Surface comparison of active and inactive protein kinases identifies a conserved activation mechanism

Alexandr P. Kornev*, Nina M. Haste†, Susan S. Taylor†‡§, and Lynn F. Ten Eyck*†§

*San Diego Supercomputer Center, †Department of Chemistry and Biochemistry, and ‡Howard Hughes Medical Institute, University of California at San Diego, 9500 Gilman Drive, La Jolla, CA 92093

Contributed by Susan S. Taylor, September 18, 2006

The surface comparison of different serine–threonine and tyrosine kinases reveals a set of 30 residues whose spatial positions are highly conserved. The comparison between active and inactive conformations identified the residues whose positions are the most sensitive to activation. Based on these results, we propose a model of protein kinase activation. This model explains how the presence of a phosphate group in the activation loop determines the position of the catalytically important aspartate in the Asp-Phe-Gly motif. According to the model, the most important feature of the activation is a ''spine'' formation that is dynamically assembled in all active kinases. The spine is comprised of four hydrophobic residues that we detected in a set of 23 eukaryotic and prokaryotic kinases. It spans the molecule and plays a coordinating role in activated kinases. The spine is disordered in the inactive kinases and can explain how stabilization of the whole molecule is achieved upon phosphorylation.

protein surface | graph theory

Protein kinases represent a large and diverse class of proteins that play critical roles in intracellular signal transmission. All protein kinases catalyze the same chemical reaction in eukaryotic cells: phosphorylation of a recipient protein by transferring the γ-phosphate of ATP to a hydroxyl moiety on a protein substrate. Modulation of protein kinase activity is vitally important, and their malfunction leads to many diseases. To date, numerous protein kinase structures have been solved in both active and inactive states, which has provided a general understanding of protein kinase functionality and its regulation (1–7). The conserved protein kinase core consists of two lobes: a smaller N-terminal lobe (N-lobe) and a larger C-terminal lobe (C-lobe). The two lobes form a cleft that serves as a docking site for ATP. There are two general kinds of conformational motions associated with protein kinases; one involves conversion of an inactive conformation into a catalytically competent form. Activation typically involves changes in the orientation of the αC helix in the small lobe and the activation segment in the large lobe. The active kinase then toggles between open and closed conformations as it goes through the catalytic cycle. Fig. 1*A* shows a diagram of known interactions between the conserved residues, ATP, and a recipient peptide substrate. One of the most important features is the Lys-Glu pair (K^{72}–E^{91}: here and throughout the PKA numbering is used) which, in the active state, forms a characteristic polar contact. The lysine, positioned in β3 strand of the N-lobe, binds to the α and β phosphates of the ATP. This strand is a part of rigid β sheet formed by five β strands, which are packed together with multiple hydrophobic residues. The N-lobe forms a hydrophobic pocket which accommodates the ATP adenine ring. Another characteristic element of protein kinases is a glycine-rich loop, located between the β2 and β3 strands. The loop forms a lid on top of the ATP, the tip of the loop bridges to its γ-phosphate and poises it for the phosphoryl transfer. One of the most important residues for catalysis is the aspartate (D^{184}) from the universally conserved Asp–Phe–Gly (DFG) motif at the N terminus of the activation segment. It forms polar contacts with all three ATP phosphates, directly or via coordinating atoms of magnesium. This part of the activation segment that includes the DFG motif and the following two residues is called the magnesium-binding loop. The DFG phenylalanine makes hydrophobic contacts with the αC helix and the nearby His–Arg–Asp (HRD) motif from the catalytic loop (Tyr–Arg–Asp in PKA, Fig. 1*A*). The phenylalanine is considered to be responsible for proper positioning of the DFG aspartate and accommodation of the αC helix facilitating the Lys–Glu contact. The DFG glycine is highly conserved, but its role remains unclear.

The magnesium-binding loop is followed by the β9 strand, which forms an antiparallel β sheet with the β6 strand that precedes the catalytic loop. This portion of the sheet is typically disordered in inactive kinases and is believed to be important for the correct magnesium-binding loop configuration. The activation loop, the most flexible and diverse part of the activation segment, follows the β9 strand. Activation of most protein kinases typically requires phosphorylation of a residue in this loop. The phosphorylation usually leads to a rearrangement of the loop and increase in enzymatic activity. In some kinases, the loop can also regulate the binding of ATP and/or substrate (8, 9).

The HRD triad is a motif conserved throughout the protein kinase family. The HRD aspartate (Asp^{166}) is the most conserved member of the motif. It is present in all eukaryotic, eukaryotic-like and atypical protein kinases (5). It is responsible for correct orientation of the P-site hydroxyl acceptor group in the peptide substrate (8). The HRD histidine/tyrosine is conserved through all eukaryotic and eukaryotic-like kinases. It serves as a central scaffold which binds to the carbonyl of the DFG aspartate and makes a hydrophobic contact to the DFG phenylalanine. The HRD arginine is conserved only in the eukaryotic kinases, therefore some authors identify the HRD motif as an "HxD motif" (5). Protein kinases which have arginine at this position are known as RD kinases (4). Most require phosphorylation of the activation loop, and the role of HRD arginine is to support the configuration of the activation segment (10). It provides a link between the catalytic loop, phosphorylation site and the magnesium-binding loop (Fig. 1*A*). There are kinases that do not require phosphorylation, but belong to the RD-kinases. Typically kinases lacking the HRD arginine are not phosphorylated in the activation loop.

