

Can sequencing shed light on cell cycling?

Andrew W. Murray^{*†} & Debora Marks[‡]

^{*}Center for Genomics Research, and [†]Molecular and Cellular Biology, Harvard University, Cambridge, Massachusetts 02138, USA

[‡]BCMP and Cell Biology, Harvard Medical School, Boston, Massachusetts 02115, USA

Every organism must have cells that can replicate indefinitely. Can the draft human genome sequence tell us how the cell cycle works and how it evolved? We studied two protein families—the cyclins and their partners the cyclin-dependent kinases (Cdks)—and a conserved regulatory circuit, the spindle checkpoint. Disappointingly, we discovered a few novel cyclins and no new Cdks or components of the spindle checkpoint, and could shed little light on the organization of the cell cycle.

Oscillations in the activity of the Cdk family of protein kinases drive the eukaryotic cell cycle. These enzymes consist of a catalytic (Cdk) and a regulatory subunit (cyclin)¹. Their activity is regulated by phosphorylation of Cdks, binding of stoichiometric kinase inhibitors and regulated proteolysis of the cyclins. Cdk activity is regulated by events inside and outside the cell and controls DNA replication, chromosome segregation and cell division. Despite wide variation in the speed and appearance of the cell cycle, the biochemical engine that regulates it has been remarkably conserved.

Cdks and cyclins

The Cdk/cyclin family contains five Cdks (Cdk1, 2, 3, 4 and 6) and four cyclin classes (cyclin A, B, D and E) that regulate the cell cycle; three Cdks (Cdk7, 8 and 9) and three cyclin classes (cyclin C, H and T) that regulate transcription; and other Cdk and cyclin classes with less defined roles (see Supplementary Information). We searched the predicted human proteins for Cdks and cyclins. We found no new Cdks. This is not surprising, as many labs have used polymerase chain reaction from conserved sequences to search for Cdks. Their success indicates that all the members of a highly conserved, well studied family are likely to be known before a genome is fully sequenced.

The draft human genome does answer one question about Cdk activation. Activating the cell-cycle Cdks requires phosphorylation of a conserved threonine as well as binding of cyclin. In mammals, the activating kinase is another Cdk/cyclin complex (Cdk7/cyclin H), in budding yeast it is a non-Cdk kinase (Cak1)¹, and in fission yeast both types of kinase are present and can activate Cdk/cyclin

complexes². There is no homologue of Cak1 in the human genome, suggesting that humans and budding yeast use different kinases for the same reaction. It appears that this step is not an important way of regulating Cdk activity, as no physiological regulation of Cdk7/cyclin H or Cak1 has been found.

The sequences of the cyclins are less well conserved than those of the Cdks. Cyclins have been found by three approaches: characterizing proteins that regulate processes such as the cell cycle or transcription; sequencing genes that are regulated in interesting ways; and as components of whole genome sequences. We searched the predicted proteins of the human genome for homology to known cyclins from humans and yeasts, focusing on genes that were annotated as novel. These candidates were compared against several databases, eliminating those that were fragments of known human cyclins. The remainder included the human homologue of chicken cyclin B3 and three novel cyclins. First, cyclin P, whose sequence is deduced from a combination of gene prediction on the genomic sequence and the sequence of a complementary DNA clone. It is related to but distinct from the A and B cyclins (see Fig. 1a and Supplementary Information). Second, a new cyclin, cyclin M, which appears to be most closely related to cyclin L (Fig. 1b). Its biological function is unknown, but it is related to cyclins that regulate transcription. Third, a protein previously reported as uracil-DNA glycosylase 2, whose similarity to existing cyclins has been noted³. Because a recombinant protein has not been shown to have glycosylase activity, this protein may be a novel glycosylase-associated cyclin rather than a catalytically active glycosylase. We suggest that this protein be called cyclin O (Fig. 1a).

