# Crystal Structure of the Catalytic Subunit of Cyclic Adenosine Monophosphate–Dependent Protein Kinase

DANIEL R. KNIGHTON, JIANHUA ZHENG, LYNN F. TEN EYCK, VICTOR A. ASHFORD, NGUYEN-HUU XUONG, SUSAN S. TAYLOR, JANUSZ M. SOWADSKI*

**The crystal structure of the catalytic subunit of cyclic adenosine monophosphate–dependent protein kinase complexed with a 20–amino acid substrate analog inhibitor has been solved and partially refined at 2.7 Å resolution to an *R* factor of 0.212. The magnesium adenosine triphosphate (MgATP) binding site was located by difference Fourier synthesis. The enzyme structure is bilobal with a deep cleft between the lobes. The cleft is filled by MgATP and a portion of the inhibitor peptide. The smaller lobe, consisting mostly of amino-terminal sequence, is associated with nucleotide binding, and its largely antiparallel β sheet architecture constitutes an unusual nucleotide binding motif. The larger lobe is dominated by helical structure with a single β sheet at the domain interface. This lobe is primarily involved in peptide binding and catalysis. Residues 40 through 280 constitute a conserved catalytic core that is shared by more than 100 protein kinases. Most of the invariant amino acids in this conserved catalytic core are clustered at the sites of nucleotide binding and catalysis.**

PROTEIN PHOSPHORYLATION AS A MECHANISM FOR REGULating protein activity was first recognized with glycogen phosphorylase (*1, 2*). Now after over three decades, it is clear that this mechanism for the reversible post-translational modification of proteins is widespread and affects nearly all aspects of growth and homeostasis in the eukaryotic cell (*3*). The enzymes catalyzing this transfer of the γ-phosphate of magnesium adenosine triphosphate (MgATP) to a protein substrate, the protein kinases, constitute a large and diverse family of enzymes. Although these enzymes differ in size, substrate specificity, mechanism of activation, subunit composition, and subcellular localization, all, nevertheless, share a homologous catalytic core that has been conserved (*4*).

The first protein kinase to be purified was phosphorylase kinase (*5*). The second was phosphorylase kinase kinase, later renamed cyclic adenosine monophosphate (cyclic AMP)–dependent protein kinase (E.C. 2.7.1.37: ATP:protein serine phosphotransferase) (*6*). Not only was cyclic AMP–dependent protein kinase (cAPK) one of the first protein kinases to be characterized, it also is one of the simplest and is certainly the best understood biochemically (*7–10*). Its simplicity is due primarily to its mechanism of activation which involves subunit dissociation. With the exception of the oncogenic enzymes, all protein kinases typically are maintained in an inactive state in the absence of the appropriate activating signal. In the case of cAPK, the ligand triggering activation is cyclic AMP, one of the first recognized second messengers for hormone signaling. In the absence of cyclic AMP, the enzyme is sequestered as an inactive holoenzyme containing two regulatory (R) and two catalytic (C) subunits. When intracellular cyclic AMP is elevated, it binds to the R subunit, thus causing the complex to dissociate into an $R_2$ dimer and two free and active C subunits. The general consensus sequence recognized by the C subunit is Arg-Arg-X-Ser[Thr]-Y, where X is any small residue and Y is a large hydrophobic group (*8, 11, 12*). The conserved catalytic core in all protein kinases is contained within this relatively simple monomeric C subunit (*4*).

Owing to its simplicity as well as its relative ease of purification, the C subunit has been a prototype for identifying functional sites that are important for substrate recognition and catalysis. Affinity labeling, group specific labeling, and fluorescence energy transfer have provided clues about regions important for peptide recognition, MgATP binding, and catalysis (*8–10*). The expression of the C subunit in *Escherichia coli* (*13*) not only facilitated the structural studies, but also allows recombinant approaches to be used for probing active site regions. A wealth of information can also be extracted from the sequence similarities in the large protein kinase family (*4*). Such sequence comparisons have resulted in the identification of conserved regions, variable regions, and places where inserts and deletions can be tolerated. Both the chemical and sequence information provide a framework for interpreting the structure of the C subunit.

We now describe the crystal structure of the catalytic subunit of cyclic AMP–dependent protein kinase. Just as the chemical information gleaned from the C subunit serves as a framework for

D. R. Knighton, J. Zheng, and S. S. Taylor are in the Department of Chemistry, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093–0654. L. F. Ten Eyck is in the Department of Chemistry, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093–0654, and the San Diego Supercomputer Center, P.O. Box 85608, San Diego, CA 92186–9784. V. A. Ashford is in the Department of Biology, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093–0317. N-h. Xuong is in the Departments of Chemistry, Biology, and Physics, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093–0319. J. M. Sowadski is in the Departments of Medicine and Biology, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093–0654.

*To whom correspondence should be addressed.

interpreting the entire kinase family, this structure provides a template for viewing the conserved catalytic core of all eukaryotic protein kinases.

**Structure solution**. The C subunit structure described below is of a binary complex of the recombinant mouse Cα subunit with a bound, high-affinity ($K_i$ = 2.3 nM) inhibitor peptide. The peptide, PKI(5–24), is derived from the $NH_2$-terminal region of the naturally occurring thermostable protein kinase inhibitor protein (PKI) (*14*). The steady-state kinetics of the C subunit, purified from *E. coli*, are identical to the mammalian C subunit, although the *E. coli* protein is more labile to heat denaturation (*13*). Unlike the mammalian enzyme, the recombinant C subunit lacks a myristoyl group at its amino terminus.

