PLOS | BIOLOGY

RESEARCH ARTICLE

The Structure of an NDR/LATS Kinase–Mob Complex Reveals a Novel Kinase–Coactivator System and Substrate Docking Mechanism

Gergő Gógl[1,2☯], Kyle D. Schneider[3☯], Brian J. Yeh[3], Nashida Alam[3], Alex N. Nguyen Ba[4], Alan M. Moses[4], Csaba Hetényi[5], Attila Reményi[1‡], Eric L. Weiss[3‡*]

1 Lendület Protein Interaction Group, Institute of Enzymology, Research Centre for Natural Sciences, Hungarian Academy of Sciences, Budapest, Hungary, **2** Department of Biochemistry, Eötvös Loránd University, Budapest, Hungary, **3** Department of Molecular Biosciences, Northwestern University, Evanston, Illinois, United States of America, **4** Department of Cell and Systems Biology, University of Toronto, Toronto, Ontario, Canada, **5** MTA-ELTE Molecular Biophysics Research Group, Eötvös Loránd University, Hungarian Academy of Sciences, Budapest, Hungary

☯ These authors contributed equally to this work.
‡ These authors are joint senior authors on this work.
* elweiss@northwestern.edu

OPEN ACCESS

Citation: Gógl G, Schneider KD, Yeh BJ, Alam N, Nguyen Ba AN, Moses AM, et al. (2015) The Structure of an NDR/LATS Kinase–Mob Complex Reveals a Novel Kinase–Coactivator System and Substrate Docking Mechanism. PLoS Biol 13(5): e1002146. doi:10.1371/journal.pbio.1002146

Editor: Mike Tyers, Mount Sinai Hospital, CANADA

Received: August 18, 2014

Accepted: April 2, 2015

Published: May 12, 2015

Data Availability Statement: Crystallographic data not contained in the paper have been deposited in PDB and can be found under the identifiers 4LQS, 4LQP, and 4LQQ. The information is publicly accessible via http://www.rcsb.org/pdb/home/home.do. All other data are contained within the paper and/or its Supporting Information files.

Funding: ELW received funding from NIH-NIGMS, grant #LP2013- 57, and from the American Cancer Society (RSG-06-0164-01-GMC). These organizations can be found online at www.nih.gov and www.cancer.org. AR received support from the Hungarian Academy of Sciences ("Momentum"

Abstract

Eukaryotic cells commonly use protein kinases in signaling systems that relay information and control a wide range of processes. These enzymes have a fundamentally similar structure, but achieve functional diversity through variable regions that determine how the catalytic core is activated and recruited to phosphorylation targets. "Hippo" pathways are ancient protein kinase signaling systems that control cell proliferation and morphogenesis; the NDR/LATS family protein kinases, which associate with "Mob" coactivator proteins, are central but incompletely understood components of these pathways. Here we describe the crystal structure of budding yeast Cbk1–Mob2, to our knowledge the first of an NDR/LATS kinase–Mob complex. It shows a novel coactivator-organized activation region that may be unique to NDR/LATS kinases, in which a key regulatory motif apparently shifts from an inactive binding mode to an active one upon phosphorylation. We also provide a structural basis for a substrate docking mechanism previously unknown in AGC family kinases, and show that docking interaction provides robustness to Cbk1's regulation of its two known in vivo substrates. Co-evolution of docking motifs and phosphorylation consensus sites strongly indicates that a protein is an in vivo regulatory target of this hippo pathway, and predicts a new group of high-confidence Cbk1 substrates that function at sites of cytokinesis and cell growth. Moreover, docking peptides arise in unstructured regions of proteins that are probably already kinase substrates, suggesting a broad sequential model for adaptive acquisition of kinase docking in rapidly evolving intrinsically disordered polypeptides.

program LP2013- 57; http://mta.hu/english) and was a Wellcome Trust International Senior Research Fellow (www.wellcome.ac.uk). BJY was a Damon Runyon research fellow (LP2013- 57; www.damonrunyon.org), and KDS was an NSF predoctoral fellow (www.nsf.gov). AR and CH were also supported by MedinProt (medinprot.chem.elte.hu) The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

Abbreviations: CLT, C-lobe tether; FP, fluorescence polarization; HM, hydrophobic motif; MD, molecular dynamics; MEN, mitotic exit network; RAM, regulation of Ace2 and morphogenesis; NCS, noncrystallographic symmetry; PDB, Protein Data Bank; WT, wild-type

Author Summary

The core organization of systems that relay information inside cells is preserved across vast evolutionary distances. Thus, detailed characterization of these systems' crucial modules can provide insight into the emergence and adaptation of signaling pathways, and illuminate broadly relevant mechanisms that control cells' diverse processes. In this paper we describe the first three-dimensional structure of a protein kinase–coactivator complex from budding yeast that is a key component of "hippo" signaling pathways, which direct cell proliferation, fate, and architecture in a wide range of eukaryotes. We show that this kinase–coactivator complex is a dynamic switch controlled by binding events distant from its active site, and that the kinase recognizes specific short motifs in disordered regions of target proteins by a previously unknown mechanism. This substrate docking interaction provides in vivo robustness to the kinase's regulation of its known targets, and identifies likely new substrates that expand our view of this hippo pathway's role in cell division. Moreover, during the course of evolution, the short motif that interacts with the kinase's docking surface appears in rapidly changing intrinsically disordered regions of a number of proteins that are probably already in vivo substrates. Thus, our findings support the idea that proteins evolve more robust functional links to the signaling networks that control them by acquiring short peptide motifs that interface with key conserved signaling modules.

Introduction

"Hippo" signaling pathways control diverse aspects of cell proliferation, survival, and morphogenesis in eukaryotes. The core organization of these networks is conserved over a billion years of evolution, with related forms described in animals and fungi [1–3]. In these systems, MST/hippo kinases activate NDR (nuclear Dbf2-related) or LATS (large tumor suppressor) kinases (Fig 1), which are closely related members of the large AGC family of protein kinases. The NDR/LATS kinases bind to highly conserved Mob coactivators, forming a regulatory complex that controls a diverse set of in vivo effector proteins.

In animals, a form of hippo signaling inhibits cell proliferation and controls tissue architecture [4,5]. In humans and *Drosophila*, LATS kinases in complex with Mob1 family proteins phosphorylate YAP/Yki family transcriptional coactivators, which promote cell division and resistance to apoptosis (Fig 1, left). This sequesters YAP/Yki proteins in the cytoplasm, blocking their activity and thus suppressing cell proliferation and increasing sensitivity to programmed cell death. Notably, activation of metazoan LATS by MST/hippo kinases appears to depend on the kinases' recruitment to a scaffold protein, and is linked to engagement with cytoskeletal elements and structures at the cell cortex [6]. *Drosophila* and human cells also use another form of hippo signaling, in which MST/hippo activates NDR kinases that form complexes with a different Mob coactivator [3,7]. These pathways control morphogenesis of cell extensions and participate in cell proliferation control, but are much more dimly understood. In *Drosophila*, this "NDR branch" of hippo signaling is important both for proper organization of actin-rich cell extensions and organization of sensory neuron dendrites [8,9]. In mammalian cells, NDR kinases appear to help drive the transition through the G1 restriction point [10].

Hippo pathways are strongly associated with cell proliferation in the budding yeast *Saccharomyces cerevisiae*. One such system called the mitotic exit network (MEN) is centrally important for the transition from mitosis to G1 [11] (Fig 1, right). In this pathway, the yeast hippo-

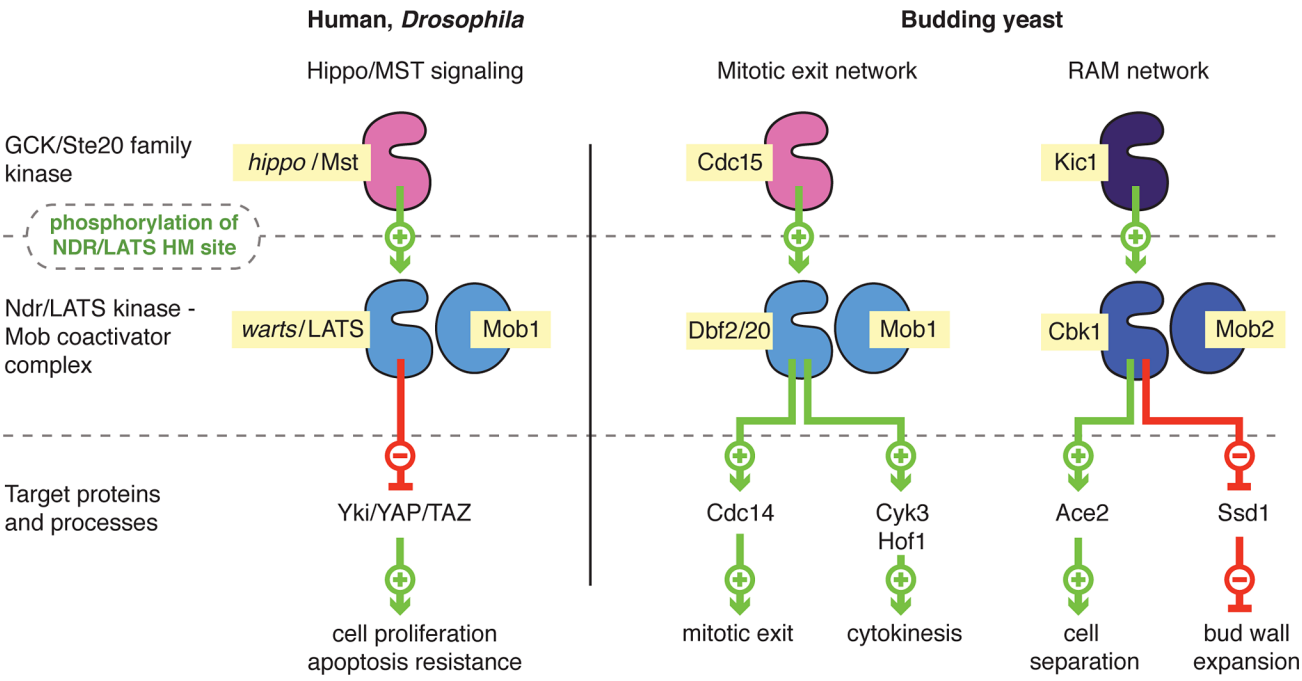

Fig 1. Role of the NDR/LATS kinase–Mob complex in hippo signaling pathways. Left: A metazoan hippo pathway in which a LATS–Mob1 complex is activated by MST/hippo kinases and inhibits the Yki/YAP transcriptional coactivator, based on research from *Drosophila* and mammalian cells. Right: Budding yeast have two distinct hippo pathways, the mitotic exit network (MEN), in which the LATS-related Dbf2 or Dbf20 kinase in complex with Mob1 controls mitotic exit and cytokinesis, and the RAM network, in which the NDR-related Cbk1 kinase in complex with Mob2 controls cell separation and morphogenesis.

doi:10.1371/journal.pbio.1002146.g001

like kinase Cdc15 activates the LATS-related kinases Dbf2 and Dbf20, which bind the Mob1 coactivator protein. The MEN controls key cytokinetic processes and helps flood the cytosol with the phosphatase Cdc14, reversing phosphorylation performed by mitotic CDK [12,13]. MEN activation involves colocalization of Cdc15 and the Dbf2/20–Mob1 complex to the Nud1 scaffold protein at the yeast spindle poles; this recruitment is closely coordinated with the onset of anaphase [12,14].

Budding yeast cells also use an NDR branch of hippo signaling, the RAM network, during the M to G1 transition (Fig 1, right) to initiate the final event of cell division: destruction of an extracellular septum that forms between mother and daughter cells during cytokinesis [15]. In the RAM network, the hippo-like kinase Kic1 activates the NDR-related kinase Cbk1, which forms a complex with the coactivator protein Mob2 [16,17]. Cbk1–Mob2 directly drives nuclear accumulation of the transcription factor Ace2, which turns on expression of septum-destroying hydrolases [18,19]. Intriguingly, Ace2 strongly localizes only to nuclei of newly born daughter cells [20,21]. This produces a sharp expression peak of mother/daughter separation genes that happens only once in a cell's lifetime, when it is newly born. The RAM network is also required for normal growth and morphogenesis of proliferating and mating cells. While the mechanisms of this regulation are incompletely understood, it partly involves Cbk1's inhibitory phosphorylation of the RNA-binding protein Ssd1. This allows translation of secreted enzymes that open up the cell wall lattice, allowing its expansion and remodeling in actively growing cell regions [20–23].

NDR/LATS kinase–Mob coactivator modules are centrally important in all known hippo pathways. The catalytic domains and flanking regions of these kinases are extraordinarily well conserved (S1 Fig), as are Mob family proteins [24]. Thus, the NDR/LATS–Mob complex

probably has distinctive biochemical characteristics that are specifically adapted to the molecular organization and output requirements of hippo signaling systems in a wide range of eukaryotes. However, these complexes are poorly understood at a structural and mechanistic level. For example, the role of Mob coactivator binding is not understood, and it is uncertain how MST/hippo kinase activation of NDR/LATS kinase–Mob modules works at a biochemical level. For reasons that are unclear, this activation appears to be extremely inefficient in the absence of scaffolding proteins that bring the pathway components together [6,12,14,16,25]. Moreover, relatively few NDR/LATS phosphorylation targets are known, and understanding how these complexes select substrates could identify new ways in which they regulate cellular processes. Furthermore, in addition to illuminating the biology of this ancient pathway, a deep mechanistic understanding of this complex could suggest ways conserved signaling systems change and adapt to control diverse functions as organisms evolve.

Here we describe the crystal structure of the yeast Cbk1–Mob2 complex, providing what is to our knowledge the first structural template for an NDR/LATS kinase–Mob coactivator complex. Combined with molecular dynamics (MD) simulation, this analysis indicates how members of this structurally unexplored kinase group are activated. We find that Mob protein association creates a novel binding pocket that participates in the formation of the active state of NDR/LATS kinases after they are phosphorylated by MST/hippo kinases. This structure also allows us to define how Cbk1's kinase domain associates with a short peptide motif in its known in vivo phosphorylation targets, a form of substrate docking not previously seen in the broader AGC family of protein kinases. We find that substrate docking confers robustness to Cbk1's regulation of these proteins. Furthermore, conservation of the short peptide that docks to the Cbk1 kinase domain identifies known targets and strongly predicts an expanded set of kinase substrates, most of which are expected to be involved in cell morphogenesis and cytokinesis. Our analysis also indicates that Cbk1–Mob2 regulatory targets first acquired NDR/LATS phosphorylation consensus sites and subsequently evolved docking motifs in unstructured protein regions, supporting a sequential model for the evolution of docking interactions in kinase–substrate networks.

Results

The Structure of Cbk1–Mob2 Reveals a Novel Kinase–Coactivator Complex

Mechanistic insight into the regulation and protein–protein interactions of NDR/LATS kinase–Mob coactivator complexes is limited by absence of crucial information about this important regulatory module's three-dimensional structure. The budding yeast NDR/LATS kinase Cbk1 and the coactivator protein Mob2 contain characteristic sequence features and regulatory inputs that are preserved across evolution, and have a number of well-characterized in vivo functions. The complex formed by these proteins is therefore a suitable model for Mob coactivator binding, kinase activation, and substrate phosphorylation. To provide the first structural template for NDR/LATS kinase–Mob coactivator complexes, we determined the structure of yeast Cbk1 bound to Mob2 in three different crystal forms (S2 Fig; S1 Table).

We used two variants of catalytically inactive Cbk1(D475A) for crystallography. These differ at position 743, a highly conserved threonine and the phosphorylation site through which MST/hippo kinases (in this case, Kic1) activate NDR/LATS kinases [1,16,26,27]. This phosphorylation site is within a C-terminal catalytic domain extension, and is part of a regulatory hydrophobic motif (HM) that is commonly present in AGC group kinases [28,29]. In addition to Cbk1 intact in its HM region, we used a variant in which threonine 743 was replaced by glutamic acid, an activating substitution that is at least partially phosphomimetic [30,31]. Kic1

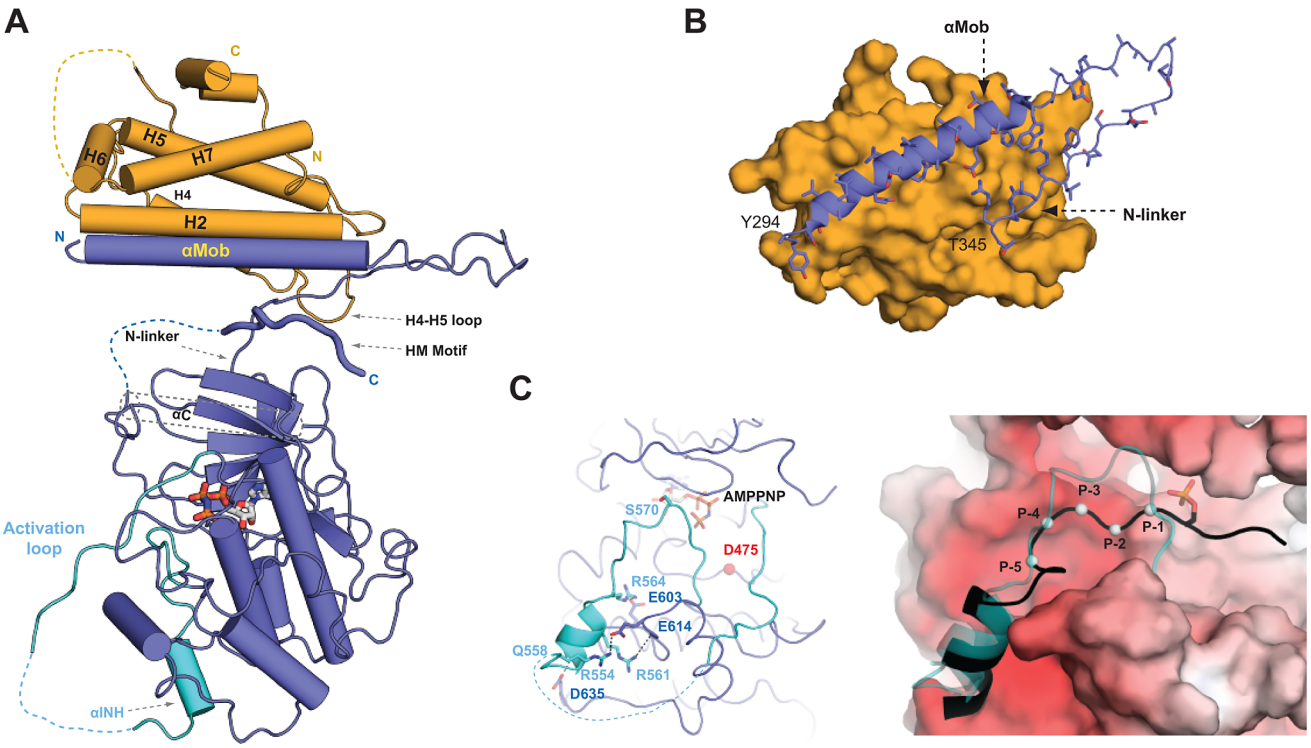

Fig 2. Structure of the Cbk1–Mob2 complex. (A) An overview of the Cbk1–Mob2 complex. Mob2 (orange) binds to the N-terminal region of Cbk1 (blue) through a large surface formed by the H2 and H7 helix and the H4–H5 loop. "N" and "C" denote the N- and C-terminal ends of the protein constructs. Flexible regions that could not be located in the electron density maps are shown with dashed lines. HM is the AGC kinase HM, and "αINH" denotes the inhibitory alpha helix that is part of the activation loop (shown in cyan). (B) The N-terminal region of the Cbk1 forms a bipartite Mob2-binding surface comprised of a long αMob helix and a highly conserved arginine rich N-linker region. (C) An activation loop segment (cyan) of Cbk1 containing the major autoregulatory site (S570) blocks access to the active site (D475) (left panel). An inhibitory helix (αINH in [A]) is anchored in the substrate-binding groove by several salt bridges (shown with black dotted lines). The substrate binds in the negatively charged "open" binding pocket that replaces αINH (shown in cyan), as seen in the state with the bound αINH (right panel). The model for the Cbk1–substrate complex was generated by superimposing the crystallographic model of the "open" state of Cbk1 with PKA binding to a phosphorylated substrate (Protein Data Bank [PDB] ID: 1JLU). The black cartoon with gray spheres (from P-1 to P-5 Cα positions) shows the PKA substrate superimposed in the Cbk1 substrate-binding pocket, while the activation loop of Cbk1, which displays the characteristics of a pseudo-substrate region, is shown with the semitransparent cartoon representation in cyan. The surface of Cbk1 is colored according to its electrostatic potential, where red indicates negatively charged surface. The nucleotide cofactor (AMPPNP) is shown in stick representation.

doi:10.1371/journal.pbio.1002146.g002

phosphorylates this threonine extremely inefficiently in vitro [16], making production of Cbk1 with phosphothreonine at position 743 unfeasible in quantities suitable for biochemical and structural analysis. For clarity, we refer to the HM in which threonine 743 is not phosphorylated as "HM-T," the glutamic acid substitution at this site as "HM-E," and the form in which threonine 743 is phosphorylated as "HM-P."

We collected crystallographic data on one Cbk1(HM-T)–Mob2 complex that diffracted to 3.3 Å, with one Cbk1–Mob2 complex in the asymmetric unit. We also solved the structure of Cbk1(HM-E)–Mob2 in a crystal form that diffracted to 3.6 Å, containing two Cbk1(HM-E)–Mob2 complexes in the asymmetric unit related by noncrystallographic symmetry (NCS). Another crystal form of the Cbk1(HM-E)–Mob2 complex diffracted to 4.5 Å, with one complex in the asymmetric unit.

The domain arrangements for the kinase and its coactivator are the same in all final crystallographic models (S2A Fig). In each structure, Mob2 binds to the conserved N-terminal extension of the Cbk1 AGC kinase domain, a region previously identified as the site by which NDR kinases bind Mob proteins [32] (Fig 2A). A portion of Cbk1's N-terminal extension forms a long helix (termed "αMob"), which is a major site of contact with Mob2 (Fig 2B). The αMob

helix is followed by a sharp turn and then a short segment, here termed the "N-linker," that connects to the core kinase catalytic domain. Mob2 also interacts with the N-linker, which together with the αMOB helix forms a >1,100 Å^2 bipartite interaction surface between the kinase and its coactivator. All structures of the Cbk1–Mob2 complex reveal a deep cleft between the kinase N-terminal lobe and the Mob-binding interface formed by αMOB and the N-linker. The crystal structure of human and budding yeast Mob1 proteins have been solved [33,34], and the fundamental structure of Mob2 in complex with Cbk1 is essentially similar.

In addition to their Mob-binding N-terminal kinase domain extensions, NDR/LATS kinases typically have notably elongated activation loop segments [35]. Superimposing our Cbk1 structures onto those of ROCK and PKA [36,37] shows that both of these Cbk1 regions adopt distinctive conformations (S2B Fig). Cbk1's activation loop forms multiple different structures: the higher resolution Cbk1(HM-E)–Mob2 crystallographic model captures two of these conformations displayed by the two NCS-related models (Fig 2C). In one form, structured parts of the activation loop occupy the substrate-binding pocket and thus occlude access to the catalytic site, similar to the structure of PKA complexed with an alpha helical peptide inhibitor (S2C Fig) [35,38]. In the other complex, the activation loop is not present in the substrate-binding region. Cbk1(HM-T)–Mob2 complex crystallized in only one of these states, and its substrate-binding pocket was blocked. We suggest that this is an auto-inhibited form of the kinase, while the other state seen in one of the two complexes present in Cbk1(HM-E) crystals may be competent to bind substrates containing basic amino acids (Fig 2C, right). Although crystal packing interactions around the kinase activation loop are clearly different for the two complexes of the Cbk1(HM-E)–Mob2 model, it is likely that the extended activation loop of NDR/LATS kinases switches them between an inhibited and open state, with phosphorylation of the conserved autoregulatory site S570 (near AMPPNP in Fig 2C, left) alleviating the pseudo-substrate-based inhibition exerted by the inhibitory helix (αINH). The activation loop displays dramatically different organization between each crystal form as well as between NCS-related molecules of the same crystal (S2D Fig), and none of these contain an intact catalytic center that can phosphorylate the autoregulatory site, preventing further speculation. Thus, further work will be necessary to decipher the regulatory role of activation loop inhibition and alleviation by autophosphorylation, as well as to identify the structural elements important for substrate binding.

In many AGC kinases, the C-terminal HM region associates with the catalytic domain's N-terminal lobe in a conformation that stabilizes the important C helix, promoting enzyme activation. In contrast, the Cbk1–Mob2 complex structures we solved exhibit a novel organization of the HM region and activating phosphorylation site. In all three structures, the Cbk1 HM peptide is wedged into a binding slot in the deep cleft formed by the kinase's αMob helix, N-linker region, and N-terminal catalytic domain lobe (Figs 2A and S2E and S2F). Mob2 does not contact the HM region directly, but rather holds the HM-binding cleft formed by Cbk1's N-terminal extension open. In this location the HM peptide probably cannot productively structure the C helix; consistent with this, the C helix region is disordered in all structures we solved (Figs 2A and S2A and 3A).

Cbk1 HM Phosphorylation Suggests an Activation Mechanism Unique to NDR/LATS Kinases

How does the conserved phosphorylation site in the NDR/LATS HM region regulate these kinases? While replacement of the phosphorylated threonine with glutamic acid at this site increases Cbk1's activity in vitro and in vivo, it is uncertain how closely this mutation mimics the biochemical effect of phosphorylation. In our crystal structures, unphosphorylated wild-type (WT) HM region (HM-T) and HM region with glutamic acid substitution (HM-E) associate

with the Mob2-supported binding site in a similar manner. If these closely related structures capture the HM region in its kinase-activating conformation, then phosphorylation of the HM site might be expected to strongly increase the HM region's affinity for Cbk1. We measured the affinity of HM peptides added in trans to Cbk1–Mob2 complex in which Cbk1's native HM region was truncated. Phosphorylated peptide (HM-P) bound with an affinity of approximately 6 μM, while the dephosphorylated peptide (HM-T) bound with an affinity of about 33 μM (S3 Fig). The HM–Cbk1 interaction is intramolecular, putting the HM peptide's effective local concentration in the millimolar range. The modest difference in affinity of HM and HM-P peptides argues that HM site phosphorylation does not achieve its regulatory effect by enhancing the region's ability to bind at a key regulatory site on the kinase.

Since it was not possible to produce Cbk1 with phosphorylated HM (HM-P) for biochemical and crystallographic characterization, we performed MD simulations using the Cbk1–Mob2 structural template to develop a plausible model for the function of HM phosphorylation. Since the αC helix of Cbk1 was not resolved, we used homology modeling to build a folded αC based on the structure of activated PKB [39]. We then conducted 125-ns MD simulations using models with either HM-T or HM-P (Figs 3A and S4). To estimate the mobility and position of Cbk1's HM region in these two models, we calculated the distance between the centers of mass of peptide segments encompassing Cbk1's N-linker, αMob, αC, and HM regions (Fig 3B).

This analysis clearly predicts that phosphorylation of Cbk1's HM site significantly affects both the HM region's mobility and the structure of the site it associates with in the Cbk1–Mob2 complex. As noted in Fig 3C, the mobility of HM-P is significantly lower than the mobility of HM-T, and HM-P and the αC, N-linker, and αMob segments are closer to each other (Fig 3C). Cbk1's N-linker region is compressed as the HM-P fragment binds in a region analogous to the PIF pocket and directly interacts with αMob, producing a bend in αMob that does not occur in the HM-T model (Fig 3D and 3E). The simulations predict that phosphorylated T743 compresses the N-linker pocket by making salt bridge contacts with αMob. Notably, two residues (R746 and F747) make direct contacts with αMob and αC (Fig 3E). These are invariant in the NDR/LATS lineage, but are not conserved in AGC kinases that lack Mob cofactors. Despite αMob reorganization, MD simulation indicates that the Mob interaction interface probably remains largely unchanged when HM-P interacts (S4A Fig). Moreover, αMob's interaction with HM-P occurs only when the Mob coactivator is present (S4B Fig). Similar MD analysis of the Cbk1(HM-E)–Mob2 structure indicates that HM-E also has more restrained movement of the HM region within the binding slot than HM-T, but slot compression was significantly less pronounced compared to HM-P (S4C Fig). This suggests that HM-E is only a partial substitute for HM-P.

MD simulation suggests that HM site phosphorylation promotes formation of multiple electrostatic salt bridge interactions. We therefore conducted in vivo charge swap experiments with two of these putative salt bridges, R307/pT743 and E314/R746 to test the validity of the simulations. As outlined in Fig 1, Cbk1 is required for mother/daughter separation; pathway output can be measured by counting the number of cells present in physically associated groups in liquid culture [19]. WT budding yeast cells separate well (about 90% of cell groups contain one to four cells), while *cbk1Δ* strains fail to separate: they grow in large clumps with about 80% of groups containing five or more cells (Fig 3F). We introduced single and double mutations of the putative HM salt bridge interactions (R307E/T743R and E314R/R746E) and measured cell group sizes. Both charge swaps showed recovery in the double mutant relative to the single mutants, with E314R/R746E returning to 50% in groups with one to four cells per group and $<5\%$ with 21 or more cells per group relative to the R746E single mutant levels of 20% and 30%, respectively. The R307E/T743R mutant also displayed some recovery, with

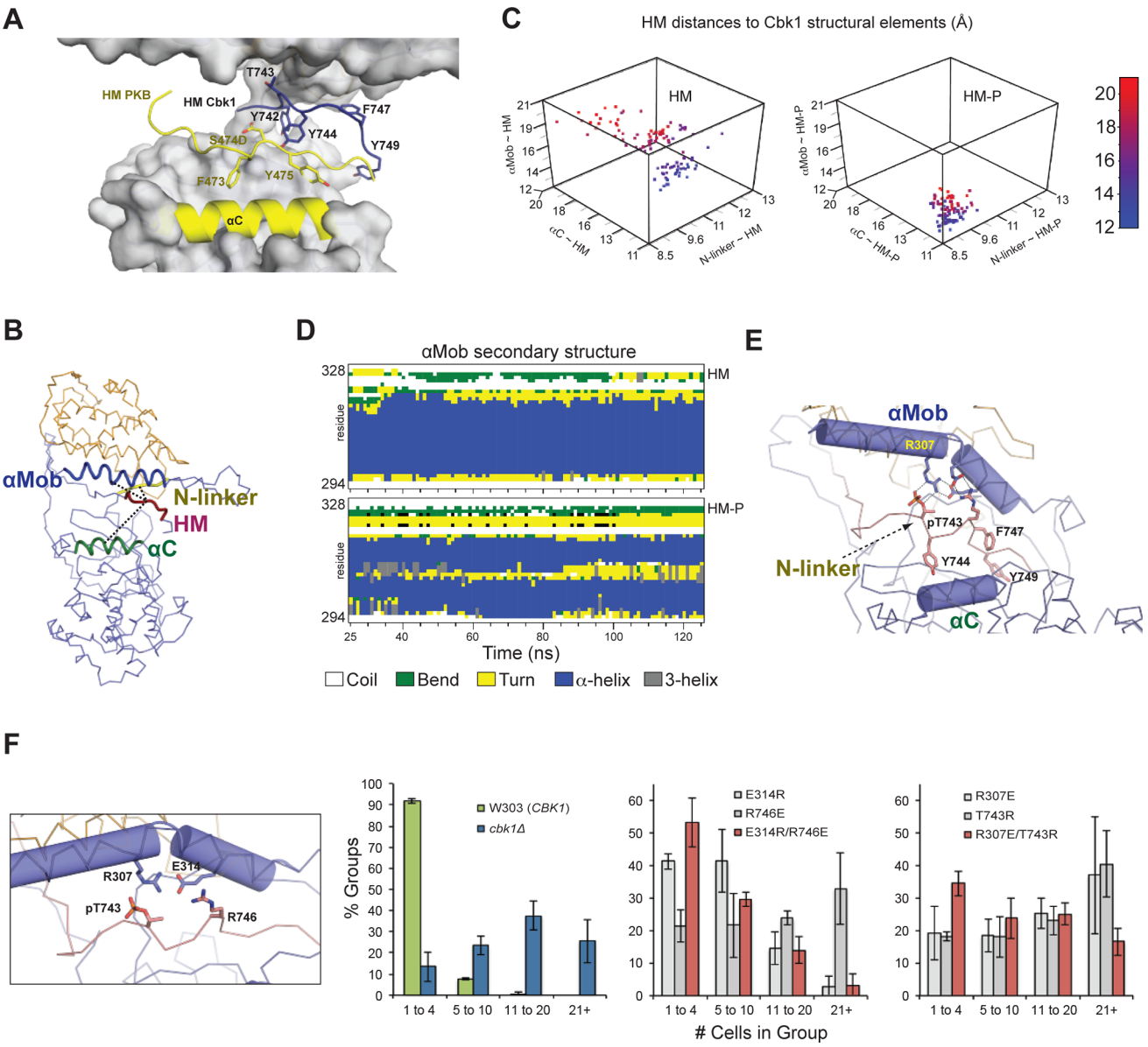

Fig 3. Molecular dynamics simulation indicates phosphorylation-induced rearrangement of the NDR/LATS kinase HM-binding slot to favor enzyme activation. (A) Comparison of the Cbk1 HM with the HM of PKB. The panel shows the Cbk1–Mob2 complex superimposed with PKB, where the αC helix and the HM of the superimposed PKB are shown in yellow (PDB ID: 1O6K). Side chains of important HM residues are shown with sticks. Superimposition of PKB and Cbk1–Mob2 reveals that the main chains of the HM motifs are >7 Å apart. The Cbk1 HM motif is shifted upwards to bind to the upper part of the HM-binding slot, while the PKB HM binds close to the N-terminal PKB kinase lobe. (B) Cbk1–Mob2 complexes in which T743 in the C-terminal HM is either dephosphorylated (HM-T) or phosphorylated (HM-P) were subjected to 125-ns MD simulations. Here we represent indicated regions of Cbk1 as centers of mass of the following: HM, amino acids 742–749; N-linker, amino acids 341–346; αC, amino acids 393–406; αMob, amino acids 297–317. We started distances of both HM forms to the N-linker region, αC, and αMob at [8.7 Å; 18.0 Å; 14.9 Å], corresponding to the conformational state captured in the Cbk1–Mob2 complex crystal structure, and recalculated the distances between HM forms, N-linker, αC, and αMob after each simulation step. (C) Three-dimensional scatter plots show 100 ns of the simulations, with each dot indicating N-linker–HM, αC–HM, and αMob–HM distances (*x*-, *y*-, and *z*-axes) for each nanosecond. For clarity, dots are colored from blue to red according to their αMob–HM distances (*z*-axis). The complex with HM-P exhibits shorter distances and smaller changes in position, indicating that T743 phosphorylation compresses the HM-binding slot and constrains the HM region association. This occurs only in Cbk1 bound to Mob coactivator (see S5 Fig). (D) HM phosphorylation promotes αMob bending and compresses the HM binding slot. Secondary structure analysis of αMob illustrates that this region remains helical (blue) when HM-T is bound, but acquires a short turn (yellow) when HM-P is bound. (E) An enlarged view of the final MD-simulated Cbk1(HM-P)–Mob2 complex indicates the HM-P's interactions that bend αMob, as well as van der Waals contacts with αC. (F) Ionic interactions suggested by the MD model (E314/R746 and R307/pT743) (left) were tested by charge swapping and analysis of cell separation. Single mutants (grey) displayed larger group sizes, indicating defective RAM network function, but were partially recovered when both charged sites were swapped (red). Data for (C) and (F) can be found in S1 Data.

doi:10.1371/journal.pbio.1002146.g003

~35% in groups with one to four cells per group and ~20% in groups of 21 or more. The lower degree of recovery for this latter mutant is consistent with MD simulation, which suggests that pT743 engages in multiple interactions that the R307E/T743R variant would not recapitulate. In summary, our findings support an activation model specific to NDR/LATS kinases that requires Mob cofactor binding, in which HM site phosphorylation induces compression of the N-linker region to produce a PIF-pocket-like binding site.

A Conserved Substrate Docking Motif Associates with a Specific Site on Cbk1's Kinase Domain

We sought to use the crystal structures of Cbk1–Mob2 to better understand how the complex recognizes targets of regulatory phosphorylation in vivo. The sites that Cbk1 and other NDR/LATS kinases phosphorylate provide some specificity: many kinases prefer phospho-acceptor sites with specific amino acids immediately nearby (HxRxx[ST] in the case of Cbk1 and other NDR/LATS kinases) [18,19,40,41]. However, these sequence motifs are generally not sufficient to distinguish in vivo substrates from proteins that contain them by chance. Notably, it has been discovered that some protein kinases bind additional short specificity-enhancing motifs (referred to as docking motifs) in substrate proteins [42,43]. We previously found a conserved short peptide motif ([YF][QK]FP) on the Cbk1 substrates Ace2 and Ssd1 that associates with bacterially expressed Cbk1 in vitro [44] (Fig 4A). As described below, we find that specific features of peptides of this motif associate with a site on Cbk1's catalytic domain, and that the segments' presence in substrate proteins significantly enhances phosphorylation by Cbk1 in vitro. We therefore consider them bona fide Cbk1 docking motifs, and refer to them as such [42,43].

We first characterized binding of Ace2 and Ssd1 peptide segments to Cbk1's catalytic domain and the Cbk1–Mob2 complex in vitro. This showed that a 15-amino-acid peptide from Ace2 with the sequence YQFP at its C-terminal end associates with the Cbk1–Mob2 complex, and that two shorter peptides from Ssd1 containing the motif FKFP also interact with this complex (Figs 4B and S5A and S5E). Ace2 and Ssd1 peptides compete for binding, indicating that they likely associate with the same kinase surface (S5G Fig). Fluorescence polarization (FP) analysis using labeled synthetic peptides shows that the Ace2 peptide binds Cbk1–Mob2 with a ~6 μM K_d, and a peptide from Ssd1 containing the FKFP motif binds with a ~20 μM K_d (S5H Fig). Both the Ace2 and Ssd1 fragments interact with Cbk1 lacking N- and C-terminal extensions, indicating that their association occurs within the catalytic domain (S5D and S5I Fig).

To define the Ace2 and Ssd1 peptide features that mediate association with Cbk1, we systematically introduced amino acid substitutions into the peptides and assessed their binding to the Cbk1 catalytic domain and the Cbk1–Mob2 complex. Substitution of alanine for either of the aromatic amino acids or for the proline in the [YF]xFP motif eliminates in vitro association with Cbk1, demonstrating that this amino acid sequence is critical for the peptides' interaction with the Cbk1 kinase domain (Fig 4C and 4D). We thus refer to [YF]xFP as the "core motif" required for Cbk1 association.

The [YF]xFP core motif is conserved and critical for Cbk1 binding, but the Ace2 and Ssd1 peptides exhibit distinct preferences for amino acids within it and in the surrounding sequence. Several lines of evidence suggest that Ace2's Cbk1-binding peptide has a functionally bipartite structure. First, a short region at the N-terminal end of this 15-amino-acid-long peptide is important for maximal association with Cbk1–Mob2 (Fig 4C); the requirement for these N-terminal amino acids is more pronounced in assays performed with the catalytic domain alone (S5B Fig). Second, replacement of amino acids near the center of Ace2's Cbk1-binding peptide with alanine does not affect interaction with the kinase. Third, the Ace2 peptide contains two consecutive glycine residues N-terminal to the core motif, and variants containing one to four

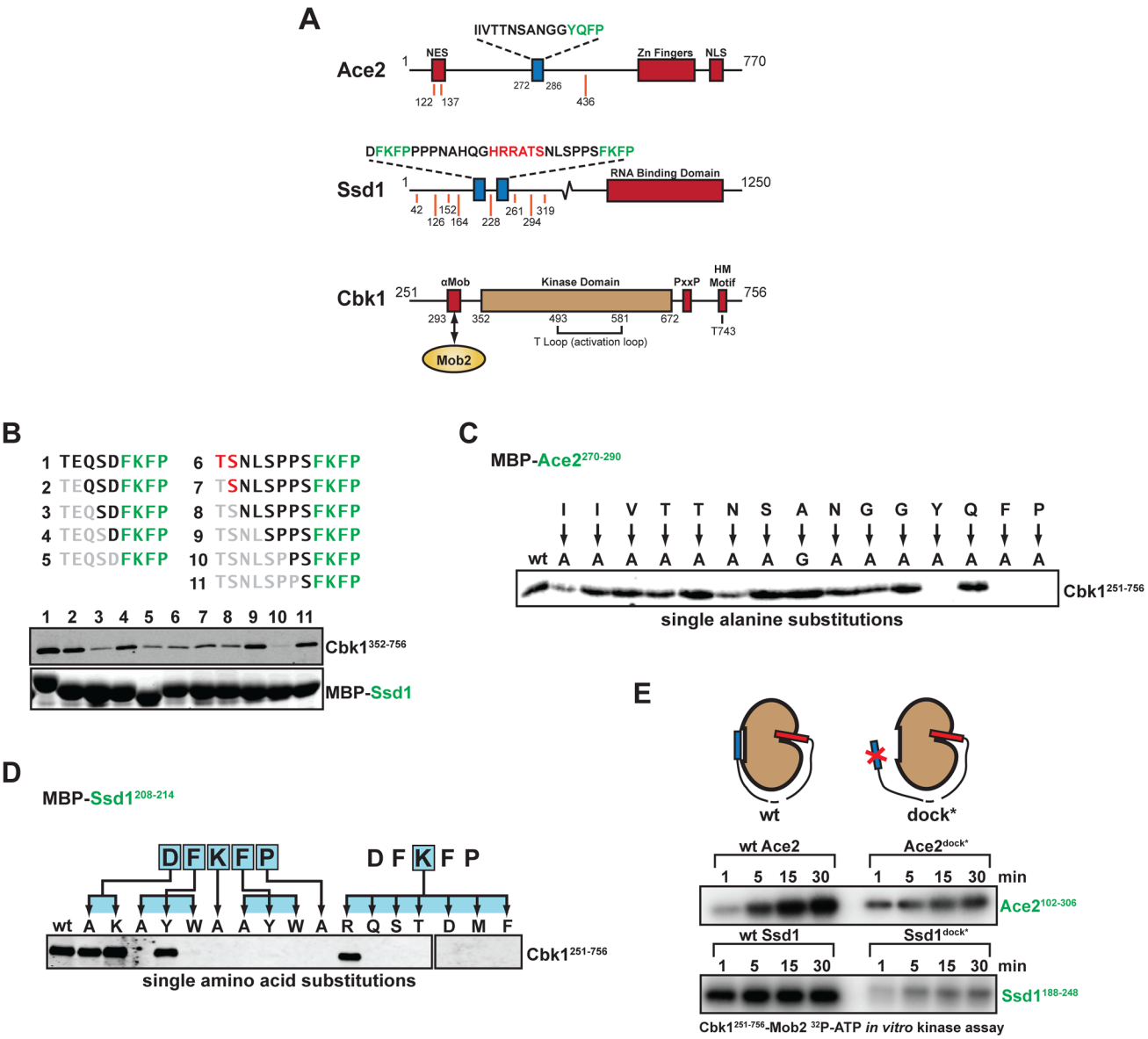

Fig 4. Molecular analysis of Cbk1 docking peptides in Ace2 and Ssd1 highlights importance of a [YF]xFP motif. (A) Protein schematics with points of interest highlighted. Orange lines denote Cbk1 consensus sites, and blue boxes denote docking motifs. (B) Pulldown of Cbk1 kinase domain by truncated Ssd1 N-terminal (1–5) and C-terminal (6–11) docking motifs suggests that only the FKFP motif is required for interaction. (C) Alanine scan of the Ace2$^{270–290}$ docking motif suggests that residues N-terminal to the YQFP motif aid in Cbk1 binding. (D) Mutational analysis of Ssd1 N-terminal docking motif highlights the sequence stringency of the core motif and suggests a consensus docking motif of [YF][KR]FP. (E) Cbk1 in vitro kinase assay with Ace2 or Ssd1 fragments containing a phosphorylation site (HxRxx[ST]) and either a WT docking motif (left) or mutated docking motif (dock*, right). Phosphorylation is enhanced in the presence of the WT docking motif.

doi:10.1371/journal.pbio.1002146.g004

glycines still bind Cbk1, while eliminating these residues abolishes interaction (S5C Fig). Thus, the Ace2 docking peptide probably has two distinct short Cbk1-binding epitopes connected by a flexible linker.

