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.

AMP)–dependent protein kinase (cAPK). In addition, it contains other features that convey specific high affinity binding.

The general requirements for peptide recognition by the C subunit, characterized originally with the use of synthetic peptides (*4–6*), revealed a general consensus sequence that includes two basic residues, typically arginines, followed by one intervening small residue preceding the site of phosphotransfer. The phosphate acceptor site, either Ser or Thr, is followed immediately by a large hydrophobic residue. The typical peptide substrate for this kinase is based on the phosphorylation site in pyruvate kinase (Table 1). Replacement of the Ser with an Ala generates an inhibitor peptide with poor affinity. In contrast is the inhibitor peptide, PKI(5–24), derived from the thermo-stable protein kinase inhibitor (*3*). This peptide (Table 1) encompasses the above-described consensus sequences and contains additional features that convey high affinity binding and that have been identified by comparison with peptide analogs (*7, 8*). A general structure of the peptide in solution was deduced from circular dichroism (CD) and nuclear magnetic resonance (NMR) spectroscopy data (*8–10*). This peptide, PKI(5–24), was co-crystallized with the C subunit, and the structure of that peptide as well as its interaction with the protein are discussed below.

The nomenclature used for the peptide designates the phosphorylation site of

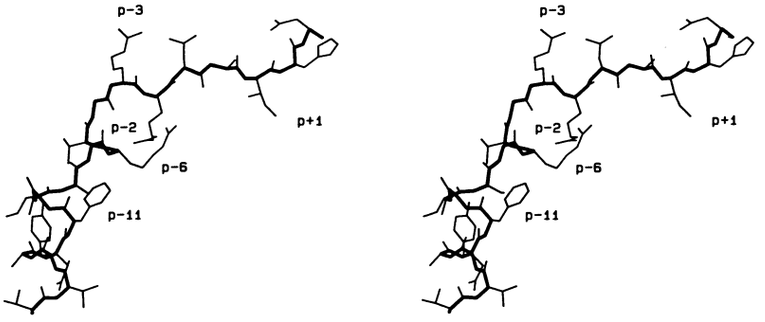



**Fig. 1.** Bound structure of PKI(5–24). Backbone C and N atoms are shown in bold. Residues particularly important for binding are labeled according to the nomenclature of Table 1. Weak electron density makes ambiguous the side chain positions of the two COOH-terminal residues (His-Asp).

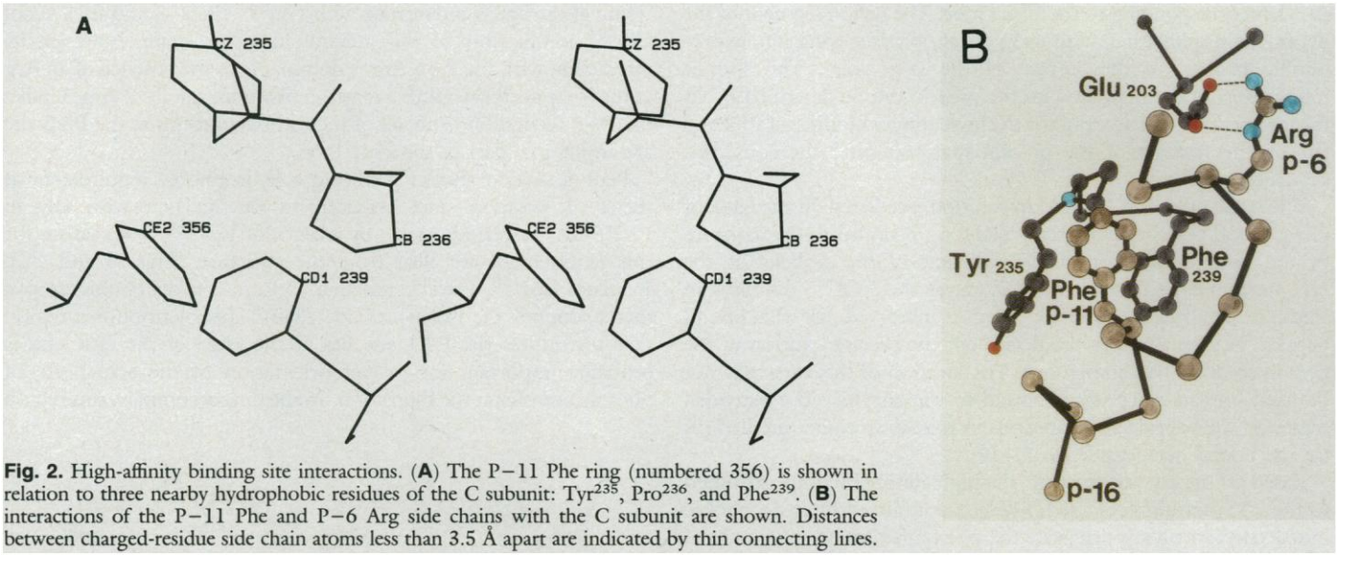
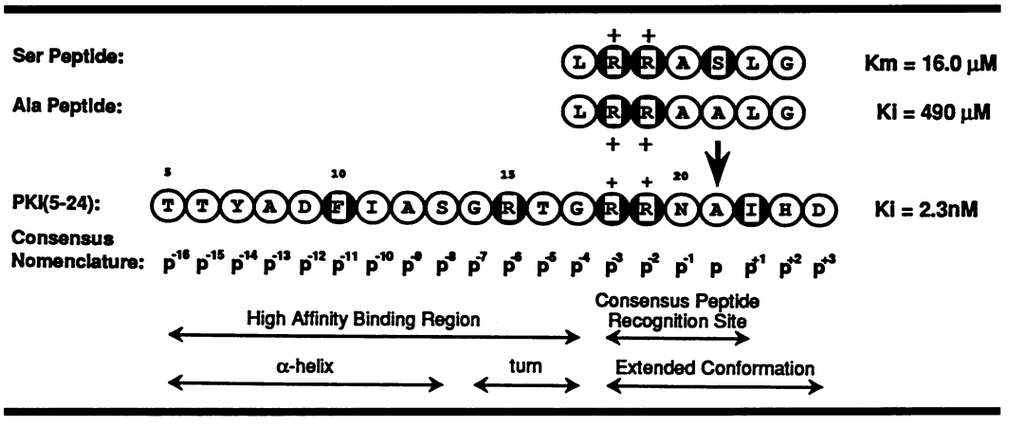