Certain protein kinases can be activated or inhibited by specific polypeptide cofactors. For example, Src tyrosine kinase can be inactivated by a complex of Src-homology domains SH2 and SH3 that bind to the αC helix and to an inhibitory *p*-tyrosine in the C-terminal tail, stabilizing it in the "inactive conformation" (3). On the other hand, cyclin-dependent kinase Cdk2 becomes partially active after binding to cyclin, which orients the αC helix in the "active conformation" (11).

Despite the great amount of information on protein kinase activation accumulated to date, several important parts of its mechanism remain unclear. For example, it was shown that phos-

Author contributions: A.P.K. designed research; A.P.K., S.S.T., and L.F.T.E. performed research; N.M.H. performed bioinformatics and visualization; A.P.K., N.M.H., S.S.T., and L.F.T.E. analyzed data; and A.P.K. wrote the paper.

The authors declare no conflict of interest.

§To whom correspondence may be addressed. E-mail: staylor@ucsd.edu or lteneyck@sdsc.edu.

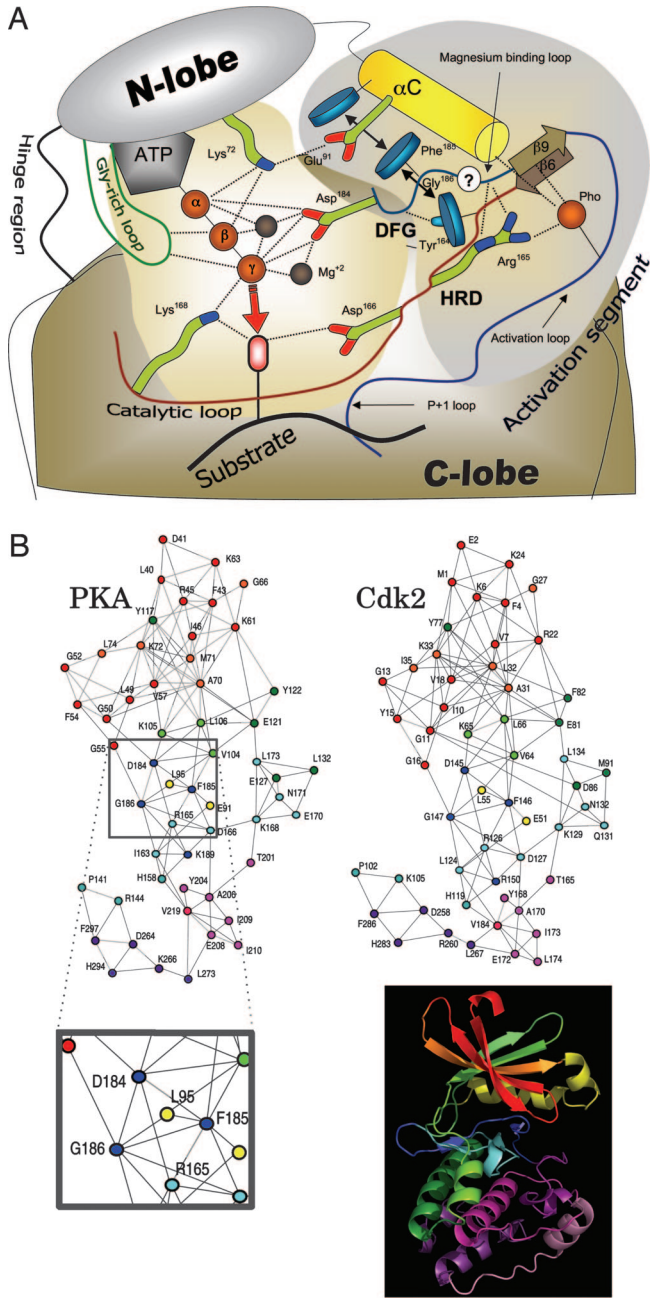

Fig. 1. Surface comparison of two active protein kinases (PKA and Cdk2) is capable to detect the most conserved, functionally important residues. (*A*) Diagram of known interactions between the protein kinase catalytic core, ATP and a substrate (PKA numbering is used). Red arrow indicates catalyzed transfer of the ATP γ-phosphate to hydroxyl group of a protein substrate. Catalytically important residues which are in contact with ATP and/or substrate are shaded yellow. Secondary structures and residues which are known to be involved in regulation of the catalytic activity are shaded gray. Hydrophobic interactions between the HRD motif, the DFG motif and the αC helix are shown by black arrows. The important polar contacts are shown by dashed lines. (*B*) Two graphs representing surface similarity between PKA and Cdk2. To indicate positions of the detected residues in protein kinase sequence, we color-coded the vertices according to the Hanks and Hunter (24) designation for subdomains. The locations of the subdomains are shown as color-coded ribbons. (*Inset*) Close-up of the highly interconnected DFG motif residues (blue vertices).

phorylation of the MAP kinase activation loop leads to substantial changes of the main chain flexibility all around the ATP binding site (12). Similar results were obtained for PKA: phosphorylation of the activation loop caused global stabilization of the active site (13). Molecular dynamics simulation of Cdk2 has demonstrated a decrease of B-factors throughout the molecule upon phosphorylation (14). To date there is no explanation for such a stabilizing role of the activation loop. Interestingly, the conserved lysine (K^{72}) in the N-lobe was found to be important for the stabilization. The mutant, with K^{72} substituted by histidine demonstrated only transient stabilization (13). There is no evident reason for conservation of the DFG glycine, which is an integral part of the most important motif in protein kinases.

