Molecular and functional insights into the regulation of D-galactonate metabolism by a GntR family transcriptional regulator, DgoR in *Escherichia coli*

BHUPINDER SINGH

A thesis submitted for the partial fulfillment of

the degree of the Doctor of Philosophy

Department of Biological Sciences

Indian Institute of Science Education and Research Mohali

Knowledge city, Sector 81, SAS Nagar, Manauli PO, Mohali 140306, Punjab, India.

December 2019

Dedicated to my family

& science

Declaration

The work presented in this thesis has been carried out by me under the guidance of Dr. Rachna Chaba at the Indian Institute of Science Education and Research Mohali. This work has not been submitted in part or in full for a degree, a diploma, or a fellowship to any other university or institute. Whenever contributions of others are involved, every effort is made to indicate this clearly, with due acknowledgement of collaborative research and discussions. This thesis is a bona fide record of original work done by me and all sources listed within have been detailed in the bibliography.

Bhupinder Singh

In my capacity as the supervisor of the candidate's thesis work, I certify that the above statements by the candidate are true to the best of my knowledge.

Dr. Rachna Chaba

Acknowledgements

First and foremost, I owe a great deal of gratitude to my Ph.D. advisor, Dr. Rachna Chaba for her guidance and support throughout my research work. Among her many qualities, I admire her will power and hard work the most. Many a times, I have witnessed her putting family responsibilities aside for work. And with every passing year, she has only raised the bar in the department of dedication. She has gone to great lengths to provide me with various forms of help within and beyond her reach. It is due to her credit that I got the best exposure to scientific community one could ask for. I consider myself very fortunate to be one of her first students, as I received first-hand training and the know-hows of the research. I whole-heartedly thank her for instilling work ethics and scientific aptitude in me. To put it simply, I could not have asked for a better guide.

I profoundly acknowledge my thesis committee members, Prof. Anand K. Bhachhawat and Dr. Kausik Chattopadhyay for timely reviewing my progress. Their valuable suggestions gave a new perspective to my research work. Also, my most sincere thanks goes to Prof. Anand K. Bhacchawat for his ever-caring and encouraging disposition towards me.

I express my gratitude to IISER Mohali directors, Prof. N. Sathyamurthy, Prof. Debi P. Sarkar, and Dr. J. Gowrishankar, for providing me an excellent academic environment and state of the art lab facilities for untroubled research activities.

I would also like to thank Dr. Dipak Dutta for radioactive facility at IMTECH, Chandigarh. His student and my dear friend Dr. Vineet along with others offered a very supportive and cheerful environment at IMTECH for conducting very crucial experiments. I would not have been able to complete this thesis without their help. I also thank Dr. Christopher V. Rao (Univ. of Illinois) for kindly providing resources for making reporter constructs. I duly thank Dr. Samrat Mukhopadhyay and Dr. Shravan Kumar Mishra for sharing their knowledge with me regarding CD spectroscopy and protein purification. I am also thankful to Dr. Arunalanda Babu and Dr. Sugumar Venkataramani and their students, Dr. Nayyar, Debapriya and Pravesh for NMR analysis. I am forever grateful to Dr. Shashi Bhushan Pandit and Dr. Vishal Agrawal for bioinformatics analysis. Dr. Vishal Agrawal holds a very special place in my regards as he played an important role by sharing his wisdom at various stages of my research. I also offer special thanks to Tecan team especially Mr. Piyush, for their help in running robotics facility. I also extend my gratitude to Dr. S. Kumaran and Dr. Dibyendu Sarkar from IMTECH for discussions regarding important experiments.

Many colleagues helped me in multiple ways to learn different techniques. A special mention goes to Dr. Prerna, Dr. Kanika, Dr. Prince Tiwari, and Arpita Mrigwani from Dr. Purnananda Guptasarma's lab; Dr. Nidhi Kundu and Amritha from Dr. Kausik's lab and Dr. Karishma, Dr. Hema, Priyanka Madhu, Aishwarya, and Sayanta from Dr. Samrat's lab; Dr. Krishnaprasad, and Dr. Sanica from Dr. Arunika's lab and Dr. Arashdeep from Dr. Kuljeet's lab.

I duly acknowledge financial support from IISER Mohali and Department of Biotechnology, Govt. of India.

I am really at a loss of words to express my gratitude for my fellow lab members, so here I simply mention their names, Dr. Shashank Agrawal, Kanchan Jaswal, Garima, Neeladrita, Swati, Megha and Deeptodeep. Garima along with Neeladrita Kundu and previous graduate students Akshay Sangwan, Shachikanta Nongthombam and Yatendra also worked in collaboration on this project. Akshay, Shachikanta and Yatendra worked with me during their BS-MS dissertation, I will always cherish the brotherly bonding, I share with them. I also acknowledge the present and past post-docs of our lab, Dr. Tapas Patra, Dr. Prakaram and Dr. Mohinder Pal for guiding me both on academic as well as personal matters. I express my gratitude to Vidya for his joyful and persistent aptitude towards providing help in various forms. A special heart-felt mention goes to Shaswath, Twinkle, Anagha, Navneet, Bandhu, Shweta, and Bhargesh, the past members of our lab. I also thank Aanya Agrawal for her cheerful presence at various lab gatherings.

My journey at IISER Mohali started in a space shared by four other labs, dearly known as 3TL2 lab. Working in the shared lab space offered me a much steeper learning curve. It is due to the cordial and scholarly members of the 3TL2, I got an opportunity to absorb knowledge about various model organisms.

I will always cherish the memories and accomplishments I made with the fellow iGEM team members, Dr. Manisha, Prateek, Ashwin and Nitish.

A long list of friends made my stay at IISER Mohali truly adventurous including Dr. Nagesh, Mr. Mayank, Deepinder, Dr. Devashish, Dr. Simran, Dr. Manisha, Dr. Prince, Kiran, Dr. Ritu, Muskan, Dr. Shivani, Dr. Krishna, Dr. Prashant, Dr. Aastha, 'Ashke bhangra team' among many others. I will always be thankful to Ravinder sir, our Yoga guru, for imbuing the art of yoga and mindfulness in me.

I thank my past teachers and mentors Dr. Brij Mohan Sarwal, Dr. Rupinder Tewari, and Dr. Rohit Sharma for their encouraging words and guidance.

And in the end, I would like to express my appreciation for my loving parents and brother, Rajwinder for their unconditional and incomparable love and care. I will always owe an infinite debt of gratitude towards my mother for putting my education first despites several hardships. I also express my gratitude to my aunt Gurmail, uncle Jasbir and uncle Harpal for coaching and supporting my fundamental education.

Bhupinder Singh

Publications

A part of the work embodied in this thesis has been published in:

Singh B, Arya G, Kundu N, Sangwan A, Nongthombam S, Chaba R. 2019. Molecular and functional insights into the regulation of D-galactonate metabolism by the transcriptional regulator DgoR in *Escherichia coli*. **J Bacteriol** 201:e00281-18. PMID: 30455279

https://doi.org/10.1128/JB.00281-18

Abstract

D-galactonate, a widely prevalent hexonate sugar acid, is used as a carbon and energy source by *Escherichia coli*. Although the structural *dgo* genes involved in the transport and metabolism of D-galactonate have been investigated, there is limited information of its regulatory aspect. Using various genetic, biochemical and bioinformatics approaches, we investigated the regulation of D-galactonate metabolism in *E. coli* K-12. In this work, we have presented the first molecular and functional insights into the regulation of D-galactonate metabolism by the transcriptional regulator, DgoR. We find that the *dgo* operon is transcriptionally repressed by DgoR and induced specifically in the presence of D-galactonate. Deletion of *dgoR* accelerates the growth of *E. coli* in medium containing D-galactonate as a carbon source concomitant with the strong constitutive expression of *dgo* genes. DgoR exerts a strong repression over *dgo* genes by binding two closely spaced inverted repeats overlapping the putative D-galactonate responsive *dgo* promoter. D-galactonate itself rather than any of its metabolic intermediates acts as the true effector to relieve the DNA bound by DgoR. Multiple findings from our work firmly place DgoR in the FadR subfamily within the GntR family of transcriptional regulators: DgoR is a majorly α-helical protein with GntR-type N-terminal wHTH domain and a predicted FadR-like all helical C-terminal FCD domain, binds [5'-TTGTA(G/C)TACA(A/T)-3'] operator sequence matching the signature of GntR family members that recognize inverted repeats [5'-(N)yGT(N)xAC(N)y-3'], and shares critical protein-DNA contacts conserved in the GntR family. Additionally, we identified features in DgoR that are otherwise less common in the regulators of GntR family. Multiple reports from the last couple of decades have implicated the physiological significance of D-galactonate metabolic pathway in the interaction of enteric bacteria with their host. Importantly, in a recent *in vivo* evolutionary study, *E. coli* adapted to the mammalian gut was found to accumulate multiple missense mutations in *dgoR*. Our results show these mutants to be DNA-binding-defective emphasizing that mutations in the *dgo*-regulatory elements are selected in the host to allow simultaneous induction of *dgo* genes. Considering that DgoR and its binding sites have been predicted in several enterobacterial strains, the present study sets the basis to explore the regulation of D-galactonate metabolic pathway in these strains and its possible role in mediating host-bacterial interactions.

Thesis Synopsis

Title – Molecular and functional insights into the regulation of D-galactonate metabolism by a GntR family transcriptional regulator, DgoR in *Escherichia coli*

Supervisor – Dr. Rachna Chaba

Department – Department of Biological Sciences

Institute – Indian Institute of Science Education and Research (IISER)-Mohali

Sugar acids, oxidized derivatives of sugars are used as carbon and energy source by several bacteria, including pathogens. D-galactonate is a widely prevalent hexonate sugar acid used as a nutrient source by enteric bacteria. Since the intial studies conducted in 1970s, the details are available only on some of the structural *dgo* genes involved in the transport and metabolism of D-galactonate. Prior to this work, classical studies from 1970s proposed DgoR as a regulator of D-galactonate metabolism in *E. coli* K-12, however, the detailed characterization of DgoR was lacking. Several genome-scale studies in the last couple of decades have illustrated the physiological importance of D-galactonate metabolism for the interaction of bacteria with their hosts. Importantly, a recent study aimed at identifying mutations in the genome that enable adaptation of an *E. coli* strain isolated from a urinary tract infection patient in the mouse gut and another study which provided a comparison of a panel of 340 naturally occuring *E. coli* isolates collectively suggested that the D-galactonate metabolic genes and its putative regulator, DgoR are under selection pressure, and can potentially play an important role in bacterial adaptation to their natural habitat. Furthermore, studies in the last couple of years have emphasized that genetic variations in the regulators of carbon source metabolism dramatically affect the interaction of bacteria with their host. For example, a point mutation in TreR, a transcriptional repressor of trehalose metabsolism was found to increase the sensitivity of certain ribotypes of *Clostridium difficile* to trehalose enabling them to metabolize this carbohydrate even at lower concentrations and attain hypervirulence. In another example, single nucleotide polymorphism in RafR, a transcriptional activator of raffinose metabolism in *Streptococcus pneumoniae*, was shown to be associated with differential metabolism of raffinose among blood and ear isolates, and served as an important determinant for their distinct tropism for lungs vs

brain and ear in mice. Taken together, the above instances stress the need to study the molecular details of the regulation of D-galactonate metabolism by DgoR, and its contribution to host-bacterial interactions.

In the present study, we used various genetic, biochemical, and bioinformatics approaches to investigate the regulation of D-galactonate metabolism in *E. coli*. We showed that the transcription of *dgo* genes is strongly repressed by DgoR and is de-repressed specifically in the presence of D-galactonate. Our results that DgoR binds to two closely-spaced inverted repeats overlapping the putative *dgo* promoter suggest that DgoR binding occludes the binding of transcriptional machinery to initiate transcription of D-galactonate metabolic genes. Using both *in vitro* and *in vivo* approaches, we tested the role of several N-terminal amino acid residues of wHTH domain of DgoR for interaction with the operator DNA. Our results suggest that DgoR displays both conserved as well as semiconserved features of GntR family regulators in interacting with the DNA. Importantly, during our investigation, we showed that the previously reported missense mutations selected in DgoR during *in vivo* adaptation of urinary tract isolate of *E. coli* to the mouse gut result in loss of DNA binding suggesting that upregulation of D-galactonate metabolism plays a role in the adaptation of *E. coli* to the mammalian gut. Finally, we identified D-galactonate as a true effector of DgoR that binds DgoR and releases it from the operator DNA. Considering that DgoR and its binding sites have been predicted in several enterobacterial strains, fundamental insights gained from our work set the basis to explore the regulation of D-galactonate metabolic pathway in these enteric bacteria and its potential role in mediating host-bacterial interactions.

This thesis is divided into six chapters organised as below:

Chapter I includes an extensive review of literature on topics relevant to this study**.** The present work deals with investigating the role of DgoR in regulating metabolism of an important sugar acid, D-galactonate, as a carbon source in *E. coli*. To align the readers to the importance of the study, the first section of Chapter I highlights the relevance of carbon source metabolism and its regulation in *E. coli* and the next two sections describe metabolism of various carbon sources utilized by *E. coli*. A separate section gives information on general mechanisms employed by *E. coli* for regulating carbon source metabolism. As DgoR belongs to the GntR family of transcriptional regulators, the next section gives a general introduction as well as details on DNA- and effector-binding characteristics of the regulators of the GntR family. The subsequent section outlines the

sugar acids used as carbon source by *E. coli*, their general metabolic routes as well as the regulation exerted on them. Further section gives a detailed account on D-galactonate metabolism and its regulation to inform the readers about the existent literature. The last section provides an outline of the objectives and the important findings from the present study.

Chapter II describes the materials and methods that have been used in this work.

Chapter III, IV and **V** cover the major results obtained in this study.

Chapter III: DgoR negatively regulates D-galactonate metabolism in *E. coli*

Early classical mutagenesis studies coupled with biochemical analysis of crude lysates prepared from mutants defective in growth on D-galactonate implicated the role of *dgo* genes (*dgoKADT*) organized in a putative operon in metabolizing D-galactonate in *E. coli*. A mutation isolated in the same study exhibited increased expression of all the Dgo enzymes and was found to map closer to the mutations in the other *dgo* genes, thus DgoR, a regulator of D-galactonate metabolism was proposed. In this section, we validated the role of various *dgo* genes in D-galactonate metabolism. Our results that *dgoR* deletion strain grows faster in medium with D-galactonate as carbon and energy source suggest that DgoR negatively regulates *dgo* genes. By measuring *dgo* transcripts in *dgoR* deletion strain grown in non-inducing conditions, we showed that DgoR strongly represses *dgo* genes at a transcriptional level. Further, WT strain showed simultaneous transcriptional induction of all the *dgo* genes only when grown in the presence of D-galactonate indicating a specific regulation on the *dgo* genes. We performed EMSA and *in vivo* fluorescent reporter assays using WT *dgo cis*-acting element (DNA upstream of *dgo* genes) as well as its mutant variants. Our results showed that DgoR represses *dgo* genes by binding to two closely spaced inverted repeats overlapping with the putative *dgo* promoter, suggesting that DgoR prevents the trancription of *dgo* genes by occluding the binding of transcriptional machinery to the putative *dgo* promoter. The inverted repeat sequence recognized by DgoR matches the consensus of operators recognized by GntR family members, validating that DgoR is a member of the GntR family.

Chapter IV: DgoR shares conserved and semiconserved protein-DNA contacts with GntR family of transcriptional regulators

GntR family of regulators bind cognate DNA using their N-terminal wHTH domain. In this section, we investigated the molecular details of interaction of DgoR with its operator DNA. Comparison of solved structures of GntR family regulators in their DNA-bound state shows that despite low sequence similarity in their N-terminal wHTH DNA binding domain, the GntR family members employ conserved amino acid residues to make specific contacts with their cognate DNA. We rationally chose to mutate three amino acid residues from each α-helix of the wHTH domain of DgoR that are either conserved or semiconserved in the GntR family regulators. Besides, we also created missense mutations selected in the *in vivo* adaptation study described above. Our results show that these mutants are defective in DNA-binding and unable to repress the *dgo* promoter suggesting their importance in DgoR-DNA interaction. Our results that DgoR carrying missense mutations selected in the *in vivo* adaptation study are DNA-binding defective, indicate that upregulation of D-galactonate metabolism plays a role in the adaptation of *E. coli* to the mammalian gut.

The CD spectrum analysis of DgoR showed that similar to other members of the FadR subfamily within the GntR family, DgoR is a majorly α-helical protein. Our *in vivo* reporter data that the DNA-binding defective mutants are dominant negative coupled with *in vitro* data that the immobilized His-tagged DgoR pulls down MBP-tagged DgoR, suggest that DgoR forms oligomers in solution, a feature similar to other GntR family members. Subsequent chemical crosslinking of purified DgoR showed that the major oligomeric form of DgoR is a dimer. Collectively these data, along with findings from the previous chapter that the inverted repeats recognized by DgoR are highly similar to that of GntR family regulators, establish that DgoR is a member of the FadR subfamily within the GntR family of transcriptional regulators.

Chapter V: D-galactonate is the physiological effector of DgoR

Multiple biochemical and structural studies show that the FadR subfamily regulators involved in regulating the metabolism of carbon sources such as sugar acids usually bind a functionally relevant small organic molecule as an effector. Although, the cognate effector molecule which can be either sugar acid substrate itself, its metabolic intermediate, or both have been determined for these sugar acid metabolism regulators, the molecular details of effector-regulator interaction and underlying mechanism to

release operator DNA bound by the regulator are still not known. In this chapter, we analyzed the binding of *dgo cis*-acting element by DgoR in the presence of either D-galactonate, its various metabolic intermediates or structurally similar carbohydrates such as sugars D-glucose, D-galactose, and an epimeric sugar acid, D-gluconate. Of all the tested compounds, only D-galactonate interfered with the binding of DgoR with its *cis*-acting element suggesting that D-galactonate is a true effector of DgoR.

Amongst the FadR subfamily members, apo structures along with its effector-, and DNA- bound forms are solved only for FadR from *E. coli* and *Vibrio cholerae* where the effector binding characteristics are similar except that an additional amino acid stretch is found in the FadR from *V. cholerae*, which provides a second site for binding the effector. To initiate probing the molecular details of DgoR-D-galactonate interaction, we perfomed a structure based sequence alignment of DgoR with FadR from *E. coli*. This analysis showed that among the various FadR subfamily members, DgoR exclusively shares a serine residue with FadR which is known in the latter protein to directly interact with its effector, long-chain fatty acyl-CoA. The serine amino acid mutant of *E. coli* FadR (S219N) has been reported to behave as a super-repressor with a compromised effector mediated release of bound DNA and subsequent de-repression of the FadR regulon members. Our *in vivo* complementation and reporter assays showed that similar to FadR S219N mutant, the corresponding DgoR S221N mutant also behaves as a super-repressor. The mutant is defective in D-galactonate mediated de-repression of the putative *dgo* promoter indicating that the serine residue in question may be important for D-galactonate mediated release of DNA bound by DgoR.

Chapter VI summarizes the findings of the present study along with concluding remarks. A section on the future prospects of this study has also been included.

Table of Contents

List of Figures

Chapter I: Review of Literature

List of Tables

Chapter I: Review of Literature

Chapter II: Materials and Methods

Abbreviations

Weight and measures

°C	Degree Celsius
g	Gram
µg	Microgram
µl	Microliter
µm	Micrometre/micromolar
kDa	Kilodaltons
mg	Milligrams
ml	Milliliter
mm	Millimetre
mM	Millimolar
MQ	Milli-Q
min	Minutes
nm	Nano-meter
nM	Nanomolar
OD	Optical density
rpm	Revolutions per minute
V	Volts
w/v	Weight (of solute) per volume (of solvent)

Symbols

*	Asterisk
:	Colon
%	Percent
~	Approximately
α	Alpha
β	Beta
γ	Gamma
λ	Lambda
δ	Delta
Δ	Delta
Å	Angstrom

Techniques

CD	Circular Dichroism
EMSA	Electrophoretic mobility shift assay
HPLC	High-performance liquid chromatography
NMR	Nuclear magnetic resonance
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
PFG	Pulsed field gradient
qRT-PCR	Quantitative reverse transcription PCR

Chemicals

KDPG	2-keto-3-deoxy-D-gluconate 6-phosphate
MPD	2-methyl-2,4-pentanediol
MOPS	3-(N-morpholino) propanesulfonic acid
HEPES	4-(2-hydroxyethyl)-1-piperazineethane sulfonic acid
KDG	5-keto-4-deoxy-D-glucarate/galactarate
APS	Ammonium persulfate
Co-NTA	Cobalt nitrilotriacetic acid
CoA	Coenzyme A
cDNA	Complementary deoxyribo nucleic acid
dATP	Deoxyadenosine triphosphate
DNA	Deoxyribo nucleic acid
dNTPs	Deoxyribonucleotide triphosphate
DMSO	Dimethyl sulfoxide
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
HCl	Hydrochloric acid
IPTG	Isopropyl β- d-1-thiogalactopyranoside
LB	Lysogeny broth
mRNA	Messenger ribonucleic acid
GlcNAc	N-acetylglucosamine
ManNAc	N-acetylmannosamine
Neu5Ac	N-acetylneuraminic acid
PMSF	Phenylmethylsulfonyl fluoride
PVDF	Polyvinylidene fluoride

PLP	Pyridoxal phosphate
RNA	Ribonucleic acid
sRNA	Small ribonucleic acid
SDS	Sodium dodecyl sulfate
SgrS	Sugar transport-related sRNA
TEMED	Tetramethylethylenediamine
TPP	Thiamine pyrophosphate
TBE	Tris Borate EDTA

Proteins/Enzymes

Cra	Catabolite repressor/activator
ASKHA	Acetate and Sugar Kinases/Hsc70/Actin
ACS	Anion:Cation Symporter
ArcB-ArcA	Anoxic redox control two component system
ABC	ATP-binding cassette
CsrA	Carbon storage regulator
cAMP-CRP	Cyclic adenosine monophosphate-cAMP receptor
CTD	C-terminal domain
DBD	DNA-binding domain
FCD	FadR C-terminal domain
FNR	Fumarate and nitrate reduction regulatory
HU	Heat unstable
H-NS	Histone-like nucleoid-structuring
HRP	Horseradish peroxidase

IHF	Integration host factor
IT	Ion transporter
MFS	Major facilitator superfamily
MBP	Maltose-binding protein
MR	Mandalate racemase
PBP	Periplasmic binding protein
RNaseE	Ribonuclease E
RNAP	RNA polymerase
SLC17	Solute carrier family 17
βgal	β-D-galactosidase

Miscellaneous

UV	Ultraviolet
N- terminus	Amino-terminus
ABU	Asymptomatic bacteriuria
attλ	Attachment site of phage λ
bp	Base pair
C- terminus	Carboxyl-terminus
dgo	D-galactonate operon
DSSP	Dictionary of protein Secondary Structure
K_D	Dissociation constant
DALI	Distance Alignment Matrix Method
FLAG	DYKDDDDK
E/O	Effector-binding or oligomerization domain

ED	Entner Doudoroff
HTH	Helix-turn-helix
IR	Inverted repeat
LPS	Lipopolysaccharide
LCFA	Long-chain fatty acid
MEP	Methylerythritol phosphate pathway
PTS	Phosphotransferase system
PDB	Protein Data Bank
Pfam	Protein Families
RCSB	Research Collaboratory for Structural Bioinformatics
RBS	Ribosome binding site
rsp	Regulatory in stationary phase
SD	Standard deviation
SCOP	Structural Classification of Proteins
TCA	Tricarboxylic acid cycle
UTRA	UbiC transcription regulator-associated
UPEC	Uropathognenic *E. coli*
WT	Wild type
wHTH	Winged helix-turn-helix

CHAPTER I

Review of Literature

The current work deals with the investigation of molecular and functional details of the regulation of metabolism of D-galactonate, an aldonic sugar acid, by a GntR/FadR-type transcriptional regulator in *E. coli*. Hence in this review of literature, after a brief introduction to *E. coli*, I have highlighted the importance of carbon source metabolism in *E. coli* and have listed the compounds utilized as carbon sources by this bacterium. After discussing the importance and levels of regulation of carbon source metabolism in *E. coli*, I have focused on the GntR family transcriptional regulators and their functions across bacterial species. Description of the GntR family has been made with regard to DNA- and effector-binding characteristics with major focus on the FadR subfamily of the GntR family of transcriptional regulators. Further, I have discussed the general scheme and components involved in regulation of sugar acid transport and metabolism in *E. coli*. Finally, I have reviewed the information available on abundance, physiological importance, transport and metabolism of D-galactonate.

1.1 Introduction

E. coli, a facultative, heterotrophic gram negative bacterium is one of the main components of the microbiota inhabiting the mammalian gut including that of humans. Besides this, *E. coli* spends substantial amount of time in abiotic habitats. Frequently, *E. coli* is one of the first microbes to colonize the mammalian host and maintains a stable life-long association. On an average there are about 5 different strains of *E. coli* inhabiting the human gut (Apperloo-Renkema et al., 1990). Thus for a long term establishment in the gut, the bacterium must compete with other gut colonizers including its other strains. Freter's nutrient-niche hypothesis suggests that during this competition, efficient utilization of at least a single limiting nutrient in the gut can provide competitive edge to the bacteria in colonization (Conway and Cohen, 2015; Freter, 1983, 1988; Freter et al., 1983). Importantly, metabolism of carbon and energy source has been suggested to

be crucial for bacteria to cause infections (Fuchs et al., 2012; Passalacqua et al., 2016). The success of *E. coli* in maintaining such a stable relationship with the host stems from its tremendous potential to utilize alternative carbon and energy sources. Out of the entire *E. coli* gene pool comprising of over 18,000 genes (pan genome), only around 2000 genes (core genome) are shared among multiple *E. coli* strains (Touchon et al., 2009). A comparison of genome-scale metabolic models constructed for fully sequenced 55 *E. coli* and *Shigella* strains, led to characterization of core and pan metabolic capacities of the *E. coli* species. Core constituents reflect the essential components, while the elements in the pan signify the reactions which are either variable or unique among different *E. coli* strains. Maximum diversity in carbohydrate metabolism with its 64% reactions being part of the pan reactome suggests that *E. coli* strains harbor diverse complement of components to metabolize alternate carbon sources (Fig. 1.1). These diverse capabilities may resonate with the importance of alternate carbon metabolism and its contribution in adaption of various *E. coli* strains to different nutritional environments (Monk et al., 2013).

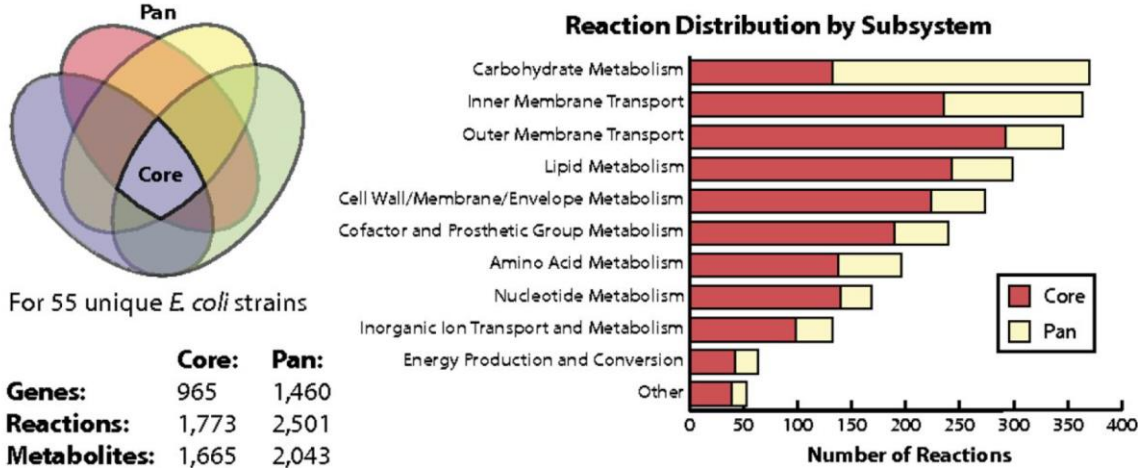

Figure 1.1 Core and pan metabolic reactome of various strains of *E. coli*. The core and pan metabolic reactions of various *E. coli* strains and its functional classification is shown on left and right, respectively (used with permission from (Monk et al., 2013) Proc Natl Acad Sci U S A *110*, 20338-20343. PMID: 24277855, https://doi.org/10.1073/pnas.1307797110)

E. coli faces a constantly changing environment. To successfully colonize the human small intestine, its major niche, following ingestion, the bacteria must resist the acidic environment in the stomach, reach colon, must find and adjust according to the nutrients to snap out of lag phase, compete with resident microbes for limited resources, proliferate and penetrate the mucus layer. The bacteria must grow actively to outpace the mucus turnover rate. As a regular process, some cells are sloughed into the lumen and eliminated in the host faeces to a nutrient-deprived abiotic environment, where again the bacteria must sense and endure a variety of stresses until a new host is found. Since the *E. coli* habitats are distinct in terms of variety of the available carbon sources, the bacterium must continuously sense and adapt its metabolic state according to the prevalent conditions. For this, a precise control of genes involved in uptake and metabolism of nutrients especially carbon sources in the bacterium is essential.

1.2 Carbon sources utilized by *E. coli*

E. coli exhibits a heterotrophic mode of nutrition and hence derives its carbon and energy by catabolism of organic compounds. In its environment, *E. coli* encounters a mélange of chemically diverse carbon sources. For its survival, the bacterium must be able to sense and metabolize a wide range of carbon sources. Multiple factors that influence the utilization of a compound as a sole carbon and energy source involve presence of biologically active metabolic components such as transporters, catalytic enzymes and suitable regulatory components. Differences amongst the bacterial species or strains for these components may account for their differential usage of the various compounds as their sole carbon source. Phenotypic microarray based analyses reported that *E. coli* K-12 is capable of utilizing more than 80 compounds as its sole carbon source (Table 1.1) (Yoon et al., 2012). The same study revealed that the spectrum of compounds that can be used as their sole source of carbon varies among *E. coli* B and K-12 strains.

The list of compounds that can be used as their sole carbon source by *E. coli* K-12 includes chemically diverse species (Table 1.1). These compounds after being transported, either directly or after being metabolized by specific routes into intermediates enter the central metabolic pathways including glycolysis, pentose phosphate pathway, gluconeogenesis, tri-carboxylic acid cycle (TCA) and glyoxylate shunt (Fig. 1.2). Depending upon the pathway and mode of respiration, ATP is generated by either substrate-level or oxidative phosphorylation. Carbohydrates such as sugars and their derivatives either directly or after going through specific metabolic routes enter at the level of glycolysis, Entner Doudoroff (ED) or pentose phosphate pathways. Amino acids undergo deamination or decarboxylation reactions to enter either the lower part of glycolysis or TCA cycle. The acetyl CoA generated from either acetate or degradation of long chain fatty acids through β-oxidation feeds into TCA cycle for metabolism. Another short chain fatty acid, propionic acid is processed through a unique pathway to generate organic acids such as succinate, pyruvate or oxaloacetate, which along with other organic acids such as malic acid, fumaric acid and α-keto-glutaric acid are metabolized in TCA cycle.

Table 1.1 List of compounds used as carbon and energy source by *E. coli* K-12 (Clark and Cronan, 2005; Kreth et al., 2013; Yoon et al., 2012).

Category	**Name of carbon source**
Sugar	α-D-Glucose, D-Ribose, Dextrin, D-Galactose, D-Saccharic acid, L-Fucose, D-Trehalose, L-Galactonic acid-y-lactone, L-Rhamnose, L-Lyxose, D-Mannose, Maltotriose, D-Xylose, Maltose, D-Fructose, β-D-Allose, D-Psicose,
Sugar acid	D-Gluconic acid, D-Glucuronic acid, D-Galacturonic acid, D-Galactonic acid-γ-lactone, Mucic acid, m-Tartaric acid
Sugar alcohol	D-Sorbitol, D-Mannitol
Phosphorylated sugar	D-Fructose-6-phosphate, D-Glucose-6-phosphate, D-Glucose-1-phosphate
Amino sugar	D-Glucosamine
Methylated sugar	α-Methyl-D-galactoside, β-Methyl-D-galactoside, β-Methyl-D-glucoside
Methylated sugar acid	β-Methyl-D-glucuronic acid
Ketosugar acid	5-Keto-D-gluconic acid
Acetylated amino sugar	N-Acetyl-neuraminic acid,
Acetylated hexosamine monosaccharide	N-Acetyl-D-glucosamine, N-Acetyl-β-D-mannosamine
Disaccharide	D-Melibiose, α-D-Lactose, 3-O-β-D-Galacto-pyranosyl-D-arabinose, Lactulose, Melibionic acid
Amino acid	L-Aspartic acid, D-Alanine, L-Serine, D-Serine, L-Alanine, L-Arabinose, L-Asparagine, L-Glutamine, L-Threonine
Dipeptide	Glycyl-L-aspartic Acid, Glycyl-L-glutamic acid, L-Alanyl-glycine, Glycyl-L-proline
Short chain fatty acid	Acetic acid, Propionic acid
Short chain fatty acid derivative	α-Hydroxy butyric acid
Long chain fatty acid	Lauric acid, Myristic acid, Palmitic acid, Stearic acid, Oleic acid
Poly-alcohol	Glycerol
Glycerol derivative	D,L-α-Glycerol- phosphate
Dicarboxylic acid	D,L-Malic acid, Succinic acid, L-Malic acid, Fumaric acid, D-Malic acid
Dicarboxylic acid derivative	Bromo-succinic acid
α-keto acid	α-Keto-glutaric acid, α-Keto-butyric acid, Pyruvic acid

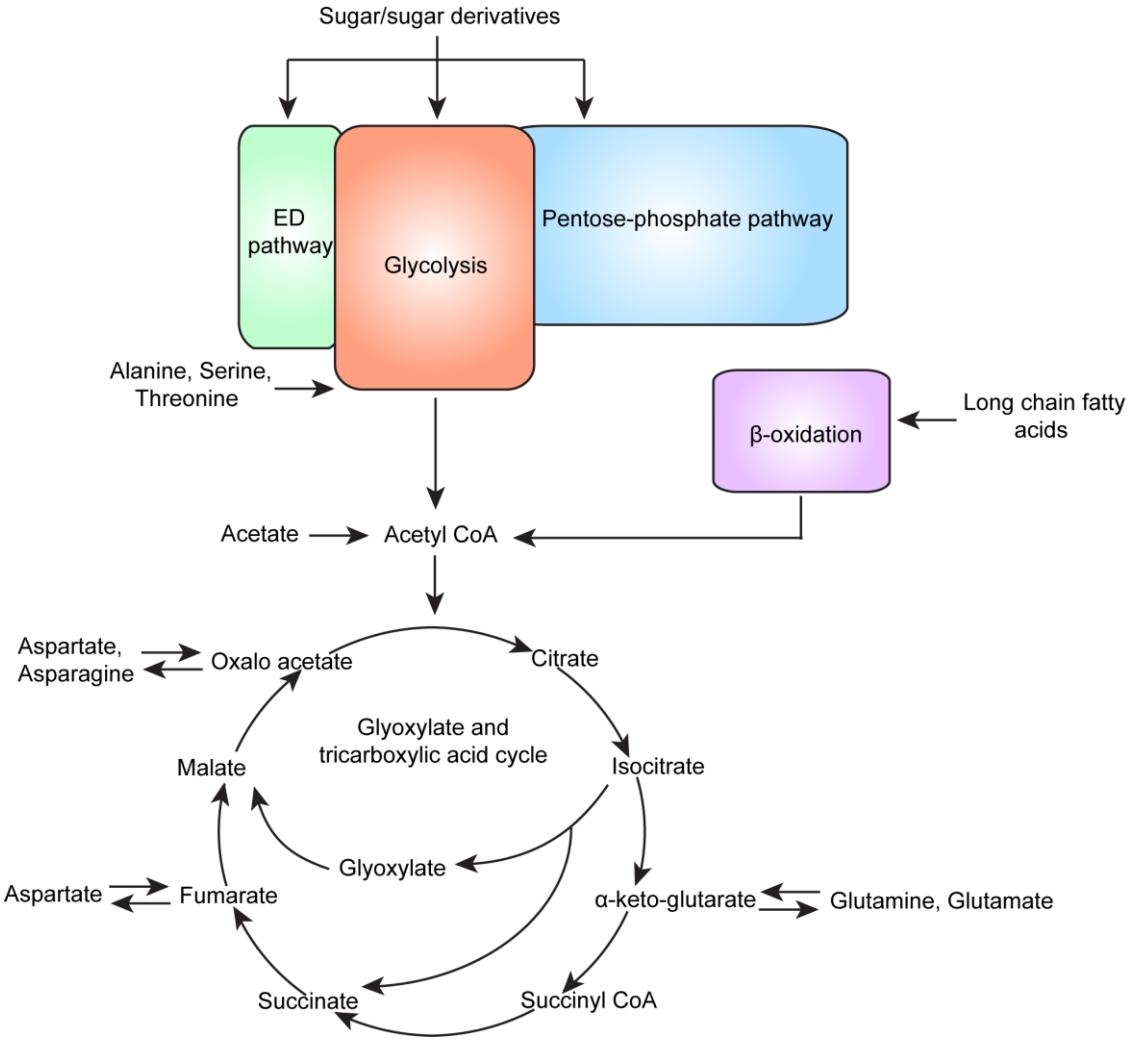

Figure 1.2 Carbon source metabolism in *E. coli* K-12. Various carbon sources enter the central metabolic pathways through indicated entry points.

1.3 Regulation of carbon source metabolism in *E. coli*

In order to survive in environment constantly changing with respect to the carbon sources, bacteria such as *E. coli* must be able to rapidly metabolize a variety of carbon sources. To avoid the cost associated with synthesizing components for catabolizing all the potential carbon sources simultaneously, bacteria have sophisticated regulatory mechanisms to selectively synthesize the catabolic components according to the available carbon source. For example, components for degradation of lactose are synthesized only when this carbon source is present in the environment. These regulatory mechanisms also assist the

bacteria to preferentially utilize a favored carbon source from a mixture. For example, in a medium supplemented with D-glucose and lactose, *E. coli* selectively keeps the synthesis of catabolic components of lactose in an off state to prioritize utilization of the simpler carbon source, D-glucose. Furthermore, reversible nature of these regulatory components allows quick adaptation to the changing availability of the carbon source. Decades of research on *E. coli* has revealed the molecular components and underlying mechanisms for regulation of carbon source metabolism. The regulatory control over carbon source metabolism is exerted at various levels: transcriptional, post-transcriptional, and post-translational (Fig. 1.3). Some of the important mechanisms involved in regulation and control of metabolic activities in *E. coli* are discussed below.

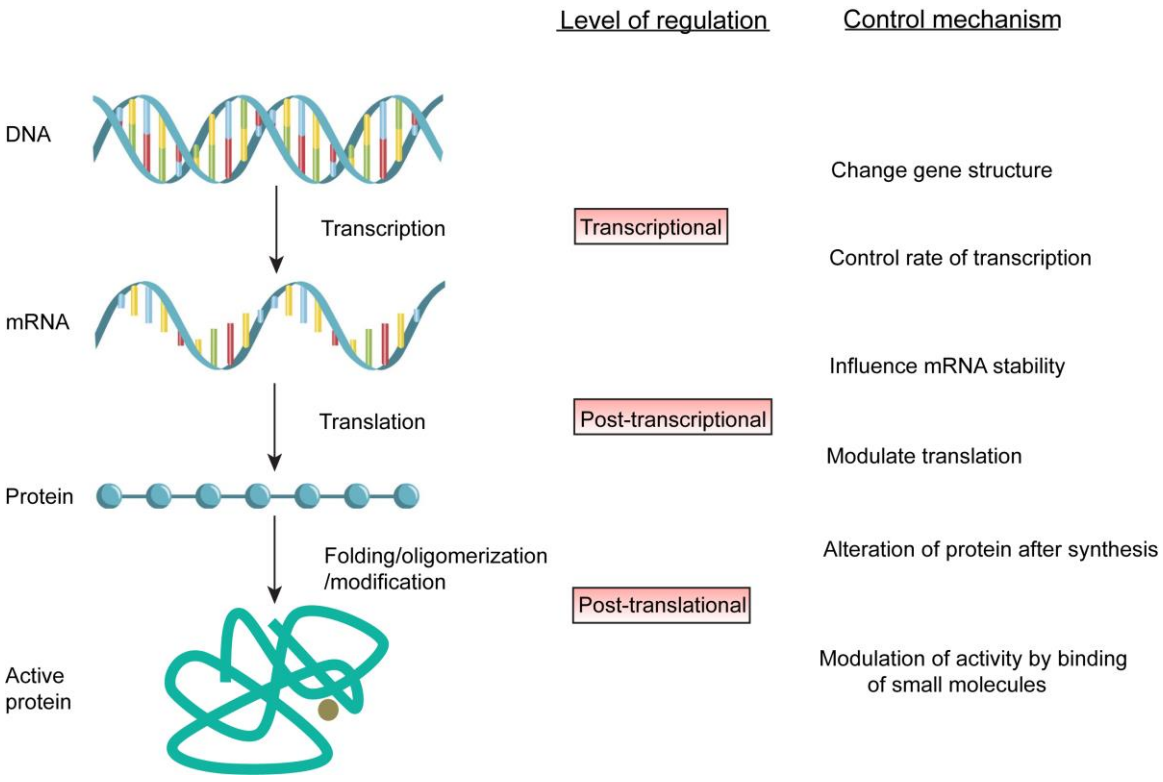

Figure 1.3 Regulation of carbon source metabolism. Metabolism of various carbon sources in bacteria is regulated at different levels such as transcriptional, post-transcription and post-translational through the indicated mechanism.

1.3.1 Regulation at transcriptional level

Non-specific DNA binding proteins such as IHF, HU and H-NS bind the DNA affecting the availability of the promoter for transcription of enzymes and transporters required for metabolism of a particular carbon source. Thus, these proteins play an important role in response of *E. coli* to the availability of carbon sources. For example, *E. coli* O157:H7 mutated for H-NS is unable to utilize 42 carbon sources as compared to its WT counterpart (Erol et al., 2006). H-NS causes condensation of genomic DNA by binding to bent non-specific DNA to affect their availability for transcription (Dame et al., 2000). Similarly HU binds and bends the intervening DNA between two operators individually bound by GalR repressor in upstream region of *gal* operon. Binding of HU provides architectural assistance in DNA loop formation to synergize weaker dimer-dimer interaction between GalR bound to distant operator sites to repress *gal* operon involved in metabolism of D-galactose (Semsey et al., 2002). IHF, another nucleoid associated protein, binds A/T rich sequences to bend the DNA regulating the condensation state of the regulatory sequences (Hales et al., 1996). IHF activates the expression of *glc* operon for metabolism of glycolate by directly binding to region upstream of *glc* operon and *E. coli* lacking IHF are unable to utilize glycolate as sole carbon and energy source (Pellicer et al., 1999). Besides this a wide variety of global and specific transcriptional regulators with a repressive/activating or dual role are involved in the regulation of carbon sources at the transcriptional level. These include global regulators such as cAMP-CRP system, involved in exerting carbon catabolite repression over a number of genes required for the metabolism of other less-preferred carbon sources. cAMP-CRP binds to CRP binding sites upstream of promoters of genes required for metabolism of alternative carbon sources such as D-galactose, to activate their transcription (reviewed in (Kolb et al., 1993)). Phosphotransferase System (PTS) mediated transport of D-glucose lowers cAMP

levels by dampening the activity of adenylate cyclase, thus reducing the transcription from *gal* operon, till D-glucose is exhausted (Ishizuka et al., 1993). Other global regulators such as Cra (Catabolite repressor/activator protein) are involved in regulating the carbon flow through central metabolic pathways (Ramseier et al., 1995). Yet other global transcription regulators such as FNR (Fumarate and nitrate reduction regulatory protein), and ArcB-ArcA (Anoxic redox control protein) two-component system assists in regulating metabolism when the bacteria switches to anaerobic or microaerophilic conditions (Levanon et al., 2005). Besides global regulators, a wide variety of pathway specific transcriptional regulators also regulate expression of metabolic genes. These metabolic transcriptional regulators can exert negative, positive or dual regulation over their regulon members (Perez-Rueda and Collado-Vides, 2000). Most of these regulators bind operator sites present upstream of metabolic genes through their DNA binding domain, which can be present at their N- or C- terminus. Depending upon whether the binding of regulator causes upregulation or downregulation of the downstream genes, they are designated as activators or repressors. However some regulators which can perform both the functions are known as dual regulators. Binding of metabolites to effector binding domains of these regulators reverses their regulatory role allowing them to act as molecular switches to regulate the expression of metabolic genes only when a relevant metabolite is present.

1.3.2 Regulation at post-transcriptional level

After the synthesis of mRNA from the metabolic genes, another layer of regulatory molecules act upon the mRNA to regulate the effective amount of proteins synthesized from them. These components may include small RNAs (sRNA) such as SgrS (Sugar transport-related sRNA) or protein molecules such as CsrA (Carbon storage regulator protein) (Liu et al., 1995; Maki et al., 2010). These regulatory components usually

interact with mRNA affecting the transcriptional termination, initiation of translation, and or its stability. For example, thiamine pyrophosphate (TPP)-sensing riboswitch regulates expression of *thiMD* operon involved in synthesis of thiamine pyrophosphate, an important cofactor for carbohydrate metabolism. Besides, Rho-dependent transcriptional termination, direct binding of TPP stabilizes an aptamer domain, resulting in sequestration of the RBS/AUG thus inhibiting initiation of translation (Bastet et al., 2018; Chauvier et al., 2017). Certain riboswitches such as lysine riboswitch of *lysC* involved in regulation of lysine biosynthesis can simultaneously block translation initiation and modulate transcript decay in an RNaseE dependent manner (Caron et al., 2012).

1.3.3 Regulation at post-translational level

The proteins involved in metabolic pathways are subject to regulation by allostery or covalent modifications such as phosphorylation, succinylation, glutarylation, acetylation and adenylation etc (reviewed in (Macek et al., 2019; Pisithkul et al., 2015). For example, acetylation of Acetyl CoA synthetase plays an important role in assimilation of acetate in *E. coli* (de Diego Puente et al., 2015).

Being the first step in synthesis of protein machinery dedicated to metabolism of carbon sources, a vast majority of regulation is exerted at the transcriptional level as a more metabolic cost-effective measure. Transcriptional regulators belonging to different families carry out the specific regulation of metabolic pathways at the transcriptional level. In our work we have characterized DgoR, proposed to be a member of GntR family of transcriptional regulators, thus below I have discussed the GntR family of regulators in detail.

1.4 GntR family regulators

Transcriptional regulators are recognized and classified according to their conserved DNA-binding motifs e.g. leucine zipper, zinc finger domain, cold shock domain, homeodomain, helix-turn-helix, RNA-binding like motifs, and β-ribbon *etc* (Aravind et al., 2005; Ishihama et al., 2016; Madan Babu and Teichmann, 2003; Pabo and Sauer, 1992; Perez-Rueda and Collado-Vides, 2000). The helix-turn-helix or HTH represents the most extensively studied and well-characterized DNA-binding structural motif. The HTH motif is recruited by a wide range of transcriptional regulators from the three super-kingdoms, bacteria, archaea and eukarya. Pioneering investigations in 1982 identified HTH motif as a tri-helical DNA-binding domain common to the phage lambda transcriptional regulators, cI and cro as well as LacI, the lactose operon regulator (Aravind et al., 2005). Further secondary structure analysis showed that the HTH motif is present in several other bacterial transcriptional regulators such as CRP and sigma factors. The primary HTH motif comprises of three core helices arranged in a right-handed helical bundle, where the second and the third helices are connected by a characteristic sharp turn. The third helix serves as the 'recognition helix' as it is directly inserted into the major grove of DNA and makes the largest contribution in making DNA-protein contacts. The other structural features serve auxiliary roles by making additional contacts or stabilizing the protein fold. Further, the HTH motif adopts diverse structural forms by acquiring a variety of N- and C-terminal extensions which serve as the basis for further classification of HTH clan into SCOP superfamilies (reviewed in (Aravind et al., 2005)). Among these, winged helix-turn-helix (wHTH) forms the second largest superfamily. In wHTH motif the C-terminus of the third helix extends into a β-hairpin unit referred to as 'wings'. As per SCOP classification there are currently 84 families of proteins bearing the wHTH domains including various important prokaryotic

transcriptional regulator families such as ArsR, BirA, CitB, DtxR-FurR, GntR, HrcA-RuvB, IclR, LexA, Lrp/AsnC, LysR MarR, PadR, PurR etc. The simplest version of wHTH motif comprises of the HTH motif followed by a two strand β-hairpin and is present in DNA-binding domains of some of the largest prokaryotic families of transcriptional regulators, including the GntR family, one of the largest and diverse family of transcriptional regulators. The GntR family of transcriptional regulators was identified by Haydon and Guest in early 1990s, where a matrix-based analysis led to identification of transcriptional regulators sharing sequence and structural similarity in their first 69 N-terminal amino acids constituting the DNA-binding wHTH motif (Haydon and Guest, 1991). Thus, a new family of transcriptional regulators having a remarkable similarity in their wHTH motif (recognized as GntR-type wHTH, Pfam 00392) was suggested. The family was named after *Bacillus subtilis* GntR, the most well-characterized member at that time. The amino acid sequence similarity did not extend beyond the N-terminal wHTH domain, suggesting that the C-terminal regions of then identified GntR family members are structurally variable. Afterwards multiple studies involving characterization of GntR family transcriptional regulators showed that the C-terminus domain of GntR family regulators is involved in binding effector molecules which allosterically drive the DNA-binding behavior of the regulator facilitating them to act as molecular switches to regulate the transcription of their regulon members. Additionally, the C-terminal domain is crucial for the GntR family regulators to form functional oligomers necessary for their DNA-binding. Thus, the GntR family regulators adopt a domain based architecture with a much similar GntR-type wHTH DNA-binding domain at the N-terminus and a more diverse effector-binding or oligomerization domain (E/O) at the C-terminus (Fig. 1.4). GntR family regulators share multiple common as well as variable features in terms of binding their cognate DNA or effectors and mode of

regulation in general establishing them as a reference for understanding the general rules governing protein-DNA interactions. Due to their enormous phylogenetic distribution, GntR family serves as a target for evolutionary studies concerned with the origin and evolution of transcriptional regulators. Further, the specific and quick responsive nature combined with their modular architecture makes GntR family of regulators amenable to construction of tailored chimeric regulators, which can bind to desired set of DNA and effectors. This feature offers opportunities to exploit the GntR family of regulators for biotechnological interventions.

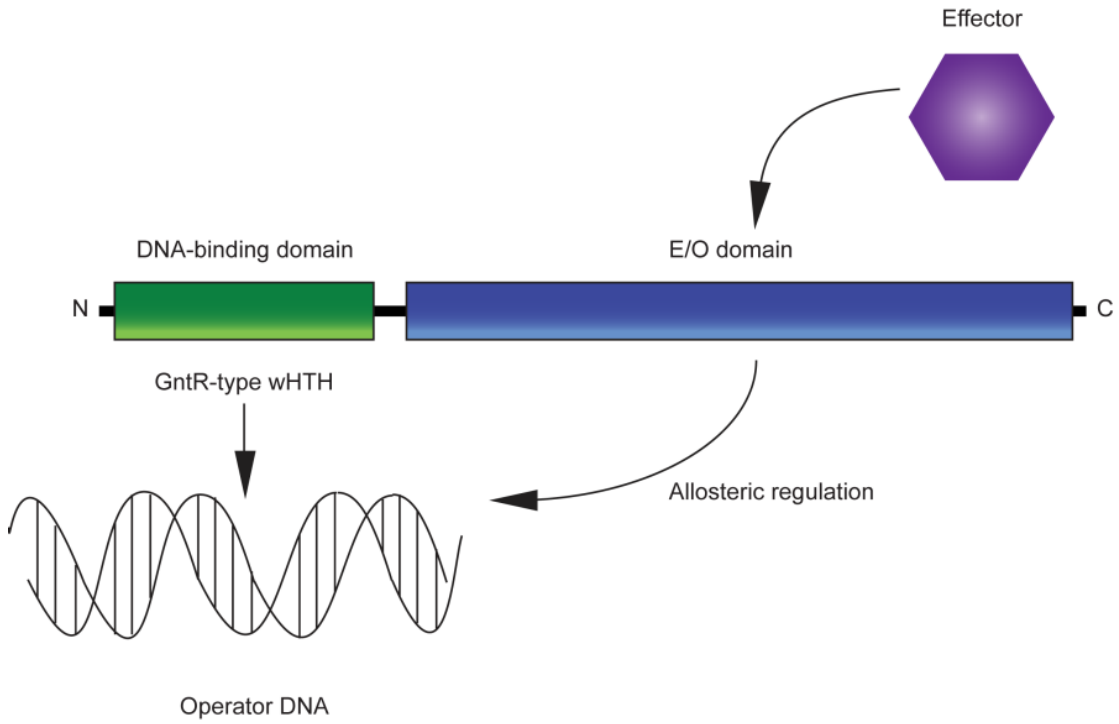

Figure 1.4 Modular domain organization in GntR family transcriptional regulators. The DNA-binding GntR-type wHTH domain is indicated in green while the effector-binding/ oligomerization domain is shown in blue.

1.4.1 Origin and distribution

GntR family constitutes one of the most abundant and diverse groups of HTH superclass. Currently, there are 93,135 sequences in the Pfam database (September, 2019) belonging to this family (El-Gebali et al., 2019) with its members represented in all three super-

kingdoms of life, bacteria, archaea and eukarya suggesting that this fold has been widely recruited for regulatory purposes. Bacteria super-kingdom contributes the bulk of the proteins with GntR type wHTH domain (92,148 sequences among 5,428 species). Of these, the proteobacteria, actinobacteria and firmicutes are the predominant phyla possessing proteins belonging to the GntR family. However, the eukaryotes (83 sequences, 23 species) and archaea (55 sequences, 40 species) have a much smaller contribution to the GntR family (Fig. 1.5). The minor fraction of GntR family proteins in bacteriophages such as *Streptomyces* phage ϕC31, TG1, and enterophage ϕP27 are supposedly derived from the bacterial host organisms. Pan-bacterial and pan-archaeal phylogenetic analysis shows that the GntR family members were present in some of the earlier representatives of these super-kingdoms. The diversity in the functions performed by the extant bacterial and archaeal members indicates that the ancestral forms of this family had a generic nucleic-acid binding role and only secondarily they were engaged as specific transcriptional regulators. Frequent recruitment of the effector-binding/oligomerization domain to the common GntR-type wHTH led to the evolution of various GntR family regulators with diverse functions. The swapping and/or fusion of the C-terminal E/O domains led to diversification of GntR transcriptional regulators into subfamilies for responding to new metabolites and regulating the cellular physiology to adapt to the changing environment. Very early multiple horizontal transfer events led to vast proliferation and sharing of GntR family members in bacteria as transcriptional regulators. Duplication followed by mutations served as another opportunity for evolution of GntR transcriptional regulators with novel regulatory potential. The complement of GntR-like transcriptional regulators possessed by an organism reflects the properties of its ecological niche as well as the manner in which the organism responds to it. For example, the microbes living in a more complex and variable environment possess more diversity

in metabolite-responsive GntR family transcriptional regulators than the obligate intracellular parasites (Hoskisson and Rigali, 2009).