**Fig. 2.** High-affinity binding site interactions. (**A**) The P−11 Phe ring (numbered 356) is shown in relation to three nearby hydrophobic residues of the C subunit: $Tyr^{235}$, $Pro^{236}$, and $Phe^{239}$. (**B**) The interactions of the P−11 Phe and P−6 Arg side chains with the C subunit are shown. Distances between charged-residue side chain atoms less than 3.5 Å apart are indicated by thin connecting lines.

**Table 1.** Substrate and inhibitor peptides of cyclic AMP–dependent protein kinase. The Ser peptide is based on the in vivo phosphorylation site in pyruvate kinase (*39*). PKI(5–24) is a segment of the heat-stable protein kinase inhibitor (*3*). Residues important for peptide recognition are shaded and were identified with the use of synthetic peptide analogs of the Ser peptide (*4, 40*) and analogs of PKI(5–24) and other related peptides from the inhibitor domain of PKI (*7, 15*). Several approaches indicate that the affinity of PKI for C is enhanced in the presence of MgATP (*41, 42*), but no actual $K_d$'s for PKI and PKI(5–24) alone are available. While MgATP does enhance the affinity of PKI for C, our results show that in the absence of MgATP, the $K_d$ is still less than 50 nM and certainly in the high affinity range. The difference Fourier map of the binary and ternary complexes indicates that there are no major changes in the conformation for the bound peptide.



pseudophosphorylation site residue as P. In the case of substrates, P is Ser or Thr; in the case of PKI(5–24), P is Ala. The residues flanking this site are designated as P+1, P−1, and so forth, as indicated in Table 1. This nomenclature provides a common frame of reference for all peptide substrates and inhibitors and can be applied to other protein kinases.

**Conformation of the bound inhibitor peptide**. In the conformation of bound PKI(5–24) (Fig. 1). The amino terminus extending from the P−16 Thr through the P−8 Ala forms an amphipathic α helix. This helix is followed by a turn flanked by glycines at the P−7 and P−4 positions. The glycines may be important for accommodating the turn or for providing flexibility to facilitate binding of the Arg that follows each Gly. The remainder of the peptide is in an extended conformation, and the density corresponding to the region at the carboxyl terminus, the P+3 Asp and P+2 His, is poorly defined.

The C subunit itself consists of two lobes—a smaller lobe, associated primarily with magnesium adenosine triphosphate (MgATP) binding, and a larger lobe (*2*). Nearly all of the features necessary for peptide recognition are found within the larger lobe, although the specific residues involved are widely dispersed both in the linear sequence and on the surface of the enzyme. The extended portion of the peptide, which includes the consensus region for recognition of all substrates and inhibitors, lies along the surface of the cleft corresponding to the larger lobe. The helical segment of the peptide is amphipathic, with its hydrophobic face lying in a hydrophobic pocket on the surface of the large lobe. The specific interactions of the peptide with the protein can be described by (i) the interactions that account for the high affinity binding of PKI and (ii) by the features of the protein that recognize the consensus sequence.

The structure of PKI(5–24) in solution, predicted on the basis of CD (*9*), and NMR spectroscopy data (*10*), is similar to the structure of the bound inhibitor. The prediction of the α helix at the $NH_2$-terminus and an extended chain at the COOH-terminus in solution parallels, in general, what is observed for the bound peptide. A turn also was predicted, but the precise location of the turn in solution was ambiguous. The location of this turn may not be fixed until the peptide is bound to the enzyme. The extended region of the peptide also conforms in general to earlier predictions for the bound Ser peptide (*6, 11–14*).

Based on the crystal structure, the high affinity binding attributed to the $NH_2$-terminus of PKI(5–24) is dominated by hydrophobic interactions involving primarily the phenylalanine side chain at the P−11 position. The importance of a large hydrophobic group at this position was recognized from analog studies (*8*). For example, replacement of this Phe with an Ala caused a 100-fold increase in $K_i$ (7), while replacement with 1′-napthylalanine, a residue that is considerably larger and more hydrophobic than Phe, actually decreased the $K_i$ by a factor of 4 (*15*). A hydrophobic pocket on the surface of the C subunit nicely complements the hydrophobic face of the helix in the inhibitor peptide (Fig. 2). This hydrophobic pocket is lined by residues 235 through 239, Tyr-Pro-Pro-Phe-Phe, with the phenyl ring in the inhibitor peptide sandwiched between the side chains of $Tyr^{235}$ and $Phe^{239}$. Based on the structure and consistent with analog studies, the Tyr at the P−14 position is not as important for this hydrophobic interaction although phosphorylation of this Tyr does increase the $K_i$ by a factor of 10 (*16*). In addition to the hydrophobic interactions associated with the helix, the orientation of the high affinity binding region of PKI(5–24) is fixed by ionic contacts involving the P−6 Arg. Two nitrogens in the guanidine side chain of this Arg undergo ion pairing with the two oxygens of the carboxyl group of $Glu^{203}$ (Fig. 2).

Certain kinds of interactions contribute to recognition of the common consensus region of the inhibitor peptide (Fig. 3). Electrostatic interactions dominate the portion of the peptide proximal to the site of phosphotransfer, while hydrophobic interactions dominate the COOH-terminal region distal to the phosphotransfer site. Once again, the interactions revealed by the structure are totally consistent with previous chemical and genetic predictions.

