Published in final edited form as:

Sci Signal. ; 8(405): rs13. doi:10.1126/scisignal.aaa6711.

Identifying three-dimensional structures of autophosphorylation complexes in crystals of protein kinases

Qifang Xu¹, Kimberly L. Malecka¹, Lauren Fink¹, E. Joseph Jordan², Erin Duffy¹, Samuel Kolander¹, Jeffrey Peterson¹, and Roland L. Dunbrack Jr.^{1,*}

¹Institute for Cancer Research, Fox Chase Cancer Center, Philadelphia, PA.

²Department of Bioengineering, University of Pennsylvania, Philadelphia, PA.

Abstract

Protein kinase autophosphorylation is a common regulatory mechanism in cell signaling pathways. Crystal structures of several homomeric protein kinase complexes have a serine, threonine, or tyrosine autophosphorylation site of one kinase monomer located in the active site of another monomer, a structural complex that we call an “autophosphorylation complex.” We developed and applied a structural bioinformatics method to identify all such autophosphorylation kinase complexes in X-ray crystallographic structures in the Protein Data Bank (PDB). We identified 15 autophosphorylation complexes in the PDB, of which 5 complexes had not previously been described in the publications describing the crystal structures. These 5 consist of tyrosine residues in the N-terminal juxtamembrane regions of colony stimulating factor 1 receptor (CSF1R, Tyr⁵⁶¹) and EPH receptor A2 (EPA2, Tyr⁵⁹⁴), tyrosine residues in the activation loops of the SRC kinase family member LCK (Tyr³⁹⁴) and insulin-like growth factor 1 receptor (IGF1R, Tyr¹¹⁶⁶), and a serine in a nuclear localization signal region of CDC-like kinase 2 (CLK2, Ser¹⁴²). Mutations in the complex interface may alter autophosphorylation activity and contribute to disease; therefore we mutated residues in the autophosphorylation complex interface of LCK and found that two mutations impaired autophosphorylation (T445V and N446A) and mutation of Pro⁴⁴⁷ to Ala, Gly, or Leu increased autophosphorylation. The identified autophosphorylation sites are conserved in many kinases, suggesting that, by homology, these complexes may provide insight into autophosphorylation complex interfaces of kinases that are relevant drug targets.

INTRODUCTION

Protein kinases play important roles in many cellular signaling pathways, such as cell cycle regulation and apoptosis (1). Problems in kinase regulation can lead to diverse illnesses ranging from cancer (2) to obesity (3). Activity of most kinases is partially regulated by the phosphorylation status and position of the activation loop, which begins with the highly

*To whom correspondence should be addressed: Roland.Dunbrack@fccc.edu.

Author contributions: RD conceived the project. RD and JP supervised the project. QX and RD wrote the paper. QX, RD, ED, EJJ, and SK performed calculations and analysis. KM and LF performed experiments on LCK and analyzed the experimental data.

Competing interests: No competing interests.

Data and materials availability: All materials and data are available from the authors. The coordinates of the structures described in this paper are available as supplemental material from *Science Signaling*.

conserved DFG (Asp-Phe-Gly) motif and ends with a sequence usually similar to APE (Ala-Pro-Glu) (4). In many kinases, the nonphosphorylated activation loop occupies a position that interferes with substrate binding. When phosphorylated, usually by trans-autophosphorylation (meaning by a second instance of the same kinase), the activation loop becomes repositioned, providing access to the active site for substrates and rearranging several residues required for catalysis (5). Many kinases contain additional sites outside the activation loop that are also trans-autophosphorylated (6).

Several kinase structures have been reported in which a serine, threonine, or tyrosine autophosphorylation site of one kinase monomer is present in the active site of another monomer of the same protein in the crystal (7-15). In these structures, the position of the phosphorylation site and adjacent residues resembles those of substrates in structures of substrate peptides bound to kinases (16-18). Phosphorylation sites reported in autophosphorylation complexes in crystals include a tyrosine in the juxtamembrane region that is N-terminal to the kinase domain of the receptor tyrosine kinase c-KIT [Protein Data Bank (PDB: 1PKG)] (7), a tyrosine in the “kinase insert region” of fibroblast growth factor receptor 1 (FGFR1) [PDB: 3GQI (8)] and of FGFR3 [PDB: 4K33 (14)], a tyrosine in the C-terminal tail of FGFR2 [PDB: 3CLY (9)] and of epidermal growth factor receptor (EGFR) [PDB: 4I21 (19)], and a tyrosine in the activation loop of insulin-like growth factor 1 receptor (IGF1R) [PDB: 3D94 (10)]. In all of these, the tyrosine side chain of the substrate kinase is hydrogen-bonded to the catalytic Asp side chain of the active site HRD (His-Arg-Asp) motif of the enzyme kinase [the site in FGFR1 (PDB: 3GQI) has been mutated to Phe but is correctly positioned if it were Tyr]. Furthermore, each of these residues is an experimentally verified autophosphorylation site in these kinases. For serine/threonine kinases, autophosphorylation complexes of the activation loop Thr residues of p21-activated kinase (PAK1) [PDB: 3Q4Z (11)] and interleukin-1 receptor-associated kinase 4 (IRAK4) [PDB: 4U97; PDB: 4U9A (15)] have been described, as have autophosphorylation complexes of the C-terminal regulatory regions of human [PDB: 2WEL (12)] and *Caenorhabditis elegans* [PDB: 3KK8; PDB: 3KK9 (13)] calcium/calmodulin-dependent kinase II (CaMKII).

Because of the importance of understanding kinase activation processes and kinase-substrate recognition, we sought to identify undetected autophosphorylation complexes in crystals of kinases in the PDB using a structural bioinformatics approach. Employing the symmetry information for each crystal provided by the PDB, we constructed all distinct interfaces between monomers in 3,525 kinase crystals in the PDB (as of October 24, 2015) and measured the distance between the Asp oxygen atoms of the HRD motif in one monomer and the hydroxyl groups on Ser, Thr, and Tyr of the other monomer, and vice versa.

This approach properly identified the ten previously described autophosphorylation complexes listed above, and identified five more that were not described as such in the relevant papers. The newly identified autophosphorylation events include: (i) the activation loop Tyr of human nonreceptor tyrosine kinase LCK [PDB: 2PL0 (20)], which is similar to the IGF1R structure (10); (ii) a second tyrosine autophosphorylation site (Tyr¹¹⁶⁶) in the activation loop of human IGF1R [PDB: 3LVP (21)]; (iii) a Tyr in the N-terminal juxtamembrane region of human colony stimulating factor 1 receptor (CSF1R) [PDB: 3LCD

(22)] that is homologous to the tyrosine observed in the c-KIT complex (7); (iv) a Tyr in the N-terminal juxtamembrane region of Ephrin receptor A2 (EPHA2) [PDB: 4PDO (23)] that represents a phosphorylation site near but distinct from that in the c-KIT and CSF1R complexes; and (v) a Ser in the C-terminal tail of human CDC2/CDC28-like kinase 2 (CLK2) [PDB: 3NR9 (24)]. We also identified several additional structures of autophosphorylation complexes of PAK1: PDB: 4O0R, 4O0T (25); 4P90 (26); 4ZY4, 4ZY5, 4ZY6 (27); and 4ZLO, 4ZJI, and 4ZJJ (28). Two of these structures, PDB: 4ZY4 and 4ZY5, contain an autophosphorylation complex with a fully ordered activation loop in the substrate kinase; whereas the others, including PDB: 3Q4Z (11), are missing the coordinates of 8-11 residues within the activation loop.

Comparison of all of the newly identified sites to the structures of previously published peptide substrate-kinase complexes in the PDB indicated that these sites are consistent with catalysis of phosphorylation. Comparison of one of the previously described autophosphorylation complexes, EGFR [PDB: 4I21, (19)], however, was not consistent with recently published structures of EGFR in complexes with peptide substrates.

Because of the central importance of activation loop autophosphorylation in the regulation of tyrosine kinases, we performed site-directed mutagenesis of the autophosphorylation complex interface of LCK that we identified in PDB: 2PLO, changing several residues in the G helix that contacts the activation loop of the opposing monomer, and analyzed the phosphorylation of these when expressed in cells. A mutation in Pro⁴⁴⁷ in LCK has previously been identified as activating and is associated with T cell leukemia (29). Whereas some mutations in the interface region impaired autophosphorylation, mutation of Pro⁴⁴⁷ to Gly, Ala, or Leu maintained or increased autophosphorylation activity.

By comparing all of the autophosphorylation complexes with each other and with the structures of peptide-bound kinases, we discovered common structural elements in subsets of the complexes, which provide insights regarding substrate specificity of kinases in general and autophosphorylation sites in particular and the importance of domain-domain interactions in the phosphorylation of some substrates in the context of full-sized proteins as distinct from peptide substrates. Furthermore, because several of the autophosphorylation sites are conserved in other kinases, the relevance of these complexes likely extends to other clinically relevant drug targets. We show that this is particularly true of the IGF1R, LCK, and PAK1 activation loop sites, which are found at homologous positions within the activation loop in the crystal structures of a large number of kinases.

RESULTS

Potential autophosphorylation complexes in crystals of protein kinases

As of 24 October 2015, the PDB contained 3,525 crystal structures of proteins that contain a Ser or Thr kinase domain or a Tyr kinase domain as identified with the associated hidden Markov models provided by Pfam (30) (Pfam models “Pkinase” and “Pkinase_Tyr”, respectively). These PDB entries comprise structural information on 237 human kinases, 25 mouse kinases, and 103 kinases from other species.

To identify possible autophosphorylation complexes in each crystal, we needed to build coordinates for a portion of the crystal large enough to contain at least one example of all distinct protein-protein interactions that make up the crystal, using methods described previously (31, 32). We started our analysis with the asymmetric unit, the smallest part of the crystal structure to which symmetry rules can be applied to produce coordinates for the entire crystal (33). We used information on the symmetry of each crystal provided by the PDB to build a unit cell of the crystal, a cuboid object (or parallelepiped) consisting of multiple copies of the asymmetric unit. Once we built a compact unit cell for a crystal, we built a $3\times 3\times 3$ collection of unit cells of the crystal by copying each unit cell and translating it in the $\pm x$, $\pm y$, and $\pm z$ directions as needed. This was done in order to guarantee that we have all possible interactions of the original asymmetric unit with all other neighboring asymmetric units in the crystal.

Each protein in the assembly of 27 unit cells is given an identifier in a standard format, such as “A:2_555.” The letter A is the chain identifier of the original asymmetric unit, which might consist of chains A, B, and C, for instance if there are three such chains in the asymmetric unit. The first number after the colon is the identifier of the symmetry operator (a rule to copy, rotate, and translate the coordinates of the asymmetric unit) used to orient the asymmetric units within the unit cell (“2” in this example). The number 555 indicates the unit cell in the $3\times 3\times 3$ unit cell structure. 555 designates the center of our $3\times 3\times 3$ box; neighboring unit cells of the center cell in the x, y, and z directions are designated by adding or subtracting 1 from these numbers. As another example, “B:3_655” indicates author chain B, symmetry operator 3 for the space group of the crystal form, and a neighboring unit cell one step in the positive x-direction. After building a $3\times 3\times 3$ collection of unit cells for each crystal, we determined the list of distinct protein-protein interfaces in the crystal for further analysis.

Formation of a hydrogen bond between the substrate hydroxyl group of the Ser, Thr, or Tyr side chain and the catalytic aspartic acid residue side chain of the active site HRD (His-Arg-Asp) motif is a critical step in the mechanism of phosphorylation by protein kinases (34). Therefore, we calculated the distances between hydroxyl oxygen atoms of Thr, Ser, or Tyr of each monomer with the Asp carboxylate atoms (O δ 1 and O δ 2) of the HRD motif for each of the other monomer in each of the distinct homodimers in each crystal. Because some kinase structures in the PDB contain phosphorylated amino acids and phosphomimetic mutations of phosphorylation sites to Asp or Glu, we also measured the distances of the appropriate oxygen atoms of these side chains and the active site Asp. Additionally, substrate residue side chains often form hydrogen bonds with the side chain of arginine or lysine residues located two or four amino acids after the HRD motif (HRD+2 and HRD+4) (35), and where this residue was present we calculated potential hydrogen bond distances between hydrogen-bond-acceptor oxygen atoms in the substrate side chain and hydrogen-bond-donor nitrogens in the Arg or Lys side chain of HRD+2 or HRD+4 (usually only one of these is present in the sequence).

With this procedure, we identified 15 potential unique autophosphorylation complexes in 26 PDB entries. These structures and the autophosphorylation sites contained in them are listed in Table 1, which provides details on each autophosphorylation complex including the

chains involved and their symmetry relationship, the relevant substrate and enzyme residues, the substrate sequence about the phosphorylation site, the surface area of interaction of the two kinases, and the hydrogen bond distance. Some crystals contain more than one unique instance of the autophosphorylation complex, and these are indicated in Table 1. Often the complexes are very similar, but in some cases they may have distinct structural features, such as the locations and extent of domain-domain interactions between the kinases. We organized and described the structures in groups on the basis of similarity of the complexes. The mean and median hydrogen bond lengths (oxygen-to-oxygen atom distances) are 3.16 and 2.91 Å, respectively, (picking only the shortest distance for each phosphosite, if there are multiple structures), which are typical of hydroxyl/carboxylate hydrogen bonds. The mean and median interface surface areas are 913 and 848 Å², respectively, indicating extensive interactions between the kinases beyond the residues adjacent to the phosphorylation site in the sequence. For each structure, we mapped the interaction areas separately to the surfaces of the enzyme and substrate kinases (fig. S1).

Table 1 also lists the symmetry operators required to determine the position of the enzyme and substrate kinase in each autophosphorylation complex. We found that the 26 PDB entries contain 33 autophosphorylation complexes, 18 of these 33 are between kinases in different copies of the asymmetric unit, and 13 are between kinase monomers in different unit cells, demonstrating the importance of considering the full symmetry of each crystal in identifying potential autophosphorylation complexes.

We expect that, in autophosphorylation complexes, the enzyme kinase should be in a conformation consistent with known active kinase structures in the PDB. The substrate kinases may be the active or inactive conformation and the state observed in the crystal may be informative in determining under what conditions trans-autophosphorylation might take place. For these reasons, in Table 2 we list three features that distinguish active from inactive kinases for both the enzymes and substrates listed in Table 1: (i) the position of the phenylalanine ring of the DFG motif; (ii) the presence or absence of a salt bridge between a Lys residue of the N-terminal β sheet and a Glu of the C-helix; (iii) the existence (or lack thereof) of van der Waals interactions of residue side chains of the “regulatory spine” (36), which comprises the His side chain of the HRD motif, the Phe of the DFG motif, and two additional hydrophobic residues in the N-terminal domain. The presence of an intact regulatory spine is associated with active enzyme structures (36). Most of the enzyme kinases in the autophosphorylation complexes are consistent with the active conformation (Table 2): the DFG motif is labeled DFG-in, meaning that the Phe ring is under the C-helix and the activation loop is positioned such that a substrate may bind in the enzyme active site; the distance of the Nζ atom of the Lys side chain and Oε1 or Oε2 atoms of the Glu side chain are within hydrogen-bonding distance (<3.5 Å); and the four regulatory spine residues are in contact along the spine, meaning that the first residue contacts the second (within 4.5 Å), the second contacts the third, and the third contacts the fourth. Table 2 also lists the ligands bound to the enzyme and substrate kinases with the three-letter codes provided by the PDB and the distance of the HRD+2 or HRD+4 Arg or Lys side-chain hydrogen bond donors (if present) with the phosphorylation site hydroxyl atom. In complexes of kinases and peptide substrates, this hydrogen bond (distance < 3.5 Å) is typically present (35). The

residues within 9 residues of the phosphorylation site of the substrate kinase that contact the enzyme kinase are highlighted in Table 3.

For each of the kinase autophosphorylation structures listed in Tables 1, 2, and 3, we provide coordinates for the complexes (Data File_S1_KinaseAutophosphorylationInterfaces.zip) and describe multiple properties of the complexes. In particular, we examined several aspects of the structures within each group: (i) the activity state of the enzyme kinase – whether the activation loop and C-helix are in positions consistent with active kinases (36) (Table 2); (ii) the position of the substrate loop or N or C terminal segment in the substrate kinase compared to non-substrate structures of the same or closely related kinases; (iii) apparent relevance in some cases of specific domain-domain contacts between the enzyme and substrate kinases; (iv) indications of specific, important determinants of substrate specificity at sequence positions neighboring the phosphorylation position (Table 3) and comparison with peptide-kinase crystal structures of closely related kinases; and (v) conservation and annotations of the phosphorylation sites at homologous sites in other kinases in addition to the autophosphorylation complex listed in Table 1.

Autophosphorylation sites in the juxtamembrane regions of tyrosine kinase receptors

In our search of interfaces between kinases in the PDB, we identified the structures of two previously undescribed autophosphorylation complexes similar to the previously described autophosphorylation complex of c-KIT [Fig. 1A, PDB: 1PKG (7)]: CSF1R [Fig. 1B, PDB: 3LCD (22)] and EPHA2 [Fig. 1C, PDB: 4PDO (23)]. Although not described by Mol *et al.* (7), we determined that the crystal of c-KIT (PDB: 1PKG) contains two distinct versions of the autophosphorylation complex of Tyr⁵⁶⁸ with different orientations of the substrate kinase domain (colored magenta and pink in Fig. 1A) relative to the enzyme kinase (colored blue and periwinkle). The asymmetric unit of this crystal contains two copies of c-KIT (chains A and B) that are in the active “DGF-in” conformation (4) (Table 2), both of which contain phosphate groups on Tyr⁵⁶⁸ and Tyr⁵⁷⁰, which are located in the juxtamembrane region of the kinase between the membrane spanning helix (residues 525-545) and the kinase domain (residues 589-937). Phosphorylated Tyr⁵⁶⁸ (pTyr⁵⁶⁸) of chain B is located in the active site of chain A within the asymmetric unit, with atom O η of Tyr⁵⁶⁸ within hydrogen-bonding distance of the side chain of Asp⁷⁹² of the HRD motif (a hydrogen bond would be formed if the Asp was protonated). In chain B, pTyr⁵⁶⁸ forms a hydrogen bond with Arg⁷⁹⁶ at the HRD+4 position of the enzyme kinase (Table 2). The pTyr⁵⁶⁸ of chain A of the asymmetric unit is located in the active site of chain B of another copy of the asymmetric unit in an adjacent unit cell (chain B:2_655).

Although not mentioned in the report of the crystal structure of CSF1R [PDB: 3LCD (22)], we found that Tyr⁵⁶¹ of the sole monomer (chain A) in the asymmetric unit is located in the active site of a neighboring asymmetric unit of the CSF1R crystal in the same unit cell (chain A:4_555) (Fig. 1B). Like the monomers in the c-KIT autophosphorylation complex (PDB: 1PKG), the kinase in CSF1R (PDB: 3LCD) is in an active DFG-in conformation (Table 2). The other structures of both c-KIT (PDB: 1T45, 1T46, 3G0E, 3G0F, 4HVS, 4U0I) and CSF1R (PDB: 2I0V, 2I0Y, 2I1M, 2OGV, 3BEA, 3DPK, 3KRJ, 3KRL, 3LCO,

4HW7, 4R7H, 4R7I) are all in inactive DFG-out conformations. Superposition of c-KIT monomers from available PDB entries illustrated that in inactive structures, the N-terminal juxtamembrane region blocks the substrate-binding site in an autoinhibitory manner; whereas in the autophosphorylation complexes, the N-terminal juxtamembrane region extends from the surface of the kinase and would be accessible for phosphorylation (Fig. 1D). Similarly, in the inactive structures of CSF1R, the N-terminal juxtamembrane region is not disordered and blocks the substrate-binding site by displacing the activation loop (Fig. 1E). Conversely, in the autophosphorylation complex, the N-terminal juxtamembrane region of CSF1R extends from the surface of the kinase, accessible for phosphorylation (Fig. 1E). The differences between the active and inactive structures for c-KIT and CSF1R are similar with the exception that the inactive structure of c-KIT (PDB: 3G0F) forms an intermolecular inhibitory complex and the other inactive structures form intramolecular inhibitory interactions.

