FEBS Letters 513 (2002) 58–66 FEBS 25675

Minireview

The FHA domain

Daniel Durocher[a,b,*], Stephen P. Jackson[c,d]

[a] *Samuel Lunenfeld Research Institute, Programme in Molecular Biology and Cancer, Mount Sinai Hospital, 600 University Ave., Toronto, ON, Canada M5G 1X5*
[b] *Department of Medical Genetics and Microbiology, University of Toronto, King's College Circle, Toronto, ON, Canada*
[c] *Wellcome/CRC Institute of Cancer and Developmental Biology, Tennis Court Road, Cambridge CB2 1QR, UK*
[d] *Department of Zoology, University of Cambridge, Cambridge, UK*

Received 12 November 2001; revised 29 November 2001; accepted 3 December 2001

First published online 20 December 2001

Edited by Gianni Cesareni and Mario Gimona

Abstract The forkhead-associated (FHA) domain is a small protein module recently shown to recognize phosphothreonine epitopes on proteins. It is present in a diverse range of proteins in eukaryotic cells, such as kinases, phosphatases, kinesins, transcription factors, RNA-binding proteins, and metabolic enzymes. It is also found in a number of bacterial proteins. This suggests that FHA domain-mediated phospho-dependent assembly of protein complexes is an ancient and widespread regulatory mechanism.

Key words: Phosphorylation; Signal transduction; Protein:protein interaction; Kinase; Forkhead-associated domain; Kinesin; Checkpoint; DNA repair

1. Introduction

Modular phosphopeptide recognition is a central component of metazoan signal transduction and, until recently, was generally considered to have evolved in parallel with metazoan evolution. Indeed, as organisms became more complex, a sophisticated apparatus needed to be developed to coordinate organismal physiology and development. This signaling apparatus had to be plastic in order to be evolvable and adaptable and also needed to have a high degree of specificity such that the cell could integrate and compute environmental cues without confusion. Protein modules provided part of the solution to this problem and the discovery of SH2 and SH3 domains initiated an important paradigm shift in the signaling field [1].

Signaling protein modules are small independently folded units having a variety of functions usually centered around being protein- or ligand-binding domains [1,2]. An important class of signaling modules is those able to recognize phosphorylated epitopes on proteins. This class of signaling modules comprises the SH2, PTB and WW domains, 14-3-3 proteins, and the subject of this review, the forkhead-associated (FHA) domain [3]. The structural characteristics of protein modules enable them to be shuffled around the genome, providing a powerful means for the rapid, in evolutionary terms, assembly of new multi-domain proteins with novel functions [1,2]. As well as creating new functionalities, the assembly of multi-domain proteins facilitates the connection of signaling pathways and the assembly of highly specific signaling complexes.

This issue of FEBS Letters contains many striking examples of the use of protein modules in signal transduction. However, given their great functional versatility, it is not surprising that protein modules are also used in processes other than signal transduction. For instance, the modular nature of transcriptional regulators has long been recognized and modular phosphopeptide recognition by the FHA domain is associated with many such processes. In this review, we will describe the recent work that has identified the FHA domain as a phosphopeptide-binding module and will summerize some of the various roles that the FHA domain plays, from bacteria to humans.

2. The FHA domain: a phosphopeptide-binding module

The FHA domain, discovered in 1995 by Hofmann and Bucher [4], has recently been shown to be a modular phosphopeptide recognition domain with a striking specificity for phosphothreonine (pT)-containing epitopes [5–11]. So far, the FHA domain has been found in more than 200 different proteins (cf. http://smart.embl-heidelberg.de/) in species ranging from prokaryotes to higher eukaryotes, and it is associated with proteins involved in numerous processes including intracellular signal transduction, transcription, protein transport, DNA repair and protein degradation [8,11], some of which are schematically depicted in Fig. 1.

The first indications of FHA domain involvement in phospho-dependent protein:protein interactions emerged from studies on two very different systems: plant signal transduction and checkpoint signaling [12,13]. In the first case, a search for proteins interacting with receptor-like protein kinase RLK5 led Walker and colleagues to the cloning of the kinase-associated protein phosphatase (KAPP) in 1994 [12]. Importantly, the interaction between KAPP and RLK5 was found to be dependent on phosphorylation and the region of KAPP required for this interaction was termed the kinase-interaction (KI) domain [12]. Four years later, genetic studies in *Arabidopsis* indicated that KAPP is a negative regulator of CLAVATA 1 (CLV1)-dependent signaling [14]. As with the

*Corresponding author. Fax: (1)-416-5868869.
E-mail addresses: durocher@mshri.on.ca (D. Durocher), spj13@mole.bio.cam.ac.uk (S.P. Jackson).

