## NOTICE

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Reproduction in full or in part of this microform is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30, and subsequent amendments.

## AVIS

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

La reproduction, même partielle, de cette microforme est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30, et ses amendements subséquents.

Canada

**Molecular characterization of the *sdaCB* operon**

**in *Escherichia coli* K-12**

Zhongqi Shao

A Thesis

in

The Special Individual

Program  .

Presented in Partial Fulfilment of the requirements

for the Degree of Doctor of Philosophy at

Concordia University

Montreal, Quebec, Canada

June 1993

Canada

**ABSTRACT**

Molecular Characterization of the *sdaCB* Operon

in *Escherichia coli* K-12

Zhongqi Shao, Ph.D.

Concordia University. 1993.

A 5.3-kbp DNA fragment of *E.coli* chromosome around 60.2 minutes has been sequenced. Three complete open-reading frames were carried on this fragment, with *sdaC* and *sdaB* in the same operon. The *sdaB* gene was shown to be very similar to *sdaA*, suggesting that it is the structural gene for the second L-serine deaminase (L-SD2). The nature of the product of the *sdaC* gene and its similarity with TdcC protein indicated that *sdaC* codes for a transport protein. *OrfX*, which is located downstream of the *sdaCB* operon, remains unidentified. The 1.2-kbp *Pvu*II-*Pst*I fragment downstream of *orfX* was found to be identical to *fucO*, a gene in the fucose operon. This establishes the exact map position of the *sdaCB* operon at 60.2 minutes.

The expression of the second L-SD was found to be regulated completely differently from that of the first one (L-SD1). The major transcriptional control governing the *sdaCB* operon is catabolite repression, which does not affect *sdaA* gene expression. The products of *lrp* and *ssd* regulate *sdaA* expression, but do not affect *sdaCB*. The environmental factors

which influence the expression of the first L-SD showed no effect on the second one.

The sdaB mRNA is translated inefficiently, because of its poor ribosome-binding site. A mutation which established synthesis of L-SD2 in minimal medium was demonstrated to be a change in the Shine-Dalgarno sequence of the *sdaB* gene.

## ACKNOWLEDGEMENT

I wish to express my deep appreciation to Dr.E.B.Newman for her guidance, financial support and providing me the best academic support throughout this study. Without this I would never have begun and certainly never have completed this project.

My sincere thanks go also to Dr.C.Cupples and Dr.R.Roy for being on my committee and providing helpful advice and suggestions.

I would like to thank Dr.M.B.Herrington for her help and guidance in my early study at Concordia. I also would like to thank Dr.R.K.Storms and Dr.P.Gulick for advice and for providing strains and plasmids.

My special thanks to H.S.Su and R.T.Lin for wonderful friendship and helpful discussions.

Many thanks to : Y.Shen, J.Zhang, J.Moniakis, M.San Martano, G.Ambartsoumian, J.Basso, Y.Li, K.Shanmugan for help and friendship.

I must also thank Concordia University for providing me a Special Fellowship, a Teaching Assistant Fellowship and an International Student Fee Remission.

Finally, I want to thank my wife, my daughter and my parents for their understanding and support.

**Table of Contents**

viii

# LIST OF FIGURES

# LIST OF TABLES

# INTRODUCTION

L-serine deaminase (L-SD), the *E.coli* enzyme which converts L-serine to pyruvate and ammonia in *E.coli*, was first described by Pardee and Prestidge in 1955 (Pardee and Prestidge, 1955). The activity of L-SD is present in cells grown in glucose-minimal medium and can be induced by glycine or L-leucine, but by not its substrate, L-serine (Pardee and Prestidge, 1955; Newman and Walker, 1982). This activity can be induced to a higher level by growing cells in rich medium (Newman et al., 1982a).

