

Gama-Castro S., Santos-Zavaleta A., Peralta-Gil M., Weiss V., Salgado H., Solano H. and Collado-Vides J.

Programa de Genómica Computacional, Centro de Ciencias Genómicas, Universidad Nacional Autónoma de México, A.P. 565-A, Cuernavaca, Morelos 62100, México.



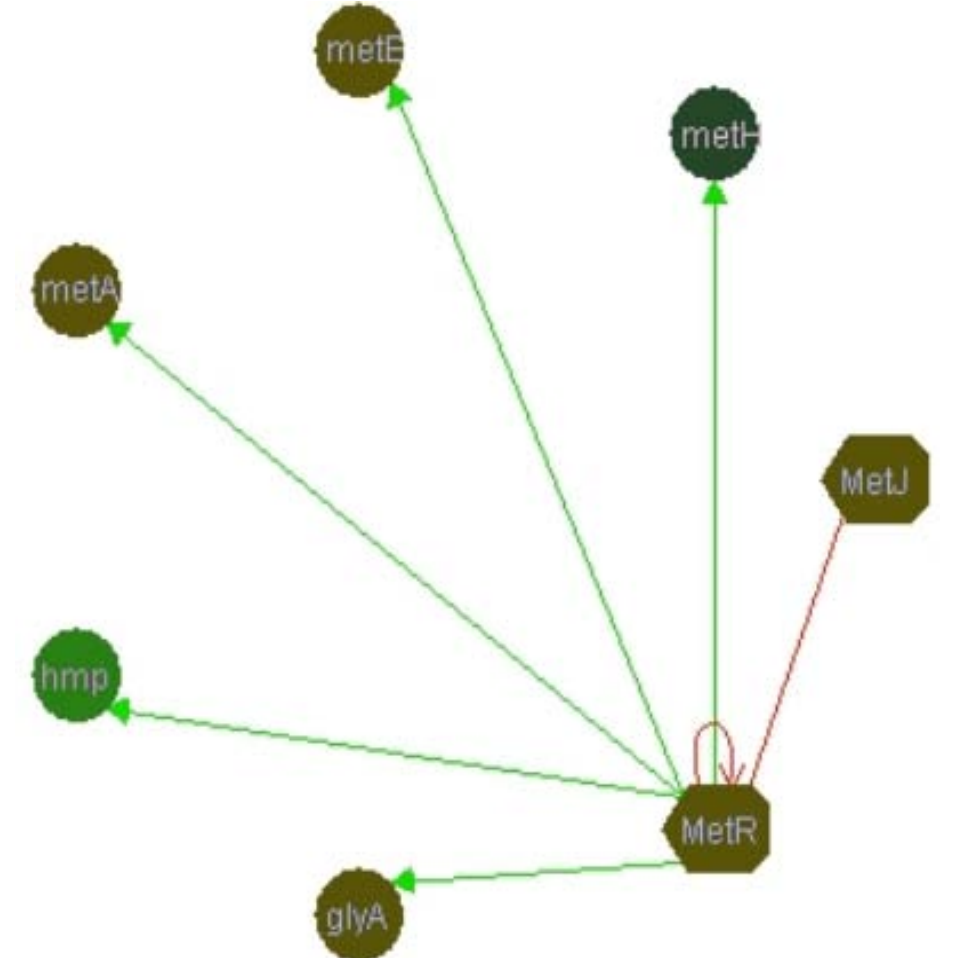
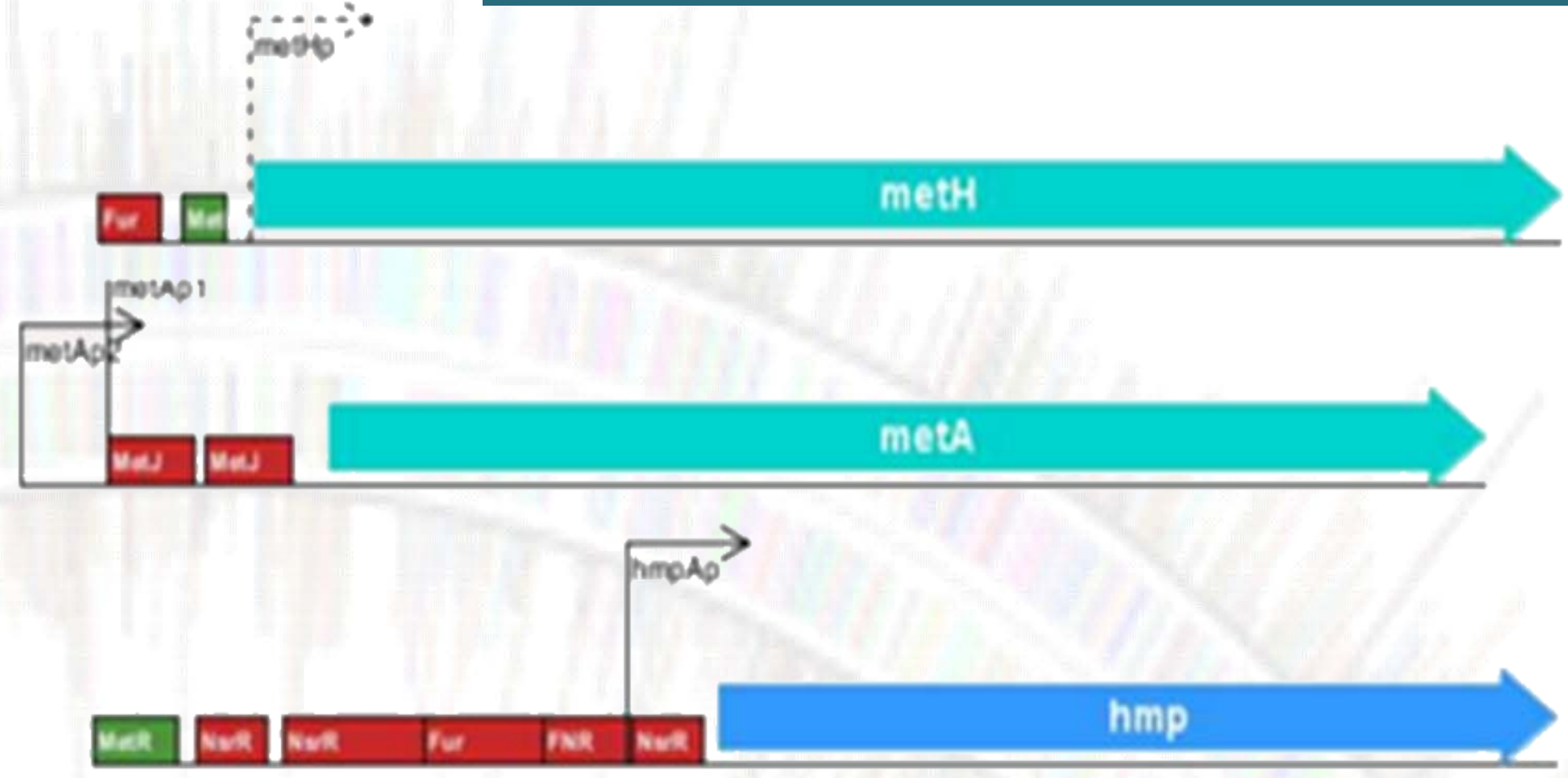
## ABSTRACT