0014-5793/02/$22.00
PII: S0014-5793(01)03294-X

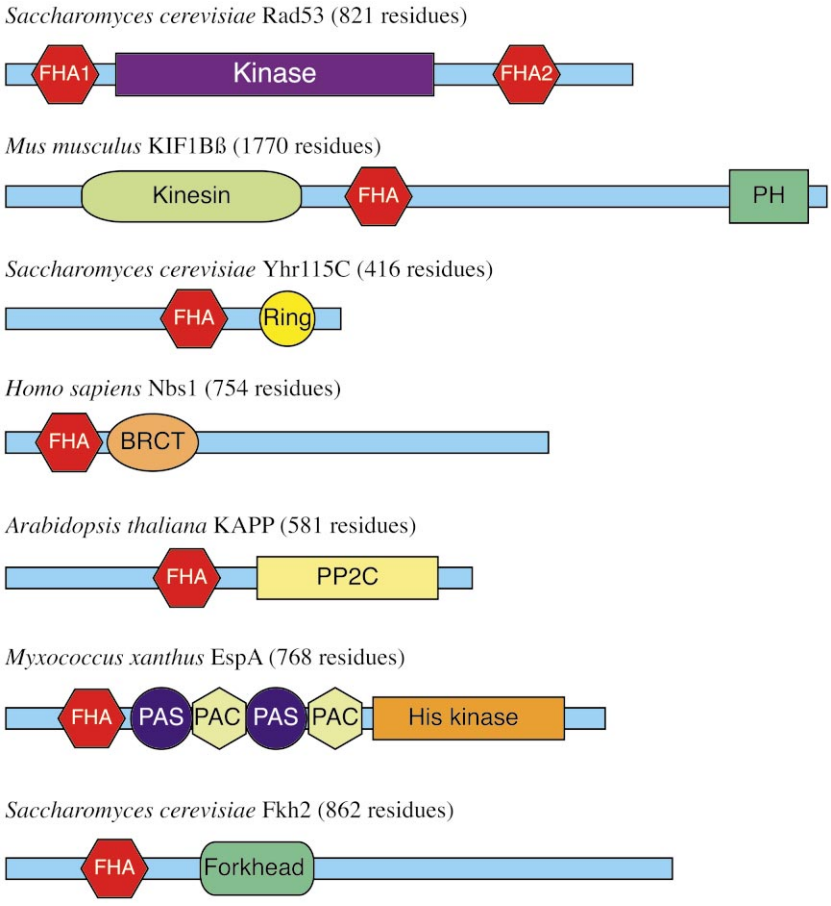

Fig. 1. The FHA domain is associated with proteins of diverse function. FHA-containing proteins Rad53 (gi: 6325104), KIF1Bβ (gi: 4512330), Yhr115c (gi: 6321907), Nbs1 (gi: 14751716), KAPP (gi: 1076344), EspA (gi: 5713126) and Fkh2 (gi: 6324261) are schematically represented along with their domain organizations. For more information on the domains, please refer to pfam at http://www.sanger.ac.uk/software/pfam/). The schematized domains are (pfam accession number in brackets): FHA domain (PF00498), BRCA1 C-terminus domain (BRCT; PF00533), kinesin motor domain (PF00225), PH domain (PF00169), protein kinase domain (kinase, PF00069), protein phosphatase type 2C domain (PP2C; PF00481), ring-finger domain (Ring; PF00097). PAS domain (PF00989), PAC motif (PF00785) and the histidine kinase domain (comprising pfam motifs PF00512, PF02518 and PF00072).

KAPP–RLK5 interaction, it was found that KAPP bound directly to activated CLV1, a plant receptor-like kinase [14]. These genetic and physical interactions suggest that recruitment of KAPP by activated receptors during plant signal transduction is very similar to the recruitment of SH2-containing tyrosine phosphatases to activated receptor tyrosine kinases [15], and that phosphorylation-dependent attenuation of receptor-mediated signaling is a common feature of both plant and animal cells.

These studies preceded the discovery of the FHA domain, but in a follow-up study, Walker and colleagues demonstrated that residues conserved in KAPP's FHA domains were critical for KI domain activity [7]. This suggested that the FHA domain of KAPP mediates phospho-dependent protein:protein interactions but did not resolve the important issue of whether these interactions directly involved the recognition of a phosphorylated residue or whether they were initiated by a conformational change triggered by phosphorylation.

In parallel, similar conclusions were drawn from studies on the budding yeast DNA damage checkpoint. In response to DNA damage, a protein kinase cascade that controls both the inhibition of cell cycle progression and the DNA repair process itself is initiated. In *Saccharomyces cerevisiae*, this cascade is, in large part, controlled by the Rad53 protein kinase [16]. Rad53, which contains two FHA domains flanking a central catalytic protein kinase domain, is a member of a family of closely related FHA domain-containing protein kinases, other members of which include the human orthologue, the *hCHK2* tumor-suppressor [17], and its *S. cerevisiae* paralogues, the meiotic checkpoint kinases Mek1/Mre4 [4,18] and Dun1 [19,20].

The presence of FHA domains on Rad53 was highly suggestive of a role in protein:protein interactions and, consistent with this, Rad53 was found to interact with Rad9 following DNA damage [13,21,22]. The interaction between Rad9 and Rad53 is dependent on Rad9 phosphorylation [5,13] and was shown to require at least the C-terminal FHA domain (FHA2) of Rad53 [13], although the N-terminal FHA domain (FHA1) is likely to also play an important role in binding [5].

The first definite demonstration that FHA domains recognize directly phosphorylated instead of epitopes whose unmasking is triggered by phosphorylation entailed the identification of peptides able to compete the interaction between the FHA1 domain of Rad53 ($Rad53^{FHA1}$) and the phosphorylated forms of Rad9 that occur after DNA damage [5]. A pT-containing peptide derived from the N-terminus of p53, but not its unphosphorylated counterpart, was found to abrogate the $Rad53^{FHA1}$–Rad9 interaction [5]. This pT-containing peptide was further shown to directly bind $Rad53^{FHA1}$ and mutation

of the conserved Arg70 and His88 residues, but not the variable Asp117 residue, in Rad53^{FHA1} abolished its interaction with both the synthetic phosphopeptide and with phosphorylated Rad9. Furthermore, the Rad53^{FHA1} domain was shown to protect the phosphorylated residue from phosphatase treatment, suggesting that the phosphate moiety is in intimate contact with the FHA domain [5].

Perhaps surprisingly, this study also demonstrated that the phosphopeptide–Rad53^{FHA1} interaction does not take place when the pT residue is substituted with phosphoserine (pS) or the acidic amino acid, Asp [5]. However, the phosphopeptide–Rad53^{FHA1} interaction tolerated substitution of the residue at positions pT−3 to pT−1 (see Fig. 2B) or pT+1 to pT+2 but did not tolerate substitution of the Asp residue at pT+3. This latter result indicated that pT+3 is critical for binding and may be involved in binding selectivity. Importantly, by examining the ability of many other FHA domains to bind, in a phospho-dependent manner, to a pT-containing