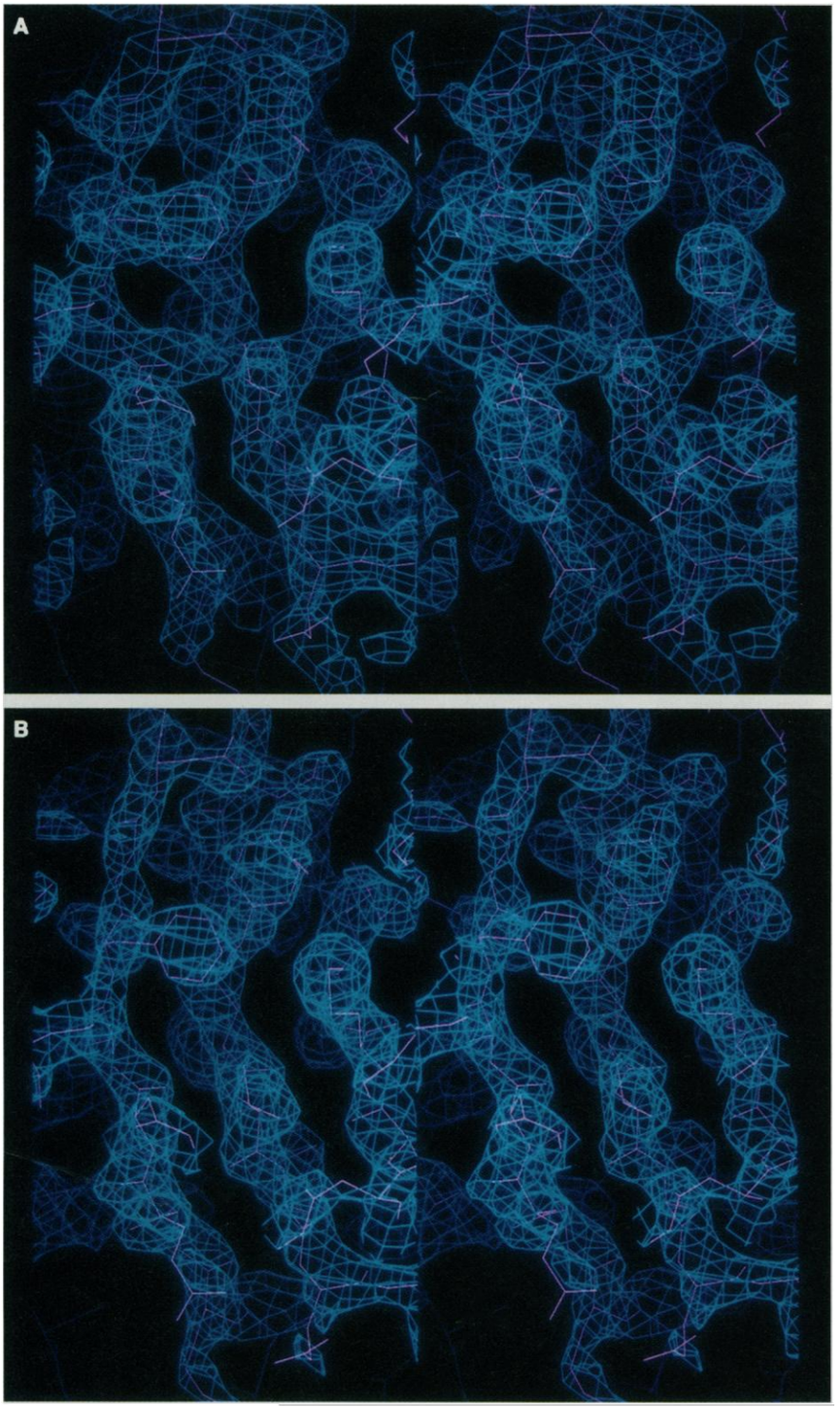


**Fig. 1.** Stereo view of the electron density for the structure determination. Portions of the latest refined model of three β strands are shown (top to bottom from left): residues 112 to 106, 114 to 121, 75 to 69. (**A**) Experimental density at 1.5σ calculated to 2.7 Å with the use of phases after Wang improvement and extension. (**B**) The ($2F_0 - F_c$) density at 1.5σ calculated with 10 to 2.7 Å X-PLOR model phases.

The crystallization of the C subunit has been reviewed (*15*). Porcine heart C subunit was co-crystallized first in space group $P6_122$ as a ternary complex with MgATP and PKI(5–24) (*16*). Slight modification of the $P6_122$ conditions allowed co-crystallization of recombinant mouse C subunit in space group $P2_12_12_1$, likewise a ternary complex with MgATP and PKI(5–24) (*17*). Only after small-angle neutron scattering studies (*18*) indicated that PKI(5–24) could effectively form a complex with C subunit in the absence of MgATP, was the recombinant mouse C subunit co-crystallized under the same conditions as a binary complex with PKI(5–24), and only this binary C:PKI(5–24) crystal was successfully derivatized for structure solution (Table 1).

The binary complex co-crystallized in space group $P2_12_12_1$ with $a$ = 73.62 Å, $b$ = 76.52 Å, and $c$ = 80.14 Å (*15*). The asymmetric unit contained one C:PKI(5–24) complex with a calculated solvent content of 0.53 and $V_m$ = 2.62 Å$^3$/D (*19*). Diffraction data for this crystal form and for the mouse recombinant C:MgATP:PKI(5–24) ternary complex co-crystal (space group $P2_12_12_1$, $a$ = 73.67 Å, $b$ = 76.56 Å, $c$ = 80.39 Å) used in difference Fourier work were measured to 2.7 Å with Xuong-Hamlin area detectors (*20*). Data from two binary complex crystals were merged to provide a more complete native data set for the initial phase computation, phase extension, and initial refinement cycles, but in later refinement more precise data from only a single crystal were used. Derivative co-crystals were prepared by exposing the C:PKI(5–24) complex in solution to 1 mM 4-(hydroxymercuri)benzoic acid (PHMB) for 6 hours prior to crystallization (Table 1). Data from two PHMB co-crystals, one with measured Bijvoet mates, were kept separate and used to compute phases, which were improved at 3.5 Å resolution with the solvent flattening approach of Wang (*21*). Phases from inversion of the solvent-flattened map were added in six steps from 3.5 Å to a final resolution of 2.7 Å to form the initial phase set. Starting with minimap Cα coordinates, the program TOM-FRODO (*22, 23*) was used with the resulting map (Fig. 1) to model in the sequence of PKI(5–24) and 257 of the 350 residues in the C subunit. This partial model was refined with X-PLOR (*24*) to improve the coordinates, and calculated structure factors were combined with the

use of equally weighted ABCD coefficients (*25*) with Wang structure factors in a resolution-dependent manner (*26, 27*) to yield improved maps. Iterations of refinement, combination, and building in areas outside the partial model used in refinement completed the C subunit except for the 14 $NH_2$-terminal residues, which have no electron density and are presumed disordered.

The current model consists of C subunit residues 15 to 350 and 1 to 20 of PKI(5–24) and has been partially refined with X-PLOR (*24*) and TNT (*28*) to an *R* factor of 0.212 against all data to 2.7 Å. A native anomalous difference Fourier calculation with X-PLOR phases confirmed with 2.5-σ peaks the sulfur atom position in six of eight sulfur-containing residues, located between 71 to 343 in C subunit residue number, and also confirmed the location of the phosphorus of phosphorylated $Thr^{197}$ with the highest peak in the map. PHMB bound to the protein at $Met^{58}$ and $Cys^{343}$ (major site). The MgATP binding site (Fig. 2) was located by difference Fourier synthesis with the C:MgATP:PKI(5–24) ternary complex co-crystal data. The structure of the peptide inhibitor and its interaction with the C subunit are described by Knighton *et al.* (*29*). Here we focus on the C subunit structure and its implications for the protein kinase family.

**Overall architecture.** Examination of the backbone structure of the C subunit with the bound peptide (Fig. 2) indicates, from the overall dimensions of the monomer (65 by 45 by 45 Å), a slightly elongated molecule. Earlier hydrodynamic measurements showing a Stokes radius of 26.1 Å, a frictional coefficient ratio ($f/f_0$) of 1.19, and a radius of gyration of 20 Å are consistent with this structure (*18, 30*). The most striking feature of the overall molecular architecture is its bilobal shape with a deep cleft between the two lobes. The core of the small lobe is associated primarily with the $NH_2$-

