Genomes & Developmental Control

Developmental timing in *Dictyostelium* is regulated by the Set1 histone methyltransferase


Received for publication 3 November 2005; accepted 30 December 2005

Available online 15 February 2006

**Abstract**

Histone-modifying enzymes have enormous potential as regulators of the large-scale changes in gene expression occurring during differentiation. It is unclear how different combinations of histone modification coordinate regimes of transcription during development. We show that different methylation states of lysine 4 of histone H3 (H3K4) mark distinct developmental phases of the simple eukaryote, *Dictyostelium*. We demonstrate that the enzyme responsible for all mono, di and tri-methylation of H3K4 is the *Dictyostelium* homolog of the Set1 histone methyltransferase. In the absence of Set1, cells display unusually rapid development, characterized by precocious aggregation of amoebae into multicellular aggregates. Early differentiation markers are abundantly expressed in growing *set1* cells, indicating the differentiation program is ectopically activated during growth. This phenotype is caused specifically by the loss of Set1 catalytic activity. *Set1* mutants induce premature differentiation in wild-type cells, indicating Set1 regulates production of an extra-cellular factor required for the correct perception of growth conditions. Microarray analysis of the *set1* mutants reveals genomic clustering of mis-expressed genes, suggesting a requirement for Set1 in the regulation of chromatin-mediated events at gene clusters.

© 2006 Elsevier Inc. All rights reserved.

**Keywords:** Chromatin; Dictyostelium; Set1; Lysine 4 methylation; Gene cluster; Clock

**Introduction**

Spatially and temporally coordinated gene expression requires structural reorganization of chromatin architecture, to permit or block access of regulatory proteins to DNA. To remodel chromatin architecture during development, eukaryotic cells use enzymes that covalently modify histones by phosphorylation, acetylation, methylation and ubiquitinylation. Another mechanism for changing chromatin involves the deposition of histone variants, which displace existing histones, and the modifications they carry. In addition, there are ATP-hydrolyzing chromatin-remodeling enzymes capable of sliding or displacing nucleosomes.

A key question is how chromatin modifications operate to regulate developmental gene expression. A chromatin-modifying enzyme could act at individual genes that are dispersed in...
the genome, or might act locally, for instance at co-regulated gene clusters. Indeed, large proportions of the genomes of eukaryotes consist of clusters of co-expressed genes (Boutanaev et al., 2002; Lercher et al., 2002; Spellman and Rubin, 2002). Some modifications spread locally along chromatin, in part due to the self-interaction of chromatin proteins and the potential for some modifying enzymes to interact with their cognate modification while marking nearby nucleosomes (Danzer and Wallrath, 2004). Other modifications are more punctate in their genomic distribution (Bernstein et al., 2005), so their effects could be distributed throughout the genome, rather than involving specialized chromatin events at gene clusters. One view is that a chromatin modification or combination of modifications marks a battery of genes specific to a particular developmental program of cellular differentiation (Kurdistani et al., 2004). This reflects the histone code hypothesis (Strahl and Allis, 2000), which takes into account the large number of post-translational modifications of histones and the combinatorial potential of these modifications to specify different programmes of gene expression.

To address the requirement for different chromatin modifications in the regulation of developmental gene expression ideally requires a genetically tractable developmental model. Genetic studies in higher eukaryotes have proven difficult to interpret because of the complexity of these developmental systems. In contrast, analysis in yeast is hampered by a lack of developmental options. We therefore decided to address the role of histone modification in development using the simple eukaryote Dictyostelium. It has a simple and well-defined developmental programme, and a small and fully sequenced genome (Eichinger et al., 2005) that encodes chromatin proteins and modifying enzymes conserved with those in higher eukaryotes. In addition, the organism is highly amenable, with powerful haploid molecular genetics (De Lozanne and Spudich, 1987). Individual Dictyostelium cells enter their program of differentiation upon nutrient depletion. Upon starvation, the cells aggregate, using cAMP as a chemoattractant, into a multicellular mound. After a series of morphogenetic steps, cells in the aggregate adopt predominantly only two developmental fates. The final developed structure consists of a spore head, containing approximately 80% of the cells, held over the substrate by a thin stalk structure, containing the remaining 20% of cells.

In this study, we have analyzed the developmental role of the histone methyltransferase (HMTase) required for methylation of the lysine 4 residue of histone H3 (H3K4). Methylation of H3K4 is associated with active chromatin in a wide range of eukaryotes. Lysine methylation of histones can occur in three forms: mono-, di- and tri-methylated. We have identified the Dictyostelium homologue of Set1, and show that set1 mutant cells completely lack mono, di and tri-methylation of H3K4. Loss of Set1 gives rise to cells displaying unusually rapid development, characterized by precocious aggregation of cells into multicellular aggregates. Early starvation markers are abundantly expressed during growth, indicating the mutants inappropriately enter the differentiation program, even in the presence of nutrients. We demonstrate a strong non-autonomous component to the developmental timing phenotype, indicating Set1 regulates production of an extra-cellular factor required for the correct perception of growth conditions. Microarray analysis of the set1 mutants indicates many of the mis-expressed genes are tightly grouped in the genome, suggesting Set1 may act to regulate chromatin-mediated events at gene clusters.