To approach these problems we have studied a set of protein kinases that have been crystallized in active and inactive states. Instead of analyzing their tertiary structures. We compared surfaces of these molecules using a recently developed surface matching method based on the edge comparison and combination algorithm. We have detected a set of 30 residues whose positions on the protein surface were highly conserved for different serine–threonine and tyrosine kinases. This set includes both the residues that were known to be functionally important and the residues that were not appreciated previously. We have analyzed these residues and their possible roles in the activation process. As a result, we propose a model that correlates phosphorylation of the activation segment with catalytic activity of the enzyme. This model explains the role of the DFG glycine, the phosphorylation-dependent stabilization of a protein kinase molecule and the role of the Lys–Glu bond in the stabilization.

Results and Discussion

Supporting Information. For further details, see Figs. 7 and 8, Tables 2 and 3, and *Supporting Text*, which are published as supporting information on the PNAS web site.

Surface Comparison. Our surface matching method is based on a graph theoretical approach (the detailed algorithm is given in *Supporting Text*). A protein molecule is represented by a graph, whose vertices correspond to the amino acid residues. The graph contains information on the type of the residues and orientation of their C_α–C_β vectors. Comparison of two graphs by the edge comparison and combination algorithm is made in three steps: first, from each graph a set of subgraphs containing two vertices is derived. The subgraphs, called "edges," are created for all pairs of residues if their C_α atoms are within a predefined distance (10 Å in the current work) (the definition of predefined parameters is described in *Supporting Text*). Only water-accessible residues are considered here. This is not a limitation of the method, but rather a heuristic added solely to reduce the size of the graphs in the steps that follow. In the second step, the two sets of edges are compared. Two edges are considered similar if they contain pairs of residues that can be matched by physico-chemical properties and their C_α–C_β vectors are congruent. In the third step, the matching pairs of edges are combined into connected graphs, which represent similarity of the protein surfaces.

Earlier, this method was used to study cAMP-binding proteins and proved to be effective in detecting conserved surface motifs that cannot be found readily by other methods (15). This method is rapid and does not require sequence or structure alignments, nor any preliminary knowledge of the protein function or ligand-binding localization. A comparison of two proteins containing ≈300 residues takes ≈30 s.

Two graphs obtained after comparison of serine-threonine protein kinases PKA and Cdk2, both in their active conformation, are shown in Fig. 1*B*. A pair of PKA residues are connected if Cdk2 has a similar pair of residues with the same mutual orientation of their C_α–C_β vectors. The graphs were laid out in a way which demonstrates correspondence between PKA and Cdk2 residues. Multiple connections of a residue on the graph indicate the fact that such similarity was detected several times when the residue was combined in a pair with different residues. The residues, which are

Table 1. Description of the representative structures analyzed in surface comparison: PKA (c-AMP-dependent protein kinase), Cdk2 (cyclin-dependent kinase 2), Lck (human lymphocyte kinase), Src (human tyrosine-protein kinase), and Irk (human insulin receptor)

Name	PDB ID	Chain	Functional state	Other components	Phosphorylated residues	Ref.
PKA	2CPK	E	Active	Peptide inhibitor	S-139, T-197, S-338	25
Cdk2	1FIN	A	Active	Bound to cyclin, ATP	—	11
Cdk2	1HCL	—	Inactive	—	—	26
Lck	3LCK	—	Active	SO_4	Y-394	27
Src	2SRC	—	Inactive	AMP-PNP	Y-527	28
Irk	1IR3	A	Active	AMP-PNP, Mg	Y-1158, Y-1162, Y-1163	29
Irk	1IRK	—	Inactive	—	—	30

The PDB ID numbers of the structures, functional configuration, bound ligands and phosphorylated residues are shown. AMP-PNP, 5′-adenyl-imido-triphosphate; PDB, Protein Data Bank.

known to play important roles in the catalysis (K^{72}, K^{168}, D^{184}, etc.), have numerous connections on the graph as their relative location in space has to be conserved in different kinases. Such conservation is natural as they provide precise positioning of the ATP γ-phosphate and the accepting hydroxyl group. In addition some residues, whose importance was not appreciated earlier, also have multiple connections in the graph: e.g., A^{70} or L^{106}.

To define the residues whose positions are conserved on the surface of protein kinases and to study changes related to the activation process, we compared the active form of PKA to active and inactive Cdk2, Src, and IRK. Because Src was not crystallized in its active form, a close homolog of Src, Lck, was used (Table 1). The number of connections detected for the residues in PKA-related graphs was used as a score that reflects conservation of their positions. We accumulated these scores for the three active-active comparisons (AA score). The same was done for the three active–inactive comparisons (AI score). We suggest that, whereas the AA score corresponds to a general conservation of residues, the difference between AA and AI scores shows which residues are most sensitive to the activation state of the molecule.

The results of calculations for the top 30 residues that had at least seven edges in three active-active comparisons are presented on Fig. 2. Most of the residues that are currently known to be important for catalysis are detected by the surface comparison: glycines from the glycine-rich loop, E^{91} from the αC-helix, R^{165} and D^{166} from the conserved HRD motif, E^{127}, K^{168} and N^{171} from the catalytic loop and, finally, the whole DFG motif and K^{189}, which contacts the primary phosphorylation site in the activation loop. One very important residue, Y^{164} from the HRD motif, was not detected because it is not exposed on the surface.

The fact that we have detected the most important, highly conserved residues shows that our method is a reliable tool for recognition of functionally important elements. As such, it can serve as a rapid first step for alignment. The most surprising fact was that 18 of the 30 discovered residues were hydrophobic (including four glycines). Previously, the major focus has been on polar interactions, whereas hydrophobic residues were considered to be passive constituents of the protein hydrophobic core, with the exception of the adenine ring binding pocket. Five of the most outstanding residues with the highest AA score were V^{57}, A^{70}, K^{72}, L^{106}, and G^{186}: of these, only K^{72} is well recognized. G^{186} is a part of the highly conserved DFG motif with unclear functionality, and the rest are hydrophobic residues from the small lobe. V^{57} and A^{70} are situated directly above the adenine ring of ATP and apparently contribute to its hydrophobic binding. L^{106} is located on the N terminus of the β4-strand and is adjacent to the αC-helix. To the best of our knowledge, there have been no reports attributing any importance to this residue.

