# Defining the Conserved Internal Architecture of a Protein Kinase

**Alexandr P. Kornev**[1,2,3] and **Susan S. Taylor**[1,3,4,*]

[1]Howard Hughes Medical Institute University of California at San Diego, 9500 Gilman Drive, La Jolla, CA 92093, USA

[2]San Diego Supercomputer Center University of California at San Diego, 9500 Gilman Drive, La Jolla, CA 92093, USA

[3]Department of Pharmacology University of California at San Diego, 9500 Gilman Drive, La Jolla, CA 92093, USA

[4]Department of Chemistry and Biochemistry University of California at San Diego, 9500 Gilman Drive, La Jolla, CA 92093, USA

## Summary

Protein kinases constitute a large protein family of important regulators in all eukaryotic cells. All of the protein kinases have a similar bilobal fold, and their key structural features have been well studied. However, the recent discovery of non-contiguous hydrophobic ensembles inside the protein kinase core shed new light on the internal organization of these molecules. Two hydrophobic "spines" traverse both lobes of the protein kinase molecule, providing a firm but flexible connection between its key elements. The spine model introduces a useful framework for analysis of intramolecular communications, molecular dynamics, and drug design.

### Keywords

Protein kinase; structure; phosphorylation; hydrophobic motifs

## Introduction

The protein kinase superfamily, one of the largest gene families encoded by the human genome, regulates multiple biological processes by posttranslational phosphorylation of serine, threonine and tyrosine residues [1]. At least 50% of all cellular proteins are likely modified by protein kinases, and malfunction of this regulatory machinery often leads to various diseases and disorders. When the cAMP-dependent protein kinase (PKA) structure was first solved in 1991 [2], it provided a template for the entire family, and as the structural kinome has unfolded over the subsequent years, that initial prediction has been repeatedly validated. The fold of the kinase core and the positioning of key residues at the active site is a conserved feature of almost every kinase. However, most protein kinases are comprised of more than just the kinase core. Either they are flanked by other domains which tend to be quite dynamic or they are part of a multi-subunit complex such as the CDKs which require an activating cyclin or phosphorylase kinase which is part of a large $\alpha_4\beta_4\gamma_4\delta_4$ complex.

*Correspondence: staylor@ucsd.edu Ph: (858) 534-8190 Fax: (858) 534-8193.

Others such as the receptor tyrosine kinases are anchored to membranes and often have long segments that tether them to the membrane as well as long C-terminal tails.

As we begin to fill in the structural kinome, we have an opportunity to elucidate some of the features that define how an active kinase is assembled and also how the inactive kinase is activated by phosphorylation [3-6]. Our analysis of spatial positions of hydrophobic residues in different kinases showed that all of them have a strictly organized interior, arranged around a single helix [7, 8]. This analysis demonstrated that the conventional interpretation of the hydrophobic core as an amorphous set of random residues is not accurate, at least in the case of protein kinases. We showed that the whole protein kinase molecule is joined together by two hydrophobic non-contiguous motifs termed "spines". The spines connect all critically important elements of the protein kinase molecule to a single hydrophobic αF-helix, providing their precise positioning in space. A characteristic feature of the spines is that they can be dynamically assembled or disassembled switching the protein kinase activity on and off. The "spine" model of protein kinase structure provides a framework for understanding of their regulation and interaction with flanking regions or other regulatory proteins that often define the specificity of different kinases.

## General organization of a protein kinase molecule

As revealed by the first structure of PKA [2], a conserved core of a typical protein kinase consists of two lobes, the N-terminal small lobe (N-lobe) and C-terminal large lobe (C-lobe) (Figure 1a). The two lobes form a deep cleft, corresponding to the active site, that accommodates a molecule of ATP bound to one or two divalent cations: magnesium or manganese. These cations compensate for the strong negative charge of the ATP phosphates[9] and provide their coordination inside the Active Site [4]. The N-lobe usually includes five β-strands and an α-helix (αC) (Figure 1b). Despite the fact that the β-strands form a relatively rigid antiparallel β-sheet, the N-lobe is very dynamic and malleable. Typically inactivation of a protein kinase involves a significant swing motion of the αC-helix directed outward from the active site (Figure 1c). Such motion disrupts multiple interactions both within the N-lobe and between the lobes, incapacitating protein kinase activity. In some kinases a short α-helix (αB) precedes αC. It is relatively weak and can easily unfold to accommodate the αC-helix motion. The N-lobe also contains a flexible loop between the β1 and β2-strands that usually includes three conserved glycines and thus was termed the glycine-rich loop (G-loop) (Figure 1b). This loop covers the β and γ-phosphates of the ATP and plays an important role in both phosphoryl transfer and ATP/ADP exchange during the catalytic cycle.