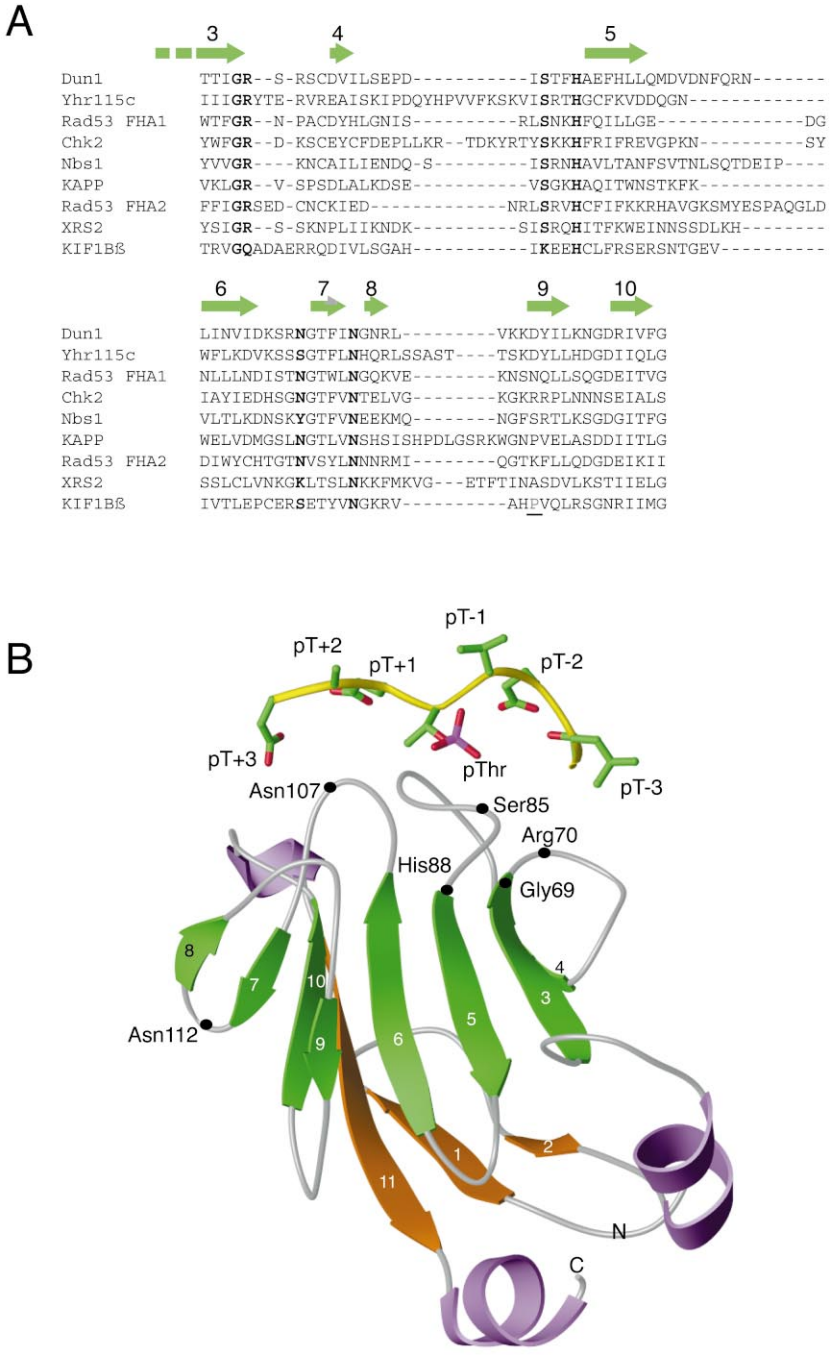

Fig. 2. FHA domain homology and structure. A: Modified sequence alignment of the core FHA homology region of various FHA domains from Rad53 (gi: 6325104), KIF1Bβ (gi: 4512330), Yhr115c (gi: 6321907), Nbs1 (gi: 14751716), KAPP (gi: 1076344), Chk2 (gi: 6005850), Dun1 (gi: 6320102) and Xrs2 (gi: 2133087). The alignment is based on structural and biochemical data and is based on a similar alignment published in [6]. The secondary structure – in green above the alignment, arrows represent β-sheets – is based on the Rad53 FHA domain structures [6,9]. Conserved residues are bolded and residue Pro578 of KIF1Bβ is in blue and underlined. B: Ribbons representation of the Rad53^{FHA1} domain in complex with a pT-containing peptide (shown in a ball-and-stick model). The core FHA homology (sheets β3–β10) is colored in green.

library by surface plasmon resonance, it was concluded that phospho-dependent binding is a common feature of the FHA domain [5]. However, the observation that only Rad53^{FHA1} could bind to a phosphopeptide containing an Asp residue at pT+3 suggested that FHA domains are likely to recognize different sets of phosphopeptide sequences and that selectivity might be dictated primarily by the pT+3 residue. This hypothesis was confirmed in later studies where anchored peptide libraries were screened against a variety of FHA domains and consensus binding sequences were thus identified [6,10].

3. The FHA domain: structure and function

The recent solution of three FHA domain-phosphopeptide structures by X-ray crystallography and nuclear magnetic resonance (NMR) has shed important light on the mechanisms by which FHA domains recognize phosphorylated epitopes. The FHA domain itself, as defined by the minimal autonomously folded unit, is significantly larger than the previously described FHA homology region [6,9,23–25]. It spans approximately 80–100 amino acid residues folded into an 11-stranded β sandwich, which sometimes contains small helical insertions between the loops connecting the strands (e.g. between β2 and β3 in Rad53^{FHA1} and between β4 and β5 in Chk2^{FHA}; Fig. 2B and [6,9,24]). The FHA homology region spans only eight β-strands on one face of the sandwich (strands β3–β10, Fig. 2) [6]. Interestingly, the fold of the FHA domain is not unique and is highly similar to the fold of the Smad MH2 domain [6,9]. The significance of this homology is discussed later.

In both the Rad53^{FHA1} and the Chk2^{FHA} structures, the phosphopeptide is bound by the loops connecting the β3/β4, β4/β5 and β6/β7 strands of the FHA domain in an extended conformation highly reminiscent of the binding of antigen epitopes on the complementary determining region of antibodies [6]. The determinants of phosphopeptide recognition by the FHA domain are all embedded in the FHA homology region and many of the most conserved residues among the FHA domains play a role in the direct binding of the phosphopeptide, either by contacting the phosphopeptide backbone (Asn107, Arg70 in Rad53^{FHA1}) or via the pT residue through an extensive network of hydrogen bonds (Arg70, Ser85 in Rad53^{FHA1}). In the structures determined to date, the phosphate moiety of the pT residue does not make salt bridges with any FHA domain residue and the very conserved and functionally essential Arg residue (Arg70 in Rad53^{FHA1} and Arg117 in Chk2) only makes an hydrogen bond with the γ-oxygen of pT [3,24,26]. This mechanism of recognition of the pT residue is strikingly different to that by which pS or phosphotyrosine (pY) residues are recognized by 14-3-3 proteins, SH2 and PTB domains. In these latter cases, the interaction with the phosphorylated residue is stabilized by at least one Arg residue making a salt bridge with the phosphate moiety [3,26]. Lastly, two other strongly conserved residues (Gly69 and His88 in Rad53^{FHA1}, Fig. 2B) stabilize the architecture of the binding site [6]. We note that these residues have often been substituted in studies looking at FHA domain function and we urge caution in using such substitutions in the future as they may result in the disruption of FHA domain tertiary structure. Indeed, mutation of His88 renders Rad53 unstable in vivo (D.D. and S.P.J., unpublished).

Taken together with the available biochemical data, the striking conservation of the amino acid residues involved in phosphopeptide binding, indicated in the structural studies of FHA domains, strongly suggests that all FHA domains bind to phosphopeptides [5,6]. The structures also help to rationalize the molecular basis of phosphopeptide sequence selectivity by FHA domains. As discussed above, FHA domains generally show strong selection for residues at the pT+3 position. Thus, Rad53^{FHA1} selects strongly for Asp, whereas Chk2 selects for Ile. Both bind their optimal target phosphopeptides with similar affinity [5,6,10,24]. By comparing the structures of the Chk2^{FHA} domain and the Rad53^{FHA1} domain in a complex with their targets, it can be seen that the extensive network of conserved contacts between the phosphopeptide and the FHA domain results in a nearly identical extended peptide conformation in both cases [24]. Perhaps surprisingly, however, the region of the FHA domain next to the pT+3 residue is structurally very variable. In the Rad53^{FHA1} structure, the pT+3 residue (Asp) makes a salt bridge with the guanidium moiety of Arg83, which is located in the loop between β4 and β5 [24]. However, Chk2^{FHA} has an eight-amino acid residue helical insertion in the β4/β5 loop which results in a structural conformation radically different to that of Rad53^{FHA1}. As a consequence of this insertion, the β4/β5 loop is steered away from the pT+3 residue, and thus, the pT+3 residue (Ile) is closely apposed to a region composed of residues from the β10/β11 loop. Thus, the large variations observed in the length and composition of both the β4/β5 and β10/β11 loops among different FHA domains may explain the diversity in binding selectivity and rationalize the observation that the phosphopeptide amino acid residue pT+3 appears to be the major determinant for binding selectivity [24].

