# Does DNA topography coordinate intra- and inter-chromosomal Galactose (*GAL*) gene expression?

**This thesis is presented in partial fulfillment of the requirements for the degree of Masters of Science (MSc) in Molecular Biology at Massey University, Albany, New Zealand**

**Shuqiao Yu**

**2008**

## Abstract

For a long time, DNA had been considered as a stabilized, rigid, and “linear” structure, which acts as a platform for molecular regulators to function. However, genome structure in living cells is far more complex than the linear representation of the primary DNA sequence implies. This thesis aims to investigate whether the position of a gene within the genome plays a role in the regulation of its activity. The galactose (*GAL*) gene family of *Saccharomyces Cerevisiae* is used as a model. This gene family enables yeast cells to utilize galactose as an alternative carbon source; and it consists of structural and regulatory genes. Structural genes *GAL1*, *GAL10* and *GAL7* exist in a cluster on yeast chromosome II. The products of the regulatory genes, *GAL3*, *GAL4*, and *GAL80*, regulate the expression of the *GAL* structural genes, depending on the availability of carbon sources. Specifically, *GAL* gene expression is repressed by glucose, paused for induction (noninduced) by glycerol/lactate, and fully induced by galactose.

The aim of this project was to study the relative position of the *GAL* structural genes within the nucleus, and whether any chromosomal interactions at the *GAL* locus help to regulate their activation. These were tested in accordance with the expression status of the *GAL* genes *(i.e.* repressed, noninduced or induced). Followed confirmation of the existence of any chromosomal interactions, protein/protein complexes that mediate these interactions were attempted to identify.

The methods applied in this project were Chromosome Conformation Capture (3C) and Circular Chromosome Conformation Capture (4C), which applied in combination to map the positions of the GAL genes in the context of the overall genome structure. The results indicated that the GAL locus on chromosome II was divided into two “interaction zones”. DNA loops formed around these interaction zones to form an S-shape structure in a carbon source-independent manner. Two novel inter-chromosomal interactions between chromosomes II and XVI, *i.e. SVL3-GAL7* and *HOS1-GAL10*, were also identified. Although these interactions occurred regardless of the *GAL* gene activities, it was suggested by real-time PCR that the interaction frequency for *SVL3-GAL7* declined as the *GAL* genes being activated. Unfortunately no protein/protein complexes were identified to play an important role in mediating either intra- or inter-chromosomal interactions.

Future work will be needed to identify the protein/protein complexes that play a role in mediating the S-Shape structure at the *GAL* locus and the two inter-chromosomal interactions. Additional works could also focus on the understanding of the functional implication of the interactions between chromosomes II and XVI.

## Acknowledgement

Firstly and most importantly I would like to say a big thank you to my supervisor, Professor David Lambert, who helped me through the most difficult time in my Masters studies. Without him, this thesis would not exist. There is no way I could fully express my appreciation. I guess the finalization and completion of this thesis is the best return for his unconditional supports.

I would also like to take this opportunity to thank you a lot of people who have been offering me enormous helps and supports throughout the years. Without you, this thesis would not exist either. I would like to give my sincere appreciation to Tim Coolbear and Alan Clark for their generous supports, especially Tim, thanks for being there for me at all time. Justin O'Sullivan for his helps in early project development and methodologyand financial supports. Leon Huynan for offering me with great suggestions and feedback for both my project and this thesis. A lot of other people in the lab, Jarod, Eli, Martina, Dharani, Katie, Tom, etc., thank you so much for the happy time we shared together. Also Rebecca, Tracey and Daniela who have already left. My friends Sha, Wei, Soyeon, Sara, Kylie, Nancy and a lot more if I missed your names. Thank you so much for the continuous supports.

Finally, I would like to thank my parents for believing in me from the very beginning and for their unconditional love and support, both financially and emotionally. My final and special thank is to Kai, who has been listening to me, comforting me when I was down and most importantly, share the best and worst moments with me. A very deep thank you!

# Lists of Figures

# Lists of Tables

# Abbreviations

In addition to the chemical symbols shown in the Periodic Table of Elements and the *Systeme international d'unites* (SI), the following abbreviations are used:

| | |
|---|---|
| SC | Synthetic complete |
| DNA | Deoxyribonucleic acid |
| DTT | Dithiothreitol |
| DMSO | Dimethyl sulfoxide |
| EDTA | Ethylenediaminetetraacetic ccid |
| PCR | Polymerase chain reaction |
| PBS | Phosphate buffered saline |
| PMSF | Phenylmethanesulfonyl flouride |
| SDS | Sodium dodecyl sulfate |
| SOC | Super optimal catobolite |
| BSA | Bovine serum albumin |
| TE | Tris EDTA |
| TBE | Tris borate EDTA. |
| TEMED | Tetramethylethylenediamine |
| LPA | Linear acrylamide |
| H | Histidine |
| Kb | Kilobase |
| w/v | Weight/volume |
| v/v | Volume/volume |
| RT | Room temperature |
| O/N | Overnight |
| Min | Minute |
| S | Second |
| U | Unit |
| 3C | Chromosome Conformation Capture |
| 4C | Circular Chromosome Conformation Capture |
| CHO | Carbohydrate |
| C | Cut/restriction enzyme digested sample |
| CL | Cut/restriction enzyme digested and ligated sample |

# Table of Contents

# Chapter 1----Introduction

## 1.1 Preview

Traditionally DNA was described and portrayed as a linear representation of genes. During cellular processes, such as DNA transcription, replication, recombination, and repair, DNA functions as a platform on which regulatory proteins operate (Johnston, 1987). On some occasions, transcription activation can take place when regulatory DNA elements are separated by kilobases of DNA (Johnston, 1987). Due to the large distances between some regulatory sites, it has been proposed that two widely separated loci interact by "looping out" intervening DNA (Matthews, 1992).

DNA looping has been used to explain how the widely separated regulatory elements, such as enhancers, promoters and terminators, come together to direct downstream transcriptional processes efficiently (Hittinger *et al*., 2004). The "looping" effect is known to occur in both eukaryotic and prokaryotic organisms, *e.g.* at the mouse β-globin locus and the *E. coli lac* operon, respectively (Rippe *et al*., 1995). Therefore DNA looping is considered to have functional implications for the understanding of cellular process, especially in transcription regulation. Furthermore, DNA looping challenges traditional view of DNA strands being linearized and rigid.

Even though changes in DNA structure appear to play a role in the regulation of gene expression, the concept of "looping" and how looping influences gene activation is still relatively unknown. It is generally difficult to apply such idea to contemporary studies, *i.e.* most of the studies concerning gene activities did not take into account of the changes in either DNA structure or gene positioning. However, recent work has shed lights on this area by showing that overall genome shape and the position of a particular gene within the nucleus (spatial positioning) can influence a gene's activity (de Laat *et al.*, 2003).

This project studies how DNA structural changes are influenced by the expression of a well-characterized gene family in the yeast *Saccharomyces cerevisiae.* This work could potentially help increase our understanding of eukaryotic genomes by investigating the role that genome structure plays in coordinating gene expression.

This introductory chapter focuses on several key elements that constitute this project. Firstly our current understanding of nuclear structure, organization and gene order will be addressed. Then the existing evidence for DNA looping and their involvement in various cellular processes will be discussed. Finally the gene family and the methodologies used in this work will be described.

## 1.2 Nuclear Structure, Chromosome Territories and Gene Positioning

In eukaryotic cells, the nucleus is a membrane-enclosed organelle that contains most of the cell's genetic material. The genetic materials are packaged with proteins into chromosomes. A chromosome is a continuous piece of DNA, which contains genes, regulatory elements and other nucleotide sequences. DNA is wrapped with histones, the DNA-bound proteins that package DNA and control its function, to form nucleosomes. Nucleosomes are organized in their most compact form to become a 30 nm condensed chromatin fibre. Chromatin is the complex of DNA and protein that makes up chromosomes. The genes that "reside" on the cell's chromosomes constitute the cell's nuclear genome. A cell's nucleus functions to control the network of biochemical processes that are essential for cell activity. The positioning of gene loci at specific regions in the nuclear appears to play an important role in both the regulation and activation of transcription in a variety of species (Brown *et al*., 1999; Volpi *et al*., 2000; Feuerbach *et al*., 2002).

Early studies proposed that the chromosomes containing fewer active genes are sequestered at the nuclear periphery, an area of putative transcriptional silencing; whereas the chromosomes containing more active genes are localized within the nuclear interior, an area considered to be more transcriptionally active (Ferreira *et al*., 1997; Croft *et al*., 1999). In support of this theory, Zink *et al.* (2004) analysed the nuclear positioning and transcriptional activity of three adjacent, but functionally unrelated, genes mapped to the human chromosome 7q31. Using fluorescent *in situ* hybridisation (FISH) and quantitative reverse transcriptase polymerase chain reaction (qRT-PCR), the authors found that when inactive these three genes were associated with the nuclear periphery, whereas when active they associated with euchromatin in the nuclear interior (Zink *et al.,* 2004). However, when Casolari *et al.* (2004) examined the association of the *S. cerevisiae* genome and nuclear pore proteins using FISH and confocal microscopy, they found that highly transcribed genes (*e.g.*, *GAL1* upon galactose induction) relocated to the nuclear periphery and associated with nuclear pore proteins. These contradictory findings might be due to cell type and gene-

specific differences. Nonetheless, the observations still suggested a differential gene activity in response to nuclear positioning, but with no clear pattern.

In the nucleus, individual chromosomes maintain their identity by occupying discrete regions of the nucleus called "chromosome territory" during the entire cell cycle (Cremer *et al*., 1993). Using FISH, Cremer *et al.* (1993) proposed an "inter-chromosome domain (ICD) compartment" model suggesting that DNA transcription and RNA processing predominantly occur at the periphery of the chromosome territories. Zirbel *et al.* (1993) showed that specific gene transcripts and RNA splicing machinery were concentrated at the borders of chromosome territories, supporting the "ICD" model. In contrast, others have provided evidence against this model. For example, Verschure *et al.* (1999) showed (using FISH) that nascent RNA was located at the inner region of chromosome territories. In addition, Mahy *et al*. (2002), also using FISH, showed that transcription of both ubiquitous and tissue-specific genes were not restricted to the periphery of chromosome territories, and transcription factors could gain access to the chromosome interior. Hence transcription is not restricted to the periphery of chromosome territory. Similarly, Zink *et al.* (2004) claimed they failed to find a correlation between the transcription activity of genes and their positioning in relation to the chromosome territories. There is therefore still considerable debate regarding correlation between nuclear positioning and gene expression, and whether this correlation is cell- or gene-type specific, or simply a random pattern.

### 1.2.1 Gene Organization within Chromosomes and Regulation of Transcription

The organization of genes (gene order) within a genome can be divided into two categories: within chromosomes (intra-chromosomal), or between chromosomes (inter-chromosomal) (Hurst *et al*., 2004). This section will focus on gene organization within chromosomes (intra-chromosomal). In bacteria, the discovery of tightly regulated gene groups (operons) suggesting that these genes are in an ordered arrangement (Song *et al*., 1996). In eukaryotes, although it was once thought that genes were randomly distributed along the chromosomes, gene clusters that occur naturally, *e.g.* mammalian homeobox (*Hox)* and globin genes, suggested an ordered arrangement (Duboule *et al*., 1989). The availability of whole-genome sequences accelerated the discovery of the co-expressed gene clusters in eukaryotic organisms. For instance, in *S. cerevisiae* 25% of genes involved in the cell cycle clustered together (Cho *et al*., 1998); in *Caenorhabditis elegans*, tissue-specific genes that are co-expressed also show significant degrees of clustering (Roy *et al*., 2002); a similar situation was found for *Drosophila melanogaster* where 45% of genes expressed in testes were found

to be closely linked (Boutanaev *et al*., 2002), and in *Homo sapiens* where clustering occurs in housekeeping and highly expressed genes, such as those in prenatal and postnatal tissues (Lercher *et al*., 2002).

In addition to being co-expressed, clustered genes may also be functionally related, *e.g.* those whose gene products are in the same pathways, gene products that interact, or gene alleles that affect the same trait (Hurst *et al*., 2004). In *S. cerevisiae*, 20% of genes whose products have similar functions tend to be closely positioned on the same chromosome (Cohen *et al*., 2000). Other eukaryotic organisms also showed a high level of functionally related gene clustering (58% in *C. elegans*, and 65% in *H. sapiens*: Lee *et al*., 2003).

Exploring the reason for gene clustering, it is likely that regulatory and promoter sequences act in concert to enable simultaneous gene expression. For example, in the yeast and human genomes, there is a particular promoter called a "bi-directional promoter" that resides between two adjacent genes and controls their co-expression (Cho *et al*., 1998; Cohen *et al*., 2000). In some extreme cases, co-expression of a group of linked genes is achieved by fusing all of the genes to make a single protein product (Hawkins, 1987). This "promoter-driven" expression model, however, is too simple to account for every case of gene clusters in different species. In the yeast genome, upstream activating sequences (UAS) have been shown to be important for the regulation of gene pairs (Cohen *et al*., 2000). UAS are usually located large distances (> 1000 bp) away from the expressed genes they regulate; and it remains unclear how these sequences operate.

Previous studies have shown that chromatin higher-order structure, *i.e.* chromosome organization beyond the primary linear nucleosomal model, plays a critical role in eukaryotic gene regulation (Bonifer, 1999; Wolffe *et al*., 2000; Woodcock *et al.*, 2001). Spellman and Rubin (2002) studied hundreds of gene microarray profiles of *Drosophila* and mapped each expressed gene to their chromosomal location. They found that the *Drosophila* genome contains multiple large groups of adjacent genes that are co-expressed but are not related in function or sequence. The authors proposed that this observation might be occurring as a result of the local chromatin structural changes, allowing the expression of large groups of linked genes. The authors also proposed that there is a dominant gene within each chromosomal territory and its expression or repression dictates whether chromatin is "open" (accessible to transcription regulators) or "closed".

For transcription regulation, Iborra *et al.* (1996) used light microscopy to visualize the location of active RNA polymerases in permeabilized HeLa cells. The authors found that

active RNA polymerases were concentrated at sites that contained newly synthesized transcripts and regulatory proteins. In particular, there are approximately 2100 such sites in each HeLa cell nucleus. These sites were "transcription factories" (Iborra *et al.,* 1996; Cook, 1999) and consisted of pre-assembled groups of transcription factors that enabled transcription en masse of multiple genes from different chromosomes. In particular, the locus control region (LCR) of DNA loops around over the desired gene of expression, so that the binding site of the transcriptional machinery is in physical close proximity to bridge and affect the transcription of that specific gene (Figure 1; Cook, 1999). LCR are defined by their ability to enhance the expression of linked genes by establishing an open chromatin configuration that inhibits the normal repression of transcription. In a further study, Jackson *et al.* (1998) found approximately 5500 transcription factories in each HeLa nucleus, and they observed each transcription factory contained ~14 active RNA polymerase II. By comparing the association rate between transcribed genes and RNA polymerase II via RNA and DNA FISH, Osborne *et al*. (2004) showed that distal genes collected at same transcription factories upon activation and repositioned themselves away from the factories when they became inactive, suggesting that active genes become mobile to allow them to be recruited to a transcription factory.

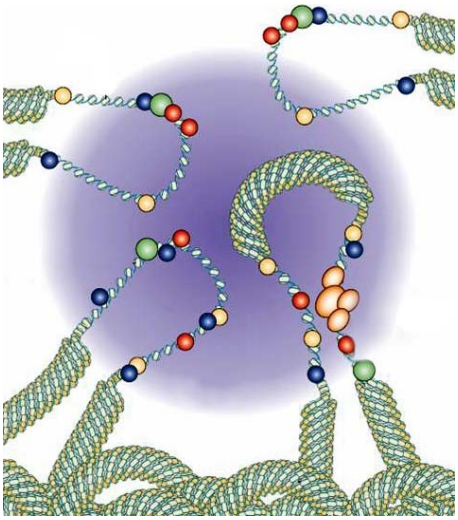


**Figure 1:** Analogue model for a transcription factory. In this model, genes are "released" from their chromosome territories to access a shared transcription factory rich in RNA polymerase II (large light purple circle). The coiled lines constitute chromatin structure. Colored (blue, red, green) spheres represent DNA binding factors. The bright orange ovals are locus control region (LCR) gene complex.

Modified from Chakalova *et al*. (2005)

### 1.2.2 Inter-Chromosomal Association and Regulation of Transcription

Although eukaryotic chromosomes occupy distinct territories in the cell nucleus, they do occasionally "overlap" with other chromosomes. Any physical contact between different chromosome loci is referred to "chromosome kissing" (Cavalli, 2007), and may have resulted from shared transcriptional machineries, or shared regulatory functions.

A prominent study that demonstrates the inter-chromosomal contacts was conducted by Spilianakis *et al*. (2005), who studied human immune cells using Chromosome Conformation Capture (3C) analysis and DNA microscopy techniques. 3C is a novel molecular technique that studies physical interactions between DNA fragments in a formaldehyde cross-linking condition (Section 1.5). The authors studied the inter-chromosomal interaction in relation to transcriptional activation between two loci that were expressed in two different cell types in the human genome. $T_H2$ locus on chromosome 11 and IFN-γ gene (*Ifng*) promoter on chromosome 10 are involved in the specification of T helper cells 1 and 2 ($T_H1$ and $T_H2$) pathways. $T_H2$ cells produce interleukins (IL4, IL5 and IL13) through its LCR, and naive $CD4^+$ T cells are induced along the $T_H2$ pathway. The induction of the IFN-γ guided naive $CD4^+$ T cells to $T_H1$ pathway. The authors found that the $T_H2$ locus interacts with the *Ifng* promoter on chromosome 10 of naive T cells. After differentiation into the $T_H1$ and $T_H2$ states, the two chromosomes were shown to move away from each other, and led to intra-chromosomal interactions. Deletion of a portion of the $T_H2$ LCR reduced the inter-chromosomal interaction and down regulated the expression of IFN-γ gene. The authors proposed that inter-chromosomal interaction "poises" the $T_H2$ and IFN-γ loci for transcription and subsequent $CD4^+$ T cell differentiation.

Similarly, Ling *et al*. (2006) applied 3C and FISH analysis to mouse fibroblast cells showing that the *H19*/*Igf* locus on chromosome 7 made contact with the genes *Wsb1* and *Nf1* on chromosome 11, and this physical contact resulted in the genes regulation of each other's expression. The factor mediating this contact was shown to be CTCF, a factor that regulates DNA methylation in mammals by simultaneously binding the imprinting control region of the maternal *H19*/*Igf2* locus and the paternal *Wsb1*/*Nf1*locus (Ling *et al*., 2006).

Perhaps the most striking study has been reported was the activation-dependent inter-chromosomal interaction occurred for olfactory receptor (OR) genes. This large superfamily consists of over 1300 genes that are organized in multiple clusters on chromosomes 18 and Y of the mouse genome. A recent study Lomvardas *et al.* (2006) using 3C analysis, followed by a combination of RNA FISH and DNA FISH, has shown that a single enhancer (H

enhancer), located on chromosome 14, is capable of multiple contacting with clusters of OR genes. The H enhancer activates OR genes at a very high frequency, and supports the view that a single regulatory element may control the expression of a large number or all members of a gene superfamily. However, in a follow-up study conducted by Fuss *et al.* (2007), deletion of the H enhancer was shown to only affect the OR genes in *cis*. Its effect was decreased on distant *trans*. OR genes, and it was undetectable on the expression of independent OR genes (not clustering), which seemed to be contradictory to what Lomvardas (2006) proposed. Fuss *et al.* (2007) suggested that high level of contacting between distant OR genes and H elements might simply be due to a positioning matter with regard to transcription, with no actual regulatory function. They also suggested that other elements similar to H enhancer might regulate OR genes. Taken together, the complexity of gene regulation suggests the need to study gene expression in depth to reveal the possible functional implications of any chromosomal interactions.

## 1.3 Gene Looping and Transcription Regulation

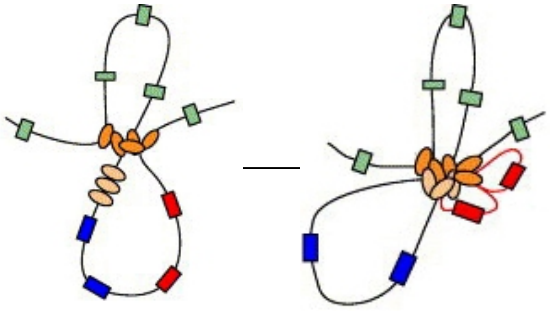
Transcription activation requires coordination between regulatory elements may be separated by long stretches of intervening DNA (Banerji *et al*., 1981), *e.g.* in the yeast genome upstream activating sequences (UAS), equivalent to enhancers, can locate more than 1 Kb away from the promoter they regulate (Barberis *et al*., 1995). There are three mechanisms have been suggested to explain this "distance effect" in transcription regulation, "looping", "scanning", and "linking". The "looping" model suggests a direct contact with intervening DNA looping out. The "scanning" model hypothesizes that the DNA bound transcription-activating complex linearly tracks along chromatin until it encounters a competent promoter (Herendeen *et al.,* 1992), which requires certain motor activity of the protein complex. The "linking" model suggests that a facilitator protein mediates communication between an enhancer and its cognate gene (Bulger and Groudine, 1999), which requires an extra helper. The looping model has long been considered to be more favoured over the other two models due to its simplicity, *e.g.* Ptashne (1986). Vilar and Saiz (2005) carried out a series of studies, using mathematical models, to specifically study the advantages of DNA looping. While DNA looping is common to many systems, the function of looping can differ between organisms. Commonly, DNA looping is regarded as a mechanism that allows the interaction between distantly bound regulatory proteins. By calculating the energies that contribute to stabilize DNA looping, the authors suggested that DNA looping is an adaptive cellular mechanism, with many advantages. Firstly, since the numbers of different molecular species are expected to differ from cell to cell, DNA looping

decreased the sensitivity of transcription to changes in the concentration of regulatory protein molecules, thereby suppressing cell-cell variability. Secondly, DNA looping reduces transcription fluctuation. If transcription switches slowly between active and inactive states, protein production over long periods of time will be with either full or no production. Therefore the number of protein molecules would fluctuate strongly between high and low values. The authors showed that DNA looping reduces transcription fluctuations by allowing a faster recycling of the regulatory elements. Finally, most biochemical reactions that occur within the nucleus are diffusion-limited reactions, *e.g.* molecules do not diffuse between each other, which may lead to transcription fluctuation. DNA looping bypasses diffusion-limited constraints that are imposed on reaction rates. The authors proposed that looping is the only valid mechanism that helps the cell to promote a fast-time scale in transcription processes naturally. These results provided evidence to support the "looping" mechanism as being more plausible than the other two hypotheses when it comes to understanding distant regulatory effects.

Looping has long been observed at the chromosome level, where two telomeres of certain chromosomes interact to form the whole chromosome loop (see Bystricky *et al.*, 2005) Chromosome looping provided some explanation for chromosome folding and telomere positioning, and is also an important mechanism to define a chromosomal territory (Bystricky *et al.*, 2005). The first evidence of the "DNA looping" came from the studies of the *Escherichia coli arabinose* operon (Englesbe *et al*., 1969), where it was shown that under certain conditions the *araBAD* promoter was repressed due to the simultaneous binding of regulatory proteins to sequences both proximal and distal to the promoter (Dunn *et al.*, 1984). DNA looping is now known to be present in both prokaryotes and eukaryotes, and generally involves the simultaneous binding of proteins or protein complexes to different sites of a DNA molecule, with the intervening DNA looping out to allow contact (Schleif *et al.*, 1992). This mechanism allows quick and efficient communication between regulatory sequences and their promoters.

DNA looping is also proposed to be the underlying mechanism of the formation of "transcription factories". de Laat *et al.* (2003) studied co-expression of clustering genes in murine β-globin locus. By applying *in vivo* 3C, in which live cells were treated with formaldehyde to fix the DNA structure, the authors found that in expressing tissue transcription initiation requires the formation of "active chromatin hub" (ACH) in which distal regulatory elements are brought together, with the intervening DNA looping out (Figure 2; de Laat *et al.,* 2003). The ACH is an analogue of a transcription factory. ACH

formation is mediated by the LCR of β-globin locus. The LCR locates at the upstream of the globin locus and contains four DNase I hypersensitive sites. A further study (Palstra *et al.,* 2003) showed that the "chromatin hub" (CH) was preserved when genes are paused for expression. Once activated, the CH transforms into an ACH enabling gene co-expression to take place (Figure 2). The presence of a CH suggested that the looping construct is stable and allows for fast resumption of the expression. This phenomenon, however, was not observed in nonexpressing loci, *e.g.* brain. It is therefore evident that DNA looping plays an important role in transcription regulation. Some trans-acting factors, such as Erythroid Kruppel-like factor (EKLF) (Drissen *et al.,* 2004), are required for the formation and/or stabilization of the looping.



"Paused" Chromatin Hub (CH) Active Chromatin Hub (ACH)

**Figure 2:** Analogue models for DNA loop formation of the murine β-globin locus. The thin black line represents chromatin and the coloured rectangles represent genes. The orange ovals and light brown ovals indicate hypersensitive sites and regulatory elements, respectively. For the ACH, the red rectangles represent the active genes, while green and blue rectangles represent the inactive genes.

Modified from Palstra *et al.* (2003)

DNA looping has been studied extensively in the budding yeast *S. cerevisiae*. O'Sullivan *et al.* (2004) used chromatin immunoprecipitation (CHIP) and 3C to reveal that the transcription factor Kin28 mediates the loop formation between promoter and terminator regions of two long genes, *FMP27* and *SEN1*, and looping was essential for transcription initiation. The authors also found that phosporylation of the C-terminal domain (CTD) of RNA polymerase II is crucial for loop formation. In a following study, Pestrascheck *et al.* (2005) showed that the binding of RNA polymerase II to an UAS and a promoter is facilitated by the formation of a DNA loop. Furthermore, Ansari and Hampsey (2005) studied another set of yeast genes and showed that upstream transcription factors interact with a 3'end-processing complex to induce transcription. All these studies proposed that DNA looping, in some cases, is essential for transcription activation by "recycling" transcription machinery to achieve a rapid transcription initiation rate.

## 1.4 Galactose Gene (*GAL*) Family in *Saccharomyces Cerevisiae*

The work reported in this thesis uses the budding yeast, *Saccharomyces cerevisiae* as a model system. *S. cerevisiae* is easy to grow and is biochemically and genetically well characterized, making this organism a perfect system for this type of study. The galactose (*GAL*) gene family of *S. cerevisiae* was particularly studied in this project because its expression is strictly regulated by the availability of carbohydrate. In the absence of glucose, the primary carbon source, the *GAL* gene family produces a series of proteins that coordinate to enable budding yeast cells to utilize galactose as an alternative carbon source (St. John and Davis, 1979a, b; Johnston, 1987). The *GAL* gene family is conserved and can be found in several other yeast species that are closely related to *S. cerevisiae* (Hittinger *et al.,* 2004). Since *GAL* expression is tightly controlled by nutrient availability, regulation of these genes is easy to manipulate experimentally (Johnston, 1987) and this, together with our current understanding from genetic and biochemical analyses over the past two decades, has placed this gene family at the forefront of model systems used for the study of eukaryotic gene regulation (Johnston, 1992; Ostergaard *et al*., 2000).

Members of the *GAL* gene family in *S. cerevisiae* can be subdivided into two categories based on their functions: structural or regulatory. The structural genes consist of *GAL1, GAL2, GAL7, GAL10,* and *MEL1* (*GAL5*), whereas the regulatory genes consist of *GAL3, GAL4,* and *GAL80* (St. John and Davis, 1979a, b). The structural genes *GAL1, GAL7* and *GAL10* are located on chromosome II, *GAL2* on chromosome XII, and *GAL5* on chromosome XIII (Johnston and Davis, 1984). The structural genes work in a coordinated manner to monitor the cell's use of galactose (Figure 3). The *GAL2* gene product transports galactose across the membrane into the cell, where it is converted to glucose 1-phosphate by the sequential action of the enzymes encoded by *GAL1, GAL7,* and *GAL10*. The *GAL5* gene product then converts glucose 1-phosphate to glucose 6-phosphate, which subsequently enters the glycolytic pathway (Johnston, 1992) (Figure 3). The *GAL* genes are tightly regulated by the carbon sources; their expression is strongly down-regulated in the presence of glucose, strongly induced in the presence of galactose, and paused for induction ("poised") in the presence of glycerol and lactate, where the genes are expressed at basal levels (St. John and Davis, 1979b). Their expression in the presence of galactose is extremely robust: the mRNA levels of *GAL1, GAL7* and *GAL10* each become 0.25-1% of total mRNA in the cell (Johnston, 1987). On the other hand, in the presence of glucose and glycerol/lactate, expression of *GAL1, GAL7* and *GAL10* is not detectable. Thus the mechanisms that activate and inhibit the expression of these genes are very powerful and effective.

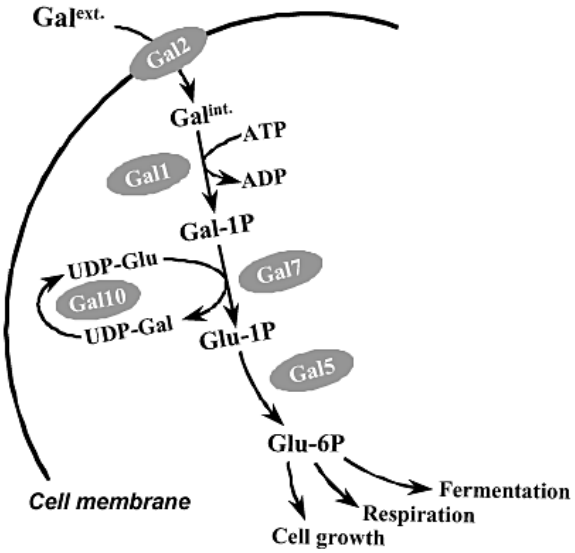



**Figure 3:** The galactose utilization pathway. Extracellular galactose is transported into the cell by Gal2 and subsequently converted to glucose-6-phosphate after several enzymatic steps, which are mediated by the products of *GAL* structural genes. Glucose-6-phosphate may be directed toward respiratory metabolism, fermentative metabolism, or conversion to various precursors required for cell growth. Gal2, galactose permease; Gal1, galactokinase; Gal7, galactose-1-phosphate uridylyltransferase; Gal10, UDP-glucose 4-epimerase; Gal5, phosphoglucomutase.

Modified from Ostergaard *et al*. (2000)

The *GAL* regulatory genes, *GAL3, GAL4,* and *GAL80*, are the major regulators of the *GAL* structural genes. These three genes are located on chromosomes IV, XVI, and XIII, respectively. The interplay of their gene products (Gal3p, Gal4p and Gal80p, respectively) determines the transcriptional status of the *GAL* genes in response to carbon source (Johnston, 1992; Ostergaard *et al*., 2000). Gal4p, the major transcription activator, is required for galactose-induction of *GAL* gene expression. Gal4p binds to upstream activating sequences ($UAS_{gal}$) of the *GAL* structural genes. This $UAS_{gal}$ element is essential for galactose-dependent gene activation. All genes known to be inducible by galactose contain one or more of these elements (Bash and Lohr, 2001). Gal80p, the major transcription inhibitor, binds to and masks Gal4p transcription activation in the absence of galactose. Gal3p, in turn, inhibits Gal80p repression in response to galactose, freeing Gal4p to activate *GAL* transcription (Ostergaard *et al*., 2000) (Figure 4). As part of the transcription activation mechanism, Gal4p forms a transcription complex comprising SAGA (Spt-Ada-Gcn5-Acetyltransferase) (Dudley *et al*., 1999), Mediator (Thompson *et al*., 1994), RNA polymerase II (Ptashne, 1997) and other general transcription factors (Johnston, 1992). The SAGA complex is a multi-protein complex, which is important for the transcription of many genes (Grant *et al*., 1998). The Mediator is a large multi-subunit complex that contains a TATA (a DNA sequence in the promoter region of the gene)-binding protein (TBP) and interacts with the carboxyl-terminal domain (CTD) of RNA polymerase II

(Thompson *et al*., 1994). RNA polymerase II may locate near the *GAL* promoter under repressed conditions (Cook, 1999). It has been suggested that Gal4p is able to localize genes preferentially to nuclear sites of facilitated transcriptional activity (Cook, 1999). Another activation function of Gal4p is to disrupt the TATA-bound nucleosome, which exposes the start site for the binding of transcriptional factors (Biggar, 1999). TATA-bound nucleosome disruption is a key aspect in *GAL* gene expression (Figure 4).

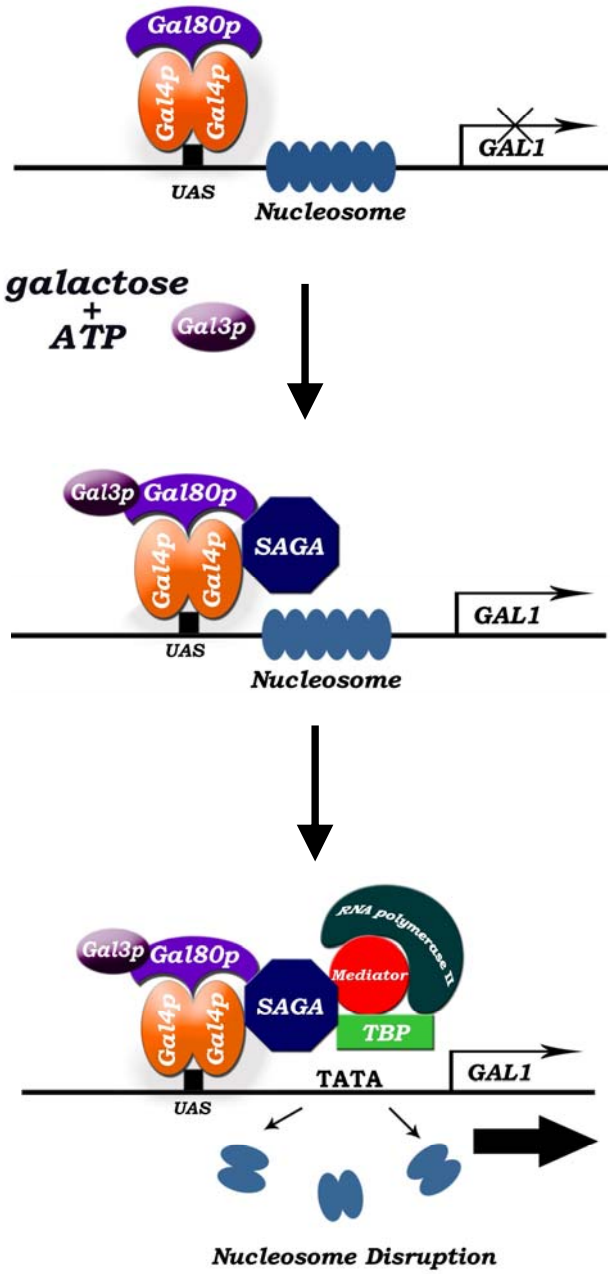



**Figure 4**: *GAL* gene regulation system. Gal4p acts as a transcriptional activator of the *GAL* genes by binding to upstream activating sites (UAS). During expression and repression of the *GAL* genes, Gal80p is bound to Gal4p. Intracellular galactose plays a pivotal role, being the signal molecule that (by an ATP-dependent mechanism) is bound to Gal3p, which acts as a transducer for galactose induction by interacting with Gal80p. Gal80p undergoes a conformational change and thus frees Gal4p. Gal80p and Gal4p remain in contact. The released Gal4p recruits coactivator SAGA and the Mediator complex. SAGA recruits TBP along with RNA Polymerase II to TATA binding site. This results in disruption of TATA-bound nucleosome and initiation of *GAL* transcription. Thin horizontal arrows above *GAL* indicate the basal level of *GAL* expression, while the thick horizontal arrow below *GAL* indicates robust *GAL* expression.

The expression patterns of *GAL3, GAL4* and *GAL80* are also dependent upon the availability of the carbon source. In the presence of galactose, the expression levels of *GAL3, GAL4* and *GAL80* are all several-fold higher than the basal level of transcription (Shimada, 1985; Giniger, 1988). *GAL4* expression is specifically controlled by the cell via the promoter elements (Griggs and Johnston, 1993), whereas *GAL80* is regulated by Gal4p and its own product, Gal80p (Shimada, 1985), and expression of *GAL3* is Gal4p and $UAS_{gal}$-dependent (Bajwa, *et al.*, 1988). Gal4p was shown to be phosphorylated in a manner that correlates with its ability to activate transcription (Mylin *et al.,* 1989, 1990). In cells expressing Gal80p, Gal4p is unphosphorylated in non-inducing conditions (*i.e.* presence of glycerol/lactate) but becomes phosphorylated upon induction by galactose. This was experimentally shown using sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) in which phosphorylated Gal4p migrated more slowly than unphosphorylated Gal4p (Mylin *et al.,* 1989). Gal4p is multiple phosphorylated at both the C and N terminal regions, and phosphorylation at the serine 669 position was suggested to be vital for galactose-inducible transcription to occur (Sadowski *et al*., 1996). Some phosphorylation of Gal4p is carried out by a cyclin-dependent kinase associated with RNA polymerase II (Hirst *et al*., 1999).