While the physiological roles of L-serine deaminase in the cells are still not understood completely, the activity of this enzyme has been shown to be highly regulated. Two global regulatory proteins, the products of the *ssd* and *lrp* genes, have been shown to be involved in the regulation of expression of L-SD (Newman et al., 1981; Lin et al., 1990). A number of environmental factors also affect the activity of this enzyme (Newman et al., 1982a; Newman, personal communication). Moreover, the activity of L-SD is subject to regulation at the post-transcriptional level. At least three different mutations have been found to cause a decrease in L-SD activity *in vivo*, but activity from the mutant cells could still be seen *in vitro* if activated by iron and dithiothreitol (DTT) (Newman et al., 1985a; Feng, 1990; Su, 1991).

1

Early studies in Dr.Newman's laboratory demonstrated that there are two L-serine deaminases in *E.coli*, which are coded by the *sdaA* and *sdaB* genes (Su & Newman, 1991). The *sdaA* gene has been cloned and sequenced (Su et al., 1989). By constructing an *sdaA::lacZ* fusion protein and showing that the purified fusion protein has both $\beta$-galactosidase and L-serine deaminase activity, *sdaA* was demonstrated to be the structural gene for L-SD1 (Su et al., 1993). The *sdaA* gene is expressed in cells grown in both glucose-minimal medium and LB medium, and is regulated by almost all the mutations or factors which had been shown to affect L-SD activity (Su, et al., 1989). The *sdaB* gene, which codes for L-SD2, is only expressed in LB medium (Su & Newman, 1991). This gene was cloned from a mutant strain in which *sdaB* is expressed in glucose-minimal medium (Su & Newman, 1991). Both L-SD1 and L-SD2 showed very similar biochemical properties (Su & Newman, 1991; Su, 1991).

The *sdaA* gene and its product have been studied in great detail in the last few years. In this thesis, I will focus on the molecular aspects of the *sdaB* gene and its regulation. This work concerns the structure and the regulation of the *sdaB* gene.

**The sequence of the *sdaB* gene and the region surrounding it:** The entire region of the *sdaB* gene has been sequenced as part of this work. The open-reading frame for the *sdaB* gene is determined by its high similarity with the *sdaA* gene. The two proteins showed great similarities not only in biochemical

properties but also in protein structure. Another open-reading frame, which is named the *sdaC* gene and seems to be in the same operon as the *sdaB* gene, is located immediately upstream of *sdaB*. The high similarity of the *sdaC* gene product with the TdcC transport protein, and the map location next to *sdaB*, both suggest that the *sdaC* gene codes for a transport protein for L-serine. The *sdaCB* operon is closely linked to the fucose operon with a possible open-reading frame, *orfX*, between them.

**The regulation of the *sdaB* gene:** This work shows that the expression of the *sdaB* gene is regulated at both transcriptional and translational levels. A mutation in the ribosome-binding site results in expression of *sdaB* in glucose-minimal medium. Transcription of the *sdaB* gene is induced in rich medium and is under catabolite repression. The activity of SdaB, unlike SdaA, is not induced in the *ssd* or *lrp* mutant cells and is not affected by the environmental factors which had been shown to affect the activity of SdaA protein.

In this part of the thesis, I will first provide the background of L-serine metabolism in *E.coli* and then review the early studies on L-serine degradation. Since L-SD has been suggested to function as a catabolic enzyme, I will also discuss the studies of catabolite repression in *E.coli*. Early evidence has implied that *sdaA* and *sdaB* might be two genes with high similarity, I will therefore review some related genes in *E.coli*.

3

**Part 1: L-serine metabolism in *E.coli***

## 1.1. The L-serine biosynthesis pathway in *E.coli*

L-serine is synthesized in *E.coli* cells from 3-phosphoglycerate, an intermediate in the Embden-Meyerhof pathway (Fig.1). Three enzymes are involved in L-serine synthesis, coded by the *serA*, *serC* and *serB* genes listed in the order in which the enzymes operate *in vivo*. 3-phosphoglycerate is first oxidized to 3-phosphohydroxy-pyruvate by 3-phosphoglycerate dehydrogenase, followed by transamination to 3-phosphoserine by 3-phosphoserine aminotransferase and dephosphorylation to L-serine by 3-phosphoserine phosphatase.

