor Fernández Almagro 3, 28029 Madrid, Spain
[3]Centre for Cancer Research, E18-580, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA
[4]Division of Mycobacterial Research, MRC National Institute for Medical Research, The Ridgeway, London NW7 1AA, UK
[5]Bioinformatics Institute (A*STAR), 30 Biopolis Street, #07-01 Matrix, Singapore 138671
[6]School of Biological Sciences, Nanyang Technological University, 60 Nanyang Drive, Singapore 637551
[7]Department of Biological Sciences, National University of Singapore, 14 Science Drive 4, Singapore 117543
[8]Present address: Architecture et Fonction des Macromolécules Biologiques UMR6098, CNRS, Universités d'Aix-Marseille I & II, Case 932, 13288 Marseille cedex 9, France
[9]These authors contributed equally to this work
*Correspondence: spennel@nimr.mrc.ac.uk (S.P.), ssmerdo@nimr.mrc.c.uk (S.J.S.)
DOI 10.1016/j.str.2010.09.014

SUMMARY

FHA domains are well established as phospho-dependent binding modules mediating signal transduction in Ser/Thr kinase signaling networks in both eukaryotic and prokaryotic species. Although they are unique in binding exclusively to phosphothreonine, the basis for this discrimination over phosphoserine has remained elusive. Here, we attempt to dissect overall binding specificity at the molecular level. We first determined the optimal peptide sequence for Rv0020c FHA domain binding by oriented peptide library screening. This served as a basis for systematic mutagenic and binding analyses, allowing us to derive relative thermodynamic contributions of conserved protein and peptide residues to binding and specificity. Structures of phosphopeptide-bound and uncomplexed Rv0020c FHA domain then directed molecular dynamics simulations which show how the extraordinary discrimination in favor of phosphothreonine occurs through formation of additional hydrogen-bonding networks that are ultimately stabilized by van der Waals interactions of the phosphothreonine γ-methyl group with a conserved pocket on the FHA domain surface.

INTRODUCTION

Cascades of protein phosphorylation are, perhaps, the most common feature of eukaryotic signaling pathways. The human genome has been estimated to encode around 500 protein kinases that play central roles in almost every cellular signaling process (Manning et al., 2002). The effects of phosphorylation are generally manifested as modification-dependent conformational changes in target proteins, or the generation of binding sites for phospho-specific binding domains that mediate the assembly of a variety of multiprotein signaling complexes. One such phospho-specific binding motif is the forkhead-associated (FHA) domain. As the name implies, FHA domains were first described as a region of homology within a subset of Forkhead transcription factor family members and have since been identified in a wide variety of proteins in yeast, plant, mammalian, and bacterial systems (Hofmann and Bucher, 1995). FHA domains have been shown to be involved in a wide range of signal transduction events such as recruitment of DNA damage repair protein complexes in eukaryotes and metabolic regulation in prokaryotes (Mahajan et al., 2008). Uniquely among phospho-binding domains, FHA domains recognize and bind only to sequences containing phosphothreonine (pThr), yet the molecular basis of this pronounced discrimination against phosphoserine (pSer) remains obscure.

Five highly conserved residues appear to be essential for function and for structural stability (Liang and Van Doren, 2008). Outside of these, there is great diversity in sequence between FHA domain examples but, nonetheless, structural studies have demonstrated high conservation of tertiary structure. Almost all FHA domains have been shown to be organized into a twisted β sandwich consisting of two β sheets of five and six strands each with variation between domains found in the lengths of the interstrand loops.

Functionally, FHA domains can be divided into distinct classes dependent on their ligand specificities. As a rule, the major determinant of specificity is the identity of the position +3 to the phosphothreonine. Two major groups have been identified showing preference for either pTXXD or pTXX(I/V/L) and the structural basis for this specificity has been examined in some cases (Durocher et al., 2000). However, examples also exist of FHA domains with more complex binding requirements, for example,

displaying affinity for positions N-terminal to the pThr site, binding to multiply phosphorylated sites, binding to an extended surface (Bernstein et al., 2005; Byeon et al., 2005; Lee et al., 2008; Lloyd et al., 2009; Becherel et al., 2010) and, more recently, phospho-independent interactions (Nott et al., 2009).

Most studies to date have focused on FHA domain target specificity from the perspective of ligand peptide sequences with particular reference to the pThr +3 position (Durocher et al., 2000). Indeed, in spite of the acknowledged importance of FHA domains in many pro- and eukaryotic signaling pathways, a systematic analysis of FHA binding specificity has not, to our knowledge, been reported. Here, we have taken a multipronged approach to analyze more global protein and peptide contributions to FHA domain binding function, and to determine the structural basis of the unique discrimination in favor of phosphothreonine (pThr) over phosphoserine (pSer). To this end, we have focused on the C-terminal FHA domain from *Mycobacterium tuberculosis* Rv0020c, part of the PknB-mediated signaling network, as a model system for crystallographic, biochemical and molecular dynamics analysis due to its structural simplicity.

RESULTS

Definition of a High-Affinity Phosphopeptide Motif for the Rv0020c FHA Domain

Oriented peptide library screening (Yaffe and Smerdon, 2004) was used to define a high-affinity phosphopeptide motif targeted to the Rv0020c FHA domain. A pThr-based degenerate library (X-X-X-X-pThr-X-X-X-X) showed some selection of small/medium hydrophobic side chains at the pThr +3 position. As observed for all other FHA domains thus far examined, no significant binding of pSer- or pTyr-based libraries was observed. Based on these data, a second round of selection was carried out in which Ile was fixed at the +3 position while all other positions were varied. After this second round of selection, a significant preference for proline at the pThr −1 position was observed (Figure 1A).

On the basis of the selection data, an optimal phosphothreonine peptide was synthesized along with nonphosphorylated, phosphorylated serine, and phosphorylated tyrosine peptides and binding was measured by isothermal calorimetry (ITC). Consistent with other FHA domains, no detectable binding to Rv0020c was observed either for the pSer, pTyr, or the nonphosphorylated optimal peptide (data not shown). The optimal pThr peptide bound stoichiometrically with an apparent dissociation equilibrium constant (Kd) of ~100 nM, one of the highest affinities yet determined for a FHA-phosphopeptide interaction (Figure 1B and Table 1).

