

support, the optimal motif phosphorylated by the kinase can be determined^{5–8}. This technique, however, has also proven problematic owing to nonspecific association of radiolabeled ATP or phosphate with the support, and artifactual results caused by noncovalent associations between the kinase and particular peptide sequences on the support, resulting in high local concentrations of nonideal substrates. Consequently, the optimal motif determined using this technique sometimes correlates poorly with the sequences surrounding actual *in vivo* phosphorylation sites.

The technique of Turk and colleagues uses the best aspects of both the solution-phase and the solid-support approaches, essentially overcoming the limitations of each. These investigators synthesized 198 separate degenerate peptide libraries, each containing a fixed serine or threonine residue and a second single fixed amino acid in any one of nine flanking positions (Fig. 1). This second fixed position included any one of all 20 naturally occurring amino acids, as well as phosphothreonine and phosphotyrosine. The remaining flanking positions contained a degenerate mixture of all the amino acids except for serine, threonine, tyrosine and cysteine.

Importantly, all of the libraries contain a C-terminal biotin tag so that they can be captured onto avidin-coated membranes after phosphorylation in solution. The use of solution-phase phosphorylation eliminates many of the artifacts seen with immobilized substrates, whereas the solid-phase peptide capture and quantification, which obviates the requirement for Edman sequencing, eliminates the problem of phosphopeptide purification and allows phosphorylated amino acids to be included in motif determination. Using this approach, Turk and colleagues were able to obtain and/or confirm motifs for the kinases PKC θ , PKD and type II TGF- β receptor kinase, and identify a new motif for the proto-oncogene kinase Pim2 in reasonably short order.

What comes next? This technique should open the door to determining the motifs for the entire 'kinome' of selected organisms. There is often a general correlation between how closely the sequence of a substrate matches the optimal motif for a particular kinase and how well it is phosphorylated. Information about the optimal phosphorylation motif for a protein kinase derived by the Cantley and Turk group's approach, particularly when combined with bioin-

formatic analysis⁹, yeast two-hybrid data, BIND database annotation¹⁰ and global protein subcellular localization data, should help researchers to rapidly build and annotate signaling pathways in the future. Even exceptions to the peptide library motif rules may prove biologically enlightening—that is, substrates that are subsequently shown to be phosphorylated *in vivo* on sequences that poorly match the optimal peptide motifs would imply that other proteins and/or domains are likely to modulate the phosphorylation process. This new combined solution- and solid-phase peptide library technique promises to be a central pillar upon which proteomic studies of cell-signal transduction can be built.

1. Hutti, J.E. *et al. Nat. Methods* **1**, 27–29 (2004).
2. Kemp, B.E., Bylund, D.B., Huang, T.S. & Krebs, E.G. *Proc. Natl. Acad. Sci. USA* **72**, 3448–3452 (1975).
3. Daile, P., Carnegie, P.R. & Young, J.D. *Nature* **257**, 416–418 (1975).
4. Songyang, Z. *Curr. Biol.* **4**, 973–982 (1994).
5. Wu, J., Ma, Q.N. & Lam, K.S. *Biochemistry* **33**, 14825–14833 (1994).
6. Rychlewski, L., Kschischo, M., Dong, L., Schutkowski, M. & Reimer, U. *J. Mol. Biol.* **336**, 307–311 (2004).
7. Rodriguez, M., Li, S.S., Harper, J.W. & Songyang, Z. *J. Biol. Chem.* **279**, 8802–8807 (2004).
8. Luo, K., Zhou, P. & Lodish, H.F. *Proc. Natl. Acad. Sci. USA* **92**, 11761–11765 (1995).
9. Yaffe, M.B. *et al. Nat. Biotechnol.* **19**, 348–353 (2001).
10. Bader, G.D. *et al. Nucleic Acids Res.* **29**, 242–245 (2001).

Centromeric heterochromatin comes clean with DNA methylation

Peter E Warburton

A sequence-independent procedure for rapid identification of plant centromeric DNA provides a new tool for genome assembly and genomic evolution studies.