Another striking feature of the FHA domain is its apparent specificity for pT residues over pS residues. Here again, the crystal structures of the FHA domains of Rad53 (FHA1) and Chk2 bound to a high-affinity binding pT-containing peptide illustrate how this specificity is imparted. In both structures, the pT γ-methyl group makes clear side chain interactions with a conserved Asn residue (Asn107 for Rad53^{FHA1} and Asn166 for Chk^{FHA1}) and main chain contacts with a number of residues (Ser82–Ile84 for Rad53^{FHA1} and Thr138–Tyr139 for Chk^{FHA1}) [3,6,24,26]. These contacts are thought to largely explain the specificity for pT over pS but experimental confirmation of this explanation is still required.

Another intriguing feature of the FHA domain is its ability to bind to peptides containing a pY [27]. It is not clear what the biological significance of this finding is, especially in organisms such as budding yeast where there is a paucity of tyrosine kinases, but the relatively high binding affinity of a pY-containing peptide for the FHA2 domain of Rad53 [27] warrants further examination and suggests that FHA domains may even recognize other types of phosphoamino acids such as phosphoaspartate or phosphohistidine.

4. Eukaryotic FHA-containing proteins

Although the FHA domains of several eukaryotic proteins are now being analyzed, this area of investigation is far from having reached maturity and the data available amount to little more than a collection of interacting proteins. Nevertheless, the presence of an FHA domain has a predictive value, i.e. it strongly indicates that the FHA-containing protein will interact with a protein partner in a process regulated by reversible protein phosphorylation. It is therefore not surprising

that FHA domains are now being used as a tool to decipher the molecular circuitry of cellular pathways. Below, we review some of the cellular pathways where FHA-containing proteins have been identified to play a role and discuss the potential significance of the presence of the FHA domains in these events.

4.1. DNA damage repair and signaling

The process by which DNA damage is recognized and then signaled to the cell cycle and DNA repair machineries can be considered as a bona fide signal transduction cascade [28]. It is initiated by a signal (DNA damage) which is detected by a receptor or sensor, the precise nature of which remains elusive to date, and which triggers activation of a proximal set of kinases (ATM and ATR in mammals; Tel1 and Mec1 in budding yeast [29,30]). These kinases, in turn, activate a series of more distal kinases, such as budding yeast Rad53, Dun1 and Chk1, which then phosphorylate and regulate a number of protein effectors of the checkpoint and DNA damage responses [16].

FHA-dependent peptide recognition plays a critical role at many levels in this cascade. In budding yeast, FHA domains are present on the checkpoint kinases Rad53 and Dun1, and on Xrs2, the orthologue of Nbs1, the protein encoded by the gene mutated in the Nijmegen-breakage syndrome (see below). In these three proteins, the FHA domains seem to play a critical role in integrating upstream signals ([13,20] and our unpublished results), but most of their cellular partners remain to be identified.

In humans, the orthologs of *RAD53* (*CHK2*) and *XRS2* (*NBS1*) play a conserved role in the DNA damage checkpoint response. Chk2 is required to stabilize p53 in response to DNA damage by phosphorylating it on Ser20 [31–33], whereas Nbs1 is part of the hMre11/hRad50/Nbs1 (MRN) complex which plays a pleiotropic role during the S-phase checkpoint and in DNA repair [34–36]. Mutations in both of these genes have been associated with an increased cancer risk. Germline mutations in the *CHK2* gene lead to a variant form of the cancer syndrome Li-Fraumeni [17] and germline mutations in the *NBS1* gene lead to the cancer-prone developmental disorder Nijmegen-breakage syndrome [34,35,37]. The FHA domains of both proteins seem to play an important role in preventing tumorigenesis. In the case of Chk2, two tumor-associated missense mutations map to the FHA domain, Ile157Thr and Arg145Trp [17]. Functional and structural analyses of these mutations have revealed that R145W destabilizes the structure of the FHA domain with an apparent concomitant reduction in protein stability [24,38,39]. As a result, this mutated Chk2 is unable to engage in phosphorylation-dependent protein:protein interactions and in the HCT15 cell line, which carries this allele of *CHK2*, DNA synthesis continues unchecked after ionizing radiation. This phenomenon is called radio-resistant DNA synthesis (RDS) and is a hallmark of a profound S-phase checkpoint defect [39]. By contrast, the functional outcome of the I157T mutation is more subtle. Despite being unable to complement the RDS phenotype of the HCT15 tumor cell line [39], Chk2-I157T is able to bind to phosphorylated epitopes and its kinase activity is still activated after DNA damage [24]. However, the I157T mutation has the intriguing consequence of abrogating Chk2 binding to a series of proteins including Cdc25C, p53 and Brca1, all of which are bona fide substrates for Chk2 kinase activity [24,39,40]. We therefore propose that the I157T mutation is located in an additional protein interaction interface of the FHA domain, which may act as a substrate docking domain for Chk2 substrates. It will be interesting to test if trans-phosphorylation of biologically relevant substrates is dependent on the integrity of this residue and to solve, if possible, the structure of the Chk2 kinase domain in relation to the FHA domain. The absence of effect on the phosphopeptide-binding ability of the Chk2-I157T mutant is explained by the location of Ile157, which is at the C-terminal end of the β5/β6 loop of the FHA domain, some 25 Å away from the phosphopeptide-binding region. Interestingly, the residue, which is largely solvent exposed, lies in an area corresponding to the SARA (Smad anchor for receptor activation)-binding interface on the MH2 domain of Smad [24]. This additional link between the FHA and MH2 domains is compelling and suggests that some FHA–protein interactions have a binding component that is not phosphorylation-dependent.

Cells from individuals suffering from Nijmegen-breakage syndrome also display a defect in the S-phase checkpoint. This role of Nbs1 and its orthologues has clearly been conserved throughout eukaryotic evolution (see above and [36,41–43]). NBS cells are usually homozygous for a frameshift mutation in the *NBS1* gene, that results in a premature termination codon C-terminal to the FHA and BRCT domains [34,35,37]. Conventional wisdom would predict that the nature of the defect in NBS cells is due to the loss of the C-terminus of the protein. However, in a recent study, Maser and colleagues [44] elegantly demonstrated that, in cells from some NBS patients, an alternative mode of translation dependent on a cryptic internal ribosome entry site results in the production of an N-terminally truncated variant of the Nbs1 protein ($p70^{Nbs1}$) in addition to the expected C-terminally truncated variant ($p26^{Nbs1}$) [44]. Furthermore, it was found that $p70^{Nbs1}$ could associate with Rad50 and Mre11 whereas $p26^{Nbs1}$ could not [44]. Satisfyingly, this study reconciles the existence of people with Nijmegen-breakage syndrome with the observation that deletion of the *NBS1* gene leads to early embryonic lethality [45] – just like the deletion of *hRAD50* [46] – and suggests that the essential function of NBS1 is embedded in the C-terminal region. Thus, the defect observed in NBS cells might be the consequence of the dissociation of the N-terminus of Nbs1 from the MRN complex. As the only recognizable motifs in this segment are an FHA and a BRCT domain, it is therefore likely that loss of either or both domains from the MRN complex could explain the aetiology of the disease.