Comparison of the AA and AI scores illustrates the protein surface changes which occur upon inactivation. The most significant decrease of the score was registered for the activation segment (F^{185}, G^{186}, K^{189}), several residues from the catalytic loop (R^{165}, D^{166}) and the αC-helix (E^{91}, L^{95}); these changes are in correspondence with the well known movements in protein kinases during activation. Here, a specific feature of our surface comparison method can be noted: only substantial rearrangements of the residues are registered; if a pattern formed by several residues move as a whole it still will be detected. For example, protein kinases in an inactive state usually are crystallized in an open conformation, when the distance between two the glycine rich loop and the C-lobe changes significantly (4–5 Å). However, this does not change their positions on the surface; accordingly, the groups of residues in the small lobe (L^{49}–K^{72}), the hinge region (E^{121}–E^{127}) and the large lobe (K^{168}–L^{173}) retained their high AI scores. This allows us to distinguish opening/closing changes from the activation/deactivation of the kinase.

DFG Motif. G^{186} is a residue which has both one of the highest AA scores and a zero AI score. Only K^{189} showed similar results. Because both residues belong to the activation segment, such dependence on the activation state of the molecule is reasonable. In fact, the role of K^{189} in activation is well recognized. It belongs to the β9 strand that is a characteristic of the active state (6), but as we mentioned before, the role of the DFG glycine is not clear. To understand its importance for the process, we compared the DFG motif configuration of both active and inactive forms of Cdk2 (Fig. 3*A*). The most obvious difference was that, after inactivation, the glycine performs an extreme twist, changing its dihedral φ-angle

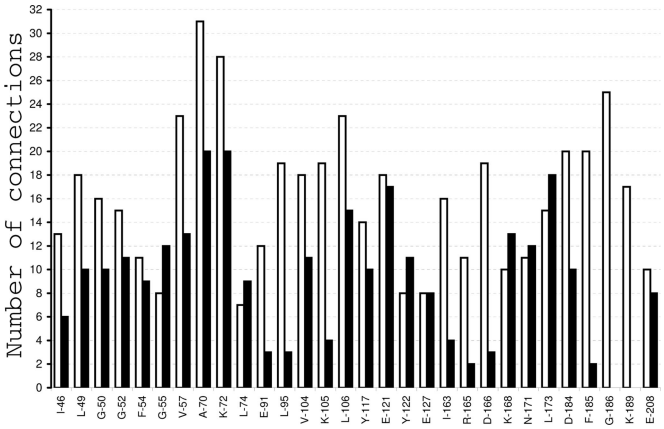

Fig. 2. Total number of connections detected for different PKA residues on similarity graphs after comparison of active PKA structure to sets of active and inactive protein kinases. The open bars show the AA score, from comparison of active PKA to other active protein kinases (see text). The filled bars show the AI score, from comparison of active PKA to a set of inactive protein kinases. The list of compared structures is presented in Table 1.

CELL BIOLOGY

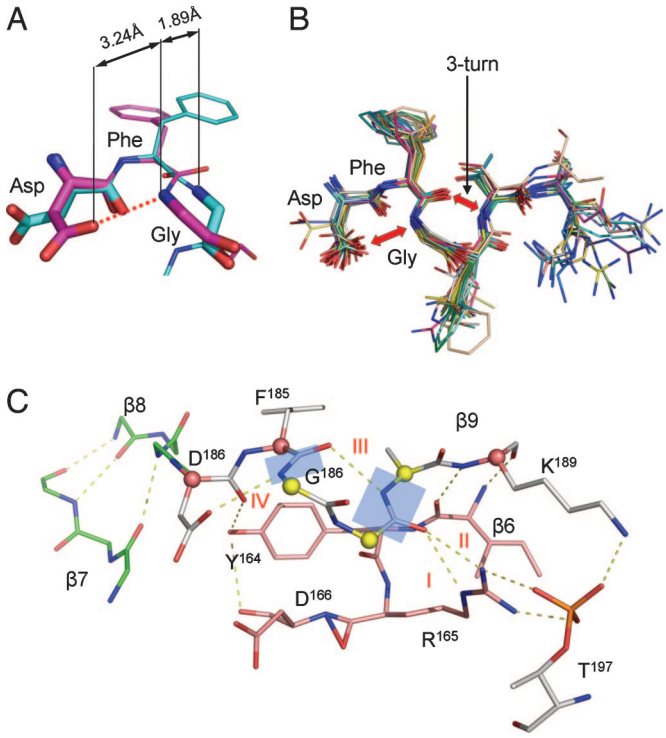

Fig. 3. Proposed model for the correlation between phosphorylation of the activation loop and the active configuration of the DFG motif. (*A*) DFG -glycine moves out its amide nitrogen upon inactivation breaking a conserved hydrogen bond to the DFG aspartate (dashed red line). Cdk2 structures were used for the illustration (magenta, active; cyan, inactive; see Table 1). (*B*) Alignment (main chain atoms only) of the magnesium-binding loops of 23 active kinases (Table 2) demonstrates conserved geometry for this part of the activation segment. Two conserved hydrogen bonds are shown as red arrows. One is between DFG aspartate and the DFG glycine; the other is a 3-turn between the DFG phenylalanine and the DFG + 2 residue. (*C*) A cascade of conserved hydrogen bonds connecting the phosphorylation site of PKA and the catalytically important DFG aspartic acid (I–IV). Carbons of the activation loop are colored white; carbons of the catalytic loop are colored tan; carbons of the β7–β8 sheet, which precedes the DFG motif, are colored green. The six C_α atoms of the magnesium-binding loop are shown as spheres, the immobilized atoms are colored tan, and the rest are colored yellow. Two peptide bonds, which are able to rotate around C_α atoms, are shown as blue planes.

