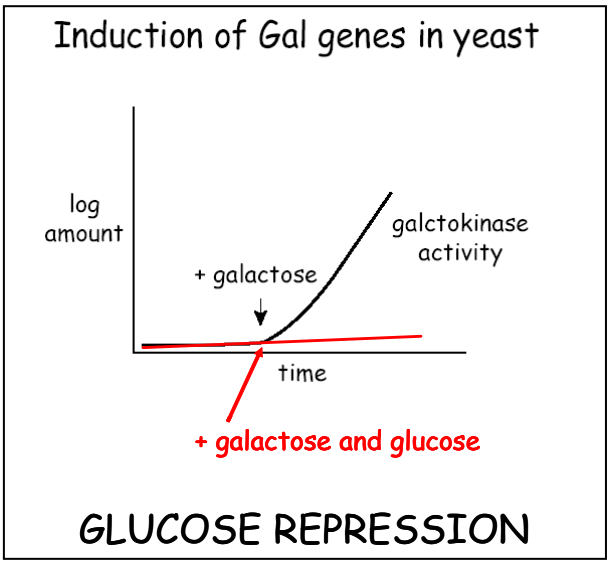
Lecture 21
Eukaryotic Genes and Genomes III

Cis-acting sequences

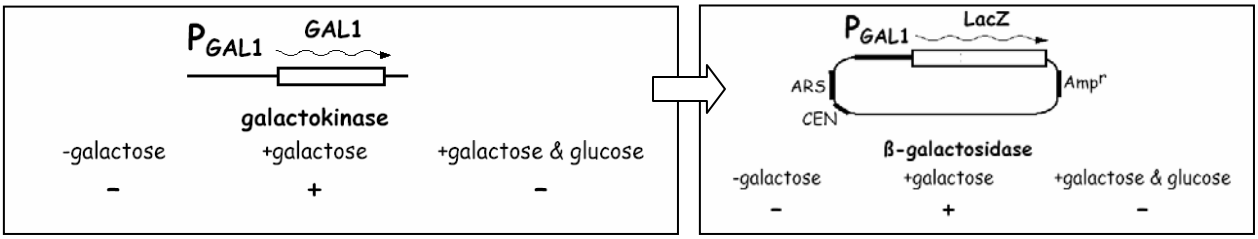
In the last lecture we considered a classic case of how genetic analysis could be used to dissect a regulatory mechanism. This analysis was contingent upon having "clean" phenotypes associated with the isolated mutants; e.g., mutations in the **Gal80** gene produce a phenotype of constitutive **Gal1** expression. However, it is sometimes very difficult to identify regulatory proteins by isolating mutants, because regulators that influence the expression of a wide variety of genes might be essential (i.e., mutations in these could be lethal), or their mutant phenotypes may be extremely complex and difficult to interpret.

One solution to this has been to work backwards from the cis-acting promoter sequences for particular genes to identifying the proteins that bind to them. Let's take the **Gal1** gene as an example. We have considered the fact that in the presence of galactose the **Gal1** gene is transcriptionally upregulated (along with other **Gal** genes). What I haven't told you is the fact that if glucose is present in addition to galactose, the induction of the **Gal** genes simply does not occur! This is known as **glucose repression**. This makes physiological sense because glucose is a more efficient energy source for yeast, and is therefore the preferred carbon source over galactose. Why bother metabolizing galactose as long as glucose is present? In fact, glucose represses a very large number of genes whose products metabolize a wide range of carbon sources (sucrose, maltose, galactose etc) that are less energy efficient than glucose, as well as repressing a whole host of other genes.



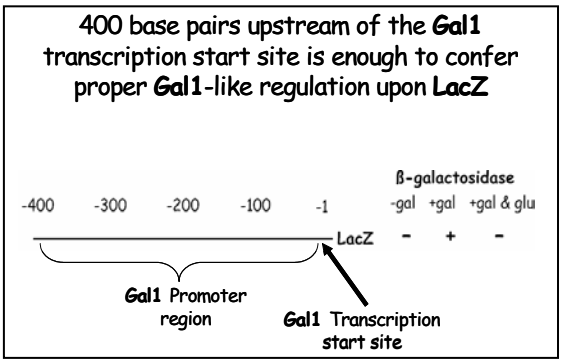
It seems reasonable to expect that there is a transcriptional repressor that responds to glucose levels; this repressor would be ineffective when glucose is low or absent, and effective when glucose is present. It also seems reasonable that one could isolate trans-acting mutants that fail to repress galactose-induced **Gal** gene expression in the presence of glucose. However, it turns out that the very fact that glucose represses such a large number of different genes made it difficult to identify such mutants.

