# MOLECULAR MECHANISM OF TRANSCRIPTIONAL REGULATION OF THE *PUT* REGULON IN *ESCHERICHIA COLI*

**By**

**Yuzhen Zhou**

**A DISSERTATION**

**Presented to the Faculty of**

**The Graduate College at the University of Nebraska**

**In Partial Fulfillment of Requirements**

**For the Degree of Doctoral of Philosophy**

**Major: Biochemistry**

**Under the Supervision of Professor Donald F. Becker**

**Lincoln, Nebraska**

**December, 2007**

UMI Number: 3284262



UMI Microform 3284262



ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

# MOLECULAR MECHANISM OF TRANSCRIPTIONAL REGULATION OF THE *PUT* REGULON IN *ESCHERICHIA COLI*

Yuzhen Zhou, Ph. D.

University of Nebraska, 2007

Advisor: Donald F. Becker

PutA from *Escherichia coli* functions both as a membrane-associated multifunctional enzyme and a transcriptional repressor of the *put* regulon, which encodes the *putA* and *putP* genes. The intracellular localization and function of PutA is regulated by the availability of proline.

In this study, we investigated how PutA represses the *put* regulon in the absence of proline. First, we localized the PutA-DNA binding domain to the N-terminal 47 amino acids and clearly demonstrated that PutA belongs to a member of the ribbon-helix-helix (RHH) family transcription factors. Second, we identified five PutA binding sites on the *put* control DNA region and studied the roles of these binding sites in the regulation of the *putA* or *putP* genes. We found that PutA binds to two of the binding sites to repress the *putP* gene and the other three sites to repress the *putA* gene expression.

We also explored how PutA switches its localization and function in the presence of proline. Proline reduces the PutA flavin cofactor which induces PutA-membrane binding, where PutA catalyzes the two-step oxidation of proline to glutamate. The PutA-membrane interaction probably leads to a conformational change so that PutA can't interact with DNA, therefore, releasing the repression of the *put* genes.

Further, we studied how the *put* genes are regulated by carbon and nitrogen sources. We provided the first evidence that the cAMP receptor protein (CRP) regulates the *put* genes as an activator by directly interacting with the *put* control DNA. We also explored whether the nitrogen assimilation control protein (NAC) regulates the *put* genes.

# ACKNOWLEDGEMENTS

First of all, I would like to gratefully and sincerely thank my advisor, Dr. Donald Becker, for his constant support, guidance, patience and encouragement throughout my graduate studies. His patience and support helped me overcome many difficulties and finish this dissertation. He encouraged me to not only grow as an experimentist and a biochemist but also as an instructor and an independent scientific thinker. His enormous scientific attitude and scientific ethics will continuously guide me in the rest of my career life.

I appreciate very much my committee members: Dr. Audrey L. Atkin, Dr. Vadim N. Gladyshev, Dr. Stephen W Ragsdale, Dr. Greg A. Somerville and Dr. Julie M. Stone for their invaluable comments and suggestions during my annual committee meetings and my comprehensive exam. All of them except Dr. Somerville taught my classes, which expanded my knowledge in biochemistry, molecular biology and bioinformatics.

I would also like to give my special thanks to Dr. John J. Tanner, Dr. Robert Powers, Dr. John D. Larson and Steve Halouska for a fruitful collaboration and as co-authors. They taught me the knowledge in structural biology.

I am also very grateful to have good lab-mates. Their success in PutA research made my research faster. They gave me my first hands-on lab experience in molecular biology and the protein world. Their friendship and assistance made my PhD study more enjoyable and unforgettable.

Finally, I would like to thank my husband, Peng, for his understanding, encouragement and endless care in both my everyday life and my research. He can always find a way of thinking that makes our life easier. I would also thank my son, Max, for choosing and coming to us. His coming made our family a more complete one and made me understand life better. I also really appreciate my mother-in-law for taking care of Max. My parents should receive deepest gratitude and love for their silent dedication and support all these years. I am also grateful to my sisters, my brother and their family for their sincere care and support.