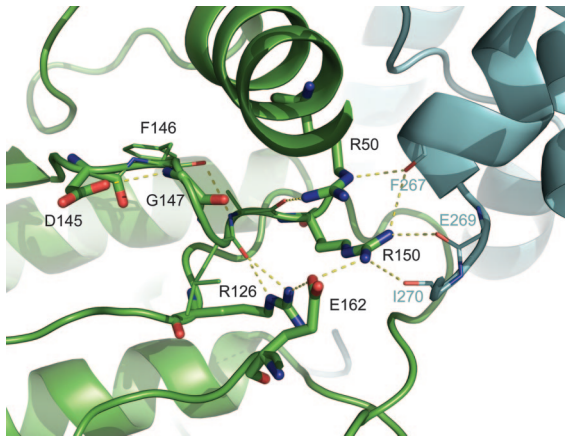

Fig. 4. Mechanism of Cdk2 partial activation induced by the cyclin binding. The cyclin ribbon and its residue labels are colored cyan; Cdk2 ribbons and residues are colored green. Cyclin is involved in configuring the magnesium-binding loop through a set of hydrogen bonds to R^{150} (K^{189} in PKA). In addition, the PSTAIRE arginine (R^{50}) is coordinated by cyclin binding to main chain of the DFG residue. The absence of negatively charged phosphate is compensated by a glutamic acid residue (E^{162}).

by 140° and disrupting a hydrogen bond to the D^{184} γ-oxygen. In addition, the aspartate (D^{184}) side chain turns away from the catalytic site, making its coordinating function impossible. The unique characteristic of glycine is the lack of a side chain, allowing it to make such a twist without steric hindrance. Thus, moving its amide nitrogen by almost 2 Å, the glycine can play the role of a bipositional switch that reorients the DFG aspartate into active or inactive positions. This finding implies that, in all active configurations, Asp (D^{184}) and Gly (G^{186}) have to form a hydrogen bond. To test this suggestion, we selected 23 kinases based on the list of known active structures (6). This selection represents a wide range of kinases covering five classes of eukaryotic protein kinases, plus two structures of prokaryotic kinases (Table 2). Fig. 3*B* shows the alignment of their magnesium-binding loops. The DFG region and the next two residues have a highly conserved configuration of the backbone and form a 3-turn between the carbonyl oxygen of the DFG phenylalanine and the amide nitrogen of the DFG + 2 residue. We must note here that the unusual orientation of two DFG aspartic acid side chains (Fig. 3*B*) are related to experimental artifacts: in the case of MAP kinase ERK2, D^{165} forms an ion pair with R^{357} from the neighboring molecule (16); and in the case of CDK5, D^{144} was mutated to asparagine (17). This analysis shows that the turn, as well as the neighboring β9 strand, is a highly conserved characteristic of the active state that was not appreciated earlier.

We know that, in active serine/threonine and tyrosine kinases, there is a conserved polar contact between the magnesium-binding loop and the HRD arginine (R^{165} in PKA), which also interacts with the primary phosphorylation site (14). Our results (Fig. 2) verify that this arginine is an important residue, and its spatial orientation strongly depends on the activation state. Thus, we have a set of features around the DFG motif that characterize the active state of the enzyme: (*i*) The hydrogen bond between δ-oxygen of Asp and amide nitrogen of Gly in the DFG motif; (*ii*) the 3-turn formed by the Phe and DFG + 2 residues; (*iii*) the β9-strand which starts at the DFG + 3 position; and (*iv*) the hydrogen bond between the HRD arginine and the DFG + 1 residue main chain.

Based on those features, we propose a model that correlates phosphorylation of the activation loop with the conformation of the DFG aspartate side chain. The general scheme of the model for the PKA case is presented in Fig. 3*C*. We know that the configuration of the activated DFG + 3 segment is highly conserved (Fig. 3*B*) and that rotation of the main chain can occur only around C_α atoms. Of the six considered residues, three have immobilized C_α atoms: D^{184}, which is anchored between the β7–β8 sheet and a conserved hydrogen bond of its carbonyl oxygen to Y^{164} from the HRD motif; F^{185} is known to be packed against both Y^{164} and the αC-helix; and K^{189}, which is a member of the rigid β6–β9 sheet (Fig. 3*C*). The C_α atoms whose positions are not anchored have to be restrained in the active configuration. Here we propose the following mechanism of activation: (*i*) The negatively charged phosphothreonine-197 in the activation loop interacts with R^{165}, which is immobilized on the catalytic loop due to its interaction with D^{220}; (*ii*) phosphothreonine also attracts K^{189} forcing it to form the β6–β9 sheet; (*iii*) the peptide bond between the DFG + 1 and DFG + 2 residues turns its carbonyl oxygen to R^{165} and forms two hydrogen bonds (I and II); (*iv*) the F^{185}–G^{186} peptide bond turns its carbonyl oxygen and forms the 3-turn (bond III); and (*v*) the G^{186} amide nitrogen rotates and forms a hydrogen bond (IV) to the D^{184} δ-oxygen, providing correct orientation of the aspartate.