Materials and methods

Disruption of the set1 locus

Dictyostelium AX2 cells and their derivatives were grown, developed and transformed as described (Chubb et al., 2000). To disrupt the set1 locus, we inserted a blasticidin resistance cassette into the SwaI site of Dictyostelium genomic clone JC1b237c03, which spans the 3’ end of the set1 gene. The SwaI site is upstream of the SET domain, so insertion renders the Set1 protein catalytically inactive. BglII linkers were inserted into the SwaI site, and a BamHI fragment containing the blasticidin resistance cassette was inserted into the linkered site. A targeting fragment was released by digestion of BamHI and KpnI sites in the polylinker. To assemble a Set1 genomic clone, we inserted the SacI/ClaI fragment from Dict_I_V660d03 into Dict_IV_V649d06 and the Sty/EcoRI fragment from Dict_IV_V71c10 into Dict_IV_V590g11. A KpnI/SpeI fragment from the first product and a SpeI/EcoRI fragment from the second, were cloned together into EcoRUKpnI-digested pBluescriptII. Catalytic dead N1425Q and C1474A mutations were generated by PCR, then spliced into the full genomic clone using ClaI/acet digestions. For expression of these clones in cells, we place an EcoRI linker into the Nhel site just upstream of the Set1 ATG, then digested out the full length set1 clones, using EcoRI, into the pDEXH 82 vector, 3’ of the GFP coding sequence, thus generating GFP-Set1 fusions.

Analysis of cellular DNA, RNA and protein

Southern, Northern and Western blotting was carried out as described (Chubb et al., 2000). For confirming disruption of the set1 locus, we Bell-digested genomic DNA from transformants and probed Southern blots with a BamHI/KpnI fragment from JC1b237c03. For Northern, the following probes were used: car1 and ACA (from cDNA clones), rasG and V18 (PCR products) and discoidin Ia (from clone SLB855). RNA FISH experiments were carried out as described (Femino et al., 1998) with cells plated on acid-washed coverslips fixed in 4% paraformaldehyde prior to lysis with 0.5% triton.
For Western blotting, we used rabbit antisera against histone H3 and their modifications. We used a commercial C-terminal histone H3 antibody (ab1791, Abcam) at 1:2000 dilution. Antiserum to mono-, di- or tri-methylated H3K4 were raised in rabbits by immunization with linear synthetic peptides conjugated to ovalbumin, using the procedures described by White et al. (1999). The peptide sequence used for immunizations corresponded to the ten N-terminal residues of H3 and incorporated mono-, di- or tri-methyl lysine at position 4 (A.R.T.mek.Q.T.A.R.K.S.C.). The C-terminal cysteine was added to facilitate coupling to affinity gels. Specificity was tested by inhibition ELISA (White et al., 1999). By this criterion, each of the antisera used here was specific for one of the three possible H3K4 methylation states (Breiling et al., 2004). The H3K4 antisera were used at dilution of 1:1000. GFP was detected with the JL-8 monoclonal antibody (Sigma). We used HRP-conjugated anti-secondary antibodies (Jackson) for detection.

Immuno-fluorescence on H3 modifications was carried out on cells fixed in 4% paraformaldehyde (K4) or 3.7% formaldehyde (K9) on acid-washed coverslips. Lysis was carried out using 0.2% Triton X100 and antibodies were used at 1:500. For K9 di-methylation, we used a rabbit IgG directed against the H3K9me2 epitope (Upstate). A Cy3-conjugated anti-rabbit secondary was used for detection. ChIP assays were carried out using anti-H3K9me2 sera from Abcam.

ChIP assays

10^6 Cells were resuspended in 50 ml KK2 buffer (20 mM potassium phosphate, pH 6.2), formaldehyde was added to 1.2% and the mixture was shaken for 15 min. Fixation was quenched with 360 mM glycine for 5 min. Cells were pelleted, washed in 10 ml RLB lysis buffer (0.32 M sucrose, 10 mM Tris 7.5, 5 mM MgCl2, 1% Triton X-100) and antibodies were used at 1:500. For K9 di-methylation, we used a rabbit IgG directed against the H3K9me2 epitope (Upstate). A Cy3-conjugated anti-rabbit secondary was used for detection. ChIP assays were carried out using anti-H3K9me2 sera from Abcam.

Genome-wide expression profiling

Total RNA was extracted from two separate vegetative samples each of two independent set1 mutant clones and the parental strain AX2 using RNeasy columns (Qiagen). 25 μg of each sample was primed with anchored oligo(dT) and separately labelled with Cy3 and Cy5 using Superscript III reverse transcriptase (Invitrogen). Each set of mutant labelled cDNA was then paired with the parental cDNA labelled with the complementary fluorophore, and the mixture hybridized to a DNA microarray. The microarrays comprised 7346 PCR products designed to be specific to predicted *Dictyostelium discoideum* genes, plus controls, printed in duplicate on Codelink slides (Amersham); the design and construction of the arrays will be described in detail elsewhere, and detailed protocols are available at http://www.sanger.ac.uk/PostGenomics/Pathogen-Arrays/protocols/Dicy. Images were acquired and quantified using a Genepix 4000B scanner and its associated software (Axon Instruments).

Data were read into the R environment and normalization and further analysis carried out using the Bioconductor package LIMMA (Smyth, 2004; Gentleman et al., 2004). Briefly, background fluorescence was subtracted using the method of Cooperberg et al. (2002) then the resulting intensities were adjusted using print-tip loss normalization. A simple linear model was fitted to estimate the log-ratios of each gene, comparing mutant with parent across all slides, and an empirical Bayes method was used to assess whether genes were differentially expressed. After ranking, genes with a P value less than 0.05 were provisionally accepted as having altered expression in the mutant cell lines. P values were adjusted for multiple testing using the procedure of Benjamini and Hochberg (1995).