# TABLE OF CONTENTS

# ABBREVIATIONS

*put*, proline utilization

PutA, proline utilization protein A

FAD, flavin adenine dinucleotide

$NAD^+$, nicotinamide adenine dinucleotide

PRODH, proline dehydrogenase

PRODH2, proline dehydrogenase isoform 2

P5C, $\Delta^1$-pyrroline-5-carboxylate

P5CDH, P5C dehydrogenase

P5CS, P5C synthetase

P5CR, P5C reductase

GSA, γ-glutamic semialdehyde

OAT, ornithine δ-aminotransferase

ROS, reactive oxygen species

TCA cycle, tricarboxylic acid cycle

PutC, *put* control DNA region

CRP, cyclic AMP (cAMP) receptor protein

NAC, nitrogen assimilation control protein

SPR, surface plasmon resonance

SDS-PAGE, sodium dodecyl sulfate polyacrylamide gel electrophoresis

HTH, helix-turn-helix

HLH, helix-loop-helix

RHH, ribbon-helix-helix

ChIP, chromatin immunoprecipitation

EMSA, Electrophoretic mobility shift assay

CD, circular dichroism

PCR, polymerase Chain Reaction

DCPIP, dichlorophenolinindophenol

PMS, phenazine methosulfate

MES, 2-(N-morpholino) ethanesulfonic acid

THFA, tetrahydro-2-furoic acid

NMR, nuclear magnetic resonance

HSQC, heteronuclear single quantum correlation

ONPG, *o*-nitrophenyl-β-D-galactopyranoside

HEPES, 4-(2-hydroxyethyl) piperazine-1-ethanesulfonic acid

PriA, Primosomal protein N'

EDTA, ethylenediaminetetraacetic acid

PVDF, polyvinylidene difluoride

# Chapter 1

## Introduction

## 1.1 General information about proline

### 1.1.1 Proline and its physiological roles

The name "proline" was derived from "pyrrolidine" by Emil Fisher in 1904. Proline is formally not an amino acid, but an imino acid. It is one of the 22 amino acids that are used in living organisms as building blocks of proteins. Proline is the only cyclic amino acid and lacks the hydrogen on the α amino group, so it cannot donate a hydrogen bond to stabilize a α-helix or a β-sheet and is often found at the end of a α-helix or in β-turns or loops.

Proline plays important roles in organisms from prokaryotes to eukaryotes. First, proline catabolism provides an important energy source in different organisms. In enteric bacteria, proline can be used as the only carbon, nitrogen and energy source (2, 25, 73, 74, 91, 116, 125). In nitrogen-fixing soil microbes, such as *Bradyrhizobium japonicm*, proline catabolism is required for colonization and therefore for the nodulation efficiency and competitiveness (131). Gastrointestinal pathogens utilize proline in the gut environment where 10-fold higher levels of proline have been detected in gastric juice of patients infected with *Helicobacter pylori* relative to non-infected individuals (95). *Saccharomyces cerevisiae* can utilize proline as the sole nitrogen source (22). Proline is the sole source of energy for certain insects during flight and is stored in high concentration in flight muscles such as in tsetse flies (59) and the mosquito *Aedes aegypti* females (115). Proline is also an important energy source for *Trypanosoma brucei*, which causes African sleeping sickness (19, 76).

It has been known for a long time that the concentration of proline increases in a large variety of plants under different stresses up to 100 times the normal level (81),

which makes up to 80% of the total pool of amino acids. The different stresses include cold (135), heat (143)**,** salt (120, 148), drought (11), UV (114), and heavy metals (49). In the literature, the function of proline in stressed plants is often explained by its property as an osmolyte, capable of balancing water stress (114, 133). In addition, other possible positive roles of proline during stress have been proposed, including stabilization of proteins, membranes and subcellular structures (7); scavenging of hydroxyl radicals (121, 129); freeze tolerance (129); regulation of cytosolic pH (132) and regulation of the intracellular $NAD^+$/NADH ratio (5). The exact molecular mechanism of proline-induced protection in plant stress is still unknown.

Proline has also been shown to be involved in protein chaperoning by promoting the correct refolding of protein both in vitro and in the cell by protecting native proteins from heat denaturation (8, 20, 130, 142). Proline is believed to act through a combination of solvophobic backbone interactions and favorable side-chain interactions that are not specific to particular sequence or structure (33, 44, 86). Thus, the osmolyte proline may be protective against biomedically important protein aggregates that are hallmarks of several late-onset neurodegenerative diseases including Huntington's, Alzheimer's and Parkinson's (62).

