Volume 304, number 1, 73–77 FEBS 11140 June 1992

Exon–intron organization of the *Arabidopsis thaliana* protein kinase genes *CDC2a* and *CDC2b**

Yoshiro Imajuku, Takashi Hirayama, Hideki Endoh and Atsuhiro Oka

Laboratory of Molecular Genetics, Institute for Chemical Research, Kyoto University, Uji, Kyoto, Japan

Received 8 April 1992

We have previously shown by cDNA cloning that a higher plant, *Arabidopsis thaliana*, possesses at least two *CDC2* genes (*CDC2a* and *CDC2b*) similar to the cell-cycle-controlling *cdc2* gene of *Schizosaccharomyces pombe*. To understand the exon–intron organization of these genes, genomic clones were isolated and their nucleotide sequences determined. The coding and 5′-untranslated regions of *CDC2a* were interrupted by seven and one introns, respectively, whilst *CDC2b* contained three introns within the coding portion. These intron positions partly overlapped with each other and with those of the yeast *cdc2* gene, nevertheless the lengths and sequences of the corresponding introns were diverse.

Cell cycle control; Evolution; Gene structure; Higher plant; Nucleotide sequence

1. INTRODUCTION

Mitotic cell division in higher plants is usually restricted to meristems, but its control system is little investigated. By contrast, there has been extensive study of the cell division cycle of the fission yeast *Schizosaccharomyces pombe*, in which the *cdc2* gene product ($p34^{cdc2}$) has a principal role by modulating its protein serine/threonine kinase activity (for reviews, see [1,2]). Defects in the *cdc2* gene lead to arrest at late G1 (start point for commitment to the mitotic cell cycle) and at the G2–M transition [3]. Functional homologs that can complement yeast *cdc2* mutations have been found from a variety of plants and animals [4–9]. Therefore, there seems to be a $p34^{cdc2}$-centered universal system for cell cycle control in every eukaryote.

Though the yeast chromosome contains only one *cdc2* gene [10,11], in a higher plant, *Arabidopsis thaliana*, there is a functional homolog (*CDC2a*) and at least one more *cdc2*-like gene (*CDC2b*), whose role has not been tested because of the lack of a full-length cDNA clone [4]. It has also been shown that several other organisms have two or more *cdc2*-like genes; some are functional homologs and others are not [6,9,12–14]. From their structural similarity these *cdc2* and *cdc2*-like genes seem to have originated from a common prototype. However, since demonstration of their ubiquity has all been done by cloning of cDNAs except for yeast *cdc2* [10], no information is available on correspondence of their genomic structures. To investigate the exon–intron organization of the *A. thaliana CDC2a* and *CDC2b* genes, together to deduce the complete structure of the *CDC2b* gene product, we have now cloned and sequenced genomic DNA fragments containing *CDC2a* and *CDC2b*.

Abbreviations: bp, base-pair(s); cDNA, DNA complementary to mRNA; kb, kilobase(s) or 1000 bp; nt, nucleotide position(s); $p34^{cdc2}$, 34-kDa protein encoded by *cdc2* or *cdc2*-like genes; PCR, polymerase chain reaction.

*DDBJ/EMBL/GenBank accession number: The DDBJ/EMBL/GenBank accession numbers for the sequences reported in this paper are D10850 and D10851.

Correspondence address: A. Oka, Laboratory of Molecular Genetics, Institute for Chemical Research, Kyoto University, Uji, Kyoto 611, Japan. Fax: (81) (774) 33 1247.

2. MATERIALS AND METHODS

Standard procedures for DNA experiments and most of the materials used were previously described [4]. A genomic library of *A. thaliana* (Columbia ecotype) constructed with λEMBL3 [15] was purchased from Clontech Laboratories, Inc. A mini library with λEMBL4 [15] was made by inserting 17-kb DNA fragments that were purified from *Eco*RI digests of *A. thaliana* genomic DNA, since the entirety or majority of the *CDC2a* coding sequence is on a single 17-kb *Eco*RI fragment [4].