The C-lobe is much more stable and predominantly α-helical. It contains seven helices (αD through αI) and four very short β-strands (β6 through β9) (Figure 1d). The C-lobe serves as a docking site for substrate peptides/proteins. The N-terminal part of the peptide lies in a groove between the αD and αF-helices on one side and the αG-helix on the other side. At the C-terminus the residue right after the phosphorylation site is buried in a pocket formed by the so called "P+1 loop". This loop is located in the C-terminal part of an extended Activation Segment. This segment is the most important regulatory element in protein kinases. Its conformation can influence both substrate binding and catalytic efficiency [3, 5, 6]. The segment contains a magnesium binding loop at its N-terminus and the Activation Loop in the middle. The five residue long Magnesium Positioning Loop starts with the conserved Asp-Phe-Gly (DFG) motif that is critically important for catalysis. Its name reflects the fact that the DFG-aspartate directly binds to both magnesiums in the Active Site. The middle part of the Activation Segment, the Activation Loop (also known as Phosphorylation Lip in MAP kinases), is the most diverse part of the segment in terms of length and sequence [6]. Typically this loop contains a serine/threonine or tyrosine residue

that can be phosphorylated by other protein kinases or autophosphorylated via a *trans* mechanism. *Cis* mechanisms for Activation loop autophosphorylation were reported recently, but they are relatively rare and require additional scaffolding proteins [10]. This site is called a primary phosphorylation site and is located in a strategically important spot close to the Magnesium binding loop, the N-terminus of the αC-helix and a conserved His/Tyr-Arg-Asp (HRD) motif in the Catalytic Loop (Figure 1d). The negatively charged phosphate makes a set of strong hydrogen bonds and serves as an organizer both for the Active site and the substrate binding surface. Some kinases have second or even third phosphorylation sites in the Activation loop and perform functions specific for these kinases. There are also some kinases that do not have the phosphorylation site and these usually are constitutively active. They also lack the HRD-arginine, that is typically bound to the phosphate and are called "non RD-kinases" [5].

## Conservation of non-contiguous hydrophobic motifs

By 2006 the general architecture of protein kinases was well understood, and all the key polar and charged residues were discovered. However, several important questions remained unanswered. Multiple authors reported an effect of long distance signaling in different kinases (for a review see Emrick et al. [11]). Mutation of a single residue or binding a small ligand could have a global effect for the whole protein kinases molecule. It was clear that there are intramolecular "networks of connectivity" inside protein kinase molecules, possibly similar to the allosteric networks in G protein-couple receptors reported by the Ranganathan group [12]. Unfortunately, the nature of these networks in protein kinases remained unknown. It was also not clear if such internal connectivity is a universal feature for all kinases.

To study this problem a new bioinformatics method called Local Spatial Pattern (LSP) alignment [7] was developed. A distinctive feature of the method is that it aligns two protein structures without consideration of connectivity between residues. Every residue is treated as a separate object positioned in space. Thus, instead of finding conserved motifs in protein sequences, or a conserved geometry of protein main chains, the LSP alignment detects conserved spatial patterns formed by single residues. The most important and unique characteristic of the method is that not only does it detect the spatial patterns that are similar in both proteins, it also provides ranking of the residues within the pattern. The rank, referred to as an "Involvement Score" (IS), reflects the participation of each residue in formation of the pattern. The higher the score, the more neighbors in the conserved pattern surround this residue, the higher is the probability for this residue to be functionally important.