Analysis of genomic clustering of mis-expressed genes

Grouping of affected genes within the genome was assessed separately using the CLAC method of Wang et al. (2005). Cluster trees were constructed along entire chromosomes without data smoothing (window size of one) for all four Set1 versus Ax2 slides, in comparison with three reference slides on which RNA from the parental strain labelled with both dyes had been hybridized. Clusters were reproduced across all mutant/parental slides and had an FDR score far less than 0.01 and thus accepted as significant.

We also investigated the distribution of significantly mis-expressed genes on chromosome 2. A 100-gene window was moved along chromosome 2 (1963 unique chromosome 2 genes on the array) to find regions on the chromosome enriched with mis-expressed genes (Cohen et al., 2000) in the set1 mutant. The probability of finding at least the number of target genes in a given window was calculated using the cumulative hyper-geometric distribution (Rice, 1995). At around 6.3 Mb, we found that 6 of the 32 target genes were in the 100-gene window (P < 0.005). To account for the multiple-testing problem, we performed a Monte Carlo simulation and verified the statistical significance of the clustering (P < 0.01). To get rid of the effect of possible recent gene duplication, we performed a BLASTN (Altschul et al., 1990) search on all genes on chromosome 2 (against themselves) and identified 69 families from 186 highly similar genes (E value < 1e-30). We then repeated the simulation by randomly picking one gene as the representative of each family.

Results

Chromatin proteins in *Dictyostelium*

The *Dictyostelium* sequence databases reveal a large complement of chromatin proteins and histone modifiers known in higher eukaryotes. Linker histones (Hauser et al., 1995), histone deacetylases and acetyl transferases, ATP-dependent chromatin-remodeling enzymes, RNAi proteins (Martens et al., 2002), chromodomains, SET domain proteins and variant histones can all be found. The *Dictyostelium* genome encodes a homolog of the histone H3 lysine 9 (H3K9) methyltransferase su(var)3–9 (DDB0190352), a protein absent.
in the yeast \textit{S. cerevisiae}. Homologs of all core nucleosomal histones were identified. The \textit{Dictyostelium} H3 family comprises three genes with 85\% identity to human H3 (DDB0191157 and DDB0216291 — two adjacent genes) and two divergent variants. The three H3 genes similar to human H3 are identical in their extreme N-termini to the human H3 (N-ARTKQTARKSTG-). The H3 residues known in higher eukaryotes to be post-translationally modified (lysine, arginine and serine/threonine) are highly conserved. Substitutions in their N-terminal and histone fold domains indicate all three are H3.3-type histones.

The developmental regulation of H3K4 methylation in \textit{Dictyostelium}

We chose to study the modification of an amino acid in the conserved N-terminus of histone H3, methylation of H3K4. H3K4 can exist in three methylation states: mono-, di- and tri-methyl. In other models, meH3K4 is considered to be a mark of active chromatin.

If histone modification regulates batteries of genes associated with developmental transitions, we should see differences in the modification states of genomes at different developmental states. We addressed this hypothesis by looking at the total genomic levels of the different methylation states of H3K4 through the entire program of \textit{Dictyostelium} development, from growing cells to the mature fruiting body (Figs. 1A, B). The levels of H3, relative to total cellular protein, increase as differentiation proceeds towards the final culmination. This may reflect the loss of cytoplasmic contents that occurs during spore formation, or increased DNA packaging during dormancy. The levels of H3K4me3 diminish considerably during the process of differentiation. In contrast, the level of H3K4me1 becomes significantly enhanced during differentiation. We see a slight

Fig. 1. Different H3K4 methylation states mark distinct phases of differentiation. (A) Western blots of different methylation states of H3K4, of extracts taken throughout the development of wild-type \textit{Dictyostelium}. Equal amounts of protein were loaded. (B) Quantification of the intensity of the chemiluminescent signals from the blots in A. (C) H3K4me3 at the 5’ end of the ACA gene correlates with transcriptional activation (green). (D) H3K4me3 at the ras\textit{G} gene is lost as the gene is inactivated. Levels of the lower modification states increase as trimethylation disappears (H3K4me2 in red, H3K4me1 in black). (C and D) Amount of DNA immunoprecipitated by antibodies against meH3K4 was assessed by quantitative PCR using primers amplifying the regions marked by black bars. Gene expression is shown by Northern blots of RNA from the same cells.
but reproducible dip in H3K4me2 around the time of aggregation (8 h) and the level of this mark again dips during the final stages of spore formation (20–24 h).

In budding yeast and vertebrates, H3K4me3 is enriched at the 5′ ends of active genes (Ng et al., 2003; Bernstein et al., 2005). To validate the Dictyostelium for the study of H3K4 methylation, we analyzed H3K4Me at genes whose expression changes during Dictyostelium development, using chromatin immunoprecipitation (ChIP) assays and quantitative PCR. We see a strong correlation between H3K4 trimethylation and gene activation, in a manner similar to other eukaryotic models. The adenylate cyclase gene (ACA) is induced a few hours after the onset of differentiation (Pupillo et al., 1992), which parallels levels of H3K4me3 at the 5′ end of the gene (Fig. 1C). The level of enrichment of H3K4me3 2 kb downstream of the transcription start site is very low (Fig. 1C). We also see a correlation between gene activity and H3K4me3 for genes that are inactivated during early differentiation. The expression of the growth stage GTPase, rasG (Robbins et al., 1992), declines a few hours after the onset of differentiation, which correlates with a significant drop in the level of H3K4me3 at the rasG locus (Fig. 1D).

