2. MATERIALS AND METHODS

This chapter contains all the materials and methods utilised to generate the results presented in chapters 3, 4 and 5. All strains used are listed in Appendix I, all buffers, solutions and media in Appendix II, all primers in Appendix III, and a table summarizing the conditions for all the different experiments in Appendix IV.

2.1 FISSION YEAST GROWTH AND MAINTENANCE

2.1.1 Propagation and storage of fission yeast strains

All strains used in this thesis (see Appendix I) were maintained in solid or liquid Yeast Extract (YE). Before conducting any experiment, the mating type and phenotype of each strain have been checked (Moreno S. *et al.*, 1991). When used for an experiment all strains were grown in liquid Edinburgh Minimal Media (EMM) (unless specified - see Appendix IV) supplemented with the required amino acids (225 mg/L), at the temperature specified in the description of each experiment (see Appendix IV) and shaken at 170 rpm. Cells were normally grown to optical density (OD_{595}) of 0.25 -0.3 corresponding to 5 x 10^6 – 6.5 x 10^6 cells/ml. For long term storage strains were stored at -70°C in 50% glycerol - 50 % YE.

2.1.2 Experimental conditions

Elutriation and 'block and release' experiments were performed as described in section 2.3.1. The *nda3* experiment was carried out as described by Baum *et al.* (1998) collecting samples at 2, 4 and 6 hours after the temperature shift to 20 ºC. A zero hour timepoint collected before the temperature shift was used as reference. The hydroxyurea (HU) experiment was done according to Baum *et al.* (1997) collecting samples at 3 and 4 hours during the HU block (11mM HU) and 1 and 2 hours after the release from the block. A zero hour timepoint collected before HU addition was used as reference. The cdc10Δ::ura4 experiment was carried out as described by Ayte *et al*. (1995) collecting samples at 2, 4 and 6 hours after addition of thiamine (2 μM). A zero hour timepoint collected before thiamine addition was used as reference. Overexpression experiments

were carried out as explained in section 2.2.3. For all deletion mutants, cells were grown to OD_{595} of 0.2-0.3 and one single timepoint collected. Wild type cells grown in the same conditions were used as reference. More details concerning the experimental conditions for each mutant can be found in Appendix IVb.

2.2 FISSION YEAST MOLECULAR GENETICS

2.2.1 PCR-based gene deletions

A PCR-based approach was used for specific *S. pombe* gene deletions (Bahler J. *et al.*, 1998). The PCR primers used (for *ace2*, *fhl1*, *meu3* and *meu19* genes) were 100 nucleotides long (see Appendix III for the list of primers used); each primer contained at the 5'-end a gene-specific sequence and at the 3'-end a stretch of sequence homologous to the plasmid multiple cloning site. DNA fragments were amplified using the Advantage 2 polymerase mix (BD Biosciences) and the plasmid pFA6a-kanMX6 as template. A detailed map of the plasmid is shown in Fig. 2.1.

Fig. 2.1 Map of the pFA6a-kanMX6 plasmid, ~ 1.6 Kb size.

PCR reactions were performed in 96-well plates (Corning) and the reaction mix was prepared as follows: 2 µl pFA6a-kanMX6 template, 5 µl 10X Advantage PCR buffer, 1 µl 10 mM dNTPs, 1 µl 60 µM Forward Primer, 1 µl 60 µM Reverse Primer, 1 µl 50 X Advantage 2 DNA polymerase and 40 µl H_2O. The programme used for the amplification consists of 20 cycles at 95°C for 1 min, 55°C for 1 min and 68°C for 2 min. The PCR products were then checked on 1% agarose gels before proceeding. The product of the PCR reaction, of approximately 1.6 kb in size, is the kan^r cassette flanked by a gene-specific sequence. When *S. pombe* is transformed with such a construct, due to

homologous recombination, the kan^r cassette replaces the gene whose flanking sequences are next to the kan^r gene, resulting in a specific gene deletion. Once checked on agarose gel, the products from four PCR reactions were pooled, extracted with an equal volume of phenol:chloroform:isoamyl alcohol (25:24:1), precipitated with 3 volumes of 100% ethanol and dissolved in 20 µl of TE (pH 8). This concentrated DNA was used directly for the transformation as described in section 2.2.3.

2.2.2 Cloning for overexpression

Overexpression strains were constructed using a multistep procedure. The genes were first amplified by PCR, then cloned into the pPCR-Script Cam SK(+)vector (Stratagene), re-cloned into the pREP3X expression vector and then transformed into *S. pombe* cells.

PCR reaction

PCR reactions with gene specific primers (for *ace2*, *fhl1*, *fkh2* and *sep1* genes – Appendix III) were used to amplify the gene of interest using the corresponding cosmid DNA as template (Wood V. *et al.*, 2002). All the forward primers contain an *Xho* I restriction site for the following cloning step. PCR reactions were performed with 1 µl of cosmid DNA, 1 µl of 60 µM forward primer, 1 µl of 60 µM reverse primer, 1 µl of dNTPs mix, 1 µl of 50 X Advantage 2 DNA polymerase and 40 µl of water. The programme used consists of 25 cycles at 95°C for 1 min, 55°C for 1 min and 68°C for 2 min.

Preparation of the insert

The first cloning step was done using the PCR-Script Cam Cloning Kit (Stratagene). In order to separate the PCR product from primers, unincorporated nucleotides, buffer components and enzymes, the amplified fragment was purified using the QIAquick PCR purification kit (Qiagen) according to the manufacturer's instructions and treated with Pfu DNA polymerase (Stratagene) to generate a blunt-end insert for the first step of cloning. The reaction was prepared by adding 10 µl of the purified PCR product, 1 µl of 10 mM dNTPs, 1.3 µl of 10 X polishing buffer and 1 µl of cloned Pfu DNA polymerase (0.5 U)

and incubated for 30 min at 72°C. The fragment was then directly used for the ligation reaction.