When the LSP alignment was systematically applied to a diverse set of both serine/threonine and tyrosine kinases, we were surprised by the domination of hydrophobic residues in the patterns conserved in all of the kinases [8]. Along with the very well-studied charged residues from the Activation and Catalytic loops, multiple hydrophobic residues received extremely high Involvement Scores. Some of them were known to be conserved, for example Ala(70) in AxK-motif from β-strand 3, or Phe(185) from the DFG-motif, but the reason for their conservation was not clear (here and for further comparisons the protein kinase A sequence is used for numbering). Some of the highly scored residues were never appreciated before, for example Val(57) from β-strand 2, or Met(128) from the α-D helix. These residues do not form any sequential motifs with neighboring residues and could not be detected by conventional sequence alignment methods.

As a consequence of our analysis, the pool of functionally important residues was significantly increased, but the most challenging task was to dissect this group into

biologically relevant spatial motifs that would help to understand the logic of the internal protein kinase architecture. We suggested that, in order to perform catalysis in a robust and effective way, every protein kinase has to secure the positions of two objects: the phosphate donor (ATP) and the phosphate acceptor (substrate peptide). The positioning of these elements have to be extremely precise as the phosphotransfer process is very sensitive to the distance between the reactants; displacement by a fraction of an Angstrom could impair catalysis [13]. Such precision can be achieved by a reliable connection of both ATP and the substrate to a rigid central element. After in depth analysis of the highly scored residues, we concluded that this role can be played by the long, mainly hydrophobic, helix αF. It is positioned in the middle of the Large lobe and almost all residues in this helix received exceptionally high scores in our analysis.

## Substrate binding

Substrate binding sites in serine/threonine and tyrosine kinases are slightly different as the side chain of tyrosine is about 4Å larger [14]. However, the general organization of each substrate binding site is very similar. The position of the phosphorylation (P) site is secured by the "P+1 loop" that accommodates the neighboring (P+1) residue (Figure 2). The P+1 loop, inturn, is firmly anchored to the αF helix via the Ala-Pro-Glu (APE) motif at the very C-terminus of the Activation Segment. This motif is almost universally conserved in all protein kinases. Ala(206) and Pro(207) of this motif bind to Trp(222) in the αF helix, which is also highly conserved. The N-terminal portion of the substrate peptide in serine/threonine kinases is usually bound directly to the opposite end of the αF helix. Glu(230), conserved in many kinases, binds to positively charged residues in the P-2 and P-5 positions of the substrate [15]. These interactions provide exact positioning of the substrate peptide with respect to the αF helix. Glu(230), in turn, is often bound to the P+1 loop thereby providing a link between both sides of the peptide binding site.

## ATP binding

Binding of ATP is controlled by the αF-helix through a non-contiguous hydrophobic motif that is also conserved in all protein kinases. We termed this structure a "Catalytic spine" (C-spine) as it spans the whole molecule and controls catalysis via ATP localization. It starts at the very C-terminus of the αF-helix with two large hydrophobic residues Met(231) and Leu(227) (Figure 3). The second level of the motif is formed by a large hydrophobic side chain from the αD-helix: Met(128). This residue is always capped by a short β7-strand that contains three hydrophobic residues: Leu(172), Leu(173) and Ile(174). The two border resides, Leu(172) and Ile(174), dock to the αD-helix, while the middle sidechain, Leu(173), serves as a foundation for the adenine ring of the ATP. It was shown that mutation of the middle residue to alanine inactivates PKA [16] and CDK2 [17]. On the opposite side of the adenine ring two conserved hydrophobic residues from the N-lobe Val(57) and Ala(70) complete the spine, forming an uninterrupted hydrophobic motif that spans the entire kinase molecule from its N-lobe to the rigid center of the C-lobe.

## Regulatory spine

The N-terminus of the aF-helix anchors another non-contiguous hydrophobic motif that can be found in all active protein kinases [7]. It is attached to this helix by a conserved aspartic acid Asp(220) that makes a strong hydrogen bond to the main chain of His/Tyr(164) from the HRD-motif (Figure 3). In all active kinases the DFG-phenylalanine (185) adopts the so called "in" position and binds to the HRD-histidine side chain. On the other side of the phenylalanine there is another bulky hydrophobic residue that comes from the αC-helix: Leu(95). It is always four residues apart from Glu(91), a well known partner of Lys(72) in the AxK-motif in β3-strand. These two residues form a salt bridge that is considered to be a

signature of the active kinase conformation. The importance of Leu(95), however, was never appreciated , as it is conserved in terms of hydrophobicity and its position in space, but not in terms of exact residue identity. Finally, the hydrophobic stack is completed by a residue from β4-strand: Leu(106), which also was not recognized as an important residue. This non-contiguous motif has been termed as the "Regulatory spine" (R-spine), as it traverses both lobes of the protein kinase and its middle part contains residues from the Activation segment and the αC-helix. As we described earlier, these parts of the molecule can undergo substantial rearrangements depending on the activation state of the kinase. The R-spine is present in all active protein kinases but disturbed in inactive kinases. Most often creation of the R-spine is achieved as a consequence of phosphorylation of the Activation Segment.