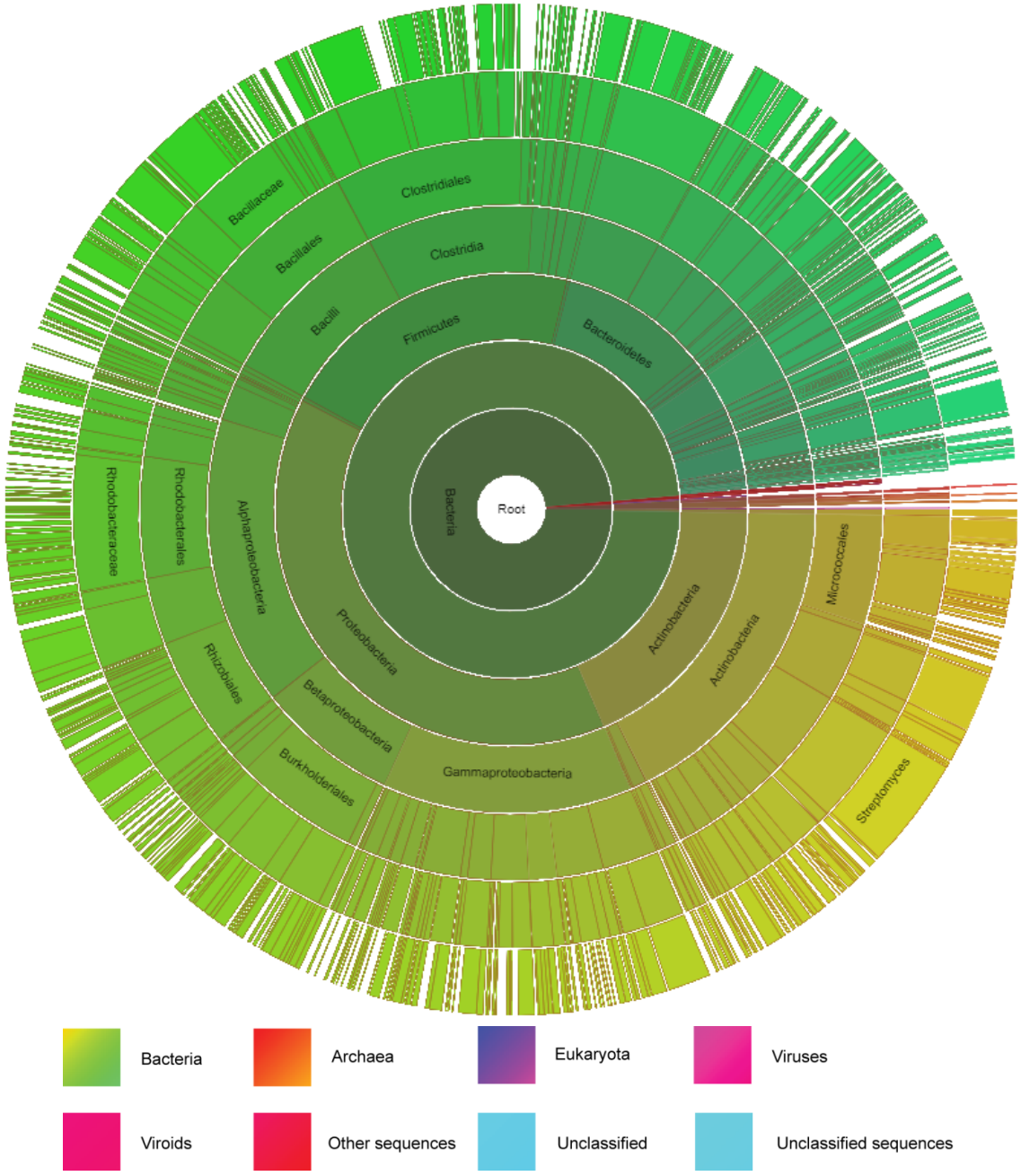

Figure 1.5 Phylogenetic distribution of proteins bearing GntR-type wHTH domain. The chart was prepared with the help of Pfam database (El-Gebali et al., 2019) and represents number of proteins containing GntR-type wHTH domain in each taxa (taxonomic ranks from inside to outside: superkingdom, kingdom, phylum, class, order, family and genus). The length of the arc indicates the number of species in a particular taxa harboring proteins with GntR-type wHTH domain.

1.4.2 Functions of GntR family regulators

The GntR regulators modulate transcription in bacteria in response to a variety of environmental signals, and hence perform an assortment of biological functions. The representative examples of GntR family regulators covering the range of functions performed by this family are tabulated below (Table 1.2).

Table 1.2 Functions performed by various GntR regulators

Metabolism			
Regulator	**Source organism**	**Function**	**Reference**
GntR	*B. subtilis*	D-gluconate metabolism	(Miwa and Fujita, 1988)
HutC	*Pseudomonas putida*	Histidine metabolism	(Allison and Phillips, 1990)
MocR	*Rhizobium meliloti*	Rhizopine metabolism	(Rossbach et al., 1994)
PtsJ	*Salmonella enterica* serovar Typhimurium	Regulates PTS for sugar metabolism	(Titgemeyer et al., 1995)
AraR	*B. subtilis*	L-arabinose metabolism	(Sa-Nogueira and Mota, 1997)
ExuR	*Erwinia chrysanthemi*	D-galacturonate catabolism	(Valmeekam et al., 2001)
LldR	*Corynebacterium glutamicum*	L-lactate and sugar utilization	(Gao et al., 2008)
CitO	*Enterococcus faecalis*	Citrate utilization	(Blancato et al., 2008)
NagQ	*Xanthomonas campestris*	N-acetylglucosamine utilization	(Boulanger et al., 2010)
YvoA	*B. subtilis*	N-acetylglucosamine utilization	(Resch et al., 2010)
MatR	*Rhizobium leguminosarum*	Malonate and propionate catabolism	(Suvorova et al., 2012)
NagR	*B. subtilis*	N-acetylglucosamine	(Fillenberg et al., 2015)
SCO1678	*Streptomyces coelicolor*	D-gluconate metabolism	(Tsypik et al., 2017)
GguR	*Ralstonia picketti* and *Polaromonas*	hexuronate/aldarate utilization	(Bouvier et al., 2019)
Plasmid maintenance and transfer			
Regulator	**Source organism**	**Function**	**Reference**
KorA	*Streptomyces lividans*	Plasmid replication and transfer	(Kendall and Cohen, 1988)
PtrFA	*Frankia*	Plasmid maintenance	(Lavire et al., 2001)
PlmA	*Anabaena* sp. *Strain PCC7120*	Plasmid maintenance	(Lee et al., 2003)
TraR	*Streptomyces venezuelae*	Plasmid transfer	(Reuther et al., 2006)
RctR	*Sinorhizobium meliloti*	Plasmid transfer	(Nogales et al.,

	1021		2013)
Degradation of aromatic/xenobiotic compounds			
Regulator	**Source organism**	**Function**	**Reference**
BphR	*Pseudomonas pseudoalcaligenes*	Biphenyl-related compound	(Watanabe et al., 2000)
PaaX	*Pseudomonas* sp. *Strain Y2*	Styrene catabolism, a toxic aromatic compound	(del Peso-Santos et al., 2006)
MhpR	*P. putida*	Chloroaromatic degradation, predicted	(Kunze et al., 2009)
BphS	*Tn4371 isolated from Ralstonia eutropha*	Biphenyl degradation	(Mouz et al., 1999)
AphS	*Comamonas testosteroni*	Phenol metabolism	(Arai et al., 1999)
Development and sporulation			
Regulator	**Source organism**	**Function**	**Reference**
WhiH	*S. ceolicolor*	Sporulation	(Ryding et al., 1998)
DasR	*Streptomyces griseus*	Morphological differentiation	(Seo et al., 2002)
DevA	*S. coelicolor*	Development	(Hoskisson et al., 2006)
SCO6974	*S. coelicolor*	Myo inositol catabolism and sporulation	(Yu et al., 2015)
Virulence			
Regulator	**Source organism**	**Function**	**Reference**
CpsR	*Streptococcus pneuomniae*	Capsular polysaccharide synthesis and virulence	(Wu et al., 2016)
YtrA	*Xanthomonas citri*	Virulence	(Zhou et al., 2017)
Mce1R	*Mycobacterium tuberculosis*	Virulence	(Casali et al., 2006)
Gnt10	*Brucella melitensis*	Virulence	(Haine et al., 2005)
HpaR1	*Xanthomonas campestris* pv. Campestris	Hypersensitive response and virulence	(An et al., 2011)
Antibiotic production			
Regulator	**Source organism**	**Function**	**Reference**
PtmR1	*Streptomyces platensis*	Production of platensimycin and platencin	(Smanski et al., 2009)
DasR	*Streptomyces cinnamonensis*	Antibiotic production	(Zhang et al., 2016)
SCO6256	*S. coelicolor*	Myoinositol and antibiotic production	(Yu et al., 2016)
Adaptation to stressful environment			
Regulator	**Source organism**	**Function**	**Reference**
DR0265	*Dienococcus radiodurans*	Regulation of radioresistance	(Dulermo et al., 2015)
PMM1637	*Prochlorococcus MED4*	Regulation of nitrogen depletion and high-light stress conditions	(Lambrecht et al., 2018)
BusRTha	*Tetragenococcus halophilus*	glycine betaine ABC transporter system for adapting to saline conditions	(Lin et al., 2019)
Antibiotic resistance			
Regulator	**Source organism**	**Function**	**Reference**
NorG	*Staphylococcus aureus*	Resistance to both quinolones and β-lactams	(Truong-Bolduc and Hooper,

			2007)
YtrA	*B. subtilis*	Cell-wall stress, response to cell wall antibiotics ramoplanin and moenomycin	(Salzberg et al., 2011)
Ms0535	*Mycobacterium smegmatis*	Isoniazid resistance	(Hu et al., 2015)
Rv1152	*M. tuberculosis*	Vancomycin resistance	(Zeng et al., 2016)
Biofilm formation			
Regulator	**Source organism**	**Function**	**Reference**
EbrA	*E. faecalis*	Biofilm formation	(Ballering et al., 2009)
LbrA	*Listeria monocytogenes*	Biofilm formation	(Wassinger et al., 2013)
McbR	*E. coli*	Biofilm formation	(Lord et al., 2014)
StsR	*Streptococcus mutans*	Biofilm formation	(Lin et al., 2019)
Ms5576	*M. smegmatis*	Biofilm formation and mannitol metabolism	(Hu et al., 2018)
Motility			
Name	**Source organism**	**Function**	**Reference**
swrZ	*Vibrio parahaemolyticus*	Motility	(Jaques and McCarter, 2006)

Although the GntR family members are involved in performing diverse functions, in *E. coli* majority of these are involved in regulation of carbon source metabolism (Ishihama et al., 2016) (Table 1.3).

Table 1.3 GntR regulators in *E. coli* and their functions

Protein name	**Function**	**Reference**
CsiR	Regulates the activity of *csiD* promoter during early stationary phase and potentially regulates catabolismof γ-aminobutyrate (GABA) and accumulation of glutamate for general stress adaptation	(Metzner et al., 2004)
DgoR	Negative regulator of D-galactonate metabolism	(Singh et al., 2019)
ExuR	Negatively regulates metabolism and transcription of D-galactouronate and D-glucouronate	(Robert-Baudouy et al., 1981; Tutukina et al., 2016a)
FadR	Negatively regulates fatty acid degradation, positively upregulates transcription genes of unsaturated fatty acid biosynthesis, and also regulates acetate metabolism through IclR	(Campbell and Cronan, 2001; DiRusso et al., 1992; Maloy and Nunn, 1981)
FrlR	Putative regulator of utilization of fructoselysine	(Wiame et al., 2002)
GlcC	Positively regulates metabolism of glycolate	(Pellicer et al., 1999)
LgoR	Plays essential role in L-galactonate metabolism	(Reed et al., 2006)
LldR	Dual regulator of L-lactate metabolism	(Aguilera et al., 2008; Dong et al., 1993)

McbR	Negatively regulates biofilm formation and mucoidity	(Lord et al., 2014; Zhang et al., 2008)
MngR	Negatively regulates 2-O-α-mannosyl-D-glycerate utilization and has been shown to be important in responding to heat shock	(Krisko et al., 2014; Sampaio et al., 2004)
NanR	Negatively regulates transport and metabolism of N-acetyl-neuraminic acid (or sialic acid)	(Kalivoda et al., 2003)
PdhR	Negatively regulates genes of pyruvate dehydrogenase complex	(Haydon et al., 1993)
PhnF	Negatively regulates metabolism of alkyl-phosphonate	(Hove-Jensen et al., 2010)
UxuR	Negatively regulates metabolism of β-D-glucuronides, D-glucuronate, and D-gluconate	(Bates Utz et al., 2004; Ritzenthaler and Mata-Gilsinger, 1982; Rodionov et al., 2000; Tutukina et al., 2016a)
YdfH	Negatively regulates *rsp* operon (Regulatory in stationary phase)	(Sakihama et al., 2012)
YieP	Involved in tolerance to 3-hydroxypropionic acid (3-HP) and proposed to regulate cellular membrane synthesis or structure, stress responses	(Gao et al., 2018; Nguyen-Vo et al., 2019)
YdcR	Uncharacterized transcriptional regulator	
YegW	Uncharacterized transcriptional regulator	
YidP	Uncharacterized transcriptional regulator	
YihL	Uncharacterized transcriptional regulator	
YjiR	Uncharacterized transcriptional regulator	

1.4.3 Domain architecture: classification into subfamilies

The pioneering work by Haydon and Guest in 1991 suggested that the GntR family regulators share a N-terminal wHTH domain with amino acid sequence similarity in their N-terminal 69 amino acids (Haydon and Guest, 1991). Further, matrix-based similarity search in 23,196 proteins across the protein length found additional 5 proteins with GntR like sequence. Multiple sequence alignment for these proteins showed that the similarity did not extend beyond first 69 amino acid residues. Several studies involving individual GntR family regulators suggested the importance of C-terminal E/O region in binding effectors and oligomerization. About a decade later, a study made a rigorous comparison of structural, phylogenetic and functional aspects of nearly 270 members of GntR family. Full-length multiple sequence alignment of these regulators was used to construct a phylogenetic tree clearly showing four clades heterogeneous in their C-terminal E/O domain. Prediction of secondary structure revealed that the different clades possessed

discrete structural topologies of the C-terminal domain. Subsequently, the members of each clade based on their similar E/O domain architecture were classified into four distinct subfamilies: FadR, HutC, MocR and YtrA (Rigali et al., 2002). Later, three additional subfamilies were identified viz. AraR, PlmA, and DevA (Franco et al., 2006; Hoskisson et al., 2006; Lee et al., 2003). The detailed features of each subfamily with regard to their nomenclature, abundance, solved crystal structures, topology of E/O domain and general functions are discussed.

1.4.3.1 FadR subfamily

FadR subfamily named after FadR, the dual regulator of long chain fatty acid metabolism in *E. coli*, is the most represented (~40% of the total family members) (Rigali et al., 2002). The FadR subfamily members are characterized by the presence of an all α-helical C-terminal E/O domain (Fig. 1.6) (Jain, 2015; Rigali et al., 2002). Based on the presence of either seven or six helices in the C-terminal domain, the subfamily is further divided into two subfamilies, FadR and VanR respectively. Besides this, an alternative classification involving members with a longer and comparatively variable C-terminal domain (FCD, PF07729) or a shorter and relatively similar C-terminal domain (FadR_C, PF07840) have been suggested (Zheng et al., 2009). The α-helices are organized in similar antiparallel orientation to form a bundle (Fig. 1.6). Most of the regulators of FadR subfamily are involved in regulating metabolism of oxidized substrates used as carbon sources such as sugar acids (UxuR, D-glucuronate), organic acids (PdhR, Pyruvate), and amino acid (AnsR, Aspartate) *etc* (reviewed in (Hoskisson and Rigali, 2009)). FadR from *E. coli*, *V. cholerae* and *Vibrio alginolyticus* represent the well-characterized members of the subfamily with their crystal structures solved in apo-, DNA- and/or effector- bound states (Gao et al., 2017; Shi et al., 2015; van Aalten et al., 2001; van Aalten et al., 2000; Xu et al., 2001). Additional structures for nine FadR subfamily members have been

solved (Table 1.4). Multiple individual biochemical studies have yielded information on the cognate operators and effectors as well as their mode of interaction with these regulatory components.

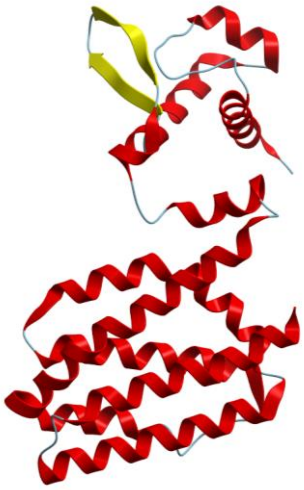

Figure 1.6 Solved structure of FadR. Solved crystal structure of FadR from *E. coli* (PDB ID: 1HW1) (Xu et al., 2001) is shown. The α-helices are shown in red, while β-strands are highlighted in yellow. The image was created by using ICM browser (www.molsoft.com).

1.4.3.2 HutC/FarR subfamily

HutC subfamily comprises the second highest fraction (~25-30% of the total GntR family members) of the GntR family regulators (Jain, 2015; Suvorova et al., 2015). The namesake member HutC negatively regulates histidine utilization in *P. putida* (Allison and Phillips, 1990). The C-terminal E/O domain with an average length of 170 amino acids harbors both α/β secondary structures (Fig. 1.7). C-terminal E/O domain of HutC subfamily members have a structural homology to chorismate lyase fold (Pfam PF04345 Chor_lyase) found in UbiC of *E. coli* now termed as UTRA (UbiC Transcription Regulator-Associated) domain (Pfam PF07702 UTRA) (Aravind and Anantharaman, 2003). Later, crystal structures solved for multiple members of the subfamily

demonstrated the presence of chorismate lyase-like fold with three α-helices surrounding a core of six-stranded β-sheets arranged in anti-parallel orientation (Fig. 1.7). Solved crystal structures of YvoA from *B. subtilis* in apo- and effector- bound states have illustrated the effector binding mode of the regulator (Fillenberg et al., 2015; Resch et al., 2010). Crystal structures of other HutC subfamily members have been solved and are listed in Table 1.4. HutC subfamily members are involved in diverse functions such as antibiotic production, conjugative plasmid transfer (KorA and TraR), regulation of sialic acid metabolism (NagR, DasR, and NagQ) *etc* (reviewed in (Hoskisson and Rigali, 2009)).

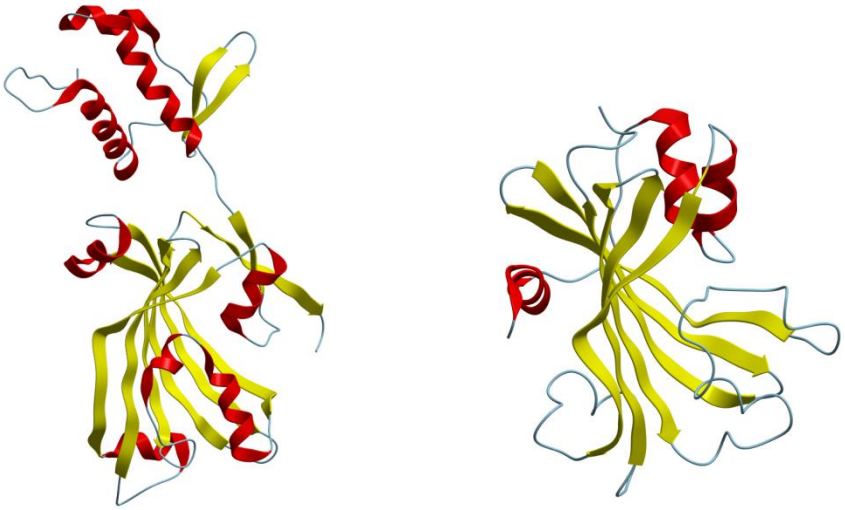

Figure 1.7 Solved structure of YvoA. Solved crystal structure of YvoA from *B. subtilis* (PDB ID:4U0V) (Fillenberg et al., 2015) is shown on the left while the structure of UbiC (PDB ID:1TT8) with chorismate lyase fold (Smith et al., 2006) is shown on the right. The α-helices are shown in red, while β-strands are highlighted in yellow. The image was created by using ICM browser (www.molsoft.com).

1.4.3.3 MocR/GabR subfamily

MocR subfamily takes its name from MocR, transcriptional regulator of rhizopine metabolism in *R. meliloti* (Rossbach et al., 1994). Subfamily members possess a noticeable longer C-terminal domain of about ~300-350 amino acid residues. The E/O domain bears a remarkable homology to aminotransferases class I (Pfam PF00155 Aminotran_1_2) domain (Fig. 1.8). The aminotransferases containing the similar domain are known to catalyse enzymatic reactions using PLP as a cofactor. Solved structure of GabR from *B. subtilis* shows that regulators bind PLP molecule covalently (Edayathumangalam et al., 2013; Okuda et al., 2015; Park et al., 2017; Wu et al., 2017) (Table 1.4). The structure also shows a dimeric arrangement in a head-to-tail configuration suited to bind direct repeats on DNA (Discussed later). Based on the similarity of their E/O fold with corresponding folds from different families of type-1 PLP dependent enzymes the MocR subfamily members have been characterized into three subgroups, GabR, PtsJ and EnuR (Stefanovski et al., 2019). The MocR subfamily regulators are known to regulate reactions requiring PLP as a cofactor, for example, TauR from *Rhodobacter capsulatus*, GabR in *B. subtilis*, PtsJ in *S. enterica* serovar Typhimurium and PdxR from *Bacillus clausii* (reviewed in (Hoskisson and Rigali, 2009)).

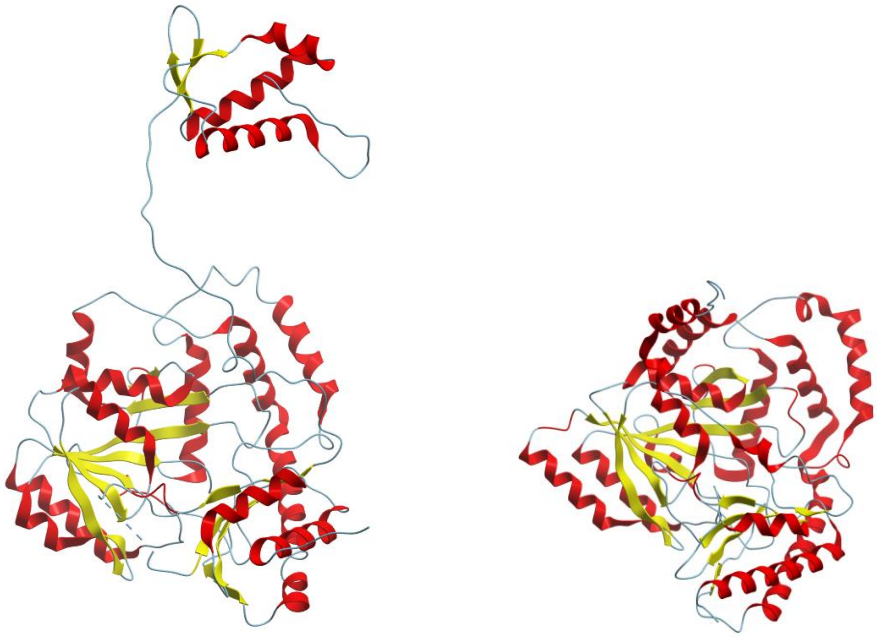

Figure 1.8 Crystal structure of GabR. Image on the left shows solved crystal structure of GabR from *B. subtilis* (PDB ID: 4N0B) (Edayathumangalam et al., 2013) while the on right solved crystal structure of putative aminotransferase (Tm1131) of *Thermotoga maritima* (PDB ID: 1VP4) (unpublished, deposited to RCSB by Joint Center for Structural Genomics (JCSG)) is shown. The α-helices are shown in red, while β-strands are highlighted in yellow. The image was created by using ICM browser (www.molsoft.com).

1.4.3.4 YtrA subfamily

YtrA subfamily is named after the transcriptional regulator involved in regulation of acetoin utilization in *B. subtilis* and represents about 6% of the GntR family regulators (Suvorova et al., 2015). The unusually shorter C-terminal E/O domain of an average 50 amino acid residues comprises of two short α-helices with a 'fish-hook' arrangement (Gao et al., 2007) (Fig. 1.9). Most of the YtrA subfamily regulators are involved in regulation of ATP-binding cassette (ABC) transport systems (Hoskisson and Rigali, 2009). Crystal structures of multiple YtrA family members have been determined (Table 1.4).

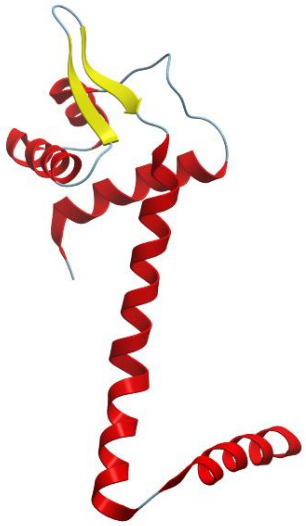

Figure 1.9 Crystal structure of CGL2947. The solved structure of CGL2947 from *C. glutamicum* (PDB ID: 2EK5) is shown (Gao et al., 2007). The α-helices are shown in red, while β-strands are highlighted in yellow. The image was created by using ICM browser (www.molsoft.com).

1.4.3.5 AraR subfamily

AraR is another minor subfamily of GntR family, with its namesake member involved in regulating metabolism of multiple carbon sources in *B. subtilis* (Sa-Nogueira and Mota, 1997). The subfamily members represent a peculiar case with a GntR-type N-terminal wHTH and a C-terminal E/O domain highly similar to the periplasmic binding protein (PBP) fold of LacI family. AraR from *B. subtilis* is the only well-studied member with multiple crystal structures solved for N-terminal DNA binding domain in complex with DNA and C-terminal E/O domain with L-arabinose (Jain and Nair, 2013; Jain et al., 2016). The C-terminal E/O domain has two structural subdomains with α/β topologies where two α helices surround a core of six parallel β-strands (Fig. 1.10).

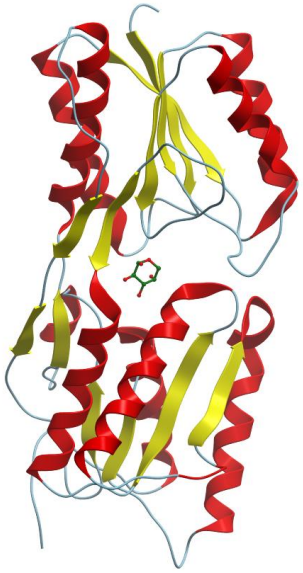

Figure 1.10 Crystal structure of C-terminal E/O domain of AraR. The solved structure of C-terminal E/O domain of AraR in complex with its effector, L-arabinose (PDB ID: 3TB6) (ball and stick representation) is shown (Prochazkova et al., 2012). The α-helices are shown in red, while β-strands are highlighted in yellow. The image was created by using ICM browser (www.molsoft.com).

1.4.3.6 PlmA subfamily

PlmA is another small group of GntR-type regulators with its namesake member involved in regulating maintenance of plasmid in *Anabaena* (Lee et al., 2003). The regulator was characterized as a member with a distinct domain. The regulators seem to be present only in cyanobacteria. The PlmA from *Synechococcus elongatus* was shown to interact with PII-PipX complex to regulate nitrogen assimilation by a novel mechanism (Labella et al., 2016). Till date, no structure for any of the subfamily member is available.

1.4.3.7 DevA subfamily

With its only characterized member, DevA, this subfamily represents another minor subfamily. DevA was identified as a GntR-type transcriptional regulator with a novel E/O domain required for proper development of *S. coelicolor* (Hoskisson et al., 2006). Based

on the similarity, DevA subfamily members have been proposed to have originated from HutC subfamily (Tsypik et al., 2016). Currently, no solved structures of any of the subfamily member are available.

Table 1.4 GntR family members with solved structures. 'YtrA*' indicates that the structure of the regulator was deposited in RCSB and assigned to GntR family, but not to any subfamily. However, from the structure they appear to be members of YtrA family. 'RCSB' in the reference column indicates that the information has been taken from structure submitted to RCSB and it does not accompany any published research article.

Protein name	**Organism**	**CTD subfamily**	**PDB ID**	**Description**	**Reference**
$FadR_{EC}$	*E. coli*	FadR	1H9T	DNA/Full length	(van Aalten et al., 2001)
			1H9G	Effector/Full length	
			1HW1	Apo/ Full length	(Xu et al., 2001)
			1HW2	DNA/ Full length	
			1E2X	Apo/ Full length	(van Aalten et al., 2000)
$FadR_{VC}$	*V. cholerae*	FadR	4PDK	Effector/Full length	(Shi et al., 2015)
			4P96	Apo/ Full length	
			4P9U	DNA/Full length	
$FadR_{VA}$	*Vibrio alginolyticus*	FadR	5XGF	Apo/Full length	(Gao et al., 2017)
			5DV5	Effector/Full length	
GntR	*Streptococcus agalactiae*	FadR	6AZ6	Apo/Full length	(Little et al., 2018)
LldR (CGL2915)	*C. glutamicum*	FadR	2DI3	Metal ion /Full length	(Gao et al., 2008)

McbR/YncC	*E. coli* UMEA 3718-1	FadR	4P9F	Apo/ Full length	(Lord et al., 2014)
PdhR	*E. coli* CFT073	FadR	5KVR	Apo/ Full length	RCSB
PdhR	*E. coli K-12*	FadR	5TPM	Apo/ CTD	RCSB
MouR	*L. monocytogenes*	FadR	6EP3	Apo/ Full length	(Pinheiro et al., 2018)
TM0439	*Thermotoga maritima*	FadR	3SXY	Apo/ Full length	RCSB
			3FMS	Apo/ Full length	(Zheng et al., 2009)
			3SXM	Metal-free FCD	RCSB
			3SXZ	Metal-free FCD	RCSB
			3SXK	Metal bound FCD	RCSB
RHA1_ro03477	*Rhodococcus* sp. RHA1	FadR	2HS5	Apo/ Full length	RCSB
PS5454	*Pseudomonas syringae* pv. tomato str. DC3000	FadR	3C7J	Apo/ Full length	RCSB
YP_298823.1	*R. eutropha* JMP134	FadR	3IHU	Apo/ Full length	RCSB
YvoA	*B. subtilis*	HutC	4U0V	Effector/Full length	(Fillenberg et al., 2015)
			4U0W	Effector/Full length	
			4U0Y	DNA/ Full length, DNA/ DBD	
			4WWC	DNA/ Full length, DNA/ DBD	

			2WV0	Apo/ Full length	(Resch et al., 2010)
DasR	*S. coelicolor*	HutC	4ZS8	Apo/ Full length	(Fillenberg et al., 2016)
			4ZSB	Apo/ CTD	
			4ZSI	Effector/ CTD	
			4ZSK	Effector/ Apo/ CTD	
PhnF	*E. coli*	HutC	2FA1	Apo/ CTD	(Gorelik et al., 2006)
PhnF	*M. smegmatis*	HutC	3F8M	Apo/ Full length	RCSB
			3F8L	Apo/ CTD	RCSB
TraR	*Streptomyces phaeochromogenes*	HutC	1V4R	Apo/DBD	RCSB
YydK	*B. subtilis*	HutC	3BWG	Apo/ Full length	RCSB
TreR	*.B subtilis*	HutC	2OGG	Apo/ CTD	(Rezacova et al., 2007)
YurK	*B. subtilis*	HutC	2IKK	Apo/ CTD	RCSB
HutC	*P. syringae* pv. tomato str. DC3000	HutC	2PKH	Apo/ CTD	RCSB
EF_1328	*E. faecalis V583*	HutC	3DDV	Apo/ Full length	RCSB
Lin2111	*Listeria innocua* Clip11262	HutC	3EDP	Apo/ Full length	RCSB
SAVERM_3189	*Streptomyces avermitilis*	HutC	3EET	Apo/ Full length	RCSB
SCO6256	*S. coelicolor*	HutC	2RA5	Apo/ Full length	RCSB

GBAA_3458	*Bacillus anthracis str. Sterne*	HutC	3LHE	Apo/ CTD	RCSB
BCE_3424	*Bacillus cereus*	HutC	3L5Z	Apo/ Full length	RCSB
Cgl0157	*C. glutamicum*	HutC	2P19	Apo/ Full length	RCSB
SA0254	*S. aureus subsp.* aureus *N315*	HutC	2OOI	Apo/ Full length	RCSB
GabR	*B. subtilis*	MocR	4MGR	Apo/ Full length	(Edayathumangalam et al., 2013)
			4N0B	Effector/ Full length	
			4TV7	Apo/ Full length	(Okuda et al., 2015)
YtrA	*Sulfolobus acidocaldarius*	YtrA	6SBS	Apo/ Full length	(Lemmens et al., 2019)
OEOE_1803	*Oenococcus oeni*	YtrA	3BY6	Apo/ Full length	RCSB
CBU_0775	*Coxiella burnetii*	YtrA*	3TQN	Apo/ Full length	(Franklin et al., 2015)
BT_1304	*Bacteroides thetaiotaomicron*	YtrA*	3IC7	Apo/ Full length	RCSB
Cgl2947	*C. glutamicum*	YtrA	2EK5	Apo/ Full length	(Gao et al., 2007)
			2DU9	Effector/ Full length	
Lmo2241	*L. monocytogenes* serovar 1/2a	YtrA*	4HAM	Apo/ Full length	RCSB
Lmo0741	*L. monocytogenes* serovar 1/2a	YtrA*	4R1H	Apo/ Full length	RCSB
Lin1836	*L. innocua* Clip11262	YtrA*	3NEU	Apo/ Full length	RCSB
AraR	*B. subtilis*	AraR	5D4R	DNA/ DBD	(Jain et al., 2016)
			5D4S	DNA/ DBD	

			4EGY	DNA/ DBD	(Jain and Nair, 2013)
			4EGZ	DNA/ DBD	
			4H0E	DNA/ NTD	
			3TB6	Effector/ CTD	(Prochazkova et al., 2012)

1.4.4 Oligomerization

Generally, GntR family regulators bind DNA with inverted repeats, although some exceptions including cases of binding direct repeats or DNA with no symmetry have also been observed (Hoskisson and Rigali, 2009; Rigali et al., 2002; Suvorova et al., 2015). Often oligomerization or conformational changes induced by binding of effector molecules allows the proper orientation of the wHTH domain for efficient binding of the cognate operator DNA. Majority of GntR family regulators exist as homodimers where each monomer recognizes one half of the 2-fold symmetric inverted or direct repeats. Multiple reports suggest that the C-terminal domain participates in dimerization although some of the regions in DNA binding domain may also contribute in forming the dimeric interface (Edayathumangalam et al., 2013; Gao et al., 2008; van Aalten et al., 2000). In FadR subfamily, the regulators belonging to FadR subgroup form dimers where besides the all-helical C-terminal E/O domain, α3 from the DNA binding domain as well as the connecting α4 linker also contributes to formation of dimers (van Aalten et al., 2000). Although the FadR subfamily members form parallel dimers, LldR from *C. glutamicum*, a member of the FadR subgroup presents as interesting case of cross dimers formed by swapping wHTH domains between the two monomers (Gao et al., 2008). In case of McbR from *E. coli*, a VanR subgroup member, the N-terminal α1 helix of wHTH domain extends close to the C-terminal E/O domain (Lord et al., 2014), although this feature is not exhibited by other VanR subgroup members such as TM0439 from *T. maritima* (PDB

ID: 3SXY, unpublished, deposited to RCSB by Czelakowski, G.P et al, 2011) and MouR from *L. monocytogenes* (Pinheiro et al., 2018). The oligomeric arrangement of the FadR subfamily members enables them to contact their cognate DNA with inverted repeats (Gao et al., 2017; Jain, 2015; Shi et al., 2015; Suvorova et al., 2015; van Aalten et al., 2001) (Fig. 1.11 and Fig. 1.12). Members from other subfamilies such as HutC and YtrA which also predominantly bind inverted repeats in DNA adopt similar arrangement. However, the MocR subfamily members such as GabR from *B. subtilis* dimerize in head-to-tail fashion likely to bind their cognate DNA with largely spaced direct repeats by looping (Edayathumangalam et al., 2013). Members from HutC subfamily, for example, YvoA from *B. subtilis* also dimerize in solution through their C-terminal E/O domain (Fillenberg et al., 2015), although TreR from *B. subtilis* has been shown to exhibit a unique tetrameric assembly (Rezacova et al., 2007). Solved structures of various YtrA subfamily members such as CGL2947 from *C. glutamicum*, also show dimeric assembly formed by inter-wining of the short C-terminal domain comprising of two 'fish-hook' shaped α-helices (Gao et al., 2007). Although the full length structure of any AraR subfamily member has not been solved, the structure of C-terminal E/O domain of AraR from *B. subtilis* suggests that in contrast to the parallel orientation of E/O domains of homologs belonging to the LacI family, the AraR forms dimers in which monomers are tilted at an angle of 40° rendering the regulator unsuited to binding the DNA (Prochazkova et al., 2012). Full-length structure of AraR in complex with its operator DNA will be required to understand the mode of oligomerization favourable for binding its cognate DNA. Oligomeric status of other subfamilies PlmA and DevA is yet to be explored.

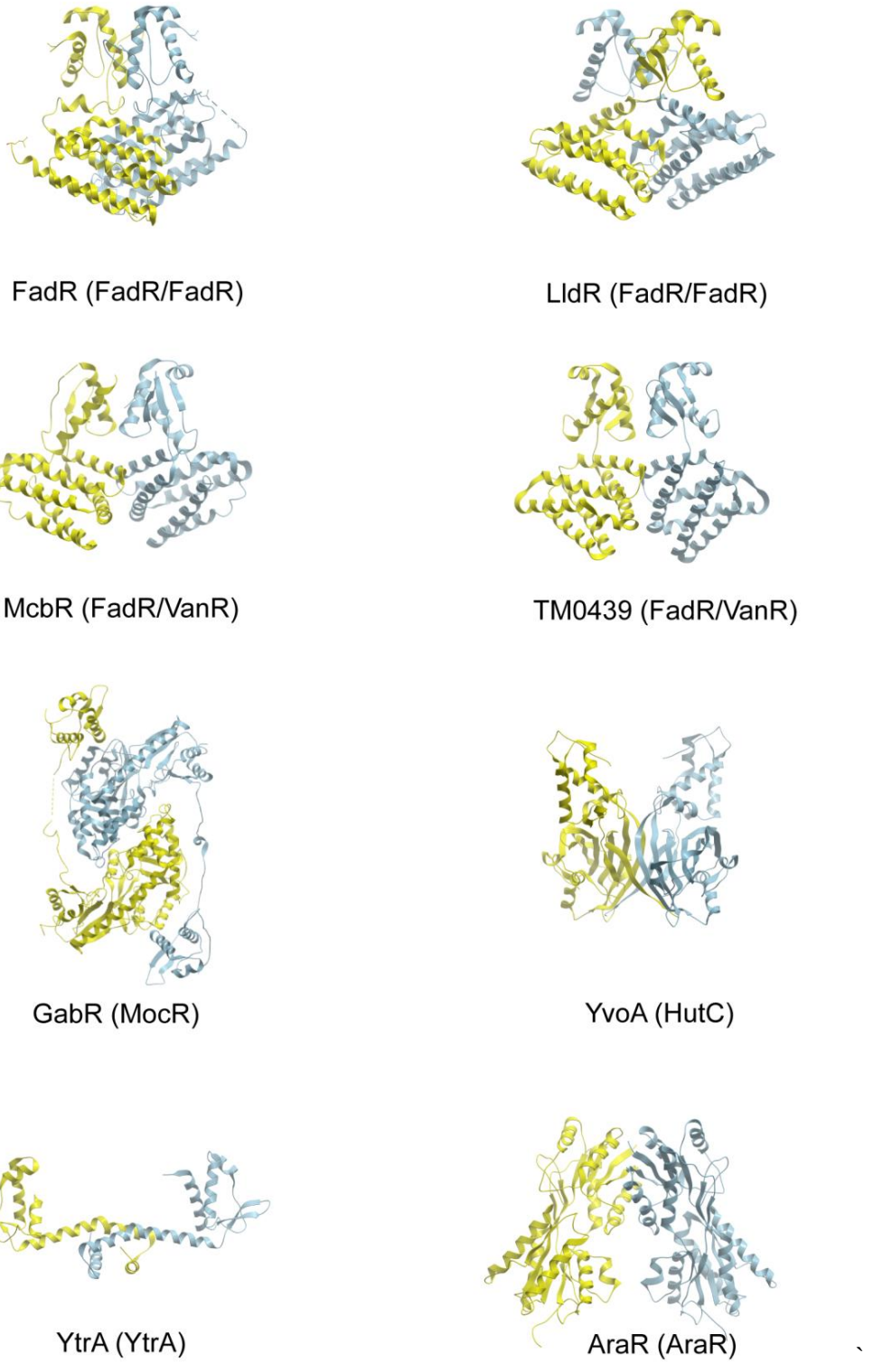

Figure 1.11 Oligomerization in the GntR family. The diverse arrangement of monomers in dimer of GntR family regulators belonging to different subfamilies is shown. The individual monomers of each dimer are colored yellow and blue. The image was created by using ICM browser (www.molsoft.com).

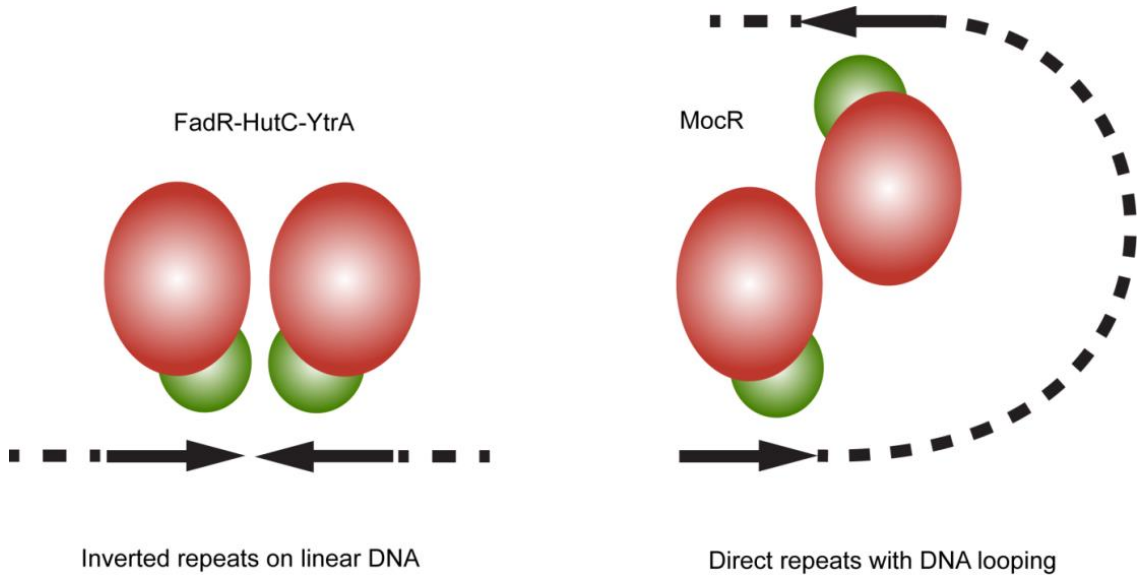

Figure 1.12 Oligomeric arrangement and mode of DNA-binding in GntR family members. The oligomeric arrangement of FadR, HutC, and YtrA members involved in binding inverted repeats in DNA is compared with that of MocR family, known to bind direct repeats. The N-terminal DNA-binding and C-terminal effector-binding domains are shown in green and red respectively. The arrows indicate DNA motifs recognized by the regulators. The image has been modified from (Rigali et al., 2002).

1.4.5 DNA binding characteristics

The N-terminal wHTH domain of GntR family regulators recognizes operator DNA harboring distinct topologies. Although the C-terminal domain is not involved in binding DNA, its oligomerization imposes steric constraints on the wHTH domain thus affecting the spatial orientation of the structural elements involved in interaction with DNA (Rigali et al., 2002; Suvorova et al., 2015). Despite a very low conservation at sequence level (average amino acid sequence identity ~25%), structural topology of the wHTH domain of the whole GntR family is remarkably similar (Fig. 1.13) (Jain, 2015; Rigali et al., 2002). The GntR-type wHTH domain (recognized as PF00392 in the Pfam database) comprises of three α-helices and two very short β-strands in α1-α2-α3-β1-β2 topology as observed from the solved structure for FadR from *E. coli* (Fig. 1.13) (van Aalten et al.,

2000). The three helices fold around a central hydrophobic core to form a tri-helical bundle where the α2 and α3 helices connected by a turn constitute the HTH motif while the two β-strands form a short 'wing'. The winged-HTH motif makes extensive contact with the operator DNA while the α1 helix extends stability to the overall fold. The relative positions of the secondary structural elements are conserved throughout the GntR family with a few anomalies such as absence of α1 (EmoR from BNC1, an EDTA-degrading bacterium and NtaR from *Chelatobacter heintzii*), presence of additional α-helix before α1 (PdxR from *S. venezuelae* and WhiH from *Streptomyces qureofacients* suggested from their predicted secondary structures) (Rigali et al., 2002). An additional β' strand after α1 is occasionally observed such as in the solved crystal structures of GabR from *B. subtilis*, and CGL2947 from *C. glutamicum*. The additional β' strand forms a β sheet by interacting with β2 strand of the wing (Fig. 1.13) (Edayathumangalam et al., 2013; Gao et al., 2007).

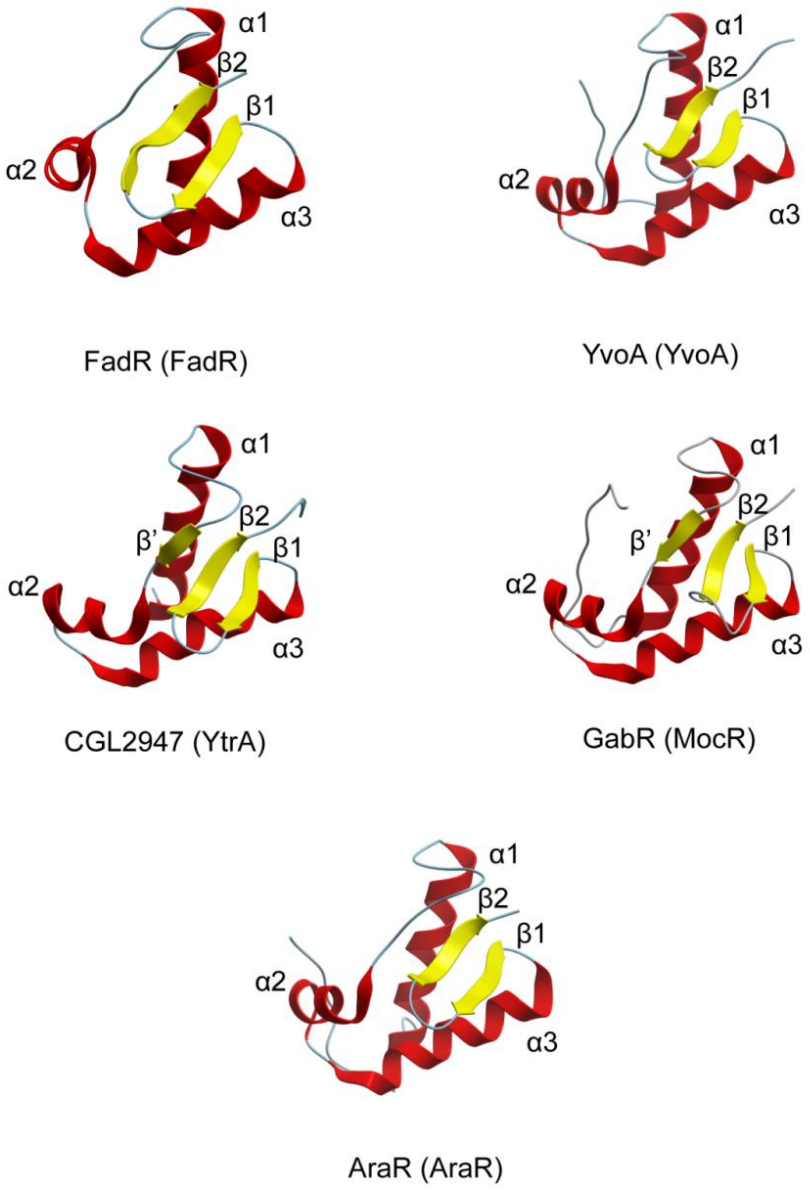

Figure 1.13 N-terminal wHTH domain of various GntR family members. The DNA-binding wHTH domains of GntR regulators belonging to different subfamilies is shown. The α-helices are shown in red, while β-strands are highlighted in yellow. The secondary structural elements of the domain are labelled. The image was created by using ICM browser (www.molsoft.com).

Although the GntR-type wHTH domain exhibits low sequence identity at the family level, when compared at the subfamily level, the average amino acid sequence identity increases to ~40% suggesting that the diversity in the C-terminal E/O domains of various subfamily members reflects upon the wHTH domain (Rigali et al., 2002). Even with its relatively simpler structural scaffold, the wHTH domain exhibits functional

versatility in terms of binding to various operators bearing distinct topologies. As mentioned earlier, the oligomeric arrangement or allosteric changes induced by binding of an appropriate effector molecule allows the correct spatial disposition of the DNA binding domain (DBD) for binding operators with contrasting features. The majority of GntR family regulators dimerize in solution to bind operator sites having conserved inverted repeats with 2-fold symmetry (Hoskisson and Rigali, 2009; Jain, 2015; Rigali et al., 2002; Suvorova et al., 2015), where each monomer from the dimer interacts with one half of the inverted repeat. The DNA-protein contacts are more or less identical for the respective halves of the nucleoprotein complex (Fillenberg et al., 2015; Gao et al., 2017; Shi et al., 2015; van Aalten et al., 2001). Many GntR family regulators recognize operator DNA having direct repeats (reviewed in (Hoskisson and Rigali, 2009; Jain, 2015; Rigali et al., 2002; Suvorova et al., 2015). The operators are A/T rich sequences with a relatively conserved center and variable periphery (Jain, 2015; Suvorova et al., 2015). Several of the experimentally determined operators recognized by GntR family regulators match with the palindromic or pseudo-palindromic NyGTNxACNy consensus sequence, where nature (N) and number (x or y) of the nucleotides varies around the conserved GT and AC pairs (Suvorova et al., 2015). Comparative studies at the subfamily level suggest that the consensus for the FadR and HutC subfamilies is NyGTM-$N_{0–1}$-KACNy and NyGTMTAKACNy, respectively (Suvorova et al., 2015). Although FadR subfamily members usually bind inverted repeats, some transcriptional regulators such as NanR (TGGTATAW) (Condemine et al., 2005) and BioR (TTATMKATAA) (Feng et al., 2013; Rodionov and Gelfand, 2006) bind DNA with direct repeats while others bind DNA lacking any symmetry (Rigali et al., 2002; Suvorova et al., 2015). Similarly, HutC subfamily members are also known that bind DNA sites other than inverted repeats such as direct repeats (NagQ: TGGTATT (Yang et al., 2006) and FarR: TGTATTAWTT

(Quail et al., 1994)) or DNA with additional symmetry (SdhR: TCTTATGTCTTATATAAGACATAAGA) (Suvorova et al., 2012). So far the consensus for other GntR subfamilies could not be derived. The spacing between the two half-sites seems to be important for recognition by the regulator and varies to different degrees amongst the subfamilies. The distance varies only weakly among the FadR and HutC subfamily members, but is quite different from the members of the YtrA subfamily where the GT and AC bases are positioned far from the middle of the palindrome (Rigali et al., 2002). Among the FadR subfamily transcriptional regulator-binding sites the distance between the conserved GT and AC pairs in respective halves of the inverted repeat is either 3 or 2 nucleotides, and seem to be critical for binding with their cognate regulators (Suvorova et al., 2015).

Majority of information regarding molecular details of interaction of GntR family members with their cognate DNA is derived from the solved crystal structures of FadR from *E. coli* (van Aalten et al., 2001) and *V. cholerae* (Shi et al., 2015) (FadR subfamily), AraR from *B. subtilis* (AraR subfamily) (Jain and Nair, 2013; Jain et al., 2016), and YvoA from *B. subtilis* (Fillenberg et al., 2015) (HutC subfamily). Comparison of the various structures suggests that the mode of interaction of various secondary structures of the wHTH domain of the regulators belonging to different subfamilies with their operator DNA is highly conserved. The amino acid contacts of FadR from *E. coli* and *V. cholerae* are identical. The DNA-protein interaction is mediated by specific amino acid-nucleotide base or non-specific contacts with the sugar phosphate backbone of the DNA. The α2 of the HTH motif orients the interaction of N-terminal amino acid in the first few turns of the α3 helix (recognition helix) deep into the major groove of the DNA. Notably a glutamic acid residue (E34, E27 and E30 in FadR, YvoA and AraR respectively) in α2 contacts two conserved arginine residues (R45 and R49 in FadR; R38 and R48 in YvoA;

and R41 and R45 in AraR) in the α3 helix to facilitate their interaction with nucleotides in their operator DNA. Of these the first arginine is usually involved in specific interaction with a conserved guanine residue in major grove. The amino acids at the tip of the wing dock into the minor grove and can make specific contacts with bases in the DNA for example H65 in FadR, G69 in YvoA and Q61 in AraR make direct contacts with the bases in the minor grove of the DNA. The amino acid residues in the α1 helix usually make non-specific contacts with the DNA and provide stability to the contacts make by the downstream amino acids. The specific contacts made by various amino acids of FadR with its cognate DNA are shown in Fig. 1.14.

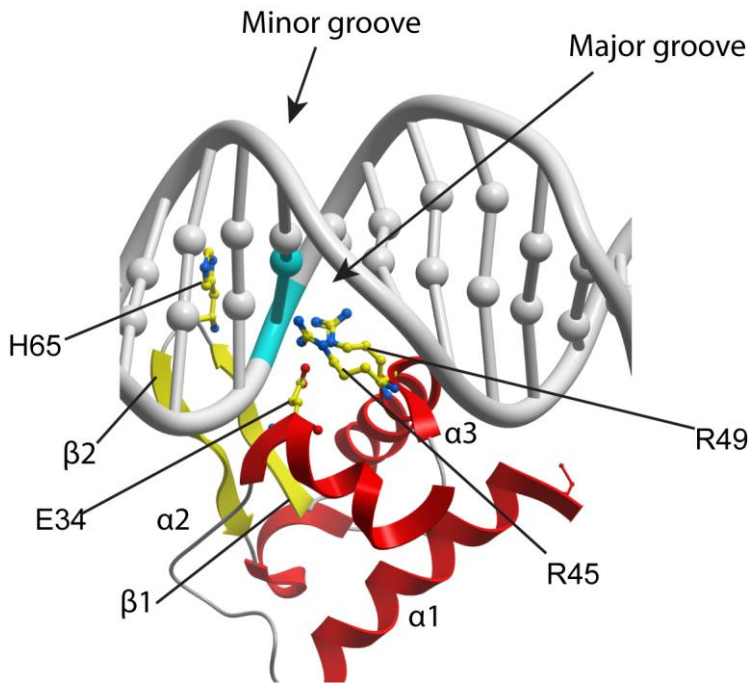

Figure 1.14 Protein-DNA contacts in FadR-DNA complex. The critical amino acids of FadR involved in making specific contacts with its cognate DNA in the FadR-DNA complex (PDB: 1HW2) (Xu et al., 2001) are labelled. The α-helices are shown in red, while β-strands are highlighted in yellow. The DNA strands are represented as grey ribbons. The conserved 'G' base involved in interaction of GntR family regulators with their DNA is highlighted in cyan color. The image was created by using ICM browser (www.molsoft.com).