Ligation reaction into pPCR-Script vector

The vector used at this stage was the pPCR-Script Cam SK(+) (Fig. 2.2).

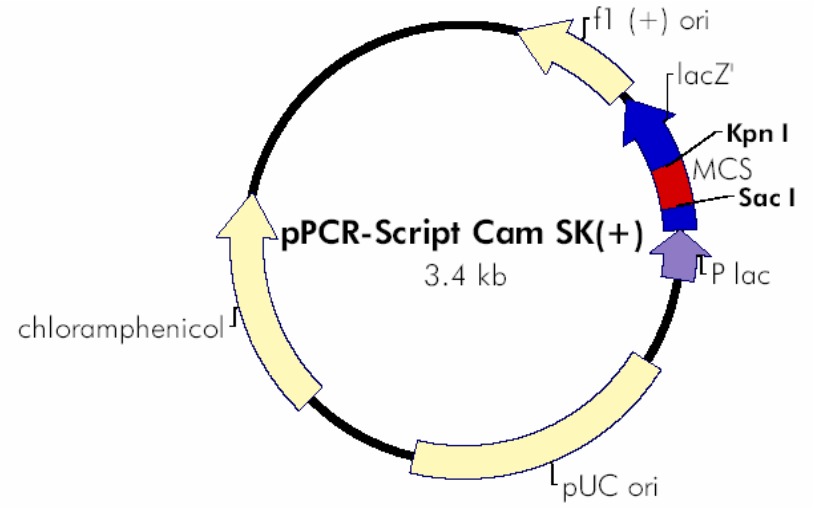

Fig. 2.2 Map of the pPCR-Script Cam SK(+) vector.

Ligation reactions were performed overnight at 15°C with 4 µl of the blunt-ended PCR product, 0.9 µl of the pPCR-Script Cam SK(+) vector, 1 µl of PCR-script 10X buffer, 0.5 µl of 10 mM rATP, 0.9 µl of *Srf*I restriction enzyme (5 U/ µl), 0.9 µl of T4 DNA ligase (4 U/ µl) and 1.8 µl of distilled water. After the overnight incubation the ligation reaction was heated up to 65°C for 10 minutes to inactivate the enzyme and stored on ice until transformation.

E. coli transformation

Transformations were carried out using XL10-Gold Ultracompetent *E. coli* cells (Invitrogen). Before starting the procedure, 1.6 µl of β-mercaptoethanol was added to 40 µl of competent cells to disrupt the membrane and facilitating plasmid DNA uptake. Cells

were then incubated on ice for 10 min, swirling gently every 2 minutes and then 2 µl of ligation mix was added to the cells. The cell/ligation mix was placed on ice for 30 minutes before being heat shocked for 30 seconds at 42°C and immediately quenched on ice for 2 minutes. 450 µl of pre-warmed LB medium was added to each sample and the samples incubated for 1 hour at 37°C with shaking. Each transformation was plated on LB plates supplemented with chloramphenicol (50 mg/ml) and the plates were incubated overnight at 37°C. The following morning, transformed colonies were picked into 2 ml of LB/chloramphenicol medium and the clones grown overnight at 37°C for plasmid DNA minipreps.

Plasmid DNA minipreps

Plasmid DNA was extracted from the overnight culture using the QIAprep Spin Miniprep Kit (Qiagen) and checked by restriction digest for the presence of an insert. 20U of *BamH*I and *Xho*I were used per reaction containing 5 µl of plasmid DNA, 1.5 µl of 10X buffer and distilled water to a final volume of 15 µl. Reactions were incubated for 4 hours at 37°C and the digestion product run out on a 1% agarose gel. Clones were selected among those carrying the insert for the next cloning step.

Preparation of the insert and ligation reaction into pREP-3X vector

The insert cloned into the pPCR-Script Cam SK(+) vector was then transferred to an expression vector, the pREP-3X (Fig. 2.3), capable of being expressed in *S. pombe* cells. The pREP-3X vector contains the inducible *nmt1* promoter (Maundrell K., 1993). Presence of thiamine (15 µM) in the medium represses the promoter. Once thiamine is removed, the promoter gets activated resulting in the transcription of the gene cloned downstream.

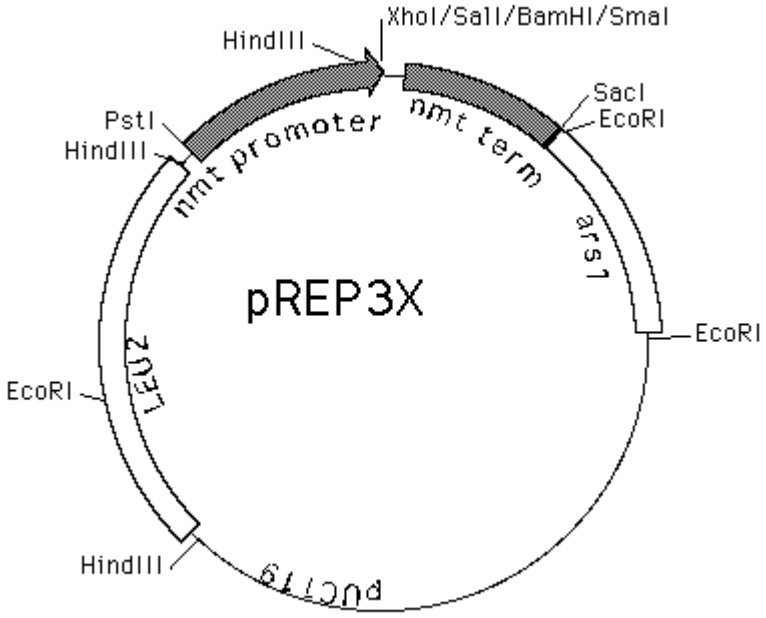

Fig. 2.3 Map of the pREP-3X expression vector

Plasmid DNA from selected clones (20 µl) and pREP-3X DNA (5 µl) were digested using *BamH*I and *Xho*I as described above (doubling the reaction volume). After incubation the samples were loaded onto 1% agarose gels made with 1X TAE and crystal violet (10 µg/ml). This DNA dye is very useful for isolating DNA fragments without exposing the nucleic acid to UV light, as normally occurs when ethidium bromide is used as a marker. This should improve cloning efficiency since the DNA damaging effect of UV light is removed (Hartman P.S., 1991).