## Assembly and stabilization of an active protein kinase molecule

Consideration of the conserved hydrophobic spines showed that they connect the N-and the C-lobes together, make the kinase structure stable, and able to secure the positions of both ATP and substrate. The "spine" model provides important insight into the internal architecture of a kinase molecule. First, as was described earlier [18], during the catalytic cycle protein kinases undergo significant conformational changes. Such flexibility is important for effective binding of substrate/ATP, for transfer of the phosphate, and subsequent release of product/ADP. The hydrophobic nature of the spines ensure a firm but flexible connection between the two protein kinase lobes (Figure 4). This cannot be provided by charge/polar interactions as they are very sensitive to the distance and angle between the binding partners.

Second, it is rather obvious that hydrophobic contacts between a few hydrophobic residues cannot effectively solidify such a relatively large molecule as a protein kinase. This indicates that the spines are necessary but not sufficient elements for the assembly of an active kinase. They constitute a structural core whose stability has to be secured by neighboring residues that are specific for each kinase. The spine model, however, helps to identify the residues that occupy critically important and conserved spatial positions. One such residue is Met(120) which is located in the β5-strand between the C- and R-spines (Figure 3). This residue was termed "the gatekeeper", as it was used to control binding of unnatural ATP analogs [19]. The gatekeeper is not conserved in different kinases, and its side chain in many cases can be removed without any loss of kinase activity. However, substitution of the relatively small threonine in Src and Abl kinases to the bulkier methionine or isoleucine is sufficient to lock the R-spine in its active form thereby deregulating the enzymes and creating constitutively active oncogenic kinases [20]. Introduction of the R-spine concept helped to explain the oncogenic effect of the mutations and was used for rational drug design.

As we mentioned earlier, all protein kinases have similar R-spines, but they assemble and stabilize them by different, kinase specific mechanisms. Many kinases control the R-spine configuration by stabilization of the αC-helix in either an active or inactive conformation. This is achieved by filling in cavities around this helix. As was demonstrated recently [14], mechanisms for filling in these cavities can be very diverse in different serine/threonine and tyrosine kinases, but the functional consequence of filling in a spatially defined cavity to stabilize the active conformation of the αC-helix is conserved. The interacting partners, for example, can come from N-terminal and/or C-terminal tails (AGC, MAP and Death Associated kinases), from other domains (SH3 in Src kinase, SH2 in Fes kinase) or from other proteins (cyclins in Cyclin-dependent kinases). It has to be noted, however, that binding to these cavities is not sufficient for activation. It can control the αC-helical part of the R-spine, but cannot effectively stabilize the DFG-phenylalanine from the activation loop [21]. This final step is achieved most often by phosphorylation of the Activation Loop.

## Conclusions

All protein kinases have a very conserved core comprised of N- and C-lobes that evolved into a highly effective catalytic tool for phosphoryl transfer from ATP to a protein substrate. The architecture of an active kinase includes two spatially conserved but non-contiguous hydrophobic motifs that contain residues from both the N- and C-lobes. These motifs, termed "spines", integrate the protein kinase core and regulate its activity. The Regulatory spine is typically assembled after phosphorylation of the activation loop whereas the Catalytic spine is completed by the adenine ring of ATP. the two spines are anchored to the buried hydrophobic αF-helix in the C-lobe. These three hydrophobic motifs provide the basic architecture of the core. This core, however, cannot perform optimal phosphorylation on its own. Generally, a catalytically competent core requires additional structural elements. Complementation of the core by peripheral elements, such as tails or other proteins, is unique for each kinase and is frequently regulated by phosphorylation.

