A Predictive Scale for Evaluating Cyclin-dependent Kinase Substrates

A COMPARISON OF p34<sub>cdc2</sub> AND p33<sub>cdk2</sub> 

(Received for publication, June 18, 1996, and in revised form, July 25, 1996)

Jennifer K. Holmes and Mark J. Solomon‡

From the Department of Molecular Biophysics and Biochemistry, Yale University School of Medicine, New Haven, Connecticut 06520-8024

Protein phosphorylation by members of the Cdk (cyclin-dependent kinase) family of protein kinases is necessary for progression through the cell cycle. However, the primary sequence determinants of Cdk substrate specificity have yet to be examined quantitatively. We have used a panel of glutathione S-transferase peptide fusions to investigate the fine-structure specificity of p34<sub>cdc2</sub> and p34<sub>cdk2</sub>. Our data indicate that the generally held consensus sequences for p34<sub>cdk2</sub> represent a significant oversimplification of its true specificity and that this specificity is conserved between species. p33<sub>cdk2</sub> and p34<sub>cdk2</sub> have similar but distinct substrate specificities that are affected modestly by the associated cyclin subunit. We derive specific values of phosphorylation efficiencies by these enzymes that can be used to estimate the phosphorylation potential of proposed Cdk substrates.

The cell cycle consists of a series of strictly ordered steps, requiring the completion of one event before the next can occur. The protein kinases that control entry into and progression through various stages of the cell cycle are members of the Cdk (cyclin-dependent kinase) subfamily of protein kinases. Cdk activities fluctuate as a result of post-translational modifications and protein-protein interactions. An active Cdk is forming after binding to a cyclin partner and phosphorylation on a key threonine (Thr-161 in human p34<sub>cdk2</sub>). In vertebrates, Cdk4-cyclin D is necessary for passage through G1, p33<sub>cdk2</sub>-cyclin E necessary for the transition from G<sub>1</sub> to S phase, p33<sub>cdk2</sub>-cyclin A is necessary for progression through S, and p34<sub>cdk2</sub>-cyclin B is necessary for the transition from G<sub>2</sub> to M phase (1).

Crucial to our understanding of the cell cycle is the ability to identify for the various Cdk-cyclin complexes the key substrates whose phosphorylation leads to the progression through a particular cellular event. Many of these downstream effects could be caused directly by the Cdk; for example, p34<sub>cdk2</sub>-cyclin B can phosphorylate lamins thus leading to their disassembly (2–4), an important event in the initiation of mitosis. Other effects could be indirect, the result of a cascade of events initiated by the Cdk; for example, Cdk4-cyclin D phosphorylates Rb, thus releasing E2F to promote the transcription of many genes important for DNA replication (5).

An understanding of the basis of substrate specificities of different Cdk-cyclin complexes is of central importance as specificity can be influenced by many factors. Obviously the choice of a phosphorylation target site will be influenced strongly by inherent differences in the substrate binding region of a particular Cdk (6–9). In addition, the cyclin subunit could influence substrate specificity in any of the following ways: by binding a potential substrate and bringing it into contact with the Cdk; by targeting the Cdk to a particular subcellular location where it has access to only a limited number of potential substrates (10–12); or by restricting Cdk activities to a narrow window within the cell cycle so that the kinase can only affect those substrates present and able to be activated during that stage (1). Most likely, the substrate is recognized by a combination of the Cdk substrate binding pocket and long range interactions with surface residues of the cyclin subunit (13).

The majority of substrates would be recognized by the Cdk in association with any cyclin, but certain subsets might be recognized or preferred by a specific Cdk-cyclin pair (14). Several recent studies have indeed demonstrated that the identity of the cyclin partner can influence substrate specificity significantly (14–17).

Several loose consensus substrate sequences have been reported for p34<sub>cdk2</sub> based on a limited number of known in vivo and in vitro p34<sub>cdk2</sub> substrates (for review, see Ref. 18). These include (K/R)(S/T)(P)(K/R), where X is any amino acid (18) or a polar amino acid (19), and (S/T)(P)(K/R), where X is any amino acid (20). It has generally been assumed that p33<sub>cdk2</sub> has a similar specificity. The few studies investigating the substrate specificity of the Cdns have been performed primarily on p34<sub>cdk2</sub> (14, 18, 20–23) and have examined only a small number of peptides or sites in diverse proteins. A systematic study of protein kinase substrate specificity was carried out recently by Songyang et al. (9) using a peptide library containing approximately 2.5 billion unique peptides, with a fixed serine as the phosphate acceptor, as substrates for various kinases including p33<sub>cdk2</sub>-cyclin A and p34<sub>cdk2</sub>-cyclin B. This method identified a sequence similar to one of the consensus sites as the optimal substrate for p34<sub>cdk2</sub>-cyclin B, (K/R)(S/T)(P)(K/R).