Similar to c-KIT and CSF1R, we found that EPHA2 forms a potential autophosphorylation complex in a crystal structure of the kinase domain [PDB: 4PDO (23)], with a hydrogen bond between juxtamembrane region Tyr⁵⁹⁴ and Asp⁷³⁹ of the HRD motif (Fig. 1C). Both chains are in the active DFG-in conformation (Table 2), as are the other available structures of EPHA2 (PDB: 1MQB, 4P2K, 4TRL) (Fig. 1F), none of which contains a visible Tyr⁵⁹⁴ in the electron density. In the EPHA2 (PDB: 4PDO) structure, the salt bridge hydrogen atom-atom distance in the N-terminal domain is too long to form a hydrogen bond (Table 2), but the backbone positions of the two residues are in the same positions as in the other EPHA2 structures. A small change of side-chain conformation would bring them into hydrogen-bonding distance, consistent with an active kinase conformation. This EPHA2 crystal (PDB: 4PDO) contains two nearly identical autophosphorylation complexes (RMSD of 0.03 Å calculated over the backbone atoms of 480 residues, 240 in each chain). One complex contains two copies of chain A in different unit cells (asymmetric monomer chains A and A: 2_764), and the other contains two copies of chain B in different unit cells (chains B and B: 2_654) (Table 1). The biological assembly deposited in the PDB by the authors is the same as the asymmetric unit, which is composed of chains A and B, and does not represent a potential autophosphorylation complex. Although Wei *et al.* (23) note that the juxtamembrane segments of each chain are ordered, point away from the kinase domain, and contact neighboring chains in the crystal, they do not mention the presence of the autophosphorylation site of these segments in the active site of the neighboring kinase protein.

We compared the autophosphorylation complex of EPHA2 that we identified in PDB: 4PDO with two crystal structures of ephrin receptor A3 (EPHA3) bound to optimized substrate peptides [PDB: 3FXX and 3FY2 (18)]. Structure alignment of the peptides in EPHA3 (PDB: 3FXX and 3FY2) with the substrate kinase residues in the autophosphorylation complex of EPHA2 (PDB: 4PDO) revealed very similar backbone conformations of the substrate from P-4 to P+3 (Fig. 1G) [RMSD values of 0.48 and 0.98 Å over backbone atoms for EPHA3 (PDB: 3FXX and 3FY2) to EPHA2 (PDB: 4PDO) respectively]. Both the peptide substrates and the autophosphorylation substrate contain a Type VIII β turn consisting of residues P-2, P-1, P0, and P+1 (where P0 is the phosphorylation site). A β turn is defined as four

consecutive residues where the C α atoms of the first and fourth residues are less than 7.0 Å apart. A Type VIII turn occurs when residues 2 and 3 are in conformations in the α and β regions of the Ramachandran map respectively (37). The same type VIII β turn occurs in the CSF1R structure (PDB: 3LCD, residues 559-562 of the substrate kinase).

Superimposition of the enzyme kinase monomers of c-KIT, CSF1R, and EPHA2 showed that the substrate sequence adjacent to the phosphorylation site assumed similar positions in all three autophosphorylation complexes (Fig. 1H). The substrate regions of CSF1R, c-KIT, and EPHA2 make direct contacts with the enzymatic kinase from positions P-2 to P+9 in c-KIT and CSF1R and positions P-4 to P+9 of EPHA2 (see Table 3 for direct contacts). The c-KIT and CSF1R substrate sequences and structures in particular are very similar, and there are direct interactions of residues P+7, P+8 and P+9 (sequence Gln-Leu-Pro) of the substrate with the G helix of the enzyme in both structures (Fig. 1H, Table 3). In contrast, the kinase domains of the substrate kinases adopted different positions in both complexes of c-KIT (Fig. 1A, PDB: 1PKG), CSF1R (Fig. 1B, PDB: 3LCD), and EPHA2 (Fig. 1C, PDB: 4PDO), suggesting that interactions between the kinase domain of the enzyme monomer and the kinase domain of the substrate monomer are unlikely to be critical for autophosphorylation of these juxtamembrane Tyr residues.

To determine whether the autophosphorylation sites identified in these structurally related complexes are conserved in other kinases, we utilized BLAST (38) to compare the sequences of c-KIT, CSF1R, and EPHA2 with the sequences of all other human kinases, which we collected from the UniProt web site (39). With the program Clustal W (40), we aligned sequences with tyrosines in similar positions to those in the query sequences (Fig. 1I). The alignment showed that Tyr⁵⁶⁸ of c-KIT and Tyr⁵⁶¹ of CSF1R are in homologous positions, whereas Tyr⁵⁹⁴ of EPHA2 aligns with Tyr⁵⁷⁰ of c-KIT, which is also an autophosphorylation site (7). Although the kinase domains of CSF1R and c-KIT are 59% identical, in the juxtamembrane region surrounding the autophosphorylated tyrosine residues the proteins are 67% identical (18/27 identical residues), including many of the residues of the substrate region that contact the enzyme kinase (Table 3).

The autophosphorylation complexes illustrated in Fig. 1 provide information on 24 potential autophosphorylation sites in 20 different human kinases (Fig. 1I), many of which are annotated as autophosphorylation sites and known to be involved in the regulation of protein-protein interactions in cell signaling. In this aligned region, c-KIT, FLT3, and platelet-derived growth factor receptors α and β (PDGFRA, PDGFRB) contain two conserved tyrosines that are two residues apart, and autophosphorylation of these sites regulates their interactions with various proteins (41-53). The first of these sites corresponds to c-KIT Tyr⁵⁶⁸ and CSF1R Tyr⁵⁶¹, and the second is homologous to EPHA2 Tyr⁵⁹⁴. Inspection of the sequence alignment of the EPHA and EPHB family members showed that the Tyr⁵⁹⁴ site is conserved in human type A and Type B ephrin receptors (A1-A8 and B1,2,3,4, and 6) with the exception of the pseudokinase ephrin receptor A10 (EPHAA_HUMAN in Fig. 1I). Autophosphorylation at Tyr⁵⁹⁴ and Tyr⁵⁸⁸ in EPHA2 is required for the interaction of EPHA2 with the SH2 domains of the guanine nucleotide exchange factors VAV2 and VAV3 and for transduction of extracellular signals in vascular endothelial cells during tumor angiogenesis (54). Mutation of Tyr⁶⁰⁰ in EPHB1

(homologous to Tyr⁵⁹⁴ in EPHA2) to Phe abrogates the interaction of the receptor with kinase c-SRC and the adaptor Shc (55). The kinases KSYK and ZAP70 are autophosphorylated at Tyr³⁵² (56) and Tyr³¹⁹ (57) respectively, positions that are homologous to the EPHA2 Tyr⁵⁹⁴ site (Fig. 1I) and that control their activity in cell signaling events.

The ephrin kinases have another autophosphorylation site [Tyr⁵⁸⁸ in EPHA2 (54)] six residues before Tyr⁵⁹⁴. These two sites share a common sequence motif {TVIP}XYXDP. The P+3 residue at both these sites is proline (Pro⁵⁹¹ in EPHA2); therefore, the PDB: 4PDO structure of EPHA2 with may serve as a model for the phosphorylation of similar nearby sites in the ephrin receptor kinases.

A comparison of the c-KIT, CSF1R, and EPHA2 autophosphorylation structures, as well as the EPHA3:substrate-peptide complex, indicates a particular element that defines the specificity of these kinases for their phosphorylation sites (Fig. 2). In all four structures, the residue at P+3 is a hydrophobic amino acid that is in van der Waals contact with several hydrophobic amino acids near the C-terminal end of the activation loop and at least one residue on the G helix of the enzyme kinase. In c-KIT, the P+3 residue is Ile⁵⁷¹, making contact with Val⁸³³ of the activation loop, Ile⁸⁴¹ immediately following the activation loop, and Tyr⁸⁸⁰ of the G helix (Fig. 2A). In CSF1R, Ile⁵⁶⁴ at P+3 contacts activation-loop residues Val⁸¹⁹ and Met⁸²², Ile⁸²⁷ after the activation loop, and G-helix residue Tyr⁸⁶⁶ (Fig. 2B). Similarly, in EPHA2, the P+3 residue is Pro⁵⁹⁷, which contacts Ile⁷⁷⁹ and Ile⁷⁸¹ of the activation loop, Ile⁷⁸⁹ following the loop, and Met⁸²⁷ of the G-helix (Fig. 2C). Similarly, the in the structures of EPHA4 with optimized peptides (PDB: 3FY2 and 3FXX), the P+3 residue is an isoleucine making contact with Ile⁷⁸⁸ of the activation loop, Ile⁷⁹⁶ just after the activation loop, and Ile⁸³⁴ of the G helix (Fig. 2D).

Autophosphorylation sites in the activation loops of kinases

In our search of kinase interfaces in the PDB, we identified an autophosphorylation complex of LCK [PDB: 2PL0 (20)] in which Tyr³⁹⁴ of the activation loop of the sole monomer in the asymmetric unit (chain A) is positioned in the active site of chain A:2_655 and vice versa in a head-to-tail dimer (Fig. 3A). Jacobs *et al.* mention this interaction but dismiss its biological relevance because of the bound inhibitors; they did not show the complex and they deposited a monomer as the biological assembly in the PDB (20). The complex that we identified strongly resembles a previously described structure of an autophosphorylation complex of Tyr¹¹⁶⁵ in the activation loop of IGF1R [PDB: 3D94 (10)], which is also a symmetric head-to-tail structure with Tyr¹¹⁶⁵ of one monomer in the active site of another monomer in a neighboring asymmetric unit in the same unit cell (chain A:4_555) and vice versa (Fig. 3B). Wu *et al.* verified the importance of the intramolecular hydrogen-bonding interactions within the activation loop of IGF1R in this structure by introducing the mutations E1162A, R1167, and R1167A in the activation loop, which all reduced autophosphorylation of Tyr¹¹⁶⁵ (10).

In both the LCK and IGF1R structures, the activation loops are in unusual, extended conformations that are not typical of other either active or inactive structures of LCK (Fig. 3C) or IGF1R (Fig. 3D). The majority of the activation loop of LCK (PDB: 2PL0) is in a

conformation that does not resemble inactive (PDB: 2OFV, 2OG8, 3BYS, 3BYU) or active structures of LCK (PDB: 1QPC, 1QPD, 1QPE, 1QPJ, 2OF2, 2OF4, 2OFU, 2ZM1, 2ZM4, 2ZYB, 3AX1, 3AC1, 3AC2, 3AC3, 3AC4, 3AC5, 3AC6, 3AC8, 3ACJ, 3ACK, 3AD4, 3AD5, 3AD6, 3BYM, 3BYO, 3KMM, 3KXZ, and 3LCK). Instead, it is extended away from the body of the kinase domain (green in Fig. 3C). For IGF1R, inactive kinases (PDB: 1M7N, 1P40, 2OJ9, 3I81, 3LW0, 3NW5, 3NW6, 3NW7, 3QQU, and 4D2R) usually have the activation loop folded against the kinase with Tyr¹¹⁶⁵ intramolecularly hydrogen bonded to the active site Asp¹¹³⁵ of the HRD motif (58). Active IGF1R kinase domains (PDB: 1K3A, 2ZM3, 3F5P), like other kinases, have the activation loop pointed toward the right when the kinase is oriented with the N-terminal domain above the C-terminal domain and looking into the active site. The activation loop of IGF1R (PDB: 3D94) (green in Fig. 3D) points outward, away from both active and inactive kinase loops (magenta and yellow, respectively, in Fig. 3D).

The symmetric autophosphorylation complexes of LCK (PDB: 2PL0) and IGF1R (PDB: 3D94) do not contain active kinase monomers in conformations that we would expect of a potential autophosphorylation complex (Table 2). For example, they are not in the DGF-in conformation like the conformations observed for the c-KIT (PDB: 1PKG), CSF1R (PDB: 3LCD), and EPHA2 (PDB: 4PDO) complexes (Table 2). The regulatory spines are not intact due to the position of the phenylalanine side chain of the DFG motif. In the imatinib-bound LCK structure (PDB: 2PL0), the starting residues of the activation loop are in an inactive DFG-out conformation, as noted by Jacobs *et al.* (20), and the regulatory spine is therefore not intact (Table 2). In the IGF1R structure, the DFG motif is in what is sometimes called a “DFG-up” conformation (59), in which the phenylalanine ring points upwards into the N-terminal domain instead of pointing inwards and under the C-helix as it does in active kinase structures. The Glu-to-Lys distance of the N-terminal domain is 9.25 Å in monomer A (enzyme and substrate kinase in different copies of the asymmetric unit), which is inconsistent with an active kinase.

We hypothesized that the autophosphorylation complexes for LCK and IGF1R likely consist of asymmetric dimers with one active kinase monomer and one kinase in a conformation resembling one of the monomers in the symmetric structures in PDB: 2PL0 and PDB: 3D94. To investigate this, we built models of asymmetric complexes by superposing active monomers onto one of the proteins in the symmetric dimer complexes followed by energy minimization with the program Gromacs (60). We used the active monomer from PDB: 1QPC (61) to successfully model LCK (Fig. 3E) and from PDB: 1K3A (17) to successfully model IGF1R (Fig. 3F) without significant clashes and with very minimal structural changes during energy minimization (data file S2: Data File_S2_IGF1R_LCK_models.zip). Table 2 provides information on the structural features of these asymmetric complexes (data marked with asterisks).

Given the physiological importance of the autophosphorylation of the activation loop of tyrosine kinases, we performed mutagenesis experiments to test the biological relevance of the interface between the LCK monomers that were present in both the structure from PDB: 2PL0 (Fig. 3A) and the modeled structure (Fig. 3E). The primary contact in this interface, apart from the interaction of the two activation loops, is between the activation loop of one

monomer and the G-helix of the other monomer and vice versa (Fig. 4A). Because mutations of the activation loop might affect autophosphorylation, we made the following point mutations on the G-helix side of the interface: T445V, N446A, P447A, P447L, P447G, and Q451E.

To assess LCK autophosphorylation activity, we expressed full-length LCK tagged with Myc in HEK293 cells, which lack endogenous LCK (62), and monitored the production of the protein by Western blot with an antibody against the Myc epitope and the amount of phosphorylated LCK with an antibody recognizing LCK phosphorylated at Tyr³⁹⁴, which is in the activation loop (Fig. 4B). Because the amounts of the LCK proteins varied, we normalized the signal obtained for the phosphorylated LCK to the signal for the total protein (Myc antibody), setting the ratio of these in the samples from cells expressing wild-type LCK as 1 (see table S1 for complete data). To calculate the average ratio and a confidence interval for each mutant relative to wildtype, we transformed the ratios by taking the log, calculated the average log and its 95% confidence interval, and transformed back with the anti-log (63). With this approach, we calculated the relative autophosphorylation activity of the mutants to that of the wild-type LCK expressed in HEK293 cells (Fig. 4C).

Thr⁴⁴⁵ is a helix N-cap residue, forming a side-chain to backbone hydrogen bond with Glu⁴⁴⁸ in the asymmetric unit monomer of LCK in PDB: 2PL0 (Fig. 4A). Threonine is a common residue in this position of an N-cap, and valine is the least favored amino acid at this position (64). A mutation to Val is likely to disturb the conformation of the helix N-cap and, therefore, the interactions of the G-helix with the activation loop of the other monomer. Indeed, we found that the T445V mutant had substantially reduced autophosphorylation activity (Fig. 4B, C). Asn⁴⁴⁶ of the G-helix forms a hydrogen bond with the backbone oxygen of Thr³⁹⁵ of the activation loop (Fig. 4A). Consistent with the importance of this interaction in stabilizing the autophosphorylation complex, the N446A mutant exhibited a substantial reduction in autophosphorylation activity (Fig. 4B, C).

The second residue of the G-helix is Pro⁴⁴⁷. This residue is mutated from Pro to Leu in a T-cell leukemia cell line and is associated with constitutive activation of LCK (29). We individually mutated this residue to Ala, Leu, or Gly. The 95% confidence levels show that each of the three mutations resulted in an increase in autophosphorylation activity relative to wildtype (Fig. 4C, table S1). This is possibly due to a loosening of the helix N-cap so that specific interactions were more easily formed. We also mutated Gln⁴⁵¹ to Glu. This residue forms a hydrogen bond with Glu³⁹⁸ of the activation loop of the other monomer; thus, we expected charge-charge repulsion to reduce the formation or stability of the complex, which we predicted would reduce autophosphorylation activity. However, this mutant had similar, although more variable, autophosphorylation activity as the wild-type LCK (Fig. 4C). We hypothesize that this autophosphorylation activity resulted from a conformational change of the side chain of the mutated residue enabling access to solvent. The data from the mutant LCK autophosphorylation analysis indicated that the interaction between Gln⁴⁵¹ and Glu³⁹⁸ was not necessary for formation of the autophosphorylation complex, but that the interaction between Asn⁴⁴⁶ and Thr³⁹⁵ was important for formation of the autophosphorylation complex.

IGF1R is autophosphorylated at Tyr¹¹⁶¹, Tyr¹¹⁶⁵, and Tyr¹¹⁶⁶ of the activation loop (17, 65). We identified a previously undescribed autophosphorylation complex of IGF1R [PDB: 3LVP (21)] in which Tyr¹¹⁶⁶ of the activation loop of chain B in the asymmetric unit is in the active site of chain C (Fig. 5A). This structure has four IGF1R monomers in the asymmetric unit. Monomers C and D have activation-loop conformations similar to the conformations of the triphosphorylated activation loops in active IGF1R kinases [PDB: 1K3A (17), PDB: 2ZM3 (66), PDB: 3F5P (67)], and have properties consistent with active structures (Table 2). Nemecek *et al.* (21) described the IGF1R structure in PDB: 3LVP as a bis-phosphorylated IGF1R in complex with a bis-azaindole compound. However, the coordinates for the phosphate groups are not in the PDB file and the electron density map (68) does not indicate the presence of phosphate groups. The activation loop of monomer B (substrate) is in a conformation not observed in other structures of IGF1R (Fig. 3D). The loop conformation extends away from the kinase domain such that Tyr¹¹⁶⁶ points outwards and sits in the active site of monomer C of the asymmetric unit in the manner of a kinase substrate. The activation loop of monomer A is disordered between residues 1161 and 1169, although it has a similar conformation to monomer B for the residues that have coordinates in the PDB file. A comparison of the substrate:enzyme complex B:C and a complex of chain A and a symmetry copy of chain D (chain D:4_555) showed that the two complexes are similar (RMSD = 0.68 Å calculated for 503 residues in the two chains) (Fig. 5B). Thus, this crystal exhibits two versions of the same autophosphorylation complex, although in one of them the activation loop and the phosphorylation site of the substrate chain A are disordered.

The orientations of the monomers in the autophosphorylation complex of Tyr¹¹⁶⁶ of IGF1R (PDB: 3LVP) differ substantially from that of the autophosphorylation complex of Tyr¹¹⁶⁵ IGF1R (PDB: 3D94). The Tyr¹¹⁶⁵ complex is a head-to-tail dimer with the substrate activation loop extended away from the kinase monomers (Fig. 3B). The Tyr¹¹⁶⁶ complex is an asymmetric head-to-head dimer with interactions between the two C-terminal domains of the kinases (Fig. 5A). The superposition of the IGF1R Tyr¹¹⁶⁵ and Tyr¹¹⁶⁶ and the LCK tyrosine autophosphorylation complexes, along with a peptide-IGF1R complex [PDB: 1K3A (17)] and two peptide-INSR complexes [PDB: 3BU5 (69) and 1IR3 (70) (INSR and IGF1R are 80% identical in the kinase domain)], reveals the similarity of the position of the Tyr phosphorylation site and its two neighboring residues on either side (Fig. 5C).

However, the positions of the residues further away from the Tyr site in the autophosphorylation complexes differ substantially from those in the peptide-substrate structures. The peptide substrates C-terminal to the phosphosite (PDB: 1K3A, 3BU5, and 1IR3) form extended β sheets with the activation loop (green, cyan, and yellow β sheet strands in Fig. 5C), whereas the substrates in the autophosphorylation complexes loop away from the enzyme kinase. This is likely due to differences in the P+3 amino acid type in the peptides versus the activation loops (Fig. 5D). We noted above that the juxtamembrane substrates and an EPHA3-optimized substrate contain a hydrophobic residue at position P+3 that interacts with a pocket consisting of three hydrophobic residues from the activation loop, a region just after the activation loop, and a residue in the G-helix (Fig. 2). These same interactions occur for the IGF1R (PDB: 1K3A) and INSR peptide substrates (PDB: 3BU5 and 1IR3), but do not occur for the activation loops of LCK and IGF1R (Fig. 5D). For LCK

(PDB: 2PL0), the P+3 residue is Arg³⁹⁷, which makes only intramolecular interactions with the substrate activation loop. For IGF1R-1165 (PDB: 3D94) and IGF1R-1166 (PDB: 3LVP), the P+3 residues are Lys¹¹⁶⁸ and Gly¹¹⁶⁹ respectively, neither of which makes direct interactions with the enzyme kinase (Fig. 5D).