The levels of H3K4me1 and H3K4me2 rise during the inactivation of rasG that occurs after the onset of differentiation (Fig. 1D). The level of dimethylation at this locus peaks coinciding with the loss of H3K4me3. This enrichment of dimethyl H3K4 declines as the rise in the level of H3K4me1 continues. We see a similar trend for the growth gene V18 (data not shown, Singleton et al., 1989).

**Identification of the H3K4 HMTase in Dictyostelium**

The Set1 methyltransferase is required for di- and trimethylation of H3K4 in the yeast *S. cerevisiae* (Santos-Rosa et al., 2002). We searched Dictyostelium sequence databases with the Set1 sequence and identified a gene predicted to encode a 1486 amino acid protein with a C-terminus showing high homology to the catalytic (SET) domain of Set1 (DDB0188336, Fig. 2). Alignment of its SET domain with other SET proteins indicates it is member of the Set1 family and is distinct from other SET proteins (Fig. 2B). We named this gene *Dictyostelium set1*. The closest homolog to the Dictyostelium Set1 catalytic domain is the human Set1/Ash2 HMTase (KIAA0339, Fig. 3) a protein known to methylate H3K4 (Wysocka et al., 2003). Unlike some Set1 proteins, there is no RNA recognition motif (RRM). The Dictyostelium genome appears to encode at least 25 additional SET proteins.

To investigate the role of Set1 in chromatin regulation in Dictyostelium, we mutated the set1 locus by homologous recombination. A blasticidin resistance cassette (Sutoh, 1993) was integrated into the set1 coding sequence upstream of the SET domain, (Fig. 3A) confirmed by Southern analysis (Fig. 3B). We obtained a high frequency (95%) of recombinants carrying the disrupted allele, indicating Set1 is dispensable for...
cell viability. Western blots show that mono-, di- and trimethylation of H3K4 are absent throughout the life cycle of the mutant (Fig. 3C), confirmed by immunofluorescence (Fig. 3D). In the wild-type cells, all three meH3K4 marks distribute throughout the nucleus. In the mutant cells, all three modifications are absent. We can also demonstrate the loss of the H3K4 methylation by ChIP (data not shown). At the gross nuclear level, the distribution of dimethyl H3K9 is unchanged in the set1 cells (Fig. 3E). In both wild-type and mutant cells, H3K9me2 staining is localized to a single spot adjacent to the nuclear envelope.

Precocious development of set1 mutants

The growth rates of wild-type and set1 cells are not significantly different during short term culture, although when wild-type and set1 mutant cells were co-cultured for a week, wild-type cells were observed to slowly take over the culture (as monitored by immunofluorescence), suggesting the set1 mutants have a mild growth disadvantage. At the gross morphological level, Dictyostelium development is normal in the absence of Set1. When developed on non-nutrient agar, set1 mutants form normal-sized fruiting structures with a wild-type morphology (Supplementary Fig. 1).

The set1 mutants are accelerated in their early development (Fig. 4A). After nutrient removal, wild-type cells enter a program of differentiation and approximately 6 h later, chemotax to form a multicellular mound. Set1 cells aggregate approximately 2 h more rapidly than wild type. This result was seen for set1 clones derived from independent transformations (Fig. 4A).

Rapid development phenotypes are observed in two classes of Dictyostelium mutant. Some mutants activate the developmental gene expression program early or rapidly, and others...
have aberrant chemotactic signaling (Luo et al., 2003). We studied the RNA expression of two genes normally induced a few hours after nutrient removal, cAR1, which encodes an aggregation stage cAMP receptor (Sun and Devreotes, 1991) and ACA, which encodes the cAMP synthesis enzyme, adenylate cyclase (Pupillo et al., 1992). Both genes are induced precociously in the set1 mutant (Fig. 4B). ACA shows a strong induction by 2 h of development in the mutant, to a level comparable to that shown at 4 h in the wild-type strain. The expression of cAR1 RNA in the set1 mutant peaks early. We can also detect enhanced expression of ACA and cAR1 in set1 mutants at the single cell level, using RNA FISH on 4 h developed cells (Fig. 4C). In addition, the expression of the growth phase gene, V18 (Singleton et al., 1989), declines early in the set1 mutants (Fig. 4B). These data indicate the timing phenotype is caused by early or rapid activation of the differentiation program.