Since its inception, RegulonDB (<http://regulondb.ccg.unam.mx/>) has been a database that compiles information about the regulation of transcription initiation of *Escherichia coli* K-12. However, we are aware that transcriptional regulation is not an isolated process; instead, it is the response to the different environmental conditions that trigger a series of concatenated reactions that end in transcriptional regulation, and it implies an adequate response in terms of induced and repressed gene products. We are working now to include all these new data in RegulonDB. As a consequence, transcriptional regulation in RegulonDB will be part of a unit that initiates with the signal, continues with the signal transduction to the core of regulation to modify expression of the affected set of target genes, and ends with an adequate response. We refer to these units as genetic sensory response units, or geSorgans.

The inclusion of geSorgans will bring a dramatic change and expansion of RegulonDB, due to the fact that we will be adding several new types of reactions and interactions. We started to collect data about signal transduction of the sigma factors, the two-component systems, of some transcription factors involved in carbon source utilization, and of genes involved in the synthesis of amino acids. We plan a high-level curation with super-pathways summarizing concatenated sets of reactions linked to those other databases that curate such information, while enabling with RegulonDB a compilation of complete geSorgans.

In addition, the number of DNA binding sites for some transcription factors has grown considerably, and therefore we decided to review systematically those sites whose lengths range from 40 to 60 bp with orientation and consensus sequences that are not easy to identify. The current version of RegulonDB is the beginning of a higher-level curation of gene regulation information, and eventually our database will include all regulatory mechanisms and their regulated genes.