## 1.2. The regulation of L-serine synthesis

The synthesis pathway of L-serine is highly regulated in the cell. The first enzyme in this pathway is inhibited by both L-serine and glycine, with 50 % feedback inhibition occurring at the concentration of $4 \times 10^{-5}$M for L-serine and $4.8 \times 10^{-3}$M for glycine (Pizer, 1963).

Transcription of the *serA* gene is repressed by L-leucine and several other amino acids which are not directly related to serine biosynthesis. Repression also occurs during growth in rich medium (McKitrick & Pizer, 1980). The repression by L-leucine occurs by interaction of L-leucine and a regulatory protein, Lrp, with the promoter of the *serA* gene (Lin et al.,

4

Fig.1  L-serine metabolism in *E.coli*

L-serine is synthesized from 3-phosphoglycerate, an intermediate in the Embden-Meyerhof pathway. It is the precusor of glycine and C1 units via serine hydroxymethyltransferase and is also a precusor of other amino acids. This diagram is adapted from Stauffer, 1987.

Abbreviations: SerA, 3-phosphoglycerate dehydrogenase; SerC, 3-phosphoserine aminotransferase; SerB, 3-phosphoserine phosphatase; SHMT, serine hydroxymethyltransferase; GCV, glycine cleavage enzymes; L-SD, L-serine deaminase; 5,10-mTHF, 5,10-methylenetetrahydrofolate.

Fig.1  L-serine metabolism in *E.coli*

6
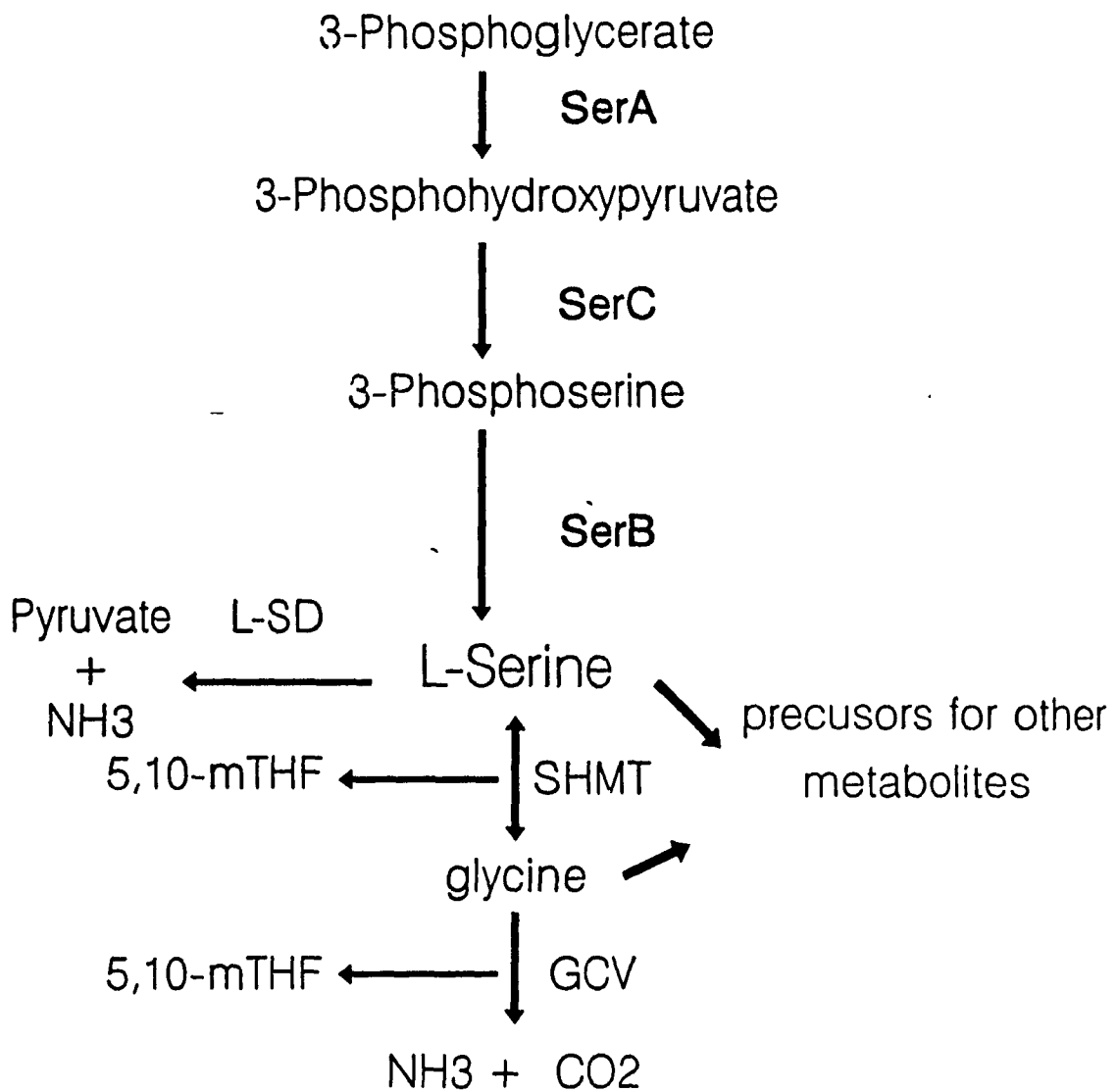
1990; Lin, 1992). Because Lrp activates expression of *serA*, expression of the *serA* gene is decreased more than 5-fold in an *lrp* mutant (Lin et al., 1990). The *lrp* mutant cells still manage to grow without adding exogenous L-serine. However, the decreased ability to synthesize serine is seen in the fact that the *lrp* mutant cell does not grow at 42°C in glucose-minimal medium but does grow on addition of exogenous serine or with a plasmid carrying the *serA* gene in the strain (Lin, 1992).