Two important, although not absolute, requirements for peptide recognition by cAPK are basic residues at the P−3 and P−2 positions. Replacing either Arg in the Ser peptide substrate leads to a $K_m$ that is 16 to 400 times higher, even when the Arg is replaced with a Lys (*4*). The environment flanking the P−3 and P−2 arginines in the structure explains these results since each Arg interacts with more than one carboxyl side chain.

At the anionic P−3 site (Fig. 4A), we find that the guanidinium nitrogens of the P−3 Arg are approximately 3 Å from the carboxyl side chains of $Glu^{127}$ and $Glu^{331}$. The tip of $Asp^{329}$ is approximately 5 Å away. Thus, the position of the guanidinium moiety is fixed. In addition, the side chain of this P−3 Arg comes close to the backbone carbonyl of $Thr^{51}$ in the glycine-rich loop and to the hydroxyls of the ribose ring. The side chain of $Glu^{333}$ lies close to $Lys^{47}$ in β strand 1, and the side chain $Glu^{334}$ is approximately 3 Å from the hydroxyl group of $Thr^{48}$.

The P−2 Arg is also fixed by more than one carboxyl group with the ε-nitrogen forming an ion pair with $Glu^{170}$, and one of the terminal nitrogens interacting with $Glu^{230}$ (Figs. 3 and 4B). Also, $Glu^{203}$ comes close to this guanidinium side chain; however, its interaction with the P−6 Arg is dominant. In the absence of an Arg at the P−6 position, $Glu^{230}$ may ion pair with the P−2 Arg. Unlike the P−3 recognition site, all of the carboxyl groups at the P−2 site are an integral part of the large lobe.

Peptide analog studies predicted a hydrophobic requirement at the P+1 position since replacement of the Leu with Gly in PKI(5–24) increased the $K_m$ by a factor of 150 (7). The reasons for this requirement are clear from the structure (Figs. 3 and 4C). Residues $Leu^{198}$, $Pro^{202}$, and $Leu^{205}$ form a hydrophobic groove that surrounds the isoleucine side chain. This hydrophobic region that constitutes the P+1 site lies at the edge of the cleft and is probably important for proper orientation of the actual site of phosphotransfer at the P position. In the binary complex, this region

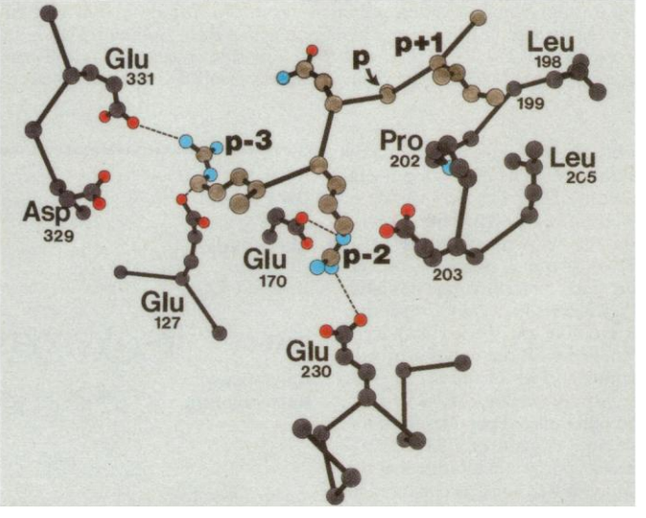



**Fig. 3.** Consensus recognition site. The interactions of the P−3 and P−2 Arg residues and the P+1 Ile residue with C subunit residues are shown. Lines are drawn between charged-residue side chain atoms <3.5 Å apart. The P+1 Ile side chain projects into the hydrophobic area formed by $Leu^{198}$, $Pro^{202}$, and $Leu^{205}$.

begins to align in an antiparallel β-like configuration with the carbonyl of the P+1 Ile coming less than 4 Å from the backbone amide of Gly[200] and the carbonyl of Gly[200] coming within approximately 3 Å of the backbone amide of this Ile at the P+1 position (Fig. 5B). Substitution of a Pro for Leu at the P+1 position in the Ser peptide (Table 1) yields an extremely poor substrate, consistent with the strict steric requirements observed (*17*). Nevertheless, a depsipeptide analog of this peptide lacking an amide proton at this P+1 site is still a good substrate for the C subunit (*13*).

The addition of a hydroxyl group would place the side chain of the residue at the P position close enough for a direct transfer of the γ-phosphate for MgATP (Fig. 5A). The side chain of the P−1 Asn also interacts with the glycine-rich loop as seen in Fig. 5A.

Thr[197], one of the two stable phosphorylation sites in this enzyme (*18*), also flanks the P+1 site. Multiple electrostatic interactions (Fig. 5C) hold this phosphate in place and account for its resistance to removal by phosphatases. Fixing this phosphate moiety contributes conformational stability, not only to Thr[197] but also to the adjacent hydrophobic residues important for recognition at the P+1 site and for the proper orientation of the site of phosphotransfer. Based on the crystal structure, this anionic group appears to be important for the final correct assembly of the structure.