For PCR primers, seven oligodeoxyribonucleotides were synthesized by a Beckman System 1 DNA Synthesizer: the forward primers 1 to 6 had sequences of nt 67–86, 276–292, 332–351, 446–466, 532–551, and 604–623 in the 5′-untranslated region of *CDC2b*, and the backward primer carried a sequence of nt 1389–1370 in the 3′-half of the *CDC2b* coding region (for nt, see Fig. 1, lower part). PCR was done on a mixture of *A. thaliana* cDNAs using a DNA thermal cycler (Perkin Elmer/ Cetus Corp) and various sets of the primers.

Complementation tests were performed with *S. pombe* (*cdc2-33*ts) as described previously [4]. Plasmids used for such experiments were pNH290 carrying *A. thaliana CDC2a* cDNA [4] and pIMA1 bearing *A. thaliana CDC2b* cDNA (see the results and discussion), both of which were derivatives of the expression vector pREP1 containing the *nmt1* promoter inducible by thiamine deprivation [16].

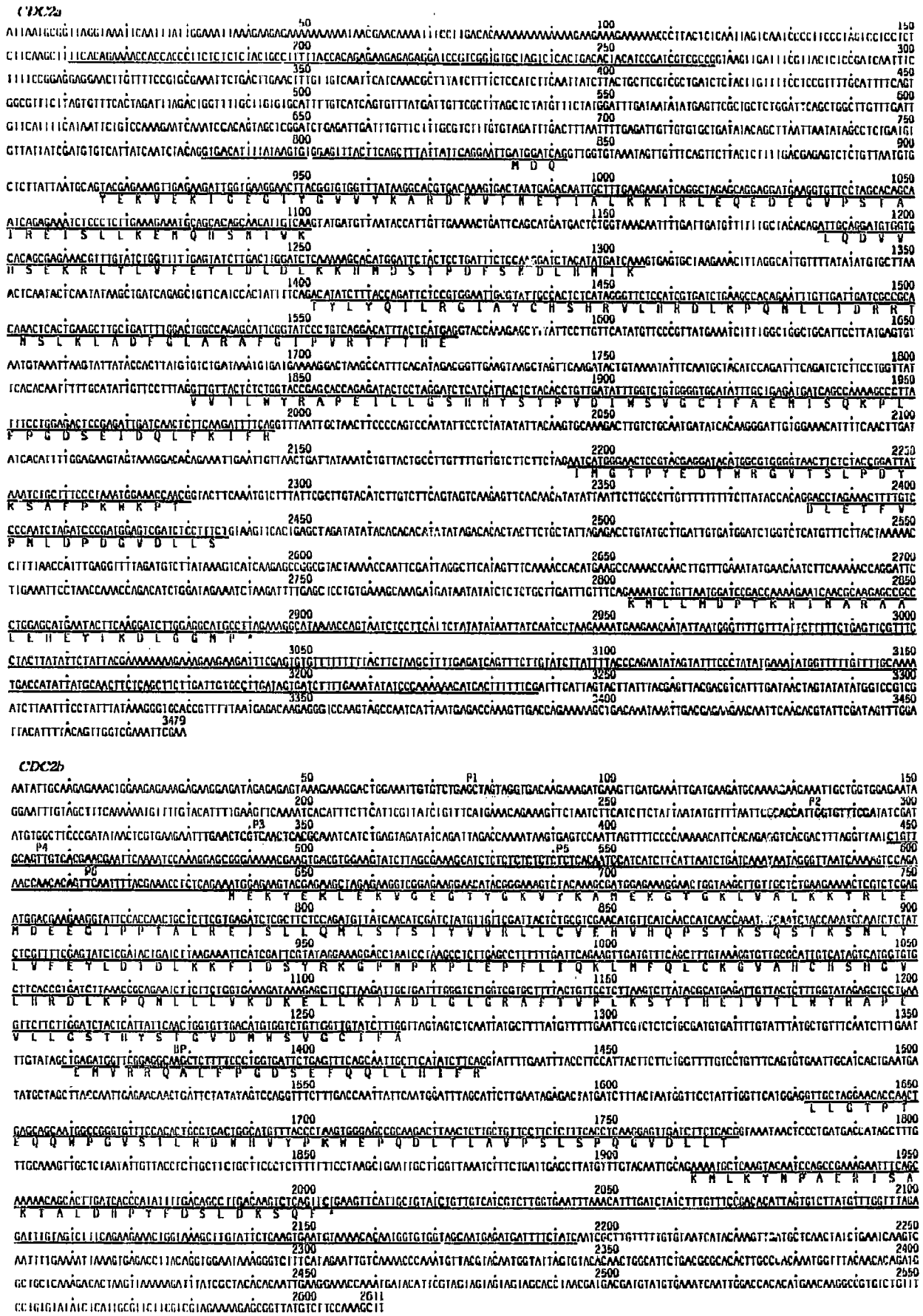
CDC2a
CDC2b

Fig. 1. Genomic sequence of *A. thaliana CDC2a* (upper part) and *CDC2b* (lower part). Nucleotides are numbered from the first residue, the *Ase*I site (for *CDC2a*) and the *Ssp*I site (for *CDC2b*). Exons are underlined, and the predicted amino acid sequences are presented below the nucleotide sequences. The 5′-end of each first exon is temporary assigned as the 5′-end of the longest cDNA. The locations of seven PCR primers are marked by shadowing on the *CDC2b* sequence: P1 to P6 and BP indicate the forward primers 1 to 6 and the backward primer, respectively.
←

3. RESULTS AND DISCUSSION

3.1. *Cloning of the* A. thaliana CDC2a *gene*

Using the *CDC2a* cDNA as a probe, the mini library was screened to isolate *CDC2a* genomic clones because a screening with the Clontech genomic library had been unsuccessful for unknown reasons. With one positive clone, restriction cleavages coupled with DNA hybridization were analyzed. The results indicated that the whole *CDC2a* cDNA sequence is involved in a 3.5-kb region (*AseI–AsuII*) on a middle portion of this *Eco*RI fragment (data not shown). This DNA region was sequenced entirely for both strands by the chain-termination procedure [17], and the results are shown in Fig. 1 (upper part). It was found that a sequence identical to the previously described *CDC2a* cDNA sequence [4] is split into nine exons by eight introns (I to VIII), one at the 5′-untranslated region and seven within the coding region, as schematically depicted in Fig. 2 (upper part). All of the intron sequences started with GT and ended with AG, and some of their 5′- and 3′-splice sites were similar to the proposed consensus sequences for plant splice junctions, (C/A)AG↓GTAAGT and TTT(T/Pu)TT(T/Pu)$_4$TGCAG↓G, respectively (vertical arrows mean splice junctions) [18]. Among the seven introns within the coding region, three introns, III, V, and VIII, were located at the same sites as those of the *S. pombe cdc2* gene introns II, III, and IV, respectively [10], although their lengths and sequences were dissimilar with the two genes.