We have investigated the substrate specificity of p33<sub>cdk2</sub> bound to cyclin A or E and of p34<sub>cdk2</sub> bound to cyclin A or B using a systematic series of specifically defined peptide substrates appended to the COOH terminus of glutathione S-transferase, constructed by polymerase chain reaction using degenerate oligonucleotides. These substrates allowed us to determine quantitatively the role of the primary sequence of a target site in substrate utilization. Our panel of altered target

* This work was supported by the Searle Scholars Program/The Chicago Community Trust (to M. J. S.) and by National Institutes of Health Grant GM47830 (to M. J. S.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

‡ Leukemia Society of America Scholar. To whom correspondence should be addressed: Dept. of Molecular Biophysics and Biochemistry, Yale University School of Medicine, 333 Cedar St., New Haven, CT 06520-8024. Tel.: 203-737-2792; Fax: 203-785-6404; E-mail: Mark.Solomon@Yale.edu.

1 The abbreviations used are: Cdk, cyclin-dependent kinase; GST, glutathione S-transferase.
sites has allowed us to compare the inherent differences in substrate recognition between p33<sub>cdk2</sub> and p34<sub>cdk2</sub> as well as to examine the effects of the cyclin regulatory subunits on specificity. In addition, we have found that the data generated from these experiments can be used to predict the potential utilization of novel phosphorylation sites.

MATERIALS AND METHODS

Production of GST Fusion Substrates—Substrates were constructed by polymerase chain reaction using pGEX-3X (24) or a previously made substrate as template. The 5′ primer, which included nucleotides 67–106 of GST, introduced an internal XhoI site (underlined): TCG ACT TCT GCT CGA GTA TCT GGA AGA AAA ATA TGA AGA G, and the 3′ primer produced a 45-mer (underlined) that either a degenerate or a specific oligonucleotide based on the following sequence: CGA TGA ATT CCC XNY XNN XNY XNN ACC CCC AGC ACC TTC GAT CAG, where X = G/C and N = G/A/T/C. The amplified products were cloned into pGEX-3X containing an introduced XhoI site at nucleotide 77 of GST using XhoI and EcoRI and sequenced over the peptide region. The terminal sequence of the wild type fusion substrate was GCGGGSGGG.

Purification of GST Fusion Substrates—The constructs were transformed into Escherichia coli strains TG1 or BL21 for protein expression. 100-ml bacterial cultures were grown in LB containing 0.1 mg/ml ampicillin at 37 °C until they reached an A<sub>600</sub> of 0.6–1.0. Isopropyl-1-thio-
β-n-p-galactopyranoside was added to 0.4 mM, and cells were incubated for 15 h at 23 °C. Cells were pelleted and washed twice with 0.9% NaCl before resuspension in 2 ml of lysis buffer (150 mM NaCl; 5 mM EDTA; 50 mM Tris (pH 8.0); 10% (v/v) glycerol; 5 mM dithiothreitol; 10 μg/ml each of leupeptin, chymostatin, and pepstatin; and 0.5 mg/ml lysozyme). After a 30-min incubation on ice, cells were lysed by the addition of Nonidet P-40 to 0.5% followed by sonication for two 25-s periods. The lysate was clarified by centrifugation at 40,000 rpm for 30 min at 4 °C using a TL100.2 rotor in a Beckman TL100 ultracentrifuge. The supernatant was applied to a column containing 200 μl of glutathione-agarose (Sigma) which had been prewashed with 3 ml of lysis buffer containing 0.5% Nonidet P-40. For binding of at least 30 min, the column was washed with 3 ml of lysis buffer containing 0.5% Nonidet P-40 followed by a wash with 3 ml of buffer H (50 μM Heps (pH 7.5); 100 mM NaCl; 50 μM Tris; 5 mM dithiothreitol; and protease inhibitors as above). The GST-peptide fusion was eluted with 600 μl of buffer H containing 5 mM glutathione and concentrated in a Centricon-10 concentrator (Millipore, Bedford, MA) for 60–90 min at 5,000 rpm in an SA-600 rotor at 4 °C. The final concentrations of the fusion proteins were determined spectrophotometrically (assuming an A<sub>280</sub> of 1.0 for a 1 mg/ml solution) ranging from 10 to 127 mg/ml.

Purification of Cdk-Cyclin Complexes—Sea urchin GST-cyclin B was expressed in E. coli and purified as described (25). The cyclin protein was added to a Xenopus egg extract arrested in interphase, and the activated p34<sub>cdk2</sub>-cyclin complexes were obtained on glutathione-agarose beads, eluted, and concentrated as described (26). The final concentration of purified Xenopus p34<sub>cdk2</sub>-cyclin B complexes was 184 nM, and their activity was 3.62 ± 0.10 pmol of phosphate transferred per min/μg of p34<sub>cdk2</sub> at a saturating concentration of the wild type substrate. Human Cdk-cyclin complexes were purified from SF9 insect cells coinfected with baculoviruses expressing GST-cyclin and Cdk. The complexes were retrieved on glutathione-Sepharose resin and eluted as described (27). The concentrations and the specific activities of the complexes toward the wild type substrate were as follows: p33<sub>cdk2</sub>-cyclin A, 455 nM and 4.77 ± 10<sup>3</sup> pmol of phosphate transferred/min/μg of p33<sub>cdk2</sub>; p33<sub>cdk2</sub>-cyclin E, 121 nM and 4.66 ± 10<sup>3</sup> pmol of phosphate transferred/min/μg of p33<sub>cdk2</sub>; p34<sub>cdk2</sub>-cyclin A, 12 nM and 1.58 ± 10<sup>3</sup> pmol of phosphate transferred/min/μg of p34<sub>cdk2</sub>; and p34<sub>cdk2</sub>-cyclin B, 12 nM and 1.13 ± 10<sup>3</sup> pmol of phosphate transferred/min/μg of p34<sub>cdk2</sub>.