The activation loop autophosphorylation complexes and the peptide substrates of tyrosine kinases do have one feature in common – a polar side chain at P-1 that makes hydrogen-bonding interactions with positively charged residues of the enzyme kinase (Fig. 5C). In LCK (PDB: 2PL0), this is a salt-bridge of Glu³⁹³ at P-1 with Arg³⁶⁶, which is located two residues after the HRD motif (HRD+2). The substrate Tyr residue also forms a hydrogen bond with Arg³⁶⁶. LCK shares an Arg at the HRD+2 position and a Glu at the P-1 position of the activation loop Tyr autophosphorylation site with other type A and type B SRC family kinases (FGR, FYN, HCK, LYN, SRC, YES, and BLK). In the two IGF1R autophosphorylation complexes (PDB: 3D94 and 3LVP) (Fig. 5C), the interaction of the P-1 residue (Glu¹¹⁶⁴ and Tyr¹¹⁶⁵, respectively) is with a lysine residue in the D helix (Lys¹⁰⁸⁸). The peptide substrates exhibit a homologous interaction of P-1 with the same Lys in the D helix of IGF1R (Lys¹⁰⁸⁸ in PDB: 1K3A) and a homologous Lys in the D helix of INSR (Lys¹⁰⁸⁵ in PDB: 3BU5 and 1IR3) (Fig. 5C).

Most tyrosine kinases contain one or more tyrosine residues in the activation loop that are autophosphorylated during the course of kinase activation (71, 72). We produced an alignment of the activation loops of all 90 human tyrosine kinases beginning with the DFG motif and ending with the APE motif (fig. S2). The two similar phosphorylation structures, LCK in PDB: 2PL0 and the Tyr¹¹⁶⁵ autophosphorylation complex of IGF1R in PDB: 3D94, contain a substrate tyrosine at position 13 (counting the DFG motif as residues 1-3); the Tyr¹¹⁶⁶ in the IGF1R complex (PDB: 3LVP) is at position 14 of the activation loop. Sixty-five of the 90 human tyrosine kinases contain a tyrosine at sites +13, +14, or both positions (fig. S2).

For 21 kinases with a tyrosine at +13, but not at +14, 17 are annotated in UniProt as autophosphorylated, two are phosphorylated by other kinases, and two are unannotated. Of the 16 kinases with a tyrosine at +14, but not at +13, 14 of them are annotated as autophosphorylated, one is unspecified, and one is unannotated (the pseudokinase EPHA10). Six kinases have a third tyrosine in this region at +15: SYK, TYK2, ZAP70, JAK1, JAK2, and JAK3. Of the 27 kinases with tyrosines at both positions 13 and 14, 17 have both tyrosines annotated as phosphorylation sites (all but one by autocatalysis), 5 have annotations of autocatalysis at one site but not the other, and only 3 are unannotated at both positions. The two IGF1R structures (Fig. 3B and 5A) and the LCK structure (Fig. 3A) provide hypothetical models for the autophosphorylation at sites +13 and +14 (relative to the DFG motif) in a majority of human tyrosine kinases. It is possible that kinases with Tyr residues at positions 13 or 14, or both positions, may use either of the two orientations that we observed: head-to-tail for IGF1R (PDB: 3D94) and LCK (PDB: 2PL0) or head-to-head for IGF1R (PDB: 3LVP). It is not necessarily the case that position 13 is head-to-tail and position 14 is head-to-head for all Tyr kinases.

In our search of kinase interfaces, we identified a set of 10 crystal structures that contain autophosphorylation complexes of the activation loop Thr⁴²³ of p21-activated kinase 1 (PAK1) (Fig. 6, Table 1, Table 2). For one of these, PAK1 (PDB: 3Q4Z), Wang *et al.* described the autophosphorylation complex, a head-to-head asymmetric complex of PAK1 in the crystal asymmetric unit (11), whereas the others were not shown or discussed in the relevant publications (25-28). Wang *et al.* (11) made a number of mutations of both the enzyme kinase (PAK1-KD^{K299R}) and a kinase-dead substrate kinase (PAK1-KD^{K299R,D389N}), demonstrating asymmetric effects on autophosphorylation rates that were consistent with the structure of the interface in PAK1 (PDB: 3Q4Z). The newly identified structures provide additional detail and insights not available in PAK1 (PDB: 3Q4Z).

These structures fall into two crystal forms, which we refer to as crystal form 1 and crystal form 2. Within each crystal form, the PAK1 monomers form lattices of protein-protein interactions that are similar within each crystal. Crystal form 1 (Fig. 6A) comprises PDB: 3Q4Z (11); 4O0R, 4O0T (25); 4P90 (26); 4ZY4, 4ZY5, 4ZY6 (27); and 4ZLO (28). Crystal form 2 (Fig. 6B) comprises PDB: 4ZJI and 4ZJJ (28). In Fig. 6A and 6B, the enzyme monomers (blue) are oriented identically. Comparison of the two images shows that the position of the substrate kinase domains (magenta) in crystal form 2 are shifted downwards relative to their position in crystal form 1, whereas the substrate autophosphorylation site of the activation loop still reaches the active site of the enzyme kinase. The contacts between the G-helices (residues 469-478) are different in the two crystal forms: In crystal form 1, the side chains of Pro⁴⁶⁹, Leu⁴⁷⁰, and Leu⁴⁷³ of the G-helix of the enzyme kinase are in contact with the side chains of Leu⁴⁷³, Tyr⁴⁷⁴, and Ala⁴⁷⁷ of the substrate kinase; in crystal form 2, Tyr⁴⁷⁴, Ala⁴⁷⁷, and Thr⁴⁷⁸ of the enzyme kinase make contact with Pro⁴⁶⁹, Leu⁴⁷⁰, and Leu⁴⁷³ of the G-helix of the substrate kinase.

For the two crystal forms, we examined the positions and conformations of the activation loops of the enzyme (Fig. 6C) and substrate monomers (Fig. 6D). In crystal form 1, the activation loops of the enzyme kinases are in active DFG-in conformations in all of the structures, and the entire loop is fully ordered except in PAK1 (PDB: 4P90) (Fig. 6C). In crystal form 2, however, the activation loops of the enzyme kinases are in DFG-out conformations due to the presence of inhibitors in the Type II pocket between the C-helix and the N-terminal domain β sheet (Fig. 6C); they are also all partially disordered. The activation loop of the substrate kinases is fully ordered in two of the newly recognized structures of crystal form 1 [PDB: 4ZY4 and 4ZY5 (27)], but not in the others, including the original structure PAK1 (PDB: 3Q4Z) (Fig. 6D). The C-terminal region of the activation loop of the crystal form 1 substrate kinases are all similar to one another, whereas the N-terminal residues 405-408 exhibit some differences. Five of the structures are in DFG-up inactive conformations (PDB: 3Q4Z, 4O0R, 4ZY4, 4ZY5, 4ZY6), two are DFG-out (PDB: 4P90, 4ZLO), and one is DFG-in (PDB: 4O0T). This may indicate that autophosphorylation of the activation loop of PAK1 can occur no matter which conformational state the beginning of the activation loop is found in. The activation loops of the substrate kinases of crystal form 2 are all DFG-out conformations and are partially disordered from residue ~411 to 415. They differ substantially in the C-terminal segment and beyond the end of the activation loop (position 434) all the way to position 443 (Fig. 6D).

To verify the substrate interactions within the PAK1 autophosphorylation complexes, we compared the substrate segments of representative structures from each of the crystal forms of PAK1 (PDB: 4ZY4 and 4ZJI) with PAK4 structures with bound peptides [PDB: 2Q0N, 4JDH, 4JDI, 4JDJ, and 4JDK (73)] with hydrogen bonds between the phosphorylation site Ser or Thr and the HRD motif Asp⁴⁴⁰ (Fig. 6E). After superposition of the enzyme kinases, the backbone positions of the substrates in the PAK1 autophosphorylation complexes and the PAK4-bound peptides overlap very well from positions P-3 to P+1, but deviate substantially before and after this region. The PAK1 and PAK4 structures share one feature of specificity, which is an interaction of Arg at P-2 with two negatively charged residues in the enzyme kinase, which is a common feature of many Ser/Thr kinases (74) (Fig. 6E).

We compared enzyme and substrate conformations of crystal form 1 of PAK1 to determine where and how the substrate kinase's activation loop differs from active structures (Fig. 6F). Working backwards from the end of the activation loop, the PAK1 substrate activation loop deviates from the enzyme activation loop at Pro⁴²⁸ and Gly⁴²⁶, which are 7 and 9 residues from the end of the loop, respectively. The values of the backbone dihedrals ϕ and ψ of these residues are consistent with this: For instance, ψ of Pro⁴²⁸ in the enzyme and substrate kinases is -61° and $+61^\circ$, respectively, in PAK1 (PDB: 4ZY4), and the ϕ of Gly⁴²⁶ is 129° and 77° in the enzyme and substrate kinases, respectively.

For some purposes, such as molecular dynamics simulations, it is important to have complete coordinates without chain breaks in the backbone due to disorder. Most of the structures in crystal form 1 have such breaks within the N-terminal domain between or adjacent to β -sheet strands. Only the enzyme kinase of PAK1 (PDB: 4O0R, chain A) and the substrate kinases of PAK1 (PDB: 4O0T, chain B and 4ZY6, chain B) are complete in the N-terminal domain.

Given the inactive characteristics of the enzyme kinase structures (Table 2) and displacement of the N-terminal region of the enzyme kinase activation loop (Fig. 6C) and the C-terminal regions of the activation loops of both the enzyme and substrate kinases activation loop (Fig. 6C and 6D), it appears as if the crystal form 1 structures are better models of the autophosphorylation complex, but we cannot rule out that both crystal forms may contain relevant information for the autophosphorylation of PAK1 Thr⁴²³.

Two human IRAK4 structures (PDB: 4U97 and PDB: 4U9A) (15) also have an asymmetric enzyme-substrate dimer in the asymmetric unit with the autophosphorylation of an activation loop threonine (Thr³⁴⁵) (Fig. 6G), as discussed by Ferrao *et al.* (15). The sequence alignment of IRAK family members showed that this threonine is conserved in all four IRAKs (fig. S3). The enzyme kinases are similar to an active structure of IRAK4 [PDB: 2NRY (75)] and the substrate kinases are in the DFG-up conformations (Table 2).

The autophosphorylation complexes of PAK1 in crystal form 1 (Fig. 6A) and crystal form 2 (Fig. 6B) can be compared with that of IRAK4 (PDB: 4U97 and 4U9A) (Fig. 6G), because the enzyme kinases are oriented similarly in all three figures. Ferrao *et al.* noted the difference between PAK1 (PDB: 3Q4Z) of crystal form 1 and their structures of IRAK4 (PDB: 4U97 and 4U9A) (15). We determined that the IRAK4 and PAK1 crystal form 2

structures are also different. The phosphorylation sites of the PAK1 and IRAK4 structures are 12 and 14 residues before the end of the activation loop, respectively, indicating that the sites are not completely homologous. In PAK1, the conformational change of the activation loop between the enzyme and substrate kinases occurs at Pro³²⁸ and Gly³²⁶, which are located 7 and 9 residues from the end of the activation loop, respectively. For IRAK4, working backwards from the end of the activation loop, the first large changes occur at Thr³⁵¹ and Gly³⁵⁰; the ϕ dihedral of Thr³⁵¹ changes from -121° to -64° from enzyme to substrate (PDB: 4U9A), and the ϕ dihedral of Gly³⁵⁰ changes from $+155^\circ$ to -56° from enzyme to substrate (Fig. 6H). This Gly-Thr motif (residues 350-351) located at -8 and -9 residues before the end of the activation loop is conserved in many kinases. The change in conformation of PAK1 occurs at a similar location (Fig. 6F).

As we noted above, the autophosphorylation sites across many tyrosine kinases are at consistent locations (positions 13 and 14) relative to the start of the activation loop at the DFG motif. We sought to determine if this was also the case for Ser/Thr kinases, which fall into several different families. A sequence alignment of the activation loop of Ser/Thr kinases (fig. S4) shows that 212 human kinases contain a Ser or Thr residue at a position aligned with the autophosphorylation site Thr⁴²³ of PAK1. This position is 12 residues from the end of the activation loop (counting the APE motif as residues 3, 2, and 1, respectively). Forty-three of the 212 human kinases with Ser or Thr at this position are annotated as having these sites as autophosphorylation sites in Uniprot (fig. S4). Thus, regardless of insertions and deletions at positions before the phosphorylation site, the crystal form 1 (and possibly crystal form 2) structures of PAK1 may represent autophosphorylation complexes for many of the kinases in fig. S4.

Autophosphorylation complexes of FGFR kinase domains

Several receptor tyrosine kinases have a region called the kinase insert domain, which is a long loop segment inserted between the D and E helices of the C-terminal domain. Bae *et al.* [PDB: 3GQI (8)] and Huang *et al.* [PDB: 4K33 (14)] described autophosphorylation complexes of homologous sites in the kinase insert domains of FGFR1 and FGFR3, respectively. The asymmetric dimer structure of FGFR1 with mutations in three autophosphorylation sites (Y583F, and Y585F) [PDB: 3GQI (8)] contains an autophosphorylation site mutated from Tyr⁵⁸³ to Phe⁵⁸³ of the kinase insert domain of one monomer that sits in the active site of a second monomer in a neighboring unit cell (A: 4_456). Using site-directed mutagenesis, Bae *et al.* demonstrated that domain-domain contacts between each monomer (PDB: 3GQI) are required for autophosphorylation of the kinase insert domain (76). Huang *et al.* (14) reported that the FGFR3 structure (PDB: 4K33) contains an autophosphorylation complex of Tyr⁵⁷⁷ of the kinase insert loop ([a known autophosphorylation site (77)], consisting of two symmetry copies of chain A of the asymmetric unit in neighboring unit cells (chains A:3_645 as substrate and chain A:1_555 as enzyme kinase).

We compared the autophosphorylation complexes of FGFR1 and FGFR3 and observed a previously unrecognized similarity of these two structures, both around the substrate autophosphorylation site and in domain-domain interactions distal to the substrate-binding

site (Fig. 7A). The RMSD between the complexes is 0.84 Å (calculated for the backbone atoms of 510 residues, 255 in each chain). Both complexes contain an ATP analog, adenosine-5'-[β,γ-methylene]-triphosphate, and exhibit DFG-in conformations and intact salt bridges in both monomers (Table 2). In both structures, the long loop that connects the F and G helices in the C-terminal domain of the substrate kinase makes extensive contact with the loops of the N-terminal domain β sheet of the enzyme kinase. The sequences of the αF-αG loop are similar and make similar contacts in the two complexes (Fig. 7B). An alignment of sequences of the kinase insert domains shows that FGFR1-Tyr⁵⁸³, FGFR3-Tyr⁵⁷⁷, as well as FGFR2-Tyr⁵⁸⁶ (a known autophosphorylation site) are homologous to one another (fig. S5). FGFR4 does not contain a tyrosine at this position.

Huang *et al.* compared the substrate binding in FGFR3 (PDB: 4K33) with their earlier structure of an autophosphorylation complex of Tyr⁷⁶⁹ of the C-terminal tail of FGFR2 [PDB: 3CLY (9)] (Fig. 7C), a site that is important for binding of phospholipase C γ (PLCγ) (78). All four FGFR protein sequences contain a tyrosine phosphorylation site at the same position as Tyr⁷⁶⁹ of FGFR2 (fig. S6). Indeed, the autophosphorylation complex of the kinase insert of FGFR1 (PDB: 3GQI) also contains PLCγ bound to the C-terminal tail of FGFR1 with a phosphorylated Tyr⁷⁶⁶, which is homologous to Tyr⁷⁶⁹ of FGFR2. A superposition of FGFR1 and FGFR2 shows the different positions of the phosphorylation site in the C-terminal tail in the PLCγ-bound structure (PDB: 3GQI) and the autophosphorylation complex (PDB: 3CLY) (Fig. 7D). We also superposed the only other FGFR kinase domain structure that contains the C-terminal phosphorylation site in the coordinates, FGFR1 [PDB: 3GQL (8)], which contains the Tyr⁷⁶⁶ bound to the kinase insert of the same monomer (Fig. 7D). We also superposed structures of RON [PDB: 3PLS (79)] and MET [PDB: 1R0P (80)], which contain Tyr¹³⁵³ and Tyr¹³⁴⁹ residues, respectively, at homologous positions to Tyr⁷⁶⁶ of FGFR1. The superposition showed that in the structures in which the Tyr is not bound to another protein (either kinase or PLCγ), the C-terminal Tyr residues are in the same position up against the E helix. In the two bound structures (PDB: 3GQI and 3CLY), the Tyr residues occupy different positions from each other and from the unbound Tyr residues (Fig. 7D).

Autophosphorylation of long disordered tails away from the kinase domain

Five structures listed in Table 1 contain autophosphorylation complexes of N- and C-terminal tails such that the phosphosite is at least 20 amino acids away from the start or end of the folded kinase domain (that is the first β strand of the N-terminal domain or the I-helix, which is the last helix of the C-terminal domain). These structures have in common that the folded domains of the enzyme and substrate kinases make little or no contact with each other; the substrates bind as if they were peptides. These structures are an autophosphorylation complex of Ser¹⁴² from an N-terminal region of CLK2 (24); three autophosphorylation complexes of the C-terminal tail of CAMKII, one from human CAMKII subunit δ (12) and two from *C. elegans* CaMKII (13); and an autophosphorylation complex of Tyr¹⁰¹⁶ in the C-terminal tail of EGFR (19). The CLK2 structure has not previously been discussed as an autophosphorylation complex (24), whereas the CAMKII structures (12, 13) and the EGFR (19) structure have been previously described as such.

CLK2 is a dual-specificity kinase that is autophosphorylated at Ser¹⁴² (in the human sequence; Ser¹⁴¹ in the mouse sequence). When unphosphorylated at this site, mouse CLK2 resides in nuclear speckles, whereas autophosphorylation at Ser¹⁴¹ results in localization within the nucleoplasm (81). This region of the protein is N-terminal to the beginning of the kinase domain, which begins at residue 163. We identified a structure of human CLK2 [PDB: 3NR9 (24)] that contains a symmetric intermolecular interaction of Ser¹⁴² of chain C of the central asymmetric unit with the active site of chain B of a neighboring asymmetric unit in the same unit cell of the crystal (chain B:5_555), and vice versa (Fig. 8A). There is a similar dimer between chain A of the asymmetric unit and chain A of chain A:4_645 (backbone RMSD 1.0 Å calculated for 680 residues in each complex). The enzyme-substrate complex comprising chains B and C has a slightly longer distance (5.36 Å) between the hydrogen-bond donor (Ser¹⁴²) and the hydrogen-bond acceptor (Asp²⁹⁰) than the other autophosphorylation complexes (Table 1). However, the position of the main chain of Ser¹⁴² is consistent with a kinase substrate (Fig. 8B), as made evident by superposition of chain B: 5_555 with the kinase domain of a structure of human dual-specificity tyrosine-phosphorylation-regulated kinase 1A (DYRK1A) with a bound peptide from crumbs homolog 2 (CRB2) [PDB:2WO6 (82)]. This structure is the closest peptide-kinase complex of known structure to CLK2 in the PDB (35% identity). A common element of substrate-enzyme interaction is the P-3 position, which is an Arg residue in both CLK2 (Arg¹³⁹) and in the DYRK1A-bound peptide, and this Arg interacts with homologous Glu residues [Glu²⁹⁴ and Glu³⁷³ in CLK2 (PDB: 3NR9); Glu²⁹¹ and Glu³⁵³ of DYRK1A (PDB: 2WO6)]. This interaction has been noted for several Ser/Thr kinases (74).