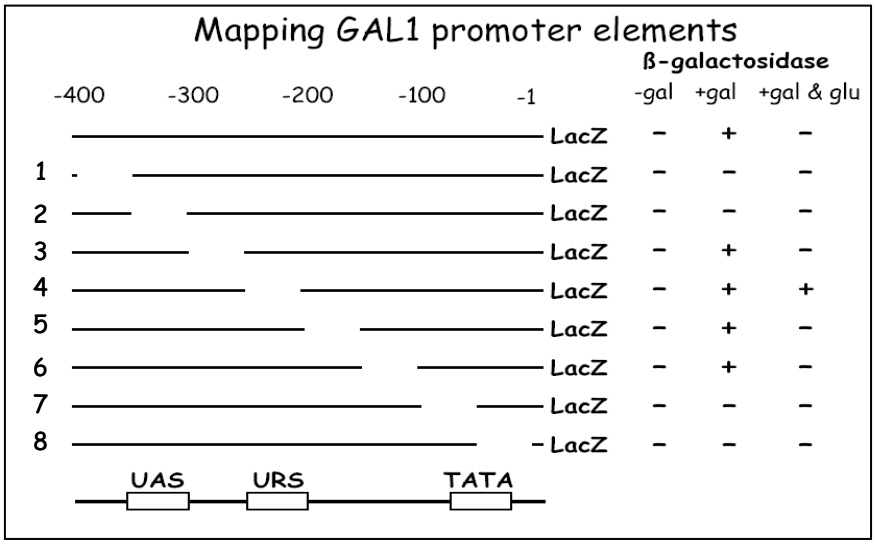
Instead of looking for mutants that fail to execute glucose repression at the **Gal1** gene, studies of the **Gal1** promoter region itself provided the key to dissecting the mechanism of glucose repression. Specifically, the **Gal1** promoter region was fused to the *E. coli* **LacZ** gene, on a plasmid that can



replicate autonomously in *S. cerevisiae*. It was first important to establish that regulation of **LacZ** (β-galactosidase) from the plasmid mirrored the regulation of **Gal1** (galactokinase) from its chromosomal locus; i.e., that β–galactosidase was induced by galactose in the absence of glucose, but not in its presence.

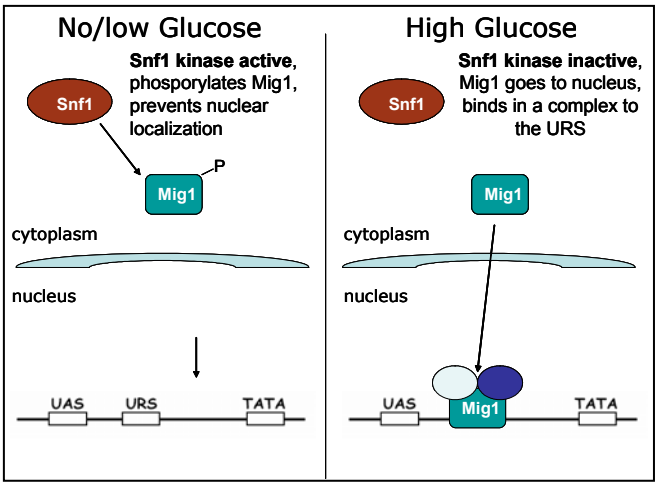
Having established that, it was possible to go on and interrogate subdomains of the **Gal1** promoter region for their role in induction of **Gal1** by galactose, as well as repression of **Gal1** by glucose. The minimal length of DNA stretching upstream into the promoter region from the **Gal1** transcription start site (designated as adjacent to -1) was 400bp DNA. Once this functional promoter region was delineated, systematic deletions of 50bp or so could be made all across the 400 bp region; this is easy to do with some recombinant DNA tricks that are not important to know about here. Suffice to say that this "**deletion analysis**" revealed two regions critical for transcriptional control, as well as the location of the TATA sequence that is required for loading of the basal transcription machinery.





The expression of β–galactosidase from each of these promoter deletion constructs under minus-galactose, plus-galactose, and plus galactose & glucose, are show. From these data we can deduce the location of cis-acting regulatory sequences for the **Gal1** gene.