Substrate Specificity of p34<sub>cdk2</sub> and p33<sub>cdk2</sub>

In Vivo Kinase Assays and Data Analysis—Xenopus p34<sub>cdk2</sub>-cyclin B kinase assays were performed by incubating 5 μl of substrate with 5 μl containing 3 nm [γ-<sup>32</sup>P]ATP, 0.25 μM C1bA (γ-P)ATP, 0.4 mM ATP, 15 mg MgCl<sub>2</sub>, 50 mM EDTA, 10 μg/ml diithiothreitol, 80 mM potassium β-glycerophosphate (pH 7.3), 1 mg/ml ovalbumin, and protease inhibitors as above at 25 °C. Human Cdk-cyclin kinase assays were performed by incubating 5 μl of substrate with 5 μl of enzyme containing 4.6 nm p34<sub>cdk2</sub>-cyclin A, 6.1 nm p33<sub>cdk2</sub>-cyclin E, 2.4 nm p34<sub>cdk2</sub>-cyclin A, or 2.4 nm p33<sub>cdk2</sub>-cyclin B, in a mixture with 0.25 μC/μl (γ-P)ATP, 0.4 mM ATP, 15 mM MgCl<sub>2</sub>, 60 mM Heps (pH 8.0), 0.1 mg/ml ovalbumin, and protease inhibitors as above at 25 °C. Phosphorylation was quantified by autoradiography of SDS-PAGE gels.

In vitro kinase assays were performed using purified Xenopus p34<sub>cdk2</sub>-cyclin B increased linearly with time over a 30-min period and a 1,000-fold range of concentrations (data not shown). A phosphorilation site mutant (KAPRK) was essentially unphosphorylated. Conditions within this linear range were chosen for all further experiments.

We replaced the charged residues at positions −1, +2, and +3 in the Michaelis-Menten equation.

RESULTS

Fine Substrate Specificity of p34<sub>cdk2</sub>—To refine the substrate specificity of p34<sub>cdk2</sub>, we constructed a series of substrates based on a histone H1 phosphorylation site, KSPRK (wild type), attached to the COOH terminus of glutathione S-transferase via gene fusion. This format allowed us to avoid the expense of producing synthetic peptides while examining a large number of sites within a single context. This approach differs from previous work investigating substrate specificity in which only a small number of sites within different protein contexts were examined, making direct comparisons of phosphorylation efficiencies difficult. We found using the wild type substrate that phosphorylation by purified Xenopus p34<sub>cdk2</sub>-cyclin B increased linearly with time over a 30-min period and a 1,000-fold range of concentrations (data not shown). A phosphorylation site mutant (KAPRK) was essentially unphosphorylated. Conditions within this linear range were chosen for all further experiments.

We replaced the charged residues at positions −1, +2, and +3 in the Michaelis-Menten equation.
with respect to the phosphorylated serine or threonine with alanines in single, double, and triple combinations to determine the overall importance of each position to substrate recognition. Phosphorylation of these alanine substitution mutants by *Xenopus* p34cdc2 was carried out over a wide range of substrate concentrations. Substitution at the −1 position (ASPRK) had only a small effect on phosphorylation efficiency, substitution at the +2 position (KSPAK) had a more significant effect, and substitution at the +3 position (KSPRA) had a severe effect (Fig. 1). The data were fit to the Michaelis-Menton equation, and kinetic parameters were determined for the four substrates that approached saturation within the concentration range of the experiment. The $K_m$ values (in µM) for these substrates were as follows: KSPRK, 98.0; ASPRK, 108; KSPAK, 446; and ASPAK, 976. The $V_{max}$ values were all between 3,190 and 4,120 pmol of phosphate transferred/min/µg of p34cdc2. Thus, all of the variation in substrate utilization was accounted for by substrate binding ($K_m$). The $K_m$ for histone H1 was 25 µM under the same assay conditions (data not shown), indicating that the $K_m$ values for our best peptide substrates were close to those of actual p34cdc2 substrates. Alanine substitutions at positions −1 and +2 had only moderate effects, increasing the $K_m$ to 1.1-fold and 4.6-fold the wild type value, respectively. However, mutation of the +3 position had a severe effect on $K_m$, even greater than the ASPAK double substitution substrate, which had a 10-fold effect, confirming previous suggestions that the identity of the +3 position was more important than the identity of the −1 or +2 position (9, 18–20).

We also examined a KTPRK substrate to determine the utilization of threonine compared with serine. The $K_m$ for this substrate was 153 µM, indicating that serine was slightly preferred (by 1.6-fold) as the phosphate acceptor.