Proline plays a dual role in modulating the level of endogenous reactive oxygen species (ROS) due to its radical scavenging properties as a free amino acid and its ability to couple of the oxidation to the mitochondrial membrane by proline dehydrogenase (PRODH). The five-membered ring of proline, pyrrolidine, has a low ionization potential that allows effective quenching of singlet oxygen and scavenging of hydroxyl radicals (63). Evidence that proline oxidation increases ROS levels has come from studies on

mammals that PRODH is a p53-inducible gene (43, 105). p53 has numerous roles in cell regulation, genome stability, apoptosis and reducing carcinogenesis (64, 80). So proline may play a critical role in helping regulate the intracellular redox environment by impacting ROS levels and mitochondrial-linked apoptotic pathways.

Proline also plays an important role in humans, even though proline is a nonessential amino acid and can be synthesized from glutamate and ornithine. Proline is shown to be important for wound healing, and an increase of the extracellular pool of free proline has been noted in burn and traumatized patient (4). The involvement of proline in human disease was first recognized just over 40 years ago. The PRODH2 (proline dehydrogenase isoform 2) gene is located on chromosome 22q11, which is a region of the human genome that long has been thought to contain a predisposition locus for psychotic illness (65). Up-regulation of PRODH2 and proline oxidation in lung and colon carcinoma cells has been shown to generate ROS and induce cell death by mitochondrial-dependent processes (80). Different genotypes of the PRODH2 gene cause varying degrees of type 1 hyperproteinemia (65), in which plasma proline concentrations are elevated from normal levels of ~ 0.2 mM to ~ 2 mM in schizophrenia patients.

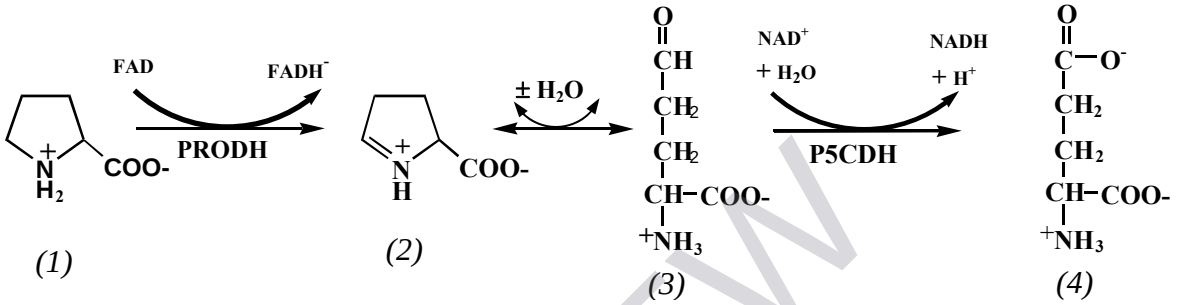


### 1.1.2 Proline utilization in bacteria

**Proline metabolism in general.** All organisms convert proline to glutamate in two enzymatic steps (Figure 1.1a). In the first step, a flavin-dependent proline dehydrogenase (PRODH) catalyzes the oxidation of proline to $\Delta^1$-pyrroline-5-carboxylate (P5C) by transferring two electrons to the FAD cofactor (reductive half-reaction). Electrons from reduced FAD ($FADH_2$) are then transferred to an acceptor in prokaryotes.

to complete the catalytic cycle (oxidative half-reaction). Next, P5C is hydrolyzed nonenzymatically to γ-glutamate semialdehyde which is oxidized to glutamate by $NAD^+$-