**Table 1.** Structure solution statistics. Recombinant mouse C subunit crystals as binary complex with PKI(5–24) or as ternary complex with MgATP and PKI(5–24) were prepared as described according to Zheng *et al.* (*17*). Briefly, C subunit was brought to a final concentration of 10 mg/ml and dialyzed against 50 mM Bicine buffer (pH 8.0) and 150 mM ammonium acetate. Crystals were grown at 4°C by hanging drop vapor diffusion. The reservoir contains 10 mM dithiothreitol (DTT) and 8 percent (w/v) polyethylene glycol 400. Methanol was added to the reservoir solution to a concentration of 15 percent before sealing the cover slip. The drop contains equal volumes of protein solution, reservoir solution without methanol, and 10 mM DTT solution containing PKI(5–24) or PKI (5–24) and MgATP. Binary complex co-crystals (native and PHMB) were mounted from mother liquor, but the ternary complex co-crystal for the difference Fourier was soaked for 96 hours in reservoir solution containing ten times more $Mg^{2+}$ (4.4 mM) to stabilize the position of the β- and γ-phosphates prior to data collection. Diffraction data were measured at 4°C with graphite-monochromated CuKα x-rays (Mark III Rigaku RU-200 rotating anode diffractometer of the UCSD Research Resource (*59*) equipped with two Xuong-Hamlin multiwire area detectors (*20*). Paired runs starting from settings (ω,ϕ,χ) and (ω,ϕ+180,−χ) were used to collect Bijvoet mates (inverse beam method). Data reduction and derivative-to-native scaling were done with the UCSD area detector data processing programs (*60*). $R_{sym} = \Sigma|I_{obs} - I_{avg}|\Sigma I_{avg}$ and is shown for merged Friedel pairs. The mean fractional difference between binary and ternary complex amplitudes was 0.173 for data to 2.7 Å. *Phasing.* Native-1 was used for native. Hg positions of the PHMB co-crystal derivatives were found from a difference *Patterson synthesis.* Positional and relative occupancy refinement of two common sites (relative occupancies 2.66, 1.87 for PHMB-1 and 3.26, 1.55 for PHMB-2), and calculation of native phases and corresponding ABCD coefficients (*25*), were done with the program HEAVY (*61*). Solvent flattening was done with the Wang programs (*21*) on imported initial ABCD coefficients (*25*) and phases to 3.5 Å. Molecular envelopes were calculated with a slightly conservative solvent content of 0.50. After three envelopes at 3.5 Å, phases were extended in six steps of resolution, each with a new molecular envelope, to a final resolution of 2.7 Å. After convergence at 3.5 Å, the mean phase change per reflection was 36.6° and the mean figure of merit was 0.84; the map inversion *R* factor was 0.181. Phase extension added 6786 phases from 5914 in the 3.5 Å starting set; 261 unobserved reflections were estimated by map inversion in the 2.7 Å set. Definitions: $f_h$, calculated heavy-atom structure factor amplitude; $F_p$, measured native structure factor amplitude; $F_{ph}$, measured derivative structure factor amplitude; $\Delta F_{anom}$, calculated Bijvoet difference; $E_{iso}$, rms isomorphous lack-of-closure; $E_{anom}$, rms anomalous lack-of-closure; $R_c = \Sigma||F_{ph} \pm F_p| - f_h|/\Sigma|F_{ph} - F_p|$. *Refinement.* X-PLOR refinement began with the partial model of stage A to improve the coordinates for structure factor combination. Combined maps were calculated with the Hendrickson-Lattman scheme (*25*) following others (*26, 27*). Wang phases were used to 6 Å, combined ones between 6 Å and 3.5 Å or 3.0 Å, and calculated phases between 3.5 or 3.0 Å and 2.7 Å. The corresponding weighted amplitudes were $m_{Wang}F_o$, $m_{comb}$ ($2F_o - F_c$), and $m_{Sim}$ ($2F_o - F_c$) (*62*). X-PLOR refinement and *R* factor ($\Sigma|F_o - F_c|/\Sigma F_o$) calculations were made with $F/\sigma > 2$ reflections in the 10 to 2.7 Å range (12014 Native-1 reflections; 10194 Native-2 reflections beginning with stage B). The coordinates corresponding to the latest X-PLOR *R* factor of 0.195 had rms deviations from ideal bond length of 0.024 Å and from ideal angles of 4.3°. Positional TNT refinement was run on these coordinates with the use of all data to 2.7 Å and improved the *R* factor while improving the geometry to rms deviations from ideal bond length of 0.0016 Å and from ideal angles of 3.6°. Ramachandran plots of the latest X-PLOR coordinates and the TNT coordinate set are similar; the TNT plot has 24 of 336 non-glycine residues outside of allowed regions, about half of which are concentrated in the surface loop between residue 308 to 326 and the nucleotide loop between 54 and 57 ($Met^{58}$ is a heavy-atom binding site, and there is clear omit map density for $Phe^{54}$ side chain). The model has 2939 atoms with rms. B = 17.6 $Å^2$. Because the refinement is still in progress, no solvent molecules have been included yet.

| | Crystals (No.) | $d_{min}$ (Å) | Measurements (No.) | Reflections (No.) | Completeness | | |
|---|---|---|---|---|---|---|---|
| | | | | | $<I/\sigma(I)>$ | % | $R_{sym}$ |
| Diffraction data | | | | | | | |
| Native-1 | 2 | 2.7 | 58889 | 12713 | 12.9 | 98.1 | 0.061 |
| Native-2 | 1 | 2.7 | 27067 | 11291 | 13.2 | 87.3 | 0.040 |
| PHMB-1 | 1 | 3.0 | 30973 | 7233 | 13.3 | 76.1 | 0.063 |
| PHMB-2 | 1 | 3.0 | 23476 | 8809 | 6.9 | 92.1 | 0.075 |
| MgATP | 1 | 2.7 | 26464 | 11840 | 11.1 | 91.1 | 0.048 |

| | Average shell resolution (Å) | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Overall | 11.72 | 7.70 | 6.11 | 5.22 | 4.63 | 4.20 | 3.87 | 3.61 |
| SIRAS phasing statistics | | | | | | | | | |
| Mean figure of merit | 0.47 | 0.74 | 0.75 | 0.68 | 0.62 | 0.55 | 0.53 | 0.47 | 0.43 |
| PHMB-1 | | | | | | | | | |
| Acentric rms $f_h/E_{iso}$ | 2.73 | 3.27 | 3.80 | 3.00 | 2.54 | 2.56 | 2.26 | 2.05 | |
| rms $\Delta F_{anom}/E_{anom}$ | 0.96 | 1.53 | 1.68 | 1.41 | 1.19 | 1.05 | 0.89 | 0.78 | 0.60 |
| $R_c$ | 0.50 | 0.34 | 0.38 | 0.45 | 0.51 | 0.78 | 0.64 | 0.70 | 0.81 |
| PHMB-2 | | | | | | | | | |
| Acentric rms $f_h/E_{iso}$ | 2.26 | 3.89 | 3.72 | 3.35 | 2.81 | 2.34 | 2.08 | 1.83 | 1.51 |
| $R_c$ | 0.60 | 0.37 | 0.53 | 0.58 | 0.64 | 0.75 | 0.71 | 0.65 | 0.61 |

| | Residues | Chains | *R* factor Initial | *R* factor Final | B | Data selection |
|---|---|---|---|---|---|---|
| Refinement | | | | | | |
| Models | | | | | | |
| (A) First unrefined partial | 275 | 4 | 0.473 | 0.304 | Overall | 10–2.7 Å, $F/\sigma > 2$ |
| (B) First unrefined full | 356 | 2 | 0.434 | 0.228 | Overall | 10–2.7 Å, $F/\sigma > 2$ |
| (C) Latest X-PLOR | 356 | 2 | | 0.195 | Individual | 10–2.7 Å, $F/\sigma > 2$ |
| (D) TNT | 356 | 2 | 0.221 | 0.212 | Individual | 40–2.7 Å |

terminus, while the core of the large lobe corresponds to the COOH-terminal region of the protein. The cleft between the lobes is filled by a portion of the bound inhibitor peptide in the binary complex. A difference Fourier map of the ternary complex containing both peptide and MgATP places MgATP at the base of that cleft (Figs. 2 and 3). The cleft is clearly the site of catalysis, and the peptide-induced conformational changes, observed by both small-angle neutron scattering (*18*) and circular dichroism (*30*), may be associated with a closing of this cleft. Neutron scattering, in particular, established that the apoform of the enzyme adopts a more expanded conformation than the ternary complex containing MgATP and PKI(5–24) (*18*). Furthermore, PKI(5–24) alone, but not MgATP, was sufficient to induce this conformational change. Whether this change in conformation correlates with a closing of the cleft must await a crystallographic solution of the apoenzyme.