Rv0020c FHA Domain Structure

The structure of Rv0020c FHA domain was determined by MAD-phasing methods using diffraction data collected from crystals of a selenomethionine-substituted protein. The Rv0020c/optimal peptide complex was then solved by molecular replacement (Table 2). The overall core tertiary structure of the Rv0020c FHA domain is similar to FHA domains solved previously (Mahajan et al., 2008) and is among the smallest and most compact FHA domains characterized thus far (Figure 2A; see Figure S1 available online). In the complex, the phosphopeptide binds in

A

-4	-3	-2	-1		1	2	3
D (2.2) H(2.1)	T (1.8) S (1.7)	A (1.3) L (1.1)	P (1.6)	pT	E (1.5) T (1.5) Q (1.4) V (1.3)	K (1.3) R (1.3) Q (1.3) V (1.3) P (1.3) I (1.2) L (1.1)	I

B

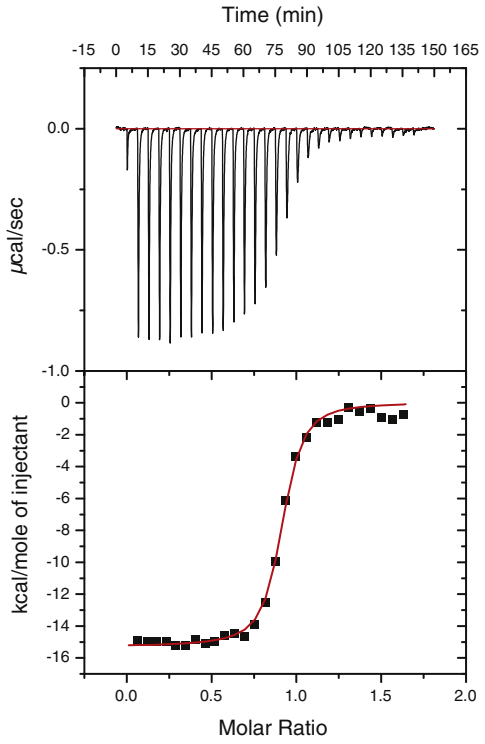

Figure 1. Oriented Library Screening

(A) Oriented peptide library selectivities for the second round of selection with the +3 position fixed as isoleucine. Numbers in brackets are the relative abundance of each amino acid compared with their abundance in the starting library mixture. Selection values ≥ 1.3 indicate moderate selection, while values ≥ 1.7 indicate strong selection based on prior experience with the technique.

(B) Binding of optimal pThr peptide to the Rv0020c FHA domain determined by ITC. The top section shows the raw calorimetric data for the interaction and the bottom section the integrated heat changes, corrected for heat of dilution, and fitted to a single site binding model.

an extended conformation making contacts with residues from loops β3-β4, β4-β5, and β6-β7 (Figure 2B). The pThr side chain is held by five hydrogen bonds to the FHA domain, mediated by side- and main-chain atoms from Ser473, Asn495, Arg474, and Thr494 (Figure 2C). These interactions are conserved among FHA domains with the exception of the last two (Arg474, Thr494). Of these, Arg474 is structurally equivalent to Lys141 in Chk2 that makes hydrogen bond contacts via its side chain with the phosphopeptide phosphate (Li et al., 2002). Additional FHA-peptide hydrogen bonds are mediated by the side chains of Arg459, Arg474, and Asn495 and main-chain atoms from Thr470 and Thr494. The cleft created by Asn495 (β6-β7 loop) and His519 (β10-β11 loop) acts as a specificity pocket for the pThr +3 position (Figure 2D). Least-squares superposition of the apo and

Table 1. Thermodynamic Parameters of the Binding of Mutant Peptides to the Wild-Type Rv0020c FHA Domain Measured by ITC

Peptide	Sequence	K_d (10^{-6} M)	N	ΔH (kCal M^{-1})	TΔS (kCal M^{-1})	ΔG (kCal M^{-1})
Optimal + pThr	D-T-A-P-pT-E-K-I	0.1	0.90	−14.3	−4.9	−9.4
Optimal – pThr	D-T-A-P-T-E-K-I	NDB	-	-	-	-
Optimal + pSer	D-T-A-P-pS-E-K-I	NDB	-	-	-	-
Optimal + pTyr	D-T-A-P-pY-E-K-I	NDB	-	-	-	-
Ile mut	D-T-A-P-pT-E-K-A	0.6	0.99	−10.7	−2.4	−8.3
Pro mut	D-T-A-A-pT-E-K-I	0.1	0.96	−14.7	−5.3	−9.3
Ile-Pro mut	D-T-A-A-pT-E-K-A	0.9	0.97	−12.6	−4.6	−8.0
pT-A	A-A-A-A-pT-A-A-A	7.0	0.97	−11.5	−4.7	−6.8
pT-AI	A-A-A-A-pT-A-A-I	1.5	0.89	−14.9	−7.2	−7.7

NDB, no detectable binding. All peptides additionally terminate with the sequence AYKK to allow spectroscopic quantification and to improve solubility.

bound structures shows minimal deviation between them with an rmsd of 0.3 Å for 93 Cα atoms with no gross structural changes seen to occur upon peptide binding of Rv0020c (Figure 2E). Nonetheless, comparison of the normalized B factors between the free and bound structures is strongly suggestive of a reduced thermal motion in the peptide-binding loops upon phosphopeptide binding (Figure S2).

Binding Contributions of Peptide Residues

We next undertook a systematic analysis of wild-type FHA domain binding affinity to a series of peptide "mutants." All of the measured interactions were enthalpy driven, with large negative ΔH values of the order of −10 kcal mol-1 comparable to values obtained in previous studies of Rad53 FHA1 (−15 kcal mol-1) and Chk2 (−12.1 kcal mol-1) (Durocher et al., 2000) (Table 1). All peptides bound to Rv0020c with 1:1 stoichiometry.

Mutation of the +3 isoleucine to alanine resulted in a 6-fold reduction in binding affinity consistent with this site being a significant specificity determinant as suggested by library screening (Durocher et al., 2000). Mutation of Pro −1 to alanine showed no clear effect on peptide affinity for Rv0020c. However, a peptide containing a +3/−1 double mutation showed a 9-fold reduction in affinity suggesting that the effects of the two mutations are not directly additive. This is consistent with the fact that clear Pro −1 selectivity was only apparent in the second round of selection in which Ile was fixed at the pThr +3 position, and implies that the interactions at the +3 position may be modulated, at least in part, by the identity of the residue at pThr −1 and vice versa. To uncouple the overall contributions of residues other than the phosphothreonine, a peptide was synthesized consisting of alanine at all other positions (pT-A) and another of the same sequence but with the +3 Ile retained (pT-AI). The pT-A peptide displayed the weakest interaction with a Kd of 7 μM, corresponding to an ~80-fold reduction in affinity when compared with wild-type. Restoration of the +3 isoleucine in the pT-AI peptide resulted in a more modest 17-fold decrease in affinity.