At a cellular level, the MRN complex strikingly localizes to sites of DNA strand-breaks after irradiation in a manner that is dependent on checkpoint kinase activity [47,48]. This relocation to sites of DNA damage is lost in NBS cells and a recent report supports the hypothesis that the Nbs1 FHA domain is necessary for this process [49]. However, in this latter study, the Nbs1 mutant used to ascribe a role to the FHA domain in MRN complex relocation was a deletion of the whole FHA domain, and therefore, one cannot exclude the possibility that the deletion may have altered the Nbs1 structure or conformation. A mouse mutant specifically mutated within the FHA domain of Nbs1 may resolve the role of this domain in the cellular pathology of the Nijmegen-breakage syndrome as well as the potential role of the FHA domain

in mediating MRN complex recruitment to sites of DNA damage. Needless to say, it will be fascinating to identify proteins that interact with the FHA domain of Nbs1 and to understand their role in the various functions of the MRN complex.

4.2. Kinesins

The kinesin superfamily is the largest family of FHA-containing proteins in eukaryotes. More than 20 different kinesin-like proteins present in species from worms to humans contain FHA domains. This subset of kinesin-like proteins is part of the N3 class of kinesins [50]. FHA-containing members of this class comprise kinesins of the UNC104 family (KIF1A, KIF1B, KIF1C and its orthologs such as *Caenorhabditis elegans* orthologue unc-104) and kinesins of the KIF14 family (including KIAA0042 and its orthologue from *Drosophila melanogaster*, KLP38b/*nebbish*/*tiovivo*) [51]. These kinesins are involved in vesicular transport (e.g. KIF1A/unc-104 and KIF1Bβ; [51–55]) or, in the case of KLP38b, in chromosome segregation [56–58]. Interestingly, a mutation in the gene encoding the KIF1Bβ protein is responsible for the Charcot-Marie-Tooth disease type 2A, indicating that FHA-containing kinesins may have a role in the development of human neuropathies [55].

Recent biophysical measurements have established that the kinesin FHA domain is an independently folded structural unit with a topology likely to be identical to that of the FHA domains of known structures [25]. Interestingly, the structure of the kinesin motor domain of KIF1A has been solved in both its ATP- and ADP-bound forms. The FHA domain is located close to the active site of the kinesin and this has led to the proposition that the FHA domain might play a role in regulating the catalytic cycle by binding to tubulin, the substrate of the kinesin [59]. Another equally interesting possibility is that the FHA domain is a site of phosphorylation-dependent cargo docking. As *C. elegans* UNC-104 and its mammalian counterpart, KIF1A/ATSV, transport synaptic vesicles, it is probable that the uptake of mature vesicles by kinesins is a regulated process. In any case, it will be exciting to test whether the transport process is regulated by FHA-mediated phospho-dependent interactions, whether it be via binding to tubulins, cargo proteins or other regulators.

Although no FHA-interacting proteins have yet been identified that interact with UNC-104-type family kinesins, a recent report suggests that the FHA domain of mouse KIF1C is functionally important [60]. In this study, mutation of the Pro residue at position 578 to Leu in the FHA domain of murine KIF1C is closely linked to resistance to anthrax lethal factor. Homology mapping of this residue on the available structures (Fig. 2A) suggests that Pro578 might play a role in the turn occurring in the β8/β9 loop and that substitution of Pro578 to Leu might result in a significantly altered FHA domain. It will be interesting to examine whether this mutation alters KIF1C motility or whether it affects the transport of a yet-to-be identified cargo.

4.3. Ring-finger proteins containing FHA domains: a role in protein degradation?

Ring-finger-containing FHA domains account for another class of FHA-containing proteins that is conserved from yeast to man. Two members of this family have been characterized so far: *Schizosaccharomyces pombe* dma1$^+$ and human Chfr. Fission yeast dma1$^+$ has been identified as a multi-copy suppressor of a *cdc16* mutation, which encodes the *S. pombe* homolog of the budding yeast *BUB2* spindle checkpoint/mitotic exit regulator [61]. Deletion of dma1$^+$ leads to sensitivity to microtubule-destabilizing agents, suggesting that, like *BUB2*, it may play a direct role in the spindle checkpoint pathway or the fission yeast equivalent of the mitotic exit pathway, the septation initiation network [61].

Exciting work from the Halazonetis laboratory has lead to the identification of Chfr as a protein required for a novel checkpoint that regulates the integrity of centrosome separation and prevents entry into metaphase in the presence of mitotic stress [62]. Interestingly, Chfr expression is undetectable in some cancer cells and the *CHFR* gene carries missense mutations in others, suggesting that inactivation of *CHFR* might play a role in cancer progression [62]. Studies on primary tumors and matched controls are required to definitely ascribe a role for Chfr in carcinogenesis [63]. Nevertheless, we know that the FHA domain of Chfr is required for its prophase checkpoint activity, as deletion of the FHA domain results in a dominant-negative protein that is unable to restrain metaphase entry in response to mitotic stress [62].

The presence of the ring-finger in this family of FHA-containing proteins is highly suggestive of a role in protein ubiquitination and/or degradation since it is the catalytic motif of a variety E3 ubiquitin ligases [64]. Furthermore, the inclusion of the FHA domain in these putative E3 ubiquitin ligases suggests that they will be recruited either to upstream regulators (such as an E2) and/or to potential substrates by phosphorylation-dependent protein:protein interactions.

4.4. Forkhead transcription factors

The term 'forkhead-associated' domain was originally given to the domain discussed in this review to underline the presence of this domain on a small subset of forkhead-type transcription factors [4]. In *S. cerevisiae*, four members of this subfamily exist (Fkh1, Fkh2, Flh1 and Flh2) and they form the largest family of FHA-containing proteins in this organism. FHA-containing forkhead-type transcription factors can be found in many eukaryotes; in humans, the MNF and ILF1/2 transcription factors are of this subfamily. Fkh1 and Fkh2 were recently found to be master regulators of G2-specific transcription in budding yeast [65–69]. They are components of the Swi5-factor (SFF) and regulate transcription of the *CLB2* and *SIC1* cluster [65–68]. The role(s) of the Fkh1 and Fkh2 FHA domains is still unknown, but a possible signal-dependent association with their cofactors Ndd1 and Mcm1 is worth investigating. Given the importance of these genes in budding yeast and the presence of FHA-containing forkhead-type transcription factors in mammals, it is tempting to speculate that mammalian counterparts of Fkh1/2 might play a similar role in regulating some aspects of cell cycle control.

