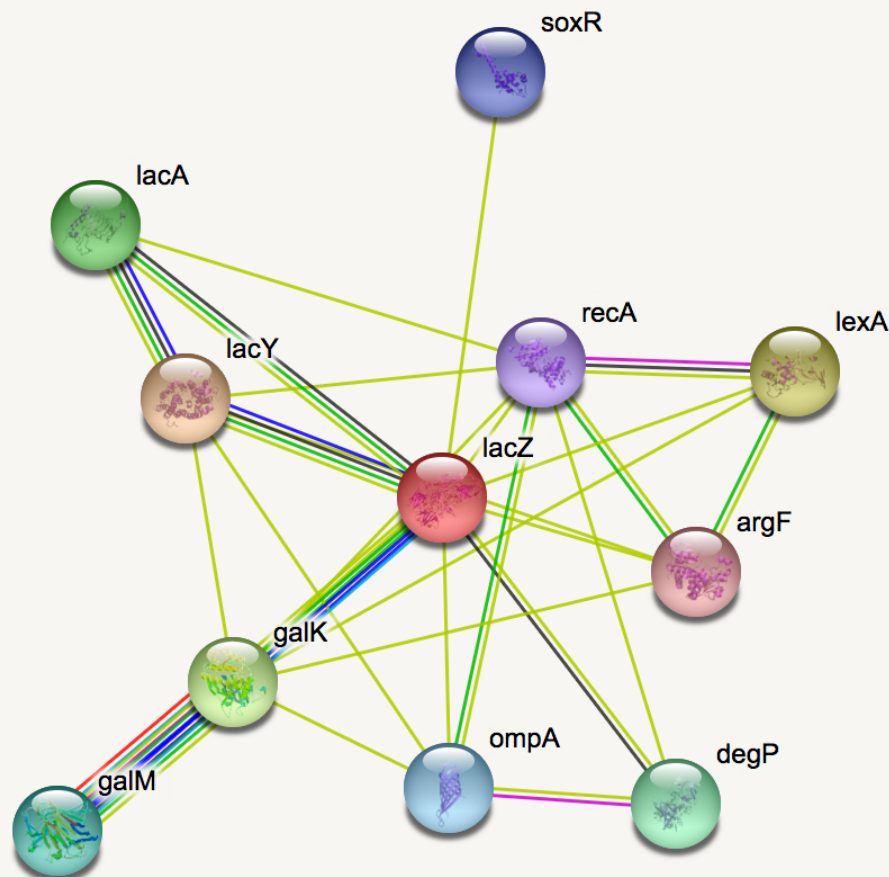


Homework 6

BCB570

This exercise takes place in *E. coli* K-12 (any substrain of K-12 is acceptable)

1. Locate the LacZ protein in the STRING database <http://string.embl.de>



lacY, lacA, galK, galM, ompA, degP, soxR, recA, argF, lexA

2. Which genes is lacZ strongly associated with? Why? To answer this question, read about the lac operon in Wikipedia. Write down the names of the genes and associated proteins of the lac operon and the other two genes lacZ is strongly associated with.

According to STRING database, 10 genes are associated with lacZ having varying weights according to the evidence. Out of those 10 proteins lacY, galK, lacA, galM are strongly associated

with lacZ written according to their STRING score. From previous literature, it has been identified that lacZ, lacY, and lacA are the three genes of the lac operon. As these genes are part of the same operon, they all get coded by the same transcription factor in a single transcript which in turn translated into multiple proteins associated with the metabolism of lactose in E.Coli. Due to this closeness, they are strongly associated with each other. The other two genes lacZ is strongly associated are galK and galM. As we know, gal genes encode for enzymes of the Leloir pathway for galactose metabolism, and lacA encodes a galactoside acetyltransferase which involves in galactose production. This relation explains the strong association of lacZ with galM and galK.

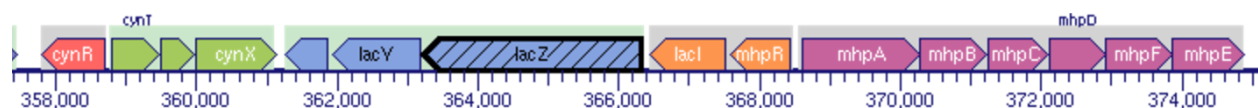
3. Locate lacZ in KEGG. Which of the STRING associated genes are there? Which are not? In which pathway do you see lacZ appear according to KEGG?

LacZ is associated with three specific pathways according to KEGG database including galactose metabolism, other glycan degradation, and sphingolipid metabolism, subsets of metabolic pathways. From this information, we can say that lacZ, in general, is associated with metabolic pathways. But from the previous literature and lac operon knowledge, we know that it is primarily associated with galactose metabolism pathway. Although it also plays an important role in other glycan degradation and sphingolipid metabolism pathways which also have galM and galK genes associated with it. We searched for the STRING genes associated with lacZ and could only find galM and galK in these pathways. The other proteins including lacY and lacA didn't appear in these pathways because they encode for enzymes which are not directly involved in the metabolism but plays an indirect role.

4. Repeat (3) for lacY and lacA. Why do you think they are not in the KEGG pathway as lacZ?

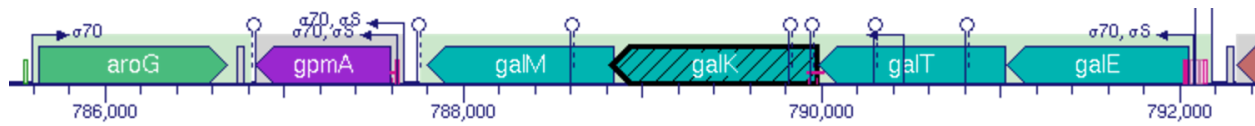
We couldn't find lacA or lacY in any of the KEGG pathways. This may be due to the fact that they are not directly involved in the lactose metabolic pathway, in fact, they help the Beta-galactosidase, product of lacZ, to carry of lactose metabolism by helping in lactose transportation and preparation.

5. Now locate lacZ in ecocyc.org Find it in the ecocyc genome browser as well.



6. Repeat (3)-(5) with galK or galM

GalK is associated with two specific pathways including the galactose metabolism pathway and the amino sugar and nucleotide sugar metabolism, a subset of the metabolic pathways. The genes we found in these pathways includes galM and lacY. It is also known that galM and galK genes are part of the same operon which is also seen on the ecocyc genome browser (shown below) which explains their relation. Also, it is also associated with lacY as explained earlier.



7. Why do you think argF is associated with lacZ in STRING? Is this association biologically meaningful? What would be your conclusion from such a reported association?

According to the STRING database, the string association between argF and lacZ is based on the the co-appearance of these terms in a five research articles on Pubmed. The association is purely based on text mining in which they and their putative homologs in other species co-appear. There is no experimental evidence for this. I think this association is biologically not meaningful as both of these proteins have different functions which are not related to each other. It also shows that relying completely on the computational techniques is not always good. We need to have some experimental validations. Similar to this case, two things which are completely opposite to each other can come as closely related. I think appropriate weights should be given to types of evidence based on its accuracy.