- Deletions 7 and 8 do not express the reporter under any conditions because the deletions have removed some of the **TATA** sequence that is required for assembly of the basal transcription machinery.
- Deletions 1 and 2 eliminate the ability of galactose to increase expression from the **Gal1** promoter, and since expression is not induced there is nothing for glucose to repress. It turns out that the 75bp sequence between -310 and -385 is the DNA binding site for **Gal4** and this kind of region is generally called a **UAS** (upstream activation sequence) and in this case **UAS_{GAL}**. We will come back to thinking about **Gal4** binding to the **UAS** recognition sequence later.
- Deletions 3, 5 and 6 have no effect on the ability of galactose to induce expression because the **UAS** remains intact. Note that shortening the distance between the **UAS** and the **TATA** region is not detrimental to induction. Indeed increasing the distance by inserting extra DNA between the **UAS** and the **TATA** sequence also has little effect on inducibility. This has led to the idea that **UAS** sequences can work at long distances (1,000 – 10,000 bp) away from the **TATA** sequence and the transcription start sites. (In mammalian cells regions containing binding sites for transcriptional activators are called **enhancers**; we will come to these in a later lecture)
- Deletion 4 turns out to reveal information about **glucose repression.** For this construct, while galactose induces expression, glucose is unable to repress that expression. The deleted region defines the position of a sequence element needed for glucose repression, and a sequence element that behaves this way (i.e., are required for repression) is generally called a **URS** (upstream repressor sequence), and in this case **URS_{GAL}.**

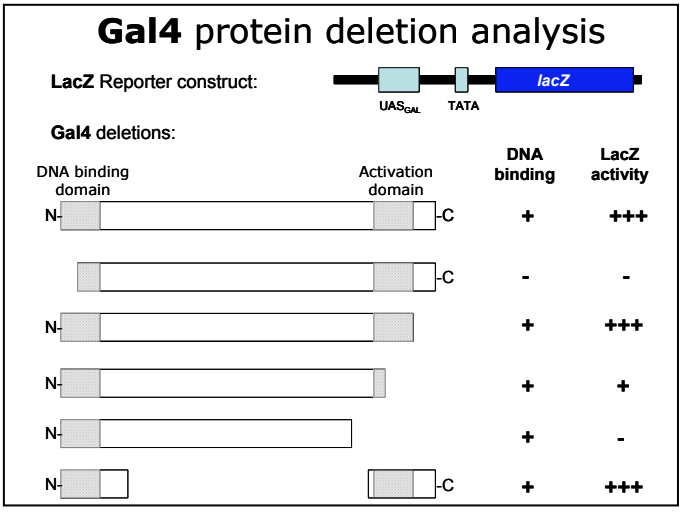


After determining that there was a **URS** element controlling glucose repression at the **Gal1** gene promoter, it was possible to go on to find the **Mig1** protein that binds the **URS_{GAL}** sequence (which turns out to lie in the promoter regions of many genes besides **Gal** genes). The Snf1 complex is a kinase that under low glucose conditions actively phosphorylates the **Mig1** repressor, preventing it from entering the

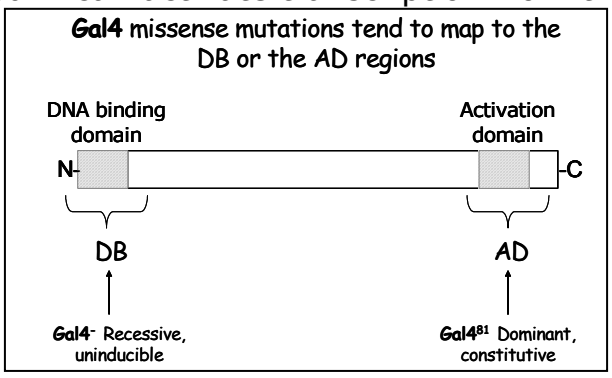
nucleus. This situation (low glucose) is permissive for galactose induction of **Gal1** gene expression via the **UAS**. In high glucose the Snf1 kinase is inactivated, so **Mig1** is not phosphorylated, and the unphoshorylated Mig1 enters the nucleus, to bind its **URS** sequence where it recruits two other proteins that together achieve repression of **Gal1** expression.