**a, Proline Oxidation:**




**b, Proline Biosynthesis:**

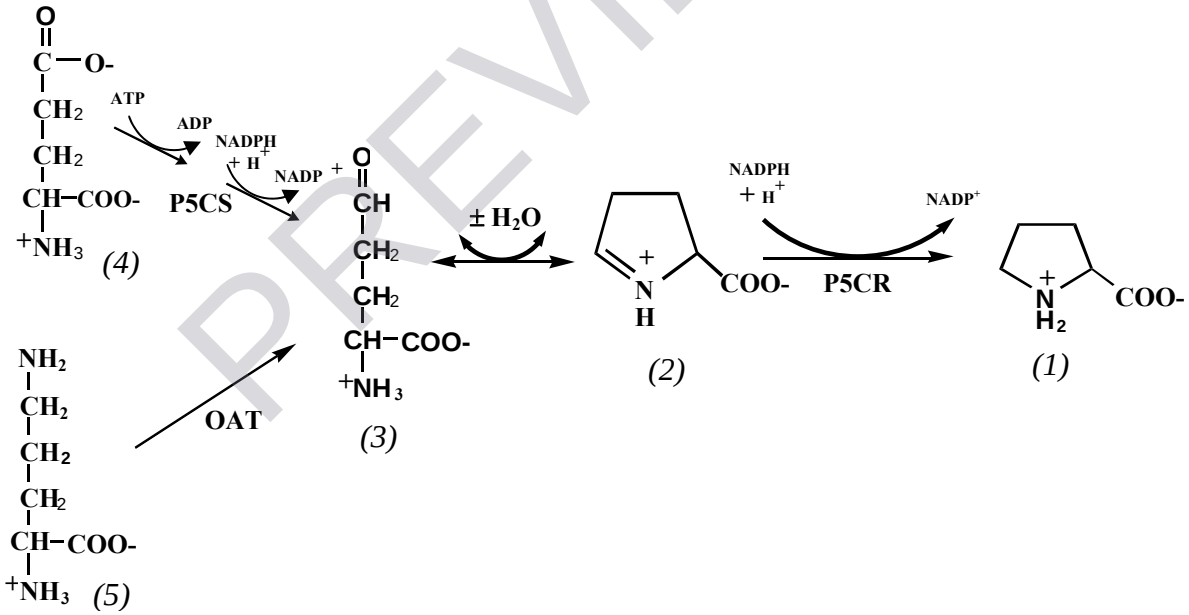



**Figure 1.1 Proline metabolic pathways.** (1), proline, (2), $\Delta^1$-pyrroline-5-carboxylate, (3), γ-glutamate semialdehyde, (4), glutamate, (5), ornithine. PRODH, P5CDH, P5CS, OAT and P5CR stand for proline dehydrogenase, P5C dehydrogenase, P5C synthesis, ornithine aminotransferase and P5C reductase, respectively.

dependent P5C dehydrogenase (P5CDH). Glutamate that is formed by proline oxidation eventually enters that TCA cycle via α-ketoglutamate. For proline biosynthesis (Figure 1.1b), glutamate or ornithine is converted to γ-glutamic semialdehyde (GSA) by $\Delta^1$-pyrroline-5-carboxylate synthetase (P5CS) or by ornithine δ-aminotransferase (OAT, EC 2.6.1.13), respectively. P5CS is a bifunctional enzyme that phosphorylates and reduces GSA, which is spontaneously converted to P5C. P5C is then reduced to proline by P5C reductase (P5CR, EC 1.5.1.2) (92). PRODH and P5CS catalyze the rate limiting steps in proline oxidation and biosynthesis, respectively.

**Proline utilization genes in bacteria**. In Gram-positive bacteria, archaea and eukaryotes, PRODH and P5CDH are distinct enzymes encoded on separate genes (134). In Gram-negative bacteria (25, 90), such as *Rhodobacter capsulatus* (71), *B. japonicm* (125), *Photobacterium leiognathi* (79), and *Agrobacterium tumefaciens* (37), both steps for proline utilization are catalyzed by a single polypeptide encoded by the *putA* gene. Proline utilization in Gram-negative bacteria also requires PutP, a high-affinity proline transporter encoded by the *putP* gene. In enteric bacteria, like *Escherichia coli* and *Salmonella enterica* serovar *typhimurium*, the *putA* and *putP* genes which comprise the *put* regulon map between the *pyrC* and *pyrD* genes at min 22 of the chromosome maps (108). The *putA* and *putP* genes are transcribed in opposite directions from different promoters in the *put* control region (PutC) which separates the *putA* and *putP* gene (Figure 1.2). Although the enzyme properties of PutA (having PRODH and P5CDH activities) are highly conserved among different bacteria, the genetic organization and control of expression of the *putA* and *putP* genes are quite divergent (see below).


**Figure 1.2 Organization of *put* regulon.**

**Regulation of the *put* genes in bacteria**. In enteric bacteria, like *E. coli, S. typhimurium, Klebsiella aerogenes and Pseudomonas putida,* the PutA protein functions as a transcriptional repressor of both the *putA* and *putP* genes in addition to its enzymatic activity. The transcriptional repressor activity of PutA is regulated by proline (25, 55, 101, 139). In the absence of proline, PutA binds to the *put* control region and represses both *putA* and *putP* genes. In the presence of proline, PutA is reduced and binds to the cytoplasmic membrane, therefore releasing the repression of both *putA* and *putP* genes. The *putA* and *putP* genes from these bacteria are under catabolic repression via the cyclic AMP (cAMP) receptor protein, CRP (34). The *putA* gene from *K. aerogenes* and *K. pneumoniae* is also positively regulated by the NAC protein (34, 85). CRP and NAC are two global transcriptional factors that are involved in carbon and nitrogen metabolism, respectively. Since proline is used as both a carbon and a nitrogen source, it is not surprising that the *putA* and *putP* genes are under the control of CRP and NAC even though it has not been shown explicitly that the *put* genes are under the control of the CRP and NAC in *E. coli*.