Phosphopeptide Binding by Site-Directed Mutants of Rv0020c

Next, we performed mutagenesis on both absolutely and highly conserved residues and used the resultant proteins for ITC titrations with the optimal peptide (Table 3). Mutation of the least-conserved peptide-interacting residues Thr494 and Arg474, resulted in the highest residual binding affinities (1.2 and 1.5 μM, respectively) among the mutants analyzed. Conversely, the largest decreases in affinity were observed for mutations of the conserved arginine (Arg459), serine (Ser473), and asparagine (Asn495). As stated, the S473A mutation resulted in no detectable binding while R459A and N495A display affinities for the optimal peptide in the region of 20 μM with associated large decreases in ΔH when compared to wild-type protein. Ser473 mediates a structural link between β3-β6 in the Rv0020c FHA structure. A similar "bridging" interaction is commonly observed in previously determined FHA structures although it is absent in others such as the FHA domain from aprataxin (Becherel et al., 2010). Nonetheless, we wished to preclude the possibility that the dramatic loss of affinity for the S473A mutant was not merely a result of gross structural disruption and comparison of the far UV circular dichroic spectra of this mutant and the wild-type FHA showed no significant differences (Figure S3).

Structural Basis of pThr-Specificity

In order to further delineate the discrimination between pThr- and pSer-containing peptides by the Rv0020c FHA domain at a molecular level, we used the X-ray structure of the FHA/peptide complex as a starting model for molecular dynamics (MD) simulations. The Rv0020c FHA domain represents an excellent model system for MD simulation due to its extremely compact fold. Stable simulations were carried out on the FHA domain and both pSer and pThr peptides alone in solution and in complex (Figure S4). Movies of the pThr- and pSer-complex simulations are available online at http://web.bii.a-star.edu.sg/bmad/FHA-2010/. Full results of the binding energy analysis are in Table S1; selected values from this analysis are cited below and derive from calculations across the whole protein/peptide system. The dynamics results can be most conveniently summarized in terms of the behavior of the free pThr and pSer phosphopeptides and for each in complex with the Rv0020c FHA domain.

Peptide Dynamics

In simulations of the free pThr and pSer peptides, several significant differences emerge. The pThr peptide appears more structured, demonstrating comparatively reduced fluctuation in atomic positions over time (Figure 3A). This is partly due to two

Table 2. Crystallographic Statistics

	Peptide-free			Peptide-bound
	Peak - SeMet	*Remote - SeMet*	*Native*	
Data Statistics				
Beamline	ESRF ID14-4	ESRF ID14-4	SRS 14.2	SRS 14.1
Wavelength (Å)	0.9791	0.9500	0.98	1.2
Space group	$C222_1$	$C222_1$	$C222_1$	$C222_1$
Cell parameters (Å)	56.87, 69.32, 49.14	56.87, 69.32, 49.14	56.80, 68.90, 48.88	57.02, 74.43, 63.02
Resolution range (Å)[a]	30.0–2.9 (3.03–2.9)	30.0–2.9 (3.03–2.9)	15–1.5 (1.57–1.5)	15–2.0 (2.07–2.0)
Number of observations	15791[b]	15398[b]	107952	61645
Unique reflections	4194 (530)[b]	4169[b]	15858 (1934)	8693 (648)
Completeness	99.6 (99.1)	99.7 (100.0)	99.8 (99.8)	93.7 (71.5)
Mean I/I(σ)	27.6 (17.0)	22.9 (13.2)	15.8 (4.1)	26.2 (5.0)
Multiplicity	3.8	3.7	6.8	7.1
R_{merge} (%)[c]	4.6 (6.4)	5.7 (9.9)	10.5 (38.9)	6.6 (20.6)
R_{meas} (%)[d]			11.8 (40.8)	7.1 (23.4)
R_{pim} (%)[e]			4.0 (20.5)	2.4 (10.9)
Refinement Statistics				
Protein atoms			758	803
Waters			258	234
Phosphate ions			1	-
Zinc ions			-	1
R_{cryst} (%)[f]			18.8	17.9
R_{free} (%)			23.9	23.7
Rmsd bond-lengths (Å)			0.011	0.009
Rmsd bond-angles (°)			1.4	1.3
Ramachandran Plot				
Most favored (%)			95.8	93.9
Additional allowed (%)			4.2	6.1
Disallowed (%)			0	0

See also Figure S3.

[a] Values for highest resolution shell in parentheses.

[b] Bijvoets not merged.

[c] $Rmerge = \Sigma_{hl} | I_{hl} - < I_h > | / \Sigma_{hl} < I_h >$.

[d] $Rmeas = \Sigma_{hl} [nh/(nh-1)]^{1/2} | I_{hl} - < I_h > | / \Sigma_{hl} < I_h >$.

[e] $Rpim = \Sigma_{hl} [1/(nh-1)]^{1/2} | I_{hl} - < I_h > | / \Sigma_{hl} < I_h >$.

[f] $R_{cryst} = \Sigma_h |F_o - F_c| / \Sigma_h F_o$,
where nh is the number of observations of reflection h.

transient H-bonds formed (~55% of simulation time) between the +1 Glu and −3 Thr residues. The pSer peptide also forms a longer lived (~90%) H-bond between the +2 Lys and the pSer phosphate moiety. Both peptides show an attenuation of almost 50% in flexibility upon complex formation. In addition the apparent energetic contributions of each peptide residue to binding show a significant enhancement for pThr residue contributions over pSer for the reasons described below (Figure 3B).

pThr Peptide Complex

The FHA domain becomes more structured when complexed and displays greatest rigidity with the pThr peptide (Figure 3C). Arg459, Ser473, and Arg474 all participate in long-lived H-bonds to the pThr phosphate. Arg459 is further involved in H-bonds with the backbone carbonyl of the peptide −1 Pro and Val472. Asn495 H-bonds to the peptide via the backbone carbonyl of the +1 Glu and the amide group of the +3 Ile. Peptide binding is enthalpically driven with the dominant contribution derived from electrostatic interactions ($\Delta G_{ele} = -42$ kcal mol^{-1}). Energetic contributions for each FHA domain residue are shown in (Figure 3D). The reduced flexibility of the free pThr peptide results in a lower entropic penalty upon complex formation. Additional stabilization is provided by van der Waals interactions arising between the pThr methyl group and a pocket formed by Gly471, Val472, Ser473, Thr494, and Asn495 ($\Delta E_{vdw} = -29$ kcal mol^{-1}). Thus, the γ-methyl binding pocket anchors the pThr side chain allowing the formation of stable H-bonds to the surrounding residues (Figure 3E).