We suggest that the geometrical arrangement of R^{165} and K^{189} around the phosphate moiety, described in the *i* and *ii* above, is a key event in the process of activation and precedes the following steps. We have no evidence for ordering of the subsequent steps,

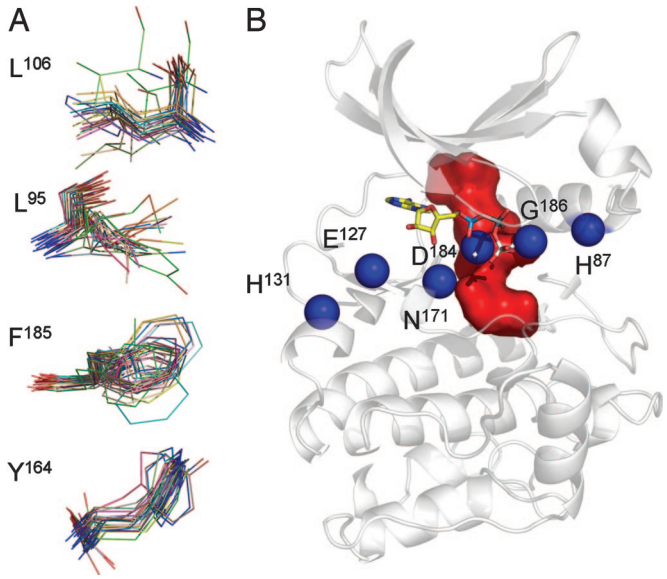

Fig. 5. Hydrophobic "spine" assembled in active kinases consists of four hydrophobic residues corresponding to four PKA residues L^{95}, L^{106}, Y^{164} and F^{185}. (*A*) Alignment of hydrophobic residues which form the spine in 23 active eukaryotic and prokaryotic kinases (Table 2). (*B*) Position of the spine in PKA. The formation is shown as a red Connolly surface with probe radius of 1.4 Å. Blue spheres represent C_α atoms of the residues that lie along the "rocking" motion axis detected in molecular dynamics simulation (22).

and expect that they may well be different in different protein kinases, or may have no fixed order.

Dephosphorylation of the T^{197} destroys the entire ensemble because the positively charged R^{165} and K^{189} repel, disrupting the $\beta6$–$\beta9$ sheet, the bonds I and II, and the 3-turn, thereby switching the DFG glycine into the inactive position, releasing DFG aspartate, which turns away from the catalytic site.

A review of inactive kinases shows that there is a wide spectrum of inactive conformations of the magnesium-binding loop. Our model explains how inactivation can be achieved by a single twist of the DFG glycine. Such a minimally disturbed inactive conformation can be observed in, for example, Cdk2 and Src kinases. Other kinases undergo substantial rearrangements of the loop with the DFG phenylalanine flipping out of the active site. An example of that is the so called "DFG out" conformation in Abl (18) and p38 kinases (19). Our model suggests that, after inactivation, the DFG motif retains the "DFG in" configuration. Considering the fact that DFG glycine is a highly conserved residue, we propose that the "glycine swing" is a universal primary event in the deactivation process, which can be followed by additional rearrangements. Disruption of the hydrogen bond cascade (Fig. 3*C*) can loosen the magnesium-binding loop providing a possibility for the "DFG flip." This suggestion is supported by the recent results showing that unphosphorylated Abl kinase can be crystallized in both "DFG in" and "DFG out" forms (20), suggesting that the inactive "DFG in" conformation is an intermediate state between the active and the "DFG out" conformation. NMR study of the p38 kinase apo form in its dephosphorylated state has also demonstrated that it dynamically changes its conformation between "DFG in" and "DFG out" configurations, proving that the magnesium-binding loop in inactive kinases can be remarkably flexible (21).

A modification of the mechanism can be seen in the case of Cdk2, which is partially activated by cyclin without the activation loop phosphorylation. It is known that cyclin binds to the highly conserved PSTAIRE motif in the αC-helix and draws it into the active conformation (11). Analysis of the cyclin bound Cdk2 structure shows that PSTAIRE-arginine (R^{50}) binds to the carbonyl oxygen of the DFG + 2 residue stabilizing both the conserved 3-turn and the $\beta6$–$\beta9$ sheet (see Fig. 4). R^{150} (corresponding to K^{189} in PKA) is also stabilized by three hydrogen bonds to the cyclin carbonyl oxygens (F^{267}, E^{269} and I^{270}) and to a glutamate from the activation loop E^{162}, which binds both to R^{150} and R^{126} (R^{165} in PKA). Obviously, the negatively charged E^{162} together with the main chain carbonyl groups of cyclin mimic presence of phosphorylated threonine. Major switch to active state is thus mediated by cyclin, whereas phosphorylation of the activation loop shifts the equilibrium to the state of maximum activity of the kinase.

Hydrophobic Interactions. Analysis of the important hydrophobic residues (those with high AA scores) that have been detected by our method (Fig. 2) shows that most of them form a hydrophobic pocket surrounding the ATP: L^{49}, F^{54}, V^{57}, A^{70}, L^{74}, V^{104}, Y^{122}, and L^{173}. Two residues, I^{46} and Y^{117}, are distant from the pocket and, apparently, represent the N-lobe rigid hydrophobic core. I^{163} also was not considered earlier to be an important residue; it precedes the HRD histidine and is a member of the conserved $\beta6$ strand. It can make a hydrophobic contact with the R^{165} side chain and, possibly, plays a role in coordination of the phosphorylation site geometry.