Linearised plasmid and insert bands were cut out and purified using GFX PCR DNA and Gel Purification Kit (Amersham Pharmacia Biotech), eluting with 30 µl of elution buffer. The purified pREP-3X (2 µl) and the insert (2 µl) were then ligated with 1.5 µl of 10X ligase buffer, 1 µl of T4 DNA ligase (4 U/µl, Stratagene) and 8.5 µl of water. Ligation was carried out overnight at 15°C.

Second E. coli transformation

The ligation reaction was then used to transform ElectroMAX DH10B cells (Invitrogen) by electroporation. 2 µl of each ligation was placed in a microcentrifuge tube, 20 µl of competent cells added and the mixture pipetted into a chilled 0.1 cm *E. coli* Pulser

Cuvette (Biorad). Samples were electroporated using a Biorad GenePulser II electroporator at 2.0 kV. 1 ml of SOC medium was then added, the solution transferred to a Falcon tube and incubated at 37°C for one hour. Transformations were then plated onto LB plates supplemented with ampicillin (50 μg/ml) and plates incubated overnight at 37°C. Transformed colonies were picked into 2 ml of LB/ampicillin medium and the clones grown overnight at 37°C.

Plasmid DNA minipreps

As previously described, plasmid DNA was then extracted from an overnight culture using the QIAprep Spin Miniprep Kit (Qiagen) and checked by *BamH* I and *Xho* I restriction digest for the presence of an insert.

Plasmid DNA was again extracted from the selected clones using the QIAprep Spin Miniprep Kit (Qiagen) and transformed into *leu1-32 h*$^-$ *S. pombe* cells using the lithium acetate procedure.

2.2.3 Yeast transformation (Lithium acetate procedure)

S. pombe cells were transformed as described in Bähler *et al*. (1998). Cells were grown at 30°C in YE medium to a concentration of ~ 10^7 cells/ml. Around 20 ml of this culture were used for each transformation. Cells were washed once with an equal amount of water, the cell pellet was resuspended in 1 ml of water, washed once with 1 ml of LiAc/TE. The cell pellet was then resuspended in LiAc/TE at 2 X 10^9 cells/ml. 100 μl of the concentrated cells were mixed with 2 μl sheared herring testis DNA and 10 μl of the transforming DNA. After 10 min of incubation at room temperature, 260 μl of 40% PEG/LiAc/TE were added. The reaction was gently mixed and incubated at 30°C for 1 hour. After this second incubation 43 μl of DMSO were added and the cells heat shocked for 5 min at 42°C. Transformed cells were then washed once with 1 ml of water, resuspended in 500 μl of water and plated onto two YE plates (250 μl/plate). These plates were incubated at 30°C for 18 hours.

If the plasmid used for transformation was the pFA6a-kanMX6 (as described in section 2.2.1 for the PCR–based gene deletion experiments) colonies were checked for kanamicin resistance by replica plating them onto YE plates containing 100 mg/l G418

Geneticin (Life Technologies). The replica plates were incubated at 30°C for 2-3 days and large colonies were restreaked onto fresh YE plates containing G418. G418-resistant colonies were checked by colony PCR for integration of the DNA fragment. The PCR reaction mix was prepared as described above and a small amount of each colony used directly as template, resuspending it in the PCR reaction mix already aliquoted in 96-well plates (Corning). The primers used are listed in appendix III. The programme used for the amplification consists of 35 cycles of 95°C for 30 sec, 55°C for 45 sec minus 0.3°C per cycle and 72°C for 45 sec.

For the overexpression experiments (section 2.2.2) transformations were plated onto EMM plates containing thiamine (15 μM) (when the *nmt1* promoter is inactive). To induce overexpression, selected strains were then plated on EMM plates without thiamine (when the *nmt1* promoter is active) and grown for 18 hours at 32ºC for full induction. Similarly, liquid cultures were first prepared growing the strains for 12/16 hours in EMM/thi. Cells were then washed twice with EMM and grown for 18 hours in EMM before collecting a 50 ml sample. Cells carrying the pREP-3X control vector were grown and treated in the same conditions as reference.

2.2.4 Construction of double mutant strains

Double mutant strains *ace2Δ sep1Δ* and *fhl1Δ sep1Δ* were constructed as follows. *Ace2* and *fhl1* were deleted in a *sep1::ura4 ade6-M210/M216 ura4-D18 leu1-32* diploid strain (kindly provided by M. Sipiczki) using the PCR-based approach described in section 2.2.1 followed by *S. pombe* cell transformation as explained in section 2.2.3. Primers used for the deletion are listed in Appendix III. Kanamicin-resistant colonies were sporulated on EMM without nitrogen for 1/2 days at 25°C and tetrad analysis carried out on the meiotic progeny using the Singer MSM System (Singer Instruments). For each strain, several asci produced from sporulation were individually picked using a micromanipulator and placed in a line about 3-5 mm apart on YE plates. The ascus walls were left to break down at 32°C for 2/3 hours and each ascus was then micromanipulated to give a line of four isolated spores, separated by about 3-5mm. Spores were incubated until colonies form at the appropriate temperature for the strain. Colonies carrying the double mutation were then identified by colony PCR and by replica plating on selective medium, as previously explained.