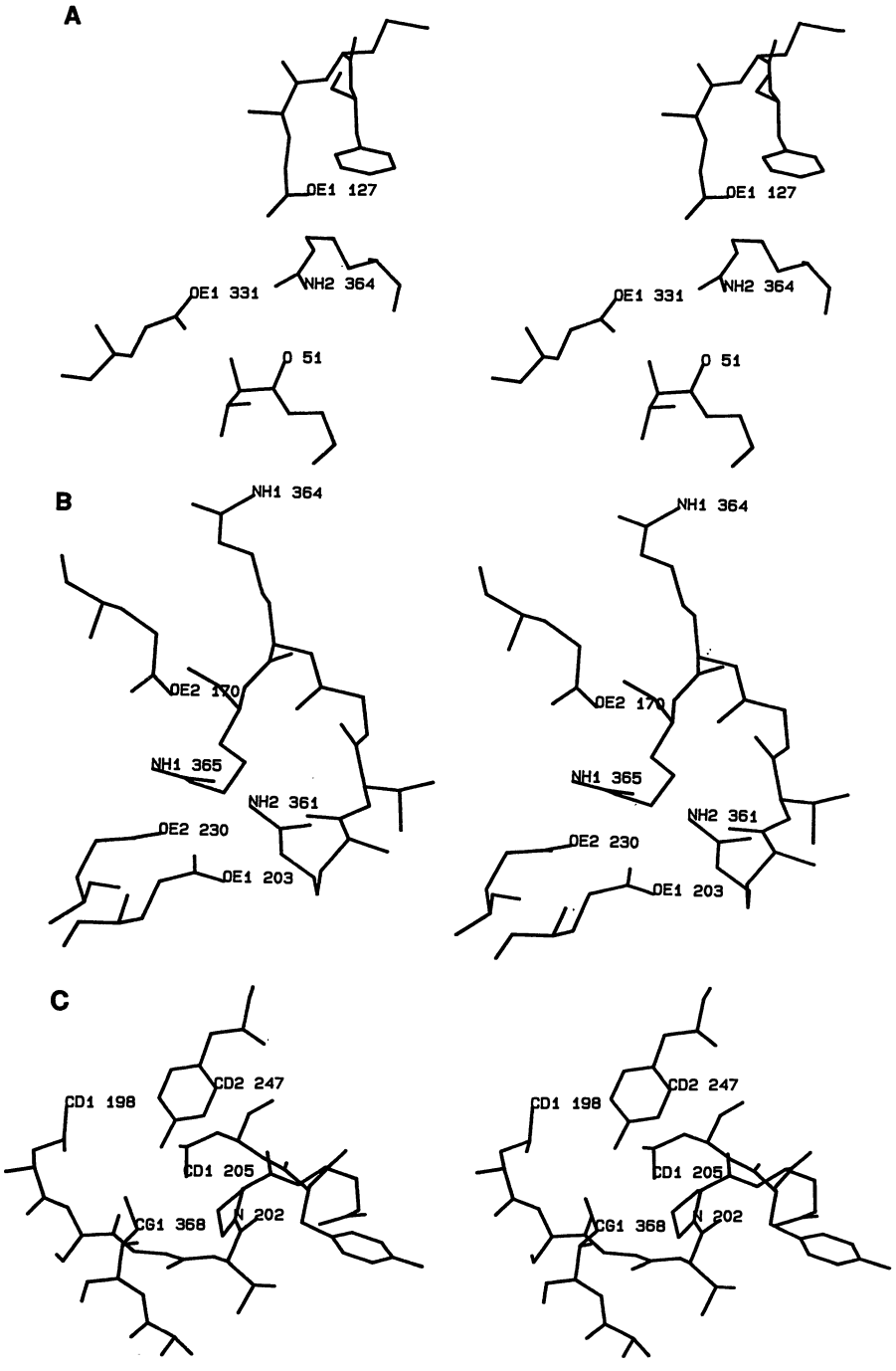



**Fig. 4.** Consensus recognition site binding interactions. Residue numbers 361, 364, 365, and 368 correspond, respectively, to PKI(5–24) P−6, P−3, P−2, and P+1 residues. (**A**) The P−3 Arg side chain tip is shown in proximity to Thr[51] carbonyl in the glycine-rich loop, and Glu[127] and Glu[331] side chain carboxylates of the domain-linking region. (**B**) The P−2 Arg side chain is shown in proximity to side chain carboxylates of Glu[170] of the catalytic loop and Glu[230]; the P−6 Arg side chain is shown near side chain carboxylate of Glu[203]. (**C**) The P+1 Ile side chain is shown projecting into hydrophobic pocket consisting of residues Leu[198], Pro[202], and Leu[205].

Amino acid side chains that contribute to peptide recognition have been identified chemically in several ways. Differential labeling with a water-soluble carbodiimide, 1-ethyl-3(3-dimethylaminopropyl)carbodiimide (EDC), targeted two sites containing solvent-accessible carboxyl groups that were accessible in the free C subunit but protected in the presence of substrate (*19*). Glu[170] was very reactive in the absence of peptide, but fully protected in the presence of peptide. The other region was the cluster of carboxyl groups near the COOH-terminus, Asp[328]-Asp-Tyr-Glu-Glu-Glu-Glu[334]. The Glu[170] interacts with the P−2 while the cluster of carboxyl groups flanks the P−3 site (Fig. 4).

A genetic approach also was used to identify carboxyl groups that contribute to peptide recognition in the yeast C subunit, TPK1 (*20*). This method, "charged-to-alanine" mutagenesis, depends on the selective replacement of all charged residues with Ala and then searching among the mutant proteins for those replacements that influenced peptide recognition. Using this approach, Gibbs and Zoller demonstrated that Glu[171], the yeast homolog of Glu[127] in the mouse C subunit, participates in recognition of the P−3 Arg and that Glu[214] and Glu[274], homologous to Glu[170] and Glu[230], are determinants for recognition of the P−2 arginine.

Several different methods indicated that Cys[199] is close to the peptide recognition site and to the γ-phosphate subsite of ATP (*6, 21–23*), and the structure confirms this prediction. In the binary complex, Cys[199] does not appear to participate in peptide binding other than to contribute to the general hydrophobic environment around the P+1 site.

**Conserved and variable sites in protein kinases.** Substrate-induced conformational changes are associated with peptide binding to the C subunit. Global changes in conformation, first observed with CD, showed both a loss of α-helical content and an

increase in β structure after peptide binding (*24*). Small angle neutron scattering demonstrated a reduction of the radius of gyration ($R_g$) after substrate binding and established, furthermore, that the inhibitor peptide alone, but not MgATP, was sufficient to cause the reduction in $R_g$ (*25*). The substrate-induced reduction in $R_g$ indicates that the apoenzyme corresponds to an open configuration of the protein while the binary and ternary complexes represent a closed configuration.