4.5. Ki-67

Ki-67, a widely used proliferation marker [70], is a rather large protein containing an FHA domain in its N-terminus, followed by a large number of repeats reminiscent of PEST sequences which are involved in protein decay [70,71]. Ki-67 seems to be an essential component of cellular proliferation since antisense oligonucleotides directed against Ki-67 inhibit

cellular proliferation [71]. Its cellular localization is very dynamic during the cell cycle. During interphase, Ki-67 is mainly found associated with the nucleolus [70]. However, during mitosis, Ki-67 relocalizes to condensed chromosomes, suggesting that it my have a role in chromosome condensation. It will be interesting to examine whether the FHA domain is involved in the dynamic redistribution of Ki-67 at mitosis.

In an attempt to understand the function of the FHA domain of Ki-67, a yeast two-hybrid screen was undertaken by the group of Yoneda and colleagues. Ten positive clones representing two cDNAs were obtained, identifying the kinesin-like protein Hklp2 and a putative nucleolar RNA-binding protein, NIFK1, as specific interactors with the FHA domain of Ki-67 [72,73]. Both proteins are able to interact with a recombinant GST–FHA(Ki-67) domain fusion protein in a phospho-dependent manner, and both seem to partially overlap with Ki-67 in terms of cellular localization. However, the functional significance of these interactions remain to be demonstrated. Interestingly, after phosphorylation with cyclin-cdk kinases, recombinant portions of NIFK1 and Hklp2 were able to interact with recombinant FHA domain of Ki-67 [72,73]. Mapping of the NIFK1 and Hklp2 sites responsible for this interaction revealed that they contained a Thr–Pro consensus motif. This result suggests that the Ki-67 FHA domain may form a new subclass of FHA domains specialized in recognizing mitotic phosphoproteins via a pT–proline motif.

4.6. Nuclear inhibitor of protein phosphatase 1 (NIPP-1)

The NIPP-1 is a metazoan regulator of type 1 protein phosphatases (PP1). PP1 dephosphorylates Ser/Thr residues on a variety of proteins and the timely regulation of these dephosphorylation events is thought to be controlled by the association of PP1 with small regulatory subunits such as NIPP-1 [74,75]. The subcellular localization of NIPP-1 is striking being mainly in nuclear speckles, which are sites of splicing or storage of splicing factors [74,75]. It has therefore been proposed that NIPP-1 specifically targets PP1 to the splicing apparatus and that PP1 acts as a regulator of pre-mRNA splicing. Interestingly, mapping studies have revealed that the FHA domain of NIPP1 is required for the subnuclear localization of NIPP1 to the spliceosomal compartment [76], suggesting that the NIPP-1 FHA domain might have a targeting function.

Based on the conservation of the FHA domain among NIPP-1 orthologues and its newly identified role as a phosphopeptide-binding domain, Boudrez et al. [77] used the FHA domain of NIPP-1 as a bait in a yeast two-hybrid screen and identified the human homolog of *S. pombe* cdc5$^+$, CDC5L, as a specific interactor. The interaction is phospho-specific and cdk2/cyclin E can promote the CDC5L–NIPP-1 interaction in vitro. Mounting evidence supports a role in pre-mRNA splicing for CDC5L and its orthologues and, despite the absence of a perfect colocalization of NIPP1 with CDC5L, this interaction may be of physiological relevance, especially given the suspected role of the FHA domain of NIPP1 in regulating its subnuclear targeting [76].

5. A bacterial origin for the FHA domain: speculation and implications

A large number of prokaryotic genomes have been or are being fully sequenced. This endeavor has revealed a previously unsuspected but widespread presence of FHA-containing proteins in eubacteria but not, so far, in archaea. Although there is a lack of functional data on most of these proteins, the information we have suggests that FHA domains are most common in organisms having complex life cycles such as *Myxococcus xanthus* or in pathogenic bacteria such as mycobacteria. For example, in *M. xanthus*, the espA histidine kinase, which contains an FHA domain, is involved in the highly regulated process of sporulation [78].

The presence of genes encoding FHA-containing proteins in eubacterial and eukaryotic genomes, but not in archaeal genomes, is intriguing. It strongly suggests that lateral gene transfer has occurred to enable one kingdom to obtain the FHA domain from the other. However, if transfers have occurred, the directionality of the transfer remains unclear. For intracellular parasites, some of the first prokaryotes to be sequenced, the presence of putative FHA-containing proteins might be the result of these organisms capturing genes encoding FHA-containing proteins from their eukaryotic hosts. However, FHA-containing proteins are also found in a wide variety of bacterial species, some of them being free-living, such as *M. xanthus*, *Deinococcus radiodurans*, *Anabaena*, *Bacillus halodurans*, *Synechocystis*, *Streptomyces coelicolor*, *Pseudomonas aeruginosa*, and *Agrobacterium tumefaciens* (cf. http://smart.embl-heidelberg.de/). These are representatives of both Gram-positive and Gram-negative bacteria are found in diverse ecosystems, and the presence of FHA domains in proteobacteria [79] (the proposed ancestors of the mitochondrion) and cyanobacteria [80] (the ancestors of chloroplasts) raises the possibility that eukaryotic FHA domains have evolved from nuclear-transferred genes from endosymbionts. It is also worth noting that in plants such as *Arabidopsis thaliana*, the FHA domain-containing enzyme zeaxanthin epoxidase, which is a key enzyme in the production of the commercially important plant hormone abscisic acid, is localized in the chloroplast [81], further suggesting a possible endosymbiotic origin for eukaryotic FHA domains.

Whatever the origin of bacterial FHA domains, it remains intriguing that modular phosphopeptide recognition has evolved many times during evolution. This may indicate that some modules have biophysical properties that are best suited for some biological functions but not for others. For example, it is not hard to imagine how a phosphopeptide-binding module on signal transduction proteins may need properties that are distinct from those present on kinesins. Since the FHA domain appears to be an ancient module, one might also find instances where the FHA domain has been replaced by or has replaced another phosphopeptide-binding module during the course of evolution. This may be the case for the PinA protein of *Dictyostelium discoideum* (DdPinA, gi:1688322), a peptidyl-prolyl isomerase of the Pin1 family. Unlike its other eukaryotic counterparts, DdPinA does not contain a phosphopeptide-binding WW domain in its N-terminus, but instead has an FHA domain. It is not clear whether this represents a case where the FHA domain has replaced the WW domain, whether the ancestral Pin1 contained an FHA domain that was subsequently supplanted by the WW domain, or whether convergent evolution has occurred. In any case, it will be interesting to examine more closely the full extent of the binding selectivity and biophysical characteristics of the pT–FHA interaction for a large number of diverse FHA-containing proteins from

both prokaryotes and eukaryotes in order to glean some answers.