We also identified possible vertebrate homologues of the large Pcl



Figure 1 Sequence alignments. **a**, Alignments of the most conserved regions of the cyclin box for A and B type cyclins. The alignments of cyclin O (Genbank accession no. NM_021147) and P (IPI protein set¹⁰ accession number IGI_M1_ctg17876_3) are shown beneath. Green letters show residues conserved in vertebrate A and B type cyclins. **b**, Alignment of the maximally conserved region of the transcriptional cyclins. Residues that are identical or conservative substitutions between cyclin L1 and cyclin M (IPI protein set¹⁰ accession number IGI_M1_ctg16669_8) are shown in red, as are residues that also show conservation in cyclin I and K. IPI, <http://www.ensemble.org/IPI/>

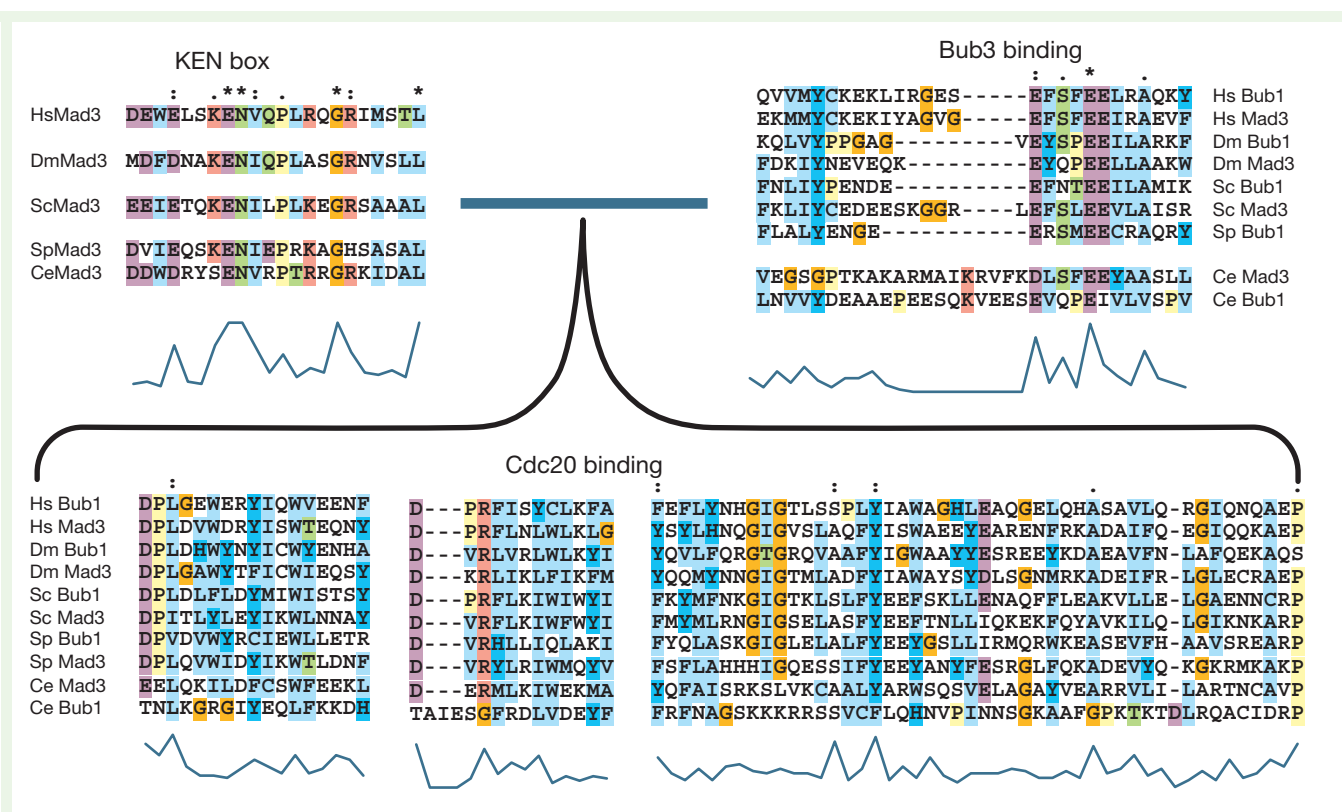


Figure 2 Sequence alignment, created using Clustal X, of portions of the Bub1 and Mad3 proteins. Includes the KEN box uniquely present in the N termini of Mad3 proteins, three stretches of conserved amino acids in the region that binds Cdc20, and conserved sequence in the region that binds Bub3. Note the divergence of the worm sequences from those of the other eukaryotes. Note that the *Drosophila* Mad3 protein is labelled as Bub1 in databases and vice versa.

family of yeast cyclins⁴. A small set of related human proteins encodes proteins with very strong homology to fly and worm proteins, and the proteins from all three organisms show weak homology to the yeast Pcl proteins (see Supplementary Information), but it is unknown whether this homology has biological significance.