3.2. *Cloning of the* A. thaliana CDC2b *gene*

Since no full-length *CDC2b* cDNA clone has been obtained, only its 3′-half sequence data are available [4]. The longest cDNA clone previously isolated was used as a probe for screening the Clontech genomic library. Several positive overlapping clones were isolated, one of which was then sequenced. The results (Fig. 1, lower part) showed that a 2.6-kb region (*Ssp*I–*Hin*dIII) is enough for encoding a p34^{cdc2}-like protein (see the next section). The presumed start codon was present at a position approximately corresponding to the start codon in *CDC2a*. Sequence comparison between the genome and the incomplete *CDC2b* cDNA indicated the presence of three introns within the 3′-half of the *CDC2b* coding region. All of these three introns had sequence characteristics similar to those found in the *CDC2a* introns as described above. To see whether more introns are present in the 5′-half region for which no cDNA clone is available, PCR was done on an *A. thaliana* cDNA mixture using six different sets of primers: the forward primers 1–6 carried sequences corresponding to various 5′-untranslated positions arranged in the upstream-to-downstream order, and the backward primer had a sequence of nt 1389–1370 in an exon part (see shaded regions in Fig. 1). PCR fragments were produced with the forward primers 4–6 but not with the forward primers 1–3. Nucleotide sequences of the PCR products were precisely the same as that of the genomic DNA (nt 446–1389) except the lack of the intron sequence present in the 3′-half (nt 1265–1358). It was thus concluded that the *CDC2b* coding portion totally contains three introns (I to III) as depicted in Fig. 2 (lower part). Based on the size (1.4 kb) of *CDC2b* mRNA previously estimated from Northern blot [4] together with the *CDC2b* exon–intron organization, the 5′-end of *CDC2b* mRNA can be deduced to be around nt 400. This assignment is consistent with the above experimental results that no PCR fragments were made with the forward primer 3 (nt 332–351) and more upstream primers. However, we can not completely rule out, at present, another possibility, that an additional intron exists immediately upstream of the forward primer 4 site and thereby the actual transcriptional initiation site is far upstream of nt 400.