Finally, another implication of the widespread presence of FHA domains in bacteria is the obligatory corollary that phosphorylation-dependent assembly of protein complexes in bacteria should also be a widespread process. It is not yet clear what kind of phosphorylated amino acid residues bacterial FHA domains recognize in bacteria but it will be fascinating to examine how phosphorylation regulates protein complex assembly in these organisms.

6. The MH2 domain: a divergent FHA domain?

One of the most intriguing observations arising from the NMR and crystallography studies of FHA domains is the striking similarity between in the core β-sheet topology of the FHA and Smad MH2 domains [6,9]. The MH2 domain of Smad transcriptional regulators plays a critical role in Smad-dependent TGFβ signaling and acts as a protein:protein interaction interface [82]. In a recent report, Huse et al. [83] proposed that the conserved fold might indicate functional conservation and that the Smad MH2 domain serves as a binding interface for the tetraphosphorylated type I TGFβ receptor (TβR-I) [83]. The mapping of the interaction determinants on the MH2 domain revealed that the region required for the interaction of MH2 with the TβR-I-derived phosphopeptide coincides with the phosphopeptide interaction region of FHA domains, i.e. in the loops connecting the β-strands of the sandwich [83]. This striking similarity leads to the exciting possibility that the FHA and Smad MH2 domains might share a common ancestor that possessed phosphopeptide-binding activity. We await the co-crystallization of the MH2 domain bound to the phosphorylated TβR-I peptide, which should shed important light on how this conserved structure binds to multiply phosphorylated peptides.

7. Conclusion

The widespread structural spectrum of phosphopeptide recognition modules suggests that this strategy has been discovered multiple times during evolution, providing a testament to the selective advantage afforded by the strategy. Many major challenges remain for those of us interested in FHA domain function, including questions about the function of the FHA domain in the major classes of FHA-containing proteins, the role of FHA domains in prokaryotes, the exciting relation between FHA and Smad MH2 domains, and the molecular and structural determinants of ligand selectivity. We also note that, with the recent progress in developing efficacious anti-SH2 compounds [84], there is an obvious interest in identifying similar anti-FHA molecules that could block signal-dependent protein:protein interactions. These could prove to be of therapeutic value in a variety of contexts and might also provide valuable tools to probe FHA-dependent processes in vivo.

Acknowledgements: We are grateful to Frank Sicheri and Jane Bradbury for critical reading of the manuscript and for help in generating Fig. 2. D.D. is a recipient of a Hitchings-Elion fellowship of the Burroughs-Wellcome Fund and holds a Canada Research Chair in Proteomics, Bioinformatics and Functional Genomics. Work in the Durocher laboratory is supported by the Canadian Institutes for Health Research (Grant MOP49435) and work in the Jackson laboratory by grants from the Cancer Research Campaign.

References

[1] Kuriyan, J. and Cowburn, D. (1997) Annu. Rev. Biophys. Biomol. Struct. 26, 259–288.
[2] Pawson, T. and Nash, P. (2000) Genes Dev. 14, 1027–1047.
[3] Yaffe, M.B. and Smerdon, S.J. (2001) Structure (Camb.) 9, R33–R38.
[4] Hofmann, K. and Bucher, P. (1995) Trends Biochem. Sci. 20, 347–349.
[5] Durocher, D., Henckel, J., Fersht, A.R. and Jackson, S.P. (1999) Mol. Cell 4, 387–394.
[6] Durocher, D., Taylor, I.A., Sarbassova, D., Haire, L.F., Westcott, S.L., Jackson, S.P., Smerdon, S.J. and Yaffe, M.B. (2000) Mol. Cell 6, 1169–1182.
[7] Li, J., Smith, G.P. and Walker, J.C. (1999) Proc. Natl. Acad. Sci. USA 96, 7821–7826.
[8] Li, J., Lee, G.I., Van Doren, S.R. and Walker, J.C. (2000) J. Cell Sci. 23, 4143–4149.
[9] Liao, H., Byeon, I.J. and Tsai, M.D. (1999) J. Mol. Biol. 294, 1041–1049.
[10] Liao, H. et al. (2000) J. Mol. Biol. 304, 941–951.
[11] Durocher, D., Smerdon, S.J., Yaffe, M.B. and Jackson, S.P. (2000) Cold Spring Harbor Symp. Quant. Biol. LXV, 423–431.
[12] Stone, J.M., Collinge, M.A., Smith, R.D., Horn, M.A. and Walker, J.C. (1994) Science 266, 793–795.
[13] Sun, Z., Hsiao, J., Fay, D.S. and Stern, D.F. (1998) Science 281, 272–274.
[14] Stone, J.M., Trotochaud, A.E., Walker, J.C. and Clark, S.E. (1998) Plant Physiol. 117, 1217–1225.
[15] Neel, B.G. (1993) Semin. Cell Biol. 4, 419–432.
[16] Zhou, B.B. and Elledge, S.J. (2000) Nature 408, 433–439.
[17] Bell, D.W. et al. (1999) Science 286, 2528–2531.
[18] Roeder, G.S. and Bailis, J.M. (2000) Trends Genet. 16, 395–403.
[19] Zhou, Z. and Elledge, S.J. (1993) Cell 75, 1119–1127.
[20] Huang, M. and Elledge, S.J. (2000) Cold Spring Harbor Symp. Quant. Biol. LXV, 413–421.
[21] Emili, A. (1998) Mol. Cell 2, 183–189.
[22] Vialard, J.E., Gilbert, C.S., Green, C.M. and Lowndes, N.F. (1998) EMBO J. 17, 5679–5688.
[23] Pike, B.L., Hammet, A. and Heierhorst, J. (2001) J. Biol. Chem. 276, 14019–14026.
[24] Li, J., Williams, B., Haire, L.F., Goldberg, M., Durocher, D., Yaffe, M.B., Jackson, S.P. and Smerdon, S.J. (2001) in preparation.
[25] Westerholm-Parvinen, A., Vernos, I. and Serrano, L. (2000) FEBS Lett. 486, 285–290.
[26] Zhou, M.M. (2000) Nat. Struct. Biol. 7, 1085–1087.
[27] Wang, P., Byeon, I.J., Liao, H., Beebe, K.D., Yongkiettrakul, S., Pei, D. and Tsai, M.D. (2000) J. Mol. Biol. 302, 927–940.
[28] Bentley, N.J. and Carr, A.M. (1997) Biol. Chem. 378, 1267–1274.
[29] Durocher, D. and Jackson, S.P. (2001) Curr. Opin. Cell Biol. 13, 225–231.
[30] Abraham, R.T. (2001) Genes Dev. 15, 2177–2196.
[31] Shieh, S.Y., Ahn, J., Tamai, K., Taya, Y. and Prives, C. (2000) Genes Dev. 14, 289–300.
[32] Chehab, N.H., Malikzay, A., Appel, M. and Halazonetis, T.D. (2000) Genes Dev. 14, 278–288.
[33] Hirao, A. et al. (2000) Science 287, 1824–1827.
[34] Varon, R. et al. (1998) Cell 93, 467–476.
[35] Carney, J.P. et al. (1998) Cell 93, 477–486.
[36] Petrini, J.H. (2000) Curr. Opin. Cell Biol. 12, 293–296.
[37] Matsuura, S. et al. (1998) Nat. Genet. 19, 179–181.
[38] Wu, X., Webster, S.R. and Chen, J. (2001) J. Biol. Chem. 276, 2971–2974.
[39] Falck, J., Mailand, N., Syljuasen, R.G., Bartek, J. and Lukas, J. (2001) Nature 410, 842–847.
[40] Falck, J., Lukas, C., Protopopova, M., Lukas, J., Selivanova, G. and Bartek, J. (2001) Oncogene 20, 5503–5510.
[41] Usui, T., Ogawa, H. and Petrini, J.H. (2001) Mol. Cell 7, 1255–1266.