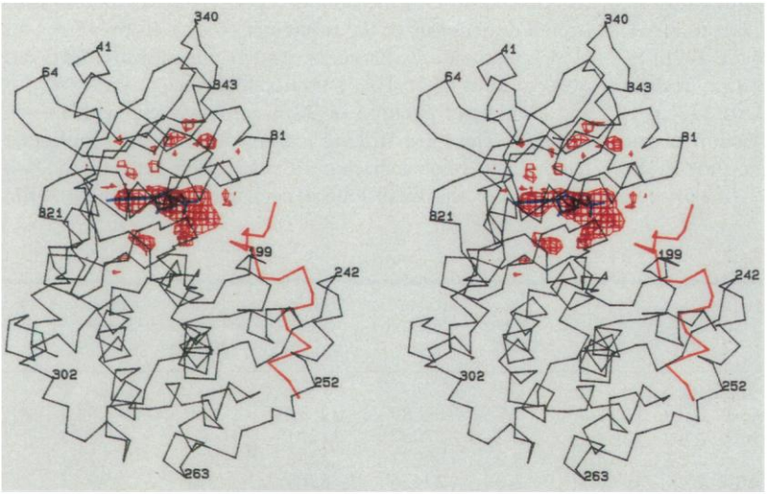


**Fig. 2.** Cα backbone with MgATP location and orientation. 3.5-σ positive density contours (red) for the ($F_{ternary} - F_{binary}$) difference Fourier calculated with X-PLOR model phases in 10 to 2.7 Å range are superimposed on the X-PLOR Cα backbone of C subunit residues 15 to 350 (black) and the 20 residues of PKI(5–24) (red). A model of AMP is shown inside the difference density with the phosphate at right.

Although most of the predictions of secondary structure in the C subunit have been quite inaccurate and do not correlate well with the actual structure (*8, 31–34*), the recent prediction by Benner and Gerloff is an exception. Their prediction of the secondary structure, based on chemical information and homologies within the protein kinase family, is remarkably accurate, particularly for the small lobe (*35*).

The $NH_2$-terminus of the C subunit begins with an amphipathic α helix that lies primarily along the surface of the larger lobe. This $NH_2$-terminal region differs in the recombinant and mammalian enzymes, since the recombinant protein lacks a myristoyl group at the $NH_2$-terminal glycine (*13*). In the crystal structure, the first 14

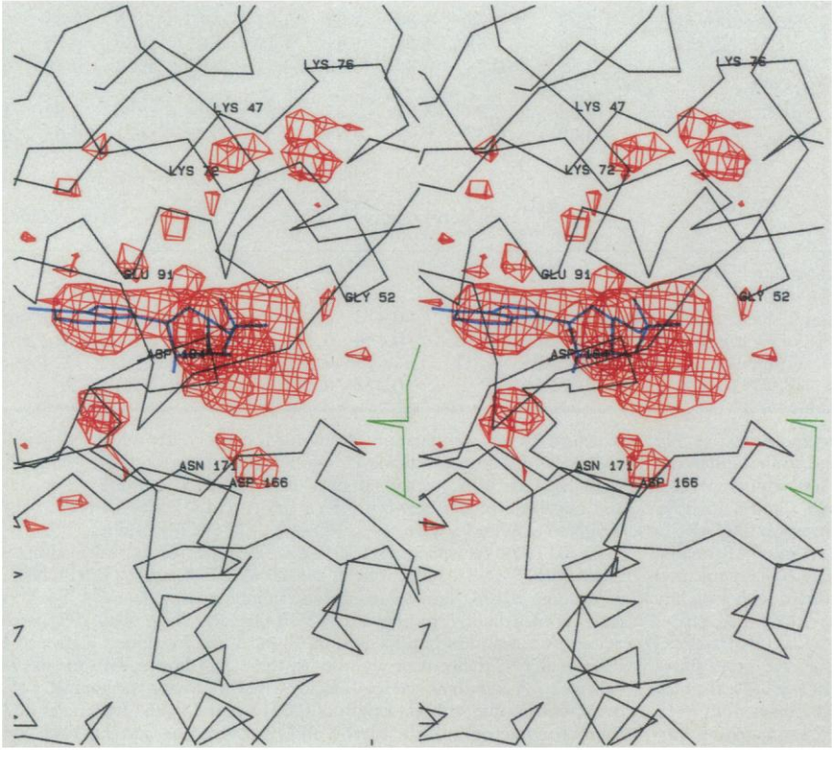


**Fig. 3.** MgATP binding site. An enlarged view of the difference density of Fig. 2 (red) is shown superimposed on the Cα backbone near the MgATP binding site. AMP is modeled in the difference density (blue) and the backbone of PKI(5–24) is shown, in part, in green. The β- and γ-phosphates of ATP are not modeled because their positions, along with any metal ion or ions, have not been unambiguously established, although sufficient density exists to accommodate them.

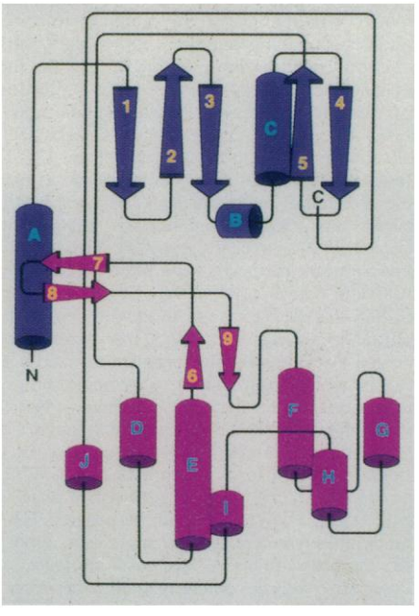


**Fig. 4.** C subunit topology. Residues corresponding to the secondary structure elements are as follows. For the β strands, 1, 43 to 48; 2, 57 to 63; 3, 67 to 75; 4, 106 to 111; 5, 115 to 120; 6, 161 to 164; 7, 171 to 175; 8, 178 to 183; 9, 188 to 191; for the α helices, A, 15 to 31; B, 76 to 82; C, 84 to 97; D, 128 to 135; E, 140 to 159; F, 218 to 233; G, 244 to 252; H, 263 to 272; I, 288 to 293; J, 301 to 307. N, amino terminal; C, carboxyl terminal.

amino acids are not visible. However, the surface of the enzyme in this $NH_2$-terminal region is hydrophobic, suggesting a possible site for the $NH_2$-terminal myristoyl moiety of the mammalian enzyme. The myristoyl group stabilizes the C subunit but does not promote association with membranes (*36*).

The smaller lobe, consisting of residues 40 through 125, is associated primarily with the binding of the nucleotide and is characterized by a dominance of β structure. Five antiparallel β strands comprise the core of this domain. The only helical element in the small lobe is inserted between β strands 3 and 4 and lies on one side of the plane of the β sheet. It consists of two parts—a two-turn helix, helix B, followed by a sharp break and a five-turn helix, helix C. On the basis of (i) a difference Fourier map (Fig. 3) with a ternary complex of the recombinant C subunit containing MgATP and PKI(5–24) and (ii) chemical evidence discussed below, this small lobe is the primary site for interaction with MgATP. The density based on the difference map is consistent with the adenine moiety of the nucleotide oriented toward the base of the cleft beneath the β sheet, with the phosphate facing outward, toward the edge of the cleft (Fig. 3). This structure is distinct from the Rossmann fold, which is characteristic of many nucleotide binding proteins (*37*).

The larger lobe, in contrast, is predominantly helical and has seven α helices. Particularly unusual are the antiparallel hydrophobic helices, helix E (residues 140 through 159) and especially helix F (residues 218 through 233), which extend through the core of this domain. The only region of β structure in this lobe is located on the surface of the cleft at the interface between the two lobes where four antiparallel β strands form a sheet. Most of the regions important for peptide recognition, as well as some conserved residues likely to be involved in catalysis, are located within this larger lobe.