## CURRENT DATA IN RegulonDB



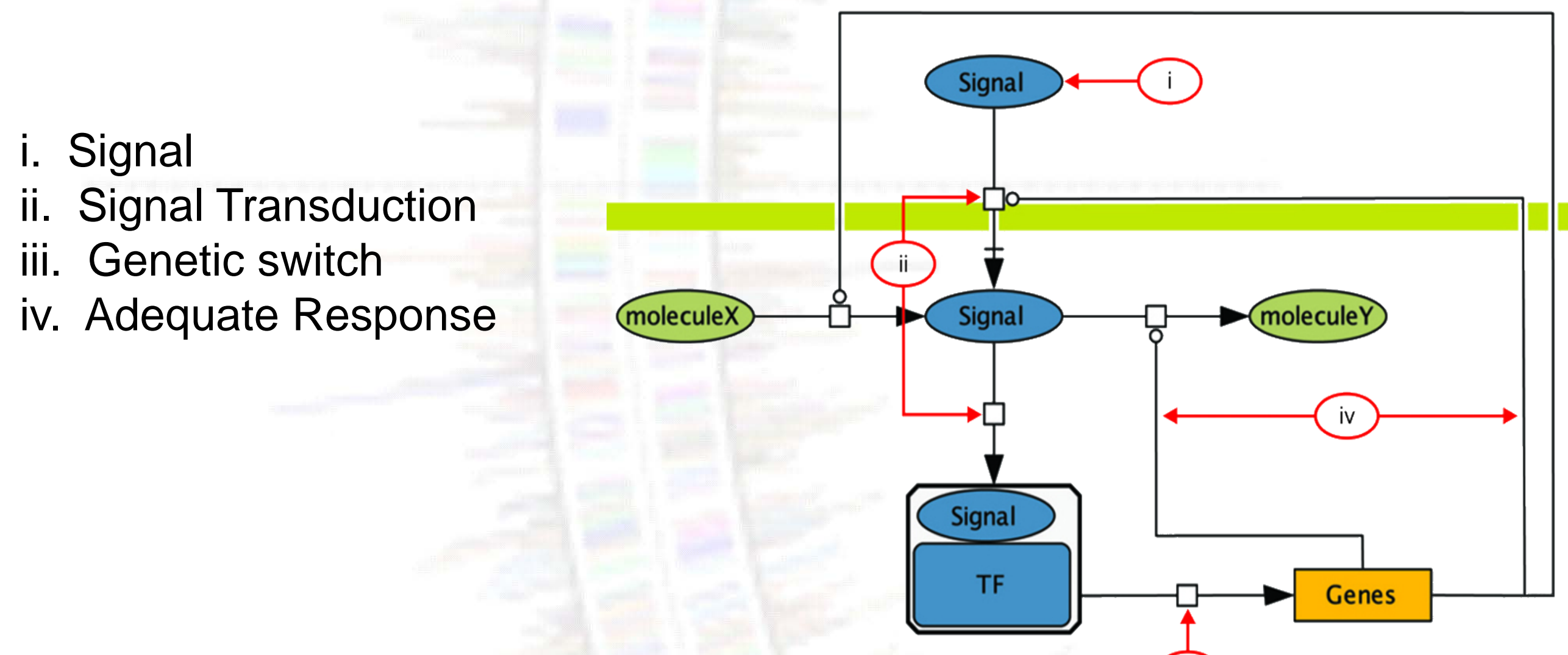
OPERON: *meth*  
Name: *meth*  
Transcription Unit: *meth*  
Gene(s): *methJ* *methK* *methL* *methS* *methH* *hmp* *hmpAp*

Transcription Factor	Function	Promoter	LeftPos	RightPos	CenterPos	Sequence	Evidence(s)	Reference(s)
MethJ	repressor	methP2	4011019	4011028	4011023	gaaataatgacgagcttcccaaac	[BPP]	[1]
MethK	repressor	methP2	4011034	4011041	4011037	gcaatagatctcttaccatggg	[BPP]	[1]
MethL	repressor	methP2	4011040	4011052	4011046	gtagcatgctatctcctcaggggga	[BPP]	[1]