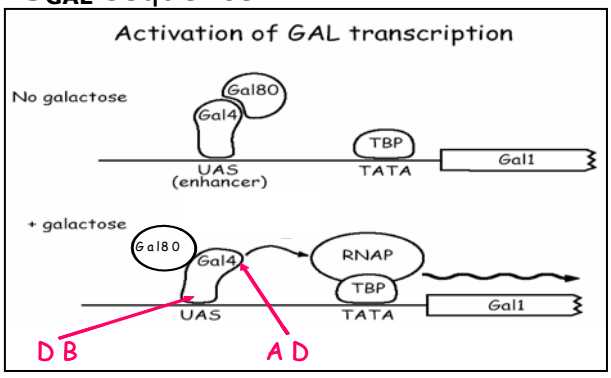
Modular properties of Transcription Activators

The **Gal4** transcriptional activator turns out to be one of the most well studied proteins to carry out this kind of function. Once again, a **LacZ** reporter was used in an imaginative way to establish that the **Gal4** protein has two functional domains that are separated by a flexible region in the protein. This time, the **Gal1** promoter region remains intact upstream of the **LacZ** reporter, but deletions are made across the **Gal4** protein; the inverse of keeping **Gal4** intact and making deletions along the promoter, as described above.



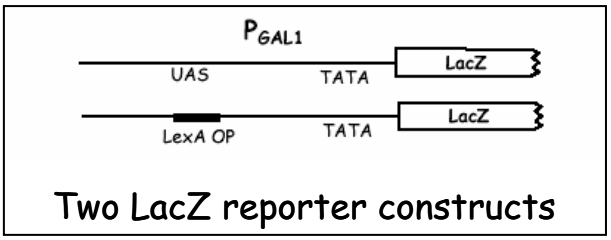
Essentially, if the N-terminal domain of the Gal4 protein is deleted, the protein can not bind to the **UAS$_{GAL}$** sequence, and so is unable to activate transcription of the reporter gene. But, in addition to DNA binding, **Gal4** must have a region near the C-terminal end that is responsible for recruiting and activating the RNA polymerase, thus allowing expression of the reporter gene. The most remarkable thing of all, was that a large region in the center of **Gal4** can be deleted; as long as the **DNA binding domain** is present at the N-terminus, and the **activating domain** is present at the C-terminus, **Gal4** can activate transcription from the **UAS$_{GAL}$** sequence.



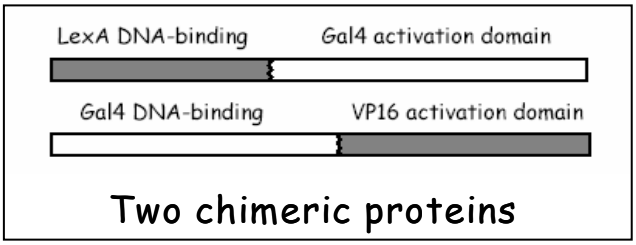


This remarkable separation of function between these two domains of **Gal4** was dramatically demonstrated by a series of experiments called **domain swapping**. Essentially, using recombinant DNA techniques, the **Gal4**

transcription activation domain (**AD**) was fused to the DNA binding (**DB**) domain of an *E. coli* protein called **LexA; LexA** is a repressor that binds to a known DNA sequence, the **LexA** operator (**LexA OP**). Also, the **Gal4 DB** domain was fused to the **AD** transcription activation domain of a viral protein know to be a strong activator, **VP16.** These **chimeric** proteins were introduced into yeast cells with the appropriate **LacZ** reporter gene constructs and the results of these domain swapping experiments were dramatic.