The COOH-terminal 70 amino acids, residues 281 through 350, extend over a large portion of the surface of the enzyme from the bottom of the large lobe to the top of the small lobe. The part of this extended chain that passes through the region linking the two lobes appears to participate in recognition of both the peptide and the nucleotide, even though these amino acids are outside the conserved catalytic core. The other extended chain connecting the two lobes of the enzyme, residues 120 through 127, likewise, passes through this linker region between the small and large lobe and also participates in peptide recognition. Hence, this linking region consisting of both chains may contribute in part to the observed peptide-induced conformational changes described above (Fig. 4).

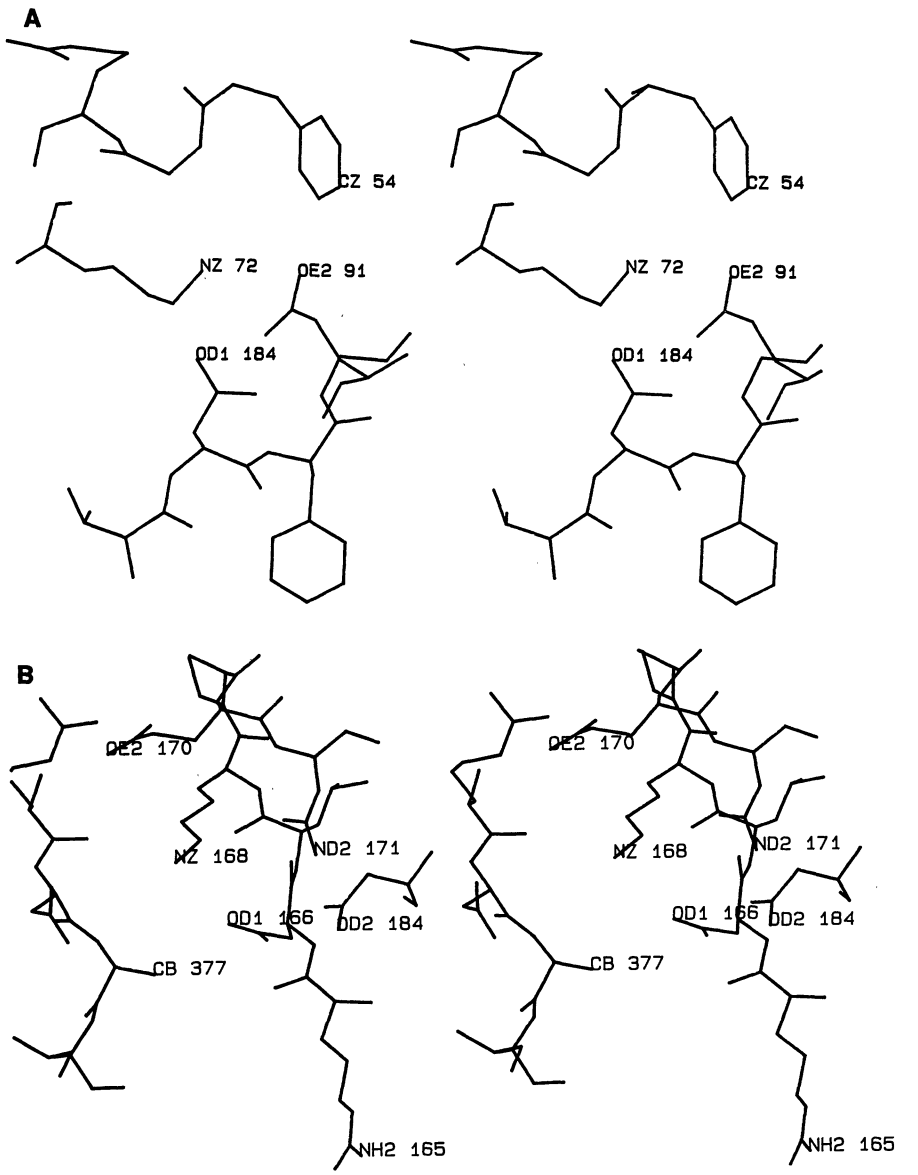


**Fig. 5.** Stereo views of selected conserved areas. (**A**) The side chains of the invariant $Lys^{72}$, $Glu^{91}$, and $Asp^{184}$ are shown in proximity to each other, along with the $Phe^{54}$ of the glycine-rich sequence ($Gly^{50}$, $Thr^{51}$-$Gly^{52}$-$Ser^{53}$-$Phe^{54}$-$Gly^{58}$) of the nucleotide binding loop. (**B**) The catalytic loop, $Arg^{165}$-$Asp^{166}$-$Leu^{167}$-$Lys^{168}$-$Pro^{169}$-$Glu^{170}$-$Asn^{171}$, is shown together with part of PKI(5–24). The PKI(5–24):$Ala^{21}$ is numbered 377 and is a Ser or Thr in normal substrates.

**Correlation with chemical data.** Many approaches have been used to chemically define this enzyme. In nearly all cases, the three-dimensional structure of the enzyme-inhibitor complex provides a solid explanation for these earlier chemical results, and, conversely, this correlation confirms the correct interpretation of the electron density map.

Evidence for localizing the nucleotide binding site near the $NH_2$-terminus first came from affinity labeling with an analog of MgATP, fluorosulfonylbenzoyl adenosine (FSBA). Reaction with FSBA leads to inactivation due to the covalent modification of $Lys^{72}$ (*38*) which lies near the β,γ-phosphate subsite (*39*). Labeling with a hydrophobic carbodiimide, DCCD, identified two carboxyl groups near the MgATP binding site, $Asp^{184}$ and $Glu^{91}$, and, furthermore, established that $Asp^{184}$ could be readily cross-linked to $Lys^{72}$ in the apoenzyme (*40, 41*). The structure of the binary complex without bound MgATP confirms that all three residues are localized in close proximity to one another (Fig. 5), while the difference Fourier map with the ternary complex places these residues close to the γ-phosphate region of MgATP although only the α phosphate is shown in Figs. 2 and 3. $Lys^{72}$ is on β strand 3, and $Glu^{91}$ lies along the edge of the C helix that faces the cleft. $Asp^{184}$ is located on the loop connect-