Sequence analysis can only hint at the functions and Cdk partners of these cyclins. A first hypothesis is that the novel cyclins will have similar functions to their closest relatives; the closer the relationship, the more likely this is to be true. The structure of one Cdk/cyclin complex is known⁵, so it should be possible to model the pairwise interactions of all Cdks with all cyclins and deduce which Cdks partner a given cyclin by looking for pairs whose hypothetical interface lacks either steric clashes or unfilled space.

The spindle checkpoint

The spindle checkpoint monitors the kinetochore, the proteinaceous complex that assembles on the centromeric DNA and attaches chromosomes to the microtubules of the spindle. Kinetochore that are not correctly attached to microtubules recruit the components of the checkpoint, initiating a signalling pathway that inhibits the anaphase-promoting complex (the enzyme that triggers chromosome segregation and the exit from mitosis⁶). The checkpoint proteins and pathway have been conserved during evolution, and mutations in two of the proteins have been implicated in the chromosomal instability that is widespread in human cancers⁷.

We examined the genomes of completely sequenced organisms for homologues of five known checkpoint genes (*Mad1–3*, *Bub1*, *Bub3*). All the organisms contain a single copy of each gene, but there were some surprises. The first is that the checkpoint proteins in worms have diverged more from those of yeasts, flies and mammals than the mammalian or fly proteins have diverged from

those of yeasts. In the worm, protein motifs that are conserved among other organisms have diverged or been lost altogether (Fig. 2). We speculate that this deviation reflects the unusual behaviour of worm chromosomes. In most eukaryotes, from yeast to human, each chromosome has only one functional kinetochore; worms have an elongated array of kinetochores that spans the length of the chromosome. Adapting to this different architecture may have imposed special challenges on the spindle checkpoint, causing its components to evolve more rapidly in the nematode lineage.

Correctly identifying two of the checkpoint proteins, Mad3 and Bub1, was remarkably difficult. In budding yeast, where the proteins were discovered, they share two regions of homology that allow them to interact with Bub3 and Cdc20 (the target to which the checkpoint binds when it arrests the cell cycle), but Bub1 has a carboxy-terminal protein kinase domain that Mad3 lacks. Other organisms contain a sequence with a similar organization to Bub1, as well as a related protein that can be as large as Bub1 and carry a protein kinase domain (mammals, fly and worm) or be even shorter than the budding yeast protein (fission yeast).

Which protein is Bub1 and which Mad3? Attempts to deduce the overall phylogeny of these proteins give confusing results, as the most conserved regions of the amino termini of the budding yeast Bub1 and Mad3 proteins are more closely related to each other than they are to either protein in fission yeast (Figs 2, 3). This may indicate that the N termini have co-evolved with their binding partners, Cdc20 and Bub3. Despite this overall similarity, there is one distinguishing feature. In every organism, only one of the paired proteins contains a conserved sequence near the N terminus that matches the KEN box that targets proteins for destruction at the end of anaphase⁸. This is an attractive finding, as once anaphase begins, the tension at the kinetochores starts to fall. Destroying one of the checkpoint components would eliminate the possibility that the