1.4.6 Effector binding characteristics

Multiple studies on individual transcriptional regulators belonging to GntR family shows that their C-terminal domain (E/O) is involved in oligomerization and binding effector molecules (reviewed in (Jain, 2015; Rigali et al., 2002)). Effectors bound by this family are generally small organic molecules that bind the transcriptional regulator to alter its DNA-binding properties. For the transcriptional regulators that govern metabolism of carbon sources, the effector bound can be either the substrate or the metabolic intermediate of the pathway (Bates Utz et al., 2004; Bouvier et al., 2019; DiRusso et al., 1992; DiRusso et al., 1998; Hoskisson and Rigali, 2009; Jain, 2015; Kalivoda et al., 2003; Lee et al., 2000; Miwa and Fujita, 1988; Tutukina et al., 2016a). Although the effectors for a large number of GntR family regulators are known, the molecular details of their interaction with the proteins are underexplored. Only a limited number of structures with bound effectors have been solved that provide information of the effector-binding pocket of the regulators (Table 1.4). Among the FadR subfamily members, the structures for effector- bound dimeric FadR from *E. coli*, *V. cholerae* and *V. alginolyticus* have been solved (Gao et al., 2017; Shi et al., 2015; van Aalten et al., 2001). The cognate effector, long chain fatty acyl CoA enters the protein through α5 and α10 in the FadR C-terminal E/O domain. The pantothenate group of CoA makes multiple contacts with the amino acids in pocket on the relatively exposed surface, whereas the long acyl chain enters a groove making a large number of contacts with the protein (Gao et al., 2017; Shi et al., 2015; van Aalten et al., 2001). In FadR from *V. cholerae* and *V. alginolyticus* insertion of additional 40 amino acids forms binding site for another effector molecule in each monomer (Gao et al., 2017; Shi et al., 2015). Comparison of the apo-, DNA- and/or effector bound structures for various FadR proteins revealed mechanism of effector-mediated change in the binding of regulator to its cognate DNA. The apo- and DNA-

bound structures show a large degree of similarity whereas binding of the effector molecule induces conformational change in the E/O domain which allosterically changes the orientation of the whole N-terminal DNA-binding domain (Discussed in detail later). Allosteric change induced by binding of effector increases the relative distance between the two recognition helices in the wHTH domain rendering them unsuitable to dock into their operator DNA. Although the solved crystal structures for other FadR subfamily members have been used to predict the effector binding pocket, experimental evidences are still lacking. The comparison of their predicted effector binding site with that of FadR shows that they may have similar or dissimilar effector binding cavity or even effector mediated allosteric change to affect the DNA-binding properties may be different (Gao et al., 2008; Lord et al., 2014; Zheng et al., 2009). Multiple reports suggest that some of the FadR subfamily members bind a metal ion through relatively conserved amino acid residues (Blancato et al., 2016; Lord et al., 2014; Zheng et al., 2009). The metal has been proposed to sit in the effector binding pocket and assist in effector-repressor interaction.

1.5 Sugar acid metabolism and its regulation

Sugar acids, oxidized derivatives of sugars, are used as carbon source by bacteria. Sugar acids are derived from sugars by oxidation of either the hydroxyl or the carbonyl group into carboxylic group. Depending upon the nature and degree of oxidation, sugar acids are categorized into aldonic, uronic, aldaric and ulosonic acids (Bhagavan, 2002; Robyt, 1998). Besides this methyl- derivatives of sugar acids are also used by bacteria (Yoon et al., 2012). The carboxyl group of sugar acids reacts with a hydroxyl group within the same molecule to form cyclic lactones (Isbell and Frush, 1933). The open chain and lactone forms are known to be spontaneously interconvertible. Among these sugar acids and their derivatives, aldonic and uronic acids are considered as the most physiologically relevant sugar acids (Bhagavan, 2002; Mandrand-Berthelot et al., 2004). Sugar acids are

produced by enzymatic oxidation or epimerization reactions during metabolism of various sugars. Besides this, many sugar acids occur naturally. D-galacturonate is a monomer of pectin, an important constituent of plant cell wall and its free form is produced by enzymatic degradation by gut microbial flora of herbivores (Jayani et al., 2005; Richard and Hilditch, 2009).

1.5.1 Physiological importance of sugar acid metabolism

Multiple studies illustrate the physiological relevance of sugar acid metabolism for various organisms. Sugar acids such as D-glucuronate are used for detoxification of harmful xenobiotics in mammalian liver (Perreault et al., 2013). D-galactonate is produced by human tissues to dispose-off excess non-metabolizable D-galactose in galactosemic patients (Lai and Klapa, 2004). Multiple plant pathogens including various bacteria and fungi are known to utilize sugar acids released by enzymatic digestion of plant cell wall to facilitate and sustain their interaction with their host (Prade et al., 1999). Soft rot pathogen *Pectobacterium carotovorum* mutant defective in D-gluconate metabolism was shown to have diminished virulence in *Arabidopsis thaliana* and *Solanum tuberosum*, suggesting the importance of D-gluconate metabolism in the virulence of bacteria (Mole et al., 2010). *Botrytis cinerea* is a necrotrophic fungus which infects multiple plant species. Its mutant defective in D-galacturonate metabolism was found to have selectively compromised virulence on *A. thaliana* and *Nicotiana benthamiana* leaves rich in this sugar acid suggesting the role of D-galacturonate metabolism in the virulence of *B. cinerea* (Zhang and van Kan, 2013). Sugar acids such as D-gluconate have been demonstrated to be important for colonization of mammalian gut by *E. coli (Sweeney et al., 1996)*. Metabolism of yet other sugar acids D-galactarate and D-glucarate was reported to play an important role in post-antibiotic expansion of gut microbiota in mammals (Faber et al., 2016).

1.5.2 Components involved in the metabolism of sugar acids in *E. coli*

E. coli K-12 is capable of utilizing sugar acids including various hexonates, hexuronates and hexuronides as sole source of carbon and energy (Mandrand-Berthelot et al., 2004). For them to be used by bacteria, the sugar acid or its derivative must be transported across the envelope into the bacterial cytoplasm, where it itself or its metabolic products must interact with its regulatory machinery to enhance the availability/activity of enzymatic components required for its metabolism. In contrast to their related sugars, most of the sugar acids are transported by specific sugar acid transporters belonging to the Ion Transporter (IT) or Major Facilitator Superfamily (MFS) superfamilies (Table 1.5). In certain cases the same transporter is employed for transporting multiple sugar acids, for example ExuT is involved in the transport of both D-galacturonate and D-glucuronate (Table 1.5). Sugar acids are metabolized by a limited number of enzymatic activities involving isomerases, oxidoreductases, dehydratases, kinases and aldolases (Table 1.6). Almost all of the sugar acids are metabolized to produce either 6-carbon 2-keto-3-deoxy 6-phosphogluconate (KDPG) or α-ketoglutarate or glyceraldehyde-3 phosphate and pyruvate, which enter the glycolytic pathway for energy generation by substrate-level phosphorylation (Mandrand-Berthelot et al., 2004; Rodionov et al., 2000; Suvorova et al., 2011). Hexuronides and hexuronates are metabolized by Ashwell pathway while other sugar acids such as D-gluconate, L-idonate or 2-keto-3-deoxy gluconate (KDG) are processed by Entner-Doudoroff (ED) pathway as shown in Fig. 1.15. D-galactonate is metabolized through a modified form of ED pathway (Cooper, 1978; Deacon and Cooper, 1977) and galactarate is metabolized by a separate pathway to produce pyruvate and 2-phosphoglycerate as final products to enter glycolysis and tricarboxylic acid cycle (Mandrand-Berthelot et al., 2004).

Table 1.5 Transporters involved in the uptake of various sugar acids in *E. coli* (Mandrand-Berthelot et al., 2004)

Family	Transporter	Function	Family	Substrate/s
IT	GntP	D-gluconate transporter	IT	D-gluconate
IT	GntT	D-gluconate transporter	IT	D-gluconate
IT	GntU	D-gluconate transporter	IT	D-gluconate
IT	IdnT	L-idonate transporter	IT	L-idonate
IT	KdgT	KDG transporter	IT	Keto-deoxygluconates
MFS	DgoT	D-galactonate transporter	MFS	D-galactonate
MFS	ExuT	Hexuonate transporter	MFS	D-galacturonate, D-glucuronate
MFS	GarP	D-galactarate transporter	MFS	D-glucarate, D-galactarate
MFS	GudP	D-glucarate transporter	MFS	D-glucarate, D-galactarate
MFS	UidB	D-glucuronide transporter	MFS	D-glucuronides

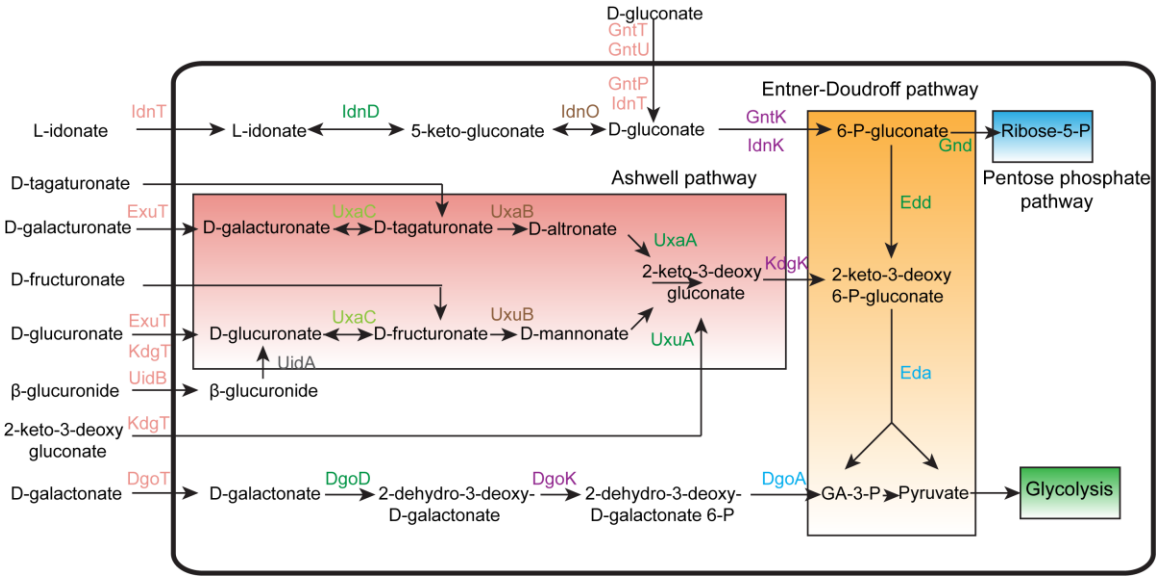

Figure 1.15 Pathways for transport and metabolism of hexuronides, hexuronates and hexonates in *E. coli.* The metabolic routes are indicated by arrows. Ashwell, Entener-Doudoroff, pentose phosphate and glycolytic pathways are shown in red, orange, blue and green boxes respectively. Various transporters and enzymes involved at various metabolic steps are mentioned above arrows. The image was modified from (Peekhaus and Conway, 1998).

Table 1.6 Metabolic enzymes involved in the metabolism of various sugar acids in *E. coli* (Mandrand-Berthelot et al., 2004)

Primary catabolic pathway	**Gene**	**Function**
2,5-diketo-D-gluconate	DkgA	2,5-diketo-D-gluconate reductase
2,5-diketo-D-gluconate	DkgB	2,5-diketo-D-gluconate reductase
2,5-diketo-D-gluconate	GhrB	2-keto-L-gulonate/2,5-diketo-D-gluconate reductase
5-keto-4-deoxyuronate	KduD	2,5-diketo-3-deoxy-D-gluconate reductase
5-keto-4-deoxyuronate	KduI	5-keto-4-deoxyuronate isomerase
Entner-Doudoroff	Eda	2-keto-3-deoxy-D-gluconate 6-phosphate aldolase
Entner-Doudoroff	Edd	D-gluconate 6-phosphate dehydratase
D-galactarate	GarD	D-galactarate dehydratase
D-galactonate	DgoA	2-keto-3-deoxy-D-galactonate 6-phosphate aldolase
D-galactonate	DgoD	D-galactonate dehydratase
D-galactonate	DgoK	2-keto-3-deoxy-D-galactonate kinase
D-galacturonate	UxaA	D-altronate dehydratase
D-galacturonate	UxaB	D-altronate oxidoreductase
D-galacturonate/ D-glucuronate	UxaC	Hexuronate isomerase
D-glucarate	GudD	D-glucarate dehydratase
D-glucarate/galactarate	GarK	D-glycerate kinase
D-glucarate/galactarate	GarL	5-keto-4-deoxy-D-glucarate aldolase
D-glucarate/galactarate	GarR	Tartronate semialdehyde reductase
D-gluconate	GntK	D-gluconate kinase
D-glucuronate	UxuA	D-mannonate dehydratase
D-glucuronate	UxuB	D-mannonate oxidoreductase
Hexuronate/2-keto-3-deoxy-D-gluconate	KdgK	2-Keto-3-deoxy-D-gluconate kinase
L-Idonate	IdnD	L-idonate dehydrogenase
L-Idonate	IdnK	D-gluconate kinase
L-Idonate	IdnO	5-keto-D-gluconate reductase
β-D-glucuronide	UidA	β-D-glucuronidase

1.5.3 Components involved in regulation of sugar acid metabolism in *E. coli*

In *E. coli*, majority of genes encoding components for transport and metabolism of various sugar acids are under regulation by global as well as specific regulators (Table 1.7). Most of the sugar acid metabolic genes are under cAMP-CRP mediated carbon catabolite repression with a few exceptions such as the *gntP*, the D-gluconate transporter and *edd* and *eda* genes of ED pathway (Mandrand-Berthelot et al., 2004). Sugar acid regulators belong to different families such as GntR, LacI, IclR, TetR and CdaR (Table 1.7). Most of the regulation is dominated by repressors with exceptions of CdaR and IdnR involved in regulation of D-glucarate/galactarate and L-idonate, respectively. Majority of the regulators exhibit autoregulation for rapidly responding to the changing environmental conditions. DgoR, ExuR, UidR, and UxuR negatively regulate themselves while CdaR is under its positive autoregulation. However, GntR and IdnR, which regulate uptake and metabolism of D-gluconate and L-idonate, respectively, are constitutively expressed. Due to extensive isomerization, sharing of inducers or the transcriptional regulators, the sugar acid metabolic pathways tend to cross-communicate with each other (Mandrand-Berthelot et al., 2004; Rodionov et al., 2000). One of the prominent examples includes cross-talk among hexuronide and hexuronate pathways by cooperative regulation of D-glucuronidase gene, *uidA* by two regulators, UidR, the specific regulator as well as UxuR, the primary regulator of D-glucuronate metabolism. The complete induction of *uidA* requires derepression of UidR and UxuR by their cognate effectors, D-glucuruonide/D-glucuronate and D-fructuronate, respectively (Novel and Novel, 1976; Ritzenthaler et al., 1983). Additional examples of cross-regulation among sugar acid metabolic pathways include cross-talk between L-idonate and D-gluconate metabolic pathways (Tsunedomi et al., 2003a; Tsunedomi et al., 2003b) as well as UxuR and ExuR

mediated co-regulation of hexuronate metabolism by formation of UxuR/ExuR hetero-dimers (Tutukina et al., 2016a).

Table 1.7 Transcriptional regulators involved in the regulation of sugar acid metabolism in *E. coli* (Mandrand-Berthelot et al., 2004)

Family	Regulator	Regulation type	Regulon members
CdaR	CdaR	Activation	*cdaR, garD, garPLRK, gudPXD*
GntR	DgoR	Repression	*dgoRKADT*
GntR	ExuR	Repression	*exuR, exuT, uxaB, uxaCA, (uxuAB, uxuR)*
GntR	UxuR	Repression/Activation	*uxuAB, uxuR, (uidABC), gntP, uidR'*
IclR	KdgR	Repression	*kdgT, kdgD, KduID*
LacI	GntR	Repression	*edd-eda, gntKU, gntT, (idnDOTR, idnK)*
LacI	IdnR	Activation	*idnDOTR, idnK,(gntKU, gntl)*
TetR	UidR	Repression	*uidABC, uidR*

1.6 D-galactonate sugar acid

1.6.1 Introduction

D-galactonate, a conjugate base of D-galactonic acid, is an aldonic sugar acid derivative of D-galactose where the aldehyde group (-CHO) is oxidized into carboxylic group (-COOH) (Fig. 1.16). L-galactonate, the enantiomer of D-galactonate, differing in spatial orientation of hydroxyl group (-OH) at the C5 carbon is also known to exist in nature and is produced by enzymatic degradation of pectin by various fungi (Kuivanen et al., 2014; Kuivanen et al., 2012; Martens-Uzunova and Schaap, 2008). D-galactonate can undergo intramolecular esterification to form 5- or 6- membered cyclic lactones, D-galactono 1-4 lactone (D-galactonic acid-γ-lactone) or D-galactono 1-5 lactone (D-galactonic acid-δ-lactone) (Fig. 1.16). The D-galactonic acid-δ-lactone is hyrolyzed fairly rapidly into D-galactonate as compared to its 5-membered counterpart, D-galactonic acid-γ-lactone.

Both linear chain and cyclic lactones are spontaneously interconvertible in their aqueous solutions although living systems such as bacteria and fungi have been reported to harbour lactonases to mediate this interconversion (De Ley and Doudoroff, 1957; Isbell and Frush, 1933; Shimizu et al., 1992). D-galactonate is a commercially attractive compound finding its application in polyester based polymer synthesis (Romero Zaliz and Varela, 2003, 2006). Recently, there has been a surge in biochemical synthesis of D-galactonate using economically cheaper raw materials (Liu et al., 2014a; Liu et al., 2014b; Ramos et al., 2017; Zhou et al., 2018). In a separate study, the end products of De Ley-Doudoroff pathway known to produce endogenous D-galactonate as an intermediate were combined with methylerythritol phosphate pathway (MEP) in an engineered *E. coli* strain to overproduce isoprenoids, another class of commercially valuable compounds (Ramos et al., 2014).

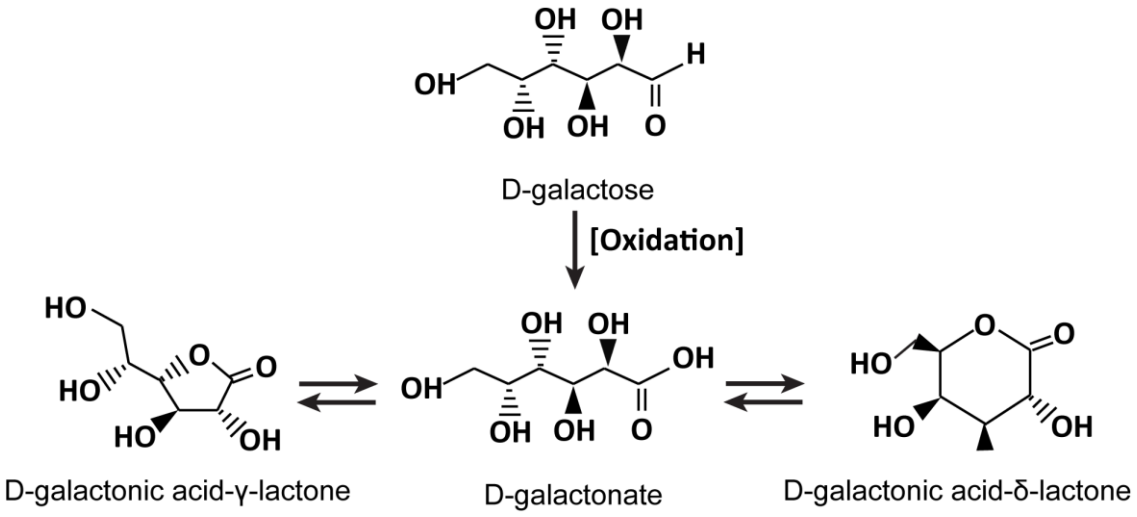

Figure 1.16 Structure of D-galactonate and its lactone forms. D-galactonate can be derived from oxidation of –CHO group of D-galactose in to –COOH group. D-galactonate can undergo intramolecular esterification to form either γ- or δ- lactones.

1.6.2 D-galactonate prevalence in nature

Several reports suggest the wide prevalence of D-galactonate in nature. D-galactonate is produced by a variety of organisms including various bacterial and fungal species as an intermediate of D-galactose metabolism through De Ley-Douduroff pathway. Bacteria such as *Stenotrophomonas maltophilia* (isolated from intestinal crypts and mucous membranes of mammalian hosts) (Saffarian et al., 2015), *Gluconobacter liquefaciens* (present in sugar-rich habitats) (Gupta et al., 2001), *P. saccharophila* (a mud bacterium), *Azotobacter vinelandii* (a free living diazotroph found in soil) (Lipman, 1903; Noar and Bruno-Barcena, 2018), *Caulobacter crescentus* (an oligotrophic aquatic bacterium) (Poindexter, 1964, 2015), and *S. meliloti* and *R. meliloti* (nitrogen fixing symbionts of legumes) (Sulieman and Tran, 2014) produce D-galactonate by enzymatic oxidation of D-galactose followed by isomerization to D-galactono-γ-lactone by a lactonase as an initial step of the De Ley-Doudoroff pathway (A.H., 1961; Arias and Cervenansky, 1986; Brechtel et al., 2002; De Ley and Doudoroff, 1957; Geddes and Oresnik, 2012; Kurn et al., 1978; Wong and Yao, 1994). Fungi such as *Aspergillus niger* also produce D-galactonate from D-galactose through a non-phosphorylative De Ley-Doudoroff pathway (Elshafei and Abdel-Fatah, 2001). A recent study detected the presence of D-galactonate in cheese curd and the inoculation of cheese curd with *Debaryomyces hansenii*, a common yeast found in food items such as dairy products (Banjara et al., 2015), increased its concentration nearly five times, suggesting that *D. hansenii* can produce D-galactonate (Pham et al., 2019).

Various reports have demonstrated the presence of D-galactonate in mammalian tissues as well as body secretions (Berry et al., 1998; Rancour et al., 1979; Rogers et al., 1984; Schadewaldt et al., 2004). Humans produce D-galactonate especially after consumption of lactose-rich diet such as cheese and milk. A recent report has proposed

D-galactonate in human serum and urine as dietary-lactose intake marker (Münger et al., 2017; Trimigno et al., 2018). Another recent report demonstrated postprandial increase in D-galactonate levels in human serum and urine as one of good proxies for genetically driven lactase activity suggesting its measurement in human urine can be used for developing a non-invasive lactose digestion test for diagnosing individuals with hypolactasia, a condition with poor lactose digestion (Vionnet et al., 2019). D-galactonate is also produced by an alternative metabolic route as a means to dispose-off excess D-galactose in galactosemic patients, who lack the enzyme(s) of D-galactose metabolism (Lai and Klapa, 2004). Patients with galactosemia produce increased levels of D-galactonate in blood and urine (Ficicioglu et al., 2005; Yager et al., 2004).

1.6.3 D-galactonate metabolism and its regulation in *E. coli*

Reports as early as 1970s suggested that exogenous D-galactonate is rapidly utilized as a sole source of carbon and energy by various bacteria such as non-pathogenic mycobacteria, *Mycobacterium butyricum*, and enteric bacteria including *E. coli* and *S. enterica* serovar Typhimurium LT-2 (Deacon and Cooper, 1977; Szumilo, 1981, 1983). In a classical mutagenesis based study, *E. coli* K-12 mutants defective in utilizing D-galactonate as a sole carbon source were isolated and biochemically analyzed for the phosphorylation, hydrolysis and aldol cleavage enzymatic activities usually required for metabolism of sugar acids thus delineating the metabolic pathway of D-galactonate metabolism in *E. coli* (Cooper, 1978). Biochemical insights from the study suggested that D-galactonate is metabolized by a modified form of ED pathway, by enzymatic action of three enzymes designated as DgoD, DgoK and DgoA with D-galactonate dehydratase, kinase and phosphorylase activities, respectively. A putative transporter DgoT, was also suggested for transport of exogenous D-galactonate. All of the Dgo enzymes were found to be simultaneously produced when the cells were grown in the presence of D-

galactonate (Cooper, 1978; Deacon and Cooper, 1977). One particular mutation led to simultaneous loss of all the D-galactonate metabolic activities and was proposed to affect the common putative promoter of the *dgo* genes, suggesting expression of all *dgo* genes from a common promoter. Further, the mutations corresponding to various D-galactonate metabolic enzymes of *E. coli* were genetically mapped by conjugation and were found to occur close to each other (Cooper et al, 1978). Based on these observations it was suggested that the D-galactonate metabolic genes constitute a putative operon (*dgo*) located at min 81.7 of the *E. coli* linkage map. Later, sequencing of the corresponding region and annotation found that five genes encoding a putative repressor, DgoR; a kinase, DgoK; an aldolase, DgoA; a dehydratase, DgoD and a putative transporter, DgoT are located next to each other (Babbitt et al., 1995; Burland et al., 1993). However, no experimental proof for the operon organization of the *dgo* genes was available. Although the enzymatic activities were determined from crude lysates prepared from cells grown in the presence of D-galactonate and assigned to putative genes in the original study (Cooper, 1978), purified proteins were used to understand structural and biochemical details of D-galactonate metabolism and transport much later and these studies are described below.

1.6.3.1 Transport of exogenous D-galactonate

D-galactonate is transported into the cytoplasm by a D-galactonate:H^+ symporter DgoT. According to Transporter Classification Database DgoT belongs to Anion:Cation Symporter (ACS) family of the MFS superfamily of transporters (Pao et al., 1998; Saier et al., 2016). Purified and reconstituted DgoT from *E. coli* was shown to mediate D-galactonate: proton symport in an electrogenic fashion. A fluorescent dye based assay revealed that DgoT is a specific transporter of D-galactonate as it does not transport its epimer, structurally varying only in the orientation of –OH group around single carbon

atom. Solved crystal structures of WT open to the cytoplasmic site (PDB ID: 6E9N) and transport-defective mutant (E133Q) in a galactonate-bound form open to the periplasmic side (PDB ID: 6E9O) shed light on the specificity and mechanism of D-galactonate transport. In contrast to other related transporters, both N- and C- terminal domains of DgoT play important role in substrate recognition (Fig. 1.17). Arg47 plays a specific role in interaction with carboxyl group of D-galactonate and further specificity in substrate recognition is provided by set of 9 hydrogen bonds formed between the sugar acid and the transporter. In substrate free-state, Glu133 and Arg47 in DgoT form a charge pair interaction. Asp46 and Glu133 in the periplasmic oriented protein are protonated, allowing the interaction of Arg47 with the carboxyl group of D-galactonate substrate which triggers allosteric changes in the protein from periplasmic to cytoplasmic facing conformation. Both the substrate and the proton are released into the cytoplasm. Subsequently, deprotonation of Glu133 restores its interaction with Arg47 to form a charge pair, reorienting the transporter from cytoplasmic to periplasmic-facing conformation in the process. In the next cycle, the transporter protein again accepts a proton for subsequent binding of another D-galactonate molecule and its transport across the bacterial membrane (Leano et al., 2019).

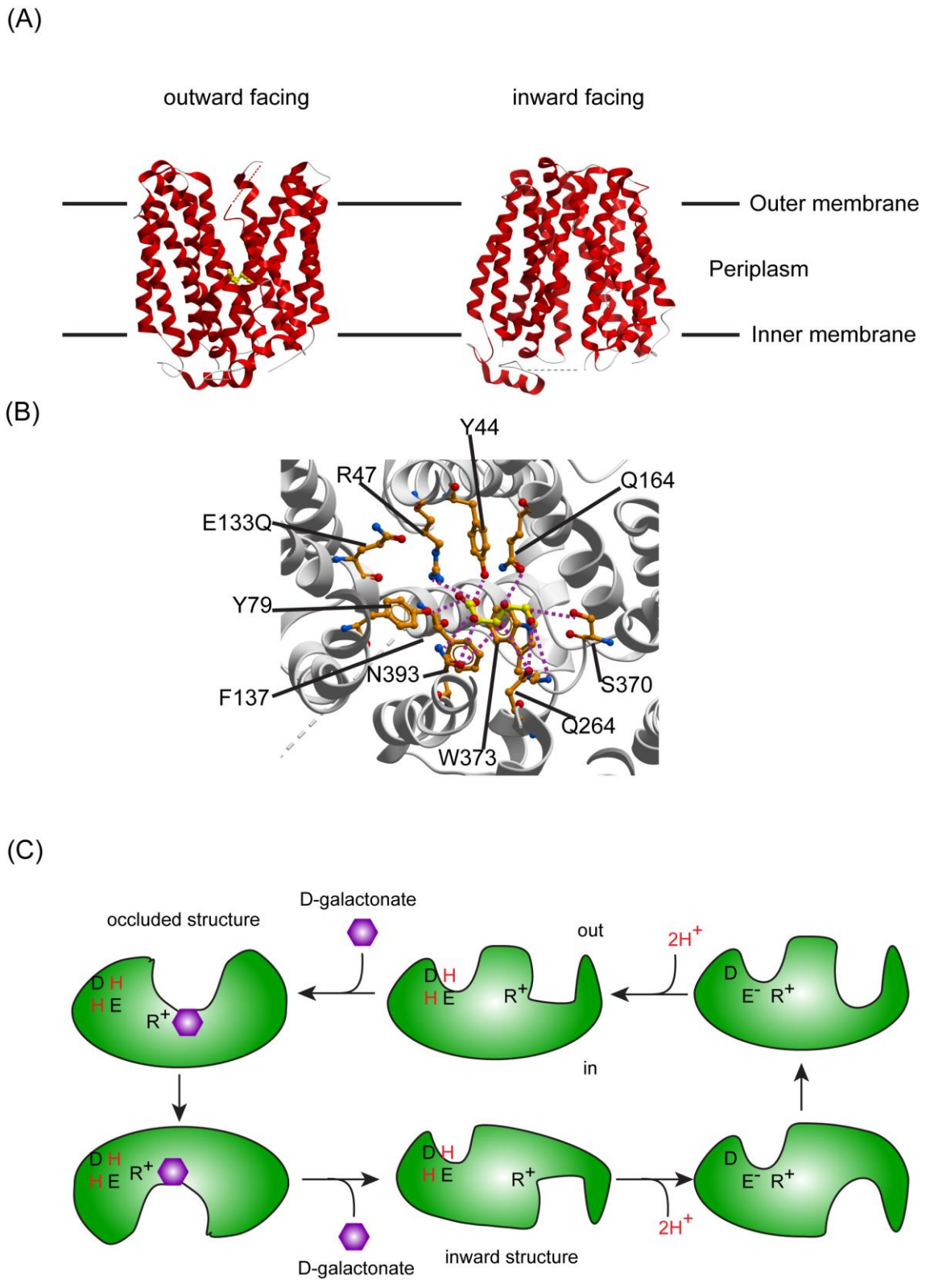

Figure 1.17 Transport of D-galactonate by DgoT. (A) Structure of D-galactonate transporter, DgoT, solved in inward (PDB ID: 6E9N) and D-galactonate-bound outward facing (PDB ID: 6E9O) conformations. The α-helices in the structure are shown in red. The structures were made by using ICM browser (www.molsoft.com). (B) Close up view showing interaction of D-

galactonate with DgoT. Amino acid residues involved in interaction (shown in orange) with D-galactonate (shown in yellow) are labelled. The image was created by using ICM browser (www.molsoft.com). C) Mechanism of D-galactonate transport by DgoT (Leano et al., 2019). Glu133 and Arg47 amino acid residues of DgoT form a charge pair interaction and through proton mediated mechanism transport D-galactonate across the cell membrane. The image was modified from (Leano et al., 2019).

1.6.3.2 Dehydration of D-galactonate

Following transport into the cytoplasm, D-galactonate is dehydrated by β-elimination of OH^- by D-galactonate dehydratase, DgoD (EC:4.2.1.6) belonging to the mandalate racemase (MR) subgroup within the enolase superfamily to form an α-ketoacid derivative, 2-dehydro-3-deoxy-D-galactonate (Babbitt et al., 1995; Gerlt et al., 2005) (Fig. 1.18A). The structure for DgoD from *E. coli* could be solved only at a poor resolution and was used to deduce information for the biochemical behavior of mutants with compromised dehydratase activity, (Wieczorek et al., 1999) (Fig. 1.18B). Based on the structure and biochemical assays with mutants a conserved His185 in DgoD active site has been proposed to act as a general acid/base catalyst to abstract the α-proton to form an enolic intermediate, which undergoes β-elimination of the 3-hydroxyl group to form an α,β-unsaturated enol (Fig. 1.18A). The subsequent α,β-unsaturated enol tautomerizes to yield the final α-ketoacid product, 2-dehydro-3-deoxy-D-galactonate (Wieczorek et al., 1999). Besides His185, His285 and Glu310 were also shown to be important for catalytic activity of DgoD. However, structure for a similar putative D-galactonate enolase PRK14017 from *R. pickettii*, an important nosocomial infectious bacteria (Ryan et al., 2006), was solved (PDB ID: 3RR1, 3RRA) and submitted in RCSB (with no accompanied publications) (Fig. 1.18C). The protein sequence alignment using L-align (Huang and Miller, 1991) shows that the two proteins are 83.8% identical and

95.0% similar. The amino acid residues His185, His285 and E310 crucial for DgoD activity are conserved in PRK14017.

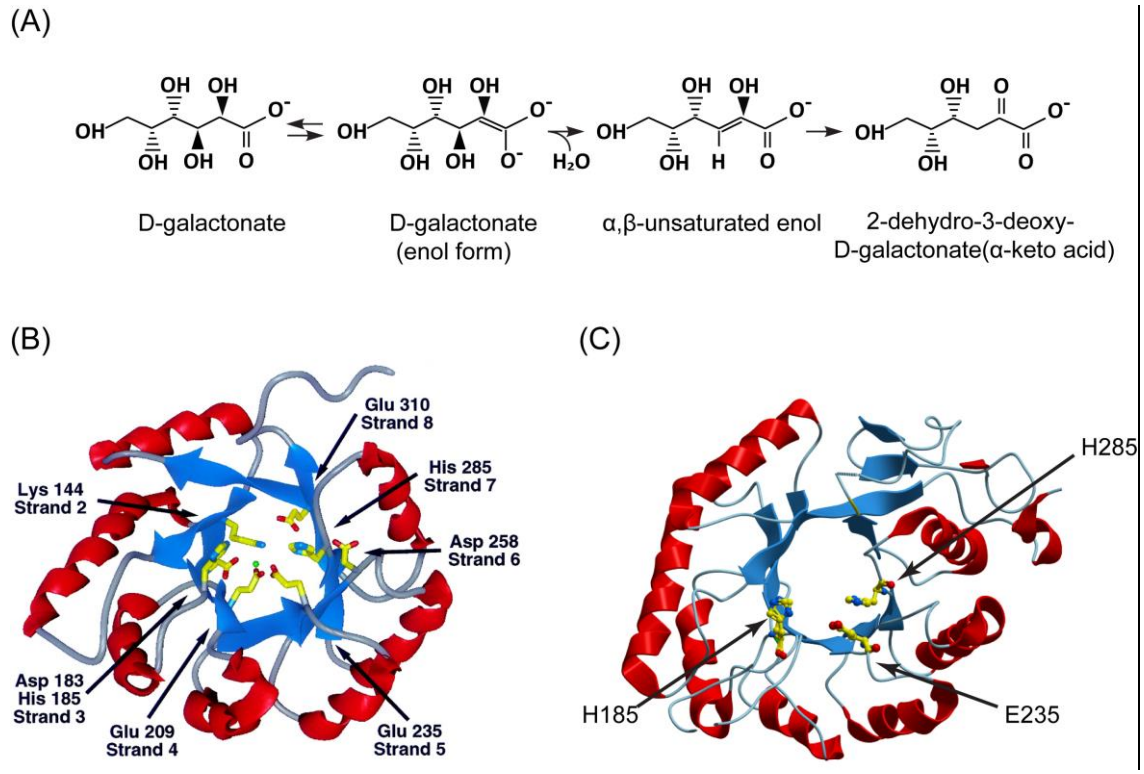

Figure 1.18 Dehydration of D-galactonate by DgoD (A) Proposed mechanism of dehydration by DgoD (Wieczorek et al., 1999). His185 in DgoD active site removes the α-proton to form an enolic intermediate. Further β-elimination of the 3-hydroxyl group leads to formation of an α,β-unsaturated enol. The subsequent α,β-unsaturated enol tautomerizes to yield the final α-ketoacid product, 2-dehydro-3-deoxy-D-galactonate (B) Structure of DgoD from *E. coli* [reprinted (adapted) with permission from (Wieczorek, S.J., Kalivoda, K.A., Clifton, J.G., Ringe, D., Petsko, G.A., and Gerlt, J.A. (1999). Evolution of enzymatic activities in the enolase superfamily: identification of a "New" general acid catalyst in the active site of D-galactonate dehydratase from *Escherichia coli*. J Am Chem Soc *121*, 4540-4541). Copyright (1999) American Chemical Society.] The structure was not submitted to RCSB due to poor resolution. (C) Solved structure of DgoD from *R. pickettii* (PDB ID: 3RR1) submitted to RCSB). The α-helices are shown in red, while β-strands are highlighted in blue. The amino acids critical for functioning of DgoD are labelled. The image was created by using ICM browser (www.molsoft.com).

1.6.3.3 Phosphorylation of 2-dehydro-3-deoxy-D-galactonate

2-dehydro-3-deoxy-D-galactonate produced from D-galactonate is phosphorylated using ATP as phosphoryl-group donor to produce 2-dehydro-3-deoxy-D-galactonate 6-phosphate by 2-dehydro-3-deoxygalactonokinase, DgoK (EC:2.7.1.58) (Fig. 1.19A). Experiments demonstrating enzyme activity with purified DgoK are still obscure. Although the crystal structure of DgoK from *E. coli* has not been determined, structure of the corresponding kinase from *Klebsiella pneumoniae* ($DgoK_{KP}$) has been determined (PDB ID: 3R1X) (Michalska et al., 2011) (Fig. 1.19B). Pairwise protein sequence alignment using L-Align (Huang and Miller, 1991) shows both proteins share 82.2% identity and 94.5% similarity. The $DgoK_{KP}$ crystallized as a dimer of dimers while the analytical gel filtration suggests the dimeric configuration of the protein. Structural comparison with closely related members suggests that the $DgoK_{KP}$ belongs to Acetate and Sugar Kinases/Hsc70/Actin (ASKHA) family of phosphotransferases. Similar to other family members, the kinase exhibits a two-domain architecture, where the N- and C-terminal domains are separated by deep groove harboring the putative catalytic site (Fig. 1.19B). Dimer interface majorly created by the unique fragments of N-terminal domain constitutes the probable substrate binding site. Structural comparison with other members of the family suggests that the nucleotide docks a putative pocket in the C-terminal domain, where a conserved loop might be involved in interacting with the phosphoryl groups (Michalska et al., 2011).

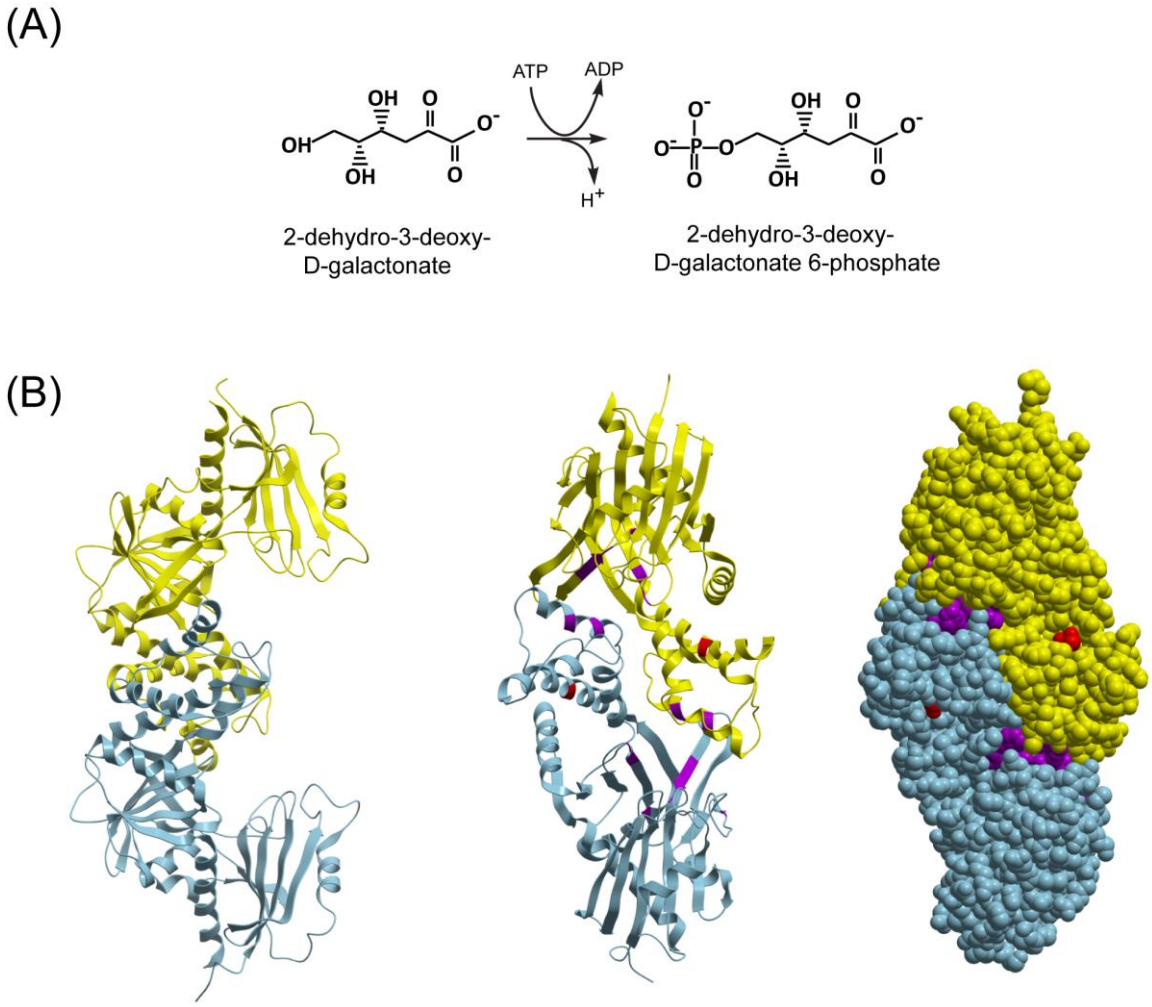

Figure 1.19 Phosphorylation of 2-dehydro-3-deoxy-D-galactonate by DgoK. (A) 2-dehydro-3-deoxy-D-galactonate is phosphorylated with ATP as phosphoryl-group donor to yield 2-dehydro-3-deoxy-D-galactonate 6-phosphate (B) Structure of DgoK from *K. pneumoniae* ($DgoK_{KP}$) showing N and C terminal domains (PDB ID: 3R1X, (Michalska et al., 2011)). Location of substrate (purple) and ATP binding site (red) is highlighted. The individual monomers of $DgoK_{KP}$ are colored yellow and blue. The image was created by using ICM browser (www.molsoft.com).

1.6.3.4 Retro-aldol cleavage of 2-dehydro-3-deoxy-D-galactonate 6-phosphate

In the final step, the phosphorylated derivative 2-dehydro-3-deoxy-D-galactonate 6-phosphate undergoes a retro-aldol cleavage mediated by 2-dehydro-3-deoxy-D-galactonate 6-phosphate aldolase, DgoA aka KDPGal aldolase (EC:4.1.2.21) to produce D-glyceraldehyde 3-phosphate and pyruvate (Fig. 1.20A). The aldolase activity of DgoA

from *E. coli* has been demonstrated with purified enzyme (Ran and Frost, 2007; Walters et al., 2008). DgoA can catalyze the above reaction in reverse by using pyruvate as a nucleophile and D-glyceraldehyde 3-phosphate as an electrophile. KDPGal aldolase as well as KDPG aldolase (2-dehydro-3-deoxy-D-gluconate 6-phosphate) from *E. coli* despite their very low sequence homology (25% sequence identity and 43% sequence similarity) catalyze similar aldol cleavage and condensation reactions, the only difference lying in orientation of –OH group around a single stereocenter (C4) of the substrate. The crystal structures for apo- and 2-dehydro-3-deoxy-D-galactonate 6-phosphate- bound KDPGal aldolase from *E. coli* have been solved (PDB ID: 2V81 and 2V82, respectively). The solved structure of KDPGal aldolase from *E. coli* is shown in Fig. 1.20B. Both KDPG aldolase and KDPGal aldolase have $(\alpha/\beta)_8$ barrel topology and exist as trimers in solution. The amino acid residues (K133/K126) and (E45/E37) important for catalysis for the both aldolases (KDPG /KDPGal) appear to occupy similar positions. A conserved lysine in the active site suggest that both KDPG and KDPGal aldoses are type I pyruvate aldolases as they seem to use lysine instead of zinc ion to activate the nucleophilic substrate, pyruvate as the Schiff base for formation of enol(ate). The substrate, 2-dehydro-3-deoxy-D-galactonate 6-phosphate, forms a Schiff base adduct with Lys126 where a water molecule mediates removal of proton from C4 –OH group of the substrate by Glu37 amino acid residue (Fig. 1.20A). Consequently, the C3-C4 bond breaks to form pyruvyl-enamine and D-glyceraldeyde 3-phosphate. Acidic Glu37 protonates pyruvyl-enamine through a water molecule to form pyruvyl-Schiff base, which is finally hydrolyzed to form pyruvate and free enzyme. Although the amino acid residues constituting the substrate binding pocket show very little conservation, the amino acid residues relevant for catalytic activity seem to be conserved in space. Notably, Thr161 amino acid residue in KDPG aldolase is replaced by Val154 in KDPGal aldolase.

Mutagenic swapping of these residues in both proteins suggests that Val154 plays important role in stereoselectivity of KDPGal aldolase (Walters et al., 2008). Another study showed that both Val154 and Ile12 amino acids are crucial for stereoselectivity of KDPGal aldolase and the corresponding regions can be grafted to related aldolases to modify their specificity (Bisterfeld et al., 2016).

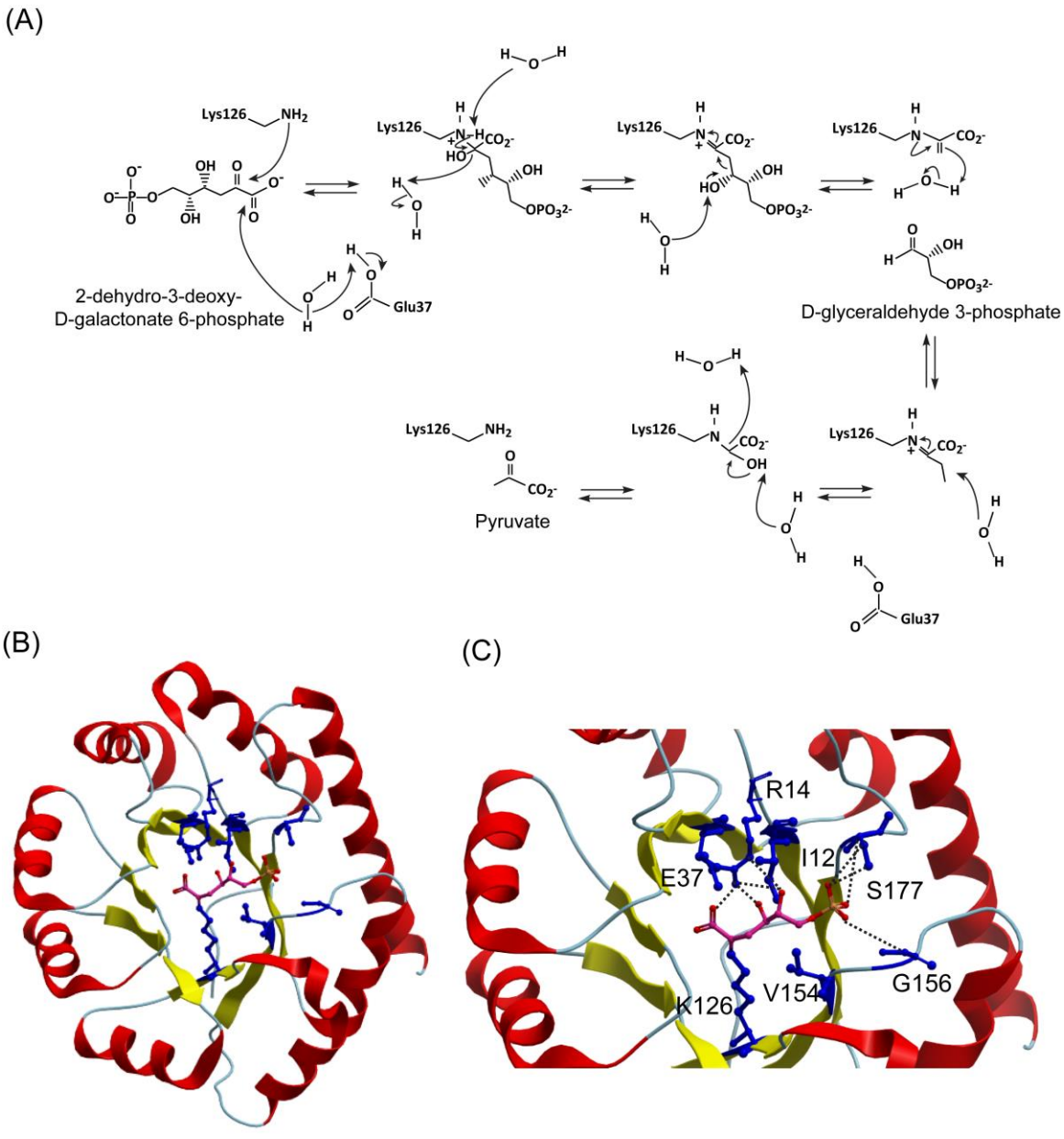

Figure 1.20 Aldol cleaveage by DgoA from *E. coli*. (A) Mechanism of retro-aldol cleavage by DgoA. A water and Glu37 mediated Schiff base is formed by 2-dehydro-3-deoxy-D-galactonate 6-phosphate and Lys126. The subsequent cleavage of C3-C4 bond results in formation of

pyruvyl-enamine and D-glyceraldeyde 3-phosphate. Protonation of the pyruvyl-enamine through acidic Glu37 and water molecule forms a pyruvyl-Schiff base, which upon hydrolysis forms pyruvate and free enzyme. (B) Structure of DgoA from *E. coli* (PDB ID: 2V82) (Walters et al., 2008). The image was created by using ICM browser (www.molsoft.com). C) Close up view showing amino acid contacts made by DgoA with substrate, 2-dehydro-3-deoxy-D-galactonate 6-phosphate (Walters et al., 2008). The α-helices are shown in red, while β-strands are highlighted in yellow. The image was created by using ICM browser (www.molsoft.com).

D-glyceraldehyde 3-phosphate and pyruvate generated after aldol condensation enter the central metabolic pathway for further metabolism (Fig. 1.21).

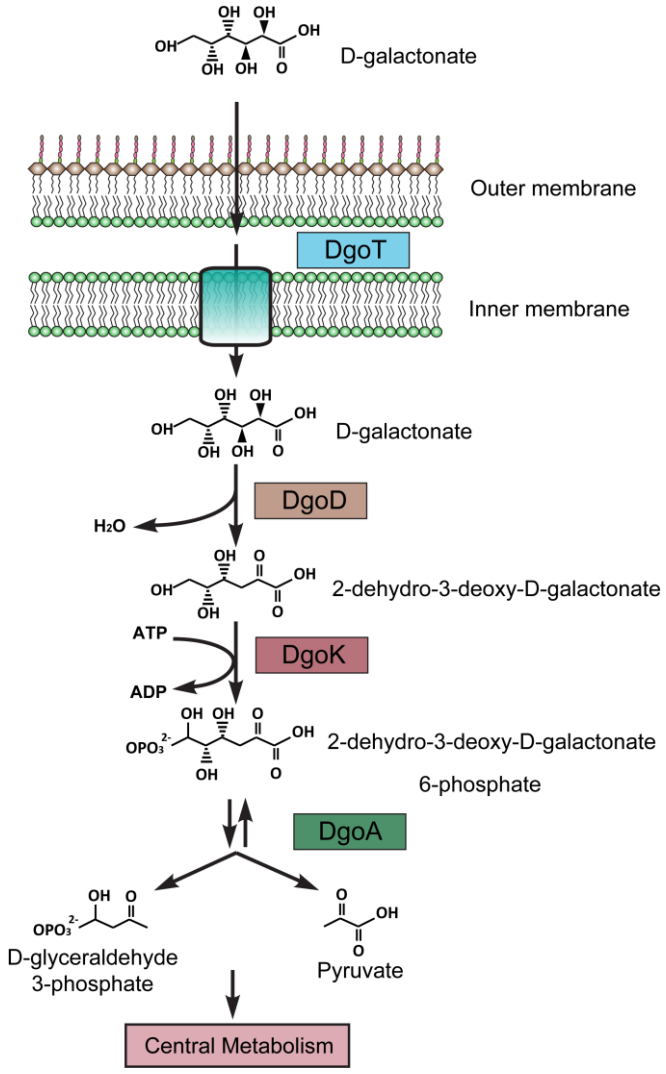

Figure 1.21 Pathway for D-galactonate metabolism in *E. coli* K-12. D-galactonate is transported across the cell membrane by DgoT, followed by its enzymatic dehydration, phosphorylation and finally aldol cleavage by the action of DgoD, DgoK and DgoA respectively

into D-glyceraldehyde 3-phosphate and pyruvate for further metabolism by glycolytic and tricarboxylic acid cycle pathways.

1.6.4 Regulation of D-galactonate metabolism

Almost all of the sugar acid metabolic genes are under dual regulation by specific as well as global regulators (Mandrand-Berthelot et al., 2004). The specialized transcriptional regulator is either transcribed as a part of the same operon or is a part of divergently transcribed transcriptional unit. However, in some cases the regulator gene is located at a distant site altogether. Here, in this section I have described the limited knowledge available on the regulation of D-galactonate metabolism prior to the current study.

1.6.4.1 Regulation of D-galactonate metabolism in *E. coli*

In the original classical biochemical study it was shown that the D-galactonate metabolic enzymes are induced only when the cells are grown in the presence of D-galactonate and the corresponding enzymatic activities are absent from the cells grown in other carbon sources such as glycerol, D-gluconate, or D-galactose, suggesting that the expression of Dgo enzymes is under regulation (Deacon and Cooper, 1977). In a follow-up mutagenesis based study, a mutant with constitutive expression of all the Dgo enzymes was isolated and subsequent genetic mapping showed that the corresponding mutations map very close to the structural *dgo* genes. Thus, DgoR a putative regulator of D-galactonate metabolism was proposed (Cooper, 1978). Later sequencing and annotation lead to identification of *dgoR* as the first gene of the putative *dgo* operon, encoding a protein with topology of a typical transcriptional regulator (Babbitt et al., 1995; Burland et al., 1993). In later studies, bioinformatics analysis of the transcriptional regulators with N-terminal wHTH DNA-binding domains suggested DgoR to be a member of GntR family (Rigali et al., 2002). Further, on the basis of its predicted all α-helical C-terminal

E/O domain, DgoR was placed in the FadR subfamily within the GntR family (Rigali et al., 2002). However, the experimental evidence for DgoR as a transcriptional regulator, including its regulatory nature and mechanism were still obscure. A global transposon mutagenesis based study in *S. enterica* serovar Typhimurium showed that a translational fusion of DgoT with LacZ harboring MudK transposon was expressed constitutively in Δ*dgoR* strain, and expression of DgoR from the plasmid abolished DgoT induction suggesting that DgoR is a repressor (Cenens et al., 2013). In the classical studies performed in *E. coli*, the activities of Dgo enzymes were drastically decreased when medium containing D-galactonate was supplemented with D-glucose, suggesting that the genes involved in D-galactonate metabolism are also subjected to carbon catabolite repression (Cooper, 1978).