This phosphorylation site is conserved in the four human CLK kinases (fig. S7). Although the only structure of CLK2 available is PDB: 3NR9, structures of CLK3 include Ser²⁸³, which is homologous to Ser¹⁴² of CLK2. A superposition of these CLK3 structures [PDB: 2WU6, 2WU7 (83) and 3RAW (84)] with CLK2 (PDB: 3NR9) demonstrated that the phosphorylation site moves 27 Å from its position in the nonautophosphorylation complexes to its position in the potential autophosphorylation complex in CLK2 (PDB: 3NR9) (Fig. 8C). The N-terminal tails (residues 136-155) about the phosphorylation site (Ser¹⁴²) are sandwiched between the enzyme and substrate monomers, which otherwise make essentially no contact with each other apart from the tails (Fig. 8A).

Three structures have been described that exhibit an autophosphorylation complex of the Thr residue of the regulatory region of CaMKII. PDB entries 3KK8 and 3KK9 (13) are structures of *C. elegans* CaMKII that contain the kinase domain and a portion of the R1 regulatory region, whereas human CaMKII subunit δ (PDB: 2WEL) contains the structure of a longer construct that includes the calmodulin-binding site of the regulatory region (12). The autophosphorylation site in the regulatory region is conserved within the four human CaMKII isoforms (fig. S8).

The substrate kinase domains are in three distinct positions relative to the enzyme kinase domains in these structures (Fig. 8D), indicating that interaction of the kinase domains, apart from the C-terminal tails is not likely to be important for autophosphorylation. As with the other N- or C-terminal tails that are distant from the kinase domain in sequence (CLK2

described above and EGFR described below), the enzyme and substrate kinase domains make little contact in these structures, particularly in CaMKII PDB: 3KK9 and 2WEL.

Chao *et al.* compared their structure of *C. elegans* CaMKII autophosphorylation with a structure that they determined of calcium/calmodulin-dependent protein kinase II inhibitor 1 bound to *C. elegans* CaMKII [PDB: 3KL8 (13)]. This inhibitor binds like a substrate peptide but has an Arg residue (instead of Ser or Thr) at the position of the phosphorylation site. Instead, we compared the interaction of the residues adjacent to the autophosphorylation site of the substrate kinases with the enzyme kinases in these structures with a peptide bound to rabbit phosphorylase kinase [PDB: 2PHK (85)], which is the closest kinase-peptide-substrate complex in the PDB to both *C. elegans* and human CaMKII (44% sequence identity). A common substrate-enzyme interaction in the peptides and the CaMKII autophosphorylation complexes occurs at the P-3 positions, which consists of an Arg residue that interacts with one or both of two Glu residues in or near the D helix (Fig. 8E, colored as in Fig 8D; peptide substrate in pink). Furthermore, hydrophobic residues at the P+1 position in the peptides and autophosphorylation substrates interact with hydrophobic residues near the end of the activation loop (4 and 8 residues from the end of the activation loop).

In structures of these enzymes that are not autophosphorylation complexes, the phosphorylation site of CaMKII interacts with the substrate kinase domain, rather than extending away from the kinase and toward an enzyme kinase within the crystal. Indeed, the phosphorylation sites in the nonautophosphorylation complexes of *C. elegans* CaMKII [PDB: 2BDW (86)] and human CaMKII subunit δ [PDB: 2VN9 (12)] are buried between the H and I helices (Fig. 8F).

EGFR catalyzes the trans-autophosphorylation of several tyrosine residues in a disordered region C-terminal to the kinase domain (residues 712-979), including Tyr⁹⁹⁸, Tyr¹⁰¹⁶, Tyr¹⁰⁹², Tyr¹¹¹⁰, Tyr¹¹⁷², and Tyr¹¹⁹⁷ (87-90). The structure of residues 694-1022 of EGFR with the inhibitory protein Mig-6 (PDB: 4I21) contains a potential autophosphorylation complex described as such by Gajiwala *et al.* (19). In this structure, the C-terminal tail of chain B in the asymmetric unit sits in the active site of chain A, such that Tyr¹⁰¹⁶ forms a hydrogen bond with Asp⁸³⁷ of the HRD motif in the active site (Fig. 8G). Additionally, the C-terminal tail of the enzyme kinase in the asymmetric unit (chain A) sits in the active site of chain B of a neighboring unit cell (chain B:1_655). These two complexes are similar (RMSD 2.7 Å calculated for backbone atoms of both chains excluding the C-terminal tails), and only the one contained in the asymmetric unit is shown in Fig. 8G. This phosphorylated tyrosine site in EGFR (Tyr¹⁰¹⁶) is conserved in ErbB2 (Tyr¹⁰²³) and ErbB4 (Tyr¹⁰²²) but is not present in ErbB3 (fig. S9). These conserved Tyr sites in ErbB2 and ErbB4 may be autophosphorylation sites or may be phosphorylated by ErbB family members when heteromeric complexes are present (91).

Gajiwala *et al.* compared their structure with a structure of the kinase domain of EGFR with a peptide covalently bonded to ATP [PDB: 2GS6 (92)]. They observed that the Tyr substrate residue was shifted by one amino acid compared to the peptide conjugate's Tyr residue, which was covalently attached to ATP through a nitrogen atom at the O η hydroxyl position. They proposed a mechanism in which the hydrogen of the Tyr OH group is removed by the

HRD Asp residue, which is followed by a shift of the peptide by one position to facilitate phosphorylation. This peptide conjugate is artificial, however, and we investigated whether noncovalent substrates bound to EGFR adopted conformations that were more similar to those of the substrate in the autophosphorylation complexes found in PDB: 4I21.

Structures of the EGFR kinase domain bound to SHC1 peptide [PDB: 5CZI (93)], an optimized peptide [PDB: 5CZH (93)], or Mig6 [PDB: 4R3P and 4R3R (94)] are available in the PDB. A superposition of the enzyme kinases of the EGFR autophosphorylation complex (PDB: 4I21) and the kinase domains of these peptide-bound structures reveals similarities and differences (Fig. 8H). The peptide-ATP conjugate conformation in PDB: 2GS6 is different from the noncovalently bound peptide substrates in EGFR (PDB: 5CZH, 5CZI, 4R3R, and 4R3P), which are similar to other tyrosine kinase structures that we described, for example, EPHA3 (PDB: 3FXX and 3FY2) (Fig. 2D and IGF1R (PDB: 1K3A) (Fig. 5C). The backbone conformation in EGFR (PDB: 4I21) is similar to the noncovalently bound substrates except that it is shifted by one amino acid, so that the Tyr (P0) residue overlaps with the P-1 position of the peptide substrates. The potential autophosphorylation complex, EGFR (PDB: 4I21), is missing the interaction of the substrate Tyr residue with the HRD+4 Arg, an interaction that is present in all of the peptide structures. Thus, it seems likely that the interaction of substrate with kinase in EGFR (PDB: 4I21) is an artifact of crystallization and not an authentic autophosphorylation complex.

Nevertheless, we examined the position of Tyr¹⁰¹⁶ in other structures of EGFR and observed that this residue is bound to the N-terminal domain with a hydrogen bond to the side chain of Glu⁷³⁶ (between the second and third β sheet strands) in almost all structures in which it is present in the coordinates. This usually occurs intramolecularly but in several structures (PDB: 1M14, 1M17, 4G5J, 4I23, 4JQ7, 4JQ8, 4JR3, 4JRV, 4LI5) appears as a contact between different monomers of EGFR within the crystal. For Tyr¹⁰¹⁶ to reach the position it occupies in PDB: 4I21, it must move a distance of approximately 25 Å compared with its position in the other structures (Fig. 8I).

Kinase homodimers with swapped activation loops

Several published structures of Ser/Thr kinases are described as potential autophosphorylation complexes in which a portion of the activation loops are symmetrically swapped between the two monomers (95-100). In all of these Ser/Thr kinase structures approximately 10 residues of the C-terminal end of the activation loop of one monomer displaces the same residues in the structure of another monomer and vice versa (Fig. 9A). These structures are members of a cluster of similar interfaces in ProtCID (32) that contains 13 unique kinases, 17 distinct crystal forms, and 58 PDB entries (table S2).

In none of these complexes, however, does the annotated autophosphorylation site 12 residues from the end of the activation loop (from the APE motif) make a hydrogen bond with the Asp O δ atoms of the HRD motif nor does it seem possible that they could do so. For example, Thr²⁸⁸ of Aurora A kinase in PDB: 4C3P (99) is 7.7 Å from Asp²⁵⁶ of the HRD motif (Fig. 9B). In addition, if the -12 site was a substrate in these structures, the polypeptide chain is running in the wrong direction across the surface of the potential enzyme kinase when compared to substrate peptides or the autophosphorylation structures

described above. In these structures, the residues adjacent (± 3) to the potential phosphorylation site run right-to-left (N to C terminus) across the surface (as oriented in Fig. 9B); whereas substrates run from left-to-right, as shown in the figure by the PAK1 autophosphorylation complex substrate (PDB: 4ZY4, also shown in Fig. 9B). The PAK1 structure contains the hydrogen bond between the phosphosite residue (Thr⁴²³) and the HRD motif Asp (D389N in PDB: 4ZY4). Given the large number of such activation-loop-swap structures (table S2) and the absence of an interaction of an autophosphorylation Ser or Thr hydrogen bond with the HRD motif Asp side chain in all of them, it seems unlikely that these common dimer structures represent autophosphorylation complexes.

Nine of these entries have an interchain hydrogen bond between a threonine side chain at position -8 of the activation loop of one monomer (typically in a sequence Gly-Thr) and the aspartic acid side chain of the HRD motif of the other monomer, shown for human Aurora A kinase (PDB: 4C3P) (table S2, Fig. 9C). This highly conserved Gly-Thr sequence is often referred to as the “GT motif” (101, 102). Structures with an intermolecular GT-motif hydrogen bond include human MKNK2 [PDB: 2AC3 and 2AC5 (95); PDB: 2HW7 (96)], human STK10 [PDB: 2J7T (97) and 4EQU (100)], human OXSR1 [PDB: 2VWI (98)], human Aurora A (AURKA) [PDB: 4C3P (99)], and mouse Aurora A [PDB: 3DJ5 (103) and 3DJ6 (104)] (Fig. 9C). This interaction is typically intramolecular and present in many kinase structures (Fig. 9D). The intramolecular hydrogen bond occurs in active structures, including those with substrates bound to the active site, including the structures described above [for example, PAK1 (PDB: 4ZY4) and CLK2 (PDB: 3NR9)]. A few kinases are annotated as exhibiting phosphorylation at the -8 position (in the GT motif) (105-107), but none of them is present in the substrate-binding site of these activation-loop swapped structures in a manner consistent with autophosphorylation (Fig. 9A).

DISCUSSION

In this paper, we have used a structural bioinformatics approach to identify the structures of potential autophosphorylation complexes in crystals of protein kinases. Of the 15 unique autophosphorylation sites in such complexes, 10 had been described in the relevant papers as autophosphorylation structures and 5 were either not detected by the authors or described in the publications presenting the structures. The observed autophosphorylation complexes now comprise N-terminal juxtamembrane tyrosines of c-KIT, CSF1R, and EPHA2; the N or C-terminal tails of FGFR2, CLK2, EGFR, *C. elegans* CaMKII, and human CaMKII subunit δ ; activation loop residues of LCK, IGF1R (Tyr¹¹⁶⁵ and Tyr¹¹⁶⁶), PAK1, and IRAK4; and kinase insert loop tyrosines of FGFR1 and FGFR3. Given the rapidly increasing number of kinase structures released by the PDB (414 from October 1, 2014 through September 30, 2015), we expect on average the availability of 2-3 new autophosphorylation complexes per year.

By sequence alignment, we associated the 15 complexes with phosphorylation sites in 170 other kinases that are either annotated as phosphorylation or autophosphorylation sites. Many more potential sites in homologous positions in other kinases exist but are not currently annotated as phosphorylation sites. The autophosphorylation complexes described

here can serve as hypothetical structures for these sites to be tested both computationally and experimentally to ascertain the physiological relevance of the identified structures.

A number of observations can be made when considering the complexes collectively. The position of the substrate phosphorylation site and residues immediately adjacent to it in the substrate kinase are generally consistent with the structures of peptide substrates bound to kinases. Comparison with the peptide-kinase complexes was highly informative. In several cases, we identified similar interactions of substrate side chains with specific residues within the enzyme kinase domain: (i) the P+3 hydrophobic residues of the c-KIT, CSF1R, and EPHA2 substrate kinases, which interact in a similar manner with the enzyme kinase as do similar residues in this position in peptide substrates bound to EPHA3, INSR, and IGF1R; (ii) polar P-1 residues in the activation loops of IGF1R (both Tyr¹¹⁶⁵ and Tyr¹¹⁶⁶) and LCK, which interact with enzyme kinase in a manner similar to the interactions of the residues in peptide substrates of IGF1R and INSR; (iii) P-2 Arg residues of PAK1 autophosphorylation complexes, which had similar interactions with the enzyme kinase as did P-2 Arg residues in PAK4-bound peptides; (iv) the P-3 Arg residue of CLK2, which interacts with the enzyme kinase in a manner similar to the P-3 Arg of a peptide bound to the closely related kinase DYRK1A; and (v) P-3 Arg residues of the CaMKII structures, which interact with the enzyme kinase in a manner similar to P-3 Arg residues in a peptide bound to the related kinase rabbit PHGK.

There are some distinct differences in the positions of substrate residues in the autophosphorylation complexes compared to those of peptides bound to closely related kinases. For example, the P+3 site of phosphorylation sites in the activation loops of LCK and IGF1R (Tyr¹¹⁶⁵ and Tyr¹¹⁶⁶) are polar and do not interact with the enzyme kinase, whereas similar kinases with bound peptides have hydrophobic residues at this position and these hydrophobic residues fill a pocket formed by the G-helix and the end of the activation loop of the enzyme kinase. Chen *et al.* compared their structure of an autophosphorylation complex of the C-terminal tail of FGFR2 [PDB: 3CLY (9)] with a structure of a peptide of the same site (residues 764-778, including the phosphosite Tyr⁷⁶⁹) bound to FGFR2 [PDB: 2PVF (108)]. The homodimer binds residues from P-6 to P+3 of the substrate kinase, whereas the peptide complex makes contacts only with residues P-2 to P+1. In the homodimer structure, the Glu residue in the P-2 position relative to Tyr⁷⁶⁹ makes specific hydrogen bonds to Arg⁵⁷³ and Ser⁷⁰² of the enzyme kinase. In the peptide complex, this Glu in the P-2 position points away from the enzyme kinase. It is likely that the autophosphorylation complex of a protein substrate is a more accurate structural depiction than the peptide-substrate structure.

For all of the autophosphorylation complexes we identified, it was possible to compare the locations of the autophosphorylation sites within the substrate kinase structure compared to the positions of the same residues (or homologous residues in closely related kinases) in nonautophosphorylation structures. As might be expected, in the autophosphorylation complexes, the modified site has to bind to the peptide-binding groove of the enzyme kinase and, therefore, is typically extended away from the body of the substrate kinase in a way that it is not in the nonsubstrate structures of the same or related kinases. For the activation loop substrates, the autophosphorylation sites in nonsubstrate kinases typically form hydrogen

bonds with other parts of the activation loop in both the active and inactive kinases. In the autophosphorylation structures, they extend away from the body of the kinase domain. For Ser/Thr kinases, the conformational change seems to initiate at a common position, the conserved GT motif located 8-9 residues from the end of the activation loop. The conservation of these residues may be due to the requirements of dynamic motion of the activation loop prior to autophosphorylation. For the N- and C-terminal autophosphorylation sites, the residues and their surrounding segments are bound closely to the kinase domain in the nonsubstrate structures but extended outwards in the autophosphorylation complexes.

An important issue in phosphorylation of proteins (as opposed to peptides) is whether domain-domain contacts distant from the substrate phosphosite are specific and whether they contribute to the affinity of the enzyme kinase for the substrate kinase. Information on whether a particular interface (or portion of an interface) is relevant for binding may come from conservation of the interface in the structures of homologous protein complexes (31, 32) and from direct experiments, including mutation of interface residues or cross-linking and mass spectrometry. Only two pairs of structures in Table 1 are from homologous autophosphorylation sites that share similar interfaces, which we identified in this work. These consist of the activation loop autophosphorylation complexes of IGF1R (PDB: 3D94) and LCK (PDB: 2PL0), and the kinase insert autophosphorylation complexes of FGFR1 (PDB: 3GQI) and FGFR3 (PDB: 4K33).

Mutation experiments support all four of these complexes. Wu *et al.* mutated two residues involved in a salt bridge within the activation loop of IGF1R (PDB: 3D94), and found significant decreases in autophosphorylation (10). Here, we presented mutation data of the LCK (PDB: 2PL0) G-helix of the enzyme kinase, which is in contact with the activation loop of the substrate kinase that is autophosphorylated at Tyr³⁹⁴. Two mutations resulted in significant loss of autophosphorylation activity as detected by an antibody recognizing the phosphorylated Tyr³⁹⁴ site, and mutations at Pro⁴⁴⁷ resulted in increases in autophosphorylation, consistent with the identification of the activating mutation P447L in two different cancer cell lines, HSB2 (29) and CTV1 (109).

Bae *et al.* (76) and Huang *et al.* (9) mutated the P-6 residues of the kinase insert phosphorylation sites in FGFR1 (PDB: 3GQI) and FGFR3 (PDB: 4K33), respectively. This residue is near the phosphorylated site in sequence but makes hydrogen bonds with an aspartic acid residue in the enzyme kinase in the loop between the third β sheet strand of the N-terminal domain and the C-helix. Both mutations in the similar FGFR1 and FGFR3 interfaces reduced autophosphorylation at the kinase insert site.

Of the other sites listed in Table 1, only three of them-- PAK1 (PDB: 3Q4Z), IRAK4 (PDB: 4U97 and 4U9A), and FGFR2 (PDB: 3CLY)-- have been tested experimentally. Wang *et al.* showed that mutations D393A or E456I of PAK1 completely prevented Thr⁴²³ autophosphorylation in its activation loop (PDB: 3Q4Z) (11). Charge reversal mutations R361E, E401K, and K440E of IRAK4 impaired dimerization (PDB: 4U97, 4U9A), resulting in loss of autophosphorylation at Thr³⁴⁵ (15). Chen *et al.* mutated residues in the large interface of the autophosphorylation of the FGFR2 C-terminal segment (PDB: 3CLY) and found that N730A on the substrate kinase side slowed down trans-phosphorylation (9). They

also mutated several sites near in sequence to the phosphorylation site, but N730A was the only one distant in sequence from the phosphosite, Tyr⁷⁶⁹.

From comparison of structures of autophosphorylation complexes of homologous positions, we can also infer that domain-domain interactions are not important for the interaction of the enzyme and substrate kinases in some cases. The c-KIT (PDB: 1PKG) contains two autophosphorylation complexes of Tyr⁵⁶⁸ that exhibit two very different orientations of the enzyme and substrate kinases with one another. The newly identified complex of a homologous site in CSF1R (PDB: 3LCD) contains yet another orientation of the two kinase domains. We hypothesize that the domain-domain interaction may not be relevant for this phosphorylation event as well.

Similarly, the kinase domains of the *C. elegans* CaMKII complexes are oriented differently from one another in the two structures (PDB: 3KK8 and 3KK9) and differently from the structure of the homologous complex of human CaMKII subunit δ (PDB: 2WEL). In two of these structures (PDB: 3KK9 and 2WEL), there is little interaction of the enzyme and substrate kinases away from the phosphorylation site. This is also true of the EGFR (PDB: 4I21) C-terminal tail complex (PDB: 4I21) and the CLK2 N-terminal tail complex (PDB: 3NR9), which indicates that domain-domain interactions are not important for these reactions.

From these observations, the general trend is that autophosphorylation of internal loops, whether the activation loop or the kinase insert loop, utilizes domain-domain interactions away from the phosphorylation site. Conversely, the autophosphorylation of N- and C-terminal residues generally does not require a specific interaction of the kinase domains. The only exception is potentially the C-terminal tail of FGFR2 (PDB: 3CLY), but in this case the phosphorylation site is only two residues from the end of helix I of the substrate kinase domain. The other N or C terminal sites in Table 1 (in c-KIT, CSF1R, EPHA2, CLK2, CaMKII, and EGFR) are 10 or more residues away from ordered secondary structural elements of the substrate kinase domain.