How *GAL3* activates *GAL* gene expression has long been debated and has been made difficult by the complex phenotypes exhibited by *GAL3* mutants in early days, *e.g.* the *gal3* strain is unable to use galactose efficiently (Douglas and Pelroy, 1963). Since then Gal3p was considered to have a catalytic role that converts galactose to an inducer or co-inducer to activate transcription of *GAL* gene (Tsuyumu and Adams, 1974). However, later evidence suggested that Gal3p itself is an inducer to activate transcription (Bhat *et al.,* 1990). Furthermore, Gal1p was identified as having Gal3p-like activity (Meyer *et al.,* 1990), therefore in the absence of Gal3p, *GAL* gene induction still occurs, but at a much slower rate (Bhat and Hopper, 1992). The currently most accepted model suggests that Gal3p is an inducer that exists in a "two-state" form: inactive and active (Bhat and Murthy, 2001) Galactose is thought to activate Gal3p and allow it to bind to Gal80p (Bhat and Hopper, 1992). Platt and Reece (2000) proposed that Gal3p is introduced to the upstream of the *GAL* gene in association with galactose and ATP, and interacts with Gal4p-Gal80p complex to form a tripartite Gal4p-Gal80p-Gal3p complex. In response, Gal80p undergoes a conformational change that releases Gal4p to recruit other regulatory proteins, allowing the activation of transcription (Platt and Reece 2000) (see Figure 4). However, this "two-state" model has been challenged by a recent successful *GAL3* mutation experiment (Lakshminarasimhan *et al.,* 2005). As a result of constructing a mutant *GAL3* with a non-

functional galactose recognition site, the authors proposed that Gal3p constantly interacts with Gal80p even in the absence of galactose. This interaction is weak until galactose is present (Lakshminarasimhan *et al.,* 2005). Other study also proposed that instead of forming a tripartite complex with Gal4p and Gal80p, Gal3p takes Gal80p away from Gal4p to activate transcription (Peng and Hopper, 2002). Gal80p is mobile to shuttle between the cytoplasm and the nucleus. In the presence of galactose, Gal3p sequesters Gal80p in the cytoplasm, resulting in the translocation of Gal80p from the nucleus to the cytoplasm (Peng and Hopper, 2002). The concentration of Gal80p in the nucleus is reduced, thus de-repressing Gal4p. Therefore, in summary, no agreement has been reached on how Gal3p regulates transcription activation.

Regarding the Gal4p transcription complex (Figure 4), recent work has shown that SAGA functions as a coactivator for *GAL* gene transcription in a Gal4p-dependent manner by recruiting TBP to the *GAL* gene upon galactose induction (Larschan *et al.*, 2001). Bryant *et al.* (2003) suggested that a sequential "recruitment" of different protein complexes at the *GAL1* promoter. Gal4p first recruits SAGA, then the mediator complex, and finally RNA polymerase and several other proteins including TBP, TFIIB and TFIIE. Interestingly, if SAGA was disrupted by a deletion mutation, the Mediator was still recruited, but transcription did not take place. Possibly due to that the RNA polymerase and associated proteins (*e.g.* TBP) are unable to bind to promoter sequences in the absence of SAGA. Bryant *et al.* (2003) also suggested that Gal4p directly contacts, but independently recruits SAGA and Mediator complexes. Another group, Cheng *et al.* (2004), tested the abilities of fusion proteins to activate transcription from an integrated reporter bearing a *GAL1* promoter with Zif (zinc finger DNA binding domain) sites upstream. The researchers constructed fusion proteins comprising Zif attached to one or another component of SAGA, Mediator, or TFIID. By comparing the abilities of tripartite fusion proteins, it was shown that Zif-Mediator-SAGA activated transcription as efficiently as Zif-Gal4p. Other combinations of fusion proteins, such as with Mediator or SAGA only, showed less efficiency in transcription activation. The results reinforced the idea that Gal4p activates transcription by requiring co-existence of SAGA and Mediator. Furthermore, these results also suggested that Gal4p activates transcription by making contacts with multiple transcription factors with a different contribution to the activated level of transcription, varying from critical to not functionally important. The experiment described by Cheng *et al.* (2004) using reporters that were integrated into the chromosome instead of a plasmid. However, the modifications that occur at the chromosomal level, such as chromosome folding, chromatin looping and gene copy numbers were not taken into considerations. The

authors observed a Zif-Mediator fusion protein works well on its own when the template is carried on a plasmid but not when integrated into the chromosome. Furthermore, Zif-Mediator-SAGA works approximately twice as efficiently as Zif-SAGA-Mediator does in terms of activating transcription. The authors failed to give an explanation on these matters. This might be due to that Mediator binds less efficiently to RNA polymerase II in Zif-SAGA-Mediator, or is possibly a result of the alteration of the chromosome structure.

## 1.5 Methodology

In the study reported in this thesis, the major technique applied is Chromosome Conformation Capture (3C), which is a novel molecular technique to detect the frequency of interaction between any two loci. It was initially proposed by Dekker *et al*. (2002, 2003) and has been widely applied and yielded reliable and reproducible data (Spellman *et al.,* 2002; Palstra *et al*., 2003; O'Sullivan *et al.*, 2004; Ansari and Hampsey, 2005; Bystricky *et al*., 2005; Spilianakis *et al.*, 2005).

Higher order structural features of chromosomes, such as the chromatin fibre, chromatin looping and inter-chromosomal association, are important for chromosome morphogenesis and have roles in gene expression and recombination (Cohen *et al.,* 2000). As many of the traditional methods used to studying chromosome conformation have gradually shown their limitations, better and efficient techniques have seen in high demand. High-resolution electron microscopy requires hard work to study specific loci, while light microscopy has insufficient resolution to study at the chromosome level; fluorescence labelling of DNA-protein either only permits the examination of a few positions in gene loci (GFP), or uses a severe treatment that may affect chromosome organization (FISH) (Vilar and Saiz, 2005). It was given these circumstances that 3C (see Chapter 2: section 2.3.5) was invented by Dekker *et al.* (2002) to overcome the shortcomings of the traditional methods. The principle of 3C is to fix DNA structure by cross-linking DNA and proteins with formaldehyde treatment. To quantify the cross-linking frequency of different sites, cross-linked DNA is digested by a restriction enzyme and subjected to ligation at a low DNA concentration. These conditions enable the preferential intra-molecular ligation of the cross-linked restriction fragments, rather than ligation of random fragments. Cross-linking is then reversed and individual ligation products are detected and quantified by the polymerase chain reaction (PCR) using locus-specific primers. A control template is generated in which all-possible ligation products are present (Dekker *et al.,* 2003). The original 3C method

proposed by Dekker *et al*. (2002) was used to analyse chromosomal interactions in yeast genome by cross-linking in isolated nuclei. Lately, this method has been applied more frequently to study transcription-dependent conformational changes in intact cells, which is transient and can avoid the artefacts that result from the isolation of nuclei. Thus, formaldehyde cross-linking was performed in actively growing cells prior to digestion and ligation (Palstra *et al*., 2003; O'Sullivan *et al.*, 2004).

Two modified 3C methods: Circular Chromosome Conformation Capture (4C) (described in Chapter 2: section 2.3.6) that detects circular ligated products, and 4C pulldown (described in Chapter 2: section 2.3.7) in which target protein is isolated by antibody after ligation, were also employed in the study reported in this thesis. Furthermore, real-time PCR was applied to study the frequencies of observed interactions in addition to standard PCR (see Chapter 2: section 2.4).

## 1.6 Research Aims and Objectives

The aim of this project was to study potential interactions occurring at the *GAL* gene locus on chromosome II in response to different carbon source induction. These interactions can be of either intra- or inter-chromosomal. Should any such interaction be identified; further study was to identify the regulators/ transcription factors that are responsible for such an interaction.

**Objective 1:** To identify if any intra-chromosomal interactions occur at the *GAL* locus on chromosome II when their expression are regulated by the presence of glucose, glycerol/lactate or galactose.

**Objective 2:** To identify if any inter-chromosomal interactions occur at the *GAL* locus with other chromosomes when their expression are regulated by the presence of glucose, glycerol/lactate or galactose.

**Objective 3:** To identify the transcriptional factors that involve in mediating the observed interactions. Target proteins include Gal4p, Gal80p and others in transcription regulation.

# Chapter 2----Materials and Methods

## 2.1 Cultures and Plasmid

### 2.1.1 Cultures and Plasmid used

The strains and plasmid used are detailed in Table 1. *Saccharomyces cerevisiae* strains were stored as glycerol stocks (final concentration of glycerol 20%, v/v) at -80°C.

**Table 1:** *S. cerevisiae* strains and the plasmid used in this study

| ***S. cerevisiae* Strains** | **Genotype** | **Description** | **Source or Reference** |
|---|---|---|---|
| W303-1a | *MATa ade2-1 his3-11,15 leu2-3,112 can1-100 trp1-1 ura3-1* | Wild-type strain used in most of the studies | This study |
| YPL248C | *MATa ade2 arg4 leu2-3 112 trp1-289 ura3-52* | TAP-tagged Gal4p strain | Open Biosystems® |
| BY4741 | *MATa his3Δ1 leu2Δ0 met15Δ ura3Δ0* | Wild-type strain for *Δgal80p* control | Research Genetics |
| YML051W | *MATa his3Δ1 leu2Δ0 met15Δ ura3Δ0 gal80::Kan$^R$* | *Δgal80p* deletion mutant strain | Research Genetics |
| FY23 | *MAT*a *trp1-Δ63 leu2-Δ1 ura3-52* | Wild-type strain for *Δrat1-1* control | Cole, C (1992) |
| DAT1-2 | *MAT*a *trp1-Δ63 leu2-Δ1 ura3-52 rat1-1* | *RAT1-1* temperature-sensitive strain | Cole, C (1992) |
| ***E. coli* Strain** | | | |
| *DH5α* | *F− Φ80d lacZΔM15 Δ(lacZYA-argF)U169 endA1 recA1 hsdR17(r−K12 m+K12) deoR supE44 thi-1 λ-gyrA96 relA1* | | Invitrogen (Carlsbad, CA, USA) |
| **Plasmid** | | | |
| pUC18 | Amp$^R$, *lac* promoter | | Invitrogen (Carlsbad, CA, USA) |

### 2.1.2 Growth of Fungal Cultures

*S. cerevisiae* strains were selected for 3C, 4C and 4C pulldown assays. These cultures were grown at 30°C (wild-type) or 23°C and 37°C (temperature sensitive) on Synthetic Complete (SC) agar plates or SC in broth with shaking at 160 rpm (Appendix 1.0) for 1-3 days or until growth was visible. SC medium is based upon Yeast Nitrogen Base (YNB) supplemented with a complete mixture of amino acids and vitamins and carbon source (glucose, glycerol/lactate, or galactose). Selected yeast strains grew in SC-X agar/medium that is based upon YNB supplemented with a "drop-out" mixture in which certain amino acids (X) are missing. Cultures were generally grown in the presence of glucose unless otherwise stated. The plates were then stored at 4°C.

For liquid cultures, 50 ml flasks containing 10 ml of SC/SC-X medium with the appropriate amino acid supplement and carbon source were inoculated from stock culture grown on plates. Incubations were at the selected temperature, shaking (160 rpm) overnight or until sufficient growth was obtained. These starter cultures were sub-inoculated the next day into 250 ml flasks containing 100 ml of SC/SC-X medium with the appropriate amino acid supplement and carbon source. These subcultures were grown, with shaking, overnight or until $OD_{600}$=0.6.

## 2.2 Nucleic Acid Extractions

### 2.2.1 Genomic DNA

Genomic DNA was prepared from *S. cerevisiae* W303-1a according to Amberg's protocol (http://www.upstate.edu/biochem/amberg/protocols/yeast_genomic_DNA.php). Briefly, 10 ml cultures were grown overnight (O/N) in SC medium in a 50 ml conical flask to a density $1x10^8$ cells per ml, and then centrifuged at 4500 × *g* for 3 min. The pellet was then washed with 6 ml of MilliQ $H_2O$, and aliquoted into 4 × 1.5 ml microcentrifuge tubes to centrifuge down. The supernatant was decanted and the pellet was vortexed until it was suspended. 200 µl of solution A (2% Triton-X-100, 1% SDS, 0.1 M NaCl, 0.01 M Tris-HCl, pH 8, and 1 mM EDTA), 200 µl of phenol/chloroform (1:1 v/v) and 0.3 g of acid washed glass beads were then added to the sample. The tubes were vortexed for 2 min, then 200 µl of TE buffer (0.01 M Tris-HCl, and 1mM EDTA, pH 8) was added and the tubes were centrifuged at 13,000 × *g* for 5 min. The supernatant was transferred to a new tube and DNA was precipitated with the addition of 1 ml of 100% EtOH, followed by inversion to mix well.

The samples were then centrifuged again for 2 min, the supernatant was decanted and the nucleic acid pellet was resuspended in 0.4 ml of TE buffer containing 1 µl of 30 µg/ml RNase A. After incubation for 5 min at 37°C, the DNA was re-precipitated with the addition of 18 µl 5 M $NH_4OAc$ and 1 ml 100% EtOH and incubated at -20°C for several hours. DNA was collected by centrifugation at 13,000 × *g* for 10 min and washed with 70% EtOH. The DNA pellet was then re-suspended in 40-60 µl of water, and stored at – 20°C.

### 2.2.2 Plasmid Preparation

Cultures of *E. coli* transformed with pUC18 (Appendix 4.0) were grown with shaking (160 rpm) at 37°C O/N in Luria Broth (LB) media supplemented with 100 µg/ml ampicillin (50 ml). Plasmid DNA was then extracted and purified using a Geno Pure Plasmid Midi Kit (Roche, Basel, Switzerland) according to the manufacturer' s instructions. The concentration and purity of the plasmid was determined using a Nanodrop ND-1000 (Nanodrop Technologies, Wilmington, DE, USA).

## 2.3 Molecular Techniques

### 2.3.1 Polymerase Chain Reaction (PCR)

General reaction conditions for amplifying DNA were: an initial step of denaturing DNA double strands at 95°C for 2 min, followed by 40 cycles of denaturation for 15 s, primer annealing for 30 s, and primer elongation for 30 s. A final elongation step of 2 min was included to ensure complete extension of the PCR products. Specific primer annealing and extension conditions depended on the primer pair used and the length of DNA being amplified. Specific conditions included variations in annealing/extension temperature, number of cycles and the addition of reaction enhancers, such as bovine serum albumin (BSA), dimethylsulphoxide (DMSO) and the detergent Tween 20. The variations applied are specified later in the text. All PCR reactions were set up on ice using a cocktail (master mix) that contained all the reagents required for n+1 PCR reactions. The PCR machine being used was ICYCLER (Bio-Rad, Hercules, CA, USA). Following amplification, reactions were stored at 4°C and the products were visualized over UV light after agarose gel electrophoresis and ethidium bromide staining (Section 2.3.3).

### 2.3.2 Primer Design and Restriction Enzyme Identification

PCR primers were designed using Primer 3 (primer3/input.htm version 0.3.0 modified for WI) (Rozen and Skaletsky, 2000). Sequence analysis of yeast genome (*www.yeastgenome.org*) was used to identify the restriction enzymes that were suitable to be applied in 3C, 4C, and 4C pull-down.

### 2.3.3 Gel Electrophoresis

DNA fragments were separated according to sizes using gel electrophoresis. Gels typically consisted of 0.75, 1.5 or 2% (w/v) agarose in 0.5 × Tris/Borate/EDTA (TBE), dependent upon the expected PCR product size. 0.1 volume of gel loading buffer was added to DNA samples before loading. Samples were run against a standard (1.5 µl of 100 bp DNA ladder; Fermentas, Maryland, USA). Electrophoresis was generally carried out at 80-100 volts for approximately 40 min or until the loading dye had migrated about 3/4 of the way along the gel. After electrophoresis, agarose gels were stained in 0.1 mg/ml ethidium bromide for 10-20 min before being briefly destained in water. The DNA fragments were then visualized over UV light and photographed using a Bio-Rad Gel Documentation system (Bio-Rad, Hercules, CA, USA).

### 2.3.4 DNA Extraction from Agarose Gels

DNA was extracted from agarose gels using the QIAquick Gel Extraction kit (Qiagen, Valencia, CA, USA) according to the manufacturer's instructions. The yield and purity of the extracted DNA was determined by agarose gel (0.75% for genomic DNA, or 1.5% for PCR products) electrophoresis and Nanodrop analysis.

### 2.3.5 DNA Purification

Tris-HCl (pH 8.0) saturated phenol and chloroform were combined at a 1:1 (v/v) ratio, and an equal volume of this mixture was added to the sample to be extracted. The sample was then mixed and the phases separated by centrifugation at 7500 × *g* for seven min. The aqueous phase was transferred to a fresh tube, and then extraction was repeated (×3) until the organic/aqueous interface appeared free of protein. The DNA in the aqueous phase was precipitated by the addition of one tenth volume of 3 M sodium acetate (pH 5.5), 5 µl of

0.5% w/v linear polyacrylamide (LPA) and 2 volumes of cold absolute ethanol. To enable precipitation, the reactants were stored at -20°C for at least two hours. Precipitated DNA was recovered by centrifugation at 13,000 × *g* for 20 min, the supernatant was decanted and the pellet was washed once with 70% EtOH and air dried before being resuspended in sterile water.

### 2.3.6 Chromosome Conformation Capture (3C) Analysis

Chromosome Conformation Capture (3C) analysis was based on a previously described protocol (Dekker *et al.,* 2002), and is diagrammatically represented in Figure 5 (1). The essential elements are chromatin extraction, chromatin digestion, ligation, and purification. These elements are described in detail below.

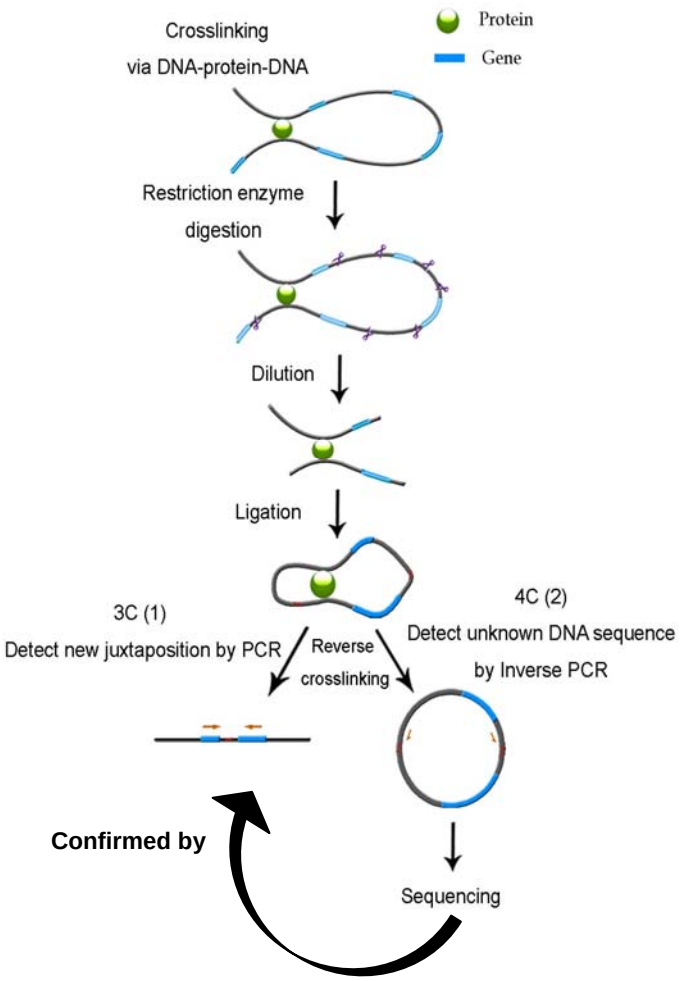



**Figure 5:** Schematic representation of the (1) 3C and (2) 4C assays. The grey folded line represents a section of chromosome and the blue short ribbons are individual genes. Purple scissors denote restriction enzyme sites. Dilution prior to ligation favours intramolecular ligation. Orange arrows indicate the directions of primer pairs for PCR analyses. The sequencing derived from 4C is confirmed by 3C PCR.

### *2.3.6.1 Chromatin Extraction*

An overnight culture of yeast strain was grown in 10 ml of SC/SC-X medium supplemented with 3% v/v glycerol, 2% w/v [*DL*] lactic acid and 0.05% w/v glucose, pH 5.7, served as the starting culture. This culture was used to inoculate 50 ml of SC/SC-X medium supplemented as above and grown at required temperature with shaking until log phase ($OD_{600}$=0.6; approximately $2.5\times10^7$ cells per ml). *GAL* gene expression was manipulated by dilution with an equal volume of pre-warmed SC/SC-X medium supplemented with 4% (w/v) glucose (*repressed*), 6%/4% (w/v) glycerol/lactate (*noninduced*) or 4% (w/v) galactose (*induced*) as sole carbon source. The cultures were then grown at the required temperature for the required time before chromatin cross-linking. Chromatin was cross-linked in live cells by treating the cells with formaldehyde (1% v/v) for 2 min with shaking, and then quenched by the addition of glycine to 125 mM for 5 min. The treated cells were collected and washed once with ice cold PBS with 1% (v/v) Triton X-100, before $0.95\times10^9$ cells were resuspended in 400 µl of FA-Lysis Buffer (Appendix 2.1). The cells were disrupted by vortexing (eight periods of 30 s, with intervals of 1 min on ice) in the presence of 500 µl of acid washed glass beads. The cell extract was released through a needle hole, and centrifuged in a bench-top centrifuge at 13,000 × *g* for 10 min. The supernatant was decanted and the pellet was washed once with 500 µl of FA-Lysis Buffer and then resuspended in 500 µl of Chromatin Digestion Buffer (Appendix 2.1). The sample was then treated with 1% (v/v) SDS at 37°C for 10 min to relax the chromatin structure and the SDS treatment was stopped by the addition of 1% (v/v) Triton X-100. The freshly made chromatin was stored at -80°C.

### *2.3.6.2 Chromatin Digestion*

Chromatin was digested using the selected restriction enzyme *Csp*61 (Promega) at 37°C for 2 hrs in the buffer specified by the manufacturer (10 × Digestion Buffer). The digestion (100 µl total volume) comprised the following components (added in the order presented): chromatin sample ($0.95\times10^9$ cells), 52.6 µl; 10 × Digestion Buffer, 10 µl; MilliQ water, 27.4 µl; *Csp*61 (10 U/µl), 10 µl. The restriction enzyme was inactivated by adding SDS to 1% (v/v) and incubation at 65°C for 20 min. The SDS was then quenched by the addition of 182 µl 11% (v/v) Triton X-100.

### *2.3.6.3 Chromatin Ligation*

Digested chromatin samples were diluted (1:20) in the buffer specified by the manufacturer (Invitrogen) prior to ligation to enable the preferential intramolecular ligation of the cross-linked restricted fragments. The digestion (2 ml total volume) comprised the following components (added in the order presented): digested sample containing 1% SDS, 110 µl; 11% Triton X-100, 182 µl; MilliQ $H_2O$, 1280 µl; 5 × Ligase Buffer, 400 µl; T4 DNA ligase (1 U/µl;), 10 µl. Ligation was performed at 16°C for 2 hrs. The reaction was stopped by inactivating the ligase enzyme with the addition of 10 µl (final concentration 1 %) EDTA at room temperature (RT).

### *2.3.6.4 Chromatin Reverse Cross-linking*

Following chromatin ligation, chromatin with linked proteins were reverse cross-linked by incubating at 65°C O/N with 0.5 µl of Proteinase K (10 mg/ml) (Roche Diagnostic, Mannheim, Germany) in high salt conditions (0.2 M NaCl, 4 mM Tris-HCl, pH 7.5).

### *2.3.6.5 DNA Isolation and Purification*

Reverse cross-linked chromatin was then treated with RNase to 2 ng/µl at 37°C for 15 min, and the DNA was isolated and purified by standard phenol/chloroform extraction and ethanol precipitation (Section 2.3.5).

### *2.3.6.6 3C PCR Analysis*

PCR was used to detect any ligated products. Positive controls were made by cutting and ligating genomic DNA to generate non-specific cross-linked fragments (Sections 2.3.6.7). A no template control (NTC), *i.e.,* water as template, was used in all reactions. PCR reaction volumes were 30 µl and comprised the following: DNA template, 1 µl; 10 µM Forward Primer, 0.3 µl; 10 µM Reverse Primer, 0.3 µl; 10 × PCR Buffer, 3 µl; dNTPs, 4.8 µl; BSA (100 mg/ml), 3 µl; MilliQ $H_2O$, 17.4 µl; *Taq* polymerase (1 U/µl), 0.2 µl. In each reaction the final concentrations of reagents were therefore 1 × PCR buffer, 0.2 mM dNTPs, 1 mg/ml BSA, and 0.2 U *Taq* DNA polymerase.

The thermocycling conditions were standardized to an initial step of 95°C for 2 min, followed by 40 cycles of denaturation at 95°C for 15s, annealing at 55-59°C for 30 s and elongation at 75°C for 1 min, followed by a final elongation step at 75°C for 5 min as shown below. The annealing temperature was varied for different primer pairs.

PCR products were examined by agarose gel electrophoresis (Section 2.3.3)

.

#### *2.3.6.7 3C PCR Positive Controls*

Purified W303-1a genomic DNA (Section 2.3.1) was used as the template in the PCR reactions to amplify the 2kb products. The PCR reaction mix that was used to amplify the region between *GAL10-GAL1* were: dNTPs, 16 µl; 10× PCR Buffer, 10 µl; primer pair mix, 2 µl; MilliQ water, 66 µl; *Taq* polymerase (5 U/µl), 1 µl; PCR template (100 ng/µl), 5 µl. The reaction conditions were modified from the standard PCR program (Section 2.3.1) to 30 cycles of 95 °C for 15 s, 56 °C for 30 s, and 72 °C for 2 min. Extension time was increased to ensure the 2kb DNA product was fully amplified.

PCR amplification of the region between *GAL7-GAL10* was optimised by the addition of BSA, which increases the yield of PCR product (Leon Huynem, personal communication). The reaction mix comprised: dNTPs, 4.8 µl; 10 × PCR Buffer, 3 µl; primer pair mix, 0.6 µl; MilliQ water, 17.4 µl; BSA (10 mg/ml), 3 µl; *Taq* polymerase (5 U/µl), 0.2 µl; PCR template (100 ng/µl), 1 µl. The reaction conditions were similar to those above but with an annealing temperature at 54°C, rather than 56°C.

The 2kb PCR products used as positive controls were fractionated using 0.07% (w/v) agarose gel electrophoresis. DNA was isolated, purified, and their concentrations were determined (Section 2.3.4). The purified DNA was the starting material for the amplification of the positive control and was digested with *Csp*61, ligated and then purified through phenol/chloroform extraction (Section 2.3.5.5).

### 2.3.7 Circular Chromosome Conformation Capture (4C) Analysis

4C analysis was adapted from the previously described protocol (Zhao *et al.,* 2006) with some modifications (Figure 5 (2)). Briefly, chromatin digestion, dilution, ligation, and purification were performed (as outlined in section 2.3.5).

### *2.3.7.1 4C PCR Analysis*

Nested PCR was used in 4C studies to increase the specificity of the first round PCR. PCR was carried out with two pairs of degenerate primers that were designed to the targeted DNA fragment. The orientation of the primer pairs relative to each other was deliberate and enabled nesting of primers. The products of the 1st PCR were diluted 1:100 and 2 µl was used as the template for the 2nd round PCR. The PCR conditions were outlined as follows.

For the outer PCR (*i.e.* first round), the PCR master mix comprised a total volume of 15 µl containing: template, 1 µl; 10 µM Forward Primer, 0.225 µl; 10 µM Reverse Primer, 0.225 µl; 10 × PCR Buffer, 1.5 µl; dNTPs (10mM each), 2.4 µl; 20% Tween 20, 0.15 µl; 100% DMSO, 0.45 µl; MilliQ $H_2O$, 8.95 µl; *Taq* polymerase (1 U/µl), 0.1 µl. The PCR program comprised an initial step of 98°C for 30 s, followed by 30 cycles of 98°C for 15 s, 59°C for 30 s and 72°C for 1.5 min.

For the nested PCR (*i.e.* second round), the PCR master mix comprised a total volume of 30 µl containing: template, 2 µl; 10 µM Forward Primer, 0.3 µl; 10 µM Reverse Primer, 0.3 µl; 10 × PCR Buffer, 3 µl; dNTPs, 4.8 µl; MilliQ $H_2O$, 19.4 µl; *Taq* polymerase (1 U/µl), 0.2 µl. The PCR program comprised an initial step of 95°C for 2 min, followed by 30 cycles of 95°C for 30 s, 59°C for 30 s and 72°C for 1.5 min. On completion, the reactants were held at 20°C.

### *2.3.7.2 4C Generation of Library Inserts*

4C libraries produced by nested PCR were purified using a Roche High Pure PCR Purification kit (Roche Diagnostic, Mannheim, Germany) according to the manufacturer's instructions, and eluted in 50 µl of elution buffer before digestion. Digestion was performed at 37°C for 2 hr and comprised a total volume of 50 µl containing: 4C library, 44 µl; 10 × React3 Buffer (Invitrogen), 5 µl; *Eco*RI (1 U; Invitrogen), 1 µl. The reaction was stopped by inactivating the restriction enzyme through heating at 65°C for 20 min.

### *2.3.7.3 Cloning-Digestion and Dephosphorylation of pUC18*

Plasmid pUC18 was obtained from Invitrogen. Digestion of pUC18 was performed at 37°C for 4 hr and comprised a total volume of 20 µl containing: pUC18 (0.28 µg/µl), 10 µl; 10×

React3 Buffer, 2 µl; MilliQ $H_2O$, 7 µl; *Eco*RI (1 U/µl), 1 µl. The reaction was stopped by inactivation of the restriction enzyme through heating at 65°C for 20 min.

pUC18 was then dephosphorylated at 37°C for 30 min and comprised a total volume of 25 µl containing: pUC18 (0.28 µg/µl), 20 µl; 10 × CIAP Buffer (Fermentas, Vilnius, Lithuania), 2.5 µl; MilliQ $H_2O$, 1.5 µl; CIAP (1 U/µl), 1 µl. The reaction was inactivated by heating at 85°C for 15 min.

#### *2.3.7.4 Ligation of 4C Library inserts to Dephosphorylated pUC18*

*Eco*RI digested 4C library inserts were ligated into dephosphorylated pUC18 using a reaction that comprised: 4C digested sample, 5 µl; pUC18, 1 µl; MilliQ $H_2O$, 9 µl; 5× Ligase Buffer, 4 µl; ligase (10 U/µl; Invitrogen), 1 µl. The reaction was incubated O/N at 4°C, and the pUC18-4C ligation product was then purified using a Zymo Clean and Concentrator kit (Zymo, Orange, CA, USA) following the manufacturer's instructions with elution in 15 µl of MilliQ $H_2O$.

#### *2.3.7.5 Transformation of Electro-competent E. coli DH5α Cells*

Electro-competent *E. coli DH5α* cells were prepared (Appendix 5.0). 5 µl of purified pUC18–4C ligation products were mixed with 20 µl of *DH5α* competent cells in a pre-chilled electroporation cuvette. The cells were subjected to one pulse (settings: 2000 volt, 25 µF, 200 ) in a Bio-Rad Gene Pulsar II electroporator, and then allowed to recover in 1 ml SOC media (Appendix 1.0) by incubating at 37°C for 1 hr with shaking (160 rpm). The transformed cells were then streaked onto LB plates containing Ampicillin (200 ug/ml) and BCIG (40 µg/ml). The plates were incubated at 37°C until colonies were visible (usually O/N). The transformation controls included uncut pUC18 plasmid and cut pUC18 plasmid (control for ligation), dephosphorylated cut pUC18 plasmid (control for phosphatase), and MilliQ $H_2O$ (control for contamination). See Appendix 11.0 for GMO approval.

Blue-white selection was used to identify white colonies indicating the presence of plasmids with inserts. White colonies were picked using autoclaved toothpicks, and then streaked onto numbered screen plates for later reference. The toothpick was then dipped into 25 µl of MilliQ $H_2O$ to release any remaining cells. These cells in 25 µl of MilliQ $H_2O$ were heated at 95°C for 10 min to release DNA for PCR analysis using plasmid-specific primer M13F.

### *2.3.7.6 Analysis of Cloned 4C Library*

Clones that contained inserts, as determined by PCR analysis, were analysed further by digestion with the restriction enzyme *Eco*RI (1 U/µl). This extra step ensured that the insert sizes observed were not the result of multiple insert ligation. The reaction comprised a total volume of 50 µl containing: PCR product, 43 µl; 10 × NE Buffer 3, 5 µl; MilliQ $H_2O$, 1 µl; *Eco*RI, 1 µl, and was performed at 37°C for 2-3 hr before size fractionation by gel electrophoresis.

Following confirmation of the identities of the clones that were indeed positive for the desired insert, these positive plasmid containing colonies were picked from the numbered screen plate and grown at 37°C, O/N, in 10 ml of LB with ampicillin resistance ($Amp^R$). Plasmid DNA was extracted from these colonies using a Roche High Pure Plasmid Isolation kit as outlined by the manufacturer. Purified plasmid DNA was sent for DNA sequencing (AWCGS, Albany Campus, Auckland, New Zealand). The sequencing reaction comprised 300 ng plasmid, 2 µl M13F primer (3.2 pmol) and up to 15 µl MilliQ $H_2O$. Sequences were analysed in Sequencher™ and compared with the *S. cerevisiae* genome using BLAST (NCBI).

## 2.3.8 4C Pulldown

The 4C pulldown assay (Figure 6; O'Sullivan *et al.*, unpublished) is similar to 3C analysis but with an additional antibody isolation step. The antibody, rabbit anti-peroxidase IgG, targets a Tandem Affinity Purification (TAP) tag consisting of a calmodulin binding peptide (CBP), a tobacco etch virus protease (TEV) cleavage site, and *Staphylococcus aureus* protein A that binds tightly to IgG, as well as a selectable marker. The TAP tag fuses to the C-terminus of the protein under study. Whole cell extracts of TAP-tagged yeast strains are immunoblotted and probed with IgG.

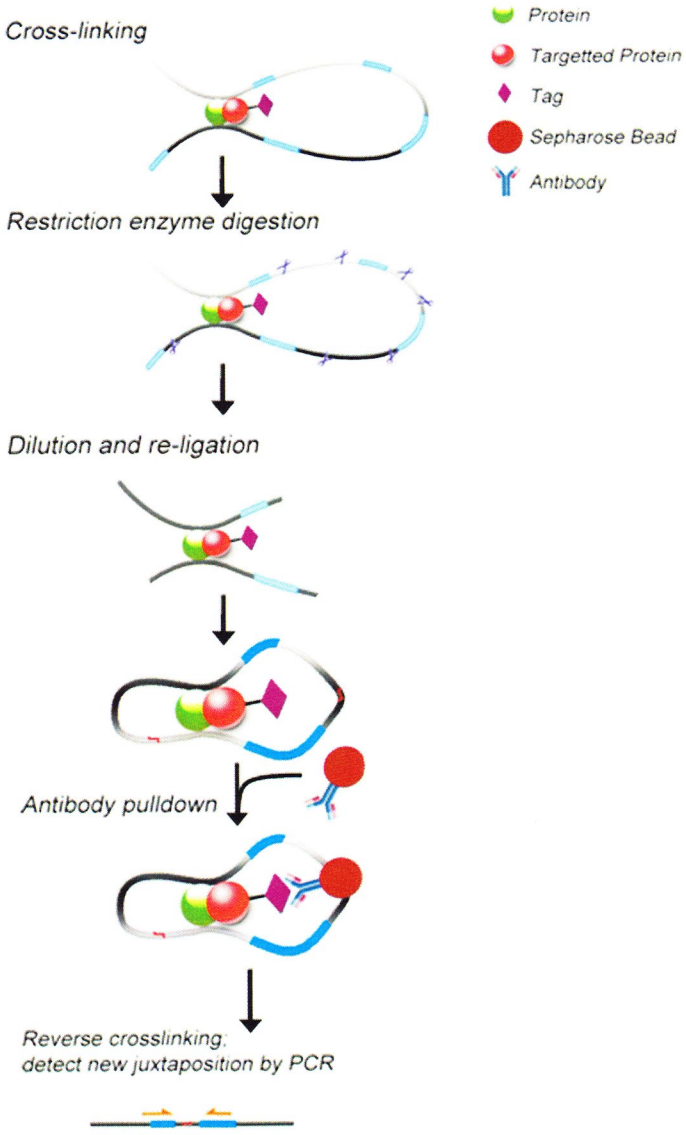



**Figure 6:** Schematic representation of the 4C pulldown assay. This technique is essentially related to 3C assay (Figure 5). An extra step of antibody isolation of the tagged targeted protein follows ligation. PCR analyses using appropriate primer pairs to detect the interacting loci. The grey folded line represents a section of chromosome and the short blue ribbons are individual genes. The orange arrows indicate the directions of primer pairs.