Notes: Both sites of the MethJ regulator in the intergenic region of *methE* and *methF* could negatively regulate both the *methP2* and *methP1* promoters. Right sites of the MethJ regulator in the intergenic region of *methE* and *methF* could negatively regulate both the *methP2* and *methP1* promoters. MethR may be responsible for both the activation of expression of *methE* as well as the autoregulation of the *methR* gene.  
[BPP] Binding of purified proteins

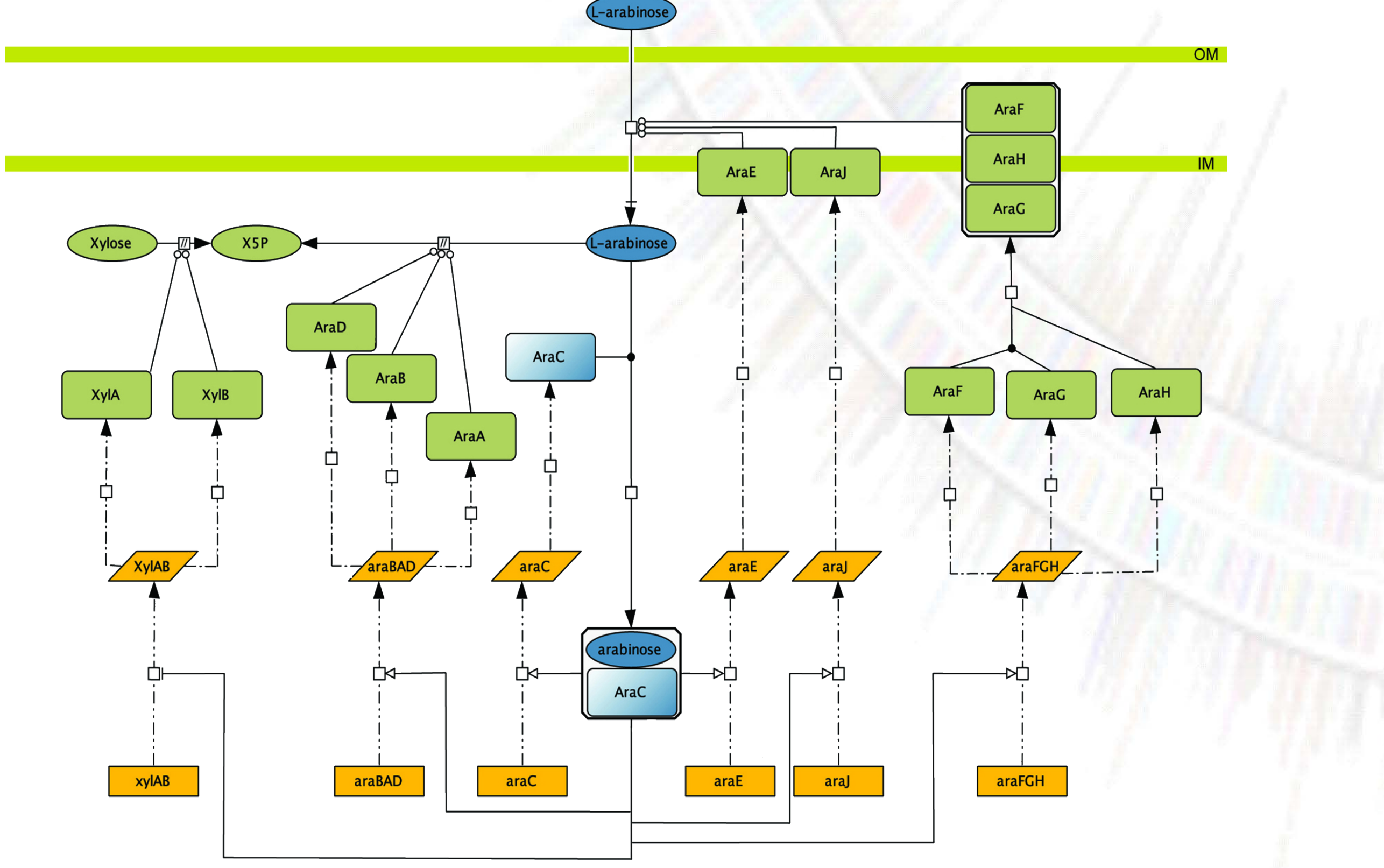
## NEW DATA: GENETIC SENSORY RESPONSE UNITS (geSorgans)