pSer Peptide Complex

pSer peptide complex formation also results in an increase in FHA domain stability albeit to a lesser extent than the pThr peptide (Figure 3C). However, the pSer peptide is able to

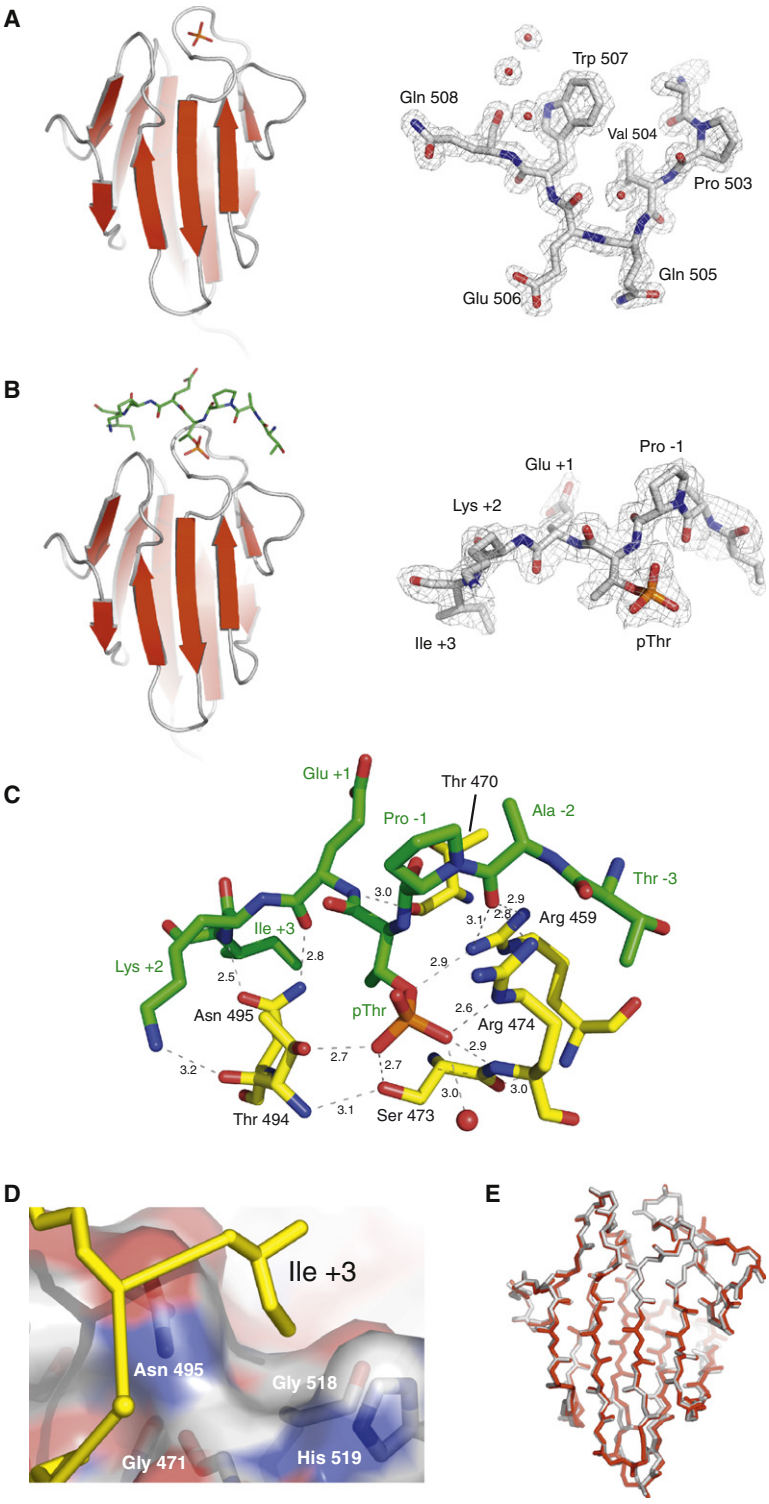

Figure 2. Rv0020c Structure

(A–C) Structures of the free (A) and phosphopeptide complexed (B) Rv0020c FHA domain along with representative m2Fo-DFc "omit" density contoured at 1.0σ. In the absence of the phosphopeptide, Rv0020c only crystallized in conditions using a phosphate-based buffer. The structure shows that a buffer-derived phosphate moiety is situated in the region expected to be associated with phosphopeptide binding. (C) The FHA-peptide interface showing an extensive hydrogen-bonding network from main- and side-chain atoms to highly conserved FHA domain residues.

(D) Ile +3 is accommodated by van der Waals interactions with a shallow pocket formed by Gly471, Asn495, Gly518, and His519.

(E) Least-squares superposition of main-chain atoms of the free and bound Rv0020c FHA domains.

See also Figures S1 and S2.

participate in far fewer interactions with FHA domain residues. Arg459 retains its ability to engage in a persistent H-bond with the pSer phosphate moiety and also contacts Val472. However, the H-bond to the −1 Pro seen for pThr is lost. Ser473 and Arg474 make transient H-bonds with the pSer phosphate representing only 65% and 50% of the simulation time respectively. Thr494 fails to interact with the phosphate at all and additionally the Asn495-peptide backbone interactions observed in the pThr complex are absent. pSer peptide binding is disfavored due to destabilized electrostatics (ΔG_{ele} = 11 kcal mol^{-1}) and a greater entropic deficit upon binding due to its inherently higher conformational flexibility in the free state. Energetic contributions for each FHA domain residue are again shown in Figure 3D. Most significantly, the lack of a γ-methyl group in pSer means that there is no interaction with the binding pocket adjacent to the phosphate binding site with concomitant loss of van der Waals stabilization (ΔE_{vdw} = −26 kcal mol^{-1}). This absence fails to restrict pSer side-chain movement and consequently leads to the loss of the network of H-bonds to these FHA domain residues present in the pThr complex (Figure 3F).