The transcription of the *serA* gene also seems to be under the control of another global regulatory protein, the product of the *ssd* gene. This is indicated by the fact that 3-phosphoglycerate dehydrogenase activity is much lower in the *ssd* mutant cell (Newman et al., 1981), as is β-galactosidase activity from cells containing a *serA::lac* gene fusion. Furthermore, β-galactosidase activity from the *serA::lacZ* fusion is extremely low in a double mutant cell deficient in both *ssd* and *lrp* as compared with either single mutant. Indeed, the expression of the *serA* gene is so low in the *ssd*, *lrp* double mutant that it grows very poorly in glucose-minimal medium. Addition of exogenous serine improves the growth of the *ssd*, *lrp* double mutant significantly (unpublished observation; Ambartsoumian et al.).

## 1.3. Glycine as a source of L-serine

L-serine is the only source of glycine in *E.coli* (Newman and Magasanik, 1963, Pizer, 1965). Synthesis of glycine from L-serine is catalysed by the enzyme serine hydroxymethyl-transferase (SHMT), the product of the *glyA* gene (Stauffer et al., 1981). This reaction is also the major source of C1 units (Mudd & Cantoni, 1964). The expression of the *glyA* gene is regulated by the end-products of pathways which use C1 units in the biosyntheses they catalyze, and by L-serine and glycine (Miller & Newman, 1974).

An L-serine auxotroph can use exogenous glycine as its source of L-serine, in a reaction catalysed by the same enzyme which makes glycine from L-serine (Stauffer & Brenchley, 1974). This reaction will also need a C1 unit, often produced by cleavage of glycine by the glycine cleavage enzymes (Mudd & Cantoni, 1964). That this is the major source of C1 units other than serine itself is shown by the fact that an L-serine auxotroph can not use glycine as a source of L-serine when glycine cleavage is blocked (Newman et al., 1974).

## 1.4. Formation of glycine from L-threonine

While it seems that L-serine is the only source of glycine in wild-type cells, *E.coli* can also derive its glycine from L-threonine through a cryptic pathway. There are two enzymes involved in this pathway, threonine dehydrogenase(TDH) and $\alpha$-amino-ketobutyrate ligase (Chan & Newman, 1981).