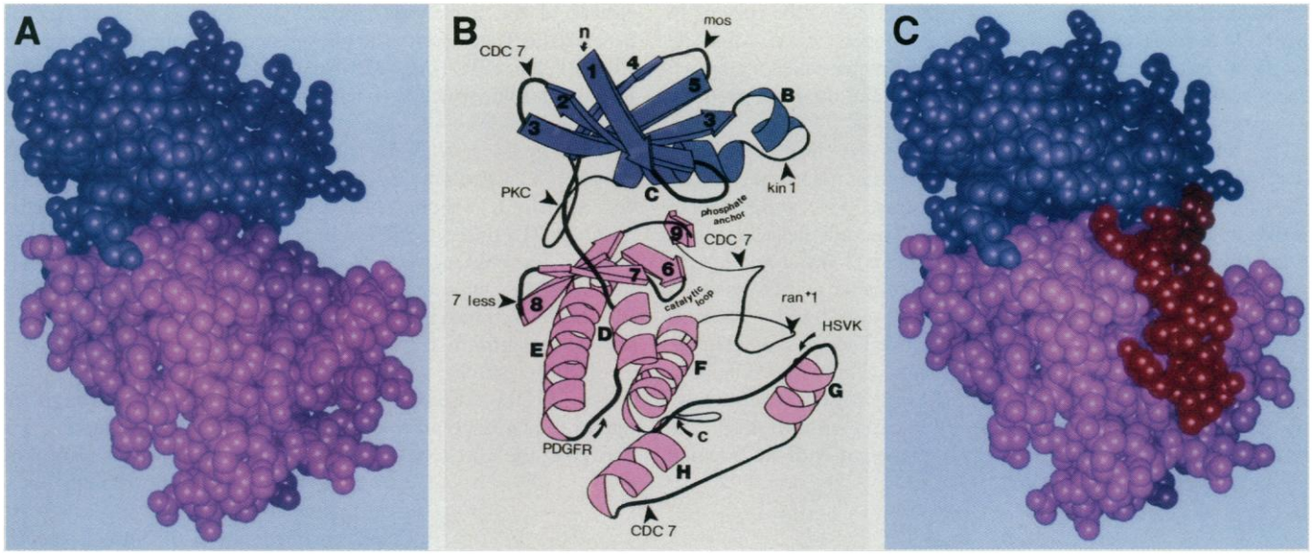


**Fig. 6.** Conserved catalytic core. (**A**) A space-filling model of the catalytic core (residues 40 to 280) shared by all protein kinases is shown. The small lobe corresponding to the nucleotide binding fold (residues 40 to 126) is indicated in purple; the larger lobe (residues 127 to 280) is shown in pink. In this model, the bound peptide is not shown. (**B**) Diagram of the conserved catalytic core obtained with the RIBBON (*57*) program of PAP package (*58*). Positions of the observed inserts in the protein kinase catalytic domain from the Hanks *et al.* alignment (*4*) are indicated. The protein kinase having the largest insert at each position is designated by the following notation to define each insert: Gene/Protein name: $NH_2$-terminal C subunit residue number (length of insert), COOH-terminal C subunit residue number. The inserts are CDC7, 64(14)65; kin1 83(26)84; PKC-γ, 98(6)99; c-mos, 113(5)114; PDGFR, 137(99)138; CDC7, 196(82)197; $ran^{+1}$, 210(23)211; HSVK, 240(11)241; CDC7, 260(93)261; 7less, 178(7)179. (**C**) This panel is identical to (A), but includes PKI(5–24), shown in red. Inserts were identified from the alignment of Hanks *et al.* (*4*); c, carboxyl terminal; n, amino terminal.

ing β strands 8 and 9, and this loop also lines the cleft. All three residues are invariant in the protein kinase family. A precise description of the location of the β- and γ-phosphates of MgATP should follow when the complete structure of the ternary complex containing MgATP, peptide, and C subunit is described (*15*).

The MgATP binding site was defined more globally by differential labeling with acetic anhydride. By describing the reactivity of each lysine side chain in the presence and absence of substrates, it has been shown that the specific protection afforded by MgATP was localized exclusively to residues in the small lobe (*42*). In addition to $Lys^{72}$, MgATP protected $Lys^{76}$ and $Lys^{47}$ against modification by acetic anhydride. These protected lysines also flank the conserved glycine-rich loop that lies between β strands 1 and 2. On the basis of the difference Fourier (Figs. 2 and 3), this loop is close to the phosphates of MgATP.

The peptide binding site also was localized initially with a peptide affinity analog that led to the stoichiometric modification of $Cys^{199}$ (*43*). Several independent studies placed $Cys^{199}$ near the active site, specifically near the γ-$PO_4$ subsite (*39*). Modification of $Cys^{199}$ leads to loss of activity, and MgATP protects against inactivation (*44, 45*); the other cysteine in the C subunit, $Cys^{343}$, can be covalently modified with no concomitant loss of activity. $Cys^{199}$ is on the surface of the cleft that interacts with the COOH-terminus of the inhibitor peptide, and $Cys^{343}$ is on the surface of the small lobe. Modification of one cysteine side chain with a fluorescence donor and the other with a fluorescence acceptor enabled us to measure the distance between the two. This minimum distance of 31 Å is consistent with the 24 Å distance measured between the two α carbons of $Cys^{199}$ and $Cys^{343}$ in the crystal structure (*46*).

**Conserved regions and their functions.** The fact that all known protein kinases share a conserved catalytic core that is homologous to the C subunit provides us with additional information that independently highlights important regions. This conserved catalytic core begins with the β 1 strand in the small lobe and extends at least through $Arg^{280}$ in the large lobe (*4*). The two lobes comprising this conserved catalytic core can be seen clearly in Fig. 6. Within this conserved core are nine invariant amino acids, as well as several highly conserved residues. Most of these conserved residues contribute directly to either MgATP binding or catalysis (*29*). Others, such as $Arg^{280}$ and $Asp^{208}$, exist as ion pairs and link two segments of the polypeptide chain that are widely separated in the linear sequence.

In addition to providing information on conserved residues, sequence comparisons among protein kinases also identify inserts, sometimes quite sizable, that lie within the catalytic core (*4*). Those inserts (Fig. 6) invariably are located at loops on the surface of the protein and can be accommodated within the tertiary structure. Two highly conserved loops, as well as a triad of invariant charged residues, appear to be particularly important for nucleotide binding and catalysis, while the regions important for recognition of the peptide substrate are quite variable.

The conserved glycine-rich segment, $Gly^{50}$-Thr-$Gly^{52}$-Ser-Phe-$Gly^{55}$, was identified originally as part of the MgATP binding site because of (i) its proximity to $Lys^{72}$ (*38*) and (ii) differential labeling with acetic anhydride, since all of the lysines flanking this region—$Lys^{47}$, $Lys^{72}$, and $Lys^{76}$—are protected in the presence of MgATP (*42*). The specific structural explanation for the protection of $Lys^{47}$ is due to ionic pairing with the side chain of $Glu^{333}$ while $Lys^{76}$ ion-pairs with $Glu^{346}$. $Lys^{76}$ is not very reactive even in the absence of substrates, whereas $Lys^{72}$ and $Lys^{47}$ are reactive in the absence of MgATP and peptide. The differential labeling thus highlights several components of the conformational changes that occur around the glycine-rich loop as a consequence of MgATP and peptide binding.

A glycine-rich motif is associated with many nucleotide binding sites, and this region has been the subject of much model building (*32, 34, 47–49*). The Rossmann fold, found in many nucleotide binding sites, contains a sheet of mostly parallel β strands with

helices above and below the plane of the sheet (*37, 48*). This motif begins with a β strand followed by a sharp turn and an α helix, and the glycine-rich loop is typically located at this first turn. The glycine-rich loop is the only region of the Rossmann fold motif that shows sequence conservation, and it typically interacts with nontransferable phosphates in the nucleotide with the dipole of the helix pointing toward the phosphate (*50*). The base of the nucleotide lies along the edge of the sheet. A similar motif containing a glycine-rich loop is found in other proteins such as adenylate kinase and $p21^{ras}$ (*49, 51*). This loop also lies between a β strand and a helix although, in general, the loop is somewhat longer than in the dehydrogenases. Also, in this so-called P loop a conserved Lys directly follows the loop, and its side chain folds over to interact with the α carbonyl of the first glycine in the loop.

In contrast, the protein kinase fold found in the C subunit and conserved in more than 100 protein kinases, does not conform to either the Rossmann fold or P loop motif; it forms a unique nucleotide binding site. (i) The glycine-rich segment lies at a sharp turn that joins two antiparallel strands at the beginning of the β sheet. (ii) The phosphate binding site is not dominated by a helix whose dipole points toward the phosphate. (iii) The nucleotide does not lie along the edge of the β sheet. Instead, it is buried under the β sheet. (iv) An invariant Lys does not immediately follow this loop. Instead, the invariant Lys in the protein kinases, $Lys^{72}$, is located in the β3 strand and is a part of the stable scaffold of the structure. A basic residue, $Arg^{56}$, does follow the glycine-rich loop in the C subunit; however, it is not conserved in the overall protein kinase family (*4*). A β strand–loop–β strand motif is present in hexokinase (*52*), HSC70 (*53*), and actin (*54*); however, only the actin three-dimensional structure provides a detailed analysis of the bound nucleotide. In this case, the location of the nucleotide with respect to the antiparallel β strand motif is different from that in the C subunit. A preliminary comparison of actin with HSC70 suggested that the nucleotide base binds the same in both proteins but that phosphate binding differs (*54*). The single conserved element in all of these nucleotide binding motifs is the glycine-rich loop whose apparent function is to serve as a phosphate anchor.