2.3 FISSION YEAST PHYSIOLOGY

2.3.1 Synchronized cultures

Elutriation

In the elutriation process, cells are pumped into a specially designed centrifugal chamber (Beckman), where they are subjected to two opposing forces, the centrifugal force generated by the rotor and the counterflow of the fluid that is pumped in the opposite direction. As a result, each cell migrates to a zone where the two forces are balanced, forming a gradient of cell size along the chamber, as shown in Fig. 2.4.

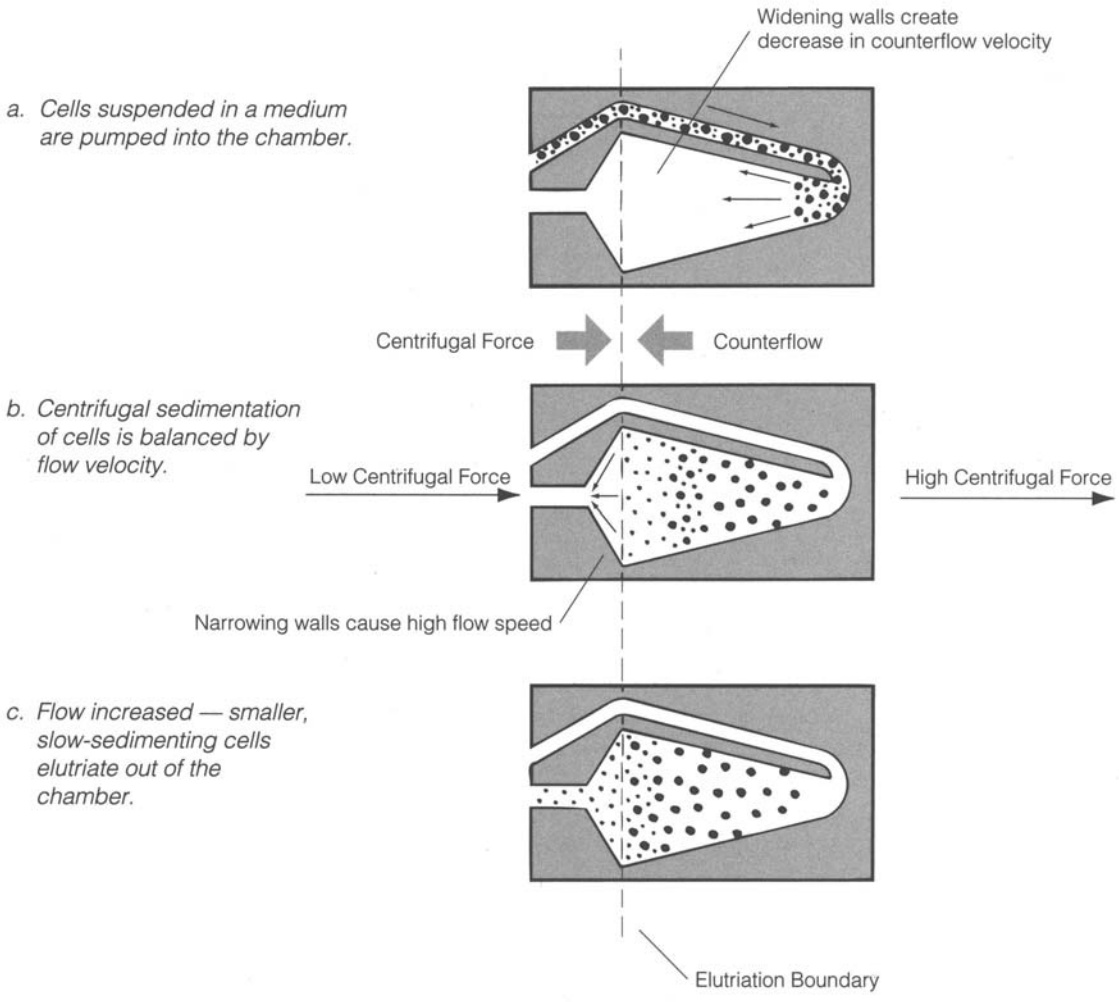

Fig. 2.4 The elutriation process.

From "The JE 5.0 Elutriation System – instruction manual", Beckman Coulter

The smallest, slowest sedimenting cells are held at the top of the chamber and can be specifically eluted by a small increase in the flow rate. This yields a homogeneous population of cells at the beginning of the cell cycle that will progress into division in a synchronous manner.

Five litres of wild-type fission yeast strain *972h*$^{-}$ were grown in minimal medium (EMM) at 30ºC to a cell density of around 5 X 10^6 cells/ml. This volume was loaded into a dual-chamber elutriation chamber (Beckman Instruments) in a Beckman J-6 M/E centrifuge equipped with a JE-5.0 elutriation rotor. A control sample of 300 ml was collected before starting the loading procedure. During the loading step the medium in which the cells were grown was collected and was referred to as conditioning medium. The elution was achieved using minimal medium (containing 20% of conditioning medium, 0.005% YE) pre-warmed to 30ºC. Small cells (650 ml at 2X10^6 cells/ml) were collected and grown at 30ºC for an hour to allow recovery. After that, samples were taken every 15 min for the next 6-7 h in order to cover two full cell cycles, with independent samples for DAPI and Calcofluor staining (1 ml), FACS analysis (1 ml), cell counting (1.4 ml) and RNA preparation (25 ml).

Arresting cells with temperature-sensitive mutants

Temperature-sensitive cell cycle mutants can be used to block cells at a particular point in the cell cycle. On return to the permissive temperature, cells will then proceed synchronously through the subsequent phases of the cycle. The temperature-sensitive strains used in this study are: *cdc25-22* which arrests in late G2 when shifted at the restrictive temperature, *cdc10-129* which arrests in G1 and *nda3-KM311*, a β-tubulin mutant that arrests in metaphase.