In this exciting new era of genomics, the availability of DNA sequence from a growing number of eukaryotic genomes promises to revolutionize biology. However, despite our increasing knowledge about genomic maps, large gaps of unknown DNA sequence still remain in these 'finished' genomes. These gaps correspond to blocks of heterochromatin, which typically contain large amounts of repetitive DNA, such as tandem satellites and transposed elements, that confound sequencing and assembly efforts^{1,2}. The work of Preuss and colleagues³ in this inaugural issue of *Nature Methods* provides a method that could help fill these gaps and allow heterochromatin to take its rightful place in genome maps.

Embedded in heterochromatic regions are the centromeres, the parts of chromosomes responsible for proper segregation during cell division and crucial to the stability and genetic inheritance of the genomes in which

we are so interested. Adding to the problem of sequencing heterochromatic regions is the remarkably rapid evolution and lack of conservation of the repetitive DNA found at centromeres of different species, which make the identification of centromeric DNA extremely tenuous^{4,5}.

In a method that is elegant in its simplicity, Preuss and colleagues set out to rapidly identify centromeric DNA from plant genomes not by the DNA sequence itself but by the extent of cytosine methylation, a secondary epigenetic modification of heterochromatin DNA that is conserved across many species. Using *Arabidopsis thaliana* as a model system, they digested genomic DNA with the methyl-sensitive enzyme *HpaII*, which cuts DNA only at unmethylated CCGG recognition sites. Two fractions of >10 kb (*HpaII* resistant, densely methylated) and <3 kb (*HpaII* sensitive, sparsely methylated) were used for differential

Peter E Warburton is in the Department of Human Genetics at the Mount Sinai School of Medicine, New York, New York 10029, USA.
e-mail: peter.warburton@mssm.edu

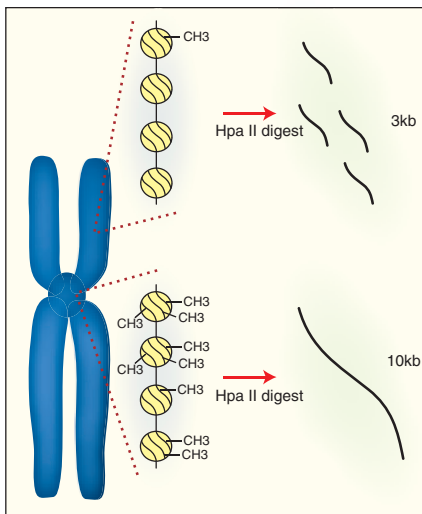


Figure 1 | Differences in DNA methylation levels allow for identification of centromeric regions.

hybridization to genomic DNA arrayed in bacterial artificial chromosome (BAC) libraries, which identified about 7% of BACs as containing islands of dense methylation (Fig. 1). Because the centromeric DNA of *A. thaliana* chromosomes had been previously characterized through the painstaking efforts of several laboratories over several years⁵, the putative centromeric BACs were rapidly verified, validating the method.

These BACs primarily comprised 178-bp centromeric satellite repeats, 5S rDNA repeats and sequences from the flanking pericentromeres, including the heterochromatic knob on chromosome 4. Importantly, this method could distinguish the centromeres from other highly repetitive but less methylated DNA, such as the nucleolar organizing regions, and from the remainder of the genome. The authors then extended the method to other species of the Brassicaceae family, whose genomes are only poorly characterized. In these species, too, BACs containing both centromeric satellite DNAs and pericentromeric sequences were identified and confirmed, demonstrating the potential of this new method to facilitate the inclusion of centromeric regions into genome assemblies.

The application of this method to Brassicaceae, which have diverged from the Arabidopsis family 10–14 million years ago, also illustrate how powerful a tool this method can become in the hands of biologists studying the structure and evolution of centromeres. The rapid identification of these sequences will permit comparative studies across species, which may uncover

key DNA elements and their relationships with rapidly evolving centromere proteins that ensure centromere function in the face of rapid evolution⁴. These studies may also permit more rapid isolation and better understanding of candidate centromere DNA sequences for use in the construction of artificial chromosomes.