The transcriptional regulators involved in the regulation of sugar acid metabolism bind either the substrate sugar acid or its metabolic intermediate as a cue to modulate expression of its regulon members (Bates Utz et al., 2004; Bouvier et al., 2019; Miwa and Fujita, 1988; Tutukina et al., 2016a). The original mutagenesis based studies found that the D-galactonate kinase and aldolase enzymatic activities are still induced if the dehydratase, the first enzymatic activity for D-galactonate metabolism is compromised (Cooper, 1978). Since, no D-galactonate metabolic intermediates are produced in the strain lacking the dehydratase activity, it was suggested that D-galactonate itself, rather than its metabolic intermediates was the inducer of the *dgo* operon. However, the direct evidence for only D-galactonate as the effector for DgoR was not available.

1.6.4.2 Regulation of D-galactonate metabolism in *Salmonella* by accessory factors

In *S. enterica* serovar Typhimurium LT-2 the β-galactosidase activity of a strain with *dgoA* interrupted with MudJ, a LacZ expressing transposon, was found to be increased

approximately 500-fold by PmrA, a response regulator of a two-component regulatory system PmrA-PmrB, which regulates modifications in LPS to tackle the environmental conditions faced *in vivo* such as antimicrobial peptides, low magnesium and mild acidic pH etc. Lack of canonical PmrA binding site upstream of *dgoA* suggested that PmrA mediated induction of *dgoA* may be indirect (Tamayo et al., 2002). Further, in another global transposon mutagenesis based study, to identify genes required for maintaining P22 phage in carrier state in *S. enterica* serovar Typhimurium, Pid (phage P22 encoded instigator of *dgo* expression) protein expressed from an uncharacterized late region of P22 genome was found to deregulate the *dgo* operon (Cenens et al., 2013). In yet another study, in *S. enterica* serovar Enteritidis, the *dgo* operon was reported to be induced in cells exposed to high temperature (45°C) and chicken egg-white (Baron et al., 2017). The induction of *dgo* operon by egg-white, which does not contain D-galactonate but D-glucose, a strong catabolite repressive carbon source, suggested that the *dgo* operon can be induced in a D-galactonate-independent manner as well, likely due to stressful conditions encountered during the experiment. The above examples show that at least in *Salmonella* sp. besides DgoR and cAMP-CRP, there might be auxiliary regulation on the *dgo* locus.

1.6.5 Physiological importance of D-galactonate metabolism

Multiple reports emphasize the role of *dgo* genes and D-galactonate metabolism in bacterial physiology. Exogenous D-galactonate is readily utilized as a sole source of carbon and energy by diverse microorganisms (Deacon and Cooper, 1977; Szumilo, 1981, 1983). Further, a comparison of metabolic models of 55 fully sequenced *E. coli* and *Shigella* strains to correlate their metabolic potential with adaptation to different environments, predicted 40 strains to be capable of using exogenous D-galactonate, further supporting the significance of this sugar acid as a relevant carbon and energy

source for these enteric bacteria (Monk et al., 2013). D-galactonate operon seems to have importance in adaptation to specific habitats. For example, *S. enterica* serovar Typhimurium strains can utilize D-galactonate as a carbon source while this capability is lost in *Salmonella enterica* serovar Typhi strains (Seif et al., 2018) Another study shows that metabolism of D-galactonate may help bacteria to adapt to a new niche; *Arthrobacter* strains isolated from soil lack *dgo* operon and hence cannot metabolize D-galactonate. Interestingly, *Arthrobacter arilaitensis*, a major bacterial inhabitant of cheese surface possesses functional *dgo* operon acquired by horizontal gene transfer from unknown gram-negative bacterium and can utilize exogenous D-galactonate. This differential ability to utilize D-galactonate has been suggested to help *A. arilaitensis* adapt to the cheese surface (Monnet et al., 2010).

Multiple global transcriptomic studies highlight the importance of D-galactonate metabolism in host-bacterial interactions. An Asymptomatic Bacteriuria (ABU) *E. coli* 83972 strain able to show long term growth in human urinary tract out-competes uropathognenic (UPEC) strains of *E. coli* in human urine. Comparison of global gene expression profile of the ABU strain grown in human urine revealed a significant induction of *dgo* genes as compared to strain grown in MOPS-glucose medium, indicating that D-galactonate may be used as a carbon source in human urine for growth (Roos et al., 2006). In another study, co-infection of cilantro leaves with *S. enterica* serovar Typhimurium SL1344 and a plant tissue macerating bacterial pathogen, *Dickeya dadantii* led to ~3000- fold increase in the *Salmonella* population as compared to ~60-fold when *D. dadantii* was not included suggesting that the maceration of plant tissue by *D. dadantii* facilitates the growth of *Salmonella* in the lesions. Transcriptome analysis of *Salmonella* grown on the leaf tissue macerated by *D. dadantii* showed a significant upregulation of *dgo* genes as compared to *Salmonella* grown in LB medium indicating

that metabolism of D-galactonate possibly generated by digestion of plant cell wall material by *D. dadantii* plays an important role in growth of *Salmonella* on plant soft-rot lesions caused by the tissue degrading *D. dadantii* (Goudeau et al., 2013). Similarly, induction of *dgoT* in plant-grown *E. chrysanthemi* 3937, a soft-rot causing plant pathogen, indicates the importance of D-galactonate metabolism in plant-bacterial interaction (Okinaka et al., 2002). In another study *S. enterica* serovar Typhimurium growing in murine macrophages showed induction of *dgo* genes indicating a connection between D-galactonate metabolism and growth in phagocytic cells (Eriksson et al., 2003). *S. enterica* serovar Enteritidis, the major cause of egg-borne infections in humans when exposed to egg white at high temperatures induces multiple genes to combat stressful conditions. Surprisingly, a 15- to 30- fold upregulation of *dgo* genes was also observed, indicating an additional uncharacterized role of D-galactonate metabolic genes in the bacterial physiology. Further, signature-tagged mutagenesis of *Salmonella enterica* serovar Choleraesuis, revealed that the strain mutated for *dgoT,* the D-galactonate transporter, is attenuated for paratyphoid in swines indicating is role in host-pathogen interaction (Ku et al., 2005).

In a recent *in vivo* study, an *E. coli* 536 strain originally isolated from a patient suffering from urinary tract infection was adapted to streptomycin-treated mouse gut for a year. The samples isolated from the mouse feces during the course of adaptation were sequenced to determine the mutations acquired in the *E. coli* genome for adjusting to the mammalian gut (Lescat et al., 2016). During the study 11 independent missense mutations in *dgoR* were isolated. Four of these mutant strains harbored additional missense mutations in *dgoD* while two strains having mutations in both *dgoR* and *dgoD* further acquired missense mutations in *dgoT*. Thus, mutations in *dgo* genes convergent at the gene level (5 exclusive mutations at the same position in *dgoR*) or at the metabolic

pathway level (additional mutations in *dgoD* and *dgoT*) accumulated during the adaptive process. The various mutant strains (*dgoR dgoD dgoT* > *dgoR dgoD* > *dgoR*) were found to grow faster in medium containing D-galactonate as the carbon source in that order as compared to the parent strain suggesting the role of *dgo* genes in bacterial colonization of the mammalian gut. Although literature survey does not suggest the presence of D-galactonate in the mammalian gut, microbes present in mammalian gut such as *S. maltophilia* (Saffarian et al., 2015), have the potential to produce D-galactonate as an intermediate of D-galactose metabolic pathway (Brechtel et al., 2002). Further, *D. hansenii*, one of the most dominant fungal species present in human offspring during the breast-feeding period (Schei et al., 2017) as well in human adults (Suhr and Hallen-Adams, 2015), has been demonstrated to produce D-galactonate in cheese through De Ley-Doudoroff pathway (Pham et al., 2019). Interestingly, D-galactonate produced by *D. hansenii* was found to be utilized by other bacteria sharing the niche such as *Brevibacterium aurantiacum* and *Hafnia alvei*. This example supports the hypothesis that D-galactonate produced by organisms known to inhabit mammalian gut can be used by other microbes.

The above examples illustrate the physiological importance of *dgo* genes as well as D-galactonate metabolic pathway for bacteria stressing the need to investigate the molecular details of D-galactonate metabolism as well as its regulation and their role in interaction with their hosts. Recent reports show that genetic variations in the regulators of carbon source metabolism can affect the interaction of bacteria including pathogens, with their hosts. For example, an epidemic ribotype of *C. difficile* was found to have acquired single point mutation in TreR, the repressor of trehalose metabolism, enabling the bacteria to sense and metabolize the disaccharide at a lower concentration and exhibit enhanced virulence (Collins et al., 2018). In another example single nucleotide

polymorphism in RafR the activator of raffinose metabolism (Rosenow et al., 1999), in *Streptococcus pneumoniae*, was found to result in differential metabolism of raffinose by blood and ear isolates as well as influence the tropism of the bacteria for lungs vs brain and ear (Minhas et al., 2019). These examples alongwith the *in vivo* adaptation study which identified missense mutations in *dgo* genes in gut-adapted *E. coli* strain (Lescat et al., 2016) suggest that the mutations in regulators of carbon source metabolism can have enormous contribution in host-pathogen interaction, making a compelling case for investigating the molecular details of regulation of D-galactonate metabolism.

1.7 Thesis objective

Recent reports showing that single point mutations in the regulators involved in regulating metabolism of carbon sources drastically affect host-bacterial interaction underscore the importance of studying carbon source regulation (Collins et al., 2018; Lescat et al., 2016; Minhas et al., 2019). Metabolism of sugar acids has been shown to play a critical role in bacterial physiology such as colonization of mammalian gut (Sweeney et al., 1996). Multiple studies from past couple of decades suggest that metabolism of D-galactonate plays an important role in physiology of bacteria inhabiting diverse habitats (Eriksson et al., 2003; Goudeau et al., 2013; Ku et al., 2005; Lescat et al., 2016; Okinaka et al., 2002; Roos et al., 2006). Although the structural *dgo* genes involved in the transport and metabolism of D-galactonate have been investigated, there is only limited information of its regulatory aspect.

In the current work, we have used various genetic, biochemical and bioinformatics approaches to investigate the molecular and functional details of regulation of D-galactonate metabolism in *E. coli* K-12 by DgoR. As our first objective, we validated that the *dgo* genes are required for metabolizing D-galactonate as sole carbon source by *E. coli* and further investigated the regulation exerted by DgoR on transcription of *dgo*

genes. Our findings from the first objective show that DgoR strongly represses transcription of *dgo* genes by binding two closely spaced inverted repeats overlapping the putative D-galactonate responsive *dgo* promoter. In our second objective we probed the molecular details of DgoR-DNA interaction. We tested several amino acids in the N-terminal wHTH domain of DgoR for their role in interaction with its cognate DNA both *in vitro* and *in vivo*. Several findings from our first two objectives firmly place DgoR in the FadR subfamily within GntR family of transcriptional regulators: DgoR is a majorly α-helical protein with GntR-type N-terminal wHTH domain and a predicted FadR-like all helical C-terminal FCD domain, binds [5'-TTGTA(G/C)TACA(A/T)-3'] operator sequence matching the signature of GntR family members that recognize inverted repeats [5'-(N)yGT(N)xAC(N)y-3'], and shares critical protein-DNA contacts conserved in the GntR family. Additionally, we identified features in DgoR that are otherwise less common in the regulators of GntR family. In the final objective of the study we have investigated the effector-binding characteristics of DgoR, where we report that D-galactonate itself, rather than its metabolic intermediate acts as the true effector of DgoR to relieve the DNA bound by DgoR. We also found a serine residue in C-terminal E/O domain of DgoR whose mutation makes DgoR less responsive to D-galactonate.

CHAPTER II

Materials and Methods

2.1 Materials

2.1.1 Chemicals and Reagents

All analytical grade chemicals used in this study were procured from commercial sources. Fine chemicals, media components, HPLC grade solvents and reagents were purchased from Sigma Aldrich, HIMedia, Merck Millipore India Ltd., Rankem, USB, and Difco. Enzymes: restriction enzymes, Calf Intestinal Phosphatase (CIP), T4 DNA ligase, T4 polynucleotide kinase and dNTPs were purchased from New England Biolabs. Standard Taq polymerase and Phusion DNA polymerase was purchased from New England Biolabs and Thermo Fisher Scientific. Antibodies: anti-FLAG (Sigma Aldrich), anti-His (Thermo Fischer Scientific), HRP-conjugated anti-MBP (New England Biolabs), HRP-conjugated anti-Mouse (Sigma Aldrich) were used. Gel-extraction kits and plasmid miniprep kits were obtained from Thermo Scientific or Qiagen. Sequencing services were obtained from 1st Base (Malaysia) or Eurofins (Bangalore, India). DNA/Protein molecular ladders were obtained from New England Biolabs or Thermo Fisher Scientific. All the carbon sources used in this study were procured from commercial sources (D-glucose from Fisher Scientific, and Sodium D-gluconate, D-galactose and glycerol from Sigma Aldrich). Calcium D-galactonate was purchased from MP Biomedicals and converted to D-galactonate crystals (see below).

2.1.2 Bacterial strains and plasmids

Table 2.1 provides a list of bacterial strains and plasmids used in the study. *E. coli* BW25113 and various deletion strains in this background were used for dilution spotting experiments, growth curve assays, qRT-PCR and fluorescence reporter assays. For deletion strains obtained from the Keio collection (Baba et al., 2006), either both independent clones and/or fresh transductants made by P1 transduction were analyzed to

rule out genetic errors. Strain DH5α was used for cloning in plasmids pACYC177, pRC10, and pMAL-c2. BL21(DE3) was used for protein expression and purification. Strain carrying deletion of the entire *dgo* operon was constructed using the λ Red-recombinase method. The Δ*dgoR* strain (derived by flipping out the kanamycin cassette from the *dgoR*::*kan* strain) was used since the Venus reporter construct carried a kanamycin marker.

Medium-copy promoter-less plasmid pACYC177 was used for making constructs for complementation experiments. IPTG-inducible plasmid pRC10 and L-arabinose-inducible pMAL-c2 were used as vectors for cloning *dgoR* for high-level expression and purification of C-terminally 6XHis-tagged DgoR (DgoR-6XHis) and N-terminally MBP-tagged DgoR (MBP-DgoR) respectively.

Table 2.1 Strains and plasmids used in this study

Strains/plasmids	**Relevant genotype**	**Source/reference**
Strains		
DH5α	F-, Δ(*argF-lac*)*169*, ϕ80d*lacZ58*(M15), *glnX44*(AS), λ^-, *rfbC1*, *gyrA96*(NalR), *recA1*, *endA1*, *spoT1*, *thiE1*, *hsdR17*	NEB
BL21(DE3)	F-, *lon-11*, Δ(*ompT-nfrA*)*885*, Δ(*galM-ybhJ*)*884*, *λDE3* [*lacI*, *lacUV5-T7 gene 1*, *ind1*, *sam7*, *nin5*], Δ*46*, $[mal^+]_{K-12}(\lambda^S)$, *hsdS10*	NEB
BW25113	F-, Δ(*araD-araB*)*567*, Δ*lacZ4787* (::rrnB-3), λ^-, *rph-1*, Δ(*rhaD-rhaB*)*568*, *hsdR514*	CGSC[a], (Baba et al., 2006)
dgoR::*kan*	BW25113 *dgoR*::*kan*, Kan^R	Keio collection, (Baba et al., 2006)
dgoK::*kan*	BW25113 *dgoK*::*kan*, Kan^R	Keio collection, (Baba et al., 2006)
dgoA::*kan*	BW25113 *dgoA*::*kan*, Kan^R	Keio collection, (Baba et al., 2006)
dgoD::*kan*	BW25113 *dgoD*::*kan*, Kan^R	Keio collection, (Baba et al., 2006)
dgoT::*kan*	BW25113 *dgoT*::*kan*, Kan^R	Keio collection, (Baba et al., 2006)

RC2132	BW25113 *dgo*::*kan*, Kan^R	This work
RC3067	BW25113 *dgoR*:3XFLAG-*kan*, Kan^R	YA
RC2069	Δ*dgoR* (Kan cassette flipped from BW25113 *dgoR*::*kan*)	This work
RC12022	BW25113 *attλ*::[Kan promoterless-*venus oriR6K*], Kan^R	AS
RC12023	Δ*dgoR attλ*::[Kan promoterless-*venus oriR6K*], Kan^R	AS
RC12018	BW25113 *attλ*::[Kan P_{dgo}-*venus oriR6K*], Kan^R	AS
RC12020	Δ*dgoR attλ*::[Kan P_{dgo}-*venus oriR6K*], Kan^R	AS
RC12035	BW25113 *attλ*::[Kan $P_{dgo(IR1L)}$-*venus oriR6K*], Kan^R	AS
RC12036	Δ*dgoR attλ*::[Kan $P_{dgo(IR1L)}$-*venus oriR6K*], Kan^R	AS
RC12040	BW25113 *attλ*::[Kan $P_{dgo(IR1R)}$-*venus oriR6K*], Kan^R	AS
RC12041	Δ*dgoR attλ*::[Kan $P_{dgo(IR1R)}$-*venus oriR6K*], Kan^R	AS
RC12028	BW25113 *attλ*::[Kan $P_{dgo(IR2L)}$-*venus oriR6K*], Kan^R	AS
RC12029	Δ*dgoR attλ*::[Kan $P_{dgo(IR2L)}$-*venus oriR6K*], Kan^R	AS
RC12031	BW25113 *attλ*::[Kan $P_{dgo(IR2R)}$-*venus oriR6K*], Kan^R	AS
RC12032	Δ*dgoR attλ*::[Kan $P_{dgo(IR2R)}$-*venus oriR6K*], Kan^R	AS
RC12043	BW25113 *attλ*::[Kan $P_{dgo(IR2L')}$-*venus oriR6K*], Kan^R	AS
RC12047	Δ*dgoR attλ*::[Kan $P_{dgo(IR2L')}$-*venus oriR6K*], Kan^R	AS
RC14053	BW25113 *attλ*::[Kan $P_{dgo(IR3L)}$-*venus oriR6K*], Kan^R	NK
RC14056	Δ*dgoR attλ*::[Kan $P_{dgo(IR3L)}$-*venus oriR6K*], Kan^R	NK
RC12025	BW25113 *attλ*::[Kan $P_{dgo(IR3R)}$-*venus oriR6K*], Kan^R	AS

RC12026	Δ*dgoR attλ*::[Kan $P_{dgo(IR3R)}$-*venus oriR6K*], Kan^R	AS
Plasmids		
pKD13	*oriR6K*, FRT-flanked Kan^R, pANTSγ, PS1 PS4, Kan^R	(Datsenko and Wanner, 2000)
pKD46	*pSC101 ori araC* repA101ts P_{araBAD}-λred, Amp^R	(Datsenko and Wanner, 2000)
pCP20	*pSC101 ori cI*857 λ-P_R *flp* ts, Amp^R, Cam^R	(Datsenko and Wanner, 2000)
pACYC177	*p15A ori*, Amp^R, Kan^R	NEB
pBS13	*dgo* promoter and *dgoR-6XHis* in pACYC177, Amp^R, Kan^R	This work
pBS14	*dgo* promoter and *dgoR*(L34A)-*6XHis* in pACYC177, Amp^R, Kan^R	This work
pBS15	*dgo* promoter and *dgoR*(R46A)-*6XHis* in pACYC177, Amp^R, Kan^R	This work
pBS22	*dgo* promoter and *dgoR*(D7A)-*6XHis* in pACYC177, Amp^R, Kan^R	This work
pBS24	*dgo* promoter and *dgoR*(S221N)-*6XHis* in pACYC177, Amp^R, Kan^R	This work
pBS35	*dgo* promoter and *dgoR*(S51L)-*6XHis* in pACYC177, Amp^R, Kan^R	This work
pBS36	*dgo* promoter and *dgoR*(R42C)-*6XHis* in pACYC177, Amp^R, Kan^R	This work
pBS37	*dgo* promoter and *dgoR*(T40I)-*6XHis* in pACYC177, Amp^R, Kan^R	This work
pRC10	*pBR322 ori*, -10 box of P_{trc} changed to P_{lac} in pTrc99a, Δ*Nco*I, Amp^R	(Chaba et al., 2007)
pBS2	*dgoR-6XHis* in pRC10, Amp^R	This work
pBS19	*dgoR*(L34A)-*6XHis* in pRC10, Amp^R	This work
pBS20	*dgoR*(R46A)-*6XHis* in pRC10, Amp^R	This work
pBS25	*dgoR*(D7A)-*6XHis* in pRC10, Amp^R	This work
pBS32	*dgoR*(S51L)-*6XHis* in pRC10, Amp^R	This work
pBS33	*dgoR*(R42C)-*6XHis* in pRC10, Amp^R	This work
pBS34	*dgoR*(T40I)-*6XHis* in pRC10, Amp^R	This work
pMAL-c2	*pBR322 ori*, leaderless *malE* MCS, Amp^R	NEB
pBS11	*dgoR* cloned in pMAL-c2, Amp^R	This work
pINT-ts	*oriR6K*, *Int*, Amp^R	Rao lab, (Haldimann and Wanner,

		2001)
pVenus	*oriR6K, MCS venus t0 attλ*, Kan^R	Rao lab, (Saini et al., 2009)
pAS1	*oriR6K, MCS* P_{dgo}-*venus t0 attλ*, Kan^R	AS
pAS2	*oriR6K, MCS* $P_{dgo(IR1L)}$-*venus t0 attλ*, Kan^R	AS
pAS3	*oriR6K, MCS* $P_{dgo(IR1R)}$-*venus t0 attλ*, Kan^R	AS
pAS4	*oriR6K, MCS* $P_{dgo(IR2L)}$-*venus t0 attλ*, Kan^R	AS
pAS5	*oriR6K, MCS* $P_{dgo(IR2R)}$-*venus t0 attλ*, Kan^R	AS
pAS6	*oriR6K, MCS* $P_{dgo(IR2L')}$-*venus t0 attλ*, Kan^R	AS
pAS7	*oriR6K, MCS* $P_{dgo(IR3L)}$-*venus t0 attλ*, Kan^R	AS
pAS8	*oriR6K, MCS* $P_{dgo(IR3R)}$-*venus t0 attλ*, Kan^R	AS
[a]Coli Genetic Stock Center 'This work' refers to strains/plasmids that were made as part of this thesis. 'AS', 'NK', and 'YA' refer to strains/plasmids obtained from Akshay Sangwan (Sangwan, 2017; Singh et al., 2019), Neeladrita Kundu (Singh et al., 2019), and Yatendra Arya (Arya, 2014), respectively.		

2.1.3 Primers

Primers used for plasmid construction, qRT-PCR and for confirmation of gene disruption are listed in Table 2.2. Primers were designed using Gene Runner program (version 3.04; Hastings Software, Inc.). De-salted primers were purchased from Integrated DNA technologies (IDT) or Sigma Aldrich.

Table 2.2 List of primers used in this study

Primers	Sequence (from 5' to 3')	Purpose
Primers used for deleting genes		
YA15	GTAAAGTAGAGAAGAACATACAGAGCACAAG GACTCTCCATGATTCCGGGGATCCGTCGACC	Forward primer for deleting *dgo* operon
YA16	GCACATATTCCACAGTTGAAGGATTAGCCAAC GCGCTTCACATCTGTAGGCTGGAGCTGCTTCG	Reverse primer for deleting *dgo* operon
Primers used for verification of deletion strains		
BS13	TACAAAGTTGCCGCGTTATG	Forward primer for verification of *dgo*::*kan*
BS53	TGGATTATGCGCAGTGGTG	Reverse primer for verification of *dgo*::*kan*
SAK1	GAGGCTATTCGGCTATGACTG	Forward primer specific to kanamycin cassette
SAK2	TTCCATCCGAGTACGTGCTC	Reverse primer specific to kanamycin cassette
Primers used for cloning and verification		
BS23	ACCGGAATTCGAAGGAGATATACATATGACTC TCAATAAAACCGATCGCATTGTCATTAC	Forward primer for cloning *dgoR*-6XHis in pRC10
BS24	CGTGGATCCTCA**GTGATGATGATGATGATG**T GTGATTTCCTTTAACCTTCGTGTCGAGC	Reverse primer for cloning *dgoR*-6XHis in pRC10
BS25	GCTGTGGTATGGCTGTGCAGG	Sequencing/verification primer for cloning in pRC10
BS26	GCCAGGCAAATTCTGTTTTATCAG	Sequencing/verification primer for cloning in pRC10
BS137	ACCGGAATTCATGACTCTCAATAAAACCGATC	Forward primer for cloning *dgoR* in pMAL-c2
BS138	ACGCGTCGACTCATGTGATTTCCTTTAACC	Reverse primer for cloning *dgoR* in pMAL-c2
BS54	GCGTACTGCGGTGATCAAC	Sequencing/verification primer for cloning in pMAL-c2
BS55	TTCAGGCTGCGCAACTGTTG	Sequencing/verification primer for cloning in pMAL-c2
BS97	ACCTGACGTCTATCTTTTGCCTGCGATAGCCCA G	Forward primer for cloning *dgo* promoter and *dgoR*-6XHis in pACYC177
BS24	CGTGGATCCTCA**GTGATGATGATGATGATG**T GTGATTTCCTTTAACCTTCGTGTCGAGC	Reverse primer for cloning *dgo* promoter and *dgoR*-6XHis in

		pACYC177
BS98	ATCAGTACCGACGGTGATATGG	Sequencing/verification primer for cloning in pACYC177
BS99	GGGTTATTGTCTCATGAGCGG	Sequencing/verification primer for cloning in pACYC177

Restriction sites are underlined.
Initiation codon of *dgoR* is highlighted in grey box.
6XHis tag sequence is shown in bold font.

Primers used for creating amino acid changes in DgoR		
BS42/ BS43	CGGCTGAGGCGGAAGCGTGTGAGGAGTTTGCA/ TGCAAACTCCTCACACGCTTCCGCCTCAGCCG	Internal mutagenic primers for creating L34A mutation
BS48/ BS49	CGCTTACCATGATCGCCAGCAACACACGAAGGTTAAAGGAAATC/ GATTTCCTTTAACCTTCGTGTGTTGCTGGCGATCATGGTAAGCG	Internal mutagenic primers for creating S221N mutation
BS93/ BS94	CGCGCAACATCATCGCAGAGGTGTTCCGTTCG/ CGAACGGAACACCTCTGCGATGATGTTGCGCG	Internal mutagenic primers for creating R46A mutation
BS95/ BS96	CTCTCAATAAAACCGCGCGCATTGTCATTACG/ CGTAATGACAATGCGCGCGGTTTTATTGAGAG	Internal mutagenic primers for creating D7A mutation
BS131/ BS132	CATCCGTGAGGTGTTCCGTTTGCTGATGGCGAAGCGGCTG/ CAGCCGCTTCGCCATCAGCAAACGGAACACCTCACGGATG	Internal mutagenic primers for creating S51L mutation
BS133/ BS134	GAGGAGTTTGCAACCTCGTGCAACATCATCCGTGAGGTG/ CACCTCACGGATGATGTTGCACGAGGTTGCAAACTCCTC	Internal mutagenic primers for creating R42C mutation
BS135/ BS136	CTGTGAGGAGTTTGCAATCTCGCGCAACATCATCCGTG/ CACGGATGATGTTGCGCGAGATTGCAAACTCCTCACAG	Internal mutagenic primers for creating T40I mutation
Mutated codon is underlined.		
Primers used for amplifying *dgo cis*-acting element for EMSA		
BS104	CGGGGTACCGCATTGTTCTTTTTGTGATC	Forward primer for amplification of *dgo cis*-acting element for EMSA

BS105	CCGGAATTCGACAATGCGATCGGTTTTATTGAG	Reverse primer for amplification of *dgo cis*-acting element for EMSA
Primers used for qRT-PCR		
BS58	CGACACTGACGTACTGCAATG	Forward primer for dgoR
BS59	CGATTCAATCTGCGCCAG	Reverse primer for dgoR
BS60	ACGTCGGCTGGAAAGTTG	Forward primer for dgoK
BS61	ATTGTGTTTCTTCGCCGC	Reverse primer for dgoK
BS73	GTTGAAATCCCGCTGAATTCC	Forward primer for dgoA
BS74	TATTGGGCGTAACGATGAGCTG	Reverse primer for dgoA
BS75	GCACAAACGCATATTCCACTG	Forward primer for dgoD
BS76	AAGGGTCACGTCATAGGCTTCTG	Reverse primer for dgoD
BS77	CATTTGTTATGTCGACCGCG	Forward primer for dgoT
BS78	ATAAGTCACGCGAGAACCTACG	Reverse primer for dgoT
RS3	GTCATGCCAACCAAAATTCC	Forward primer for gyrA
RS4	ATGTGTTCCATCAGCCCTTC	Reverse primer for gyrA
RS5	TCTACCGGTTCGCTTTCACT	Forward primer for recA
RS6	GCGTGTTCAGCATCGATAAA	Reverse primer for recA

2.1.4 Antibiotics

Ampicllin (100 mg/ml) and Kanmycin (30 mg/ml) stocks were prepared in autoclaved type I MQ water; filter-sterlized using 0.22μm syringe filters (MDI) and stored at -20°C.

2.1.5 Media

All the media, buffers and stock solutions were prepared using type I MQ water from Millipore synergy system. Sterilization of the various solutions was done as recommended, either by autoclaving at 15 lb/inch2 (psi) pressure at 121°C for 15 minutes, or by using membrane (Millipore)/ syringe filters (MDI) of pore size (0.45 μm or 0.22 μm).

2.1.5.1 Lysogeny-Broth (LB)

5 g/liter yeast extract, 10 g/liter Bacto Tryptone, and 5 g/liter NaCl. The components were dissolved in 1 liter type I MQ water and autoclaved.

2.1.5.2 Rich media

5 g/liter yeast extract, 10 g/liter Bacto Tryptone, 2 g/liter D-glucose, and 5 g/liter NaCl. The components were dissolved in 1 liter type I MQ water and autoclaved.

2.1.5.3 M9 minimal medium

The following stocks required for M9 minimal medium were prepared separately.

a) 10X M9 salts

The following components were dissolved in 1 liter type I MQ water; sterilized by autoclaving and stored at room temperature.

Table 2.3 Composition of 10X M9 salts

Component	**Amount (per liter)**
Na_2HPO_4	53 g
KH_2PO_4	30 g
NaCl	5 g
NH_4Cl	10 g

b) 250X vitamin stock

The following vitamins were dissolved in 200 ml autoclaved type I MQ water; filter sterilized and stored at -20°C.

Table 2.4 Composition of 250X vitamin stock

Component	**Amount (for 200 ml)**
Biotin	100 mg
Nicotinamide	100 mg
Riboflavin	10 mg
Thiamine	100 mg

c) 1 M $MgSO_4$

24.65 g MgSO4.7H2O was dissolved in 100 ml type I MQ water; sterilized by autoclaving and stored at room temperature.

To prepare M9 minimal medium, the above components were added according to following composition in 1000 ml autoclaved type I MQ water.

Table 2.5 Composition of 1X M9 minimal medium

Component	**Working concentration (per liter)**
10X M9 Salts	100 ml
1M $MgSO_4.7H_2O$	1 ml
250X vitamins	2.5 ml

2.1.5.4 Carbon sources

Stocks of following carbon sources used in this thesis were prepared in type I MQ water separately; autoclaved and stored at room temperature. 20% (w/v) D-glucose, 1M D-gluconate, 1 M D-galactose, 50% glycerol and 100 mM D-galactonate. Wherever required, M9 minimal medium was supplemented with one of the carbon sources: D-glucose (10 mM), Sodium D-gluconate (10 mM), D-galactose (10 mM), D-galactonate (10 mM), glycerol (0.4% v/v), or both glycerol and D-galactonate.

Media were solidified using 1.5% (w/v) bacto agar (Difco) when needed. When required, ampicillin (100 μg/ml) or kanamycin (30 μg/ml) was used.

2.1.6 Buffers and solutions for EMSA

Following buffers and solution were prepared for setting up EMSA.

a) 40% Acrylamide:bis-Acrylamide solution

The following components were dissolved in 100 ml type I MQ water and stored in amber colored bottle at 4°C.

Table 2.6 Composition of 40% (w/v) Acrylamide:bis-Acrylamide solution

Component	Amount (for 100 ml)
Acrylamide	4 g
N,N'-methylene bis-Acrylamide	1.06 g

b) 4X-Separating buffer (1.5 M Tris HCl)

9.075 g Tris-base was dissolved in 30 ml type I MQ water and its pH was adjusted to 8.8. The volume was made up to 50 ml and stored at 4°C.

c) 0.5 M EDTA

14.6 g of tetra sodium EDTA was dissolved in 70 ml type I MQ water and its pH was adjusted to 8.8. The final volume was made up to 100 ml and stored at 4°C.

d) Tris Borate EDTA (TBE) Buffer

Following components were dissolved in 700 ml type I MQ water and its pH was adjusted to 8.3. Final volume was made up to 1 liter and used immediately.

Table 2.7 Composition of 0.5X TBE buffer

Component	Amount (per liter)
Tris-base	5.4 g
Boric acid	2.7 g
0.5 M EDTA	2 ml

e) 8% Native PAGE gel

Following components were added to a 50 ml centrifuge tube mixed gently. APS and TEMED were added and the mixture was immediately poured in the BioRad plate assembly. The gel was allowed to polymerize and pre-run at 4°C for two hours using 0.5X TBE at 60V using BioRad protein electrophoresis apparatus before loading samples.

Table 2.8 Composition of 8% native PAGE gel

Component	Amount (for 1 gel)
40% Acrylamide: bis-Acrylamide	1.4 ml
4x Separating buffer	1.75 ml
50% glycerol	1.75 ml
Type I MQ water	2.1 ml
10% APS	35 µl
TEMED	7 µl

2.2 Methods

2.2.1 Preparation of D-galactonate crystals

D-galactonate was prepared from its calcium salt as described elsewhere (Isbell and Frush, 1933). Briefly, equivalent amounts of calcium D-galactonate (2.6 g) and Oxalic acid (520 mg) were added to 3 ml boiling water and vortexed for 3 minutes. The milky solution was filtered through a 0.22 µm PVDF filter (MDI) and the filtrate was immediately transferred to 4°C for 10-15 minutes. Excess water was removed and

crystals of D-galactonate were dried overnight at room temperature. D-galactonate crystals were stored at room temperature. The purity of D-galactonate preparation was assessed by NMR spectroscopy. D-galactonate crystals (5-10 mg) were dissolved in 500 µl DMSO-d6 (Sigma Aldrich) and analyzed by recording ^{13}C-NMR spectra on a 400 MHz Bruker Biospin Advance III FT-NMR spectrometer analyses.

2.2.2 Recombinant DNA techniques and gel electrophoresis

Protocols for PCR, manipulation of DNA, cloning, agarose gel electrophoresis, SDS-PAGE, and transformation were modified from 'Sambrook Molecular cloning: A Laboratory manual'. Plasmid isolation, PCR product purification and agarose gel extraction of DNA were performed with commercial kits procured from Thermo scientific or Qiagen as per manufacturer's instructions.

2.2.3 Deletion of *dgo* operon

The entire *dgo* operon was deleted from *E. coli* BW25113 strain using the λ Red-recombinase method (Datsenko and Wanner, 2000).For the deletion of *dgo* operon, kanamycin cassette was PCR amplified from the plasmid pKD13 using primers YA15 and YA16. The amplified cassettes were transformed into BW25113 expressing λ Red-recombinase enzymes from the pKD46 helper plasmid. Recombinants were selected on LB-kanamycin plates. Strains were PCR verified for deletion using primers specific to the antibiotic resistance marker (SAK1 and SAK2) and site-specific primers, BS13 and BS53 for *dgo* operon deletion. *dgo* operon deletion was subsequently transduced into a clean BW25113 background using P1 phage, yielding strains RC2132.

2.2.4 Construction of plasmid constructs for complementation and protein purification

For complementation experiments, C-terminally 6XHis-tagged *dgoR* was cloned in the medium-copy promoter-less plasmid pACYC177. The coding region of *dgoR* along with its putative promoter (446 bp upstream of the *dgoR* start codon) was PCR amplified using primers BS24 and BS97. The resulting PCR fragment was digested with *Aat*II and *Bam*HI, and ligated at the corresponding restriction sites of pACYC177 to generate plasmid pBS13. Various *dgoR* mutants in pACYC177 were created by overlap extension PCR using pBS13 as template, external primers BS24 and BS97, and the internal mutagenic primers listed in Table 2.2.

Plasmid pBS2 was constructed for high-level expression and purification of C-terminally 6XHis-tagged DgoR (DgoR-6XHis). Briefly, the coding region of *dgoR* was PCR amplified using primers BS23 and BS24. The PCR product was digested with *Eco*RI and *Bam*HI, and ligated at the corresponding restriction sites of IPTG-inducible plasmid pRC10. For expression and purification of various DgoR mutant proteins, mutations were generated by overlap extension PCR using pBS2 as template, external primers BS23 and BS24, and the internal mutagenic primers listed in Table 2.2. Plasmid pBS11 was constructed for expression and purification of MBP-DgoR by amplifying the coding region of *dgoR* using primers BS137 and BS138. The PCR product was digested with *Eco*RI and *Sal*I, and ligated at the corresponding sites of pMAL-c2 to create an in-frame fusion with MBP.

2.2.5 Growth and maintenance of bacteria

All strains were maintained as frozen stocks containing glycerol (15% v/v) stored at -80°C. Whenever required, desired strain was streaked on appropriate media and single

colonies were used for setting up cultures. Besides the specific requirement of the protocol, cultures were incubated at 37°C. Primary cultures were grown in 3 ml LB liquid medium. Secondary cultures were set-up either in LB or in M9 minimal medium containing the desired carbon source with an initial OD of ~0.01, unless indicated otherwise.

2.2.6 P1 Lysate preparation and transduction

P1 lysate was prepared and used for moving selectable mutations of interest into desired genetic background according to protocol described in (Miller, 1972) with minor modifications.

2.2.6.1 P1 lysate preparation:

Overnight cultures of the donor strains grown in 3 ml LB at 37°C were diluted 1:100 into 5 ml LB supplemented with 5 mM $CaCl_2$ and incubated at 37°C for 90 minutes. 20 µl of P1 phage stock was added and incubation at 37°C was continued till the culture lysed. 50 µl chloroform (Sigma Aldrich) was added to the lysed cultures and shaken for 2 minutes. Cell debris was separated by centrifugation at room temperature. The clear supernatant (P1 lysate) was transferred to fresh centrifuge tubes containing 50 µl chloroform and stored at 4°C.

2.2.6.2 P1 transduction:

1ml overnight culture of recipient strain in LB was pelleted and re-suspended in 500µl solution containing 5 mM $MgSO_4$ and 10 mM $CaCl_2$. To 100 µl aliquot of re-suspended cells 100µl of desired P1 lysate was added, while another 100 µl aliquot was left untreated. Both aliquots were incubated at 30°C in a water bath for 30 minutes. Cells were pelleted down and re-suspended in 1 ml LB supplemented with 10 mM sodium

citrate and incubated at 37°C for 1 hour in a water bath. Cells were pelleted, washed with 1 ml LB twice and re-suspended in 100 µl LB. Re-suspended cells were spread on LB agar plates supplemented with 10 mM sodium citrate and appropriate antibiotic and incubated at 30°C for 16-18 hours. 60 µl of P1 lysate was spread as a control. Colonies obtained after incubation were re-streaked twice on LB agar plates with sodium citrate and suitable antibiotic and incubated at 37°C. Transductants were confirmed by colony PCR using primers mentioned in Table 2.2.

2.2.7 Dilution spotting

Overnight cultures of various strains in LB were pelleted and re-suspended in M9 minimal medium without any carbon source and normalized to a similar OD_{450}. 5 µl of serially diluted re-suspended cells were spotted on M9 minimal agar plates containing the carbon source of choice. Plates were incubated at 37°C and imaged using the Gel Doc XR^{+} imaging system from Bio-Rad.

2.2.8 Growth curves

Cells grown overnight in LB were pelleted, washed and re-suspended in M9 minimal medium. For growth curves in shake flasks, cells were re-inoculated in 25 ml M9 minimal medium supplemented with the desired carbon source (in 125 ml flasks) to an initial OD_{450} of ~0.01. Cultures were incubated in water bath (New Brunswick Scientific) maintained at 37°C, samples were withdrawn at various time points and OD_{450} was measured. For growth curves in 96-well plates (Falcon, Corning technologies), cells were re-inoculated in 200 µl M9 minimal medium supplemented with the desired carbon source to a starting OD_{450} of ~0.03, using a robotic liquid handling system (Tecan). Plates were incubated in a shaker at 37°C, and OD_{450} of the cultures was measured at hourly intervals (Tecan Infinite M200 monochromator). The incubator shaker and microplate

reader were integrated with the liquid handling system, and the transfer of plates between shaker and reader was automated.

2.2.9 RNA isolation, cDNA preparation and quantitative RT-PCR

Secondary cultures were grown in 25 ml M9 minimal medium with desired carbon source to OD_{450} ~0.5. 20 ml samples were collected. Chilled ethanol containing 5% water-saturated phenol was added to the samples followed by centrifugation for 2 min. Cell pellets were stored at -80°C after flash freezing in liquid nitrogen until use. Total RNA was extracted by hot phenol method as described previously (Agrawal et al., 2017). Cell pellets were suspended in 1300 µl lysis solution (320 mM Na acetate pH 4.6, 8% SDS and 16 mM EDTA, Ambion). The lysate was added to 2 ml phenol (Ambion) and incubated at 65°C for 5 min with intermittent vortexing. The samples were then chilled on ice for 5 min and centrifuged for 10 min at 14,000 rpm at 4°C. The supernatant was extracted with phenol:chloroform mixture twice and RNA was precipitated with 2.5 volumes of pre-chilled absolute ethanol (Merck). RNA was pelleted and washed with 70% ethanol and air-dried. RNA pellet was suspended in 35 µl nuclease free water and treated with Turbo DNA-free Turbo DNAse (Applied Biosystems) according to the manufacturer's protocol to avoid genomic DNA contamination. 4 µg RNA was used to prepare cDNA using SuperScript III Reverse Transcriptase (Applied Biosystems) using protocol described previously. qRT-PCR was carried out with Power Sybrgreen PCR master mix (Applied Biosystems) as per the manufacturer's recommended protocol using 50 ng cDNA as template and 5 pmol of each forward and reverse primer listed in Table 2.2. qRT-PCR was performed using Quant Studio 6 Flex system (Applied Biosystems). Data were analyzed as described previously using *recA* and *gyrA* as internal controls.

2.2.10 Western blotting

The expression of C-terminally 3XFLAG-tagged DgoR (DgoR-3XFLAG), DgoR-6XHis, and MBP-DgoR was monitored by Western Blot analysis. Samples were electrophoresed on SDS-PAGE and transferred to nitrocellulose membrane (Bio-Rad). The membrane was blocked with 5% (w/v) skimmed milk prepared in TBST buffer (pH 7.6) at 4°C overnight and probed with anti-FLAG (1:1000, Sigma Aldrich) or anti-His (1:1000, Thermo Fisher Scientific) primary antibody and HRP-conjugated anti-Mouse (1:5000, Sigma Aldrich) secondary antibody. Wherever required anti-MBP conjugated to HRP (1:5000, New England Biolabs) was used. Blots were developed with the SuperSignal West Dura Extended Duration Substrate (Pierce). Signal was captured with LAS4000 (GE healthcare) or on an X-ray film (Fujifilm).

2.2.11 Protein overexpression and purification

2.2.11.1 Overexpression and purification of DgoR-6XHis

BL-21(DE3) bearing plasmids constructed for overexpression of WT or mutant DgoR-6XHis proteins (Table 2.1. described previously) were cultured in 400 ml LB at 37°C till OD_{600} reached ~0.5 and induced with 0.5 mM IPTG overnight at 18°C. Cells were harvested by centrifugation at 10,000X g for 10 min at 4°C and stored at -80°C until use. Cell pellet was re-suspended in 20 ml lysis buffer (50 mM sodium phosphate (pH 7.5), 300 mM NaCl, 1 mM PMSF and 20 mM imidazole) and disrupted using sonicator (QSonica, 20 Amplitude, 15 sec On and 30 sec off for 5 minutes) on ice. The cell debris was removed by centrifugation at 15,000 X g for 10 min at 4°C and the supernatant was incubated with 1 ml Co-NTA beads (Pierce) on ice for 90 min with continuous shaking. All purification procedures were carried out with pre-chilled buffers. Beads were washed with 40 bed volumes of wash buffer A (50 mM sodium phosphate (pH 7.5), 1 M NaCl

and 20 mM imidazole), followed by washing with 40 bed volumes of wash buffer B (50 mM sodium phosphate (pH 7.5), 300 mM NaCl and 60 mM imidazole), and finally the protein was eluted using 5 ml elution buffer (50 mM sodium phosphate (pH 7.5), 300 mM NaCl and 500 mM imidazole). The eluted protein was dialysed against buffer protein dialysis buffer (20 mM Tris (pH 8.0), 300 mM NaCl, 1 mM DTT and 10% (w/v) glycerol) at 4°C.

2.2.11.2 Overexpression and purification of MBP-β-gal α fragment and MBP-DgoR

BL-21 DE3 bearing either pMAL-c2 or pBS11 (described previously) was cultured in 400 ml rich media at 37°C till OD_{600} reached ~0.5 and induced with 1 mM IPTG for 4 hours at 37°C. The cells were harvested by centrifugation and stored at - 80°C until use. Cell pellet was re-suspended in column buffer (20 mM HEPES (pH 7.5), 300 mM NaCl and 1 mM EDTA) according to 5 ml/g cell pellet and lysed by sonication. Cell debris was removed by centrifugation and the supernatant was incubated with 200 µl amylose resin (New England Biolabs) for 90 min at 4°C. Amylose beads were washed with five bed volumes of column buffer and finally eluted with one bed volume of column buffer containing 10 mM maltose (Fisher Scientific).

Purity of the proteins was assessed by SDS-PAGE and stored at -80°C until use. The proteins were quantified by Bradford reagent, using bovine serum albumin as standard, as per the manufacturer's protocol (Sigma Aldrich).

2.2.12 Circular dichroism

Far-UV CD spectra of WT and mutant DgoR-6XHis were measured in a Chirascan spectrophotometer (Applied Photophysics). The spectra were collected at 25°C using 5 µM DgoR-6XHis in a quartz cuvette with a 1 mm path length. Spectra were recorded in

triplicate within 200 nm to 280 nm range with a step size of 1 nm. The average values of each spectrum were corrected for contribution of buffer. The millidegree ellipticity at each point was plotted as a function of wavelength. Deconvolution of CD spectra was performed using the BeStSel server (Micsonai et al., 2018; Micsonai et al., 2015). Secondary structure assignment to modelled structure of DgoR was made by using 2Struc server using DSSP method (Kabsch and Sander, 1983; Klose et al., 2010).

2.2.13 Protein cross-linking

WT or mutant DgoR-6XHis proteins at 10 μM final concentration were mixed with increasing concentrations of glutaraldehyde (50% w/v, Sigma Aldrich) in a 20 μl reaction prepared in protein dialysis buffer. Samples were incubated at room temperature for 5 min. The reaction was terminated by boiling the sample in reducing SDS sample buffer at 95°C for 10 min. Samples were run on a 15% SDS-PAGE and stained with Coomassie brilliant blue.

2.2.14 Protein pull-down

DgoR-6XHis was overexpressed in BL-21 DE3 cells as mentioned previously and the cells were harvested by centrifugation. The cell pellet was re-suspended in lysis buffer (50 mM sodium phosphate pH 7.5, 300 mM NaCl, 1mM PMSF and 20 mM imidazole) and lysed by sonication. Cell debris was removed by centrifugation and the supernatant was incubated with 60 μl Co-NTA beads (Thermo Scientific) at 4°C for 90 min with gentle shaking. Beads were washed with 35 bed volumes of chilled lysis buffer and equally split into two fractions. Equimolar amounts of purified MBP-DgoR or MBP-β-gal α fragment were added to individual fractions and a portion from each suspension was saved as ‘input’. The remaining suspension was incubated at 4°C for 20 min. The beads were then washed with 25 bed volumes of column buffer (20 mM Tris pH 7.4, 200 mM

NaCl and 1 mM EDTA). Finally, the proteins were eluted with 120 µl elution buffer (50 mM sodium phosphate pH 7.5, 300 mM NaCl and 300 mM imidazole) and saved as 'pull-down'. Input and pull-down samples were boiled at 95°C for 10 min in reducing Laemelli buffer and subjected to 15% SDS-PAGE followed by western blotting.

2.2.15 Electrophoretic Mobility Shift Assay (EMSA)

Primer BS105 was 5' end-labeled with P^{32} using T4 polynucleotide kinase (NEB) and [γ - ^{32}P]dATP. 194 bp wild-type or mutated *dgo cis*-acting elements were PCR amplified with primers BS104 and ^{32}P-labeled BS105 using appropriate pVenus constructs (refer to TableXX) as the template. The 5' end-labeled DNA fragments were gel purified using QIAquick Gel Extraction Kit (Qiagen) and quantified. 20 µl EMSA reactions were set up in protein dialysis buffer (20 mM Tris pH 8.0, 300 mM NaCl, 10% glycerol, 1mM DTT) supplemented with 10 mM $MgCl_2$ and 1 mM PMSF. In each reaction, 20 nM WT or mutated labeled DNA was incubated with purified WT or mutant DgoR-6XHis in the presence of 1 µg herring sperm DNA (Promega) at 27°C for 30 min. The samples were loaded onto a pre-run 8% native polyacrylamide gel and electrophoresed at 60 V for 6 hours at 4°C in 0.5X TBE buffer. The electrophoresed gel was exposed to a phosphorimager screen overnight, and the band patterns were visualized using a PhosphorImager (Fuji or Bio-Rad). To test for the possible effectors, DgoR-6XHis was incubated with various carbohydrates and D-galactonate catabolic intermediates at 37°C for 20 min before incubation with DNA.

2.2.16 Fluorescence reporter assay

1 ml overnight cultures of reporter strains set up in LB were pelleted and re-suspended in 1 ml M9 minimal medium without any carbon source. Re-suspended cells were normalized to OD_{450} 4 and used to inoculate 200 µl M9 minimal media (initial OD_{450}

0.03) supplemented with glycerol alone or glycerol and D-galactonate in a 96-well, black, clear-bottom plate (Costar, Corning). The plates were incubated at 37°C with shaking in a shaker-incubator integrated with liquid handling system (Tecan). Fluorescence was measured in exponential phase with microplate reader (Tecan Infinite M200 monochromator) in top mode with excitation and emission wavelength 498 nm and 568 nm, respectively. OD_{450} and fluorescence measurements were corrected for background. Averaged fluorescence normalized to OD_{450} from three biological replicates was plotted in a bar graph along with their standard deviation.

2.2.17 Bioinformatics analysis

The amino acid sequence of full-length DgoR was submitted to the Robetta server (Kim et al., 2004) for model building. The model build by Robetta was used for analysis. The PDB database ws searched for proteins with a structure similar to that of the N-terminal wHTH domain of DgoR (from amino acids 1 to 73) using the DALI server (Holm and Laakso, 2016). The DALI server was also used to generate structure-based sequence alignment of DgoR with selected members of the GntR family. ICM-Browser (www.molsoft.com) was used for model visualization, structure comparisons and analysis, and preparation of related figures. Sequences corresponding to ~300 bp upstream of *dgoR* initiation codon and 194 bp *dgo cis*-acting element used for constructing reporter constructs were analyzed for promoter elements using bTSSfinder (Shahmuradov et al., 2017) and BPROM (Solovyev and Salamov, 2011). The 194 bp *dgo cis*-acting element was used for prediction of cAMP-CRP binding sites using Virtual footprint program (Münch et al., 2005).

CHAPTER III

DgoR negatively regulates D-galactonate metabolism in *E. coli*

3.1 Introduction

The common gut bacterium *E. coli* can utilize a variety of sugar acids, i.e., hexonates, hexuronates, hexuronides, and aldarates, as carbon and energy sources (Mandrand-Berthelot et al., 2004). Utilization of sugar acids has been implicated in the colonization of *E. coli* in the mammalian gut. The gut microbiota liberates sugar acids from polysaccharides present in nutrients ingested by the host and the mucosal layer that lines the intestinal epithelial cells (Conway and Cohen, 2015; Peekhaus and Conway, 1998; Sweeney et al., 1996). Certain gut microbes also produce sugar acids from simple sugars as catabolic intermediates of metabolism (Brechtel et al., 2002; Hommes et al., 1989). Antibiotic treatment has also been reported to induce the host to oxidize sugars present in the gut into sugar acids (Faber et al., 2016). *E. coli* degrades sugar acids via the Entner-Doudoroff (ED) or Ashwell pathway into glyceraldehyde 3-phosphate and pyruvate, which further enter the central metabolism through glycolysis and the tricarboxylic acid cycle, respectively (Ashwell et al., 1958; Mandrand-Berthelot et al., 2004; Peekhaus and Conway, 1998).

In *E. coli*, the sugar acid metabolic pathways are regulated by specific transcriptional regulators belonging to CdaR (e.g., CdaR, D-glucarate/galactarate), GntR (e.g., ExuR, D-galacturonate/D-glucuronate; UxuR, D-glucuronate/β-D-glucuronides/D-gluconate; LgoR, L-galactonate), IclR (e.g., KdgR, 2-keto-3-deoxy-D-gluconate), LacI (e.g., GntR, D-gluconate/L-Idonate; IdnR, L-Idonate/D-gluconate), or TetR (e.g., UidR, β-D-glucuronides) families of transcriptional regulators. Majority of these regulators exert negative control over their regulon members, although IdnR and CdaR are known to mediate a positive regulation over their respective metabolic pathways. Besides their dedicated specific regulators, sugar acid metabolism is positively regulated by cAMP-CRP mediated carbon catabolite repression system (Mandrand-Berthelot et al., 2004).

D-Galactonate, a hexonate sugar acid, was first reported as a carbon source for *E. coli* in studies conducted in the 1970s (Deacon and Cooper, 1977). In these studies, through chemical mutagenesis screening coupled with biochemical assays and genetic mapping, it was shown that D-galactonate is metabolized via modified ED pathway. Briefly, D-galactonate is transported into the cytoplasm by a putative major facilitator superfamily (MFS) transporter, DgoT and dehydrated to 2-dehydro-3-deoxy-D-galactonate by the D-galactonate dehyradatase, DgoD. The dehydrated product is phosphorylated by a kinase, DgoK, in an ATP dependent manner to form 2-dehydro-3-deoxy-D-galactonate 6-phosphate. Finally, the phosphorylated product is cleaved by an aldolase, DgoA to form glyceraldehyde 3-phosphate and pyruvate, which subsequently enter the central metabolism (Fig. 3.1A) (Cooper, 1978). Deacon and Cooper observed high activity of the D-galactonate catabolic enzymes in the crude lysates prepared from *E. coli* grown in D-galactonate but not in other carbon sources, viz., D-glucose, D-gluconate or glycerol, suggesting that the D-galactonate catabolism is under specific regulation (Deacon and Cooper, 1977). In a follow-up study by Cooper, mutations in the genes encoding D-galactonate metabolic enzymes and the transporter were found to be in close proximity. In the same study a mutation in the putative promoter was also isolated that simultaneously stopped the synthesis of all the catabolic enzymes of D-galactonate pathway. Based on these observations, it was proposed that the *dgo* genes constitute a putative operon (Cooper, 1978). Later, whole genome sequencing led to the annotation of this putative operon as *dgo* operon consisting of five genes; *dgoR*, *dgoK*, *dgoA*, *dgoD* and *dgoT* in that order (Fig. 3.1B) (Babbitt et al., 1995; Burland et al., 1993). In the study by Cooper a mutation mapping close to the predicted *dgo* operon was also identified which resulted in the constitutive expression of all three Dgo enzymes; thus a regulator of D-galactonate metabolism, DgoR was suggested (Cooper, 1978). In later studies,

bioinformatics analysis of transcriptional regulators harboring N-terminal winged helix-turn-helix (wHTH) DNA-binding domains predicted DgoR to be a member of the GntR family (Rigali et al., 2002). On the basis of the similarity of the C-terminal effector-binding and oligomerization (E-O) domain, DgoR has been placed in the FadR subfamily within the GntR family (Jain, 2015; Rigali et al., 2002).