### *2.3.8.1 Protein Isolation ("Pulldown")*

Chromatin digestion, dilution and ligation were performed according to the 3C protocol (Section 2.3.6). After ligation, the sample was split into 2 × 1000 µl aliquots to allow subsequent ligation in 1.5 ml microcentrifuge tubes (Axygen, Union City, CA, USA). Each sample (1000 µl) was pre-cleaned by rotation at 4°C for 1 hr with 40 µl Sephacryl 5300 Beads (Pharmacia, Uppsala, Sweden) in 330 µl IP-like Buffer (Appendix 2.1) to remove all non-specific proteins and DNA. The resulting preparations were termed pre-cleaned ligated samples, and were transferred to beads coated with IgG (antibody that isolates TAP-tagged

protein) that had been treated to block non-specific binding sites. The beads were prepared as follows.

Two sets of IgG beads (Sigma, St. Louis, MO, USA), 30 µl and 20 µl respectively, were used in order to maximize protein isolation efficiency. They were washed by rotation at RT for 30 min with 5% BSA and 3 mg sonicated λ DNA in 150 µl IP Buffer (Appendix 2.1). The reaction for 30 µl IgG beads comprised 15 µl BSA, 10 µl sonicated λ DNA, and 150 µl IP Buffer; the reaction for 20 µl IgG beads comprised 10 µl BSA, 10 µl sonicated λ DNA, and 150 µl IP Buffer. These washed beads were collected by centrifuging at 2,200 × *g* for 1 min, and the supernatants were decanted. The two sets of beads were then washed once with 150 µl of IP buffer.

The pre-cleaned ligated sample was mixed with the 1$^{st}$ set of pre-blocked IgG beads (30 µl) and incubated, with rotation, at 4°C for 2 hr. Following centrifugation at 2,200 × *g* for 1 min, the supernatant was then mixed with the 2$^{nd}$ set of pre-blocked IgG beads (20 µl) and incubated for a further hour, with rotation, at 4°C. Following centrifugation at 2,200 × *g* for 1 min, the supernatant was transferred to a clean tube, and referred to as the depleted sample.

Two sets of IgG beads were combined and were subjected to a series of washing steps with Buffers A-E (Appendix 2.1). This series comprised 1 wash with 1 ml Buffer A, 2 washes with 1 ml Buffer B, 4 washes with 1 ml Buffer C, 2 washes with 1 ml Buffer D, and 2 washes with 1 ml Buffer E. Each wash comprised 5 min incubation at RT before centrifugation at 2,200 × *g* for 1 min.

The washed IgG beads were then resuspended in 300 µl of Buffer E. DNA fragments attached to TAP-tagged proteins were released by digestion with TEV protease (1 U) at 25°C for 1 hr. The beads were then centrifuged at 2,200 × *g* for 1 min and the supernatant was transferred into clean tubes: these were the pulldown samples.

The pulldown samples, containing linked protein/DNA fragments, were reverse cross-linked as described in 3C (Section 2.3.6.4).

#### *2.3.8.2 PCR Analyses for 4C Pulldown*

The presence of intra/inter-molecular ligation products was determined by PCR. Positive and negative controls were included as for 3C (Section 2.3.6.6).

## 2.4 Determination of Interaction Frequency by Real-time PCR

### 2.4.1 *SyBr* Green Real-time PCR Application

The DNA template concentration of the PCR reaction was determined using a *SyBr* green mix package from Invitrogen, consisting of *SyBr* Mix (comprises SYBR Green 1 Dye, AmpliTaq Gold® DNA Polymerase, dNTPs with dUTP) and Rox reference dye. The primer pairs used for analysis were designed within the *GAL10* gene on chromosome II. The forward primer *(GAL10 intragenic F)*, (5' TGGCGTATTTCGTATGACCA 3'), and the reverse primer *(GAL10 intragenic R)*, (5' CCAAGCATCACATTCCCTTC 3'), were designed using the yeast genome database (www.yeastgenome.govt). These primers were chosen to amplify a 333-bp fragment on the intragenic region of the *GAL10* gene. The standards used for *SyBr* green real-time PCR analyses were dilution of W303-1a genome DNA digested with the restriction enzyme *Msp*1, with DNA concentrations determined using Nanodrop 1000 (Appendix 10.1; courtesy of Justin O'Sullivan's group). All *SyBr* green real-time PCR reactions were carried out in 96-well PCR plates (ABI®, Applied Biosystems, Foster City, CA, USA) using a master mix. The master mix was set up on ice and comprised the following: *SyBr* Mix, 7.5 µl; Rox reference dye, 0.1 µl; MilliQ $H_2O$, 5.1 µl or 3.9 µl; Primer pair mix, 0.3 µl (200 nM) or 1.5 µl (1000 nM); sample, 2 µl.

All reactions and standards were performed in duplicate or triplicate. Amplification, data acquisition, and analysis were carried out using an ABI® Prism 7000 Sequence Detection System (Applied Biosystems, Foster City, CA, USA). The PCR programme comprised an initial step at 50°C for 2 min, a second step of 95°C for 2 min followed by 40 cycles of 95°C for 15 s and 60°C for 1 min, and finally holding at 72°C.

The results of *SyBr* green real-time PCR were exported to Microscoft® Excel. Amplified DNA from each sample was quantified and standardized according to the standards.

### 2.4.2 *Taqman*® Real-time PCR

Interaction frequency of the inter-chromosomal interaction between *SVL3* gene and the promoter region of *GAL7* was determined by *Taqman*® real-time PCR. PCR primers and the *Taqman*® probe for this interaction were designed using Oligo (National Biosciences, Plymouth, Minn.) and Primer Express (PE Applied Biosystems, Foster City, CA) software programs, and checked for non-specific annealing by a BLAST® search of GenBank. The

forward primer, *SVL3-GAL7F* (5' CCGTCCATATCTTTCCATAGATTTTC 3'), and the reverse primer, *SVL3-GAL7R* (5' TTCTCAGTGCTGGACTTACTTTCG 3'), were chosen to amplify a 102-bp fragment spanning the junction of the *SVL3* gene and the promoter of the *GAL7* gene (Figure 7). The internal *Taqman*® probe (5' TCTCCTTTGCAGTACCAC 3', on the reverse strand) was designed following the general rules outlined by the manufacturer (Figure 7). The standards used in *Taqman*® real-time PCR were dilution of 3C PCR product of *SVL3-GAL7* interaction, and DNA concentrations determined using a Nanodrop 1000 (Appendix 10.1)

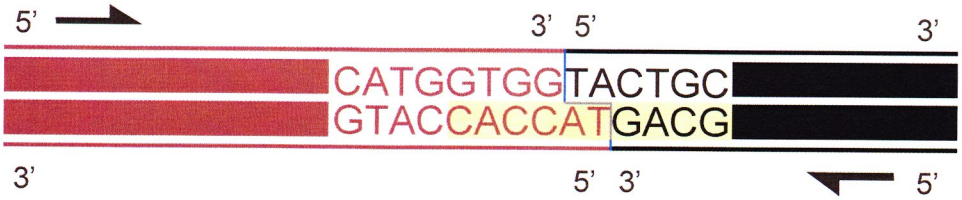



**Figure 7:** Diagrammatical representation of primers and probe for *Taqman*® real-time PCR. Primer pair is designed across the junction of the *SVL3* gene (brown colour region) and the promoter of *GAL7* gene (black colour region). These two regions are joined at the compatible *Csp*61 restriction ends (GTAC sequence that is separated by the blue lines). Coloured bars are omitted nucleotide sequences. Black arrows indicate the designed primers that amplify across the junction, and the region that is highlighted by yellow bar denotes a *Taqman*® probe sequence which is on the reverse strand.

*Taqman*® probe required in *Taqman*® real-time PCR containing a non-extendable oligonucleotide complementary to the target sequence that is labelled with a 5'-reporter (R) dye and a 3' quencher (Q) dye. During PCR amplification, forward and reverse primers hybridise to a specific sequence of the target DNA and *Taqman*® probe hybridises to the target sequence internal to the primer sequences. This enables *Taq* polymerase to cleave the reporter dye and quencher dye as a result of the 5'-3' exonuclease activity. Such cleavage allows the separation of both dyes and results in increased fluorescence of the reporter (Figure 8). This process occurs in every cycle and the increase in fluorescence, which is proportional to the amount of target amplification during PCR, is measured. A passive reference dye is also required to provide an internal reference for normalization of the reporter fluorescence (Rn). The threshold cycle (Ct) value is the number of cycles before the fluorescence emitted passes a fixed limit (Holland *et al.*, 1991).

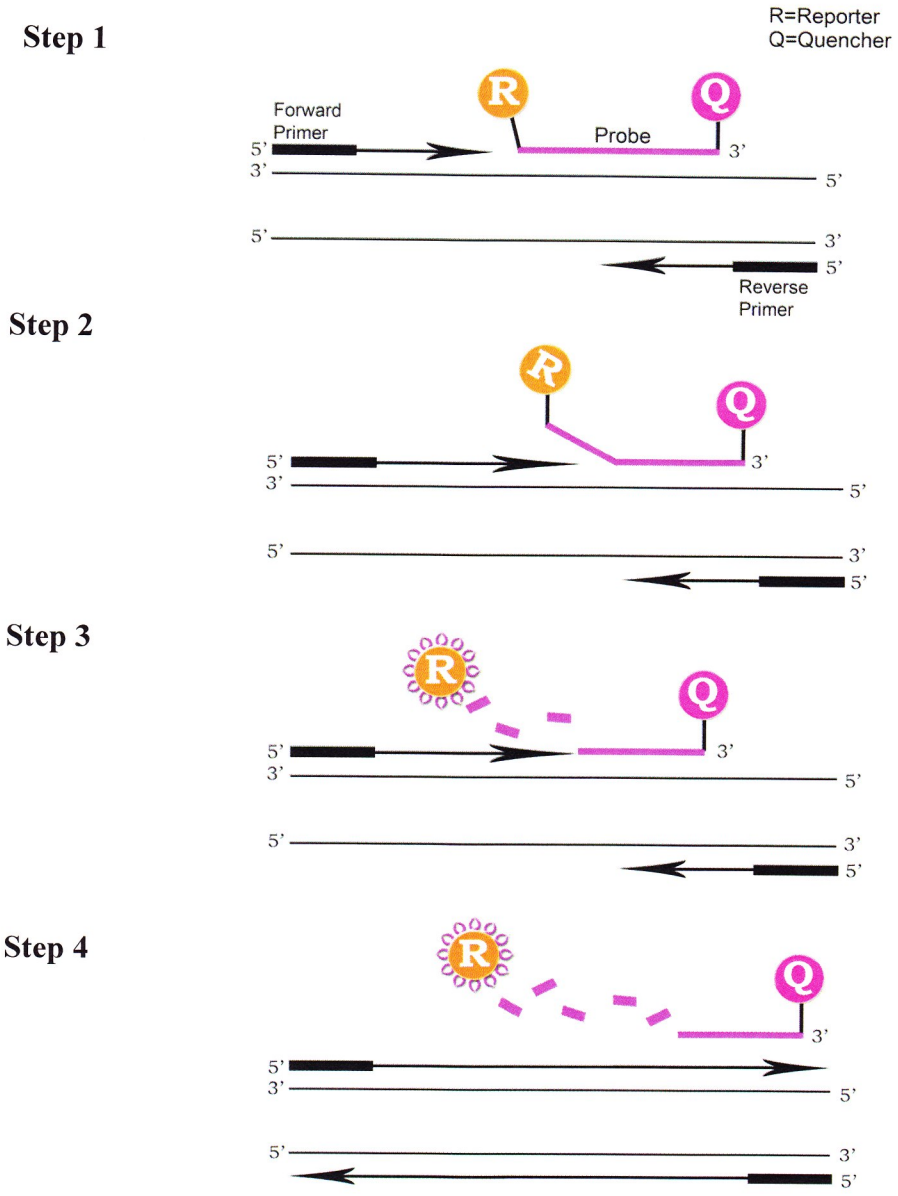



**Figure 8:** Diagrammatic representation of the *Taqman*® Real-time PCR application. Double thin lines represent double-stranded DNA sequence. Black arrows represent the primer pair used. Step 1: polymerisation, step 2: strand displacement, step 3: cleavage, and step 4: polymerization completion.

The 20 µl *Taqman*® real-time PCR mixture for *SVL3-GAL7* interaction amplification consisted of 10 µl of *Taqman*® Gene expression Master Mix (Lot. No. 0705011, Foster City, CA), 2 µl each of *SVL3-GAL7F* and *SVL3-GAL7R* (each at 900 nM), 2 µl of *Taqman*® Probe (at 2000 nM probe), 2 µl of MilliQ $H_2O$ and 2 µl of sample. The reaction comprised an initial step at 50°C for 2 min, a second step of 95°C for 10 min followed by 40 cycles of 95°C for 15 s and 60°C for 1 min.

Amplification, data acquisition, and analysis were performed using the ABI Prism 7000 Sequence Detection System. All standards, controls and samples were run in duplicate or triplicate. These results were then exported to Microsoft® Excel, and the amplified DNA

was quantified and standardized according to the standards. The standardized DNA quantities of each sample were then recalculated according to the results of *SyBr* green real-time PCR, and the values were converted to $Log_{10}$ value to quantify the frequency of *SVL3-GAL7* interaction.

## 2.5 Bioinformatics

### 2.5.1 BLAST Searches

The searches undertaken for establishing sequence identity all used the Basic Local Alignment Search Tool Algorithm (BLAST) (Altschul *et al.*, 1990), and were completed on either the NCBI website (www.ncbi.nlm.nih.gov/BLAST/) or on the *Saccharomyces* Genome Database (seq.yeastgenome.org/cgi-bin/blast-sgd.pl).

### 2.5.2 Sequence Alignments

The nucleotide sequences alignment for comparison was mainly performed using ClustalW from EMBL-EBL (www.ebi.ac.uk/clustalw/index.html) program.

# Chapter 3----Intra-chromosomal Interactions at the *GAL* Locus of Chromosome II

## 3.1 Introduction

The galactose (*GAL*) gene family of *Saccharomyces cerevisiae* is required for galactose metabolism in the yeast cells. This gene family, consisting of structural and regulatory genes, is an important model to study transcription regulation and the accumulation of mRNAs. The structural genes *GAL1*, *GAL7* and *GAL10* are located as a cluster on chromosome II of the yeast genome, whereas *GAL2* and *GAL5* are located on chromosome XII and XIII, respectively. The products of the regulatory genes *GAL3*, *GAL4* and *GAL80*, which are located on chromosomes IV, XVI, and XIII, respectively, coordinate to regulate the expression of the *GAL* structural genes.

The genes that comprise the *GAL* locus on chromosome II span a distance of seven kb, with intervals between each two genes of 600-700bp. The expression of *GAL* structural genes is strictly regulated by the availability of carbon sources. Glucose represses *GAL* gene expression, glycerol/lactate induces a basal level of expression, and galactose induces robust level of expression. Thus manipulating *GAL* gene expression allowing us to study the DNA structural changes that might be occurring during transcriptional regulation.

The regulation and function of members of the GAL gene family have been studied extensively over the last few decades (Mylin *et al.*, 1989, 1990; Johnston, 1987, 1992; Ostergaard *et al.*, 2000). However to date, no study of *GAL* gene positioning in relation to its expression has been conducted. The project described in this thesis was set up to provide information to help to fill this knowledge gap and assist in the understanding of *GAL* gene activity in relation to DNA structural change.

This study, which focuses on the interactions between DNA structural changes at the *GAL* locus and *GAL* gene regulation, has been conducted from two perspectives. This chapter examines possible interactions between DNA loci at the *GAL* locus (intra-chromosomal), while Chapter 4 examines a range of interactions between the *GAL* locus on chromosome II and other chromosomes (inter-chromosomal).

## 3.2 Chromatin Isolation and Preliminary 3C Studies

### 3.2.1 Restriction Enzyme Selection

The regulatory sequences of a gene, *e.g.*, enhancer, promoter, and terminator, are essential components for gene regulation. When studying potential changes in gene positioning in relation to gene activity, it is important to take into account the positioning changes of these regulatory sequences. Experimentally, it was hoped that an entire piece of chromosome could be physically broken down into DNA pieces. The targeted DNA fragment would then be examined specifically to determine whether it interacted with any other DNA fragments within the genome. Once the identities of the interacting DNA fragments were determined, the position of the targeted DNA fragment could in turn be determined. Because the yeast genome has been completely sequenced (www.yeastgenome.org), it becomes relatively straightforward to retrieve DNA sequence information from the database and manipulate it. The physical location of the three structural genes *GAL1*, *GAL7* and *GAL10* on chromosome II is shown in Figure 9. Table 2 contains detailed information on these structural genes with respect to chromosomal location and interval distances. The strain used in this study was W303-1a (Table 1: Section 2.1.1), which was cultivated in SC medium.

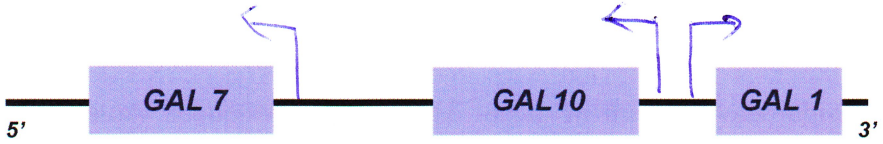



**Figure 9:** Diagrammatical representation of *GAL* structural genes *1, 7* and *10* on chromosome II. The purple rectangles represent the DNA sequence of *GAL* genes and the length of rectangles reflects the spans of each *GAL* sequence. The diagram is not drawn to scale. Information obtained from www.yeastgenome.org.

**Table 2.** Information of the *GAL* structural genes on chromosome II. Information regarding the chromosomal coordinates and relative position of each *GAL* gene is present; intervals between the genes and the total distance across all genes are also given[a].

| Gene[b] | Chromosome Coordinates | Coding Sequence Length |
|---|---|---|
| *GAL* 7[c] | 275527-274427 | 1101 |
| *GAL* 10[d] | 278352-276253 | 2100 |
| *GAL* 1 | 279021-280607 | 1587 |

[a]All information obtained from www.yeastgenome.org;. [b]Total span of *GAL* genes is 6179 bp; [c]Interval between *GAL7* and *GAL10* is 725 bp; [d]Interval between *GAL10* and *GAL1* is 1668 bp.

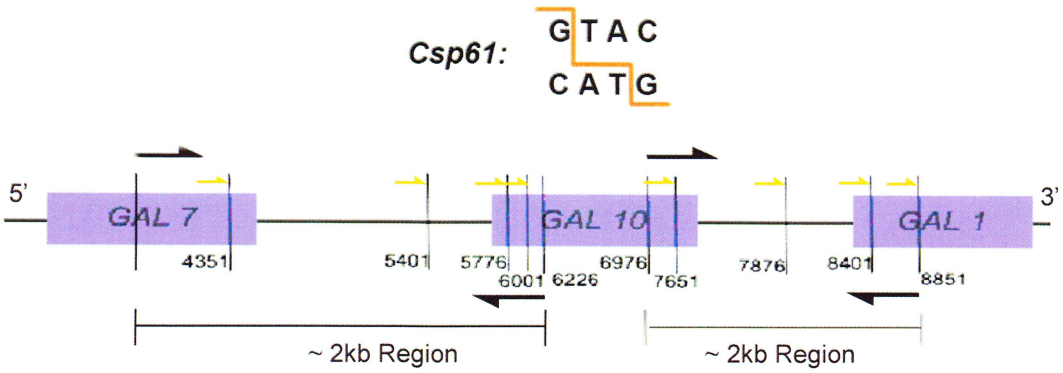



**Figure 10:** Diagrammatical representation of primer design at the *GAL* locus on chromosome II. The purple rectangles represent the DNA sequence of *GAL* genes. Restriction enzyme *Csp*61 cleave sequence is illustrated and the restriction sites at the *GAL* locus are shown by black vertical lines, and the restricted fragments are named according to their nucleotide positions. Large black arrows indicate the location of primers that designed to amplify the 2kb regions between *GAL7-GAL10* and *GAL10-GAL1* genes. Small orange arrows are internal primers for 3C analyses, and they are facing the same direction.

These positive controls were tested with primer pairs within the regions, *e.g.,* primer pairs for fragments 5776-5401, 5776-6001, 7876-8401, and 7876-7651 (Figure 11 and data not shown). These primers are illustrated in Figure 10 as the orange arrows and they were paired by random selection in these positive control testing experiments.

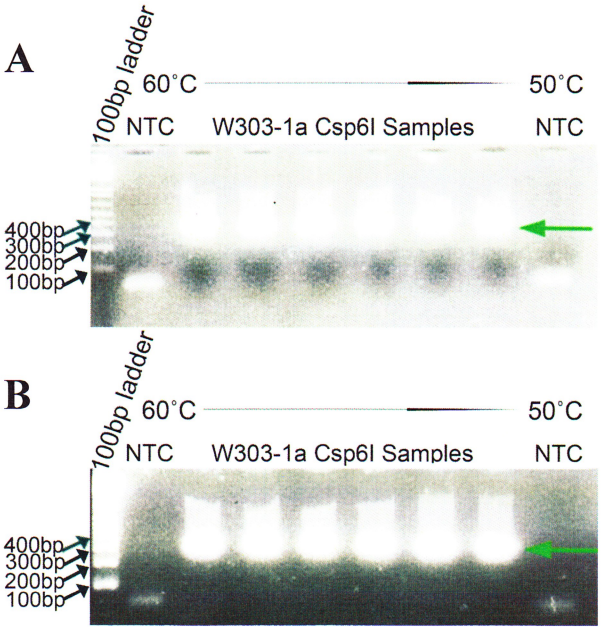



**Figure 11:** Positive control testing. W303-1a genomic DNA was digested with *Csp*61, ligated and amplified using a temperature gradient (50˚C-60˚C) PCR on primer pairs (A) 5776-6001 and (B) 5776-5401. The expected product sizes for these primer pairs were 414 bp and 328 bp, respectively, which are indicated by green arrows. Negative controls (NTCs), in which water was used as template, were included at 50˚C and 60˚C. 20 µl of PCR products were loaded on 1.5% agarose gel, which was run at 80 volts for 40 min and then photographed.

### 3.2.3 Preliminary 3C Analyses

As *GAL* gene expression is strictly regulated by the availability of a carbon source, it was hypothesized that DNA may undergo conformational changes in response to gene activity. DNA conformation changes could be manifested by an interaction between DNA regulatory elements, *e.g.,* the promoter and terminator, via DNA looping, to facilitate the regulation of gene activity. To examine how DNA conformation is altered at the *GAL* locus on chromosome II, experiments were performed to detect any interactions between the promoter and terminator regions of the *GAL* genes. This was hoped to provide some information on the interaction pattern with respect to their gene activity status, which is regulated by the availability of carbon sources, *i.e.,* glucose, glycerol/lactate, or galactose.

First, potential interactions between adjacent DNA fragments on the region of *GAL10* and *GAL1* genes were studied (Figure 10). This was a preliminary study mainly to determine the efficiency of 3C method in this type of study. In 3C analysis, each set of experiments consisted of an uncut sample (chromatin not digested by enzyme), a cut sample (chromatin digested by enzyme but not ligated), and a cut-ligated sample (chromatin digestion by enzyme and ligated). Uncut and cut samples are a form of control in relation to the cut-ligated sample. As shown in Figure 12, PCR analyses using various primer combinations were performed. For the primer pair 7876-8401, negative results appeared in the presence of all carbon sources (glucose, glycerol/lactate and galactose, Figure 12A: lanes 3, 6 and 9); for the primer pair 7876-7651, positive outcomes shown in the presence of glycerol/lactate and galactose (Figure 12B, lanes 6 and 9), as well as a possible positive shown in the presence of glucose (Lane 3). Positive and negative controls validated PCR analyses, whereas the uncut and cut samples confirmed that 3C assays were performed appropriately (Figure 12B: compare lanes 4 and 5 with lane 6). As fragments 7651, 7876 and 8401 are in adjacent positions, these results implied that these products may simply be due to inter-molecular ligation and were not influenced by protein binding. For other primer combinations, mainly negative results were obtained under all induction conditions (data not shown). Similarly, the examination of the potential interactions between the adjacent fragments on the region of *GAL7* and *GAL10* genes showed negative results under all induction conditions (fragments 4351, 5401 and 5776, data not shown) even when PCR conditions were related.

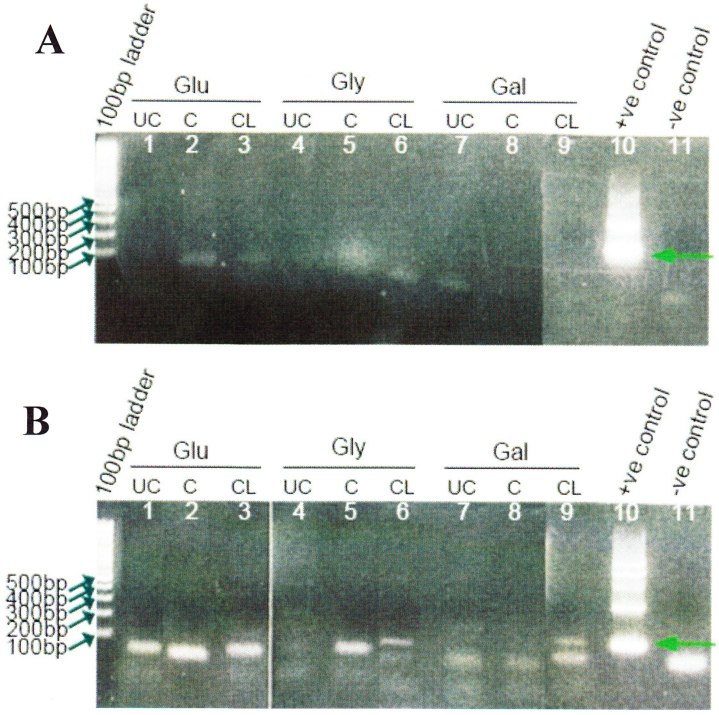



**Figure 12**: Preliminary 3C analyses of interactions at the *GAL* locus. Adjacent DNA fragments at the *GAL* locus were tested for interactions. W303-1a strains were induced by glucose (Glu), glycerol/lactate (Gly) or galactose (Gal). PCR amplifications were performed on primer pairs (A) 7876-8401 and (B) 7876-7651. The expected product sizes were (a) 278 bp and (b) 119 bp, respectively, which are indicated by green arrows. UC denotes uncut sample, C denotes *Csp*61 cut sample, while CL denotes *Csp*61 cut and ligated sample. 20 µl of PCR product were loaded on a 1.5% agarose gel, which was run at 80 volts for 40 min, then photographed.

## 3.3 Switch to 4C Analysis

### 3.3.1 4C Principle and Application

With the exception of one positive result (Figure 12B), the preliminary 3C analyses generated a high level of negative results in the search of interactions at the *GAL* locus. At this stage searching for interactions at the *GAL* locus on chromosome II was performed on DNA fragments that were randomly chosen due to the lack of information on any potential interactions. 3C is thus considered to be of low efficiency in the search of potential interactions. Therefore, a different approach termed Circular Chromosome Conformation Capture (4C) that had been recently published (Zhao *et al.*, 2006), was adapted and applied prior to 3C analysis for the initial detection of any intra- or inter-chromosomal interactions. Although 3C analysis has been proven to be very useful in determining close physical interactions between sequences from remote locations, it is not suitable for screening interactions without prior knowledge of the existence of the linking complex (Zhao *et al.*, 2006). The novel 4C method, which is essentially based on 3C analysis, has been shown to

overcome the shortcomings of 3C analysis. The principle of the 4C method is proximity ligation, in which DNA-protein/protein-DNA complexes form circular DNA molecules under prolonged ligation times. After reversal of cross-linking, inverse primers within the known genomic fragment, but proximal to the unknown sequence during ligation, will amplify the unknown DNA sequences with physical proximity, without any prior knowledge of their identities (Figure 5, Chapter 2). These DNA sequences are subsequently cloned into a plasmid and transformed into *E. coli.* The identity of these unknown sequences was reviewed by DNA sequencing and studied using the BLAST algorithm. Following the interaction being understood, the intention was then to revisit 3C analysis to confirm the interaction.

For 4C analysis, inverse primer pairs were designed within the terminator and promoter regions of the *GAL1, GAL7* and *GAL10* genes (Figure 13). Without the prior knowledge of potential interacting DNA sequences, the possibility for the unknown sequence is numerous. Thus nested PCR that is less likely to amplify non-specific products was employed in 4C analysis. This increases the specificity of the degenerate PCR.

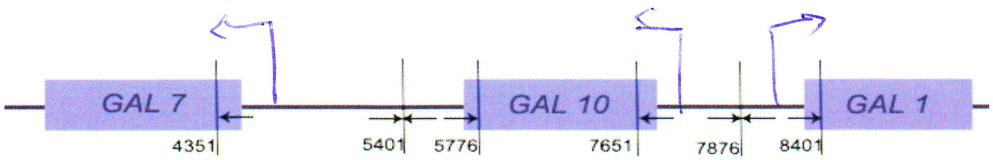



**Figure 13:** Diagrammatical representation of primer design at the *GAL* locus on chromosome II for 4C analyses. The promoter and terminator regions of *GAL1, GAL7* and *GAL10* were the targeted regions for primer design. These primers are named according to their corresponding nucleotide positions, which were determined by *Csp*61 restriction sites (shown in black vertical lines). All primer pairs are inversed as indicated by the direction of the black arrows.

### 3.3.2 4C Results at the *GAL* Locus

The first step was to study any potential interactions occur between the regulatory elements at the *GAL* locus on chromosome II under different induction conditions (glucose, glycerol/lactate, and galactose). These regulatory sequences and their corresponding designated PCR primer pairs included: the promoter of *GAL*10 (corresponding to the inverse primer pair 7876-7651), the promoter of *GAL1* (inverse primer pair 7876-8401), the promoter of *GAL*7 (inverse primer pair 4351-5401), and the terminator of *GAL1* (inverse primer pair 5401-5776) (Figure 13; for primer information see Appendix 3.0). For the sake

of clarity of presentation, only cut-ligated (CL) samples are presented here. As shown by nested PCR, there was more than one product present under all induction conditions (Figure 14, ligated products). The PCR products with various sizes, including smears, suggested that these regions of *GAL* gene might have interacted with multiple regions, either on the same chromosome or other chromosomes. The PCR products with smaller molecular sizes (at around 100 bp) derived from the self-ligation products, which were formed by the circularization of the known DNA sequence during prolonged ligation process. Their PCR products migrated closer to the bottom of the gel due to their smaller sizes (Figure 14, Self-ligated product). Because they are formed easily and are readily amplified by PCR, they appear to be the brightest bands in the gel. Furthermore, there was no apparent difference in terms of the PCR product formation under different induction conditions. As shown in Figure 14, in the presence of glucose (repressed state), glycerol/lactate (non-induced state) or galactose (induced state), the PCR pattern appeared to be very similar. These 4C assays were repeated three times.

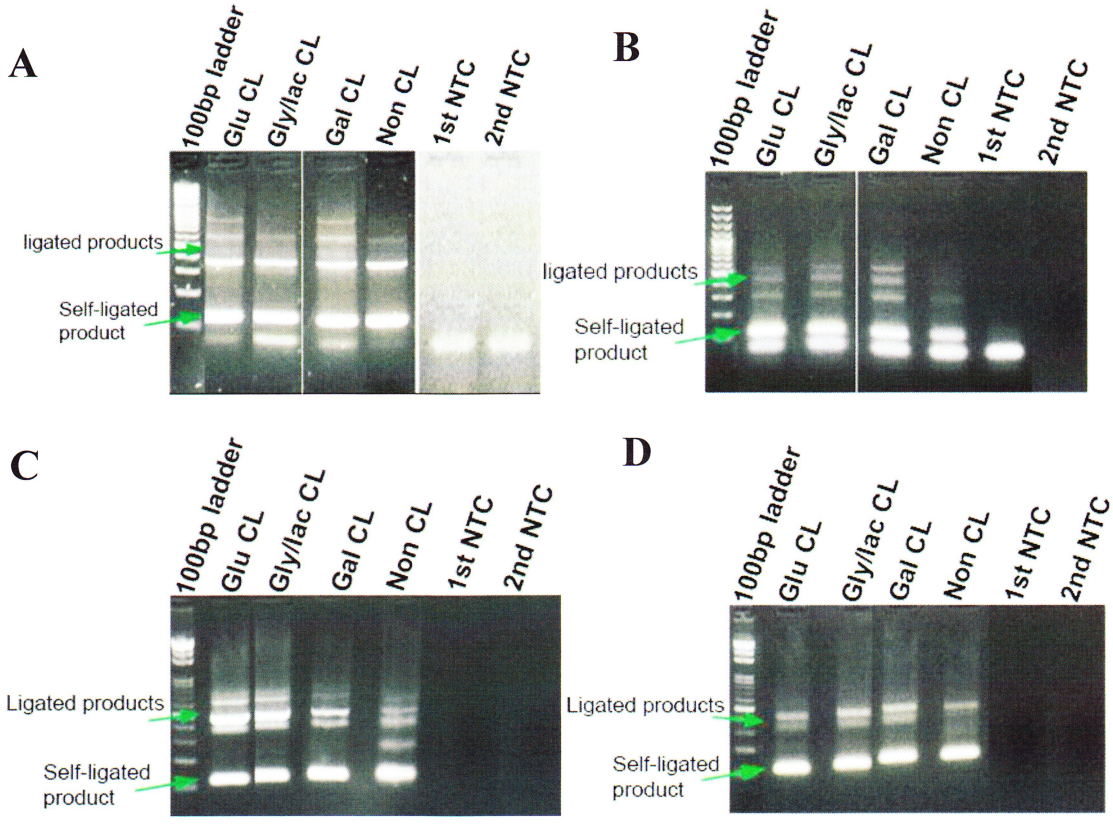



**Figure 14:** 4C nested PCR analyses of interactions at the *GAL* locus of chromosome II. W303-1a cells were cross-linked in the presence of glucose (repressed), glycerol/lactate (noninduced), or galactose (induced). Another set of cells, not cross-linked in the presence of glycerol/lactate, was included as controls for the cross-linked cells. Glu, Gly/lac and Gal denote glucose, glycerol/lactate, and galactose, respectively. CL denotes *Csp*61 cut and ligation sample. Negative controls (NTCs) were included for both rounds of nested PCRs. PCR amplifications were performed using primer pairs (A) 7876-7651, (B) 7876-8401, (C) 4351-5401, and (D) 5401-5776. 20 µl of PCR reaction samples were loaded on a 1.5% agarose gel, which was run at 80 volts for 40 min, then photographed. Green arrows indicate ligated products and self-ligated products.

There is no positive control required for 4C analyses due to the lack of information on potential ligated sequences. However, the presence of self-ligation products is a form of validation for the PCR reactions. The density of the self-ligation products also served as an indication of the effectiveness of the PCR. A temperature gradient PCR reaction was performed to optimise the primer conditions; genomic DNA that was digested by *Csp*61 and ligated was used as a template. In such a PCR, the primer pair was tested over a range of annealing temperatures with the other reaction conditions remaining the same (see Appendix 7.0).

In addition, the potential interactions occur at the *GAL* locus under non-cross-linking conditions, *i.e.*, protein and DNA structure is not fixed by formaldehyde treatment, were studied. This was included as a control for the cross-linked samples. This study begun with a non-cross-linking culture in the presence of glycerol/lactate (noninduced state), the procedure followed a 4C protocol but with no formaldehyde treatment step. It was showed by nested PCR that the non-cross-linked samples displayed a pattern similar to that of the cross-linked samples, dependent on the primers used (Figure 14: compare B and C with A and D). This was a very interesting result as it indicated that certain interactions might be present without cross-linking, in which case protein and DNA interaction/fixation stays in place as a stable construct.