We systematically replaced residues at positions −1, +2, and +3 with each possible amino acid to define any specific substrate preferences at these sites. The relative specificity of *Xenopus* p34cdc2-cyclin B toward these single amino acid substitutions was determined at a fixed substrate concentration of 50 µM, which is well below the $K_m$ value of the wild type substrate (Fig. 1) and thus within the linear range. The analysis of these substrates agreed well with the findings from the alanine substitution substrates regarding the overall sensitivity of each position, although there were marked preferences.

The first position was relatively insensitive to amino acid substitutions (Fig. 2A). All but five of the substrates were phosphorylated at least 80% as well as the wild type substrate. The peptide with proline at the first position was a relatively poor substrate, phosphorylated at only 46% of the wild type level. Thus we find that all substitutions at the −1 position are tolerated, but that there is a distinct variation in preference. In contrast, the conventional consensus sequences specify only lysine or arginine at the first position (18) or indicate that all amino acids are equivalent (20).

The +2 position tolerated a much more limited number of amino acid substitutions (Fig. 2B). All but five of the substrates were phosphorylated at least 80% as well as the wild type substrate. The peptide with proline at the first position was a relatively poor substrate, phosphorylated at only 46% of the wild type level. Thus we find that all substitutions at the −1 position are tolerated, but that there is a distinct variation in preference. In contrast, the conventional consensus sequences specify only lysine or arginine at the first position (18) or indicate that all amino acids are equivalent (20).

The +2 position tolerated a much more limited number of amino acid substitutions (Fig. 2B). Only two substrates approached wild type levels, lysine at 108% and methionine at 80.1%. All other substrates were phosphorylated less than 65% as well as the wild type substrate. The most poorly tolerated substitutions were aspartic acid, glutamate, and proline, which reduced phosphorylation to less than 10% of wild type. The traditional consensus sequences either do not indicate any specificity at this position (18) or require a polar side chain (19).

**Fig. 2.** Effects of amino acid substitutions at each of the three charged positions in the canonical substrate (KSPRK) on substrate utilization by *Xenopus* p34cdc2-cyclin B. The three panels show the effects of substitutions at position −1 (panel A), +2 (panel B), or +3 (panel C) of the KSPRK sequence. Assays were performed as described under "Materials and Methods" at a substrate concentration of 50 µM. The relative ability of a peptide to be utilized as a substrate is expressed as percent of wild type (KSPRK). Values represent the means ± S.E. from five separate experiments.
In fact, substitutions at the +2 position show a full range of tolerance, from excellent to poor. Moreover, several polar amino acid side chains form poor substrates (e.g. aspartate and glutamine), whereas some nonpolar side chains yield excellent substrates (e.g. methionine).

The +3 position was the most sensitive to substitution (Fig. 2C). Only arginine and lysine were well tolerated. Peptides containing histidine or proline were utilized at about 20% of the wild type level, which although low, was still considerably greater than the rest of the substrates. Aspartic acid and glutamic acid were the least tolerated changes at this position, resulting in peptides that were phosphorylated at less than 0.5% of the wild type level. The results for the +3 position agree well with the consensus sites, which specify only lysine or arginine (18–20). However, our results expand this definition by identifying histidine and proline as tolerable substitutions and by indicating that very few amino acids are entirely excluded at this position.

Based on molecular modeling, Songyang et al. (9) have proposed that the proline residue directly following the serine is necessary to anchor the substrate in the correct orientation for phosphorylation. Mutation of this proline to asparagine (KSNRK) abolished all detectable phosphorylation (data not shown). We also tested the utilization of the substrate KSTRK (data not shown). This sequence is based on the nonconventional p34\(^{cdc2}\) phosphorylation sites in myosin light chain (SSKR and KTTKK) (23, 28). This substrate was utilized at 0.03% of wild type (data not shown), which agrees well with the findings of Yamakita et al. (23) who saw extremely low levels of phosphorylation of myosin light chain by p34\(^{cdc2}\) relative to phosphorylation of histone H1 (23).

**Substrate Specificity of p34\(^{cdc2}\) versus p33\(^{cdk2}\)**—We turned to purified human Cdk-cyclin complexes, rather than Xenopus enzymes, for studies comparing the specificities of p33\(^{cdk2}\) and p34\(^{cdc2}\) since these enzymes were more readily available. We first verified that human p34\(^{cdc2}\) had essentially the same specificity as that from Xenopus. Both enzymes were most sensitive to alanine substitution at the +3 position, followed by the +2 and −1 positions, respectively (Figs. 1 and 3). They also showed the same general pattern of fine structure specificity based on the single amino acid substitutions (Figs. 2 and 4). The most significant difference was seen with the RSPRK.
substrate, which was utilized at 63.7% of the wild type level with the Xenopus enzyme (Fig. 2A) and at 140% of the wild type level for the human enzyme (Fig. 4A). This result may reflect slight differences in the structures of the substrate binding pockets of the two enzymes that allow the human enzyme to accommodate the bulky arginine residue at the −1 position more readily. The overall similarity in substrate specificity between human and Xenopus p34cdc2-cyclin B provides reassurance that the innate specificities of these enzymes have been well conserved through evolution.