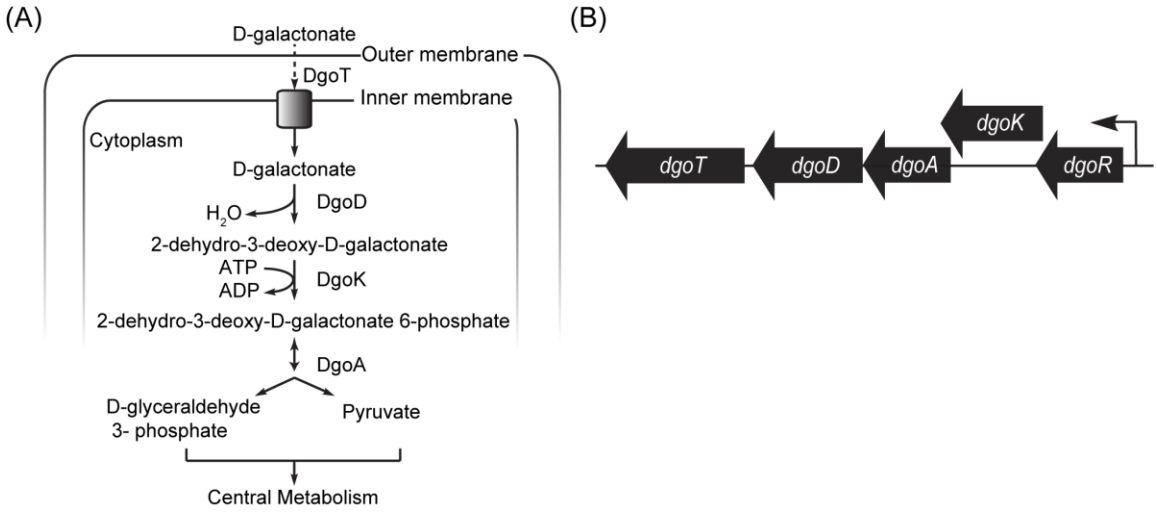

Figure 3.1 D-galactonate metabolism in *E. coli* K-12. (A) Pathway of D-galactonate transport and degradation in *E. coli* K-12. (B) A schematic of the chromosomal organization of the putative *dgo* operon in *E. coli* K-12. Filled arrows (not drawn to scale) show the direction of *dgo* genes, and the bent arrow indicates the direction of transcription.

D-galactonate is a widely prevalent aldonate sugar acid. Bacteria such as *Stenotrophomonas maltophilia*, an inhabitant of intestinal crypts and mucous membranes of mammalian hosts (Saffarian et al., 2015); *Azotobacter vinelandii*, a free living soil bacterium (J.G., 1903); *Gluconobacter liquefaciens*, a dweller of sugar-rich habitats (Gupta et al., 2001); and *Pseudomonas saccharophila*, a mud bacterium (Gomila et al., 2007), enzymatically produce D-galactonate as an intermediate in D-galactose catabolic pathway (A.H., 1961; Brechtel et al., 2002; De Ley and Doudoroff, 1957; Wong and Yao, 1994). Several studies have reported the presence of D-galactonate in mammalian tissues as well as in body fluids especially when individuals are fed a D-galactose rich diet.

Although humans produce D-galactonate in small quantities, its levels increase in galactosemic patients, who are deficient in enzyme/s required for D-galactose assimilation. In such individuals, D-galactose is converted into D-galactonate and galactitol by alternative metabolic pathways where only a small fraction of D-galactonate is metabolized in the liver and the majority of the non-metabolized fraction is excreted in the urine. (Ficicioglu et al., 2005; Lai and Klapa, 2004; Münger et al., 2017; Rancour et al., 1979; Trimigno et al., 2018; Vionnet et al., 2019; Yager et al., 2004).

Several genome-scale studies from past couple of decades emphasize the physiological importance of D-galactonate metabolic pathway in enteric bacteria. Transcriptome analysis of *S. enterica* strains growing in diverse niches such as, in soft-rotted leaves, in murine macrophages or exposed to egg white showed significant upregulation of *dgo* genes (Baron et al., 2017; Eriksson et al., 2003; Goudeau et al., 2013). Similarly, *dgo* genes were found to be significantly induced in an asymptomatic bacteriuria *E. coli* strain cultured in human urine (Roos et al., 2006). Genome-scale metabolic models of 55 fully sequenced *E. coli* and *Shigella* strains predicted 40 of these strains to grow on D-galactonate (Monk et al., 2013). In another study, a mutant library screening of swine paratyphoid causing *S. enterica* serovar Choleraesuis revealed that mutation in *dgoT* leads to attenuation for infection in pigs (Ku et al., 2005). Importantly, in a study where an *E. coli* strain isolated from a patient suffering from urinary tract infection was evolved in the streptomycin-treated mouse gut for a year and sequenced to decipher changes leading to *in vivo* adaptation, *dgo* operon was found to be one of the dominant targets of the adaptation process implying the role of *dgo* genes in the bacterial colonization of the mammalian gut (Lescat et al., 2016).

The above-mentioned examples of D-galactonate prevalence in nature and the proposed significance of *dgo* genes emphasize the importance of investigating D-

galactonate metabolism in enteric bacteria. Although there are a few biochemical studies on Dgo enzymes (Babbitt et al., 1995; Ran et al., 2004; Walters et al., 2008; Wieczorek et al., 1999), and very recently the structure of DgoT has revealed it to belong to the SLC17 family of the anion:cation symporter MFS superfamily (Leano et al., 2019), there are no details on the regulation of D-galactonate metabolism by DgoR. In this chapter, we aimed at investigating the role of DgoR in the regulation of D-galactonate metabolism in *E. coli* K-12. We showed that the putative *dgo* operon is under strong negative regulation by DgoR and that this repression is lifted specifically in the presence of D-galactonate. DgoR binds two inverted repeats overlapping the putative *dgo* operon promoter suggesting that DgoR occludes binding of RNA polymerase (RNAP) to repress the transcription of the putative *dgo* operon. Finally, the DgoR operator sequence matches the signature of GntR family members that recognize inverted repeats.

3.2 Results

3.2.1 The putative *dgo* operon is required for D-galactonate metabolism

Earlier studies showed the involvement of the putative *dgo* operon (Fig. 3.1B) in D-galactonate metabolism by isolating mutants unable to grow on D-galactonate as the sole carbon source (Cooper, 1978). However, the growth phenotype of strains carrying a clean deletion of *dgo* genes had not been reported. Therefore, before performing a detailed investigation on the role of DgoR, we validated the involvement of the predicted *dgo* operon in metabolizing D-galactonate.

When we initiated our studies, D-galactonate was commercially available only as a hemi-calcium salt. However, the hemi-calcium D-galactonate resulted in the formation of white insoluble precipitates in the M9 minimal medium which we desired to use to restrict carbon source available to *E. coli*. Therefore, using a previously described method

we titrated hemi-calcium D-galactonate with oxalic acid to get crystals of D-galactonate (Isbell and Frush, 1933). To determine the quality of the crystalline D-galactonate preparation, D-galactonate sample was dissolved in DMSO-d6 and analysed by ^{13}C-NMR spectroscopy. NMR spectra indicated that the preparation of D-galactonate was pure and free of any lactone contamination (Fig. 3.2) (Bock K. and C., 1983; Liu et al., 2014a; Wehrli et al., 1997). The D-galactonate crystals prepared in the lab were readily soluble in the M9 minimal medium. We tested the utilization of crystalline D-galactonate by *E. coli* K-12 BW25113 by assessing the growth of bacteria on solid M9 minimal medium supplemented with D-galactonate. As a control, the growth of *E. coli* was also monitored on solid M9 minimal medium supplemented with D-glucose. *E. coli* was able to grow on M9 minimal medium supplemented with D-galactonate as the carbon source (Fig. 3.3A). Further, growth analysis in liquid M9 minimal medium supplemented with D-galactonate showed that *E. coli* grows in D-galactonate with a generation time of ~1.8 hours (Fig. 3.3B). Thus D-galactonate crystals prepared in the lab were used as a carbon source for further experiments. A separate batch of D-galactonate crystals prepared in the lab was also independently tested for purity by NMR, and used in some of the experiments along side to ensure reproducibility (Nongthombam, 2015).

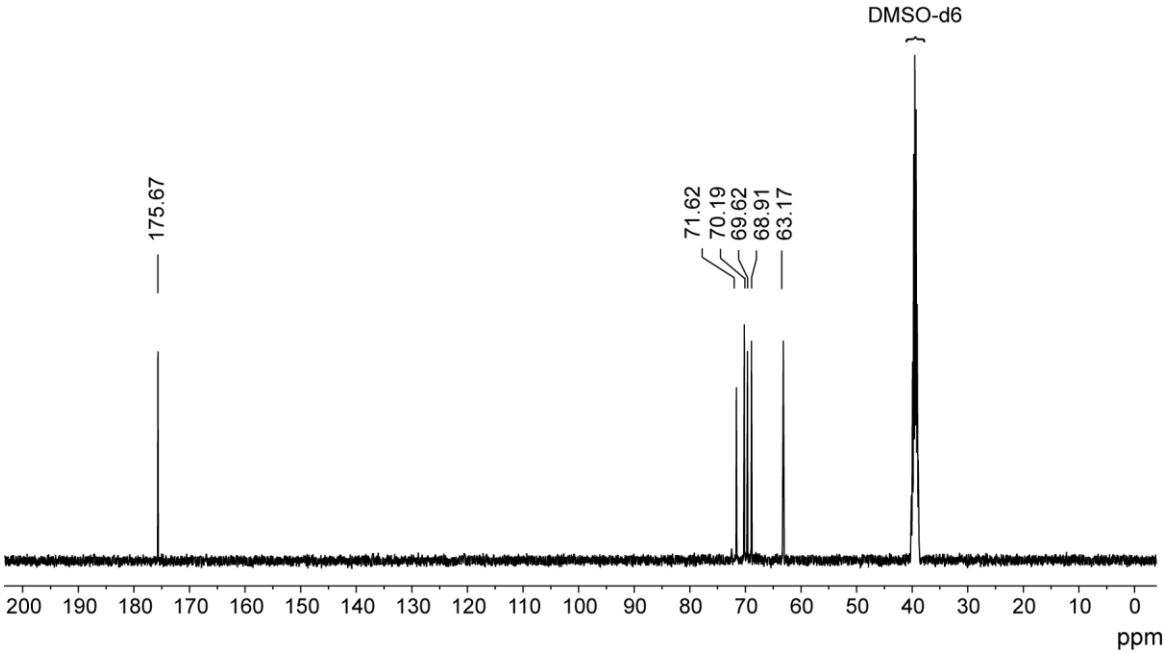

Figure 3.2 Crystalline D-galactonate prepared in the lab is pure. Crystalline D-galactonate prepared in the lab was analysed by ^{13}C-NMR spectroscopy. Crystalline D-galactonate preparation is in acidic form and free of any lactone intermediate contamination.

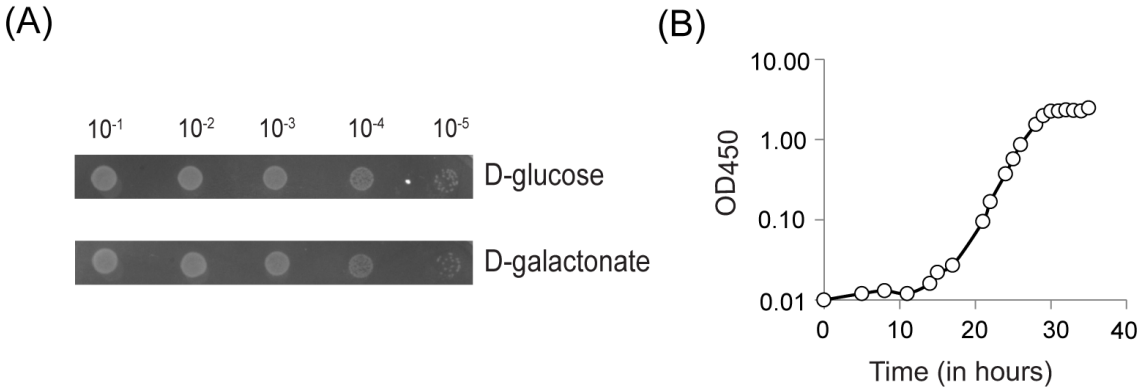

Figure 3.3 Crystalline D-galactonate prepared in the lab is used as a carbon source by *E. coli*. (A) Crystalline D-galactonate supports the growth of WT strain on solid minimal medium. Dilutions of cultures were spotted on M9 minimal medium containing either D-glucose or D-galactonate as the carbon source. Experiment was repeated 3 times. A representative data set is shown. (B) WT strain can use crystalline D-galactonate as a carbon source in liquid medium. WT strain was grown in shake flask in minimal medium containing D-galactonate as the carbon source. The experiment was performed 3 times. A representative data is shown.

Strains with individual *dgo* genes replaced by kanamycin cassette (*dgoR*::*kan*, *dgoK*::*kan*, *dgoA*::*kan*, *dgoD*::*kan* and *dgoT*::*kan*) were obtained from the Keio collection

(Baba et al., 2006) and verified by PCR. In addition, in the present study, the complete putative *dgo* operon was replaced with kanamycin cassette using λ Red recombinase system to obtain *dgo*::*kan* strain. Growth of various *dgo* strains was assessed on M9 minimal agar medium containing either D-glucose or D-galactonate as the carbon source. In contrast to normal growth of strains on D-glucose, none of the *dgo* strains except *dgoR*::*kan* strain showed growth on D-galactonate (Fig. 3.4). Hence we confirmed that the *dgo* gene products are required for metabolizing D-galactonate.

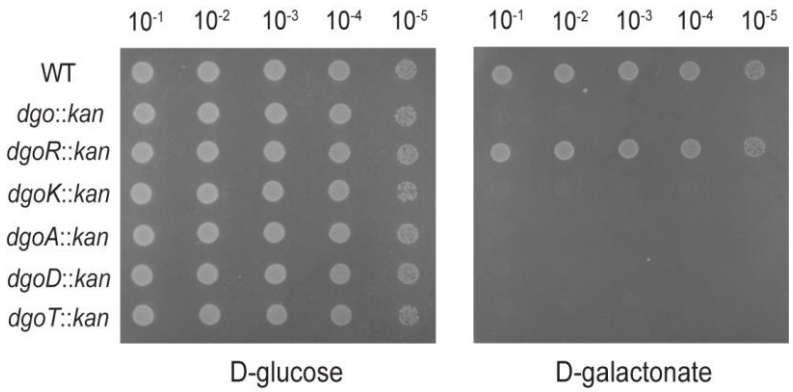

Figure 3.4 Putative *dgo* operon is involved in the metabolism of D-galactonate. Dilutions of the cultures were spotted on M9 minimal medium containing either D-glucose or D-galactonate as the carbon source. The experiment was repeated 3 times. A representative data set is shown.

3.2.2 DgoR negatively regulates D-galactonate metabolism

To investigate the regulatory role of *dgoR* in D-galactonate metabolism, we assessed the growth of *dgoR* deletion strain in medium containing D-galactonate as the carbon source. Since as mentioned in the previous section, *dgoR* deletion strain did not display any phenotype on the solid medium (Fig. 3.4), we assessed the growth of *dgoR*::*kan* strain in liquid medium supplemented with D-galactonate. The *dgoR*::*kan* strain showed significantly faster growth in D-galactonate as compared to the WT. Growth comparison of WT and *dgoR*::*kan* in additional carbon sources i.e., D-glucose, D-gluconate and D-galactose showed that the faster growth phenotype of *dgoR* deletion strain is specific to D-galactonate (Fig. 3.5). The *dgoR* alongwith ~400 bp upstream region carrying the

putative promoter was cloned in pACYC177 for complementation experiments. DgoR was His-tagged at the C-terminus. Expression of DgoR from this construct (pBS13) complemented the accelerated growth phenotype of *dgoR*::*kan* in D-galactonate (Fig. 3.6). Collectively, these data indicate that DgoR is a negative regulator of D-galactonate metabolism.

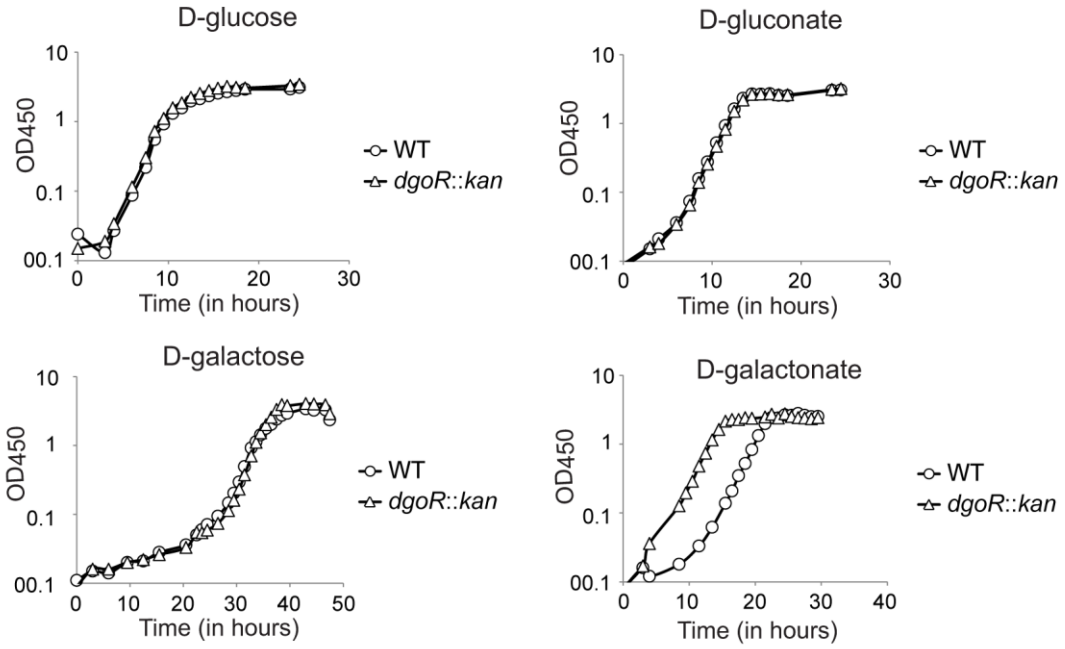

Figure 3.5 Deletion of *dgoR* leads to faster growth of *E. coli* in D-galactonate. WT and *dgoR*::*kan* strains were grown in shake flasks in minimal medium containing one of the indicated carbon sources, and OD_{450} was measured. The experiment was performed 3 times. A representative data set is shown.

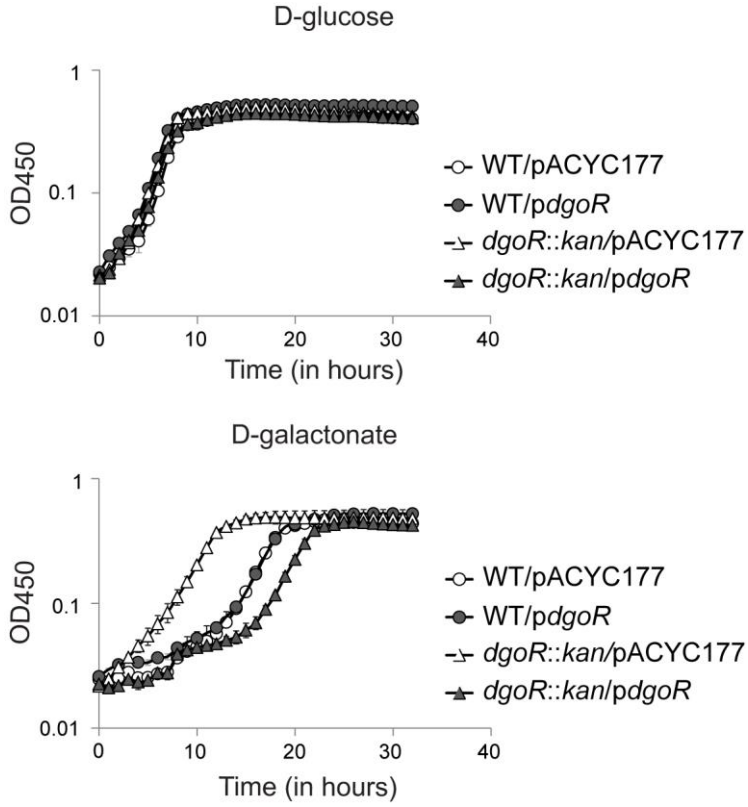

Figure 3.6 *dgoR* cloned in plasmid complements the faster growth phenotype of *dgoR*::*kan* in D-galactonate. WT and *dgoR*::*kan* strains carrying either empty plasmid (pACYC177) or pACYC177 with *dgoR* (p*dgoR*) were grown in 96-well plates in minimal medium containing either D-glucose or D-galactonate as the carbon source, and OD_{450} was measured. The experiment was done 3 times; each experiment had 3 technical replicates. A representative data set, with average and SD of technical replicates, is shown.

Transcriptional regulators usually bind a regulatory DNA sequence to modulate the transcription of their regulon members. Depending upon whether the binding of transcriptional regulator to the DNA enhances or reduces the expression of the regulon member, they are classified as positive or negative regulators, respectively, although some transcriptional regulators are known to exhibit dual behaviour (Madan Babu and Teichmann, 2003; Perez-Rueda and Collado-Vides, 2000). In order to investigate the regulatory nature of DgoR over the putative *dgo* operon, we sought to compare the transcript levels of various *dgo* genes in WT and *dgoR* deletion strains. Certain carbon

sources are known to suppress the metabolic genes of other carbon sources by carbon catabolite repression (Görke and Stülke, 2008), thus in order to avoid complications due to carbon catabolite repression, we cultured WT and *dgoR*::*kan* strains in M9 minimal medium supplemented with glycerol, owing to its mild carbon catabolite repression (Chan et al., 2002; Hogema et al., 1997). Transcript levels of various *dgo* genes were determined by qRT-PCR. In comparison to WT, all the *dgo* gene transcripts were constitutively expressed in the *dgoR*::*kan* strain (Fig. 3.7). The expression of *dgo* genes ranged from ~100- to 1200- fold. The considerable expression of *dgo* genes in the absence of DgoR established that the putative *dgo* operon is under strong negative regulation by DgoR.

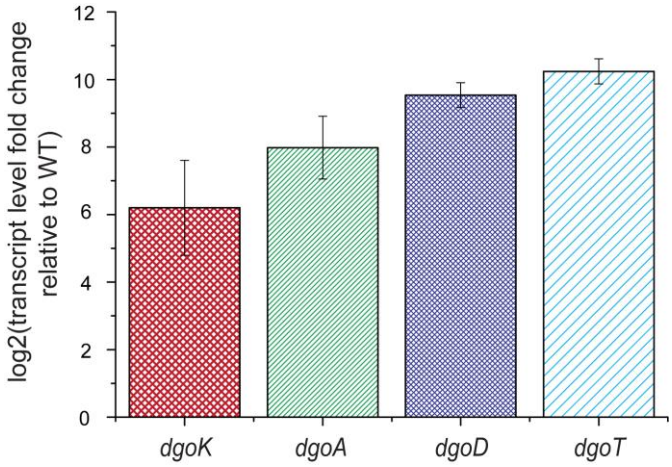

Figure 3.7 DgoR negatively regulates transcription of the putative *dgo* operon. WT and *dgoR*::*kan* strains were grown in minimal medium containing glycerol as carbon source until exponential phase. RNA was isolated, and transcript abundance of *dgo* genes was assayed by qRT-PCR. Data were normalized to the transcript levels of *dgo* genes in WT and represent the average (± SD) of 3 independent experiments.

3.2.3 The *dgo* genes are transcriptionally induced by D-galactonate

In an earlier study, the enzymatic activities corresponding to D-galactonate dehydratase (DgoD), 2-dehydro-3-deoxygalactonate kinase (DgoK) and 2-dehydro-3-deoxygalactonate 6-phosphate aldolase (DgoA) were found at high levels in the cell-free extracts prepared from WT cells grown in D-galactonate but not in the extracts prepared

from cells grown in glycerol, D-gluconate or D-galactose (Deacon and Cooper, 1977). To futher investigate the regulation of putative *dgo* operon at a transcriptional level, we employed qRT-PCR to determine the levels of various *dgo* transcripts in WT cells grown in M9 minimal medium supplemented with different carbon sources viz. D-glucose, D-gluconate, D-galactose, D-galactonate, glycerol or glycerol supplemented with D-galactonate. We found that of all the tested carbon sources, the *dgo* genes are induced at a transcriptional level specifically in the presence of D-galactonate (Fig. 3.8A). Also, in contrast to glycerol or D-galactose grown cells, the expression of various *dgo* genes was much lower in cells grown in D-glucose or D-gluconate which can be attributed to much stronger carbon catabolite repression of the latter carbon sources (Chan et al., 2002; Hogema et al., 1997). To validate our above findings at a protein level, we used a strain where the chromosomal *dgoR* was tagged with a 3XFLAG tag at the C-terminus end (Table 2.1, strain RC3067). Consistent with the qRT-PCR data, C-terminally 3XFLAG-tagged DgoR was observed at a higher level specifically in the presence of D-galactonate (Fig. 3.8B). Taken together, our data show that the *dgo* genes are specifically induced in the presence of D-galactonate.

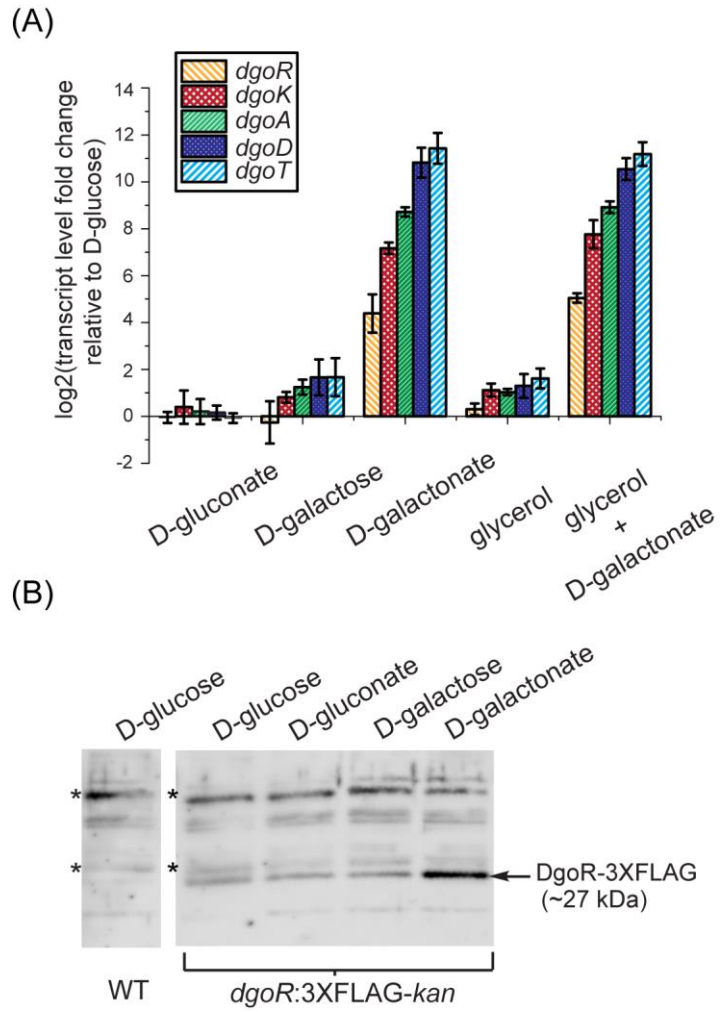

Figure 3.8 Putative *dgo* operon is significantly induced by D-galactonate. *(*A) Putative *dgo* operon is induced at a transcriptional level by D-galactonate. WT strain was grown in minimal medium supplemented with carbon sources indicated in the figure. RNA was isolated in the exponential phase and processed for qRT-PCR. Data were normalized to the transcript levels of *dgo* genes in WT grown in minimal medium containing D-glucose. Data represent the average (± SD) of three independent experiments. (B) DgoR protein levels are higher in D-galactonate-grown cells. The strain expressing C-terminally 3XFLAG-tagged DgoR from the chromosome was grown in minimal medium supplemented with one of the indicated carbon sources. Cells were harvested in the exponential phase and processed for Western blotting using anti-FLAG antibody. WT cells grown in D-glucose were taken as negative control. The band corresponding to DgoR-3XFLAG is indicated (Mol. wt. ~27 kDa). '*' indicates non-specific band detected by anti-FLAG antibody and serves as a control for equal loading of samples.

3.2.4 DgoR binds to the DNA sequences upstream of the putative *dgo* operon

Recognizing and binding to a suitable regulatory DNA sequence as well as the modulation of this DNA binding in response to appropriate signal constitutes the fundamental activities central to the functioning of a typical transcriptional regulator. A majority of GntR family transcriptional regulators bind inverted repeat sequences on DNA (Hoskisson and Rigali, 2009; Jain, 2015; Rigali et al., 2002; Suvorova et al., 2015). A search in the RegPrecise database, a manually curated database comprising of transcriptional regulons reconstructed on the basis of comparative genomics, found three inverted repeats upstream of the *dgoR* start codon in *E. coli* (hereafter designated as IR1, IR2 and IR3) (Fig. 3.9A) suggesting that DgoR binds the DNA sequences upstream of putative *dgo* operon to regulate itself and other genes in the putative *dgo* operon (Novichkov et al., 2013). Each inverted repeat comprises of two conserved half sites separated by a central non-conserved C/G base. Whereas IR1 is a perfect inverted repeat, IR2 and IR3 have one mismatch each in either right or left half of the inverted repeat, respectively. This arrangement of inverted repeats seems to be more or less conserved in other members of Enterobacteriaceae (Novichkov et al., 2013).

To test whether DgoR binds to the DNA sequences upstream of putative *dgo* operon, we performed radioactive EMSAs. For this, we overexpressed and purified C-terminally 6×His tagged DgoR (Fig. 3.9B). The DgoR-6×His could be purified to a final concentration of ~15 μM; attempts to purify the protein at a higher concentration led to aggregation of the protein. The pAS1 construct where ~200 bp DNA sequence upstream of the putative *dgo* operon harbouring the three inverted repeats (hereafter, referred to as the *dgo cis*-acting element) (Fig. 3.9A) is cloned upstream of a yellow fluorescent reporter in the plasmid pVenus (Table 2.1), was used for amplification of *dgo cis*-acting element using ^{32}P-labelled primer. In the EMSA, the labelled *cis*-acting element was

incubated with different concentrations of purified DgoR-6XHis. DgoR shifted the labelled *dgo* cis-acting element in a concentration dependent manner to form a single DNA-protein complex (Fig. 3.9C). Further, addition of unlabelled *dgo cis*-acting element in 200-fold excess of labelled DNA completely abolished the interaction of DgoR with labelled DNA (Fig. 3.9C). These results show that DgoR binds to sequence upstream of the putative *dgo* operon.

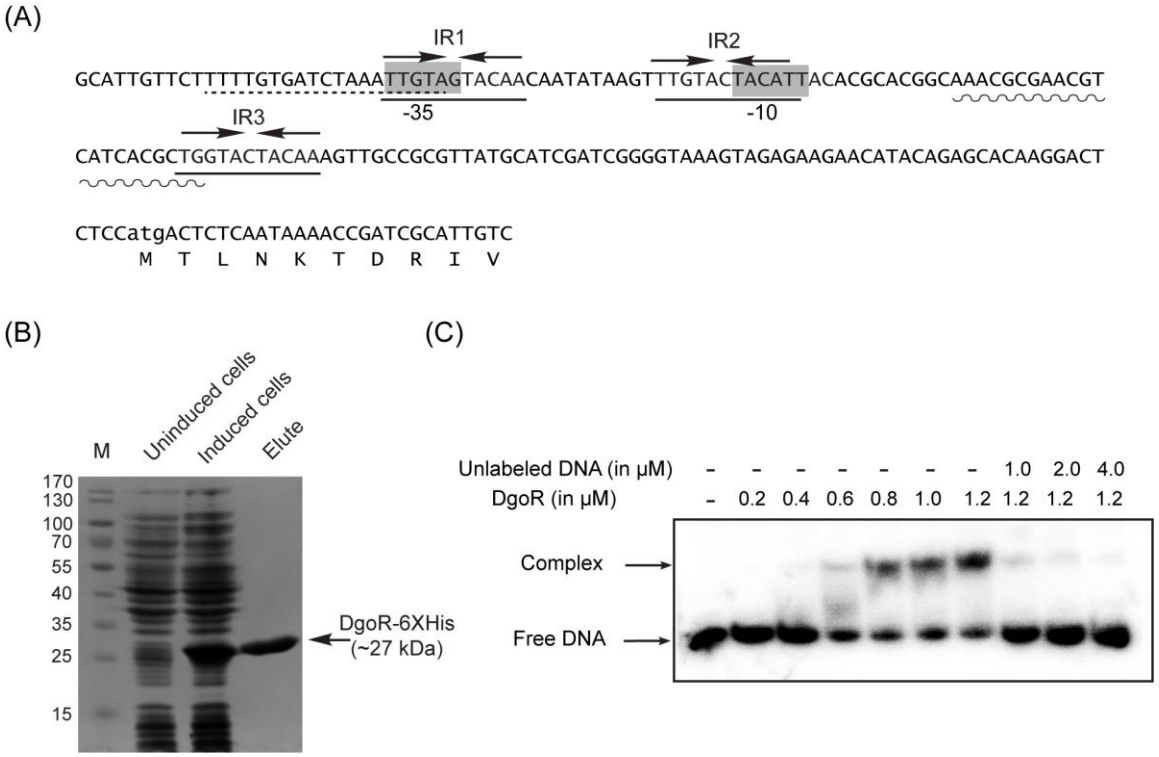

Figure 3.9 DgoR binds its *cis*-acting element. (A) The sequence of the ~200 bp *cis*-acting element used in EMSAs. Lowercase letters are used to denote the start codon of *dgoR*. The sequence of promoter elements (-35 and -10 boxes) in the upstream region of *dgoR*, predicted by both BPROM and bTSSfinder, is highlighted in gray boxes. The cAMP-CRP binding sites, site 1 and site 2 predicted by vitrtual footprinting are underlined by dashed line and wavy line, respectively. The three inverted repeats (IR1, IR2 and IR3) are underlined using solid lines. Arrows depict the left and right halves of the repeats. (B) SDS-PAGE of purified DgoR-6XHis. WT strain carrying pBS2 (C-terminally His-tagged *dgoR* cloned in pRC10, Table 2.1) was induced with IPTG. Cells were harvested and processed for protein purification using Co-NTA resin. Lane M represents the molecular mass markers (in kDa). The band corresponding to DgoR-6XHis is indicated (Mol. wt. ~27 kDa). The tagged protein could be purified to a final concentration of ~15 μM. (C) EMSA shows the binding of purified DgoR to the *cis*-acting

element. 20 nM ^{32}P-labeled DNA fragment was incubated with different concentrations of purified DgoR-6XHis with or without the unlabeled *cis*-acting element. Samples were resolved on native PAGE and subjected to autoradiography.

3.2.5 DgoR binds to two closely spaced inverted repeats overlapping the putative *dgo* promoter in the *dgo cis*-acting element

We next investigated the importance of three inverted repeats in the *dgo cis*-acting element. In this direction, we used plasmid constructs carrying WT or mutated *dgo cis*-acting elements cloned upstream of a yellow fluorescent reporter in the pVenus plasmid (pAS1-pAS8; Table 2.1) for amplification of ^{32}P- labeled fragments for EMSA. The mutant *cis*- elements had the inverted repeats disrupted individually by creating substitution or deletion mutations in either left or right half of the inverted repeat (Fig. 3.10A). The ^{32}P-labelled WT or mutant *dgo cis*-elements were incubated with purified DgoR-6XHis in an EMSA. Whereas similar to previous result, DgoR shifted the WT *cis*-acting element, the protein was unable to shift the DNA mutated for either IR1 or IR2 (DNA fragments: IR1L, IR1R, IR2L, IR2R and IR2L') (Fig. 3.10B). In contrast, DNA fragments bearing mutation in IR3 (DNA fragments: IR3L and IR3R) formed protein-DNA complexes similar to the WT *cis*-acting element (Fig. 3.10B). Taken together, these *in vitro* findings suggest that out of the three inverted repeats, only IR1 and IR2 are critical for DgoR to bind the *dgo cis*-acting element.

Further, we investigated the significance of inverted repeats *in vivo*. For this, we used the WT and Δ*dgoR* strains carrying either the promoter-less pVenus plasmid or the Venus reporter constructs, pAS1 to pAS8 integrated at the chromosomal *attλ* site in a single copy (Table 2.1). We first ensured that our reporter construct mimics the regulation at the *dgo* locus by comparing the fluorescence of the reporter expressed from the WT *dgo cis*- acting element in WT and Δ*dgoR* strains in both noninducing (M9 minimal

containing glycerol) and inducing (M9 minimal containing glycerol and D-galactonate) medium. Whereas in the WT strain fluorescence from the reporter was found to be increased in the presence of D-galactonate as compared to the noninducing condition (Fig. 3.10C), the same reporter construct displayed constitutive fluorescence in the noninducing medium in *dgoR* deletion background validating that the reporter construct mimics the regulation at the *dgo* locus (Fig. 3.10C).

One of the mechanisms by which transcriptional repressors downregulate transcription is that they bind to a site overlapping the promoter thereby interfering with the binding of RNAP (Rojo, 1999). Hence, in order to interpret the results from *in vivo* fluorescence reporter assay for IR mutants, it was important to consider the location of the *dgo* promoter. We used BPROM and bTSSfinder programs to predict the -10 and -35 promoter elements in the *dgo cis*-acting element (Shahmuradov et al., 2017; Solovyev and Salamov, 2011). Interestingly, the common -35 and -10 promoter elements predicted by both the programs were found to overlap with IR1 and IR2 sites, respectively (Fig. 3.10A). A separate set of experiments conducted in our lab (Neeladrita Kundu, Ph.D. student) showed that the predicted -35 and -10 promoter elements indeed constitute the *dgo* promoter. Both WT and Δ*dgoR* reporter strains carrying *dgo cis*- acting element truncated either for -35 element or both the -10 and -35 elements showed near-background fluorescence in noninducing (M9 minimal containing glycerol) as well as inducing (M9 minimal containing glycerol and D-galactonate) medium (Fig. 3B, (Singh et al., 2019)).

The left half of IR1 overlaps with the predicted -35 element, whereas the right half of IR1 overlaps a few base pairs of the spacer region between the predicted -35 and -10 elements (Fig. 3.10A). In the Δ*dgoR* background, both the IR1L and IR1R constructs exhibited fluorescence similar to that of the reporter construct harboring the WT *cis*-

acting element (Fig. 3.10C). These data show that mutations in IR1 do not abrogate promoter activity. Importantly, both constructs showed constitutive fluorescence in the WT strain, suggesting a complete loss of DgoR repression (Fig. 3.10C). These results correlate with the defective binding of purified DgoR with IR1L and IR1R constructs in EMSA (Fig. 3.10B). Together these data reveal that IR1 is important for DgoR to bind the *cis*-acting element to repress transcription of the putative *dgo* operon.

The right half of IR2 overlaps with the predicted -10 box of the *dgo* promoter (Fig. 3.10A). In the *dgoR* deletion background, the IR2R construct exhibited lower fluorescence than the construct harbouring the WT *cis*-acting element, suggesting that mutations in the predicted -10 box compromise promoter activity (Fig. 3.10C). Interestingly, mutations in the left half of IR2 (IR2L and IR2L') which lie outside the predicted -10 box also exhibited low fluorescence in the *dgoR* deletion background, suggesting that the region adjacent to -10 also contributes to the promoter activity (Fig. 3.10C). Further, in WT cells grown under noninducing conditions, all IR2 mutants exhibited slightly higher fluorescence than the WT *cis*-acting element, indicating that mutations in IR2 compromise DgoR binding (Fig. 3.10C). These observations are again in line with that of our *in vitro* EMSA findings, where *cis*-acting element carrying mutations in IR2 did not bind DgoR (Fig. 3.10B). These results show that IR2 is essential for binding of DgoR to the *dgo cis*-acting element.

Consistent with the observation that IR3 does not overlap with the promoter (Fig. 3.10A), in the *dgoR* deletion background, both IR3L and IR3R reporter constructs exhibited fluorescence equivalent to that of the construct carrying the WT *cis*-acting element. Similarly, the WT strain carrying IR3L and IR3R reporter constructs showed regulatory behaviour analogous to that of the strain carrying WT fragment (Fig. 3.10C). These data correlate with the ability of purified DgoR to bind IR3L and IR3R fragments

in EMSA (Fig. 3.10B). Thus, the IR3 site is not necessary for the interaction of DgoR with its *cis*-acting element.

Collectively, from our *in vitro* EMSA and *in vivo* fluorescent reporter studies, we find that DgoR binds IR1 and IR2 sites in the *dgo cis*-acting element for repression of the putative *dgo* operon. The fact that the DgoR binding site overlaps with the putative *dgo* promoter suggests that DgoR occludes the binding of RNAP for repressing the transcription of the putative *dgo* operon.

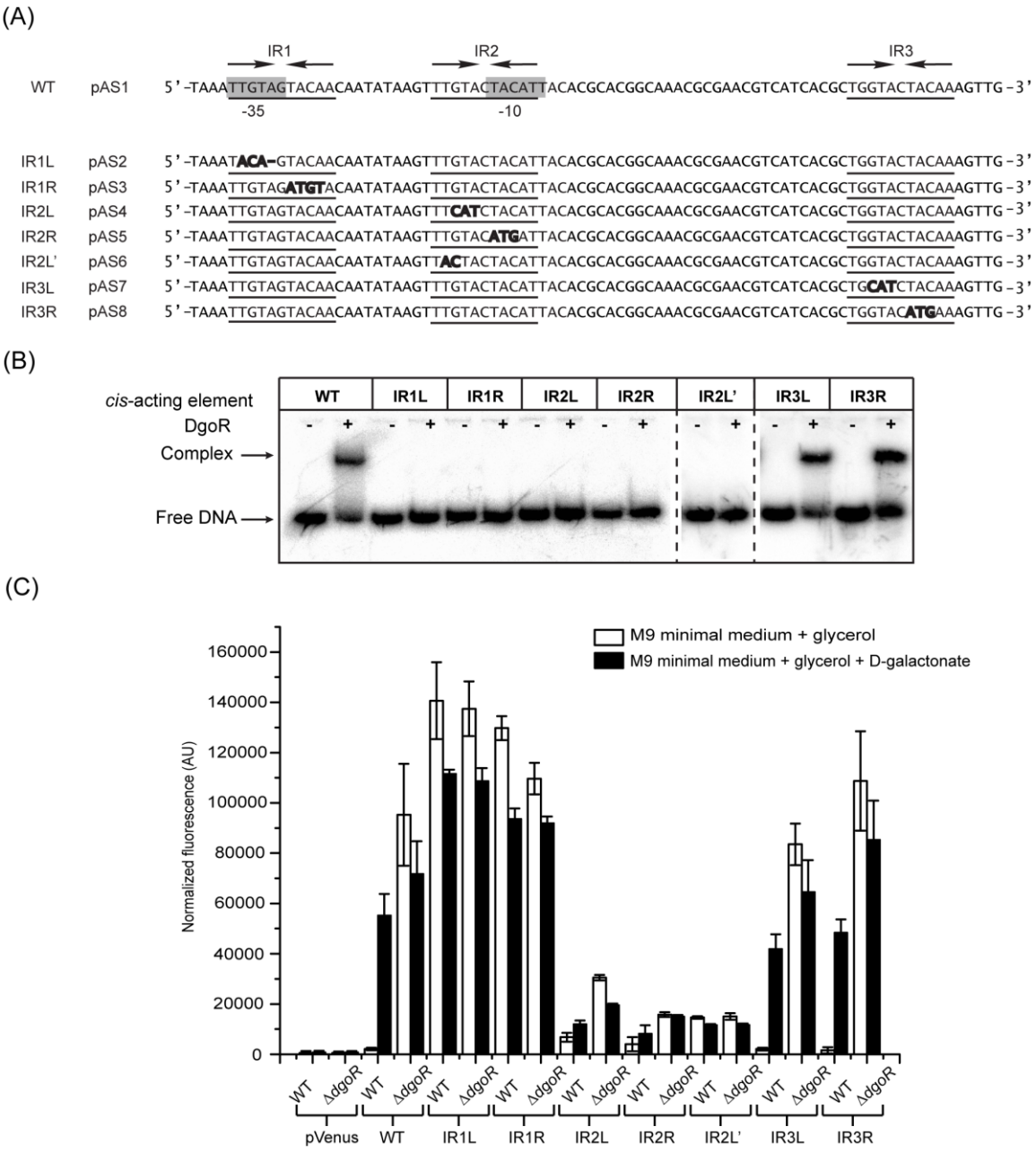

Figure 3.10 Inverted repeats, IR1 and IR2, overlap with the promoter and are critical for binding of DgoR to its *cis*-acting element. (A) The sequence of a region of the *dgo cis*-acting element showing mutations created within the three inverted repeats. The sequence of a region of the WT *cis*-acting element and its various mutants (IR1L, IR1R, IR2L, IR2R, IR2L', IR3L and IR3R) are shown where the predicted promoter core elements (-35 and -10 boxes) are highlighted in gray boxes, the three inverted repeats (IR1, IR2 and IR3) are underlined, left and right halves of the repeats are shown by arrows, nucleotide substitutions in the inverted repeats are shown in bold and nucleotide deletions in the inverted repeats are shown by a line (-). L and R in the names of the inverted repeat mutants denote whether the mutations are in the left or right half of the inverted repeat. For example, IR1L and IR1R denote that the mutations are in the left and right half of IR1, respectively. (B) EMSA shows that IR1 and IR2 are critical for binding of DgoR to

its *cis*-acting element *in vitro*. 20 nM ^{32}P-labeled WT and various mutant fragments were incubated with (+) or without (-) 1.2 μM purified DgoR-6XHis. Samples were resolved on native PAGE and subjected to autoradiography. (C) Fluorescence assay reveals the requirement of IR1 and IR2 in the *cis*-acting element for binding of DgoR *in vivo*. Reporter strains were grown in M9 minimal medium supplemented either with glycerol or with glycerol and D-galactonate to exponential phase. Fluorescence was measured and normalized to OD_{450} of the samples. Data represent the average (± SD) of 3 independent experiments. (One biological replicate of the fluorescence reporter assay for a subset of the IR mutants was also perfomed in (Sangwan, 2017)).

3.3 Discussion

Despite considerable evidence on the importance of D-galactonate metabolic pathway in the physiology of enteric bacteria, in-depth details of its regulation even in *E. coli*, which is known to utilize D-galactonate since the 1970s, were still inadequate. Since the classical genetic and biochemical studies conducted in 1970s in *E. coli*, although a few studies biochemically characterized DgoD and DgoA enzymes involved in D-galactonate metabolism (Babbitt et al., 1995; Ran et al., 2004; Walters et al., 2008; Wieczorek et al., 1999), and very recently, DgoT was characterized for its role in transporting D-galactonate across the bacterial membrane (Leano et al., 2019), investigations on the role of DgoR in D-galactonate metabolism was still missing. In the work presented in this chapter, we investigated the molecular and functional insights into the regulation of D-galactonate metabolism by DgoR in *E. coli*.

Both the faster growth in D-galactonate and a considerable increase in *dgo* transcript levels in cells lacking *dgoR* (Figs. 3.5 and 3.7), suggest the strong repression of D-galactonate metabolism by DgoR. These results are in line with constitutive expression of a putative *dgo* promoter-reporter construct in a Δ*dgoR* strain observed in a study reported at the time our current work was in progress (Lescat et al., 2016). Further, the *dgo* transcript levels increase in the presence of D-galactonate (Fig. 3.8A), consistent with

earlier reports that the activity of Dgo enzymes (Deacon and Cooper, 1977) and the expression of a *dgo* promoter-reporter construct (Lescat et al., 2016) are induced under similar conditions. In addition, similar to other sugar acid metabolic pathways (Mandrand-Berthelot et al., 2004), D-galactonate metabolism is also positively regulated by cAMP-CRP driven carbon catabolite repression system. An earlier study showed that D-glucose represses the expression of D-galactonate catabolic enzymes (Cooper, 1978). We also find that *dgo* transcript levels are higher (~1.5- fold to 3- fold) in *E. coli* grown in a non-catabolite-repressing (glycerol) or a mild catabolite-repressing (D-galactose) carbon source in comparison to their levels in catabolite-repressing substrates (D-glucose and D-gluconate) (Chan et al., 2002; Hogema et al., 1997) (Fig. 3.8). In fact, an analysis of the *dgo cis*-acting element by virtual footprint predicts the presence of two cAMP-CRP binding sites (Fig. 3.9A) (Münch et al., 2005). While our work was in progress, a high-throughput plasmid-based reporter study confirmed the presence of predicted cAMP-CRP binding site 1 (Fig. 3.9) (Belliveau et al., 2018). Collectively, these results suggest that the putative *dgo* operon is under dual regulation: negative regulation by DgoR and positive regulation by cAMP-CRP system. Additional experiments such as *in vitro* binding of cAMP-CRP complex to the *dgo cis*-acting element and *in vivo* reporter assay with D-glucose grown cells in the presence and absence of exogenous cAMP supplementation are required for detailed understanding of regulation by cAMP-CRP.

Our work shows that DgoR forms a single complex at the *dgo* operator comprising of two inverted repeats that are likely on the same face of the DNA (the distance between the inverted repeats is 10 bp, i.e., 1 helix turn) (Figs. 3.9, A and C). Importantly, the inverted repeat sequences recognized by DgoR [5'-TTGTA(G/C)TACA(A/T)-3'] (Fig. 3.10), match the signature of GntR family regulators that also bind inverted repeats [5'-(N)*y*GT(N)*x*AC(N)*y*-3'], where *x* and *y* indicate the

number of nucleotides, which varies (Hoskisson and Rigali, 2009; Jain, 2015; Rigali et al., 2002; Suvorova et al., 2015). Detailed studies involving determination of stoichiometry of DgoR-DNA complex would be required to delineate the mode of binding of DgoR to the *dgo cis*-acting element. In addition to identifying the cAMP-CRP binding site, the above mentioned high-throughput plasmid-based reporter study reported a region in the *dgo cis*-acting element to harbor putative overlapping DgoR- and RNAP-binding sites (Belliveau et al., 2018). Importantly, the region identified in this study encompasses the sites determined by our single-copy reporter assays and EMSAs, providing strong support to our findings. Using fluorescent reporter and RT-PCR based experiments performed in our lab, we also validated the putative *dgo* promoter in the *dgo cis*-acting element (Neeladrita Kundu, Ph.D. student, unpublished data and Fig. 3, (Singh et al., 2019)). The DgoR-binding sites overlap the *dgo* promoter indicating that DgoR represses transcription of the putative operon by occluding the binding of RNAP (Fig. 3.10A). Further experiments such as competition assays of RNAP holoenzyme and DgoR for binding *dgo cis*-acting element are required to validate this model (Hernandez-Arriaga et al., 2009).

Information from the work presented in this thesis, as well as from work published in another report (Belliveau et al., 2018) suggests that the D-galactonate responsive *dgo* promoter is under complex regulation. The -35 as well as -10 sequences of the *dgo* promoter overlap with the two inverted repeats IR1 and IR2 recognized by DgoR, respectively. Besides, the -35 region also overlaps with cAMP-CRP binding site (Figure 3.9A) thus classifying the *dgo* promoter as a class-II promoter. Interestingly, we note that the nucleotide substitutions in the IR1L site although change the DgoR binding site drastically by breaking the symmetry of the inverted half required for the binding of DgoR, simultaneously lead to changes in the -35 element as well as the overlapping

cAMP-CRP binding site. We find that although mutations in IR1L discrupt the binding of DgoR to the *dgo cis*-acting element (Fig. 3.10B), it does not negatively affect the *dgo* promoter activity due to changes in the -35 element (Fig. 3.10C). Further, we find that mutations in IR1L does not impact much on the integrity of the cAMP-CRP binding site (Virtual footprint server originally used to predict the site identifies the same site as cAMP-CRP binding site in both WT and IR1L DNA fragments). It is well reported that cAMP-CRP complex enhances the promoter activity at class II promoters by actively interacting with the transcriptional machinery to facilitate initial binding of RNAP to the promoter as well as transition into open complex (Savery et al., 1998). The fact that reporter construct lacking -35 element as well as the cAMP-CRP binding site (Δ-35 construct) completely loses the promoter activity suggests a strong positive role of the cAMP-CRP on expression of *dgo* promoter (Fig. 3B, (Singh et al., 2019)). It seems that mutations in -35 element are compensated by the stronger positive role of cAMP-CRP complex (Figure 3.10C). Further *in vitro* and *in vivo* experiments involving carefully planned mutations in the *dgo cis*-acting element will be required to understand the complex regulation at the *dgo* promoter.

CHAPTER IV

DgoR shares conserved and semiconserved protein-DNA contacts with the GntR family of transcriptional regulators

4.1 Introduction

The transcriptional regulators are categorized into families based on the presence of conserved motifs and their mode of DNA binding (Ishihama, 2009). GntR family transcriptional regulators are characterized by similarity in their N-terminal winged helix-turn-helix (wHTH) domain (Pfam accession number PF00392) (Haydon and Guest, 1991; Rigali et al., 2002; Suvorova et al., 2015). Further, based on the topology of C-terminal effector-binding and oligomerization (E/O) domain, the GntR family members have been subdivided into seven subfamilies: FadR, HutC, MocR, YtrA, AraR, PlmA and DevA (reviewed in (Hoskisson and Rigali, 2009; Jain, 2015)). FadR subfamily is characterized by an all helical C-terminal E/O domain [recognized as FCD (Pfam accession number PF07729) or FadR_C- domain (Pfam accession number PF07840)] consisting of either six (VanR subgroup) or seven (FadR subgroup) antiparallel helices organized in a bundle. HutC subfamily bears a α/β C-terminal [UTRA domain (Pfam accession number PF07702)] comprising of three α-helices surrounding a core of beta sheet made of six anti-parallel beta strands. MocR members have a comparatively longer α/β C-terminal domain with homology to aminotransferase class I domain (Pfam accession number PF00155). YtrA subfamily has a shorter fish-hook shaped domain containing only two helices. AraR presents a peculiar case of fusion between a GntR-type wHTH (Pfam accession number PF0039) domain with C-terminal having homology to that of LacI family of transcriptional regulators (InterPro IPR028082 Peripla_BP_I). C-terminal E/O domains of DevA and PlmA harbor topologies distinct from other subfamily members which are not well characterized.