Ectopic activation of the differentiation program during growth

Set1 mutants could develop precociously because they proceed rapidly through the differentiation program after nutrient removal. Alternatively, they might be already partially differentiated during growth. To distinguish between these possibilities, we studied the expression of the early developmental marker discoidin I (Devine et al., 1982). This gene is normally induced immediately after nutrient depletion, and also in growth media at high culture densities (Clarke et al., 1987). We assessed discoidin I RNA levels in wild-type and set1 cells at a range of cell densities (Fig. 5A). In wild-type cells, discoidin I becomes significantly induced at a density of 2 \times 10^6 cells/ml. In set1 cells, discoidin I is strongly induced even at very low culture densities (5 \times 10^5 cells/ml). Indeed, the level of discoidin RNA at this density far exceeds the expression level seen in the wild-type cells at a 10-fold higher density (Fig. 5A).
In our developmental timing experiments, cells were taken out of media at a density of $1 \times 10^6$ cells/ml. We conclude that set1 cells enter the differentiation program even under conditions of nutrient abundance and low cell density. Inappropriate activation of developmental gene expression occurs during growth. There are several signals, in addition to nutrient depletion, which trigger differentiation. We addressed whether these signals were perturbed in set1 cells. The bacterial metabolite, folate, represses differentiation of Dictyostelium (Blusch and Nellen, 1994). We inoculated cultures of wild-type and set1 cells with 1 mM folate, and monitored discoidin I expression. Although set1 cells have a higher starting level of discoidin I than wild-types, after 24 h of folate treatment the level of this marker is negligible (Fig. 5B) indicating the mutants are responsive to folate. Early Dictyostelium development requires signaling through protein kinase A (PKA) dependent pathways (Firtel and Chapman, 1990). We assessed the requirement of set1 cells for PKA signaling during early development (Fig. 5C). Both wild-type and set1 cells show a strong increase in discoidin I induction after 5 h of differentiation. If PKA signaling is inhibited using the inhibitor H89, discoidin induction is impaired in both wild-type and set1 cells. We conclude that set1 cells still require PKA for differentiation.

Another differentiation trigger in Dictyostelium is high cell density. One pathway for sensing cell density is the presuppression pathway (Clarke et al., 1988). Growing cells continually secrete a presuppression factor (PSF) and when a critical level of PSF is present (at the critical population density) the cells differentiate (Rathi et al., 1991). One possibility is that set1 cells secrete an excess PSF or fail to secrete a PSF antagonist. We therefore assessed the ability of set1 mutants to induce the starvation response in wild-type cells. Cultures of set1 and wild-type cells were seeded with reporter wild-type cells expressing luciferase under the control of a discoidin promotor. The level of luciferase expression, and therefore the degree to which the reporter cells are differentiated, is shown at different culture densities. Both set1 clones (red and green lines) induce reporter cell differentiation more strongly than wild-type cells (black) at all cell densities.

In our developmental timing experiments, cells were taken out of media at a density of $1 \times 10^6$ cells/ml. We conclude that set1 cells enter the differentiation program even under conditions of nutrient abundance and low cell density. Inappropriate activation of developmental gene expression occurs during growth. There are several signals, in addition to nutrient depletion, which trigger differentiation. We addressed whether these signals were perturbed in set1 cells. The bacterial metabolite, folate, represses differentiation of Dictyostelium (Blusch and Nellen, 1994). We inoculated cultures of wild-type and set1 cells with 1 mM folate, and monitored discoidin I expression. Although set1 cells have a higher starting level of discoidin I than wild-types, after 24 h of folate treatment the level of this marker is negligible (Fig. 5B) indicating the mutants are responsive to folate. Early Dictyostelium development requires signaling through protein kinase A (PKA) dependent pathways (Firtel and Chapman, 1990). We assessed the requirement of set1 cells for PKA signaling during early development (Fig. 5C). Both wild-type and set1 cells show a strong increase in discoidin I induction after 5 h of differentiation. If PKA signaling is inhibited using the inhibitor H89, discoidin induction is impaired in both wild-type and set1 cells. We conclude that set1 cells still require PKA for differentiation.

Another differentiation trigger in Dictyostelium is high cell density. One pathway for sensing cell density is the presuppression pathway (Clarke et al., 1988). Growing cells continually secrete a presuppression factor (PSF) and when a critical level of PSF is present (at the critical population density) the cells differentiate (Rathi et al., 1991). One possibility is that set1 cells secrete an excess PSF or fail to secrete a PSF antagonist. We therefore assessed the ability of set1 mutants to induce the starvation response in wild-type cells. Cultures of set1 and wild-type cells were seeded with reporter wild-type cells expressing luciferase under the control of a discoidin promotor (Fig. 5D). The differentiation of the reporter cells was monitored as an increase in luciferase production. When wild-type cells are cultured with the reporter cells, a surge in discoidin promotor activity occurs at a density of $1.5 \times 10^6$ cells/ml. When reporter cells were cultured with two different set1 clones, the discoidin promotor activity was considerably enhanced over wild type levels at all cell...
densities. Therefore, there is a strong non-autonomous component to the developmental timing phenotype of set1 cells. The set1 cells might secrete an excess of a starvation-promoting factor, such as PSF. Alternatively, Set1 could be required to express an inhibitor of the starvation response.

Genome organization of Set1-regulated genes

To evaluate the role of Set1 in orchestrating changes in gene expression during the growth to development transition, we compared the genome-wide transcript differences between growth phase wild-type and set1 cells using DNA microarrays. A surprisingly small number of genes are significantly mis-expressed in the mutant. 44 genes are overexpressed and 28 genes underexpressed ($P < 0.05$. These data can be accessed at http://www.sanger.ac.uk/Users/alicat/chubb.html). Of these 72 genes, only 19 (26%) are part of the normal early developmental response, in that they are normally altered in wild-type cells after 2 h of starvation. A comparison of growing wild-type cells with 2 h starved wild-type cells reveals 605 genes significantly different in their expression profiles ($P < 0.05$). Therefore, although the set1 mutants are precociously differentiated with respect to certain genes and cellular responses, the initial developmental response is neither complete nor concerted. A survey of the classes of gene mis-expressed in the set1 mutants and a comparison with other studies using the same DNA arrays, reveals no obvious bias in the functional classes of the mis-expressed proteins. To validate the microarray data, the expression changes for 5 mis-expressed genes were assessed by Northern blots of independently derived RNA. The alterations detected by microarray were confirmed for all 5 genes (data not shown).