We next compared the sensitivities of p33cdc2-cyclin A and p34cdc2-cyclin A with single alanine substitutions at the −1, +2, and +3 positions of our substrate (Fig. 3). The enzymes were qualitatively similar in that each was barely sensitive to substitution by alanine at the −1 position, more sensitive to substitution at the +2 position, and very sensitive to substitution at the +3 position. Quantitatively, however, p33cdc2 was much more sensitive to the +2 and +3 substitutions than was p34cdc2 and showed almost no detectable activity toward the KSPRA substrate. We obtained $K_m$ and $V_{max}$ values for all four human enzymes using the three substrates (KSPRK, KTPrK, and ASPRK) that approached saturation closely enough to permit accurate fits to the Michaelis-Menton equation (Table I). The p34cdc2 enzymes consistently had about a 2-fold or more higher affinity for the substrates than the p33cdc2 enzymes. Interestingly, p33cdc2-cyclin E had a significantly lower $K_m$ for the substrates than did p33cdc2-cyclin A. We assume that this effect represents a slight interaction between cyclin E and the GST substrates, although it could also represent a very subtle alteration in the substrate binding region of p33cdc2 induced by the cyclin-binding partner. Although the p33cdc2 enzymes had approximately equal affinities for substrates containing either Ser or Thr as the phosphorylation target, the p34cdc2 enzymes preferred Thr by approximately 2-fold, again suggesting that there may be minor differences in the substrate binding interfaces of these enzymes. We were struck that the $V_{max}$ values for the p34cdc2 enzymes were significantly higher than those for the p33cdc2 enzymes since we normalized the amounts of each enzyme we used in the assays based on histone H1 kinase activity. One interpretation is that the p33cdc2 enzymes may have a higher binding affinity for histone H1 and that we therefore underestimated the amounts of these enzymes to use in our substrate assays.

We examined the sensitivities of all four human enzymes to our full panel of substitution substrates in order to probe their site preferences more systematically. The four enzymes displayed remarkably similar fine specificity at the −1 position (Fig. 4A). As with Xenopus p34cdc2, the human enzymes tolerated all substitutions at this position, but Pro was again the most poorly tolerated at approximately 35–45% of the wild type level. Although there were some statistically significant differences in sensitivity between p34cdc2 and p33cdc2, these were relatively modest, especially compared with some of those seen at the +2 and +3 positions (see below). The greatest differences were a relative preference of p33cdc2 for Trp and of p34cdc2 for Ile. Emphasizing the tolerance at this position, phosphorylation of a number of the substrates by all four enzymes was near or even exceeded that of the wild type substrate.

As with Xenopus p34cdc2, human p34cdc2 and p33cdc2 showed a wide range of substrate preferences at the +2 position. Again, the most poorly tolerated substitutions were Asp, Pro, and Gln, each phosphorylated at less than 10% of the wild type level. However, unlike the case at the −1 position, there were marked differences at the +2 position in the sensitivities of p33cdc2 and p34cdc2 to some substitutions. In particular, the p34cdc2 enzymes phosphorylated the Gly-containing substrate (KSPGK) at only 1.0 and 3.2% of the wild type level, whereas p34cdc2 phosphorylated the same substrate at 14.8 and 46.9% of the wild type level (Fig. 4B). Similarly, the p33cdc2 enzymes phosphorylated the Pro-containing substrate (KSPPK) at only 0.1 and 0.5% of the wild type level, whereas p34cdc2 phosphorylated the same substrate at 3.0 and 5.1% of the wild type level (Fig. 4B).

Substitutions at the +3 position produced the greatest effects. For the p34cdc2 enzymes, substitutions of basic amino acids were clearly preferred, whereas acidic substitutions were tolerated most poorly. As with Xenopus p34cdc2, His and Pro stood out somewhat above the other amino acids that made poor, but tolerated, substrates (Fig. 4C). The p33cdc2 enzymes showed the same pattern of sensitivity but to a much greater extent, thus elaborating on the heightened sensitivity seen earlier with the single alanine substitutions (Fig. 3). In particular, there were only two substrates that were phosphorylated at better than 1% of the wild type level, KSPRK (the wild type), at 100%, and KSPPR, at 4.3 and 5.0% (Fig. 4C). This sensitivity of p33cdc2 to substitution of Lys with even another basic residue was in marked contrast to the sensitivity of the p34cdc2 enzymes, which phosphorylated the KSPRK substrate at 21.1 and 59.9% of the wild type level (Fig. 4C). Most +3 substitution substrates were not detectably phosphorylated at all by the p33cdc2 enzymes, and only six (Ala, Cys, His, Lys, Pro, and Arg) were phosphorylated at greater than 0.1% of the wild type level by either enzyme (Fig. 4C), whereas we could always detect at least a low level of phosphorylation of the same substrates by the p34cdc2 enzymes (Fig. 4C).