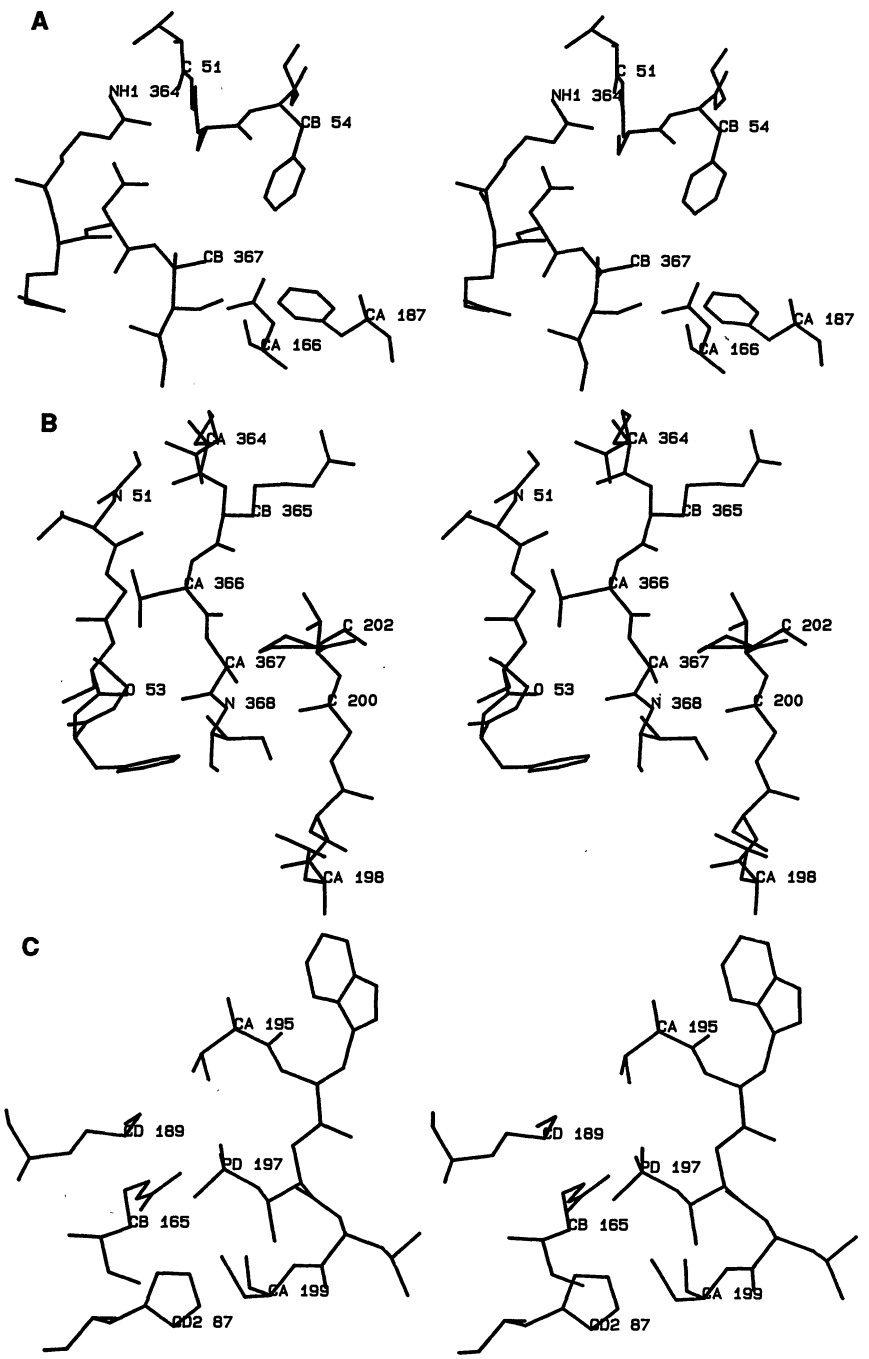



**Fig. 5.** Catalytic site area. Residue numbers 364 and 367 correspond to the P−3 Arg and the P Ala. (**A**) The site of catalysis is shown with the possible catalytic base side chain of $Asp^{166}$ near the Cβ of the P Ala. The $Thr^{51}$ of the glycine-rich loop is shown near the P−3 Arg side chain. The hydrophobic side chains of residues $Phe^{54}$ (at the loop apex) and $Phe^{187}$ are shown near the site of phosphotransfer. These residues may help exclude water from the active site. (**B**) The consensus recognition site residues Arg-Arg-Asn-Ala-Ile are shown with the glycine-rich phosphate anchor loop to the left and residues 198 to 202 to the right. The carbonyl of $Gly^{200}$ can be seen pointing to the amide N of the P+1 Ile. (**C**) The phosphate of $Thr^{197}$ is shown buried among side chains of cationic and hydrogen-bonding residues ($His^{87}$, $Arg^{165}$, $Lys^{189}$, and $Thr^{195}$). The $Cys^{199}$ is also shown nearby.

An extension of the CD studies established that recognition of the peptide by the C subunit was a multistep process (*26*), and this sequential pathway was reinforced by kinetic studies (*27*). The initial step, associated with a loss in α-helical structure, was induced by both the substrate and inhibitor heptapeptides (Table 1). The second step, presumably corresponding to the final orientation of the peptide into the correct position at the active site, was associated with an increase in β structure and could only be accomplished with the substrate peptide, not by the Ala peptide inhibitor (*26*). This finding is strengthened by observations that the basal adenosine triphosphatase (ATPase) activity of the enzyme is not blocked by the Ala peptide inhibitor (*28*). The increase in β structure is probably due, in part, to the P+1 region of the peptide interacting with the protein. Understanding these substrate-induced conformational changes will eventually require a detailed comparison of the apoenzyme structure with binary and ternary complexes containing inhibitors and substrate peptides both in the presence and absence of MgATP.