The genes mis-expressed in the set1 mutants are non-randomly distributed in the genome. To assess possible clusters of differentially expressed genes along chromosomes we used the CLAC (Cluster Along Chromosomes) method implemented in the CGH-Miner software of Wang et al. (2005). The method uses agglomerative clustering to group together adjacent genes on the basis of their log-ratios and assesses how long strings of up and down-regulated genes are. The arrays tested were compared against control arrays in which no significant clustering should be seen to estimate the false discovery rate (FDR) among the positive clusters (Benjamini and Hochberg, 1995). We identified 12 gene clusters reproduced in all arrays tested that passed our FDR cutoff of 0.01 (Fig. 6, Supplementary Fig. 2). Of the 12 clusters, 6 were predominantly composed of genes with significant sequence homology to their neighbors (Fig. 6, green bars). The other 6 clusters are comprised of non-homologous genes (Fig. 6, red bars), so their clustering does not reflect co-expression of gene family members or cross hybridization.

We took an alternative approach to assessing the clustering of mis-expressed genes, using only genes identified as significantly altered ($P = 0.05$ or less). We concentrated our analysis on chromosome 2, which encodes nearly one half of the mis-expressed genes (32/72), and provides a large pool for

Fig. 6. Genomic clustering of Set1-regulated genes. DNA microarrays were used to identify expression changes in the Set1 mutant. The figure shows plots of log-ratio of expression against chromosomal location for the four arrays used in the analysis. We used the CLAC (Cluster Along Chromosomes) method to assess the significance of the strings of up- and down-regulated genes. Strings passing the cut-off FDR (False Discovery Rate) of 0.01 were accepted as significant. Of 12 such strings identified, all of which were reproduced in all arrays, six (in green) contain predominately closely-related sequences. The remaining 6 are highlighted in red.
rigorous quantitative analysis. We moved a window of 100 genes along the chromosome to identify portions enriched with differentially expressed genes. A region at 6.3 Mb was filtered out and inspection of that region revealed 4 genes within 30 kb ($P < 7 \times 10^{-6}$) and a total of 6 genes falling in the 100-gene window ($P < 0.005$). These four genes also show the most significant perturbations of gene expression in the set1 mutants and this clustering was also detected by CLAC (Supplementary Fig. 2, cluster 8). We then restricted our analysis to dissimilar genes. We performed a BLASTN search (Altschul et al., 1990) on all chromosome 2 genes against themselves and identified 69 gene families ($E$ value < $1e^{-30}$). By randomly picking one gene to representative of each family, we repeated the simulation and found the clustering at 6.3 Mb was still significant ($P < 0.02$).

Fig. 7. Set1 catalytic activity and the set1 mutant phenotype. (A) Methyl H3K4 detected on genes mis-expressed in the absence of Set1. ChIP studies on wild-type cells were carried out using antibodies against H3K4me2 and H3K4me3 for the following genes — DDB0169498, DDB0203518, DDB0169496, DDB0217768 (all from the 6.3 Mb cluster) and also DDB0188438 and DDB0185548. Results were normalized to H3 antibody controls. The background levels obtained from normalized ChIP values from set1 mutant cells were subtracted from these data. (B) No increase in H3K9Me2 on genes repressed in set1 cells. ChIP experiments were carried out to detect changes in H3K9Me2 levels at genes repressed in set1 cells. These were compared to the changes in H3K9Me2 levels observed at the LTR and at two points in the reading frame of the DIRS transposon (w = wild-type, s = set1 mutant). (C) Western blots showing rescue of H3-K4 methylation by expression of GFP-Set1, but not catalytically dead N1425Q and C1474A GFP-Set1 proteins. (D) Rescue of the set1 mutant phenotype by a wild-type GFP-Set1 clone, but not catalytically dead GFP-Set1 proteins. An empty vector set1 control was used so all cells experience the same selective growth conditions (10 μg/ml G418).
We also found 3 pairs of genes within 3 kb and 4 pairs within 7 kb. To test whether these pairs could arise randomly, we ran a simulation that randomly picked 32 genes from the chromosome, and assessed how often pairs fall within 3 kb and 7 kb. We found significant clustering of our data relative to random pairs using both a 3 kb and 7 kb window ($P < 0.003$ and $P < 0.009$, respectively).

**H3K4 methylation of Set1-regulated genes**

To assess whether the genes mis-expressed in the set1 mutant cells are normally methylated by the Set1 enzyme, we carried out ChIP experiments to assess the levels of H3K4me2 and H3K4me3 at these loci in growing wild-type cells (Fig. 7A). We studied six loci, three over-expressed homologous genes at the 6.3-Mb cluster (DDB0169498, DDB0203518 and DDB0169496), one under-expressed gene adjacent to this cluster (DDB0217768) and two other repressed genes (DDB0188438 and DDB0185548). The data were normalized by subtraction of the background levels of immunoprecipitated chromatin from set1 cells. All the down-regulated genes show high levels of both H3K4me2 and H3K4me3. The up-regulated gene DDB0169498 is strongly dimethylated. The other up-regulated genes show more modest dimethylation, but the levels we detect are consistently above background.