We generally saw no or only very modest effects of the cyclin partner on the phosphorylation of the various substrates by either p34cdc2 or p33cdc2. For instance, the largest effects at the −1 position involved the ISPRAK substrate, which was phosphorylated 1.4 times as well by p33cdc2-cyclin E as by p33cdc2-cyclin A, and WSPRK, which was phosphorylated 1.5 times as well by p33cdc2-cyclin A as by p33cdc2-cyclin E (Fig. 4A). At the +2 position, KSPGK was phosphorylated 3.3 times as well by p33cdc2-cyclin A as by p33cdc2-cyclin E and 3.2 times as well by p34cdc2-cyclin B as by p34cdc2-cyclin A (Fig. 4B). Of those substrates phosphorylated at all well at the +3 position, the greatest effect of the cyclin partner was seen with the KSPRK substrate, which was phosphorylated 2.8 times as well by p34cdc2-cyclin B as by p34cdc2-cyclin A (Fig. 4C). However, since all of the +3 position substrates were phosphorylated approximately 2–3-

### Table I

| Enzyme complex | KSPRK | KTPrK | ASPRK |
|----------------|-------|-------|-------|
|                | $K_m$ | $V_{max}$ | $K_m$ | $V_{max}$ | $K_m$ | $V_{max}$ |
| p33cdc2-cyclin A | 213 | 4,770 | 230 | 4,340 | 234 | 5,890 |
| p33cdc2-cyclin E | 135 | 4,680 | 115 | 5,110 | 141 | 5,410 |
| p34cdc2-cyclin A | 84.5 | 15,840 | 32.7 | 10,640 | 108 | 17,290 |
| p34cdc2-cyclin B | 75.3 | 11,310 | 40.2 | 8,520 | 71.8 | 12,000 |
fold as well by p34<sup>cd2</sup>-cyclin B as by p34<sup>cd2</sup>-cyclin A, relative to the wild type substrate to which we normalized all of our data, we suspect that the p34<sup>cd2</sup>-cyclin A enzyme actually prefers the wild type KSPRK sequence relative to p34<sup>cd2</sup>-cyclin B. Although all these effects could reflect subtle alterations in the substrate binding region of p33<sup>cd2</sup>-cyclin and of p34<sup>cd2</sup>-cyclin caused by binding different cyclins, we are more inclined to view the generally sporadic effects as being due to relatively weak longer range interactions of the cyclins with some individual substrates. The physiological relevance of 2-fold differences in phosphorylation efficiency is doubtful.

**Prediction of Substrate Utilization**—Further analysis of the alanine substitution substrates indicated that the single amino acid substitutions showed an additive effect that could be used to predict the substrate utilization of a double or triple alanine substitution substrate by *Xenopus* p34<sup>cd2</sup>-cyclin B. We confirmed the generality of this finding by testing our ability to predict the phosphorylation efficiency of random multiple substitution substrates based on the data from the single amino acid substitutions. We predicted the utilization of each substrate by multiplying the percent wild type phosphorylation of the respective single amino acid substitutions. This predicted value was then compared with the experimental value for that substrate (a ratio of 1 represents a perfect prediction) (Table II). For example, to predict the utilization of the SSPNL triple mutant, we multiplied the percent wild type phosphorylation of the three single amino acid substitutions, SSPRK (75.1%) and KSPNK (24.9%) and KSPLL (2.6%). The result (0.50%) was then compared with the actual utilization for this substrate (0.32%) to produce a ratio of 1.56.

We chose the substrates to reflect a broad range of amino acid substitutions. These substrates had predicted phosphorylation efficiencies ranging from 42.6 to 0.07% of wild type. There was a good correlation between the actual and predicted values; only two of the substrates had a predicted to actual ratio of greater than 5 or less than 0.2 (Table II). We repeated this analysis with the human cyclins and obtained similar results (data not shown).

**DISCUSSION**

We have used a panel of GST fusion proteins containing systematic alterations of a canonical p34<sup>cd2</sup> phosphorylation site to determine the fine specificity of p34<sup>cd2</sup> and p33<sup>cd2</sup> bound to various cyclins. Understanding the similarities and differences in the specificities of these enzymes is an essential first step toward evaluating potential substrates that could play important roles in cell cycle progression. Previous studies of p34<sup>cd2</sup> phosphorylation sites have involved compilations of sites found in diverse proteins and examination of modest numbers of synthetic peptide substrate variants. Recently a peptide selection approach has been used to define a p34<sup>cd2</sup> consensus site as (K/R)(S/P/R/F)/(R/K/H). Although this method is extremely useful for rapidly determining optimal phosphorylation sites, it is not as well suited for determining which amino acids are poorly tolerated or excluded, it does not analyze all 20 amino acids, and it systematically overestimates the phosphorylation of suboptimal substrates (9). Our approach benefits from using a comprehensive collection of variant substrates within the same protein context. The fusion proteins are inexpensive and readily purified, and additional mutant phosphorylation sites can be engineered quite easily. We also expect that our panel of substrates will prove useful for determining the substrate specificity of other Cdk's, including those involved in cell cycle control as well as those involved in other processes.