The *cdc25-22 h-* strain was grown to an OD_{600} of 0.15-0.2 in EMM medium at 25°C shaking at 170-200 rpm. The culture was then shifted to a 36°C water bath and held at that temperature for 3.5 hours. By this time cell density reached an OD_{600} value of 0.3-0.5. The cells were then released from the cdc25 arrest by shifting the culture to a 25°C water bath. 25 ml samples were taken every 15 min for 360 min after the shift. A control sample of 300 ml was collected before the temperature shift. Synchrony and progression through the cell cycle were monitored by FACS analysis (to measure DNA replication),

by DAPI staining of DNA (to determine mitotic cells), by cell counting and by Calcofluor staining of the cell wall (to determine cell division septa).

Two experiments, one using *cdc25-22* and another using *cdc10-129* were performed combining the two synchronization methods. Cells were first elutriated as explained above and subsequently blocked at 36ºC for 2.5 hours. Samples were collected every 15 minutes during the block and after the release following one full cell division. The *nda3-KM311* strain was grown at 32ºC for 12/16 hours and then shifted to 20ºC. Samples were collected at 2, 4 and 6 hours after the shift. A reference sample was collected before the shift to 20ºC. A wild type *972 h-* culture was grown in parallel and a single timepoint collected 6 hours after the shift to the restrictive temperature.

2.3.2 Cell number measurement

For cell counting cells (1.4 ml) were fixed by adding 2.8 ml of formal saline solution. Samples were diluted between 20X to 100X in ISOTON II (Beckman Coulter) in order to get an average count of 25000. Cell number was measured using a Beckman Coulter Z Series.

2.3.3 DNA content measurement

DNA content determination samples were prepared using propidium iodide staining as previously described (Sazer S. and Sherwood S.W., 1990). 500 μl of the ethanol fixed cells are washed in 1 ml of 10mM EDTA, cells are centrifuged for 1 min at 14000 rpm, the supernatant discarded and the pellet resuspended in 500 μl of 10 mM EDTA containing RNase A at 0.1 mg/ml final concentration. Cells are incubated at 37°C for at least 2 hours. After incubation 500 μl of 10 mM EDTA containing propidium iodine (4 μg/ml) are added to each sample. DNA content is measured using the Becton Dickinson FACScan.

2.4 FISSION YEAST MICROSPOPY

DAPI and Calcofluor staining were performed as described in Moreno *et al.* (1991). A separate 1 ml aliquot was taken at each time point of the time course experiments, cells were centrifuged for 15 sec at 14000 rpm and the pellets resuspended in 1 ml of cold 70% ethanol. Alternatively cells were fixed using a formaldehyde solution. For this fixation 10 µl of formaldehyde were added to 900 µl of sample.

2.4.1 DAPI staining

For DAPI staining, 50 µl of the ethanol fixed cells were washed in 1 ml of water, centrifuged for 1 min at 14000 rpm and the pellet resuspended in water containing DAPI at 1 µg/µl final concentration. Cells were then placed on a glass slide and visualized by fluorescence microscopy, using a Zeiss microscope.

2.4.2 Calcofluor staining

For Calcofluor staining, formaldehyde-fixed cells were centrifuged for 1 min at 14000 rpm, the supernatant discarded and the pellet resuspended in 10 µl of PBS. 1-2 µl of a Calcofluor saturated solution was added to 5 µl of sample and cells visualized using a Zeiss microscope.

2.5 MICROARRAY EXPERIMENTS

General overview

The microarrays used in this study were DNA arrays. They were fabricated by printing on a glass microscope slide many gene-specific probes obtained by PCR amplification from *S. pombe* genomic DNA. After fixation of the DNA to the glass surface, the slides were then hybridized with fluorescently tagged cDNA synthesized from total RNA of sample and reference. The fluorescent signal was then detected using a laser scanner, an image

acquired for each slide and subsequently processed to extract the data and analyze them (see Fig. 2.5).