DISCUSSION

Although many examples have been described of varied posttranslational modifications promoting protein-protein recognition, protein assembly as a result of phosphorylation remains the most commonly employed method in both pro- and eukaryotic systems. Phosphorylation-dependent binding was first observed some 20 years ago when a small, independently folding protein module, the SH2 domain, was shown to bind partner proteins through short, phosphotyrosine-containing motifs. This simple but potent mechanism has since been extended to Ser/Thr kinase signaling networks and a number of structurally unrelated pSer/pThr-dependent binding proteins and domains have now been identified. For the majority of pSer/pThr binders (14-3-3, BRCT-repeats, Polo-boxes, WD40 repeats etc), no significant preference for pSer or pThr has been shown. The FHA is an exception and early studies revealed a remarkable binding specificity for pThr (Durocher et al., 2000). In spite of this, and the fact that any pSer-dependent FHA

Table 3. Thermodynamic Parameters of the Binding of Rv0020c FHA Domain Mutants to the "Optimal" pThr Phosphopeptide Measured by ITC

Protein	K_d (10^{-6} M)	N	ΔH (kCal M^{-1})	TΔS (kCal M^{-1})	ΔG (kCal M^{-1})
Wild-type	0.1	0.9	−14.3	−5.0	−9.4
T494A	1.2	0.83	−10.5	−2.6	−7.9
N495A	22.0	1.0	−2.3	4.0	−6.3
R459A	22.0	1.1	−8.6	−2.4	−6.2
R474A	1.5	0.93	−14.9	−7.2	−7.7
S473A	NDB	—	—	—	—

NDB, no detectable binding.

domain interaction has yet to be reported, the structural determinants of pThr specificity remain unclear. In addition, although general sequence preferences for most phospho-specific interacting domain families have been investigated, most often by the use of oriented peptide library methods (Yaffe and Smerdon, 2004), detailed analyses of the contributions of residues adjacent to the phosphoresidues, particularly for pS/pT binders have been lacking.

Overall, and as expected from previous work, the main determinant of peptide binding by the Rv0020c FHA domain was the presence of phosphothreonine. No FHA domain binding to peptides containing unmodified threonine, phosphotyrosine, or phosphoserine was detected by ITC. Beyond this, and consistent with our previous studies on a panel of pro- and eukaryotic FHA domains, the major determinant of overall sequence specificity was provided by the +3 position. The contributions of other residues were more subtle, in line with the observation that many peptide residues do not interact with the FHA domain directly. Similarly, the effects of mutagenesis of the conserved FHA domain residues were dominated by mutations affecting those residues making direct contacts with the phosphothreonine residue. In addition, mutation of the conserved Asn495 residue had a dramatic effect on binding. Hydrogen bonds between Asn495, the main chain C = O of the +1 and the NH of the +3 residues of the bound peptide have been observed in almost all FHA domain complexes described to date. Indeed, the thermodynamic data show that binding to this mutant is associated with a ~12 kCal M^{-1} decrease in enthalpy that is, nonetheless, partially compensated by a favorable entropic term. This presumably reflects a lower entropic penalty on binding to the N495A mutants due to the absence of main-chain hydrogen bonds that tether the C-terminal peptide residues in the wild-type complex. Together, these interactions have been shown to be a major factor in binding partner selection through orientation of the peptide +3 side chain such that it can bind to the structurally variable, specificity-determining "pocket" on the FHA surface (Li et al., 2002) (Figure 2D).

Comparison of peptide-bound and free Rv0020c structures shows no significant structural rearrangements of the protein upon peptide binding. This supports the view that the FHA domains present a preformed binding surface that is essentially conserved between FHA domains despite differences in the orientation of loop structures at the peptide binding site and the notable lack of sequence homology in the ligand sequences themselves. Our molecular dynamics simulations do, however, show an overall reduction in FHA domain dynamics upon pThr peptide binding (Figure 3C; Figure S5). These observations are in accord with the observation of reduced conformational exchange upon phosphopeptide binding in NMR studies of the *Arabidopsis thaliana* Ki67 FHA domain (Ding et al., 2005). Of particular note is the marked reduction in mean fluctuation for the loop containing Asn495 in our pThr peptide complex simulations that is recapitulated in the comparison of crystallographic temperature factors (Figure 3C; Figure S2).

Finally, MD calculations have allowed us to further investigate the remarkable discrimination between phosphothreonine and phosphoserine displayed by FHA domains. It was clear from early studies on FHA domain structure that a surface pocket is present to accommodate the γ-methyl group of pThr (Figure 4). Indeed, NMR analyses of the Rad53 FHA1 domain showed that around 30% of intermolecular NOEs observed in a pThr peptide complex were associated with pThr γ-methyl protons (Yuan et al., 2001). However, it seemed unlikely to us that van der Waals interactions alone could account for the high degree of pThr versus pSer discrimination that has been repeatedly observed biochemically and in vivo. Rather, the MD calculations show that specificity arises from a number of contributing factors. The simulations show that the free pThr peptide is significantly less flexible than the pSer peptide, conferring an advantage through a reduced loss of conformational entropy upon binding to the FHA domain. More importantly, the γ-methyl group clearly engages in favorable but weak van der Waals interactions with atoms that form its binding pocket. However, it seems that the largest contribution to FHA domain discrimination in favor of pThr is that the protein-peptide hydrogen-bonding network is highly stabilized in the pThr complex, especially those H-bonds contacting the pThr phosphate. Thus, the interactions formed at the γ-methyl-binding pocket stabilize a specific pThr rotamer that, in turn, positions the phosphoryl group to maximize the stability of hydrogen-bonding/salt-bridging interactions with conserved FHA domain residues.

In general, 90% of all Ser/Thr kinase activity is directed toward serine phosphorylation (Mann et al., 2002). Thus, through the simple but remarkably effective structural mechanism shown here, FHA domains have evolved a means to effectively decrease the number of potential interaction sites in the cellular milieu by 10-fold with obvious implications for overall binding specificity and precision in FHA-mediated signaling pathways.

EXPERIMENTAL PROCEDURES

Protein Expression and Purification

The coding region of the FHA domain corresponding to residues 430–527 of Rv0020c was amplified using the polymerase chain reaction (PCR) and ligated into the BamH1 and XhoI sites of a pGEX-6P1 vector to create pGST-Rv0020c. This plasmid was transformed into *Escherichia coli* BL21 (DE3) CodonPlus cells (Agilent Technologies) for overexpression. GST-Rv0020c fusion protein was purified using a glutathione-Sepharose column (GE Healthcare), the GST tag removed using 3C protease cleavage, and the Rv0020c FHA domain further purified by size-exclusion chromatography using an S75 size-exclusion column (GE Healthcare). Seleno-methionine-substituted protein was prepared by expressing the protein in the *E. coli* methionine auxotroph B834 (DE3) grown on seleno-methionine-substituted media. Protein molecular weights were confirmed by electrospray mass spectrometry.