## REFERENCES

[1]. Manning G, Plowman GD, Hunter T, Sudarsanam S. Evolution of protein kinase signaling from yeast to man. Trends in Biochemical Sciences. 2002; 27:514–520. [PubMed: 12368087]

[2]. Knighton DR, Zheng JH, Ten Eyck LF, Ashford VA, Xuong NH, Taylor SS, Sowadski JM. Crystal structure of the catalytic subunit of cyclic adenosine monophosphate-dependent protein kinase. Science. 1991; 253:407–414. [PubMed: 1862342]

[3]. Adams JA. Activation loop phosphorylation and catalysis in protein kinases: Is there functional evidence for the autoinhibitor model? Biochemistry. 2003; 42:601–607. [PubMed: 12534271]

[4]. Johnson DA, Akamine P, Radzio-Andzelm E, Madhusudan, Taylor SS. Dynamics of cAMP-Dependent Protein Kinase. Chem. Rev. 2001; 101:2243–2270. [PubMed: 11749372]

[5]. Johnson LN, Lewis RJ. Structural basis for control by phosphorylation. Chemical Reviews. 2001; 101:2209–2242. [PubMed: 11749371]

[6]. Nolen B, Taylor S, Ghosh G. Regulation of protein kinases; controlling activity through activation segment conformation. Mol Cell. 2004; 15:661–675. [PubMed: 15350212]

[7]. Kornev AP, Haste NM, Taylor SS, Ten Eyck LF. Surface comparison of active and inactive protein kinases identifies a conserved activation mechanism. P Natl Acad Sci USA. 2006; 103:17783–17788.

[8]. 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]

[9]. Herberg FW, Doyle ML, Cox S, Taylor SS. Dissection of the nucleotide and metalphosphate binding sites in cAMP-dependent protein kinase. Biochemistry. 1999; 38:6352–6360. [PubMed: 10320366]

[10]. Lochhead PA. Protein kinase activation loop autophosphorylation in cis: overcoming a Catch-22 situation. Sci Signal. 2009; 2:pe4. [PubMed: 19155529]

[11]. Emrick MA, Lee T, Starkey PJ, Mumby MC, Resing KA, Ahn NG. The gatekeeper residue controls autoactivation of ERK2 via a pathway of intramolecular connectivity. P Natl Acad Sci USA. 2006; 103:18101–18106.

[12]. Suel GM, Lockless SW, Wall MA, Ranganathan R. Evolutionarily conserved networks of residues mediate allosteric communication in proteins. Nat Struct Biol. 2003; 10:59–69. [PubMed: 12483203]

[13]. Valiev M, Yang J, Adams JA, Taylor SS, Weare JH. Phosphorylation reaction in cAPK protein kinase-free energy quantum mechanical/molecular mechanics simulations. J Phys Chem B. 2007; 111:13455–13464. [PubMed: 17983217]

[14]. Thompson EE, Kornev AP, Kannan N, Kim C, Ten Eyck LF, Taylor SS. Comparative surface geometry of the protein kinase family. Protein Sci. 2009; 18:2016–2026. [PubMed: 19610074]

[15]. 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]

[16]. Yang J, Kennedy EJ, Wu J, Deal MS, Pennypacker J, Ghosh G, Taylor SS. Contribution of non-catalytic core residues to activity and regulation in protein kinase A. J Biol Chem. 2009; 284:6241–6248. [PubMed: 19122195]

[17]. Elphick LM, Lee SE, Child ES, Prasad A, Pignocchi C, Thibaudeau S, Anderson AA, Bonnac L, Gouverneur V, Mann DJ. A quantitative comparison of wild-type and gatekeeper mutant cdk2 for chemical genetic studies with ATP analogues. Chembiochem. 2009; 10:1519–1526. [PubMed: 19437469]

[18]. Johnson DA, Akamine P, Radzio-Andzelm E, Madhusudan M, Taylor SS. Dynamics of cAMP-dependent protein kinase. Chem Rev. 2001; 101:2243–2270. [PubMed: 11749372]

[19]. Liu Y, Shah K, Yang F, Witucki L, Shokat KM. A molecular gate which controls unnatural ATP analogue recognition by the tyrosine kinase v-Src. Bioorg Med Chem. 1998; 6:1219–1226. [PubMed: 9784863]