### 3.3.3 Identify of the Unknown Sequence Derived from 4C Analyses

The ligation products of 4C nested PCR indicated that there were potentially numerous interactions occurring between sequences at the *GAL* locus on chromosome II (Figure 14). DNA sequencing is required to reveal the identities of these unknown sequences, and therefore to recover these potential interactions. Specifically two approaches were used: (i) direct sequencing, and (ii) clonal sequencing.

*(i) Direct sequencing*

4C nested PCR samples were visualized by electrophoresis using an agarose gel of appropriate concentration. For direct sequencing, the samples were run for an extended period of time (at least 2 hours at 80 volts) until all bands were well separated. Examples are shown in Figure 15 in which the individual bands were labelled according to their positions.

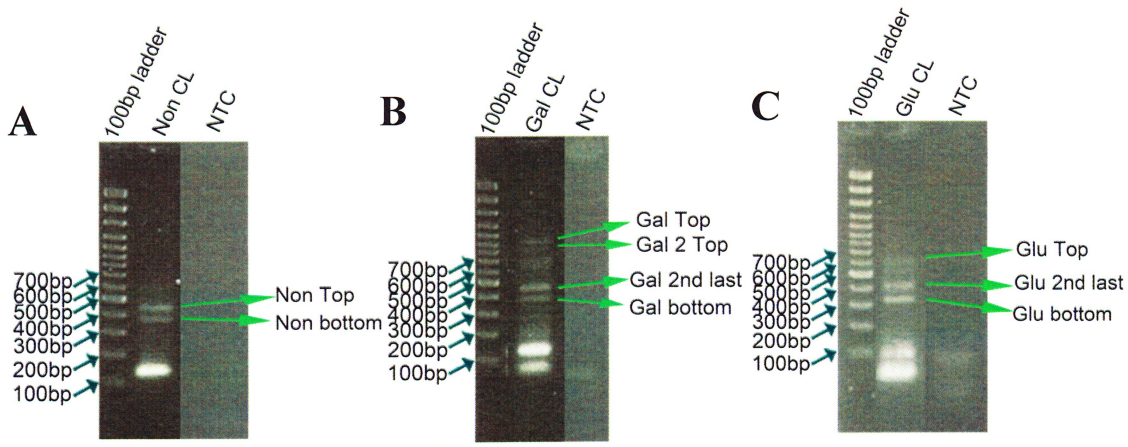



**Figure 15:** PCR analyses for direct sequencing. 4C nested PCR analyses were performed at the *GAL* locus using primer pairs (A & B) 7651-7876 and (C) 7876-8401. W303-1a cells were either cross-linked in the presence of galactose (A) and glucose (C), or not cross-linked in the presence of glycerol/lactate (B). 20 µl of PCR reactions were loaded on a 1.5% agarose gel, which were run at 80 volts for 2 hr or more and then photographed. Visible bands were isolated (green arrows) and underwent further treatment for DNA recovery.

The visible bands were excised with a scalpel and the DNA was purified by agarose gel purification (Section 2.3.4). These fragments were sequenced directly and analysed using the BLAST algorithm. A few 4C samples were studied in this way to get a general idea of the nature of these interacting sequences. The sequencing results are presented in Figure 16.

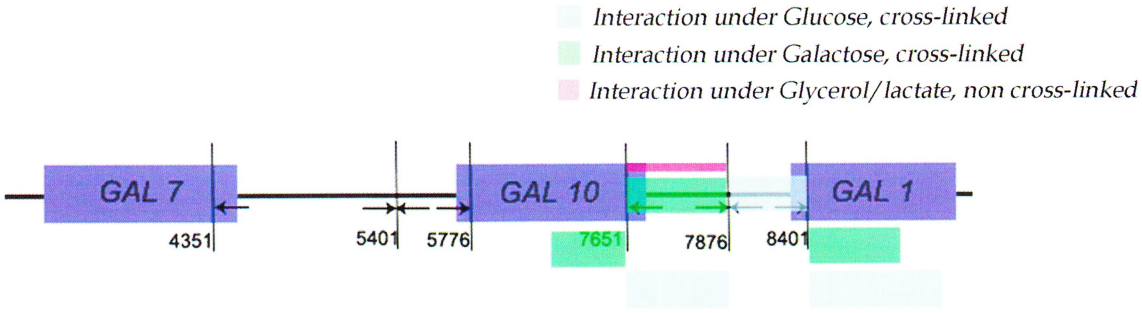



**Figure 16:** Diagrammatical representation of the direct sequencing results. The purple rectangles represent the *GAL* genes on chromosome II. Black vertical lines represent *Csp*61 restriction sites. Black arrows indicate inverse primer pairs. The remaining coloured rectangles indicate corresponding interaction zones.

As shown in Figure 16, most of the interactions that occurred at the *GAL* locus on chromosome II were between the adjacent fragments, *e.g.*, fragment 7876 interacts with fragment 7651 in the presence of galactose; fragment 8401 interacts with fragment 7876 and another fragment locates immediately downstream in the presence of glucose. Products derived from interactions between adjacent fragments implied a partial *Csp*61 digestion.

However, the interaction between fragment 7876 and a fragment that is located to the downstream of fragment 8401 (Figure 16, green rectangles) presented as an exception. This suggested that the observed interactions could be resulting from partial digestion or is real interactions that happened to occur in close adjacent positions.

The induction conditions made little difference to the pattern of the interactions. However, in the sample of non-cross-linked DNA, there seemed to be fewer interactions compared with the ones of cross-linked DNA. It appeared that in non-cross-linked samples there was only self-interaction occurring (Figure 16, pink rectangle).

In direct sequencing, only the visible bands were isolated and sequenced. The lighter bands or the smear, which contained numerous DNA fragments that existed at low densities, were almost impossible to retrieve directly from the agarose gel. Thus direct sequencing is greatly limited in terms of reflecting all possible interactions. An alternative approach of sequencing clones offers an opportunity to gain a better understanding of the interactions between regions of the genome.

*(ii) Clonal sequencing*

In the clonal sequencing approach, 4C nested PCR products were subjected to another restriction enzyme (*Eco*RI) digestion before being cloned into *E. coli* plasmid, which was also pre-digested by the same enzyme to generate a complementary cohesive end. The cloned plasmid was then transformed into the host cells to generate a 4C clone library that included all possible recombination. The library was screened and the selected clones were sequenced (Section 2.3.7.6). Although labour intensive, sequencing of clones dramatically increases the possibility of identifying interacting sequences. Table 3 provides a summary of the clone sequences with different known DNA fragments under various induction conditions. The self-ligation clones constituted an average of 78% of the total clones. Only a few clones, *i.e.* 2-5, were eventually sequenced due to the fact that a large number of the clones contained multiple inserts of the same fragment. Clonal sequencing generated the majority of the results that were used to identify the unknown interacting sequences. On the other hand, direct sequencing provided preliminary information as well as verifying the results obtained from clonal sequencing.

**Table 3:** Summary of the cloning sequencing at the *GAL* locus of chromosome II. Clones obtained from interactions with fragments 8401, 5401, and 5776 were studied under different induction conditions: glucose, glycerol/lactate and galactose. Total clones denote the total number of clones obtained in each of the target fragments, self-ligated denotes the clones that contain self-ligated products, sequenced denotes the clones that contained inserts that were sequenced.

| Fragment | Glucose | | | Glycerol/lactate | | | Galactose | | |
|---|---|---|---|---|---|---|---|---|---|
| | Total clones | Self-ligated | Sequenced | Total clones | Self-ligated | Sequenced | Total clones | Self-ligated | Sequenced |
| **8401** | 115 | 76 | 2 | 119 | 82 | 2 | 116 | 83 | 3 |
| **5401** | 133 | 112 | 2 | 116 | 103 | 2 | 107 | 92 | 5 |
| **5776** | 171 | 130 | 0 | 158 | 127 | 1 | 138 | 111 | 3 |

The results obtained from the clonal sequencing and its subsequent analyses using BLAST are illustrated below in Figure 17:

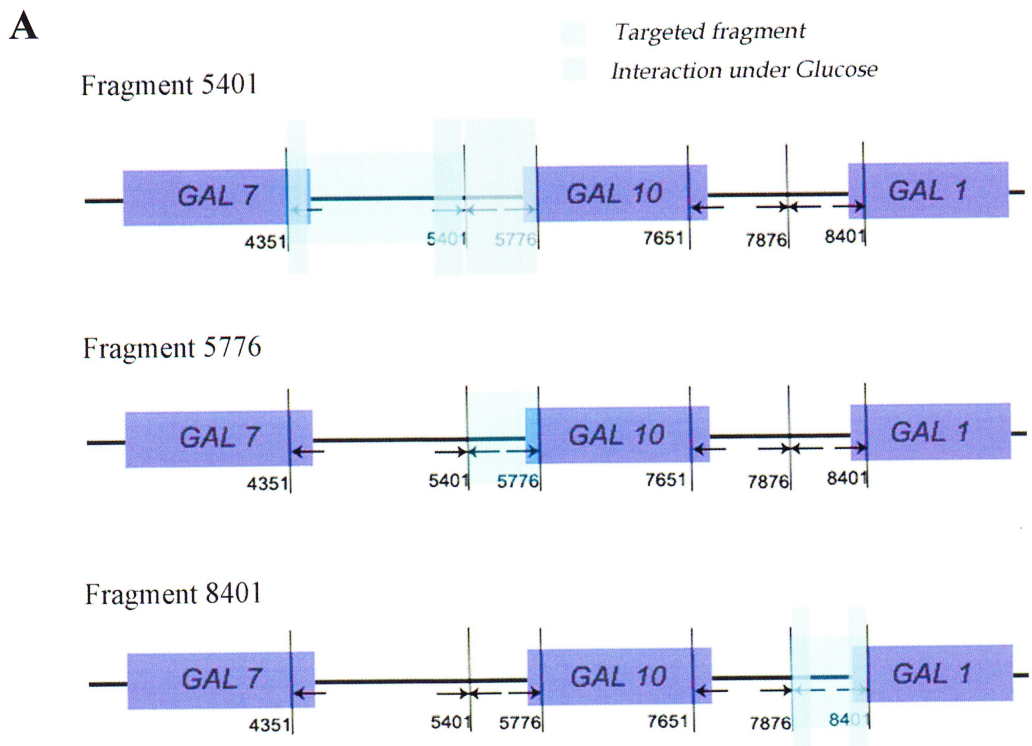



Figure 17: see next page for legend.

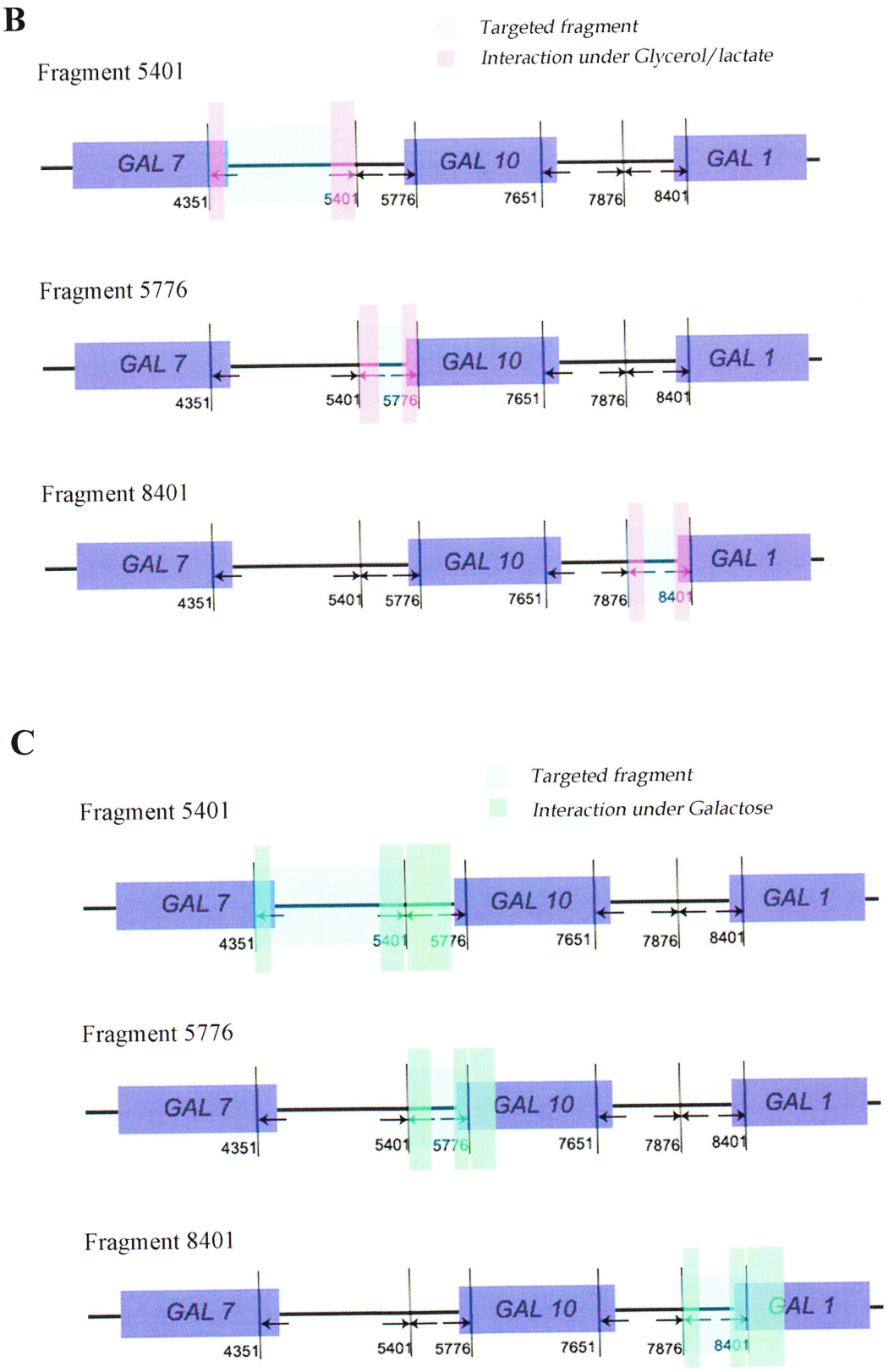



**Figure 17:** Diagrammatical representation of the clonal sequencing results. The intra-chromosomal interactions at the *GAL* locus were revealed by clonal sequencing. The purple rectangles represent the *GAL* genes on chromosome II. The black vertical lines are *Csp*61 restriction sites, and black arrows indicate inverse primer pairs. The interactions with the targeted DNA fragments (the light blue rectangles locate between two adjacent *Csp*61 sites) were studied in the presence of (A) glucose, (B) glycerol/lactate, and (C) galactose. The corresponding interacting DNA sequences under each condition were illustrated by coloured rectangles: (A) grey, (B) red, and (C) green.

The results showed that interactions occurred mostly within (self-ligated) or immediately adjacent to the targeted (known) fragments. It seemed that the interactions between adjacent fragments occurred mostly in the presence of galactose, that is, when the *GAL* genes are expressed, and the interactions all occurred to the downstream of the targeted DNA fragments (Figure 17C, green areas). Similar to direct sequencing, these interactions possibly indicated a partial *Csp*61 digestion. However, as only a few clones being sequenced (Table 3), it is possible that other interactions between non-adjacent fragments have not been reviewed.

The results of clonal sequencing, combined with those from direct sequencing, provided some information regarding the interactions occurring at the *GAL* locus on chromosome II. To understand how these interactions are associated in order to build up the overall genome structure, 3C analyses needed to be revisited in order to study the interactions in more detail.

## 3.4 3C Studies of Intra-chromosomal Interactions

### 3.4.1 3C Analyses on Adjacent Fragments

3C studies were applied to confirm and further identify interactions occurring at the *GAL* locus on chromosome II based on the results of 4C assays. As DNA sequencing suggested a pattern of interactions with adjacent fragments, a "walking–down" strategy was firstly applied to study the interactions systematically (Figure 18). Briefly, the interaction between every two immediately adjacent fragments was studied at one time under all induction conditions. Starting with the two adjacent fragments on the far left-hand side, with a gradual move down the scale to the right-hand side, all adjacent fragments were studied in an overlapping manner (Figure 18, blue lines). These 3C assays were repeated three times.

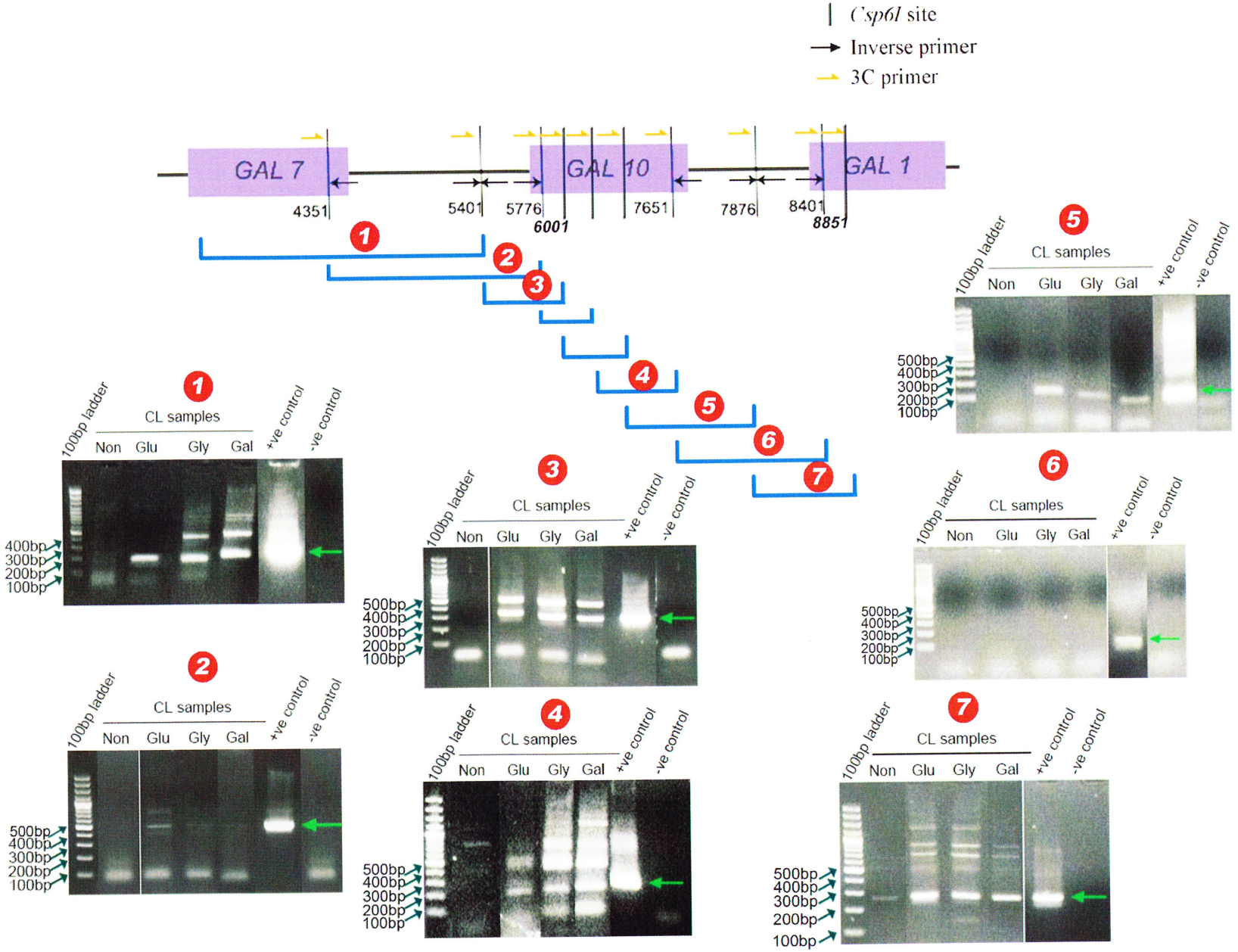

| Csp6I site
→ Inverse primer
→ 3C primer
GAL 7
GAL 10
GAL 1
4351
5401
5776
6001
7651
7876
8401
8851
1
2
3
4
5
6
7
100bp ladder
CL samples
Non
Glu
Gly
Gal
+ve control
-ve control
500bp
400bp
300bp
200bp
100bp

**Figure 18:** Diagrammatical representation of the "walking-down" strategy in 3C studies at the *GAL* locus. These interactions occurred between adjacent fragments. The purple rectangles represent the *GAL* gene on chromosome II. Black arrows indicate inverse primer pairs. 3C primers are indicated by small orange arrows, which all face the same direction. The blue lines indicate that every two fragments being studied at one time. The expected sizes for these 3C PCR analyses were (1) 392 bp, (2) 414 bp, (3) 328 bp, (4) 201 bp, (5) 119 bp, (6) 295 bp, and (7) 264 bp. These products are indicated by green arrows. 20 µl of PCR reactions were loaded on a 1.5% agarose gel, which were run at 80 volts for 40 min, then photographed.

The results of 3C PCR demonstrated a continuous interaction pattern across the *GAL* locus region with an exception between fragments 7876 and 8401. The interactions occurred under all induction conditions (Figure 18). The non-cross-linking samples showed negative results in PCR for most primer pairs, except for 8401-8851 (Figure 18, number 7). Therefore a "two interaction zone" was proposed to form at the *GAL* locus, with a "break-point" at the interval of the *GAL10* and *GAL1* genes (Figure 19).

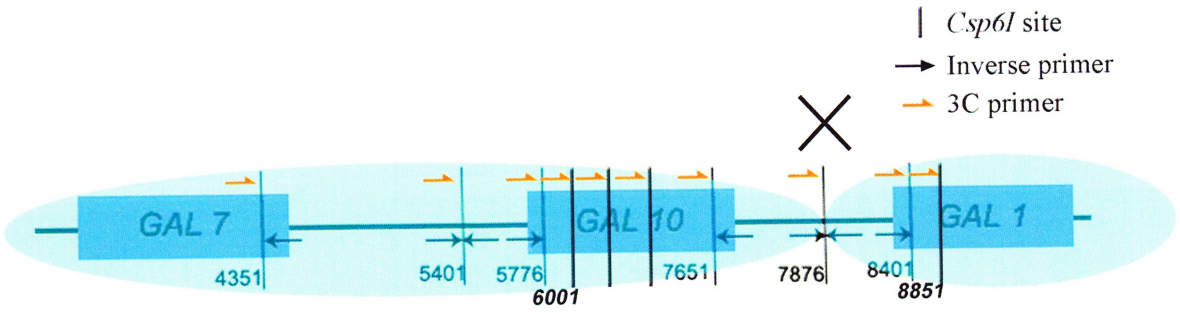



**Figure 19:** Diagrammatical representation of the "two interaction zone" at the *GAL* locus. The blue rectangles represent the *GAL* genes on chromosome II. The light blue ovals cover the linking areas at the *GAL* locus, with a break point occurred between fragments 7876 and 8401; a black cross denotes the break point.

### 3.4.2 3C Analyses on Non-adjacent Fragments

Followed the proposal of the "two interaction zones"; the longer distance interactions at the *GAL* locus were examined in order to determine how *GAL* genes were positioned with respect to the genome. Longer distance interactions referred to those interactions between the non-adjacent, distant fragments of the *GAL* locus. The candidate fragments are illustrated in Figure 20.

Several fragments (4351, 7876 and 8401) were analysed under different induction conditions. Fragment 7651 was selected as a 'reference fragment' as it is located close to the break point between fragments 7876 and 8401. Since there was no difference observed in the interaction pattern under all induction conditions, attention was drawn on comparing glucose (repressed) and galactose (induced) conditions, which represented two opposite gene expression status. These 3C assays were repeated for three times.

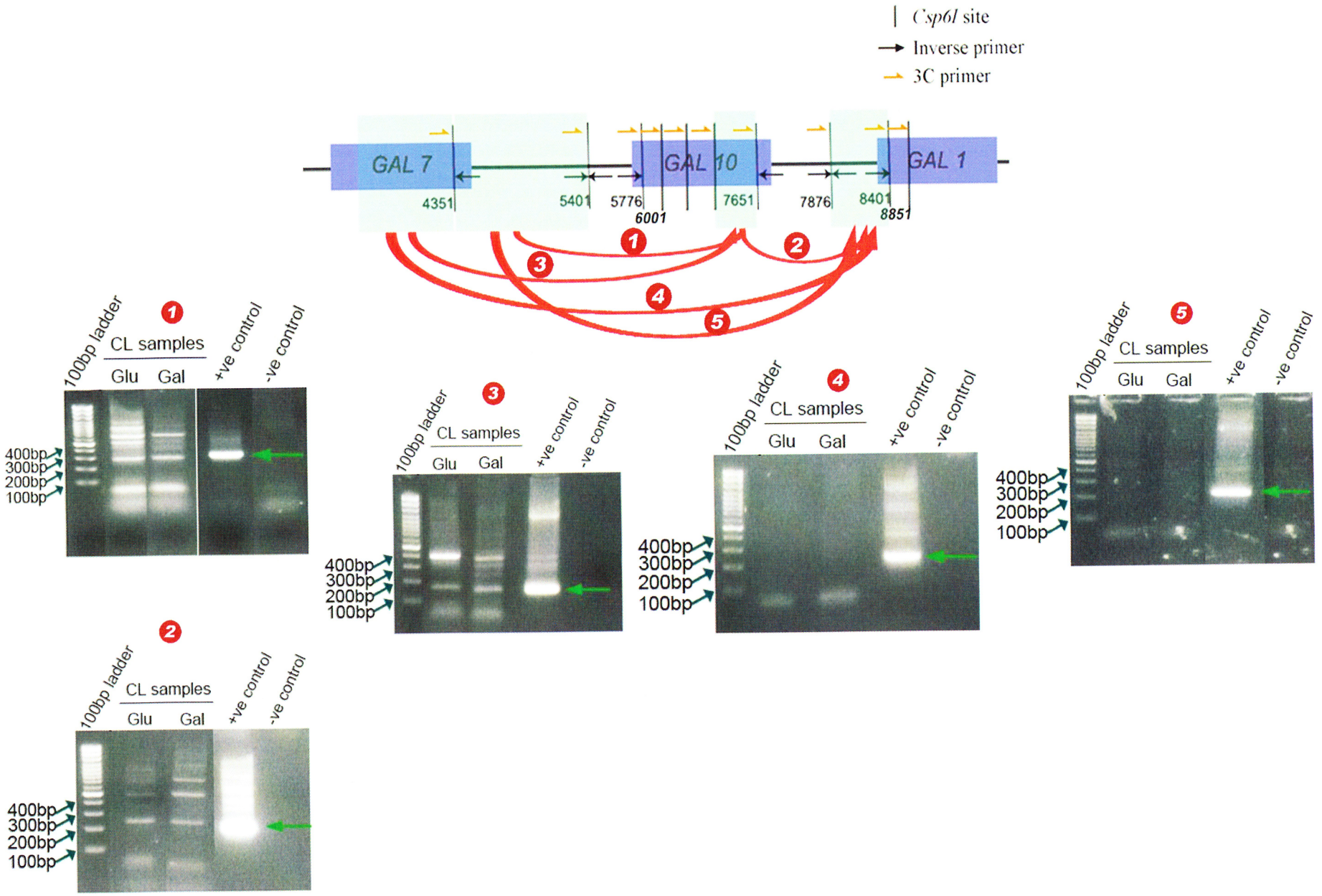

CspδI site
Inverse primer
3C primer
GAL 7
GAL 10
GAL 1
4351
5401
5776
6001
7651
7876
8401
8851
1
2
3
4
5
100bp ladder
CL samples
Glu
Gal
+ve control
-ve control
400bp
300bp
200bp
100bp

**Figure 20:** Diagrammatical representation of the "long distance" 3C analyses at the *GAL* locus. These interactions occurred between non-adjacent, distant fragments. The purple rectangles represent the *GAL* gene on chromosome II. Black arrows indicate inverse primer pairs. 3C primers are indicated by small orange arrows, which all face the same direction. The red ribbons connected every two non-adjacent DNA fragments being studied at one time. The expected sizes of these 3C PCR analyses were (1) 287 bp, (2) 219 bp, (3) 174 bp, (4) 259 bp, (5) 193 bp. 20 µl of PCR reaction samples were loaded on a 1.5% agarose gel, which were run at 80 volts for 40 min, then photographed.

3C PCR showed that most of the non-adjacent fragments interacted in the presence of glucose or galactose when referred to fragment 7651 (Figure 20, number 1, 2 and 3). As these interactions seemed to occur despite their positions on the chromosome, it was suspected that these interactions occurred by random collisions rather than specific interaction. The fragments located at the far ends, *e.g.*, fragments 4351, 8401 and 8801, were therefore examined for interactions. The negative results shown by PCR, in the presence of glucose or galactose (Figure 20, number 4 and 5), suggested that the observed pattern of interactions between distant DNA sequences did not happen randomly.

Considering the results so far, an S-shape structure, which is constituted at the *GAL* locus on chromosome II, was proposed as the model (Figure 21). This model reflected all induction conditions, *i.e.* the *GAL* gene activities do not influence the formation of the structure.

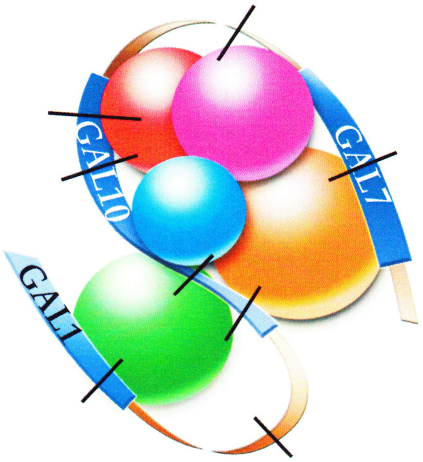



**Figure 21:** Modelling of an S-shape structure at the *GAL* locus of chromosome II. The brown ribbon represents the DNA sequence of chromosome II, while the blue ribbons represent the *GAL* genes. The coloured spheres represent protein/protein complexes that facilitate the formation of the S-shape structure. Black short lines represent the *Csp*61 restriction enzyme digestion sites.

## 3.5 Exploration of the Formation of S-shape Structure

The 3C and 4C studies regarding the intra-chromosomal interaction at the *GAL* locus on chromosome II have led to the proposal of an "S"-shape structure. Although this structure seemed to form regardless of the *GAL* gene expression status, it was still necessary to explore the regulatory factors that mediate the formation of the structure. It has been shown in previous studies (O'Sullivan *et al.*, 2004; Ansari and Hampsey, 2005) that protein/protein

complex mediated DNA interactions. It was therefore appropriate to selectively study proteins that might be involved in mediating the observed intra-chromosomal interactions, which contribute to S-shape structure formation at the *GAL* locus. The focus was on the critical interactions, including (A) 4351-8401, (B) 5401-7651, (C) 4351-7651, and (D) 7651-8401, as shown in Figure 22. All PCR amplifications in this section were performed using primer pairs corresponding to the above interactions. Two experiments were performed in this section.

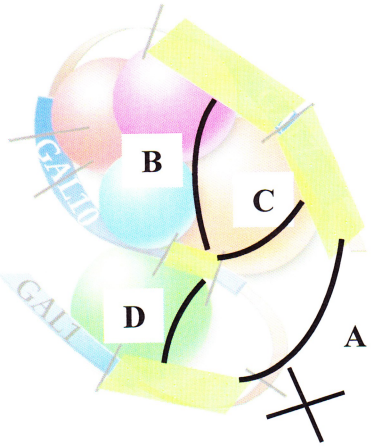



**Figure 22:** Diagrammatic representation of the critical interactions at S-shape structure. These interactions contribute to the S-shape structure formation at the *GAL* locus. The brown ribbon represents DNA sequence of chromosome II, while the blue ribbons represent the *GAL* genes. The coloured spheres represent the protein/ protein complexes that facilitate the formation of the S-shape structure. Light grey lines are the *Csp*61 restriction enzyme digestion sites. DNA fragments being studied are joined by the black curved lines and numbered. The black cross denotes to the break point occurred between fragments 7876 and 8401.

### 3.5.1 Experiment 1--△*gal80p* Strain

The first candidate protein was Gal80p. The regulation of *GAL* gene expression relies heavily on three *GAL* proteins, Gal3p, Gal4p and Gal80p (St. John and Davis, 1979a, b; Johnston and Davis, 1984; Johnston, 1987). Gal4p, the activator of *GAL* gene expression, is generally sequestered by Gal80p, the inhibitor, in the absence of galactose. Gal80p masks the function of Gal4p. The introduction of galactose triggers Gal3p, the inducer, to bind to Gal80p and causes a conformational change on Gal80p, thus releasing Gal4p that in turn allow the activation of *GAL* genes to take place. A generally accepted model is that Gal4p, Gal80p and Gal3p form a tripartite complex on the upstream of the *GAL* structural genes (St. John and Davis, 1979; Johnston and Davis, 1984), in which Gal80p plays an important facilitating role in the formation of the complex. The deletion of Gal8p that disrupts such a

formation might have an effect on DNA conformational change at the *GAL* locus. A Gal80p deletion mutant strain (YML02W) was used, and a wild-type strain (BY4741) with Gal80p not deleted was included as a control. Both were grown in SC medium. The BY4741 and YML02W strains were firstly validated by PCR on a primer pair that amplifies a 513 bp sequence at the intragenic region of *GAL80* gene. The primer pair used was: forward primer (gal80F), GACACATTACCCCGCCATAC; reverse primer (gal80R), TCTACACCGTTCCCGATTTC. This experiment was to confirm that the deletion of *GAL80* gene was present in the strain (YML02W) used.

As shown in Figure 23, the absence of PCR product in YML02W ( *gal80p)* strain proved that the *GAL80* gene was completely deleted. The wild-type strain (BY4741) served as a positive control to validate PCR.

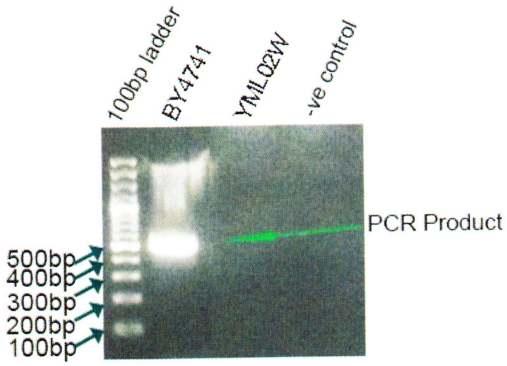



**Figure 23**: Gal80p deletion confirmation. PCR analysis confirmed that the *GAL80* gene was completely deleted in YML02W ( *gal80p*) strain. BY4741 is a wild-type strain that used as a positive (+ve) control. The expected PCR product size was 513 bp, which is indicated by green arrow. 20 µl of PCR reaction were loaded on a 1.5% agarose gel, which was run at 80 volts for 40 min and then photographed.

Following confirmation of the authenticity of the strains, a 3C study was conducted to explore whether Gal80p plays a role in the formation of S-shape structure. Because Gal80p is present at all time, all experiments were carried out under different induction conditions. As before, only the results obtained from *Csp*61 cut-ligated samples under glucose and galactose conditions are presented. These experiments were repeated at least twice.