Another residue that had one of the highest AA scores in our comparison, but was never before appreciated, was L^{106}. Closer analysis shows that it is packed against another highly scored residue from the αC-helix, L^{95}, which is in contact with the DFG phenylalanine. We also know that the latter forms a hydrophobic contact with the HRD histidine/tyrosine, which did not show up in our analysis because it is not exposed on the surface. So, these residues (L^{95}, L^{106}, F^{185}, and Y/H^{164}) form a hydrophobic stack that goes from the rigid core of the C-lobe, up to the rigid core of the N-lobe with the middle part formed by the residues whose positions depend on the activation state. To test whether this is a universal feature for active protein kinases, we have analyzed the 23 structures that we used earlier (Fig. 3*B*). We selected the four residues whose positions correspond to Y^{164}, F^{185}, L^{95}, and L^{106} in PKA, and aligned their main chains. As shown on Fig. 5*A* all 23 kinases had the charac-

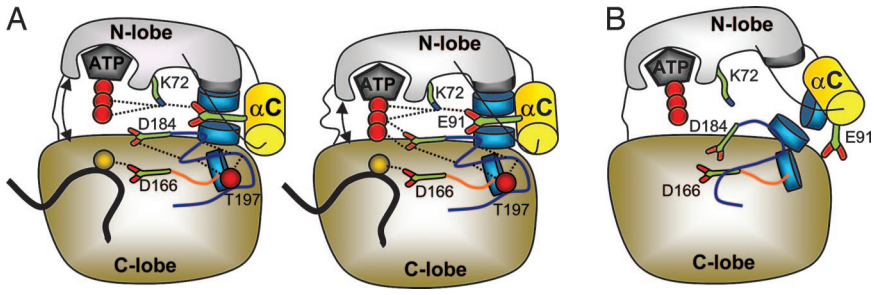

Fig. 6. General model of protein kinase activation (PKA example). (*A*) Active conformation. Phosphothreonine T^{197} arranges the magnesium-binding loop positioning the DFG aspartate for interaction with the ATP and the DFG phenylalanine for building up the hydrophobic "spine." The αC helix flips to complete the spine formation and simultaneously secures it with the K^{72}–E^{91} polar contact. The spine residues are shown as blue disks and the shaded gray portion of the N-lobe. The spine stabilizes the protein kinase molecule, which can perform coordinated motions going through open (*Left*) and closed (*Right*) conformations during the catalytic cycle. (*B*) Inactive conformation. The magnesium-binding loop and the spine are distorted, destabilizing the molecule. The lobes move independently. The unconstrained magnesium-binding loop becomes flexible and can attain different inactive configurations.

teristic formation, where the four hydrophobic residues stacked against each other. This finding leads us to suggest that such a set of hydrophobic contacts plays an important role in the protein kinase structure. Here we propose a model that explains the significance of the hydrophobic contact cascade and provides a new perspective to protein kinase activation.

We know that Y^{164} is a part of the rigid C-lobe domain; its position is strictly defined by another conserved residue, D^{220} from the αF-helix. The side chain of the tyrosine/histidine binds to the DFG phenylalanine, whose orientation can be changed during the activation process. F^{185}, in turn, binds to another activation-dependent residue, L^{95}, which binds to the rigid hydrophobic ensemble in the N-lobe via L^{106}. We suggest that this stack can unify both lobes of the molecule, providing a flexible "spine" that coordinates movements of the lobes in the active conformation, when the enzyme shuttles through catalysis performing the "rocking" or "breathing" motion observed in the molecular dynamics simulations (14, 22). Support for the suggestion is that the detected rotation axis for the breathing movement (22) is almost perpendicular to the spine (Fig. 5*B*) and crosses the latter next to the ATP γ-phosphate.

Because the N-lobe hydrophobic pocket is rather large, a single L^{95}–L^{106} contact would not be able to restrict its movement. However, the additional bonding can be provided by the K^{72}–E^{91} contact (note that L^{95} and E^{91} are bound by the rigid αC helix).

Conclusions

The surface comparison between different protein kinases has shown that hydrophobic interactions play a significant role in protein kinase functionality. Hydrophobic residues were dominant in the list of highly conserved water accessible residues (Fig. 2). In addition, we have detected a unique formation within the protein kinases made up of hydrophobic residues, which belong to the different parts of the molecule. They do not form a primary sequence motif or a particular secondary structure, so the formation is not detectable by the other methods. It is highly conserved through different types of protein kinases and is a characteristic of their active state. We propose the term "spine" for these residues because it may play an important structural role, binding together different parts of the molecule and coordinating their motions. Because it exists only in active kinases, the assembly of the spine can be an important regulatory element.

Fig. 6 shows the proposed model for protein kinase activation. As we have shown, the phosphate moiety can be a driving force for the correct folding of the activation segment. We have found a cascade of conserved hydrogen bonds, which connect the primary phosphorylation site and the catalytically important DFG aspartate (Fig. 3*C*). The DFG glycine, according to the model, plays the important role of a bipositional switch that induces the correct orientation of the aspartate. The "activated" configuration of the DFG phenylalanine facilitates the inward movement of the αC-helix, thereby locking the "spine" in place and completing the activation process. The active form is characterized by the tight connection between the two lobes through the hydrophobic spine and the K^{72}–E^{91} salt bridge, which binds the αC-helix to the middle of the rigid N-lobe domain. We suggest that such connections provide coordination of the two lobes motions, which may be important for precise positioning of the ATP with respect to the catalytic site on the C-lobe.