In contrast, Ssd1's Cbk1-binding peptide requires only the core [YF]xFP motif to associate with the kinase. Phylogenetic analysis, however, indicates that arginine and lysine are conserved in the second position of the motif ([YF]KFP or [YF]RFP), and these basic amino acids are critical for Cbk1 binding in vitro (Fig 4D). Basic amino acids are not present in the core

motif of the Ace2 peptide, which has the sequence YQFP. Thus, the extended binding configuration of Ace2's peptide likely changes the constraints on the sequence of the core motif, relieving the need for basic amino acids in the second position (as seen in the Ssd1 peptides). Intriguingly, the Ssd1's core FKFP motif does not bind Cbk1's kinase domain when present in Ace2 peptides with mutated N-terminal regions, indicating that the peptide context in which the core motif is displayed influences its ability to bind the kinase (S5F Fig).

Docking motifs in substrate proteins can dramatically increase their rate of phosphorylation by kinases that bind to them [43,45]. We therefore determined if Cbk1-binding peptides from Ace2 and Ssd1 enhance phosphorylation of these substrates, comparing phosphorylation of Ace2 and Ssd1 protein fragments containing either native Cbk1-binding peptides or mutated sequences. The presence of WT Cbk1-binding segments in Ace2 and Ssd1 fragments significantly increases their in vitro phosphorylation by both the Cbk1–Mob2 complex and the Cbk1 kinase domain alone (Fig 4E).

Our attempts to use crystallography and crosslinking to understand how Cbk1 interacts with the Ace2 and Ssd1 peptides were not successful. We therefore used unbiased blind docking of the Ssd1 DFKFP peptide to the Cbk1 kinase domain using AutoDock [46]. This computational approach strongly supports docking motif association with a pocket formed by a PxxP region in Cbk1's C-terminal extension (where Y687 follows the first proline) and a surface loop connecting the D and E helices in the C-terminal kinase lobe (W444, F447) (see S1 Fig). This computationally supported peptide docking site is consistent with the crucial importance of aromatic amino acids in the peptide and predicts that the aromatic amino acids F447, W444, and Y687 in Cbk1 are involved in the interaction (Fig 5A). We tested these predictions by substituting alanine at these sites individually and assessing in vitro binding of docking peptides to the mutant proteins. As predicted, F447A, W444A, and Y687A substitutions abolish or greatly reduce binding of Cbk1–Mob2 to Ace2 and Ssd1 docking peptides. However, HM-P binding in trans stays intact, indicating correct folding and structural integrity of all tested mutants (Fig 5B and 5C).

Cbk1's PxxP motif is in the C-lobe tether (CLT) region that is conserved in AGC kinases [28]. In other AGC kinases, mutation of the PxxP motif can disorder the CLT, which compromises catalytic activity. We therefore determined if mutation of Cbk1's PxxP motif has such an effect by measuring the kinase activity of Cbk1(Y687A). This mutation eliminates enhancement of substrate phosphorylation conferred by the presence of an FKFP docking motif, as expected from direct binding experiments, but, importantly, does not compromise basal kinase activity on substrates lacking a functional docking motif (Fig 5D).

Substrate Docking Confers Robustness to Functional Kinase–Substrate Interactions In Vivo

To determine the in vivo function of Cbk1's docking interactions with its known substrates, we evaluated phenotypic effects of mutations that eliminate the peptides' in vitro interaction with Cbk1's kinase domain. We introduced mutations into Ace2's Cbk1 docking motif that abrogate the peptide's in vitro interaction with the kinase (an allele termed $ace2^{dock^*}$). In cells expressing $ace2^{dock^*}$ there was a shift towards larger groups of cells—60% of groups contained 1–2 cells, and 20% of groups had five or more cells—though not nearly to the extent of *ace2*Δ strains (<5% of groups contained 1–2 cells and >95% of groups contained five or more cells per group) (Fig 6A).

These results show that $ace2^{dock^*}$ is a partial loss-of-function allele under ideal growth conditions. We therefore hypothesized that the docking interaction enhances robustness of this kinase–substrate interaction, buffering the system against variability in RAM network activity

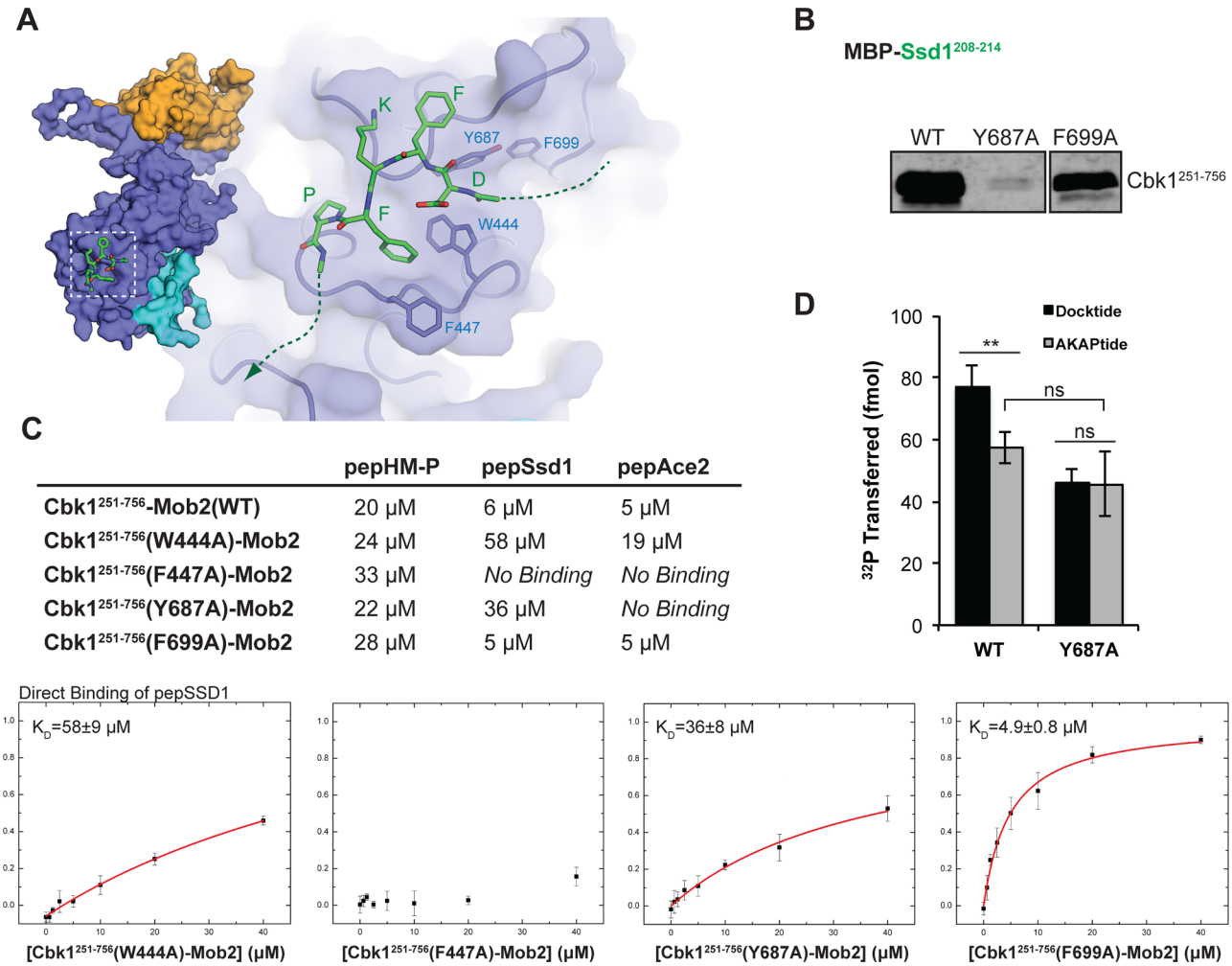

Fig 5. Structural model of the Cbk1–docking peptide complex. (A) AutoDock binding simulation of Ssd1 docking fragment (DFKFP) identifies an interaction surface on Cbk1 and suggests important aromatic interactions with residues W444, F447, and Y687. (B) Y687A but not nearby F699A abrogate Cbk1 pulldown by Ssd1$^{208–214}$ as predicted by the model. (C) FP analysis of putative docking motif binding site residues. The binding of all mutants to pepHM-P was similar to that of WT Cbk1, indicating that the mutants are folded properly (see first column in the table summarizing peptide binding affinities). W444A, F447A, and Y687A were reduced in their binding affinity for Ssd1 and Ace2 docking peptides, while F699A was not. (D) Peptide kinase assay of WT and Y687A mutant of Cbk1$^{251–756}$–Mob2. Peptides are based on Ssd1 (210–229), which contains the N-terminal docking motif (Docktide: FKFP, AKAPtide: AKAP) and a Cbk1 consensus site. The presence of the WT docking motif enhances phosphorylation by WT Cbk1 relative to a peptide containing a mutant docking motif (similar to Fig 4F). Cbk1(Y687A) loses this enhancement when unable to bind WT docking motif. Data is plotted as mean ± standard error of the mean. Student's unpaired *t*-test: $^{**}p < 0.01$; ns, not significant. Data for (C) can be found S2 Data, and data for (D) can be found in S3 Data.

doi:10.1371/journal.pbio.1002146.g005

and maintaining constant signaling output. This predicts that docking interaction should be far more important when Cbk1 activity is compromised. We tested this by measuring cell separation in strains expressing the hypomorphic Cbk1-as2 (M429A) mutant protein. This mutant protein, which allows inhibition of Cbk1 by modified ATP analogs, has significantly reduced intrinsic kinase activity in the absence of drug (we performed experiments without drug addition) [21,46]. Notably, the catalytically weakened *cbk1-as2* allele has absolutely no phenotype in cells with WT *ACE2*. However, combining the *cbk1-as2* and *ace2^{dock*}* alleles dramatically disrupts cell separation: only ~25% of groups contained 1–2 cells, while ~54% contained five or more cells (Fig 6B). We also measured transcription of three Ace2 target genes. Expression of the Ace2-driven genes *CTS1*, *DSE1*, and *SCW11* was reduced in *ace2^{dock*}* cells, and was nearly absent in *ace2^{dock*} cbk1-as2* cells (Figs 6C and 6D and S6).

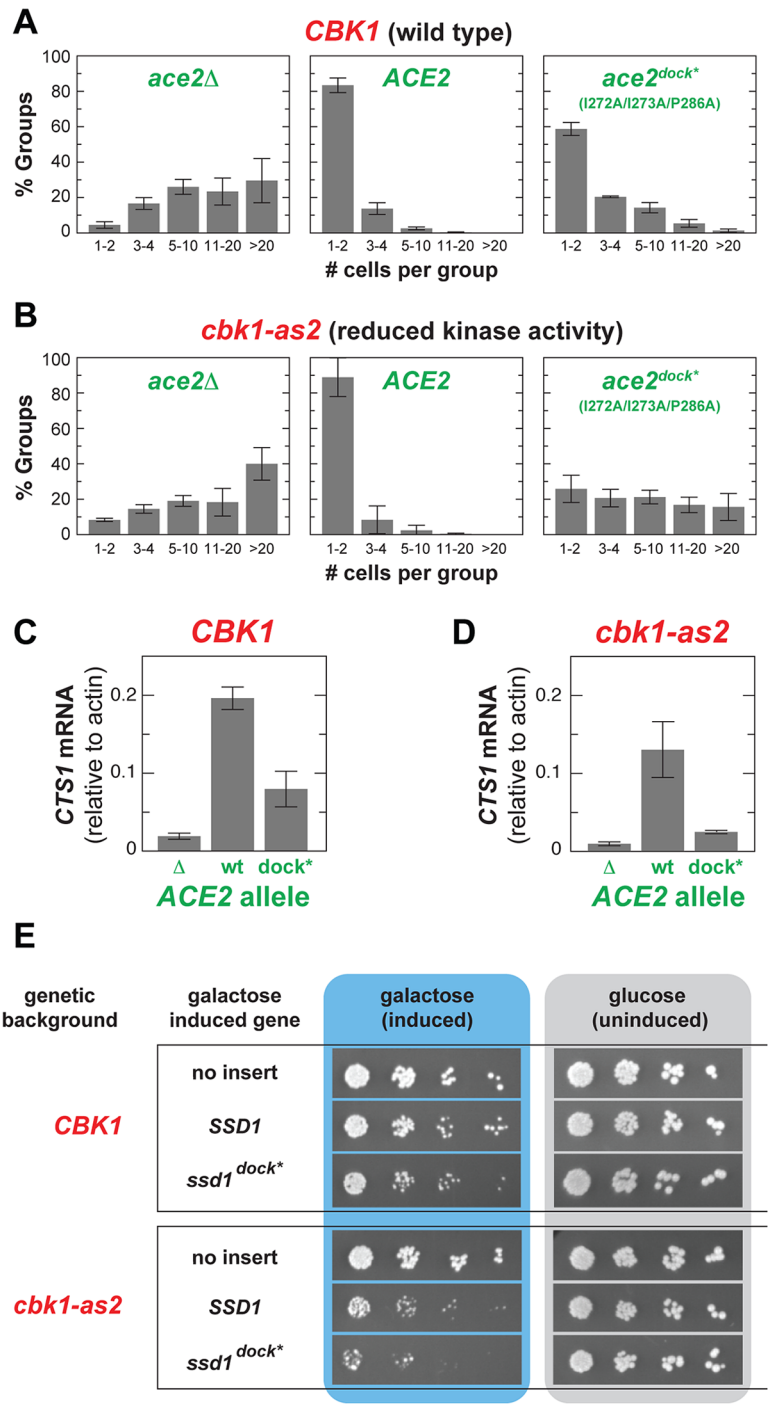

Fig 6. Docking sites increase robustness of Cbk1 control of in vivo substrates. (A) Mutation of the Cbk1 docking motif in Ace2 (*ace2*^{dock*}) confers a modest defect in cell separation in cells with WT *CBK1*. (B) The *ace2*^{dock*} allele has a marked cell separation defect in cells carrying the *cbk1-as2* allele, which is catalytically weakened. Note that *cbk1-as2* cells exhibit no cell separation defect when the WT *ACE2* allele is present. (C and D) *CBK1* WT cells carrying *ace2*^{dock*} have a slight reduction in *CTS1* transcript levels (C), while *cbk1-as2* cells carrying *ace2*^{dock*} exhibit strongly reduced *CTS1* transcription (D). (E) Overexpression of the *ssd1*^{dock*} allele, carrying mutations that eliminate docking interaction with Cbk1, affects viability of WT cells, and this is far more dramatic in *cbk1-as2* cells. Cells were 10-fold serially diluted from left to right and plated on galactose (inducing) or glucose (repressing) media: reduced colony formation in serial dilutions reflects impaired viability. Data for (A) and (B) can be found in S4 Data, and data for (C) and (D) can be found in S5 Data.

doi:10.1371/journal.pbio.1002146.g006

Cbk1's phosphorylation of the mRNA binding protein Ssd1 allows translation of proteins required for cell wall expansion in the growing bud (Fig 1), and loss of RAM network activity is lethal in cells with functional Ssd1. Moreover, overexpression of WT *SSD1* is mildly toxic to WT cells but dramatically deleterious in *cbk1-as2* cells, and overexpression of an Ssd1 variant lacking Cbk1 phosphorylation consensus sites is lethal [23,47]. Thus, the strength of Cbk1's regulation of Ssd1 can be assessed using viability effects. We compared viability of strains overexpressing either WT *SSD1* or an allele in which Cbk1 docking motifs are mutated (termed $ssd1^{dock^*}$). Overexpression of $ssd1^{dock^*}$ subtly compromises the viability of WT cells, and this effect is markedly worse in *cbk1-as2* cells (Fig 6E).

Our findings show that docking motifs in Ace2 and Ssd1 stabilize their regulation by Cbk1. This suggests that Cbk1–substrate docking with target proteins in vivo confers signaling system robustness, maintaining pathway output when input signal strength fluctuates or is compromised.

The Cbk1 Docking Motif Predicts an Expanded System of RAM Network Regulatory Targets

The docking motifs in Ace2 and Ssd1 are conserved in these proteins over hundreds of millions of years of fungal evolution [44], suggesting that the signaling robustness they provide confers a selective advantage. It therefore seems reasonable that this kind of kinase–substrate docking in other Cbk1 regulatory targets might also increase fitness. We therefore evaluated conservation of both the core [YF]xFP Cbk1 docking motif and the distinctive Cbk1 phosphorylation consensus sequence (HxRxx[ST]) in all budding yeast proteins. We used the algorithm ConDens [48], which assigns a conservation score to each instance of a specified short motif based on its preservation relative to surrounding amino acids and its degree of enrichment within a protein region above random expectation. This approach quantifies motif conservation independent of precise sequence alignment. Notably, however, motifs that occur in regions of high background conservation are given a low score.

We assigned two ConDens scores to each protein encoded by the *S. cerevisiae* proteome: one for Cbk1 core docking peptides and another for Cbk1 phosphorylation consensus sequences. Comparing these scores identifies six proteins with clear conservation of both short linear motifs (Fig 7). There is experimental evidence for in vivo phosphorylation of Cbk1 consensus sites in all six of these proteins, and for physical interaction with Cbk1 for most of them [22,23,49,50]. This analysis strongly suggests that, in addition to Ace2 and Ssd1, the proteins Boi1, Fir1, Irc8, and Bop3 are in vivo targets of Cbk1 that engage in docking interactions with the kinase. Tao3, a RAM network component that physically interacts with Cbk1, contains a core docking motif with a high ConDens score, as well as Cbk1 phosphorylation consensus sequences whose ConDens scores are relatively low because they occur in a region of high background conservation. Intriguingly, the proteins Dsf2, Mpt5, and Sec3 lack a Cbk1 docking motif but have Cbk1 phosphorylation consensus sequences that are clearly conserved; there is evidence for in vivo phosphorylation of these sites in Mpt5 and Sec3 [49,50].

We found that proteins with a match to the [YF]xFP Cbk1 core docking motif are significantly more likely to contain matches to the Cbk1 phosphorylation consensus motif (S9 Fig) (Fisher's exact test, $p = 0.016$). In proteins in which the core docking motif is conserved, the probability is dramatically higher that Cbk1 phosphorylation consensus sequences are present, and these consensus sites are themselves far more likely to be conserved (Fisher's exact test, $p < 10^{-6}$ and $p < 10^{-10}$, respectively). Furthermore, consensus site matches in proteins with conserved Cbk1 core docking motifs are significantly more likely to be phosphorylated in vivo (PhosphoGrid annotation [50]; $p = 0.01$) (S7 Fig). Thus, Cbk1's core docking motif appears to

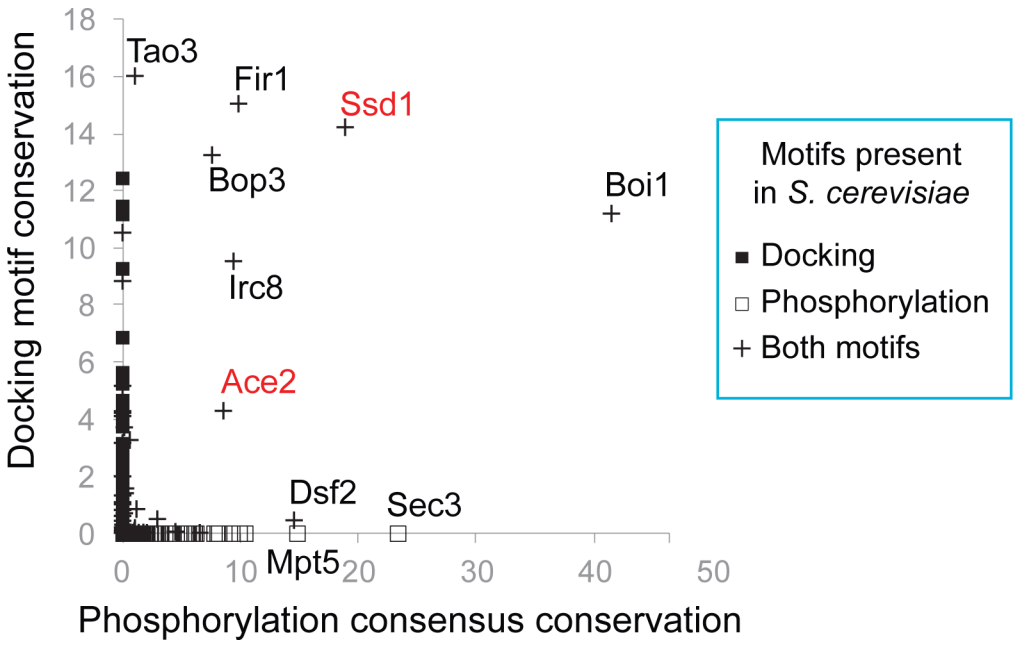

Fig 7. Co-conservation of docking and phosphorylation sites identifies known and likely Cbk1 regulatory targets. Conservation of Cbk1 docking motifs and phosphorylation consensus sites in individual proteins, calculated using ConDens [48]. For each motif type, the most significant score among the matches in each protein is assigned as the score for that protein. The *x*-axis plots conservation scores of [YF]xFP sequences (Cbk1 docking), and the *y*-axis plots conservation scores of Hx[RK]xx[ST] sequences (Cbk1 phosphorylation consensus). The known Cbk1 substrates Ssd1 and Ace2 are highlighted in red. Proteins without docking motifs that score very highly for phosphorylation consensus conservation are noted. ConDens scores can found in S6 Data.

doi:10.1371/journal.pbio.1002146.g007

be a useful predictor of functionally important Cbk1 phosphorylation consensus motifs and previously unappreciated regulatory targets of the budding yeast RAM network.

Discussion

Cbk1–Mob2, a Structural Template for an NDR/LATS Kinase–Coactivator Complex, Provides Insight into Activation Mechanisms

Structural and biochemical analysis of the yeast Cbk1–Mob2 complex presents a picture of a dynamic and allosterically regulated molecular switch, in which a distinctive activation loop and N-terminal kinase domain extension superimpose regulatory mechanisms that are unique to NDR/LATS kinases on a highly conserved AGC kinase catalytic core. For example, regions of Cbk1's activation loop conserved among the NDR/LATS family may directly block substrate binding when the kinase is inactive, while autophosphorylation of the flexible region promotes a restructuring that brings the kinase into an "open" state that can bind substrates. Additionally, the NDR/LATS N-terminal extension adopts a novel conformation upon association with Mob coactivator proteins, forming a dramatic cleft where the two proteins come together.

Our analysis indicates that Mob coactivator association creates a mechanism that discriminates between the phosphorylated and unphosphorylated states of the NDR/LATS HM. In all of our crystal structures, Cbk1's HM region is bound within the cleft formed by Mob association with Cbk1's N-terminal extension. This organization differs from that of PKB/Akt, in which HM site phosphorylation promotes its association with the PIF pocket on the kinase N-lobe, shifting the αC helix to a conformation that organizes the catalytic site in an optimally active state [39,51]. Our MD simulations suggest that phosphorylation of the NDR/LATS HM site leads to compression of the N-linker region and allows bidirectional interactions between

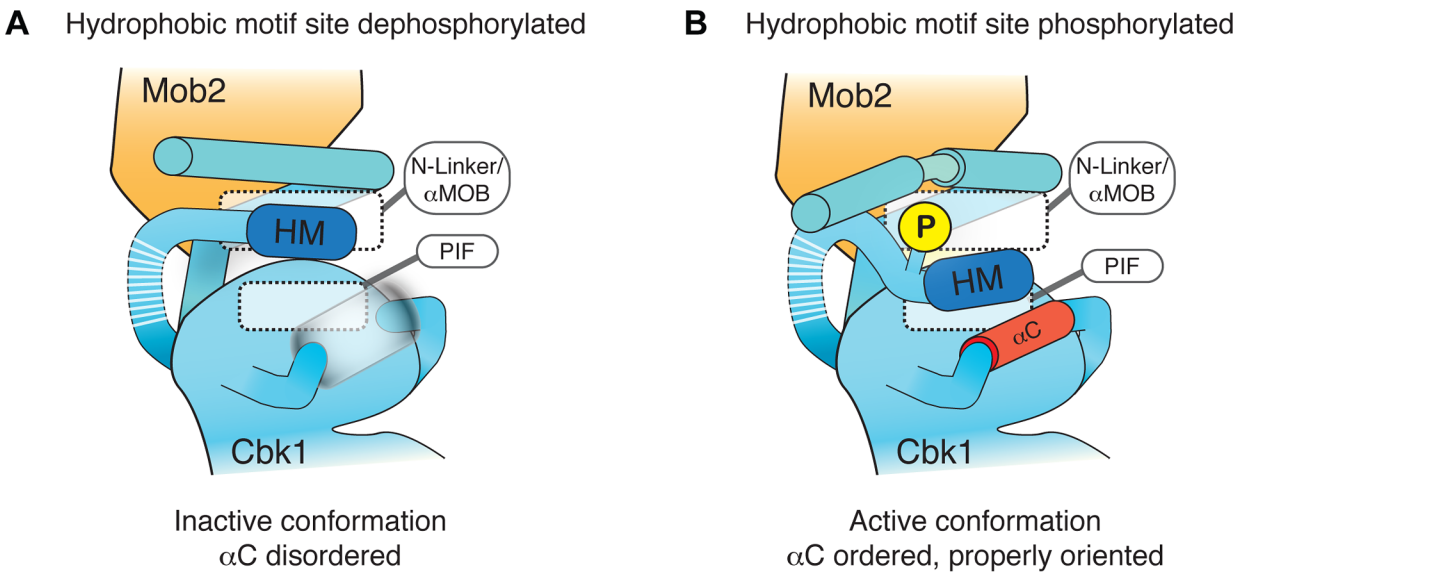

Fig 8. Model for role of Mob binding and HM phosphorylation in NDR/LATS kinase activation. Mob supports the NDR/LATS kinase N-terminal extension in a conformation that creates a binding slot for both HM-T and HM-P. (A) HM-T remains associated with the N-linker region in a manner that does not promote an ordered conformation of the αC helix. (B) HM-P promotes reorganization of αMob, compressing the HM binding slot and holding aromatic side chains of the HM in a configuration that favors ordering of αC.

doi:10.1371/journal.pbio.1002146.g008

the αMob helix, the HM, and a site analogous to the PIF pocket. This interaction depends on the presence of the Mob cofactor.

Based on these data, we propose that the unphosphorylated NDR/LATS HM (HM-T) binds the kinase N-linker region, but cannot shift to a conformation that orders αC. The phosphorylated form of the HM site (HM-P) can also bind in this manner, but causes significant reorganization of the cleft formed by Mob2 association, leading to active kinase (Fig 8). Intriguingly, replacing human NDR's HM region with the corresponding segment of PKB/Akt yields a hyperactive kinase that does not appear to require its Mob coactivator [52]. In the context of our findings, this suggests that the NDR/LATS kinases have evolved an activating motif that requires Mob interaction. Ordering of αC in the activated state of Cbk1 is a key element of this activation model. Although our inactive Cbk1–Mob2 crystallographic complexes do not suggest how this mechanistically may happen, it is likely that the HM motif will dock onto the αC helix, based on other activated AGC kinase structures. Cbk1 activation may be reminiscent of the structurally characterized activation model of RSK, where activating phosphorylation grossly remodels the αC of its AGC kinase domain [53].

Hippo pathways, characterized in animals and yeast, require large scaffold proteins to spatially coordinate network components and efficiently activate the NDR/LATS kinase [1,6]. The yeast MEN pathway, for example, requires the scaffold Nud1 for Cdc15 to activate Dbf2, whose recruitment is facilitated by a Mob1 phospho-recognition region that Mob2 lacks [12,14]. The scaffold for the RAM network is currently unknown, although the large protein Tao3 that is related to metazoan furry family proteins has genetic interactions with all RAM components and may fulfill that role. The need for scaffolding is bolstered by the fact that the HM site is a poor substrate for the upstream hippo kinase [25]. Our structures show that there is likely an autoinhibitory HM binding site in the N-linker region, which would further hinder HM phosphorylation. We suspect scaffolding may not only function to bring hippo components together, but is likely also necessary to allosterically regulate HM N-linker region dynamics to facilitate efficient hippo phosphorylation and thus NDR/LATS activation.

Overall, the structures of Cbk1–Mob2 complexes suggest that Mob coactivator binding by the NDR/LATS kinases enhances the general HM allosteric mechanism that is a hallmark of AGC kinase regulation, making it more switch-like. This provides an elegant example of a common regulatory mechanism altered and adapted by the presence of extra regulatory factors. This fits well with the physiological roles of NDR/LATS kinases as central regulators of cell-cycle-dependent processes. Therefore, physiologically and mechanistically, the NDR/LATS–Mob coactivator system is reminiscent of the far better characterized cyclin-dependent kinase (CDK)–cyclin system [45,54].

Cbk1's Docking Mechanism Suggests a Surface at Which New Substrate Interaction Specificity Evolves

We have shown that a site on the C-terminal lobe of Cbk1's catalytic domain interacts with a short peptide docking motif in the kinase's known in vivo substrates, Ace2 and Ssd1. AGC kinase catalytic domains have not thus far been demonstrated to bind short linear docking motifs in substrate proteins in this way. In one case, the AGC kinase PDK1 uses a form of substrate docking to associate with PKB/Akt [29]. This system, however, is an adaptation of an intramolecular regulation mechanism for recruitment of a regulatory target: PDK1 lacks its own cis-regulatory C-terminal HM region, making available a site where the corresponding HM region of PKB/Akt associates. Our findings describe a form of substrate docking that has not been seen previously in an AGC group kinase, and thus raise the possibility that similar mechanisms may be more widely distributed.

Cbk1's docking peptide binding site is built from conserved structural elements and, in a broad sense, is located similarly to the well-characterized D motif binding region of MAPKs such as ERK and JNK [55,56]. As in the MAPKs, part of Cbk1's peptide-binding site is formed by a short loop that links the D and E alpha helices of the core catalytic domain. Another crucial part of Cbk1's docking peptide binding site is contributed by the kinase's conserved AGC kinase C-terminal extension, in the CLT [28]. Here, aromatic amino acids near a pair of prolines in a conserved PxxP motif play an important role in Cbk1's docking motif binding surface.

The similarities between the Cbk1 peptide docking site and the D motif binding site of MAPKs, which are members of the relatively distantly related CMGC group, are intriguing. However, D motif peptides that associate with MAPKs are unrelated to the core docking motif that binds Cbk1, and the topography of the peptide–kinase interaction is different (S8A Fig). Furthermore, the amino acids in the D–E loop are divergent between Cbk1 and the MAPKs. Thus, the relationship of the docking motif interaction sites of Cbk1 and the MAPKs indicates that this kinase surface is probably functionally well suited for peptide interaction, with distinct sequence specificity arising in different lineages.

The involvement of Cbk1's CLT in docking motif binding suggests that other AGC kinases might also use this region to associate with peptide motifs in target proteins. The overall organization of Cbk1's docking surface is fairly similar to those of PKB/Akt and ROCK [36,51] (S8B Fig). In fungal NDR/LATS kinases, the combination of bulky aromatic side chains important for Cbk1's docking peptide binding is clearly present. However, while the kinase regions that underlie the Cbk1 docking surface are present in metazoan orthologs, there is considerable evolutionary plasticity within them (S9 Fig). For example, the PxxP motif that occurs in the CLT of LATS and most AGC kinases has been lost in metazoan NDR orthologs. The D–E loop is also evolutionarily divergent. Thus, while the protein surface Cbk1 uses for docking peptide binding is commonly present in a wide range of protein kinases, its ability to specifically bind the [YF]xFP motif probably arose in the evolution of this NDR family kinase in fungi. The

variability of amino acids clustered at this kinase surface suggests that it is a place where new peptide interaction specificities can evolve. We therefore propose that other members of the NDR/LATS family and the broader group of AGC kinases adopt analogous docking solutions, albeit with different specificity.

New Cbk1 Targets Implicate Additional Processes in Cell Separation and Morphogenesis

Shared conservation of Cbk1's docking and consensus motifs can be used to predict proteins that are controlled by this NDR/LATS kinase. This analysis identifies seven proteins that form a putative system of docking substrates (Fig 9A, green lines). As noted, there are also proteins with highly conserved Cbk1 consensus sites but no docking motif, and we propose that these are also regulatory targets of the kinase as a group of non-docking substrates (Fig 9A, blue lines). For nearly all of these proteins, there is evidence that Cbk1 consensus sites are phosphorylated in vivo.

This expanded system of proposed RAM-network-regulated proteins forms a remarkable nexus around the processes of cell morphogenesis, cytokinesis, and cell separation. Several are involved in bud growth and the remodeling of the cell cortex, consistent with their localization. Boi1 acts with proteins involved in cell polarity to help maintain cell polarity, and Ssd1 modulates translation of proteins required for both cell separation and normal expansion of the cell wall [22,57,58]. Additionally, Boi1 has been proposed to function in a regulatory system that prevents completion of cytokinesis when chromosome segments remain in the zone of cell division [59]. Sec3, a predicted Cbk1 substrate that lacks a docking motif, localizes to the bud cortex to recruit proteins that promote polarized secretion and thus localized cell growth. Ace2, which is activated by Cbk1, has a well-established role in the separation of mother and daughter cells [18,19]. While Irc8's function is unknown, it is a putative membrane protein that localizes to the cortical region of growing buds [60].

The Cbk1 substrate system we propose is also connected to processes of gene expression, mRNA stability, and translational control. In addition to Ssd1's suppression of translation, Mpt5 promotes mRNA decapping and deadenylation, and genetic interaction suggests that Cbk1 inhibits Mpt5 [61,62]. Fir1's function is unknown, but it may play a role in mRNA polyadenylation; Bop3 is similarly uncharacterized, although its nuclear localization may indicate a role in transcriptional regulation. Intriguingly, there is evidence for functional integration of some of these proteins. Ace2 drives expression of cell separation genes that are translationally controlled by Ssd1, and this may form a Cbk1-controlled feed-forward loop that is modulated by kinase docking interactions [59].

A Sequential Model for Adaptive Acquisition of Kinase Docking

The amino acids that are important for Cbk1's docking motif binding are generally conserved among fungal NDR kinases, suggesting that this interaction mechanism emerged early in the evolution of this branch of life and conferred fitness. Consistent with this, Ssd1 orthologs from basidiomycetes to ascomycetes have the core docking motif and Cbk1 phosphorylation consensus motifs in a conserved organization, although in some lineages the docking motif is absent (Fig 9B).

While Cbk1's docking interaction appears ancient in fungi, the kinase's probable substrates do not invariably have docking motifs (Figs 7 and 9A). It is therefore plausible that proteins first become kinase substrates and then subsequently evolve docking. In this view, kinase docking motifs arise by random evolution of unconstrained regions in proteins that are already Cbk1 substrates, and are retained if they enhance the regulatory interaction. This is consistent

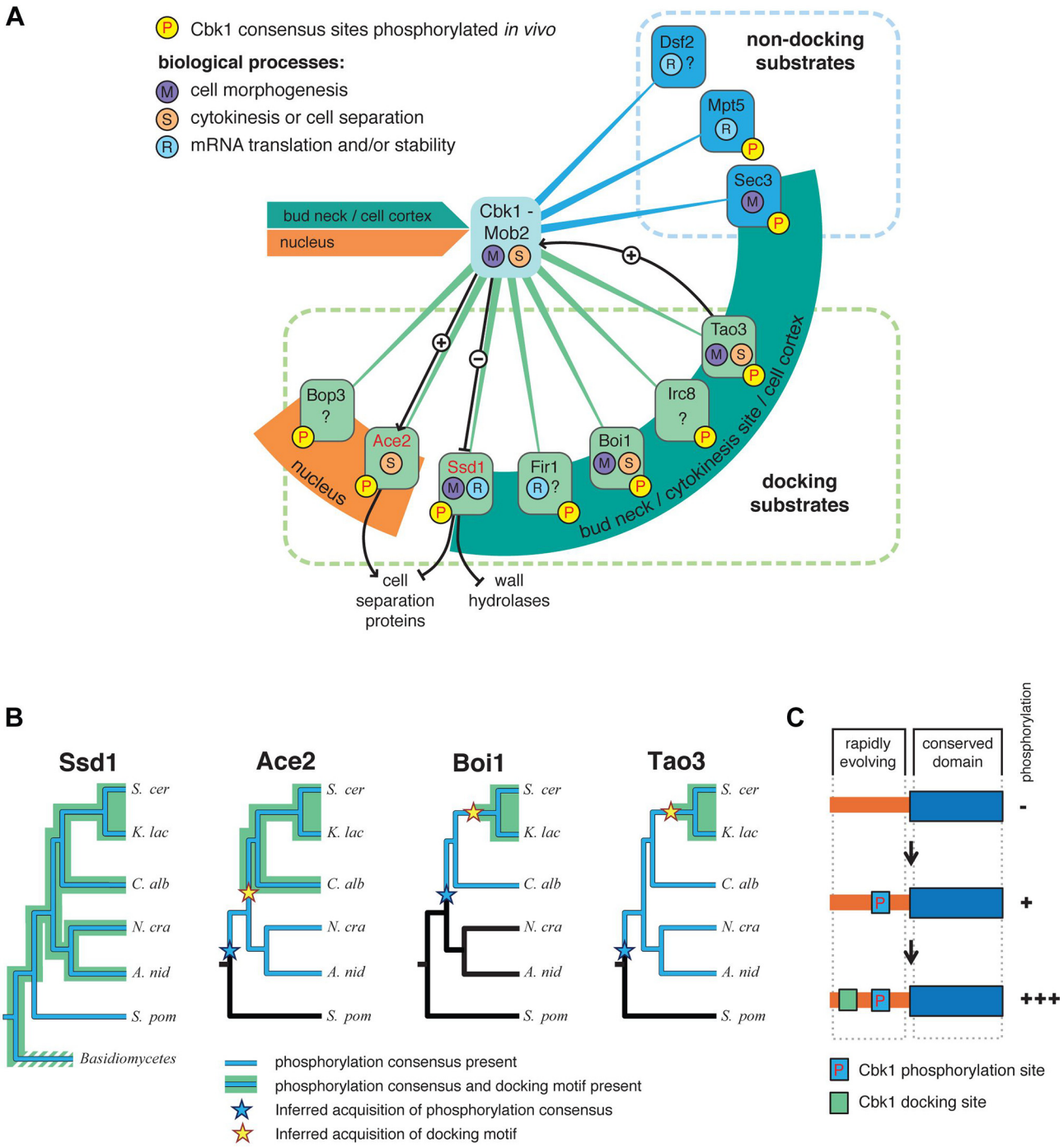

Fig 9. A probable network of proteins controlled by the yeast RAM network, and sequential evolution of docking in Cbk1 substrates. (A) An outline of known and likely Cbk1 substrates in which the core docking motif is conserved (green lines) and likely substrates that lack a docking motif (blue lines). Ace2 and Ssd1 (red type) are known Cbk1 phosphorylation targets, and black lines indicate regulatory interactions inferred from molecular genetic analysis. Orange and green arcs indicate localization of target proteins to either the nucleus or sites of cell growth and cortical remodeling (bud neck/cytokinesis site/ cell cortex): Cbk1 displays both of these localization patterns. (B) Distribution of phosphorylation consensus sites and docking motifs in known and likely Cbk1 substrates. Orthologs of Ace2, Boi1, and Tao3 are present in the lineages drawn. Blue indicates lineages in which the proteins have conserved phosphorylation consensus sites, while in those drawn in black, consensus sites are absent. Inferred acquisition of consensus sites is represented by a blue star. Lineages in which the proteins have conserved Cbk1 docking motifs are highlighted green, with inferred acquisition of this motif labeled by a yellow star. (C) Sequential addition model for acquisition of regulatory motifs during signaling system evolution. Substrates first become phosphorylation targets of a given upstream kinase. Docking motifs are subsequently acquired in unstructured, rapidly evolving sequence. The robustness of substrate phosphorylation increases upon docking motif acquisition. *A. nid*, *Aspergillus nidulans*; *C. alb*, *Candida albicans*; *K. lac*, *Kluyveromyces lactis*; *N. cra*, *Neurospora crassa*; *S. cer*, *Saccharomyces cerevisiae*; *S. pom*, *Schizosaccharomyces pombe*.

doi:10.1371/journal.pbio.1002146.g009

with the observation that short functional motifs evolve rapidly in intrinsically disordered segments of proteins [63]. To test this model we inferred the ancestral configuration of Cbk1 phosphorylation consensus sequences and docking motifs in known and likely Cbk1 substrates, where we could confidently identify orthologs across a wide span of fungal evolution.

For at least three known or likely Cbk1 substrates, consensus phosphorylation sites were clearly present before the core docking motif appeared in unstructured protein regions (Fig 9B). In these cases, there are orthologs in species that diverged from the lineage leading to *S. cerevisiae* that have phosphorylation sites, but no docking motifs. In other likely Cbk1 substrates, phosphorylation sites and docking motifs appear simultaneously. We never found cases in which the appearance of docking motifs predates the acquisition of phosphorylation sites. We surmise that a surface capable of binding to a [YF]xFP docking peptide emerged in fungal NDR and provided selection for subsequent acquisition of cognate docking motifs in phosphorylation targets, exemplifying co-evolution of kinases and substrates. This kind of adaptive evolution of kinase docking (Fig 9C) is consistent with an increase in signaling complexity that has been proposed as a molecular mechanism underlying the evolution of animal development [64,65].

Materials and Methods

Protein Purification

His6-Cbk1$^{251-756}$ was co-expressed with GST-Mob2 (full length or 45–287). Cbk1 kinase domain alone (352–756 or 352–692) was expressed as GST-Cbk1-His6. All constructs were bacterially expressed and purified on a Ni-NTA resin (Qiagen) followed by glutathione Sepharose beads (GE Biosciences). For crystallization, tags were cleaved with TEV protease followed by cation exchange chromatography on a Resource S column. All Cbk1 preparations were subsequently dialyzed into Cbk1 buffer (20 mM Tris-HCl, 150 mM NaCl, 2 mM DTT [pH 8.0]).