The recognition of a protein substrate by the C subunit is not unlike the recognition of a protein antigen by the variable domain of an immunoglobulin. In both cases recognition involves a relatively large surface area comprised of largely charged and hydrophilic residues. Also, the binding sites of both structures are dominated by interfacing β sheets surrounded by loops that participate in recognition of the protein. Although the C subunit also has helical regions, both β sheets converge at the active site, and the loops play the dominant role in peptide recognition and catalysis. In the case of the protein kinases, two of the loops are essential for catalysis and are highly conserved, unlike the immunoglobulins, whose primary function is binding and not catalysis.

The two essential conserved loops that assemble at the site of catalysis in the C subunit (Figs. 6 and 7) are the glycine-rich loop in the small lobe and the catalytic loop in the large lobe. Both lie on the surface that lines the cleft between the two lobes. The glycine-rich loop serves as an anchor for the phosphates of MgATP, whereas the catalytic loop is essential for peptide binding and catalysis. Seven of the nine invariant amino acids conserved in all protein kinases are located here (Fig. 6), either in the loops themselves or connecting directly with loop residues. The single invariant glycine, $Gly^{52}$, lies in the phosphate anchoring loop. The proposed catalytic base, $Asp^{166}$, as well as $Asn^{171}$, are in the catalytic loop. It is remarkable how thoroughly interconnected this region is with multiple ion pairs providing a finely tuned scaffolding for communication at the active site.

The three invariant residues in the small lobe participate in nucleotide binding. Un-

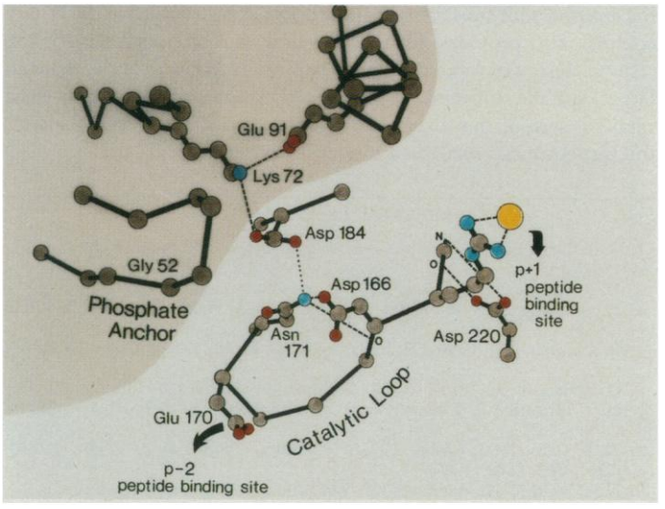



**Fig. 6.** Relations of invariant amino acids at the active site. Key features of the active site of the C subunit are shown. Nine of the amino acids that are nearly invariant in all protein kinases are indicated. Gly[186], another invariant residue, is not shown. The Cα's are colored gray, the oxygens red, and the nitrogens blue. The position of the phosphorylation site at Thr[197] is indicated in yellow. The portion of the active site associated with the small lobe is shaded in gray and includes three of the invariant amino acids, Gly[52], Lys[72], and Glu[91]. The remaining six are located in the large lobe. Residues close enough for hydrogen bonding or ion pairing are indicated by a dashed line while residues within 4 to 5 Å of one another are connected by a dotted line.

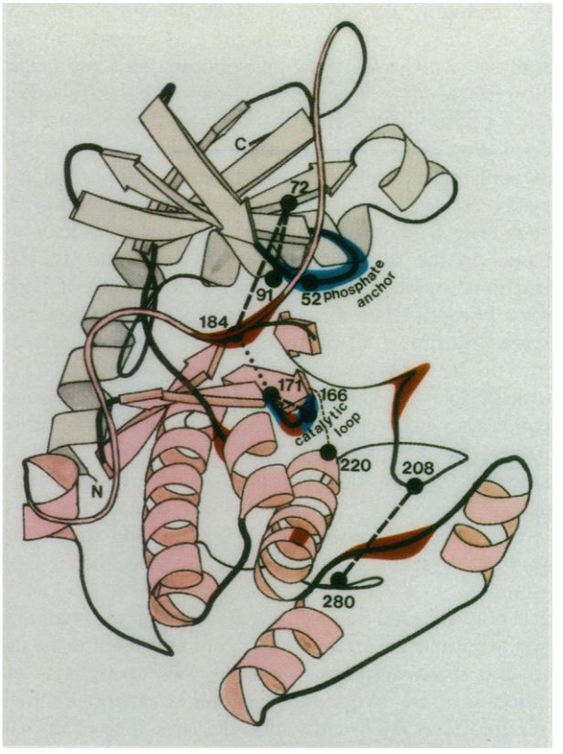



**Fig. 7.** Conserved and variable regions of the catalytic (C) subunit. The ribbon diagram shows the folding of the C subunit. The coloring of the chain depicts the $NH_2$-terminus through residue 123 in gray and the COOH-terminus in blue. Conserved regions include two loops—the glycine-rich loop and the catalytic loop—and are indicated in blue. The variable peptide binding sites are shown in red. Invariant amino acids Gly[52], Lys[72], Glu[91], Asp[166], Asn[171], Asp[184], Glu[208], Asp[220], and Arg[280] are indicated by a large dot and are numbered. Dashed lines indicate residues that are close enough to pair, while the dotted line extends from Arg[165] to the Thr[197].

like Gly[52], which is part of a flexible loop, both Lys[72] and Glu[91] are anchored to defined parts of the secondary structure; for example, Lys[72] to β strand 3 and Glu[91] to the helix C. Much chemical evidence is consistent with these residues playing a key role in recognition of the phosphates of MgATP (*23, 29*), and the difference Fourier (*2*) shows the phosphate density near these residues, with the presumed γ-phosphate density close enough to the P Ala Cβ for phosphotransfer were it a Ser(Thr).