The three regulatory components involved in regulation by various GntR family members; DNA-binding domain, E/O domain and the cognate operator impose spatial and structural constraints on each other dictating conserved or semiconserved features of

the family. Variations in the linker connecting the DNA-binding wHTH to the C-terminal E/O domain as well as the manner of oligomerization of monomers in the regulator contribute to variations in the operator sequence recognized by different transcriptional regulators even within the same subfamily (Hoskisson and Rigali, 2009; Jain, 2015; Rigali et al., 2002; Suvorova et al., 2015). Most of the GntR family members bind A/T rich palindromic or pseudo-palindromic DNA. A comparison of predicted or experimentally determined binding sites recognized by the GntR family transcriptional regulators shows that nearly all bind sites resembling the consensus sequence [5'-$(N)_y$GT$(N)_x$AC$(N)_y$-3'] where the number (x,y) and the nucleotide (N) surrounding the constant GT/AC pair varies. The centre of the palindrome appears to be more conserved than the A/T rich peripheral regions. The number of A and T residues vary among the different GntR subfamilies, for FadR subfamily the operator consensus deduces to [5'-$(N)_y$GTM-$(N)_{0-1}$-KAC$(N)_y$-3'], where 'M' represents A or C nucleotide and 'K' represents G or T nucleotide (reviewed in (Suvorova et al., 2015)). The typical distance between conserved GT/AC pair in the DNA binding sites of most of the FadR-subfamily members is 3 nucleotides (for example FadR of *E. coli* and LldR of *C. glutamicum*), or 2 nucleotides (for example GntR of *B. subtilis*, and UxuR of *E. coli*) (Suvorova et al., 2015). Many of the FadR subfamily members do not bind the operators matching the above mentioned consensus such as AphS of *Comamonas testosteroni* (Q9RHW8), ATAAAACTATCGATAAAA; BioR of *Agrobacterium tumefaciens* (A9CJW2), TTATMKATAA; CitO of *Enterococcus faecalis* (Q2KKB8), TGTTTTTTTGATGTTCTTTTTGTTT; while NanR of *E. coli* (P0A8W0), TGGTATAW is known to recognize direct repeats (Arai et al., 1999; Blancato et al., 2008; Condemine et al., 2005; Kalivoda et al., 2013; Rodionov and Gelfand, 2006). A deluge of experimental studies on GntR family regulators show that these transcriptional regulators

generally form homodimers by making a number of contacts in the C-terminal E/O domain, although TreR, a HutC subfamily member was shown to adopt a tetrameric organization (Jain, 2015; Rezacova et al., 2007). The spatial arrangement of monomers in the dimers differs even amongst the same subfamily members and affects how they interact with their cognate operator DNA (Gao et al., 2008; Jain, 2015). However, most of the GntR family members including the FadR subfamily transcriptional regulators bind as dimers where wHTH domain of each monomer contacts one half of the symmetrical inverted DNA repeat.

The majority of details on DNA-protein interaction in the GntR family have been derived from a limited number of solved DNA-bound structures (Fillenberg et al., 2015; Jain, 2015; Jain and Nair, 2013; Shi et al., 2015; Suvorova et al., 2015; van Aalten et al., 2001; Xu et al., 2001). The structural data shows that despite the weak amino acid sequence similarity in the wHTH domain of the GntR family (~25%), the base specific interactions are conserved. The DNA-protein complex is dominated by non-specific interactions between the amino acid residues of the regulator and the sugar-phosphate backbone of DNA, and only limited base-specific contacts are made. The canonical wHTH domain is comprised of a HTH motif followed by a 'wing' motif involving two β-strands connected by a small loop. The HTH motif contains two helices, α2 and α3 connected by a loop where α3 helix termed as the 'recognition' helix makes sequence specific contacts with cognate DNA (Fig. 4.1). The N-terminal amino acid residues of α1 helix along with some of the preceding amino acids, usually make nonspecific contacts with the sugar-phosphate backbone of the DNA and provides stability to the contacts made by the downstream amino acids (van Aalten et al., 2001). The recognition helices of the two monomers are inserted in the same central major groove to make base-specific contacts, while the two wings are docked into the flanking minor grooves of the DNA. In

this arrangement, the recognition helix is oriented roughly orthogonal to the DNA, where only residues present in its N-terminus contact the major groove. Notably, a conserved arginine in the N-terminus of recognition helix participates in forming bidentate hydrogen bond with a guanine residue in the major groove of DNA serving as a conserved feature of the GntR family. On similar lines, presence of a smaller amino acid such as glycine at the tip of the wing, which facilitates deeper contact into the minor groove, is another conserved feature of this family (van Aalten et al., 2001).

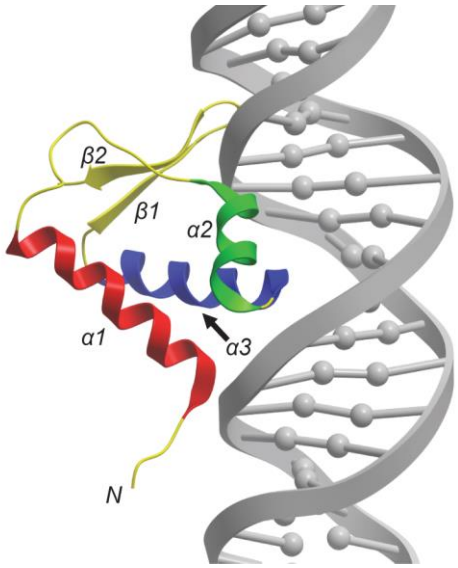

Figure 4.1 Model showing secondary structural elements in wHTH domain of AraR bound to its operator DNA (PDB ID: 4EGY). The information regarding structure of DNA bound AraR was taken from (Jain and Nair, 2013). The image was created in ICM browser (www.molsoft.com).

In this chapter, using a combination of genetics, biochemical, biophysical and bioinformatics approaches, we probed the features of DgoR involved in interaction with its cognate operator DNA. We showed that similar to other FadR subfamily members, DgoR is a largely helical protein, oligomerizes independent of DNA and exists majorly as dimers. The N-terminal wHTH domain of DgoR is involved in binding to the *dgo cis*-acting element. DgoR exhibits various features of DNA-protein interaction that are conserved or semiconserved in the FadR subfamily members.

4.2 Results

4.2.1 DgoR is a majorly α-helical protein

Previous bioinformatics analysis based on the presence of N-terminal wHTH domain and all helical C-terminal domain (Pfam accession number PF07729) predicted DgoR to be a member of the FadR subfamily within the GntR family of transcriptional regulators (Rigali et al., 2002; Suvorova et al., 2015). To confirm the above annotation we predicted the tertiary structure of DgoR using the Robetta server (Kim et al., 2004). The predicted structure showed the presence of an N-terminal wHTH DNA-binding domain and a C-terminal seven helical domain (Fig. 4.2A). Further, we recorded the Far-UV CD spectrum of purified DgoR. Purified DgoR exhibited a signature of a majorly helical protein (Fig. 4.2B). We used the server, BeStSel (Micsonai et al., 2018; Micsonai et al., 2015) to analyze the Far-UV CD spectrum of DgoR. Deconvolution suggested a composition of 74.9% helix, 0.0% antiparallel β-sheet, 0.0% parallel β-sheet, 9.3% turn and 15.8% other. We also calculated the secondary structure content of DgoR using 2Struc (DSSP) (Kabsch and Sander, 1983; Klose et al., 2010), which indicated a composition of 76.2% helix, 3.5% sheet and 20.3% other. Thus both CD spectra and the predicted tertiary structure suggest that DgoR is a majorly helical protein, a signature of FadR subfamily of transcriptional regulators.

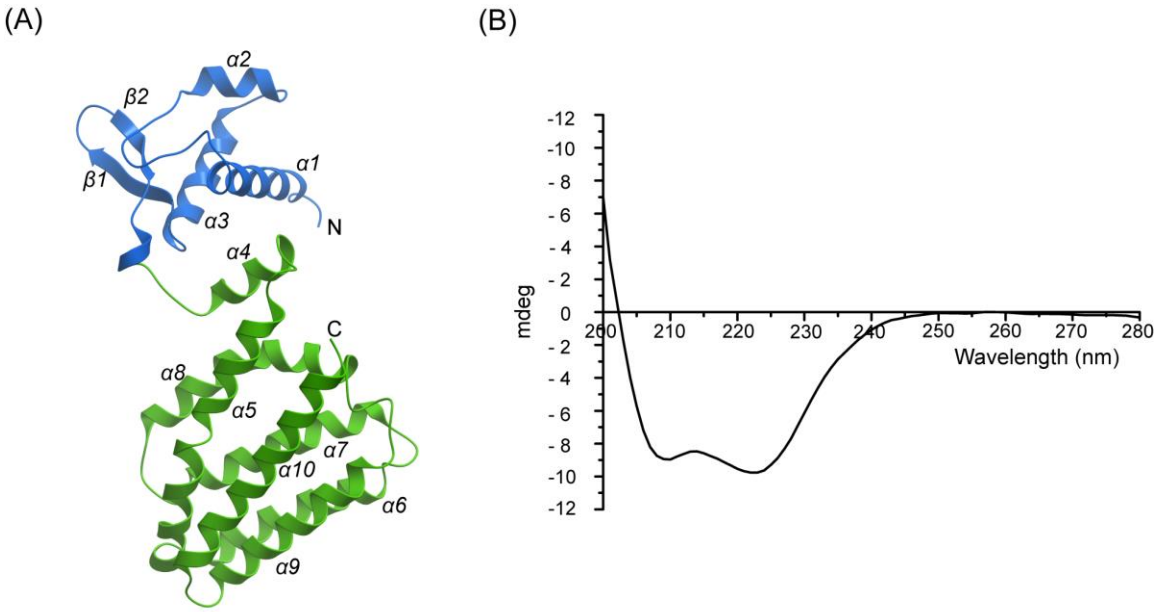

Figure 4.2 DgoR is a majorly helical protein. (A) Predicted structure of DgoR. The full-length amino acid sequence of DgoR was submitted to Robetta server for model building. The structure of a monomer of DgoR is shown. The N- and C- termini of the protein and its secondary structure elements are labeled. The N-terminal wHTH DNA-binding domain is shown in blue, and the C-terminal E-O domain is represented in green. (B) DgoR shows Far-UV CD spectrum of a mostly α-helical protein. Far-UV CD spectrum of purified DgoR-6XHis was recorded to determine its secondary structure. The experiment was performed 3 times. A representative data is shown.

4.2.2 Purified DgoR mainly exists as a dimer in solution

Most GntR family proteins exist as oligomers (Jain, 2015). To determine whether DgoR forms oligomers *in vitro*, DgoR-6XHis was immobilized on Co-NTA beads and incubated either with MBP-β-gal α fragment or MBP-DgoR. DgoR-6XHis was eluted using imidazole. Whereas MBP-β-gal α fragment did not co-elute with DgoR-6XHis, MBP-DgoR was pulled down by DgoR-6XHis suggesting that DgoR molecules interact with each other to form oligomers (Fig. 4.3A). Further, to identify the composition of oligomers, purified DgoR-6XHis was cross-linked with increasing concentrations of glutaraldehyde. Chemical cross-linking yielded dimer as the major oligomeric form. The higher order crosslinked species of DgoR-6XHis obtained with increasing concentrations of glutaraldehyde could be due to the non-specific crosslinking capacity of glutaraldehyde

(Fig. 4.3B). In subsequent studies, the size exclusion gel chromatography and analytical ultracentrifugation experiments conducted in our lab have also confirmed that DgoR exists as a dimer in solution (Garima Arya, unpublished data).

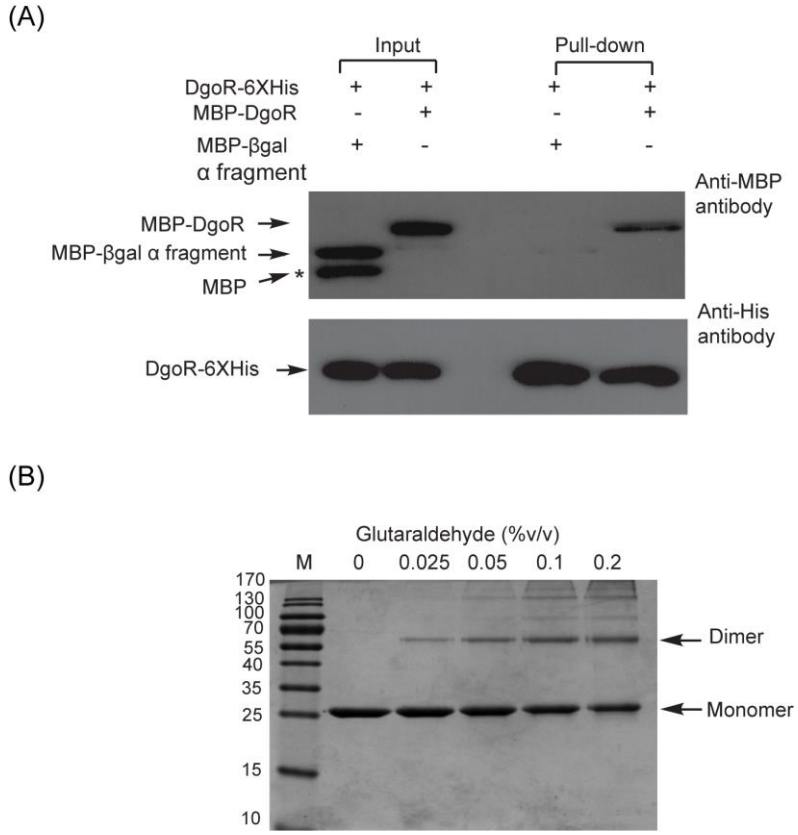

Figure 4.3 DgoR mainly exists as a dimer in solution. (A) DgoR oligomerizes *in vitro*. DgoR-6XHis from cell lysate was immobilized on Co-NTA beads to which either purified MBP-DgoR or MBP-β-gal α fragment was added. A portion of these suspensions was saved as input. The remaining suspension was incubated and proteins were finally eluted as pull-down. Samples were run on SDS-PAGE and processed for Western blotting using anti-MBP (upper panel) and anti-His (lower panel) antibody. '*' indicates band corresponding to MBP which likely arises due to spontaneous degradation of MBP-β-gal α fragment. Three independent experiments were performed. A representative data is shown. (B) Dimer is the major oligomeric form of DgoR. 10 µM DgoR-6XHis was cross-linked with different concentrations of glutaraldehyde and the cross-linked products were separated on 15% SDS-PAGE. Lane M represents the molecular mass markers (in kDa). Bands corresponding to the monomer (Mol. wt. ~27 kDa) and dimer (Mol. wt. ~ 54 kDa) forms of DgoR-6XHis are indicated. Three independent experiments were performed. A representative data is shown.

4.2.3 Missense mutations in the DNA-binding domain abrogate interaction of DgoR with the *dgo cis*-acting element

The N-terminal wHTH domain involved in DNA binding is composed of a tri-helical core followed by two beta strands connected through a small loop, designated as the "wing" motif (Fig. 4.2A). To ascertain whether secondary structure elements in the predicted DNA-binding domain of DgoR are involved in protein-DNA interaction, we created a missense mutation in each of the three α-helices (D7A in α1, L34A in α2 and R46A in α3), which are either conserved across the GntR family or reported to be functionally important in other characterized members (Rigali et al., 2002) (see discussion below). In addition, in a recent study where a human urinary tract isolate of *E. coli* was introduced in the streptomycin-treated mouse gut and mutations in the evolved lineages were followed over a span of a year, three missense mutations were independently recovered in the predicted α2-α3 region of DgoR. These data suggested the importance of de-repression of *dgo* operon in bacterial adaptation in the mouse gut. Indeed, the evolved lineages grew faster than the parental strain in D-galactonate implying a potential benefit to the evolved strains in utilizing this carbon source (Lescat et al., 2016). However, because there were mutations at additional loci in the evolved strains, a re-examination of *dgoR* mutations in a clean background coupled with testing their DNA-binding ability was imperative to assess their true contribution to the upregulation of *dgo* operon and adaptation in the host. To this end, we created mutations in *dgoR* corresponding to the three missense mutations (T40I, R42C and S51L) isolated in the above study.

To test the DNA binding ability of the DgoR mutants, we cloned *dgoR* carrying these mutations in IPTG inducible plasmid pRC10. The resulting plasmids (Table 2.1, pBS19, pBS20, pBS25, pBS32, pBS33 and pBS34) were used to overexpress and purify mutant DgoR as C-terminal His-tagged proteins. Far-UV CD spectra of the purified

proteins showed that the mutants are folded (Fig. 4.4). Deconvolution of the CD spectra using BeStSel suggested that the mutations do not significantly alter the secondary content of the protein except for S51L mutant which showed ~16 % decrease in the helical content compared to the WT protein (Table 4.1). In addition, crosslinking of purified DgoR mutants with glutaraldehyde showed that similar to WT, they dimerize in solution (Fig. 4.5).

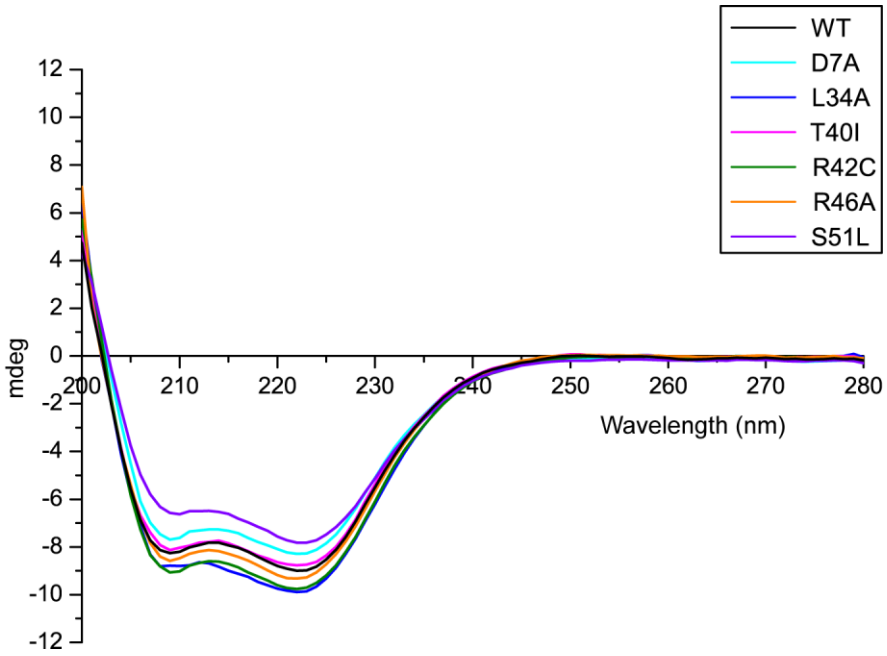

Figure 4.4 Various DNA-binding domain mutants of DgoR are folded. Far-UV CD spectra of WT DgoR-6XHis and its various mutants (5 µM) were recorded. The experiment was performed 3 times. A representative data is shown.

Table 4.1 Secondary structure content of WT and mutant DgoR proteins deconvoluted from their CD spectra

DgoR / % Secondary structure content	*WT*	*D7A*	*L34A*	*T40I*	*R42C*	*R46A*	*S51L*
Helix	72.9	70.1	74.9	74.7	76.9	70.4	61
Antiparallel	0.1	2.5	0	0	0	0	0
Parallel	0	0	0	0	0	0	1.4
Turn	10	11.1	7.5	9.4	9.3	8.5	11.8
Other	17	16.4	17.6	15.8	13.7	21.1	25.8

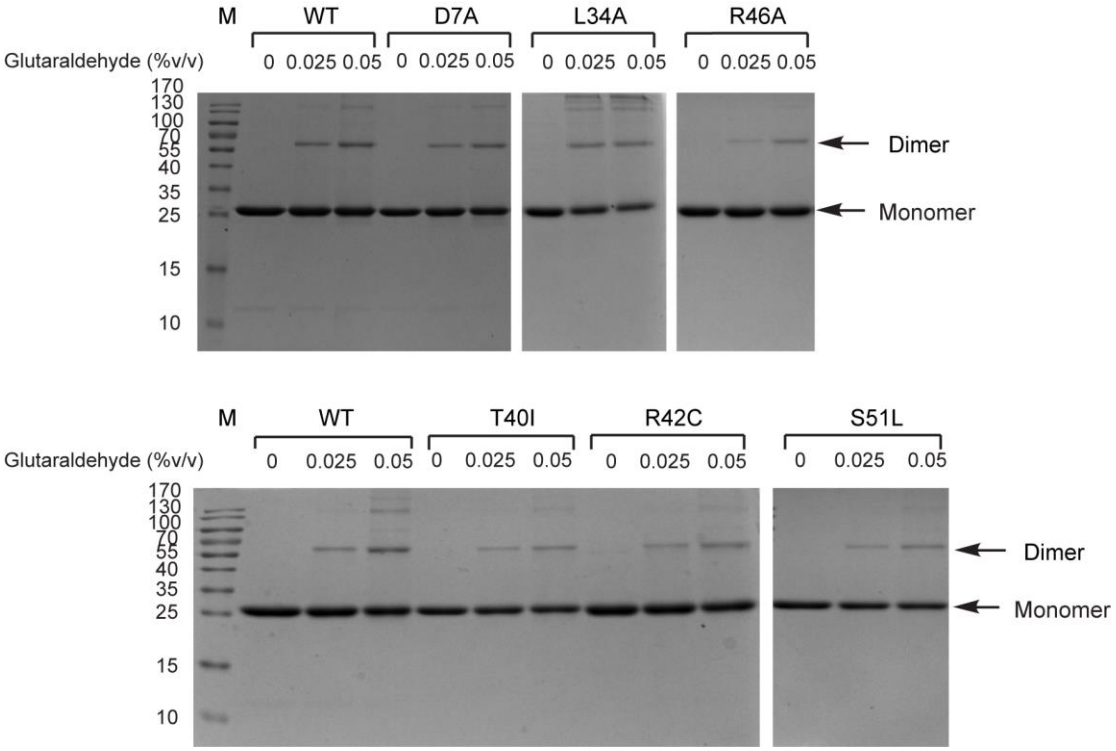

Figure 4.5 DNA-binding domain mutants form dimer as the main product upon glutaraldehyde cross-linking. 10 μM WT DgoR-6XHis and its various mutants were cross-linked with different concentrations of glutaraldehyde and the cross-linked products were separated on 15% SDS-PAGE. Lane M represents the molecular mass markers (in kDa). Bands corresponding to the monomer (Mol. wt. ~27 kDa and dimer (Mol. wt. ~54 kDa) forms of DgoR-6XHis are indicated.

To determine if the chosen mutations in the three α-helices ((D7A, L34A and R46A), and missense mutations (T40I, R42C and S51L) isolated in (Lescat et al., 2016) affect DgoR binding to its *cis*-acting element, the radiolabelled *dgo cis*-acting element was incubated with either WT DgoR-6XHis or its various mutants. In contrast to WT, all the tested mutants of DgoR were unable to shift the labelled *dgo cis*-acting element suggesting that the above mutations abrogate the binding of DgoR with the *cis*-acting element (Fig. 4.6).

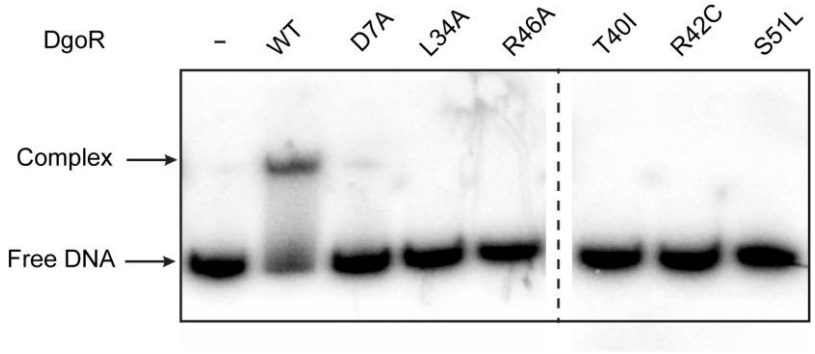

Figure 4.6 Mutations in the N-terminal wHTH domain of DgoR abrogate its binding to the *dgo cis*-acting element. DNA-binding domain mutants do not bind the *cis*-acting element *in vitro*. 20 nM ^{32}P-labeled DNA fragment was incubated either with 1.2 µM WT DgoR-6XHis or its various mutants. Samples were resolved on native PAGE and subjected to autoradiography.

4.2.4 Amino acid substitutions in the DNA binding domain leads to loss of DgoR mediated repression over D-galactonate metabolism

Our work from previous chapter showed that deletion of *dgoR* leads to elimination of repression over D-galactonate metabolism which manifests as faster growth of *dgoR*::*kan* strain in medium containing D-galactonate as the carbon source (Fig. 3.5, Chapter 3). Further, expression of WT DgoR from a plasmid with *dgoR* cloned with its promoter (Table 2.1, pBS13) complements the accelerated growth phenotype (Fig. 3.6, chapter 3). To investigate the ability of DgoR carrying mutations in the DNA binding domain to repress D-galactonate metabolism, we cloned various *dgoR* mutants along with *dgo*

promoter in pACYC177, and transformed the resultant plasmid constructs (Table 2.1, pBS22, pBS14, pBS15, pBS37,pBS36, and pBS35) and pBS13 carrying WT *dgoR* in *dgoR*::*kan* strain. The *dgoR*::*kan* strain expressing either WT or mutant DgoR was cultured in M9 medium containing D-galactonate as the carbon source. Consistent with the loss of DNA-binding ability observed in EMSA (Fig. 4.6), the mutants could not complement the growth phenotype of *dgoR*::*kan* in D-galactonate (Fig. 4.7). Further, we determined the expression of DgoR variants in the complementation strains in the non-inducing medium (M9 medium supplemented with glycerol), using Western blotting. The expression of DgoR mutants from their native promoter largely correlated with their repressor ability (Fig. 4.8). Whereas WT DgoR could not be detected due to auto-repression, D7A, T40I, R42C, R46A and S51L exhibited significant expression. Although the L34A mutant could not be detected likely due to its rapid degradation, its low-level expression inside the cell is evident from the ability of this mutant to exhibit weak dominant negative phenotype (see below, Fig. 4.10).

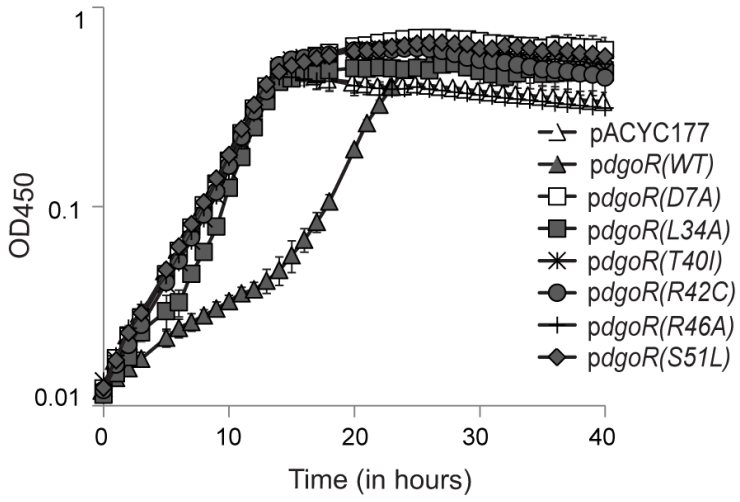

Figure 4.7 DNA-binding domain mutants fail to complement the faster growth phenotype of *dgoR* deletion strain in D-galactonate. WT and various *dgoR* mutants cloned in pACYC177 [p*dgoR(WT)*: pBS13, p*dgoR(D7A)*: pBS22, p*dgoR(L34A)*: pBS14, p*dgoR(R46A)*: pBS15, p*dgoR(T40I)*: pBS37, p*dgoR(R42C)*: pBS36, and p*dgoR(S51L)*: pBS35], were individually

transformed in a *dgoR*::*kan* strain. Cultures were grown in 96-well plates in minimal medium containing D-galactonate as the carbon source, and OD_{450} was measured. The experiment was performed 3 times; each experiment had 3 technical replicates. A representative data set, with average and SD of technical replicates, is shown.

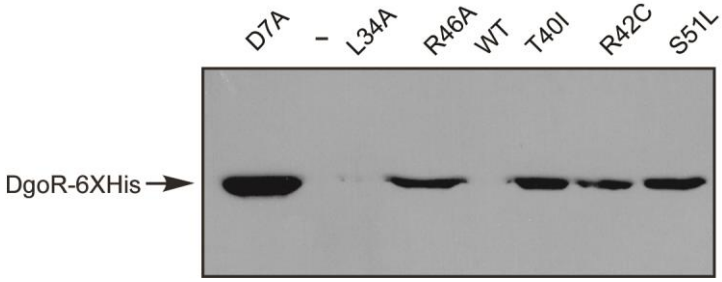

Figure 4.8 The expression of WT DgoR and various DNA-binding mutants from the native promoter largely correlates with their repressor ability. Strains described in the legend to figure 4.7 were grown in minimal medium supplemented with glycerol. Cells were harvested and processed for Western blotting using anti-His antibody. Two independent experiments were performed. A representative data is shown.

4.2.5 DNA-binding defective mutants of DgoR are incapable of repressing expression of the fluorescent reporter from the *dgo* promoter

We further tested the ability of the DNA-binding defective mutants of DgoR to repress expression from the *dgo* promoter *in vivo*. For this, we expressed either WT or mutant DgoR proteins from pACYC177 in a Δ*dgoR* reporter strain (Result section 3.2.5, Chapter 3). In the non-inducing medium, in contrast to the significant repression of the reporter in Δ*dgoR* expressing WT DgoR from the plasmid, the mutants behaved similar to Δ*dgoR* carrying empty plasmid, i.e., exhibited constitutive expression of the reporter (Fig. 4.9). In the inducing medium, as expected, all strains exhibited comparable reporter expression (Fig. 4.9). Taken together, our results emphasize that D7, L34, T40, R42, R46 and S51 amino acid residues are important for DgoR to bind its *cis*-acting element to repress *dgo* operon.

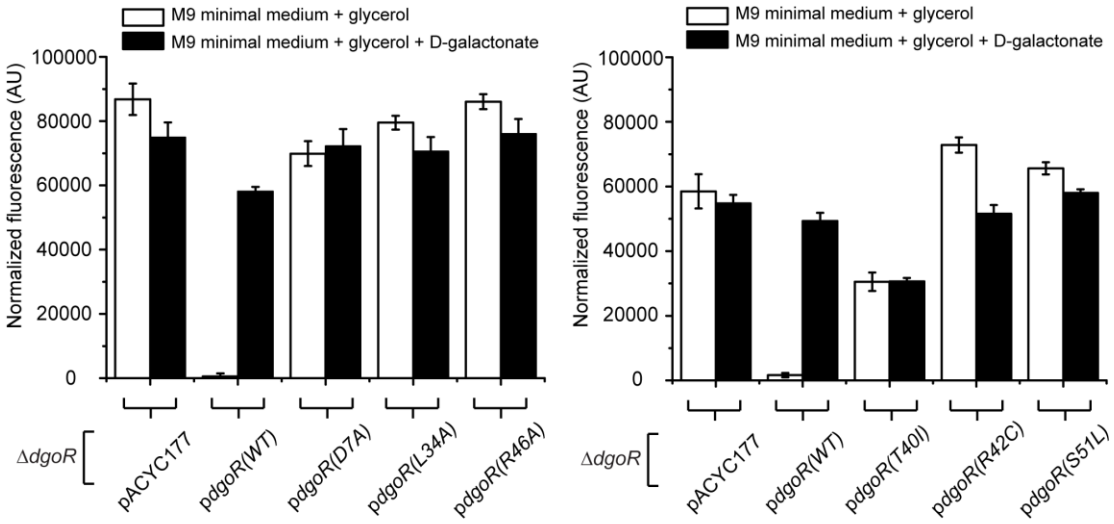

Figure 4.9 Fluorescence assay shows that DNA-binding domain mutants are unable to repress expression from the *dgo* promoter *in vivo*. Plasmids [p*dgoR(WT)*: pBS13, p*dgoR(D7A)*: pBS22, p*dgoR(L34A)*: pBS14, p*dgoR(R46A)*: pBS15, p*dgoR(T40I)*: pBS37, p*dgoR(R42C)*: pBS36, and p*dgoR(S51L)*: pBS35] were individually transformed in a Δ*dgoR* strain carrying fluorescent Venus reporter on the chromosome under the control of *dgo* promoter. Strains were grown in minimal medium supplemented either with glycerol or with glycerol and D-galactonate to exponential phase. Fluorescence was measured and normalized to OD_{450} of the samples. Data represent the average (± SD) of 3 independent experiments.

4.2.6 DgoR oligomerizes *in vivo*

In our *in vitro* assays, we had observed that DgoR forms oligomers (Fig. 4.3). We made use of the DNA-binding-defective mutants, D7A, L34A, R46A, T40I, R42C and S51L to investigate whether DgoR also oligomerizes *in vivo*. For this, we expressed either WT or mutant DgoR proteins from pACYC177 in a WT reporter strain. In the non-inducing medium, whereas the WT protein from the plasmid did not affect the expression of Venus reporter, mutant DgoR proteins increased reporter expression (Fig. 4.10). These results suggest that the DNA-binding-defective mutants expressed in *trans* oligomerize with WT DgoR from the chromosome interfering with its binding to the *cis*-acting element thereby exhibiting dominant negative phenotype. On the other hand, as expected, under inducing conditions, all strains exhibited similar expression of the Venus reporter (Fig. 4.10).

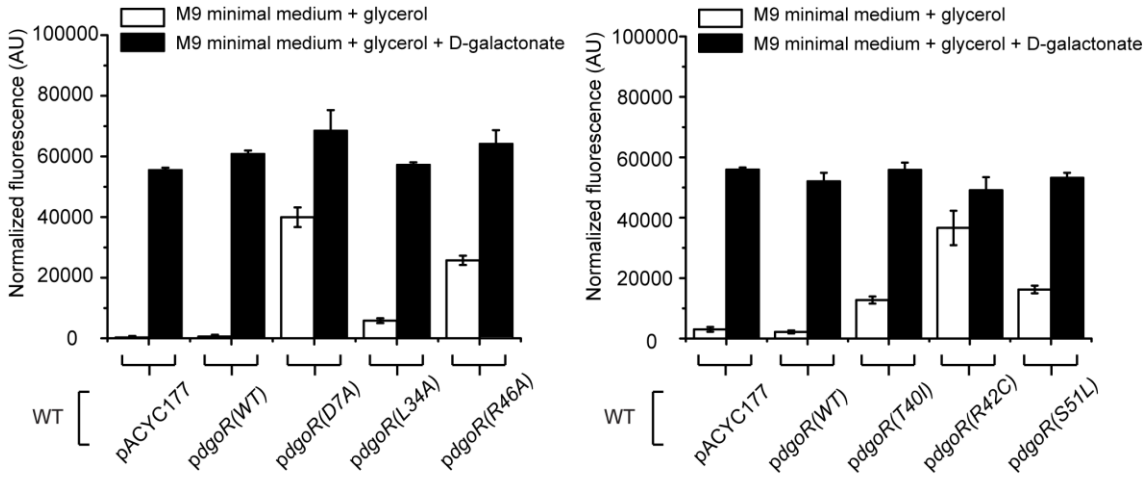

Figure 4.10 DgoR oligomerizes *in vivo.* DNA-binding-defective mutants exhibit dominant negative phenotype. The empty plasmid pACYC177 and various constructs, described in the legend to figure 4.7, were individually transformed in the WT strain carrying Venus reporter (under the control of native *dgo* promoter) on the chromosome. Strains were grown in minimal medium supplemented either with glycerol or with glycerol and D-galactonate to exponential phase. Fluorescence was measured and normalized to OD_{450} of the samples. Data represent the average (± SD) of 3 independent experiments. (One biological replicate of the fluorescence reporter assay for the mutants shown in the left panel was also perfomed in (Sangwan, 2017)).

4.3 Discussion

Transcriptional regulators belonging to the GntR family share multiple features such as structural similarity in their N-terminal DNA-binding domain, oligomeric status as well as similarities in the sequence of their cognate operator DNA. A number of DNA-protein contacts are conserved among the GntR family regulators. However, heterogeneity in the topological architecture of the C-terminal E/O domain among the various subfamilies and the organization of monomers in the oligomers also result in DNA-protein recognition features that are either semiconserved or unique even among the same subfamily members. In this chapter we established DgoR as a member of the FadR subfamily within the GntR family of transcriptional regulators. Further, as discussed below we compared

DgoR with other GntR family members for attributes that are conserved or semiconserved in the family.

Prediction of tertiary structure suggested that along with a typical N-terminal GntR-type wHTH domain, DgoR harbors an all helical C-terminal E/O domain. Predicted structure showed that the E/O domain of DgoR bears seven α-helices arranged in a bundle suggesting it to be a member of the FadR subgroup within the FadR subfamily (Fig. 4.2A). Further, a majorly α-helical far-UV CD profile of purified DgoR, similar to another FadR subfamily member McbR, validated the placement of DgoR in the FadR subfamily (Fig. 4.2B) (Lord et al., 2014). GntR family members including FadR subfamily members have been shown to dimerize independent of DNA (reviewed in (Jain, 2015)). The *in vivo* oligomerization of GntR family regulators such as FadR from *E. coli*, and GntR and AraR from *B. subtilis* were demonstrated by dominant negative behavior of their DNA-binding defective mutants (Franco et al., 2006; Raman et al., 1997; Yoshida et al., 1993). Here, the DNA-binding defective mutants of DgoR also exhibited negative dominance when expressed in *trans* (Fig. 4.10). Also, similar to TraR from *Streptomyces nigrifaciens*, pull down experiments showed that DgoR can oligomerize independent of DNA (Fig. 4.3A) (Kataoka et al., 2008). Finally, chemical crosslinking of purified protein showed that DgoR majorly exists as dimers (Fig. 4.3B). The dimeric status of DgoR has also been validated by size exclusion gel chromatography and analytical ultracentrifugation experiments conducted in our lab (Garima Arya, unpublished data).

The mechanistic details of the interaction between the wHTH domain and the operator is derived mainly from the DNA-bound structures of GntR family regulators belonging to different subfamilies [AraR subfamily: *B. subtilis* AraR (Jain and Nair, 2013); FadR subfamily: *E. coli* FadR (van Aalten et al., 2001; Xu et al., 2001) and *V.*

cholerae FadR (Shi et al., 2015); and HutC subfamily: *B. subtilis* NagR (Fillenberg et al., 2015)], and genetic/biochemical characterization of these and additional GntR members (Arai et al., 1999; Franco et al., 2006, 2007; Gao et al., 2008; Kalivoda et al., 2013; Lord et al., 2014; Raman et al., 1997). DgoR binds the operator sequence [5'-TTGTA(G/C)TACA(A/T)-3'] matching the signature of FadR subfamily members that recognize inverted repeats [5'-$(N)_y$GTM-$(N)_{0-1}$-KAC$(N)_y$-3'] (Hoskisson and Rigali, 2009; Jain, 2015; Suvorova et al., 2015). To understand how amino acid residues in DgoR mutated in this study affect its DNA-binding ability, we obtained the DNA bound state of DgoR by superimposing wHTH domain of a DgoR monomer onto the wHTH domain of *E. coli* FadR bound to its operator (PDB ID: 1H9T) (Fig. 4.11).

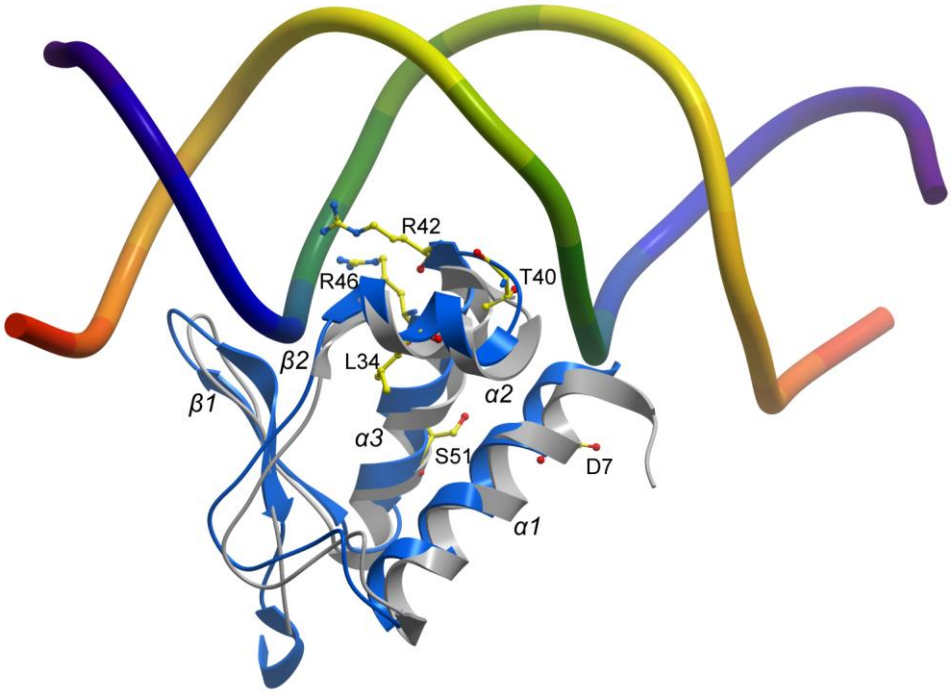

Figure 4.11 Proposed DNA-binding model of DgoR. The DNA bound state of DgoR was obtained by superposition of N-terminal wHTH domain of DgoR (blue) onto the operator-bound structure of N-terminal wHTH domain of *E. coli* FadR (grey) (PDB ID: 1H9T). Amino acid residues in the wHTH domain of DgoR chosen for mutation in this study are shown as sticks and labeled.

In the characterized GntR family proteins, amino acid residues in α1 make non-specific contacts with the phosphate backbone and support the specific contacts made by the downstream residues (Fillenberg et al., 2015; Gao et al., 2008; Jain, 2015; Shi et al., 2015; van Aalten et al., 2001; Xu et al., 2001). In the DgoR-DNA model, although the side chain of D7, an α1 residue, points towards the DNA, it does not make any contact (Fig. 4.11). However, similar to the DNA-binding defect of D7A mutant of DgoR, a previous study on another FadR subfamily member, AphS, reported a loss of repression by a mutant harboring mutation in the corresponding α1 residue, E17 (Arai et al., 1999). A Glu or Asp residue is present at the analogous position in several GntR family members (Rigali et al., 2002), thus an acidic residue at the start of α1 may play an important role in the repressor ability of these regulators.

Residues in α2 and α3 form specific contacts in the major groove with α3 defining a larger part of the specificity and hence termed as the 'recognition helix' (Jain, 2015). The information that amino acid substitution of several residues in α2-α3 abrogated the interaction of DgoR with its operator (Fig. 4.6), further underscores the importance of wHTH motif in DNA recognition by GntR family regulators. Corresponding to the α2 residue, L34 in DgoR, a hydrophobic amino acid is present in a majority of the GntR family members (Fig. 4.12) (Rigali et al., 2002). Mutation of an equivalent Leu in FadR (L37A) and AraR (L33S) led to a weak dominant negative phenotype and poor expression of the protein, respectively, suggesting that this hydrophobic amino acid is important for protein stability (Franco et al., 2006, 2007; Raman et al., 1997). On similar lines, we observed a weak dominant negative phenotype of L34A mutant of DgoR and were unable to detect its expression from the native *dgo* promoter (Fig. 4.10 and Fig. 4.8). In the DgoR-DNA model, we find that L34 is part of a

hydrophobic core further emphasizing that a hydrophobic amino acid at this position is structurally important in GntR family regulators (Fig. 4.13). Additionally, the importance of L34 residue in DNA-binding is evident from both the inability of L34A to form protein-DNA complex *in vitro* and negative dominance *in vivo* (Fig. 4.6 and Fig. 4.10). In the DgoR-DNA model, L34 interacts with two hydrophobic residues I45 and F49 located in α3 (Fig. 4.13). Because mutating L34 to A will reduce hydrophobicity and increase the distance of I45 and F49 from α2, we suggest that the recognition helix might get re-oriented affecting its contact with DNA.

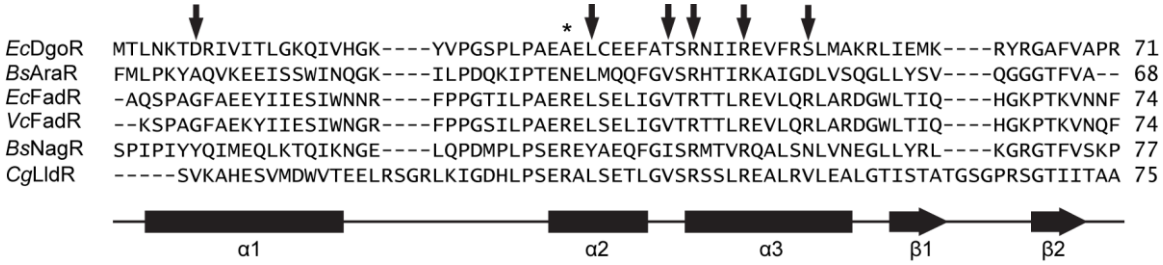

Figure 4.12 Structure-based sequence alignment of N-terminal wHTH DNA-binding domain of GntR family members. Amino acid sequence of the N-terminal wHTH domain of DgoR is aligned with that of GntR family members for which either the DNA-bound structures are solved (*Bs*AraR, PDB ID: 4EGY; *Ec*FadR, PDB ID: 1H9T; *Vc*FadR, PDB ID: 4P9U; and *Bs*NagR, PDB ID: 4WWC) or the structure of their apo-form is modeled on DNA (*Cg*LldR, PDB ID: 2DI3), using DALI server (Holm and Laakso, 2016). Vertical arrows indicate amino acid residues in the wHTH domain of DgoR mutated in this study. Secondary structure elements are indicated in bars (α-helices) and horizontal arrows (β-strands). '*' shows the presence of Arg residue at this position in GntR members that recognize 'TGGT' motif in the operator (see below Fig. 4.15). Abbreviations: *Ec, E. coli*; *Bs, B. subtilis*; *Vc, V. cholerae*; and *Cg, C. glutamicum*.

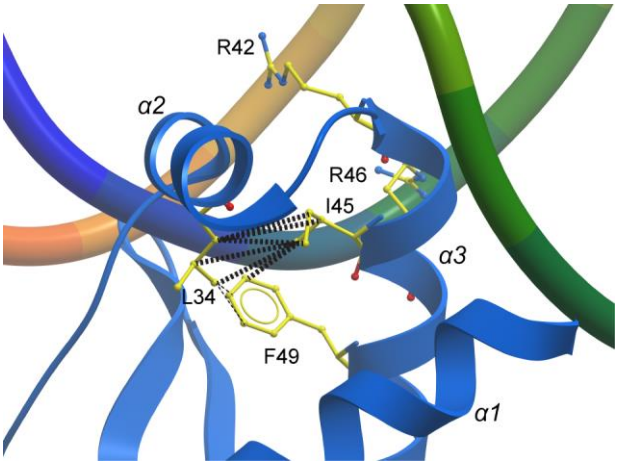

Figure 4.13 Amino acid contacts of L34 in wHTH of DgoR. Model shows hydrophobic interaction of L34 in α2 with I45 and F49 in α3 (recognition helix).

Similar to the DNA-binding defect of R42C and R46A in DgoR, mutation of Arg residues at equivalent positions in α3 in several other GntR family members compromises their repressor ability (Franco et al., 2006, 2007; Gao et al., 2008; Raman et al., 1997). From the available DNA-bound structures, it is evident that both Arg residues interact with a common Glu residue in α2, which orients them such that arginine corresponding to R42 makes a specific contact with guanine (G_{-3}) while arginine equivalent to R46 interacts non-specifically with the phosphate backbone (Fillenberg et al., 2015; Gao et al., 2008; Jain and Nair, 2013; Shi et al., 2015; van Aalten et al., 2001; Xu et al., 2001). In the DgoR-DNA model, we observe the interaction of R42 and R46 with an equivalent Glu residue, E31 in α2 (Fig. 4.14), which leads us to speculate that these Arg residues make similar contacts with DNA as observed in other characterized members (Fig. 4.12). GntR family regulators that recognize 'TGGT' motif in the operator make a specific contact with 'G_{-4}' through an Arg residue conserved in these members (Fig. 4.15 and Fig. 4.12) (Fillenberg et al., 2015; Gao et al., 2008; Rigali et al., 2002; Shi et al., 2015; van Aalten et al., 2001; Xu et al., 2001). However, GntR family proteins including DgoR that recognize 'TTGT' motif lack the analogous Arg (Fig. 4.15 and Fig.

4.12) (Jain and Nair, 2013; Rigali et al., 2002), emphasizing that differences in the operator sequence are reflected in the amino acid sequence of their cognate regulators.

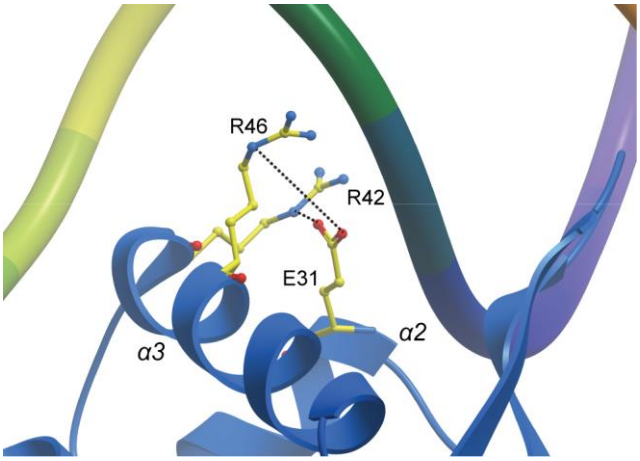

Figure 4.14 Amino acid contacts of R42 and R46 in wHTH of DgoR. The model shows E31 in α2 forming a salt bridge with R42 and R46 in α3.

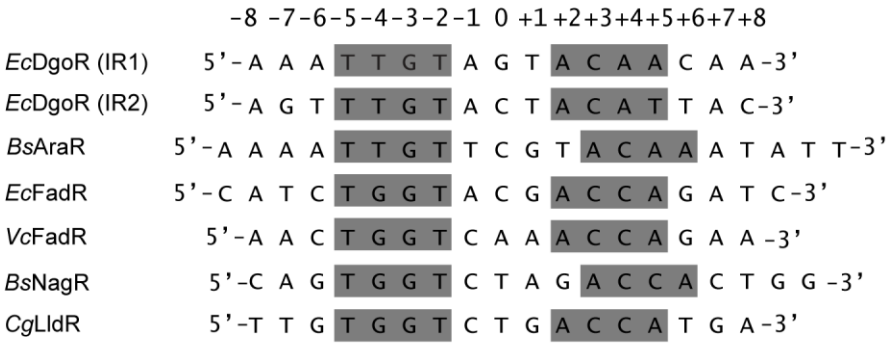

Figure 4.15 Comparison of operator sequences of GntR family members. Binding sites of DgoR (IR1 and IR2) are compared with the operator sequences of GntR family members used to either solve the DNA-bound structures of their wHTH domain (*Bs*AraR, *Ec*FadR, *Vc*FadR, and *Bs*NagR) or biochemically test the binding to its transcriptional regulator (*Cg*LldR). Nucleotides in the inverted repeat that are conserved and/or involved in making specific contacts with the protein are highlighted in gray boxes. Abbreviations: *Ec*, *E. coli*; *Bs*, *B. subtilis*; *Vc*, *V. cholerae*; and *Cg*, *C. glutamicum*.

The DNA-binding defect of T40I and S51L cannot be explained from the DgoR-DNA model. Although the side chain of T40 points towards the DNA, there is no

interaction with the DNA (Fig. 4.11). In a majority of the GntR family regulators, a hydrophobic amino acid is present at a similar position (Fig. 4.12) (Rigali et al., 2002). The residue S51 does not face the DNA (Fig. 4.11), and the nature of amino acid at a similar position is highly variable across the GntR family (Fig. 4.12) (Rigali et al., 2002). The Far-UV CD spectra shows that there is ~16 % decrease in helical content of S51L mutant (Fig. 4.3). The altered conformation of S51L could be responsible for its loss of DNA binding. Future studies aimed at solving the structure of DgoR alone and in complex with its operator might explain the importance of these amino acid residues in the functioning of the repressor.

Our work in this chapter, besides providing a fundamental understanding of the interaction between DgoR and its cognate operator DNA, also emphasizes that upregulation of *dgo* genes caused by disruption of the interaction of DgoR with its operator determines the colonization efficiency of *E. coli* isolates. This is evident from our observation that missense mutations in *dgoR* (T40I, R42C and S51L) selected in a urinary tract isolate of *E. coli* adapated to the mammalian gut (Lescat et al., 2016) are DNA-binding defective and result in constitutive expression of *dgo* genes. The above information combined with the study where a comparison of a panel of 340 natural isolates of *E. coli* revealed variations in *dgoR* (Galardini et al., 2017) stress the importance of studying the effect of genetic variations in *dgoR* in the natural isolates on their colonization in biological niches.

CHAPTER V

D-galactonate is the physiological effector of DgoR

5.1 Introduction

In order to sense change in the environment, transcriptional regulators bind relevant metabolites, ions, or drug molecules as effectors. Binding of effectors usually brings about a conformational change in the protein, thus altering its DNA-binding ability. The binding of effector can either increase (activator) or decrease (repressor) the affinity of the regulator to its cognate DNA. This allows transcriptional regulators to act as molecular switches to turn on or off the expression of its regulon members in response to an appropriate signal. GntR family transcriptional regulators are involved in regulating diverse physiological processes such as metabolism, virulence, biofilm production, plasmid transfer, etc (Hoskisson and Rigali, 2009; Rigali et al., 2002). GntR family regulators display modularity in their function, where the N-terminal wHTH domain and the C-terminal E/O domain are responsible for binding cognate operator DNA and the effector, respectively. GntR family members involved in regulating various metabolic pathways in bacteria bind either the substrate itself or its metabolic intermediate or both as their effectors (Bates Utz et al., 2004; Bouvier et al., 2019; DiRusso et al., 1992; DiRusso et al., 1998; Hoskisson and Rigali, 2009; Jain, 2015; Kalivoda et al., 2003; Lee et al., 2000; Miwa and Fujita, 1988; Tutukina et al., 2016a). Binding of the effectors induces an allosteric change in the E/O domain, which in turn alters the conformation of the DNA-binding domain thus modulating the DNA binding properties of the regulator (reviewed in (Jain, 2015)).

Within the GntR family, the various subfamilies have topologically distinct E/O domains. Thus, the molecular mechanism underlying effector mediated regulation of DNA binding is quite different among the subfamilies (reviewed in (Jain, 2015)). For example *B. subtilis* NagR (YvoA), a member of HutC subfamily, forms homodimers and has a canonical E/O domain bearing an α/β topology [UTRA domain (Pfam accession

number PF07702)] (Aravind and Anantharaman, 2003; Fillenberg et al., 2016; Fillenberg et al., 2015). In the DNA-bound state, the amino acid residues in the region connecting the DNA-binding and the effector-binding domains constitute a partially disordered loop. Upon binding its effector, this region undergoes a dramatic loop-to-helix transition ultimately leading to forced separation of DNA-binding domain in a 'jumping jack'-like fashion due to its 122° rotation with respect to the E/O domain (Resch et al., 2010). The *C. glutamicum* CGL2947, a member of YtrA subfamily harbors a typically shorter 'fish-hook' shaped E/O domain comprised of two α-helices. Comparison of CGL2947 bound to 2-methyl-2,4-pentanediol (MPD) (a component of the crystallization buffer) with that of its apo- form led to the suggestion regarding its effector binding mode (Gao et al., 2007). Two monomers interact through their E/O domain to form a functional homodimer, where the α-helices in the E/O domain of the two monomers intertwine to form an effector-binding cavity. In the apo- form, the two recognition helices in the N-terminal DNA-binding domain are 64 Å apart ideal for binding two adjacent major grooves of the operator DNA. Binding of MPD leads to conformational changes in the effector binding pocket, which are transmitted to the DNA-binding domain such that the effective distance between the two α-helices is decreased to 56 Å, making the regulator unsuitable to bind the DNA (Gao et al., 2007).