An advantage of having a larger number of autophosphorylation complexes to study is that, as a group, these 15 complexes provide a new information on the determinants of substrate specificity of kinases, and in particular of tyrosine kinases. These 15 autophosphorylation complexes join the 20 heterodimeric structures of kinase substrate/enzyme interactions as identified in the ProtCID database (32). The peptide-kinase complexes consist of 13 Ser or Thr phosphorylation sites and 7 Tyr phosphorylation sites, and so the autophosphorylation complexes, which comprise 10 Tyr and 5 Ser or Thr sites, provide a substantial increase in the amount of structural information on tyrosine phosphorylation. Most previous analyses of kinase-substrate specificity have been performed on a small number (<5) structures of kinase domains with peptide substrates (110-114).

Each of the structures that we described provide insight into the mechanisms of autophosphorylation, because our criterion was the presence of a hydrogen bond between the side chain of the autophosphorylation site in the substrate kinase and the aspartic acid of the active site HRD motif of the enzyme kinase (or a direct contact of a mutated

autophosphorylation site in a few cases). We also looked for other evidence favoring a phosphorylation reaction, such as an active conformation of the enzyme kinase and interaction of the phosphorylation site of the substrate kinase with the HRD+2 or HRD+4 Arg or Lys residue.

Other crystal structures have been used as evidence of potential autophosphorylation complexes without the presence of a direct interaction of the substrate side chain and the active site Asp. For example, PAK1 as represented by PDB: 1YHV lacks this direct interaction (115), and the relative orientations of the enzyme and substrate kinase domains are different from the structures that do contain the hydrogen bond between the phosphosite and the Asp side chain of the active site HRD motif (for example, PDB: 3Q4Z, 4ZY4). Several structures of Ser/Thr kinases with swapped activation loops have been described as autophosphorylation complexes (105-107), but our analysis indicated that none of them contains a direct interaction of the phosphosite side chain and the Asp in the active site HRD motif. In these structures, the conserved phosphosite at position -12 of the activation loop (12 residues from the end of the loop) is 10 Å or more from the Asp side chain, and the orientation of the main chain about the Ser or Thr site is opposite that of bona fide peptide and autophosphorylation substrates. Thus, we predict that these are not autophosphorylation complexes and the relevance of these structures remains to be determined.

The structures described in this paper may provide information that leads to understanding the roles of certain driver mutations in the development of cancer. It is likely that some mutations affect the structure, dynamics, and affinity of these kinase autophosphorylation complexes. As an example, we found evidence that the activating mutation P447L of LCK in a T-cell leukemia cell line is located in the interface of the autophosphorylation complex (PDB: 2PL0) and significantly increased the rate of autophosphorylation of LCK in vitro. Oncogenic c-Kit mutations result in ligand-independent kinase activation (116). Mutations (K550I, V559D and V560G) in the c-Kit juxtamembrane region around the two main autophosphorylation sites (Tyr⁵⁶⁸ and Tyr⁵⁷⁰) are associated with human gastrointestinal stromal tumors and present in HMC-1, a human mast leukemia cell line (117).

Finally, the majority of kinase inhibitors target the ATP-binding site, which is highly conserved, thus leading to difficulties in the development of inhibitors specific for individual kinases (118). A number of other locations in kinase domains have been targeted as binding sites for allosteric inhibitors (119). The structures here may be used to identify pockets that could be targeted by an inhibitor to prevent autophosphorylation of specific sites and the downstream effects of phosphorylation at that site.

We anticipate that as more kinase structures are deposited in the PDB, application of the bioinformatics approach described here should readily reveal the structures of additional autophosphorylation complexes, thereby providing further insights into the mechanisms of kinase regulation, kinase substrate recognition, and opportunities for understanding the role of cancer-driving mutations and the development of anticancer therapeutics.

MATERIALS AND METHODS

Procedure to identify autophosphorylation complexes

PDB structures with kinase domains were collected from the PDBfam database (120), which is produced by searching the PDB with hidden Markov models (HMMs) defined by Pfam (30). There are two Pfam families, Pkinase and Pkinase_Tyr, defined for protein kinases that we used to identify the structures of kinases in the Protein Data Bank (PDB). This list was organized by UniProt (<http://uniprot.org>) (39) codes and the atom coordinates in the PDB coordinate files were renumbered to reflect UniProt numbering according to the SIFTS database (121).

The PDB provides coordinates of the asymmetric unit (ASU) of the crystal. The ASU is the smallest element of structure which when copied, rotated, and translated many times can be used to build up a model of the crystal lattice consistent with the X-ray experimental data. The symmetry operators for building up a single unit cell of the crystal are provided by the PDB for each entry in the form of 3×3 rotation matrices and translation vectors. We built a single unit cell for each entry with the symmetry operators of the appropriate space group. The unit cell is a cuboid or parallelepiped-shaped object which when copied and translated in the X, Y, and Z directions will produce a model of the full crystal lattice. From each unit cell, we produced a 3×3×3 ensemble of 27 unit cells, which is sufficient for generating all possible protein-protein interfaces that exist within the crystal. This was performed as described previously (32) to produce a set of files for each entry with the coordinates of a single homodimer of the kinase with a potential autophosphorylation interface.

The HRD motif in each structure was identified in each kinase by sequence alignment of each chain's sequence to the sequence of AURKA with the program PSI-BLAST (115) and by structure alignment to the structure of Aurora A kinase in PDB: 3E5A (122) for ambiguous cases. For the coordinates of each homodimer for each entry, we calculated the distances between all hydroxyl oxygens of one protein with the Oδ1 and Oδ2 atoms of the active site HRD Asp of the other protein and vice versa. This was performed with a custom-written Perl script. Because some potential autophosphorylation sites are mutated to Asp or Glu side chains or are already phosphorylated, we also calculated distances between the oxygen atoms of these side chains with the carboxylate atoms of the HRD Asp residue side chain.

Those interfaces with OH to Oδ distances less than 6 Å were examined and compared with structures of known kinase substrates bound to kinase domains to verify that the interaction was consistent with the known structures of substrate-kinase complexes. The structures that we used for this purpose included PAK4 with a bound consensus peptide substrate [PDB: 2Q0N, (33)], IGF1R with a peptide substrate from the IRS1 protein [PDB: 1K3A (17)], and CDK2 with a bound CDC6 substrate peptide [PDB: 1GY3 (123)]. The kinase acting as enzyme – the one containing the HRD motif of the interaction – was superposed on the kinase domains of the peptide-substrate/kinase complexes. Those for which the autophosphorylation site and one residue on either side visually coincided with the same positions in the peptide substrate were considered potential autophosphorylation complexes and examined further. For each kinase with a potential autophosphorylation site, we checked

the UniProt web page for annotation and experimental data confirming that the site was experimentally validated as an autophosphorylation site.

Multiple structure alignment of the kinase domain structures for each Uniprot were aligned with the program Theseus (124). Structures were visualized with the program (125). Surface areas of the complexes and of the individual monomers were calculated with the program NACCESS (126). The interface surface area was calculated as the surface area of the complex minus the surface areas of the two monomers divided by 2.

Mutation experiments of LCK

Residues 1 – 509 of human LCK were cloned into the pCMV-myc vector (Clontech). Primers were designed to generate the indicated mutants. We used a two-stage PCR protocol for site-directed mutagenesis adapted from Stratagene (127). Following DpnI digest and heat inactivation, PCR products were transformed into DH5alpha cells. Purified plasmids from colonies were sequenced to confirm the mutations. Plasmids were transfected into HEK293 (ATCC) with Lipofectamine 2000, according to the manufacturer's protocol (Invitrogen). Twenty-four hours following transfection, cells were lysed with 50 mM tris, pH 7.4, 150 mM NaCl, 2 mM EDTA, 1% NP-40, 1% sodium doxylcholate, 0.1% SDS, protease inhibitor cocktail (Roche), and 1 mM Na3VO4 [adapted from Chiang and Sefton, 2004 (128)]. Loading buffer was added to lysates and proteins were separated by SDS-PAGE. Total LCK and phosphorylated LCK bands were visualized by Western blot with antibodies recognizing myc or Y505pLck (Cell Signaling). Bands were quantitated with ImageJ. For each LCK protein, four independent transfections were performed to obtain an average and standard deviation for the degree of autophosphorylation achieved by that protein.

To average ratios that may be less than or greater than 1.0, we took the average of their logs and then the anti-log (for example, the mean of 2 and ½ is 1.0 using this method, but 1.25 using the arithmetic mean). For each transfection, we calculated the ratio of pLCK to Myc for each mutant and the wild-type and calculated the ratio of ratios:

$$\frac{pLCK(Mut)}{Myc(Mut)} / \frac{pLCK(WT)}{Myc(WT)}$$

The average (geometric mean) and the 95% confidence interval (CI) for this ratio is calculated as

$$\mu = \exp(\mu_{log})$$

$$CI_{\pm} = \exp\left(\mu_{log} \pm 1.96 \frac{\sigma_{log}}{\sqrt{N}}\right)$$

where μ_{log} and σ_{log} are the mean and standard deviation of $\log\left(\frac{pLCK(Mut)_i}{Myc(Mut)_i} / \frac{pLCK(WT)_i}{Myc(WT)_i}\right)$.

Sequence alignment

Homologous sites of the autophosphorylation sites observed in kinase crystals were identified by a PSI-BLAST (115) search of all human kinases downloaded from UniProt

(39) with the relevant kinase domain as query. Kinases with Ser/Thr or Tyr aligned to the query (for Ser/Thr and Tyr kinases, respectively) were aligned with Clustal W (116). The structure alignment program FATCAT (117) was used to resolve some ambiguous identifications of the DFG and APE motifs in some kinases.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Jinhua Wu for useful discussions and Vivek Modi for preparing the models of the asymmetric LCK and IGF1R dimers.

Funding: This work was funded by NIH grant R01 GM084453.

REFERENCES

1. Chang F, Steelman LS, Shelton JG, Lee JT, Navolanic PM, Blalock WL, Franklin R, McCubrey JA. Regulation of cell cycle progression and apoptosis by the Ras/Raf/MEK/ERK pathway (Review). *Int. J. Oncol.* 2003; 22:469–480. [PubMed: 12579299]
2. Sebolt-Leopold JS, Herrera R. Targeting the mitogen-activated protein kinase cascade to treat cancer. *Nat. Rev. Cancer.* 2004; 4:937–947. [PubMed: 15573115]
3. Rudolph MJ, Amodeo GA, Bai Y, Tong L. Crystal structure of the protein kinase domain of yeast AMP-activated protein kinase Snf1. *Biochem. Biophys. Res. Commun.* 2005; 337:1224–1228. [PubMed: 16236260]
4. Taylor SS, Kornev AP. Protein kinases: evolution of dynamic regulatory proteins. *Trends Biochem. Sci.* 2011; 36:65–77. [PubMed: 20971646]
5. Hubbard SR, Mohammadi M, Schlessinger J. Autoregulatory mechanisms in protein-tyrosine kinases. *J. Biol. Chem.* 1998; 273:11987–11990. [PubMed: 9575136]
6. Mohammadi M, Dikic I, Sorokin A, Burgess WH, Jaye M, Schlessinger J. Identification of six novel autophosphorylation sites on fibroblast growth factor receptor 1 and elucidation of their importance in receptor activation and signal transduction. *Mol. Cell. Biol.* 1996; 16:977–989. [PubMed: 8622701]
7. Mol CD, Lim KB, Sridhar V, Zou H, Chien EY, Sang BC, Nowakowski J, Kassel DB, Cronin CN, McRee DE. Structure of a c-kit product complex reveals the basis for kinase transactivation. *J. Biol. Chem.* 2003; 278:31461–31464. [PubMed: 12824176]
8. Bae JH, Lew ED, Yuzawa S, Tome F, Lax I, Schlessinger J. The selectivity of receptor tyrosine kinase signaling is controlled by a secondary SH2 domain binding site. *Cell.* 2009; 138:514–524. [PubMed: 19665973]
9. Chen H, Xu CF, Ma J, Eliseenkova AV, Li W, Pollock PM, Pitteloud N, Miller WT, Neubert TA, Mohammadi M. A crystallographic snapshot of tyrosine trans-phosphorylation in action. *Proc. Natl. Acad. Sci. U.S.A.* 2008; 105:19660–19665. [PubMed: 19060208]
10. Wu J, Li W, Craddock BP, Foreman KW, Mulvihill MJ, Ji QS, Miller WT, Hubbard SR. Small-molecule inhibition and activation-loop trans-phosphorylation of the IGF1 receptor. *EMBO J.* 2008; 27:1985–1994. [PubMed: 18566589]
11. Wang J, Wu JW, Wang ZX. Structural insights into the autoactivation mechanism of p21-activated protein kinase. *Structure.* 2011; 19:1752–1761. [PubMed: 22153498]
12. Rellos P, Pike AC, Niesen FH, Salah E, Lee WH, von Delft F, Knapp S. Structure of the CaMKII δ /calmodulin complex reveals the molecular mechanism of CaMKII kinase activation. *PLoS Biol.* 2010; 8:12.

13. Chao LH, Pellicena P, Deindl S, Barclay LA, Schulman H, Kuriyan J. Intersubunit capture of regulatory segments is a component of cooperative CaMKII activation. *Nat. Struct. Mol. Biol.* 2010; 17:264–272. [PubMed: 20139983]
14. Huang Z, Chen H, Blais S, Neubert TA, Li X, Mohammadi M. Structural mimicry of a-loop tyrosine phosphorylation by a pathogenic FGF receptor 3 mutation. *Structure.* 2013; 21:1889–1896. [PubMed: 23972473]
15. Ferrao R, Zhou H, Shan Y, Liu Q, Li Q, Shaw DE, Li X, Wu H. IRAK4 dimerization and trans-autophosphorylation are induced by Myddosome assembly. *Mol. Cell.* 2014; 55:891–903. [PubMed: 25201411]
16. Huber K, Brault L, Fedorov O, Gasser C, Filippakopoulos P, Bullock AN, Fabbro D, Trappe J, Schwaller J, Knapp S, Bracher F. 7,8-dichloro-1-oxo-beta-carbolines as a versatile scaffold for the development of potent and selective kinase inhibitors with unusual binding modes. *J. Med. Chem.* 2012; 55:403–413. [PubMed: 22136433]
17. Favellyukis S, Till JH, Hubbard SR, Miller WT. Structure and autoregulation of the insulin-like growth factor 1 receptor kinase. *Nat. Struct. Biol.* 2001; 8:1058–1063. [PubMed: 11694888]
18. Davis TL, Walker JR, Allali-Hassani A, Parker SA, Turk BE, Dhe-Paganon S. Structural recognition of an optimized substrate for the ephrin family of receptor tyrosine kinases. *FEBS J.* 2009; 276:4395–4404. [PubMed: 19678838]
19. Gajiwala KS, Feng J, Ferre R, Ryan K, Brodsky O, Weinrich S, Kath JC, Stewart A. Insights into the aberrant activity of mutant EGFR kinase domain and drug recognition. *Structure.* 2013; 21:209–219. [PubMed: 23273428]
20. Jacobs MD, Caron PR, Hare BJ. Classifying protein kinase structures guides use of ligand-selectivity profiles to predict inactive conformations: structure of LCK/imatinib complex. *Proteins.* 2008; 70:1451–1460. [PubMed: 17910071]
21. Nemecek C, Metz WA, Wentzler S, Ding FX, Venot C, Souaille C, Dagallier A, Maignan S, Guilloteau JP, Bernard F, Henry A, Grapinet S, Lesuisse D. Design of potent IGF1-R inhibitors related to bis-azaindoles. *Chem. Biol. Drug Des.* 2010; 76:100–106. [PubMed: 20545947]
22. Meyers MJ, Pelc M, Kamtekar S, Day J, Poda GI, Hall MK, Michener ML, Reitz BA, Mathis KJ, Pierce BS, Parikh MD, Mischke DA, Long SA, Parlow JJ, Anderson DR, Thorarensen A. Structure-based drug design enables conversion of a DFG-in binding CSF-1R kinase inhibitor to a DFG-out binding mode. *Bioorg. Med. Chem. Lett.* 2010; 20:1543–1547. [PubMed: 20137931]
23. Wei Q, Liu J, Wang N, Zhang X, Jin J, Chin-Sang I, Zheng J, Jia Z. Structures of an Eph receptor tyrosine kinase and its potential activation mechanism. *Acta Crystallogr. Sect. D-Biol. Crystallogr.* 2014; 70:3135–3143. [PubMed: 25478832]
24. Chaikuad A, Savitsky P, Krojer T, Muniz JRC, Filippakopoulos P, Rellos P, Keates T, Fedorov O, Pike ACW, Eswaran J, Berridge G, Phillips C, Zhang Y, von Delft F, Weigelt J, Arrowsmith CH, Edwards AM, Bountra C, Knapp S. Structural Genomics Consortium, Structure of human CDC2-like kinase 2 (CLK2). 2009 (10.2210/pdb3nr9/pdb).
25. Staben ST, Feng JA, Lyle K, Belvin M, Boggs J, Burch JD, Chua CC, Cui H, DiPasquale AG, Friedman LS, Heise C, Koeppen H, Kotey A, Mintzer R, Oh A, Roberts DA, Rouge L, Rudolph J, Tam C, Wang W, Xiao Y, Young A, Zhang Y, Hoeflich KP. Back pocket flexibility provides group II p21-activated kinase (PAK) selectivity for type I 1/2 kinase inhibitors. *J. Med. Chem.* 2014; 57:1033–1045. [PubMed: 24432870]
26. McCoull W, Hennessy EJ, Blades K, Box MR, Chuaqui C, Dowling JE, Davies CD, Ferguson AD, Goldberg FW, Howe NJ, Kemmitt PD, Lamont GM, Madden K, McWhirter C, Varnes JG, Ward RA, Williams JD, Yang B. Identification and optimisation of 7-azaindole PAK1 inhibitors with improved potency and kinase selectivity. *MedChemComm.* 2014; 5:1533–1539.
27. Crawford JJ, Lee W, Aliagas I, Mathieu S, Hoeflich KP, Zhou W, Wang W, Rouge L, Murray L, La H, Liu N, Fan PW, Cheong J, Heise CE, Ramaswamy S, Mintzer R, Liu Y, Chao Q, Rudolph J. Structure-Guided Design of Group I Selective p21-Activated Kinase Inhibitors. *J. Med. Chem.* 2015; 58:5121–5136. [PubMed: 26030457]
28. Karpov AS, Amiri P, Bellamacina C, Bellance MH, Breitenstein W, Daniel D, Denay R, Fabbro D, Fernandez C, Galuba I, Guerro-Lagasse S, Gutmann S, Hinh L, Jahnke W, Klopp J, Lai A, Lindvall MK, Ma S, Mobitz H, Pecchi S, Rummel G, Shoemaker K, Trappe J, Voliva C, Cowan-