As can be seen in Fig. 2, the intron sites within the

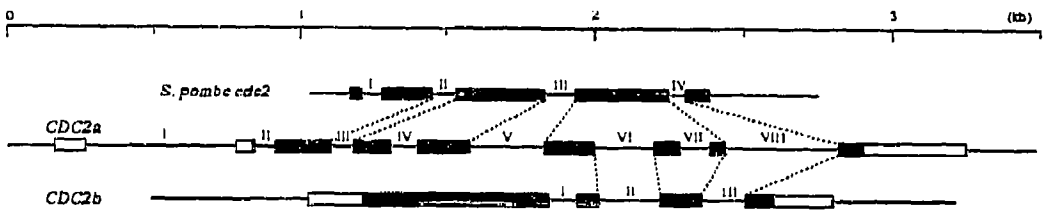

Fig. 2. Schematic exon–intron organization of *CDC2a* and *CDC2b* of *A. thaliana* and *S. pombe cdc2*. At the top is the scale in kb. Bars refer to exons, in which the coding regions are shaded. Lines indicate introns (named by Roman numerals) or neighboring DNA regions. Dotted lines show the corresponding positions of exon–intron junctions. The organization of *S. pombe cdc2* is based on the previous data [10], and its exons do not incorporate unclear 5′- and 3′-untranslated regions.

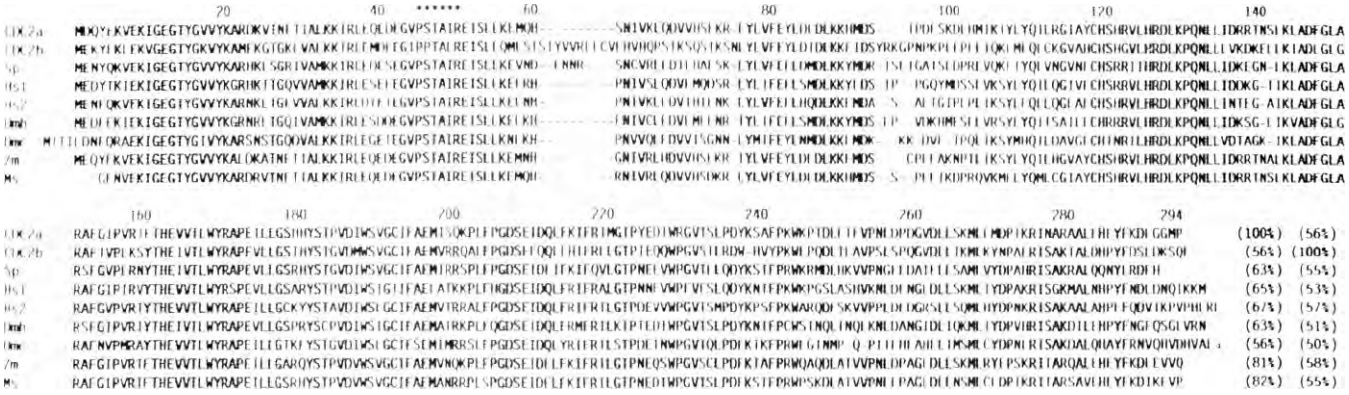

Fig. 3. Amino acid sequence comparison of various p34^{cdc2} homologs. CDC2a and CDC2b, *A. thaliana* p34^{CDC2a} and p35^{CDC2b} ([4] and this paper); Sp, *S. pombe* [10]; Hs1 and Hs2, *Homo sapiens* p34^{CDC2} [8] and p34^{CDK2} [14]; Dmh and Dme, *Drosophila melanogaster* functional homolog and non-functional cognate variant [9]; Zm, *Zea mays* [6]; Ms, alfalfa (*Medicago sativa*) [7]. All of these except *A. thaliana* p35^{CDC2b} and Dme are functional homologs. Hyphens are introduced for the best matching. Shaded letters indicate identity in at least 6 of the 9 sequences. The P-S-T-A-I-R motif is marked by asterisks, and the C-terminal 16 residues of Dme are simplified by a double arrow. Similarity of each p34^{cdc2} with *A. thaliana* p34^{CDC2a} or p35^{CDC2b} is represented by % identical residues in parentheses.