8

Early evidence showed that the formation of glycine from L-threonine occurs only in the presence of exogenous L-threonine (Roberts et al., 1955). However, the presence of exogenous L-threonine itself does not support the growth of a glycine auxotroph unless exogenous L-leucine is also present (Fraser & Newman, 1975). L-leucine was shown to increase the expression of threonine dehydrogenase (Newman et al., 1976). Regulation of threonine dehydrogenase also involves the Lrp protein, which represses the expression of threonine dehydrogenase with L-leucine releasing this repression (Lin, et al; 1990). The activity of threonine dehydrogenase is also increased in the ssd mutant cells (Lin et al., 1990).

## 1.5. L-serine as a biosynthetic precursor

The L-serine biosynthetic pathway is one of the major metabolic pathways in *E.coli* cells. It has been estimated that 15% of the carbon from glucose passes through the pathway (Pizer & Potochny, 1964).

Besides L-serine, this pathway is the only source of glycine and the major, if not only, source of one-carbon units. The importance of this pathway is also seen in that L-serine, glycine and C1 units are the precursors for a number of metabolites.

The complete L-serine molecule is incorporated into indole glycerolphosphate to make L-tryptophan (Yanofsky, 1960). All three carbons of cysteine are derived from L-serine

(Kredich & Tomkins, 1966). L-serine is also needed for synthesis of L-methionine, since L-cysteine is the direct precursor for L-methionine (Tran et al., 1983). L-methionine is involved in the synthesis of S-adenosylmethionine (SAM) which is a C1 donor for numerous other methylation reactions (Mudd & Cantoni, 1964).

Glycine is incorporated into purines and heme-containing compounds (Pizer & Potochny, 1964).

There are two steps in this pathway which produce C1 units, glycine synthesis by SHMT and glycine cleavage (Fig.1). C1 units are involved in purine, thymine, histidine, methionine and fMet-tRNA biosynthesis (Blakley, 1969).

## 1.6. L-serine degradation through L-SD

L-serine deaminase converts L-serine to pyruvate and ammonia (Pardee & Prestidge, 1955). Despite continuing studies since its first discovery almost 40 years ago, the physiological role of L-serine deaminase in the cell still remains unclear. The synthesis of L-serine requires an expenditure of energy. Therefore it would be strange for cells to degrade L-serine without some important reason. This may be why the activity of L-SD is highly regulated in the cell.

Two possible physiological functions for L-serine deaminase have been suggested in the past. First, since a high concentration of L-serine in the cells has been known to be toxic (Cosloy & McFall, 1970; Hama et al., 1990; Uzan &

10

Danchin, 1978), L-serine deaminase might function to regulate the L-serine pool in the cell (Su, 1991). This has been proven to be true at least insofar as cells with deficient L-SD activity have been shown to be more sensitive to exogenous L-serine than wild-type cells (Su, 1991). Second, the products of L-serine deamination, pyruvate and ammonia, can be used as carbon and nitrogen sources. Therefore, L-serine deaminase may function as a catabolic enzyme to convert excess L-serine to carbon and nitrogen sources (Su & Newman, 1989; Su, 1991).

**Part 2: Regulation of L-SD expression in *E.coli***

**2.1. Environmental factors which affect L-SD activity**

L-SD is present in *E.coli* cells grown in glucose-minimal medium, but at low level. This activity can be induced slightly by addition of glycine and significantly by addition of L-leucine. The combination of glycine and L-leucine increases the L-SD activity to a still higher level (Pardee & Prestidge, 1955). L-SD activity in cells grown in LB medium is even higher (Newman et al., 1982a).

L-SD activity has been shown to be induced by several other environmental factors. Cells exposed to DNA damaging agents, such as ultraviolet irradiation and nalidixic acid, or to alcohol shock, as well as those grown at high temperature, or in anaerobic conditions all show increased L-SD activity (Newman et al., 1982a; Newman, personal communication).

11

## 2.2. High L-SD activity in the *ssd* mutant

The *ssd* mutation has been shown to be a pleiotropic mutation, which leads to many changes in the cell, such as decreased growth rate and yield; inability to use succinate or α-ketoglutarate as a carbon source; resistance to certain antibiotics and colicin; deficiency in transport of some amino acids; extreme sensitivity to UV irradiation and inability to grow anaerobically (Newman et al., 1981).