- Jacob SW, Marzinzik AL. Optimization of a Dibenzodiazepine Hit to a Potent and Selective Allosteric PAK1 Inhibitor. *ACS Med. Chem. Lett.* 2015; 6:776–781. [PubMed: 26191365]
29. Wright DD, Sefton BM, Kamps MP. Oncogenic activation of the Lck protein accompanies translocation of the LCK gene in the human HSB2 T-cell leukemia. *Mol. Cell. Biol.* 1994; 14:2429–2437. [PubMed: 8139546]
30. Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, Pang N, Forslund K, Ceric G, Clements J, Heger A, Holm L, Sonnhammer EL, Eddy SR, Bateman A, Finn RD. The Pfam protein families database. *Nucleic Acids Res.* 2012; 40:D290–D301. [PubMed: 22127870]
31. Xu Q, Canutescu AA, Wang G, Shapovalov M, Obradovic Z, Dunbrack RL Jr. Statistical analysis of interface similarity in crystals of homologous proteins. *J. Mol. Biol.* 2008; 381:487–507. [PubMed: 18599072]
32. Xu Q, Dunbrack RL Jr. The protein common interface database (ProtCID)--a comprehensive database of interactions of homologous proteins in multiple crystal forms. *Nucleic Acids Res.* 2011; 39:D761–770. [PubMed: 21036862]
33. <http://www.rcsb.org/pdb/static.do?p=help/glossary.html>.
34. Adams JA. Kinetic and catalytic mechanisms of protein kinases. *Chem. Rev.* 2001; 101:2271–2290. [PubMed: 11749373]
35. Kornev AP, Haste NM, Taylor SS, Eyck LF. Surface comparison of active and inactive protein kinases identifies a conserved activation mechanism. *Proc. Natl. Acad. Sci. U.S.A.* 2006; 103:17783–17788. [PubMed: 17095602]
36. Kornev AP, Taylor SS, Ten Eyck LF. A helix scaffold for the assembly of active protein kinases. *Proc. Natl. Acad. Sci. U.S.A.* 2008; 105:14377–14382. [PubMed: 18787129]
37. Wilmot CM, Thornton JM. Beta-turns and their distortions: a proposed new nomenclature. *Protein Eng.* 1990; 3:479–493. [PubMed: 2371257]
38. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 1997; 25:3389–3402. [PubMed: 9254694]
39. UniProt Consortium. UniProt: a hub for protein information. *Nucleic Acids Res.* 2015; 43:D204–212. [PubMed: 25348405]
40. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG. Clustal W and Clustal X version 2.0. *Bioinformatics.* 2007; 23:2947–2948. [PubMed: 17846036]
41. Price DJ, Rivnay B, Fu Y, Jiang S, Avraham S, Avraham H. Direct association of Csk homologous kinase (CHK) with the diphosphorylated site Tyr568/570 of the activated c-KIT in megakaryocytes. *J. Biol. Chem.* 1997; 272:5915–5920. [PubMed: 9038210]
42. Lennartsson J, Blume-Jensen P, Hermanson M, Ponten E, Carlberg M, Ronnstrand L. Phosphorylation of Shc by Src family kinases is necessary for stem cell factor receptor/c-kit mediated activation of the Ras/MAP kinase pathway and c-fos induction. *Oncogene.* 1999; 18:5546–5553. [PubMed: 10523831]
43. Masson K, Heiss E, Band H, Ronnstrand L. Direct binding of Cbl to Tyr568 and Tyr936 of the stem cell factor receptor/c-Kit is required for ligand-induced ubiquitination, internalization and degradation. *Biochem. J.* 2006; 399:59–67. [PubMed: 16780420]
44. Wollberg P, Lennartsson J, Gottfridsson E, Yoshimura A, Ronnstrand L. The adapter protein APS associates with the multifunctional docking sites Tyr-568 and Tyr-936 in c-Kit. *Biochem. J.* 2003; 370:1033–1038. [PubMed: 12444928]
45. Wrobel CN, Debnath J, Lin E, Beausoleil S, Roussel MF, Brugge JS. Autocrine CSF-1R activation promotes Src-dependent disruption of mammary epithelial architecture. *J. Cell Biol.* 2004; 165:263–273. [PubMed: 15117969]
46. Rosenkranz S, Ikuno Y, Leong FL, Klinghoffer RA, Miyake S, Band H, Kazlauskas A. Src family kinases negatively regulate platelet-derived growth factor alpha receptor-dependent signaling and disease progression. *J. Biol. Chem.* 2000; 275:9620–9627. [PubMed: 10734113]
47. Avrov K, Kazlauskas A. The role of c-Src in platelet-derived growth factor alpha receptor internalization. *Exp. Cell Res.* 2003; 291:426–434. [PubMed: 14644164]

48. Choi MH, Lee IK, Kim GW, Kim BU, Han YH, Yu DY, Park HS, Kim KY, Lee JS, Choi C, Bae YS, Lee BI, Rhee SG, Kang SW. Regulation of PDGF signalling and vascular remodelling by peroxiredoxin II. *Nature*. 2005; 435:347–353. [PubMed: 15902258]
49. Mori S, Ronnstrand L, Yokote K, Engstrom A, Courtneidge SA, Claesson-Welsh L, Heldin CH. Identification of two juxtamembrane autophosphorylation sites in the PDGF beta-receptor; involvement in the interaction with Src family tyrosine kinases. *EMBO J*. 1993; 12:2257–2264. [PubMed: 7685273]
50. Wang J, Dai H, Yousaf N, Moussaif M, Deng Y, Boufelliga A, Swamy OR, Leone ME, Riedel H. Grb10, a positive, stimulatory signaling adapter in platelet-derived growth factor BB-, insulin-like growth factor I-, and insulin-mediated mitogenesis. *Mol. Cell. Biol*. 1999; 19:6217–6228. [PubMed: 10454568]
51. Razumovskaya E, Masson K, Khan R, Bengtsson S, Ronnstrand L. Oncogenic Flt3 receptors display different specificity and kinetics of autophosphorylation. *Exp Hematol*. 2009; 37:979–989. [PubMed: 19477218]
52. Heiss E, Masson K, Sundberg C, Pedersen M, Sun J, Bengtsson S, Ronnstrand L. Identification of Y589 and Y599 in the juxtamembrane domain of Flt3 as ligand-induced autophosphorylation sites involved in binding of Src family kinases and the protein tyrosine phosphatase SHP2. *Blood*. 2006; 108:1542–1550. [PubMed: 16684964]
53. Arora D, Stopp S, Bohmer SA, Schons J, Godfrey R, Masson K, Razumovskaya E, Ronnstrand L, Tanzer S, Bauer R, Bohmer FD, Muller JP. Protein-tyrosine phosphatase DEP-1 controls receptor tyrosine kinase FLT3 signaling. *J. Biol. Chem*. 2011; 286:10918–10929. [PubMed: 21262971]
54. Fang WB, Brantley-Sieders DM, Hwang Y, Ham AJ, Chen J. Identification and functional analysis of phosphorylated tyrosine residues within EphA2 receptor tyrosine kinase. *J. Biol. Chem*. 2008; 283:16017–16026. [PubMed: 18387945]
55. Vindis C, Cerretti DP, Daniel TO, Huynh-Do U. EphB1 recruits c-Src and p52Shc to activate MAPK/ERK and promote chemotaxis. *J. Cell Biol*. 2003; 162:661–671. [PubMed: 12925710]
56. Bohnenberger H, Oellerich T, Engelke M, Hsiao HH, Urlaub H, Wienands J. Complex phosphorylation dynamics control the composition of the Syk interactome in B cells. *Eur. J. Immunol*. 2011; 41:1550–1562. [PubMed: 21469132]
57. Di Bartolo V, Mege D, Germain V, Pelosi M, Dufour E, Michel F, Magistrelli G, Isacchi A, Acuto O. Tyrosine 319, a newly identified phosphorylation site of ZAP-70, plays a critical role in T cell antigen receptor signaling. *J. Biol. Chem*. 1999; 274:6285–6294. [PubMed: 10037717]
58. Munshi S, Kornienko M, Hall DL, Reid JC, Waxman L, Stirdivant SM, Darke PL, Kuo LC. Crystal structure of the Apo, unactivated insulin-like growth factor-1 receptor kinase. Implication for inhibitor specificity. *J. Biol. Chem*. 2002; 277:38797–38802. [PubMed: 12138114]
59. Dodson CA, Kosmopoulou M, Richards MW, Atrash B, Bavetsias V, Blagg J, Bayliss R. Crystal structure of an Aurora-A mutant that mimics Aurora-B bound to MLN8054: insights into selectivity and drug design. *Biochem. J*. 2010; 427:19–28. [PubMed: 20067443]
60. Pronk S, Pall S, Schulz R, Larsson P, Bjelkmar P, Apostolov R, Shirts MR, Smith JC, Kasson PM, van der Spoel D, Hess B, Lindahl E. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics*. 2013; 29:845–854. [PubMed: 23407358]
61. Zhu X, Kim JL, Newcomb JR, Rose PE, Stover DR, Toledo LM, Zhao H, Morgenstern KA. Structural analysis of the lymphocyte-specific kinase Lck in complex with non-selective and Src family selective kinase inhibitors. *Structure*. 1999; 7:651–661. [PubMed: 10404594]
62. Dumaual CM, Steere BA, Walls CD, Wang M, Zhang ZY, Randall SK. Integrated analysis of global mRNA and protein expression data in HEK293 cells overexpressing PRL-1. *PLOS ONE*. 2013; 8:e72977. [PubMed: 24019887]
63. Bland JM, Altman DG. Transformations, means, and confidence intervals. *BMJ*. 1996; 312:1079. [PubMed: 8616417]
64. Keedy DA, Georgiev I, Triplett EB, Donald BR, Richardson DC, Richardson JS. The role of local backrub motions in evolved and designed mutations. *PLOS Comput. Biol*. 2012; 8:e1002629. [PubMed: 22876172]

65. Kaushansky A, Gordus A, Chang B, Rush J, MacBeath G. A quantitative study of the recruitment potential of all intracellular tyrosine residues on EGFR, FGFR1 and IGF1R. *Mol. Biosyst.* 2008; 4:643–653. [PubMed: 18493663]
66. Mayer SC, Banker AL, Boschelli F, Di L, Johnson M, Kenny CH, Krishnamurthy G, Kutterer K, Moy F, Petusky S, Ravi M, Tkach D, Tsou HR, Xu W. Lead identification to generate isoquinolinedione inhibitors of insulin-like growth factor receptor (IGF-1R) for potential use in cancer treatment. *Bioorg. Med. Chem. Lett.* 2008; 18:3641–3645. [PubMed: 18501599]
67. Miller LM, Mayer SC, Berger DM, Boschelli DH, Boschelli F, Di L, Du X, Dutia M, Floyd MB, Johnson M, Kenny CH, Krishnamurthy G, Moy F, Petusky S, Tkach D, Torres N, Wu B, Xu W. Lead identification to generate 3-cyanoquinoline inhibitors of insulin-like growth factor receptor (IGF-1R) for potential use in cancer treatment. *Bioorg. Med. Chem. Lett.* 2009; 19:62–66. [PubMed: 19041240]
68. Klewegt GJ, Harris MR, Zou JY, Taylor TC, Wahlby A, Jones TA. The Uppsala Electron-Density Server. *Acta Crystallogr. Sect. D-Biol. Crystallogr.* 2004; 60:2240–2249. [PubMed: 15572777]
69. Wu J, Tseng YD, Xu CF, Neubert TA, White MF, Hubbard SR. Structural and biochemical characterization of the KRLB region in insulin receptor substrate-2. *Nat. Struct. Mol. Biol.* 2008; 15:251–258. [PubMed: 18278056]
70. Hubbard SR. Crystal structure of the activated insulin receptor tyrosine kinase in complex with peptide substrate and ATP analog. *EMBO J.* 1997; 16:5572. [PubMed: 9312016]
71. Lemmon MA, Schlessinger J. Cell signaling by receptor tyrosine kinases. *Cell.* 2010; 141:1117–1134. [PubMed: 20602996]
72. Neet K, Hunter T. Vertebrate non-receptor protein-tyrosine kinase families. *Genes Cells.* 1996; 1:147–169. [PubMed: 9140060]
73. Chen C, Ha BH, Thevenin AF, Lou HJ, Zhang R, Yip KY, Peterson JR, Gerstein M, Kim PM, Filippakopoulos P, Knapp S, Boggon TJ, Turk BE. Identification of a major determinant for serine-threonine kinase phosphoacceptor specificity. *Mol. Cell.* 2014; 53:140–147. [PubMed: 24374310]
74. Zhu G, Fujii K, Liu Y, Codrea V, Herrero J, Shaw S. A single pair of acidic residues in the kinase major groove mediates strong substrate preference for P-2 or P-5 arginine in the AGC, CAMK, and STE kinase families. *J. Biol. Chem.* 2005; 280:36372–36379. [PubMed: 16131491]
75. Wang Z, Liu J, Sudom A, Ayres M, Li S, Wesche H, Powers JP, Walker NP. Crystal structures of IRAK-4 kinase in complex with inhibitors: a serine/threonine kinase with tyrosine as a gatekeeper. *Structure.* 2006; 14:1835–1844. [PubMed: 17161373]
76. Bae JH, Boggon TJ, Tome F, Mandiyan V, Lax I, Schlessinger J. Asymmetric receptor contact is required for tyrosine autophosphorylation of fibroblast growth factor receptor in living cells. *Proc. Natl. Acad. Sci. U.S.A.* 2010; 107:2866–2871. [PubMed: 20133753]
77. Hart KC, Robertson SC, Donoghue DJ. Identification of tyrosine residues in constitutively activated fibroblast growth factor receptor 3 involved in mitogenesis, Stat activation, and phosphatidylinositol 3-kinase activation. *Mol. Biol. Cell.* 2001; 12:931–942. [PubMed: 11294897]
78. Ceridono M, Belleudi F, Ceccarelli S, Torrisi MR. Tyrosine 769 of the keratinocyte growth factor receptor is required for receptor signaling but not endocytosis. *Biochem. Biophys. Res. Commun.* 2005; 327:523–532. [PubMed: 15629145]
79. Wang J, Steinbacher S, Augustin M, Schreiner P, Epstein D, Mulvihill MJ, Crew AP. The crystal structure of a constitutively active mutant RON kinase suggests an intramolecular autophosphorylation hypothesis. *Biochemistry.* 2010; 49:7972–7974. [PubMed: 20726546]
80. Schiering N, Knapp S, Marconi M, Flocco MM, Cui J, Perego R, Rusconi L, Cristiani C. Crystal structure of the tyrosine kinase domain of the hepatocyte growth factor receptor c-Met and its complex with the microbial alkaloid K-252a. *Proc. Natl. Acad. Sci. U.S.A.* 2003; 100:12654–12659. [PubMed: 14559966]
81. Nayler O, Schnorrer F, Stamm S, Ullrich A. The cellular localization of the murine serine/arginine-rich protein kinase CLK2 is regulated by serine 141 autophosphorylation. *J. Biol. Chem.* 1998; 273:34341–34348. [PubMed: 9852100]
82. Soundararajan M, Roos AK, Savitsky P, Filippakopoulos P, Kettenbach AN, Olsen JV, Gerber SA, Eswaran J, Knapp S, Elkins JM. Structures of Down syndrome kinases, DYRKs, reveal

mechanisms of kinase activation and substrate recognition. *Structure*. 2013; 21:986–996. [PubMed: 23665168]

83. Fedorov O, Huber K, Eisenreich A, Filippakopoulos P, King O, Bullock AN, Szklarczyk D, Jensen LJ, Fabbro D, Trappe J, Rauch U, Bracher F, Knapp S. Specific CLK inhibitors from a novel chemotype for regulation of alternative splicing. *Chem. Biol.* 2011; 18:67–76. [PubMed: 21276940]
84. Debdab M, Carreaux F, Renault S, Soundararajan M, Fedorov O, Filippakopoulos P, Lozach O, Babault L, Tahtouh T, Baratte B. Leucettines, a class of potent inhibitors of cdc2-like kinases and dual specificity, tyrosine phosphorylation regulated kinases derived from the marine sponge leucettamine B: modulation of alternative pre-RNA splicing. *J. Med. Chem.* 2011; 54:4172–4186. [PubMed: 21615147]
85. Lowe ED, Noble ME, Skamnaki VT, Oikonomakos NG, Owen DJ, Johnson LN. The crystal structure of a phosphorylase kinase peptide substrate complex: kinase substrate recognition. *EMBO J.* 1997; 16:6646–6658. [PubMed: 9362479]
86. Rosenberg OS, Deindl S, Sung RJ, Nairn AC, Kuriyan J. Structure of the autoinhibited kinase domain of CaMKII and SAXS analysis of the holoenzyme. *Cell.* 2005; 123:849–860. [PubMed: 16325579]
87. Margolis BL, Lax I, Kris R, Dombalagian M, Honegger AM, Howk R, Givol D, Ullrich A, Schlessinger J. All autophosphorylation sites of epidermal growth factor (EGF) receptor and HER2/neu are located in their carboxyl-terminal tails. Identification of a novel site in EGF receptor. *J. Biol. Chem.* 1989; 264:10667–10671. [PubMed: 2543678]
88. Shao H, Cheng HY, Cook RG, Tweardy DJ. Identification and characterization of signal transducer and activator of transcription 3 recruitment sites within the epidermal growth factor receptor. *Cancer Res.* 2003; 63:3923–3930. [PubMed: 12873986]
89. Dephoure N, Zhou C, Villen J, Beausoleil SA, Bakalarski CE, Elledge SJ, Gygi SP. A quantitative atlas of mitotic phosphorylation. *Proc. Natl. Acad. Sci. U.S.A.* 2008; 105:10762–10767. [PubMed: 18669648]
90. Jura N, Endres NF, Engel K, Deindl S, Das R, Lamers MH, Wemmer DE, Zhang X, Kuriyan J. Mechanism for activation of the EGF receptor catalytic domain by the juxtamembrane segment. *Cell.* 2009; 137:1293–1307. [PubMed: 19563760]
91. Fan YX, Wong L, Johnson GR. EGFR kinase possesses a broad specificity for ErbB phosphorylation sites, and ligand increases catalytic-centre activity without affecting substrate binding affinity. *Biochem. J.* 2005; 392:417–423. [PubMed: 16122376]
92. Zhang X, Gureasko J, Shen K, Cole PA, Kuriyan J. An allosteric mechanism for activation of the kinase domain of epidermal growth factor receptor. *Cell.* 2006; 125:1137–1149. [PubMed: 16777603]
93. Begley MJ, Yun CH, Gewinner CA, Asara JM, Johnson JL, Coyle AJ, Eck MJ, Apostolou I, Cantley LC. EGF-receptor specificity for phosphotyrosine-primed substrates provides signal integration with Src. *Nat. Struct. Mol. Biol.* 2015 (doi:10.1038/nsmb.3117).
94. Park E, Kim N, Ficarro SB, Zhang Y, Lee BI, Cho A, Kim K, Park AK, Park WY, Murray B, Meyerson M, Beroukhim R, Marto JA, Cho J, Eck MJ. Structure and mechanism of activity-based inhibition of the EGF receptor by Mig6. *Nat. Struct. Mol. Biol.* 2015; 22:703–711. [PubMed: 26280531]
95. Jauch R, Jakel S, Netter C, Schreiter K, Aicher B, Jackle H, Wahl MC. Crystal structures of the Mnk2 kinase domain reveal an inhibitory conformation and a zinc binding site. *Structure*. 2005; 13:1559–1568. [PubMed: 16216586]
96. Jauch R, Cho MK, Jakel S, Netter C, Schreiter K, Aicher B, Zweckstetter M, Jackle H, Wahl MC. Mitogen-activated protein kinases interacting kinases are autoinhibited by a reprogrammed activation segment. *EMBO J.* 2006; 25:4020–4032. [PubMed: 16917500]
97. Pike AC, Rellos P, Niesen FH, Turnbull A, Oliver AW, Parker SA, Turk BE, Pearl LH, Knapp S. Activation segment dimerization: a mechanism for kinase autophosphorylation of non-consensus sites. *EMBO J.* 2008; 27:704–714. [PubMed: 18239682]
98. Villa F, Deak M, Alessi DR, van Aalten DM. Structure of the OSR1 kinase, a hypertension drug target. *Proteins*. 2008; 73:1082–1087. [PubMed: 18831043]