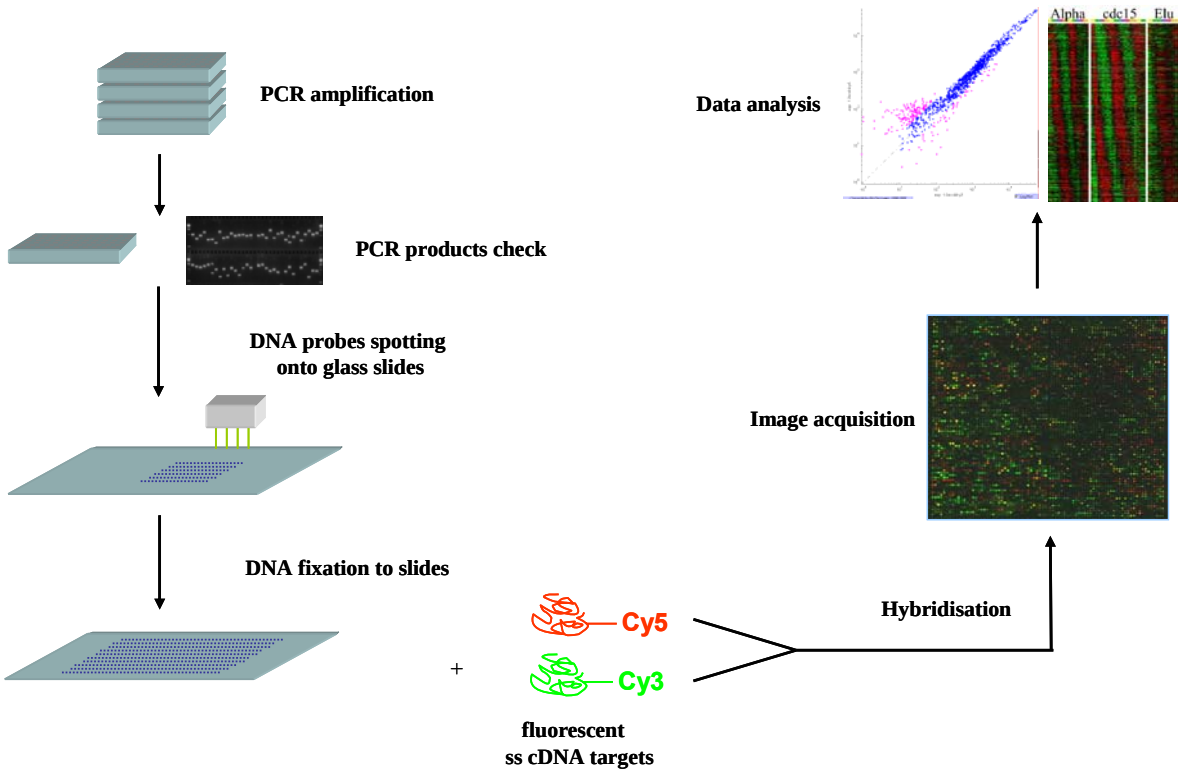

Fig. 2.5 **Overview of a typical microarray experiment**

2.5.1 Microarray fabrication

Each array contains probes for all the known *S. pombe* ORFs (the number of which varies between 4900 and 5270 depending on the generation of arrays), for 11 mitochondrial genes, 19 pseudogenes, several RNA genes, 114 hypothetical ORFs, 33 large introns as well as several control spots such as 22 *S. cerevisiae* genes, markers and tagging sequences (Kan-MX, GFP, GST, Myc, etc) and five *Bacillus subtilis* genes (lysA, pheB, dapB, thrB and trpC).

Two rounds of PCR amplification were performed for each array element: a first round using gene-specific primers with *S. pombe* genomic DNA, and a second round using a combination of a gene-specific reverse primer and a universal forward primer with the product of the first amplification. The universal primer contains a 5'-amino

modification that allows covalent attachment of the DNA to the glass slide. Array elements were printed onto amine-binding slides (Codelink, Amersham) using a MicroGrid II Total Array System (Biorobotics) at The Sanger Institute Microarray Facility. All elements were printed in duplicate, in separate halves of the slide, for a total number of 13000 spots. More details concerning microarray fabrication, printing and post-processing of the slides can be found in Lyne *et al*. (2003) or on the lab website (http://www.sanger.ac.uk/PostGenomics/S_pombe/).

2.5.2 RNA preparation

S. pombe total RNA was isolated according to Jang *et al.* (1995) with minor modifications. 25 ml aliquots of yeast culture were collected during a time course experiment every 15 min, harvested by centrifugation and frozen at -70ºC. Cells were then thawed on ice and resuspended in 1 ml of DEPC-treated water and pelleted by centrifugation. 750 µl of TES extraction buffer was added to the pelleted cells, cells resuspended and 750 µl of acidic phenol-chloroform (5:1, Sigma) added. The mixture was vortexed for 10 sec and incubated for 1 h at 65°C, vortexing for 10 sec every 10 min. The mixture was then rapidly chilled on ice and centrifuged for 20 min at 14000 rpm in a microfuge at 4°C to separate the aqueous and organic phases. The aqueous phase was transferred to a phase-lock tube (Eppendorf) and extracted with 700 µl of acidic phenol-chloroform. After centrifugation, a third extraction was similarly carried out with 700 µl of chloroform/isoamyl alcohol (25:1, Sigma). The aqueous phase collected after the centrifugation was precipitated in 2.5 volumes of absolute ethanol and 1/20th of 3M Na acetate pH 5.2. Afterwards the RNA pellet was washed with ice-cold 70% ethanol and air-dried to evaporate the residual ethanol. The pellet was resuspended in 100 µl of DEPC-treated water, the OD was measured and 2 µl checked on a 1% agarose gel.

2.5.3 Labelling protocol for total RNA

Total RNA (10-20 µg) and 2 µg of anchored oligo-dT/random primers were combined in a total volume of 16.4 µl. RNA/primer mixtures were heated to 70°C for 10 min and then cooled on ice for 1 minute. 2 µl of 25 mM Cy3- or Cy5-conjugated dCTP (NEN), 3 µl of 0.1 M DTT, 6 µl of first-strand buffer (GibcoBRL), 0.6 µl of dNTPs (25 mM each of

dATP, dTTP, and dGTP and 10 mM dCTP), and 2 µl of Superscript II reverse transcriptase (Gibco BRL) were added. Each sample was then incubated at 42°C for 1.5 hours to generate Cy-labelled cDNA. Starting RNA was hydrolysed by addition of 1.5 µl of 1 M NaOH and incubation at 70°C for 15 min. Samples were neutralized by addition of 1.5 µl of 1 M HCl. Labelled cDNA was separated from unbound fluorochromes by separation in an AutoSeq G-50 column according to the manufacturer's instructions (Amersham). After ethanol precipitation, the probe was resuspended in 27 µl/reaction of hybridisation buffer and 3 µl/reaction of polyA DNA (2 µg/µl, Sigma). The hybridisation mixture was then denatured at 100°C for 5 min, cooled down at room temperature for 10 min and hybridised to a microarray.

2.5.4 Hybridisation and posthybridisation

The microarray was covered by a 25X60 mm coverslip (Menzel-Glaser) and placed in a hybridisation chamber. 2 ml of 15X SSC were placed inside the hybridisation chamber before sealing, and the chamber placed in a 49°C oven. The microarrays were hybridised for 12-16 h. Microarrays were removed from the chambers and placed in standard histochemistry slide holders where they were washed at room temperature once for 5 min in washing solution 1, twice for 15 min in washing solution 2 and once for 5 min in washing solution 3 (see appendix II for details on washing solutions). Slides were then dried by spinning for 2 min at 1200 rpm.