Will this method, shown to work in plants, be applicable to the genomes of mammalian species such as mouse, rat or human? Will the extent of cytosine hypermethylation in mammalian heterochromatin, and particularly centromeres, allow a similar approach? Mammalian and plant centromeres share many similarities, such as the presence of tandem satellite DNA flanked by transposon-rich pericentromeric regions^{1,2}. In both plant and mammalian heterochromatin, it is becoming increasingly clear that cytosine methylation is linked with several other epigenetic modifications, such as histone lysine hypoacetylation and histone H3 methylation at lysine 9 (H3-mK9). The establishment and maintenance of these combinations of epigenetic marks are mediated by multiprotein complexes which appear to form reinforcing interdependent feedback loops that functionally define the transcriptionally silent and condensed heterochromatic state⁶. For example, in *A. thaliana*, mutations in the *DDM1* gene (decrease in DNA methylation), a nucleosome remodeling factor of the SWI2-SNF2 class, leads to a severe reduction in cytosine methylation specifically in the repetitive DNA sequences, probably through the depletion of H3-mK9. Although there are similarities in heterochromatin across kingdoms, plants do possess some unique DNA methyltransferases for which equivalents have not yet been found in mammals. Among these are the chromodomain-containing chromomethyltransferases, which may link cytosine methylation to histone modifications⁶. In addition, plants possess an RNA-directed DNA methylation mechanism, which may act through transcribed transposable elements⁷. Hence there may be some differences in the regulation of DNA methylation in the heterochromatin of plants and mammals.

The extent of cytosine methylation at mammalian centromeres also remains somewhat unclear. Several human chromosomes contain large blocks of heterochromatin composed of heavily methylated classical satellites. Upon mutation of the human methyltransferase DNMT3B⁸,

these classical satellites become severely hypomethylated and chromosomes become unstable. However, given that these classical satellites are near, but distinct from, human centromeres, it is unlikely that the observed chromosome instability results directly from defects in centromere function. In fact, the presence of at least some unmethylated DNA in the centromeres may also be important, because there is evidence that centromere protein B (CENP-B), an abundant, conserved centromeric satellite DNA-binding protein, will recognize its binding site only when it is demethylated⁹. Analysis of human centromeric alpha-satellite DNA showed a variable cytosine methylation profile suggestive of a distinct domain organization¹⁰. Furthermore, both plant and animal centromeres contain domains of distinct kinetochore-forming chromatin in which histone H3 is replaced with a homolog (CENP-A in mammals). The methylation state of these domains is not known⁵. Thus, although mammalian heterochromatin is heavily methylated in general, the extent to which a method relying on cytosine methylation-based whole-genome fractionation will identify certain functional domains within the centromeres, such as kinetochore regions, remains somewhat unclear.

Nevertheless, the method devised by Preuss' group will identify BACs that contain heavily methylated DNA over a relatively long range. These will certainly represent important missing pieces for finishing the puzzle presented by the large heterochromatin gaps in our current genome assemblies. This method will facilitate the actual 'finishing' of our current genome assemblies, which will provide great insight into the evolution, regulation and maintenance of complex eukaryotic genomes².

- Schueler, M.G., Higgins, A.W., Rudd, M.K., Gustashaw, K. & Willard, H.F. *Science* **294**, 109–115 (2001).
- She, X. *et al. Nature* **430**, 857–864 (2004).
- Luo, S., Hall, A.E., Hall, S.E. & Preuss, D. *Nat. Methods* **1**, 67–71 (2004).
- Malik, H.S. & Henikoff, S. *Curr. Opin. Genet. Dev.* **12**, 711–718 (2002).
- Hall, A.E., Keith, K.C., Hall, S.E., Copenhagen, G.P. & Preuss, D. *Curr. Opin. Plant Biol.* **7**, 108–114 (2004).
- Richards, E.J. *Curr. Biol.* **12**, R694–R65 (2002).
- Lippman, Z. *et al. Nature* **430**, 471–476 (2004).
- Bestor, T.H. *Hum. Mol. Genet.* **9**, 2395–2402 (2000).
- Mitchell, A.R., Nicoll, L. & Kipling, D.J. *Cell Sci.* **109**, 2199–2206 (1996).
- Miniou, P. *et al. Hum. Genet.* **99**, 738–745 (1997).