A. thaliana CDC2a and *CDC2b* genes and the *S. pombe cdc2* gene partly overlapped with one another: the position of *CDC2b* intron III coincided with those of both *CDC2a* intron VIII and *S. pombe* intron IV, and the *CDC2b* intron II and the *CDC2a* intron VI were at the same site, nevertheless their lengths and sequences were not conserved. This organization feature definitively supports the view that these three genes have evolved from a common ancestral gene.

3.3. *Functional difference between the* CDC2a *and* CDC2b *gene products*

From the exon–intron organization of *CDC2b*, we can deduce the amino acid sequence (308 residues) of the *CDC2b* gene product (p35^{CDC2b}). Its amino acid sequence is closely related to those of various p34^{cdc2} kinases (Fig. 3). However, this similarity (50–58% identical) is slightly less than that among p34^{cdc2} functional homologs (63–67% identical except for the 81–82% sim-

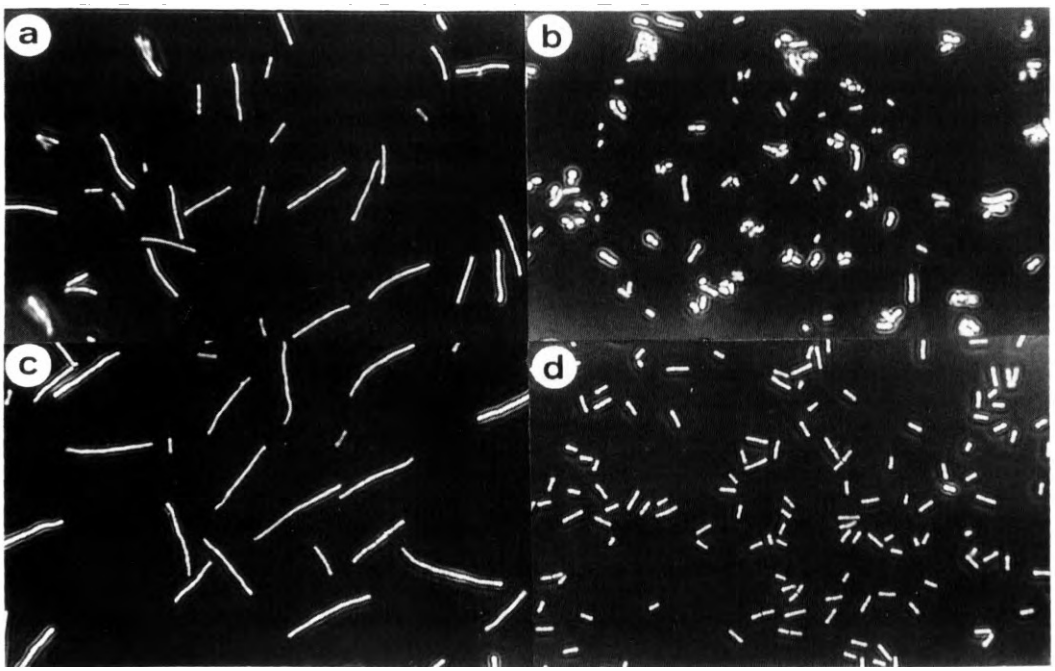

Fig. 4. Morphology of *S. pombe cdc2* mutant cells. (a), (b), and (c) show cells carrying pIMA1 (*CDC2b*), pNH290 (*CDC2a*), and pREP1, respectively, which have been incubated at the restrictive temperature under inducing conditions. (d) Presents pREP1 carriers cultivated at the permissive temperature under inducing conditions.

ilarity among plant functional homologs). The low value of similarity appears to be comparable to the instance of a non-functional cognate variant identified with *Drosophila melanogaster* (Dmc in Fig. 3) [9]. In addition, the $p34^{cdc2}$ hallmark P-S-T-A-I-R motif is somewhat modified to P-P-T-A-L-R in $p35^{CDC2b}$. These facts imply that *CDC2b* function is different from that of *CDC2a*.