Cbk1 substrates containing the docking motifs or substrate sites were bacterially expressed as MBP or GST fusions. For pulldown assays, MBP-fusion lysates were incubated with amylose resin (New England Biolabs) on ice for 15 min, followed by washing with PBS (137 mM NaCl, 2.7 mM KCl, 4.3 mM Na_2HPO_4, 1.4 mM KH_2PO_4 [pH 7.3]). GST fusions were purified on glutathione Sepharose beads followed by dialysis into Cbk1 buffer.

Crystallization and Data Collection

The crystal structure of the Cbk1–Mob2 complex was solved in three different crystal forms using Cbk1 constructs with an intact or with a mutated (T743E) HM motif (S1 Table).

To determine the structure of the Cbk1–Mob2 complexes, 4–6 mg/ml protein samples were supplemented with 2 mM AMPPNP, 2 mM $MgCl_2$, and 2 mM TCEP. All samples used for crystallization contained an inactivating mutation (D475A) in Cbk1, while some Cbk1 constructs also contained a phosphorylation-mimicking mutation in the HM (T743E). The rational for the D475A mutation was to eliminate the autophosphorylation activity of Cbk1, while the T743E mutation was introduced to capture the kinase in an activated state. Crystallization conditions were found by using a custom in-house PEG crystallization screen in microbatch under oil (1:1 silicon oil:paraffin oil) set-ups at 23°C. All crystals were treated with 10% glycerol before flash freezing in liquid nitrogen. Data on frozen crystals were collected at 100 K on the PXI and PXIII beam lines of the Swiss Light Source (Villigen, Switzerland) using a beam of 1.00 Å wavelength (S1 Table).

Single tetragonal crystals of Cbk1(D475A/T743E) complexed with Mob2 were grown in 25% PEG 20000 buffered with 0.1 M sodium citrate (pH 5.5); however, these crystals diffracted only to 4.5 Å resolution. Hoping to improve the diffraction limit of this low diffracting complex

by addition of a docking peptide, crystallization trials were set up in the presence of the Ssd1 docking peptide (TTEQSDFKFP) in approximately 4-fold excess. Plate-shaped crystals could be grown under such conditions with both the Cbk1(D475A/T743E)–Mob2 and the Cbk1 (D475A)–Mob2 complexes. These crystals diffracted to 3.6 and 3.3 Å resolution (S1 Table). Unfortunately, we could not locate the peptide in the electron density. It is likely that peptides could not bind in the crystal to Cbk1, as the docking surface was occluded in both crystal forms: Phe447 made prominent hydrophobic crystal contacts to symmetry-related molecules. All crystals contained AMPPNP.

Data Processing and Refinement

Data was processed with XDS [66], and the structure was solved by molecular replacement with PHASER [67]. The 3.6 Å resolution Cbk1(D475A/T743E) (crystal form B) dataset processed in C222 was used to search with the ROCKI kinase domain, which is highly similar to that of Cbk1 [36] (PDB ID: 2ETR). This gave a clear solution for the Cbk1 kinase domain (TFZ = 14.5, LLG = 184), and the Mob2 subunit was located using the highly similar Mob1 protein [33] (PDB ID: 2HJN) as the search model (TFZ = 6.9, LLG = 175). Later refinement, indicated that the correct space group for this crystal was C2; therefore, its asymmetric unit contains two Cbk1–Mob2 complexes. The MR search with the assembled Cbk1–Mob2 complex model then identified a single complex per asymmetric unit in the 4.5 Å resolution Cbk1 (D475A/T743E)–Mob2 crystal (crystal form A). Similarly, only one complex was located in the crystal of the Cbk1(D475A)–Mob2. In contrast, two complexes were found per asymmetric unit in the higher resolution crystals of Cbk1(D475A/T743E)–Mob2. Although all four structural models for the Cbk1–Mob2 complex are subject to different crystal packing interactions, they all display the same Cbk1 and Mob2 domain orientation (S2A Fig). Structure refinement was carried out in PHENIX [68] using secondary structure restraints and translation-libration-screw (TLS) parameterization. Modeling was done manually in Coot [69]. All models have less than 1.5% of residues falling into disallowed regions on the Ramachandran plot. Parts of the electron density maps around the HM motif and the αINH regions are shown in S2B, S2C, S2E, and S2F Fig.

For the Cbk1(D475A)–Mob2 model, which was refined to 3.3 Å resolution, 218 of the 745 residues, corresponding to close to one-third of the used protein constructs, are missing in the final model. For example, a long N-terminal Cbk1 region (251–293) and the N-terminal Mob2 region (45–110) could not be located in the electron density map. Apart from these apparently flexible N-terminal protein construct ends, there was no density for the region corresponding to the αC (392–406), the middle part of the Cbk1 activation loop (508–553), the linker region connecting the Cbk1 kinase core with the hydrophobic C-terminal region (714–739), and an internal Mob2 region (149–160) (see Figs 2A and S1). Unfortunately, all attempts to express shortened constructs failed because the yield of soluble protein was insufficient for structural studies.

Molecular Dynamics Simulations

First, we constructed a hybrid model that incorporated regions from Cbk1–Mob2 and PKB/Akt crystal structures. Cbk1's disordered region containing αC was completed with the ordered region from the PKB/Akt's crystal structure (PDB ID: 1O6K) using homology modeling followed by energy minimization to remove minor clashes. Furthermore, Cbk1's short activation loop that contains the S570 autophosphorylation site was remodeled to match the position of the phosphorylated activation loop of PKB. The long activation loop (499–570) and the kinase core–HM connector region (714–739) were built in manually in an extended conformation.

We subjected this complete model to MD simulations. Initial analyses suggested that the artificially built-in flexible loop regions (499–570 and 714–739) showed too much structural fluctuation, which precluded free MD simulations. Next, we excluded the activation loop (499–570) and performed a 100-ns MD run on the model where only the HM connector loop region (714–739) was allowed to move freely to pre-equilibrate this loop region. All MD simulations started out from this pre-equilibrated model.

Heavy atom position restraints were applied on the C-terminal kinase core region (569–670); otherwise, the structure was allowed to move freely during 125-ns-long MD runs. Protein structures were blocked with N-methyl and acetyl groups at the artificial chain break between Y569 and P670, and the nucleotide was modeled as ATP. In all calculations the GROMACS ver. 4.5.5 program package was used [70], and the Amber-03 force field [71] was applied, along with neutralizing Na+ counter-ions and numerous TIP3P [72] explicit water molecules filling a 5 Å spacing between the protein parts and the edges of the cubic simulation box. Long-range electrostatics was calculated by the PME method and a van der Waals cutoff of 9 Å was used. Prior to the productive MD runs, the systems were energy-minimized using steepest descent and conjugate gradient molecular mechanics energy minimizations with a step size of 0.1 Å, and a tolerance of 10 kJ mol^{-1} nm^{-1}. The energy-minimized systems were subjected to MD calculations. A time step of 2 fs, constant temperature of 300 K, and LINCS bond constraints [73] were applied. Neighbor lists were used and updated every 10 fs. Temperature coupling was carried out using the v-rescale scheme [74]. The resulting trajectory files were analyzed with programs of the GROMACS package.

In Vivo Analysis of RAM Network Output

For cell separation assays, *ckb1*Δ, *ace2*Δ*CBK1*, *or ace2*Δ*cbk1-as2* cells were transformed with empty, WT, or mutant vector under control of the ADH1 promoter. Cells were grown to log-phase, briefly sonicated, and mounted on slides to count cell group size. Histograms represent mean ± standard deviation of three independent cultures, with >100 groups counted for each. For quantification of transcriptional output, RNA was isolated from log-phase cultures and analyzed by real-time PCR. Bar graphs represent mean ± standard deviation of three independent cultures. For growth assays, WT *CBK1* or *cbk1-as2* strains were transformed with vectors containing Ssd1 (WT and mutants) expressed from a galactose-inducible promoter. Stationary phase cells grown in glucose were diluted to OD_{600} = 0.2. Five-fold serial dilutions (5^{-2}, 5^{-3}, 5^{-4}, 5^{-5}) were plated in the presence of glucose or galactose, then incubated for 3 d at 24°C.

In Vitro Pulldown Assays

MBP fusions were incubated on bead with 1 μM purified Cbk1 on ice for 15 min. Samples were washed 3× with PBS (transferred to fresh tube on last wash), resuspended in SDS-PAGE buffer, and separated by SDS-PAGE. Samples were directly visualized with GelCode Blue (Pierce) or transferred to PVDF membrane for immunoblotting. Cbk1 was detected using His or pS570 primary antibodies. For experiments detected using α-pS570, antibody signal with pure protein was analyzed to control for lot-dependent variations in expression as well as variations between Cbk1 mutants (no variations detected).

In Vitro Kinase Assays

Kinase assays resolved by SDS-PAGE were conducted with GST-Ace2 or GST-Ssd1 fusions. Reactions contained 1–5 μM substrate, 100–250 nM Cbk1, 5 mM $MnCl_2$, 2 mM DTT, 20 μM ATP, 10 μCi γ-^{32}P-ATP in Cbk1 buffer. Samples were removed at indicated time points and

quenched with SDS-PAGE buffer, followed by SDS-PAGE separation. ^{32}P was visualized using a Storm phosphorimager (GE Biosciences) and quantified using ImageQuant software.

For peptide kinase assays, reactions were performed in the presence of 50 nM Cbk1$^{251–756}$–Mob2, 40 μM ATP, 2 mM DTT, 5 mM $MnCl_2$, 2 μCi γ-^{32}P-ATP, and 100 μM Docktide (CDFKFPPPPNAHGGHRRATSN) or AKAPtide (CDAKAPPPPNAHGGHRRATSN). Reactions were initiated by the addition of ATP. Samples were quenched after 1 h by the addition of a 5-μl sample to P81 paper (Whatman) and subsequent submersion in 75 mM phosphoric acid to remove background γ-^{32}P-ATP. After drying, signal was determined by Čerenkov counting in an LS 6500 scintillation counter (Beckman). Counts per minute was efficiency corrected and converted to femtomoles ^{32}P followed by conversion to total ATP (labeled and unlabeled). Significance was determined by an unpaired two-tailed Student's *t*-test.

Fluorescence-Polarization-Based Protein–Peptide Binding Affinity Measurements

For FP-based binding affinity measurements, different docking peptides (pepAce2: GSGSIIVTTNSANGGYQFP; pepSSD1: TTEQSDFKFP; pepHM: IGYTYSRFDY; pepHM-P: IGYpTYSRFDY) were N-terminally (pepAce2 and pepSSD1) or C-terminally (pepHM and pepHM-P) labeled with carboxyfluorescein (CF) fluorescent dyes.

Change in the FP signal in direct binding affinity measurements was monitored as a function of increasing concentration of protein with a Synergy H4 (BioTek Instruments) plate reader in 384-well plates. The labeled peptides were at 20 nM in 20 mM Tris (pH 8.0), 100 mM NaCl, 0.05% Brij35P, 2 mM DTT. The resulting binding isotherms were fit to a quadratic binding equation. The affinity of the unlabeled peptides was measured in steady-state competition experiments: 20 nM labeled reporter was mixed with protein samples in a concentration to achieve ~60%–80% complex formation. Subsequently, increasing amounts of unlabeled peptide were added, and the FP signal was measured as described earlier for direct titration experiments. The K_d for each unlabeled peptide interaction was determined by fitting the data to a competition binding equation. Titration experiments were carried out in triplicate, and the average FP signal was used for fitting the data with OriginPro 7.

Identification of the Cbk1 Docking Surface

The structure of acetyl-DFKFP-NH-methyl ligand was built and optimized using the Tinker program package [75] and the Amber force field [76] with an automatic selection of minimum search algorithms and a gradient norm of 0.001. The target protein was the Cbk1 kinase domain from the 3.3 Å resolution Cbk1–Mob2 crystal structure where missing side chains were inserted automatically by rotamer preferences in Coot and the missing αC helix was homology modeled based on the corresponding region from PKB. The completed structure was energy-minimized using the GROMACS software package [77] using Amber force field, TIP3P explicit water surrounding, steepest descent, and conjugate gradient optimizations with tolerances of 1,000 and 100 kJ mol^{-1} nm^{-1}, respectively. Notably, in the docking calculations the highly unstructured loop between residues Y509 and S552 was not used to avoid docking artifacts, and the terminating residues were mutated into chain-blocking—NH-methyl and acetyl—ends, respectively. We prepared both target and ligand structures for docking with the aid of AutoDockTools [46]. We performed blind docking using AutoDock 4 [78] at 0.375 Å grid spacing in a docking box of 93.8 × 93.8 × 93.8 Å^3 (250 grid points in each direction) covering the entire target. The number of energy evaluations was increased to 50 million, and the population size of the genetic algorithm was set to 250 to handle the completely flexible ligand with 20 active torsions (only omega torsions of backbone amide groups were kept restrained in trans

conformation). More than 200 docking runs were performed and evaluated, and other parameters not described above were set as in previous studies [78–80]. We performed focused docking similarly with an acetyl-FKFP-NH-methyl tetrapeptide using a smaller docking box ($63.8 \times 75.0 \times 63.8$ Å^3) centered at a point (−61.65, 41.95, 5.8) above the docking interface found by blind docking. Unbiased and focused docking resulted in the same binding conformation for the FKFP core motif (Fig 4A).

Binding experiments with constructs containing mutated Ssd1 docking motifs showed that for the FxFP core motif, the first-position tyrosine and phenylalanine are equivalent, but only phenylalanine is tolerated at the third position (Fig 4D and 4E). There is also a positively charged amino acid (lysine or arginine) requirement for the second position. The final model is in excellent agreement with experimental observations on the importance of Ssd1 and Cbk1 docking surface residues for Cbk1–Ssd1 binding (see Figs 4 and S5).

Analysis of Docking Motif and Phosphorylation Consensus Site Conservation

To identify proteins with conserved phosphorylation consensus sites or conserved docking motifs we used ConDens [48] with alignments of related ascomycetes taken from the Yeast Gene Order Browser [81]. Briefly, ConDens compares the density of matches to the specified short peptide sequence in each species to the density of matches in a "reference" species (in our case *S. cerevisiae*) and evaluates the probability that each species retained as many matches (or more) to that sequence in the absence of selection. Each species and each site in the protein are then combined into a single score (shown in Fig 6) for each protein in *S. cerevisiae*.

Accession Numbers

Coordinates and structure factors for the reported crystal structures have been deposited to the PDB under accession codes 4LQP, 4LQQ, and 4LQS.

Supporting Information

S1 Data. Measurements of the distances between the centers of mass of indicated Cbk1 and Mob2 regions over the course of the molecular dynamics simulations described in Figs 3 and S4. This file also contains data quantifying the effectiveness of cell separation in different genetic backgrounds. In these experiments we counted the number of cells present in clumps of specific sizes in random microscope fields, and then binned these groups into clump size ranges: 1–4 cells, 5–10 cells, 11–20 cells, and ≥21 cells per clump. We quantified this for the indicated allele combinations, scoring three independent isolates of each allele combination. The number of cell bodies present in clumps of different sizes is a function of the efficiency of cell separation.
(XLSX)

S2 Data. Fluorescence polarization assay data that are represented in graphical form in Figs S3 and S5.
(XLSX)

S3 Data. Enzymological characterization of the effects of Cbk1 docking. Assay data from experiments that used γ-^{32}P-ATP in protein kinase reactions with synthetic substrate peptides from Ssd1 that either contained both the core Cbk1 docking motif and a Cbk1 phosphorylation consensus motif (Docktide) or contained a core Cbk1 docking motif mutated to abrogate binding to the kinase and a Cbk1 consensus motif (AKAPtide). These kinase assays were performed either with Cbk1(T743E) or with Cbk1(Y687A/T743E). The latter construct carries a mutation

that eliminates docking motif association with the kinase.
(XLSX)

S4 Data. Data quantifying the effectiveness of cell separation in different genetic backgrounds. In these experiments we counted the number of cells present in clumps of specific sizes in random microscope fields, and then binned these groups into clump size ranges: 1–2 cells, 3–4 cells, 5–10 cells, 11–20 cells, and $\geq$21 cells per clump. We quantified this for the indicated allele combinations, scoring three independent isolates of each allele combination. The number of cell bodies present in clumps of different sizes is a function of the efficiency of cell separation.
(XLSX)

S5 Data. Real-time PCR data for experiments measuring mRNA levels of the Ace2 target genes *CTS1*, *DSE1*, and *SCW11*, as well as *ACT1* mRNA in strains carrying different alleles of *CBK1* and *ACE2* (WT, as well as mutant alleles affecting different protein functions).
(XLSX)

S6 Data. Scores computed using the ConDens algorithm [48], which evaluates conservation of sequence motifs in a manner not dependent on precise alignment of orthologous sequences. This file lists proteins in which Cbk1 phosphorylation consensus and core docking motif sequences are significantly conserved, as well as proteins in which one but not the other of these two sequence motifs is significantly conserved.
(XLSX)

S1 Fig. Multiple sequence alignment of NDR kinases. Sequence comparison of Cbk1 from *Saccharomyces cerevisae* (CBK1_YEAST) with other NDR kinases from different organisms (*SCHPO: Schizosaccharomyces pombe; TRIAD: Trichoplax adherens; CAEEL: Caenorhabditis elegans; DROME: Drosphila melanogaster; DANRE: Danio rerio*). The consensus sequence is presented as a sequence logo. Residues and motifs important for Cbk1 activity or for its regulation are boxed: Asp475 is the catalytic aspartate, Ser570 is an autophosphorylation site, while Thr743 is phosphorylated by an upstream kinase. Important regions are underlined. Highlighting is based on sequence identity compared to Cbk1. Secondary structure elements from the crystallographic models are shown above the sequences. Dashed lines indicate regions that could not be built into the crystallographic model of the Cbk1–Mob2 complex.
(PDF)

S2 Fig. Additional information on the crystal structures of Cbk1–Mob2 complexes. (A) Superimposition of the crystallographic models for Cbk1–Mob2 complexes. All four complexes (from three different crystal forms) display the same Cbk1 and Mob2 domain arrangements. (B) The Cbk1 kinase domain (shown in blue) is similar to the AGC kinase domains from related kinases such as PKA (PDB ID: 1JLU; yellow) and ROCK1 (PDB ID: 2ETR; green) [36,37]. In contrast, the activation loop and N-terminal kinase domain extensions adopt markedly different structures. (AGC kinase domains are shown in thin ribbon, while activation loops and N-terminal kinase domain extensions are shown in thicker tube representation.) (C) Stereo view of the final 2Fo-Fc electron density map around the activation loop (cyan) at 3.3 Å resolution. (D) In the two different crystal forms of Cbk1(T743E), the αINH is rotated with an angle of ~30° due to different crystal packing. For clarity, Cbk1 from crystal form A is colored purple and from crystal form B is colored in blue. (E) Simulated annealing 2Fo-Fc omit map contoured at 1σ and calculated around the HM motif region for the 4.5 Å resolution Cbk1–Mob2 structure. (F) Omit map (generated with the same protocol) for the 3.3 Å resolution Cbk1–

Mob2 crystal structure.
(PDF)

S3 Fig. Characterization of the Cbk1 hydrophobic motif and its binding to the Cbk1–Mob2 complex. (A) Summary of the FP binding experiments with Cbk1-HM-containing peptides. Although the Cbk1–Mob2 and the Cbk1(T743E)–Mob2 complexes did not show major changes in the position of their Cbk1 HM motifs, we compared the binding affinity of peptides containing unphosphorylated and phosphorylated HM motifs to a $Cbk1^{\Delta730-756}$ variant complexed with Mob2 (which lacks the HM region) in trans. These in vitro binding affinity measurements indicated that the phosphorylated HM peptide bound more than 5-fold stronger in trans into the open Cbk1 HM-binding slot. Binding of a carboxyfluorescein-labeled Cbk1-HM-containing peptide was monitored and compared to the binding of the phosphorylated HM peptide. In direct titration binding experiments, the binding of the labeled peptide was monitored and binding isotherms were fit to a classical binding equation. Competitive titration experiments (starting from 60%–80% complex formation between the protein and the labeled peptide) monitored the binding of unlabeled peptide as it competed with the labeled peptide for the same binding site. Competitive titrations can indirectly report on the binding affinity of unlabeled peptides if data are fit to a competition binding equation. ND, not determined. (Note that these experiments monitored the binding of peptides in trans, and they do not report on the binding of the HM motif as part of the full-length Cbk1.) (B) Phosphorylated HM motif binds to $Cbk1^{\Delta730-756}$–Mob2 with increased affinity. (C) Competitive titration curves for monitoring HM and HM-P peptide binding to the $Cbk1^{\Delta730-756}$–Mob2 construct. The markedly reduced binding affinity for the unlabeled versus the labeled HM-P peptide is likely due to the presence of the fluorophore, which seems to artificially increase binding. (D) Competitive titration binding curve for monitoring labeled HM-P and Ssd1 docking peptide binding. This panel shows that the docking peptide could not compete with the HM peptide, indicating that the Ssd1 docking peptide does not bind to the groove around the Cbk1 N-linker region. Each measurement is representative of at least two sets of independent experiments where K_d values were calculated from triplicate data points. Triplicates were independently prepared samples that were assayed at the same time. Error bars on the measurement points indicate standard deviations from the mean. Errors in the calculated K_d values indicate uncertainty of the fit to a direct or competition binding equation. FP data can be found in S2 Data.
(PDF)

S4 Fig. Molecular dynamics simulations of Cbk1–Mob2 complexes. (A) Cbk1–Mob2 interface analysis. Upper panel shows the angle distribution between αMOB (Cbk1) and H2 (Mob2) as defined on the left panel. The small differences between the HM-P complex and the HM complex are not statistically significant. Lower panel displays interface area between Cbk1 and Mob2. These analyses show that there is no major global change at the Cbk1–Mob2 interface during MD, confirming that the changes shown in Fig 3 occur because HM-mediated local interactions change within the binding slot. (B) Three-dimensional N-linker–HM, αC–HM and αMOB–HM distance scatter plots on Cbk1 with unphosphorylated and phosphorylated HM. MD simulations were identical to those on Cbk1–Mob2 complexes, but the Mob2 protein chain was removed from the starting MD model. (C) Three-dimensional N-linker–HM, αC–HM, and αMOB–HM distance scatter plots on Cbk1 with HM-E. MD simulations were identical to those on Cbk1–Mob2 complexes, but Cbk1 T743 was mutated to glutamic acid. MD data can be found in S1 Data.
(PDF)

S5 Fig. Identification and analysis of docking motifs in Ace2 and Ssd1. (A) Pulldown of Cbk1–Mob2 by Ace2 truncation fragments (left) and Ace2 fragments centered on the docking motif (right). Smaller fragments than $Ace2^{272–286}$ abrogate Cbk1 interaction. (B) Alanine scan of $Ace2^{270–290}$ pulldown with the $Cbk1^{352–692}$ kinase domain alone highlights the importance of N-terminal hydrophobic residues in addition to the C-terminal core motif. (C) Flexibility analysis of $Ace2^{270–290}$ by glycine insertion/deletion and pulldown of Cbk1–Mob2. The bipartite motif can be extended, but deletion abrogates Cbk1 interaction. WT Ace2 contains two glycine residues. (D) $Cbk1^{352–756}$ kinase domain in vitro kinase assay with $Ace2^{102–306}$. The presence of the docking motif enhances phosphorylation 100-fold as well as enhances Cbk1 autophosphorylation. (E) Pulldown of $Cbk1^{352–756}$ by Ssd1 truncation fragments containing the N-terminal (1–6) or C-terminal (7–11) docking motif. (F) Pulldown of $Cbk1^{352–756}$ by the Ace2 docking motif with stepwise conversion to the Ssd1 docking motif. Conversion of the core motif (YQFP → FKFP) could not rescue mutation of N-terminal residues, highlighting the importance of surrounding sequence to the core motif. (G) Competition of Ace2 (left) and Ssd1 (right) docking motifs with unbound $Ace2^{270–290}$. Competition was analyzed by $Cbk1^{352–756}$ pulldown. (H) FP affinity measurements of peptides containing Ssd1 (top) and Ace2 (bottom) docking motifs with Cbk1–Mob2. (I) FP of Ssd1 with the Cbk1 kinase domain (top) or with the HM deleted (bottom). Neither Cbk1 truncation exhibited defects in docking motif interactions. FP data can be found in S2 Data.
(PDF)

S6 Fig. Cbk1–Ace2 docking confers robustness to regulation of Ace2 target gene transcription. (A) *CBK1* WT cells carrying $ace2^{dock^*}$ have no significant reduction in transcript levels of the *DSE1* gene, while *cbk1-as2* cells carrying $ace2^{dock^*}$ exhibit dramatically reduced *DSE1* transcription. (B) *CBK1* WT cells carrying $ace2^{dock^*}$ exhibit modest reduction in transcript levels of the *SCW11* gene, while *cbk1-as2* cells carrying $ace2^{dock^*}$ exhibit strongly reduced *SCW11* transcription. Transcription data can be found in S5 Data.
(PDF)

S7 Fig. Co-occurrence of docking motifs and Cbk1 phosphorylation consensus sites is extremely significant, and docking sites predict consensus site conservation and phosphorylation. (A) In proteins that contain matches to the Cbk1 docking motif [YF]xFP, there is a statistically significant enrichment of NDR/LATS phosphorylation consensus sites Hx[RK]xx[ST] compared to proteins that do not contain docking motifs (0.45 versus 0.36 matches per 1,000 amino acids; Fisher's exact test, $p = 0.016$). Similar results are obtained for phosphorylation consensus sites for which there is mass spectrometric evidence of phosphorylation as annotated in 2013 by PhosphoGrid [50] (0.040 versus 0.017 phosphorylated matches per 1,000 amino acids; Fisher's exact test, $p = 0.013$). If we consider the 50 most conserved phosphorylation consensus sites as identified by ConDens [48], the enrichment is much stronger (0.067 versus 0.011 conserved matches per 1,000 amino acids; Fisher's exact test, $p = 3.629 \times 10^{-8}$). Since many matches to the docking motif ([YF]xFP) might appear in proteins by chance, we also considered proteins that contain conserved docking motifs. In these proteins the enrichments are much stronger. For example, phosphorylation consensus matches are now >8× more like to appear (2.95 versus 0.36 matches per 1,000 amino acids; Fisher's exact test, $p < 10 \times 10^{-10}$). Similarly, phosphorylated consensus matches and the 50 most conserved phosphorylation consensus matches are >30× and >100× more likely to appear (Fisher's exact test, $p < 10 \times 10^{-6}$ and $p < 10 \times 10^{-10}$, respectively). These tests show that a conserved docking site makes consensus sites (and conserved consensus sites) more likely to appear, and that the phospho-acceptor residues within these consensus sites are significantly more likely to be among the set of phosphorylated positions annotated in the PhosphoGrid database. (B) In the

11 proteins with conserved [YF]xFP docking motifs, 14/27 (51.8%) of the Cbk1 phosphorylation consensus matches are in the 50 most conserved, which is much more than 50/887 (5.6%) in the proteome (Fisher's exact test, $p < 10 \times 10^{-10}$) Similarly, consensus matches are more likely to be phosphorylated if they are found in proteins with conserved docking motifs (18% versus 5%, $p = 0.01$) These tests show that a given Cbk1 phosphorylation consensus match is much more likely to be conserved and phosphorylated in vivo if it is found in a protein that contains a conserved docking motif. Therefore, conserved docking motifs point to the functional Cbk1 consensus sites. ConDens data can be found in S6 Data.
(PDF)

S8 Fig. Comparison of analysis of Cbk1's docking motif binding site with corresponding regions from MAPKs and other AGC kinases. (A) A comparison of Cbk1 bound to the FKFP core docking motif and ERK2 bound to the MNK1 docking motif (2Y9IQ) [55]. The loop between kinase domain D and E helices is noted for both Cbk1 and ERK2, and Cbk1's CLT region is also highlighted. An overlay of the structures shows different peptide binding modes on roughly corresponding surface regions of both kinases. (B) Overlay of the Cbk1 docking region with the corresponding surface of PKA (PDB ID: 1JLU; yellow) and ROCK1 (PDB ID: 2ETR; green) [36,37]. Side chains of amino acids differ, but are placed in a roughly similar geometry.
(PDF)

S9 Fig. Evolutionary plasticity of the site at which Cbk1 binds to the core docking peptide. A gene tree of AGC kinases related to the NDR family, with the closest Cbk1 orthologs in *S. cerevisiae* (Sc), *Sc. pombe* (Sp), *D. melanogaster* (Dm), and *Homo sapiens* (Hs) bracketed in red. The loop between the D and E helices in the kinase C-lobe and the PxxP motif in the CLT region are enclosed in orange boxes. Bulky hydrophobic amino acids involved in docking motif binding in Cbk1 are boxed in pink. Tree is not drawn to scale. PKA and MAST are used as outgroups to the NDR family.
(PDF)

S1 Table. Relevant crystallographic parameters for the three different crystal forms of the Cbk1–Mob2 complex we solved.
(DOCX)

Acknowledgments

This research was supported in part by the computational resources and staff contributions provided by the Quest High Performance Computing Cluster at Northwestern University, which is jointly supported by the Office of the Provost, the Office for Research, and Northwestern University Information Technology. We acknowledge PRACE for providing access to the Monte Rosa computation resource based in Switzerland at Centro Svizzero di Calcolo Scientifico Swiss National Supercomputing Centre, and the National Information Infrastructure Development Institute supercomputing service. We are thankful to Jennifer Brace (Northwestern University) for assistance with site-directed mutagenesis and to Marianna Rakács (Research Centre for Natural Sciences of the Hungarian Academy of Sciences) for assistance with protein expression.

Author Contributions

Conceived and designed the experiments: GG KDS BJY ANNB AMM CH AR ELW. Performed the experiments: GG KDS BJY NA CH. Analyzed the data: GG KDS BJY ANNB AMM CH AR

ELW. Contributed reagents/materials/analysis tools: GG KDS BJY ANNB AMM CH AR ELW. Wrote the paper: GG KDS BJY ANNB AMM CH AR ELW.

References

1. Avruch J, Zhou D, Fitamant J, Bardeesy N, Mou F, et al. (2012) Protein kinases of the Hippo pathway: regulation and substrates. Semin Cell Dev Biol 23: 770–784. doi: 10.1016/j.semcdb.2012.07.002 PMID: 22898666
2. Hergovich A, Hemmings BA (2012) Hippo signalling in the G2/M cell cycle phase: lessons learned from the yeast MEN and SIN pathways. Semin Cell Dev Biol 23: 794–802. doi: 10.1016/j.semcdb.2012.04.001 PMID: 22525225
3. Hergovich A, Stegert MR, Schmitz D, Hemmings BA (2006) NDR kinases regulate essential cell processes from yeast to humans. Nat Rev Mol Cell Biol 7: 253–264. PMID: 16607288
4. Pan D (2010) The hippo signaling pathway in development and cancer. Dev Cell 19: 491–505. doi: 10.1016/j.devcel.2010.09.011 PMID: 20951342
5. Zhao B, Tumaneng K, Guan KL (2011) The Hippo pathway in organ size control, tissue regeneration and stem cell self-renewal. Nat Cell Biol 13: 877–883. doi: 10.1038/ncb2303 PMID: 21808241
6. Enderle L, McNeill H (2013) Hippo gains weight: added insights and complexity to pathway control. Sci Signal 6: re7.
7. Cornils H, Kohler RS, Hergovich A, Hemmings BA (2011) Downstream of human NDR kinases: impacting on c-myc and p21 protein stability to control cell cycle progression. Cell Cycle 10: 1897–1904. PMID: 21593588
8. Emoto K, He Y, Ye B, Grueber WB, Adler PN, et al. (2004) Control of dendritic branching and tiling by the Tricornered-kinase/Furry signaling pathway in Drosophila sensory neurons. Cell 119: 245–256. PMID: 15479641
9. Geng W, He B, Wang M, Adler PN (2000) The tricornered gene, which is required for the integrity of epidermal cell extensions, encodes the Drosophila nuclear DBF2-related kinase. Genetics 156: 1817–1828. PMID: 11102376
10. Cornils H, Kohler RS, Hergovich A, Hemmings BA (2011) Human NDR kinases control G(1)/S cell cycle transition by directly regulating p21 stability. Mol Cell Biol 31: 1382–1395. doi: 10.1128/MCB.01216-10 PMID: 21262772
11. Bardin AJ, Amon A (2001) Men and sin: what's the difference? Nat Rev Mol Cell Biol 2: 815–826. PMID: 11715048
12. Rock JM, Amon A (2011) Cdc15 integrates Tem1 GTPase-mediated spatial signals with Polo kinase-mediated temporal cues to activate mitotic exit. Genes Dev 25: 1943–1954. doi: 10.1101/gad.17257711 PMID: 21937712
13. Wurzenberger C, Gerlich DW (2011) Phosphatases: providing safe passage through mitotic exit. Nat Rev Mol Cell Biol 12: 469–482. doi: 10.1038/nrm3149 PMID: 21750572
14. Rock JM, Lim D, Stach L, Ogrodowicz RW, Keck JM, et al. (2013) Activation of the yeast Hippo pathway by phosphorylation-dependent assembly of signaling complexes. Science 340: 871–875. doi: 10.1126/science.1235822 PMID: 23579499
15. Weiss EL (2012) Mitotic exit and separation of mother and daughter cells. Genetics 192: 1165–1202. doi: 10.1534/genetics.112.145516 PMID: 23212898
16. Hsu J, Weiss EL (2013) Cell cycle regulated interaction of a yeast Hippo kinase and its activator MO25/Hym1. PLoS ONE 8: e78334. doi: 10.1371/journal.pone.0078334 PMID: 24205201
17. Nelson B, Kurischko C, Horecka J, Mody M, Nair P, et al. (2003) RAM: a conserved signaling network that regulates Ace2p transcriptional activity and polarized morphogenesis. Mol Biol Cell 14: 3782–3803. PMID: 12972564
18. Doolin MT, Johnson AL, Johnston LH, Butler G (2001) Overlapping and distinct roles of the duplicated yeast transcription factors Ace2p and Swi5p. Mol Microbiol 40: 422–432. PMID: 11309124
19. Mazanka E, Alexander J, Yeh BJ, Charoenpong P, Lowery DM, et al. (2008) The NDR/LATS family kinase Cbk1 directly controls transcriptional asymmetry. PLoS Biol 6: e203. doi: 10.1371/journal.pbio.0060203 PMID: 18715118
20. Colman-Lerner A, Chin TE, Brent R (2001) Yeast Cbk1 and Mob2 activate daughter-specific genetic programs to induce asymmetric cell fates. Cell 107: 739–750. PMID: 11747810
21. Weiss EL, Kurischko C, Zhang C, Shokat K, Drubin DG, et al. (2002) The Saccharomyces cerevisiae Mob2p-Cbk1p kinase complex promotes polarized growth and acts with the mitotic exit network to

facilitate daughter cell-specific localization of Ace2p transcription factor. J Cell Biol 158: 885–900. PMID: 12196508

22. Jansen JM, Wanless AG, Seidel CW, Weiss EL (2009) Cbk1 regulation of the RNA-binding protein Ssd1 integrates cell fate with translational control. Curr Biol 19: 2114–2120. doi: 10.1016/j.cub.2009.10.071 PMID: 19962308
23. Kurischko C, Kim HK, Kuravi VK, Pratzka J, Luca FC (2011) The yeast Cbk1 kinase regulates mRNA localization via the mRNA-binding protein Ssd1. J Cell Biol 192: 583–598. doi: 10.1083/jcb.201011061 PMID: 21339329
24. Hergovich A (2011) MOB control: reviewing a conserved family of kinase regulators. Cell Signal 23: 1433–1440. doi: 10.1016/j.cellsig.2011.04.007 PMID: 21539912
25. Liu G, Young D (2012) Conserved Orb6 phosphorylation sites are essential for polarized cell growth in Schizosaccharomyces pombe. PLoS ONE 7: e37221. doi: 10.1371/journal.pone.0037221 PMID: 22629372
26. Chan EH, Nousiainen M, Chalamalasetty RB, Schafer A, Nigg EA, et al. (2005) The Ste20-like kinase Mst2 activates the human large tumor suppressor kinase Lats1. Oncogene 24: 2076–2086. PMID: 15688006
27. Stegert MR, Hergovich A, Tamaskovic R, Bichsel SJ, Hemmings BA (2005) Regulation of NDR protein kinase by hydrophobic motif phosphorylation mediated by the mammalian Ste20-like kinase MST3. Mol Cell Biol 25: 11019–11029. PMID: 16314523
28. Kannan N, Haste N, Taylor SS, Neuwald AF (2007) The hallmark of AGC kinase functional divergence is its C-terminal tail, a cis-acting regulatory module. Proc Natl Acad Sci U S A 104: 1272–1277. PMID: 17227859
29. Pearce LR, Komander D, Alessi DR (2010) The nuts and bolts of AGC protein kinases. Nat Rev Mol Cell Biol 11: 9–22. doi: 10.1038/nrm2822 PMID: 20027184
30. Brace J, Hsu J, Weiss EL (2011) Mitotic exit control of the Saccharomyces cerevisiae Ndr/LATS kinase Cbk1 regulates daughter cell separation after cytokinesis. Mol Cell Biol 31: 721–735. doi: 10.1128/MCB.00403-10 PMID: 21135117
31. Panozzo C, Bourens M, Nowacka A, Herbert CJ (2010) Mutations in the C-terminus of the conserved NDR kinase, Cbk1p of Saccharomyces cerevisiae, make the protein independent of upstream activators. Mol Genet Genomics 283: 111–122. doi: 10.1007/s00438-009-0501-3 PMID: 19967545
32. Bichsel SJ, Tamaskovic R, Stegert MR, Hemmings BA (2004) Mechanism of activation of NDR (nuclear Dbf2-related) protein kinase by the hMOB1 protein. J Biol Chem 279: 35228–35235. PMID: 15197186
33. Mrkobrada S, Boucher L, Ceccarelli DF, Tyers M, Sicheri F (2006) Structural and functional analysis of Saccharomyces cerevisiae Mob1. J Mol Biol 362: 430–440. PMID: 16934835
34. Stavridi ES, Harris KG, Huyen Y, Bothos J, Verwoerd PM, et al. (2003) Crystal structure of a human Mob1 protein: toward understanding Mob-regulated cell cycle pathways. Structure 11: 1163–1170. PMID: 12962634
35. Tamaskovic R, Bichsel SJ, Hemmings BA (2003) NDR family of AGC kinases—essential regulators of the cell cycle and morphogenesis. FEBS Lett 546: 73–80. PMID: 12829239
36. Jacobs M, Hayakawa K, Swenson L, Bellon S, Fleming M, et al. (2006) The structure of dimeric ROCK I reveals the mechanism for ligand selectivity. J Biol Chem 281: 260–268. PMID: 16249185
37. Madhusudan, Trafny EA, Xuong NH, Adams JA, Ten Eyck LF, et al. (1994) cAMP-dependent protein kinase: crystallographic insights into substrate recognition and phosphotransfer. Protein Sci 3: 176–187. PMID: 8003955
38. Narayana N, Cox S, Shaltiel S, Taylor SS, Xuong N (1997) Crystal structure of a polyhistidine-tagged recombinant catalytic subunit of cAMP-dependent protein kinase complexed with the peptide inhibitor PKI(5–24) and adenosine. Biochemistry 36: 4438–4448. PMID: 9109651
39. Yang J, Cron P, Thompson V, Good VM, Hess D, et al. (2002) Molecular mechanism for the regulation of protein kinase B/Akt by hydrophobic motif phosphorylation. Mol Cell 9: 1227–1240. PMID: 12086620
40. Mok J, Kim PM, Lam HY, Piccirillo S, Zhou X, et al. (2010) Deciphering protein kinase specificity through large-scale analysis of yeast phosphorylation site motifs. Sci Signal 3: ra12. doi: 10.1126/scisignal.2000482 PMID: 20159853
41. Ubersax JA, Ferrell JE Jr (2006) A noisy 'Start' to the cell cycle. Mol Syst Biol 2: 2006.0014.
42. Davey NE, Van Roey K, Weatheritt RJ, Toedt G, Uyar B, et al. (2012) Attributes of short linear motifs. Mol Biosyst 8: 268–281. doi: 10.1039/c1mb05231d PMID: 21909575
43. Remenyi A, Good MC, Lim WA (2006) Docking interactions in protein kinase and phosphatase networks. Curr Opin Struct Biol 16: 676–685. PMID: 17079133

44. Nguyen Ba AN, Yeh BJ, van Dyk D, Davidson AR, Andrews BJ, et al. (2012) Proteome-wide discovery of evolutionary conserved sequences in disordered regions. Sci Signal 5: rs1. doi: 10.1126/scisignal.2002515 PMID: 22416277

45. Endicott JA, Noble ME, Johnson LN (2012) The structural basis for control of eukaryotic protein kinases. Annu Rev Biochem 81: 587–613. doi: 10.1146/annurev-biochem-052410-090317 PMID: 22482904

46. Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, et al. (2009) AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility. J Comput Chem 30: 2785–2791. doi: 10.1002/jcc.21256 PMID: 19399780

47. Jansen JM, Barry MF, Yoo CK, Weiss EL (2006) Phosphoregulation of Cbk1 is critical for RAM network control of transcription and morphogenesis. J Cell Biol 175: 755–766. PMID: 17145962

48. Lai AC, Nguyen Ba AN, Moses AM (2012) Predicting kinase substrates using conservation of local motif density. Bioinformatics 28: 962–969. doi: 10.1093/bioinformatics/bts060 PMID: 22302575

49. Chatr-Aryamontri A, Breitkreutz BJ, Heinicke S, Boucher L, Winter A, et al. (2013) The BioGRID interaction database: 2013 update. Nucleic Acids Res 41: D816–D823. doi: 10.1093/nar/gks1158 PMID: 23203989

50. Sadowski I, Breitkreutz BJ, Stark C, Su TC, Dahabieh M, et al. (2013) The PhosphoGRID Saccharomyces cerevisiae protein phosphorylation site database: version 2.0 update. Database (Oxford) 2013: bat026.