Overall, we found that *Xenopus* p34<sup>cd2</sup> was least sensitive to substitutions at the -1 position of the wild type sequence KSPRK (with respect to the phosphorylated Ser), fairly sensitive to substitutions at the +2 position, and most sensitive to substitutions at the +3 position. Although this general pattern is consistent with widely held consensus sites for phosphorylation by p34<sup>cd2</sup>, our data significantly alter our view of what sequences can constitute good, fair, or poor phosphorylation targets. Our finding that the -1 position can accommodate any amino acid, but that there is about a 2-fold variation in phosphorylation efficiency, is closer to the consensus that posits no specificity than to those that place a basic residue at this position. At the +2 position, we found that neither consensus view, either specifying a polar residue or tolerating all amino acids, adequately fit the data. There was a strong degree of specificity at this position since some substitution mutants were phosphorylated almost 20-fold more efficiently than others. However, we have been unable to discern any simple pattern to explain this specificity. Clearly, though, both polar and nonpolar amino acid side chains could form excellent substrates, and some polar amino acids yielded quite poor substrates. At the +3 position our data are in agreement with the consensus view that basic residues are best. However, by focusing on the best sites, the consensus view fails to distinguish among the poorer sites. We would divide the substitutions at the +3 position into four classes: basic residues, which form excellent sites (about 100% of wild type phosphorylation efficiency); His and Pro, which can form surprisingly strong sites (about 20% of wild type); most other amino acids, which form weak but significant sites (about 5% of wild type); and acidic groups, which form virtually unphosphorylatable sites. The approximately 20-fold reduction in binding affinity on substitution of Ala and most other amino acids at the +3 position corresponds to a weakening of the interaction by about 1.8 kcal/mol, which is consistent with the loss of a single ionic interaction involving the +3 basic residue.

We observed two classes of modest effects of the cyclin partners on phosphorylation of the substrates. First, we noted that the cyclin A-containing complex of p34<sup>cd2</sup> had a consistently 2-fold higher $K_m$ for our substrates than the cyclin E-containing complex (Table I). The $K_m$ was not measured for all substrates, but if phosphorylation efficiencies reflect changes in $K_m$, and not in $V_{max}$, then this difference in $K_m$ probably applies to nearly all of the substrates and not just to those shown in

---

**Table II**

**Prediction of substrate utilization for *Xenopus* p34<sup>cd2</sup>-cyclin B**

| Substrate sequence | %WT predicted | %WT actual | Predicted/actual |
|--------------------|---------------|------------|-----------------|
| HSPLR              | 42.6          | 18.6 (± 1.73) | 2.29 |
| ASPAK              | 34.0          | 10.4 (± 0.28) | 3.27 |
| KSPRK              | 5.46          | 0.51 (± 0.04) | 10.7 |
| ASPRA              | 5.26          | 4.05 (± 0.32) | 1.30 |
| LSFRS              | 5.12          | 4.57 (± 0.33) | 1.12 |
| FSPLH              | 3.57          | 0.60 (± 0.07) | 5.95 |
| HSPTV              | 2.16          | 1.43 (± 0.08) | 1.51 |
| KSPAA              | 2.04          | 1.14 (± 0.11) | 1.79 |
| ASPAA              | 1.91          | 0.91 (± 0.09) | 2.10 |
| SSPNL              | 0.50          | 0.32 (± 0.02) | 1.56 |
| PSSPP              | 0.49          | 0.19 (± 0.04) | 2.59 |
| LSPLD              | 0.20          | 0.22 (± 0.04) | 0.91 |
| YSPQQ              | 0.15          | 0.51 (± 0.03) | 0.29 |
| PSPPQ              | 0.07          | 0.22 (± 0.08) | 0.32 |
| KSTRK              | NA            | NA         | 1.00 |
Table I. This result may indicate that cyclin E is a “better” cyclin in that it may be able to induce a better geometry of the binding pocket in p34<sup>cdk2</sup> for substrates. We saw no comparable effect of cyclin A versus cyclin B in the p34<sup>cdk2</sup>-containing complexes. We also noted a more sporadic effect of cyclin partner on phosphorylation efficiencies that we are inclined to attribute to weak longer range interactions between the cyclin and individual target sequences.

Although the specificities of p34<sup>cdk2</sup> and p33<sup>cdk2</sup> were generally similar, we were surprised to find a number of instances where p33<sup>cdk2</sup> was far more selective. p33<sup>cdk2</sup> was much less tolerant of Gly or Pro at the +2 position than was p34<sup>cdk2</sup>. This effect was approximately 10-fold or more, depending on the cyclin partner. The greatest differences were seen at the +3 position, where p33<sup>cdk2</sup> essentially did not phosphorylate (less than 0.4% of wild type efficiency) any substrate not containing Lys or Arg at this site. Even Arg, which yields a very good acid substitution mutants on phosphorylation efficiency were anticipated to guide the phosphorylation site can be taken as a strong indication that additive (Table II). We envision that a full prediction of potential phosphorylation sites on novel proteins will help to guide experiments toward the most likely physiological sites. For example, our data predict that an intuitively poor target site, YSPMH, would be phosphorylated almost twice as efficiently by Xenopus p34<sup>cdk2</sup>-cyclin B as an intuitively excellent site, KSPDR (13.0% versus 7.0% of wild type efficiency). We do not anticipate that our predictive scale will be accurate for all sites in all protein contexts. Clearly, many factors combine to determine the phosphorylation efficiency of a given target. A theoretically excellent site could be buried within a protein or folded rigidly in an unfavorable conformation. Similarly, an otherwise weak site could be folded tightly and presented in a favorable way. Additional interactions between either subunit of the Cdk and a substrate could further influence specificity. Despite these caveats, and in particular because their contributions are difficult to evaluate, we feel that our scale presents an unbiased starting point for examination of Cdk substrates. Tables showing the phosphorylation efficiencies depicted graphically in Figs. 2 and 4 are readily available from the authors.