In the catalytic loop the two invariant residues, Asp[166] and Asn[171], interact with each other. Not only are their side chains close, but, more important, the nitrogen in the amide side chain of Asn[171] is less than 3 Å from the backbone carbonyl of Asp[166]. One additional nearly invariant residue, Asp[220], contributes directly to stabilization of the catalytic loop. The two oxygens of this carboxylate come within hydrogen bonding distance of the backbone carbonyl and amide of residue 164, which immediately precedes the loop. This interaction of the catalytic loop with a conserved residue that lies deep within the large lobe fixes the loop from one side while peptide binding and interactions with the small lobe fix it from the opposite direction. The consensus region of the peptide is sandwiched between the P+1 site on one side and the glycine-rich loop on the other side (Fig. 5, A and B).

Of all the invariant residues, Asp[184] is the only one that appears to communicate with both the small lobe and the large lobe. In the binary complex, it is most closely associated with Lys[72], but it is also only 4 to 5 Å from Asn[171] and Asp[166] in the catalytic loop. Although not shown in Fig. 6, Asp[184] is itself part of a tight turn with the carboxylate located within hydrogen-bonding distance of the backbone amide of Gly[186], another invariant residue. This entire segment, Asp[184]-Phe-Gly, is highly conserved in all protein kinases (*30*), and this hydrogen bonding to stabilize the turn is probably conserved as well. Certainly Asp[184] has the potential to shuttle between the two conserved loops, and it is anticipated that the contacts of Asp[184] will differ somewhat in both the apoenzyme and in the ternary complex containing bound MgATP as well as peptide. A detailed description of Asp[184] and its interaction with MgATP will be discussed when the ternary complex is refined. If Asp[184] participates in the chelation of $Mg^{2+}$ (*2*), then this charge will be sequestered from the immediate environment of the catalytic loop. Other residues close to the conserved residues in the catalytic loop in the binary complex are Tyr[164], whose side chain is less than 3 Å from the side chain nitrogen of Asn[171] and Lys[168], whose side chain comes close to the carboxylate of Asp[166]. Either Tyr or His, another good hydrogen-bonding residue, is always found at position 164, so this contact may be conserved. Any significant change in the position of Asp[184] will necessarily change the environment of the catalytic loop. Asp[184], as well as Asn[171] and Asp[166], also were identified as a sequence motif associated with many phosphotransferases, so this may represent a common mechanism (*31*).

The versatility and importance of the catalytic loop is highlighted not only by the conserved networking of essential amino acids at the active site, but also by the special ways in which this conserved network communicates with the variable residues that compose the peptide binding sites. This communication specifically involves loop residues that are not highly conserved. For example, Glu[170] contributes directly to the anionic P−2 site. In contrast, Thr[201] in the P+1 site comes very close to the side chain of Asp[166]. These two particular regions of contact involving the peptide binding site and the catalytic loop, Lys[168]-Pro-Glu and Thr[201]-Pro-Glu-Tyr-Leu-Ala-Pro-Glu, contain sequences that differ characteristically between the kinases that transfer phosphate to the Ser(Thr) and those that transfer phosphate to tyrosine (*32*).

$Arg^{165}$ is highly conserved in most protein kinases, and it connects in a characteristic way with the P+1 peptide binding site. Specifically, it points toward the phosphothreonine and helps to fix that phosphate so that the hydrophobic groove that follows it and serves as a pocket for the side chain of the P+1 residue is firmly positioned (Fig. 5C). This is an autophosphorylation site (*33*), and it is the only phosphorylation site in the C subunit that could conceivably result from an intramolecular autophosphorylation. This phosphate is very resistant to removal by phosphatases (*18*) and, based on the structure, appears to contribute to the final conformational stability of the enzyme. It should be emphasized as well that a phosphorylation site in this region of the protein is not a conserved feature of all protein kinases. Some kinases, such as $pp60^{c\text{-}src}$, do have an autophosphorylation site nearby (*34*), but many others do not. Whether the catalytic loop communicates in particular ways with other autophosphorylation sites in other protein kinases remains to be established.

The two invariant residues that are most distant from the active site are $Asp^{208}$ and $Arg^{280}$. These residues constitute a conserved ion pair that lies just beneath the P+1 site and they appear to stabilize a very hydrophobic region that buttresses the P+1 peptide binding site.

Several points should be emphasized regarding the recognition of a peptide or protein substrate by the C subunit. First is the number of sites and their diversity. Some of these peripheral peptide recognition sites are hydrophilic and highly charged; others are hydrophobic. Most are found within the large lobe of the catalytic core shared by all protein kinases, but some can also lie outside of this boundary (Fig. 7). A second observation is that the requirements for recognition at the consensus site are not absolute. A comparison of in vivo phosphorylation sites reveals that the actual residues at each site vary somewhat as does the spacing between the positively charged side chains and the site of phosphotransfer (*5*). Thus, even in the consensus region, some variability can be tolerated. A third point is the potential for variability in recognition of different inhibitor, and presumably substrate, proteins that bind with a high affinity to the C subunit. Most of the features essential for the high affinity recognition of PKI are apparent from this structure of the binary complex. The regulatory (R) subunit, however, also binds to the C subunit with a subnanomolar affinity in the absence of cyclic AMP (*35*). The consensus region, P−3 through P+1, is shared by both molecules. However, the R subunit, cleaved at the P−5 position, still retains its high affinity binding for the C subunit (*36, 37*). In addition, the P−16 to P+1 region of the $R^I$ subunit is Pro-Pro-Pro-Pro-Asn-Pro-Val-Val-Lys-Gly-Arg-Arg-*Arg-Arg*-Gly-*Ala-Ile* (*38*), and this certainly cannot conform to the helical motif that dominates the corresponding region of PKI(5–24). Hence, an amphipathic helix is not required for the high affinity binding of R. Instead, the residues that contribute to the high affinity binding of the R subunit, specifically, must lie beyond the P−3 position and may complement a different portion of the surface of the C subunit. This variability presumably can also extend to protein substrates where the C subunit may recognize unique sequences that lie outside the consensus site.