Another highly conserved loop in the C subunit extends from $Arg^{165}$ through $Asn^{171}$ and can be termed the catalytic loop (Fig. 5B). This catalytic loop, $Arg^{165}$-$Asp^{166}$-$Leu^{167}$-$Lys^{168}$-$Pro^{169}$-$Glu^{170}$-$Asn^{171}$, contains two invariant residues, $Asp^{166}$ and $Asn^{171}$, and two highly conserved residues, $Arg^{165}$ and $Leu^{167}$. While the purpose of the glycine loop is to anchor the phosphate moiety and, in particular, to help position the γ-phosphate so that it is poised for transfer, it is the catalytic loop that appears to be the central hub that communicates to many different parts of the molecule. This loop not only directs the catalytic event, but also guides the peptide into its proper orientation so that catalysis can occur. The loop itself and, in particular, the residues that are important for catalysis are highly conserved, while the parts of the loop that direct the peptide binding are not.

$Asp^{166}$ is one of four invariant carboxyl groups in the protein kinase family. It is the only one that is oriented toward the Ala side chain at the pseudo-phosphorylation site in the bound inhibitor peptide. Hence, if there is a catalytic base, as kinetic studies suggest (*55*), this is the most likely candidate. Ho *et al.* demonstrated that catalysis occurs at a direct in-line transfer with inversion of configuration confirming that no enzyme-bound phospho intermediate occurs during catalysis (*56*).

The triad composed of the side chains of $Lys^{72}$, $Asp^{184}$, and $Glu^{91}$ (Fig. 5A), is conserved throughout the protein kinase family and is close to the γ-phosphate of MgATP. Although $Asp^{184}$ was a candidate for the catalytic base, the structure indicates that a more plausible role is participation in the chelation of $Mg^{2+}$ in the MgATP complex. The side chain of $Asp^{184}$ also comes within 4 to 5 Å of the side chain of $Asn^{171}$. This cluster, $Asp^{184}$, $Asn^{171}$, and $Asp^{166}$, thus forms a second triad of invariant amino acids. A component of both triads, $Asp^{184}$ has the potential to shuttle between the two conserved loops, the glycine-rich loop in the small lobe and the catalytic loop in the larger lobe. Hence, if the position of $Asp^{184}$ changes after the binding of MgATP, as it probably would because of its location in the structure relative to the MgATP binding site, the consequences should have a direct impact on both conserved loops. If, for example, $Asp^{184}$ participates in the chelation of $Mg^{2+}$, its negative charge would be sequestered from the catalytic loop, thus allowing the other residues to rearrange in order to maximize the nucleophilicity of the serine hydroxyl moiety that is poised to receive the phosphate from ATP.

The protein kinases represent a large family of more than 100 enzymes that includes growth factor receptors as well as many oncoproteins. In spite of the tremendous diversity of these enzymes, all share a conserved catalytic core that retains the same essential features of secondary and tertiary structure and the same general mechanism of catalysis. The essential hallmarks of this conserved core include (i) two lobes with a cleft between that is occupied by the substrates, (ii) a characteristic nucleotide binding fold dominated by β structure, (iii) a large helical domain associated with peptide binding and catalysis, (iv) two β sheets that converge at the active site near the domain interface, and (v) two conserved loops (one in each lobe) that converge at the active site. In marked contrast to these conserved features shared by all protein kinases, recognition of the peptide by the catalytic subunit involves nonconserved amino acids, and the peptide binding sites extend over diverse and widely separated regions on the surface of the enzyme.

REFERENCES AND NOTES


1. E. H. Ficher and E. G. Krebs, *J. Biol. Chem.* **216**, 121 (1955).
2. E. W. Sutherland and W. D. Wosilait, *Nature* **175**, 169 (1955).
3. E. G. Krebs, *Biochem. Soc. Trans.* **13**, 813 (1985).
4. S. K. Hanks, A. M. Quinn, T. Hunter, *Science* **241**, 42 (1988).
5. E. G. Krebs, D. J. Graves, E. H. Fischer, *J. Biol. Chem.* **234**, 2867 (1959).
6. D. A. Walsh, J. P. Perkins, E. G. Krebs, *ibid.* **243**, 3763 (1968).
7. S. J. Beebe and J. D. Corbin, in *Cyclic Nucleotide–Dependent Protein Kinases*, E. G. Krebs and P. D. Boyer, Eds., *The Enzyme: Control by Phosphorylation*, part A (Academic Press, New York, 1986), vol. 17.
8. H. N. Bramson, E. T. Kaiser, A. S. Mildvan, *CRC Crit. Rev. Biochem.* **15**, 93 (1984).
9. S. S. Taylor, J. A. Buechler, Y. Yonemoto, *Annu. Rev. Biochem.* **59**, 971 (1990).
10. S. S. Taylor, J. A. Buechler, D. R. Knighton, *CRC Crit. Rev. Biochem.* (1990), pp. 1–42.
11. B. E. Kemp, D. J. Graves, E. Benjamini, E. G. Krebs, *J. Biol. Chem.* **252**, 4888 (1977).
12. Ö. Zetterqvist, U. Ragnarsson, L. Engström, *CRC Crit. Rev. Biochem.* (1990), p. 171.
13. L. W. Slice and S. S. Taylor, *J. Biol. Chem.* **264**, 20940 (1989).
14. H.-C. Cheng, S. M. van Patten, A. J. Smith, D. A. Walsh, *Biochem. J.* **231**, 655 (1986).
15. J. Zheng, D. R. Knighton, J. Parello, S. S. Taylor, J. M. Sowadski, *Methods Enzymol.*, in press.
16. D. R. Knighton, N.-h. Xuong, S. S. Taylor, J. M. Sowadski, *J. Mol. Biol.*, in press.
17. J.-h. Zheng *et al.*, *ibid.*, in press.
18. J. Parello, P. A. Timmins, J. M. Sowadski, S. S. Taylor, *ibid.*, in press.
19. B. W. Matthews, *ibid.* **33**, 491 (1968).
20. R. Hamlin *et al.*, *J. Appl. Cryst.* **14**, 85 (1981).
21. B.-C. Wang, *Methods Enzymol.* **115**, 90 (1985).
22. T. A. Jones, *J. Appl. Cryst.* **11**, 268 (1978).
23. C. Cambillau and E. Horjales, *J. Mol. Graphics* **5**, 174 (1987).
24. A. T. Brunger, J. Kuriyan, M. Karplus, *Science* **235**, 458 (1987).
25. W. A. Hendrickson and E. E. Lattman, *Acta Cryst.* **B26**, 136 (1970).
26. S. Remington, G. Wiegard, R. Huber, *J. Mol. Biol.* **158**, 111 (1982).
27. V. S. Allured, R. J. Collier, S. F. Carroll, D. B. McKay, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 1320 (1986).
28. D. E. Tronrud, L. F. Ten Eyck, B. W. Matthews, *Acta Cryst.* **A43**, 489 (1987).
29. D. R. Knighton *et al.*, *Science* **253**, 414 (1991).
30. J. Reed and V. Kinzel, *Biochemistry* **23**, 1357 (1984).
31. S. Shoji, L. H. Ericsson, K. A. Walsh, E. H. Fischer, K. Titani, *ibid.* **22**, 3702 (1983).
32. M. J. E. Sternberg and W. R. Taylor, *FEBS Lett.* **175**, 387 (1984).
33. S. S. Taylor *et al.*, *Cold Spring Harbor Symp. Quant. Biol.* **53**, 121 (1989).