51. Yang J, Cron P, Good VM, Thompson V, Hemmings BA, et al. (2002) Crystal structure of an activated Akt/protein kinase B ternary complex with GSK3-peptide and AMP-PNP. Nat Struct Biol 9: 940–944. PMID: 12434148

52. Cook D, Hoa LY, Gomez V, Gomez M, Hergovich A (2014) Constitutively active NDR1-PIF kinase functions independent of MST1 and hMOB1 signalling. Cell Signal 26: 1657–1667. doi: 10.1016/j.cellsig.2014.04.011 PMID: 24747552

53. Malakhova M, Kurinov I, Liu K, Zheng D, D'Angelo I, et al. (2009) Structural diversity of the active N-terminal kinase domain of p90 ribosomal S6 kinase 2. PLoS ONE 4: e8044. doi: 10.1371/journal.pone.0008044 PMID: 19956600

54. Jeffrey PD, Russo AA, Polyak K, Gibbs E, Hurwitz J, et al. (1995) Mechanism of CDK activation revealed by the structure of a cyclinA-CDK2 complex. Nature 376: 313–320. PMID: 7630397

55. Garai A, Zeke A, Gogl G, Toro I, Fordos F, et al. (2012) Specificity of linear motifs that bind to a common mitogen-activated protein kinase docking groove. Sci Signal 5: ra74. doi: 10.1126/scisignal.2003004 PMID: 23047924

56. Lee T, Hoofnagle AN, Kabuyama Y, Stroud J, Min X, et al. (2004) Docking motif interactions in MAP kinases revealed by hydrogen exchange mass spectrometry. Mol Cell 14: 43–55. PMID: 15068802

57. Bender L, Lo HS, Lee H, Kokojan V, Peterson V, et al. (1996) Associations among PH and SH3 domain-containing proteins and Rho-type GTPases in Yeast. J Cell Biol 133: 879–894. PMID: 8666672

58. Matsui Y, Matsui R, Akada R, Toh-e A (1996) Yeast src homology region 3 domain-binding proteins involved in bud formation. J Cell Biol 133: 865–878. PMID: 8666671

59. Norden C, Mendoza M, Dobbelaere J, Kotwaliwale CV, Biggins S, et al. (2006) The NoCut pathway links completion of cytokinesis to spindle midzone function to prevent chromosome breakage. Cell 125: 85–98. PMID: 16615892

60. Shepard KA, Gerber AP, Jambhekar A, Takizawa PA, Brown PO, et al. (2003) Widespread cytoplasmic mRNA transport in yeast: identification of 22 bud-localized transcripts using DNA microarray analysis. Proc Natl Acad Sci U S A 100: 11429–11434. PMID: 13679573

61. Bourens M, Panozzo C, Nowacka A, Imbeaud S, Mucchielli MH, et al. (2009) Mutations in the Saccharomyces cerevisiae kinase Cbk1p lead to a fertility defect that can be suppressed by the absence of Brr1p or Mpt5p (Puf5p), proteins involved in RNA metabolism. Genetics 183: 161–173. doi: 10.1534/genetics.109.105130 PMID: 19546315

62. Goldstrohm AC, Seay DJ, Hook BA, Wickens M (2007) PUF protein-mediated deadenylation is catalyzed by Ccr4p. J Biol Chem 282: 109–114. PMID: 17090538

63. Tompa P, Davey NE, Gibson TJ, Babu MM (2014) A million peptide motifs for the molecular biologist. Mol Cell 55: 161–169. doi: 10.1016/j.molcel.2014.05.032 PMID: 25038412

64. Jin J, Pawson T (2012) Modular evolution of phosphorylation-based signalling systems. Philos Trans R Soc Lond B Biol Sci 367: 2540–2555. doi: 10.1098/rstb.2012.0106 PMID: 22889906

65. Richter DJ, King N (2013) The genomic and cellular foundations of animal origins. Annu Rev Genet 47: 509–537. doi: 10.1146/annurev-genet-111212-133456 PMID: 24050174

66. Kabsch W (2010) XDS. Acta Crystallogr D Biol Crystallogr 66: 125–132. doi: 10.1107/S0907444909047337 PMID: 20124692

67. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, et al. (2007) Phaser crystallographic software. J Appl Crystallogr 40: 658–674. PMID: 19461840

68. Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, et al. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr D Biol Crystallogr 66: 213–221. doi: 10.1107/S0907444909052925 PMID: 20124702

69. Emsley P, Lohkamp B, Scott WG, Cowtan K (2010) Features and development of Coot. Acta Crystallogr D Biol Crystallogr 66: 486–501. doi: 10.1107/S0907444910007493 PMID: 20383002

70. Pronk S, Pall S, Schulz R, Larsson P, Bjelkmar P, et al. (2013) GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. Bioinformatics 29: 845–854. doi: 10.1093/bioinformatics/btt055 PMID: 23407358

71. Duan Y, Wu C, Chowdhury S, Lee MC, Xiong G, et al. (2003) A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. J Comput Chem 24: 1999–2012. PMID: 14531054

72. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) Comparison of simple potential functions for simulating liquid water. J Chem Phys 79: 926–935.

73. Hess B, Bekker H, Berendsen HJC, Fraaije JGEM (1998) LINCS: A Linear Constraint Solver for Molecular Simulations. J Comput Chem 18: 1463–1472.

74. Bussi G, Donadio D, Parrinello M (2007) Canonical sampling through velocity rescaling. J Chem Phys 126: 014101. PMID: 17212484

75. Ren P, Ponder JW (2003) Polarizable atomic multipole water model for molecular mechanics simulation. J Phys Chem B 107: 5933–5947.

76. Wang J, Cieplak P, Kollman PA (2000) How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? J Comput Chem 21: 1049–1074.

77. Hess B, Kutzner C, van der Spoel D, Lindahl E (2008) GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. J Chem Theory Comput 4: 435–447.

78. Hetenyi C, van der Spoel D (2002) Efficient docking of peptides to proteins without prior knowledge of the binding site. Protein Sci 11: 1729–1737. PMID: 12070326

79. Hetenyi C, van der Spoel D (2006) Blind docking of drug-sized compounds to proteins with up to a thousand residues. FEBS Lett 580: 1447–1450. PMID: 16460734

80. Hetenyi C, van der Spoel D (2011) Toward prediction of functional protein pockets using blind docking and pocket search algorithms. Protein Sci 20: 880–893. doi: 10.1002/pro.618 PMID: 21413095

81. Byrne KP, Wolfe KH (2005) The Yeast Gene Order Browser: combining curated homology and syntenic context reveals gene fate in polyploid species. Genome Res 15: 1456–1461. PMID: 16169922

Structural assembly of the signaling competent ERK2–RSK1 heterodimeric protein kinase complex

Anita Alexa[a,1], Gergő Gógl[a,b,1], Gábor Glatz[a], Ágnes Garai[b], András Zeke[a], János Varga[b], Erika Dudás[c], Norbert Jeszenői[d], Andrea Bodor[c], Csaba Hetényi[e], and Attila Reményi[a,2]

[a]Lendület Protein Interaction Group, Institute of Enzymology, Research Centre for Natural Sciences, and [e]MTA-ELTE Molecular Biophysics Research Group, Hungarian Academy of Sciences, 1117 Budapest, Hungary; Departments of [b]Biochemistry and [d]Genetics, Eötvös Loránd University, 1117 Budapest, Hungary; and [c]Institute of Chemistry, Laboratory of Structural Chemistry and Biology, 1117 Budapest, Hungary

Edited by Robert M. Stroud, University of California, San Francisco, CA, and approved January 27, 2015 (received for review September 12, 2014)

Mitogen-activated protein kinases (MAPKs) bind and activate their downstream kinase substrates, MAPK-activated protein kinases (MAPKAPKs). Notably, extracellular signal regulated kinase 2 (ERK2) phosphorylates ribosomal S6 kinase 1 (RSK1), which promotes cellular growth. Here, we determined the crystal structure of an RSK1 construct in complex with its activator kinase. The structure captures the kinase–kinase complex in a precatalytic state where the activation loop of the downstream kinase (RSK1) faces the enzyme's (ERK2) catalytic site. Molecular dynamics simulation was used to show how this heterodimer could shift into a signaling-competent state. This structural analysis combined with biochemical and cellular studies on MAPK→MAPKAPK signaling showed that the interaction between the MAPK binding linear motif (residing in a disordered kinase domain extension) and the ERK2 "docking" groove plays the major role in making an encounter complex. This interaction holds kinase domains proximal as they "readjust," whereas generic kinase domain surface contacts bring them into a catalytically competent state.

protein kinase | signal transduction | structural biology | ERK2 | RSK1

Protein kinase activity is controlled by phosphorylation at its activation loop by upstream kinases (1, 2). Therefore, a catalytically competent kinase–kinase pair must involve surface contacts around the catalytic center of the upstream kinase binding to the activation loop of the downstream kinase. Because of the transient and presumably highly dynamic nature of these enzyme–substrate interactions, little is known about the structural assembly of cognate kinase–kinase pairs. For example, the pivotal role of mitogen-activated protein kinase (MAPK)→MAPK-activated protein kinase (MAPKAPK) signaling events propagating mitogenic and stress signals is well established, but it is structurally not known how a catalytically competent MAPK–MAPKAPK enzyme–substrate complex forms.

Extracellular signals or mitogen stimulation activate the extracellular signal regulated kinase (ERK) pathway, which comprises a hierarchically organized kinase cascade (3, 4). ERK2 becomes phosphorylated by upstream MKK1/2 kinases on a threonine (Thr185) and a tyrosine (Tyr187) residue located in its activation loop (5). In turn, activated ERK1/2 activates ribosomal S6 kinase 1 (RSK1) by sequential phosphorylation events where double-phosphorylated ERK1/2 (ppERK1/2) first phosphorylates the C-terminal RSK1 kinase domain at its activation loop (on Thr573). This is required for the activation of the N-terminal AGC kinase-type domain that will in turn become capable of phosphorylating cell growth promoting substrates (6). Other MAPKAPKs such as MAPKAPK2 (MK2) or MAP kinase-interacting serine/threonine-protein kinase 1 (MNK1) have only one kinase domain that directly phosphorylates downstream substrates. The three proteins are evolutionarily related, activated by MAPKs similarly, but they play markedly different physiological roles (7).

In addition to the transient interactions forming between enzyme–substrate kinase domain pairs, efficient phosphorylation of all 11 mammalian MAPKAPKs by MAPKs (e.g., ERK1/2 and p38 kinases) requires an intact ∼20- to 30-amino-acid-long extension following the C-terminal kinase domain (8–10). This region harbors a MAPK binding consensus sequence referred to as a linear motif (LM) (11). MAPKAPKs all contain a domain related to the kinase domain of calcium/calmodulin-dependent kinases (CAMKs), which is phosphorylated in its activation loop by activated MAPKs. This is the first step in MAPKAPK activation (12, 13). Different MAPKAPKs have diverse sets of substrates but the first step of MAPK→MAPKAPK activation may share a common mechanism.

In the present study, we determined the crystal structure of unphosphorylated ERK2 in complex with an RSK1 construct composed of the C-terminal kinase domain and the linear motif (hereafter referred to as RSK1). The complex is in a precatalytic quaternary arrangement where the activation loop of the downstream kinase (RSK1) faces the enzyme's (ERK2) catalytic site.

Results

Crystal Structure of ERK2–RSK1 Captures a MAPK–MAPKAPK Complex in a Precatalytic State. To structurally elucidate protein–protein interactions involved in the first step of RSK1 activation, we determined the crystal structure of unphosphorylated ERK2 bound to RSK1 at 2.15-Å resolution (R_{free} = 20.8%; *SI Appendix*, Table S1). This complex captured the quaternary structure of

Significance

Signaling pathways often use kinase cascades, but structural characterization of catalytic complexes between heterodimeric kinase pairs has been elusive. For MAPK–MAPKAPK binary complexes, a high-affinity "docking" interaction holds kinase domains proximal within a tethered complex. This heterodimer provided a unique opportunity to shed light on kinase domain–domain contacts that play a role in the assembly of the transient catalytic complex. Starting out from a new precatalytic extracellular signal regulated kinase 2–ribosomal S6 kinase 1 (ERK2–RSK1) crystallographic complex, where the activation loop of the downstream kinase (RSK1) faced the enzyme's (ERK2) catalytic site, we used molecular dynamics simulation to show how the catalytic ERK2–RSK1 complex forms. Our findings reveal the dynamic process through which transient, physiologically relevant kinase heterodimers form in a prototypical kinase cascade.

Author contributions: A.A., G. Gógl, and A.R. designed research; A.A., G. Gógl, G. Glatz, Á.G., A.Z., J.V., E.D., and C.H. performed research; A.A., G. Gógl, G. Glatz, Á.G., A.Z., E.D., N.J., A.B., C.H., and A.R. analyzed data; and A.R. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

Data deposition: The atomic coordinates and structure factors have been deposited in the Protein Data Bank, www.pdb.org (PDB ID code 4NIF).

[1]A.A. and G. Gógl contributed equally to this work.

[2]To whom correspondence should be addressed. Email: remenyi.attila@ttk.mta.hu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1417571112/-/DCSupplemental.

a MAPK–kinase substrate pair in which the activation loop of the downstream kinase is positioned next to the catalytic site of its activator kinase (Fig. 1*A*). The face-to-face stoichiometric complex displays a bipartite protein–protein interface with a buried surface area of ~1,500 Å^2. The RSK1 linear motif region forms several hydrogen bonds, salt bridges, and side-chain–specific van der Waals interactions in the MAPK docking groove (interface 1, IF1) (Fig. 1*B*). This part of the new protein–protein complex shows an excellent agreement to the crystal structure of ERK2 bound to the chemically synthesized RSK1 linear motif peptide (11). The second interface (interface 2, IF2) forms between the two kinase domains and it makes up half of the total interaction surface. IF2 is dominated by van der Waals interactions and it forms between generic kinase features that are highly conserved across different kinase families (e.g., Ala-Pro-Glu, APE motif, the conserved segment of the kinase activation loop or the P loop involved in ATP cofactor binding) (Fig. 1*B*). In contrast, IF1 involves the so-called MAPK docking groove, which is a protein surface that shows topographical diversity even between closely related MAPK family members (e.g., ERK2, p38α) (11).

To test whether the observed face-to-face arrangement of the complex plays a role in RSK1 activation, we analyzed the impact of binding surface disrupting mutations: (*i*) RSK1 was mutated in a central linear motif position (the leucine in φ_A was changed to glutamate, L714E) or (*ii*) the linker connecting the linear motif and the kinase domain was shortened by 2, 4, and 6 residues (RSK1_ΔL2,4,6). RSK1 activation by preactivated ERK2 was monitored in in vitro kinase assays (Fig. 2*A*). Whereas phosphorylation of RSK1_ΔL2 and RSK1_ΔL4 was only slightly affected, the initial rate of Thr573 phosphorylation was greatly decreased with RSK1(L714E) and RSK1_ΔL6 compared with the wild-type substrate (*SI Appendix*, Fig. S1*A*). Differences were not due to impaired structural integrity of RSK1 constructs as mutants and wild-type displayed identical circular dichroism spectra (*SI Appendix*, Fig. S1*B*). The reduced rate of RSK1(714E) activation is likely due to its impaired capacity to bind to ERK2 as their binding affinity is greatly reduced (>50 μM) (*SI Appendix*, Fig. S1*C*). In contrast, RSK1_ΔL6 binds ERK2 with similar affinity compared with wild type (K_d ~ 0.2 μM), but it is a suboptimal substrate because its reduced linker length presumably limits formation of contacts in the face-to-face ERK2–RSK1 heterodimer at IF2.

Next, phosphorylation of RSK1 mutants was examined in a cell-based assay to examine the physiological relevance of the crystallographic complex. HEK293 cells were transiently transfected with RSK1 mutant constructs or with wild type, the ERK pathway was stimulated by addition of epidermal growth factor (EGF), and RSK1 phosphorylation following endogenous ERK2 activation was monitored by Western blots using a phospho-RSK1(Thr573) antibody (Fig. 2*B*). This cell-based assay showed that RSK1(L714E) and the RSK1_ΔL6 both had reduced phosphorylation compared with wild type. Blocking ERK–RSK1 interaction through IF1 appeared to have a more severe impact in cells [RSK1(L714E)]; nevertheless, EGF stimulation also caused diminished RSK1_ΔL6 activation. In summary, these experiments suggest that, despite the fact that ERK2 was unphosphorylated in the complex, the ERK2–RSK1 crystal structure captured a physiologically relevant heterodimeric state that plays a role in ERK2→RSK1 activation.

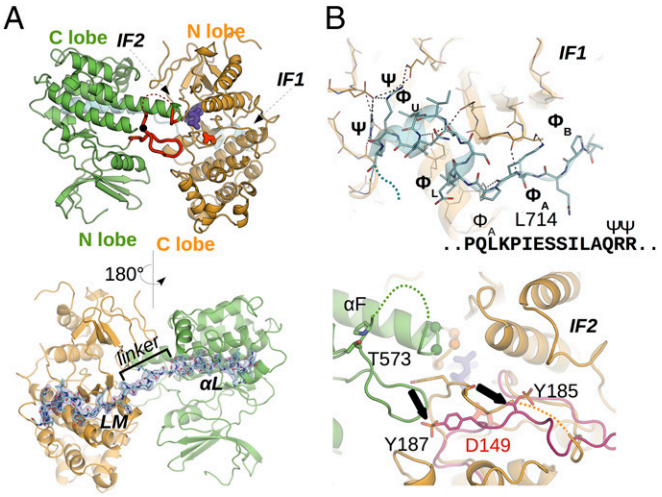

Fig. 1. Structure of the ERK2–RSK1 complex. (*A*) Crystal structure of the ERK2 (orange)–RSK1 (green) complex. The RSK1 linear motif region binds to the MAPK docking groove (interface 1, IF1). RSK1 forms a face-to-face antiparallel kinase dimer with the ERK2 kinase domain through IF2 where the catalytic aspartate (Asp149, shown in red) is located next to the RSK1 activation loop (shown in red, and Thr573 is shown with a black sphere). The nucleotide cofactor (AMPPNP) is colored blue. The unstructured part of the RSK1 loop is shown with a dashed line. *Lower* shows the complex from the back where the electron density for the linear motif (LM), αL, and the intervening linker region (colored in cyan) is shown with sigmaA-weighted omit map contoured at 1σ. (*B*) A close view of IF1 and IF2. At IF1, contacts are highlighted and labeled according to linear motif consensus (11). Black dashed lines indicate H-bond interactions. At IF2 contacts form between the APE motif of RSK1 and the P loop of ERK2 (Cα atoms of residues within van der Waals contacts are shown with spheres). The dephosphorylated activation loop of ERK2 blocks the RSK1 phosphorylation site (Thr573 in the TP motif) from accessing the ERK2 active site (Asp149, shown in stick representation in red) and substrate binding pocket. The ppERK2–RSK1 complex was generated by superposing the ppERK2 structure (PDB ID: 2ERK) with ERK2 from the crystallographic complex. The phosphorylated activation loop of ERK2 (with Tyr187 and Thr185) is shown in purple.

Molecular Dynamics on ERK2–RSK1 Complexes. The ERK2 catalytic center and the RSK1 activation loop face each other in the unphosphorylated ERK2–RSK1 complex. The enzyme's active site (Asp149), however, is shielded off from the substrate by the unphosphorylated ERK2 activation loop (Fig. 1*B*). Thus, this crystallographic model likely captures the snapshot of a precatalytic MAPK–MAPKAPK complex. In contrast, to inactive ERK2, the activated form has an open active site (5) (Fig. 1*B*). Therefore, we used the ERK2–RSK1 binary complex crystal structure as the basis for generating a phosphoERK2–RSK1 model by superpositioning double phosphorylated ERK2 (ppERK2) with unphosphorylated ERK2 within this complex. The complexes were then subjected to molecular dynamics (MD) simulations.

Unrestrained MD simulation on ppERK2–RSK1 showed markedly different movements of the activator and substrate kinase domains relative to each other compared with the complex containing unphosphorylated ERK2. Principal component analysis of MD results revealed great differences: Diverse domain orientation movements suggested a conformationally divergent ERK2–RSK1 complex, whereas intramolecular movements were dominating in the ppERK2–RSK1 complex. These latter appeared to maximize/optimize the interaction surface around ERK2's catalytic core (*SI Appendix* and Movie S1). During these simulations IF1 was highly stable, displaying only small range variations. In contrast, kinase domain–domain contacts (around IF2) changed greatly and the buried surface area increased compared with the starting state (*SI Appendix*, Fig. S2*A*). The interaction between the APE motif of RSK1 and the P loop from ERK2 was stable for both complexes during the 150-ns-long MD run, and this prominent contact surface appeared as a pivot point around which the kinases swiveled to optimize their interaction surface (*SI Appendix* and Movies S2 and S3). Although the interaction surface increased for both complexes, indicating that this can be optimized compared with what was observed in the crystallographic complex, contacts became more extensive in the ppERK2–RSK1 complex. This was because RSK1 formed a unique surface with ppERK2 in addition to contacts formed with unphosphorylated ERK2. Unique contacts formed between residues of αD and αG

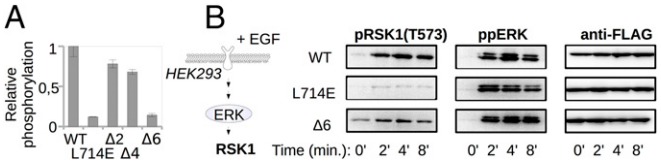

Fig. 2. Experimental validation of ERK2–RSK1 contacts. (*A*) RSK1 mutants display reduced phosphorylation by preactivated ERK2. Phosphorylation rate of RSK1 was monitored by in vitro kinase assays (*SI Appendix*, Fig. S1). L714E has a mutation at IF1, whereas linker length mutants (Δ2–6) affect contacts at IF2 indirectly. The bar graph shows relative RSK1 phosphorylation, which was calculated based on initial phosphorylation rates that were normalized to wild type (WT), and error bars show the SDs from mean value; $n = 3$ experiments (*SI Appendix*, Fig. S1*A*). (*B*) RSK1 mutants display reduced activation capacity upon EGF stimulation in HEK293 cells. (*Right*) Representative set of three independent experiments where samples were taken at the indicated time points after EGF treatment of serum-starved HEK293 cells. Heterologous RSK1 and endogenous ERK activation, the latter as a control for EGF treatment, were monitored by Western blots with phospho-RSK1 (Thr573) and ppERK2-specific antibodies, respectively. Transiently transfected RSK1 constructs had a FLAG tag, and anti-FLAG Western blotting was used to demonstrate equal load for different samples and uniform transfection efficiency for different experiments.

from the C lobe of ppERK2 and the N lobe of RSK1 (e.g., residues from the P loop, and from loops connecting β3 and αC or β4 and β5) (*SI Appendix*, Fig. S2*B*).

Next, we validated the importance of contacts suggested by the ppERK2–RSK1 MD model. Ser452 and Glu623 are located on the N- and C-terminal RSK1 kinase lobes, on the β3–αC loop and on αG, respectively (Fig. 3*A*). Note that Ser452 and Glu623 are located on contact surface patches that were implicated by MD simulations only and they do not form contacts within the ERK2–RSK1 crystal structure (*SI Appendix*, Fig. S3*A*). These two residues were mutated to bulky tryptophan amino acids. MD indicated that the RSK1 APE motif is at the center of kinase domain contacts in the ERK2–RSK1 heterodimer and this region was also subjected to amino acid replacements. The impact of these mutations or their combination were tested in in vitro kinase assays and in cell-based assays (Fig. 3 *B* and *C* and *SI Appendix*, Fig. S3 *B* and *C*). In line with the results of in vitro kinase assays, RSK1 mutants got activated less in a cell-based assay where ERK2 and RSK1 phosphorylation was triggered by stimulating cells with EGF. To demonstrate that RSK1 mutants were structurally intact and could bind ERK2 through their linear motif region, the RSK1(APE/623) construct was subjected to circular dichroism and ERK2 binding measurements (*SI Appendix*, Fig. S3*D*). This RSK1 mutant, that ppERK2 activated the least, did not show any difference compared with wild type, indicating that its structure and enzyme binding capacity stayed intact. Because amino acid replacements had to be made on an extensive, presumably dynamic and "loose" interface where van der Waals interactions appeared to dominate, residues were changed to bulky amino acids such as to tryptophan or arginine instead of alanines. We argued that some local clashes would rather impede dynamic complex assembly as opposed to mere side-chain shortening on an extensive IF2-like surface. In summary, our experimental results validated ERK2–RSK1 MD models and show that we correctly identified kinase domain contacts that govern signaling in a physiologically relevant catalytic complex.

Structural Model of the Catalytic ppERK2–RSK1 Complex. The ppERK2–RSK1 MD model clearly demonstrated that the critical RSK1 activation loop region containing Thr573 and Pro574 could flip into the ERK2 substrate binding pocket without perturbing the compact quaternary arrangement observed in the 150-ns-long MD simulation (Fig. 4). Because this structural model was conducive to a signaling competent complex, we attempted to obtain a structural model for a catalytic ppERK2–RSK1 enzyme–substrate complex. The catalytic aspartate (Asp149) is ~30 Å apart from Thr573 of RSK1 in the ERK2–RSK1 crystal structure (measured between their Cα atoms), but MD simulations indicated that the RSK1 activation loop is a highly flexible region of the kinase domain (*SI Appendix*, Fig. S4*A*). The distance between Thr573 (RSK1) and Asp149 (ERK2) indeed decreased in the course of the 1-μs-long unrestrained MD simulation on ppERK2–RSK1 (*SI Appendix*, Fig. S4*B*). However, the simulation time was presumably not long enough to capture the catalytic complex where the RSK1 TP motif binds into the ERK2 substrate pocket and becomes optimally positioned for phosphotransfer. Similarly to other so-called proline-directed serine–threonine kinases, the substrate binding pocket of MAPKs accepts serine or threonine residues that are followed by a proline (S/TP motifs). Because proline-directed kinases presumably bind their substrates similarly, an optimal distance for the two critical ERK2 and RSK1 residues could be obtained based on a related proline-directed kinase–substrate complex structure (14). The Thr-Pro motif region of the RSK1 activation loop was superimposed with the corresponding residues from the DYRK1A–substrate peptide complex and the loop conformation was minimized (*SI Appendix*, Fig. S4*B*). This approach gave a feasible "restrained" structural model for the catalytic ppERK2–RSK1 complex. Moreover, energy calculations on the starting, the unrestrained, and the restrained models indeed indicated that the catalytic interface increasingly contributed to the total computed interaction energy of the ppERK2–RSK1 complex (*SI Appendix*, Fig. S4*C*).

Role of the MAPKAPK APE Motif in Activator Kinase Binding. The major contact between kinases at IF2 forms between the glycine-rich P loop of ERK2 and the APE motif of RSK1 in the ERK2–RSK1 crystal structure. In addition, MD simulations on ppERK2–RSK1 implicated these generic regions as pivots around which kinase domain contacts get optimized during ppERK2→RSK1 phosphorylation. In kinases the P loop is involved in ATP binding and ADP release, whereas the APE motif plays a pivotal role in protein substrate binding at the P+ side (15). In calcium/calmodulin-dependent protein kinases (CAMKs) an inhibitory helix sterically blocks the binding of substrates by occluding the substrate binding pocket (16). Similarly, the αL helix plays the same role in the related C-terminal kinase domain of MAPKAPKs (Fig. 5*A*). In contrast to other known protein kinases, the APE motif occupies a noncanonical position in all inactive MAPKAPK structures (for example it is part of the extended αF helix in known RSK structures), whereas in their active state—after

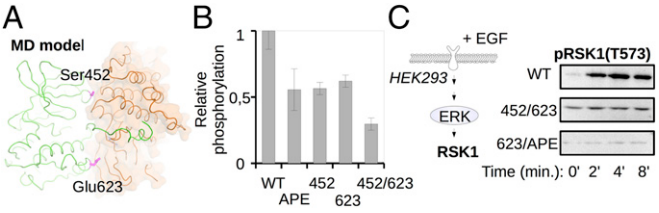

Fig. 3. Experimental validation of surface contacts from the ppERK2–RSK1 MD model. (*A*) MD predicts that Ser452 and Glu623 play a role in the catalytic ppERK2–RSK1 complex. Phosphorylation rate of RSK1 was monitored by using in vitro kinase assays (*SI Appendix*, Fig. S3*B*). (*B*) The bar graph shows relative RSK1 phosphorylation, which was calculated based on initial phosphorylation rates that were normalized to wild type (WT), and error bars show the SDs from mean value; $n = 3$ experiments. (APE: RSK1 mutant with a modified APE motif; 452 or 623: Ser452 or Glu623 were changed to tryptophans; APE/623: two mutated regions are combined within one RSK1 construct.) (*C*) Results of RSK1 activation in EGF-stimulated cells. Blots show a representative set of three independent experiments where samples were taken at the indicated time points after EGF treatment of serum-starved HEK293 cells. Heterologous RSK1 and endogenous ERK activation or equal protein load were analyzed similarly as in Fig. 2 (and control blots for ppERK and RSK1 level are shown on *SI Appendix*, Fig. S3*C*). (For comparison, the blot for WT is the same as in Fig. 2.)

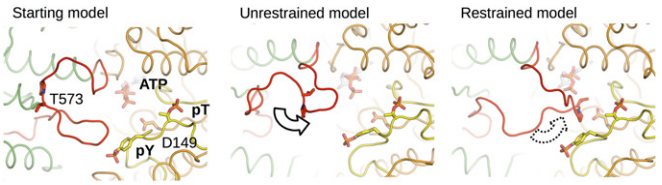

Fig. 4. MD simulations on the ppERK2–RSK1 complex. Movements of the RSK1 activation loop are highlighted (in red). "Starting model" shows a close-up around the catalytic center; the "unrestrained" model shows the same region from the MD model that displayed the shortest distance between Thr573(RSK1) and Asp149(ERK2) (*SI Appendix*, Fig. S4*B*); and the "restrained model" was generated by modeling the RSK1 activation loop based on the DYRK1A–substrate peptide complex (14) (*SI Appendix*, Fig. S4*B*).

MAPK-mediated phosphorylation on their activation loop—this region presumably displaces αL so as to play the same pivotal role in substrate binding as in all kinases (Fig. 5*B*) (17).

The APE motif region of αF may undergo a major conformational change that is triggered by phosphorylation of the MAPKAPK activation loop. These intramolecular rearrangements then create a functional substrate binding pocket on MAPKAPKs (17, 18). For RSK1, and presumably for other MAPKAPKs, the APE region is also involved in activator kinase binding in addition to its canonical role in downstream substrate binding. Thus, uniquely, the MAPKAPK APE region plays a dual role. The ERK2 P loop in the complex also has an unexpected role: In addition to ATP binding, it is involved in substrate kinase binding as it contacts the RSK1 APE motif (Fig. 1*B*).

Determinants of MAPK→MAPKAPK Signaling Specificity. Bimolecular fluorescence complementation (BiFC)-based cellular assays showed that the RSK1 linear motif is absolutely necessary to mediate the binary interaction between ERK2 and its substrate kinase in the cell (*SI Appendix*, Fig. S5). In addition, we formerly showed that linear motif containing peptides from RSK1, MK2, and MNK1 bind to MAPKs with well-defined selectivity profiles that match to MAPK→MAPKAPK biological specificity: RSK1 and MK2 linear motif peptides bound their cognate MAPKs with submicromolar affinities (<0.5 μM), whereas they bound noncognate MAPKs weaker (K_d ~ 10–20 μM) (11). In agreement with the biological role of MNK1, the linear motif containing peptide from this protein bound both to ERK2 and p38α with equal (~0.5 μM) binding affinity (11). Here, we examined the behavior of three MAPKAPK peptides in solution using NMR-based secondary chemical shift (SCS) analysis (*SI Appendix*, Fig. S6). Variation of SCS values for Hα, Cα, and Cβ resonances along the peptide chain can reveal secondary structure propensities in unfolded and partly folded proteins. Although all linear motif peptides adopt a characteristic binding conformation upon binding to MAPKs, they were found to be disordered in solution. This analysis in combination with structure solution of MAPKs in complex with linear motif containing peptides and MAPK–MAPKAPK protein–protein complexes suggests that these MAPKAPK regions undergo disorder-to-order transition upon binding to the MAPK "docking" groove (11, 19). This is also supported by the fact that these regions are disordered in monomeric crystal structures of RSK2 and MK2 (17, 20).

RSK1 and MK2 MAPKAPKs are specifically activated by ERK2 and p38 MAPKs, respectively, whereas MNK1 is activated by both MAPKs in cells. To address how distinct interfaces contribute to MAPK–MAPKAPK signaling, we monitored MAPK→MAPKAPK phosphorylation by in vitro kinase assays using purified proteins (*SI Appendix*, Figs. S7 and S8). Changing the unspecific linear motif in MNK1 into a MAPK specific motif (pepRSK1 or pepMK2) mildly shifted chimera construct phosphorylation toward corresponding MAPKs as expected (*SI Appendix*, Fig. S8 *A* and *B*). Similarly, phosphorylation of RSK1 and MK2 chimera constructs showed agreement to the MAPK binding specificity profile of their linear motif region (*SI Appendix*, Fig. S8 *C* and *D*). However, RSK1 and MK2 were phosphorylated not only by their cognate MAPKs but also by noncognate MAPKs (10). Particularly, ERK2-mediated phosphorylation of MK2 was unexpectedly high, close to half of what was observed on RSK1 (*SI Appendix*, Fig. S8*C*). Mitogen stimulus involving ERK2 leads to RSK1 but not to MK2 activation in the cell (10, 13), although MK2 was first identified as an in vitro ERK2 substrate (21), biochemical specificity of binary MAPK→MAPKAPK pairs is clearly not sufficient to explain physiological specificity. In the cell, however, MAPKs work in the context of other MAPKs. Thus, inactive p38α may efficiently hinder signaling through the ERK2–MK2 noncognate kinase pair indirectly. The mechanism is based on interfering with noncognate recruitment of ERK2 to MK2 because inactive p38α can bind to the MK2 linear motif region with higher affinity compared with activated ERK2. When similar in vitro kinase assays were carried out in the presence of inactive p38α, ERK2-mediated phosphorylation of MAPKAPKs indeed became specific and "leakage" between noncognate pairs was abolished (*SI Appendix*, Fig. S8*E*).

These results suggest that linear motif regions have a pivotal initiator role in complex formation, possibly by tethering the two kinase domains next to each other. Once an activated MAPK is recruited, MAPKAPK activation loop phosphorylation progresses in a nonselective fashion. Correct physiological specificity was achieved only in the presence of noncognate MAPKs when illicit MAPK recruitment was efficiently blocked. Thus, additional surfaces on kinase domains do not greatly influence signaling specificity. Leakage, however, is influenced by MAPKs from other signaling pathways, suggesting that higher level contextual factors also contribute to correct MAPK→MAPKAPK signaling in the cell (*SI Appendix*, Fig. S8*F*).

Discussion

Structural and biochemical characterization of MAPK–MAPKAPK complexes suggest the first mechanistic model on the structural assembly of a signaling competent kinase heterodimer (Fig. 6). This model explains the pivotal role of the linear motif region in MAPKAPKs and it highlights the role of various catalytic and noncatalytic kinase surfaces. The short MAPKAPK linear motif region likely promotes the assembly of an encounter complex in which the kinase domains are randomly oriented. This complex is tethered together through a linear motif mediated interaction

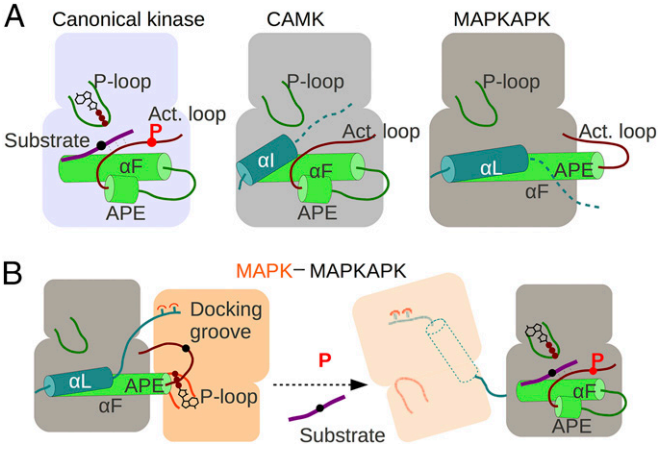

Fig. 5. Role of the MAPKAPK APE motif and the MAPK P loop in substrate or activator kinase binding. (*A*) Inactive, unphosphorylated MAPKAPKs have a unique APE motif region different from related Ca^{2+}/calmodulin-dependent protein kinases (CAMK) or from a canonical kinase. (*B*) Schematic model of a MAPK→MAPKAPK(CTD) signaling complex. The model depicts the dual role of the APE motif in activator kinase binding for unphosphorylated MAPKAPKs as well as in downstream substrate binding after MAPKAPK activation loop phosphorylation.

engaging the MAPK docking groove (IF1). The "encounter complex" provides the possibility of readjustments of kinase domain orientations for maximizing their contacts around IF2 without disassembly ("readjusting complex"). Complete alignment of kinase domains in which the substrate kinase's activation loop binds next to the upstream kinase's catalytic site requires that the MAPK activation loop is double phosphorylated. A precatalytic complex can only transition into a "signaling complex" (modeled by MD in this study) if the MAPK had been activated by upstream kinases formerly. As MAPK activation also involves the MAPK docking groove where MAPK kinases bind to MAPKs, ERK2 activation and RSK1 phosphorylation happens independently in distinct heterodimeric complexes. Acknowledgedly, the crystallographic ERK2–RSK1 complex is not on the pathway to the Michaelis complex, and it likely represents a complex that is unproductive in terms of RSK1 activation as the productive complex has to contain preactivated ERK2. Despite all this, the new ERK2–RSK1 complex was a good starting point for MD to model the ppERK2–ERK2 signaling competent (or Michaelis) complex. Although MAPK cascades are organized by scaffolding proteins that may align and assemble complexes (22), here we demonstrated that interactions between a linear motif and a dedicated docking groove is sufficient to promote the assembly of the catalytic ERK2→RSK1 binary complex. Interestingly, a similar proximity-induced catalytic mechanism was formerly suggested to facilitate efficient phosphorylation of the ERK2 target site in the Ets-1 transcription factor (23, 24).

The MAPK docking groove mediated interface engages the MAPKAPK linear motif. Despite the fact that this interaction does not involve direct contacts relevant for the catalytic enzyme–substrate complex, it is absolutely necessary for MAPK–MAPKAPK complex formation and signaling. Its pivotal role may be explained by at least two independent mechanisms: (*i*) Kinetically, a disordered interacting region may greatly increase the chance of forming an energetically favorable encounter complex because the interaction forms through induced fit. Thus, complex forming collisions require far less precise orientation of the interacting molecules at the first encounter (25, 26). The disorder-to-order transition at the MAPKAPK linear motif likely has only a small entropic cost so as to form a high-affinity encounter complex with its compact MAPK partner. (*ii*) Tethering of kinase domains via a spatially distinct interface may allow readjustments between kinase domains without dissociation of the first MAPK–MAPKAPK encounter complex. Fine tuning involves generic kinase domain regions that will then lead to a signaling competent complex; however, these generic contacts are insufficient without additional specific contacts to drive the formation of a signaling binary complex (*SI Appendix*, Fig. S5). This presumes that kinase–kinase domain orientations can vary in readjusting MAPK–MAPKAPK complexes. This is indeed supported by unrestrained MD simulations on the ERK2–RSK1 complex. In addition, a p38α–MK2 crystallographic model captured an unrelated, noncatalytic quaternary arrangement compared with what is described in this study (*SI Appendix*, Fig. S9) (19, 27). Direct comparison of available MAPK–MAPKAPK crystal structures shows that the ERK2–RSK1 complex is the first structure to our knowledge in which the activation loop of the MAPKAPK is in the vicinity of the MAPK active site. The p38α–MK2 structure captures an unproductive heterodimer as critical enzyme and substrate regions cannot meet as captured in this crystallographic complex. Moreover, this complex showed that the role of the catalytically incompetent complex is the stabilization of unphosphorylated p38. These are in contrast to the ERK2–RSK1 crystallographic complex.

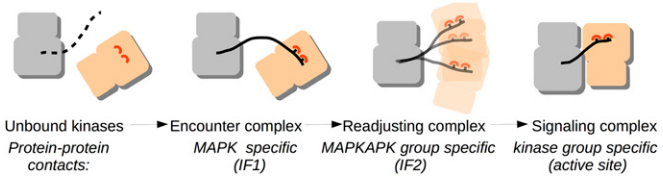

Fig. 6. Model on the structural assembly of the signaling competent MAPK–MAPKAPK complex. The signaling competent ppERK2–RSK1 complex forms through hierarchical assembly of unique, group specific, and common kinase surface contacts. (This scheme is partly speculative but it is consistent with biochemical and structural data presented in this study.)

Interaction at IF1 are highly MAPK specific, as linear motifs have their own characteristic MAPK binding specificity. Formation of contacts at IF2 are likely to be MAPKAPK group specific, as their APE motif is distinct compared with other known kinases, and it has a unique role in upstream kinase binding. MAPKAPK contacts at IF2, however, are less specific within family members as noncognate MAPK–MAPKAPK pairs formed productive complexes if their kinase domains were tethered close artificially (*SI Appendix*, Fig. S8). Overall, contacts through IF1 and IF2 collectively hold the kinase heterodimer in a precatalytic state and the topography of the active site is presumably similar in all proline-directed kinases as these phosphorylate similar target motifs.

Most of our knowledge on kinase dimerization and activation is based on crystal structures of symmetrical homodimers (28). Based on these structures, activation segment exchange for example was suggested to be a common mechanism of kinase autophosphorylation for a subset of protein kinases (29). Dimeric contacts usually form between αG helices, which are normally involved in canonical substrate binding (30–32). Interestingly, this dimerization mode is also observed in known head-to-head heterodimeric complex structures (33, 34). However, the ERK2–RSK1 complex structure revealed an unusual head-to-tail kinase dimerization mode (28). It also revealed alternative functions of well-characterized generic kinase regions, in particular for the P loop and for the APE motif. In addition to the canonical role of the APE motif in substrate binding at the P+ side, its involvement in other protein–protein interactions—such as in upstream activator kinase binding—is unique to MAPKAPKs. This is due to the special position and/or the flexible nature of this motif in inactive MAPKAPKs compared with canonical kinases. Previous structural studies showed that the P loop can directly participate in kinase dimerization. Examples include unrelated, catalytically competent homodimeric structures of the prokaryotic kinase PknB (from *Mycobacterium tuberculosis*) and human checkpoint kinase 2 (CHK2) or the heterodimeric structure of RIP3 and MLKL (31, 34, 35) (*SI Appendix*, Fig. S10). Because the P loop is directly involved in ATP binding by coordinating the β- and γ-phosphate groups for optimal phosphotransfer in all kinases, it may also be that this glycine-rich loop is an ancient allosteric hotspot in precatalytic kinase–kinase complexes.

In conclusion, MAPK→MAPKAPK signaling provides a great example of how generic kinase domain regions could combine with more divergent surface regions in a hierarchical assembly process. This synergism could be particularly important to achieve functional diversity within kinase cascades using similarly built and evolutionarily related enzymatic components. Interestingly, MAPK activation by MAPK kinases (MAPKK), which are also all evolutionarily related, depends on their linear binding motif regions as well (36). Thus, regarding the nature of interactions leading to the formation of a signaling competent kinase–kinase complex, MAPKK→MAPK and MAPK→MAPKAPK activation may be mechanistically alike.

Methods

Protein Production for Structural Studies. The cDNA of full-length human ERK2 and the RSK1 [C-terminal kinase domain (CTD)-LM] construct containing the C-terminal RSK1 region between residues 411 and 735 were cloned into modified pET expression vectors. All protein constructs were expressed in *Escherichia coli* Rosetta (DE3) pLysS (Novagen) cells with standard techniques. Dephosphorylated ERK2 with N-terminal cleavable hexahistidine tag was coexpressed with GST-tagged λ-phage phosphatase. RSK1(CTD-LM) was expressed as N-terminal GST fusion protein with a C-terminal noncleavable hexahistidine tag. Affinity-purified ERK2 was cleaved by the tobacco etch virus (TEV) protease

and samples were further purified by ion exchange on a Resource Q column (GE Healthcare). Double affinity purified RSK1 was also cleaved by the TEV protease and the sample was further purified on a HiTrap Blue-Sepharose column (GE Healthcare). Purified kinases were mixed in 1:1.2 ratio with ERK2 in excess and the sample was gel filtrated on a Superdex 75 column (GE Healthcare). Fractions corresponding to the stoichiometric complex were pooled and the sample was concentrated to 10 mg/mL.