### Model & Components



### Integration of RegulonDB Data and other databases to build Sensory Response Units

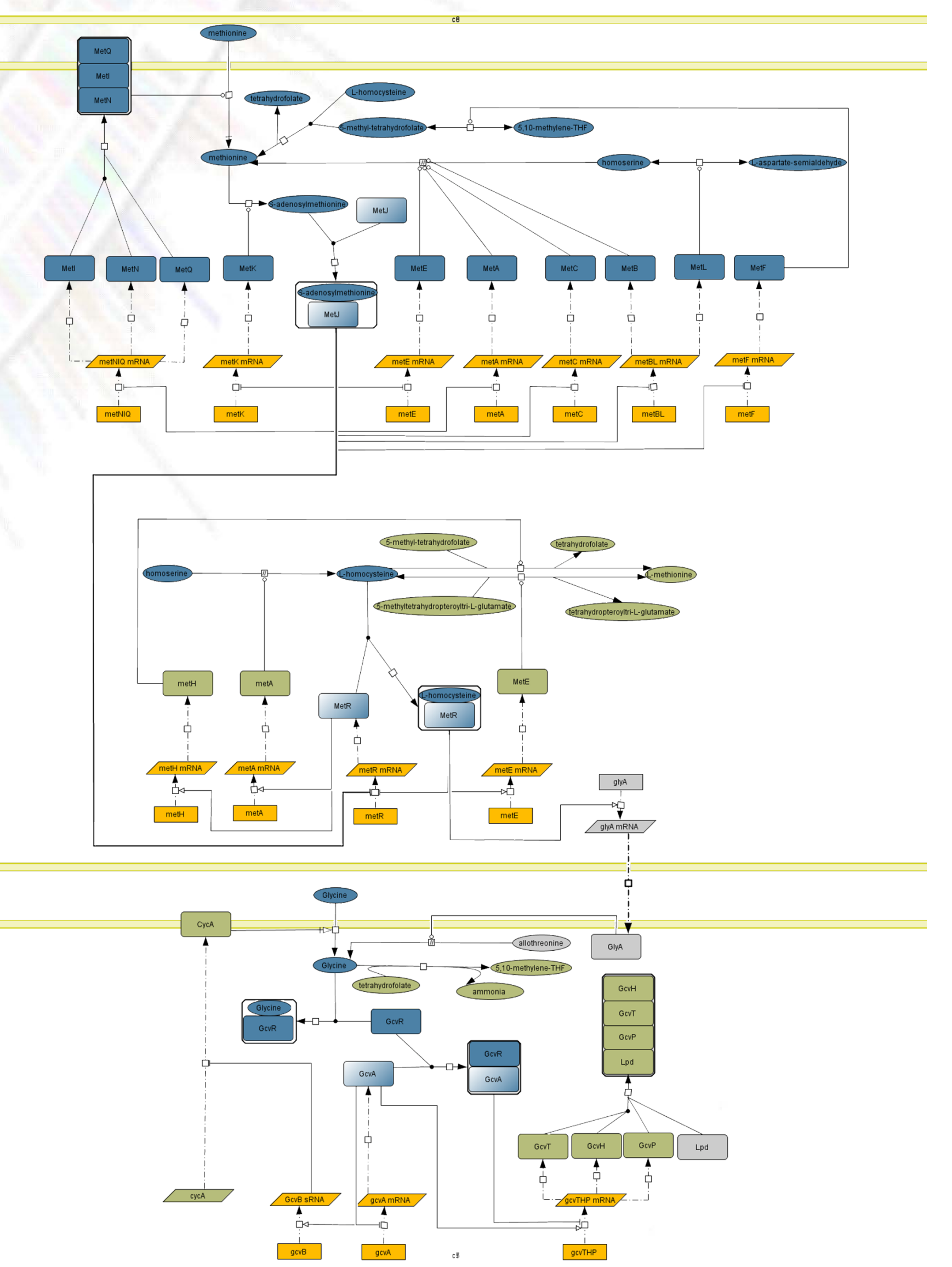
### AraC-L-arabinose Sensory Response Unit