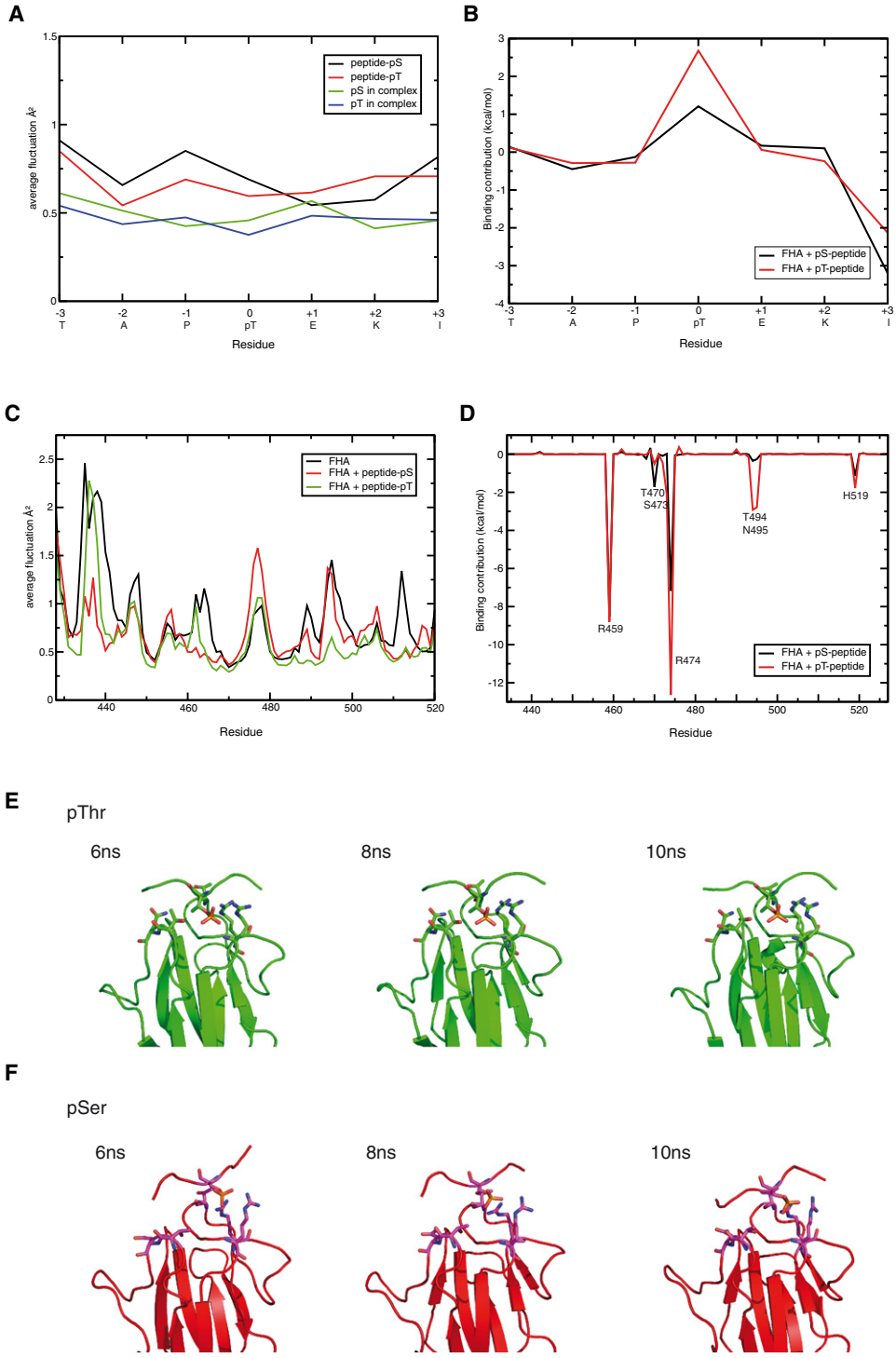

Figure 3. Molecular Dynamics Simulations

(A) Fluctuation of peptide atomic positions over the duration of the simulation. In their free states, the pThr peptide is somewhat more structured than the pSer peptide. The flexibility of both peptides is attenuated by almost 50% upon complex formation as judged by their reduced average fluctuation. The fluctuation of atomic positions of free pThr is somewhat lower than that of free pSer, although once bound this difference is reduced.

(B) Energetic contribution (enthalpy) of peptide residues to binding. The higher stability of the pThr peptide within the complex derives from the contribution of the pThr residue. This also shows, for both pThr and pSer, the solvation penalty that is paid upon the burial of a charged group.

(C) Fluctuation of FHA domain atomic positions over the duration of the simulation. Rv0020c is more structured in its complexed state displaying more rigidity in the context of the pThr complex than when complexed with the pSer peptide. Overall it is clear that peptide binding causes a reduction in mobility throughout the surface of the protein.

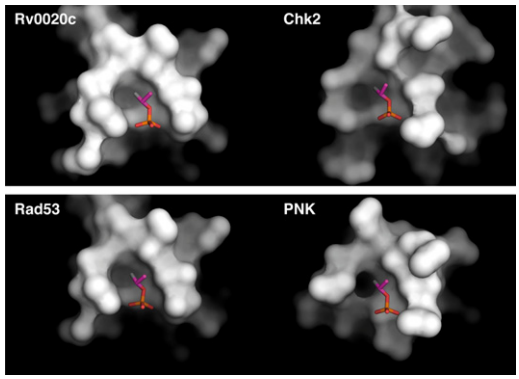

Figure 4. γ-Methyl Binding Pockets
A comparison of the Rv0020c γ-methyl binding pocket to those observed in the Chk2, Rad53 FHA1, and PNK (polynucleotide kinase 3′-phosphatase) FHA domain-phosphopeptide complexes (PDB ids 1GXC, 1G6G, and 2W3O, respectively). Only the phosphothreonine residue from bound phosphopeptides is shown for clarity. In all cases, the surface pocket is positioned so as to accommodate the γ-methyl group of pThr upon docking of its phosphate moiety.