2.5.5 Image acquisition

Slides were scanned using GenePix 4000B (Axon Instruments), a confocal scanner that uses two lasers operating at 635 nm and 532 nm to excite Cy5 and Cy3, respectively. The scanner uses a simultaneous dual-laser scanning system to produce a ratio image in real time. Each image was saved as a separate 16-bit TIFF file. These images were analyzed to calculate the relative expression levels of each gene and to identify differentially expressed genes.

2.6 MICROARRAY DATA ANALYSIS

2.6.1 Image processing

For image processing GenePix Pro 3.0-5.0 software (Axon Instruments, Inc.) was used. It applies a thresholding algorithm to separate spots from background, allowing a grid to be laid across the spots. Having found a grid, spots are found within each grid element, local background is calculated, background-subtracted and integrated intensities are calculated in both the Cy3 and Cy5 channels. GenePix Pro 3.0-5.0 computes a large number of different ratio quantities, each of which provides different insight into the raw data. They can be divided into three different subgroups: those that are ratios of quantities derived from whole features (such as 'ratio of medians' and 'ratio of means'), those that are derived from pixel-by-pixel ratios of intensities (such as 'median of ratios' and 'mean of ratios') and quality factors.

2.6.2 Data normalization and evaluation

Following image processing, the data generated for the arrayed genes must be further analysed before differentially expressed genes can be identified. Various computer programs are required to organize and evaluate the data. The first step in this process is the normalization of the relative fluorescence intensities in each of the two scanned channels. Normalization is necessary to adjust for differences in labelling and detection efficiencies for the fluorescent labels and for differences in the quantity of starting RNA from the two samples examined in the assay. These problems can cause a systematic shift in the average ratio of Cy5 to Cy3, and the intensities must be rescaled before an experiment can be properly analysed.

In this thesis, normalisation was performed using a Perl script developed by our group (Lyne R. *et al.*, 2003) that uses as imput file the GenePix result files. The script has two main functions: it first selects the spots on the array that are above certain cut-off parameters and then performs the normalisation on those selected spots. Only the spots that have >50% of pixels that are >2 standard deviation (SD) above median local background signal in one or both channels will pass the first screening. If a spot has

>95% of pixels that are > 2SD above local background in only one channel it will also be kept for the normalisation step. All discarded spots were flagged 'absent'. The normalisation is performed locally, using a sliding window including 400 spots as minimum number of features. The normalisation factor is calculated for each gene individually assuming that the medial signal ratio of all the measurable spots within the window is equal to 1. The resulting normalisation factor will then be used to correct the signal ratio for the spot on which the window was centred.

In addition, for the time course experiments, a second normalisation step was carried out using the 'per gene' normalisation option in GeneSpring. For each gene, it divides the value for each timepoint by the median of the value of all timepoints, thus centring all data to 1.

2.6.3 Identification of periodic genes

Identification of periodic genes in large datasets as obtained using microarrays is not an easy task. Spellman *et al*. (1998) pioneered this field publishing one of the first studies where periodic gene transcription in budding yeast was extensively investigated using microarrays. In the original paper, cell cycle regulated genes were identified using the Fourier transform. Since its publication, the data have been re-analysed in many different ways using increasingly sophisticated statistical methods (Cooper S. and Shedden K., 2003).

Since then several microarray studies focusing on periodic transcription have used different statistical approaches to tackle the same issue: Fourier transform (Whitfield M.L. *et al.*, 2002), fast Fourier transform [FFT - (Bozdech Z. *et al.*, 2003)], discrete cosine transform [DCT - (Laub M.T. *et al.*, 2000)], supervised clustering (Cho R.J. *et al.*, 2001) and hierarchical/K-means clustering (Ishida S. *et al.*, 2001).

It should be mentioned that a purely statistical approach to the identification of periodic genes will not always be the best option. This is because such an approach requires a relatively high number of repeats and the arbitrary choice of a cut-off point below which the data is unreliable (Cooper S. and Shedden K., 2003). For this reason, in this thesis the identification of periodic genes was done combining a simple autocorrelation approach, visual inspection and subsequently validated with a more sophisticated statistical analysis using fast Fourier transform. The traditional clustering

algorithms (Eisen M.B. *et al.*, 1998; Sherlock G., 2000) are based on a correlation algorithm. All single expression profiles in a dataset are compared to each other and a correlation score is assigned to each pair. The members of the pair with the highest correlation score will be joined to form a node of the cluster, and the process will be repeated until all profiles have been assigned to a node.

In the majority of the timecourse experiments done for this thesis two full cell cycle divisions have been followed, allowing the assumption that a gene behaving periodically would have reached a peak of expression in the first cycle and another peak in the second cycle. The first-cycle profile was then compared to the second-cycle profile and an autocorrelation score assigned according to the similarity between the two. An average autocorrelation score was calculated for each gene across three elutriation and two *cdc25-22* 'block and release' experiments, and genes were ranked according to the average autocorrelation value.

In parallel, a fast Fourier transform analysis (FFT) was run on the same dataset as well as on the randomized dataset after 100000 permutations of the expression signals. Randomisation is very important to determine whether the cyclic gene behavior arises from chance arrangements in the measurements or has a real biological significance. Part of this was done in collaboration with Pietro Lio in Alvis Brazma's group at the European Bioinformatics Institute.

For each gene a P-value (*P*) was calculated comparing the result of the FFT analysis performed on the real data with the one obtained from the randomized data. Only genes with a $P < 0.01$ were considered for further analysis. Among those only the ones with a fold change > 1.5 were again visually inspected to confirm periodicity.