[42] D'Amours, D. and Jackson, S.P. (2001) Genes Dev. 15, 2238–2249.
[43] Grenon, M., Gilbert, C. and Lowndes, N.F. (2001) Nat. Cell Biol. 3, 844–847.
[44] Maser, R.S., Zinkel, R. and Petrini, J.H. (2001) Nat. Genet. 27, 417–421.
[45] Zhu, J., Petersen, S., Tessarollo, L. and Nussenzweig, A. (2001) Curr. Biol. 11, 105–109.
[46] Luo, G., Yao, M.S., Bender, C.F., Mills, M., Bladl, A.R., Bradley, A. and Petrini, J.H. (1999) Proc. Natl. Acad. Sci. USA 96, 7376–7381.
[47] Nelms, B.E., Maser, R.S., MacKay, J.F., Lagally, M.G. and Petrini, J.H. (1998) Science 280, 590–592.
[48] Maser, R.S., Monsen, K.J., Nelms, B.E. and Petrini, J.H. (1997) Mol. Cell Biol. 17, 6087–6096.
[49] Tauchi, H. et al. (2001) J. Biol. Chem. 276, 12–15.
[50] Miki, H., Setou, M., Kaneshiro, K. and Hirokawa, N. (2001) Proc. Natl. Acad. Sci. USA 98, 7004–7011.
[51] Bloom, G.S. (2001) Curr. Opin. Cell Biol. 13, 36–40.
[52] Hall, D.H. and Hedgecock, E.M. (1991) Cell 65, 837–847.
[53] Otsuka, A.J., Jeyaprakash, A., Garcia-Anoveros, J., Tang, L.Z., Fisk, G., Hartshorne, T., Franco, R. and Born, T. (1991) Neuron 6, 113–122.
[54] Yonekawa, Y. et al. (1998) J. Cell Biol. 141, 431–441.
[55] Zhao, C. et al. (2001) Cell 105, 587–597.
[56] Ohkura, H., Torok, T., Tick, G., Hoheisel, J., Kiss, I. and Glover, D.M. (1997) J. Cell Sci. 110, 945–954.
[57] Molina, I., Baars, S., Brill, J.A., Hales, K.G., Fuller, M.T. and Ripoll, P. (1997) J. Cell Biol. 139, 1361–1371.
[58] Alphey, L., Parker, L., Hawcroft, G., Guo, Y., Kaiser, K. and Morgan, G. (1997) J. Cell Biol. 138, 395–409.
[59] Kikkawa, M., Sablin, E.P., Okada, Y., Yajima, H., Fletterick, R.J. and Hirokawa, N. (2001) Nature 411, 439–445.
[60] Watters, J.W., Dewar, K., Lehoczky, J., Boyartchuk, V. and Dietrich, W.F. (2001) Curr. Biol. 11, 1503–1511.
[61] Murone, M. and Simanis, V. (1996) EMBO J. 15, 6605–6616.
[62] Scolnick, D.M. and Halazonetis, T.D. (2000) Nature 406, 430–435.
[63] Bachant, J.B. and Elledge, S.J. (1999) Nature 398, 757–758.
[64] Tyers, M. and Jorgensen, P. (2000) Curr. Opin. Genet. Dev. 10, 54–64.
[65] Koranda, M., Schleiffer, A., Endler, L. and Ammerer, G. (2000) Nature 406, 94–98.
[66] Kumar, R., Reynolds, D.M., Shevchenko, A., Goldstone, S.D. and Dalton, S. (2000) Curr. Biol. 10, 896–906.
[67] Pic, A. et al. (2000) EMBO J. 19, 3750–3761.
[68] Zhu, G., Spellman, P.T., Volpe, T., Brown, P.O., Botstein, D., Davis, T.N. and Futcher, B. (2000) Nature 406, 90–94.
[69] Simon, I. et al. (2001) Cell 106, 697–708.
[70] Endl, E. and Gerdes, J. (2000) Exp. Cell Res. 257, 231–237.
[71] Schluter, C., Duchrow, M., Wohlenberg, C., Becker, M.H., Key, G., Flad, H.D. and Gerdes, J. (1993) J. Cell Biol. 123, 513–522.
[72] Takagi, M., Sueishi, M., Saiwaki, T., Kametaka, A. and Yoneda, Y. (2001) J. Biol. Chem. 276, 25386–25391.
[73] Sueishi, M., Takagi, M. and Yoneda, Y. (2000) J. Biol. Chem. 275, 28888–28892.
[74] Boudrez, A., Evens, K., Beullens, M., Waelkens, E., Stalmans, W. and Bollen, M. (1999) FEBS Lett. 455, 175–178.
[75] Trinkle-Mulcahy, L., Ajuh, P., Prescott, A., Claverie-Martin, F., Cohen, S., Lamond, A.I. and Cohen, P. (1999) J. Cell Sci. 112, 157–168.
[76] Jagiello, I., Van Eynde, A., Vulsteke, V., Beullens, M., Boudrez, A., Keppens, S., Stalmans, W. and Bollen, M. (2000) J. Cell Sci. 21, 3761–3768.
[77] Boudrez, A. et al. (2000) J. Biol. Chem. 275, 25411–25417.
[78] Cho, K. and Zusman, D.R. (1999) Mol. Microbiol. 34, 714–725.
[79] Emelyanov, V.V. (2001) FEBS Lett. 501, 11–18.
[80] McFadden, G.I. (1999) Curr. Opin. Plant Biol. 2, 513–519.
[81] Marin, E., Nussaume, L., Quesada, A., Gonneau, M., Sotta, B., Hugueney, P., Frey, A. and Marion-Poll, A. (1996) EMBO J. 15, 2331–2342.
[82] Attisano, L. and Wrana, J.L. (2000) Curr. Opin. Cell Biol. 12, 235–243.
[83] Huse, M., Muir, T.W., Xu, L., Chen, Y., Kuriyan, J. and Massague, J. (2001) Mol. Cell 8, 671–682.
[84] Shakespeare, W.C. (2001) Curr. Opin. Chem. Biol. 5, 409–415.