Within the FadR subfamily, effector-induced allosteric changes have been elucidated only for FadR from *E. coli*, *V. cholera*, and *V. alginolyticus* by a comparison of their apo-, DNA- and/or effector-bound structures (Gao et al., 2017; Shi et al., 2015; van Aalten et al., 2001). The effector, fatty acyl-CoA binds *E. coli* FadR in a cavity in the E/O domain more than 30 Å away from the wHTH DNA binding motif. Binding of acyl-CoA induces dramatic conformational changes in the E/O domain, causing the α8 helix to push against the α4 linker helix which in turn makes additional contacts with α1 to tilt and

displace the entire wHTH motif with respect to the E/O domain. (Fig. 5.1). Overall hinge-bending motion in the protein dimer in the opposite direction leads to a separation of ~7.2 Å of the α3 DNA recognition helices resulting in the loss of DNA binding (Fig. 5.1) (van Aalten et al., 2001). Random mutagenesis coupled with biochemical studies also yielded information on the amino acid residues involved in the interaction of *E. coli* FadR with fatty acyl-CoA (Raman and DiRusso, 1995). Likewise, a few genetic/biochemical studies have identified amino acid residues important for other FadR subfamily regulators (Blancato et al., 2016; Franco et al., 2006; Lee and Kim, 2001; Lord et al., 2014). Further, a limited number of structural studies on the apo- form have led to a prediction of the effector binding cavity in some of the transcriptional regulators belonging to the FadR subfamily (Gao et al., 2008; Little et al., 2018; Lord et al., 2014; Pinheiro et al., 2018).

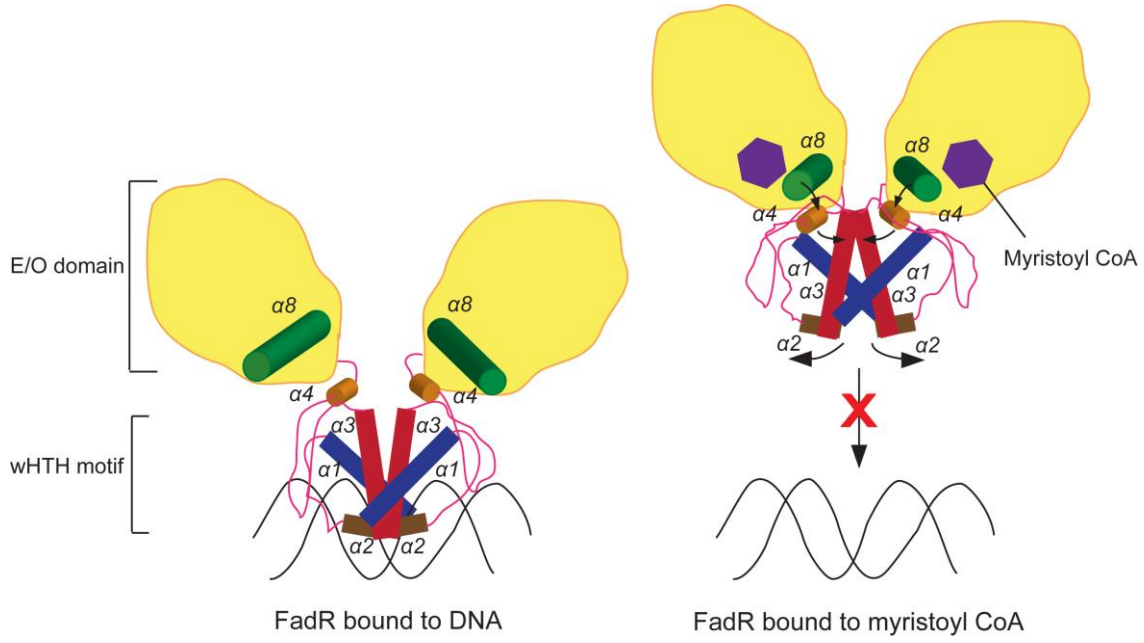

Figure 5.1 Schematic depicting molecular basis of fatty acyl-CoA mediated relief of DNA binding by FadR. The recognition helix (α3), linker helix (α4), α1 helix, α2 helix and, α8 helix of the E/O domain are shown in red, orange, blue, brown and green, respectively. The effector fatty acyl-CoA (myristoyl-CoA) is shown in purple. The arrows depict the conformational changes induced in FadR upon binding fatty acyl-CoA. The schematic was drawn using information provided in (van Aalten et al., 2001).

Amongst the FadR subfamily members involved in regulating the metabolism of various sugar acids, the associated physiological effectors are known for *B. subtilis* GntR (D-gluconate), *E. coli* UxuR/ExuR (D-glucuronate/D-galacturonate) and *Polaromonas* sp. JS666 GguR (5-keto-4-deoxy-D-glucarate/galactarate) (Bates Utz et al., 2004; Bouvier et al., 2019; Miwa and Fujita, 1988; Tutukina et al., 2016a). A random mutagenesis screen conducted in 1990s identified a couple of amino acids in the E/O domain of the *B. subtilis* GntR which were suggested to be involved in the interaction with its effector, D-gluconate (Yoshida et al., 1995). Another computational modelling and docking based study predicted the amino acid residues important for interaction of *E. coli* UxuR and ExuR with their effectors (Tutukina et al., 2016b). Thus, clearly there are not enough in-depth studies addressing the molecular details of the interaction of the FadR subfamily sugar acid metabolism regulators with their cognate effectors (Jain, 2015). *E. coli* DgoR, which we have characterized as the negative regulator of D-galactonate metabolism can serve as a paradigm for such studies (Singh et al., 2019). To initiate work in this direction, we identified the physiological effector of DgoR.

In this chapter, we investigated the effector binding characteristics of DgoR. We find that similar to *B. subtilis* GntR, the substrate of the metabolic pathway, D-galactonate, serves as the true effector of DgoR. Our data reveals that D-galactonate and not the D-galactonate catabolic intermediates or other structurally related carbohydrates interferes with the binding of DgoR to the target DNA. Further, we find that mutation of S221 amino acid in the E/O domain of DgoR leads to defects in D-galactonate mediated derepression of the *dgo* promoter.

5.2 Results

5.2.1 D-galactonate is the specific effector of DgoR

A previous report showed that the activities of DgoK and DgoA, the catabolic enzymes involved in D-galactonate metabolism, are induced in cell lysates prepared from *dgoD* mutant grown in medium supplemented with D-galactonate. Given that DgoD is involved in catalyzing the first metabolic reaction in D-galactonate metabolism, no D-galactonate catabolic intermediate are supposed to be produced in a *dgoD* mutant. Thus, it was proposed that D-galactonate itself acts as the effector to induce expression of various components involved in D-galactonate metabolism (Cooper, 1978). However, since the catabolic pathways of sugar acids cross-talk with each other either due to their enzymatic inter-conversion or shared transcriptional regulators (Mandrand-Berthelot et al., 2004), it is important to examine whether D-galactonate is the true effector of DgoR.

Amongst all tested sugars/sugar acids, only D-galactonate abrogates the binding of DgoR to its cis-acting element

To determine if D-galactonate is a specific effector of DgoR, we incubated DgoR with its *cis*-acting element in the presence of various concentrations of D-galactonate, sugars (D-glucose and D-galactose) and an epimer of D-galactonate (D-gluconate) (Fig. 5.2). Whereas D-galactonate relieved DNA bound by DgoR in a concentration-dependent manner, other tested carbohydrates did not affect the binding of DgoR to its target DNA (Fig. 5.3).

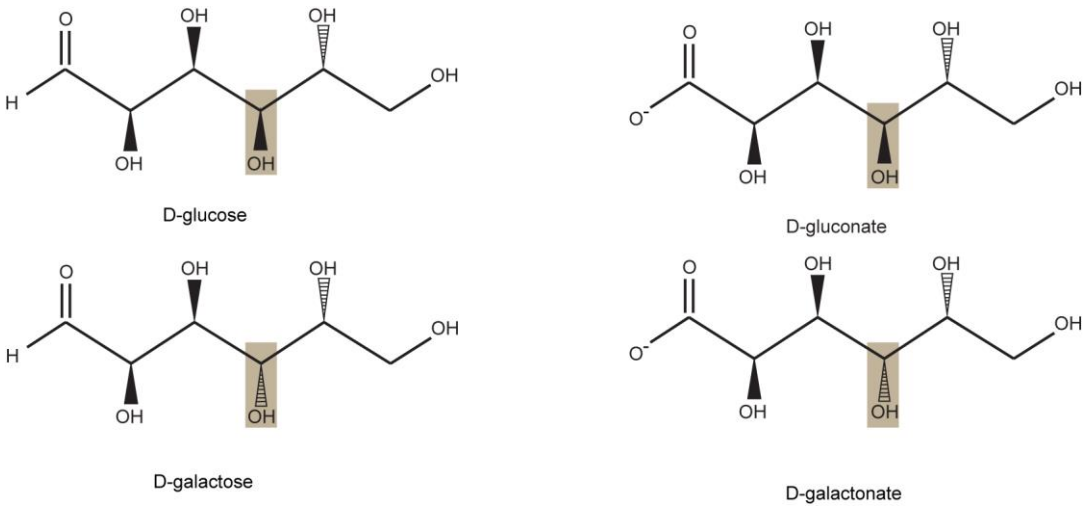

Figure 5.2 Structures of various carbohydrates incubated with DgoR in EMSA. Structures of D-glucose, D-gluconate, D-galactose and D-galactonate are shown. The –OH varying amongst the different epimers is highlighted in gray box. The structural information of the molecules was obtained from EcoCyc database (Keseler et al., 2017).

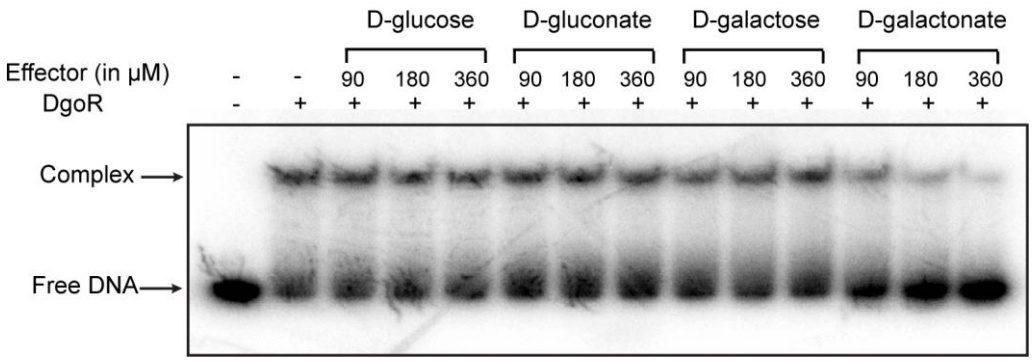

Figure 5.3 D-galactonate releases DgoR from its target DNA. 1.2 μM DgoR-6XHis was incubated with indicated concentrations of various carbohydrates for 20 min. 20 nM ^{32}P-labeled *dgo cis*-acting element was added to the samples and incubated for 30 min. Samples were resolved on native PAGE and subjected to autoradiography.

Intermediates of the D-galactonate metabolic pathway do not interfere with the binding of DgoR to its cis-acting element

Several FadR subfamily transcriptional regulators recognize both the substrate and its metabolic intermediates as effectors to regulate the transcription of their cognate regulon members (Bates Utz et al., 2004; Bouvier et al., 2019; DiRusso et al., 1992; DiRusso et al., 1998; Hoskisson and Rigali, 2009; Jain, 2015; Kalivoda et al., 2003; Lee et al., 2000; Miwa and Fujita, 1988; Tutukina et al., 2016a). To determine if in addition to D-galactonate, any of its catabolic intermediates affect the DNA-binding ability of DgoR, we incubated DgoR with its *cis*-acting element in the presence of 2-dehydro-3-deoxy-D-galactonate, 2-dehydro-3-deoxy-D-galactonate 6-phosphate, D-glyceraldehyde 3-phosphate and pyruvate (Fig. 5.4). None of the intermediates could abrogate DNA binding by DgoR (Fig. 5.5).

Taken together, our data provide evidence that D-galactonate is the specific effector of DgoR.

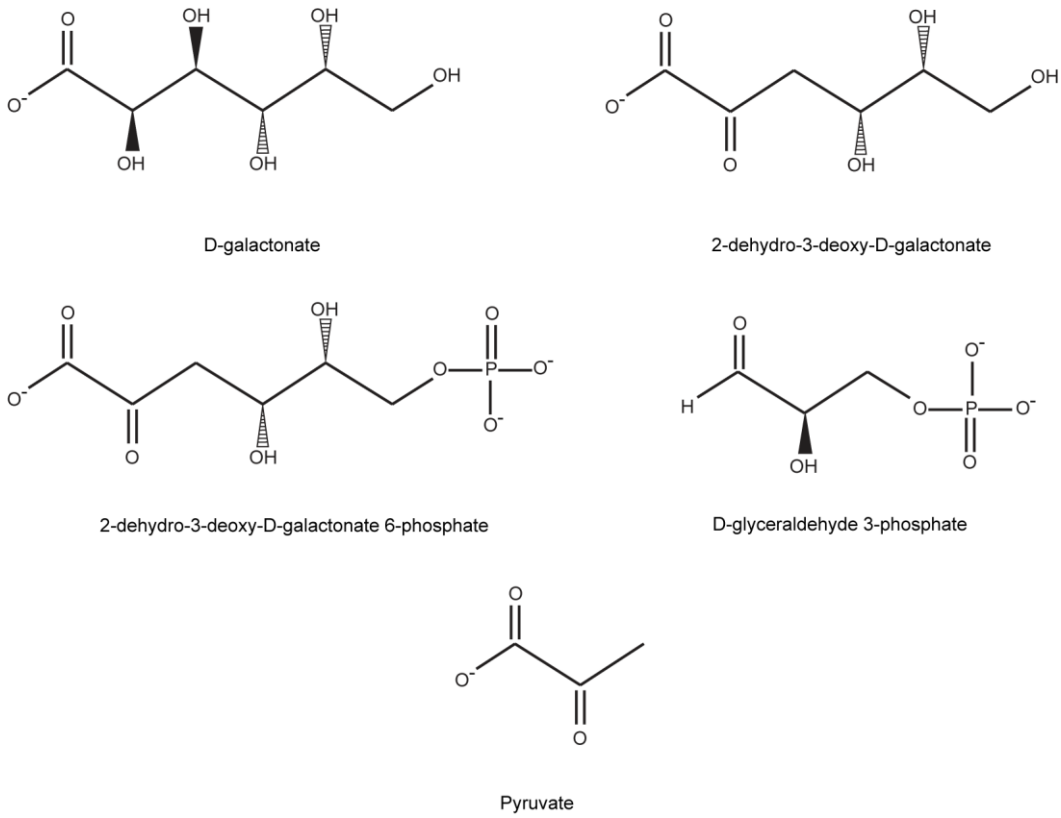

Figure 5.4 Structures of D-galactonate and intermediates of the D-galactonate metabolic pathway in *E. coli*. The structural information of the molecules was obtained from EcoCyc database (Keseler et al., 2017).

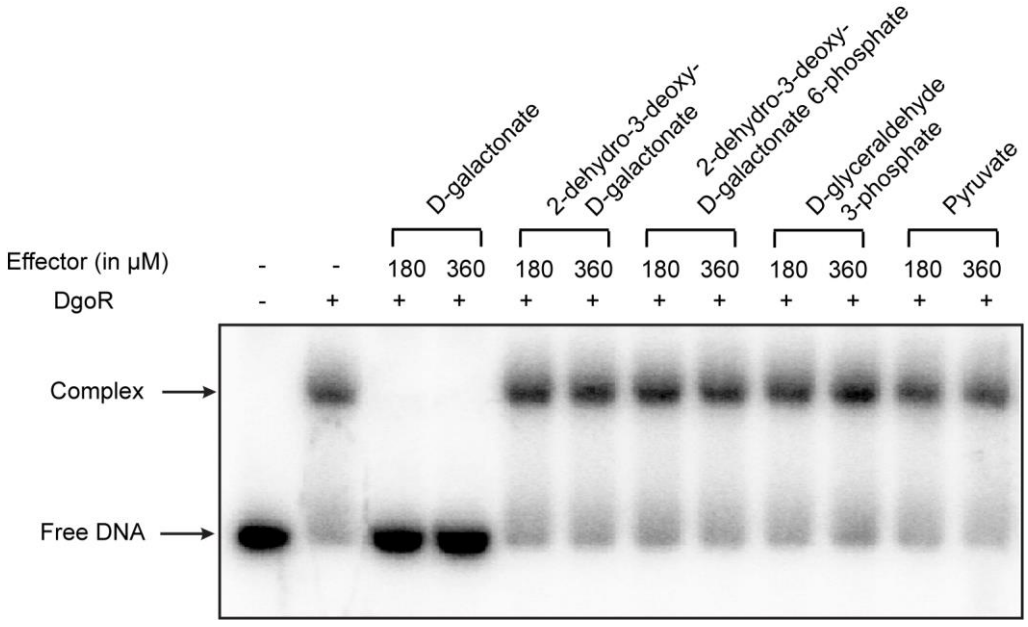

Figure 5.5 D-galactonate catabolic intermediates do not release DNA bound by DgoR. 1.2 μM DgoR-6XHis was incubated with indicated concentrations of various D-galactonate catabolic

intermediates for 20 min. 20 nM ^{32}P-labeled *dgo cis*-acting element was added to the samples and incubated for 30 min. Samples were resolved on native PAGE and subjected to autoradiography.

5.2.2 D-galactonate binds DgoR-DNA complex with µM affinity

To determine the binding affinity of DgoR with its cognate effector, purified DgoR-6XHis was mixed with various concentrations of D-galactonate and incubated with ^{32}P-labeled *dgo cis*-acting element. The fraction of DgoR-DNA complex dissociated with increasing amount of D-galactonate was used to calculate apparent K_D of D-galactonate for DgoR-DNA complex (Fig. 5.6), and was estimated to be 55 ± 6 µM.

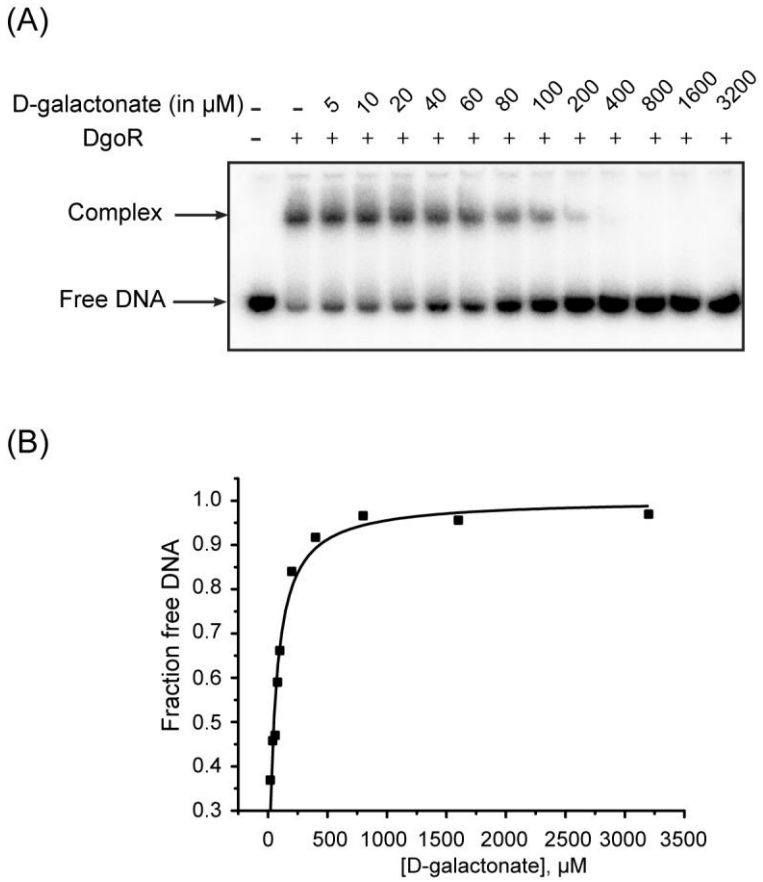

Figure 5.6 Determination of the apparent dissociation constant of D-galactonate. (A) EMSA shows titration of DgoR-DNA complex with D-galactonate. 20 nM ^{32}P-labeled DNA fragment was incubated with 1.2 µM purified DgoR-6XHis and the protein-DNA complex was titrated with indicated concentrations of D-galactonate. (B) The plot shows the fraction of free DNA as a function of D-galactonate concentration. The fraction of free DNA quantified from image shown

in panel A was plotted against D-galactonate concentration. Three independent experiments were performed to determine apparent K_D. A representative data set is shown.

5.2.3 S221N mutant of DgoR exhibits super-repressor phenotype

Structure based sequence alignment of the E/O domain of DgoR with the well-characterized member of FadR subfamily, *E. coli* FadR, revealed that DgoR shares a serine amino acid at 221 position with serine at 219 position of FadR (Fig. 5.7) (Rigali et al., 2002). S219N mutant of FadR has a reduced affinity towards its cognate effector fatty acyl-CoA and thus showed a non-inducible, super-repressor phenotype (Raman and DiRusso, 1995).

Figure 5.7 DgoR and FadR share a common serine amino acid residue in the E/O domain. Structure based sequence alignment of DgoR and FadR was performed using DALI (Holm and Laakso, 2016). The common serine amino acid is enclosed in the red box. The secondary structural elements; helices (H), β-strands (E) and loops (L) of DgoR and FadR are indicated.

To determine if similar to FadR, mutation of S221 residue in DgoR (S221N) interferes with D-galactonate mediated derepression of *dgo* genes, *dgoR* with S221N mutation was cloned with its promoter in plasmid, pACYC177 (Table 2.1, pBS24). Plasmids pBS13 and pBS24 were used to express either WT or S221N mutant of DgoR-6XHis in *dgoR*::*kan* strain, respectively. The strains were cultured in M9 medium supplemented with D-galactonate as the carbon source. Expression of S221N mutant led to slower growth in D-galactonate compared to WT indicating defective derepression of *dgo* operon by the S221N mutant (Fig. 5.8).

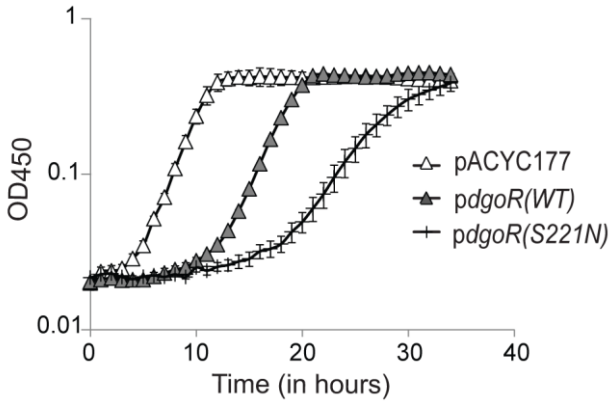

Figure 5.8 S221N mutant exhibits slower growth in D-galactonate. WT and S221N *dgoR* mutant cloned in pACYC177 [p*dgoR(WT)*: pBS13 and p*dgoR(S221N)*: pBS24], were individually transformed in a *dgoR*::*kan* strain. Cultures were grown in 96-well plates in minimal medium containing D-galactonate as the carbon source, and OD_{450} was measured. The experiment was performed 3 times; each experiment had 3 technical replicates. A representative data set, with average and SD of technical replicates, is shown.

Further, to determine the effect of S221N mutation in DgoR on the regulation of the *dgo* promoter, we expressed WT as well as S221N mutant DgoR in a Δ*dgoR* reporter strain (Result section 3.2.5, Chapter 3). Both WT DgoR and the S221N mutant repressed the reporter in the non-inducing medium (Fig. 5.9). However, in the presence of D-galactonate, whereas significant reporter expression was observed for the strain

expressing WT DgoR, the reporter strain expressing S221N mutant exhibited weak induction suggesting that S221N exhibits stronger repression over the *dgo* promoter as compared to the WT (Fig. 5.9).

Taken together, both the complementation and fluorescence reporter assays validate that S221 amino acid residue is important for the functioning of DgoR as the regulator of *dgo* operon.

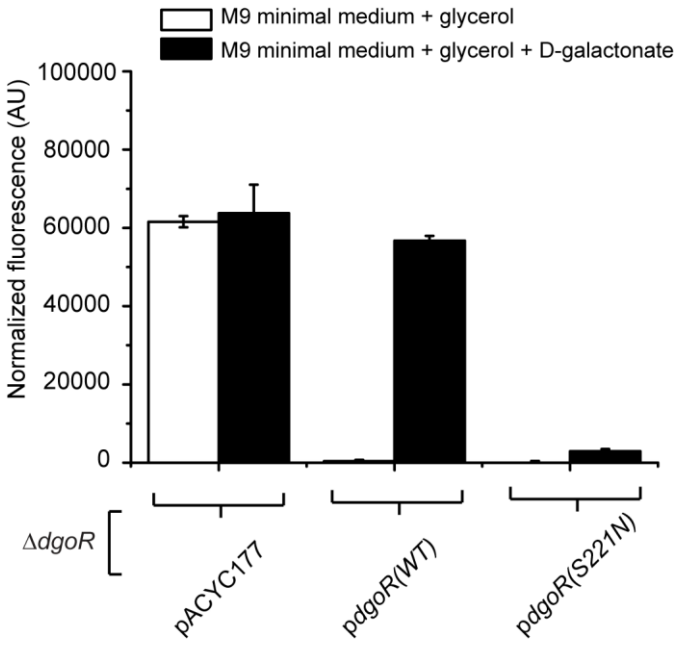

Figure 5.9 Fluorescence reporter assay shows that S221N mutant is defective in derepression of the *dgo* promoter *in vivo*. Plasmids [p*dgoR(WT)*: pBS13, and p*dgoR(S221N)*: pBS24] were individually transformed in a Δ*dgoR* strain carrying fluorescent Venus reporter on the chromosome under the control of *dgo* promoter. Strains were grown in minimal medium supplemented either with glycerol or with glycerol and D-galactonate to exponential phase. Fluorescence was measured and normalized to OD_{450} of the samples. Data represents average ± SD of technical replicates of an individual experiment.

5.3 Discussion

Several FadR subfamily regulators are known to govern metabolism of a variety of carbon sources such as fatty acids, sialic acid, L-lactate and other substrates emerging from central metabolism (Hoskisson and Rigali, 2009; Jain, 2015; Rigali et al., 2002). Besides DgoR that regulates D-galactonate metabolism in *E. coli*, FadR subfamily members also control the metabolism of sugar acids such as D-gluconate, D-glucuronate, D-galacturonate, and D-glucarate (Mandrand-Berthelot et al., 2004). These regulators employ substrate (e.g. *E. coli* NanR: N-acetylneuraminic acid, *B. subtilis* GntR: D-gluconate), metabolic intermediate (e.g. *E. coli* FadR: long-chain fatty acyl-CoA, *Polaromonas* sp. JS666 GguR: 5-keto-4-deoxy-D-glucarate/galactarate) or both (e.g. *R. leguminosarum* MatR: malonate and citrate, *E. coli* UxuR: D-glucuronate and D-fructuronate) as their effectors (Tables 5.1 and 5.2) (Bates Utz et al., 2004; Bouvier et al., 2019; DiRusso et al., 1992; DiRusso et al., 1998; Hoskisson and Rigali, 2009; Jain, 2015; Kalivoda et al., 2003; Lee et al., 2000; Miwa and Fujita, 1988; Tutukina et al., 2016a). From these examples, it is evident that whereas some FadR subfamily members (including ones that regulate sugar acid metabolism) only bind their substrates or intermediates, others bind both.

Table 5.1 Effectors of FadR subfamily members involved in regulating metabolism of various carbon sources

Regulator	Organism	Substrates and intermediates tested (+ and – indicate whether tested compounds serve as effectors or not)		Reference
		Substrate	Intermediate	
FadR	*E. coli*	Long-chain fatty acids (LCFA) (-)	Long-chain fatty acyl-CoA (+)	(DiRusso et al., 1992); (DiRusso et al., 1998)
MatR	*R. leguminosarum*	Malonate (+)	Citrate (+)	(Lee et al., 2000)
NanR	*E. coli*	*N*-acetylneuraminic acid (Neu5Ac) (+)	*N*-acetylmannosamine (ManNAc) (-) and *N*-acetylglucosamine (GlcNAc) (-)	(Kalivoda et al., 2003)

Table 5.2 Effectors of FadR subfamily transcriptional regulators involved in regulating sugar acid metabolism.

Regulator	Organism	Substrates and intermediates tested (+ and – indicate whether tested compounds serve as effectors or not)		Reference
		Substrate	Intermediate	
GntR	*B. subtilis*	Gluconate (+)	6-phosphogluconate (-)	(Miwa and Fujita, 1988)
GguR	*Polaromonas* sp. JS666	D-glucuronate (-)/ D-galacturonate (-)/ D-glucarate (-)/ *meso*-galactarate (-)	5-keto-4-deoxy-D-glucarate/galactarate (KDG) (+)	(Bouvier et al., 2019)
UxuR	*E. coli*	Glucuronate (+)	Fructuronate (+)	(Bates Utz et al., 2004; Tutukina et al., 2016a)

Our work shows that DgoR binds only the substrate of the metabolic pathway that it regulates, i.e., D-galactonate. D-galactonate prevented the binding of DgoR to its *cis*-acting element in a concentration-dependent manner. Other carbohydrates including two sugars D-glucose and D-galactose as well as the sugar acid, D-gluconate, an epimer of D-galactonate differing only in the spatial orientation of one -OH group around the fourth carbon did not affect the binding of DgoR to DNA (Fig. 5.2 and 5.3). Further, none of the D-galactonate catabolic intermediates abrogated the DgoR-operator interaction suggesting that similar to *B. subtilis* GntR, the substrate itself acts as the true effector of DgoR (Fig. 5.4 and 5.5) (Miwa and Fujita, 1988). Our observations are also in line with an earlier study which showed that the enzymatic activities of D-galactonate catabolic enzymes, the kinase (DgoK) and the aldolase (DgoA) are significantly induced in cell extracts prepared from a *dgoD* mutant grown in D-galactonate supplemented medium. Because, the dehydratase, DgoD catalyzes the first step in D-galactonate metabolism and hence no catabolic intermediates are expected to accumulate in a *dgoD* mutant, it was suggested that D-galactonate itself is the inducer of *dgo* operon (Cooper, 1978).

Binding of effectors induces a conformational change in the GntR family regulators including the various FadR subfamily members, which alters the DNA-binding ability of the regulator (reviewed in (Jain, 2015)). In limited proteolysis digestion experiments performed in our lab (Garima Arya, Ph.D. student), we found that binding of D-galactonate also induces a conformational change in DgoR. Compared to DgoR alone or DgoR incubated with D-gluconate, the protein incubated with D-galactonate was less accessible to trypsin digestion and showed a different digestion pattern (Fig. 8C, (Singh et al., 2019)) indicating that D-galactonate binds and induces a conformational change in DgoR.

Since the C-terminal effector-binding E/O domain of FadR subfamily is different from other GntR subfamilies, the comparison of DgoR with respect to its binding affinity with cognate effector can only be made with FadR subfamily members. As shown in the table below (Table 5.3), effector affinity has been determined for only a few FadR subfamily members using different techniques, and the K_D varies from nM to μM range. We determined the affinity of D-galactonate by titrating DgoR-DNA complex with D-galactonate in an EMSA experiment. The apparent K_D was estimated to be 55 ± 6 μM (Fig. 5.6). Among the FadR subfamily members involved in regulating sugar acid metabolism for which effectors are known (*B. subtilis* GntR, *E. coli* UxuR, *E. coli* ExuR, and *Polaromonas* sp. JS666 GguR), information on the affinity with effectors is available only for *B. subtilis* GntR. In the latter case, K_D has been determined by equilibrium dialysis that involved two components: effector and protein. However, in our case, we have determined the apparent K_D of D-galactonate by EMSA where in addition to protein and effector, DNA is also present. Thus, the K_D values for *B. subtilis* GntR- D-gluconate and *E. coli* DgoR- D-galactonate cannot be directly compared. To draw a meaningful comparison of effector affinities amongst FadR subfamily members that regulate sugar

acid metabolism, efforts have to be directed in future studies towards determining the binding parameters of multiple regulators with their cognate effectors using similar techniques.

Table 5.3 Effector binding affinity of FadR subfamily transcriptional regulators

Regulator	Organism	Effector	Dissociation constant (K_D)	Technique used for K_D determination	Reference
CitO	*E. faecalis*	Citrate	1.2 ± 0.2 μM	Isothermal Titration Calorimetry	(Blancato et al., 2016)
FadR	*E. coli*	Long-chain fatty acyl-CoA	5nM – 250 nM	EMSA	(DiRusso et al., 1992)
			45-369 nM	Isothermal Titration Calorimetry	(DiRusso et al., 1998)
GntR	*B. subtilis*	D-gluconate	5.3 μM	Equilibrium dialysis	(Yoshida et al., 1995)

Among the FadR subfamily regulators, *E. coli* FadR, *V. cholerae* FadR and *V. alginolyticus* FadR are the well-characterized members of FadR subfamily till date (Gao et al., 2017; Jain, 2015; Shi et al., 2015; van Aalten et al., 2001; Xu et al., 2001). A random mutagenesis screen identified the amino acid residues critical for interaction of *E. coli* FadR with its effector, fatty acyl-CoA (Raman and DiRusso, 1995). Mutation of one particular S219 amino acid in *E. coli* FadR was characterized in further detail. The transcriptional reporter assay showed that the FadR S219N mutant was defective in induction by fatty acids and behaved like a super-repressor. Further *in vitro* investigations showed that the mutant was unable to release the bound DNA even in the presence of its effector, fatty acyl-CoA (Raman and DiRusso, 1995). In a subsequent study, solved structure of FadR bound to its effector helped in assigning roles to the amino acids identified in the screen as either directly interacting with fatty acyl-CoA, or playing an important role in conducting the allosteric change required to release the bound DNA (van Aalten et al., 2001; Xu et al., 2001). Structure showed that the S219 residue directly interacts with the pantothenic moiety of fatty acyl-CoA and thus plays a critical role in

the effector mediated release of DNA bound by FadR. Along similar lines, solved structures of the effector bound FadR homologues in *V. cholerae* and *V. alginolyticus* showed that although they harbor additional effector binding sites, the known effector-regulator interactions in *E. coli* FadR are conserved in *Vibrio* FadR proteins including the interaction equivalent of S219 in *E. coli* FadR (Gao et al., 2017; Shi et al., 2015).

Structure based sequence alignment shows that DgoR and FadR share the homologous serine residue in their E/O domain (Fig. 5.7) (Rigali et al., 2002). We asked whether the corresponding S221 amino acid is important for functioning of DgoR. Similar to FadR, transcriptional reporter assay shows that the DgoR mutated for the serine (S221N mutant) is defective in induction by D-galactonate and hence behaves as a super-repressor, suggesting that the serine residue plays an important role in the functioning of DgoR (Fig. 5.8 and 5.9). There can be three mechanistic explanations for the super-repressor phenotype of the S221N mutant: 1) the mutant has increased affinity for the operator DNA, 2) the mutant cannot bind D-galactonate altogether or binds very weakly, and/or 3) the mutant can bind D-galactonate but is not conducive to conformational change necessary for releasing DNA. Future experiments involving comparison of the affinity of the mutant with operator and D-galactonate will be required to understand the mechanistic basis of the super-repressor phenotype of the S221 mutant. In the ongoing experiments in our lab, using random mutagenesis based genetic screen and computational docking, we are trying to identify the D-galactonate binding pocket in DgoR. The genetic screen followed by complementation and fluorescence reporter assay identified S221L mutant as a super-repressor. Further, computational docking of modeled DgoR with D-galactonate predicted that S221 amino acid could be involved in direct interaction with D-galactonate (Garima Arya *et al.*, unpublished data). Structure based sequence alignment shows that the concerned serine is not conserved amongst the FadR

subfamily members involved in regulating metabolism of other sugar acids. However, transcriptional reporter assays with *B. subtilis* GntR where the positional equivalent valine was mutated to leucine/phenylalanine (V220L/F) shows that this mutant also behaves as a super-repressor, suggesting that the size rather than the nature of the amino acid is important at this position. Further, similar to DgoR, computational docking of GntR with D-gluconate indicates the direct interaction of the valine residue with the effector (Garima Arya *et al.*, unpublished data).

Given the physiological importance of sugar acid metabolism and the fact that there are only a limited number of studies on the interaction of FadR transcriptional regulators involved in regulation of sugar acid metabolism with their cognate effectors, in future focus should be channeled on determining the molecular details of their interaction with the corresponding effectors. Till date, there are no solved effector bound structures of any of the FadR subfamily regulators involved in regulation of sugar acid metabolism. Solving the structures of regulators in their apo-, DNA- or effector- bound states will enable a deeper understanding of how the binding of relevant effectors modulates the DNA binding ability of the FadR transcriptional regulators dedicated to regulating sugar acid metabolism in bacteria. Such studies on investigating the structural and molecular details of the interaction of DgoR with D-galactonate can lead the way for other FadR subfamily members that regulate sugar acid metabolism.

CHAPTER VI

Summary and Future Prospects

Summary and future prospects

In the present study, we investigated interactions between the *dgo* operon repressor, DgoR with its operator and effector. The combined results provide evidence that DgoR represses the D-galactonate metabolic pathway by binding two closely spaced inverted repeats in the *dgo cis*-acting element and employs a derepression mechanism using D-galactonate as a specific effector molecule. Our findings that DgoR is a majorly α-helical protein with GntR-like N-terminal wHTH domain and a predicted all helical C-terminal FCD domain common to FadR subfamily members, binds [5'-TTGTA(G/C)TACA(A/T)-3'] operator sequence matching the signature of GntR family members that recognize inverted repeats [5'-(N)yGT(N)xAC(N)y-3'], and shares conserved as well as semi-conserved protein-DNA contacts of the GntR family, strongly place DgoR in the FadR subfamily within the GntR family of transcriptional regulators. Importantly, our work showed that the recently identified *dgoR* missense mutations in *E. coli*, adapted to the mammalian gut, abrogate the interaction of DgoR with its operator DNA, asserting that mutations in the *dgo*- regulatory elements are selected in the host to allow simultaneous induction of *dgo* genes.

Although we identified the DgoR binding site in the *dgo cis*-acting element, the binding-mode of DgoR to the two inverted repeats is still elusive. Determining whether DgoR binds the two inverted repeats as two individual dimers or as a dimer of dimers, will be important for contributing to the limited information available for binding of GntR family regulators to their cognate operator DNA. While the third inverted repeat in the *dgo cis*-acting element was not found to be critical for regulation by DgoR, based on previous reports regarding complex regulation of other operons such as *lac* and *gal* involving multiple operators, the importance of the third inverted repeat cannot be ruled out and it may be involved in DNA looping or binding to yet unidentified protein for

regulation of *dgo* operon in a different context. From our current work, and a recently published report we gather that the *dgo* promoter is under dual regulation where besides negative regulation by DgoR, it is positively regulated by cAMP-CRP. The published report has investigated the role of only one of the two predicted cAMP-CRP binding sites in the *dgo cis*-acting element. Probing the role of the third inverted repeat and the additional putative cAMP-CRP binding site may reveal additional layers of regulation on the expression and role of *dgo* genes.

Till date, in-depth details and underlying mechanism for effector mediated regulation of sugar acid metabolism by FadR subfamily members such as ExuR, UxuR, and LgoR from *E. coli*; GguR from *Polaromonas* and *Ralstonia* sp. as well as GntR from *B. subtilis* have not been explored. In an ongoing project in our lab, we are investigating the molecular details of effector-binding properties of DgoR which can serve as a template for building and comparing similar information from other sugar acid metabolism regulators belonging to FadR subfamily. The *in vivo* study where *E. coli* adapted to the mammalian gut was reported to accumulate mutations in *dgoR* and the report on the comparison of a panel of 340 natural isolates of *E. coli* which shows that there are genetic variations in *dgoR* combined together suggest that differential regulation of *dgo* operon might play an important role in the adaptation of *E. coli* isolates to their diverse natural habitats. Thus in future it will be interesting to investigate the effect of natural genetic variations in *dgoR* on the colonization of biological niches by these bacteria. Further, multiple genome-level studies from past couple of decades have implicated the role of *dgo* genes in the interaction of enteric bacteria with their hosts. The fundamental molecular and functional insights regarding the interaction of DgoR with its cognate DNA gained from the work presented in this thesis and the identification of D-galactonate as an effector of DgoR from this study coupled with the ongoing project on the investigation of DgoR-

galactonate interaction sets the stage for understanding the regulation of D-galactonate metabolism in additional enterobacterial strains and its potential role in mediating host-bacterial interactions.

CHAPTER VII

Bibliography

A.H., S. (1961). Glucose and galactose metabolism in *Gluconobacter liquefaciens*. Biochim Biophys Acta *48*, 484-500.

Agrawal, S., Jaswal, K., Shiver, A.L., Balecha, H., Patra, T., and Chaba, R. (2017). A genome-wide screen in *Escherichia coli* reveals that ubiquinone is a key antioxidant for metabolism of long-chain fatty acids. J Biol Chem *292*, 20086-20099.

Aguilera, L., Campos, E., Gimenez, R., Badia, J., Aguilar, J., and Baldoma, L. (2008). Dual role of LldR in regulation of the *lldPRD* operon, involved in L-lactate metabolism in *Escherichia coli*. J Bacteriol *190*, 2997-3005.

Allison, S.L., and Phillips, A.T. (1990). Nucleotide sequence of the gene encoding the repressor for the histidine utilization genes of *Pseudomonas putida*. J Bacteriol *172*, 5470-5476.

An, S.Q., Lu, G.T., Su, H.Z., Li, R.F., He, Y.Q., Jiang, B.L., Tang, D.J., and Tang, J.L. (2011). Systematic mutagenesis of all predicted *gntR* genes in *Xanthomonas campestris* pv. Campestris reveals a GntR family transcriptional regulator controlling hypersensitive response and virulence. Mol Plant Microbe Interact *24*, 1027-1039.

Apperloo-Renkema, H.Z., Van der Waaij, B.D., and Van der Waaij, D. (1990). Determination of colonization resistance of the digestive tract by biotyping of Enterobacteriaceae. Epidemiol Infect *105*, 355-361.

Arai, H., Akahira, S., Ohishi, T., and Kudo, T. (1999). Adaptation of *Comamonas testosteroni* TA441 to utilization of phenol by spontaneous mutation of the gene for a *trans*-acting factor. Mol Microbiol *33*, 1132-1140.

Aravind, L., and Anantharaman, V. (2003). HutC/FarR-like bacterial transcription factors of the GntR family contain a small molecule-binding domain of the chorismate lyase fold. FEMS Microbiol Lett *222*, 17-23.

Aravind, L., Anantharaman, V., Balaji, S., Babu, M.M., and Iyer, L.M. (2005). The many faces of the helix-turn-helix domain: transcription regulation and beyond. FEMS Microbiol Rev *29*, 231-262.

Arias, A., and Cervenansky, C. (1986). Galactose metabolism in *Rhizobium meliloti* L5-30. J Bacteriol *167*, 1092-1094.

Arya, Y.S. (2014). Investigating role of DgoR in enterobacteriaceae family (Thesis submitted to Indian Institute of Science Education and Research Mohali, India, for BS-MS dual degree).

Ashwell, A., Wahba, A.J., and Hickman, J. (1958). A new pathway of uronic acid metabolism. Biochim Biophys Acta *30*, 186-187.

Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K.A., Tomita, M., Wanner, B.L., and Mori, H. (2006). Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. Molecular systems biology *2*, 2006 0008.

Babbitt, P.C., Mrachko, G.T., Hasson, M.S., Huisman, G.W., Kolter, R., Ringe, D., Petsko, G.A., Kenyon, G.L., and Gerlt, J.A. (1995). A functionally diverse enzyme superfamily that abstracts the alpha protons of carboxylic acids. Science *267*, 1159-1161.

Ballering, K.S., Kristich, C.J., Grindle, S.M., Oromendia, A., Beattie, D.T., and Dunny, G.M. (2009). Functional genomics of *Enterococcus faecalis*: multiple novel genetic determinants for biofilm formation in the core genome. J Bacteriol *191*, 2806-2814.

Banjara, N., Suhr, M.J., and Hallen-Adams, H.E. (2015). Diversity of yeast and mold species from a variety of cheese types. Curr Microbiol *70*, 792-800.

Baron, F., Bonnassie, S., Alabdeh, M., Cochet, M.F., Nau, F., Guerin-Dubiard, C., Gautier, M., Andrews, S.C., and Jan, S. (2017). Global gene-expression analysis of the response of *Salmonella*

Enteritidis to egg white exposure reveals multiple egg white-imposed stress responses. Front Microbiol *8*, 829.
Bastet, L., Turcotte, P., Wade, J.T., and Lafontaine, D.A. (2018). Maestro of regulation: Riboswitches orchestrate gene expression at the levels of translation, transcription and mRNA decay. RNA Biol *15*, 679-682.
Bates Utz, C., Nguyen, A.B., Smalley, D.J., Anderson, A.B., and Conway, T. (2004). GntP is the *Escherichia coli* fructuronic acid transporter and belongs to the UxuR regulon. J Bacteriol *186*, 7690-7696.
Belliveau, N.M., Barnes, S.L., Ireland, W.T., Jones, D.L., Sweredoski, M.J., Moradian, A., Hess, S., Kinney, J.B., and Phillips, R. (2018). Systematic approach for dissecting the molecular mechanisms of transcriptional regulation in bacteria. Proc Natl Acad Sci U S A *115*, E4796-E4805.
Berry, G.T., Wehrli, S., Reynolds, R., Palmieri, M., Frangos, M., Williamson, J.R., and Segal, S. (1998). Elevation of erythrocyte redox potential linked to galactonate biosynthesis: elimination by Tolrestat. Metabolism: clinical and experimental *47*, 1423-1428.
Bhagavan, N.V. (2002). Chapter 9 - Simple carbohydrates. Medical Biochemistry (Fourth Edition), Academic Press, Science direct, 133-151.
Bisterfeld, C., Classen, T., Kuberl, I., Henssen, B., Metz, A., Gohlke, H., and Pietruszka, J. (2016). Redesigning aldolase stereoselectivity by homologous grafting. PLoS One *11*, e0156525.
Blancato, V.S., Pagliai, F.A., Magni, C., Gonzalez, C.F., and Lorca, G.L. (2016). Functional analysis of the citrate activator CitO from *Enterococcus faecalis* implicates a divalent metal in ligand binding. Front Microbiol *7*, 101.
Blancato, V.S., Repizo, G.D., Suarez, C.A., and Magni, C. (2008). Transcriptional regulation of the citrate gene cluster of *Enterococcus faecalis* involves the GntR family transcriptional activator CitO. J Bacteriol *190*, 7419-7430.
Bock K., and C., P. (1983). Carbon-13 Nuclear Magnetic Resonance spectroscopy of monosaccharides. Advances in Carbohydrate Chemistry and Biochemistry *41*, 27-66.
Boulanger, A., Dejean, G., Lautier, M., Glories, M., Zischek, C., Arlat, M., and Lauber, E. (2010). Identification and regulation of the N-acetylglucosamine utilization pathway of the plant pathogenic bacterium *Xanthomonas campestris* pv. Campestris. J Bacteriol *192*, 1487-1497.
Bouvier, J.T., Sernova, N.V., Ghasempur, S., Rodionova, I.A., Vetting, M.W., Al-Obaidi, N.F., Almo, S.C., Gerlt, J.A., and Rodionov, D.A. (2019). Novel metabolic pathways and regulons for hexuronate utilization in Proteobacteria. J Bacteriol *201*.
Brechtel, E., Huwig, A., and Giffhorn, F. (2002). L-glucitol catabolism in *Stenotrophomonas maltophilia* Ac. Appl Environ Microbiol *68*, 582-587.
Burland, V., Plunkett, G., 3rd, Daniels, D.L., and Blattner, F.R. (1993). DNA sequence and analysis of 136 kilobases of the *Escherichia coli* genome: organizational symmetry around the origin of replication. Genomics *16*, 551-561.
Campbell, J.W., and Cronan, J.E., Jr. (2001). *Escherichia coli* FadR positively regulates transcription of the *fabB* fatty acid biosynthetic gene. J Bacteriol *183*, 5982-5990.
Caron, M.P., Bastet, L., Lussier, A., Simoneau-Roy, M., Masse, E., and Lafontaine, D.A. (2012). Dual-acting riboswitch control of translation initiation and mRNA decay. Proc Natl Acad Sci U S A *109*, E3444-3453.
Casali, N., White, A.M., and Riley, L.W. (2006). Regulation of the *Mycobacterium tuberculosis mce1* operon. J Bacteriol *188*, 441-449.
Cenens, W., Mebrhatu, M.T., Makumi, A., Ceyssens, P.J., Lavigne, R., Van Houdt, R., Taddei, F., and Aertsen, A. (2013). Expression of a novel P22 ORFan gene reveals the phage carrier state in *Salmonella* Typhimurium. PLoS genetics *9*, e1003269.
Chaba, R., Grigorova, I.L., Flynn, J.M., Baker, T.A., and Gross, C.A. (2007). Design principles of the proteolytic cascade governing the sigmaE-mediated envelope stress response in *Escherichia coli*: keys to graded, buffered, and rapid signal transduction. Genes Dev *21*, 124-136.
Chan, V., Dreolini, L.F., Flintoff, K.A., Lloyd, S.J., and Mattenley, A.A. (2002). The effects of glycerol, glucose, galactose, lactose and glucose with galactose on the induction of β-galactosidase in *Escherichia coli*. J Exp Microbiol & Immunol *2*, 130-137.

Chauvier, A., Picard-Jean, F., Berger-Dancause, J.C., Bastet, L., Naghdi, M.R., Dube, A., Turcotte, P., Perreault, J., and Lafontaine, D.A. (2017). Transcriptional pausing at the translation start site operates as a critical checkpoint for riboswitch regulation. Nature communications *8*, 13892.
Clark, D.P., and Cronan, J.E. (2005). Two-carbon compounds and fatty acids as carbon sources. EcoSal Plus *1*.
Collins, J., Robinson, C., Danhof, H., Knetsch, C.W., van Leeuwen, H.C., Lawley, T.D., Auchtung, J.M., and Britton, R.A. (2018). Dietary trehalose enhances virulence of epidemic *Clostridium difficile*. Nature *553*, 291-294.
Condemine, G., Berrier, C., Plumbridge, J., and Ghazi, A. (2005). Function and expression of an N-acetylneuraminic acid-inducible outer membrane channel in *Escherichia coli*. J Bacteriol *187*, 1959-1965.
Conway, T., and Cohen, P.S. (2015). Commensal and pathogenic *Escherichia coli* metabolism in the gut. Microbiol Spectr *3*.
Cooper, R.A. (1978). The utilisation of D-galactonate and D-2-oxo-3-deoxygalactonate by *Escherichia coli* K-12. Biochemical and genetical studies. Arch Microbiol *118*, 199-206.
Dame, R.T., Wyman, C., and Goosen, N. (2000). H-NS mediated compaction of DNA visualised by atomic force microscopy. Nucleic Acids Res *28*, 3504-3510.
Datsenko, K.A., and Wanner, B.L. (2000). One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. Proc Natl Acad Sci U S A *97*, 6640-6645.
de Diego Puente, T., Gallego-Jara, J., Castano-Cerezo, S., Bernal Sanchez, V., Fernandez Espin, V., Garcia de la Torre, J., Manjon Rubio, A., and Canovas Diaz, M. (2015). The protein acetyltransferase PatZ from *Escherichia coli* is regulated by autoacetylation-induced oligomerization. J Biol Chem *290*, 23077-23093.
De Ley, J., and Doudoroff, M. (1957). The metabolism of D-galactose in *Pseudomonas saccharophila*. J Biol Chem *227*, 745-757.
Deacon, J., and Cooper, R.A. (1977). D-Galactonate utilisation by enteric bacteria. The catabolic pathway in *Escherichia coli*. FEBS letters *77*, 201-205.
del Peso-Santos, T., Bartolome-Martin, D., Fernandez, C., Alonso, S., Garcia, J.L., Diaz, E., Shingler, V., and Perera, J. (2006). Coregulation by phenylacetyl-coenzyme A-responsive PaaX integrates control of the upper and lower pathways for catabolism of styrene by *Pseudomonas* sp. strain Y2. J Bacteriol *188*, 4812-4821.
DiRusso, C.C., Heimert, T.L., and Metzger, A.K. (1992). Characterization of FadR, a global transcriptional regulator of fatty acid metabolism in *Escherichia coli*. Interaction with the *fadB* promoter is prevented by long chain fatty acyl coenzyme A. J Biol Chem *267*, 8685-8691.
DiRusso, C.C., Tsvetnitsky, V., Hojrup, P., and Knudsen, J. (1998). Fatty acyl-CoA binding domain of the transcription factor FadR. Characterization by deletion, affinity labeling, and isothermal titration calorimetry. J Biol Chem *273*, 33652-33659.
Dong, J.M., Taylor, J.S., Latour, D.J., Iuchi, S., and Lin, E.C. (1993). Three overlapping *lct* genes involved in L-lactate utilization by *Escherichia coli*. J Bacteriol *175*, 6671-6678.
Dulermo, R., Onodera, T., Coste, G., Passot, F., Dutertre, M., Porteron, M., Confalonieri, F., Sommer, S., and Pasternak, C. (2015). Identification of new genes contributing to the extreme radioresistance of *Deinococcus radiodurans* using a Tn5-based transposon mutant library. PLoS One *10*, e0124358.
Edayathumangalam, R., Wu, R., Garcia, R., Wang, Y., Wang, W., Kreinbring, C.A., Bach, A., Liao, J., Stone, T.A., Terwilliger, T.C., *et al.* (2013). Crystal structure of *Bacillus subtilis* GabR, an autorepressor and transcriptional activator of *gabT*. Proc Natl Acad Sci U S A *110*, 17820-17825.
El-Gebali, S., Mistry, J., Bateman, A., Eddy, S.R., Luciani, A., Potter, S.C., Qureshi, M., Richardson, L.J., Salazar, G.A., Smart, A., *et al.* (2019). The Pfam protein families database in 2019. Nucleic Acids Res *47*, D427-D432.
Elshafei, A.M., and Abdel-Fatah, O.M. (2001). Factors affecting D-galactonate degradation by extracts of *Aspergillus niger*. Journal of basic microbiology *41*, 149-158.