Two LacZ reporter constructs



Two chimeric proteins

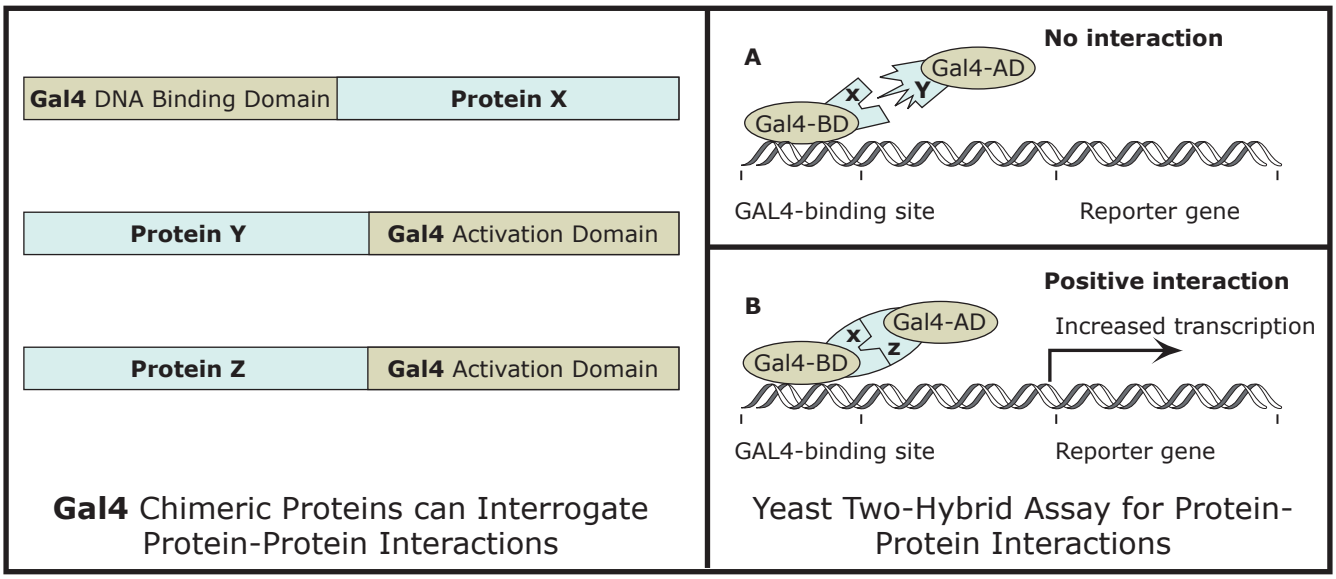
Two derivatives of a **Gal4**$^{-}$ yeast strain were created, one containing the **LacZ** reporter construct downstream of the **Gal1$_{UAS}$,** and the other containing the **LacZ** reporter construct downstream of the **LexA OP.** The two different **chimeric** proteins were expressed in each strain and the ability to induce **LacZ** activity monitored. In addition the following constructs were also introduced into the two strains: the wild type **Gal4** protein and a third **chimeric** protein with the activation domain of the **Gal4^{81}** mutant protein fused to the **LexA** DB domain. The results from these experiments clearly show that the **AD** and the **DB** domains function independently of one another.

	Gal1 UAS -gal/+gal	LexA OP -gal/+gal
Gal4^{+}	-/+	-/-
LexA-Gal4AD	-/-	-/+
LexA-Gal4AD^{81}	-/-	+/+
Gal4-VP16AD	+/+	-/-

This series of experiments, while interesting and certainly revealing about the how the **Gal** genes are regulated, have turned out to have a profound effect on all of biological research because it contributed to the development of a widely used technology called the **yeast two hybrid** assay. This assay makes it possible to determine whether two proteins interact with each other as a complex with long-lived interaction, and sometimes even when two proteins only interact transiently.

To determine whether **protein X** interacts with either **protein Y** or **protein Z** one can do the following: fuse **protein X** to the **Gal4 DB**, this chimeric protein is known as the **bait,** and it will attach to the **UAS$_{GAL}$** that lies upstream of a

reporter gene, usually a selectable marker or **LacZ,** or both. This **bait** lies in wait for an interaction with another protein. The **GAL4 AD**, is fused to either **protein Y** or **protein Z**. Should either one of these proteins be able to interact with **protein X** then the **Gal4 AD** region will become tethered to the **UAS_{GAL}** region and will recruit and activate the RNA polymerase.



Gal4 Chimeric Proteins can Interrogate Protein-Protein Interactions

Yeast Two-Hybrid Assay for Protein-Protein Interactions

Figure by MIT OCW.

Note that the **protein X, Y and Z** do not have to be yeast proteins; the only requirement is that the DNA coding sequence for the protein is available (which is now true for all of the genes from a wide variety of organisms); these sequences are then cloned such that they produce the appropriate **Gal4** chimeric proteins.

In the previous two three lectures we have looked at one particular regulatory network in *S. cerevisiae,* and have employed a wide range of tools to understand this network. In the next lecture I will be telling you how these and other tools have evolved into technologies that allow us to look globally at gene regulation in eukaryotic cells.