99. Zorba A, Buosi V, Kutter S, Kern N, Pontiggia F, Cho YJ, Kern D. Molecular mechanism of Aurora A kinase autophosphorylation and its allosteric activation by TPX2. *Elife*. 2014; 3:e02667. [PubMed: 24867643]
100. Ranjitkar P, Perera BG, Swaney DL, Hari SB, Larson ET, Krishnamurty R, Merritt EA, Villen J, Maly DJ. Affinity-based probes based on type II kinase inhibitors. *J. Am. Chem. Soc.* 2012; 134:19017–19025. [PubMed: 23088519]
101. Bayliss R, Fry A, Haq T, Yeoh S. On the molecular mechanisms of mitotic kinase activation. *Open Biol.* 2012; 2:120136. [PubMed: 23226601]
102. Taylor SS, Radzio-Andzelm E, Hunter T. How do protein kinases discriminate between serine/threonine and tyrosine? Structural insights from the insulin receptor protein-tyrosine kinase. *FASEB J.* 1995; 9:1255–1266. [PubMed: 7557015]
103. Elling RA, Erlanson DA, Yang W, Tangonan BT, Hansen SK, Romanowski MJ. Crystal structure of the mouse Aurora-A catalytic domain (Asn186->Gly, Lys240->Arg, Met302->Leu) in complex with Compound 290. 2009 (10.2210/pdb3dj5/pdb).
104. Elling RA, Erlanson DA, Yang W, Tangonan BT, Hansen SK, Romanowski MJ. Crystal structure of the mouse Aurora-A catalytic domain (Asn186->Gly, Lys240->Arg, Met302->Leu) in complex with Compound 823. 2009 (10.2210/pdb3dj6/pdb).
105. Lee CH, Chung JH. The hCds1 (Chk2)-FHA domain is essential for a chain of phosphorylation events on hCds1 that is induced by ionizing radiation. *J. Biol. Chem.* 2001; 276:30537–30541. [PubMed: 11390408]
106. Graves PR, Winkfield KM, Haystead TA. Regulation of zipper-interacting protein kinase activity in vitro and in vivo by multisite phosphorylation. *J. Biol. Chem.* 2005; 280:9363–9374. [PubMed: 15611134]
107. Hagerty L, Weitzel DH, Chambers J, Fortner CN, Brush MH, Loiselle D, Hosoya H, Haystead TA. ROCK1 phosphorylates and activates zipper-interacting protein kinase. *J. Biol. Chem.* 2007; 282:4884–4893. [PubMed: 17158456]
108. Chen H, Ma J, Li W, Eliseenkova AV, Xu C, Neubert TA, Miller WT, Mohammadi M. A molecular brake in the kinase hinge region regulates the activity of receptor tyrosine kinases. *Mol. Cell.* 2007; 27:717–730. [PubMed: 17803937]
109. Hoheisel JD, Kalender Atak Z, De Keersmaecker K, Gianfelici V, Geerdens E, Vandepoel R, Pauwels D, Porcu M, Lahortiga I, Brys V, Dirks WG, Quentmeier H, Cloos J, Cuppens H, Uyttebroeck A, Vandenberghe P, Cools J, Aerts S. High Accuracy Mutation Detection in Leukemia on a Selected Panel of Cancer Genes. *PLOS ONE*. 2012; 7:e38463. [PubMed: 22675565]
110. Ellis JJ, Kobe B. Predicting protein kinase specificity: Predikin update and performance in the DREAM4 challenge. *PLOS ONE*. 2011; 6:e21169. [PubMed: 21829434]
111. Brinkworth RI, Breinl RA, Kobe B. Structural basis and prediction of substrate specificity in protein serine/threonine kinases. *Proc. Natl. Acad. Sci. U.S.A.* 2003; 100:74–79. [PubMed: 12502784]
112. Kumar N, Mohanty D. MODPROPEP: a program for knowledge-based modeling of protein-peptide complexes. *Nucleic Acids Res.* 2007; 35:W549–555. [PubMed: 17478500]
113. Kumar N, Mohanty D. Identification of substrates for Ser/Thr kinases using residue-based statistical pair potentials. *Bioinformatics*. 2010; 26:189–197. [PubMed: 19910306]
114. Endicott JA, Noble ME, Johnson LN. The structural basis for control of eukaryotic protein kinases. *Annu. Rev. Biochem.* 2012; 81:587–613. [PubMed: 22482904]
115. Lei M, Robinson MA, Harrison SC. The active conformation of the PAK1 kinase domain. *Structure*. 2005; 13:769–778. [PubMed: 15893667]
116. Corless CL, Barnett CM, Heinrich MC. Gastrointestinal stromal tumours: origin and molecular oncology. *Nat. Rev. Cancer*. 2011; 11:865–878. [PubMed: 22089421]
117. Hirota S, Isozaki K, Moriyama Y, Hashimoto K, Nishida T, Ishiguro S, Kawano K, Hanada M, Kurata A, Takeda M, Muhammad Tunio G, Matsuzawa Y, Kanakura Y, Shinomura Y, Kitamura Y. Gain-of-function mutations of c-kit in human gastrointestinal stromal tumors. *Science*. 1998; 279:577–580. [PubMed: 9438854]

118. Liu Y, Gray NS. Rational design of inhibitors that bind to inactive kinase conformations. *Nat. Chem. Biol.* 2006; 2:358–364. [PubMed: 16783341]
119. Foda ZH, Seeliger MA. Kinase inhibitors: an allosteric add-on. *Nat. Chem. Biol.* 2014; 10:796–797. [PubMed: 25195010]
120. Xu Q, Dunbrack RL Jr. Assignment of protein sequences to existing domain and family classification systems: Pfam and the PDB. *Bioinformatics.* 2012; 28:2763–2772. [PubMed: 22942020]
121. Velankar S, Dana JM, Jacobsen J, van Ginkel G, Gane PJ, Luo J, Oldfield TJ, O'Donovan C, Martin M-J, Kleywegt GJ. SIFTS: Structure Integration with Function, Taxonomy and Sequences resource. *Nucleic Acids Res.* 2013; 41:D483–D489. [PubMed: 23203869]
122. Zhao B, Smallwood A, Yang J, Koretke K, Nurse K, Calamari A, Kirkpatrick RB, Lai Z. Modulation of kinase-inhibitor interactions by auxiliary protein binding: crystallography studies on Aurora A interactions with VX-680 and with TPX2. *Protein science : a publication of the Protein Society.* 2008; 17:1791–1797. [PubMed: 18662907]
123. Cook A, Lowe ED, Chrysina ED, Skamnaki VT, Oikonomakos NG, Johnson LN. Structural studies on phospho-CDK2/cyclin A bound to nitrate, a transition state analogue: implications for the protein kinase mechanism. *Biochemistry.* 2002; 41:7301–7311. [PubMed: 12044161]
124. Theobald DL, Wuttke DS. THESEUS: maximum likelihood superpositioning and analysis of macromolecular structures. *Bioinformatics.* 2006; 22:2171–2172. [PubMed: 16777907]
125. DeLano, WL. The PyMOL molecular graphics system. DeLano Scientific; San Carlos, CA: 2002.
126. Hubbard, SJ.; Thornton, JM. NACCESS. Department of Biochemistry and Molecular Biology. University College London; London: 1993.
127. Wang W, Malcolm B. Two-stage PCR protocol allowing introduction of multiple mutations, deletions and insertions using QuikChange Site-Directed Mutagenesis. *BioTechniques.* 1999; 26:680–682. [PubMed: 10343905]
128. Chiang GG, Sefton BM. Specific dephosphorylation of the Lck tyrosine protein kinase at Tyr-394 by the SHP-1 protein-tyrosine phosphatase. *J. Biol. Chem.* 2001; 276:23173–23178. [PubMed: 11294838]

Research Resource

Overline: Structural Biology

One-sentence summary: Structural bioinformatics reveals autophosphorylation complexes hidden in published crystals of protein kinases.

Editor's Summary

Autophosphorylation sites revealed

Three-dimensional structural data from crystals of protein kinases have aided the development of drugs and provided insights into kinase regulation and substrate recognition. Many protein kinases trans-autophosphorylate; one kinase phosphorylates another molecule of the same kinase. Anticipating that published crystallographic data may include undescribed information, Xu *et al* developed a bioinformatics method to analyze the crystals of kinases for the presence of complexes representing the conformation of kinases during autophosphorylation. The authors identified 15 autophosphorylation complexes in the Protein Data Bank, including 5 that had not been previously described. With this additional information, structural motifs involved in autophosphorylation become identifiable, which may aid in rational drug design and understanding disease-associated mutations.

Fig. 1. Autophosphorylation complexes of c-KIT, CSF1R and EPHA2 N-terminal juxtamembrane segments

(A) Superposition of two distinct c-KIT autophosphorylation complexes (PDB: 1PKG). The first homodimer is composed of chain A (enzyme, in marine) and B (substrate, in magenta) from the asymmetric unit. The substrate Tyr (in this case phosphoTyr) is in magenta spheres and the Asp of the HRD motif in the enzyme kinase is in orange spheres. ADP (Adenosine-5'-Diphosphate) is shown in sticks. The second homodimer contains chain B: 2_655 (enzyme, in light blue) and A (substrate, in pink) from neighboring unit cells. (B) Human CSF1R autophosphorylation complex (PDB: 3LCD) showing one asymmetric unit monomer and one neighbor. The bound ligand (PDB 3-letter code BDY) is shown in sticks. (C) Human EPHA2 autophosphorylation complex (PDB: 4PDO), showing one asymmetric unit monomer and one neighbor. (D) Superposition of c-KIT monomers from available PDB entries illustrates that in inactive structures, the N-terminal juxtamembrane region blocks the substrate-binding site in an autoinhibitory manner, whereas in the autophosphorylation complexes, the N-terminal juxtamembrane region extends from the surface of the kinase,

accessible for phosphorylation. Structures are colored in marine. The activation loops and juxtamembrane regions of the inactive structures and the autophosphorylation complex (PDB: 1PKG) are indicated. The phosphorylation sites (Tyr⁵⁶⁸ and Tyr⁵⁷⁰) are shown in sticks for all monomers in which they are present. **(E)** Superposition of CSF1R structures from the PDB. The autophosphorylation site, Tyr⁵⁶¹, is indicated in sticks. In the autophosphorylation complex, PDB: 3LCD, the juxtamembrane region is sticks out from the kinase domain (green). The loop coloring is the same as in panel (D). **(F)** Superposition of EPHA2 structures from the PDB. Except PDB: 4PDO, the juxtamembrane regions of the other structures are disordered. The autophosphorylation site, Tyr⁵⁹⁴, is shown in sticks. **(G)** The similarity of backbone conformations of substrate peptides (from P-4 to P+3) in EPHA2 autophosphorylation complex (PDB: 4PDO) and EPHA3 peptide complexes (PDB: 3FXX and 3FY2). The substrate regions are in stick figures with each kinase indicated by a different color. The active site Asp residues are in orange ball-and-stick and the substrate Tyr is also in protein-specific colored ball-and-stick. The helix at bottom is the G helix of the enzyme kinases. **(H)** Close-up of the active site and the substrate region of c-KIT (PDB: 1PKG), CSF1R (PDB: 3LCD), and EPHA2 (PDB: 4PDO). The EPHA2 substrate deviates from c-KIT and CSF1R at P+4, while the CSF1R and c-KIT substrates deviate from each other at P+9. **(I)** Sequence alignment of the juxtamembrane region of CSF1R, c-KIT, EPHA2, and homologues that contain a Tyr residue (in red) at the same position as the ones present in the autophosphorylation complexes shown in panels A, B, and C. The residue numbers of the red tyrosines that are homologous to the c-KIT, CSF1R, and EPHA2 structures are given after the sequence. Other tyrosine autophosphorylation sites annotated in Uniprot for each kinase are in blue bold type. The conserved Asp and Pro residues (DP) are in green bold type.

Fig. 2. The substrate specificity contacts between substrate peptides and enzymes of c-KIT, CSF1R, EPHA2 and EPHA3

(A) Hydrophobic contacts of the c-KIT (PDB: 1PKG) P+3 side chain (Ile⁵⁷¹) and Val⁸³³ of the activation loop, Ile⁸⁴¹ immediately following the activation loop, and Tyr⁸⁸⁰ of the G-helix. (B) Hydrophobic contacts of the P+3 side chain (Ile⁵⁶⁴) of CSF1R (PDB: 3LCD) and Val⁸¹⁹ and Met⁸²² of activation loop, Ile⁸²⁷ after the activation loop, and Tyr⁸⁶⁶ of G-helix. (C) Hydrophobic contacts of the EPHA2 (PDB: 4PDO) P+3 side chain (Pro⁵⁹⁷) and Ile⁷⁷⁹ and Ile⁷⁸¹ of the activation loop, Ile⁷⁸⁹ following the loop, and Met⁸²⁷ of the G-helix. (D) Hydrophobic contacts of the P+3 side chain (Ile⁹) of the peptide bound to EPHA3 (PDB: 3FY2) and Ile⁷⁸⁶ and Ile⁷⁸⁸ of the activation loop, Ile⁷⁹⁶ following the loop, and Ile⁸³⁴ of the G-helix.

Fig. 3. IGF1R and LCK autophosphorylation complexes

(A) Autophosphorylation complex of LCK (PDB: 2PL0) composed of the asymmetric unit monomer and a neighboring asymmetric unit in the crystal. The coloring scheme is the same as in Figure 1. The activation loop of one monomer lies in the active site of the other monomer and vice versa. (B) Autophosphorylation complex of IGF1R (PDB: 3D94) composed of one monomer from the asymmetric unit and one copy from a neighboring asymmetric unit. (C) Superposition of LCK monomers from the PDB (28 active, 4 inactive, 1 autophosphorylation site). The active activation loops are colored in magenta, while the

inactive activation loops are colored in yellow. The activation loop of PDB: 2PL0 is colored in green. It extends away from the kinase in a manner distinct from typical inactive or active conformations. **(D)** Superposition of IGF1R monomers from the PDB (3 active, 10 inactive). The activation loop of PDB: 3D94 is colored in green, which is not like a typical active or inactive conformation. The activation loop of PDB: 3LVP is colored in blue, which is also different from the other IGF1R monomers. **(E)** Model of an asymmetric dimer of LCK involved in autophosphorylation of Tyr³⁹⁴ of the activation loop. The model was created by superposition of an active monomer of LCK (PDB: 1QPC) (light blue, DFG-in) onto one monomer of the symmetric PDB: 2PL0 autophosphorylation complex of LCK, followed by energy minimization with Gromacs. The Phe residues of the DFG motif are shown as sticks. Phe of PDB: 2PL0 is colored in orange and Phe from PDB: 1QPC is colored in yellow. The Asp of the active-site HRD motif in PDB: 2PL0 is shown as spheres and colored in orange, and the Asp from PDB: 1QPC is shown as spheres and colored in blue, which is completely superposed with Asp in PDB: 2PL0. **(F)** Model of an asymmetric dimer of IGF1R involved in autophosphorylation of Tyr¹¹⁶⁵ of the activation loop. The model was created by superposition of an active monomer of IGF1R (PDB: 1K3A) (light blue, DFG-in) onto one monomer of the symmetric PDB: 3D94 autophosphorylation complex of IGF1R, followed by energy minimization with Gromacs. The coloring scheme is the same as in E.

Fig. 4. Extent of autophosphorylation for mutants of LCK

(A) Mutations in enzyme kinase of LCK (PDB: 2PL0). The mutation sites (Thr⁴⁴⁵, Asn⁴⁴⁶, Pro⁴⁴⁷, Gln⁴⁵¹) in the G-helix and residues Glu⁴⁴⁸ and Asp³⁶⁴ in the enzyme kinase are shown in sticks (nitrogen atoms in blue, oxygen atoms in red, carbon atoms in rainbow colors). The residues of the substrate activation loop are shown in magenta. (B) HEK293 cells were transfected with plasmids encoding wild-type and mutant Myc tagged Lck proteins. Representative Western blots with anti-Myc and anti-pLCK antibodies are shown. Equal volumes from each independent lysate were pooled for each protein and blotted with indicated antibodies to create the representative blots. (C) The phosphorylation rate of each mutant relative to wildtype (1.0, horizontal line in cyan) is plotted along with the 95% confidence interval (CI) in these values. Four experiments were conducted for each mutant and WT. Details of the calculations are provided in online methods and table S1.

Fig. 5. Autophosphorylation dimers of IGF1R Try1166 in 3LVP

(A) In the asymmetric unit of IGF1R structure PDB: 3LVP, monomer chain C (marine) is the enzyme kinase and monomer chain B (magenta) Tyr¹¹⁶⁶ is the substrate kinase. Tyr¹¹⁶⁵ of the substrate is shown in magenta sticks, and the autophosphorylation site, Tyr¹¹⁶⁶, is shown in spheres. (B) Superposition of the two dimers of PDB: 3LVP demonstrates their similarity. Asymmetric unit monomers chains D (light blue, enzyme kinase) and A (pink, substrate kinase) of PDB: 3LVP. The dimer exhibits the same interface as the C-B dimer, but the activation loop region around Tyr¹¹⁶⁶ in chain A is disordered and Tyr¹¹⁶⁶ is not present in the coordinates. (C) The substrate specificity of P-1 residues of three autophosphorylation complexes and three peptide complexes. Autophosphorylation complexes are two IGF1R structures (PDB: 3D94 and 3LVP) and one LCK structure (PDB: 2PL0). The peptide complexes consist of one IGF1R structure (PDB: 1K3A) and two INSR structures (PDB: 3BU5 and 1IR3). Substrates are indicated by different colors. The residues at P-1 are Tyr¹¹⁶⁵ of IGF1R (PDB: 3LVP), Asp¹¹⁶⁴ of IGF1R (PDB: 3D94), Glu³⁹³ of LCK (PDB: 2PL0), Glu⁷ of IGF1R-bound peptide (PDB: 1K3A), Asp⁸ of INSR-bound peptide (PDB: 3BU5), and Asp⁹ of INSR-bound peptide (PDB: 1IR3). P-1 residues have hydrogen bonds with Lys¹⁰⁸⁸ of IGF1R enzymes, and Lys¹¹²² of INSR enzymes. P-1 residue of LCK structure PDB: 2PL0 interacts with Arg³⁶⁶ (HRD+2). (D) The substrate specificity of P+3 residues. The P+3 residues of the peptide substrates are hydrophobic, Ile (PDB: 1K3A and 3LVP).

3BU5) and Met (PDB: 1IR3), while the P+3 residues of the autophosphorylation complexes are hydrophilic or neutral, Arg (PDB: 2PL0), Gly (PDB: 3LVP) and Lys (PDB: 3D94).

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Fig. 6. Autophosphorylation of serine/threonine kinases in the PDB

(A) PAK1 asymmetric autophosphorylation dimers in crystal form 1. Crystal form 1 comprises PDB: 3Q4Z; 4O0R, 4O0T; 4P90; 4ZY4, 4ZY5, 4ZY6; and 4ZLO. The interacting residues from enzyme G-helix and substrate G-helix are shown in light-blue and pink sticks respectively. The substrate activation loops are colored in rainbow from N to C terminus. (B) PAK1 asymmetric autophosphorylation dimers in crystal form 2. Crystal form 2 comprises two distinct complexes each in PDB: 4ZJI and 4ZJJ. Contacts between enzyme G-helix and substrate G-helix are shown in light-blue and pink sticks respectively. The substrate activation loops are colored in rainbow from N to C terminus. (C) The activation

loops of the PAK1 enzyme kinases of crystal form 1 and 2. Activation loops are shown in cartoon. Activation loops of crystal form 1 are DFG-in, active conformations, and colored in magenta. Activation loops of crystal form 2 are DFG-out, inactive conformations, and colored in green. Inhibitors (light sticks) bound in the crystal form 2 structures beneath the C-helix are responsible for pushing the DFG phenylalanine ring out of the Type II pocket, resulting in the DFG-out conformations. **(D)** The activation loops of PAK1 substrates kinases. The activation loops of crystal form 2 are DFG-out, while the activation loops of crystal form 1 substrates are DFG-in (PDB: 4O0R), DFG-out (PDB: 4P90 and 4ZLO) and DFG-up (PDB: 3Q4Z, 4O0T, 4ZY4, 4ZY5 and 4ZY6). Two of the newly identified crystal form 1 structures contain coordinates for the complete activation loop and are indicated (PDB: 4ZY4 and 4ZY5). **(E)** The substrate specificity of P-2 residues of PAK1 (PDB: 4ZY4) and peptide-bound PAK4 structures (PDB: 2Q0N, 4JDH, 4JDI, 4JDJ, and 4JDK). The P-2 residue in these structures is Arg and forms hydrogen bonds to conserved Glu residues in PAK1 and PAK4, as indicated. **(F)** The conformation change between activation loops of PAK1 crystal form 1 enzymes and substrates occurs at positions 426-428 (7-9 residues before the end of the activation loop). ϕ and ψ were measured in PyMOL. **(G)** IRAK4 autophosphorylation dimers (PDB: 4U97 and 4I9A) of activation loop Thr³⁴⁵. **(H)** The conformation change between activation loops of IRAK4 enzymes and substrates occurs at positions 350-351 (8-9 residues before the end of the activation loop). The activation loops are orientated similarly to (F).

Fig. 7. Autophosphorylation of FGFR1, FGFR3 and FGFR2

(A) Superposition of autophosphorylation complex of FGFR1 (PDB:3GQI) and of FGFR3 (PDB:4K33). The substrate residues, Y583F (mutant) of FGFR1 and Tyr⁵⁷⁷ of FGFR3, are positioned as a substrate in the enzyme active sites of monomers in neighboring unit cells of the crystals. (B) Specific contacts between substrate α F- α G loops and enzyme β 1- β 2 loops in the FGFR1 and FGFR3 autophosphorylation complexes. (C) Autophosphorylation complex of FGFR2 in PDB:3CLY. The C-terminal segment of the substrate sits in the active site of the enzyme chain. The loop between α G and α H is colored in pink and makes contact with the α loop (colored light blue) from the enzyme between the β 1 and β 2 strands of the N-terminal β sheet. (D) Structural superposition of FGFR2 (PDB:3CLY), FGFR1 (PDB: 3GQI), FGFR1 (PDB:3GQL), MET (PDB:1R0P) and RON (PDB:3PLS), which contain homologous autophosphorylation sites in their C-terminal tails. The conserved regions from the sequence alignment (fig. S6) are colored in light blue. Tyr⁷⁶⁹ of FGFR2 is colored in red, Tyr⁷⁶⁶ of FGFR1 is colored in magenta, and Tyr⁷⁶⁶ of FGFR1 (PDB: 3GQL) is colored in pink. Tyr¹³⁴⁹ of MET is colored in yellow, Tyr¹³⁵³ of RON is colored in green. The autophosphorylation complex Tyr (FGFR2, Tyr⁵⁶⁹, PDB: 3CLY) is in a position distinct

from that of the PLC γ bound site (FGFR1, PDB: 3GQI) and the unbound Tyr residues in RON, MET, and a second structure of FGFR1 (PDB: 3GQL).