Oriented Peptide Library Screening

Phosphothreonine-oriented degenerate peptide libraries of general sequence Met-Ala-X-X-X-X-pThr-X-X-X-X-Ala-Lys-Lys-Lys and Met-Ala-X-X-X-X-pThr-X-X-Ile-X-Ala-Lys-Lys-Lys were synthesized using N-α-FMOC-protected amino acids and standard BOP/HOBt coupling chemistry. To screen the peptide libraries, GST-tagged FHA domain was bound to 100 μl of glutathione beads to saturation (1–1.5 mg/ml). These beads were incubated with 0.5 mg of the peptide library mixture in PBS (150 mM NaCl, 3 mM KCl, 10 mM Na_2HPO_4, 2 mM KH_2PO_4 [pH 7.2]), for 10 min at room temperature. Unbound peptides were removed by washing the column twice with PBS + 0.5% NP-40. Bound peptides were eluted with 30% acetic acid for 10 min at room temperature, lyophilized, resuspended in water, and sequenced by automated Edman degradation. Raw selection ratios for each amino acid were determined by comparing the relative mole percentage of each amino acid at a particular sequencing cycle in the recovered peptides to that of each amino acid in the original peptide library mixture at the same position. Final selectivity values were calculated by normalizing the raw selection values for the optimal amino acids to the average selection of the remaining amino acids at each position.

Isothermal Titration Calorimetry

Rv0020c protein used in ITC experiments was dialyzed overnight at 4°C into 300 mM sodium chloride, 50 mM HEPES (pH 7.0) using Amicon Millipore dialysis membrane. Peptides were dissolved in the same buffer. Protein and peptide solutions were quantified by UV/visible spectrophotometry and degassed before use. ITC experiments were performed using a VP-ITC microcalorimeter (MicroCal Inc.) and analyzed using ORIGIN software as per the manufacturer's instructions. A typical experiment involved 20–40 μM protein in the sample cell and 200–400 μM peptide in the injection syringe. All measurements were performed at 20°C.

Crystallization

Rv0020c was concentrated to 16 mg/ml in a buffer containing 50 mM Tris HCl (pH 7.8), 300 mM NaCl, 1 mM DTT, and 1 mM EDTA. Crystals were grown by vapor diffusion at 18°C using 1 μl of protein mixed with 1 μl of well solution (0.4 M NaH_2PO_4/1.6 M K_2HPO_4, 0.1 M imidazole [pH 8.0], 0.2 M NaCl). Crystals belong to orthorhombic space group $C222_1$ and contain one Rv0020c molecule per asymmetric unit. Rv0020c Se-Met protein was concentrated to 16 mg/ml in a buffer containing 50 mM Tris HCl (pH 7.8), 300 mM NaCl, 1 mM tris-carboxy ethyl phosphine (TCEP) and 1 mM EDTA and crystallized under identical conditions to those employed for the native protein.

The phosphopeptide DTAPpTEKIAYKK was mixed with Rv0020c in a 1:1.5 (protein:peptide) stoichiometric excess and concentrated to 16 mg/ml in a buffer containing 50 mM Tris HCl (pH 7.8), 300 mM NaCl, 1 mM DTT, and 1 mM EDTA. Crystals were grown using vapor diffusion at 18°C using 1 μl of protein/peptide complex mixed with 1 μl of well solution (10% v/v 2-propanol, 0.1 M sodium cacodylate [pH 6.5], 0.2 M zinc acetate). Crystals grew over 1 week and also belong to the orthorhombic space group $C222_1$ with one complex per asymmetric unit.

Structure Determination and Refinement

All data were collected from flash-cooled crystals maintained at a temperature of 100K using an Oxford Cryosystems Cryostream device and reduced using the HKL suite of programs. Native data for Rv0020c were collected on station 14.2 at the SRS (Daresbury, UK) on an ADSC Q4R CCD detector. Phase information was derived from a two-wavelength MAD experiment using selenomethionine derivitized Rv0020c protein crystals. Data for each wavelength were collected to a nominal 3.0 Å on an ADSC Q4R CCD detector at the ESRF (Grenoble) on beamline ID14EH4. One Se site was located and the phases were refined using SOLVE. Electron density maps showed a clear solvent boundary and evidence of β strand structure. Phases were further improved and extended to 2.6 Å by solvent-flattening. The resulting electron density maps were of sufficient quality to construct a polyalanine model encompassing 72 residues of Rv0020c in "O." This model was partially refined in Refmac and the improved phases used as input to the warpNtrace routine of wARP using diffraction data extending to 1.5 Å spacing. The resulting warpNtrace model contained 888 atoms comprising 95 residues (built as Ser, Gly, Ala, or Val). The remaining protein structure was built using "O" and the resulting model was refined using Refmac incorporating automated solvent building with ARP (Lamzin and Wilson, 1993).

The Rv0020c/peptide complex data were collected on station 14.1 at the SRS (Daresbury, UK) on an ADSC Q4R CCD detector. Data were collected to a high-resolution limit of 2.0 Å. The structure was solved by molecular replacement employing AMoRe and the refined structure of uncomplexed Rv0020c as a search model (Navaza, 2001). The model from molecular replacement was refined using Refmac at a resolution of 2 Å. Crystallographic statistics are shown in Table 2.

Molecular Dynamics Simulations

All the calculations were carried out using the molecular mechanical representation described by the parm99 force field (Cornell et al., 1995), as implemented within the AMBER 9 suite of programs (Case et al., 2006). Five starting structures were prepared for the simulations: the Rv0020c/pThr-peptide and Rv0020c/pSer-peptide complexes (this one was generated from the pThr complex by removal of the methyl group), along with their uncomplexed constituents, i.e., the protein and the two phosphopeptides. The LEaP module from AMBER 9 was used to patch the N-terminus of the protein with an ACE residue. LEaP was also used to place each starting structure into a truncated-octahedral periodic box of TIP3P water molecules and to attain electroneutrality by adding appropriate ions. The distance between the edges of the water boxes and the closest atom of the solute was at least 12 Å. The particle mesh Ewald (PME) method (Darden et al., 1993) was used to treat long-range electrostatic interactions, and bond lengths involving bonds to