This binary complex allows us for the first time to begin to appreciate the complexity and sophistication of the process by which a protein kinase recognizes its protein substrate. While peptide analogs provide important clues, the diversity of the peptide binding sites and their dispersion over such a wide area on the enzyme surface makes it imperative to have structural data on complexes of the enzyme with substrates or inhibitors. This structure of the binary complex also provides, for the first time, a molecular basis for the rational design of inhibitors, both peptide and non-peptide, that can target specific protein kinases. Furthermore, because the basic catalytic core of this enzyme is so conserved in all protein kinases, this structure can serve as a model for other protein kinases.

## REFERENCES AND NOTES


1. B. E. Kemp, Ed., *Peptides and Protein Phosphorylation* (CRC Press, Boca Raton, FL, 1990).
2. D. R. Knighton *et al.*, *Science* **253**, 407 (1991).
3. H.-C. Cheng, S. M. van Patten, A. J. Smith, D. A. Walsh, *Biochem. J.* **231**, 655 (1986).
4. B. E. Kemp, D. J. Graves, E. Benjamini, E. G. Krebs, *J. Biol. Chem.* **252**, 4888 (1977).
5. Ö. Zetterqvist, U. Ragnarsson, L. Engström, in (*1*), pp. 171–188.
6. H. N. Bramson, E. T. Kaiser, A. S. Mildvan, *CRC Crit. Rev. Biochem.* **15**, 93 (1984).
7. D. B. Glass, H.-C. Cheng, L. Mende-Mueller, J. Reed, D. A. Walsh, *J. Biol. Chem.* **262**, 8802 (1989).
8. D. A. Walsh, K. L. Angelos, S. M. Van Patten, D. B. Glass, L. P. Garetto, in (*1*), pp. 43–84.
9. J. Reed, V. Kinzel, H.-C. Cheng, D. A. Walsh, *Biochemistry* **26**, 7641 (1987).
10. J. Reed *et al.*, *Biochem. J.* **264**, 371 (1989).
11. J. Granot, A. S. Mildvan, H. N. Bramson, N. Thomas, E. T. Kaiser, *Biochemistry* **20**, 602 (1981).
12. P. R. Rosevear *et al.*, *ibid.* **23**, 3161 (1984).
13. N. E. Thomas, H. N. Bramson, W. T. Miller, E. T. Kaiser, *ibid.* **26**, 4461 (1987).
14. H. N. Bramson *et al.*, *ibid.*, p. 4466.
15. D. B. Glass, L. J. Lundquist, B. M. Katz, D. A. Walsh, *J. Biol. Chem.* **264**, 14579 (1989).
16. S. M. van Patten, G. J. Heisermann, H.-C. Cheng, D. A. Walsh, *ibid.* **262**, 3398 (1987).
17. J. Granot, A. S. Mildvan, K. Hiyama, H. Kondo, E. T. Kaiser, *ibid.* **255**, 4569 (1981).
18. S. Shoji, K. Titani, J. G. Demaille, E. H. Fischer, *ibid.* **254, 6211 (1979).**
19. J. A. Buechler and S. S. Taylor, *Biochemistry* **29**, 1937 (1990).
20. C. S. Gibbs and M. J. Zoller, *ibid.* **30**, 5329 (1991).
21. N. Nelson and S. S. Taylor, *J. Biol. Chem.* **258**, 10981 (1983).
22. J. S. Jimenez, A. Kupfer, V. Gani, S. Shalteil, *Biochemistry* **21**, 1623 (1982).
23. D. Bhatnager, F. T. Hartl, R. J. Roskoski, R. A. Lessor, N. J. Leonard, *ibid.* **23**, 4350 (1984).
24. J. Reed and V. Kinzel, *ibid.*, p. 1357.
25. J. Parello, P. A. Timmins, J. M. Sowadski, S. S. Taylor, *J. Mol. Biol.*, in press.
26. J. Reed, V. Kinzel, B. E. Kemp, H.-C. Cheng, D. A. Walsh, *Biochemistry* **24**, 2967 (1985).
27. S. Whitehouse, J. R. Feramisco, J. E. Casnellie, E. G. Krebs, D. A. Walsh, *J. Biol. Chem.* **258**, 3693 (1983).
28. A. Salerno, M. Mendelow, M. Provok, D. S. Lawrence, *ibid.* **265**, 18079 (1990).
29. M. J. Zoller, N. C. Nelson, S. S. Taylor, *ibid.* **256**, 10387 (1981).
30. S. K. Hanks, A. M. Quinn, T. Hunter, *Science* **241**, 42 (1988).
31. S. Brenner, *Nature* **329**, 21 (1987).
32. S. K. Hanks, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 388 (1987).
33. W. Yonemoto, M. L. McGlone, L. W. Slice, S. S. Taylor, *Methods Enzymol.*, in press.
34. T. Hunter, *Cell* **49**, 1 (1987).
35. F. Hofmann, *J. Biol. Chem.* **255**, 1559 (1980).
36. S. L. Weldon and S. S. Taylor, *ibid.* **260**, 4203 (1985).
37. J. Bubis and S. S. Taylor, *Biochemistry* **26**, 5997 (1987).
38. K. Titani *et al.*, *ibid.* **23**, 4193 (1984).
39. G. Hjelmquist, J. Anderson, B. Edlund, L. Engstrom, *Biochem. Biophys. Res. Commun.* **61**, 559 (1974).
40. B. E. Kemp, E. Benjamini, E. G. Krebs, *Proc. Natl. Acad. Sci. U.S.A.* **73** (4), 1038 (1976).
41. S. Whitehouse and D. A. Walsh, *J. Biol. Chem.* **258**, 3682 (1983).
42. S. M. van Patten, W. H. Fletcher, D. A. Walsh, *ibid.* **261**, 5514 (1986).


43. Supported by the Lucille P. Markey Foundation; 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 S. Bell, M. Montella, and G. Hasegawa for preparation of manuscript; the NIH National Research Resource at UCSD(RR01644) for data collection and staff members C. Nielsen and D. Sullivan for technical support; 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 1991; accepted 21 June 1991