[20]. Azam M, Seeliger MA, Gray NS, Kuriyan J, Daley GQ. Activation of tyrosine kinases by mutation of the gatekeeper threonine. Nat Struct Mol Biol. 2008; 15:1109–1118. [PubMed: 18794843]

[21]. Takaki T, Echalier A, Brown NR, Hunt T, Endicott JA, Noble MEM. The structure of CDK4/cyclin D3 has implications for models of CDK activation. P Natl Acad Sci USA. 2009; 106:4171–4176.
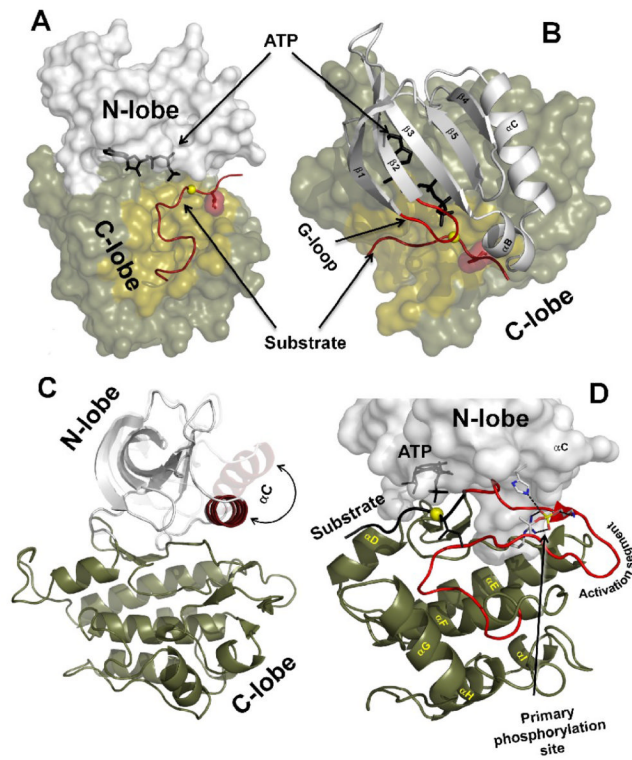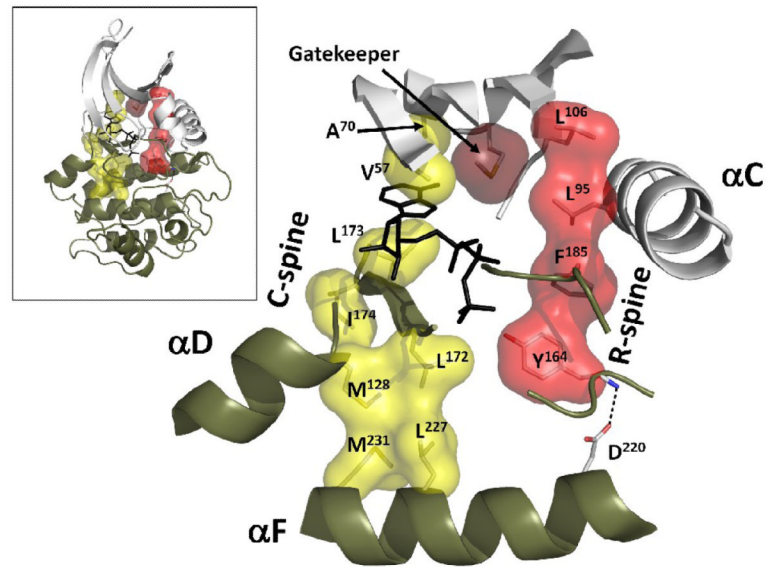
**Figure 1. General architecture of the protein kinase catalytic core**
A) Two lobes of a protein kinase molecule (PKA structure 1ATP was used). ATP is bound in a deep cleft between the lobes. The substrate binding groove on the C-lobe is colored sand with the phosphorylation (P) site shown as a yellow sphere. The residues that follow the P-site (P+1 residue) are shown as a dark red surface. B) Five β-strands and a prominent αC-helix from the N-lobe. The G-loop is colored red. C) A large swing motion of the αC-helix is associated with activation (Insulin receptor kinase structures 1IR3 and 1IRK were used). D) Helical content of the C-lobe. The Activation segment is colored red. The primary phosphorylation site, $T^{197}$ and its binding partners in PKA are shown as sticks.
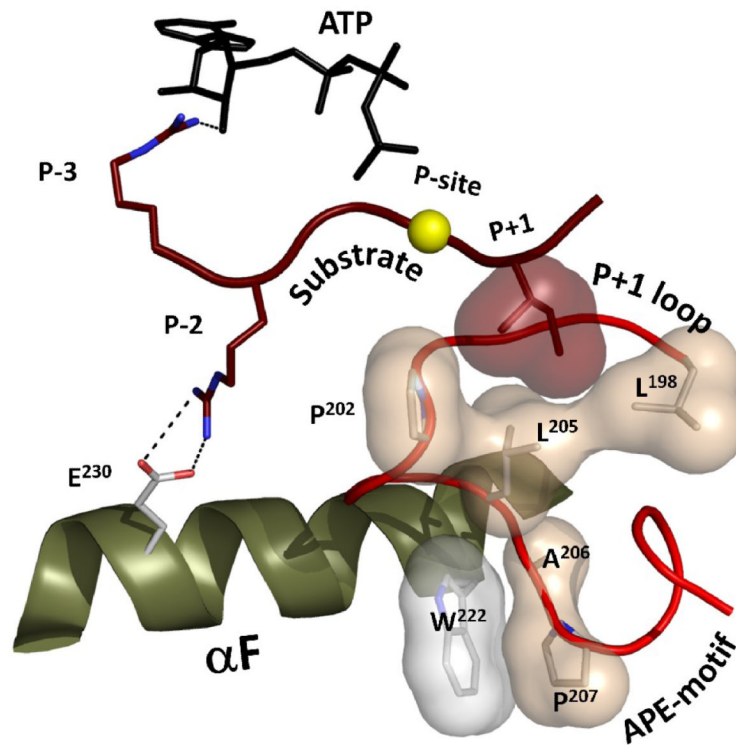
**Figure 2. The substrate peptide is positioned with respect to the αF-helix by a set of conserved hydrophobic residues in the Activation segment**
The phosphorylation site for a PKA substrate is shown as a yellow sphere. The hydrophobic pocket created by the P+1 loop is clearly anchored to the conserved $W^{222}$ in the αF-helix via the APE-motif. The P-2 arginine forms a salt bridge with $E^{230}$ in the C-terminus of the helix.