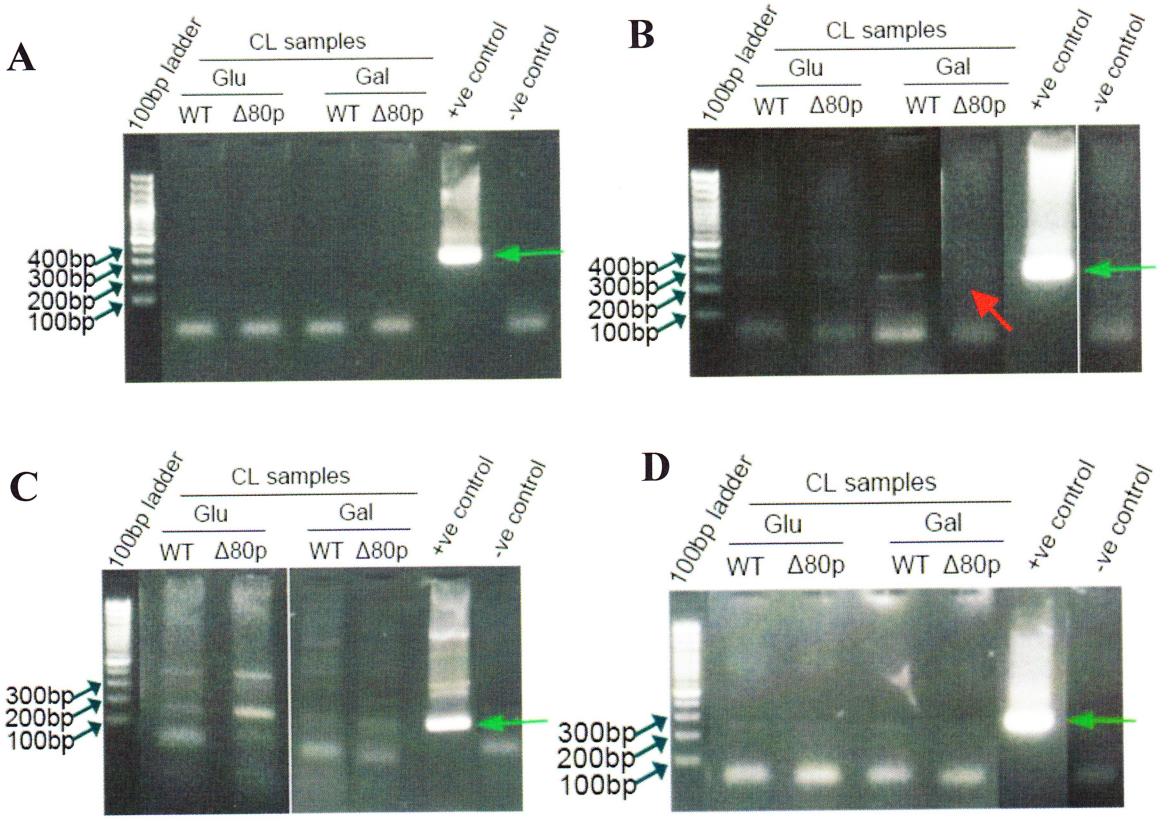



**Figure 24:** 3C analyses on Gal80p for intra-chromosomal interactions. These studies tested whether or not Gal80p was involved in mediating the formation of S-shape structure at the *GAL* locus on chromosome II. WT and 80 denote the BY4741 strain and YML02W ( *gal80p*) strain, respectively. Studies were conducted in the presence of glucose (Glu) or galactose (Gal). Only the cut and ligated (CL) samples are shown. The expected product sizes for these interactions were (A) 259 bp, (B) 287 bp, (C) 174 bp, and (D) 219 bp, which are indicated by green arrows. The red arrow points to a position where a band is absent. 20 µl of PCR reactions were loaded on a 1.5% agarose gel, which was run at 80 volts for 40 min, and then photographed.

As shown in Figure 24, it appeared that the deletion of Gal80p (YML02W) did not affect the interactions that were critical to the formation of S-shape structure at the *GAL* locus. This happened both in the presence of glucose and galactose. It is shown that interactions A, C and D occurred regardless of the yeast strains or the induction conditions. However, in interaction B, the disappearance of the band in *gal80p* strain (indicated by the red arrow), in the presence of galactose, implied that Gal80p was involved in mediating this particular interaction. As Gal80p undergoes a conformational change that releases the suppression on Gal4p during *GAL* gene activation process, deletion of Gal80p might lead to the absence of other critical proteins that play important roles in mediating this particular interaction when *GAL* genes are activated (in the presence of galactose).

### 3.5.2 Experiment 2-- *rat1-1* Temperature-sensitive Strain

The availability of a carbon source dramatically changes the mRNA level produced by the *GAL* structural genes, from an undetectable level when glucose and glycerol are present to 0.25-1% of total mRNA level when galactose is present (Johnston, 1987). In the gene transcription process, RNA polymerase II plays a critical role in catalysing the transcription of DNA to synthesize precursors of mRNA. Transcription can be divided into three stages: initiation, elongation and termination, each of which is regulated by multiple proteins/protein complexes. To understand the correlation between transcription regulation and DNA conformational changes, it was of interest to study separately the regulation systems at each transcriptional stage. As there is a distinct difference in mRNA levels of the *GAL* gene under different induction conditions, it was hypothesized that the regulation of transcription termination, *i.e.*, the capping and releasing of mRNA, had an effect on DNA conformational changes at the *GAL* locus.

*RAT1* is an essential gene required for the efficient trafficking of mRNA in *S. cerevisiae*. The *RAT1* gene product possesses a nuclear 5'-3' RNA exonuclease activity involved in RNA metabolism, including rRNA and snRNA processing as well as mRNA transcription termination (Amberg *et al.*, 1992; Johnson, 1997; Kim *et al.*, 2004). *RAT1* mutants accumulate poly (A)+ RNA in one to several intranuclear spots that lie at the nuclear periphery (Amberg *et al.*, 1992); thus causing a blockage in the termination process. A deletion mutation of *RAT1* would make yeast cells non-viable. Therefore a strain contains a mutant *rat1-1* gene that conferred temperature-sensitivity was selected. This particular strain displays mutation condition at high temperatures, namely a blockage of transcription termination (Amberg *et al.*, 1992; Kim *et al.*, 2004). RNA processing was unaffected in *rat1-1* strains, except for a defect in trimming the 5' end of the 5.8S rRNA (Kim *et al.*, 2004).

This 3C study was conducted with a *rat1-1* temperature-sensitive (TS) strain (DAT1-2) and a wild-type (FY23). Both DAT1-2 and FY23 grow in SC-H medium. In the DAT1-2 strain, the *RAT1* gene expresses normally at the permissive temperature of 23°C, but does not express at the non-permissive temperature of 37°C. The experimental set up was as follows: one set of FY23 and two sets of DAT1-2 were grown at 23°C until $OD_{600}$ reached 0.6. One set of FY23 and DAT1-2 were transferred to 37°C for 2 hr, which enabled the TS strain to be fully inactivated before induction (Amberg *et al.*, 1992). The other set of DAT1-2 was left at 23°C for this extra 2 hr. This particular set of DAT1-2 was another control that ensured any changes were not due to the change in temperature. After 2 hr inactivation, the

cultures were induced with carbon sources (glucose, glycerol/lactate or galactose) and kept at the allocated temperatures to allow gene expression. These experiments were repeated three times.

The critical interactions that contributed to the formation of the S-shape structure at the *GAL* locus were also examined (see Figure 22). As shown in Figure 25, it appeared that the defect in Rat1p, which led to the blockage of mRNA synthesis, did not appear to have any effect on the formation of S-shape structure in the presence of glucose or galactose. PCR analyses indicated that for interactions A, B, C and D, there was no difference in either the wild type or the TS mutant strains, regardless of the carbon source (Figure 25, compare WT with *rat* at 37°C). The extra control ( *rat* at 23°C) validated that the results were not due to the switch of the temperature. Therefore transcription regulation, specifically at termination stage, might not be related to the DNA conformational changes at the *GAL* locus.

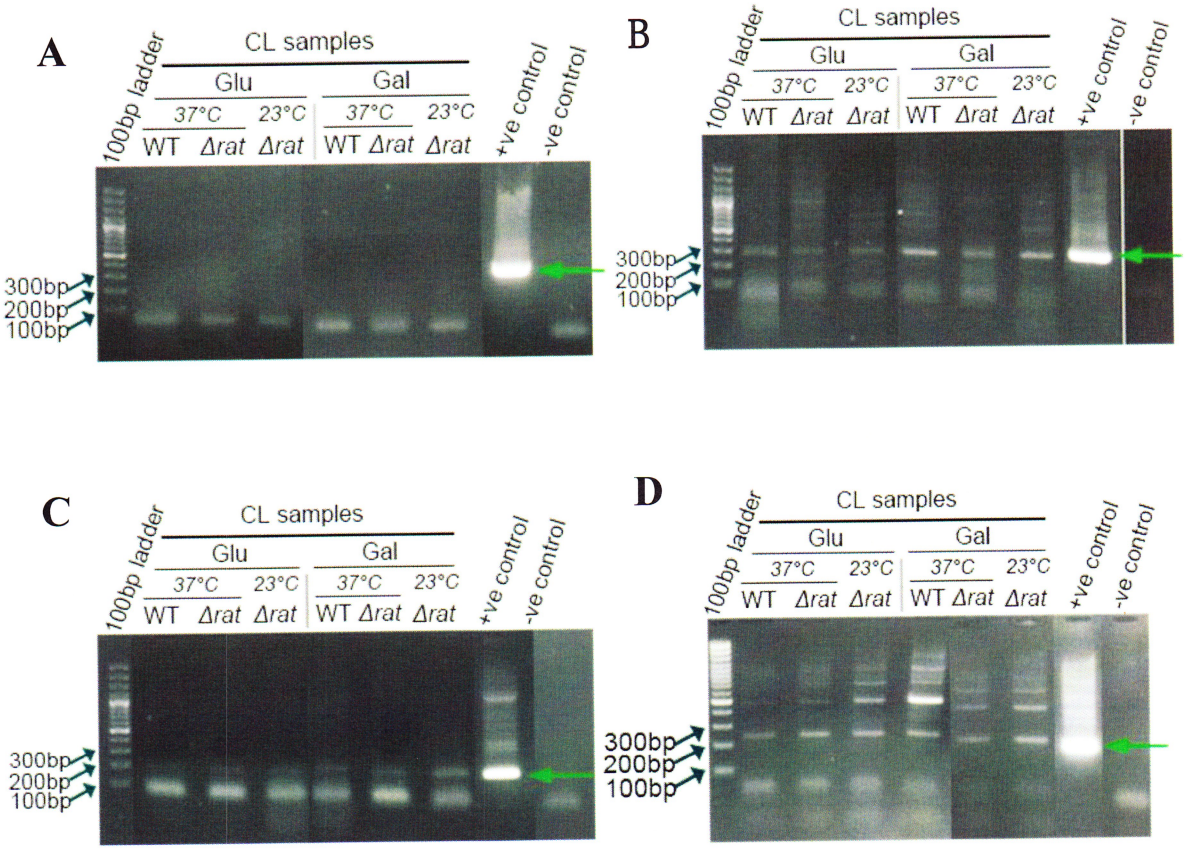



**Figure 25:** 3C analyses on Rat1p for intra-chromosomal interactions. These studies tested whether or not Rat1p was involved in the mediation of S-shape structure at the *GAL* locus. FY23 (WT) strain and DAT1-2 (TS) strain were studied in the presence of glucose (Glu) and galactose (Gal). Only the CL (cut-ligated) samples are shown. The expected product sizes for these interactions were (A) 259 bp, (B) 287 bp, (C) 174 bp, and (D) 219 bp; which are indicated by green arrows. 20 µl of PCR reactions were loaded on a 1.5% agarose gel, which was at 80 volts for 40 min, then photographed.

## 3.6 Summary

Three *GAL* structural genes, *GAL1, GAL7* and *GAL10,* are located in adjacent positions to one another on chromosome II of yeast genome. This study was designed to examine in detail the positioning of these *GAL* genes in relation to their expression. A restriction enzyme, *Csp*61, was selected to generate a desirable pattern that well separates the regulatory sequences, including the promoters and terminators, at the *GAL* locus. The novel molecular technique, 3C, was firstly employed to identify any possible interactions between DNA sequences at the *GAL* locus. Positive controls for 3C assay were generated using 2kb PCR products as starting materials. The preliminary 3C analyses, however, showed that 3C alone was inadequate to detect the unknown interactions without prior knowledge of their existence.

A more advanced approach, 4C, was then applied to detect all possible interactions occurring at the *GAL* locus using inversed nested PCR. With the aid of DNA sequencing, 4C results showed a number of intra-chromosomal interactions occurred at the *GAL* locus on chromosome II. Due to the limited number of sequenced clones, most of the observed interactions occurred between immediate adjacent fragments, which might indicate a partial restriction enzyme digestion.

The revisiting of 3C not only confirmed the observed interactions generated from the 4C analyses, but also further uncovered intra-chromosomal interactions at the *GAL* locus. Those interactions between the distant DNA restriction fragments indicated the occurrence of "DNA looping" and contributed to the modelling of an S-shape structure at the *GAL* locus.

Since the S-shape structure was present regardless of the *GAL* gene activities, it was considered to evolve at the *GAL* locus with unknown functional implications. Two proteins, Gal80p and Rat1p that each plays a distinct role in the regulation of *GAL* gene expression, were studied to determine their involvements in mediating the S-shape structure. 3C analyses were applied to study both of the proteins, of which *GAL80* gene was deleted, and Rat1p was functionally deactivated at high temperature. Although PCR analyses showed neither of them plays a prominent role in mediating the S-shape structure, Gal80p seemed to mediate one of the interactions that resulted from DNA looping when *GAL* genes are activated.

# Chapter 4----Inter-chromosomal Interactions at the *GAL* Locus of Chromosome II

## 4.1 Introduction

As previously described, the galactose (*GAL*) gene family of *Saccharomyces cerevisiae* is required for galactose metabolism in yeast cells. This gene family, consisting of structural and regulatory genes, is an important model for transcription and the build up of mRNA. The structural genes *GAL1*, *GAL7* and *GAL10* are located on chromosome II, whereas *GAL2* and *GAL5* are located on chromosome XII and XIII respectively. The products of the regulatory genes *GAL3, GAL4* and *GAL80,* which are located on chromosomes IV, XVI, and XIII, respectively, coordinate to regulate the expression of the *GAL* structural genes. Furthermore, a number of proteins are brought into close proximity at the upstream of the *GAL* structural genes to form a complex (Figure 1, Chapter 1) that is critical to the regulation of the *GAL* gene expression process.

The objective of this project was to study the positioning of *GAL* gene in response to their expression regulation. In Chapter 3, an S-shape model was proposed to form at the *GAL* locus on chromosome II due to the observed intra-chromosomal interactions. This structure does not appear to be regulated by *GAL* gene activity. In this chapter, a range of inter-chromosomal interactions between the *GAL* locus on chromosome II and other chromosomes is examined.

## 4.2 Interactions Between the *GAL* Locus and Other Chromosome

The identification of the DNA sequences that resulted from the 4C analyses is an effective approach to unmask the unknown interacting sequences. This approach not only led to the modelling of an S-shape structure at the *GAL* locus of chromosome II (Chapter 3), but also provided information with respect to the interactions between the *GAL* locus and other chromosomes.

### 4.2.1 The promoter of *GAL7* interacts with the *SVL3* gene

As stated in Chapter 3, only a few clones, *i.e.,* 2-5 out of a hundred were eventually sequenced for each sample, after confirming they indeed contained various inserts in the plasmid. For instance, five clones containing different ligated sequences from interactions with fragment 5401 (the promoter of the *GAL7* gene, see Figure 13, Chapter 3) were identified (Table 3, Chapter 3). One of these sequences showed homology to a 376 bp sequence of the *Saccharomyces* genome specifically within the intragenic region of *SVL3* open reading frame (ORF), with 99% similarity (BLAST, Appendix 9.1). The *SVL3* (YPL032C) ORF is located on the left hand side of the centromere on chromosome XVI of the yeast genome. This interaction between the promoter of *GAL7* and the *SVL3* ORF is the first evidence suggesting an interaction between chromosome II and another chromosome. This interaction was observed in the presence of galactose, *i.e.,* when the *GAL* genes are active.

### 4.2.2 The promoter of *GAL1* interacts with the *HOS1* gene

Similarly, three clones containing various ligated sequences from interactions with fragment 8401 (the promoter of the *GAL1* gene, see Figure 13, Chapter 3) were identified (Table 3, Chapter 3). One of these sequences showed homology to a 172 bp sequence of the *Saccharomyces* genome at the upstream region of *HOS1* ORF, with 95% similarity (BLAST, Appendix 9.2). The *HOS1* (YPR068C) ORF is located on the right hand side of the centromere on chromosome XVI. This interaction was also observed in the presence of galactose. This was the second evidence of an inter-chromosomal interaction occurring at the *GAL* locus. Surprisingly, this interaction also happened between chromosomes II and XVI, coincidental to the *SVL3-GAL7* interaction.

### 4.2.3 3C Confirmation of *GAL/SVL3/HOS1* interactions

These inter-chromosomal interactions were confirmed using 3C analyses. Both of the interactions, *i.e., SVL3-GAL7* and *HOS1-GAL1,* were observed in the presence of galactose (induced state). It was of interest to study whether these interactions also occurred under other conditions, *i.e.,* in the presence of glucose (repressed state) or glycerol/lactate (non-induced state).

3C primers were designed to detect interactions between the promoter of *GAL7* and *SVL3 ORF* region, and the promoter of *GAL1* and *HOS1 ORF*. For interaction between *SVL3* and *GAL7,* the *SVL3-GAL7* forward primer (within the *GAL7* promoter region) was

GCGGCTCGTGCTATATTCTT, while the *SVL3-GAL7* reverse primer (within the *SVL3* intragenic region) was GTCTGGCTCCCATTGTTTTT. For the interaction between *HOS1* and *GAL1,* the *HOS1-GAL1* forward primer (within the *GAL1* promoter region) was TGACTAAATCTCATTCAGAAGAAGTG, while the *HOS1-GAL1* reverse primer (within the *HOS1* intragenic region) was TGTCCATTTTAGAAAAACGCTATG.

3C analyses were performed using wild type W303-1a strain. The PCR results showed that the products were amplified at the expected sizes for both of the inter-chromosomal interactions (Figure 26). Furthermore, these interactions appeared to occur under all induction conditions (*i.e.,* glucose, glycerol/lactate and galactose). However, the PCR analyses in this project were conducted qualitatively rather than quantitatively, *i.e.,* the data only indicate the presence or absence of a PCR product, but not whether there were more products in one reaction than another. The conclusion so far is that both of the inter-chromosomal interactions occurred regardless of the *GAL* gene activities. The 3C assays were repeated for three times.

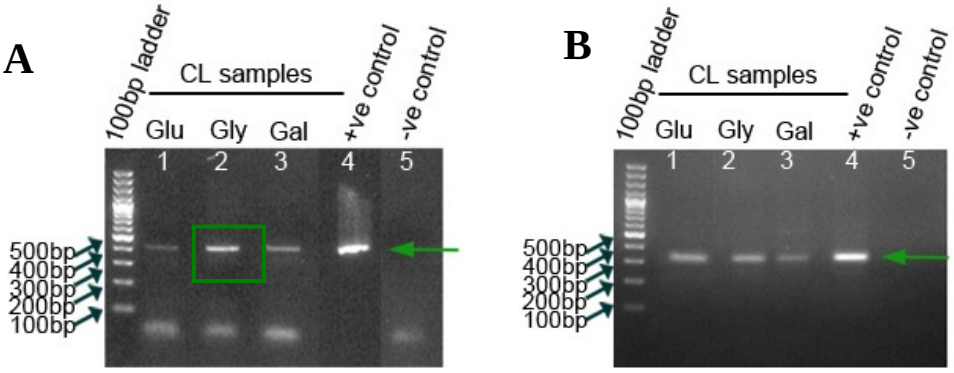



**Figure 26:** Inter-chromosomal interactions confirmation. 3C PCR analyses confirmed that the inter-chromosomal interactions between the *GAL* locus on chromosome II and two ORFs, (A) *SVL3,* and (B) *HOS1*, on chromosome XVI. W303-1a strain was cross-linked in the presence of glucose (repressed), glycerol/lactate (noninduced), and galactose (induced). Glu, Gly and Gal denote glucose, glycerol/lactate, and galactose, respectively. CL denotes *Csp61* digestion followed by ligation. Both of the positive (+ve) and negative (-ve) controls were included for PCR. PCR amplifications were performed using primer pairs (A) *SVL3-GAL7* and (B) *HOS1-GAL1*. The expected amplified product sizes were (A) 378 bp and (B) 359 bp, respectively. +ve control denotes PCR reaction using a template that can generate all possible PCR products, -ve control denotes PCR reaction using water as a template. 20 µl of PCR reactions were loaded on a 1.5% agorose gel, which was run at 80 volts for 40 min, then photographed. The band highlighted by a green square was purified, sequenced, and analyzed by BLAST.

The PCR product was isolated, gel purified and sequenced to further verify the authenticity of the interactions (Figure 26A: band indicated by a green square). BLAST analysis was also included to search for matching sequences (Appendix 9.3, with 99% similarity). As the PCR

product was made up of a *SVL3* sequence and a *GAL7* sequence at the promoter region, this further indicated the interaction is genuine.

## 4.3 Exploration of the Mediation of Interactions

Following confirmation of the inter-chromosomal interactions between the *GAL* locus on chromosome II and *SVL3/HOS1* ORFs on chromosome XVI, the next step was to study the protein/protein complexes that might facilitate the interactions. Similar to in Chapter 3, the *GAL* protein, *e.g.,* Gal80p, which represses *GAL* gene expression, and Rat1p, which plays an important role in transcription termination, were selected as candidate proteins to study their potential roles in mediating the observed inter-chromosomal interactions. Furthermore, Gal4p, the activator of the *GAL* genes, was also studied since (i) it activates *GAL* genes by forming a complex with Gal80p and Gal3p, and (ii) it recruits other regulatory proteins to form a complex to regulate *GAL* gene transcription (Figure 2, Chapter 1). In total, three experiments were performed.

### 4.3.1 Experiment 1--TAP-tagged Gal4p

The presence of Gal4p in mediating the inter-chromosomal interactions outlined above was determined using a "pull-down" assay to isolate Gal4p along with its cross-linked DNA fragments. These DNA fragments were later released from Gal4p, and PCR analyses were performed on them to determine whether Gal4p was responsible for the observed interactions. In order to be isolated, Gal4p is fused to a tandem affinity purification (TAP) tag (Section 2.3.8, Chapter 2). TAP tagged Gal4p is firstly attached to antibody by binding to beads coated with IgG. After washing away any non-specific binding substances, the TAP tag is broken by an enzyme to release Gal4p for further examination. Since the TAP-tagged yeast proteins are commercially available, we can readily obtain TAP-tagged Gal4p to carry out this study. The TAP-tagged Gal4p strain used was YPL248C, which grows in SC-Histidine medium. The wild-type (WT) strain W303-1a was included as a control. As Gal4p is, at all time, bound to Gal80p (Figure 2, Chapter 1), the study was conducted under all induction conditions followed a 4C pulldown protocol (Section 2.8.4, Chapter 2).

The principle of the 4C pulldown assay was described previously (Figure 6, Chapter 2). In a similar to the 3C and 4C assays previously described, each set of experiments consisted of

an uncut, a cut and a cut-ligated sample. After the TAP-tagged protein had been isolated using IgG coated beads (Section 2.3.8.1), each of these subsequently became divided into a pull down sample and a depleted sample. Thus each set of experiments gave rise to a total of six products, under one induction condition. The pull down samples contained linked DNA fragments/protein, and the depleted samples contained the remaining DNA and proteins after protein isolation. However, in practice, protein isolation by antibody cannot reach 100% efficiency. There some TAP-tagged proteins, along with their cross-linked DNA fragments, would still be present in the depleted samples. Depleted samples served as another form of experimental control in addition to the wild-type strain.

Firstly in order to understand whether Gal4p was indeed isolated by antibody in the experiments, primers were designed at *GAL10* $UAS_{gal}$ site. Gal4p is constantly bound at $UAS_{gal}$, it is therefore certain that $UAS_{gal}$ DNA sequence was cross-linked to Gal4p and isolated during the process. There is one $UAS_{gal}$ site located upstream of each of the *GAL* genes (Figure 2, Chapter 1), and the one that is upstream of *GAL10* gene was chosen for primers design. The primers were: $UAS_{gal}$ *GAL10* primer forward (at the *GAL10* $UAS_{gal}$ site), GGATGGACGCAAAGAAGTTTA, and $UAS_{gal}$ *GAL10* primer reverse (at the *GAL10* $UAS_{gal}$ site), GCCAGGTTACTGCCAATTTT.

As shown in Figure 27, a PCR product with the expected size was amplified in TAP pulldown samples, in the presence of glucose or galactose (Figure 27: lanes 2 and 4). In WT pulldown samples (Figure 27: lanes 1 and 3), no PCR product could be amplified, as no TAP-tagged Gal4p was present. This indicated that Gal4p isolation was indeed achieved in the process. The presence of products in depleted samples (Figure 27: lanes 5, 6, 7, and 8) indicated the experiments were working appropriately. These experiments were repeated at least twice.

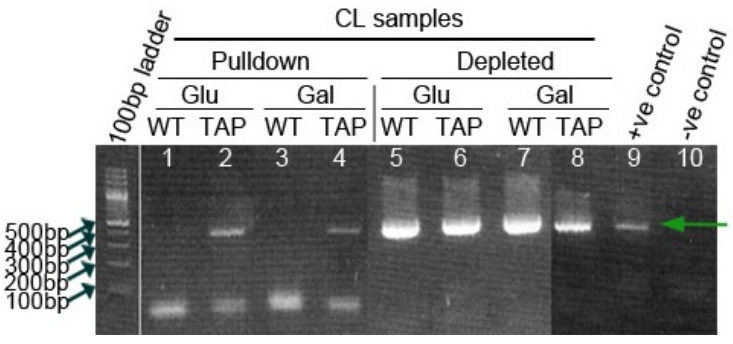



**Figure 27:** 3C PCR analysis confirm the isolation of Gal4p in 4C pulldown experiment. Wild-type (WT) and Gal4p TAP-tagged (TAP) strains were studied in the presence of glucose (Glu) and galactose (Gal).

Gal4p was isolated (Pulldown) and the remaining (Depleted) was retained. PCR amplification was performed using $UAS_{gal}$ *GAL10* primer pair. The expected product size was 363 bp, which is indicated by a green arrow. +ve control denotes PCR reaction using a template that can generate all possible PCR products, -ve control denotes PCR reaction using water as a template. 20 µl of PCR reactions were loaded on a 1.5% agarose gel, which was run at 80 volts for 40 min and then photographed.

For the observed inter-chromosomal interactions, as shown by PCR analyses, both of the *SVL3-GAL7* (Figure 28A) and *HOS1-GAL1* (Figure 28B) interactions were lost in TAP pulldown samples (Figure 28: lanes 2 and 4), but were conserved in depleted samples (Figure 28: compare lanes 2 and 4 with lanes 6 and 8). This occurred regardless of the availability of the carbon source. Therefore, it seemed that Gal4p was not responsible for the occurrence of the interactions between chromosomes II and XVI (neither *SVL3-GAL7* nor *HOS1-GAL1*). Thus Gal4p might not have a functional role in mediating these inter-chromosomal interactions. The experiments were repeated at least twice.

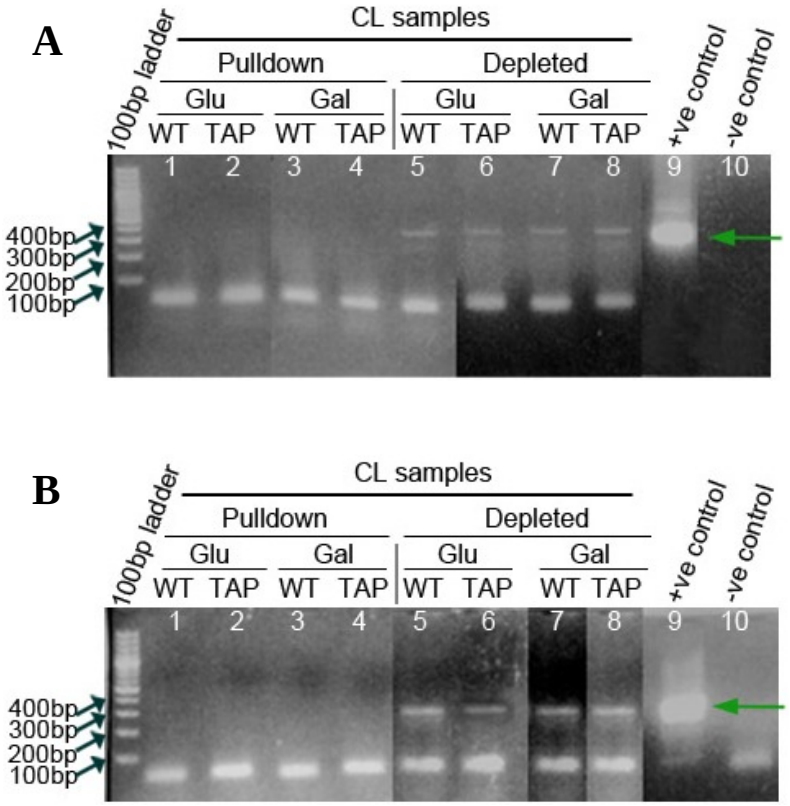



**Figure 28:** 3C analyses on Gal4p for inter-chromosomal interactions. These studies tested on whether Gal4p was involved in mediating the inter-chromosomal interactions between the *GAL* locus on chromosome II and *SVL3* ORF (A) or *HOS1* ORF (B) on chromosome XVI. Wild-type (WT) and Gal4p TAP-tagged (TAP) strains were studied in the presence of glucose (Glu) and galactose (Gal). Gal4p was isolated (Pulldown) and the remaining (Depleted) was retained. The expected product sizes for these interactions were (A) 378 bp and (B) 359 bp, respectively, which are indicated by green arrows. +ve control denotes PCR reaction using a template that can generate all possible PCR products, -ve control denotes PCR reaction using water as a template. 20 µl of PCR reactions were loaded on a 1.5% agarose gel, which was run at 80 volts for 40 min and then photographed.

### 4.3.2 Experiment 2--*Δgal80p* strain

The next study was conducted using Gal80p, the inhibitor of *GAL* gene expression. Similar to the case described in Chapter 3, it is hypothesized that Gal80p plays a role in mediating the inter-chromosomal interactions. This experiment was carried out using a deletion mutant strain (*Δgal80p*) followed the 3C protocol, and it was conducted under all induction conditions due to the fact that Gal80p is present in a complex with Gal4p at all times. In a similar manner to that described in Chapter 3, both a Gal80p deletion mutant strain (YML02W; *Δgal80p*) and a wild type strain (BY4741) were used in this study (Section 3.5.1, Chapter 3).

As shown in Figure 29, in both the BY4741 (wild type) and YML02W (deletion mutant) strains, inter-chromosomal interactions occurred with no apparent difference regardless of the induction conditions (Figure 29: compare lanes 1 and 3 to lanes 2 and 4). The PCR results indicated that Gal80p might not be responsible for mediating either of the inter-chromosomal interactions. These experiments were repeated at least twice.

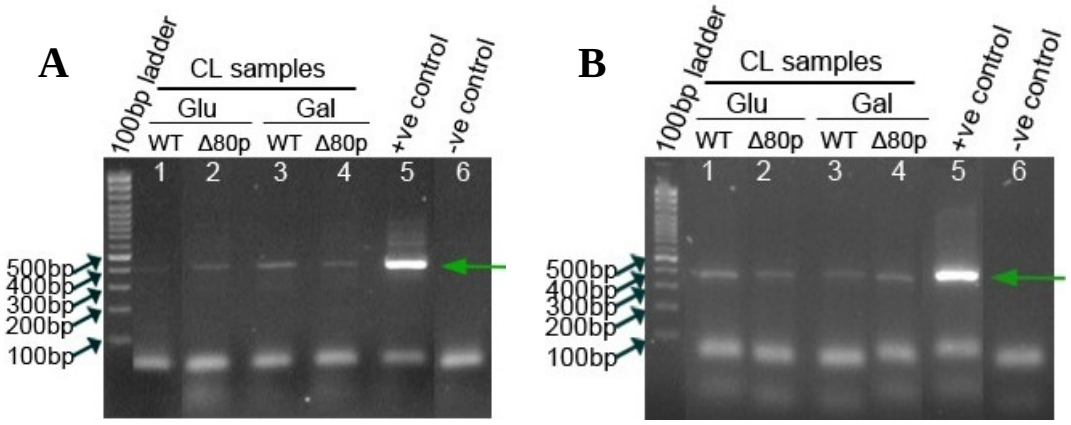


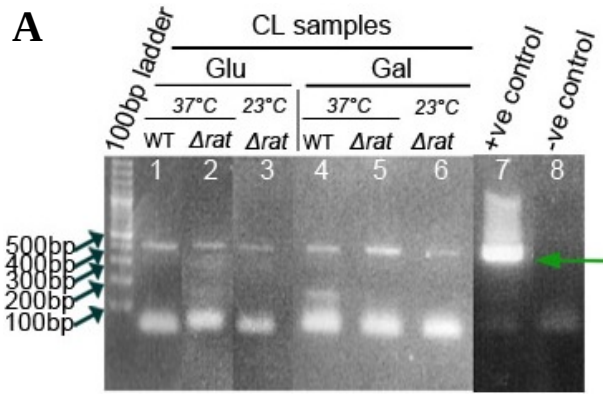

**Figure 29:** 3C analyses on Gal80p for inter-chromosomal interactions. These studies tested whether Gal80p was involved in mediating the inter-chromosomal interactions between the *GAL* locus on chromosome II and (A) *SVL3* ORF or (B) *HOS1* ORF on chromosome XVI. WT and Δ80 denote BY4741 strain and YML02W strains respectively. Studies were conducted in the presence of glucose (Glu) or galactose (Gal). Only the cut and ligated (CL) samples are shown. The expected sizes for these interactions were (A) 378 bp and (B) 359 bp, respectively, which are indicated by green arrows. +ve control denotes PCR reaction using a template that can generate all possible PCR products, -ve control denotes PCR reaction using water as a template. 20 µl of PCR reactions were loaded on a 1.5% agarose gel, which was run at 80 volts for 40 min and then photographed.

### 4.3.3 Experiment 3--*rat1-1* temperature-sensitive strain

During the gene transcription process, RNA polymerase II plays a critical role in catalysing the transcription of DNA to synthesize precursors of mRNA. The *RAT1* gene product, Rat1p, which processes a nuclear 5'-3' RNA exonuclease and functions in mRNA transcription termination, plays an important role in transcription regulation. Therefore, Rat1p was considered as a candidate protein to study DNA conformational changes in response to transcription.

Similar to the experiment outlined in Chapter 3, a temperature-sensitive mutant *rat1-1* strain was used to study whether the blockage of transcription termination had any effect on the inter-chromosomal interactions between the *GAL* locus and the *SVL3/HOS1 ORFs* of chromosome XVI. 3C studies were carried out using FY23 (wild-type) and DAT1-2 (temperature-sensitive TS) strains.

As shown in Figure 30, the interactions appear to occur regardless of the strain type and induction conditions (Figure 30: compare lanes 1 and 4 to lanes 2 and 5). The control experiment with strain DAT1-2 (*Δrat)* at 23°C confirmed that the changes of temperature did not make any differences to the interactions taking place (Figure 30: lanes 3 and 6). Therefore it seemed that the blockage of transcription termination, which resulted in incomplete mRNA production, did not affect the interactions occurring between chromosomes II and XVI. These experiments were repeated at least twice.



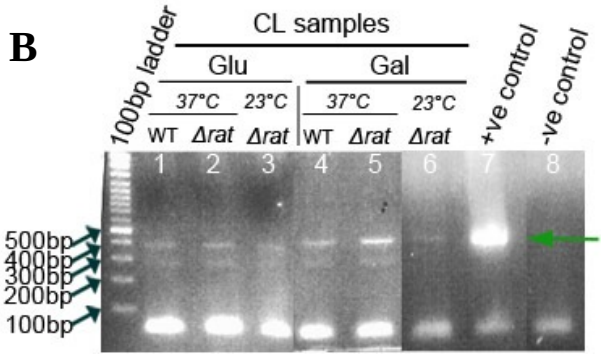



**Figure 30:** 3C analyses on Rat1p for inter-chromosomal interactions. These studies tested whether Rat1p mediates the inter-chromosomal interactions between the *GAL* locus on chromosome II and (A) *SVL3* ORF or (B) *HOS1* ORF on chromosome XVI. WT and *Δrat* denote FY23 and *DAT1-2* strains respectively. Studies were conducted in the presence of glucose (Glu) or galactose (Gal). The cut and ligated (CL) samples were shown. The expected sizes for these interactions were (A) 378 bp and (B) 359 bp, respectively, which are indicated by green arrows. +ve control denotes PCR reaction using a template that can generate all possible PCR products, -ve control denotes PCR reaction using water as a template. 20 µl of PCR reactions were loaded on a 1.5% agarose gel, which was run at 80 volts for 40 min and then photographed.

## 4.4 Quantitative Real-time PCR Analysis

Up to this point, standard qualitative PCR had been applied to determine the presence of any chromosomal interactions. However, as the interactions seemed to occur under all conditions, and there were no protein or protein complexes identified to mediate these interactions, qualitative PCR was insufficient to determine levels of interactions. Furthermore, as PCR was performed qualitatively *i.e.* the concentration of the DNA template used was not quantified prior to PCR reactions; the resulting products could not be compared by their quantities. It was therefore not possible to examine whether there is a higher frequency of the interaction occurring in one sample than in another. Hence real-time PCR was used to measure possible small changes in interaction levels between cells grown in glucose. Glycerol/lactate, and galactose.