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Fig. 8. Auto-phosphorylation of EGFR C-terminal tail

(A) CLK2 (PDB: 3NR9) symmetric autophosphorylation of an N-terminal nuclear regulatory signal sequence (Ser¹⁴²). (B) The P-3 residues of CLK2 (PDB: 3NR9) and DYRK1A peptide structure (PDB: 2WO6) form hydrogen bonds with conserved Glu residues in the enzyme kinases. (C) Superposition of CLK2 and CLK3 (PDB: 2WU6, 2WU7 and 3RAW) monomers shows that the homologous phosphorylation sites (Ser¹⁴² in CLK2, Ser²⁸³ in CLK3) are 27 Å apart in the autophosphorylation complex vs. the CLK3 monomers. The phosphorylation sites are shown in spheres, colored in pink and magenta (PDB: 3NR9). (D) Three CaMKII autophosphorylation complexes (PDB: 3KK8, 3KK9 and 2WEL) are superposed by the enzyme kinase domains (marine blue). The substrate kinases are indicated in magenta (PDB: 3KK9), green (PDB: 3KK8), and yellow (PDB: 2WEL). (E) Similar substrate specificity of P-3 and P+1 residues of the three CaMKII complexes and a structure of rabbit phosphorylase kinase with a bound peptide (PDB: 2PHK). The P-3 arginine residues interact with two Glu residues in the D-helix and one Glu residue in the HRD+4 position. The P+1 residues interact with the C-terminal segment of the activation

loop (from positions -8 to -4 from the end of the loop). **(F)** Superposition of human (PDB: 2VN9) and *C. elegans* (PDB: 2BDW) CaMKII monomers from non-autophosphorylation complexes and the substrate kinases of CaMKII (PDB: 2WEL, 3KK8, 3KK9). The autophosphorylation sites of PDB: 3KK8 and 3KK9 are colored in red and shown in spheres. The autophosphorylation site of PDB: 2WEL is colored in blue and is separated from the substrate kinase domain by a disordered segment. The phosphorylation sites in the non-substrate kinases are shown in magenta spheres, buried between the I-helix and the H-helix. **(G)** The autophosphorylation complex of the C-terminal tail of EGFR (Tyr¹⁰¹⁶) from the asymmetric unit of PDB: 4I21. **(H)** Superposition of EGFR-peptide complexes (PDB: 2GS6, 4R3P, 4R3R, 5CZH and 5CZI) (aligned by the enzyme kinases) and the EGFR autophosphorylation complex (PDB: 4I21). The substrate in PDB: 2GS6 is covalently linked to an ATP analogue. Tyrosine residues are shown in sticks in different colors for each complex. Tyr¹⁰¹⁶ of PDB: 4I21 is shifted one residue from the peptide substrates and is aligned with their P-1 residues. The distance between the HRD+4 residue (Arg⁸⁴¹) and Tyr¹⁰¹⁶ of PDB: 4I21 is 6.01 Å, while in the peptide complexes, the hydrogen bond length between the HRD+4 residue and the phosphorylation site Tyr is about 2.8 Å, consistent with a kinase substrate-enzyme interaction. **(I)** Superposition of EGFR monomers shows the intra-chain contacts between Tyr¹⁰¹⁶ (magenta) and Glu⁷³⁶ (yellow) in several non-autophosphorylation structures. In the autophosphorylation complex (PDB: 4I21), Glu⁷³⁶ (green) is exposed and Tyr¹⁰¹⁶ (red) extends away from the substrate kinase domain by about 26 Å.

Fig. 9. Activation loop swapped dimers of Serine/Threonine kinases

(A) Superposition of activation-loop swapped dimers from 17 distinct crystal forms and 58 PDB entries in ProtCID. The chains are colored in green and cyan. (B) The inter-chain hydrogen bond between Asp²⁵⁶ and Thr²⁹² in AURKA structure PDB: 4C3P. (C) The intra-chain hydrogen bond in other AURKA structures (PDB: 1OL5, 3E5A and 3HA6). (D) Comparison of the substrate segment of PAK1 (PDB: 4ZY4) bound to the enzyme kinase with the substrate segment of an activation loop swapped structure of AURKA (PDB: 4C3P). The substrate peptides are shown in ribbons, and colored from blue to red from N to C terminus. The substrate of AURKA (PDB: 4C3P) has opposite direction to the substrate of PAK1 (PDB: 4ZY4) and the potential autophosphorylation site of AURKA is 6 Å from the HRD motif Asp side chain, indicating that the activation-loop-swapped structure is not consistent with autophosphorylation.

Table 1

New and previously documented autophosphorylation complexes in the PDB

No.	Kinase	New	PDB	Location	Enzyme	Substrate	Sequence	Dist	HRD	OH	ASA	BA
1	KIT		1PKG	juxtamem	A	B	eimgNNXVXIDPT	2.82	D792	pY568	939	PISA
2	CSFIR	Y	3LCD	juxtamem	B:2_655 A:4_555	A	eimgnNXVXIDPT syegNSYTFIDPT	3.29			826	PISA
3	EPHA2	Y	4PDO	juxtamem	A	A:2_764	yvDPHTYEDPNQA	3.15	D739	Y594	819	-
		Y			B	B:2_654	yvDPHTYEDPNQA	3.21			816	-
4	IGFIR		3D94	act. loop	A:4_555	A	DIYETDYYRKGGK	2.51	D1135	Y1165	997	Auth., PISA
5	LCK	Y	2PL0	act. loop	A	A:2_655	LIEDNEYTAREGA	2.91	D364	Y394	878	PISA
6	IGFIR	Y	3LVP	act. loop	C	B	IYETDYYRKGGK	3.14	D1135	Y1166	848	-
7	PAK1		3Q4Z	act. loop	A	B	eQSKRSTMVGTPT	2.84	N389	T423	875	Auth., PISA
	PAK1	Y	4O0R	act. loop	A	B	eQSKRSEMVGTPY	2.62	N389	T423E	819	-
	PAK1	Y	4O0T	act. loop	A	B	eQSKRSEMVGTPY	2.61	N389	T423E	874	PISA
	PAK1	Y	4P90	act. loop	A	B:2_646	eqsKRSEMVGTPY	3.19	N389	T423E	813	-
	PAK1	Y	4ZLO	act. loop	B	A:1_556	eqsKRSTMVGTPT	2.54	D389	pT423	761	Auth., PISA
	PAK1	Y	4ZY4	act. loop	A	B	eQSKRSTMVGTPT	2.99	N389	T423	863	PISA
	PAK1	Y	4ZY5	act. loop	A	B	eQSKRSTMVGTPT	2.92	N389	T423	904	PISA
	PAK1	Y	4ZY6	act. loop	A	B	eqsKRSTMVGTPT	2.83	N389	T423	870	PISA
	PAK1	Y	4ZJI	act. loop	A	D	eqsKRSTMVGTPT	2.83	D389	T423	1230	PISA
	PAK1	Y	4ZJI	act. loop	B	C	eQSKRSTMVGTPT	2.93	D389	T423	1214	PISA
	PAK1	Y	4ZJI	act. loop	A	D	eQSKRSTMVGTPT	2.85	D389	T423	1169	PISA
	PAK1	Y	4ZJI	act. loop	B	C	eQSKRSTMVGTPT	3.00	D389	T423	1190	PISA
8	IRAK4		4U97	act. loop	A	B	faqtVMTSRIVGT	4.83	D311	T345	1187	Auth., PISA
	IRAK4		4U9A	act. loop	A	B	faqtVMTSRIVGT	3.78	D311	T345	1154	Auth., PISA
9	FGFR1		3GQI	kinase ins.	A	A:4_456	RPPGLEFSFNPSH	3.00	D623	Y583F	824	-
10	FGFR3		4K33	kinase ins.	A:3_645	A	RPPGLDYSFDTSE	3.35	D617	Y577	843	-
11	FGFR2		3CLY	C-term tail	A:3_655	A	LTTNEEYLDLSQP	2.82	D626	Y769	897	Auth.

No.	Kinase	New	PDB	Location	Enzyme	Substrate	Sequence	Dist	HRD	OH	ASA	BA
12	CLK2	Y	3NR9	N-term tail	B:5_555	C	sSRRAKSVEDDAE	5.36	D290	S142	1501	Auth.,PISA
		Y					sSRRAKSVEDDAE	5.40			1554	PISA
13	ceCaMKII		3KK8	C-term tail	A:5_555	A	AIHRQDTVDCL*	4.19	N134	pT284	1310	Auth.
							AIHRQDTVDCL*	4.17	N134	pT284	700	Auth.
14	CaMKII		2WEL	C-term tail	A:3_644	A	mMHRQETVDCLKK	2.71	D136	T287	668	-
15	EGFR		4I21	C-term tail	A	B	VVDADEYLIPQqg	2.47	D837	Y1016	792	-
							vvdadEYLIpqqg	2.81			558	-

All kinases in the autophosphorylation complexes listed in the Table are human kinases except ceCaMKII from *C. elegans*. Newly identified complexes are indicated in Column 3. Residues in the sequence that are disordered in the structure are in lower case. The phosphorylation site is the middle of the 13 residues of each sequence shown. Distance is in Ångstroms. The column labeled HRD provides the residue number of the catalytic site aspartic acid and OH the phosphorylation site (Y, T, or S). The enzyme and substrate columns provide the asymmetric unit chain ID given by the authors (A, B, etc.). When the enzyme or substrate kinase does not come from the original asymmetric unit (ASU) but from a copy of the ASU, the symmetry operator and unit cell identifier are given (see text). ASA is the solvent accessible surface area in Å². The last column “BA” indicates biological assemblies deposited in the PDB that contain the autophosphorylation dimer. These are defined either by the authors (“auth.”) or the program PISA. The chains listed are as defined by authors. “-” indicates that no biological assemblies deposited in the PDB contain the autophosphorylation dimer.

Table 2

Structural conformations of autophosphorylation complexes

No.	Kinase	New	PDB	Location	Enz DFG	Sub DFG	Enz K/E	Sub K/E	Enz spine	Sub spine	Enz ligand	Sub ligand	HRD+2	HRD+4
1	cKIT		1PKG	juxtamem.	in	in	3.24	2.91	Y	Y	ADP	ADP		2.87
					in	in	2.91	3.24	Y	Y	ADP	ADP		2.89
2	CSF1R	Y	3LCD	juxtamem.	in	in	2.95	2.95	Y	Y	BDY	BDY		3.12
3	EPHA2	Y	4PDO	juxtamem.	in	in	5.46	5.46	Y	Y	-	-		2.98
		Y			in	in	5.54	5.54	Y	Y	-	-		3.00
4	IGF1R*		3D94	act. loop	in*	up	4.30*	9.25	Y*	N	D94	D94		2.73
5	LCK*	Y	2PL0	act. loop	in*	out	3.20*	2.89	Y*	N	STI	STI	2.89	
6	IGF1R	Y	3LVP	act. loop	in	in	5.73	4.70	Y	Y	PDR	PDR		3.01
		Y			In	In	3.13	6.08	Y	Y	PDR	PDR		-
7	PAK1		3Q4Z	act. loop	in	up	5.85	4.86	Y	N	ANP	-	2.85	
	PAK1	Y	4O0R	act. loop	in	in	3.84	5.73	Y	N	X4Z	-	2.98	
	PAK1	Y	4O0T	act. loop	in	up	2.62	8.04	Y	N	2OL	-	3.11	
	PAK1	Y	4P90	act. loop	in	(out)	6.16	7.20	Y	N	2K0	2K0	2.75	
	PAK1	Y	4ZLO	act. loop	in	out	7.79	5.77	Y	Y	-	4PV	2.71	
	PAK1	Y	4ZY4	act. loop	in	up	3.23	8.27	Y	Y	4T3	4T3	2.87	
	PAK1	Y	4ZY5	act. loop	in	up	3.14	9.25	Y	Y	4T5	4T5	2.63	
	PAK1	Y	4ZY6	act. loop	in	up	3.42	8.37	Y	Y	4T6	4T6	2.86	
	PAK1	Y	4ZJI	act. loop	out	out	9.17	7.12	Y	Y	4OQ	4OQ	3.70	
	PAK1	Y	4ZJI	act. loop	out	out	9.49	7.13	Y	Y	4OQ	4OQ	3.74	
	PAK1	Y	4ZJJ	act. loop	out	out	10.47	9.37	Y	Y	4OR	4OR	3.66	
	PAK1	Y	4ZJJ	act. loop	out	out	9.34	11.16	Y	Y	4OR	4OR	3.89	
8	IRAK4		4U97	act. loop	in	up	2.79	2.99	Y	Y	STU	STU	3.20	
	IRAK4		4U9A	act. loop	in	up	3.22	2.92	Y	Y	STU	STU	3.26	
9	FGFR1		3GQI	kinase ins.	in	in	2.69	2.69	Y	Y	ACP	ACP	-	
10	FGFR3		4K33	kinase ins.	in	in	2.90	2.90	Y	Y	ACP	ACP	3.32	

No.	Kinase	New	PDB	Location	Enz DFG	Sub DFG	Enz K/E	Sub K/E	Enz spine	Sub spine	Enz ligand	Sub ligand	HRD+2	HRD+4
11	FGFR2		3CLY	C-term tail	in	in	2.62	2.62	Y	Y	ACP	ACP		2.54
12	CLK2	Y	3NR9	N-term tail	in	in	2.95	2.86	Y	Y	NR9	NR9	5.65	
		Y			in	in	2.83	2.83	Y	Y	NR9	NR9	6.02	
13	ceCaMKII		3KK8	C-term tail	in	in	4.50	4.50	Y	Y	-	-	2.81	
	ceCaMKII		3KK9	C-term tail	in	in	4.90	4.90	Y	Y	-	-	2.88	
14	CaMKII		2WEL	C-term tail	in	in	4.74	4.74	Y	Y	K88	K88	3.08	
15	EGFR		4I21	C-term tail	out	out	9.02	9.02	N	N	-	-	4.75	5.71
					out	out	9.02	9.02	N	N	-	-	4.86	5.15

Features indicating active or inactive kinases are given for the enzyme and substrate kinases separately, including the position of the DFG motif residues (in, out, or up), the distance between the Lys and Glu atoms typically involved in a salt bridge in active kinases (a distance < 3.5 Å indicates a salt bridge), whether the regulatory spine is intact. The PDB's three letter codes of ligands in the active site are given for the enzyme and substrate kinases. "-" indicates there is no ligand bound. The HRD+2 and HRD+4 columns indicate the distance between the phosphosite hydroxyl atom and the hydrogen bond donor site on Arg or Lys residues either 2 or 4 residues after the HRD motif. This interaction is commonly observed in peptide-kinase phosphorylation complexes, and indicates that the complex may be a phosphorylation complex when this distance is <3.5 Å. Newly identified complexes are indicated in bold type. The autophosphorylation site Tyr¹¹⁶⁶ of activation loop is disordered in the second dimer of PDB: 3LVP, and Tyr⁵⁸³ is mutated to phenylalanine in PDB: 3GQI, and so HRD+2 and HRD+4 sites are labeled "-".

Table 3

The specificity contacts of the substrate regions with the enzyme kinase

Uniprot	New	PDB	Substrate chain	Symmetry Operator	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9
KIT		1PKG	B		v	v	e	e	i	n	g	N	N	X	V	X	I	D	P	T	Q	L	P
			A		v	v	e	e	i	n	g	n	N	X	V	X	I	D	P	T	Q	L	P
CSF1R	Y	3LCD	A		i	i	e	s	y	e	G	N	S	Y	T	F	I	D	P	T	Q	L	P
EPHA2	Y	4PDO	A:2_764		h	h	h	s	q	D	P	H	T	Y	E	D	P	N	Q	A	V	L	K
	Y		B:2_654		h	h	h	s	q	D	P	H	T	Y	E	D	P	N	Q	A	V	L	K
IGF1R		3D94	A		M	T	R	D	I	Y	E	T	D	Y	Y	R	K	G	G	K	G	L	L
LCK	Y	2PL0	A:2_655		L	A	R	L	I	E	D	N	E	Y	T	A	R	E	G	A	K	F	P
IGF1R	Y	3LVP	B		T	R	D	I	Y	E	T	D	Y	Y	R	K	G	G	K	G	L	L	P
PAK1		3Q4Z	B		i	t	p	e	Q	S	K	R	S	T	M	V	G	T	P	Y	W	M	A
PAK1	Y	4O0R	B		i	t	p	e	Q	S	K	R	S	E	M	V	G	T	P	Y	W	M	A
PAK1	Y	4O0T	B		i	t	p	e	Q	S	K	R	S	E	M	V	G	T	P	Y	W	M	A
PAK1	Y	4P90	B:2_646		i	t	p	e	q	S	K	R	S	E	M	V	G	T	P	Y	W	M	A
PAK1	Y	4ZLO	A_1_556		i	t	p	e	q	S	K	R	S	T	M	V	G	T	P	Y	W	M	A
PAK1	Y	4ZY4	B		I	T	P	E	Q	S	K	R	S	T	M	V	G	T	P	Y	W	M	A
PAK1	Y	4ZY5	B		I	T	P	E	Q	S	K	R	S	T	M	V	G	T	P	Y	W	M	A
PAK1	Y	4ZY6	B		i	t	p	e	q	S	K	R	S	T	M	V	G	T	P	Y	W	M	A
PAK1	Y	4ZJ1	D		I	T	p	e	q	S	K	R	S	T	M	V	G	T	P	Y	W	M	A
PAK1	Y	4ZJ1	C		i	t	p	e	Q	S	K	R	S	T	M	V	G	T	P	Y	W	M	A
PAK1	Y	4ZJ1	D		i	t	p	E	Q	S	K	R	S	T	M	V	G	T	P	Y	W	M	A
PAK1	Y	4ZJ1	C		i	t	p	E	Q	S	K	R	S	T	M	V	G	T	P	Y	W	M	A
IRAK4		4U97	B		s	e	k	f	a	q	t	V	M	T	S	R	I	V	G	T	T	A	Y
IRAK4		4U9A	B		s	e	k	f	a	q	t	V	M	T	S	R	I	V	G	T	T	A	Y
FGFR1		3GQI	B:4_456		Q	A	R	R	P	P	G	L	E	F	S	F	N	P	S	H	N	P	E
FGFR3		4K33	A		R	A	R	R	P	P	G	L	D	Y	S	F	D	T	S	k	p	p	e
FGFR2		3CLY	A		I	L	T	L	T	T	N	E	E	Y	L	D	L	s	q	p	i	e	d

Uniprot	New	PDB	Substrate chain Symmetry Operator	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9
CLK2	Y	3NR9	C	s	s	h	s	S	R	R	A	K	S	V	E	D	D	A	E	G	H	L
EGFR		4I21	B	m	d	d	V	V	D	A	D	E	Y	L	I	P	Q	q	G	*		
			A	m	d	d	v	v	d	a	d	E	Y	L	I	P	q	q	g	*		
KCC2D_CAEEL		3KK8	A	V	A	S	A	I	H	R	Q	D	T	V	D	C	L	*				
KCC2D_CAEEL		3KK9	A	v	a	s	A	I	H	R	Q	D	T	V	D	C	*					
KCC2D		2WEL	A	v	a	s	M	M	H	R	Q	E	T	V	D	C	L	K	K	F	N	A

Residues of the substrate kinase that are in contact with the enzyme kinase are highlighted in yellow. The autophosphorylation site is residue 0. The Phe and Glu residues at position 0 are mutants. The residues in lower-case letters are disordered (i.e., present in the chain sequence but not present in the coordinates). An asterisk in the residue number after the autophosphorylation site indicates the C-terminus of the protein sequence present in the crystal.