In other bacteria, like *Vibrio vulnificus*, *R. capsulatus, A. tumefaciens* and *B. japonicm*, the expression of the *putA* gene is regulated by the regulatory gene *putR*,

located immediately upstream of the *putA* gene (37, 71, 125). PutR belongs to the class of Lrp-like activator proteins (66, 77). In the absence of proline, the expression of the *putA* gene is low due to weak PutR-DNA binding. In the presence of proline, the affinity of PutR to the *putA* promoter increases thereby inducing the *putA* gene expression. PutR negatively autoregulates its own transcription both in the absence and presence of proline (122).

## 1.2 General information about gene regulation

The information required by every living organism for obtaining and maintaining its structure and function is encoded in its DNA. At various levels of information handling such as storage (DNA packing into chromatin), maintenance (e.g. repair of damaged DNA), copying (replication) and transfer (i.e. the expression of genes involving transcription and translation), protein-nucleic acid interactions play a central role. Particularly important for metabolism, replication and development of each organism is the expression of its genes as proteins at the correct location, in proper amounts and at the correct time in response to the cellular environment and developmental cycles. The central dogma of molecular biology is that DNA directs its own replication and its transcription to RNA, which, in turn, directs its translation into proteins. So, RNA is the bridge between the genetic material and functional proteins. RNA polymerase is the enzyme that catalyzes the synthesis of the RNA. In bacteria, RNA polymerase consists of the core enzyme and the sigma factors. A RNA core polymerase is a multi-subunit complex with a general structure of $\alpha_2\beta\beta'$ that catalyzes the elongation of RNA (21). Sigma factors are needed for the initiation of RNA transcription, and they have a major influence on selection of promoters (140).

### 1.2.1 Transcription Factors

A transcription factor is a protein needed to activate or repress the transcription of a gene (26). Some transcription factors bind to cis-acting sequences (transcription factor binding sites) only; some bind to each other; others bind to both DNA and other transcription factors (26). Regulation of gene transcription in bacteria involves a complex network, in which the DNA-binding transcription factors are a key component. They regulate the transcription of specific genes by acting on the cis-regulating sequences within the promoter regions of those genes (26). When a transcription factor binds to a specific promoter, it can either activate or repress transcription initiation (26). An activator stimulates the expression of the target gene, through either interacting with the C-terminal domain of RNA polymerase $\alpha$ subunit or $\sigma$ factor, thereby recruiting the polymerase to the promoter or inducing a conformational change in the promoter region therefore increasing the polymerase-promoter binding affinity. For negative control, the transcription factor binds to DNA to prevent RNA polymerase from initiating transcription by steric hindrance or DNA looping or modulating an activator activity. Some transcription factors function solely as an activator or a repressor, whereas others can function as either (dual regulators) according to the target promoters. Some transcription factors regulate only specific functionally related genes, whereas others regulate large member of genes that belong to different functional classes. These kinds of transcription factors are called global transcriptional regulator. Global transcription factors can control a complex regulatory cascade by a mechanism of not only directly controlling the expression of specific genes, but also indirectly by regulating various cellular pathways by acting on a set of local regulators which control just one or a few

genes. It has been estimated that seven global transcriptional factors (cAMP receptor protein (CRP), fumarate nitrate reductase regulator (FNR), integration host factor (IHF), factor for inversion stimulation (Fis), aerobic respiration control A (ArcA), nitrate/nitrite response regulator protein (NarL) and leucine-responsive regulatory protein (Lrp) in *E. coli* control 50% of all regulated genes, where as about 60 transcription factors control only a single promoter (87).

### 1.2.2 AraC: a paradigm of positive and negative control by the same protein

The arabinose system enables *E. coli* to take up the pentose L-arabinose from the growth medium using products of the unlinked *araA* and *araFGH* genes (the low and the high arabinose affinity uptake system, respectively) and then convert intracellular arabinose in three steps catalyzed by the products of the *araBAD* genes to D-xylulose-5-phosphate (Figure 1.3) (118), which then enters the pentose phosphate pathway (118). AraC, a dimer of ~ 30 kDa encoded by *araC*, is the protein that regulates all of the genes involved in uptake and utilization of arabinose, but the mechanisms involved are different (Figure 1.3) (123).