**Figure 3. Two hydrophobic spines define the internal architecture of every protein kinase**
The spines are anchored to the αF-helix and traverse both lobes of the kinase. The C-spine residues are shown as a yellow surface. The R-spine is colored red. The gatekeeper residue is located between the spines (shown as a dark red surface). (Insert: global view of the spines inside the kinase core).
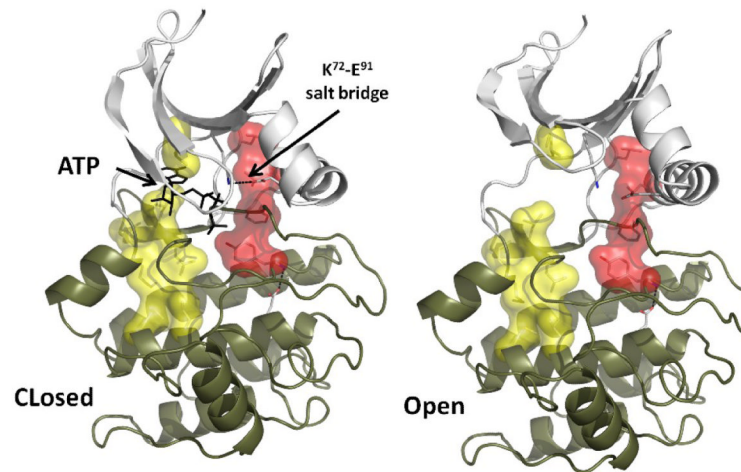
**Figure 4. The spines provide a flexible connection between the two lobes of the kinase core**
Closed and open conformations of active PKA are shown (PDB IDs 1ATP and 1J3H respectively). The conserved salt bridge between $K^{72}$ and $E^{91}$ is formed in the closed configuration and is disrupted in the open state due to the movements in the N-lobe. The spines remain intact during the "breathing" motion of the kinase, which is important for efficient catalysis.