Information concerning the biological function of each periodic gene was derived from the following databases: the PombePD (www.incyte.com/control/tools/proteome) and the *S. pombe* GeneDB (www.genedb.org/genedb/pombe/index.jsp). When a more detailed description of the gene function was needed, one of the following text books was consulted: The Molecular and Cellular Biology of the Yeast *Saccharomyces* (1997), The Yeast Nucleus (2000) and The Molecular Biology of *Schizosaccharomyces pombe* (2004). On many occasions, specific research papers were also used but references to them were omitted in the text (especially in Chapter 3, section 3.4).

Identification of genes differentially expressed in gene deletion and overexpression experiments was achieved by combining two methods. First, an intensity-

dependent standard Z-score was calculated by determining the local mean and standard deviation (SD) within a sliding window of 1000 genes (Yang I.V. *et al.*, 2002). Genes with ≥ 2 SD above the local mean in repeated experiments were considered differentially expressed. When at least three or more repeats were available for each experiment, SAM (Significance Analysis of Microarrays) was used to confirm differential expression (Tusher V.G. *et al.*, 2001).

2.6.4 Clustering periodic genes

The simplest example of classification is a case where genes are grouped according to the time of their highest peak of expression which can be easily calculated from the Fourier algorithm. Genes can also be grouped together by various clustering algorithms, based on the similarity between their expression profiles under different conditions (Sherlock G., 2000; Eisen M.B. *et al.*, 1998). This helps to suggest functions of unknown genes that are co-expressed together with genes of known function ('guilt by association'). Moreover, genes that are co-expressed can be checked for regulatory promoter sequence motifs required for the binding of transcription factors (Vilo J. and Kivinen K., 2001; Brazma A. and Vilo J., 2000).

To group genes based on the similarity in their expression profiles, we used the Gaussian mixture model with Pearson coefficient in ArrayMiner 5 (Optimal Design - (http://www.optimaldesign.com/Download/ArrayMiner/AM2whitepaper.pdf). Clustering was also performed using the K-means algorithm (Sherlock G., 2000) in GeneSpring 6 (Silicon Genetics).

2.6.5 Identification of promoter motifs

It seems reasonable to expect that genes that are co-expressed (i.e. that have very similar expression profiles) might also be co-regulated, meaning that they share some common regulatory mechanisms such as being targets of the same transcription factor. Transcription factors act by binding to a sequence-specific site in the DNA and influencing the expression of the downstream gene. These binding sites are located in promoter regions that in yeast are normally less than 1000 bp upstream of the transcription starting site.

Once a cluster of co-expressed genes has been identified, the upstream regions of those genes can be scanned by looking for common regulatory elements. Several programmes are available to perform such analysis, within microarray analysis packages such as GeneSpring as well as other applications. In this thesis the pattern discovery tool SPEXS (http://ep.ebi.ac.uk/EP/SPEXS) (Vilo J. *et al.*, 2000) was used and the analysis was run at the European Bioinformatics Institute by Katja Kivinen in Alvis Brazma's group. The up- and downstream intergenic regions (600bp and 400bp, respectively) for all coding sequences were extracted from genomic sequences from the Sanger Institute (ftp://ftp.sanger.ac.uk/pub/yeast/pombe/). 19 genes were discarded (transposons, telomeric duplications, or very short sequences). We grouped genes in overlapping clusters using a sliding window (40 and 60 genes) along the phase-ranked gene list of periodic genes, shifting the window by 10 genes at a time. Alternatively, we used lists of genes dependent on Ace2p or MBF. To estimate the expected number of pattern occurrences, we used all intergenic regions. We defined cluster-specific significance thresholds from randomized sets of sequences of respective sizes. We required that the pattern occurred in at least six sequences in the cluster, that its binomial probability was smaller than the significance threshold, and that it occurred in the cluster at least twice as often as expected using these criteria. No significant motifs in downstream regions were found. All patterns passing the criteria were grouped by similarity, and the best one in each group was reported. Putative regulatory patterns and locations were visualised using PATMATCH and SEQLOGO tools in Expression Profiler (http://ep.ebi.ac.uk/EP/). For sequence logos, we introduced pattern variations by allowing one wildcard, searched for matches to this approximate pattern and calculated the number of each base occurrence for each position. The letter sizes were rescaled according to the information content in that position. For MCB and Ace2 motif logos, we used the lists of MCB- and Ace2p-dependent genes instead of sliding windows.

2.6.6 Comparison between fission and budding yeasts

Gene lists of *S. cerevisiae* cell cycle regulated genes were downloaded from the accompanying websites of Spellman *et al.* (1998) and Cho *et al.* (1998). These lists were imported into GeneSpring (Silicon Genetics) and the overlapping genes between the two datasets identified. For each comparison a P-value (*P*) was calculated using the

hypergeometric distribution (Fig. 2.6). This equation calculates the probability of overlap corresponding to *k* or more genes between a gene list of *n* genes compared to a gene lists of *m* genes, when randomly sampled from a universe of *u* genes.

$$\frac{1}{\binom{u}{m}} \sum_{i=k}^{n} \binom{m}{i} \binom{u-m}{n-i}$$

Fig. 2.6 Hypergeometric distribution.

This formula was used to calculate the P-value associated with all gene lists comparisons

For each comparison the size of the overlap expected by chance (x) was also calculated using the following formula where y and z represent the size of the two lists compared and T the total number of *S. pombe* genes with a *S. cerevisiae* ortholog:

$$x = \frac{(y)(z)}{T}$$

The resulting core set of genes in common between the two budding yeast studies was used for comparison between the two yeast models. Genes with a prospective *S. pombe* ortholog were determined using a table of curated orthologs created by Val Wood at the Sanger Institute. The total number of orthologs included in the analysis was 2981. Genes were translated into *S. cerevisiae* homologs using this table and lists comparison was performed with Genespring. A comparison was also done with the list of human genes reported as periodic in the literature, using traditional genetic approaches.