The regulation of the *araBAD* operon involves a DNA looping mechanism. An unusual characteristic of the AraC binding sites on the promoter of *araBAD* ($P_{BAD}$) is that they form direct repeats (27, 78). So binding to the adjacent half-sites ($araI_1$ and $araI_2$, which are separated by 4 bp) would require substantial bending in AraC. Hence, it is energetically disfavored for AraC to bind to $araI_1$-$araI_2$ in the absence of arabinose, but energetically favored for AraC to bind to nonadjacent half sites ($araI_1$-$araO_2$) and form a DNA loop (Figure 1.4b). This will, therefore, sterically blocking access of RNA polymerase to the $P_{BAD}$ promoter and holding the basal level of the *araBAD* expression at

a low level (78, 82). Also in the absence of arabinose, an N-terminal arm of ~ 18 amino acids extends from the dimerization domain and binds to the side of the DNA-binding domain. This holds the DNA binding domain relatively rigid to the dimerization domain, therefore favoring binding to $araI_1$-$araO_2$ instead of $araI_1$-$araI_2$ (Figure 1.4a, b) (117).

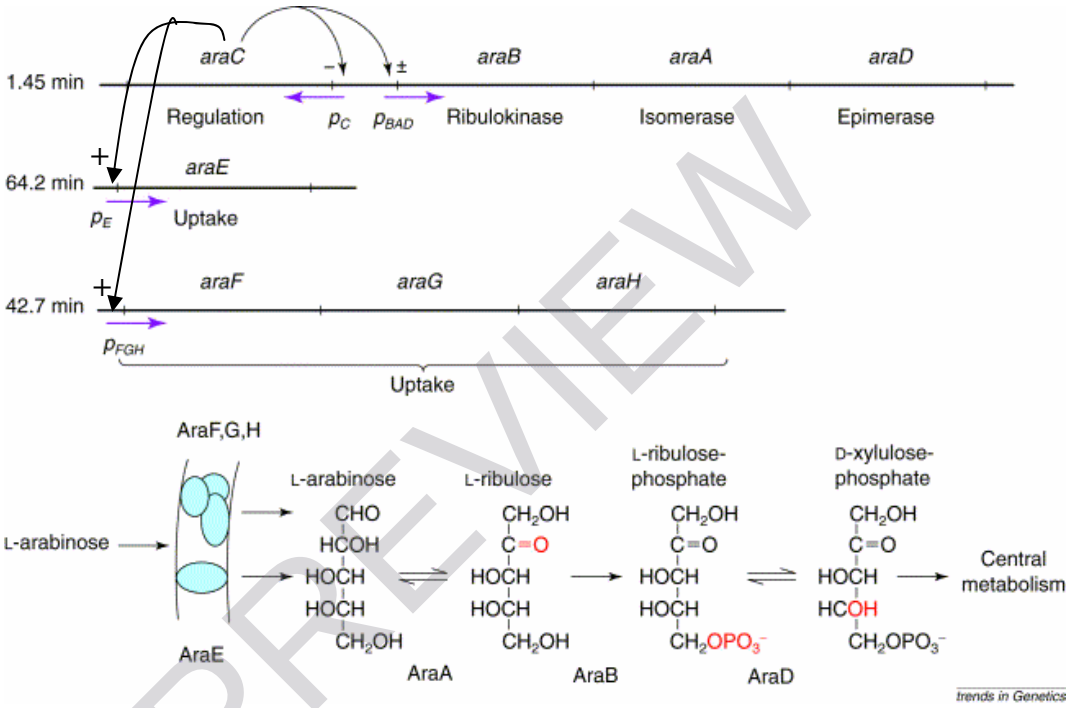



**Figure 1.3 The genes required for the uptake and catabolism of L-arabinose in *Escherichia coli* and the catabolic pathway of L-arabinose (modified from (118)).**

Binding of arabinose to AraC makes the N-terminal arm of AraC fold back to the arabinose-binding pocket (Figure 1.4c), which makes the DNA-binding domain more flexible. Arabinose binding also weakens its binding affinity for $araO_2$ and increases its affinity with $araI_2$, which then makes AraC become unable to contact those two distant sites so the loop is broken (82, 89), therefore releasing the repression of *araBAD* genes.