### Reactions of a geSorgan

REACTION DESCRIPTION	Reaction #
L-arabinose (arabinose) → L-arabinose (arabinose)	1
Transposon (Tn) → Transposon (Tn)	2
L-arabinose + glucose → L-arabinose + glucose	3
L-arabinose + glucose → L-arabinose + glucose	4
L-arabinose + glucose → L-arabinose + glucose	5
L-arabinose + glucose → L-arabinose + glucose	6
L-arabinose + glucose → L-arabinose + glucose	7
L-arabinose + glucose → L-arabinose + glucose	8
L-arabinose + glucose → L-arabinose + glucose	9
L-arabinose + glucose → L-arabinose + glucose	10
L-arabinose + glucose → L-arabinose + glucose	11
L-arabinose + glucose → L-arabinose + glucose	12
L-arabinose + glucose → L-arabinose + glucose	13
L-arabinose + glucose → L-arabinose + glucose	14
L-arabinose + glucose → L-arabinose + glucose	15
L-arabinose + glucose → L-arabinose + glucose	16
L-arabinose + glucose → L-arabinose + glucose	17
L-arabinose + glucose → L-arabinose + glucose	18
L-arabinose + glucose → L-arabinose + glucose	19
L-arabinose + glucose → L-arabinose + glucose	20
L-arabinose + glucose → L-arabinose + glucose	21
L-arabinose + glucose → L-arabinose + glucose	22
L-arabinose + glucose → L-arabinose + glucose	23
L-arabinose + glucose → L-arabinose + glucose	24
L-arabinose + glucose → L-arabinose + glucose	25
L-arabinose + glucose → L-arabinose + glucose	26
L-arabinose + glucose → L-arabinose + glucose	27
L-arabinose + glucose → L-arabinose + glucose	28
L-arabinose + glucose → L-arabinose + glucose	29
L-arabinose + glucose → L-arabinose + glucose	30
L-arabinose + glucose → L-arabinose + glucose	31
L-arabinose + glucose → L-arabinose + glucose	32
L-arabinose + glucose → L-arabinose + glucose	33
L-arabinose + glucose → L-arabinose + glucose	34
L-arabinose + glucose → L-arabinose + glucose	35
L-arabinose + glucose → L-arabinose + glucose	36
L-arabinose + glucose → L-arabinose + glucose	37
L-arabinose + glucose → L-arabinose + glucose	38
L-arabinose + glucose → L-arabinose + glucose	39
L-arabinose + glucose → L-arabinose + glucose	40
L-arabinose + glucose → L-arabinose + glucose	41
L-arabinose + glucose → L-arabinose + glucose	42
L-arabinose + glucose → L-arabinose + glucose	43
L-arabinose + glucose → L-arabinose + glucose	44
L-arabinose + glucose → L-arabinose + glucose	45
L-arabinose + glucose → L-arabinose + glucose	46
L-arabinose + glucose → L-arabinose + glucose	47
L-arabinose + glucose → L-arabinose + glucose	48
L-arabinose + glucose → L-arabinose + glucose	49
L-arabinose + glucose → L-arabinose + glucose	50
L-arabinose + glucose → L-arabinose + glucose	51
L-arabinose + glucose → L-arabinose + glucose	52
L-arabinose + glucose → L-arabinose + glucose	53
L-arabinose + glucose → L-arabinose + glucose	54
L-arabinose + glucose → L-arabinose + glucose	55
L-arabinose + glucose → L-arabinose + glucose	56
L-arabinose + glucose → L-arabinose + glucose	57
L-arabinose + glucose → L-arabinose + glucose	58
L-arabinose + glucose → L-arabinose + glucose	59
L-arabinose + glucose → L-arabinose + glucose	60
L-arabinose + glucose → L-arabinose + glucose	61
L-arabinose + glucose → L-arabinose + glucose	62
L-arabinose + glucose → L-arabinose + glucose	63
L-arabinose + glucose → L-arabinose + glucose	64
L-arabinose + glucose → L-arabinose + glucose	65
L-arabinose + glucose → L-arabinose + glucose	66
L-arabinose + glucose → L-arabinose + glucose	67
L-arabinose + glucose → L-arabinose + glucose	68
L-arabinose + glucose → L-arabinose + glucose	69
L-arabinose + glucose → L-arabinose + glucose	70
L-arabinose + glucose → L-arabinose + glucose	71
L-arabinose + glucose → L-arabinose + glucose	72
L-arabinose + glucose → L-arabinose + glucose	73
L-arabinose + glucose → L-arabinose + glucose	74
L-arabinose + glucose → L-arabinose + glucose	75
L-arabinose + glucose → L-arabinose + glucose	76
L-arabinose + glucose → L-arabinose + glucose	77
L-arabinose + glucose → L-arabinose + glucose	78
L-arabinose + glucose → L-arabinose + glucose	79
L-arabinose + glucose → L-arabinose + glucose	80
L-arabinose + glucose → L-arabinose + glucose	81
L-arabinose + glucose → L-arabinose + glucose	82
L-arabinose + glucose → L-arabinose + glucose	83
L-arabinose + glucose → L-arabinose + glucose	84
L-arabinose + glucose → L-arabinose + glucose	85
L-arabinose + glucose → L-arabinose + glucose	86
L-arabinose + glucose → L-arabinose + glucose	87
L-arabinose + glucose → L-arabinose + glucose	88
L-arabinose + glucose → L-arabinose + glucose	89
L-arabinose + glucose → L-arabinose + glucose	90
L-arabinose + glucose → L-arabinose + glucose	91
L-arabinose + glucose → L-arabinose + glucose	92
L-arabinose + glucose → L-arabinose + glucose	93
L-arabinose + glucose → L-arabinose + glucose	94
L-arabinose + glucose → L-arabinose + glucose	95
L-arabinose + glucose → L-arabinose + glucose	96
L-arabinose + glucose → L-arabinose + glucose	97
L-arabinose + glucose → L-arabinose + glucose	98
L-arabinose + glucose → L-arabinose + glucose	99
L-arabinose + glucose → L-arabinose + glucose	100