Eriksson, S., Lucchini, S., Thompson, A., Rhen, M., and Hinton, J.C. (2003). Unravelling the biology of macrophage infection by gene expression profiling of intracellular *Salmonella enterica*. Mol Microbiol *47*, 103-118.
Erol, I., Jeong, K.C., Baumler, D.J., Vykhodets, B., Choi, S.H., and Kaspar, C.W. (2006). H-NS controls metabolism and stress tolerance in *Escherichia coli* O157:H7 that influence mouse passage. BMC Microbiol *6*, 72.
Faber, F., Tran, L., Byndloss, M.X., Lopez, C.A., Velazquez, E.M., Kerrinnes, T., Nuccio, S.P., Wangdi, T., Fiehn, O., Tsolis, R.M., *et al.* (2016). Host-mediated sugar oxidation promotes post-antibiotic pathogen expansion. Nature *534*, 697-699.
Feng, Y., Zhang, H., and Cronan, J.E. (2013). Profligate biotin synthesis in alpha-proteobacteria - a developing or degenerating regulatory system? Mol Microbiol *88*, 77-92.
Ficicioglu, C., Yager, C., and Segal, S. (2005). Galactitol and galactonate in red blood cells of children with the Duarte/galactosemia genotype. Molecular genetics and metabolism *84*, 152-159.
Fillenberg, S.B., Friess, M.D., Korner, S., Bockmann, R.A., and Muller, Y.A. (2016). Crystal structures of the global regulator DasR from *Streptomyces coelicolor*: Implications for the allosteric regulation of GntR/HutC repressors. PLoS One *11*, e0157691.
Fillenberg, S.B., Grau, F.C., Seidel, G., and Muller, Y.A. (2015). Structural insight into operator dre-sites recognition and effector binding in the GntR/HutC transcription regulator NagR. Nucleic Acids Res *43*, 1283-1296.
Franco, I.S., Mota, L.J., Soares, C.M., and de Sa-Nogueira, I. (2006). Functional domains of the *Bacillus subtilis* transcription factor AraR and identification of amino acids important for nucleoprotein complex assembly and effector binding. J Bacteriol *188*, 3024-3036.
Franco, I.S., Mota, L.J., Soares, C.M., and de Sa-Nogueira, I. (2007). Probing key DNA contacts in AraR-mediated transcriptional repression of the *Bacillus subtilis* arabinose regulon. Nucleic Acids Res *35*, 4755-4766.
Franklin, M.C., Cheung, J., Rudolph, M.J., Burshteyn, F., Cassidy, M., Gary, E., Hillerich, B., Yao, Z.K., Carlier, P.R., Totrov, M., *et al.* (2015). Structural genomics for drug design against the pathogen *Coxiella burnetii*. Proteins *83*, 2124-2136.
Freter, R. (1983). Mechanisms that control the microflora in the large intetstine. Human intestinal microflora in health and disease New York, NY: Academic Press, Inc, 33-54.
Freter, R. (1988). Mechanisms of bacterial colonization of the mucosal surfaces of the gut. Virulence mechanisms of bacterial pathogens Washington, DC: American Society for Microbiology, 45-60.
Freter, R., Brickner, H., Fekete, J., Vickerman, M.M., and Carey, K.E. (1983). Survival and implantation of *Escherichia coli* in the intestinal tract. Infect Immun *39*, 686-703.
Fuchs, T.M., Eisenreich, W., Heesemann, J., and Goebel, W. (2012). Metabolic adaptation of human pathogenic and related nonpathogenic bacteria to extra- and intracellular habitats. FEMS Microbiol Rev *36*, 435-462.
Galardini, M., Koumoutsi, A., Herrera-Dominguez, L., Cordero Varela, J.A., Telzerow, A., Wagih, O., Wartel, M., Clermont, O., Denamur, E., Typas, A., *et al.* (2017). Phenotype inference in an *Escherichia coli* strain panel. Elife *6*.
Gao, R., Li, D., Lin, Y., Lin, J., Xia, X., Wang, H., Bi, L., Zhu, J., Hassan, B., Wang, S., *et al.* (2017). Structural and functional characterization of the FadR regulatory protein from *Vibrio alginolyticus*. Front Cell Infect Microbiol *7*, 513.
Gao, Y., Yurkovich, J.T., Seo, S.W., Kabimoldayev, I., Drager, A., Chen, K., Sastry, A.V., Fang, X., Mih, N., Yang, L., *et al.* (2018). Systematic discovery of uncharacterized transcription factors in *Escherichia coli* K-12 MG1655. Nucleic Acids Res *46*, 10682-10696.
Gao, Y.G., Suzuki, H., Itou, H., Zhou, Y., Tanaka, Y., Wachi, M., Watanabe, N., Tanaka, I., and Yao, M. (2008). Structural and functional characterization of the LldR from *Corynebacterium glutamicum*: a transcriptional repressor involved in L-lactate and sugar utilization. Nucleic Acids Res *36*, 7110-7123.
Gao, Y.G., Yao, M., Itou, H., Zhou, Y., and Tanaka, I. (2007). The structures of transcription factor CGL2947 from *Corynebacterium glutamicum* in two crystal forms: a novel homodimer assembling and the implication for effector-binding mode. Protein Sci *16*, 1878-1886.

Geddes, B.A., and Oresnik, I.J. (2012). Inability to catabolize galactose leads to increased ability to compete for nodule occupancy in *Sinorhizobium meliloti*. J Bacteriol *194*, 5044-5053.
Gerlt, J.A., Babbitt, P.C., and Rayment, I. (2005). Divergent evolution in the enolase superfamily: the interplay of mechanism and specificity. Arch Biochem Biophys *433*, 59-70.
Gomila, M., Bowien, B., Falsen, E., Moore, E.R., and Lalucat, J. (2007). Description of *Pelomonas aquatica* sp. nov. and *Pelomonas puraquae* sp. nov., isolated from industrial and haemodialysis water. Int J Syst Evol Microbiol *57*, 2629-2635.
Gorelik, M., Lunin, V.V., Skarina, T., and Savchenko, A. (2006). Structural characterization of GntR/HutC family signaling domain. Protein Sci *15*, 1506-1511.
Görke, B., and Stülke, J. (2008). Carbon catabolite repression in bacteria: many ways to make the most out of nutrients. Nat Rev Microbiol *6*, 613-624.
Goudeau, D.M., Parker, C.T., Zhou, Y., Sela, S., Kroupitski, Y., and Brandl, M.T. (2013). The *Salmonella* transcriptome in lettuce and cilantro soft rot reveals a niche overlap with the animal host intestine. Appl Environ Microbiol *79*, 250-262.
Gupta, A., Singh, V.K., Qazi, G.N., and Kumar, A. (2001). *Gluconobacter oxydans*: its biotechnological applications. J Mol Microbiol Biotechnol *3*, 445-456.
Haine, V., Sinon, A., Van Steen, F., Rousseau, S., Dozot, M., Lestrate, P., Lambert, C., Letesson, J.J., and De Bolle, X. (2005). Systematic targeted mutagenesis of *Brucella melitensis* 16M reveals a major role for GntR regulators in the control of virulence. Infect Immun *73*, 5578-5586.
Haldimann, A., and Wanner, B.L. (2001). Conditional-replication, integration, excision, and retrieval plasmid-host systems for gene structure-function studies of bacteria. J Bacteriol *183*, 6384-6393.
Hales, L.M., Gumport, R.I., and Gardner, J.F. (1996). Examining the contribution of a dA+dT element to the conformation of *Escherichia coli* integration host factor-DNA complexes. Nucleic Acids Res *24*, 1780-1786.
Haydon, D.J., and Guest, J.R. (1991). A new family of bacterial regulatory proteins. FEMS Microbiol Lett *63*, 291-295.
Haydon, D.J., Quail, M.A., and Guest, J.R. (1993). A mutation causing constitutive synthesis of the pyruvate dehydrogenase complex in *Escherichia coli* is located within the *pdhR* gene. FEBS letters *336*, 43-47.
Hernandez-Arriaga, A.M., Rubio-Lepe, T.S., Espinosa, M., and del Solar, G. (2009). Repressor CopG prevents access of RNA polymerase to promoter and actively dissociates open complexes. Nucleic Acids Res *37*, 4799-4811.
Hogema, B.M., Arents, J.C., Inada, T., Aiba, H., van Dam, K., and Postma, P.W. (1997). Catabolite repression by glucose 6-phosphate, gluconate and lactose in *Escherichia coli*. Mol Microbiol *24*, 857-867.
Holm, L., and Laakso, L.M. (2016). Dali server update. Nucleic Acids Res *44*, W351-355.
Hommes, R.W., Postma, P.W., Tempest, D.W., and Neijssel, O.M. (1989). The influence of the culture pH value on the direct glucose oxidative pathway in *Klebsiella pneumoniae* NCTC 418. Arch Microbiol *151*, 261-267.
Hoskisson, P.A., and Rigali, S. (2009). Chapter 1: Variation in Form and Function: The Helix-Turn-Helix Regulators of the GntR Superfamily. Adv Appl Microbiol *69*, 1-22.
Hoskisson, P.A., Rigali, S., Fowler, K., Findlay, K.C., and Buttner, M.J. (2006). DevA, a GntR-like transcriptional regulator required for development in *Streptomyces coelicolor*. J Bacteriol *188*, 5014-5023.
Hove-Jensen, B., Rosenkrantz, T.J., Zechel, D.L., and Willemoes, M. (2010). Accumulation of intermediates of the carbon-phosphorus lyase pathway for phosphonate degradation in *phn* mutants of *Escherichia coli*. J Bacteriol *192*, 370-374.
Hu, J., Zhang, H., Zhou, S., Li, W., and He, Z.G. (2018). Characterization of a novel regulatory pathway for mannitol metabolism and its coordination with biofilm formation in *Mycobacterium smegmatis*. J Genet Genomics *45*, 477-488.
Hu, J., Zhao, L., and Yang, M. (2015). A GntR family transcription factor positively regulates mycobacterial isoniazid resistance by controlling the expression of a putative permease. BMC Microbiol *15*, 214.

Huang, X., and Miller, W. (1991). A time-effecient, linear-space local similarity algorithm. Adv Appl Math *12*, 337-357.
Isbell, H.S., and Frush, H.L. (1933). Preparation and properties of aldonic acids and their lactones and basic calcium salts. BS Jour Research *11*, 649-664.
Ishihama, A. (2009). The nucleoid: an overview. EcoSal Plus *3*.
Ishihama, A., Shimada, T., and Yamazaki, Y. (2016). Transcription profile of *Escherichia coli*: genomic SELEX search for regulatory targets of transcription factors. Nucleic Acids Res *44*, 2058-2074.
Ishizuka, H., Hanamura, A., Kunimura, T., and Aiba, H. (1993). A lowered concentration of cAMP receptor protein caused by glucose is an important determinant for catabolite repression in *Escherichia coli*. Mol Microbiol *10*, 341-350.
J.G., L. (1903). Experiments on the transformation and fixation of nitrogen by bacteria. New Jersey State Agric Exp Sta Ann Rep *24*, 217-285.
Jain, D. (2015). Allosteric control of transcription in GntR family of transcription regulators: A structural overview. IUBMB Life *67*, 556-563.
Jain, D., and Nair, D.T. (2013). Spacing between core recognition motifs determines relative orientation of AraR monomers on bipartite operators. Nucleic Acids Res *41*, 639-647.
Jain, D., Narayanan, N., and Nair, D.T. (2016). Plasticity in repressor-DNA interactions neutralizes loss of symmetry in bipartite operators. J Biol Chem *291*, 1235-1242.
Jaques, S., and McCarter, L.L. (2006). Three new regulators of swarming in *Vibrio parahaemolyticus*. J Bacteriol *188*, 2625-2635.
Jayani, R.S., Saxena, S., and Gupta, R. (2005). Microbial pectinolytic enzymes: A review. Process Biochemistry *40*, 2931-2944.
Kabsch, W., and Sander, C. (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers *22*, 2577-2637.
Kalivoda, K.A., Steenbergen, S.M., and Vimr, E.R. (2013). Control of the *Escherichia coli* sialoregulon by transcriptional repressor NanR. J Bacteriol *195*, 4689-4701.
Kalivoda, K.A., Steenbergen, S.M., Vimr, E.R., and Plumbridge, J. (2003). Regulation of sialic acid catabolism by the DNA binding protein NanR in *Escherichia coli*. J Bacteriol *185*, 4806-4815.
Kataoka, M., Tanaka, T., Kohno, T., and Kajiyama, Y. (2008). The carboxyl-terminal domain of TraR, a *Streptomyces* HutC family repressor, functions in oligomerization. J Bacteriol *190*, 7164-7169.
Kendall, K.J., and Cohen, S.N. (1988). Complete nucleotide sequence of the *Streptomyces lividans* plasmid pIJ101 and correlation of the sequence with genetic properties. J Bacteriol *170*, 4634-4651.
Keseler, I.M., Mackie, A., Santos-Zavaleta, A., Billington, R., Bonavides-Martinez, C., Caspi, R., Fulcher, C., Gama-Castro, S., Kothari, A., Krummenacker, M., *et al.* (2017). The EcoCyc database: reflecting new knowledge about *Escherichia coli* K-12. Nucleic Acids Res *45*, D543-D550.
Kim, D.E., Chivian, D., and Baker, D. (2004). Protein structure prediction and analysis using the Robetta server. Nucleic Acids Res *32*, W526-531.
Klose, D.P., Wallace, B.A., and Janes, R.W. (2010). 2Struc: the secondary structure server. Bioinformatics *26*, 2624-2625.
Kolb, A., Busby, S., Buc, H., Garges, S., and Adhya, S. (1993). Transcriptional regulation by cAMP and its receptor protein. Annu Rev Biochem *62*, 749-795.
Kreth, J., Lengeler, J.W., and Jahreis, K. (2013). Characterization of pyruvate uptake in *Escherichia coli* K-12. PLoS One *8*, e67125.
Krisko, A., Copic, T., Gabaldon, T., Lehner, B., and Supek, F. (2014). Inferring gene function from evolutionary change in signatures of translation efficiency. Genome Biol *15*, R44.
Ku, Y.W., McDonough, S.P., Palaniappan, R.U., Chang, C.F., and Chang, Y.F. (2005). Novel attenuated *Salmonella enterica* serovar Choleraesuis strains as live vaccine candidates generated by signature-tagged mutagenesis. Infect Immun *73*, 8194-8203.

Kuivanen, J., Dantas, H., Mojzita, D., Mallmann, E., Biz, A., Krieger, N., Mitchell, D., and Richard, P. (2014). Conversion of orange peel to L-galactonic acid in a consolidated process using engineered strains of *Aspergillus niger*. AMB Express *4*, 33.
Kuivanen, J., Mojzita, D., Wang, Y., Hilditch, S., Penttila, M., Richard, P., and Wiebe, M.G. (2012). Engineering filamentous fungi for conversion of D-galacturonic acid to L-galactonic acid. Appl Environ Microbiol *78*, 8676-8683.
Kunze, M., Zerlin, K.F., Retzlaff, A., Pohl, J.O., Schmidt, E., Janssen, D.B., Vilchez-Vargas, R., Pieper, D.H., and Reineke, W. (2009). Degradation of chloroaromatics by *Pseudomonas putida* GJ31: assembled route for chlorobenzene degradation encoded by clusters on plasmid pKW1 and the chromosome. Microbiology *155*, 4069-4083.
Kurn, N., Contreras, I., and Shapiro, L. (1978). Galactose catabolism in *Caulobacter crescentus*. J Bacteriol *135*, 517-520.
Labella, J.I., Obrebska, A., Espinosa, J., Salinas, P., Forcada-Nadal, A., Tremino, L., Rubio, V., and Contreras, A. (2016). Expanding the cyanobacterial nitrogen regulatory network: The GntR-like regulator PlmA interacts with the PII-PipX complex. Front Microbiol *7*, 1677.
Lai, K., and Klapa, M.I. (2004). Alternative pathways of galactose assimilation: could inverse metabolic engineering provide an alternative to galactosemic patients? Metab Eng *6*, 239-244.
Lambrecht, S.J., Wahlig, J.M.L., and Steglich, C. (2018). The GntR family transcriptional regulator PMM1637 regulates the highly conserved cyanobacterial sRNA Yfr2 in marine picocyanobacteria. DNA Res *25*, 489-497.
Lavire, C., Louis, D., Perriere, G., Briolay, J., Normand, P., and Cournoyer, B. (2001). Analysis of pFQ31, a 8551-bp cryptic plasmid from the symbiotic nitrogen-fixing actinomycete *Frankia*. FEMS Microbiol Lett *197*, 111-116.
Leano, J.B., Batarni, S., Eriksen, J., Juge, N., Pak, J.E., Kimura-Someya, T., Robles-Colmenares, Y., Moriyama, Y., Stroud, R.M., and Edwards, R.H. (2019). Structures suggest a mechanism for energy coupling by a family of organic anion transporters. PLoS Biol *17*, e3000260.
Lee, H.Y., An, J.H., and Kim, Y.S. (2000). Identification and characterization of a novel transcriptional regulator, MatR, for malonate metabolism in *Rhizobium leguminosarum* bv. Trifolii. Eur J Biochem *267*, 7224-7230.
Lee, H.Y., and Kim, Y.s. (2001). Identification of amino acid residues in the carboxyl terminus required for malonate-responsive transcriptional regulation of MatR in *Rhizobium leguminosarum* bv. Trifolii. BMB reports *34*, 305-309.
Lee, M.H., Scherer, M., Rigali, S., and Golden, J.W. (2003). PlmA, a new member of the GntR family, has plasmid maintenance functions in *Anabaena* sp. strain PCC 7120. J Bacteriol *185*, 4315-4325.
Lemmens, L., Tilleman, L., De Koning, E., Valegard, K., Lindas, A.C., Van Nieuwerburgh, F., Maes, D., and Peeters, E. (2019). YtrASa, a GntR-Family transcription factor, represses two genetic loci encoding membrane proteins in *Sulfolobus acidocaldarius*. Front Microbiol *10*, 2084.
Lescat, M., Launay, A., Ghalayini, M., Magnan, M., Glodt, J., Pintard, C., Dion, S., Denamur, E., and Tenaillon, O. (2016). Using long-term experimental evolution to uncover the patterns and determinants of molecular evolution of an *Escherichia coli* natural isolate in the streptomycin-treated mouse gut. Molecular ecology.
Levanon, S.S., San, K.Y., and Bennett, G.N. (2005). Effect of oxygen on the *Escherichia coli* ArcA and FNR regulation systems and metabolic responses. Biotechnol Bioeng *89*, 556-564.
Lin, J., Zhu, Y., Tang, H., Yan, J., and Luo, L. (2019). Identification of a GntR family regulator BusRTha and its regulatory mechanism in the glycine betaine ABC transport system of *Tetragenococcus halophilus*. Extremophiles *23*, 451-460.
Lipman, J.G. (1903). Experiments of the transformation and fixation of nitrogen by bacteria. New Jersey State Agric Exp Sta Ann Rep *24*, 217-285.
Little, M.S., Pellock, S.J., Walton, W.G., Tripathy, A., and Redinbo, M.R. (2018). Structural basis for the regulation of beta-glucuronidase expression by human gut Enterobacteriaceae. Proc Natl Acad Sci U S A *115*, E152-E161.
Liu, H., Ramos, K.R., Valdehuesa, K.N., Nisola, G.M., Malihan, L.B., Lee, W.K., Park, S.J., and Chung, W.J. (2014a). Metabolic engineering of *Escherichia coli* for biosynthesis of D-galactonate. Bioprocess and biosystems engineering *37*, 383-391.

Liu, H., Valdehuesa, K.N., Ramos, K.R., Nisola, G.M., Lee, W.K., and Chung, W.J. (2014b). L-arabonate and D-galactonate production by expressing a versatile sugar dehydrogenase in metabolically engineered *Escherichia coli*. Bioresource technology *159*, 455-459.
Liu, M.Y., Yang, H., and Romeo, T. (1995). The product of the pleiotropic *Escherichia coli* gene *csrA* modulates glycogen biosynthesis via effects on mRNA stability. J Bacteriol *177*, 2663-2672.
Lord, D.M., Uzgoren Baran, A., Soo, V.W., Wood, T.K., Peti, W., and Page, R. (2014). McbR/YncC: implications for the mechanism of ligand and DNA binding by a bacterial GntR transcriptional regulator involved in biofilm formation. Biochemistry *53*, 7223-7231.
Macek, B., Forchhammer, K., Hardouin, J., Weber-Ban, E., Grangeasse, C., and Mijakovic, I. (2019). Protein post-translational modifications in bacteria. Nat Rev Microbiol *17*, 651-664.
Madan Babu, M., and Teichmann, S.A. (2003). Functional determinants of transcription factors in *Escherichia coli*: protein families and binding sites. Trends Genet *19*, 75-79.
Maki, K., Morita, T., Otaka, H., and Aiba, H. (2010). A minimal base-pairing region of a bacterial small RNA SgrS required for translational repression of *ptsG* mRNA. Mol Microbiol *76*, 782-792.
Maloy, S.R., and Nunn, W.D. (1981). Role of gene fadR in *Escherichia coli* acetate metabolism. J Bacteriol *148*, 83-90.
Mandrand-Berthelot, M.A., Condemine, G., and Hugouvieux-Cotte-Pattat, N. (2004). Catabolism of Hexuronides, Hexuronates, Aldonates, and Aldarates. EcoSal Plus *1*.
Martens-Uzunova, E.S., and Schaap, P.J. (2008). An evolutionary conserved D-galacturonic acid metabolic pathway operates across filamentous fungi capable of pectin degradation. Fungal Genet Biol *45*, 1449-1457.
Metzner, M., Germer, J., and Hengge, R. (2004). Multiple stress signal integration in the regulation of the complex sigma S-dependent *csiD-ygaF-gabDTP* operon in *Escherichia coli*. Mol Microbiol *51*, 799-811.
Michalska, K., Cuff, M.E., Tesar, C., Feldmann, B., and Joachimiak, A. (2011). Structure of 2-oxo-3-deoxygalactonate kinase from *Klebsiella pneumoniae*. Acta crystallographica Section D, Biological crystallography *67*, 678-689.
Micsonai, A., Wien, F., Bulyaki, E., Kun, J., Moussong, E., Lee, Y.H., Goto, Y., Refregiers, M., and Kardos, J. (2018). BeStSel: a web server for accurate protein secondary structure prediction and fold recognition from the circular dichroism spectra. Nucleic Acids Res *46*, W315-W322.
Micsonai, A., Wien, F., Kernya, L., Lee, Y.H., Goto, Y., Refregiers, M., and Kardos, J. (2015). Accurate secondary structure prediction and fold recognition for circular dichroism spectroscopy. Proc Natl Acad Sci U S A *112*, E3095-3103.
Miller, J.H. (1972). Experiments in molecular genetics. Cold Spring Harbon Laboratory, NY.
Minhas, V., Harvey, R.M., McAllister, L.J., Seemann, T., Syme, A.E., Baines, S.L., Paton, J.C., and Trappetti, C. (2019). Capacity to utilize raffinose dictates Pneumococcal disease phenotype. mBio *10*.
Miwa, Y., and Fujita, Y. (1988). Purification and characterization of a repressor for the *Bacillus subtilis gnt* operon. J Biol Chem *263*, 13252-13257.
Mole, B., Habibi, S., Dangl, J.L., and Grant, S.R. (2010). Gluconate metabolism is required for virulence of the soft-rot pathogen *Pectobacterium carotovorum*. Mol Plant Microbe Interact *23*, 1335-1344.
Monk, J.M., Charusanti, P., Aziz, R.K., Lerman, J.A., Premyodhin, N., Orth, J.D., Feist, A.M., and Palsson, B.O. (2013). Genome-scale metabolic reconstructions of multiple *Escherichia coli* strains highlight strain-specific adaptations to nutritional environments. Proc Natl Acad Sci U S A *110*, 20338-20343.
Monnet, C., Loux, V., Gibrat, J.F., Spinnler, E., Barbe, V., Vacherie, B., Gavory, F., Gourbeyre, E., Siguier, P., Chandler, M., *et al.* (2010). The *Arthrobacter arilaitensis* Re117 genome sequence reveals its genetic adaptation to the surface of cheese. PLoS One *5*, e15489.
Mouz, S., Merlin, C., Springael, D., and Toussaint, A. (1999). A GntR-like negative regulator of the biphenyl degradation genes of the transposon Tn4371. Molecular & general genetics : MGG *262*, 790-799.
Münch, R., Hiller, K., Grote, A., Scheer, M., Klein, J., Schobert, M., and Jahn, D. (2005). Virtual Footprint and PRODORIC: an integrative framework for regulon prediction in prokaryotes. Bioinformatics *21*, 4187-4189.

Münger, L.H., Trimigno, A., Picone, G., Freiburghaus, C., Pimentel, G., Burton, K.J., Pralong, F.P., Vionnet, N., Capozzi, F., Badertscher, R., *et al.* (2017). Identification of urinary food intake biomarkers for milk, cheese, and soy-based drink by untargeted GC-MS and NMR in healthy humans. J Proteome Res *16*, 3321-3335.
Nguyen-Vo, T.P., Liang, Y., Sankaranarayanan, M., Seol, E., Chun, A.Y., Ashok, S., Chauhan, A.S., Kim, J.R., and Park, S. (2019). Development of 3-hydroxypropionic-acid-tolerant strain of *Escherichia coli* W and role of minor global regulator *yieP*. Metab Eng *53*, 48-58.
Noar, J.D., and Bruno-Barcena, J.M. (2018). *Azotobacter vinelandii*: the source of 100 years of discoveries and many more to come. Microbiology *164*, 421-436.
Nogales, J., Blanca-Ordonez, H., Olivares, J., and Sanjuan, J. (2013). Conjugal transfer of the *Sinorhizobium meliloti* 1021 symbiotic plasmid is governed through the concerted action of one- and two-component signal transduction regulators. Environ Microbiol *15*, 811-821.
Nongthombam, S. (2015). Investigating the role of *dgo* operon in D-galactonate metabolism and its regulation in *Escherichia coli* (Thesis submitted to Indian Institute of Science Education and Research Mohali, India, for BS-MS dual degree).
Novel, M., and Novel, G. (1976). Regulation of beta-glucuronidase synthesis in *Escherichia coli* K-12: pleiotropic constitutive mutations affecting *uxu* and *uidA* expression. J Bacteriol *127*, 418-432.
Novichkov, P.S., Kazakov, A.E., Ravcheev, D.A., Leyn, S.A., Kovaleva, G.Y., Sutormin, R.A., Kazanov, M.D., Riehl, W., Arkin, A.P., Dubchak, I., *et al.* (2013). RegPrecise 3.0-a resource for genome-scale exploration of transcriptional regulation in bacteria. BMC Genomics *14*, 745.
Okinaka, Y., Yang, C.H., Perna, N.T., and Keen, N.T. (2002). Microarray profiling of *Erwinia chrysanthemi* 3937 genes that are regulated during plant infection. Mol Plant Microbe Interact *15*, 619-629.
Okuda, K., Kato, S., Ito, T., Shiraki, S., Kawase, Y., Goto, M., Kawashima, S., Hemmi, H., Fukada, H., and Yoshimura, T. (2015). Role of the aminotransferase domain in *Bacillus subtilis* GabR, a pyridoxal 5'-phosphate-dependent transcriptional regulator. Mol Microbiol *95*, 245-257.
Pabo, C.O., and Sauer, R.T. (1992). Transcription factors: structural families and principles of DNA recognition. Annu Rev Biochem *61*, 1053-1095.
Pao, S.S., Paulsen, I.T., and Saier, M.H., Jr. (1998). Major facilitator superfamily. Microbiol Mol Biol Rev *62*, 1-34.
Park, S.A., Park, Y.S., and Lee, K.S. (2017). Crystal structure of the C-terminal domain of *Bacillus subtilis* GabR reveals a closed conformation by gamma-aminobutyric acid binding, inducing transcriptional activation. Biochemical and biophysical research communications *487*, 287-291.
Passalacqua, K.D., Charbonneau, M.E., and O'Riordan, M.X. (2016). Bacterial metabolism shapes the host-pathogen interface. Microbiol Spectr *4*.
Peekhaus, N., and Conway, T. (1998). What's for dinner?: Entner-Doudoroff metabolism in *Escherichia coli*. J Bacteriol *180*, 3495-3502.
Pellicer, M.T., Fernandez, C., Badia, J., Aguilar, J., Lin, E.C., and Baldom, L. (1999). Cross-induction of *glc* and *ace* operons of *Escherichia coli* attributable to pathway intersection. Characterization of the *glc* promoter. J Biol Chem *274*, 1745-1752.
Perez-Rueda, E., and Collado-Vides, J. (2000). The repertoire of DNA-binding transcriptional regulators in *Escherichia coli* K-12. Nucleic Acids Res *28*, 1838-1847.
Perreault, M., Bialek, A., Trottier, J., Verreault, M., Caron, P., Milkiewicz, P., and Barbier, O. (2013). Role of glucuronidation for hepatic detoxification and urinary elimination of toxic bile acids during biliary obstruction. PLoS One *8*, e80994.
Pham, N.P., Landaud, S., Lieben, P., Bonnarme, P., and Monnet, C. (2019). Transcription profiling reveals cooperative metabolic interactions in a microbial cheese-ripening community composed of *Debaryomyces hansenii*, *Brevibacterium aurantiacum*, and *Hafnia alvei*. Front Microbiol *10*, 1901.
Pinheiro, J., Lisboa, J., Pombinho, R., Carvalho, F., Carreaux, A., Brito, C., Pontinen, A., Korkeala, H., Dos Santos, N.M.S., Morais-Cabral, J.H., *et al.* (2018). MouR controls the expression of the *Listeria monocytogenes* Agr system and mediates virulence. Nucleic Acids Res *46*, 9338-9352.

Pisithkul, T., Patel, N.M., and Amador-Noguez, D. (2015). Post-translational modifications as key regulators of bacterial metabolic fluxes. Curr Opin Microbiol *24*, 29-37.
Poindexter, J.S. (1964). Biological properties and classification of the *Caulobacter* group. Bacteriol Rev *28*, 231-295.
Poindexter, J.S. (2015). *Caulobacter*. Bergey's Manual of Systematics of Archaea and Bacteria.
Prade, R.A., Zhan, D., Ayoubi, P., and Mort, A.J. (1999). Pectins, pectinases and plant-microbe interactions. Biotechnol Genet Eng Rev *16*, 361-391.
Prochazkova, K., Cermakova, K., Pachl, P., Sieglova, I., Fabry, M., Otwinowski, Z., and Rezacova, P. (2012). Structure of the effector-binding domain of the arabinose repressor AraR from *Bacillus subtilis*. Acta crystallographica Section D, Biological crystallography *68*, 176-185.
Quail, M.A., Dempsey, C.E., and Guest, J.R. (1994). Identification of a fatty acyl responsive regulator (FarR) in *Escherichia coli*. FEBS letters *356*, 183-187.
Raman, N., Black, P.N., and DiRusso, C.C. (1997). Characterization of the fatty acid-responsive transcription factor FadR. Biochemical and genetic analyses of the native conformation and functional domains. J Biol Chem *272*, 30645-30650.
Raman, N., and DiRusso, C.C. (1995). Analysis of acyl coenzyme A binding to the transcription factor FadR and identification of amino acid residues in the carboxyl terminus required for ligand binding. J Biol Chem *270*, 1092-1097.
Ramos, K.R., Valdehuesa, K.N., Liu, H., Nisola, G.M., Lee, W.K., and Chung, W.J. (2014). Combining De Ley-Doudoroff and methylerythritol phosphate pathways for enhanced isoprene biosynthesis from D-galactose. Bioprocess and biosystems engineering *37*, 2505-2513.
Ramos, K.R., Valdehuesa, K.N., Maza, P.A.M.M., Nisola, G.M., Lee, W.K., and Chung, W.J. (2017). Overexpression and characterization of a novel α-neoagarobiose hydrolase and its application in the production of D-galactonate from *Gelidium amansii*. Process Biochemistry *63*, 105-112.
Ramseier, T.M., Bledig, S., Michotey, V., Feghali, R., and Saier, M.H., Jr. (1995). The global regulatory protein FruR modulates the direction of carbon flow in *Escherichia coli*. Mol Microbiol *16*, 1157-1169.
Ran, N., Draths, K.M., and Frost, J.W. (2004). Creation of a shikimate pathway variant. J Am Chem Soc *126*, 6856-6857.
Ran, N., and Frost, J.W. (2007). Directed evolution of 2-keto-3-deoxy-6-phosphogalactonate aldolase to replace 3-deoxy-D-arabino-heptulosonic acid 7-phosphate synthase. J Am Chem Soc *129*, 6130-6139.
Rancour, N.J., Hawkins, E.D., and Wells, W.W. (1979). Galactose oxidation in liver. Arch Biochem Biophys *193*, 232-241.
Reed, J.L., Patel, T.R., Chen, K.H., Joyce, A.R., Applebee, M.K., Herring, C.D., Bui, O.T., Knight, E.M., Fong, S.S., and Palsson, B.O. (2006). Systems approach to refining genome annotation. Proc Natl Acad Sci U S A *103*, 17480-17484.
Resch, M., Schiltz, E., Titgemeyer, F., and Muller, Y.A. (2010). Insight into the induction mechanism of the GntR/HutC bacterial transcription regulator YvoA. Nucleic Acids Res *38*, 2485-2497.
Reuther, J., Wohlleben, W., and Muth, G. (2006). Modular architecture of the conjugative plasmid pSVH1 from *Streptomyces venezuelae*. Plasmid *55*, 201-209.
Rezacova, P., Krejcirikova, V., Borek, D., Moy, S.F., Joachimiak, A., and Otwinowski, Z. (2007). The crystal structure of the effector-binding domain of the trehalose repressor TreR from *Bacillus subtilis* 168 reveals a unique quarternary assembly. Proteins *69*, 679-682.
Richard, P., and Hilditch, S. (2009). D-Galacturonic acid catabolism in microorganisms and its biotechnological relevance. Appl Microbiol Biotechnol *82*, 597-604.
Rigali, S., Derouaux, A., Giannotta, F., and Dusart, J. (2002). Subdivision of the helix-turn-helix GntR family of bacterial regulators in the FadR, HutC, MocR, and YtrA subfamilies. J Biol Chem *277*, 12507-12515.
Ritzenthaler, P., Blanco, C., and Mata-Gilsinger, M. (1983). Interchangeability of repressors for the control of the *uxu* and *uid* operons in *E. coli* K12. Molecular & general genetics : MGG *191*, 263-270.

Ritzenthaler, P., and Mata-Gilsinger, M. (1982). Use of in vitro gene fusions to study the *uxuR* regulatory gene in *Escherichia coli* K-12: direction of transcription and regulation of its expression. J Bacteriol *150*, 1040-1047.
Robert-Baudouy, J., Portalier, R., and Stoeber, F. (1981). Regulation of hexuronate system genes in *Escherichia coli* K-12: multiple regulation of the *uxu* operon by *exuR* and *uxuR* gene products. J Bacteriol *145*, 211-220.
Robyt, J.F. (1998). Essentials of carbohydrate chemistry. New York: Springer, 366-367.
Rodionov, D.A., and Gelfand, M.S. (2006). Computational identification of BioR, a transcriptional regulator of biotin metabolism in Alphaproteobacteria, and of its binding signal. FEMS Microbiol Lett *255*, 102-107.
Rodionov, D.A., Mironov, A.A., Rakhmaninova, A.B., and Gelfand, M.S. (2000). Transcriptional regulation of transport and utilization systems for hexuronides, hexuronates and hexonates in gamma purple bacteria. Mol Microbiol *38*, 673-683.
Rogers, S., Lichtenstein, G., Gentile, D., and Segal, S. (1984). Accumulation of galactonate in liver of suckling rats perfused with galactose. Biochemical and biophysical research communications *118*, 304-309.
Rojo, F. (1999). Repression of transcription initiation in bacteria. J Bacteriol *181*, 2987-2991.
Romero Zaliz, C.L., and Varela, O. (2003). Straightforward synthesis of derivatives of D- and L-galactonic acids as precursors of stereoregular polymers. Tetrahedron: Asymmetry *14*, 2579-2586.
Romero Zaliz, C.L., and Varela, O. (2006). Facile synthesis of a D-galactono-1,6-lactone derivative, a precursor of a copolyester. Carbohydr Res *341*, 2973-2977.
Roos, V., Ulett, G.C., Schembri, M.A., and Klemm, P. (2006). The asymptomatic bacteriuria *Escherichia coli* strain 83972 outcompetes uropathogenic *E. coli* strains in human urine. Infect Immun *74*, 615-624.
Rosenow, C., Maniar, M., and Trias, J. (1999). Regulation of the alpha-galactosidase activity in *Streptococcus pneumoniae*: characterization of the raffinose utilization system. Genome Res *9*, 1189-1197.
Rossbach, S., Kulpa, D.A., Rossbach, U., and de Bruijn, F.J. (1994). Molecular and genetic characterization of the rhizopine catabolism (*mocABRC*) genes of *Rhizobium meliloti* L5-30. Molecular & general genetics : MGG *245*, 11-24.
Ryan, M.P., Pembroke, J.T., and Adley, C.C. (2006). *Ralstonia pickettii*: a persistent gram-negative nosocomial infectious organism. J Hosp Infect *62*, 278-284.
Ryding, N.J., Kelemen, G.H., Whatling, C.A., Flardh, K., Buttner, M.J., and Chater, K.F. (1998). A developmentally regulated gene encoding a repressor-like protein is essential for sporulation in *Streptomyces coelicolor* A3(2). Mol Microbiol *29*, 343-357.
Sa-Nogueira, I., and Mota, L.J. (1997). Negative regulation of L-arabinose metabolism in *Bacillus subtilis*: characterization of the *araR* (*araC*) gene. J Bacteriol *179*, 1598-1608.
Saffarian, A., Mulet, C., Naito, T., Bouchier, C., Tichit, M., Ma, L., Grompone, G., Sansonetti, P.J., and Pedron, T. (2015). Draft genome sequences of *Acinetobacter parvus* CM11, *Acinetobacter radioresistens* CM38, and *Stenotrophomonas maltophilia* BR12, isolated from murine proximal colonic tissue. Genome Announc *3*.
Saier, M.H., Jr., Reddy, V.S., Tsu, B.V., Ahmed, M.S., Li, C., and Moreno-Hagelsieb, G. (2016). The Transporter Classification Database (TCDB): recent advances. Nucleic Acids Res *44*, D372-379.
Saini, S., Pearl, J.A., and Rao, C.V. (2009). Role of FimW, FimY, and FimZ in regulating the expression of type I fimbriae in *Salmonella enterica* serovar Typhimurium. J Bacteriol *191*, 3003-3010.
Sakihama, Y., Mizoguchi, H., Oshima, T., and Ogasawara, N. (2012). YdfH identified as a repressor of *rspA* by the use of reduced genome *Escherichia coli* MGF-01. Biosci Biotechnol Biochem *76*, 1688-1693.
Salzberg, L.I., Luo, Y., Hachmann, A.B., Mascher, T., and Helmann, J.D. (2011). The *Bacillus subtilis* GntR family repressor YtrA responds to cell wall antibiotics. J Bacteriol *193*, 5793-5801.
Sampaio, M.M., Chevance, F., Dippel, R., Eppler, T., Schlegel, A., Boos, W., Lu, Y.J., and Rock, C.O. (2004). Phosphotransferase-mediated transport of the osmolyte 2-O-alpha-mannosyl-D-

glycerate in *Escherichia coli* occurs by the product of the *mngA* (*hrsA*) gene and is regulated by the *mngR* (*farR*) gene product acting as repressor. J Biol Chem *279*, 5537-5548.
Sangwan, A. (2017). Characterization of DNA binding site and oligomerization status of DgoR: A negative regulator of D-galactonate metabolism in *Escherichia coli* (Thesis submitted to Indian Institute of Science Education and Research Mohali, India, for BS-MS dual degree).
Savery, N.J., Lloyd, G.S., Kainz, M., Gaal, T., Ross, W., Ebright, R.H., Gourse, R.L., and Busby, S.J. (1998). Transcription activation at Class II CRP-dependent promoters: identification of determinants in the C-terminal domain of the RNA polymerase alpha subunit. The EMBO journal *17*, 3439-3447.
Schadewaldt, P., Hammen, H.W., Stolpmann, S., Kamalanathan, L., and Wendel, U. (2004). Galactonate determination in urine by stable isotope dilution gas chromatography-mass spectrometry. Journal of chromatography B, Analytical technologies in the biomedical and life sciences *801*, 249-255.
Schei, K., Avershina, E., Oien, T., Rudi, K., Follestad, T., Salamati, S., and Odegard, R.A. (2017). Early gut mycobiota and mother-offspring transfer. Microbiome *5*, 107.
Seif, Y., Kavvas, E., Lachance, J.C., Yurkovich, J.T., Nuccio, S.P., Fang, X., Catoiu, E., Raffatellu, M., Palsson, B.O., and Monk, J.M. (2018). Genome-scale metabolic reconstructions of multiple *Salmonella* strains reveal serovar-specific metabolic traits. Nature communications *9*, 3771.
Semsey, S., Geanacopoulos, M., Lewis, D.E., and Adhya, S. (2002). Operator-bound GalR dimers close DNA loops by direct interaction: tetramerization and inducer binding. The EMBO journal *21*, 4349-4356.
Seo, J.W., Ohnishi, Y., Hirata, A., and Horinouchi, S. (2002). ATP-binding cassette transport system involved in regulation of morphological differentiation in response to glucose in *Streptomyces griseus*. J Bacteriol *184*, 91-103.
Shahmuradov, I.A., Mohamad Razali, R., Bougouffa, S., Radovanovic, A., and Bajic, V.B. (2017). bTSSfinder: a novel tool for the prediction of promoters in cyanobacteria and *Escherichia coli*. Bioinformatics *33*, 334-340.
Shi, W., Kovacikova, G., Lin, W., Taylor, R.K., Skorupski, K., and Kull, F.J. (2015). The 40-residue insertion in *Vibrio cholerae* FadR facilitates binding of an additional fatty acyl-CoA ligand. Nature communications *6*, 6032.
Shimizu, S., Kataoka, M., Shimizu, K., Hirakata, M., Sakamoto, K., and Yamada, H. (1992). Purification and characterization of a novel lactonohydrolase, catalyzing the hydrolysis of aldonate lactones and aromatic lactones, from *Fusarium oxysporum*. Eur J Biochem *209*, 383-390.
Singh, B., Arya, G., Kundu, N., Sangwan, A., Nongthombam, S., and Chaba, R. (2019). Molecular and functional insights into the regulation of D-Galactonate metabolism by the transcriptional regulator DgoR in *Escherichia coli*. Journal of Bacteriology *201*, e00281-00218.
Smanski, M.J., Peterson, R.M., Rajski, S.R., and Shen, B. (2009). Engineered *Streptomyces platensis* strains that overproduce antibiotics platensimycin and platencin. Antimicrob Agents Chemother *53*, 1299-1304.
Smith, N., Roitberg, A.E., Rivera, E., Howard, A., Holden, M.J., Mayhew, M., Kaistha, S., and Gallagher, D.T. (2006). Structural analysis of ligand binding and catalysis in chorismate lyase. Arch Biochem Biophys *445*, 72-80.
Solovyev, V., and Salamov, A. (2011). Automatic annotation of microbial genomes and metagenomic sequences. *In* RW L (ed), Metagenomics and its applications in agriculture, biomedicine and environmental studies Nova Science Publishers, Hauppauge, NY, 61-78.
Stefanovski, L., Triebkorn, P., Spiegler, A., Diaz-Cortes, M.A., Solodkin, A., Jirsa, V., McIntosh, A.R., Ritter, P., and Alzheimer's Disease Neuroimaging, I. (2019). Linking molecular pathways and large-scale computational modeling to assess candidate disease mechanisms and pharmacodynamics in alzheimer's disease. Front Comput Neurosci *13*, 54.
Suhr, M.J., and Hallen-Adams, H.E. (2015). The human gut mycobiome: pitfalls and potentials--a mycologist's perspective. Mycologia *107*, 1057-1073.
Sulieman, S., and Tran, L.S. (2014). Symbiotic nitrogen fixation in legume nodules: metabolism and regulatory mechanisms. Int J Mol Sci *15*, 19389-19393.

Suvorova, I.A., Korostelev, Y.D., and Gelfand, M.S. (2015). GntR family of bacterial transcription factors and their DNA binding motifs: Structure, positioning and co-evolution. PLoS One *10*, e0132618.
Suvorova, I.A., Ravcheev, D.A., and Gelfand, M.S. (2012). Regulation and evolution of malonate and propionate catabolism in proteobacteria. J Bacteriol *194*, 3234-3240.
Suvorova, I.A., Tutukina, M.N., Ravcheev, D.A., Rodionov, D.A., Ozoline, O.N., and Gelfand, M.S. (2011). Comparative genomic analysis of the hexuronate metabolism genes and their regulation in gammaproteobacteria. J Bacteriol *193*, 3956-3963.
Sweeney, N.J., Laux, D.C., and Cohen, P.S. (1996). *Escherichia coli* F-18 and E. coli K-12 *eda* mutants do not colonize the streptomycin-treated mouse large intestine. Infect Immun *64*, 3504-3511.
Szumilo, T. (1981). Pathway for D-galactonate catabolism in nonpathogenic mycobacteria. J Bacteriol *148*, 368-370.
Szumilo, T. (1983). Separation and some properties of D-galactonate pathway enzymes from *Mycobacterium* sp. 607. Acta microbiologica Polonica *32*, 47-52.
Tamayo, R., Ryan, S.S., McCoy, A.J., and Gunn, J.S. (2002). Identification and genetic characterization of PmrA-regulated genes and genes involved in polymyxin B resistance in *Salmonella enterica* serovar Typhimurium. Infect Immun *70*, 6770-6778.
Titgemeyer, F., Reizer, J., Reizer, A., Tang, J., Parr, T.R., Jr., and Saier, M.H., Jr. (1995). Nucleotide sequence of the region between *crr* and *cysM* in *Salmonella* typhimurium: five novel ORFs including one encoding a putative transcriptional regulator of the phosphotransferase system. DNA Seq *5*, 145-152.
Touchon, M., Hoede, C., Tenaillon, O., Barbe, V., Baeriswyl, S., Bidet, P., Bingen, E., Bonacorsi, S., Bouchier, C., Bouvet, O., *et al.* (2009). Organised genome dynamics in the *Escherichia coli* species results in highly diverse adaptive paths. PLoS genetics *5*, e1000344.
Trimigno, A., Munger, L., Picone, G., Freiburghaus, C., Pimentel, G., Vionnet, N., Pralong, F., Capozzi, F., Badertscher, R., and Vergeres, G. (2018). GC-MS based metabolomics and NMR Spectroscopy investigation of food intake biomarkers for milk and cheese in serum of healthy humans. Metabolites *8*.
Truong-Bolduc, Q.C., and Hooper, D.C. (2007). The transcriptional regulators NorG and MgrA modulate resistance to both quinolones and beta-lactams in *Staphylococcus aureus*. J Bacteriol *189*, 2996-3005.
Tsunedomi, R., Izu, H., Kawai, T., Matsushita, K., Ferenci, T., and Yamada, M. (2003a). The activator of GntII genes for gluconate metabolism, GntH, exerts negative control of GntR-regulated GntI genes in *Escherichia coli*. J Bacteriol *185*, 1783-1795.
Tsunedomi, R., Izu, H., Kawai, T., and Yamada, M. (2003b). Dual control by regulators, GntH and GntR, of the GntII genes for gluconate metabolism in *Escherichia coli*. J Mol Microbiol Biotechnol *6*, 41-56.
Tsypik, O., Makitrynskyy, R., Bera, A., Song, L., Wohlleben, W., Fedorenko, V., and Ostash, B. (2017). Role of GntR Family regulatory gene SCO1678 in gluconate metabolism in *Streptomyces coelicolor* M145. Biomed Res Int *2017*, 9529501.
Tsypik, O., Yushchuk, O., Zaburannyi, N., Flardh, K., Walker, S., Fedorenko, V., and Ostash, B. (2016). Transcriptional regulators of GntR family in *Streptomyces coelicolor* A3(2): analysis in silico and *in vivo* of YtrA subfamily. Folia Microbiol (Praha) *61*, 209-220.
Tutukina, M.N., Potapova, A.V., Cole, J.A., and Ozoline, O.N. (2016a). Control of hexuronate metabolism in *Escherichia coli* by the two interdependent regulators, ExuR and UxuR: derepression by heterodimer formation. Microbiology *162*, 1220-1231.
Tutukina, M.N., Potapova, A.V., Vlasov, P.K., Purtov, Y.A., and Ozoline, O.N. (2016b). Structural modeling of the ExuR and UxuR transcription factors of *E. coli*: search for the ligands affecting their regulatory properties. J Biomol Struct Dyn *34*, 2296-2304.
Valmeekam, V., Loh, Y.L., and San Francisco, M.J. (2001). Control of *exuT* activity for galacturonate transport by the negative regulator ExuR in *Erwinia chrysanthemi* EC16. Mol Plant Microbe Interact *14*, 816-820.
van Aalten, D.M., DiRusso, C.C., and Knudsen, J. (2001). The structural basis of acyl coenzyme A-dependent regulation of the transcription factor FadR. The EMBO journal *20*, 2041-2050.

van Aalten, D.M., DiRusso, C.C., Knudsen, J., and Wierenga, R.K. (2000). Crystal structure of FadR, a fatty acid-responsive transcription factor with a novel acyl coenzyme A-binding fold. The EMBO journal *19*, 5167-5177.
Vionnet, N., Munger, L.H., Freiburghaus, C., Burton, K.J., Pimentel, G., Pralong, F.P., Badertscher, R., and Vergeres, G. (2019). Assessment of lactase activity in humans by measurement of galactitol and galactonate in serum and urine after milk intake. Am J Clin Nutr *109*, 470-477.
Walters, M.J., Srikannathasan, V., McEwan, A.R., Naismith, J.H., Fierke, C.A., and Toone, E.J. (2008). Characterization and crystal structure of *Escherichia coli* KDPGal aldolase. Bioorg Med Chem *16*, 710-720.
Wassinger, A., Zhang, L., Tracy, E., Munson, R.S., Jr., Kathariou, S., and Wang, H.H. (2013). Role of a GntR-family response regulator LbrA in *Listeria monocytogenes* biofilm formation. PLoS One *8*, e70448.
Watanabe, T., Inoue, R., Kimura, N., and Furukawa, K. (2000). Versatile transcription of biphenyl catabolic *bph* operon in *Pseudomonas pseudoalcali* genes KF707. J Biol Chem *275*, 31016-31023.
Wehrli, S.L., Berry, G.T., Palmieri, M., Mazur, A., Elsas, L., 3rd, and Segal, S. (1997). Urinary galactonate in patients with galactosemia: quantitation by nuclear magnetic resonance spectroscopy. Pediatric research *42*, 855-861.
Wiame, E., Delpierre, G., Collard, F., and Van Schaftingen, E. (2002). Identification of a pathway for the utilization of the Amadori product fructoselysine in *Escherichia coli*. J Biol Chem *277*, 42523-42529.
Wieczorek, S.J., Kalivoda, K.A., Clifton, J.G., Ringe, D., Petsko, G.A., and Gerlt, J.A. (1999). Evolution of enzymatic activities in the enolase superfamily: identification of a "New" general acid catalyst in the active site of D-galactonate dehydratase from *Escherichia coli*. J Am Chem Soc *121*, 4540-4541.
Wong, T.Y., and Yao, X.T. (1994). The DeLey-Doudoroff pathway of galactose metabolism in *Azotobacter vinelandii*. Appl Environ Microbiol *60*, 2065-2068.
Wu, K., Xu, H., Zheng, Y., Wang, L., Zhang, X., and Yin, Y. (2016). CpsR, a GntR family regulator, transcriptionally regulates capsular polysaccharide biosynthesis and governs bacterial virulence in *Streptococcus pneumoniae*. Sci Rep *6*, 29255.
Wu, R., Sanishvili, R., Belitsky, B.R., Juncosa, J.I., Le, H.V., Lehrer, H.J., Farley, M., Silverman, R.B., Petsko, G.A., Ringe, D., *et al.* (2017). PLP and GABA trigger GabR-mediated transcription regulation in *Bacillus subtilis* via external aldimine formation. Proc Natl Acad Sci U S A *114*, 3891-3896.
Xu, Y., Heath, R.J., Li, Z., Rock, C.O., and White, S.W. (2001). The FadR.DNA complex. Transcriptional control of fatty acid metabolism in *Escherichia coli*. J Biol Chem *276*, 17373-17379.
Yager, C., Ning, C., Reynolds, R., Leslie, N., and Segal, S. (2004). Galactitol and galactonate accumulation in heart and skeletal muscle of mice with deficiency of galactose-1-phosphate uridyltransferase. Molecular genetics and metabolism *81*, 105-111.
Yang, C., Rodionov, D.A., Li, X., Laikova, O.N., Gelfand, M.S., Zagnitko, O.P., Romine, M.F., Obraztsova, A.Y., Nealson, K.H., and Osterman, A.L. (2006). Comparative genomics and experimental characterization of N-acetylglucosamine utilization pathway of *Shewanella oneidensis*. J Biol Chem *281*, 29872-29885.
Yoon, S.H., Han, M.J., Jeong, H., Lee, C.H., Xia, X.X., Lee, D.H., Shim, J.H., Lee, S.Y., Oh, T.K., and Kim, J.F. (2012). Comparative multi-omics systems analysis of *Escherichia coli* strains B and K-12. Genome Biol *13*, R37.
Yoshida, K., Fujita, Y., and Sarai, A. (1993). Missense mutations in the *Bacillus subtilis gnt* repressor that diminish operator binding ability. J Mol Biol *231*, 167-174.
Yoshida, K., Ohmori, H., Miwa, Y., and Fujita, Y. (1995). *Bacillus subtilis gnt* repressor mutants that diminish gluconate-binding ability. J Bacteriol *177*, 4813-4816.
Yu, L., Gao, W., Li, S., Pan, Y., and Liu, G. (2016). GntR family regulator SCO6256 is involved in antibiotic production and conditionally regulates the transcription of myo-inositol catabolic genes in *Streptomyces coelicolor* A3(2). Microbiology *162*, 537-551.

Yu, L., Li, S., Gao, W., Pan, Y., Tan, H., and Liu, G. (2015). Regulation of myo-inositol catabolism by a GntR-type repressor SCO6974 in *Streptomyces coelicolor*. Appl Microbiol Biotechnol *99*, 3141-3153.
Zeng, J., Deng, W., Yang, W., Luo, H., Duan, X., Xie, L., Li, P., Wang, R., Fu, T., Abdalla, A.E., *et al.* (2016). Mycobacterium tuberculosis Rv1152 is a novel GntR family transcriptional regulator involved in intrinsic vancomycin resistance and is a potential vancomycin adjuvant target. Sci Rep *6*, 28002.
Zhang, L., and van Kan, J.A. (2013). *Botrytis cinerea* mutants deficient in D-galacturonic acid catabolism have a perturbed virulence on *Nicotiana benthamiana* and *Arabidopsis*, but not on tomato. Mol Plant Pathol *14*, 19-29.
Zhang, X.S., Garcia-Contreras, R., and Wood, T.K. (2008). *Escherichia coli* transcription factor YncC (McbR) regulates colanic acid and biofilm formation by repressing expression of periplasmic protein YbiM (McbA). ISME J *2*, 615-631.
Zhang, Y., Lin, C.Y., Li, X.M., Tang, Z.K., Qiao, J., and Zhao, G.R. (2016). DasR positively controls monensin production at two-level regulation in *Streptomyces cinnamonensis*. J Ind Microbiol Biotechnol *43*, 1681-1692.
Zheng, M., Cooper, D.R., Grossoehme, N.E., Yu, M., Hung, L.W., Cieslik, M., Derewenda, U., Lesley, S.A., Wilson, I.A., Giedroc, D.P., *et al.* (2009). Structure of *Thermotoga maritima* TM0439: implications for the mechanism of bacterial GntR transcription regulators with Zn^{2+}-binding FCD domains. Acta crystallographica Section D, Biological crystallography *65*, 356-365.
Zhou, X., Hua, X., Zhou, X., and Xu, Y. (2018). Process for the successive production of calcium galactonate crystals by *Gluconobacter oxydans*. Bioresource technology *261*, 458-460.
Zhou, X., Yan, Q., and Wang, N. (2017). Deciphering the regulon of a GntR family regulator via transcriptome and ChIP-exo analyses and its contribution to virulence in *Xanthomonas citri*. Mol Plant Pathol *18*, 249-262.