The *ssd* mutant cell grows well on L-serine as carbon and energy source and L-SD activity in the mutant cells is 7-fold higher than that in the wildtype cells (Newman et al., 1981). Besides L-serine deaminase activity, at least two other enzyme activities are altered in the *ssd* mutant cells. L-threonine dehydrogenase (TDH) activity is increased 8-fold (Lin et al., 1990), and 3-phosphoglycerate dehydrogenase activity is significantly decreased in *ssd* mutant cells (Newman et al., 1981).

The nature of the *ssd* gene and its product still largely remains a mystery. It has been suggested that *ssd* is the same gene as *cpxA*, a gene which codes for a membrane sensor protein involved in signal transduction (Rainwater & Silverman, 1990). This is mainly based on the fact that *ssd* and *cpxA* mutations were mapped at the same position on the *E.coli* linkage map, and that the *cpxA* plasmid complements the *ssd* mutation (Lin, personal communication; Rainwater & Silverman, 1990). However, the *ssd* and *cpxA* mutants do show some difference in

12

phenotypes; indeed the *cpxA* mutant does not have increased L-SD activity (unpublished observation).

## 2.3. The *lrp*/leucine regulon

The *lrp* gene codes for the leucine-responsive regulatory protein (Lrp). *E.coli* cells coordinate the expression of many genes according to the presence of exogenous L-leucine through Lrp. The group of genes regulated by the Lrp protein and L-leucine constitute the Leucine/Lrp regulon (Lin et al., 1990).

The activity of L-SD is greatly induced by L-leucine or in the *lrp* mutant cells (Lin et al., 1990). Lrp, therefore, likely functions as a repressor for the structural gene of L-SD. L-leucine releases the repression by interacting with the Lrp protein, since leucine induction is not seen in the *lrp* mutant cells (Lin et al., 1990; Lin, 1992).

In the *ssd* and *lrp* mutant cells, both isolated by selecting cells using L-serine as sole carbon source, the expression of L-serine degradation enzyme(s) is increased while the activity of enzymes for L-serine synthesis is decreased. Therefore, the products of the *ssd* and *lrp* genes seem to be two of the major proteins involved in the regulation of L-serine metabolism.

## 2.4. L-SD is subject to post-translational modification

The fact that there is no L-SD activity detectable in extracts frozen at -20°C for more that 24 hours unless iron

13

and DTT were added during the *in vitro* assay (Newman et al., 1985a; Newman & Walker, 1990), suggests that L-SD is not stable. It also tells us that L-SD can exist in a relatively more stable inactive form and then be activated by iron and DTT.

Several new types of L-SD deficient mutants have been isolated in the past few years. As expected for L-SD deficient strains (Newman et al., 1985b), these mutants could not grow on L-serine even in the presence of L-leucine and glycine, and showed no L-SD activity *in vivo*. However in three mutants, MEW128, MEW191 and MEW84, L-SD activity in the extracts could be activated by iron and DTT (Newman et al., 1985a; Feng, 1990), suggesting that L-SD was made in the mutant cells but was in an inactive form. This suggests that post-translational modification is needed to convert L-SD to a functional form.

**Part 3: *sdaA*: the gene which codes for L-SD1**

**3.1. Isolation of mutants deficient in L-SD both *in vivo* and *in vitro***

Two types of mutants with deficient L-SD activity had been isolated by screening mutants which are unable to grow on L-serine in the presence of L-leucine and glycine. Mutants of the first type, mentioned above, are considered to be the activation-deficient mutants, strains with deficient L-SD activity *in vivo*, but with demonstrable activity *in vitro*

14

(Newman et al., 1985a). Mutants of the second type do not show L-SD activity either *in vivo* or *in vitro*, suggesting that they may carry mutations in the structural gene for L-SD (Newman et al., 1985b; Su et al., 1989).

One of the second type of mutant was isolated by H.Su using λplacMu9 insertion and was studied in detail (Su et al., 1989). The L-SD assay showed that this λplacMu9 insertion mutant had no L-SD activity when it was grown in glucose-minimal medium with or without L-leucine and glycine (Su et al., 1989). It also had no L-SD activity even in the *ssd* or *lrp* mutant background, in which L-SD activity is highly induced in wildtype cells (Su et al., 1989). The mutation, called *sdaA*, was mapped at 40.1 min (Su & Newman, 1991).