Acknowledgments—We thank Beth Egan, Deborah Enke, David Gonda, Karen Ross, and David Stern for comments on the manuscript; David Gonda for encouragement throughout this work; and Joyce Wall for constructing many of the substrates. We are extremely grateful to Wade Harper for providing the purified human Cdk-cyclin complexes.

REFERENCES

1. Pines, J. (1995) Biochem. J. 308, 697–711
2. Perez, M., Nakagawa, J., Durée, M., Labbé, J. C., and Nigg, E. A. (1990) Cell 61, 591–602
3. Desver, G., Iovcheva-Dessy, C., Bischoff, J. R., Beach, D., and Goldman, R. (1991) J. Cell Biol. 112, 523–532
4. Peter, M., Hattlinger, E., Haner, M., Aebi, U., and Nigg, E. A. (1991) EMBO J. 10, 1535–1544
5. Kato, J.-y., Matsushima, H., Hiebert, S. W., Ewen, M. E., and Sherr, C. J. (1993) Genes Dev. 7, 351–342
6. Knighton, D., Zheng, J., Eyck, L. F. T., Ashford, V. A., Xuong, N.-H., Taylor, S. S., and Sowadski, J. M. (1991) Science 253, 407–414
7. Knighton, D. R., Zheng, J., Eyck, L. F. T., Xuong, N.-H., Taylor, S. S., and Sowadski, J. M. (1991) Science 253, 414–420
8. DeBondt, H. L., Rosenblatt, J., Janzarik, J., Jones, H. D., Morgan, D. O., and Kim, S.-H. (1993) Nature 363, 595–602
9. Songyang, Z., Blechner, S., Haagland, N., Hoekstra, M. F., Piwnica-Worms, H., and Cantley, L. C. (1994) Curr. Biol. 4, 973–982
10. Pines, J., and Hunter, T. (1991) J. Cell Biol. 113, 1–17
11. Pines, J., and Hunter, T. (1994) EMBO J. 13, 3772–3781
12. Jackman, M., Firth, M., and Pines, J. (1995) EMBO J. 14, 1646–1654
13. Jeffrey, P. D., Russo, A. A., Polyak, K., Gibbs, E., Hurwitz, J., Massague, J., and Pavletich, N. P. (1995) Nature 376, 313–320
14. Minshall, J., Golsteyn, R., Hill, C. S., and Hunt, T. (1990) EMBO J. 9, 2865–2875
15. Thomas, L., Clarke, P. R., Pagano, M., and Grunenberg, J. (1992) J. Biol. Chem. 267, 6183–6187
16. Hoffmann, I., Clarke, P. R., Marcote, M. J., Karsenti, E., and Draetta, G. (1993) EMBO J. 12, 53–63
17. Pepsey, D., Parker, L. L., Ewen, M. E., Toebes, M., Hall, F. L., Xu, M., Zanetina, A., van der Eb, A. J., and Piwnica-Worms, H. (1993) EMBO J. 12, 1947–1954
18. Nigg, E. A. (1991) Semin. Cell Biol. 2, 261–270
19. Moreno, S., and Nurse, P. (1990) Cell 61, 549–551
20. Nigg, E. A. (1993) Trends Cell Biol. 3, 296–301
21. Eriksson, E., and Maller, J. L. (1989) J. Biol. Chem. 264, 19577–19582
22. Muzio, O., Meggio, F., Draetta, G., and Pinna, L. A. (1992) FEBS Lett. 301, 111–114
23. Yamakita, Y., Yamashiro, S., and Matsumura, F. (1994) J. Cell Biol. 124, 125–137
24. Smith, D. B., and Johnson, K. S. (1988) Science 241, 697–711
25. Solomon, M. J., Glotzer, M., Lee, T. H., Philippe, M., and Kirschner, M. W. (1990) Cell 63, 1013–1024
26. Solomon, M. J., Lee, T., and Kirschner, M. W. (1992) Mol. Biol. Cell 3, 13–27
27. Harper, J. W., Adams, G. R., Wei, N., Keyomarsi, K., and Elledge, S. J. (1993) Cell 73, 805–816
28. Satterwhite, L. L., Lohka, M. J., Wilson, K. L., Scherson, T. Y., Ciesek, L. J., Corden, J. L., and Pollard, T. D. (1992) J. Cell Biol. 118, 595–605