A possible reason for the silencing of genes in the set1 mutants could be spreading of redundant silencing factors in the absence of H3K4me. To address this possibility, we carried out ChIP studies using antisera against H3K9me2 (Fig. 7B). The levels of H3K9me2 at these repressed genes are at background levels (no antibody controls) in both wild-type and set1 mutant cells, so there is no evidence that deposition of H3K9me2 is responsible for the silencing. We observe robust levels of H3K9me2 at the DIRS transposon, both in the LTR and at two sites in the open reading frame. H3K9me2 is reduced, notably in the LTR, in the set1 cells, indicating that the loss of H3K4Me causes dissipation of silencing factors from some locations. These observations on H3K9me2 distribution were reproduced with different H3K9me2 antisera.

To address whether the precocious development phenotype is caused by the lack of Set1 catalytic activity in the set1 mutant cells, we transformed set1 cells with set1 genes engineered to express mutations in the SET domain of the protein. We used two substitutions which render the protein catalytically inactive (N1425Q and C1474A; Santos-Rosa et al., 2003). Expression of wild-type GFP-Set1 reverts the loss of meH3K4 in the set1 mutants, whereas transformation of the same sequence encoding the N1425Q or C1474A mutations does not recover meH3K4 (Fig. 7C). The wild-type clone retards the precocious development of the set1 mutant (Fig. 7D) while the two mutant clones fail to rescue the phenotype.

**Discussion**

We have studied the function of Set1, a histone methyltransferase homolog in Dictyostelium. Set1 mutant cells lack all methylation of the lysine 4 residue of histone H3. At the genome-wide level, the different methylation states mark different developmental phases of the organism. Depletion of Set1 gives rise to cells displaying unusually rapid development, characterized by precocious aggregation of amoebae into multicellular aggregates after nutrient removal. This phenotype is caused specifically by the loss of Set1 catalytic activity. Early developmental markers are abundantly expressed in growing set1 cells, revealing that the mutants differentiate ectopically during growth. Co-culture of wild-type cells with set1 mutants induces the starvation response in the wild-type cells, indicating Set1 regulates production of an extra-cellular signal required for the correct perception of growth conditions. Microarray analysis of the set1 mutants indicates significant genomic clustering of mis-expressed genes.

**Histone modification and development**

A number of reports have linked histone modification to the timing of developmental transitions. Rhythmic histone H3 acetylation of circadian clock genes has been reported in the mouse, with acetylation levels mimicking corresponding RNA levels (Etchegaray et al., 2003), and interfering with histone deacetylation with the drug trichostatin A impairs light-induced gene expression (Naruse et al., 2004). Vernalisation, the stimulation of flowering by long periods of cold, has been correlated to deposition of dimethyl H3K9 and H3K27 at the promoter of the FLC gene (Bastow et al., 2004). Sustained down-regulation of the FLC is required for the normal timing of flowering, and these repressive modifications are required for maintenance of a silent FLC gene. In this study, we correlate a genome-wide loss of H3K4 methylation with unusually rapid development. The role of meH3K4 in gene regulation is unclear. Although H3K4me3 appears to be a mark of recent RNA polymerase II activity (Ng et al., 2003), the functional consequences of this mark are ambiguous. A possible role of H3K4me3 is the maintenance of a transcription-competent chromatin state, but other models include H3K4me3 acting as a signal for the recruitment of RNA processing factors (Ng et al., 2003).

In this study, we have shown H3K4 methylation is distributed throughout the nucleus, implying modification of much of the genome. Although possible that the effects of loss of meH3K4 act at a single dominant locus, we must also consider a concerted transcriptional change, distributed through the genome, or more locally, at gene clusters. This is pertinent when considering the phenotype of the set1 mutants, which show accelerated, but normally organized development, consistent with the concerted mis-expression of a battery of early differentiation genes. Concerted genome-wide transcriptional regulation by histone modifiers has been suggested in a recent study on the yeast S. cerevisiae. A combination of ChIP and microarrays revealed that different states of H3 and H4 hyper- and hypoacetylation define groups of biologically related genes (Kurdjatani et al., 2004). Similar to microarray studies on the yeast set1 mutant, we see only a small percentage of the genome is mis-expressed in the Dictyostelium set1 cells. This implies redundancy in the histone code, with meH3K4, an abundant mark, required for normal expression of small proportion of genes. In addition,
only 26% of these mis-expressed genes are part of the normal early developmental response, and the normal early developmental response involves altering the expression profile of at least 8-fold more genes than are mis-expressed in the set1 mutants. The set1 phenotype therefore appears not to be concerted at the level of the whole genome. However, it is possible that loss of set1 could have a concerted effect on a subset of genes involved in the growth to development switch, and that these genes, when mis-expressed, have a penetrant effect.

One possibility we addressed was whether any of the mis-expressed genes are tightly linked on chromosomes. Several well-documented examples of gene clusters encoding proteins with similar biological functions exist, notably the Hox, globin and myosin heavy chain clusters. In addition, significant proportions of the genomes of metazoan are organized with respect to shared expression profiles (Boutanaev et al., 2002; Lercher et al., 2002; Spellman and Rubin, 2002). Genes that are clustered in the genome have a tendency to be co-expressed. This is true even of clusters comprising non-homologous sequences. The Drosophila genome contains at least 200 groups of adjacent and similarly expressed genes (Spellman and Rubin, 2002) and these groups account for over 20% of genes. A recent study on FLC indicates this gene is also part of a cluster which is coordinately modified, to give common expression profiles to adjacent and similarly expressed genes (Boutanaev et al., 2002; Lercher et al., 2002; Spellman and Rubin, 2002). Genes that are clustered in the genome have a tendency to be co-expressed. This is true even of clusters comprising non-homologous sequences. The Drosophila genome contains at least 200 groups of adjacent and similarly expressed genes (Spellman and Rubin, 2002) and these groups account for over 20% of genes. A recent study on FLC indicates this gene is also part of a cluster which is coordinately modified, to give common expression profiles to adjacent and similarly expressed genes (Boutanaev et al., 2002; Lercher et al., 2002; Spellman and Rubin, 2002). Genes that are clustered in the genome have a tendency to be co-expressed. This is true even of clusters comprising non-homologous sequences. The Drosophila genome contains at least 200 groups of adjacent and similarly expressed genes (Spellman and Rubin, 2002) and these groups account for over 20% of genes. A recent study on FLC indicates this gene is also part of a cluster which is coordinately modified, to give common expression profiles to adjacent and similarly expressed genes (Boutanaev et al., 2002; Lercher et al., 2002; Spellman and Rubin, 2002).