Crystallization, Structure Solution, and Refinement. The stock solution of the final protein sample was supplemented with 2 mM adenosine 5′-(β,γ-imido)triphosphate (AMPPNP) and 2 mM $MgCl_2$. Crystallization was done in standard sitting drop vapor-diffusion set-up at 23 °C. The crystallization solution consisted of 0.1 M Mes pH = 6.25, 15% (vol/vol) PEG4000, 0.125 M $(NH_4)_2SO_4$ and 2% (wt/vol) benzamidine. Drops with plate-shaped crystals (with an average size of 0.15 mm × 0.15 mm × 0.02 mm) were supplemented with 10% (vol/vol) glycerol before flash cooling in liquid nitrogen. Data were collected on the PXIII beam line of the Swiss Light Source (Villigen) at 100 K. Details on data collection, analysis, and structure determination are given in *SI Appendix, Methods*.

MD Simulations. Starting MD models for ERK2–RSK1 and ppERK2–RSK1 were generated from the ERK2–RSK1 crystallographic complex. The crystal structure of double-phosphorylated ERK2 (ppERK2; PDB ID: 2ERK) was superimposed with ERK2 from the unphosphorylated ERK2–RSK1 crystal structure (giving the starting model for ppERK2–RSK1). The phosphate groups on Thr185 and Tyr187 were removed from the ppERK2–RSK1 model (giving the starting model for the unphosphorylated ERK2–RSK1 complex). In all calculations, the GROMACS ver. 4.5.5 program package (37), the Amber-03 (38) force field was applied along with neutralizing Na^+ counter ions and numerous TIP3P (39) explicit water molecules filling a 5-Å spacing between the protein parts and the edges of the cubic simulation box. The lengths of the unrestrained MD runs were 1 μs (ppERK2–RSK1) or 150 ns (ERK2–RSK1). Further details on MD simulation parameters and processing of MD results are given in *SI Appendix, Methods*.

ERK2→RSK1 Activation Assays. For in vitro assays, recombinant-expressed and purified proteins were used, and RSK1 phosphorylation was monitored by P32 autoradiography or by phospho-Thr573(RSK1) Western blots. Further details on in vitro kinase assays are given in *SI Appendix, Methods*. For cell-based assays, RSK1 constructs were subcloned into modified pcDNA 3.1 vectors with N-terminal Venus fluorescent protein and C-terminal FLAG fusion tags (Invitrogen). HEK293T cells were cultured in 96-well plates as described in detail in *SI Appendix, Methods*. Cells were transfected with 0.4 μg RSK1 DNA constructs and were serum starved for 24 h. The media was removed after 40 h from DNA transfection and 100 μL PBS was added to wells. ERK pathway stimulation was started by addition of EGF (Sigma, E9644) in 100 ng/mL concentration to each well and stimulation was terminated at different time points by adding 35 μL of 4× SDS loading buffer to wells. Cells were lysed and 10 μL of each sample was subjected to SDS/PAGE. Western blots for monitoring RSK1 phosphorylation on Thr573 were done using the phospho-p90RSK (Thr573) primary antibody (Cell Signaling, 9346). The phospho-p44/42 MAPK (Erk1/2) (Thr202/Tyr204) antibody (Cell Signaling, 9101) and an anti-FLAG antibody (Sigma, F1804) were used to check endogenous ppERK2 and heterologous RSK1 protein levels, respectively.

ACKNOWLEDGMENTS. We thank József Kardos and Éva Bulyáki for their help in CD measurements, András Patthy for excellent peptide synthesis, PRACE for awarding us access to resources Monte Rosa based in Switzerland at CSCS Swiss National Supercomputing Centre, and NIIFI SC based in Hungary at NIIF National Information Infrastructure Development Institute. A.R. is supported by the "Lendület" grants from the Hungarian Academy of Sciences (LP2013-57) and by an International Senior Research fellowship from the Wellcome Trust. The work was also supported by OTKA NK101072 (to A.B.) and a MedinProt grant (to A.R. and C.H.).

1. Taylor SS, Kornev AP (2011) Protein kinases: Evolution of dynamic regulatory proteins. *Trends Biochem Sci* 36(2):65–77.
2. Endicott JA, Noble MEM, Johnson LN (2012) The structural basis for control of eukaryotic protein kinases. *Annu Rev Biochem* 81:587–613.
3. Johnson GL, Lapadat R (2002) Mitogen-activated protein kinase pathways mediated by ERK, JNK, and p38 protein kinases. *Science* 298(5600):1911–1912.
4. Raman M, Chen W, Cobb MH (2007) Differential regulation and properties of MAPKs. *Oncogene* 26(22):3100–3112.
5. Canagarajah BJ, Khokhlatchev A, Cobb MH, Goldsmith EJ (1997) Activation mechanism of the MAP kinase ERK2 by dual phosphorylation. *Cell* 90(5):859–869.
6. Pearce LR, Komander D, Alessi DR (2010) The nuts and bolts of AGC protein kinases. *Nat Rev Mol Cell Biol* 11(1):9–22.
7. Cargnello M, Roux PP (2011) Activation and function of the MAPKs and their substrates, the MAPK-activated protein kinases. *Microbiol Mol Biol Rev* 75(1):50–83.
8. Gavin AC, Nebreda AR (1999) A MAP kinase docking site is required for phosphorylation and activation of p90(rsk)/MAPKAP kinase-1. *Curr Biol* 9(5):281–284.
9. Tanoue T, Maeda R, Adachi M, Nishida E (2001) Identification of a docking groove on ERK and p38 MAP kinases that regulates the specificity of docking interactions. *EMBO J* 20(3):466–479.
10. Smith JA, et al. (2000) Creation of a stress-activated p90 ribosomal S6 kinase. The carboxyl-terminal tail of the MAPK-activated protein kinases dictates the signal transduction pathway in which they function. *J Biol Chem* 275(41):31588–31593.
11. Garai Á, et al. (2012) Specificity of linear motifs that bind to a common mitogen-activated protein kinase docking groove. *Sci Signal* 5(245):ra74.
12. Dalby KN, Morrice N, Caudwell FB, Avruch J, Cohen P (1998) Identification of regulatory phosphorylation sites in mitogen-activated protein kinase (MAPK)-activated protein kinase-1a/p90rsk that are inducible by MAPK. *J Biol Chem* 273(3):1496–1505.
13. Ben-Levy R, et al. (1995) Identification of novel phosphorylation sites required for activation of MAPKAP kinase-2. *EMBO J* 14(23):5920–5930.
14. Soundararajan M, et al. (2013) Structures of Down syndrome kinases, DYRKs, reveal mechanisms of kinase activation and substrate recognition. *Structure* 21(6):986–996.
15. Kornev AP, Taylor SS, Ten Eyck LF (2008) A helix scaffold for the assembly of active protein kinases. *Proc Natl Acad Sci USA* 105(38):14377–14382.
16. Rellos P, et al. (2010) Structure of the CaMKIIdelta/calmodulin complex reveals the molecular mechanism of CaMKII kinase activation. *PLoS Biol* 8(7):e1000426.
17. Malakhova M, et al. (2008) Structural basis for activation of the autoinhibitory C-terminal kinase domain of p90 RSK2. *Nat Struct Mol Biol* 15(1):112–113.
18. Underwood KW, et al. (2003) Catalytically active MAP KAP kinase 2 structures in complex with staurosporine and ADP reveal differences with the autoinhibited enzyme. *Structure* 11(6):627–636.
19. ter Haar E, Prabhakar P, Liu X, Lepre C (2007) Crystal structure of the p38 alpha-MAPKAP kinase 2 heterodimer. *J Biol Chem* 282(13):9733–9739.
20. Meng W, et al. (2002) Structure of mitogen-activated protein kinase-activated protein (MAPKAP) kinase 2 suggests a bifunctional switch that couples kinase activation with nuclear export. *J Biol Chem* 277(40):37401–37405.
21. Stokoe D, et al. (1992) MAPKAP kinase-2; a novel protein kinase activated by mitogen-activated protein kinase. *EMBO J* 11(11):3985–3994.
22. Good MC, Zalatan JG, Lim WA (2011) Scaffold proteins: Hubs for controlling the flow of cellular information. *Science* 332(6030):680–686.
23. Rainey MA, Callaway K, Barnes R, Wilson B, Dalby KN (2005) Proximity-induced catalysis by the protein kinase ERK2. *J Am Chem Soc* 127(30):10494–10495.
24. Lee S, et al. (2011) A model of a MAPK•substrate complex in an active conformation: A computational and experimental approach. *PLoS ONE* 6(4):e18594.
25. Pontius BW (1993) Close encounters: why unstructured, polymeric domains can increase rates of specific macromolecular association. *Trends Biochem Sci* 18(5):181–186.
26. Dunker AK, et al. (2001) Intrinsically disordered protein. *J Mol Graph Model* 19(1):26–59.
27. White A, Pargellis CA, Studts JM, Werneburg BG, Farmer BT, 2nd (2007) Molecular basis of MAPK-activated protein kinase 2:p38 assembly. *Proc Natl Acad Sci USA* 104(15):6353–6358.
28. Malecka KA, Peterson JR (2011) Face-to-face, pak-to-pak. *Structure* 19(12):1723–1724.
29. Oliver AW, Knapp S, Pearl LH (2007) Activation segment exchange: A common mechanism of kinase autophosphorylation? *Trends Biochem Sci* 32(8):351–356.
30. Wang J, Wu JW, Wang ZX (2011) Structural insights into the autoactivation mechanism of p21-activated protein kinase. *Structure* 19(12):1752–1761.
31. Mieczkowski C, Iavarone AT, Alber T (2008) Auto-activation mechanism of the Mycobacterium tuberculosis PknB receptor Ser/Thr kinase. *EMBO J* 27(23):3186–3197.
32. Nayak V, et al. (2006) Structure and dimerization of the kinase domain from yeast Snf1, a member of the Snf1/AMPK protein family. *Structure* 14(3):477–485.
33. Brennan DF, et al. (2011) A Raf-induced allosteric transition of KSR stimulates phosphorylation of MEK. *Nature* 472(7343):366–369.
34. Xie T, et al. (2013) Structural insights into RIP3-mediated necroptotic signaling. *Cell Reports* 5(1):70–78.
35. Cai Z, Chehab NH, Pavletich NP (2009) Structure and activation mechanism of the CHK2 DNA damage checkpoint kinase. *Mol Cell* 35(6):818–829.
36. Bardwell AJ, Frankson E, Bardwell L (2009) Selectivity of docking sites in MAPK kinases. *J Biol Chem* 284(19):13165–13173.
37. Pronk S, et al. (2013) GROMACS 4.5: A high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* 29(7):845–854.
38. Duan Y, et al. (2003) A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J Comput Chem* 24(16):1999–2012.
39. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) Comparison of simple potential functions for simulating liquid water. *J Chem Phys* 79:926.

Bioinformatics, 31(12), 2015, 1959–1965
doi: 10.1093/bioinformatics/btv093
Advance Access Publication Date: 13 February 2015
Original Paper

OXFORD

Structural bioinformatics

Mobility-based prediction of hydration structures of protein surfaces

Norbert Jeszenői[1], István Horváth[2], Mónika Bálint[3], David van der Spoel[4] and Csaba Hetényi[5,*]

[1]Department of Genetics, Eötvös Loránd University, Pázmány Péter sétány 1/C, 1117 Budapest, [2]Chemistry Doctoral School, University of Szeged, Dugonics tér 13, 6720 Szeged, [3]Department of Biochemistry, Eötvös Loránd University, Pázmány Péter sétány 1/C, 1117 Budapest, Hungary, [4]Uppsala Center for Computational Chemistry, Science for Life Laboratory, Department of Cell and Molecular Biology, University of Uppsala, Box 596, SE-75124 Uppsala, Sweden, and [5]MTA-ELTE Molecular Biophysics Research Group, Hungarian Academy of Sciences, Pázmány sétány 1/C, 1117 Budapest, Hungary

*To whom correspondence should be addressed.
Associate Editor: Anna Tramontano

Received on November 4, 2014; revised on January 9, 2015; accepted on February 10, 2015

Abstract

Motivation: Hydration largely determines solubility, aggregation of proteins and influences interactions between proteins and drug molecules. Despite the importance of hydration, structural determination of hydration structure of protein surfaces is still challenging from both experimental and theoretical viewpoints. The precision of experimental measurements is often affected by fluctuations and mobility of water molecules resulting in uncertain assignment of water positions.

Results: Our method can utilize mobility as an information source for the prediction of hydration structure. The necessary information can be produced by molecular dynamics simulations accounting for all atomic interactions including water–water contacts. The predictions were validated and tested by comparison to more than 1500 crystallographic water positions in 20 hydrated protein molecules including enzymes of biomedical importance such as cyclin-dependent kinase 2. The agreement with experimental water positions was larger than 80% on average. The predictions can be particularly useful in situations where no or limited experimental knowledge is available on hydration structures of molecular surfaces.

Availability and implementation: The method is implemented in a standalone C program MobyWat released under the GNU General Public License, freely accessible with full documentation at http://www.mobywat.com.

Contact: csabahete@yahoo.com

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

Water molecules located on protein surfaces play fundamental structural and functional roles in biology. For example, hydrogen bonds formed by waters stabilize protein structure (Nisius and Grzesiek, 2012) and affect folding (Cheung *et al.*, 2002; Levy and Onuchic, 2006). Surface water molecules are mediators of the assembly of β-amyloid protofilaments of Alzheimer's disease (Thirumalai *et al.*, 2011) and there is evidence that structurally conserved waters are parts of electron transfer networks (Antonyuk *et al.*, 2013) such as respiratory chain (de la Lande *et al.*, 2010). Structures of many G-protein-coupled receptors are also stabilized by hydration (Angel *et al.*, 2009). A recent study (Xu and Leitner, 2014) suggests that structural water molecules are also involved in thermal

conductance of proteins, in photochemistry, as well as playing a fundamental role in charge transfer, allostery and energy flow (Fang *et al.*, 2009).

Water molecules are often considered essential parts of the protein structure (Petsko and Ringe, 2009) and the first hydration shell is a key determinant of the solubility and aggregation of solute molecules (Israelachvili and Wennerström, 1996). Protein–protein and protein–ligand interactions are influenced by surface-bound water molecules, and therefore, knowledge of their location is of great importance during structure-based drug design (Baron *et al.*, 2012; García-Sosa, 2013). Tightly bound water molecules can affect the chemical diversity of designed ligands (García-Sosa and Mancera, 2006) leading to simple rules for the use of water molecules in drug design (García-Sosa *et al.*, 2005) and also in interpretation of ligand-based pharmacophore models (Lloyd *et al.*, 2004). Inclusion of explicit water molecules in drug design (Mancera, 2007) have been thoroughly studied and was found to be of central importance in ligand–protein docking (Roberts and Mancera, 2008; Thilagavathi and Mancera, 2010).

Although hydration structure is important, it has hitherto proven to be very difficult to determine at the atomic level by experimental means largely due to mobility and complexity of interactions of water molecules located on a protein surface. The residence of a water molecule on the surface and its exchange with bulk are affected not primarily by the strength of protein–water interactions, but it is 'rather a topography that prevents the water molecule from exchanging by a cooperative mechanism' (Halle, 2004a). Importantly, such a cooperative mechanism of exchange also governs several water–water interactions that can often be detected between surface water molecules (Finney, 1977). It is problematic to handle (and to predict) the residence of water molecules in the hydration layer of a protein using merely thermodynamic or kinetic approaches (Halle, 2004a).

Crystallography is the prime experimental method for detection of water positions, via electron density maps and used as the *de facto* standard (Savage and Wlodawer, 1986). However, there are still numerous limitations of this method coming from low resolution of large structural assemblies (Finney, 1977), assignment problems (Afonine *et al.*, 2013; Badger *et al.*, 1997), and artifacts due to cryogenic temperatures used (Halle, 2004b).

A number of computational methods have been proposed for prediction of hydration structure on protein surfaces. Such methods generally require the 'dry' protein structure as an input and provide predictions for hydration structure using a variety of algorithms. A large group of the methods uses fast and simplified approaches disregarding exchange (mobility) between surface and bulk water molecules and dynamics of the hydration structure. They assume a static picture of hydration shells and focus on finding appropriate binding sites of water molecules on the protein surface using scoring schemes, energy calculations (Schymkowitz *et al.*, 2005), prior knowledge (Pitt and Goodfellow, 1993), H-bonding information (Vedani and Huhta, 1991) or artificial neural networks (Ehrlich *et al.*, 1998). Several studies (Makarov *et al.*, 1998; Truchon *et al.*, 2014; Virtanen *et al.*, 2010) have dealt with construction and use of density distribution functions of hydration shells for different atom types occurring in proteins. Limitations of generalized, density-based approaches were discussed in detail (Henchman and McCammon, 2002). These methods ignore dynamics and cooperativity governing hydration.

With advancement of computational infrastructure and force fields, the efficiency and chemical accuracy of atomic level Monte-Carlo and molecular dynamics (MD) simulations has increased enormously in the past decades (Michel *et al.*, 2009; Pettitt and Karplus, 1987) enabling their applications in cutting edge drug design projects (Dror *et al.*, 2012). It has become a routine task to generate MD trajectories with explicit water molecules for virtually any protein of interest. Atomistic simulations of MD hold a conceptual advantage over the static or density-based (trained) methods as the mobility, a key determinant of hydration structure is described directly at atomic level. Whereas such benefits of atomic MD calculations have been extensively used in analyses (Schoenborn *et al.*, 1995), there are not many MD-based methods for prediction of the hydration structure (Abel *et al.*, 2008; Cui *et al.*, 2013; Henchman and McCammon, 2002). These approaches focus on all individual positions of hydrating water molecules and apply various evaluation schemes such as the definition of time averaged positions (Henchman and McCammon, 2002) for calculation of the hydration structure. In this study, we introduce a mobility-based atomic-level method for prediction of hydration structure of molecular surfaces using only 'dry' protein structures as input. Our method was tested on 20 proteins, and the corresponding computational procedures are provided in a program MobyWat, which can be used in conjunction with any MD software that can produce all-atom MD trajectories.

2 Algorithm

2.1 Prediction

Logging molecular movements of all water molecules during a time period provides mobility information required by the prediction process used here. Such a log-book (a trajectory) is preferably generated by MD calculations with an explicit water model. Generation of molecular trajectories was performed by the GROMACS (Hess *et al.*, 2008; Pronk *et al.*, 2013) MD package in this study. During additional post-MD and preparatory steps a standard protocol was followed (Supplementary Methods S1.2).

Mobility information of the trajectory is transformed into the hydration structure of the protein surface during the prediction process outlined in Figure 1. All predictions can be performed with the program MobyWat designed and written in C implementing the prediction protocols of this study. Detailed descriptions of the algorithms can be found in Supplementary Algorithm S2.1 and also in the User's Manual of the program.

Briefly, during the prediction procedure, MobyWat performs clustering of water molecules in candidate pools filtered from the corresponding MD frames. Besides the usual spatial position-based (POS) clustering, an identity (ID)-based algorithm was also introduced with ranking variants named all-inclusive (IDa) and elitist (IDe, Supplementary Algorithm S2.1.5). The procedure ends up in prediction lists including the coordinates and mobility values of water molecules in Protein Databank (PDB) format. A merged (MER) prediction list can be also produced combining the results of the above IDa, IDe and POS predictions.

2.2 Validation

The identification of matches between experimental and predicted water positions is used for validating algorithms of MobyWat. From the matches, a success rate (SR_X) value is calculated for a prediction list (X = IDa, IDe, POS or MER, Eq. 1). The higher the SR_X value, the more successful a prediction is in comparison with crystallographic water positions. For comparison and estimation of the effect of clustering, per frame SR_n values

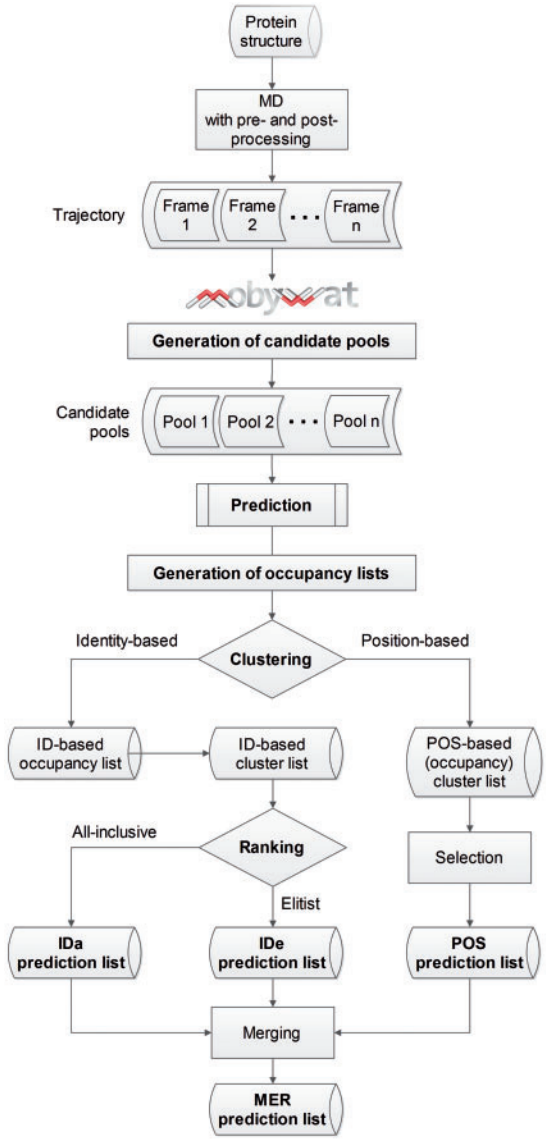

Fig. 1. The prediction process

are also calculated for each candidate pool using the analysis mode of MobyWat ($X = n$, Eq. 1).

$$\mathrm{SR}_X = 100\frac{\text{Number of matches in } X}{\text{Number of water molecules in the reference pool}}\%, \quad \text{where } X = \begin{cases} \text{IDa/IDe/POS/MER (prediction list in validation)}, \\ n \text{ (denotes the nth candidate pool in analysis)}. \end{cases} \tag{1}$$

Further details on validation including selection and calibration of tolerance values are described in Supplementary Algorithm S2.2, and Figure S1. Twenty reference protein systems used for validation and external tests are listed in Tables S1 and S8.

3 Results and Discussion

3.1 Sampling versus predictions

MobyWat predictions are based on atomic mobility data of all water molecules obtained from MD simulations. In this study, mobility of a predicted water molecule is defined by its occupancy value (Supplementary Eq. S2). Occupancy can be counted using a collection (sample) of hydrated protein structures. Such a sample can be collected as a series of hydrated experimental structures of the same protein (Carugo, 1999; Patel *et al.*, 2014), or generated by computational methods. Sample collection from experimental structures is not an option for this purpose as the number of hydrated structures is limited to available entries available in the PDB. In addition, if there are hydrated PDB structures available, then comparative analysis can be performed by other tools (García-Sosa *et al.*, 2003; Patel *et al.*, 2014) which proved to be useful for selection of consensus or conserved water molecules.

However, in most of the cases, only a single structure of the same protein is available. Thus, computational generation of hydration states of a protein is presently the only tractable approach to produce an appropriate sample even if only a 'dry' protein surface is available lacking experimentally determined positions of water molecules. Among computational techniques atomic level MD simulation with an explicit water model is the obvious choice of sampling method. The user needs to supply only a 'dry' protein structure and a series of hydrated protein structures are resulted as an MD trajectory. MD-generated, raw hydration structures are sometimes used even as references in comparison with other methods (Ross *et al.*, 2012). However, important parameters such as the minimal length of an MD simulation necessary for a predictive sampling have not been determined. To address this question, 1-μs-long MD simulations were performed for the protein systems of the validation set producing a sample of 1000 frames spaced at 1 ns. SR_n values were calculated for each pool according to Eq. 1 and plotted in Figure 2A for Alzheimer's amyloid precursor protein (system 2FMA). Descriptive statistics of SR_n values are provided for all validation systems in Supplementary Table S4. The descriptive statistics show a good performance of raw MD sampling with mean SR_n values ranging between 44.6 and 72.7. The SR_n values fluctuate randomly during the 1 μs time-scale of the trajectory (Fig. 2A). This finding can be explained by the short residence time of water molecules in the hydration shell of protein surface (Halle, 2004a). During 1 μs water molecules can change their positions many times, and occurrence of frames with large SR_n values (with a lot of matching water positions) is unpredictable and non-deterministic.

In summary, MD provides an appropriate sampling with good SR_n values. However, the performance of a 'prediction' based on a single frame (randomly) picked from a trajectory is non-deterministic. Thus, a valid prediction cannot be guaranteed if using only one frame. Processing several frames of a trajectory may be a better way to maximize SR and arrive at valid predictions. Accordingly, validation, calibration and measurement of the performance of prediction algorithms are described in the forthcoming sections.

3.2 Validation, performance and robustness

The prediction parameters dmax, ctol and ptol (Supplementary Table S3) were calibrated for all four types of prediction algorithms implemented in MobyWat. The calibration process is documented in Supplementary Results S3.2. Optimal sampling conditions were also determined, as the final step of the validation process. Using calibrated values of parameters, MobyWat predictions were performed for all proteins by processing 1000 coordinate frames from 1-μs-long trajectories. The results are shown for system 2FMA (Fig. 2A), and for all systems of the Validation set (Supplementary Table S4). The SR values yielded by the predictions were significantly higher than the mean SR_n from raw MD, and in many cases they were close to the maximal SR_n values. Thus, all four algorithms

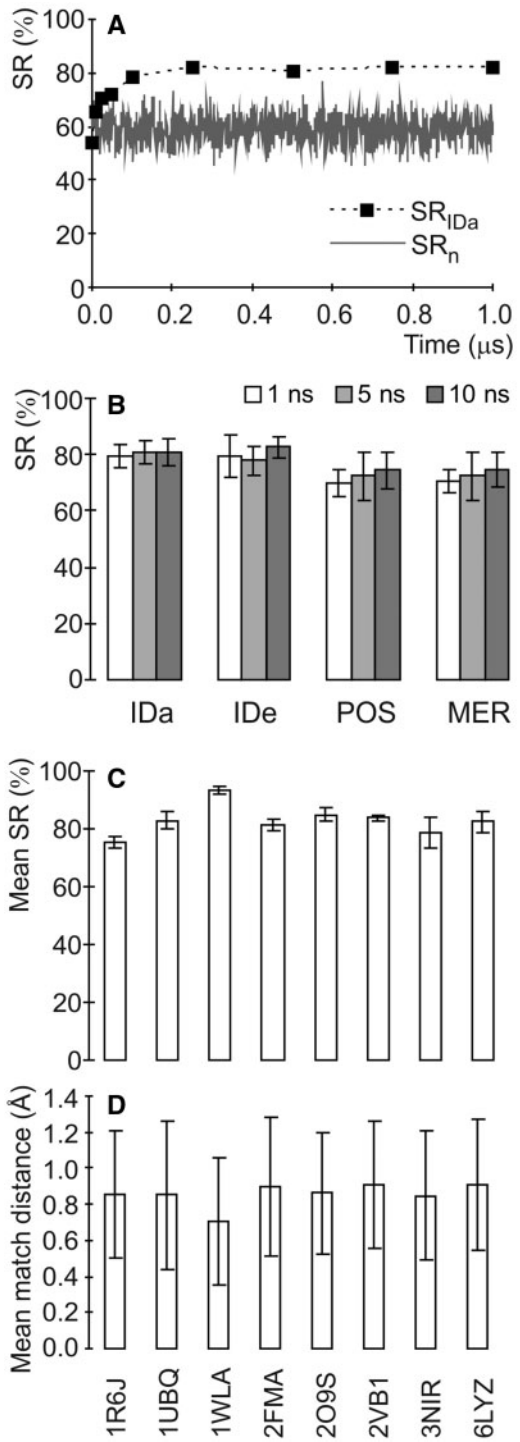

Fig. 2. (**A**) Success rates of Alzheimer's amyloid precursor protein (system 2FMA) calculated for the pools of the raw MD trajectory frames (SR_n) and resulting from IDa prediction of MobyWat (SR_{IDa}). MD trajectory of 1 μs with 1000 frames was used as a sample. (**B**) Effect of sampling time on the performance of prediction algorithms. Ten thousand frames were used for prediction with sampling times 1, 5 and 10 ns. Mean values are calculated from SRs obtained for the Validation set. Standard deviations are shown as error bars. (**C**) Reproducibility of MD sampling in terms of mean SR values calculated from three independent MD runs for each protein system. (**D**) Mean distances in matched pairs of predicted and reference water oxygen atoms plotted for all systems. Error bars denote standard deviations

Table 1. Success rates (%): statistics calculated for raw MD sampling and prediction results achieved by MobyWat

PDB ID[a]	Raw MD[b] (SR_n in Eq. 1)			MobyWat[b,c]			
	Min.	Mean	Max.	SR_{IDa}	SR_{IDe}	SR_{POS}	SR_{MER}
Validation set							
1R6J	41.4	52.4	64.1	71.8	76.2	64.6	65.8
2FMA	39.4	61.5	80.3	80.3	83.6	77.1	77.1
2O9S	46.2	62.2	77.9	87.5	85.6	78.9	78.9
2VB1	44.9	59.9	71.7	82.6	84.1	79.7	80.4
3NIR	33.9	59.0	80.4	80.4	83.9	71.4	71.4
Mean	41.2	59.0	74.9	80.5	82.7	74.3	74.7
SD^V	4.9	3.9	7.0	5.7	3.7	6.3	6.1
Test set 1							
1UBQ	28.6	53.9	82.4	85.7	80.0	68.6	74.3
1WLA	31.4	68.5	94.3	94.3	88.6	82.9	82.9
6LYZ	32.2	54.2	72.9	78.0	81.5	71.2	71.2
Mean	30.7	58.8	83.2	86.0	83.4	74.2	76.1
SD^E	1.9	8.3	10.7	8.2	4.6	7.6	6.0

[a]Mean and standard deviation (SD) values of success rates were calculated for systems of external test and validation separately. [b]Sampling conditions: 10 ns MD run time, 1.0001×10^4 frames. [c]Success rates of MobyWat predictions were calculated with default mtol = 1.5 Å, $b_{max} = 30.0$ Å^2, $d_{max} = 3.5$ Å, $p_{tol} = 2.5$ Å and c_{tol} according to Supplementary Table S5.

resulted in valid predictions. Whereas sampling of 1-μs-long trajectories provided good predictions, such simulations with explicit waters can be computationally demanding. Figure 2A shows that SR_{IDa} values exceeded the SR_n curve and reached a plateau relatively early, after 100–200 ns sampling time. This finding suggested that shortening the sampling time should be possible without a large drop in SR of the prediction. Increasing the sampling frequency (frame count) is also a logical step to achieve reliable predictions with shortened sampling time. Indeed, results in Table 1 reveal that 10-ns-long trajectories with increased frame count yielded mean SR values of >80% for the Validation set, similarly to the 1-μs-long runs (Supplementary Table S4).

Figure 2B shows that the good performance of ID-based prediction algorithms was preserved at 1, 5 and 10 ns sampling times averaged for all systems used in Validation set. In the cases of MER and POS, there is a 5% increase in average SR values if comparing trajectories of 1 and 10 ns length. In summary, the ID-based algorithms outperformed POS and MER predictions, and they provide good predictions even at 1 ns sampling time (Table 1, Fig. 2B).

To evaluate system-independence of our method, a test of the predictions was performed. Systems of Test set 1 (1UBQ, 1WLA and 6LYZ) have relatively moderate resolution and a low number of assigned water positions per protein surface area (Supplementary Table S1). The same set had been used earlier in a study (Virtanen *et al.*, 2010) applying a solvent density-based approach. Detailed comparison of our results using the standards of the earlier study (Supplementary Results S3.3) indicates that overall performance of MobyWat is good if compared with solvent density-based results. For comparability with the above validation results performance of MobyWat on Test set 1 was also evaluated using the standards of this study and the results are listed separately in Table 1. All four algorithms provide valid predictions with SR significantly higher than average values of SR_n. Moreover, the mean SR values of Test set 1 are comparable to or slightly higher than mean SR values obtained for Validation set (Table 1) indicating system-independence of the method.

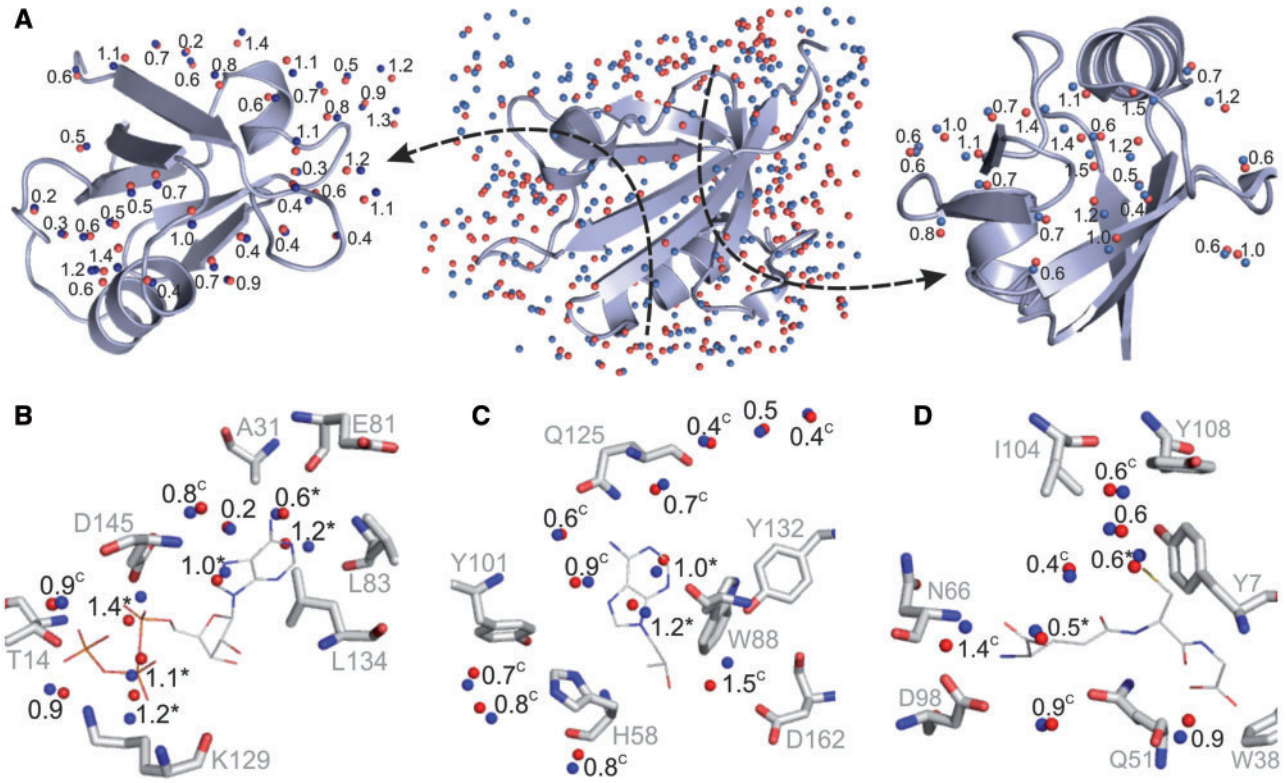

Fig. 3. (**A**) Prediction results for system 1R6J. (**B–D**) Featured binding sites of apo enzymes cyclin-dependent kinase 2 (system 1HCL, B), thymidine kinase (system 1E2H, C) and glutathione S-transferase (system 16GS, D). Ligands were inserted from superimposed ligand-bound enzyme structures (PDB codes 1HCK, 1E2I and 5GSS) for comparison with water positions. Match distances between crystallographic (red spheres) and predicted (blue spheres) water oxygen atoms are given in Å. Conserved and replaceable water molecules are marked with C and asterisk at the distance values, respectively

Reproducibility is also a key issue of robustness. As MobyWat operations are reproducible by their algorithmic definition, reproducibility tests can be performed for the MD sampling process. MD trajectories are inherently chaotic in practical applications due to hardware-dependent rounding of floating point calculations, the use of dynamic load balancing in parallel execution and so on. Therefore, it is common to repeat MD calculations with different starting atomic velocity values to test the convergence of trajectories. Practically, this can be done by selecting different seed numbers of the velocity generator routine. During the tests, three MD trajectories of all systems were produced using three different sets of initial velocities. For these trajectories, predictions were made using the top performer algorithms of Table 1.

The corresponding three SR values were averaged for all systems and plotted in Figure 2C. Their standard deviations are found to be small compared with mean values for all systems, and MD sampling is therefore shown to be reproducible in terms of SR. Improvements in the quality of force fields, in particular the introduction of polarization, may improve the reproducibility of water prediction further (Lopes *et al.*, 2013).

During validations and tests, MobyWat automatically calculated SR values using a match tolerance of 1.5 Å which is the upper limit for the detection of matches between predicted and reference water molecule pairs (Section 2.2). To further quantify the precision of matches, statistics of distances of all matched pairs of the top performer algorithms were calculated (Fig. 2D). It can be seen that mean match distances are below 1 Å for all systems. Matching water positions of one of the systems is shown in Figure 3A, and three other systems are depicted in Supplementary Figure S4.

3.3 Featured test examples

Test set 2 containing 12 proteins was assembled to further check the performance of MobyWat predictions. Using prediction algorithm IDa, a mean SR of 87% was achieved for this set. The members of Test set 2 and the resulted SR values are listed in Supplementary Table S8. Below the prediction results obtained for three enzymatic systems of Test set 2 are discussed focusing on their active sites.

Cyclin-dependent kinase 2 (Cdk2) is a key enzyme in cell cycle control and a promising drug target in oncology (Akli *et al.*, 2011) that also affects senescence (Chenette, 2010). A change of the hydration structure of the active site of Cdk2 due to ligand binding has been reported with obvious implications for drug design (Schulze-Gahmen *et al.*, 1996). A good agreement was obtained between predicted (blue spheres, Fig. 3B) and experimental reference (red spheres) water positions verifying that MobyWat accurately predicted the hydration structure of the active site of apo Cdk2 (Fig. 3B). Notably, experimental water positions were used in comparisons of Figures. 3B–D without any restrictions on their B-factors. Insertion of the ligand (ATP, thin lines in Fig. 3B) from the superimposed ATP-bound Cdk2 structure reveals that six waters (marked with asterisks in Fig. 3) are displaced by the ligand during binding. Release of such water molecules has a favorable contribution to binding entropy of the ligand, and therefore, their identification is important for thermodynamics-driven engineering of new ligands. The results were not affected by the chemical nature of ligand binding as waters replaced by both the charged phosphate moieties and the non-charged adenine ring were found correctly. This finding is in agreement with our general results showing that prediction quality is independent on the type of interacting amino acids

(Supplementary Results S3.7). The second example (Fig. 3C) features the nucleoside binding pocket of thymidine kinase from Herpes simplex type 1. This enzyme has been involved in enzyme-prodrug gene therapy of cancer (Vogt *et al.* 2000). Besides two replaceable water molecules, MobyWat precisely predicted several conserved water positions (marked with C in Fig. 3) existing in both the apo and the ligand-bound enzyme structures. Similar to the cases of replaceable water molecules, locating conserved water sites precisely is also important during the design of new ligands. A complete chain of waters leading to the active site was also predicted correctly (top-right corner of Fig. 3C). The third binding pocket in Figure 3D belongs to glutathione S-transferase, an important detoxifying enzyme (Wu and Dong, 2012). Binding chemistry of glutathione, the peptidic ligand of this enzyme is remarkably different from the previous two ligands with heteroaromatic cores (Fig. 3B and C). However, the quality of MobyWat prediction of the surrounding water positions is similarly good as it was in the other two examples.

MobyWat produces a prediction list including water positions in increasing order of mobility scores (Supplementary Algorithm S2.1.5) where experimentally verified (positive) predictions are mostly located at the top of the prediction list. It was found (Supplementary Results S3.4) that 88% of positive predictions for whole protein surfaces are located in the top 50% of the prediction list. As active sites are the most important spots on enzymes, it was also checked how mobility scores work for these specific segments of the surface. 20 of 24 (85%) of the correctly predicted water positions shown in Figures. 3B–D are located in the top 15% of the prediction lists. Thus, in the cases of active sites investigated, the mobility scores short-list the positive candidates very efficiently at the top of the prediction list. This indicates that water molecules in the active sites of enzymes are predicted with higher fidelity than other water molecules residing on the surface. This result can in part be explained by the presence of conserved water molecules surrounding the ligands, most of which are located at the top 5% of prediction lists. Notably, half of replaceable water molecules occupying active sub-sites in the apo structures were also ranked at top 10%.

4 Conclusions

MD has become an indispensable tool of prediction of structure of proteins and protein–ligand complexes (Shan *et al.*, 2011; Söderhjelm *et al.*, 2012). However, there are only a few MD-based methods for the prediction of hydration structure using explicit simulation of water contacts. Here, we presented MobyWat, a freely available program validated and tested on more than 1500 experimental water positions in 20 different protein surfaces. The prediction process of MobyWat aims at finding the least mobile (most occupied) points of the hydration structure. It was shown that MD simulation is an appropriate sampling technique for such predictions. MobyWat performs predictions using mobility information cumulated in MD trajectories. Two predictive approaches were implemented and tested. The first approach uses only spatial information (coordinates) for a candidate water position. This can be done for example by averaging trajectory frames and producing solvent densities (Virtanen *et al.*, 2010) or by clustering water molecules along the trajectory and counting frequencies of their occurrence in candidate positions. In this study, a second approach was introduced based on identification records of water molecules rather than spatial positions. On average, the identity-based predictions provided higher success rate values than positional and merged algorithms. This is probably a consequence of the position-independent philosophy of the identity-based algorithms.

Valid predictions do not require trajectories from long MD runs. The typical lifetime of a hydrogen bond is a few pico seconds only, virtually independent of the environment (van der Spoel *et al.*, 2006). Consequently, due to rapid exchange and equilibration of water positions relatively short simulations (e.g. 1–10 ns) with regular saving of coordinates suffice. Thus, with a moderate computational effort valid predictions can be achieved.

Limitations of mobility-based predictions were also investigated via an analysis of non-matched water positions of eight systems (Supplementary Results S3.7 and Appendix 2). The analysis identified location of waters above shallow protein sites and/or far from the surface to be a limiting factor in a few cases. Further work is on the way to overcome such limitations using a relative coordinate definition and testing combined MD sampling schemes.

MobyWat algorithms were coded in the portable C language. As the program has to perform calculations on numerous atoms in numerous frames (e.g. $10^4 \times 10^4$) special attention was paid to the efficient use of memory. MobyWat can be used in conjunction with any MD program as it reads frames from PDB files. However, for efficient use of memory and disk space MobyWat also reads and writes xdr-type portable binary trajectory files called xtc in GROMACS.

Mobility is often considered as a disturbing property hampering experimental determination of positions of water molecules on protein surfaces. In this study, it was shown that mobility can be utilized as an information source for prediction of hydration structure. If experimental determination of water structure is not available or incomplete, MobyWat can offer an alternative solution.