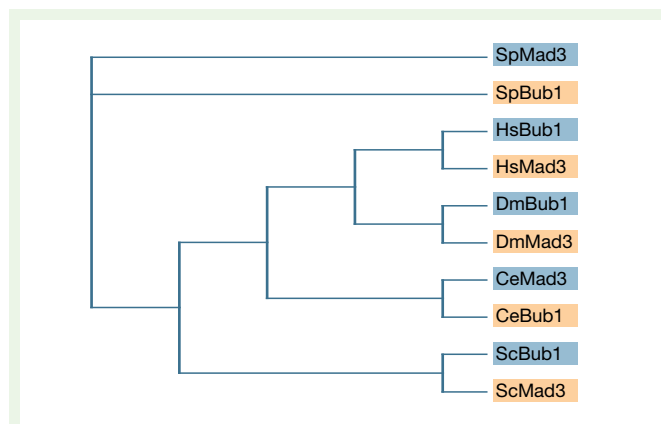


Figure 3 Phylogeny of the conserved regions of the N terminal portion of the Bub1 and Mad3 proteins. Sequences were aligned and trees created with Clustal X (<http://www-igbmc.u-strasbg.fr/BioInfo/ClustalX/>). Hs, *Homo sapiens*; Dm, *Drosophila melanogaster*; Sc, *Saccharomyces cerevisiae*; Sp, *Schizosaccharomyces pombe*; Ce, *Caenorhabditis elegans*.

checkpoint could be inadvertently reactivated, impeding the exit from mitosis. In both yeasts, this motif is in the shorter, kinase-deficient protein, suggesting that the KEN box identifies the Mad3 homologues in all eukaryotes.

This comparison illustrates two points. First, phylogenetic analysis on different regions of the same class of proteins can produce different conclusions about the relationships among the protein family, reflecting the different evolutionary pressures faced by different parts of the protein. Second, there can be large-scale reorganizations of components of a conserved pathway. In the case of the spindle checkpoint and Mad3, the functional consequences of these changes are unclear.

Conclusions

Genome sequencing has revealed surprisingly little about the cell cycle. The two main conclusions of comparative analysis were

drawn well before the first eukaryotic genome was sequenced: the machinery that regulates the cell cycle has been highly conserved in eukaryotic evolution, and the size of protein families such as the Cdks and cyclins has expanded as organisms have become bigger and acquired more cell types. Comparing sequenced genomes has strengthened but not altered these conclusions. Disappointingly, comparing fully sequenced genomes does not explain how differences have evolved between the cell cycles of different organisms. For example, a conserved pathway (the mitotic exit network) is required for cytokinesis in budding and fission yeast, but is required to complete mitosis in budding but not in fission yeast⁹. Men live, and must avoid cancer, for thirty times as long as mice, and it is much easier to induce mouse cells to proliferate indefinitely *in vitro*. Can comparing the sequence and expression of two genomes reveal the source of these differences? Getting a positive answer to such questions will require a marked improvement in our ability to turn raw sequence data into biological knowledge. □

1. Morgan, D. O. Cyclin-dependent kinases: engines, clocks, and microprocessors. *Annu. Rev. Cell. Dev. Biol.* **13**, 261–291 (1997).
2. Lee, K. M., Saiz, J. E., Barton, W. A. & Fisher, R. P. Cdc2 activation in fission yeast depends on Msc6 and Csk1, two partially redundant Cdk-activating kinases (CAKs). *Curr. Biol.* **9**, 441–444 (1999).
3. Muller, S. J. & Caradonna, S. Cell cycle regulation of a human cyclin-like gene encoding uracil-DNA glycosylase. *J. Biol. Chem.* **268**, 1310–1319 (1993).
4. Moffat, J., Huang, D. & Andrews, B. Functions of Pho85 cyclin-dependent kinases in budding yeast. *Prog. Cell Cycle Res.* **4**, 97–106 (2000).
5. Jeffrey, P. D. *et al.* Mechanism of CDK activation revealed by the structure of a cyclinA-CDK2 complex. *Nature* **376**, 313–320 (1995).
6. Amon, A. The spindle checkpoint. *Curr. Opin. Genet. Dev.* **9**, 69–75 (1999).
7. Cahill, D. P. *et al.* Mutations of mitotic checkpoint genes in human cancers. *Nature* **392**, 300–303 (1998).
8. Pfeiffer, C. M. & Kirschner, M. W. The KEN box: an APC recognition signal distinct from the D box targeted by Cdh1. *Genes Dev.* **14**, 655–665 (2000).
9. Jaspersen, S. L., Charles, J. F., Tinker-Kulberg, R. L. & Morgan, D. O. A late mitotic regulatory network controlling cyclin destruction in *Saccharomyces cerevisiae*. *Mol. Biol. Cell* **9**, 2803–2817 (1998).
10. International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Correspondence and requests for materials should be addressed to A. M. (e-mail: amurray@mcb.harvard.edu).