To verify the functional discrimination of *CDC2b* from *CDC2a* by complementation tests with yeast *cdc2*, a cDNA sequence encoding the complete $p35^{CDC2b}$ was constructed by recombining a genomic *Ase*I–*Sac*I fragment for the 5′-half (nt 561–1190) with a cDNA fragment carrying the *Sac*I site to the poly(A) tail region for the 3′-half (corresponding to nt 1190–2193). This cDNA was inserted in the pREP1 vector, giving pIMA1. Upon incubation of the *S. pombe* $cdc2^{ts}$ strain harboring this plasmid at the restrictive temperature under inducing conditions, no significant increase in cell number was detected, and cells showed the elongated shapes characteristic of *cdc2* mutant cells (Fig. 4a), as in the case of pREP1 carriers (Fig. 4c). By contrast pNH290 (*CDC2a*) carriers had non-elongated figures (Fig. 4b), similar to those of pREP1 carriers grown at the permissive temperature (Fig. 4d). These results indicate that *CDC2a* is a functional homolog, confirming the previous results [4], and that *CDC2b* is a nonfunctional variant in the *cdc2* role. Though the actual role of *CDC2b* in *A. thaliana* is unknown at present, its structural similarity to the functional homolog *CDC2a* may intimate some possibilities: (i) $p35^{CDC2b}$ controls the cell division cycle at a point different from the G2–M transition (and possibly also the start point) regulated by $p34^{CDC2a}$; (ii) $p35^{CDC2b}$ operates in specific tissues distinct from those for $p34^{CDC2a}$; and (iii) $p35^{CDC2b}$ controls the cell cycle by modulating $p34^{CDC2a}$ activity through competitive interaction with its partners such as cyclins.

Acknowledgements: This work was supported in part by a Grant-in-Aid for Scientific Research from the Ministry of Education, Science, and Culture of Japan.

REFERENCES

[1] Nurse, P. (1990) Nature 344, 503–508.
[2] Lewin, B. (1990) Cell 61, 743–752.
[3] Nurse, P. and Bissett, Y. (1981) Nature 292, 558–560.
[4] Hirayama, T., Imajuku, Y., Anai, T., Matsui, M. and Oka, A. (1991) Gene 105, 159–165.
[5] Ferreira, P.C.G., Hemerly, A.S., Villarroel, R., Van Montagu, M. and Inzé, D. (1991) Plant Cell 3, 531–540.
[6] Colasanti, J., Tyers, M. and Sundaresan, V. (1991) Proc. Natl. Acad. Sci. USA 88, 3377–3381.
[7] Hirt, H., Páy, A., Györgyey, J., Bakó, L., Németh, K., Bögre, L., Schweyen, R.J., Heberle-Bors, E. and Dudits, D. (1991) Proc. Natl. Acad. Sci. USA 88, 1636–1640.
[8] Lee, M.G. and Nurse, P. (1987) Nature 327, 31–35.
[9] Lehner, C.F. and O'Farrell, P.H. (1990) EMBO J. 9, 3573–3581.
[10] Hindley, J. and Phear, G.A. (1984) Gene 31, 129–134.
[11] Lörincz, A.T. and Reed, S.I. (1984) Nature 307, 183–185.
[12] Feiler, H.S. and Jacobs, T.W. (1991) Plant Mol. Biol. 17, 321–333.
[13] Elledge, S.J. and Spottswood, M.R. (1991) EMBO J. 10, 2653–2659.
[14] Ninomiya-Tsuji, J., Nomoto, S., Yasuda, H., Reed, S.I. and Matsumoto, K. (1991) Proc. Natl. Acad. Sci. USA 88, 9006–9010.
[15] Frischauf, A.-M., Lehrach, H., Poustka, A. and Murray, N. (1983) J. Mol. Biol. 170, 827–842.
[16] Maundrell, K. (1990) J. Biol. Chem. 265, 10857–10864.
[17] Sanger, F. (1981) Science 214, 1205–1210.
[18] Brown, J.W.S. (1986) Nucleic Acids Res. 14, 9549–9559.