Acknowledgements

Part of the simulations were carried out on resources provided by the Swedish National Infrastructure for Computing (SNIC) at the 'Abisko' supercomputer of the High Performance Computing Center North (HPC2N, Sweden, grant SNIC2013-26-6). We acknowledge PRACE for awarding us access to resources Monte Rosa based in Switzerland at CSCS Swiss National Supercomputing Centre, and NIIFI SC based in Hungary at NIIF National Information Infrastructure Development Institute.

Funding

The work was supported by the Hungarian Scientific Research Fund (OTKA K112807) and the MedinProt project of the Hungarian Academy of Sciences. We are thankful to the Gedeon Richter Pharmaceutical Plc. for a pre-doctoral scholarship (to N.J.).

Conflict of Interest: none declared.

References

Abel,R. *et al.* (2008) Role of the active-site solvent in the thermodynamics of factor Xa ligand binding. *J. Am. Chem. Soc.*, **130**, 2817–2831.

Afonine,P.V. *et al.* (2013) Bulk-solvent and overall scaling revisited: faster calculations, improved results. *Acta Cryst.*, **D69**, 625–634.

Akli,S. *et al.* (2011) Cdk2 is required for breast cancer mediated by the low-molecular-weight isoform of Cyclin E. *Cancer Res.*, **71**, 3377–3386.

Angel,T.E. *et al.* (2009) Structural waters define a functional channel mediating activation of the GPCR, rhodopsin. *Proc. Natl Acad. Sci. USA*, **106**, 147367–14372.

Antonyuk,S.V. *et al.* (2013) Structures of protein–protein complexes involved in electron transfer. *Nature*, **496**, 123–126.

Badger,J. (1997) Modeling and refinement of water molecules and disordered solvent. *Methods Enzymol.*, **277**, 344–352.

Baron,R. *et al.* (2012) Hydrophobic association and volume-confined water molecules. In:Gohlke,H. (ed.) *Protein–Ligand Interactions*. Wiley-VCH Verlag GmbH & Co. KGaA, Wennheim.

Carugo,O. (1999) Correlation between occupancy and B-factor of water molecules in protein crystal structures. *Protein Eng.*, **12**, 1021–1024.

Chenette,E.J. (2010) Senescence: a key role for CDK2. *Nat. Rev. Cancer.*, **10**, 84.

Cheung,M.S. *et al.* (2002) Protein folding mediated by solvation: water expulsion and formation of the hydrophobic core occur after the structural collapse. *Proc. Natl Acad. Sci. USA*, **99**, 685–690.

Cui,G. *et al.* (2013) SPAM: a simple approach for profiling bound water molecules. *J. Chem. Theor. Comput.*, **9**, 5539–5549.

de la Lande,A. *et al.* (2010) Surface residues dynamically organize water bridges to enhance electron transfer between proteins. *Proc. Natl Acad. Sci. USA*, **107**, 11799–11804.

Dror,R.O. *et al.* (2012) Biomolecular simulation: a computational microscope for molecular biology. *Annu. Rev. Biophys.*, **41**, 429–452.

Ehrlich,L. *et al.* (1998) Prediction of protein hydration sites from sequence by modular neural networks. *Protein Eng. Des. Sel.*, **11**, 11–19.

Fang,C. *et al.* (2009) Mapping GFP structure evolution during proton transfer with femtosecond Raman spectroscopy. *Nature*, **462**, 200–204.

Finney,J.L. (1977) The organization and function of water in protein crystals. *Philos. Trans. R. Soc. Lond. B*, **278**, 3–32.

García-Sosa,A.T. (2013) Hydration properties of ligands and drugs in protein binding sites: tightly-bound, bridging water molecules and their effects and consequences on molecular design strategies. *J. Chem. Inf. Model.*, **53**, 1388–1405.

García-Sosa,A.T., *et al.* (2003) WaterScore: a novel method for distinguishing between bound and displaceable water molecules in the crystal structure of the binding site of protein–ligand complexes. *J. Mol. Mod.*, **9**, 172–182.

Garcia-Sosa,A.T., *et al.* (2005) Including tightly-bound water molecules in de novo drug design. Exemplification through the in silico generation of poly(ADP-ribose)polymerase ligands. *J. Chem. Inf. Model.*, **45**, 624–633.

Garcia-Sosa,A.T., and Mancera,R.L. (2006) The effect of tightly-bound water molecules on scaffold diversity in the computer-aided de novo ligand design of CDK2 inhibitors. *J. Mol. Mod.* **12**, 422–431.

Halle,B. (2004a) Protein hydration dynamics in solution: a critical survey. *Philos. Trans. R. Soc. Lond. B*, **359**, 1207–1224.

Halle,B. (2004b) Biomolecular cryocrystallography: structural changes during flash-cooling. *Proc. Natl Acad. Sci. USA*, **101**, 4793–4798.

Henchman,R.H. and McCammon,J.A. (2002) Extracting hydration sites around proteins from explicit water simulations. *J. Comput. Chem.*, **23**, 861–869.

Hess,B. *et al.* (2008) GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.*, **4**, 435–447.

Israelachvili,J. and Wennerström,H. (1996) Role of hydration and water structure in biological and colloidal interactions. *Nature*, **379**, 219–225.

Levy,Y. and Onuchic,J.N. (2006) Water mediation in protein folding and molecular recognition. *Annu. Rev. Biophys. Biomol. Struct.*, **35**, 389–415.

Lloyd,D.G. *et al.* (2004) The effect of tightly bound water molecules on the structural interpretation of ligand-derived pharmacophore models. *J. Comput. Aided Mol. Des.*, **18**, 89–100.

Lopes,P.E.M. *et al.* (2013) Polarizable force field for peptides and proteins based on the classical drude oscillator. *J. Chem. Theor. Comput.*, **9**, 5430–5449.

Makarov,V.A. *et al.* (1998) Reconstructing the protein-water interface. *Biopolymers*, **45**, 469–478.

Mancera,R.L. (2007) Molecular modeling of hydration in drug design. *Curr. Opin. Drug Discov. Dev.*, **10**, 275–280.

Michel,J. *et al.* (2009) Energetics of displacing water molecules from protein binding sites: consequences for ligand optimization. *J. Am. Chem. Soc.*, **131**, 15403–15411.

Nisius,L. and Grzesiek,S. (2012) Key stabilizing elements of protein structure identified through pressure and temperature perturbation of its hydrogen bond network. *Nat. Chem.*, **4**, 711–717.

Patel,H. *et al.* (2014) PyWATER: a PyMOL plug-in to find conserved water molecules in proteins by clustering. *Bioinformatics*, **30**, 2978–2980.

Petsko,G.A. and Ringe,D. (2009) *Protein Structure and Function.* Oxford University Press Inc., New York.

Pettitt,B.M. and Karplus,M. (1987) The structure of water surrounding a peptide: a theoretical approach. *Chem. Phys. Lett.*, **136**, 383–386.

Pitt,W.R. and Goodfellow,J.M. (1991) Modelling of solvent positions around polar groups in proteins. *Protein Eng.*, **4**, 531–537.

Pronk,S. *et al.* (2013) GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics*, **29**, 845–854.

Roberts,B.C. and Mancera,R.L. (2008) Ligand-protein docking with water molecules. *J. Chem. Inf. Model.*, **48**, 397–408.

Ross,G.A. *et al.* (2012) Rapid and accurate prediction and scoring of water molecules in protein binding sites. *PLoS ONE*, **7**, e32036.

Savage,H. and Wlodawer,A. (1986) Determination of water structure around biomolecules using x-ray and neutron diffraction methods. *Methods Enzymol.*, **127**, 162–183.

Schoenborn,B.P. *et al.* (1995) Hydration in protein crystallography. *Prog. Biophys. Mol. Biol.*, **64**, 105–119.

Schulze-Gahmen,U. *et al.* (1996) High-resolution crystal structures of human cyclin-dependent kinase 2 with and without ATP: bound waters and natural ligand as guides for inhibitor design. *J. Med. Chem.* **39**, 4540–4546.

Schymkowitz,J.W.H. *et al.* (2005) Prediction of water and metal binding sites and their affinities by using the Fold-X force field. *Proc. Natl Acad. Sci. USA*, **102**, 10147–10152.

Shan,Y. *et al.* (2011) How does a drug molecule find its target binding site? *J. Am. Chem. Soc.*, **133**, 9181–9183.

Söderhjelm,P. *et al.* (2012) Locating binding poses in protein–ligand systems using reconnaissance metadynamics. *Proc. Natl Acad. Sci. USA*, **109**, 5170–5175.

Thilagavathi,R. and Mancera,R.L. (2010) Ligand–protein cross docking with water molecules. *J. Chem. Inf. Model.*, **50**, 415–421.

Thirumalai,D. *et al.* (2011) Role of water in protein aggregation and amyloid polymorphism. *Acc. Chem. Res.*, **45**, 83–92.

Truchon,J-F. *et al.* (2014) A cavity corrected 3D-RISM functional for accurate solvation free energies. *J. Chem. Theor. Comput.*, **10**, 934–941.

van der Spoel,D. *et al.* (2006) Thermodynamics of hydrogen bonding in hydrophilic and hydrophobic media. *J. Phys. Chem. B*, **110**, 4393–4398.

Vedani,A. and Huhta,D.W. (1991) An algorithm for the systematic solvation of proteins based on the directionality of hydrogen bonds. *J. Am. Chem. Soc.*, **113**, 5860–5862.

Virtanen,J.J. *et al.* (2010) Modeling the hydration layer around proteins: HyPred. *Biophys. J.*, **99**, 1611–1619.

Vogt,J. *et al.* (2000) Nucleoside binding site of herpes simplex type 1 thymidine kinase analyzed by X-ray crystallography. *Proteins*, **41**, 545–553.

Wu,B. and Dong,D. (2012) Human cytosolic glutathione transferases: structure, function, and drug discovery. *Trends Pharmacol. Sci.*, **33**, 656–668.

Xu,Y. and Leitner,M.D. (2014) Vibrational energy flow through the green fluorescent protein–water interface: communication maps and thermal boundary conductance. *J. Phys. Chem. B.*, **118**, 7818–7826.

Published online: November 3, 2015

Article

SOURCE DATA | TRANSPARENT PROCESS | OPEN ACCESS

molecular systems biology

Systematic discovery of linear binding motifs targeting an ancient protein interaction surface on MAP kinases

András Zeke[1], Tomas Bastys[2,3], Anita Alexa[1], Ágnes Garai[1], Bálint Mészáros[4], Klára Kirsch[1], Zsuzsanna Dosztányi[5], Olga V Kalinina[2] & Attila Reményi[1,*]

Abstract

Mitogen-activated protein kinases (MAPK) are broadly used regulators of cellular signaling. However, how these enzymes can be involved in such a broad spectrum of physiological functions is not understood. Systematic discovery of MAPK networks both experimentally and *in silico* has been hindered because MAPKs bind to other proteins with low affinity and mostly in less-characterized disordered regions. We used a structurally consistent model on kinase-docking motif interactions to facilitate the discovery of short functional sites in the structurally flexible and functionally under-explored part of the human proteome and applied experimental tools specifically tailored to detect low-affinity protein–protein interactions for their validation *in vitro* and in cell-based assays. The combined computational and experimental approach enabled the identification of many novel MAPK-docking motifs that were elusive for other large-scale protein–protein interaction screens. The analysis produced an extensive list of independently evolved linear binding motifs from a functionally diverse set of proteins. These all target, with characteristic binding specificity, an ancient protein interaction surface on evolutionarily related but physiologically clearly distinct three MAPKs (JNK, ERK, and p38). This inventory of human protein kinase binding sites was compared with that of other organisms to examine how kinase-mediated partnerships evolved over time. The analysis suggests that most human MAPK-binding motifs are surprisingly new evolutionarily inventions and newly found links highlight (previously hidden) roles of MAPKs. We propose that short MAPK-binding stretches are created in disordered protein segments through a variety of ways and they represent a major resource for ancient signaling enzymes to acquire new regulatory roles.

Keywords cellular signaling; linear motif; MAP kinase; protein–protein interaction

Subject Categories Computational Biology; Signal Transduction; Structural Biology

DOI 10.15252/msb.20156269 | Received 29 April 2015 | Revised 29 September 2015 | Accepted 1 October 2015

Mol Syst Biol. (2015) 11: 837

Introduction

Protein–protein interactions influence all aspects of cellular life and the most direct mechanism through which proteins can influence each other is by physical interaction. This brings them into proximity to exert control on activity or to create opportunities for posttranslational modification. Protein–protein associations often involve so-called linear binding motifs which are short (5–20 amino acid long) protein regions lacking autonomous tertiary structure. These functional sites reside in intrinsically disordered protein regions and adopt stable conformation only upon binding. Currently, we can only guess how abundant linear motif-based interactions are; nevertheless, it was recently estimated that there are ~100,000 linear binding motifs targeting dedicated protein surfaces in the human proteome (Tompa *et al*, 2014). As an example relevant to cellular signaling, mitogen-activated protein kinases (MAPKs) are prototypical enzymes that depend on short segments from partner proteins and on their dedicated protein–protein interaction hot spots. They mainly recognize their substrates not with the catalytic site but with auxiliary docking surfaces on their kinase domains (Tanoue *et al*, 2000; Biondi & Nebreda, 2003). The most important of these docking sites consists of a hydrophobic docking groove and the negatively charged CD (common docking) region (Chang *et al*, 2002) (Fig 1A). Together, they can bind the so-called D(ocking)-motifs of the target proteins. D-motifs are short linear motifs ranging from 7 to 18 amino acids in length and are typically found in intrinsically disordered segments—potentially far away from target phosphorylation sites (Garai *et al*, 2012). Such docking

1 Lendület Protein Interaction Group, Institute of Enzymology, Research Center for Natural Sciences, Hungarian Academy of Sciences, Budapest, Hungary
2 Max Planck Institute for Informatics, Saarbrücken, Germany
3 Graduate School of Computer Science, Saarland University, Saarbrücken, Germany
4 Institute of Enzymology, Research Center for Natural Sciences, Hungarian Academy of Sciences, Budapest, Hungary
5 MTA-ELTE Lendület Bioinformatics Research Group, Department of Biochemistry, Eötvös Loránd University, Budapest, Hungary
*Corresponding author. Tel: +36 1 3826613; E-mail: remenyi.attila@ttk.mta.hu

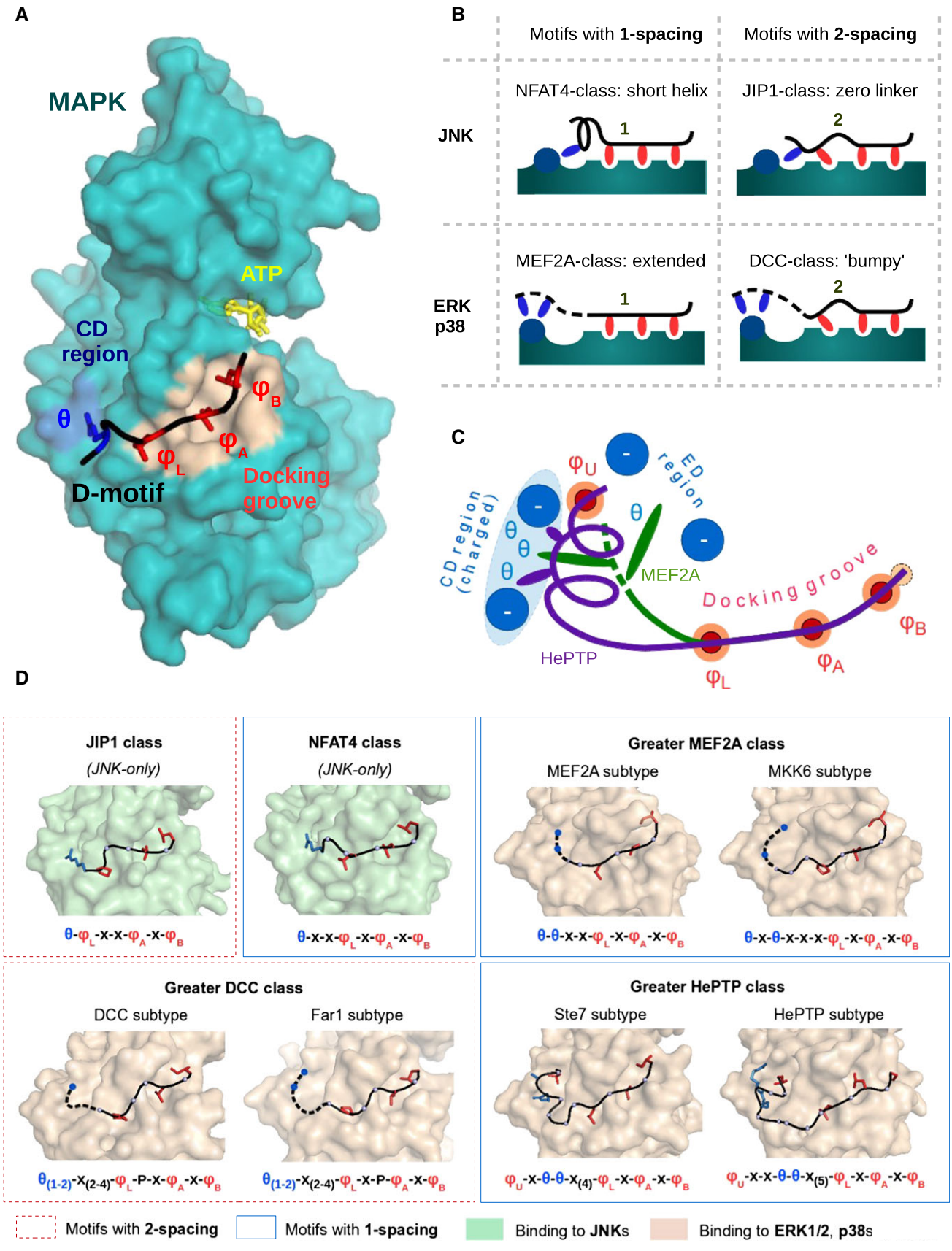

Figure 1.

Figure 1. Structural classification of MAPK-docking motifs.

A The MAPK-docking groove comprises two distinct regions: the common docking (CD) and the hydrophobic region. These are colored in blue and light brown, respectively, and are shown on the JNK1 surface from the JNK1-NFAT4 protein–peptide complex crystal structure (Garai *et al*, 2012). (The CD groove is extended by the ED region, extra negatively charged residues for ERK and p38; see (C); Tanoue *et al*, 2001.)

B Different binding modes of D-motifs. The hydrophobic docking groove binds three hydrophobic amino acids in a row, while admitting two different spacing schemes. At the same time, θ to φ linker length determines the MAPK specificity of a given motif. These two features can combine freely with each other, resulting in the four basic arrangements shown here.

C Distinct binding conformations at the CD groove. N-termini of longer D-motifs are variable and ERK2- or p38α-binding peptides may take a variety of conformations —ranging from fully linear (e.g., MEF2A, green) to highly alpha-helical (e.g., HePTP, magenta).

D Structural heterogeneity of D-motifs. The combinations of the three variable features yield structurally well-defined, distinct classes of D-motifs. Many of these models also define separate types of linear motifs, but their consensus sequences are not always exclusive. JNK kinases only admit two major types of motifs, the NFAT4 class (1-spacing, short linker) and the JIP1 class (2-spacing, short linker). On the other hand, known ERK1/2 and p38 binder peptides may belong to the greater MEF2A class (1-spacing, longer linker, linear end), the greater HePTP class (1-spacing, longer linker, helical end), or the greater DCC class (2-spacing, longer linker, linear end). A sixth class of ERK or p38 interactors is theoretically also possible (2-spacing, longer linker, helical end), but this combination can only be observed in long reverse (revD) motifs (Garai *et al*, 2012), and no classical motif of this type is known up to date. Subtypes and other variants within a given greater class are also featured wherever applicable. These are shown based on structures of MAPK-docking motif complexes. Dashed lines indicate N-terminal peptide regions that are usually not visible in the crystal structures, albeit indispensable for binding. Consensus motif of each subtype is shown below, where $\varphi_{U,}$ φ_{L}, φ_{A}, and φ_{B} letters denote positions that are filled by hydrophobic amino acids—L, A, and B refer to the lower pocket, and pockets A and B, respectively—while the θ positions are positively charged (Arg or Lys) while letter "x" denotes arbitrary amino acids.

elements are not only restricted to substrates: They are also found in MAPK-activating kinases (MAP2Ks), in MAPK-inactivating phosphatases (MKPs), and in a variety of scaffold proteins. While extracellularly regulated kinases (ERK1-2), c-Jun N-terminal kinases (JNK1-3), and the 38-kDa protein kinases (p38α-δ) control diverse physiological processes, they phosphorylate most of their substrates at Ser-Pro or Thr-Pro (S/TP) sequence motifs. Naturally, this weak consensus provided by their catalytic site is insufficient for selective target recognition, and additional linear binding motifs provide specificity (Johnson & Lapadat, 2002; Bardwell, 2006). Therefore, the MAPK D-motif protein–protein interaction system is an ideal test bed for linear binding motif discovery.

Several previous attempts were aimed at predicting MAPK-binding proteins from full proteomes by using a generic consensus of D-motifs as it had been established more than a decade ago (Sharrocks *et al*, 2000). This consensus was derived from an observation that D-motifs almost always include at least a single positively charged residue (termed the θ position: arginine or lysine) and a series of alternating hydrophobic residues (φ positions: frequently leucine), connected by a linker of a variable length (Dinkel *et al*, 2014). But despite the use of extensive multiple alignments and sophisticated algorithms, predictions had only low success rates and large-scale assessment of predicted hits was not performed (Whisenant *et al*, 2010; Garai *et al*, 2012; Gordon *et al*, 2013). Regarding experimental MAPK network discovery, ERK2 has been the most widely explored. For example, several different methods were utilized to identify ERK2 substrates by large-scale phosphoproteomics (Kosako *et al*, 2009; Carlson *et al*, 2011; Courcelles *et al*, 2013). Unfortunately, pairwise overlaps between the lists of substrates are low across studies (e.g., around ~10%), with not a single overlap between five different studies that aimed to find ERK2-phosphorylated substrates (Courcelles *et al*, 2013), suggesting great dependence on the experimental conditions used. It was noted that D-motif-like sequences are enriched in experimentally detected ERK2 substrates (Carlson *et al*, 2011), yet detection or verification of direct physical association was not performed. In addition, studies that used a high-throughput approach to identify partners of JNK1 (Chen *et al*, 2014) or p38α (Bandyopadhyay *et al*, 2010) based on direct physical interaction resulted in low number of hits. Thus, it is likely that a protein–protein interaction-based MAPK network discovery could greatly benefit from a target tailored approach, which takes into account—and possibly capitalizes on—specific biochemical and biophysical knowledge already available on known MAPK–partner protein interactions.

In recent years, the number of experimentally determined MAPK–partner protein complex structures increased considerably (Garai *et al*, 2012). This development made it possible to amend the definition of the underlying sequence motifs and it became clear that D-motifs encompass multiple classes of similarly built, but structurally distinct linear motifs (similar to SH3- or PDZ-domain-binding sequences, for example) (Lim *et al*, 1994; Tonikian *et al*, 2008). In the current study, we show that by building a strategy that can handle this conformational diversity in binding, and using structural compatibility with specific interaction surface topography as an additional criterion for prediction, the identification of novel D-motifs can be dramatically improved. This analysis in combination with tailored experimental techniques for the validation of low-affinity (1–20 μM) protein–protein interactions produced unique, molecular-level insight into physiological roles and evolution of MAPK-based protein networks.

Results

Structure-guided prediction of MAPK-binding linear motifs

MAPK–D-peptide complex structures revealed distinct D-motif binding modes in the MAPK-docking groove (Fig 1). For example, D-motifs from the JNK-binding scaffold protein JIP1 and from the JNK-regulated transcription factor NFAT4 bind to the same docking surface differently (Fig 1A and B) (Heo *et al*, 2004; Garai *et al*, 2012; Laughlin *et al*, 2012). Similarly, ERK- and p38-binding D-motifs may also be structurally distinct; nonetheless, D-motifs could be described with a common loosely defined consensus [$\theta_{1,2}$-$x_{(0\text{-}5)}$-φL-$x_{(1,2)}$-φA-x-φB; φL, φA, and φB denote positions that are typically filled by hydrophobic amino acids—L, A, and B refer to the lower pocket, and pockets A and B, respectively—while the θ denotes positively charged (arginine or lysine) and "x" denotes any amino acid]. However, the rules are much stricter for sequences that are compatible with a given MAPK-docking surface in a given binding

mode. Interestingly, D-motifs and their binding modes may be conserved from yeast to human as the docking surface is ancient and well conserved across all eukaryotes (Reményi *et al*, 2005; Grewal *et al*, 2006).

Because the CD region of ERK and p38 is wider compared to that of JNK, the N-termini of motifs binding to the two former kinases have larger conformational freedom (Fig 1C) (Garai *et al*, 2012). These can be classified as MEF2A- and DCC-type motifs named after the proteins in which they were first identified. Some motifs with longer intervening regions also exists (HePTP) (Zhou *et al*, 2006). Interestingly, the typical helical conformation at the N-terminus of HePTP-type docking motif is also characteristic to some MAPK interactors from yeast (Ste7) and peptides with such motifs are known to bind human ERK2 with high affinity (Fernandes *et al*, 2007). Therefore, we also set up a hypothetical subclass of Ste7-type motifs, hitherto unknown in humans (Fig 1D).

Simple pattern matching of motifs normally produces a large number of false positives, because motif-matching sequences may occur simply by chance. In order to drastically reduce the number of false hits, an *in silico* filtering procedure was implemented to search for putative linear motifs (Fig 2). The first step was to screen for motifs in regions with intrinsic disorder but with propensity for disorder-to-order transition (ANCHOR) (Dosztányi *et al*, 2009; Mészáros *et al*, 2009). This procedure was used in order to eliminate those consensus motif occurrences that would either be buried in folded domains or permanently locked in an unfavorable conformation. Importantly, it also removed initial hits with an inappropriate amino acid composition, not being able to adopt a stable structure upon binding to a protein surface. Motifs were then filtered for MAPK accessibility: Motifs that were predicted to lie in extracellular protein segments or in other, kinase-inaccessible compartments (e.g., ER, Golgi) were discarded. In addition, an auxiliary check was performed against structured Pfam domains. This was applied to remove all spurious motifs in ordered regions which had been retained after ANCHOR filtering. Since almost all the known motifs passed these filters (45 out of 47, with enrichment ratios between 4.1 to 6.6 depending on motif type), we concluded that the dataset was of sufficient quality for further testing.

The classification of D-motifs based on a coherent structural model enabled us to make use of structure-based scoring. As a motif occurrence could always be unambiguously matched with its corresponding MAPK-docking peptide structural model, we used FoldX, which had been validated on protein–peptide complexes, to estimate the change of the protein–peptide binding energy with respect to the energy of an experimentally resolved complex (Appendix Table S1) (Schymkowitz *et al*, 2005). This allowed the scoring of motifs according to their structural compatibility to the MAPK-docking groove. FoldX-derived binding energy estimates were also used as a guide when motifs were being chosen for later experimental screening.

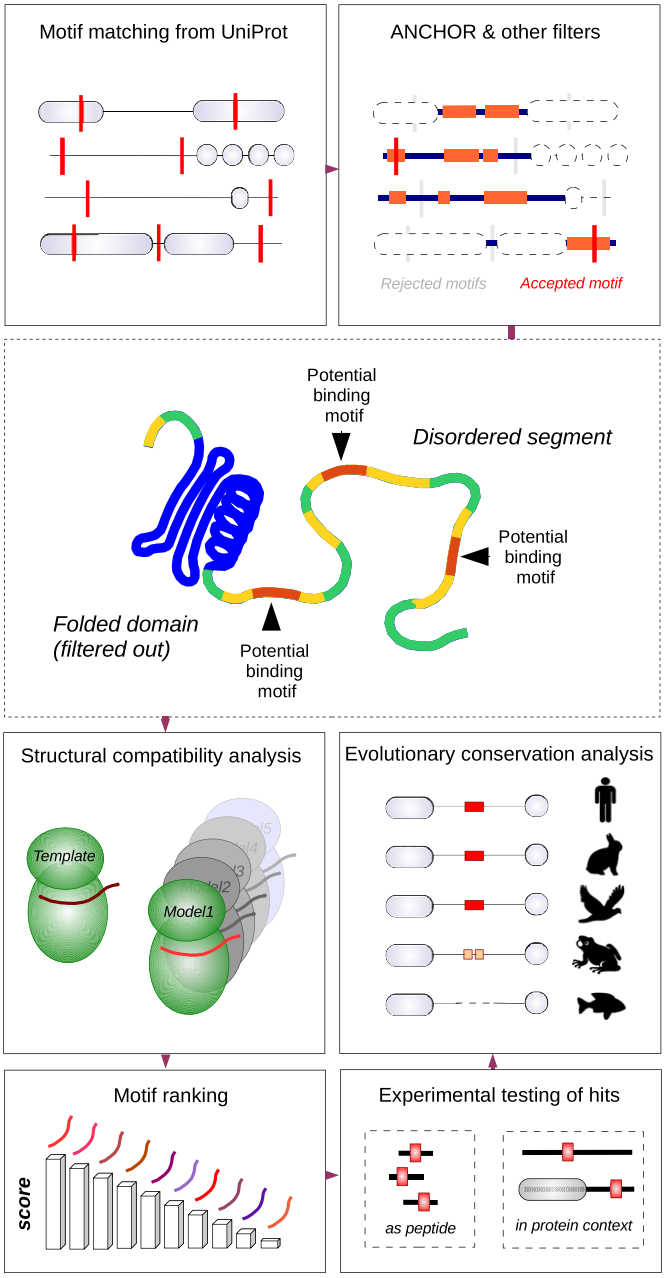

Figure 2. Motif finding work flow.
To find novel MAPK-docking motifs, the primary motif-matching step (on UniProt KB sequences) was followed by a series of filters. Valid motifs had to pass through an ANCHOR filter, a localization filter (combined from SignalP and Phobius) and an auxiliary Pfam filter, in order to be scored by FoldX homology models. ANCHOR (middle panel) had the most important role in selecting segments that can potentially act as linear motifs, while FoldX gave motif-specific binding energy estimates (see Source data). Predicted motifs were subsequently tested as short fragments and (in selected cases) as full-length proteins. Finally, an automated evolutionary analysis was performed to give information on motif conservation trends.

Source data is available online for this figure.

Experimental screening

After completion of initial lists, we chose a number of candidate proteins from each motif type to test. Expression of full-length human proteins of large size (> 1,000 amino acids) in recombinant systems can be a limiting step in experimental validation; therefore, first we opted for a fragment-based approach. Former experiments showed that simple binding assays (such as pull-downs with recombinant D-motif-containing proteins or immobilized solid-phase peptide arrays) lack the robustness to reliably detect low-affinity

(1–20 μM) protein–peptide interactions. Therefore, we developed a different assay which was based on substrate phosphorylation enhancement on a solid-phase support (Fig 3A). As the majority of known D-motif-containing proteins are MAPK substrates, this adequately captures the original function of these motifs. An artificial substrate was constructed containing the D-motifs as well as the Thr71 phosphorylation site from ATF2, which is a well-known MAPK target site (Livingstone *et al*, 1995) (Fig 3B). As linkers and substrate sites in the recombinant proteins were identical, the "docking efficiency" of the given motifs could be directly compared to each other. Phosphorylation of this reporter solely depended on the presence or absence of specific docking motifs, and phosphorylation of the target site was low without a functional D-motif (Fig 3C and D).

In the final panels, we included 70 different constructs: 63 of these were directly selected from the lists ranked by the predicted binding energy (Fig EV1 and Appendix Fig S1). We also included additional seven motifs based on sequence similarity to known motifs. This was done in order to test whether some other similar motifs not conforming to the formerly defined sequence patterns have the capacity to bind MAPKs. Out of 70, a total of 52 motifs were found to interact with at least one of the three MAPKs (ERK2, JNK1, or p38α). In particular, we were able to detect several novel interactors based on the JIP1, NFAT4, MEF2A, MKK6, and DCC models. As for our hypothetic Ste7 model, we also found a novel hit: a motif from RHDF1 that is also found in the related RHDF2 protein. Such a high number of hits suggest that D-motifs are in fact quite widespread in the human proteome (Table 1, Fig EV2, and Appendix Fig S2).

To show that the phosphorylation enhancements were indeed due to the presence of canonical MAPK D-motifs binding into the MAPK-docking groove, a set of 16 chemically synthesized peptides were titrated against fluorescently labeled control peptides known to bind at the MAPK-docking groove (Fig EV3 and Appendix Fig S3). In addition to confirming binding in the MAPK-docking groove, the dissociation constant (Kd) of unlabeled test peptides could also be calculated.

Binding affinities obtained through this fluorescence polarization (FP) based *in vitro* assay also allowed us to examine the specificity profiles of D-motifs. The tested peptides could be clustered into two groups based on their sequences and affinities. Similar to earlier results, these experiments confirmed the strong correlation between the ability of a given motif to bind ERK2 and p38α. Binding results also reflected the fundamental lack of correlation between ERK2/p38α and JNK1 association (Garai *et al*, 2012). These observations did agree well with phosphorylation enhancement results from dot blots. There was no positive correlation between the profiles of the JNK1 vs. p38α or the JNK1 vs. ERK2 pairs (Pearson's $r = 0.003$ and $r = -0.280$, respectively). At the same time, a modest correlation was observed between ERK2 and p38α ($r = 0.680$). This MAPK profiling confirmed our structural models. Practically, no strong JNK1-binding motif was found from other than the JIP1- or NFAT4-type classes. Most novel p38α interactors, on the other hand, belonged to the MEF2A, MKK6, or DCC types as expected.

To test whether docking motifs were also functional in their native protein context, we set up a bimolecular fluorescent protein fragment complementation (BiFC)-based cellular assay (Fig 4A). In this series of experiments, one fragment of YFP was fused to either ERK2, JNK1, or p38α. The other fragment was joined to the test protein, and fluorescence intensity was measured in live HEK293 cells. BiFC signals were always compared to the results obtained with the same construct but lacking the docking motif. Although the transgenes were overexpressed, comparison of the BiFC signal between wild-type and D-motif-mutated MAPK partner proteins could be reliably used to infer D-motif-dependent interactions within cells.

Well-known MAPK partners, such as MKK1, JIP1, or MKK6, displayed a pattern consistent with the specificity of their D-motifs. Such interactions are also greatly diminished or abrogated after the loss of the docking motif, similar to novel MAPK partners (Fig 4B). For example, the MEF2A-type motif-bearing AMP-activated protein kinase subunit γ2 (AAKG2) interacted with p38α (and ERK2), but not with JNK1. Interestingly, AAKG2 is known to have multiple shorter isoforms and it uses alternative initiation codons. One such variant (isoform C) is only 44 amino acids shorter. This natural deletion mutant lacking the N-terminal (MEF2A-type) docking motif showed a greatly reduced level of fluorescence for both partners. The differences in fluorescence were readily visible on cells under a fluorescent microscope (Fig 4C). All intensities, as well as their reduction in the mutants, were also comparable to those observed in control experiments (Fig 4D). These results are well in line with *in silico* predictions and *in vitro* fragment-based experiments. To this end, we tested six predicted motifs (AAKG2, MKP5, RHDF1, KSR2, DCX, APBA2), and one non-binder based on results of dot-blot arrays (FAM122A) was also included (Fig 4B and Appendix Fig S4). Results of this cell-based approach were consistent with the structural models as well as with the results of *in vitro* experiments.

D-motif-based MAPK interactomes

Next, we utilized the experimentally validated new D-motifs to further improve our initial structural models. Evolutionary conservation analysis on motifs was also used to examine sequence conservation or diversity per each position (Fig 5A). Once the consensus sequences were improved, we set out to build a sequence-based method to enable direct search for MAPK-interacting proteins from the human proteome. Position-specific scoring matrices (PSSMs) were constructed from full sets of evolutionarily related docking motifs. PSSM-based profiles have been used in multiple databases for encoding information about sequence profiles, in the search for proteins with distant similarity, and they were recently used for detecting MAPK target phosphorylation sites (Schäffer *et al*, 2001; Hulo *et al*, 2004; Gordon *et al*, 2013).

High area under curve (AUC) values of the receiver operation characteristic (ROC) curve from a fivefold cross-validation with validated binders and simulated non-binders (see Materials and Methods) imply an adequate coverage of motifs in the JIP1, NFAT4, and greater MEF2A classes: 0.98, 0.94, and 0.97, respectively (Table EV1). The correlation with the original, FoldX-based rankings was modest, but clearly present in the case of JIP1-type ($r = -0.62$) and NFAT4-type ($r = -0.59$) motifs (where best correlation is −1, due to the negative energy scale). It was lower for the DCC ($r = -0.40$), MEF2A ($r = -0.30$), and MKK6 ($r = -0.26$) models, as somewhat expected, since the structural templates of these were incomplete (as the structures of the charged N-termini of some D-motifs are not known). Unfortunately, the lack of sufficiently diverse hits among DCC and HePTP-type motifs made PSSM construction impractical. A PSSM was still built for the greater DCC

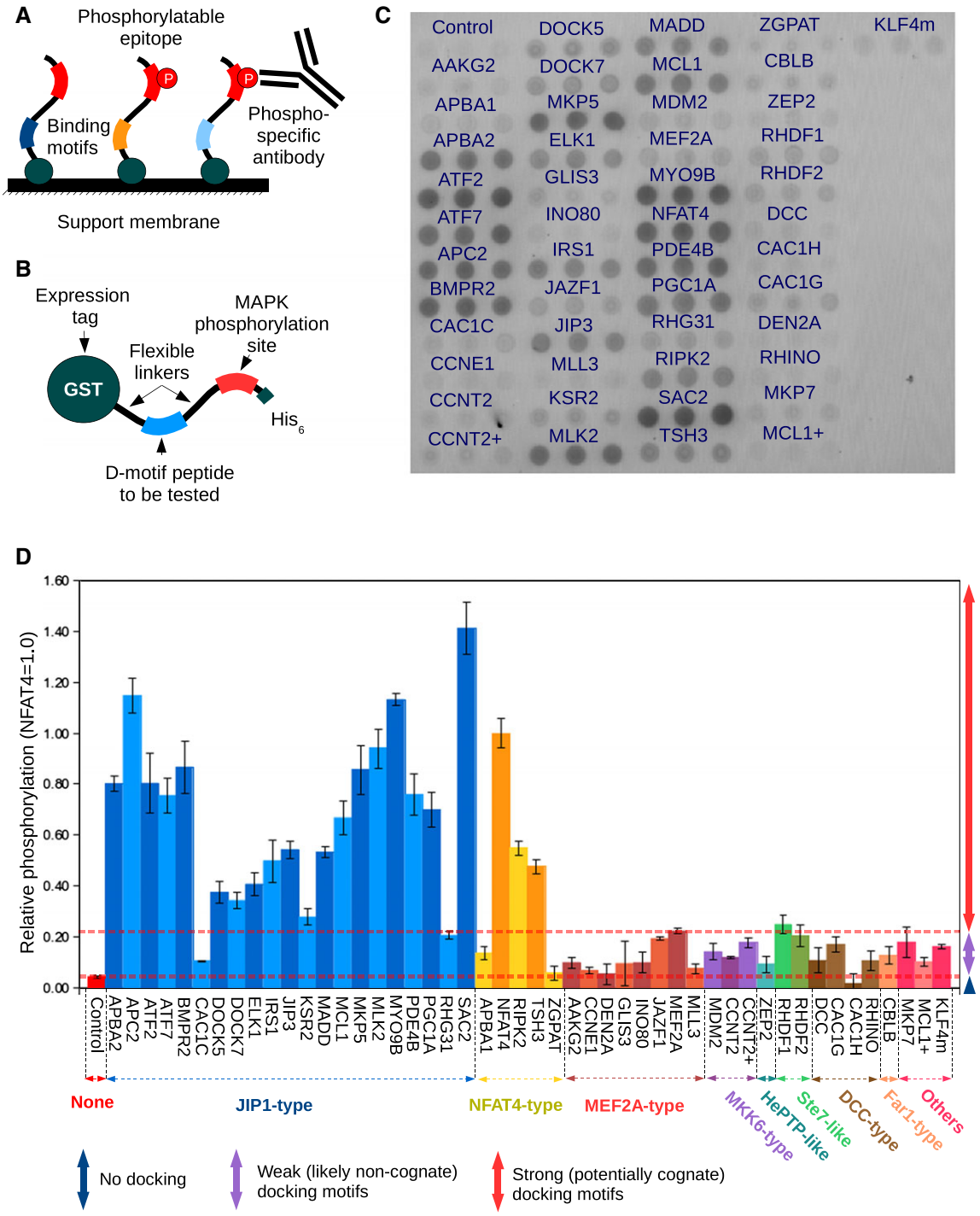

Figure 3. Dot-blot arrays of D-motifs.

A The principle of the phosphorylation enhancement dot-blot arrays. Protein constructs are immobilized onto a solid-phase support where phosphorylation takes place. Afterward, the phosphorylated epitopes are detected through standard Western blot procedures using a phosphorylation-sensitive antibody.

B The schematic structure of the artificial substrates utilized in the dot-blot arrays. All constructs share the same tags, substrate sites, and linkers: Only the docking motif-containing fragments differ.

C A sample dot-blot array for detecting JNK-binding docking motifs. This specific array contains 48 of the 70 motifs tested in total, and was incubated against activated JNK1.

D Quantitative analysis of the sample dot-blot array. All intensities are relative to that of the NFAT4 motif (positive control), error bars were derived from three parallel samples on the same membrane and show standard deviation from the mean (N = 3). "+" denotes additional, non-overlapping motifs tested from the same protein. "m" refers to murine (non-human) sequence. (The corresponding ERK2 and p38α 48 motif arrays and the 70 motif arrays for all three MAPKs are shown on Appendix Fig S1 or on Fig EV1).

Table 1. Validated sequences grouped by D-Motif class.