We detected clustering of mis-expressed genes in set1 mutant cells. These clusters were not solely comprised of duplicated genes, perhaps reflecting the co-regulation of non-homologous proteins involved in similar physiological processes. One hypothesis is that these clusters represent functional chromatin domains, which permit co-regulation of expression by concerted changes in local chromatin topology. The topology of these clusters is particularly sensitive to loss of meH3K4, hence mis-expression in the set1 mutant background. A pertinent question here is how H3K4me3, a modification restricted to a specific portion of a gene, could exert effects on neighbouring loci. Indirect models are possible, with a mis-expressed protein being responsible for activation of gene clusters via shared regulatory elements, although the cluster we detect at 6.3 Mb on chromosome 2 also includes down-regulated genes, in addition to three up-regulated homologous genes. Another possibility is that H3K4me3 allows docking of factors capable of exerting more long-range effects on transcription. Alternatively, the lower methylation states of H3K4 may have the more dominant influence on clustered expression. The 6.3 Mb cluster is only weakly trimethylated, but more robustly dimethylated. In both Drosophila and yeast, H3K4me2 appears to be more uniformly distributed through the genome, so could potentially have more long-range effects than H3K4me3 (Noma et al., 2001; Santos-Rosa et al., 2002; Schubeler et al., 2004). In mouse and human cells, the distribution of H3K4me2 appears generally more punctate, although broad regions of meH3K4 overlay the Hox clusters (Bernstein et al., 2005).

**Developmental regulation of H3K4 methylation**

Genome wide levels of the different H3K4 methylation states are enriched at different stages during Dictostelium development. Levels of H3K4me3 drop as cells enter the multicellular phase of the lifecycle, while H3K4me1 increases. A drop in total H3K4me2 occurs at the end of development. H3K4me3 and H3K4me2 are modifications associated with transcriptional activation and euchromatin, respectively, so the reductions in these marks may reflect silencing and compaction of the genome occurring as the cells approach the dormant spore state. Alternatively, the loss of these higher methylation states may reflect slowing of the cell cycle rather than differentiation, as observed for meH3K4 in resting lymphocytes (Baxter et al., 2004). The association of H3K4me1 with the transcriptionally dormant spore state is interesting, and may reflect a primed ground state.

H3K4me3 is a mark of active chromatin, yet we see an ectopic induction of the ACA gene in set1 mutants, where H3K4me3 is absent. This apparent inconsistency can be resolved. Firstly, we observe precocious starvation properties in the set1 mutants prior to nutrient removal, 2 h before ACA induction. A primary cause of the rapid development phenotype occurs during growth, so the effect of loss of H3K4me3 on ACA expression will be overridden by this dominant early response. Secondly, the microarray experiments in this work and those performed on yeast Set1 mutants indicate that a number of genes appear to be repressed (directly or indirectly) by Set1. Loss of Set1 results in the loss of all three H3K4 methylation states, so up-regulation of genes in the absence of Set1 could also be a consequence of the loss H3K4me1 or H3K4me2, or differing background modification contexts.

When H3K4me3 diminishes, as genes are inactivated, we observe a transient rise in the level of H3K4me2 and a more sustained increase in the level of H3K4me1. Our observations are consistent with the stepwise removal of methyl groups from H3K4 as transcription ceases. This could involve lysine demethylation involving a FAD-dependent monoamine oxidase (Shi et al., 2004), as Dictostelium has genes for these enzymes, one being a few kilobases from the set1 locus. Alternatively, histone replacement might be used as the mechanism for removing methylated lysines (Ahmad and Henikoff, 2002; Janicki et al., 2004). At the whole genome level, we also observe an increase in H3K4me1 as H3K4me3 diminishes during differentiation. The Dictostelium H3 proteins all appear to be H3.3 variants, so a mechanism of demethylation via histone replacement, followed by constitutive activity of Set1 methylating newly incorporated histones is certainly feasible. It will be interesting to address these possibilities in Dictostelium cells mutated for H3 variants and nuclear monoamine oxidases.

**Acknowledgments**

We acknowledge the help of Matt Ritchie of The Walter and Eliza Hall Institute of Medical Research for help with the Kooperberg function. Clones were provided by Peter DeVreotes, Robert Insall, Annette Mueller-Taubenberger and Carole.
Parent, The *Dictyostelium* cDNA project in Japan, and The *Dictyostelium* Genome Project. Harry MacWilliams kindly provided the discoidin-luciferase reporter cells. We also thank Pascale Gaudet for help in identifying *Dictyostelium* SET proteins.

**Appendix A. Supplementary data**

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.ydbio.2005.12.054.

**References**