Using quantitative/real-time PCR, the focus of this part of the project was to determine the interaction frequency between *SVL3-GAL7* under various induction conditions. *SVL3-GAL7* interaction was the first inter-chromosomal interaction discovered, but no functional implication had been shown. It was assumed that the real-time PCR approach might be useful to gain a better understanding of this particular interaction.

The real-time PCR applications consisted of two aspects. First the concentration of DNA template used, the "DNA input", was determined for each PCR reaction. This was achieved

by applying *SyBr* green real-time PCR. Second the interaction frequency between *SVL3-GAL7* was determined using *Taqman*® real-time PCR.

### 4.4.1 *SyBr* green real-time PCR

The *SyBr* green real-time PCR system is dependent upon a *SyBr* green I dye that intercalates with double-stranded (ds) DNA and produces a fluorescent signal. The intensity of the signal is proportional to the amount of dsDNA present in the reaction. Therefore, at each step of the PCR reaction, the signal intensity increases as the amount of product increases (Wittwer *et al.,* 1997). This provides a very simple and reliable method to monitor PCR productions in real time. The *SyBr* Master mixture (Invitrogen) contains all the essential components for a real-time PCR to proceed. Reaction conditions for each primer pair need to be optimised: this includes primer concentration, annealing temperature and magnesium chloride concentration (Wittwer *et al.,* 1997; Morrison, *et al.*, 1998)

An internal primer pair that spans part of the *GAL10* gene was specifically designed to facilitate the quantification of DNA template concentration into each PCR reaction. The principle of this design was that the copy number of a random internal DNA fragment remained identical for the tested samples to be monitored, and therefore the DNA input of each sample could be adjusted. This particular section of the intragenic region of *GAL10* was selected because it is free of restriction enzyme digestion sites (Figure 31); hence it avoids the digestion and re-ligation processes. As far as the *SyBr* green based amplicon detection is concerned, it was important to run a dissociation curve following the real time PCR. This is due to the fact that *SyBr* green will detect any double stranded DNA including primer dimers, contaminating DNA, and PCR product from mis-annealed primer. A dissociation curve analysis ensures the desired amplicon has been detected (Wittwer *et al.,* 1997).

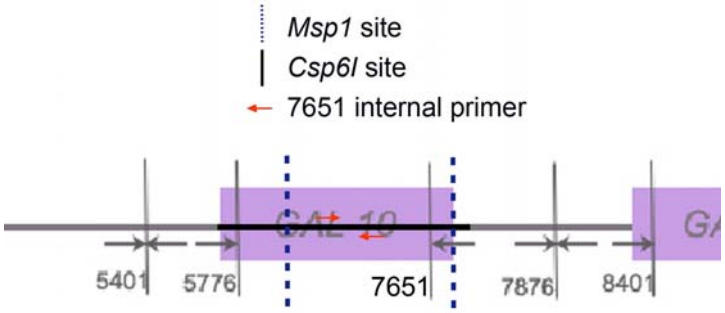
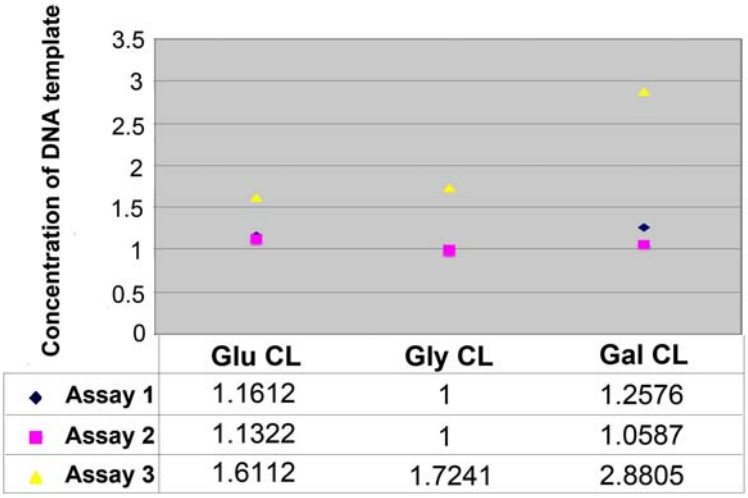



**Figure 31:** Diagrammatical representation of internal primers designs at the *GAL10* gene. The target region is within the intragenic region *GAL10* gene. This region where primers (red arrows) are designed is free of restriction digestion sites *i.e. Msp1* (blue dotted line) or *Csp61* (black solid line). The information of primer sequences is in Appendix 3.0.

In this study, the concentration of primer pair mix was selected between 200 nM and 1000 nM using dissociation curve analysis. It was determined that the lower primer concentration 200 nM, which generated a relatively smaller secondary peak, was more appropriate under the conditions (Appendix 10.3). *SyBr* real-time PCR required standards with known DNA quantity. The standards used in this study were a series of quantified dilution of W303-1a genome DNA that was digested by the restriction enzyme *Msp1*. As the DNA quantity of each standard was known, the DNA input into each PCR reaction could therefore be calculated.

Three independent 3C assays that performed in the presence of glucose, glycerol/lactate and galactose were used in this study. Only the cut-ligated samples were analysed. For every real-time PCR reaction, the concentration of DNA template used was calculated and the results were plotted (experimental raw data are shown in Appendix 10.2). Every reaction was studied in triplicate.




| | Glu CL | Gly CL | Gal CL |
|---|---|---|---|
| Assay 1 | 1.1612 | 1 | 1.2576 |
| Assay 2 | 1.1322 | 1 | 1.0587 |
| Assay 3 | 1.6112 | 1.7241 | 2.8805 |

**Figure 32:** *SyBr* green real-time PCR that indicates DNA template concentration into each PCR reaction. These PCR analyses were performed using 3C cut-ligated (CL) samples that obtained under the conditions of glucose, glycerol/lactate and galactose. The values were calculated from triplicates of three independent assays.

As shown in Figure 32, the concentration of DNA template for each induction condition *i.e.* glucose, glycerol/lactate or galactose, was similar to one another within one assay. However, in assay 3 (Figure 32: yellow line), there was a big increase in the galactose sample. It is therefore very important to quantify DNA template concentration prior to gene expression evaluation. Moreover, the purpose of this study was to ensure the values, within each assay, *e.g.,* Glu CL, Gly CL and Gal CL, are similar as they underwent the same experimental procedure. Therefore the values between each independent assay are not comparable.

### 4.4.2 *Taqman*® real-time PCR

The next experiment was to determine the interaction frequency for *SVL3-GAL7*. This was achieved with the application of *Taqman*® real-time system (Section 2.4.2, Figure 8).

Standard are included in the *Taqman*® real-time PCR. These standards are derived from a 3C PCR product of *SVL3-GAL7* interaction. The product was generated from a cut-ligated sample and was purified through a High Column PCR purification kit. The DNA template concentration was determined to be 67.1 ng/µl using Nanodrop. DNA was then diluted initially to a 2 ng/µl and subsequently through a series of dilution to make a set of appropriate standards (Appendix 10.0).

A primer pair and a *Taqman*® probe were specifically designed and accustomed for *SVL3-GAL7* interaction according to the manufacturer's instructions (Section 2.4.2; www.ambion.com). The optimal primer concentration and optimal probe concentration were also determined by following the manufacturer's instructions. A forward/reverse concentration of 900 nM/900 nM and a *Taqman*® probe concentration of 200 µM were determined as being optimal.

The results of the *SVL3-GAL7* interaction frequency using *Taqman*® real-time PCR applications are shown in Table 4. The cut-ligated samples from each induction condition were analysed in triplicate (experimental raw data are shown in Appendix 10.3). The X values in Table 4 are the original data, which were corrected according to quantification of DNA template that was described in *SyBr* green Real-time PCR.

These original values are very small and there is a big statistical difference between each assay *i.e.* the values of assays 1 and 2 are at least ten times smaller than those of assay 3. It is therefore necessary to convert them in a way that compresses the values so that they are more easily compared (Nancy Xu and Mike Anderson, personal communications). These values were therefore converted to Log algorithm using an expression of: Y = Log (X), in which X is the original value, and Y is the $Log_{10}$ value (Table 4). The Log values from three assays were analysed (Appendix 10.5) and the averages are illustrated in Figure 33.

**Table 4:** Calculation of *SVL3-GAL7* interaction frequency. The cut-ligated (CL) samples from three independent 3C assays were examined and analysed using *Taqman*® real-time PCR to study the interaction

frequency of *SVL3-GAL7*. The values X are the originals that were corrected based on the results of *SyBr* green real-time PCR, and $Log_{10}$ values Y are calculated using expression Y = Log (X).

| | Assay 1 | | Assay 2 | | Assay 3 | |
|---|---|---|---|---|---|---|
| | Value (X) | $Log_{10}$ (Y) | Value (X) | $Log_{10}$ (Y) | Value (X) | $Log_{10}$ (Y) |
| **Glu CL** | $1.99 \times 10^{-6}$ | -5.702 | $2.07 \times 10^{-7}$ | -6.685 | $7.93 \times 10^{-8}$ | -7.101 |
| **Gly CL** | $3.4 \times 10^{-7}$ | -6.531 | $1.37 \times 10^{-7}$ | -6.864 | $7.48 \times 10^{-8}$ | -7.126 |
| **Gal CL** | $5.27 \times 10^{-7}$ | -6.278 | $1.1 \times 10^{-7}$ | -6.957 | $6.49 \times 10^{-8}$ | -7.188 |

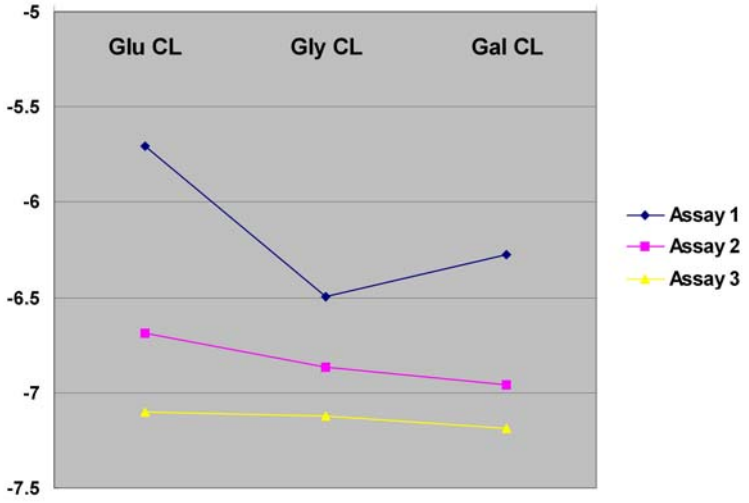



**Figure 33:** Line graph illustrates the interaction frequency between *SVL3-GAL7* under various induction conditions in $Log_{10}$ values, which are displayed as negative numbers. Three independent 3C assays are included and only cut and ligated (CL) samples are analysed.

The column graph in Figure 33 shows a decline in the interaction frequency for *SVL3-GAL7* from glucose to glycerol/lactate and galactose conditions, *i.e.* when the *GAL* genes are activated, at both basal level of expression (glycerol/lactate) and robust level of expression (galactose). The results suggested that the interaction between *SVL ORF* and *GAL 7* promoter might become "looser" when the *GAL* genes are in the process of being activated. However, the error bars are large when comparing between the samples, which might suggest the results are less significant.

## 4.5 Summary

In this section, it was shown that two inter-chromosomal interactions occurred at the *GAL* locus, *SVL3*-*GAL7* and *HOS1*-*GAL1*. These were identified by 4C analyses with DNA sequencing. Both *SVL3* and *HOS1* are open reading frames (ORFs) on chromosome XVI. Therefore a novel chromosomal interaction between chromosomes II and XVI was established. 3C analyses were used to confirm these interactions, and PCR product was also isolated and sequenced to further verify the identity of interaction between *SVL3*-*GAL7*. These inter-chromosomal interactions, *SVL3*-*GAL7* and *HOS1*-*GAL1,* appeared to occur despite the *GAL* gene activities. Their constant appearance under all induction conditions suggested that they might not be determined by the regulation of *GAL* gene expression.

Three experiments were conducted on three regulatory factors to identify the proteins that mediate these observed inter-chromosomal interactions. The candidate proteins were: Gal4p, Gal80p and Rat1p, each of them plays a role in the regulation of *GAL* gene activities. The experimental approaches used in studies of these proteins were as follows. Gal4p and its associated DNA were isolated by antibody, The *GAL80* gene was deleted and Rat1p was functionally deactivated at high temperature. Gal4p isolation was achieved using a 4C pulldown assay, whereas the other two were studied using 3C analyses; all of them were analysed by PCR. The results showed that none of these proteins had functional implications in mediating the observed inter-chromosomal interactions.

Therefore, real-time PCR applications were applied to study the interaction frequency for *SVL3*-*GAL7* under different induction conditions. The concentration of DNA template used in each PCR reaction was firstly determined using *SyBr* green real-time PCR. DNA templates were obtained from cut-ligated samples of three independent 3C assays. The interaction frequency for *SVL3*-*GAL7* was then examined using *Taqman*® real-time PCR. With the aid of a *Taqman*® probe, PCR amplification was performed across the junction between the *SVL3*-*GAL7* genes. The interaction frequency was calculated according to the concentration of DNA template used. Because they appeared to be very small, they were converted to $Log_{10}$ algorithm for a better comparison. The results showed a gradual decline in interaction frequency from glucose to glycerol/lactate and galactose. This decline might suggest a change of DNA structure at the *GAL* locus in response to *GAL* gene activation. However, large error bars indicated the results are less significant than they appear to be.

# Chapter 5----Discussion and Summary

## 5.1 Studies on DNA Structural Changes at the *GAL* Locus

### 5.1.1 Non-linear DNA Structure at *GAL* Locus on Chromosome II

The galactose (*GAL*) gene family in yeast has been a popular target for studies in molecular biology over the past few decades. More than a hundred studies have been conducted that significantly contributed to the establishment of a solid foundation for understanding the *GAL* gene regulatory mechanism. However, the majority of the studies were only concerned with functional properties of the *GAL* gene products and how the interplay of proteins regulates the *GAL* gene activity. The unique positioning of the three *GAL* structural genes on chromosome II, *GAL1, GAL7* and *GAL10* that are adjacent to each other (Figure 9), implies that the construct itself influences the regulation of gene expression. Since this gene family is strictly regulated by the availability of a carbon source, this project was designed to study whether there are any DNA structural changes at the *GAL* locus in response to *GAL* gene expression.

The entire *GAL* locus spans a distance of less than 7 kb on chromosome II. When studying intra-chromosomal interactions occurring at the *GAL* locus within chromosome II, PCR analyses suggested that DNA looping occurred at this region and an S-shape structure was modelled (Figure 21). This is direct evidence for the non-linear DNA structure at the *GAL* locus. Furthermore, as the S-shape structure exists generically despite the induction conditions for the *GAL* genes, its formation seemed not to be regulated by *GAL* gene expression. It appears that this structure has evolved as part of the chromosome conformation in the yeast genome. It is also possible that this construct is essential to the co-expression of three *GAL* structural genes, which are also functionally related. Through the loop formation, these genes are brought into close proximity and thereby gain access to the common regulatory machineries allowing efficient expression. This effect is similar to the "transcription factory" (Cook, 1995; Iborra *et al.*, 1996) in which genes extend out of their chromosome territories in *cis* and *trans* configuration in order to access to a shared transcription factory that contains a high density of RNA polymerase II. As *GAL* gene activation is robust in response to galactose (Johnston, 1987), it is possible that such a "transcription factory" would have existed around the *GAL* locus.

A surprising finding was that the *GAL* locus is divided into a "two interaction zone" with a breakpoint between the *GAL10* and *GAL1* genes (Figure 19: between adjacent fragments 7876 and 8401). This also occurred regardless of any gene activity status. This breakpoint was firstly observed in preliminary 3C studies (Figure 12A), and later confirmed by "walking-down" 3C analyses (Figure 18). Furthermore, 4C clonal sequencing results also showed that fragment 8401 did not interact with fragment 7876 (Figure 17C). Previous studies have shown that there are several hypersensitive sites (HS) at this point (Proffitt, 1985). A HS is a short region of chromatin that is specifically vulnerable to nuclease (DNase I and other restriction enzymes) attack because it is not wrapped in histone, unlike those for nucleosomes. The unusual nucleosomal structure in HS results in a loss of protection of DNA by protein. Thus there is a 100-fold increase in sensitivity to enzyme attack in comparison with bulk chromatin. The HS may result from the disruption of the normal repeated nucleosomal chromatin structure by the binding of sequence-specific proteins. It has been shown that at the *GAL* locus, HS is present during both galactose induction and glucose repression of transcription. This might be due to the binding of one or more of the regulatory proteins that act under both trancriptional states (Proffitt, 1985). It is assumed that these regulatory proteins are Gal4p and Gal80p. These proteins form a complex at the $UAS_{gal}$ site, which is overlapping with HS on chromosome II (Proffitt, 1985). Between the *GAL10* and *GAL1* genes, the $UAS_{gal}$ sites locate within fragment 8401, while HS spans across the junction of fragments 7876 and 8401 (Figure 34). It is understandable that this particular position is vulnerable to attack by restriction enzymes.

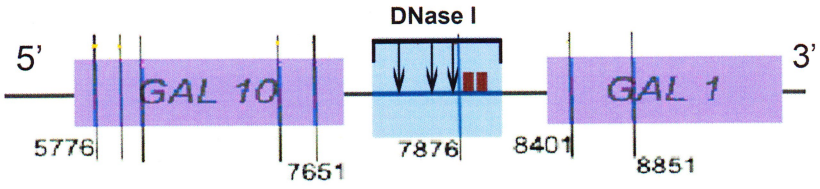



**Figure 34:** Location of the DNase I-hypersensitive sites between *GAL10* and *GAL1* on chromosome II. The purple rectangles represent the DNA sequence of the *GAL* genes. The restriction enzyme *Csp*61 cleave sites are shown by the black vertical lines, and the restricted fragments were named according to their nucleotide positions. Blue rectangle represents the DNase I-hypersensitive region. Downward arrows represent centers of DNase I cutting. Dark brown squares represent $UAS_{gal}$ sites.

An interesting observation was that a lot of intra-chromosomal interactions either resulted from self-ligation (Figure 16 and 17: compare the colored rectangles and their target fragments), or if not, the interacting fragments happened to be located to the downstream of the target fragments, as demonstrated by the clonal sequencing results. This seems to largely occur under the galactose condition (Figure 17C). This pattern was consistent for three different target fragments being studied. A possible assumption is that they are derived from

partial *Csp*61 digestion. An exception to this was that in direct sequencing an interaction occurred between non-adjacent fragments (Figure 16, green rectangles). It is likely that as only a few clones were sequenced for each induction condition, the interpretation of the results must, necessarily, be limited.

In 3C analyses that confirmed the 4C results (Figures 18 and 20), an interesting observation was the multiple PCR banding pattern shown with a number of primer pairs (Figure 18: 3, 4; Figure 20: 1, 2 and 3). Due to the time restraint, we were not able to sequence them to reveal the identities. However, based on the sizes of these extra bands, the products are likely to be derived from the co-catemerization of ligation, that is, ligation takes place between multiple copies of the same DNA fragment. If this were the case, it could be related back to the results of clonal sequencing in which the majority clones containing multiple inserts of the same fragment.

In current study both the 3C and 4C studies included samples containing non-cross-linked DNA/proteins as an extra control. The purpose of this was to understand the nature of the interactions, whether they were dependent upon the cross-linking agents. It was interesting to observe that some interactions occurred without cross-linking. For instance, in Figure 18, interaction 7 occurred in both non-cross-linked and cross-linked samples. Furthermore, in Figure 14, the pattern shown in the ligation products of both the non-cross-linked and cross-linked samples appeared to be similar (Figure 14A, 14C, and 14D). However, direct sequencing indicated that the interactions observed in the non-cross-linked samples were mostly resulting from self-ligation (Figure 16), whereas the interactions observed in the cross-linked samples mostly occurred between the adjacent DNA fragments. Therefore even though 4C PCR results showed a similar ligation pattern, further analyses suggested they did not necessarily represent the same interactions. In terms of interaction 7 in Figure 18, a valid assumption could be that the interaction was protected by certain unknown proteins. This protection was so strong and stable that no cross-linking agent was required to facilitate the interaction. It is important to notice that due to the unique primer design in the 3C assay, only DNA fragments that underwent digestion and ligation procedures could be detected by PCR analyses (Figure 10, all primers facing the same direction). Therefore in the case of Figure 18 (7), restriction enzyme digestion and ligation still took place but the initial fixation between DNA and proteins did not require formaldehyde (cross-linking agent). Furthermore, it is also possible that because 3C assays were conducted with a population of cells that were not synchronized (see Section 5.3), the interactions that occurred in non-cross-linked samples might be a reflection of cell-cell variation of a population of cells at different phases.

Furthermore, a potential technique pitfall in both the 3C and 4C analyses is that the formaldehyde treatment in the live cells can interfere with the restriction enzyme digestion procedure. However, in current study, restriction enzyme digestion took place in crude chromatin preparation. The amount of enzyme and the length of digestion time were increased dramatically (100 U and 2-3 hours) in order to accommodate with the sample conditions. Therefore the possible influence of formaldehyde is considered to be less significant.

### 5.1.2 Interactions Between Chromosomes II and XVI

When studying the inter-chromosomal interactions at the *GAL* locus, the results of 4C clonal sequencing indicated a novel interaction between chromosomes II and XVI through two independent interactions. One of these interactions was between the *SVL* gene and the *GAL7* promoter, the other was between the *HOS1* gene and the *GAL1* promoter. Subsequent 3C analyses and sequencing verified these interactions. These interactions also occurred regardless of the *GAL* gene activity. Taken together, these results provide evidence for a model comprising an S-shape structure formed at *GAL* locus on chromosome II. In addition a crossing-over structure of chromosome XVI was proposed as a model of the interactions that were observed at the *GAL* locus (Figure 35).

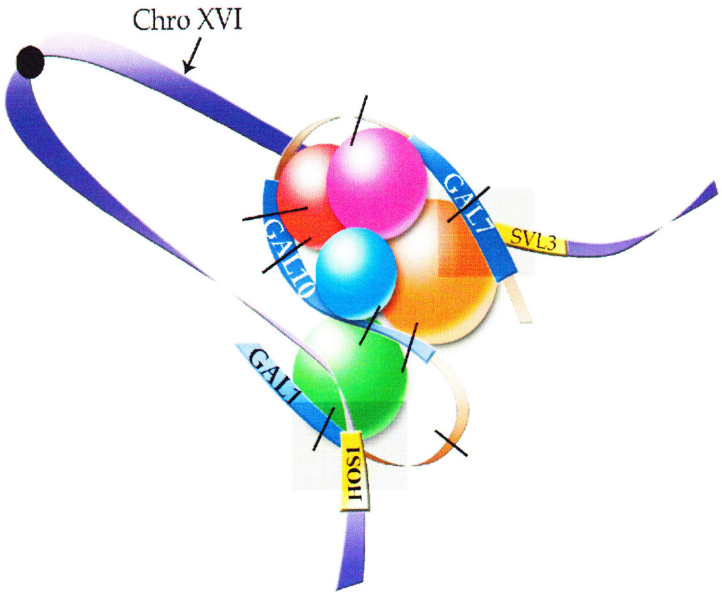



**Figure 35:** Combination model of chromosomes II and XVI. A combination model that consists of an S-shape structure formed at the *GAL* locus of chromosome II and a crossing-over structure of chromosome XVI. This model represents a generic structure formed between chromosomes II and XVI, in spite of the *GAL* gene activity. The brown ribbon represents DNA sequences of chromosome II, while the blue ribbons represent the *GAL* genes. The coloured spheres represent the protein/protein complexes that contribute to the formation of the S-shape structure. Black vertical lines are the *Csp61* digestion sites. The purple ribbon represents chromosome XVI, while yellow ribbons represent the specific genes. The black dot on the purple ribbon denotes to the centremere. This model is not drawn in scale.

Among these interactions, *SVL3* is a verified *ORF* that is located on the upstream of the centromere on chromosome XVI. The *SVL3* gene product is of unknown function. *SVL3* mutations result in large vacuoles, which suggest a potential role in vascuolar function (Zheng *et al.*, 1998). There have not been many studies focused on this particular gene type and its gene product, thus the observed interaction between *SVL3-GAL7* could not be comprehended from a functional point of view. As far as gene expression is concerned, a study conducted by Gasch *et al.* (2000), using DNA microarrays to measure changes in transcript levels for yeast genes, has shown that both the *GAL* and *SVL3* genes experience dramatic changes in their expression level in response to environmental changes. These environmental changes included temperature shocks, amino acid starvation, and nitrogen source depletion. In comparison, both of the *GAL* and *SVL3* genes express stably in response to a wide range of other conditions, including response to cell cycle change, DNA-damaging agents, histone depletion, and zinc levels (*Saccharomyces* genome database). Therefore given their physical locations are in a close proximity, it is possible that *SVL3* and *GAL7* genes share common machinery or "co-localize" at a "transcription factory" that suits their common expression pattern.

*HOS1* is located on the downstream of the centromere on chromosome XVI. *HOS1* encodes a histone deacetylase (HDAC) (Rundlett *et al.*, 1996). The deletion mutation result in an increasing histone acetylation at rDNA repeats. In yeast and other eukaryotes, covalent modifications of histones, including acetylation and deacetylation, are important for transcriptional regulation (Mizzen *et al.*, 1996). At first glance it seems possible that *HOS1* gene product plays a role in the regulation of *GAL* gene transcription. However, as the interaction between *HOS1-GAL1* occurs regardless of the *GAL* gene activities, it seemed that such interaction might not have any functional implications. Similar to the *SVL3-GAL7* interaction, *HOS1* gene expression level is also stable in response to a number of conditions (*Saccharomyces* genome database), but experiences significant changes in response to environmental changes (Gasch *et al.*, 2000). Therefore the *HOS1-GAL1* interaction might also be due to the requirement for co-expression regulation.

Previous studies have shown that each chromosome occupies a distinct area within the nucleus, which is referred as "chromosome territory" (Cremer *et al.*, 1993; Volpi *et al.*, 2000; Feuerbach *et al.*, 2002). The position of each chromosome territory is fixed with respect to the whole genome. Since the chromosomal interaction between chromosomes II and XVI occur at all times regardless of the *GAL* gene activity, it is proposed that these two chromosomes, II and XVI, and their corresponding "territories" are positioned adjacent to each other within the nucleus. Therefore the observed inter-chromosomal interactions might

demonstrate a pattern of chromosome positioning that evolves naturally. The interactions between the specific genes on these chromosomes might be due to the need of common transcription facilities, thereby allowing co-expression. If this were the case, an organization of the "chromosome territory" is not random. Instead it may have a functional implication, that is, allowing the genes that are co-regulated to locate in close proximity to each other. However, does gene expression determine its position within the genome or *vice versa*? Further studies are needed to improve our understanding of this issue.

Furthermore, if chromosomes II and XVI are located in a close proximity and share regulation facilities for their own genes' expression, it is possible that other genes of these two chromosomes also demonstrate such interactions. Since this project was focused on the interactions occurring at the *GAL* locus, the interactions between other gene loci on these two chromosomes have not been examined.

## 5.2 Functional Implications of Changes in *GAL* Gene Positioning

### 5.2.1 Exploration of S-shape Structure Formation at the *GAL* Locus

Presumably most DNA-DNA interactions are mediated by proteins or protein complexes. This also constitutes the principle of the cross-linking methodology. To determine the proteins that facilitate the S-shape structure formation at the *GAL* locus, two proteins, Gal80p and Rat1p, were chosen as candidate proteins due to their specific involvement in the regulation of *GAL* gene expression. Gal80p represses *GAL* gene expression in the absence of galactose, while Rat1p functions at the final stage of transcription, including mRNA capping and export. Rat1p was chosen because it is directly involved in transcription, thus any DNA structural changes observed with this protein might be considered as functionally critical.

The results showed that deletion/malfunction of Gal80p or Rat1p did not interrupt the key interactions contributing to the formation of an S-shape structure (Figure 24 and 25, Chapter 3). Gal80p forms a bipartite complex with Gal4p, which in turn recruits other regulatory factors on the upstream of the *GAL* structural genes (Figure 2, Chapter 1). If the S-shape structure was evolved naturally, it would be possible that these proteins/regulatory factors co-ordinate, rather than act alone, to mediate the structure formation. The interesting observation in these studies was that Figure 24B showed a loss of interaction in *gal80p*

strain under the galactose condition; this particular interaction was mediated by DNA looping (refer to Figure 22: interaction B, Chapter 3), which was potentially mediated by a group of proteins/protein complexes. Deletion of Gal80p was likely to affect the integrity of the protein group; thus resulting to the loss of this particular interaction. On the other hand, as all 3C PCR analyses were conducted qualitatively, it was also possible that there was insufficient DNA template for this particular reaction to be amplified and visualized. However, as every assay was repeated at least twice, the chance for this to happen repeatedly is very small.

As far as Rat1p is concerned particularly, there might be other proteins involved in the regulation at the transcription termination stage, either by substituting its function or forming a group with Rat1p. In the latter case, it is very difficult to understand the functional implication of a group of proteins by only focusing on a random member of it.

### 5.2.2 Exploration of the Inter-chromosomal Interactions at the *GAL* Locus

In terms of the inter-chromosomal interactions between chromosomes II and XVI, *i.e. SVL3-GAL7* and *HOS1-GAL1*, Gal80p and Rat1p were still selected as candidate proteins for the consistency in the experiments. Gal4p, the activator of the *GAL* genes, was also selected as a candidate protein. Gal4p is a critical regulator that activates *GAL* gene transcription by binding as dimers to $UAS_{gal}$ elements upstream of the *GAL* genes. Therefore it may have functional implications in facilitating inter-chromosomal interactions. However, the 3C PCR and 4C pulldown studies suggested that none of these proteins mediated these interactions (Figures 27, 29 and 30). Since Gal4p forms a transcription complex with Gal80p and a number of other transcription factors, *e.g.* SAGA complex, Mediator (Johnston, 1992; Thompson *et al.*, 1994; Dudley *et al.*, 1999), it is reasonable to suggest that any single protein of the complex might not play a significant role in determining the DNA structural formation.

Furthermore, in terms of the methodology approaches in these studies, Gal4p and Gal80p were studied via TAP-tagged protein isolation and deletion mutation, respectively. These are certain advantages and disadvantages in each approach. The advantage of using a deletion mutant strain over a TAP-tagged strain was that the former would not be compromised from a loss of the protein during antibody isolation process. However, the problem with the deletion mutant strain was that the loss of certain protein might interfere with the normal gene regulation, *e.g.*, in the current study, loss of Gal80p means loss of the repression of

*GAL* gene expression. Although the studies were carried out under all induction conditions, the *GAL* gene expression might not be regulated accurately. An addition experiment that studies the gene product of *GAL1, GAL7* or *GAL10* should have been included as a control. These experiments were carried out in the presence of glucose or glycerol/lactate thereby to confirm that the *GAL* genes are regulated according to the carbon sources.

## 5.3 Real-time PCR Studies on Interaction Frequency

### 5.3.1 Interaction Frequency for the *SVL3-GAL7* Interaction

To date no differences have been observed for the interactions occurring at the *GAL* locus under different induction conditions. In addition, no protein/protein complex has been identified to mediate either intra- or inter-chromosomal interactions. Attention was then drawn to an attempt to determine whether the frequency of particular interactions at the *GAL* locus varies in response to *GAL* gene expression, using real-time PCR. Due to time limitations, only one interaction, *SVL3-GAL7,* was studied because it was the first observed inter-chromosomal interaction.

Real-time PCR applications consisted of two aspects: 1) concentration of DNA template in each PCR reaction was firstly determined using *SyBr* green real-time PCR against an intragenic region at the *GAL* locus; 2) interaction frequency of *SVL3-GAL7* was studied using *Taqman*® real-time PCR that amplifies product across the interaction junction. The combination of these two types of real-time PCRs is a powerful tool to study the interaction frequency accurately. Three independent 3C assays were studied. It was shown in Figure 33 that although the interaction occurred under all induction conditions, a difference in the interaction frequency was observed. These interactions seemed to be the strongest in the presence of glucose and declined in the presence of glycerol/lactate and galactose. The results demonstrated a decline pattern in the interaction frequency when the *GAL* genes are activated. This is interesting since it implies that the *SVL3-GAL* interaction undergoes conformational changes during *GAL* gene activation process. For instance, high interaction frequency in the presence of glucose indicates that the repressed *GAL* genes introduce a "tightened" structure between two chromosomes, while low interaction frequency in the presence of glycerol/lactate and galactose indicates that the interaction "loosened". This potentially enables the access of transcriptional factors to the *GAL* genes allowing the onset of transcription. If this were the case, the result would be consistent with the loss of one of the intra-chromosomal interactions in a *gal80p* strain (Figure 24B: red arrow), which also

occurred under the galactose condition. It appears that when the *GAL* genes are activated, their relative positions with respect to the genome undergoes alteration; thus leading to the loss of an interaction that is mediated by Gal80p and other unknown proteins. The occurrence or loss of an interaction, or the changes in the interaction frequency, reflects a change in *GAL* gene positioning.

Despite the decline pattern illustrated by real-time PCR, however, the existence of large error bars suggested the results might not be as significant as it implied. As the real-time PCR studies were carried out on three assays, it is likely that they could not accurately reflect the actual situation. Further studies with large sample sizes are required to generate more meaningful conclusion with respect to the interaction frequency.

Nonetheless, it is necessary to bear in mind that the availability of the carbon source only has effects on the *GAL* gene, but not the *SVL* ORF. Hence any changes in the interaction frequency are considered to be the direct result of the *GAL* gene regulation. Overall, these interactions, either intra- or inter-chromosomal, reflect the inevitable, albeit non-random, juxtaposition of the regions next to each other within the genome.

## 5.4 Methodology

In this project, the core methods were 3C and 4C assays via PCR screening, DNA sequencing and BLAST analysis. In combination this was a powerful approach to detect the interacting DNA sequence, revealing the identity of any potential interactions. However, hundreds of PCR products obtained from the ligation of DNA fragments in any given 4C assay represented a major technical difficulty. PCR screening was labour intensive and it was reflected in the current study: only over one hundred clones for each detected sample were screened and a few of them were sequenced. The fact that each locus contacts not one or two, but tens of genomic regions, is likely to primarily be due to a cell-cell difference in DNA topology, with chromatin dynamics only accounting for a small proportion of interactions of a given cell (de Laat, 2007). It shows in this study that only a few DNA sequences were identified to be significant to the formation of the observed interactions. Therefore a great number of sequences need to be revealed in order to identify novel interactions at a specific gene locus. The shortcoming of manual screening is likely to be overcome by replacing DNA sequencing with a DNA microarray approach, in which a probe is used to hybridize the correct target sequences under high stringency conditions using a robotic system (Zhao *et al.*, 2006).

Furthermore, the problem with yeast cells in this project was that they were not synchronized. Synchronization is related to glycolysis, which is a step-by-step breakdown of glucose and the storing of the released energy in the form of ATP. The concentrations of the glycolytic intermediates in yeast cells are oscillated to synchronize yeast cells in a population (Bier *et al.*, 2000). Synchronization ensures that all yeast cells are grown to the same phase and undergo the same biochemical transformation. Yeast cells that were not synchronized might lead to misinterpretation of the results, *i.e.* DNA structural alteration is a result of variations displayed by yeast cells at different phases. It is necessary to include at least one group of synchronized yeast cells as a control in future studies.

## 5.5 Significance

This research project has contributed to the initiation of an increased understanding of how the position of a gene within the nucleus influences its activity by exploring DNA structural alterations under various expression conditions. For many years the gene carrier, chromatin, was considered merely as a structural device for packaging DNA. However it is clear that chromatin has a potential role in regulating gene expression, possibly by influencing the accessibility of transcription factors and cofactors to the *cis*-acting elements that control gene expression (Higgs *et al.*, 2007).

This research indicates an interesting connection between gene positioning and the regulation of gene activities, potentially by facilitating gene co-expression. This may have important implications for our understanding of disease processes. Accurate and stable expression of genes is essential for normal cellular survival and development. Diseased cells often result from mis-regulated genes and it is possible that this mis-regulation results from changes in gene positioning. If changes in gene positioning initiate or shut down a gene's activity, the re-positioning of certain DNA regions may permit the early diagnosis of diseases, *e.g.*, cancers. The results of this project constitute a definite step in the decoding of eukaryotic genomes by showing that genome structure has an important role in co-ordinating gene expression.