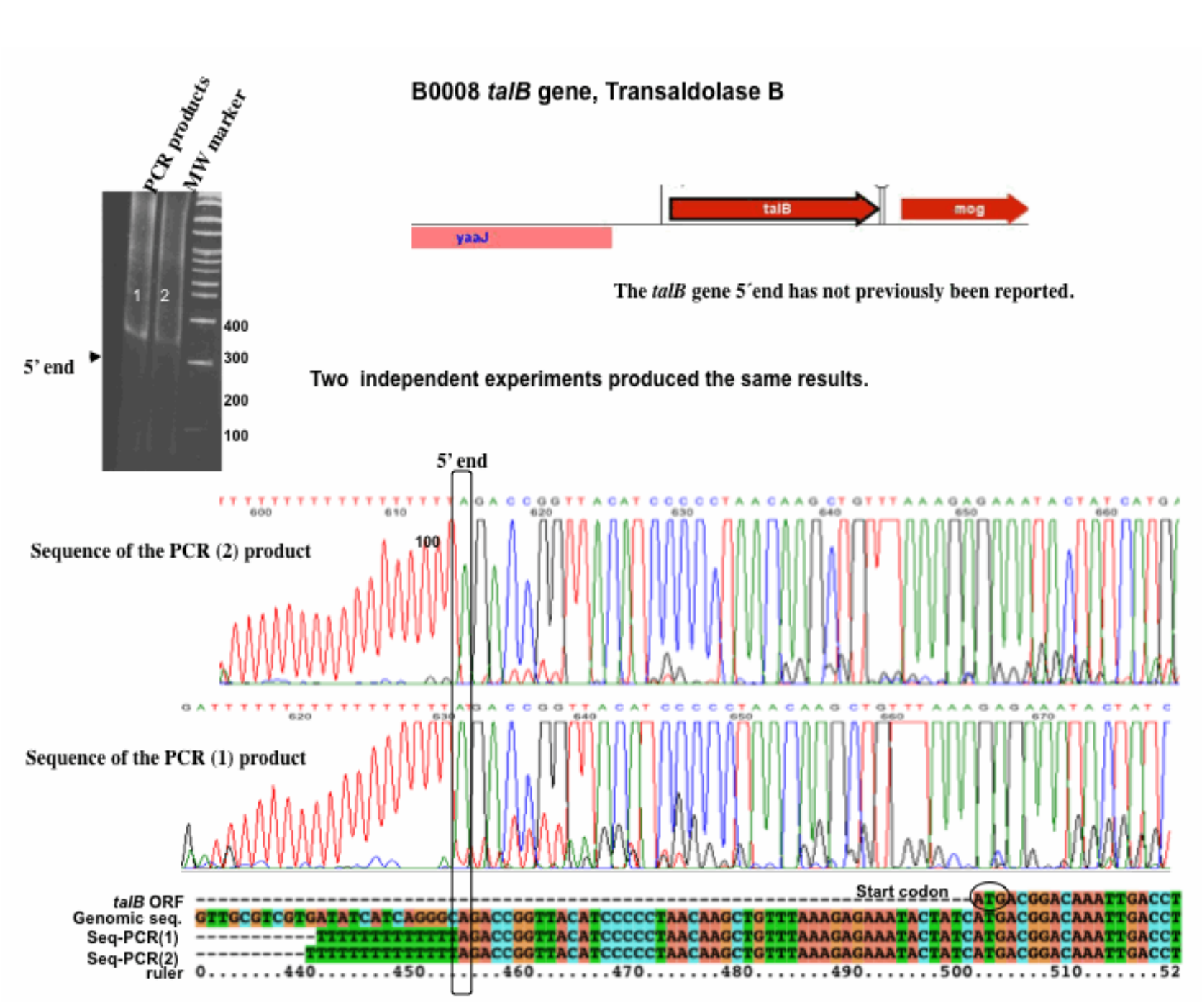
### geSorgan Connections



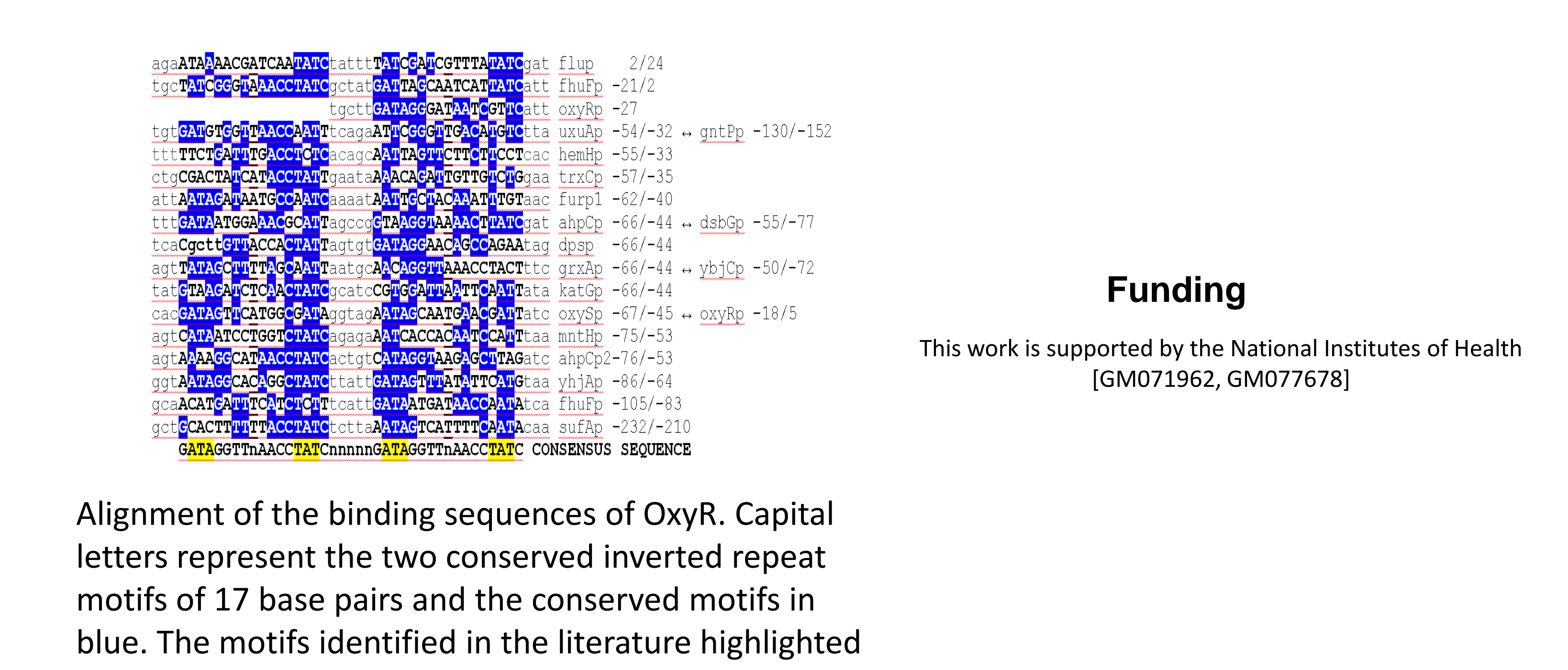
Summary: In the presence of L-arabinose, AraC activates transcription of genes that code for proteins necessary for utilization and transport of L-arabinose.

## ACTIVE CURATION

### High-Throughput Mapping of Transcription start sites



### Redefining Transcription factor binding sites



Funding  
This work is supported by the National Institutes of Health [GM071962, GM077678]