Greater MEF2A class (phosphorylation by p38α)			Greater DCC class (phosphorylation by ERK2)		Greater HePTP class (phosphorylation by ERK2)	
MEF2A-type	**MKK6-type**	**Misc.**	**DCC-type**	**Far1-type**	**Ste7-like**	**HePTP-like**
AAKG2/PRKAG2 (28–37)	CCNT2 (498–509)	AMPD1 (109–120)	**DCC (1,144–1,155)**	CBLB (489–500)	RHDF1/iRhom1 (11–24)	ZEP1/HIVEP1 (1,422–1,438)
JAZF1 (77–86)	GAB3 (363–374)	AMPD3 (79–90)	CACNA1G (1,030–1,041)	ELMSAN1 (601–612)	RHDF2/iRhom2 (18–31)	
INO80[a] (1,318–1,327)	INO80[a] (1,316–1,327)			TRERF1 (653–664)		
MEF2A (268–277)	KSR2 (330–341)			GAB1 (526–536)		
KLF3 (88–97)	KMT2C/MLL3[a] (1,195–1,206)					
KMT2C/MLL3[a] (1,197–1,206)						
RIPK2[a] (326–335)						
TSHZ3[a] (321–330)						

JIP1 class (phosphorylation by JNK1)	APBA2/MINT2 (279–285)	ATF2 (164–170)	**ATF7 (162–168)**	APC2 (962–968)	**BMPR2 (753–759)**	DOCK5 (1,762–1,768)
	DOCK7 (884–890)	DUSP10/MKP5 (18–24)	**ELK1 (314–320)**	**IRS1 (856–862)**	**JIP3 (203–209)**	M3K10/MLK2 (876–882)
	MADD (809–815)	MCL1 (136–142)	**MYO9B (1,249–1,255)**	PDE4B (72–78)	PRGC1/PPARGC1A (253–259)	SAC2/INSPP5 (1,009–1,015)
NFAT4 class (phosphorylation by JNK1)	**AKAP6/mAKAP (433–440)**	CCSER1 (573–580)	DYH12/DNAH12 (12–19)	FMN1 (672–679)	FHOD3 (506–513)	**JUND (46–53)**
	KANK2 (244–251)	M3K10/MEKK1 (1,077–1,084)	**MABP1 (1,292–1,299)**	**NFATC3/NFAT4 (145–152)**	RIPK2[a] (327–334)	TSHZ3[a] (322–329)

The table lists motifs that tested as positive ("hits") in the dot-blot arrays and it shows the most commonly used names of proteins (when necessary, two variants), together with the position of the core motif—based on the reference isoform featured in UniProt. Names in bold type denote previously known docking motifs, while the names in normal type indicate novel interactors.
[a]Denotes motifs that appear under more than one class as they satisfy multiple consensus sequences.

class, but only to compare it to the other three. This comparison showed that in several cases, positional preferences could be explained on a structural basis (Fig 5B and C).

The structurally consistent PSSM-based search method offered us a unique glimpse into the human MAPK interactome, albeit limited to D-motif-containing proteins. As it included a rather large number of proteins that have little or no formal Gene Ontology (GO) annotation, we decided to annotate the best 100 hits manually, based on UniProt labels, domain composition, and literature (Table EV2). Out of the three classes examined, the JIP1 type had by far the highest number of validated hits. Thus, the predictions for this class were deemed most reliable, shedding some light on the interactome of JNK1 (Fig 6). Among the less surprising categories discovered were the MAPK pathway components themselves (especially at the MAPK kinase kinase [MAP3K] level, as potential feedback elements), several transcription factors and other gene expression regulatory systems, or various ubiquitin ligases. A considerable number of experimentally tested or predicted JNK-interacting proteins have preferentially or exclusively neuronal functions. We predict that the axons, nerve terminals, and dendrites—especially in synapses—contain a high number of specialized JNK-interacting proteins, as do developing neuroblasts and their axonal growth cones (Appendix Fig S5A). Interestingly, the majority of JNK-associating proteins (both experimentally validated and predicted) seem to be involved in cytoskeletal regulation. We encountered numerous actin-binding or microtubule-binding proteins, molecular motors as well as small G protein partners. Docking motifs were even found on proteins localized to centrosomes, basal bodies, or those involved in the formation of primary cilia. Several other high-scoring hits suggest that JNK is intimately involved in the regulation of endo- and exocytosis.

The presence of insulin signaling pathway components in the lists may also explain many previous observations on the causative role of JNK in type II diabetes. This kinase is involved in pathways over-activated by cytokines derived from adipose tissue. JNK1 knockout mice are also known to be resistant to type II diabetes induced by obesity (Hirosumi *et al*, 2002; Sabio *et al*, 2010). Proteins bearing JIP1-type docking motifs (e.g., MADD, IRS1, PGC1A) are located in critical points of networks responsible for insulin signaling, and these are the same pathways that are also targeted by most anti-

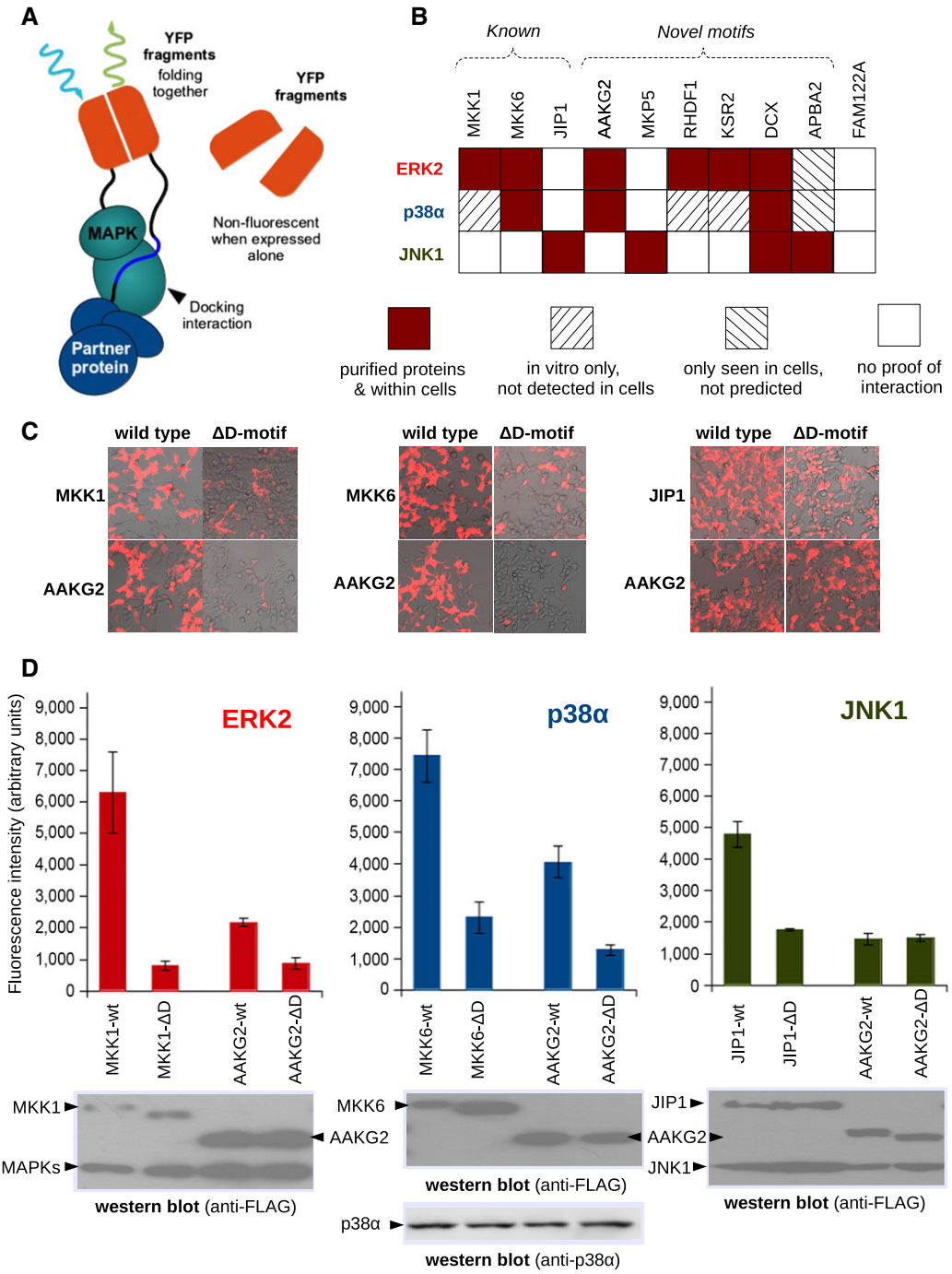

Figure 4. Bimolecular fragment complementation (BiFC) experiments.

A The principle of YFP fragment complementation driven by MAPK–partner protein interaction. The weak and transient interactions between MAPKs and its binding partners still lead to well-detectable signals.

B Summary of the BiFC experiments. In addition to successfully testing six novel docking motif-dependent interactions, three positive controls (MKK1, MKK6, and JIP1) and an extra negative control (FAM122A) were also introduced into this analysis. Red squares indicate positive BiFC results (which were mostly directly predictable from fragment-based experiments). However, some interactions suggested by dot-blot experiments and/or FP titrations were not seen in BiFC (lined squares). These were possibly too weak or absent in the cellular context.

C Bright-field image of transiently co-transfected HEK293 cells overlaid with the fluorescence image. Although expression levels and complementation efficiency vary between cells, ablation of D-motifs results in robust fluorescence intensity changes for known MAPK–partner protein pairs (MKK1-ERK2, MKK6-p38α, and JNK1-JIP1, from left to right, upper panels) similar to a novel MAPK partner (AAKG2, lower panels).

D Results of fluorescence measurements on co-transfected cell populations with positive controls and for AAKG2. (Error bars show standard deviations from the mean, $N = 6$). Similar expression levels of FLAG-tagged proteins (wild-type or D-motif lacking versions of known or putative MAPK partners) or MAPKs were confirmed by Western blotting using anti-FLAG (ERK2 and JNK1) or anti-p38α antibodies. Further BiFC results are shown on Appendix Fig S4.

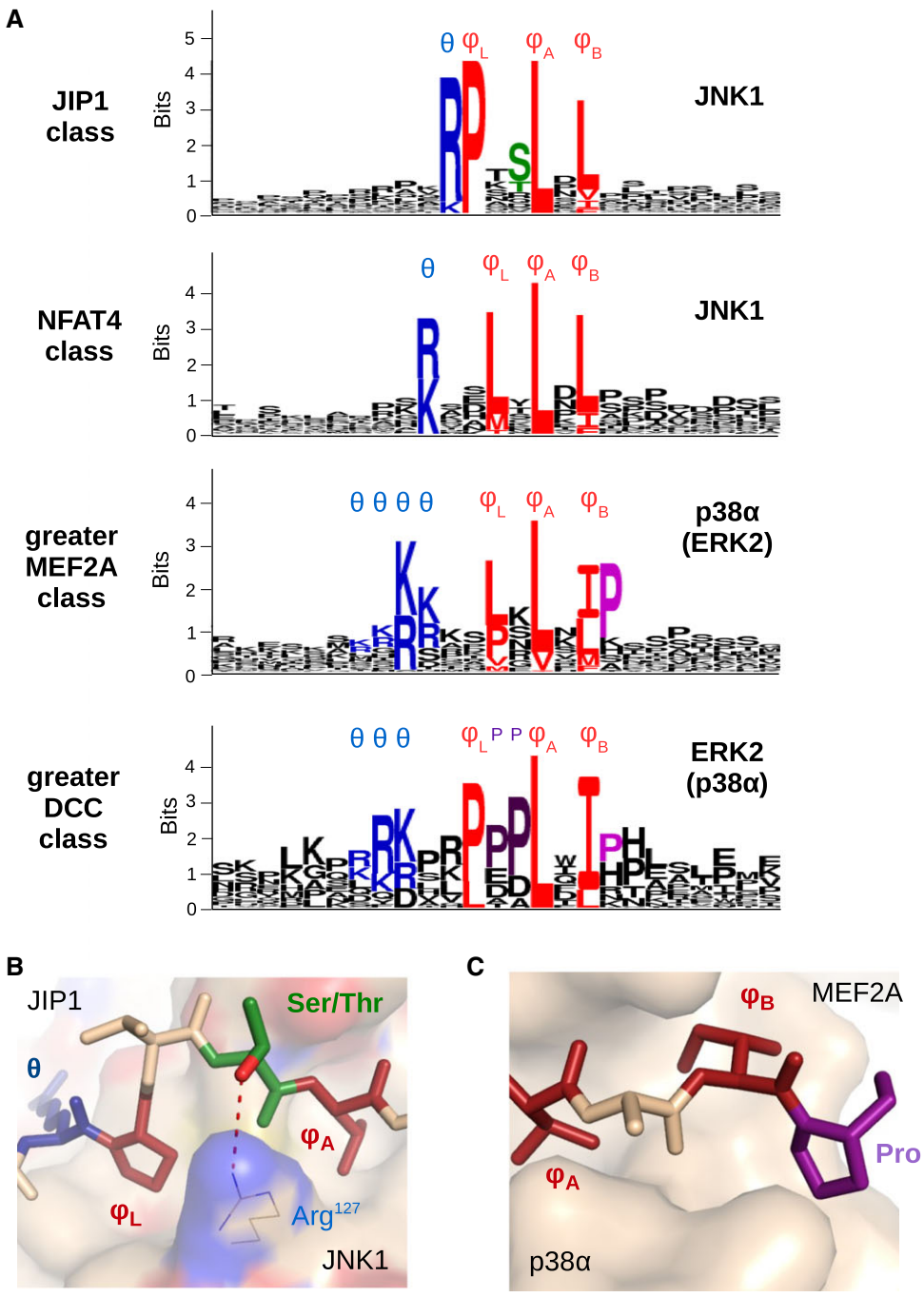

Figure 5. Position-specific scoring matrix (PSSM) logos.

A Sequence logos generated from the evolutionarily weighted PSSMs. These logos were generated for the four classes for which an adequate number of examples was available. In the core motif, the positively charged θ positions are colored blue, while the three φ hydrophobic contact points are red. The JIP1 class and the NFAT4 class were built from motifs binding to JNK1; therefore, they are directly comparable. The greater MEF2A class includes p38α-binding motifs (with a minority also binding to ERK2). At the same time, the greater DCC class contains motifs primarily binding to ERK2 (with many of its members also associating with p38α). The logo of the greater DCC class was built based on only 6 independently evolved proteins (compared to 21 for JIP1, 15 for NFAT4, and 15 for greater MEF2A classes, respectively). As the JIP1-type PSSM contains the highest number of independent examples, it is considered the most reliable, while the DCC is the most coarse.

B Positional amino acid preferences in the PSSM matrix can be explained on structural grounds even at highly variable intervening regions between core motif positions: The JIP1-type motifs frequently have serine or threonine in the position immediately preceding φA (colored in green on the logo). From the structure of the JIP1–JNK1 complex (Heo *et al*, 2004), it is clear that this amino acid has the ability to form a hydrogen bond with the underlying arginine side chain of JNK1.

C The greater MEF2A—and to a lesser extent, greater DCC—motif classes show a clear preference for proline after φB (colored in light magenta). The panel on the MEF2A-p38α complex shows that this proline can form an additional hydrophobic interaction toward the surface of p38α (Chang *et al*, 2002). Thus, the reason for this phenomenon is different from the preference for prolines in the greater DCC class (dark magenta on the logos) where the latter are required to favor a type II polyproline helix (Ma *et al*, 2010).

Figure 6. JNK interactome based on the presence of JIP1-type motifs.

A Low-level functional classification of JIP1-type motif-bearing proteins. This analysis reveals that cytoskeletal regulation is at least as important aspect of JNK1 signaling as control of gene expression. The best 100 hits on the JIP1-type motif list contain many proteins whose docking motif is already known (blue letters), or was validated in our dot-blot essays (red letters). Only categories that contain more than a single protein are shown.

B High-level functional classification of the best-ranking 100 JIP1-type motif. The analysis reveals several major functions associated with JNK kinases. The role of JNK in inflammation, neuronal development, metabolic regulation, or apoptosis is already known from cell-based experiments. Motif search results, in addition to providing the mechanistic basis for these regulatory processes, also suggest novel functions. The full functional clustering of JIP1-type D-motifs along with NFAT4- and MEF2A-type motifs is shown in Table EV2.

diabetic pharmaceuticals (Lee *et al*, 2003; Finck & Kelly, 2006; Olson *et al*, 2008; Li *et al*, 2014) (Appendix Fig S5B).

Experiments with greater MEF2A-type motif-bearing proteins provided another interesting observation. The occurrence of such motifs in proteins (KSR2, AAKG2, and AMPD1,3) with an important regulatory role in the adenosine monophosphate-activated protein kinase (AMPK) pathway implies that this system connects to the p38 and/or the ERK1/2 pathways on very specific points in specialized tissues (Costanzo-Garvey *et al*, 2009; Pearce *et al*, 2013; Rybakowska *et al*, 2014) (Appendix Fig S5C). Interaction with tissue-specific protein isoforms may thus cause highly cell-type-specific regulation by MAPKs.

The analysis of the best 100 hits for the NFAT4 class yielded results similar to the JIP1-type motifs, with some differences. In contrast, members of the greater MEF2A class were markedly dissimilar from those of the JIP1 class. Here, the proportion of cytoskeletal proteins was minimal, while the fraction of transcription factors was considerably higher. Proteins involved in other functions related to gene expression, such as chromatin remodeling or histone methylation, were also present in higher numbers. When comparing distributions of protein functions, the NFAT4 class appeared to lie between the two extremes represented by the JIP1 types (mostly cytoplasmatic targets) and greater MEF2A types (emphasizing nuclear actions) (Appendix Fig S6A). The similarity of NFAT4-type motif-containing proteins to JIP1-type bearing ones is easy to understand: Both primarily interact with JNK1. In certain protein families, one can discover closely related pairs in which one protein contains a JIP1-type docking motif and the other contains a likely independently evolved NFAT4-type docking motif (Appendix Fig S6B). On the other hand, the NFAT4-type motifs are structurally compatible with MEF2A types (unlike JIP1 types); thus, some of the predicted best binders are shared between the latter two lists. Our dot-blot experiments indeed corroborated that the overlapping motif definitions result in a naturally overlapping set of interactors for JNK1 and p38α (Appendix Fig S6C).

Evolutionary analysis of D-motifs

MAPK pathways are found in almost every eukaryotic organism, and the three-tiered kinase cascade architecture of the MAPK module core is well conserved from yeast to human. Therefore, one would naturally expect the downstream targets of these pathways to be conserved as well. However, our results do not support this and in fact suggest the opposite. Although evolutionary conservation is considered to be a predictive feature of a functional linear motif, this did hold through for D-motifs. There was no correlation between FoldX (predicted binding energy) and any of the evolutionary conservation scores. The maximum traceable distance (MTD) of a motif in evolutionarily related species could be calculated from the eggNOG alignments. Here, we also noted that most of the motifs were traceable to vertebrates only. A more thorough search, using p-Blast searches in the UniProt database, revealed that some motifs are actually more ancient than what eggNOG data would suggest. Still, a high number of experimentally validated motifs were found to be relatively recent evolutionary inventions. After mapping the most distant organisms in which the motif in question is already present, we were able to compile an evolutionary histogram of MAPK-docking motif emergence (Fig 7A and Table EV3). Despite the fact that MAPK pathways are an eukaryotic common heritage, very few human docking motifs had an ancestry among unicellular organisms. This latter was only true for the MAPK kinases (MAP2Ks) or MAPK-activated kinases (MAPKAPKs) and a few truly ancient substrates, like MEF2/MADS-box proteins. Only in multicellular animals (Metazoa) did docking motifs become detectable in a variety of phosphatases and MAP3Ks as well as in the core set of mammalian substrates (ELKs, ATFs, JUNs, etc.). However, some of these motifs were difficult to find as they were subsequently lost in several lineages, especially in arthropods. The diversification of docking motifs continued in chordates, but it was in early vertebrates where a major re-wiring and expansion of MAPK partnerships occurred. Over 50% of the motifs identified in our experiments evolved at this period. After the development of bony fishes, motif emergence events became less common, but did not stop completely: New motifs appeared in lobe-finned fishes (Sarcopterygii), in terrestrial vertebrates (Tetrapoda), and even in mammals. Comparison of the known and predicted motifs from the best 100 hits for JIP1-type motifs suggests that there are many more recently evolved motifs in mammals (Fig 7B). These findings are well in line with recent results on yeast calcineurin interactomes: Yeast phosphatase-docking motifs were found to evolve fast, and their

Figure 7. Evolutionary analysis of D-motifs.

A Analysis of MAPK-docking motif emergence (based on p-BLAST searches) paints a dynamic picture of MAPK pathway evolution. The panel was made based on 62 independent MAPK D-motif occurrences (see Table EV3). The histogram counts the number of known D-motifs (blue bars and blue numbers) and those newly identified in our experiments (red bars and red numbers). An approximate timeline is also added to give a realistic scale of the time dimension. The percentage of mammalian motifs found in selected model organisms is also indicated. Among model organisms, mice and zebrafish are relatively similar to human based on their MAPK interactomes. But fruit flies or yeast are rather poor models due to the low number of docking motifs being conserved across species.

B Distribution of the 100 best JIP1-type motifs (predicted by the PSSM) versus their eggNOG-derived maximum traceable distance. Motifs validated experimentally as binders are represented under the green columns, while the total number of predicted motifs is shown in magenta. The analysis suggests that the most recently emerged motifs are still under-explored. While a reasonable percentage of D-motifs shared between humans and zebrafish were successfully validated in experiments, there also appears to be an intriguing number of (predicted) motifs restricted to mammals only. Note that this distance metric is different from the one used in (A) and extends to bony fishes only.

C The branching pattern of closely related human proteins with MAPK-docking motifs points at rapid evolution. Most of the already-established MAPK partner proteins (blue bars) are members of families where more than one paralog carries the same motif. However, the more recently identified docking motifs (red bars) show a rather different picture. Proteins with stand-alone docking motifs thus appear to be much more common than previously expected.

D Comparison of vertebrate and invertebrate genomes suggests that most novel D-motifs may provide paralog-specific regulation. The panel traces the emergence of D-motifs within protein families. Where multiple vertebrate paralogs carry the same motif, the docking elements are overwhelmingly pre-vertebrate inventions (often subject to motif loss, upper rows). However, where only a single paralog has the motif, the trend is exactly the opposite: Most motifs have typically newly evolved and have no counterpart in invertebrates (lower rows).

Data information: * denotes human paralogs containing validated D-motifs in (C) and (D).

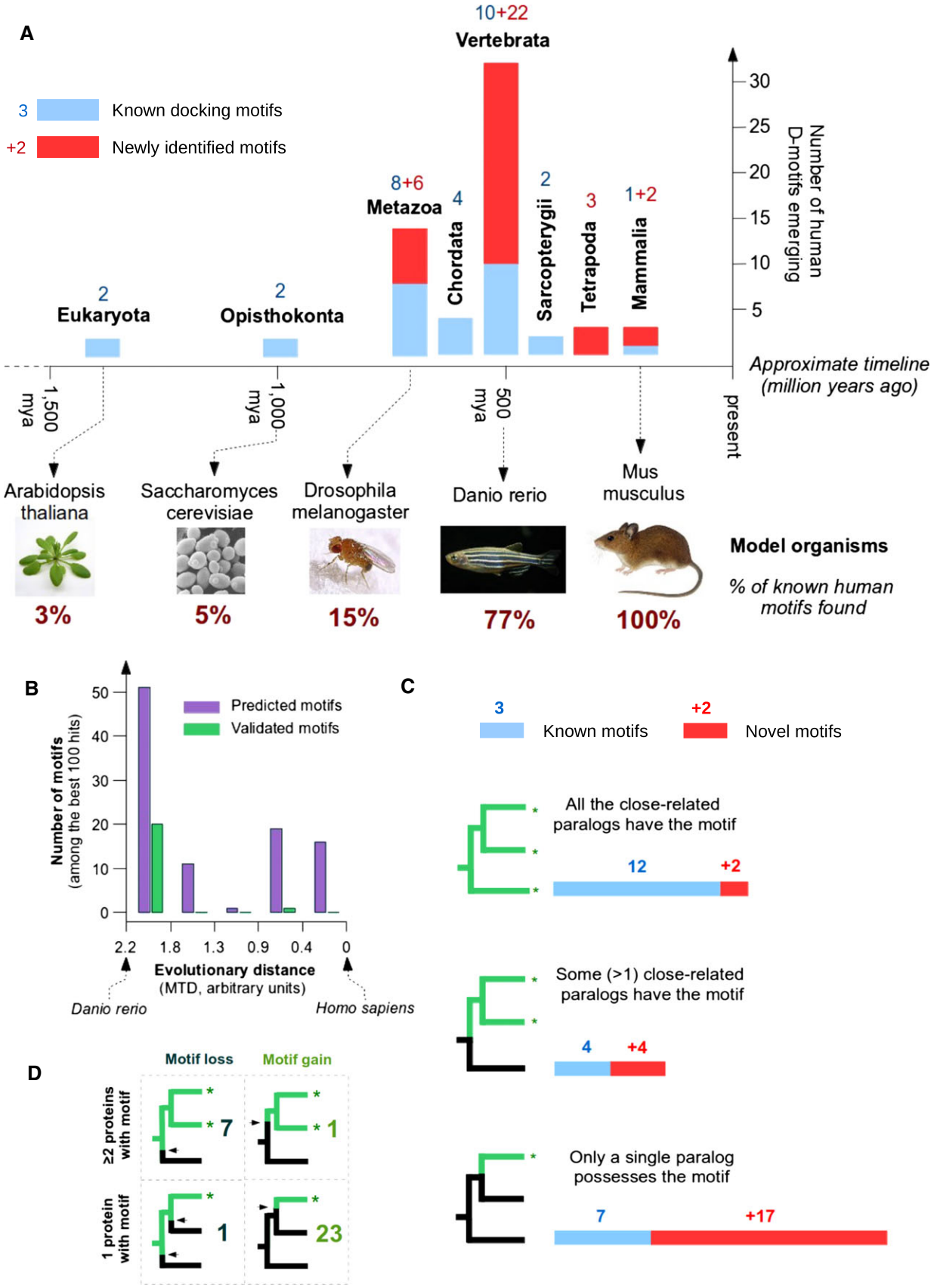

Figure 7.

distribution was highly divergent between related species (Goldman *et al*, 2014).

A further proof for the late evolution of MAPK partnerships is found when comparing paralogs (Fig 7C). These latter are closely related copies of the same ancestral gene that often preserved linear motifs from before their split. Most vertebrate proteins come in groups of 2, 3, or 4 closely related paralogs due to twin genome duplications—and subsequent gene loss—at the dawn of vertebrate evolution (Ohno, 1993; Durand, 2003). Interestingly, most of the better-known MAPK target proteins possess a D-motif in more than one vertebrate paralogs. However, the same is not true for the majority of novel partner proteins. Comparison of vertebrate proteins with those from earlier-branching genomes also helped us to determine whether a motif developed after the gene duplications or before (Fig 7D). Our analysis suggests that the presence of a motif in more than one paralog is predictive for ancient motif emergence. (Over 50% of such protein families have non-vertebrate members with the motif already in place.) In this case, motif loss appeared to be the dominant mechanism to create differences between vertebrate paralogs. Only a very few new motifs emerged in-between the two whole-genome duplication events, suggesting that this evolutionary stage was short-lived (Kuraku *et al*, 2009). On the other hand, where only a single paralog contained the motif, this motif was predominantly a new invention after the twin duplications—and not a result of an ancient motif being lost. This was the most common scenario for newly found D-motifs. Many of these novel MAPK-recruiting motifs are suspected to provide a paralog (or even isoform)-specific regulation, thereby offering unique roles to otherwise highly similar human proteins.

Having obtained a sufficient number of experimentally verified examples, we could also test some theories on the evolutionary processes creating the linear motifs. The motifs we validated (at least at a fragment level) could be classified based on their predicted origins (Fig 8 and Table EV3). Not surprisingly, the most common way of motif emergence appeared to be random mutations in an already-existing disordered segment. This can be illustrated by a known interactor, the Smoothelin-like protein 2 (SMTL2) (Gordon *et al*, 2013). Here, gradual sequence changes in terrestrial vertebrates led to the creation of the motif, which is restricted to placental mammals (Eutheria) (Appendix Fig S7A). There were also several examples for creation from scratch (i.e., from non-coding DNA). This could mean either translational start shift (translating an earlier non-translated 5′ UTR) or splicing site shift (leading to exonization of intronic sequences). While the N-terminal expansion of the protein is seen in MCL1 (as the motif-bearing segment has no counterpart in Bcl2 or in any other related protein), another newly identified partner, KSR2, serves as an example for splicing site rearrangements (Appendix Fig S7B). Here, the paralog KSR1 retains the ancestral intron–exon boundaries, which appear to have shifted in KSR2 (Appendix Fig S7C). We could even find examples for proteins where this mechanism may still be active: The motif can (in an isoform-specific way) be included or excluded due to alternative splicing or initiation. This is the case with the PDE4 genes, where most paralogs (PDE4A, PDE4B, PDE4D) still retain an ancestral, alternative exon containing a JIP1-type motif (Appendix Fig S7D).

Interestingly, linear motifs can also transmute into each other: Some examples in the dataset show potential switching between different MAPK-docking motif classes. As a result, distant organisms

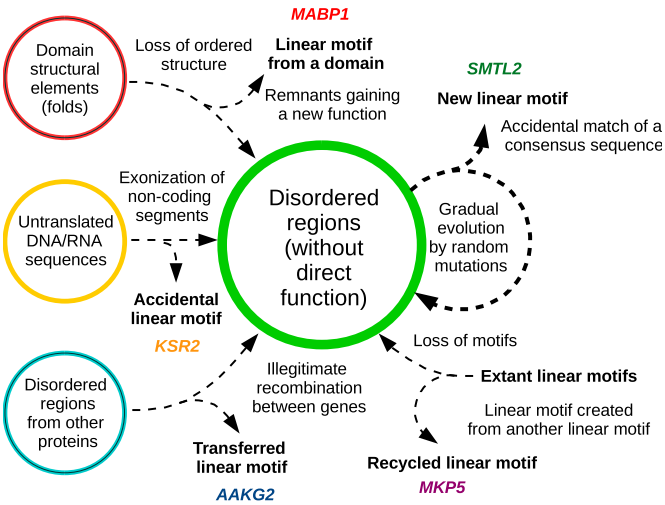

Figure 8. Mechanisms of binding motif emergence.
D-motifs have diverse evolutionary origins. They most commonly originate from disordered protein segments via gradual accumulation of point mutations (e.g., SMTL2). However, they may also be created from folded domains (e.g., MABP1), or being introduced entirely *de novo*, from a previously untranslated genomic segment (e.g., KSR2). D-motifs can also be born from existing linear motifs, through gene fusion or recombination (e.g., AAKG2). But more commonly, this involves transmutation of a (previous) linear motif into a new one (example: MKP5). For more details, see Table EV3 and Appendix Fig S7.

may show different motif types at the same location: That is, the JIP1-type motif that we identified in MKP5 corresponds to an NFAT4-type motif in distant organisms (as in arthropods). In contrast, CCSER1 has an NFAT4-type motif in humans, but a JIP1-type one in zebrafish (Appendix Fig S7E). The motif in ELK1 is of the JIP1 type in humans, but Far1 type in protostomes. Also, the motif of TAB 1 is DCC-like in most non-vertebrate organisms, unlike the human which is MEF2A-like (Appendix Fig S7F). In some cases, a "horizontal motif transfer" (i.e., recombination between unrelated genes) may have complemented *de novo* emergence of motifs. This was the likely case for AAKG2, and the N-terminal region of AAKG2 (which is unique and sets it apart from AAKG1) showed a surprising similarity to the C-termini of MEF2A or MEF2C. In addition to the core motif, the disordered segments flanking the motifs also aligned well, and this cannot be explained by convergent evolution alone as the latter regions are not subject to the same selection (Appendix Fig S7G). The creation of a new linear motif from the unfolded remnants of earlier structured domains was yet another intriguing possibility. For the WDR62/MABP1 family, the duplication of WD40 repeats and their subsequent degeneration were the most likely source of the NFAT4-like D-motif (Appendix Fig S7H).

Functional aspects of docking motif evolution

The typical purpose of docking motifs is to enable phosphorylation of recruited substrates (Reményi *et al*, 2006). Indeed, when the newly identified D-motifs from AAKG2 and DCX were mutated, their phosphorylation by their cognate MAPK was greatly reduced (Appendix Fig S8). Due to the lack of strict spatial constraints between D-motifs and phosphorylation sites, such roles can only be interpreted in the broader context of a protein. Unfortunately, most

phosphorylation target sites controlled by the novel docking motifs remain elusive. Yet in some cases, such sites either have been discovered beforehand or could be inferred based on spatial proximity and/or coevolution with D-motifs (Gordon *et al*, 2013). Examples for the latter give insight into how docking motifs emerged in relation to their target sites. In particular, both motif loss and gain are expected to have a profound effect on target sites: potentially endowing the protein with a new regulation. This is well illustrated by three cases: the nuclear factor of activated T cells (NFAT) family, with a *de novo* motif created in NFAT4 (adding on to a preexisting target site, Fig EV4A); the myocyte enhancer factor 2 (MEF2) family, displaying motif loss to a varying degree (with the concomitant loss of target sites, Fig EV4B); and the Grb2-associated binder (GAB) family, in which both events took place (Fig EV4C) (Chow *et al*, 1997; Yang *et al*, 1999; Wolf *et al*, 2015). To examine the evolutionary relationship between MAPK-binding D-motifs and phosphorylation target sites, the presence of validated D-motifs was analyzed in parallel with experimentally validated S/TP phosphorylation sites (Table EV4). This analysis was carried out in 50 proteins and compared the evolutionary conservation depth of the human motifs among vertebrate homologs (Appendix Fig S9). We found that there is a correlation in conservation of D-motifs and the most conserved putative phosphorylation sites, suggesting that D-motifs and putative target sites may have coevolved.

Our studies also support the notion that there are a number of proteins with multiple MAPK-docking elements. Apart from the case where these elements interact with different MAPKs, such as in the case of MKP5, where a rhodanese domain can bind p38 and a JIP1-type linear motif interacts with JNK, or GAB3, with separate motifs to recruit ERK1/2 or p38, the purpose of multiple D-motifs is unclear, especially because they tend to bind to the same surface and thus compete with each other for MAPK binding. Such domains or motifs are often not simple duplicates of each other and emerged independently during evolution. This happened in the case of BMPR2, where the first JIP1-type motif is found in almost all multicellular animals, but the second one is restricted to vertebrates (Podkowa *et al*, 2010) (Appendix Fig S10A). In addition, the ATF2/CREB family of transcription factors has an N-terminal motif next to a Zn-finger serving as the primary docking element in all metazoans, but several vertebrate paralogs have an additional JIP1-type motif with an unclear role (Appendix Fig S10B). As one MAPK molecule can only accommodate one motif at a time, it is probable that multiple docking motifs would allow several, mutually largely exclusive complexes—each with unique spatial orientation. As in the case of MKK7 which activates JNK1, the precise orientation of the MAPK versus the partner protein might have important implications on phosphorylating specific target sites (Ho *et al*, 2006; Kragelj *et al*, 2015).

D-motifs in proteins are known to facilitate phosphorylation of MAPK targets. However, the functional consequences of MAPK target site phosphorylation are unfortunately often not well understood. Phosphorylation may have diverse impact on protein function, and its relevance may only be revealed in the context of a signaling cascade. To this end, we have characterized the role of one of the newly found D-motif (527-RKVKPAPLEI-536) in the GAB1 signaling adapter protein (see Table 1 or Fig EV2, Far1-type) in HEK293 cell-based assays. GAB1 belongs to the insulin receptor substrate 1 (IRS1) family of adapter/scaffolding molecules playing a role in multiple signaling pathways (Holgado-Madruga *et al*, 1996). Recently, it has been shown that cell membrane recruitment of GAB1 via its PH domain is controlled by an intramolecular switch (Wolf *et al*, 2015). The PIP3-binding surface of the GAB1 PH domain is masked by intramolecular interactions and phosphorylation at Ser551 unmasks this membrane binding surface and promotes recruitment of GAB1 to the cell membrane (Appendix Fig S11A). Interestingly, Ser551 in human GAB1 is a known MAPK S/TP target site, and we tested whether the newly found D-motif in this scaffold protein is indeed important for GAB1 membrane recruitment (Appendix Fig S11B), and whether it influences EGF/Ras→ERK2 signaling (Appendix Fig S11C). We found that docking motif versions of GAB1 had diminished capacity to translocate to the cell membrane and that these mutants were more sensitive to EGF stimulation regarding ERK2 activation. These results are fully consistent with GAB1's role as a complex regulator of EGF-mediated signaling: It exerts negative feedback control on the EGF/Ras→ERK2 pathway, presumably by relying on its ERK2-binding D-motif.

Discussion

Protein kinases often use dedicated domains for substrate recognition. Known examples include the Src-family tyrosine kinases (SH2 and SH3 domains) (Alexandropoulos & Baltimore, 1996; Pellicena *et al*, 1998), SPAK/OSR kinases (unique domain) (Vitari *et al*, 2006) and Polo-like kinases (Polo-box domains) (Lee *et al*, 2014). In other cases, recruitment is provided by the catalytic domain itself, but by a distinct surface which is noncontiguous with the catalytic site. This appears to be common among relatives of MAPKs, the so-called CMGC (cyclin-dependent kinase/MAPK/glycogen synthase kinase 3/CDK-like) kinase group. However, each kinase family uses a different surface, with strikingly different recognition modes. Thus, motifs recognized by GSK3 or SRPK kinases (Dajani *et al*, 2003; Ngo *et al*, 2005) are unrelated to D-motifs or FxFP-motifs of MAPKs, or to CDK-docking motifs recognized by the cyclin subunit, for example (Lowe *et al*, 2002). Based on our results on MAPK-binding D-motifs, it may be anticipated that insights into other recruitment motif-based systems will greatly contribute to a system-level understanding of protein kinase-based intracellular signaling networks.

In the current study, it was demonstrated that canonical, D-motif-dependent partners of MAPKs are in fact quite common. However, a number of partners with atypical or "naturally defective" docking motifs do exist (e.g., MKK3, MKK7, TAB 1), and these are difficult to predict (Chang *et al*, 2002; De Nicola *et al*, 2013). Often such defective motifs act in a non-autonomous way: These weak elements may be complemented by additional protein stretches, motifs, or domains (Glatz *et al*, 2013). Besides, not all MAPK-binding elements are linear motifs. Folded domains such as the rhodanese domain of dual-specificity phosphatases may bind to the same site as intrinsically disordered docking motifs (Zhang *et al*, 2011). It should be noted that motifs other than the canonical D-motifs (e.g., the so-called FxFP-motifs) also exist (Jacobs *et al*, 1999; Fantz *et al*, 2001; Zhang *et al*, 2003). A considerable number of interactions might also be indirect, mediated by a third partner. Nevertheless, directly interacting with a MAPK solely through short linear motifs appears to be a major and widespread phenomenon in mammals. Although experimental testing of all putative MAPK D-motifs could not be performed, we suggest that the fraction of the

human proteome that harbors high-scoring D-motifs may be representative of the full interactomes for three distinct MAPKs, which may be best captured for JNK1 by the procedure presented in this study.

Some of the newly identified partners directly fit into the core of MAPK pathways. These include specific phosphatases as well as MAPK kinase kinases (MAP3Ks). While there can be little doubt that docking motifs of phosphatases would be required for MAPK dephosphorylation, the presence of docking motifs in MAP3Ks is a more intriguing observation. It is probable that phosphorylation of proteins acting on the MAP3K level (like on MEKK1, MLK1/2, or KSR2) would allow direct feedback control of MAPK pathways (Flotho *et al*, 2004). However, the majority of novel hits appear to lie outside the core MAPK pathway module, and these are probably simple downstream elements (i.e., substrates). Most of the novel proteins are expected to be either direct MAPK substrates or scaffold proteins (i.e., enabling phosphorylation of indirect MAPK substrates through protein complexes) (Fig EV5).

The wide distribution of D-motifs in a functionally diverse set of proteins explains how MAPKs can regulate such a broad spectrum of physiological processes. Previously, their specific regulatory roles were often attributed to single target proteins. For example, the role of JNK in axonal growth was attributed to the JNK–JIP1 interaction, and the association of JNK with diabetes was attempted to be explained by the JNK1–IRS1 interaction alone (Lee *et al*, 2003; Dajas-Bailador *et al*, 2008). In contrast, our results imply that these interactions may only be two examples of a substantially more complex protein network and JNK (as all MAPKs) connects to its targeted physiological systems by a large number of direct interactions. While individual connections might not be stable (especially in an evolutionary sense), multiple specific linkages could provide the key mechanism for a robust and adaptable physiological regulation.

Surprisingly, many of the newly implied MAPK partners have a restricted expression pattern enabling fine-tuned regulation in specialized tissues. Because of the latter phenomenon, a great deal of these interactions is unlikely to be discovered by large-scale protein–protein interaction screens. Easy-to-handle cell lines and mass-spectrometry-based analyses provide a powerful tool, but not for proteins that are only expressed in special, differentiated tissues (e.g., AAKG2, which is only abundant in cardiomyocytes) or restricted to certain embryonic developmental stages (e.g., DCX is almost exclusively expressed in developing neuroblasts) (Lang *et al*, 2000; Brown *et al*, 2003). Here, a modeling-driven interactome search is the most suitable tool to fill in the gaps in our knowledge. In addition, a reliable sequence-based prediction procedure sets the stage for easy examination on how MAPK signaling partners changed over time during evolution.

A comparative analysis suggests that rapid changes happened to MAPK pathways during the early evolution of vertebrates. Target proteins could have been brought under tight MAPK control by simply introducing docking motifs; however, this also necessitated target sites where phosphorylation could elicit functional effects. Similarly, existing links could have been thrown away by the loss of docking mechanisms. These processes were apparently the fastest when early vertebrates diverged from other chordates. Since then, the rate of motif emergence seemingly slowed down, but it has not completely stopped (some novel human motifs could be traced back only to mammals).

The current study explored MAPK–partner protein interactions in the human proteome. The identified D-motifs have the capacity to bind specific MAPKs *in vitro* and likely also in cells and in full protein context. Their functional relevance, however, remains largely unexplored. Undoubtedly, further studies will be required to address the relationship between these physical recruitment sites and MAPK phosphorylation target sites to understand how MAPK-mediated phosphorylation could elicit specific regulation. Nevertheless, this study suggests a rich and dramatically fast-evolving landscape for short recruitment sites and helps to explain how MAPKs could have become such widespread regulators of cellular physiology.

Materials and Methods

Motif scan and filtering by ANCHOR

Putative MAPK-binding D-motif instances were collected from the human proteome. Protein sequences were downloaded from the reviewed section of the UniProt database. The resulting Human Proteome Database (HPD) contains 20,248 sequences. The HPD was scanned for motif hits with basic pattern matching using the regular expressions of seven different D-motif classes/types (see Fig 1 and Appendix Fig S2), yielding 87,857 hits (JIP1-class, NFAT4-class, MEF2A-class-MEF2A-type, MEF2A-class-MKK6-type, DCC-class-generic, HePTP-class-Ste7-type, HePTP-class-HePTP-type). These hits were filtered using specific bioinformatics predictors aiming at selecting for biologically relevant instances. The filtering procedure followed a stepwise fashion as outlined below. Step 1: Filtering for the tendency to be part of disordered region that can undergo a disorder-to-order transition. The estimation of the interaction potential of the selected protein regions was done with the ANCHOR algorithm, a method trained to recognize binding regions in disordered protein segments (Mészáros *et al*, 2009). As described earlier, in linear motif selection, a more permissive version of ANCHOR can be used; therefore, the default 0.5 cutoff value was lowered in an adaptive way so that at least 90% of the known 47 formerly known D-motifs are retained (Mészáros *et al*, 2012). Motif hits were kept only if they overlapped with either an ANCHOR region predicted by using a 0.4 cutoff, or an ANCHOR region predicted by using a 0.3 cutoff, but in this case at least one of the 20 residue flanking regions of the motif hit had to have a sufficiently high average disorder value (above 0.45) predicted with IUPred (Dosztányi *et al*, 2005). As a result, the number of hits was reduced to 21,201. Step 2: Filtering for intracellular accessibility. Motif hits were discarded if they resided in proteins that were predicted to have a signal peptide by SignalP, and if they were also predicted not to have a transmembrane region predicted by Phobius (Käll *et al*, 2007; Petersen *et al*, 2011). These motif hits were predicted to be localized outside of the cell, which is incompatible with MAPK binding. Phobius alone was also allowed to predict signal peptides alone, if SignalP score were not too low (above 0.3). If a motif instance resided in a protein that was predicted to have a signal peptide but it also had at least one transmembrane region, the localization of the motif region was further checked. If it was entirely intracellular, it was kept, otherwise discarded. This filtering step reduced the number of motif hits to 18,952. Step 3: Filtering for correct localization. All hits that

were predicted by WoLF PSORT to be extracellular (with score ≥ 25), membrane protein (with score ≥ 25), localized to the E.R. (with score ≥ 15) or the Golgi (with score ≥ 9) were filtered out, unless they harbored transmembrane regions, and the region containing the motif was predicted to be localized in the intracellular space (Horton *et al*, 2007). There were 18,637 hits remaining after this step. Step 4: Filtering for structural accessibility. Motif hits that were determined to reside in Pfam domain regions were discarded (Finn *et al*, 2014). Some hits were also discarded in a manual curation process if they were located in Pfam Family/Repeat/Motif regions likely to have a stable structure in isolation. Furthermore, motif occurrences that overlapped with coiled-coil regions predicted by COILS were removed as well (Lupas *et al*, 1991). Finally, there were 14,062 motifs remaining for further analysis including more than 90% of the known positives. Motifs passing all filters together with known positive hits are listed in Table EV1.