In addition, this project represents a useful addition to our current understanding of the *GAL* gene regulation system. Over the past decades, with the aid of advanced genetic and biochemical techniques, there have been a number of studies aimed at improving our understanding of the regulation of *GAL* gene expression. Rapid progress has placed this gene family at the forefront of model systems that are used to study gene regulation in eukaryotic

organisms. However, most of these studies have focused on function of gene products. The importance of gene positioning and their influence on gene expression regulation have just begun to be appreciated.

There remains a series of unsolved questions about the *GAL* gene system, *e.g.*, how does Gal3p interact with the Gal4p-Gal80p complex (Section 1.1)? Which transcription factors play a critical role to activate the *GAL* genes? It seems reasonable to seek answers to these questions from a different perspective, *i.e.* understand how gene positioning is associated with the *GAL* gene regulation system. In this project, the discovery of an S-shape structure at the *GAL* locus and the interaction between the *GAL* locus and two ORFs on chromosome XVI, made such an effort worthwhile. The findings will add extra value to our understanding of *GAL* gene regulation and has provided important data of use in discriminating between alternative hypotheses.

## 5.6 Summary

This project has modelled an S-shape structure formation at the *GAL* locus on chromosome II. Although this conformation, to date, has been regarded as non-functional with respect to *GAL* gene activities, it is clearly possible that it evolved to facilitate *GAL* gene expression. Also, the novel interactions between chromosomes II and XVI, is an exciting new finding that for the first time demonstrates a direct interaction between two yeast chromosomes. Similarly, these interactions might have evolved to facilitate the regulation of co-expressed genes on these chromosomes. Although no protein/protein complex has been confirmed to mediate any inter- or intra-chromosomal interactions in this study, real-time PCR might indicate that the *SVL3-GAL7* interaction exhibits a gradual decline in interaction frequency in response to the regulation of *GAL* gene expression.

## 5.7 Future studies

The combinational application of 3C and 4C is a powerful method to study DNA structural changes *in vivo*, although time consuming and qualitative. There are a few novel approaches that allow quantitative study of the interaction between DNA fragments, *e.g.*, replacement of DNA sequencing with microarray analyses, which use robotic applications to increase efficiency, in order to reveal unknown DNA fragments (Zhao *et al.*, 2006), or 3C-on-Chip (Chromatin Immunoprecipitation) that uses real-time PCR to study gene expression level in

association with the DNA structural changes (Simonis *et al.*, 2006). These methods are "more advanced" than the current 4C assay. It might be necessary to adapt to these new methods or combine them with the current 4C for an improved approach. In particular, for short-range interactions in *cis* or in *trans* configurations, the detection of a ligation product may reflect a random collision rather than a specific interaction. Interaction frequency needs to be quantified accurately to rule out random collisions; therefore it will be necessary to apply real-time PCR from the beginning. With respect to yeast cells used, it is necessary to synchronize the cells for future studies, or include one assay of synchronized yeast cells as a control.

In this study, unfortunately, no protein or protein complex was identified as playing a role in mediating any observed intra- and inter-chromosomal interactions at *GAL* locus. The problem is that there are numerous proteins at the *GAL* locus that could play a role in the regulation of *GAL* gene expression. Studies on randomly selected individual proteins seem to be a very limited effectiveness in the identification of critical proteins. More importantly, these proteins and protein complexes are positioned so closely that they might function co-ordinately to regulate DNA structural changes. The deletion or malfunction of one or two proteins might not be enough to fully disrupt the function. Future studies may need to explore the function of a group of proteins in relation to DNA structural change. Furthermore, in terms of the 4C pulldown method that targets a particular protein, it is also necessary to include a quantitative step in order to quantify the interaction frequency and thereby identify significant interactions over random collisions. An example is as described in the CHIP-combined loop assay where a real-time PCR application is included (Kumar, 2007).

The discovery of the close residency of chromosomes II and XVI, as a result of the interactions between the *GAL* locus on chromosome II and two ORFs on chromosome XVI, suggested that "chromosome territory" might have functional implications for gene regulation. It will be of interest to explore whether other genes on these two chromosomes interact in a co-expressing and/or co-functional manner. An immediate start could be an investigation of the reciprocal interactions between *GAL* locus and two ORFs on chromosome XVI, *i.e. SVL3-GAL1,* and *HOS1-GAL7.* Moreover, the observed chromosomal interaction could be further confirmed and visualized by the application of DNA FISH.

# Appendices:

## Appendix 1.0 Media Preparation

All media were prepared with Milli-Q water and autoclaved at 121°C for 15 minutes to sterilize.

**SC (synthetic complete) Media:**
--Bacto-yeast nitrogen base without amino acids 0.67%
--Glucose 2%
--Complete amino acid mix* 2%
MilliQ water

*Contains all possible supplements, nothing is "dropped out"

**SC-X (X refers to amino acid) Media:**
--Bacto-yeast nitrogen base without amino acids 0.67%
--Glucose 2%
--Drop-out mix* 2%
MilliQ water

*Contains all possible supplements except X

**SOC Media (Hanahan, 1983):**
--Bacto Tryptone 2%
--Bacto Yeast Extract 0.5%
--NaCl 2.5%
--KCl 0.25%
--$MgCl_2$ 1%
--$MgSO_4$ 1%
--Glucose 2%
MilliQ water

**Complex amino acid mix:** a combination of the following ingredients minus the appropriate supplement. It is mixed thoroughly with the help of clean marbles. In "Drop out" mix, X is left out.

| | | | |
|---|---|---|---|
| Adenine | 0.5g | Leucine | 10g |
| Arginine | 2g | Lysine | 2g |
| Aspartic acid | 2g | Methionine | 2g |
| Glycine | 2g | Phenylalanine | 2g |
| Histidine | 2g | Threonine | 2g |
| Inositol | 2g | Tryptophan | 2g |
| Isoleucine | 2g | Tyrosine | 2g |
| Uracil | 2g | | |

## Appendix 2.0 Common Buffers and Solutions

| **5x Ligase Buffer (Invitrogen specifications 100ml)** | **/100 ml** |
|---|---|
| 250 mM Tris | 25 ml |
| 50 mM MgCl2 | 5 ml |
| 5 mM DTT | 0.5 ml |
| 5 mM Ratp | 1.0 ml |
| 25% PEG 8000 | 25 g |
| Volume to 1 l with MilliQ $H_2O$ | |

| **LB Amp$^R$ X-Gal plate*** | **/1 l** |
|---|---|
| Bacto Tryptone | 10 g |
| Bacto Yeast Extract | 10 g |
| pH to 7.0 | |
| Agar | 12 g |
| Volume to 1l with MilliQ $H_2O$ Autoclave to sterilize | |
| Ampicillin | 1 ml |
| X-Gal | 1 ml |

* Makes 20-30 plates

| **LB Medium** | **/1 l** |
|---|---|
| Bacto Tryptone | 10 g |
| Bacto Yeast Extract | 5 g |
| NaCl | 10 g |
| pH7.5 with NaOH | |
| Volume to 1 l with MilliQ $H_2O$ | |

### Appendix 2.1 Buffers Used in 3C, 4C and 4C Pulldown

#### 3C & 4C Chromatin Extraction Buffers

| **FA-lysis buffer** | **/10 ml** | Chromatin Digestion Buffer | **/5ml** |
|---|---|---|---|
| 1 M HEPES-KOH | 500 µl | 1 M Tris-HCl (Ph 8.0) | 50 µl |
| 5 M NaCl | 280 µl | 1 M MgCl2 | 25 µl |
| 0.5 M EDTA | 20 µl | 11% Triton-X-100 | 45.5 µl |
| 11% Triton-X-100 | 900 µl | MilliQ $H_2O$ | 4.88 ml |
| 10% DOC | 100 µl | | |
| MilliQ $H_2O$ | 8.2 ml | | |
| 1x Protease Inhibitor tablet per 10 ml | | | |

#### 4C Pulldown Chromatin Extraction Buffers

| **IP-like Buffer** | **/2 ml** | **IP Buffer** | **/50 ml** |
|---|---|---|---|
| 0.5 M EDTA | 19.2 µl | 1 M Tris-HCl (Ph 8.0) | 835 µl |
| 5 M NaCl | 267.2 µl | 0.5 M EDTA | 120 µl |
| 11% Triton-X-100 | 880 µl | 5 M NaCl | 1.67 ml |
| 10% SDS | 8 µl | 11% Triton-X-100 | 5 ml |

| | | | |
|---|---|---|---|
| MilliQ $H_2O$ | 825.6 µl | 10% SDS | 50 µl |
| 250 mM PMSF | 16 µl | MilliQ $H_2O$ | 42.187 ml |
| 5 mg/ml Pepstain | 1.28 µl | 250 mM PMSF | 100 µl |
| 1 mg/ml Leupeptin | 4.8 µl | 5 mg/ml Pepstain | 8 µl |
| | | 1 mg/ml Leupeptin | 30 µl |

**4C Pulldown Washing Buffers (Alekseyenko et al. 2006)**

| **Buffer A** | **/50 ml** | **Buffer B** | **/50 ml** |
|---|---|---|---|
| 1 M Tris-HCl (pH 8.0) | 2.5 ml | 1 M Tris-HCl (pH 8.0) | 2.5 ml |
| 0.5 M EDTA | 200 ul | 0.5 M EDTA | 200 µl |
| 5 M NaCl | 1.5 ml | 5 M NaCl | 8 ml |
| 10% DOC | 2.5 ml | 10% NP40 | 5 ml |
| 10% SDS | 500 µl | MilliQ $H_2O$ | 34.3 ml |
| 10% NP40 | 5 ml | | |
| MilliQ $H_2O$ | 37.7 ml | | |
| 250 mM PMSF | 100 µl | | |
| 5 mg/ml Pepstain | 8 µl | | |
| 1 mg/ml Leupeptin | 30 µl | | |

| **Buffer C** | **/50 ml** | **Buffer D** | **/50 ml** |
|---|---|---|---|
| 1 M Tris-HCl (pH 8.0) | 2.5 ml | 1 M Tris-HCl (pH 8.0) | 5 ml |
| 0.5 M EDTA | 200 µl | 0.5 M EDTA | 200 µl |
| 5 M NaCl | 5 ml | 1 M LiCl | 25 ml |
| 10% SDS | 500 µl | 10% NP40 | 5 ml |
| 10% NP40 | 5 ml | 10% DOC | 5 ml |
| MilliQ $H_2O$ | 36.8 ml | MilliQ $H_2O$ | 9.8 ml |

| **Buffer E** | **/50 ml** |
|---|---|
| 1 M Tris-HCl (pH 8.0) | 500 µl |
| 0.5 M EDTA | 50 µl |
| 5 M NaCl | 1.5 ml |
| 1 M DTT | 100 µl |
| Tween20 | 50 µl |
| MilliQ $H_2O$ | 47.8 ml |

# Appendix 3.0 Primers

Primers for amplifying 2kb product between *GAL7-GAL10* and *GAL10-GAL1* and 3C primers:

| Primer name | Sequence | Description |
|---|---|---|
| *Csp*61 Outside-Fwd | TGATCGGGGTACCATCTCTG | Forward primer between *GAL10-1* region to amplify ~2 kilobase fragment |
| *Csp*61 Outside-Rev | GCCGAAGAAGACAATCCACT | Reverse primer between *GAL10-1* region to amplify ~2 kilobase fragment |
| *Csp*61-7876f | Aaagagccccattatcttagcc | Forward primer between *GAL10* and *GAL1* promoter region |
| *Csp*61-8401f | Gcgaagcgatgatttttgat | Forward primer on *GAL1* promoter region |
| *Csp*61-7651f | Cagcaccacctgtaaccaaa | Forward primer on *GAL10* promoter region |
| GAL7-10 Fwd | CGAAACAATTGCCTCTCACA | Forward primer between *GAL7-10* region to amplify ~2kilobase fragment |
| GAL7-10 Rev | AGAAGGACACCGAATTTCCA | Reverse primer between *GAL7-10* region to amplify ~2kilobase fragment |
| *Csp*61 5776 f | ttggacccgtaagtttcacc | Forward primer on *GAL10* terminator region |
| *Csp*61 6001 f | gcattttcatccaccacaaa | Forward primer on *GAL10* terminator region |
| *Csp*61 4351 f | Cgtaatcggcaaacaaatca | Forward primer on *GAL7* promoter region |
| *Csp*61 5401 f | Caaaaggagaggcaaagcaa | Forward primer on *GAL7* promoter region |

4C primers:

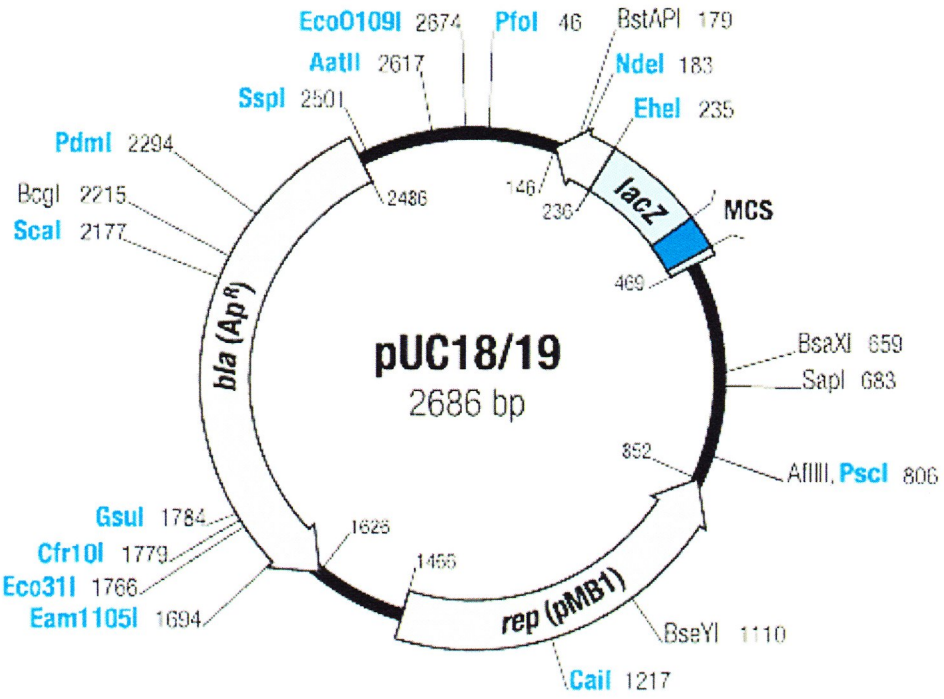
| Primer name | Sequence | Description |
|---|---|---|
| GAL10-1 7651-7876 F | Acgcttaactgctcattgct | Forward primer on *GAL 10* promoter region |
| GAL10-1 7651-7876 R | Aaacttctttgcgtccatcc | Reverse primer on *GAL 10* promoter region |
| Nested GAL10-1 7651-7876 F | ggaattcgcttaactgctcattgctatattga | Nested forward primer on *GAL 10* promoter region |
| Nested GAL10-1 7651-7876 R | Ggaattctctttgcgtccatccaaaa | Nested reverse primer on *GAL 10* promoter region |
| GAL10-1 7876-8401F | Tgactaaatctcattcagaagaagtg | Forward primer on *GAL 1* promoter region |
| GAL10-1 7876-8401R | Gaccggtgaagacgagga | Reverse primer on *GAL 1* promoter region |
| Nested GAL10-1 7876-8401F | Ggaattcatctcattcagaagaagtgatt | Nested forward primer on *GAL 1* promoter region |
| Nested GAL10-1 7876-8401R | Ggaattcaccggtgaagacgaggac | Nested reverse primer on *GAL 1* promoter region |
| GAL7-10 4351-5401 F | Gcggctcgtgctatattctt | Forward primer on *GAL 7* promoter region |
| GAL7-10 4351-5401 R | Tcataagcctttcaaatatgtcca | Reverse primer on *GAL 7* promoter region |
| Nested GAL7-10 4351-5401 F | Ggaattctgctaccgtccatatctttcc | Nested forward primer on *GAL 7* promoter region |
| Nested GAL7-10 4351-5401 R | Ggaattctaagcctttcaaatatgtcca | Nested reverse primer on *GAL 7* promoter region |
| GAL7-10 5401-5776 F | Gcttcgtaaccagcagacaa | Forward primer on *GAL 1* terminator region |
| GAL7-10 5401-5776 R | Tgtcacagcgaatttcctca | Reverse primer on *GAL 1* terminator region |
| Nested GAL7-10 5401-5776 F | Ggaattcccagcagacaagaaatcacc | Nested forward primer on *GAL 1* terminator region |
| Nested GAL7-10 5401-5776 R | Ggaattctcacagcgaatttcctcaca | Nested reverse primer on *GAL 1* terminator region |

*GAL10* internal primers:

| Primer name | Sequence | Description |
|---|---|---|
| Frag 7651 internal F | tggcgtatttcgtatgacca | Forward primer within fragment 7651 |
| Frag 7651 internal R | ccaagcatcacattcccttc | Reverse primer within fragment 7651 |

## Appendix 4.0 pUC18 Vector

Contains M13 forward and reverse priming sites, multiple cloning sites (MCS), CAP protein binding site, *lac* repressor binding site, Ampicillin antibiotic resistance. mRNA (LacZ) starts at position 507



## Appendix 5.0 Preparation of Bacterial Cells for Electroporation

Evening before:

Set up overnight culture of bacteria (*e.g., DH5α*) in 100 ml LMM. Autoclave several flasks of 2x YT broth (1 litre per flask). Aim to process at least 2 flasks. Also autoclave 3 litres of 10% glycerol water (381 g of glycerol dissolved in 3 litres of distilled water). Wash 6 large and 4 small centrifuge containers with screw caps in 0.1% SDS then rinse in distilled water and autoclave.

Next day:

Innoculate 10 ml of fresh o/n culture into each litre of 2X YT and set shaking at 37°C. Save a small aliquot of uninoculated media as a control for measuring the $OD_{600}$ later. Pre-chill 10% glycerol water in cold room. Measure $OD_{600}$ three hours later. When $OD_{600}$ reaches 0.7-0.8, Chill cells on ice for 20 min. Meanwhile pre-chill the rotors GSA and GS3.

Pour cells into 6 large and 4 small containers and spin simultaneously at 5000 x *g* for 5 min. Discant supernatant and resuspend cells in cold 10% glycerol water. Combine resuspended cells from the small containers into the 6 large containers, then spin these 6 in the GS3 rotor at 5000 x *g* for 5 min (aim to fill these to 3/4 of their capacity with 10% glycerol water). Resuspend cells in approx 45 ml of 10% glycerol water per large container and transfer to 6 sterile 50 ml tubes and centrifuge down at 3700 x *g* for 10 min. Decant supernatant and resuspend in 50 ml of glycerol water and centrifuge down. Decant supernatant and resuspend each in approx. 15 ml of glycerol water. Combine cells into two 50 ml tubes and centrifuge down at 3700 x *g* for 10 min. Decant supernatant and resuspend cells in 10% glycerol water to give a final volume of about 5 - 7 ml, depending on yield. Aliquot cells into 1.5 ml centrifuge tubes with 50µl each.

It is important to keep cells, rotors and containers well chilled at all stages of the process.

# Appendix 6.0 Amplification of the ~2kb Regions for the Production of Positive Controls in 3C Analysis

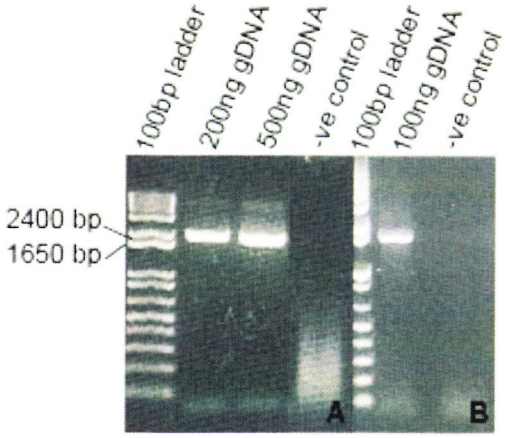



**Figure i:** Amplification of the ~2kb regions for the production of positive controls in 3C assay. These 2kb regions are between (A) *GAL10-GAL1* and (B) *GAL7-GAL10* gene loci. 200 ng and 500 ng of W303-1a genomic DNA were used as the template to amplify the region between *GAL10-GAL1*, while 100 ng genomic DNA was used to amplify the region between *GAL10-GAL7*. 5 µl of PCR products were loaded on 0.7% agarose gel, which was run at 100 volts for 30 min and then photographed.

## Appendix 7.0 Primer testing in 4C Analysis

An example of 4C primer testing is shown. Primer pair 7876-7651 was tested against W303-1a *Csp*61 cut-ligated sample with the PCR conditions remained optimum across a wide range of annealing temperatures.

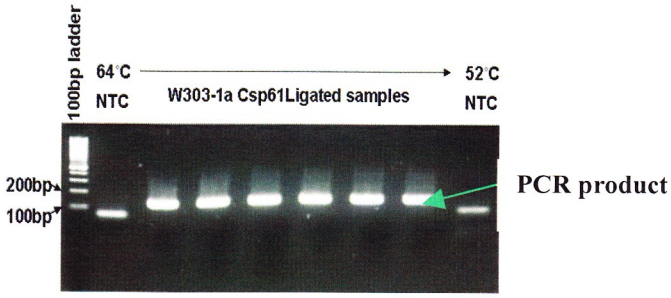



**Figure ii**: 4C primer pair validation and optimization. W303-1a DNA was digested with *Csp*61, ligated and amplified using a temperature gradient (52°C-64°C) PCR on 3C primer pair 7876-7651. The expected product size was 106 bp, which is indicated by the green arrow. Negative controls (NTCs), in which water was used as template, were included at 52°C and 64°C. 20 µl of PCR product were loaded on a 1.5% agarose gel, which was run at 80 volts for 40 min and then photographed.

## Appendix 8.0 Cross-Linking Time Determination

In both of the 3C and 4C analyses, a 2 minute cross-linking time was recommended (Dekker *et al.*, 2002; O'Sullivan *et al.*, 2004; Zhao *et al.*, 2006) to crosslink DNA and proteins. However, it was possible that 2 minutes might be insufficient for the whole genome cross-linking, thus a series of experiments were conducted to determine the most appropriate cross-linking time.

These experiments were conducted following the 4C protocol (Section 2.3.7), including non-cross-linking, cross-linking for 2 min, 10 min, 30 min, 2 hr, 4 hr, and overnight. Since the changes in DNA structure are most likely to occur when genes are active, the cross-linking time experiments were conducted in the presence of galactose. This experiment was repeated twice.

4C nested PCR analyses showed that an identical pattern was generated at all cross-linking times when examined the cut-ligated samples (Figure iii). The non-cross-linking samples showed either a pattern similar to that of the cross-linked samples or no pattern. It is clear that 2 min and 10 min cross-linking time resulted into the best pattern in PCR (Figure iii: lanes 2 and 3) in terms of banding pattern and density; this deteriorated as the cross-linking time increased (Figure iii: lanes 4, 5, 6, and 7). This pattern was consistent with every primer pair that was tested (Figure iiiA and B; additional data not shown). It therefore appeared that 2-10 min is the most appropriated cross-linking time. As all previous 3C and 4C studies were conducted using a 2 min cross-linking time, it was sensible to continue using this protocol in subsequent experiments.

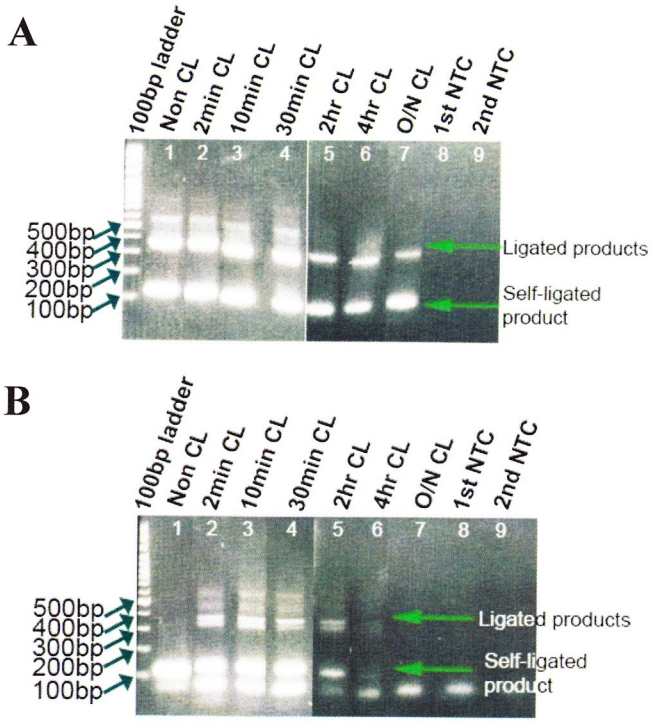
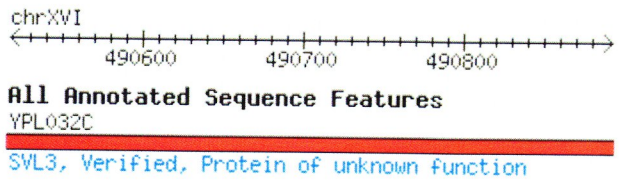



**Figure iii:** 4C studies of the cross-linking time experiment. W303-1a cells were cross-linked in the presence of galactose for the indicated time. Another set of cells, not cross-linked in the presence of galactose, was included as a control for the cross-linked cells. CL denotes *Csp*61 cut and ligated sample. Negative controls (NTCs) were included in both rounds of nested PCRs. PCR amplification using primer pairs (A) 7876-7651 and (B) 7876-8401. Green arrows indicate ligated products and self-ligated products.20 µl of PCR reactions were loaded on a 1.5% agarose gel, which was run at 80 volts for 40 min and then photographed.

# Appendix 9.0 BLAST Results

## Appendix 9.1 BLAST Analysis Showed a Matching Sequence to the *SVL3* ORF

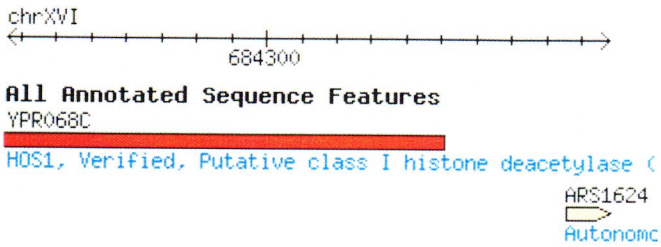



Score = 286.8 bits (1871), Expect = 2.3e-79, P = 2.3e-79 [ Retrieve Sequence / ORF Map / Genome Browser ]

Identities = 375/376 (99%), Frame = -1 / +1

```
Query: 376     CTTCCATAGAGATAAAAGAGAGTTTTCATTATCATGACTACTGTTTAGTCTGGCTCCCAT 317
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 490517 CTTCCATAGAGATAAAAGAGAGTTTTCATTATCATGACTACTGTTTAGTCTGGCTCCCAT 490576

Query: 316     TGTTTTTGCTACAGTGATGATTTCAGTTACCAGACCTGATATTAAAGGTTTGGCAATGAT 257
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 490577 TGTTTTTGCTACAGTGATGATTTCAGTTACCAGACCTGATATTAAAGGTTTGGCAATGAT 490636

Query: 256     CTGCTGGTCTAGATCACTTGGGTTTTCTTGTTCAAACATGATTAACAACGGATCAAAACA 197
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 490637 CTGCTGGTCTAGATCACTTGGGTTTTCTTGTTCAAACATGATTAACAACGGATCAAAACA 490696

Query: 196     AATTCTGGAAATCGCCAGCTTCCATTGTTGTGATAGAAACTCAATTGATGAGAAATTGCG 137
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 490697 AATTCTGGAAATCGCCAGCTTCCATTGTTGTGATAGAAACTCAATTGATGAGAAATTGCA 490756

Query: 136     CAAATTGATTTTTATGTTGGAAAACAATTTAGCAAACAGTTTTTCAAAAGTGGTTAATAA 77
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 490757 CAAATTGATTTTTATGTTGGAAAACAATTTAGCAAACAGTTTTTCAAAAGTGGTTAATAA 490816

Query: 76      TGTTATAACGCCTGATGAATATTTCTCAGTGCTGGACTTACTTTCGCCGAGGTAAATCGT 17
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 490817 TGTTATAACGCCTGATGAATATTTCTCAGTGCTGGACTTACTTTCGCCGAGGTAAATCGT 490876

Query: 16      GTTCTCCTTTGCAGTA 1
               ||||||||||||||||
Sbjct: 490877 GTTCTCCTTTGCAGTA 490892
```

## Appendix 9.2 BLAST Analysis Showed a Matching Sequence to the *HOS1* ORF



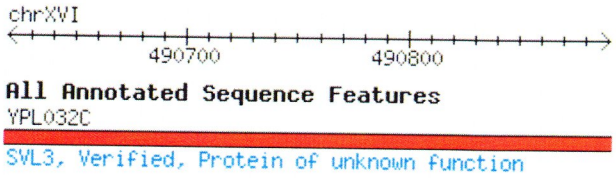

Score = 124.3 bits (788), Expect = 1.9e-30, P = 1.9e-30 [ Retrieve Sequence / ORF Map / Genome Browser ]

Identities = 164/172 (95%), Frame = -1 / +1

```
Query: 172     TTCAGAAGAAGTGATTGCGCGTTTATTAGGGAATATGTCAGCTGCGATTTTTGATGATTG 113
               | | | |||||     | ||||||||||||||||||||||||||||||||||||||||||
Sbjct: 684226 TGCTGGAGAAGATCGTACGCGTTTATTAGGGAATATGTCAGCTGCGATTTTTGATGATTG 684285
```

```
Query: 112     TTACATGGGAGTAAATCTGCTACTTGAGACTGAAATATTGACGTTGATATGACCAATTTC 53
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 684286 TTACATGGGAGTAAATCTGCTACTTGAGACTGAAATATTGACGTTGATATGACCAATTTC 684345

Query: 52      GACATAGCGTTTTTCTAAAATGGACAGGTGTTTATTAATTCATATTACAGTA 1
               ||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 684346 GACATAGCGTTTTTCTAAAATGGACAGGTGTTTATTAATTCATATTACAGTA 684397
```

## Appendix 9.3 BLAST Analysis Showed a Matching Sequence to the *SVL3* ORF and Sequence of Chromosome II

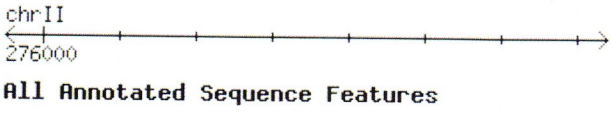



Score = 209.5 bits (1356), Expect = 4.2e-56, P = 4.2e-56 [ Retrieve Sequence / ORF Map / Genome Browser ]

Identities = 271/272 (99%), Frame = +1 / +1

```
Query: 4       TAAAGGTTTGGCAATGATCTGCTGGTCTAGATCACTTGGGYTTTCTTGTTCAAACATGAT 63
               ||||||||||||||||||||||||||||||||||||+||||||||||||||||||||
Sbjct: 490619 TAAAGGTTTGGCAATGATCTGCTGGTCTAGATCACTTGGGTTTTCTTGTTCAAACATGAT 490678

Query: 64      TAACAACGGATCAAAACAAATTCTGGAAATCGCCAGCTTCCATTGTTGTGATAGAAACTC 123
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 490679 TAACAACGGATCAAAACAAATTCTGGAAATCGCCAGCTTCCATTGTTGTGATAGAAACTC 490738

Query: 124     AATTGATGAGAAATTGCACAAATTGATTTTTATGTTGGAAAACAATTTAGCAAACAGTTT 183
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 490739 AATTGATGAGAAATTGCACAAATTGATTTTTATGTTGGAAAACAATTTAGCAAACAGTTT 490798

Query: 184     TTCAAAAGTGGTTAATAATGTTATAACGCCTGATGAATATTTCTCAGTGCTGGACTTACT 243
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 490799 TTCAAAAGTGGTTAATAATGTTATAACGCCTGATGAATATTTCTCAGTGCTGGACTTACT 490858

Query: 244     TTCGCCGAGGTAAATCGTGTTCTCCTTTGCAG 275
               ||||||||||||||||||||||||||||||||
Sbjct: 490859 TTCGCCGAGGTAAATCGTGTTCTCCTTTGCAG 490890
```

&




Score = 64.7 bits (391), Expect = 1.6e-12, P = 1.6e-12 [ Retrieve Sequence / ORF Map / Genome Browser ]

Identities = 78/79 (98%), Frame = -1 / +1

```
Query: 79      CGCGGCTCGTGCTATATTCTTGTTGCTACCGTCCATATCTTTCCATAGATTTTCAATTTT 20
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 275995 CGCGGCTCGTGCTATATTCTTGTTGCTACCGTCCATATCTTTCCATAGATTTTCAATTTT 276054

Query: 19      TGATGTCTCCATGGTGGTM 1
               ||||||||||||||||||+
Sbjct: 276055 TGATGTCTCCATGGTGGTA 276073
```

# Appendix 10.0 Real-time PCR Standards and Data

## Appendix 10.1 *SyBr* Green and *Taqman*® Standards

### *SyBr* green Real-time PCR standards

W303-1a genomiC DNA digested by *Msp*1

| Dilution fold | DNA quantity |
|---|---|
| 1/10 | 3 ng |
| 1/40 | 0.75 ng |
| 1/160 | 0.185 ng |
| 1/640 | 0.0469 ng |
| 1/1280 | 0.0234 ng |
| 1/2560 | 0.012 ng |
| 1/5120 | 0.006 ng |
| 1/10240 | 0.003 ng |

### *Taqman*® Real-time PCR standards

*SVL3-GAL7* 3C purified product

| Dilution fold | DNA quantity |
|---|---|
| 1/10 | 2 ng |
| 1/100 | 0.2 ng |
| 1/1000 | 0.02 ng |
| 1/10000 | 0.002 ng |
| 1/100000 | 0.0002 ng |
| 1/1000000 | 0.00002 ng |
| 1/10000000 | 0.000002 ng |
| 1/100000000 | 0.0000002 ng |

## Appendix 10.2 *SyBr* Green PCR Data

| Sample Name | Detector | Task | Ct | StdDev Ct | Qty | Mean Qty | StdDev Qty | Tm | Correct for mean | Correct for 1 |
|---|---|---|---|---|---|---|---|---|---|---|
| Glu CL Assay 1 | CHD | Unknown | 17.1 | 2.111 | 5.266 | 7.303 | 5.537 | 78.3 | 7.303 | 1.16116664 |
| Glu CL Assay 1 | CHD | Unknown | 14.12 | 2.111 | 9.34 | 7.303 | 5.537 | 78.3 | | |
| Gly CL Assay 1 | CHD | Unknown | 15.1 | 0.415 | 9.51 | 11.81 | 3.257 | 78.5 | 11.81 | 1 |
| Gly CL Assay 1 | CHD | Unknown | 14.51 | 0.415 | 14.12 | 11.81 | 3.257 | 78.5 | | |
| Gal CL Assay 1 | CHD | Unknown | 15.23 | 0.673 | 5.01 | 6.743 | 2.64 | 78.9 | 6.743 | 1.25760047 |
| Gal CL Assay 1 | CHD | Unknown | 14.28 | 0.673 | 8.47 | 6.743 | 2.64 | 78.9 | | |
| Glu CL Assay 2 | CHD | Unknown | 14.48 | 0. 564 | 8.67 | 7.493 | 2.09 | 78.6 | 7.49 | 1.13217623 |
| Glu CL Assay 2 | CHD | Unknown | 14.48 | 0.564 | 6.31 | 7.493 | 2.09 | 78.6 | | |
| Gly CL Assay 2 | CHD | Unknown | 14.49 | 0.291 | 7.42 | 8.48 | 1.499 | 78.9 | 8.48 | 1 |
| Gly CL Assay 2 | CHD | Unknown | 14.08 | 0.291 | 9.54 | 8.48 | 1.499 | 78.9 | | |
| Gal CL Assay 2 | CHD | Unknown | 14.74 | 0.476 | 6.38 | 8.01 | 2.3 | 78.6 | 8.01 | 1.05867665 |
| Gal CL Assay 2 | CHD | Unknown | 14.07 | 0.476 | 9.64 | 8.01 | 2.3 | 78.6 | | |
| Glu CL Assay 3 | CHD | Unknown | 19.49 | 3.522 | 0.496 | 7.33 | 9.66 | 78.5 | 7.33 | 1.6111869 |
| Glu CL Assay 3 | CHD | Unknown | 14.51 | 3.522 | 14.16 | 7.33 | 9.66 | 78.5 | | |
| Gly CL Assay 3 | CHD | Unknown | 16.94 | 1.449 | 2.76 | 6.85 | 5.787 | 78.5 | 6.85 | 1.72408759 |
| Gly CL Assay 3 | CHD | Unknown | 14.89 | 1.449 | 10.94 | 6.85 | 5.787 | 78.5 | | |
| Gal CL Assay 3 | CHD | Unknown | 17.68 | 1.426 | 1.68 | 4.1 | 3.427 | 78.5 | 4.1 | 2.8804878 |
| Gal CL Assay 3 | CHD | Unknown | 15.66 | 1.426 | 6.52 | 4.1 | 3.427 | 78.5 | | |

## Appendix 10.3 *Taqman*® PCR Dissociation Curve

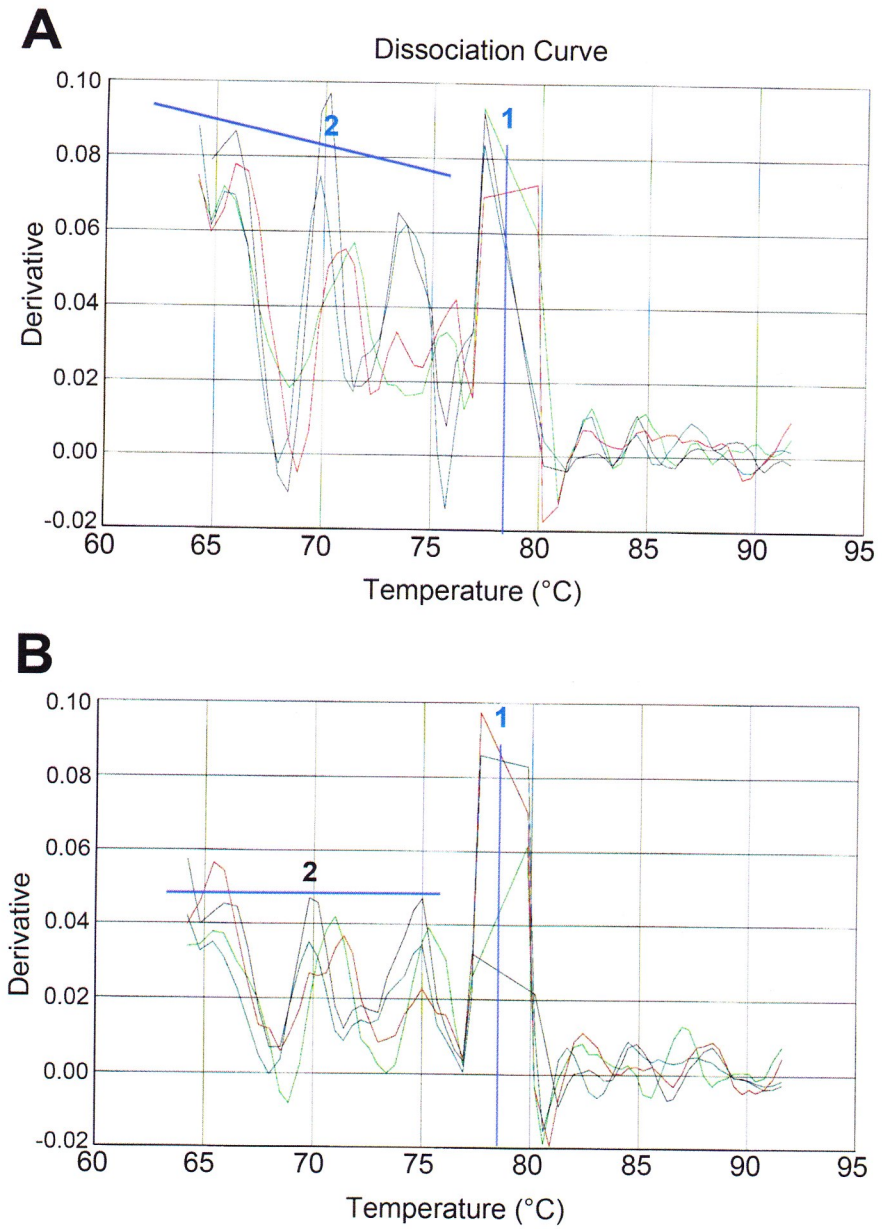



**Figure iv:** Dissociation curve analyses. These analyses confirm that there was a larger peak for secondary products at a primer concentration of (A) 1000 nM than (B) 200 nM. A dissociation curve analysis was performed after a completed PCR. Data were obtained by slowly ramping the temperature of reaction solutions from 60 to 95°C while continuously collecting fluorescence data. The increase in temperature causes PCR products to undergo denaturation, a process accompanied by a decrease in fluorescence levels for solutions containing *SyBr* Green I dye. The melting temperature peak that was generated from a specific product is demonstrated by (1), which equals 78.5°C. The melting peaks (2) for secondary products, including primer dimers, contaminating DNA, *etc.*, are much higher in (A) than (B).