Dephosphorylation of the activation loop leads to mutual repulsion of the positive charges that were bound to the phosphate. Those include the HRD arginine and the positively charged residue from the β9 strand, and often include a positively charged N terminus of the αC-helix (due to the dipole moment). This leads to destabilizing of the magnesium-binding loop, movement of the αC-helix out of the active site, destruction of the hydrophobic spine, and loosening of the N-lobe. This model provides an explanation of the protein kinase stabilization induced by phosphorylation (12–14). In addition, the involvement of the lysine-glutamate bond into the stabilization phenomenon can explain the results observed in the $K^{72}H$ mutant of PKA (13).

Recent studies (20, 21) have demonstrated that, in inactive Abl and p38 kinases, the DFG motif was extremely flexible, flipping between "DFG in" and "DFG out" conformations. Our model is in a good correspondence with these facts because it suggests not only relaxation of the magnesium-binding loop, but also disruption of a connection between the lobes, that has to facilitate such conformational changes. The DFG flip is certainly associated with activation process but not likely to be a conserved feature for different kinases as the opening and closing of the active site cleft does not require the DFG flip.

Our model has a high predictive power because it defines the functional roles for a large ensemble of residues and can be tested experimentally. Verification of the model may lead to a substantial progress in understanding of regulation such an important class of enzymes as protein kinases.

Materials and Methods

The detailed description of the surface comparison algorithm is available as supporting information. All calculations were performed on a personal computer Pentium 4 (1.8 GHz) with 1 Gb RAM under Linux OS. Molecular graphics were prepared using PyMOL (DeLano Scientific, San Carlos, CA). Graphs were laid out by BIOLAYOUT program (31).

We thank Vladimir Kotlovyi for the very useful discussions and criticism during the edge comparison and combination algorithm development. This work was supported by National Science Foundation Grant NSF-DBI 99111196, National Institute of General Medical Sciences Grant GM70996 (to L.F.T.E.), National Institutes of Health Grant GM19301, and National Science Foundation Grant DBI0217951 (to S.S.T.).

1. Engh RA, Bossemeyer D (2001) *Adv Enzyme Reg* 41:121–149.
2. Engh RA, Bossemeyer D (2002) *Pharmacol Ther* 93:99–111.
3. Huse M, Kuriyan J (2002) *Cell* 109:275–282.
4. Johnson LN, Lewis RJ (2001) *Chem Rev* 101:2209–2242.
5. Kannan N, Neuwald AF (2005) *J Mol Biol* 351:956–972.
6. Nolen B, Taylor S, Ghosh G (2004) *Mol Cell* 15:661–675.
7. Taylor SS, Kim C, Vigil D, Haste NM, Yang J, Wu J, Anand GS (2005) *Biochim Biophys Acta* 1754:25–37.
8. Adams JA (2003) *Biochemistry* 42:601–607.
9. Lew J (2003) *Biochemistry* 42:849–856.
10. Krupa A, Preethi G, Srinivasan N (2004) *J Mol Biol* 339:1025–1039.
11. Jeffrey PD, Russo AA, Polyak K, Gibbs E, Hurwitz J, Massague J, Pavletich NP (1995) *Nature* 376:313–320.
12. Lee T, Hoofnagle AN, Resing KA, Ahn NG (2005) *J Mol Biol* 353:600–612.
13. Iyer GH, Garrod S, Woods VL, Jr, Taylor SS (2005) *J Mol Biol* 351:1110–1122.
14. Barrett CP, Noble ME (2005) *J Biol Chem* 280:13993–14005.
15. Berman HM, Ten Eyck LF, Goodsell DS, Haste NM, Kornev A, Taylor SS (2005) *Proc Natl Acad Sci USA* 102:45–50.
16. Canagarajah BJ, Khokhlatchev A, Cobb MH, Goldsmith EJ (1997) *Cell* 90:859–869.
17. Tarricone C, Dhavan R, Peng J, Areces LB, Tsai LH, Musacchio A (2001) *Mol Cell* 8:657–669.
18. Nagar B, Hantschel O, Young MA, Scheffzek K, Veach D, Bornmann W, Clarkson B, Superti-Furga G, Kuriyan J (2003) *Cell* 112:859–871.
19. Pargellis C, Tong L, Churchill L, Cirillo PF, Gilmore T, Graham AG, Grob PM, Hickey ER, Moss N, Pav S, Regan J (2002) *Nat Struct Biol* 9:268–272.
20. Levinson NM, Kuchment O, Shen K, Young MA, Koldobskiy M, Karplus M, Cole PA, Kuriyan J (2006) *PLoS Biol* 4:e144.
21. Vogtherr M, Saxena K, Hoelder S, Grimme S, Betz M, Schieborr U, Pescatore B, Robin M, Delarbre L, Langer T, *et al.* (2006) *Angew Chem Int Ed Eng* 45:993–997.
22. Lu B, Wong CF, McCammon JA (2005) *Protein Sci* 14:159–168.
23. Enright AJ, Ouzounis CA (2001) *Bioinformatics* 17:853–854.
24. Hanks SK, Hunter T (1995) *FASEB J* 9:576–596.
25. Knighton DR, Zheng JH, Ten Eyck LF, Ashford VA, Xuong NH, Taylor SS, Sowadski JM (1991) *Science* 253:407–414.
26. Schulze-Gahmen U, Brandsen J, Jones HD, Morgan DO, Meijer L, Vesely J, Kim SH (1995) *Proteins* 22:378–391.
27. Yamaguchi H, Hendrickson WA (1996) *Nature* 384:484–489.
28. Xu W, Doshi A, Lei M, Eck MJ, Harrison SC (1999) *Mol Cell* 3:629–638.
29. Hubbard SR, Wei L, Ellis L, Hendrickson WA (1994) *Nature* 372:746–754.
30. Hubbard SR (1997) *EMBO J* 16:5572–5581.