Compared to classical D-motifs, a "mirror image" like orientation for certain MAPK-binding motifs has also been described. In these "reverse docking" (revD) motifs, the hydrophobic stretch is located N-terminally to the charged residues. Apart from a short revD motif in PEA-15, all other mammalian examples are from a single group of MAPK interactors, the RSK/MAPKAPK family of kinases (Mace *et al*, 2013; Alexa *et al*, 2015). The low number of known reverse docking motifs, however, precluded their analysis in any systematic way.

Scoring for structural compatibility by FoldX

Complexes of peptides with a MAPK were modeled using FoldX, similarly as it was previously described for SH2-binding peptides (Schymkowitz *et al*, 2005; Sánchez *et al*, 2008). For some motif classes, a reliable crystallographic model was available because the JNK1-pepJIP1 and JNK-pepNFAT4 complex crystal structures contained full consensus motifs (Heo *et al*, 2004; Garai *et al*, 2012). For p38α-pepMEF2A, p38α-pepMKK6, and ERK2-pepDCC crystal structures, the N-terminal positively charged amino acids binding in the CD groove were not located in the electron density, so for these cases the modeling was done only for the hydrophobic portion of the motif (Chang *et al*, 2002; Ma *et al*, 2010; Garai *et al*, 2012). The ERK2-pepHePTP complex contained artificial disulfide bond at the Φ_B position that facilitated crystallization of this protein–peptide complex; however, it also distorted the conformation of the docking peptide at the C-terminal region (Zhou *et al*, 2006). This was corrected by remodeling this part of the peptide by FlexPepDock. The model for the hypothetical Ste7 model was constructed from the yeast Fus3-pepSte7 and Fus3-pepMsg5 structures, superimposed on human ERK2, and then optimized with FlexPepDock (Reményi *et al*, 2005; London *et al*, 2011). List of the PDB files used to generate the various MAPK-docking motif structural templates are listed in Appendix Table S1. Each model contained the core motif with one extra alanine on both sides. The final FoldX score took into account the estimated energy of the complex as well as peptide stability in solution.

Automated evolutionary sequence conservation analysis

All motifs were checked for evolutionary conservation. Sequence conservation, relative motif sequence conservation with respect to its flanking regions, and the maximum traceable evolutionary distance of the motif (maximum distance between species bearing the same motif in closely related proteins) were all calculated from databases with pre-computed alignments. This was necessary to be able to compare the conservation of novel hits versus known motifs. For each protein with a potential motif, a cluster of orthologous proteins was extracted from the eggNOG database, using all vertebrates as the reference set of species (Jensen *et al*, 2008). Additionally, homologs from the inParanoid database were considered (Berglund *et al*, 2008). Here, the reference set of species consisted, in addition to human, of *Pan troglodytes*, *Mus musculus*, *Gallus gallus*, *Xenopus tropicalis*, *Danio rerio*, *Ciona intestinalis,* and *Brachiostoma floridae*. For each extracted cluster, only those sequences were retained that contained a motif instance within 10 or 50 amino acids compared to the human motif occurrence in full-length alignments in the eggNOG and inParanoid clusters, respectively. The regions containing the motif with its 10 amino acid flanks on either sides were retained and realigned using MUSCLE (Edgar, 2004).

PSSM building, sequence logos, and final scoring

Position-specific scoring matrices for JIP1, NFAT4, greater MEF2A, and greater DCC classes were built including formerly known and newly found, validated human motif instances as well as all their identifiable vertebrate orthologs. To increase the sequence space coverage, we included more than just (known or novel) human MAPK-docking motifs. A method was devised to use evolutionarily weighted sequences for each independently evolved (or sufficiently unique) motif and to collect all known vertebrate orthologs. For this purpose, alignments were built from vertebrate proteins obtained by BLAST searches. Based on the refined consensus, motifs were classified as either potentially functional or non-functional. The motifs deemed potentially functional were realigned (with no gaps allowed) to the original sequence. In the end, the sequences were weighted by their evolutionary distance (based on the phylogenetic distances of the corresponding vertebrate organisms) and the final frequencies were obtained by summing up all independent groups with equal weights. In PSSM, each row represents one of 20 possible residues, and each column represents a position in a motif. Thus, the score for residue X at position i is defined in the following way:

$$X_i = \frac{\sum_s (w_s \times I(s_i = X)) + p \times X_b}{\sum_s w_s + p},$$

where s is a peptide sequence, w_s is the weight of that sequence based on the species from which it stems, I is the indicator function which is 1 when its argument is true and 0 otherwise, p is the pseudo-count defined as square root of total number of training peptides from the class and it is used to account for residues that do not appear at position i, and X_b is the background frequency of the residue (based on UniProtKB/Swiss-Prot Release 2013.05). For computational efficiency and to account for background frequencies of residues, log-odds scores of X_i were used in the form $X_i = \log(X_i/X_b)$. The final score was calculated as the sum of the log-odds scores of individual positions.

The distance between species A and B was the direct sum of branch lengths leading from A endpoint to B endpoint along the phylogenetic tree. These distances were the same that appear in

PhyloWidget visualization of Ensembl data (Jordan & Piel, 2008). In case of species not included there, evolutionary distances were extracted from trees published in literature and numerically re-calculated to fit the scaling of Ensembl-derived distances wherever possible (Barley *et al*, 2010; Agnarsson *et al*, 2011; Betancur-R *et al*, 2013). In the remaining few species where the tree topology was available without exact distance metrics, a numeric interpolation with equal weights was used.

Sequence logos were constructed using Seq2Logo-2.0 (Thomsen & Nielsen, 2012). Height of residue X at position i is directly proportional to its PSSM score X_i (with $p = 0$) and R_i, information content of position i, defined as:

$$R_i = \log_2(20) - (H_i + e_n),$$

where e_n is small sample correction parameter, expressed as 19/(2 * ln(2) * number of peptides), and H_i is uncertainty at position i, defined as:

$$H_i = -\sum X_i \times \log_2(X_i).$$

Receiver operating characteristic (ROC) curve was constructed by adding simulated negative cases: Peptides in the human proteome conforming to the respective D-motif consensus but lying in a Pfam A-structured region were scored together with validated true-positive motifs, and at every value of the true-positive rate, the corresponding value of the false-positive rate was calculated and plotted (Finn *et al*, 2014). The area under the ROC (AUC) curve was calculated to assess the quality of the prediction: the closer it is to 1, the better is the predictor, and 0.9 indicates a very good predictor. For each motifs class, AUC calculation was performed in fivefold cross-validation setting with 100 samples, selecting 4 folds of D-motifs for training and 1 remaining fold, together with the D-motifs from Pfam A domains, for testing, while not allowing motifs from the same vertebrate orthology set to be used for both training and testing.

Analysis on D-motif and phosphorylation site conservation

Consensus MAPK phosphorylation sites from proteins, which contain an experimentally verified D-motif (Table EV3), were collected from PhosphoSitePlus database (Li *et al*, 2002). Phosphorylation sites was selected if it was more than 10 amino acids away from the core D-motif and if there was site specific or mass-spectrometry experimental evidence of phosphorylation at that site in human. Normalized average traceable evolutionary distance for these sites from eggNOG vertebrate alignments has been estimated as the sum of the weights of the organisms where the phosphorylation site was present in the homologous protein divided by the sum of all the weights from the organisms in the alignment (Table EV4). If multiple candidate MAPK phosphorylation sites were present in the protein, the site with maximum normalized average traceable evolutionary distance was selected.

Protein expression and purification

Activated ERK2, p38α, and JNK1 were produced by co-expressing the MAPKs with constitutively active forms of GST-tagged MAPK kinases (MKK1, MKK6, and MKK7, respectively) in *E. coli* (Rosetta pLysS). Full-length human MAPKs and their corresponding activator kinases were encoded in bicistronic vectors allowing the expression of hexa-histidine-tagged, phosphorylated MAPKs. Proteins were purified similar to dephosphorylated MAPKs as described earlier, using a Ni-column affinity and an ion-exchange step (Garai *et al*, 2012). The double-phosphorylated state of activated MAPK samples was confirmed by using anti-phospho MAPK antibodies with Western blots. Fragments corresponding to the D-motifs of various proteins (with appropriate flanks) were reverse-translated to synthetic DNA fragments endowed with sticky BamHI- and NotI-compatible ends. These were annealed, treated with PNK (polynucleotide kinase), and ligated into a customized pET-17 vector (pAZAD) containing flexible linkers (known to be intrinsically disordered) derived from ATF2 (Nagadoi *et al*, 1999). The linkers were incorporated directly between the GST-tag and the D-motif cassette as well as between the docking cassette and the target site. GST fusion constructs contained an ATF2-derived phosphorylatable epitope, lacking the T69 site, with only the T71 site present (ADQAPTPTRFL). All constructs were checked by DNA sequencing. The resulting GST-tagged artificial substrates with a C-terminal hexa-histidine-tag were expressed in *E. coli*, purified first on Ni-Sepharose and on glutathione-Sepharose by affinity chromatography. To prevent degradation of unstructured protein segments, lysates were treated with protease inhibitor cocktails (cOmplete EDTA-free inhibitor cocktail tablet, Roche), phenylmethanesuphonyl fluoride (PMSF) and benzamidine. All buffers following Ni-affinity purification contained PMSF (0.4 mM), benzamidine (2 mM), and EDTA (1 mM). Since detergents might influence immobilization efficiency, care was taken to use the same detergent, BOG (Octyl-β-D-glucopyranozide), in all final solutions (including dilutions) at the same concentration (0.02%).

In vitro dot-blot phosphorylation assay

Stocks containing the purified artificial substrates were diluted to equal concentrations (~1 mg/ml), and printed (1 μl) as triplicates on nitrocellulose membrane using a pipetting robot (Hamilton STARlet). Membranes were dried at room temperature for at least 1 h. Prior to phosphorylation, membranes were blocked in Tris-buffered saline and Tween-20 buffer (TBS-T) containing 3% bovine serum albumin (BSA) for 30 min and washed three times with TBS-T. Phosphorylation was performed in kinase buffer containing 50 mM TRIS–HCl (pH = 7.5), 10 mM $MgCl_2$, 2 mM DTT, 0.1% BSA, and 2 mM ATP. Activated MAPKs were applied in 100–300 nM concentration. The phosphorylation solution was pre-mixed before applying onto the membrane. The reaction took place at room temperature on a rocker, it was stopped after 10 min by the addition of EDTA (at 25 mM end concentration), and membranes were then washed three times with TBS-T. Thereafter, membranes were blocked again by 3% BSA in TBS-T and developed by standard Western blot techniques using an anti-phospho-T71 ATF2 antibody (Cell Signaling Technology, #9221S) at 1:1,000 dilution and a secondary anti-rabbit antibody (Cell Signaling Technology, #7074S) at 1:2,000 dilution. After development with the Immobilon ECL kit (Millipore), phosphorylation signal was read either by luminescence (Alpha Innotech gel documentation system) or by fluorescence (Typhoon Trio+ scanner, GE). Non-phosphorylated membranes

were also checked for protein immobilization efficiency. The C-terminal hexa-histidine epitopes of GST phosphorylation reporter constructs were detected by a standard anti-His6 Western blot. Dot-blot experiments were performed for each construct at least twice, using different protein stocks. Only those constructs that consistently performed in all experiments above the non-cognate control were regarded as positive.

In vitro protein–peptide binding assays

For fluorescence polarization (FP)-based binding affinity measurements, known MAPK-docking groove-binding peptides were N-terminally labeled with carboxyfluorescein or carboxytetramethylrhodamine (for ERK2, pepRSK1: PQLKPIESSILAQRRVRKLP STTL; for p38α, pepMK2: IKIKKIEDASNPLLLKRRKK or pepMEF2A: SRKPDLRVVIPPS; for JNK1, pep JIP3: RKERPTSLNFPL). 50 nM labeled reporter peptide was mixed with MAPKs in a concentration to achieve ~60–80% complex formation. Subsequently, increasing amounts of chemically synthesized test peptides were added, and the FP signal was measured with a Synergy H4 (BioTek Instruments) plate reader in 384-well plates. The Ki for each unlabeled peptide was determined by fitting the data to a competition binding equation. Titration experiments were carried out in triplicates, and the average FP signal was used for fitting the data with OriginPro7.

Cell-based protein–protein interaction assay

The full-length cDNA of yellow fluorescent protein (YFP) was split at position 159, and fragments (F1 and F2) were pasted into pcDNA 3.1 vectors (Invitrogen). ERK2 was expressed as C-terminal and p38α and JNK1 as N-terminal F2 fusions. To facilitate expression, JNK1 and p38α had kinase-inactivating mutations (K55R and K53R, respectively), while ERK2 was wild type. The ERK2 and JNK1 constructs contained a FLAG-tag, but similarly tagged p38α constructs could not be expressed to a comparable extent. Therefore, expression levels of F2-p38α could only be monitored by an anti-p38α antibody. MAPK partners were expressed as N-terminal and C-terminal F1 fusions with FLAG-tags. F1 and F2 fusion pairs that gave the highest BiFc signal with wild-type MAPK partners were chosen to analyze the impact of docking motif truncations or mutations. These were introduced into full-length MAPK partner constructs by PCR or by the QuickChange method. N-terminal truncations were made in proteins with N-terminal docking motifs: MKK1(14-393), MKK6(18-334), RHDF1(26-855), MKP5(28-482), and FAM122A (14-287) and a 20-amino-acid-long C-terminal truncation was used to generate the DCX(1-343) construct. Mutation of multiple residues was utilized for proteins with internal motifs: JIP1 (R160A,P161A,L164A,L166A) and APBA2(R280A,P281A,L284A, L286A), and K331, K332, K333, and L337 were mutated to alanines in the short KSR2(325-399) construct All sequences were verified by DNA sequencing. HEK293T cells were cultured in Dulbecco's modified Eagle's medium (DMEM, Lonza) containing 10% fetal bovine serum and 1% penicillin/streptomycin at 37°C in an atmosphere of 5% CO_2 in 25-cm^2 tissue culture flasks (Orange Scientific). Cells were seeded onto 96-well plate (tissue culture test plate 96F, TPP) at 60–70% confluence 24 h prior to transfection. The medium was changed to serum-free OPTI-MEM (Gibco). Transient transfections with Lipofectamine 2000 reagent (Invitrogen) were carried out according to the manufacturer's instruction. Cells were assayed 2 days post-transfection. For BiFc signal intensity measurements, cells were washed and suspended in 100 μl PBS. Twenty microliters of this cell suspension (~20,000 cells) was aliquoted into a 384-well black-sided plate. Fluorescence intensity per well was measured using a Synergy H4 (BioTek Instruments) fluorescence plate reader (excitation/emission wavelength was 515/535 nm). Cells from 50 μl of the PBS suspension were collected, and samples were subjected to Western blots using anti-FLAG-tag antibody (Sigma, F1804). For imaging, transfected cells were examined with an Olympus IX81 microscope using an Olympus FluoView 500 confocal laser scanning microscope system (Hamburg, Germany). YFP fluorescence was imaged using 514-nm excitation and a 535- to 560-nm emission filter.

Cell-based assays for EGF stimulation and monitoring GAB1 localization

GAB1 constructs were subcloned into modified pCerulean-C1 vector with N-terminal Cerulean fluorescent protein and C-terminal FLAG fusion tags. HEK293T cells were cultured and transfected similarly as described above. Cells were transfected with 0.2 μg GAB1 DNA constructs and were serum-starved for 24 h. The media were removed after 40 h from DNA transfection and 100 μl PBS was added to wells. ERK pathway stimulation was started by the addition of epidermal growth factor (EGF, Sigma, E9644) at 20 ng/ml concentration to each well, and stimulation was terminated at different time points by adding 35 μl of 4× SDS loading buffer to wells. Cells were lysed and 10 μl of each sample was subjected to SDS–PAGE and transferred to PVDF membrane (Sigma, P2563). Western blotting for monitoring total ERK1/2 and GAB1 was done by using anti-p42/44 MAPK (Cell Signaling, #4695) and anti-FLAG antibody (Sigma, F1804), respectively. After stripping the membrane, phospho-p44/42 MAPK (ERK1/2) (Thr202/Tyr204) antibody (Cell Signaling, #9101) was used to detect ppERK1/2 protein levels. For secondary antibody, anti-rabbit HRP-linked antibody (Cell Signaling, #7074) and anti-mouse HRP-linked antibody (Calbiochem, #401215) were used. The ECL signal was quantified with an Alpha Innotech Fluorchem FC2 instrument.

For visualizing GAB1 localization analysis, transfected cells were examined by Zeiss LSM 710 confocal laser scanning microscope using 405-nm laser for excitation in chambered borosilicate cover glass (Lab-Tek, #155411). Evaluation of CFP-GAB1 localization was done using ImageJ. Cells were treated by 100 ng/ml epidermal growth factor (EGF) for 10 min.

In vitro kinase assays

Activated JNK and p38α were produced by co-expressing the MAPKs with constitutively active forms of GST-tagged MAPK kinases (MKK7 and MKK6, respectively) in *E. coli* with bicistronic modified pET vectors. DCX and AAKG2 proteins and their mutants were produced in *E. coli* using modified pET vectors with C-terminal His-tag and N-terminal MBP-tag or GST-tag, respectively. Proteins were purified with double-affinity chromatography using Ni and MPB or GST column steps.

Twenty nM activated MAPK was incubated with 400 nM DCX or AAKG2 proteins at room temperature in the presence of 1 mg/ml

BSA. Kinase reactions were carried out in 50 mM HEPES pH 7.5, 100 mM NaCl, 5 mM $MgCl_2$, 0.05% IGEPAL, 5% glycerol, 2 mM DTT in the presence of 250 μM ATP, and ~5 μCi of ATP (γ32P). Reactions were stopped with protein loading sample buffer complemented with 100 mM EDTA, boiled, and then subjected to SDS–PAGE. The dried gel was analyzed by phosphorimaging on a Typhoon Trio+ scanner (GE Healthcare). Competitor docking motif peptides were chemically synthesized and used in 10 μM concentration (pepMK2, specific to p38: IKIKKIEADASNPLLLKRRKK; and pepJIP1, specific to JNK: DTYRPKRPTTLNLFP).

Expanded View for this article is available online.

Acknowledgements
We are grateful to András Patthy and Ádám Póti for excellent quality chemically synthesized peptides, to Marianna Rakács and Gergő Gógl for their help in protein production, to László Végner for help in operating the pipetting robot for dot-blot assays, and to Bálint Szeder for help with confocal laser scanning microscopy. A.R. is the recipient of the "Lendület" Grant from the Hungarian Academy of Sciences (LP2013-57). This work was also supported by the OTKA NN 114309 grant (awarded to A.R.). Z. D. acknowledges the support of the "Lendület" Grant from the Hungarian Academy of Sciences (LP2014-18) and OTKA grant (K108798) B.M. acknowledges the support from OTKA grant NK 100482.

Author contributions
AZ and AR designed the study. AZ, AA, ÁG, and KK performed experimental work. TB, BM, and OVK performed *in silico* analysis. AZ, OVK, ZD, and AR analyzed data. AZ and AR wrote the paper. All authors read and approved the manuscript.

Conflict of interest
The authors declare that they have no conflict of interest.

References

Agnarsson I, Zambrana-Torrelio CM, Flores-Saldana NP, May-Collado LJ (2011) A time-calibrated species-level phylogeny of bats (Chiroptera, Mammalia). *PLoS Curr* 3: RRN1212

Alexa A, Gógl G, Glatz G, Garai Á, Zeke A, Varga J, Dudás E, Jeszenői N, Bodor A, Hetényi C, Reményi A (2015) Structural assembly of the signaling competent ERK2–RSK1 heterodimeric protein kinase complex. *Proc Natl Acad Sci* 112: 201417571

Alexandropoulos K, Baltimore D (1996) Coordinate activation of c-Src by SH3- and SH2-binding sites on a novel p130Cas-related protein. *Sin Genes Dev* 10: 1341 – 1355

Bandyopadhyay S, Chiang C, Srivastava J, Gersten M, White S, Bell R, Kurschner C, Martin CH, Smoot M, Sahasrabudhe S, Barber DL, Chanda SK, Ideker T (2010) A human MAP kinase interactome. *Nat Methods* 7: 801 – 805

Bardwell L (2006) Mechanisms of MAPK signalling specificity. *Biochem Soc Trans* 34: 837 – 841

Barley AJ, Spinks PQ, Thomson RC, Shaffer HB (2010) Fourteen nuclear genes provide phylogenetic resolution for difficult nodes in the turtle tree of life. *Mol Phylogenet Evol* 55: 1189 – 1194

Berglund A-C, Sjölund E, Ostlund G, Sonnhammer ELL (2008) InParanoid 6: eukaryotic ortholog clusters with inparalogs. *Nucleic Acids Res* 36: D263 – D266

Betancur-R R, Broughton RE, Wiley EO, Carpenter K, López JA, Li C, Holcroft NI, Arcila D, Sanciangco M, Cureton IJ, Zhang F, Buser T, Campbell MA, Ballesteros JA, Roa-Varon A, Willis S, Borden WC, Rowley T, Reneau PC, Hough DJ *et al* (2013) The tree of life and a new classification of bony fishes. *PLoS Curr* 5: 1 – 33

Biondi RM, Nebreda AR (2003) Signalling specificity of Ser/Thr protein kinases through docking-site-mediated interactions. *Biochem J* 372: 1 – 13

Brown JP, Couillard-Després S, Cooper-Kuhn CM, Winkler J, Aigner L, Kuhn HG (2003) Transient expression of doublecortin during adult neurogenesis. *J Comp Neurol* 467: 1 – 10

Carlson SM, Chouinard CR, Labadorf A, Lam CJ, Schmelzle K, Fraenkel E, White FM (2011) Large-scale discovery of ERK2 substrates identifies ERK-mediated transcriptional regulation by ETV3. *Sci Signal* 4: rs11

Chang CI, Xu B, Akella R, Cobb MH, Goldsmith EJ (2002) Crystal structures of MAP kinase p38 complexed to the docking sites on its nuclear substrate MEF2A and activator MKK3b. *Mol Cell* 9: 1241 – 1249

Chen W-K, Yeap YYC, Bogoyevitch MA (2014) The JNK1/JNK3 interactome - Contributions by the JNK3 unique N-terminus and JNK common docking site residues. *Biochem Biophys Res Commun* 453: 576 – 581

Chow CW, Rincón M, Cavanagh J, Dickens M, Davis RJ (1997) Nuclear accumulation of NFAT4 opposed by the JNK signal transduction pathway. *Science* 278: 1638 – 1641

Costanzo-Garvey DL, Pfluger PT, Dougherty MK, Stock JL, Boehm M, Chaika O, Fernandez MR, Fisher K, Kortum RL, Hong E-G, Jun JY, Ko HJ, Schreiner A, Volle DJ, Treece T, Swift AL, Winer M, Chen D, Wu M, Leon LR *et al* (2009) KSR2 is an essential regulator of AMP kinase, energy expenditure, and insulin sensitivity. *Cell Metab* 10: 366 – 378

Courcelles M, Frémin C, Voisin L, Lemieux S, Meloche S, Thibault P (2013) Phosphoproteome dynamics reveal novel ERK1/2 MAP kinase substrates with broad spectrum of functions. *Mol Syst Biol* 9: 669

Dajani R, Fraser E, Roe SM, Yeo M, Good VM, Thompson V, Dale TC, Pearl LH (2003) Structural basis for recruitment of glycogen synthase kinase 3beta to the axin-APC scaffold complex. *EMBO J* 22: 494 – 501

Dajas-Bailador F, Jones EV, Whitmarsh AJ (2008) The JIP1 scaffold protein regulates axonal development in cortical neurons. *Curr Biol* 18: 221 – 226

De Nicola GF, Martin ED, Chaikuad A, Bassi R, Clark J, Martino L, Verma S, Sicard P, Tata R, Atkinson RA, Knapp S, Conte MR, Marber MS (2013) Mechanism and consequence of the autoactivation of p38α mitogen-activated protein kinase promoted by TAB 1. *Nat Struct Mol Biol* 20: 1182 – 1190

Dinkel H, Van Roey K, Michael S, Davey NE, Weatheritt RJ, Born D, Speck T, Krüger D, Grebnev G, Kuban M, Strumillo M, Uyar B, Budd A, Altenberg B, Seiler M, Chemes LB, Glavina J, Sánchez IE, Diella F, Gibson TJ (2014) The eukaryotic linear motif resource ELM: 10 years and counting. *Nucleic Acids Res* 42: D259 – D266

Dosztányi Z, Csizmók V, Tompa P, Simon I (2005) The pairwise energy content estimated from amino acid composition discriminates between folded and intrinsically unstructured proteins. *J Mol Biol* 347: 827 – 839

Dosztányi Z, Mészáros B, Simon I (2009) ANCHOR: web server for predicting protein binding regions in disordered proteins. *Bioinformatics* 25: 2745 – 2746

Durand D (2003) Vertebrate evolution: doubling and shuffling with a full deck. *Trends Genet* 19: 2 – 5

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32: 1792 – 1797

Estrella NL, Desjardins CA, Nocco SE, Clark AL, Maksimenko Y, Naya FJ (2015) MEF2 Transcription Factors Regulate Distinct Gene Programs in Mammalian Skeletal Muscle Differentiation. *J Biol Chem* 290: 1256 – 1268

Eulenfeld R, Schaper F (2009) A new mechanism for the regulation of Gab1 recruitment to the plasma membrane. *J Cell Sci* 122: 55 – 64

Fantz DA, Jacobs D, Glossip D, Kornfeld K (2001) Docking sites on substrate proteins direct extracellular signal-regulated kinase to phosphorylate specific residues. *J Biol Chem* 276: 27256 – 27265

Fernandes N, Bailey DE, VanVranken DL, Allbritton NL (2007) Use of Docking Peptides to Design Modular Substrates with High Efficiency for Mitogen-Activated Protein Kinase Extracellular Signal-Regulated Kinase. *ACS Chem Biol* 2: 665 – 673

Finck BN, Kelly DP (2006) PGC-1 coactivators: inducible regulators of energy metabolism in health and disease. *J Clin Invest* 116: 615 – 622

Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heger A, Hetherington K, Holm L, Mistry J, Sonnhammer ELL, Tate J, Punta M (2014) Pfam: the protein families database. *Nucleic Acids Res* 42: D222 – D230

Flotho A, Simpson DM, Qi M, Elion EA (2004) Localized feedback phosphorylation of Ste5p scaffold by associated MAPK cascade. *J Biol Chem* 279: 47391 – 47401

Garai Á, Zeke A, Gógl G, Törő I, Fördős F, Blankenburg H, Bárkai T, Varga J, Alexa A, Emig D, Albrecht M, Reményi A (2012) Specificity of linear motifs that bind to a common mitogen-activated protein kinase docking groove. *Sci Signal* 5: ra74

Glatz G, Gogl G, Alexa A, Remenyi A (2013) Structural Mechanism for the Specific Assembly and Activation of the Extracellular Signal Regulated Kinase 5 (ERK5) Module. *J Biol Chem* 288: 8596 – 8609

Goldman A, Roy J, Bodenmiller B, Wanka S, Landry CR, Aebersold R, Cyert MS (2014) The calcineurin signaling network evolves via conserved kinase-phosphatase modules that transcend substrate identity. *Mol Cell* 55: 422 – 435

Gordon EA, Whisenant TC, Zeller M, Kaake RM, Gordon WM, Krotee P, Patel V, Huang L, Baldi P, Bardwell L (2013) Combining docking site and phosphosite predictions to find new substrates: identification of smoothelin-like-2 (SMTNL2) as a c-Jun N-terminal kinase (JNK) substrate. *Cell Signal* 25: 2518 – 2529

Grewal S, Molina DM, Bardwell L (2006) Mitogen-activated protein kinase (MAPK)-docking sites in MAPK kinases function as tethers that are crucial for MAPK regulation in vivo. *Cell Signal* 18: 123 – 134

Heo Y-S, Kim S-K, Il SC, Kim YK, Sung B-J, Lee HS, Il LJ, Park S-Y, Kim JH, Hwang KY, Hyun Y-L, Jeon YH, Ro S, Cho JM, Lee TG, Yang C-H (2004) Structural basis for the selective inhibition of JNK1 by the scaffolding protein JIP1 and SP600125. *EMBO J* 23: 2185 – 2195

Hirosumi J, Tuncman G, Chang L, Görgün CZ, Uysal KT, Maeda K, Karin M, Hotamisligil GS (2002) A central role for JNK in obesity and insulin resistance. *Nature* 420: 333 – 336

Ho DT, Bardwell AJ, Grewal S, Iverson C, Bardwell L (2006) Interacting JNK-docking sites in MKK7 promote binding and activation of JNK mitogen-activated protein kinases. *J Biol Chem* 281: 13169 – 13179

Holgado-Madruga M, Emlet DR, Moscatello DK, Godwin AK, Wong AJ (1996) A Grb2-associated docking protein in EGF- and insulin-receptor signalling. *Nature* 379: 560 – 564

Horton P, Park K-J, Obayashi T, Fujita N, Harada H, Adams-Collier CJ, Nakai K (2007) WoLF PSORT: protein localization predictor. *Nucleic Acids Res* 35: W585 – W587

Hulo N, Sigrist CJA, Le Saux V, Langendijk-Genevaux PS, Bordoli L, Gattiker A, De Castro E, Bucher P, Bairoch A (2004) Recent improvements to the PROSITE database. *Nucleic Acids Res* 32: D134 – D137

Jacobs D, Glossip D, Xing H, Muslin AJ, Kornfeld K (1999) Multiple docking sites on substrate proteins form a modular system that mediates recognition by ERK MAP kinase. *Genes Dev* 13: 163 – 175

Jensen LJ, Julien P, Kuhn M, von Mering C, Muller J, Doerks T, Bork P (2008) eggNOG: automated construction and annotation of orthologous groups of genes. *Nucleic Acids Res* 36: D250 – D254

Johnson GL, Lapadat R (2002) Mitogen-activated protein kinase pathways mediated by ERK, JNK, and p38 protein kinases. *Science* 298: 1911 – 1912

Jordan GE, Piel WH (2008) PhyloWidget: web-based visualizations for the tree of life. *Bioinformatics* 24: 1641 – 1642

Jung US, Sobering AK, Romeo MJ, Levin DE (2002) Regulation of the yeast Rlm1 transcription factor by the Mpk1 cell wall integrity MAP kinase. *Mol Microbiol* 46: 781 – 789

Käll L, Krogh A, Sonnhammer ELL (2007) Advantages of combined transmembrane topology and signal peptide prediction–the Phobius web server. *Nucleic Acids Res* 35: W429 – W432

Kosako H, Yamaguchi N, Aranami C, Ushiyama M, Kose S, Imamoto N, Taniguchi H, Nishida E, Hattori S (2009) Phosphoproteomics reveals new ERK MAP kinase targets and links ERK to nucleoporin-mediated nuclear transport. *Nat Struct Mol Biol* 16: 1026 – 1035

Kragelj J, Palencia A, Nanao MH, Maurin D, Bouvignies G, Blackledge M, Jensen MR (2015) Structure and dynamics of the MKK7-JNK signaling complex. *Proc Natl Acad Sci USA* 112: 3409 – 3414

Kuraku S, Meyer A, Kuratani S (2009) Timing of genome duplications relative to the origin of the vertebrates: did cyclostomes diverge before or after? *Mol Biol Evol* 26: 47 – 59

Lang T, Yu L, Tu Q, Jiang J, Chen Z, Xin Y, Liu G, Zhao S (2000) Molecular Cloning, Genomic Organization, and Mapping of PRKAG2, a Heart Abundant γ2 Subunit of 5′-AMP-Activated Protein Kinase, to Human Chromosome 7q36. *Genomics* 70: 258 – 263

Laughlin JD, Nwachukwu JC, Figuera-Losada M, Cherry L, Nettles KW, LoGrasso PV (2012) Structural mechanisms of allostery and autoinhibition in JNK family kinases. *Structure* 20: 2174 – 2184

Lee YH, Giraud J, Davis RJ, White MF (2003) c-Jun N-terminal kinase (JNK) mediates feedback inhibition of the insulin signaling cascade. *J Biol Chem* 278: 2896 – 2902

Lee KS, Park J-E, Kang YH, Kim T-S, Bang JK (2014) Mechanisms underlying Plk1 polo-box domain-mediated biological processes and their physiological significance. *Mol Cells* 37: 286 – 294

Li J, Ning Y, Hedley W, Saunders B, Chen Y, Tindill N, Hannay T, Subramaniam S (2002) The Molecule Pages database. *Nature* 420: 716 – 717

Li L, Wang Y, Carr R, Haddad CS, Li Z, Qian L, Oberholzer J, Maker AV, Wang Q, Prabhakar BS (2014) IG20/MADD plays a critical role in glucose-induced insulin secretion. *Diabetes* 63: 1612 – 1623

Lim WA, Richards FM, Fox RO (1994) Structural determinants of peptide-binding orientation and of sequence specificity in SH3 domains. *Nature* 372: 375 – 379

Livingstone C, Patel G, Jones N (1995) ATF-2 contains a phosphorylation-dependent transcriptional activation domain. *EMBO J* 14: 1785 – 1797

London N, Raveh B, Cohen E, Fathi G, Schueler-Furman O (2011) Rosetta FlexPepDock web server–high resolution modeling of peptide-protein interactions. *Nucleic Acids Res* 39: W249 – W253

Lowe ED, Tews I, Cheng KY, Brown NR, Gul S, Noble MEM, Gamblin SJ, Johnson LN (2002) Specificity determinants of recruitment peptides bound to phospho-CDK2/cyclin A. *Biochemistry* 41: 15625 – 15634

Lupas A, Van Dyke M, Stock J (1991) Predicting coiled coils from protein sequences. *Science* 252: 1162 – 1164

Ma W, Shang Y, Wei Z, Wen W, Wang W, Zhang M (2010) Phosphorylation of DCC by ERK2 is facilitated by direct docking of the receptor P1 domain to the kinase. *Structure* 18: 1502 – 1511

Mace PD, Wallez Y, Egger MF, Dobaczewska MK, Robinson H, Pasquale EB, Riedl SJ (2013) Structure of ERK2 bound to PEA-15 reveals a mechanism for rapid release of activated MAPK. *Nat Commun* 4: 1681

Mészáros B, Simon I, Dosztányi Z (2009) Prediction of protein binding regions in disordered proteins. *PLoS Comput Biol* 5: e1000376

Mészáros B, Dosztányi Z, Simon I (2012) Disordered binding regions and linear motifs–bridging the gap between two models of molecular recognition. *PLoS One* 7: e46829

Nagadoi A, Nakazawa K, Uda H, Okuno K, Maekawa T, Ishii S, Nishimura Y (1999) Solution structure of the transactivation domain of ATF-2 comprising a zinc finger-like subdomain and a flexible subdomain. *J Mol Biol* 287: 593 – 607

Ngo JCK, Chakrabarti S, Ding J-H, Velazquez-Dones A, Nolen B, Aubol BE, Adams JA, Fu X-D, Ghosh G (2005) Interplay between SRPK and Clk/Sty kinases in phosphorylation of the splicing factor ASF/SF2 is regulated by a docking motif in ASF/SF2. *Mol Cell* 20: 77 – 89

Ohno S (1993) Patterns in genome evolution. *Curr Opin Genet Dev* 3: 911 – 914

Olson BL, Hock MB, Ekholm-Reed S, Wohlschlegel JA, Dev KK, Kralli A, Reed SI (2008) SCFCdc4 acts antagonistically to the PGC-1alpha transcriptional coactivator by targeting it for ubiquitin-mediated proteolysis. *Genes Dev* 22: 252 – 264

Pearce LR, Atanassova N, Banton MC, Bottomley B, van der Klaauw AA, Revelli J-P, Hendricks A, Keogh JM, Henning E, Doree D, Jeter-Jones S, Garg S, Bochukova EG, Bounds R, Ashford S, Gayton E, Hindmarsh PC, Shield JPH, Crowne E, Barford D *et al* (2013) KSR2 mutations are associated with obesity, insulin resistance, and impaired cellular fuel oxidation. *Cell* 155: 765 – 777

Pellicena P, Stowell KR, Miller WT (1998) Enhanced phosphorylation of Src family kinase substrates containing SH2 domain binding sites. *J Biol Chem* 273: 15325 – 15328

Petersen TN, Brunak S, von Heijne G, Nielsen H (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods* 8: 785 – 786

Podkowa M, Zhao X, Chow C-W, Coffey ET, Davis RJ, Attisano L (2010) Microtubule stabilization by bone morphogenetic protein receptor-mediated scaffolding of c-Jun N-terminal kinase promotes dendrite formation. *Mol Cell Biol* 30: 2241 – 2250

Reményi A, Good MC, Bhattacharyya RP, Lim WA (2005) The role of docking interactions in mediating signaling input, output, and discrimination in the yeast MAPK network. *Mol Cell* 20: 951 – 962

Reményi A, Good MC, Lim WA (2006) Docking interactions in protein kinase and phosphatase networks. *Curr Opin Struct Biol* 16: 676 – 685

Rybakowska I, Romaszko P, Zabielska M, Turyn J, Kaletha K, Barton PJ, Slominska EM, Smolenski RT (2014) Effect of AMP-deaminase 3 knock-out in mice on enzyme activity in heart and other organs. *Nucleosides, Nucleotides Nucleic Acids* 33: 319 – 322

Sabio G, Kennedy NJ, Cavanagh-Kyros J, Jung DY, Ko HJ, Ong H, Barrett T, Kim JK, Davis RJ (2010) Role of muscle c-Jun NH2-terminal kinase 1 in obesity-induced insulin resistance. *Mol Cell Biol* 30: 106 – 115

Sánchez IE, Beltrao P, Stricher F, Schymkowitz J, Ferkinghoff-Borg J, Rousseau F, Serrano L (2008) Genome-wide prediction of SH2 domain targets using structural information and the FoldX algorithm. *PLoS Comput Biol* 4: e1000052

Schäffer AA, Aravind L, Madden TL, Shavirin S, Spouge JL, Wolf YI, Koonin EV, Altschul SF (2001) Improving the accuracy of PSI-BLAST protein database searches with composition-based statistics and other refinements. *Nucleic Acids Res* 29: 2994 – 3005

Schymkowitz J, Borg J, Stricher F, Nys R, Rousseau F, Serrano L (2005) The FoldX web server: an online force field. *Nucleic Acids Res* 33: W382 – W388

Sharrocks AD, Yang SH, Galanis A (2000) Docking domains and substrate-specificity determination for MAP kinases. *Trends Biochem Sci* 25: 448 – 453

Tanoue T, Adachi M, Moriguchi T, Nishida E (2000) A conserved docking motif in MAP kinases common to substrates, activators and regulators. *Nat Cell Biol* 2: 110 – 116

Tanoue T, Maeda R, Adachi M, Nishida E (2001) Identification of a docking groove on ERK and p38 MAP kinases that regulates the specificity of docking interactions. *EMBO J* 20: 466 – 479

Taru H, Suzuki T (2004) Facilitation of stress-induced phosphorylation of beta-amyloid precursor protein family members by X11-like/Mint2 protein. *J Biol Chem* 279: 21628 – 21636

Thomsen MCF, Nielsen M (2012) Seq2Logo: a method for construction and visualization of amino acid binding motifs and sequence profiles including sequence weighting, pseudo counts and two-sided representation of amino acid enrichment and depletion. *Nucleic Acids Res* 40: W281 – W287

Tompa P, Davey NE, Gibson TJ, Babu MM (2014) A million peptide motifs for the molecular biologist. *Mol Cell* 55: 161 – 169

Tonikian R, Zhang Y, Sazinsky SL, Currell B, Yeh J-H, Reva B, Held HA, Appleton BA, Evangelista M, Wu Y, Xin X, Chan AC, Seshagiri S, Lasky LA, Sander C, Boone C, Bader GD, Sidhu SS (2008) A specificity map for the PDZ domain family. *PLoS Biol* 6: e239

Vitari AC, Thastrup J, Rafiqi FH, Deak M, Morrice NA, Karlsson HKR, Alessi DR (2006) Functional interactions of the SPAK/OSR1 kinases with their upstream activator WNK1 and downstream substrate NKCC1. *Biochem J* 397: 223 – 231

Whisenant TC, Ho DT, Benz RW, Rogers JS, Kaake RM, Gordon EA, Huang L, Baldi P, Bardwell L (2010) Computational prediction and experimental verification of new MAP kinase docking sites and substrates including Gli transcription factors. *PLoS Comput Biol* 6: e1000908

Wolf A, Eulenfeld R, Bongartz H, Hessenkemper W, Simister PC, Lievens S, Tavernier J, Feller SM, Schaper F (2015) MAPK-induced Gab1 translocation to the plasma membrane depends on a regulated intramolecular switch. *Cell Signal* 27: 340 – 352

Yang SH, Galanis A, Sharrocks AD (1999) Targeting of p38 mitogen-activated protein kinases to MEF2 transcription factors. *Mol Cell Biol* 19: 4028 – 4038

Ying CY, Dominguez-Sola D, Fabi M, Lorenz IC, Hussein S, Bansal M, Califano A, Pasqualucci L, Basso K, Dalla-Favera R (2013) MEF2B mutations lead to deregulated expression of the oncogene BCL6 in diffuse large B cell lymphoma. *Nat Immunol* 14: 1084 – 1092

Zhang J, Zhou B, Zheng C, Zhang Z (2003) A bipartite mechanism for ERK2 recognition by its cognate regulators and substrates. *J Biol Chem* 278: 29901 – 29912

Zhang YY, Wu JW, Wang ZX (2011) A Distinct Interaction Mode Revealed by the Crystal Structure of the Kinase p38α with the MAPK Binding Domain of the Phosphatase MKP5. *Sci Signal* 4: ra88

Zhou T, Sun L, Humphreys J, Goldsmith EJ (2006) Docking interactions induce exposure of activation loop in the MAP kinase ERK2. *Structure* 14: 1011 – 1019