## Appendix 10.4 *Taqman*® PCR Data

| Sample Name | Detector | Task | Ct | StdDev Ct | Qty | Mean Qty | StdDev Qty | Correct for mean | Correct for % |
|---|---|---|---|---|---|---|---|---|---|
| Glu CL Assay 1 | Taqman MGB set up | Unknown | 25.19 | 0.211 | 1.23E-06 | 1.23E-06 | 1.66E-07 | 1.2333E-06 | 1.99E-06 |
| Glu CL Assay 1 | Taqman MGB set up | Unknown | 25.41 | 0.211 | 1.07E-06 | 1.23E-06 | 1.66E-07 | | |
| Glu CL Assay 1 | Taqman MGB set up | Unknown | 24.99 | 0.211 | 1.4E-06 | 1.23E-06 | 1.66E-07 | | |
| Gly CL Assay 1 | Taqman MGB set up | Unknown | 29.23 | 0.368 | 9.32E-08 | 1.24E-07 | 2.71E-08 | 1.244E-07 | 3.4E-07 |
| Gly CL Assay 1 | Taqman MGB set up | Unknown | 28.62 | 0.368 | 1.38E-07 | 1.24E-07 | 2.71E-08 | | |
| Gly CL Assay 1 | Taqman MGB set up | Unknown | 28.57 | 0.368 | 1.42E-07 | 1.24E-07 | 2.71E-08 | | |
| Gal CL Assay 1 | Taqman MGB set up | Unknown | 27.38 | 1.002 | 3.05E-07 | 1.83E-07 | 1.12E-07 | 1.829E-07 | 5.27E-07 |
| Gal CL Assay 1 | Taqman MGB set up | Unknown | 29.38 | 1.002 | 8.47E-08 | 1.83E-07 | 1.12E-07 | | |
| Gal CL Assay 1 | Taqman MGB set up | Unknown | 28.4 | 1.002 | 1.59E-07 | 1.83E-07 | 1.12E-07 | | |
| Glu CL Assay 2 | Taqman MGB set up | Unknown | 28.24 | 0.257 | 1.75E-07 | 1.86E-07 | 3.14E-08 | 1.8567E-07 | 2.07E-07 |
| Glu CL Assay 2 | Taqman MGB set up | Unknown | 28.38 | 0.257 | 1.61E-07 | 1.86E-07 | 3.14E-08 | | |
| Glu CL Assay 2 | Taqman MGB set up | Unknown | 27.88 | 0.257 | 2.21E-07 | 1.86E-07 | 3.14E-08 | | |
| Gly CL Assay 2 | Taqman MGB set up | Unknown | 27.33 | 0.572 | 9.6E-08 | 1.37E-07 | 6.11E-08 | 1.3667E-07 | 1.37E-07 |
| Gly CL Assay 2 | Taqman MGB set up | Unknown | 26.27 | 0.572 | 2.07E-07 | 1.37E-07 | 6.11E-08 | | |
| Gly CL Assay 2 | Taqman MGB set up | Unknown | 27.18 | 0.572 | 1.07E-07 | 1.37E-07 | 6.11E-08 | | |
| Gal CL Assay 2 | Taqman MGB set up | Unknown | 26.96 | 0.637 | 1.25E-07 | 8.79E-08 | 3.77E-08 | 8.7767E-08 | 1.1E-07 |
| Gal CL Assay 2 | Taqman MGB set up | Unknown | 28.22 | 0.637 | 5.01E-08 | 8.79E-08 | 3.77E-08 | | |
| Gal CL Assay 2 | Taqman MGB set up | Unknown | 27.44 | 0.637 | 8.82E-08 | 8.79E-08 | 3.77E-08 | | |
| Glu CL Assay 3 | Taqman MGB set up | Unknown | 30.62 | 0.478 | 5.33E-08 | 7E-08 | 2.1E-08 | 7.0033E-08 | 7.93E-08 |
| Glu CL Assay 3 | Taqman MGB set up | Unknown | 30.33 | 0.478 | 6.32E-08 | 7E-08 | 2.1E-08 | | |
| Glu CL Assay 3 | Taqman MGB set up | Unknown | 29.68 | 0.478 | 9.36E-08 | 7E-08 | 2.1E-08 | | |
| Gly CL Assay 3 | Taqman MGB set up | Unknown | 30.2 | 0.346 | 6.84E-08 | 7.48E-08 | 1.63E-08 | 7.4833E-08 | 7.48E-08 |
| Gly CL Assay 3 | Taqman MGB set up | Unknown | 29.69 | 0.346 | 9.34E-08 | 7.48E-08 | 1.63E-08 | | |
| Gly CL Assay 3 | Taqman MGB set up | Unknown | 30.35 | 0.346 | 6.27E-08 | 7.48E-08 | 1.63E-08 | | |
| Gal CL Assay 3 | Taqman MGB set up | Unknown | 30.62 | 0.428 | 5.32E-08 | 6.18E-08 | 1.69E-08 | 6.1833E-08 | 6.49E-08 |
| Gal CL Assay 3 | Taqman MGB set up | Unknown | 29.92 | 0.428 | 8.13E-08 | 6.18E-08 | 1.69E-08 | | |
| Gal CL Assay 3 | Taqman MGB set up | Unknown | 30.69 | 0.428 | 5.1E-08 | 6.18E-08 | 1.69E-08 | | |

## Appendix 10.5 *Taqman*® PCR Data Calculation

| | Assay 1 | Assay 2 | Assay 3 | S. D.[a] | S. E.[b] | SUM | Medium |
|---|---|---|---|---|---|---|---|
| **Glu CL** | -6.6845 | -7.1008 | -5.7021 | 0.71822 | 0.41466 | -19.487 | -6.4958 |
| **Gly CL** | -6.8643 | -7.1259 | -6.4947 | 0.31713 | 0.18309 | -20.485 | -6.8283 |
| **Gal CL** | -6.9571 | -7.1876 | -6.2784 | 0.47266 | 0.27289 | -20.423 | -6.8077 |

[a]Standard Deviation, [b]Standard Error

## References:


Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J. (1990). Basic Local Alignment Search Tool. *Journal of Molecular Biology* 215: P. 403-410.

Amberg, D. C., Goldstein, A. L., Cole, C. N. (1992). Isolation and Characterization of RAT1: An Essential Gene of *Saccharomyces cerevisiae* Required For the Efficient Nucleocytoplasmic Trafficking of mRNA. *Genes Developments* 6(7): P. 1173-1189.

Ansari, A., Hampsey, M. (2005). A Role For The Cpf 3'-End Processing Machinery In Rnap Ii-Dependent Gene Looping. *Genes & Development* 19(24): P. 2969-2978.

Bajwa, W. T., Hopper, J. E. (1988). Yeast Regulatory Gene Gal3 - Carbon Regulation -Uasgal Elements In Common With Gal1, Gal2, Gal7, Gal10, Gal80, And Mel1 - Encoded Protein Strikingly Similar To Yeast And *Escherichia-Coli Galactokina. Molecular And Cellular Biology* 8(8): P. 3439-3447.

Banerji, J., Rusconi, S, Schaffner, W. (1981). Expression Of A Beta-Globin Gene Is Enhanced By Remote Sv40 Dna-Sequences. *Cell* 27(2): P. 299-308.

Barberis, A., Pearlberg, J, Simkovich, N, Farrell, S, Reinagel, P, Bamdad, C, Sigal, G, Ptashne, M. (1995). Contact With A Component Of The Polymerase-Ii Holoenzyme Suffices For Gene Activation. *Cell* 81(3): P. 359-368.

Bash, R., Lohr, D. (2001). Yeast Chromatin Structure And Regulation Of Gal Gene Expression. *Progress In Nucleic Acid Research And Molecular Biology* 65: P. 197-259.

Bhat, P., Hopper, J. E. (1992). Overproduction Of Gal1 Or Gal3 Protein Causes Galactose-Independent Activation Of The Gal4 Protein: Evidence For A New Model For Induction For The Yeast *Gal/Mel* Regulation. *Molecular And Cellular Biology* 12: P. 2701-2707.

Bhat, P., Oh, D., Hopper, J. E. (1990). Analysis Of The Gal3 Signal Transduction Pathway Activating Gal4 Protein-Dependent Transcription In *Saccharomyces Cerevisiae. Genetics* 125(2): P. 281-291.

Bhat, P., Murthy, T. V. S. (2001). Transcriptional Control Of The Gal/Mel Regulon Of Yeast *Saccharomyces Cerevisiae:* Mechanism Of Galactose-Mediated Signal Transduction. *Molecular Microbiology* 40(5): P. 1059-1066.

Bier, M., Bakker, BM., Westerhoff, HV. (2000). How Yeast Cells Synchronize their Glycolytic Oscillations: A Perturbation Analytic Treatment. *Biophysical Journal* 78: P. 1087-1093.

Biggar, S.C., G. (1999). Continuous And Widespread Roles For The Swi-Snf Complex In Transcription. *Embo Journey* 18: P. 2254-2264.

Bonifer, C. (1999). Long-Distance Chromatin Mechanisms Controlling Tissue-Specific Gene Locus Activation. *Genetics* 238(2): P. 277-289.

Boutanaev, A., Kalmykova, A. I., Shevelyou, Y. Y., Nurminsky, D. I. (2002). Large Clusters Of Co-Expressed Genes In The Drosophila Genome. *Nature* 420(6916): P. 666-669.

Brown, K., Baxter, J., Graf, D., Merkenschlager, M., Fisher, A. G. (1999). Dynamic Repositioning Of Genes In The Nucleus Of Lymphocytes Preparing For Cell Division. *Molecular Cell* 3(2): P. 207-217.

Bryant, G., Ptashne, M. (2003). Independent Recruitment *In Vivo* By Gal4 Of Two Complexes Required For Transcription. *Molecular Cell* 11(5): P. 1301-1309.

Bulger, M., Groudine, M. (1999). Looping Versus Linking: Toward A Model For Long-Distance Gene Activation. *Genes & Development* 13(19): P. 2465-2477.

Bystricky, K., Laroche, T, Van Houwe, G, Blaszczyk, M, Gasser, S. M. (2005). Chromosome Looping In Yeast: Telomere Pairing And Coordinated Movement Reflect Anchoring Efficiency And Territorial Organization. *Journal Of Cell Biology* 168(3): P. 375-387.

Caron, H., Van Schaik, B, Van Der, Mee M, Baas, F, Riggins, G, Van Sluis, P, Hermus, Mc, Van Asperen, R, Boon, K, Voute, Pa, Heisterkamp, S, Van Kampen, A, Versteeg, R. (2001). The Human Transcriptome Map: Clustering Of Highly Expressed Genes In Chromosomal Domains. *Science* 291(5507): P. 1289-1292.

Casolari, J., Brown, Cr, Komili, S, West, J, Hieronymus, H, Silver, P. A. (2004). Genome-Wide Localization Of The Nuclear Transport Machinery Couples Transcriptional Status And Nuclear Organization. *Cell* 117(4): P. 427-439.

Cavalli, G. (2007). Chromosome Kissing. *Current Opinion in Genetics & Development* 17: P. 443-450.

Chakalova, L., Debrand, E., Mitchell, JA, Osborne, CS, Fraser, P. (2005). Replication and transcription: Shaping the landscape of the genome. *Nature Reviews Genetics* 6: P. 669-677.

Cheng, J., Gandolfi, M., Ptashne, M. (2004). Activation Of The Gall Gene Of Yeast By Pairs Of 'Non-Classical' Activators. *Current Biology* 14(18): P. 1675-1679.

Cho, R., Campbell, M. J., Winzeler, E. A., Steinmetz, L, Conway, A, Wodicka, L, Wolfsberg, T. G., Gabrielian, A. E., Landsman, D, Lockhart, D. J., Davis, R. W. (1998). A Genome-Wide Transcriptional Analysis Of The Mitotic Cell Cycle. *Molecular Cell* 2(1): P. 65-73.

Cohen, B., Mitra, R. d., Hughes, J. D., Church, G. M. (2000). A Computational Analysis Of Whole-Genome Expression Data Reveals Chromosomal Domains Of Gene Expression. *Nature Genetics* 26(2): P. 183-186.

Cook, P. (1995). A Chromomeric Model For Nuclear And Chromosome Structure. *Journal Of Cell Science* 108: P. 2927-2935.

Cook, P. (1999). Molecular Biology - The Organization Of Replication And Transcription. *Science* 284(5421): P. 1790-1795.

Cremer, T., Kurz, A, Zirbel, R, Dietzel, S, Rinke, B, Schrock, E, Speicher, Mr, Mathieu, U, Jauch, A, Emmerich, P, Schertan, H, Reid, T, Cremer, C, Lichter, P. (1993). Role Of Chromosome Territories In The Functional Compartmentalization Of The Cell Nucleus. *Cold Spring Harb Symp Quant Biology* 58: P. 777-792.

Croft, J. A., Bridger, J. M., Boyle, S., Perry, P., Teague, P., Bickmore, W. A. (1999). Differences In The Localization And Morphology Of Chromosomes In The Human Nucleus. *Journal Of Cellular Biology* 145: P. 1119-1131.

de Laat, W., Grosveld, F. (2003). Spatial Organization Of Gene Expression: The Active Chromatin Hub. *Chromosome Research* 11(5): P. 447-459.

de Laat, W. (2007). Long-range DNA contacts: romance in the nucleus. *Current Opinion in Cell Biology* 19: P. 317-320.

Dekker, J., Ripple, K., Dekker, M., Kleckner, N. (2002). Capturing Chromosome Conformation. *Science* 295: P. 1306-1311.

Douglas, H.C., and Pelroy, G. (1963). A gene controlling inducibility of the galactose pathway enzymes in *Saccharomyces. Biochim. Biophys. Acta* 68: P. 155–156.

Drissen, R., Palstra, Rj, Gillemans, N, Splinter, E, Grosveld, F, Philipsen, S, de Laat, W. (2004). The Active Spatial Organization Of The Beta-Globin Locus Requires The Transcription Factor Eklf. *Genes & Development* 18(20): P. 2485-2490.

Duboule, D., Dolle, P. (1989). The Structural And Functional-Organization Of The Murine Hox Gene Family Resembles That Of *Drosophila* Homeotic Genes. *Embo Journal* 8(5): P. 1497-1505.

Dudley, A. M., Rougeulle, C., And Winston, F. (1999). The Spt Components Of Saga Facilitate Tbp Binding To A Promoter At A Post-Activator-Binding Step *In Vivo. Genes & Development* 13: P. 2940-2945.

Dunn, T., Hahn, S., Ogden, S., Schleif, R. F. (1984). An Operator At -280 Base-Pairs That Is Required For Repression Of Arabad Operon Promoter - Addition Of Dna Helical Turns Between The Operator And Promoter Cyclically Hinders Repression. *Proceedings Of The National Academy Of Sciences Of The United States Of America-Biological Sciences* 81(16): P. 5017-5020.

Englesbe, E., Squires, C, Meronk, F. (1969). L-Arabinose Operon In Escherichia Coli B/R - A Genetic Demonstration Of Two Functional States Of Product Of A Regulator Gene. *Proceedings Of The National Academy Of Sciences Of The United States Of America* 62(4): P. 1100-107.

Ferreira, J., Paolella, G., Ramos, C., Lamond, A. I. (1997). Spatial Organization Of Large-Scale Chromatin Domains In The Nucleus: A Magnified View Of Single Chromosome Territories. *Journal Of Cell Biology* 139(7): P. 1597-1610.

Feuerbach, F., Galy, V., Trelles-Sticken, E., Fromont-Racine, M., Jacquier, A., Gilson, E., Olivo-Marin, J. C., Scherthan, H., Nehrbass, U. (2002). Nuclear Architecture And Spatial Positioning Help Establish Transcriptional States Of Telomeres In Yeast. *Nature Cell Biology* 4(3): P. 214-221.

Fuss, S. H., Omura, M., Mombaerts, P. (2007). Local and cis effects of the H element on expression of odorant receptor genes in mouse. *Cell* 130: P. 373-384.

Gasch, A. P., Spellman, P. T., Kao, C. M., Carmel-Harel, O., Eisen, M. B., Storz, G., Botstein, D., Brown, P. O. (2000). Genomic expression programs in the response of yeast cells to environmental changes. *Molecular Biology of Cell* 11(12): P. 4241-4257.

Giniger, E. P., M. (1988). Cooperative DNA-Binding Of The Yeast Transcriptional Activator Gal4. *Proceedings Of The National Academy Of Sciences Of The United States Of America* 85(2): P. 382-386.

Grant, P. A., Sterner, D. E., Duggan, L. J., Workman, J. L., Berger, S. L. (1998). The Saga Unfolds: Convergence Of Transcription Regulators In Chromatin-Modifying Complexes. *Trends Cell Biology* 8: P. 193-197.

Griggs, D. W., Johnston, M. (1993). Promoter Elements Determining Weak Expression Of The Gal4 Regulatory Gene Of *Saccharomyces cerevisiae. Molecular And Cellular Biology* 13(8): P. 4999-5009.

Hanahan, D. (1983). Studies on Transformation of *Escherichia coli* with Plasmids. *Journey of Molecular Biology* 166(4): 557-580.

Hawkins, A. R. (1987). The Complex Arom Locus Of *Aspergillus Nidulans.* Evidence For Multiple Genes Fusions And Convergent Evolution. *Current Genetics* 11: P. 491-498.

Herendeen, D., Kassavetis, G. A., Geiduschek, E. P. (1992). A Transcriptional Enhancer Whose Function Imposes A Requirement That Proteins Track Along DNA. *Science* 256(5061): P. 1298-1303.

Higgs, D. R., Vernimmen, D., Hughes, J., Gibbons, R. (2007). Using genomics to study how chromatin influences gene expression. *Annual Review of Genomics Human Genetics* 8: P. 299-325.

Hirst, M., Kobor, M. S., Kuriakose, N., Greenblatt, J., Sadowski, I. (1999). Gal4 Is Regulated By The Rna Polymerase Ii Holoenzyme-Associated Cyclin-Dependent Protein Kinase Srb10/Cdk8. *Molecular Cell* 3(5): P. 673-678.

Hittinger, C.T., Rokas, A., Carroll, S. B. (2004). Parallel Inactivation Of Multiple Gal Pathway Genes And Ecological Diversification In Yeasts. *Proceedings Of The National Academy Of Sciences Of The United States Of America* 101(39): P. 14144-14149.

Hurst, L., Pal, C., Lercher, M. J. (2004). The Evolutionary Dynamics Of Eukaryotic Gene Order. *Nature Reviews Genetics* 5(4): P. 299-310.

Iborra, F., Pombo, A, Jackson, D. A., Cook, P. R. (1996). Active Rna Polymerases Are Localized Within Discrete Transcription 'Factories' In Human Nuclei. *Journal Of Cell Science* 109: P. 1427-1436.

Jackson, D., Iborra, F. J., Manders, E., Cook, P. R. (1998). Numbers And Organization Of Rna Polymerases, Nascent Transcripts, And Transcription Units In Hela Nuclei. *Molecular Biology Of The Cell* 9(9): P. 1523-1536.

Johnston, M., Davis, R. W. (1984). Sequences that regulate the divergent *GAL1-GAL10* promoter in *Saccharomyces cerevisiae. Molecular Cellular Biology* 4: P. 1521-1527.

Johnston, M. (1987). A Model Fungal Gene Regulatory Mechanism - The *Gal* Genes Of *Saccharomyces-Cerevisiae. Microbiological Reviews* 51(4): P. 458-476.

Johnson, A. W. (1997). Rat1p and Xrn1p are functionally interchangeable exoribonucleases that are restricted to and required in the nucleus and cytoplasm, respectively. *Molecular Cellular Biology* 17(10): P. 6122-6130.

Johnston, M., Carlson, M. (1992). Regulation Of Carbon And Phosphate Utilization. *Gene expression* 2: P. 193-281.

Kim, M., Krogan, N. J., Vasiljeva, L., Rando, O. J., Nedea, E., Greenblatt, J. F., Buratowski, S. (2004). The yeast Rat1 exonuclease promotes transcription termination by RNA polymerase II. *Nature* 432: P. 517-522.

Kumar, P. P. (2007). Functional interaction between PML and SATB1 regulates chromatin-loop architecture and transcription of the MHC class I locus. *Natural Cellular Biology* 9: P. 45-56.

Lakshminarasimhan A, B.P. (2005). Replacement Of A Conserved Tyrosine By Tryptophan In Gal3p Of *Saccharomyces Cerevisiae* Reduces Constitutive Activity: Implications For Signal Transduction In The *Gal* Regulation. *Molecular Genetics And Genomics* 274(4): P. 384-393.

Larschan, E. W., F. (2001). The *S-Cerevisiae* Saga Complex Functions *In Vivo* As A Coactivator For Transcriptional Activation By Gal4. *Genes & Development* 15(15): P. 1946-1956.

Lee, J., Sonnhammer, E. (2003). Genomic Gene Clustering Analysis Of Pathways In Eukaryotes. *Genome Research* 13(5): P. 875-882.

Lercher, M., Urrutia, A. O., Hurst, L. D. (2002) Clustering Of Housekeeping Genes Provides A Unified Model Of Gene Order In The Human Genome. *Nature Genetics* 31(2): P. 180-183.

Ling, J., Li, T., Hu, J. F., Vu, T. H., Chen, H. l., Qiu, X. W., Cherry, A. M., Hoffman, A. R. (2006). Ctcf Mediates Interchromosomal Colocalization Between *Igf2/H19* And *Wsb1/Nf1*. *Science* 312(5771): P. 269-272.

Lomvardas, S., Barnea, G., Pisapia, D. J., Mendelsohn, M., Kirkland, J., Axel, R. (2006). Interchromosornal Interactions and Olfactory Receptor Choice. *Cell* 126(2): P. 403-413.

Mahy, N., Perry, P. E., Gilchrist, S., Baldock, R. A., Bickmore, W. A. (2002). Spatial Organization Of Active And Inactive Genes And Noncoding Dna Within Chromosome Territories. *Journal Of Cell Biology* 157(4): P. 579-589.

Matthews, K. S. (1992). DNA Looping. *Microbiological Reviews* 56(1): P. 123-136.

Meyer, J. W., J.A. Hollenberg, C. P. (1990). Gal1 Of *Kluyveromyces Lactis* Is Required For The Induction Of The Lelior Enzymes And Can Complement The Gal3 Long Term Adaptation Phenotype In *Saccharomyces Cerevisiae*. *Yeast* 6: P. 591-596.

Mizzen, C., Kuo, M. H., Smith, E., Brownell, J., Zhou, J., Ohba, R., Wei, Y., Monaco, L., Sassone-Corsi, P., Allis, C. D. (1998). Signaling to chromatin through histone modifications: how clear is the signal? *Cold Spring Harb Symp Quant Biology* 63: P. 469-481.

Mueller, H.P.S., W. (1990). Transcriptional Enhancers Can Act *In Trans. Trends Genetics* 6: P. 300-304.

Mylin, L. M., Bhat, J. P., Hopper, J. E. (1989). Regulated Phosphorylation And Dephosphorylation Of Gal4, A Transcriptional Activator. *Genes Development* 3(8): P. 1157-1165.

Mylin, L. M., Johnston, M., Hopper, J. E. (1990). Phosphorylated Forms Of Gal4 Are Correlated With Ability To Activate Transcription. *Molecular And Cellular Biology* 10(9): P. 4623-4629.

O'Sullivan, J., Tan-Wong, S. M., Morillon, A., Lee, B., Coles, J., Mellor, J., Proudfoot, N. J. (2004). Gene Loops Juxtapose Promoters And Terminators In Yeast. *Nature Genetics* 36(9): P. 1014-1018.

Osborne, C., Chakalova, L., Brown, K. E., Carter, D., Horton, A., Debrand, E., Goyenechea, B., Mitchell, J. A., Lopes, S., Reik, W., Fraser, P. (2004). Active Genes Dynamically Colocalize To Shared Sites Of Ongoing Transcription. *Nature Genetics* 36(10): P. 1065-1071.

Ostergaard, S. O., Johnston, M., Nielsen, J. (2000). Increasing Galactose Consumption By *Saccharomyces Cerevisiae* Through Metabolic Engineering Of The Gal Gene Regulatory Network. *Nature* 18: P. 1283 - 1286.

Palstra, R. J., Tolhuis, B., Splinter, E., Nijmeijer, R., Grosveld, F., de Laat, W. (2003). The Beta-globin Nuclear Compartment in Development and Erythroid Differentiation. *Nature Genetics* 35(2): P. 190-194.

Peng, G., Hopper, J. E. (2000). Evidence For Ga13p's Cytoplasmic Location And Ga180p's Dual Cytoplasmic-Nuclear Location Implicates New Mechanisms For Controlling Ga14p Activity In *Saccharomyces Cerevisiae. Molecular And Cellular Biology* 20(14): P. 5140-5148.

Peng, G., Hopper, J. E. (2002). Gene Activation By Interaction Of An Inhibitor With A Cytoplasmic Signaling Protein. *Proceedings Of The National Academy Of Sciences Of The United States Of America* 99(13): P. 8548-8553.

Petrascheck, M., Escher, D., Mahmoudi, T., Verrijzer, C. P., Schaffner, W., Barberis, A. (2005). DNA Looping Induced By A Transcriptional Enhancer *In Vivo. Nucleic Acids*

*Research* 33(12): P. 3743-3750.

Platt, A., Reece, R. J. (1998). The Yeast Galactose Genetic Switch Is Mediated By The Formation Of A *Gal4p-Gal80p-Gal3p* Complex. *Embo Journal* 17(14): P. 4086-4091.

Proffitt, J. H. (1985). DNase I-hypersensitive sites in the galactose gene cluster of *Saccharomyces cerevisiae. Molecular Cellular Biology* 5(6): P. 1522-1524.

Ptashne, M. (1986). Gene-Regulation By Proteins Acting Nearby And At A Distance. *Nature* 322(6081): P. 697-701.

Ptashne, M. G. (1997). Transcriptional Activation By Recruitment. *Nature* 386(6625): P. 569-577.

Rippe, K., Vonhippel, P. H., Langowski, J. (1995). Action at a Distance-DNA Looping and initiation of Transcription. *Trends in Biochemical Sciences* 20(12): P. 500-506.

Roy, P., Stuart, J. M., Lund, J., Kim, S. K. (2002). Chromosomal Clustering Of Muscle-Expressed Genes In *Caenorhabditis Elegans. Nature* 418(6901): P. 975-979.

Rozen, S., Skaletsky, H. (2000). Primer3 on the WWW for general users and for biologist programmers. *Methods Molecular Biology* 132: P. 365-386.

Rundlett, S. E., Carmen, A. A., Kobayashi, R., Bavykin, S., Turner, B. M., Grunstein, M. (1996). HDA1 and RPD3 are members of distinct yeast histone deacetylase complexes that regulate silencing and transcription. *Proceedings Of The National Academy Of Sciences Of The United States Of America* 93(25): P. 14503-14508.

Sadowski, I., Costa, C., Dhanawansa, R. (1996). Phosphorylation Of Gal4p At A Single C-Terminal Residue Is Necessary For Galactose-Inducible Transcription. *Molecular And Cellular Biology* 16(9): P. 4879-4887.

Schleif, R. (1992). DNA Looping. *Annual Review Of Biochemistry* 61: P. 199-223.

Shimada, H. F., T. (1985). Controlled Transcription Of The Yeast Regulatory Gene *Gal80. Gene* 39(1): P. 1-9.

Simonis, M., Klous, P., Splinter, E., Moshkin, Y., Willemsen, R., de Wit, E., van Steensel, B., de laat, W. (2006). Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). *Natural Genetics* 38: P. 1348-1354.

Simonis, M., Kooren, J., de Laat, W. (2007). An evaluation of 3C-based methods to capture DNA interactions. *Nature Methods* 4(11): P. 895-901.

Song, J., Jensen, R. A., Phh, R. (1996). A Divergently Transcribed Activator Of The Phenylalanine Hydroxylase Gene Cluster Of *Pseudomonas Aeruginosa. Molecular Microbiology* 22(3): P. 497-507.

Spellman, P. T., Rernet, G. M. (2002). Evidence For Large Domains Of Similarly Expressed Genes In The *Drosophila Genome. Journal Of Biology* 1(1): P. 5-10.

Spilianakis, C., Lalioti, M. D., Town, T., Lee, G. R., Flavell, R. A. (2005). Interchromosomal Associations Between Alternatively Expressed Loci. *Nature* 435(7042): P. 637-645.

St. John, T. P., Davis, R. W. (1979a). Isolation of Galactose-inducible DNA sequence from *Saccharomyces cerevisiae* by differential plaque filter hybridization. *Cell* (1). 16: P. 443-452.

St. John, T. P., Davis, R. W. (1979b).The Organization and Transcription of the galactose gene Cluster of *Saccharomyces. Journey of Molecular Biology* 152: P. 285-315.

Thompson, C. M., Koleske, A. J., Chao, D. M., Young, R. A. (1994). A Multisubunit Complex-Associated With The Rna Polymerase-Ii Ctd And Tata-Binding Protein In Yeast. *Cell* 73(7): P. 1361-1375.

Tolhuis, B., Palstra, R. J., Splinter, E., Grosveld, F., de Laat, W. (2002). Looping And Interaction Between Hypersensitive Sites In The Active Beta-Globin Locus. *Molecular Cell* 10(6): P. 1453-1465.

Tsuyumu, S., Adams, B. G. (1974). Dilution Kinetic Studies Of Yeast Populations – *In vivo* Aggregation Of Galactose Utilizing Enzymes And Positive Regulator Molecules. *Genetics* 77(3): P. 491-505.

Verschure, P. J., Van Der, K. I., Manders, E. M. M., Van Driel, R. (1999). Spatial Relationship Between Transcription Sites And Chromosome Territories. *Journal Of Cell Biology* 147(1): P. 13-24.

Vilar, J., Saiz, L. (2005). Dna Looping In Gene Regulation: From The Assembly Of Macromolecular Complexes To The Control Of Transcriptional Noise. *Current Opinion In Genetics & Development* 15(2): P. 136-144.

Volpi, E., Chevret, E., Jones, T., Vatcheva, R., Williamson, J., Beck, S., Campbell, R. D., Goldsworthy, M., Powis, S. H., Ragoussis, J., Trowsdale, J., Sheer, D. (2000). Large-Scale Chromatin Organization Of The Major Histocompatibility Complex And Other Regions Of Human Chromosome 6 And Its Response To Interferon In Interphase Nuclei. *Journal Of Cell Science* 113(9): P. 1565-1576.

Wolffe, A., Guschin, D. (2000). Chromatin Structural Features And Targets That Regulate Transcription. *Journal Of Structural Biology* 129(2-3): P. 102-122.

Woodcock, C., Dimitrov, S. (2001). Higher-Order Structure Of Chromatin And Chromosomes. *Current Opinion In Genetics & Development* 11(2): P. 130-135.

Zhao, Z., Tavoosidana, G., Sjolinder, M., Gondor, A., Mariano, P., Wang, S., Kanduri, C., Lezcano, M., Sandhu, K.S., Singh, U., Pant, V., Tiwari, V., Kurukuti, S., Ohlsson, R. (2006). Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. *Nature Genetics* 38(11): P. 1341-1347.

Zheng, B., Wu, J. N., Schober, W., Lewis, D. E., Vida, T. (1998). Isolation of yeast mutants defective for localization of vacuolar vital dyes. *Proceedings Of The National Academy Of Sciences Of The United States Of America* 95(20): P. 11721-11726.

Zink, D., Amaral, M. D., Englmann, A., Lang, S., Clarke, L. A., Rudolph, C., Alt, F., Luther, K., Braz, C., Sadoni, N., Rosenecker, J., Schindelhauer, D. (2004). Transcription-Dependent Spatial Arrangements Of Cftr And Adjacent Genes In Human Cell Nuclei. *Journal Of Cell Biology* 166(6): P. 815-825.

Zirbel, R. M., Mathieu, U. R., Kurz, A., Cremer, T., Lichter, P. (1993). Evidence For A Nuclear Compartment Of Transcription And Splicing Located At Chromosome Domain Boundaries. *Chromosome Research* 1: P. 93-106.