(D) Energetic contribution (enthalpy) of FHA domain residues to binding. The highlighted residues contact the peptide through the pThr phosphate (Arg459, Ser473, Arg474, Thr494), main-chain atoms (Arg459, Thr470, Asn495) or the +3 specificity position (His519). The similar H-bonding patterns of Arg459 in both systems is reflected in almost equal contributions, with the 1 kcal mol-1 difference in favor of the complex with the pThr peptide resulting from an additional contact to the +2 Ala or the increased stability of the pThr side chain. The larger number and stability of H-bonds made by Ser473, Arg474, Thr494, and Asn495 lead to respective contributions that are of greater magnitude in the pThr complex. These effects can be clearly seen in snapshots from the MD trajectories of the pThr (E) and pSer (F) complexes.
See also Figure S4 and Movie S1 and S2.

hydrogen atoms were constrained using SHAKE (Ryckaert et al., 1977). The time-step for molecular dynamics (MD) simulations up to the equilibration was 2 fs, followed by 1 fs during the production runs. A direct-space nonbonded cutoff of 8 Å was used. The energy of each system was minimized by 500 steps of steepest descent minimization followed by up to 500 steps of conjugate gradient minimization. Harmonic restraints with force constants of 2 kcal mol^{-1} $Å^{-2}$ were applied to all solute atoms during the minimization. These restraints were maintained through a 50 ps canonical ensemble (NVT)-MD simulation, during which the systems were heated from 0 to 300 K. The restraints were also kept for a subsequent 50 ps isothermal isobaric ensemble (NPT)-MD, to adjust the solvent density. Finally, after an additional 2 ns of unrestrained simulation at 300 K with a time constant of 1.0 ps for heat-bath coupling, the following 10 ns were used to extract the snapshots for analyses (for example, see Dastidar et al., 2008) Further details of the energetic analysis can be found in Supplemental Information.

ACCESSION NUMBERS

Coordinates and structure factors for the free and complexed Rv0020c FHA domain have been deposited with the Protein Data Bank, accession numbers 3PO8, 3POA.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, five figures, one table, and two movies and can be found with this article online at doi:10.1016/j.str.2010.09.014.

ACKNOWLEDGMENTS

M.B.Y. acknowledges funding from the National Institutes of Health (GM60594, GM68762). R.S.B. and S.J.S. are grateful to the Medical Research Council, UK for its continuing support.

Received: July 6, 2010
Revised: September 14, 2010
Accepted: September 17, 2010
Published: December 7, 2010

REFERENCES

Becherel, O.J., Jakob, B., Cherry, A.L., Gueven, N., Fusser, M., Kijas, A.W., Peng, C., Katyal, S., McKinnon, P.J., Chen, J., et al. (2010). CK2 phosphorylation-dependent interaction between aprataxin and MDC1 in the DNA damage response. Nucleic Acids Res. *38*, 1489–1503.

Bernstein, N.K., Williams, R.S., Rakovszky, M.L., Cui, D., Green, R., Karimi-Busheri, F., Mani, R.S., Galicia, S., Koch, C.A., Cass, C.E., et al. (2005). The molecular architecture of the mammalian DNA repair enzyme, polynucleotide kinase. Mol. Cell *17*, 657–670.

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.

Case, D.A., Darden, T.A., Cheatham, T.E., Simmerling, C.L., Wang, J., Duke, R.E., Luo, R., Merz, K.M., Pearlman, D.A., Crowley, M., Walker, R.C., et al. (2006). AMBER 9 (San Francisco: University of California).

Cornell, W.D., Cieplak, P., Bayly, C.I., Gould, I.R., Merz, K.M., Ferguson, D.M., Spellmeyer, D.C., Fox, T., Caldwell, J.W., and Kollman, P.A. (1995). A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. J. Am. Chem. Soc. *117*, 5179–5197.

Darden, T., York, D., and Pedersen, L. (1993). Particle mesh Ewald: an Nlog(N) method for Ewald sums in large systems. J. Chem. Phys. *103*, 8577–8593.

Dastidar, S.G., Lane, D.P., and Verma, C.S. (2008). Multiple peptide conformations give rise to similar binding affinities: molecular simulations of p53-MDM2. J. Am. Chem. Soc. *130*, 13514–13515.

Ding, Z., Lee, G.I., Liang, X., Gallazzi, F., Arunima, A., and Van Doren, S.R. (2005). PhosphoThr peptide binding globally rigidifies much of the FHA domain from Arabidopsis receptor kinase-associated protein phosphatase. Biochemistry *44*, 10119–10134.

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.

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.

Lamzin, V.S., and Wilson, K.S. (1993). Automated refinement of protein models. Acta Crystallogr. D Biol. Crystallogr. *49*, 129–147.

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.

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.

Liang, X., and Van Doren, S.R. (2008). Mechanistic insights into phosphoprotein-binding FHA domains. Acc. Chem. Res. *41*, 991–999.

Lloyd, J., Chapman, J.R., Clapperton, J.A., Haire, L.F., Hartsuiker, E., Li, J., Carr, A.M., Jackson, S.P., and Smerdon, S.J. (2009). A supra-modular FHA/BRCT-repeat architecture mediates Nbs1 adaptor function in response to DNA-damage. Cell *139*, 100–111.

Mahajan, A., Yuan, C., Lee, H., Chen, E.S., Wu, P.Y., and Tsai, M.D. (2008). Structure and function of the phosphothreonine-specific FHA domain. Sci. Signal. *1*, re12.

Mann, M., Ong, S.E., Gronborg, M., Steen, H., Jensen, O.N., and Pandey, A. (2002). Analysis of protein phosphorylation using mass spectrometry: deciphering the phosphoproteome. Trends Biotechnol. *20*, 261–268.

Manning, G., Whyte, D.B., Martinez, R., Hunter, T., and Sudarsanam, S. (2002). The protein kinase complement of the human genome. Science *298*, 1912–1934.

Navaza, J. (2001). Implementation of molecular replacement in AMoRe. Acta Crystallogr. D Biol. Crystallogr. *57*, 1367–1372.

Nott, T.J., Kelly, G., Stach, L., Li, J., Westcott, S., Patel, D., Hunt, D.M., Howell, S., Buxton, R.S., O'Hare, H.M., et al. (2009). An intramolecular switch regulates phosphoindependent FHA domain interactions in Mycobacterium tuberculosis. Sci. Signal. *2*, ra12.

Ryckaert, J.-P., Ciccotti, G., and Berendsen, H.J.C. (1977). Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. J. Comp. Physiol. *23*, 327–341.

Yaffe, M.B., and Smerdon, S.J. (2004). The use of in vitro peptide-library screens in the analysis of phosphoserine/threonine-binding domain structure and function. Annu. Rev. Biophys. Biomol. Struct. *33*, 225–244.

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.