34. D. C. Fry, S. A. Kuby, A. S. Mildvan, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 907 (1986).
35. S. A. Benner and D. Gerloff, *Adv. Enzyme Reg.* **31**, 121 (1991).
36. W. Yonemoto, M. McGlone, S. S. Taylor, in preparation.
37. M. G. Rossmann, D. Moras, K. Olsen, *Nature* **250**, 194 (1974).
38. M. J. Zoller, N. C. Nelson, S. S. Taylor, *J. Biol. Chem.* **256**, 10387 (1981).
39. D. Bhatnager, F. T. Hartl, R. J. Roskoski, R. A. Lessor, N. J. Leonard, *Biochemistry* **23**, 4350 (1984).
40. J. A. Buechler and S. S. Taylor, *Biochemistry* **27**, 7356 (1988).
41. ———, *ibid.* **28**, 2065 (1988).
42. J. A. Buechler, T. A. Vedvick, S. S. Taylor, *ibid.*, p. 3018.
43. H. N. Bramson *et al.*, *J. Biol. Chem.* **257**, 10575 (1982).
44. N. Nelson and S. S. Taylor, *ibid.* **258**, 10981 (1983).
45. J. S. Jimenez, A. Kupfer, V. Gani, S. Shalteil, *Biochemistry* **21**, 1623 (1982).
46. E. A. First and S. S. Taylor, *ibid.* **28**, 10981 (1989).
47. R. K. Wierenga, P. Terpstra, W. G. J. Hol, *J. Mol. Biol.* **187**, 101 (1986).
48. C.-I. Brändén, *Q. Rev. Biophys.* **13**, 317 (1980).
49. M. Saraste, P. R. Sibbald, A. Wittinghofer, *Trends Biochem. Sci.* **15**, 430 (1990).
50. W. G. J. Hol, *Prog. Biophys. Mol. Biol.* **45**, 149 (1985).
51. G. E. Schulz *et al.*, *Eur. J. Biochem.* **161**, 127 (1986).
52. C. M. Anderson, F. H. Zucker, T. A. Steitz, *Science* **204**, 375 (1979).
53. K. M. Flaherty, C. DeLuca-Flaherty, D. B. McKay, *Nature* **346**, 623 (1990).
54. W. Kabsch, H. G. Mannherz, D. Suck, E. F. Pai, K. C. Holmes, *ibid.* **347**, 37 (1990).
55. M.-Y. Yoon and P. F. Cook, *Biochemistry* **26**, 4118 (1987).
56. M.-f. Ho, H. N. Bramson, D. E. Hansen, J. R. Knowles, E. T. Kaiser, *J. Am. Chem. Soc.* **110**, 2680 (1988).
57. J. P. Priestle, *J. Appl. Cryst.* **21**, 572 (1988).
58. T. J. Callahan, W. B. Gleason, T. P. Lybrand, Molecular Simulation Laboratory, University of Minnesota, ©1990.
59. N.-h. Xuong, D. Sullivan, C. Nielsen, R. Hamlin, *Acta Cryst.* **B41**, 267 (1985).
60. A. J. Howard, C. Nielson, N.-h. Xuong, *Methods Enzymol.* **14**, 452 (1985).
61. T. C. Terwilliger and D. Eisenberg, *Acta Cryst.* **A39**, 813 (1983).
62. G. A. Sim, *ibid.* **13**, 511 (1960).
63. CDC7, *Saccharomyces cerevisiae* cell division cycle 7 gene product; K1N1, putative *S. cerevisiae* protein kinase gene product; PKC-γ, protein kinase C, γ form; c-mos, cellular homolog of oncogene product from Moloney murine sarcoma virus; PDGFR, platelet-derived growth factor receptor; $ran^{+1}$, *Schizosaccharomyces pombe* "meiotic bypass" mutant wild-type gene product: HSVK, herpes simplex virus–US3 gene product; 7less, *Drosophila sevenless* gene product.
64. Supported by the Lucille P. Markey Foundation without which this work could not have been completed; by grants from NIH (S.S.T. and N-h.X.), the American Cancer Society (J.M.S. and S.S.T.), NSF (S.S.T. and L.T.E.), NIH training grants T32CA09523 and T32DK07233 (D.R.K.), and the University of California (J.M.S.). We thank the following individuals and resources for their contributions: S. Bell, M. Montella, and G. Hasegawa for preparation of manuscript; the NIH National Research Resource at UCSD(RR01644) and staff members Chris Nielsen and Don Sullivan for data collection facilities; the San Diego Supercomputer Center for use of the Advanced Scientific Visualization Laboratory and the Cray Y-MP/864; and J. Buechler, W. Yonemoto, and B. Driscoll for discussions and review of the manuscript. Atomic coordinates have been deposited in the Brookhaven Protein Data Bank.

3 May 1990; accepted 21 June 1991

# Structure of a Peptide Inhibitor Bound to the Catalytic Subunit of Cyclic Adenosine Monophosphate–Dependent Protein Kinase

DANIEL R. KNIGHTON, JIANHUA ZHENG, LYNN F. TEN EYCK, NGUYEN-HUU XUONG, SUSAN S. TAYLOR, JANUSZ M. SOWADSKI*

**The structure of a 20–amino acid peptide inhibitor bound to the catalytic subunit of cyclic AMP–dependent protein kinase, and its interactions with the enzyme, are described. The x-ray crystal structure of the complex is the basis of the analysis. The peptide inhibitor, derived from a naturally occurring heat-stable protein kinase inhibitor, contains an amphipathic helix that is followed by a turn and an extended conformation. The extended region occupies the cleft between the two lobes of the enzyme and contains a five-residue consensus recognition sequence common to all substrates and peptide inhibitors of the catalytic subunit. The helical portion of the peptide binds to a hydrophobic groove and conveys high affinity binding. Loops from both domains converge at the active site and contribute to a network of conserved residues at the sites of magnesium adenosine triphosphate binding and catalysis. Amino acids associated with peptide recognition, nonconserved, extend over a large surface area.**

THE PROBLEM OF HOW, DURING PROTEIN PHOSPHORYLation, a targeted protein substrate is recognized by a specific protein kinase has been particularly elusive because the determinants for peptide recognition are widely dispersed and often distant from the actual site of phosphotransfer (*1*). Information about peptide binding sites comes from several different directions. Substrate analogs have provided insights into the specific features of a given substrate that are important for recognition. Chemical approaches, such as affinity labeling and group specific labeling have identified regions and specific residues that are in close proximity to substrates. A true understanding of peptide recognition, however, requires not only a high-resolution crystal structure of the enzyme but also co-crystals containing the bound peptide. The structure of the catalytic (C) subunit presented in the previous paper (*2*) contains a bound inhibitor peptide, a fragment of the naturally occurring heat stable protein kinase inhibitor (PKI) (*3*). This peptide includes the consensus features common to all peptide substrates and inhibitors of 3′,5′-adenosine monophosphate (cyclic

D. R. Knighton, J. Zheng, and S. S. Taylor are in the Department of Chemistry, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093–0654; L. F. Ten Eyck is in the Department of Chemistry, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093–0654, and the San Diego Supercomputer Center, P.O. Box 85608, San Diego, CA 92186–9784. N-h. Xuong is in the Departments of Chemistry, Physics, and Biology, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093–0319. J. M. Sowadski is in the Departments of Medicine and Biology, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093–0654.

*To whom correspondence should be addressed.