## 3.2. Cloning and sequencing of the *sdaA* gene

The *sdaA* gene was cloned by the miniMu *in vivo* cloning method (Su et al., 1989). The cloned *sdaA* gene complemented the *sdaA* mutation and showed extremely high L-SD activity both *in vivo* and in extracts. Hybridization using a cloned *sdaA* gene as probe showed that this gene is indeed disrupted by the insertion in the *sdaA* mutant cells (Su et al., 1989).

It was originally predicted from the DNA sequence containing the *sdaA* gene that the *sdaA* open reading frame starts with ATG and codes for 448 amino acids (Su et al., 1989). This result was later revised to use a GTG as start codon 6 codons further upstream from ATG. The revised *sdaA*

15

sequence codes for 454 amino acids and this sequence is supported by the results of the Edman degradation of the N-terminal sequence of purified L-serine deaminase (Su et al., 1993).

The upstream sequence of the *sdaA* gene as determined by H.Su was found to be very similar to the sequence downstream of the *pabB* gene of *Klebsiella aerogenes* (Su, 1991). The sequence around the *pabB* gene in *E.coli* has also been sequenced and shown to be very similar to the sequence in *K.aerogenes*. The *pabB* sequence stopped at a *Sal*I site and the sequence by Su started from a *Sal*I site (Su et al., 1989). A comparison with the sequence in *K.aerogenes* suggested that the *Sal*I site in *pabB* and *sdaA* sequences was the same one (Su, 1991). This locates the *sdaA* gene downstream of the *pabB* gene with another open reading frame between the two genes (Su, 1991). This agrees with the fact that the *pabB* and *sdaA* genes were mapped at same position on the *E.coli* chromosome by conventional mapping (Goncharaff & Nichols, 1984; Su & Newman, 1991).

### 3.3. *sdaA* is the structural gene for L-serine deaminase

To prove that *sdaA* codes for L-serine deaminase, Su constructed an *sdaA-lacZ* in-frame gene fusion *in vitro* (Su et al., 1993). β-galactosidase was then purified by affinity chromatography from the cells containing a plasmid carrying this *sdaA-lacZ* fusion. Purified protein showed both β-

galactosidase activity and L-serine deaminase activity, proving that *sdaA* indeed codes for L-serine deaminase (Su et al., 1993).

The result of the Edman degradation of the purified L-SD-$\beta$-galactosidase fusion protein confirms that the open reading frame for the *sdaA* gene starts at the GTG site (Su et al.,1993).

## 3.4. Regulation of *sdaA* gene expression

The expression of L-SD activity was known to be affected by mutations and some other factors. Studies on the regulation of the *sdaA* gene expression were carried out by measuring $\beta$-galactosidase activity from the *sdaA*::$\lambda$placMu9 fusion and *sdaA*::*lacZ* fusion constructed *in vitro*. It was found that the products of both the *ssd* and *lrp* genes regulate *sdaA* expression at the transcriptional level (Su et al., 1989; Lin, 1992). Among the environmental factors, at least anaerobic condition and UV irradiation have been shown to affect *sdaA* transcription (Su, 1991).

The mechanism of Lrp protein regulation at the *sdaA* promoter was studied by R.T.Lin in considerable detail. It was found that Lrp binds to at least two sites upstream of *sdaA*, one with high affinity and the other with low affinity. The presence of L-leucine greatly reduced Lrp binding activity at the *sdaA* promoter (Lin, 1992). Therefore, it is likely that Lrp represses *sdaA* transcription by binding to its upstream

17

region, while L-leucine activates *sdaA* expression by releasing the binding of Lrp.

## 3.5. Biochemical properties of the SdaA protein

As discussed earlier, it seems that L-SD needs to be activated from an inactive form after translation. At least three genes have been found to be involved in this post-translational activation process *in vivo* (Newman et al., 1985a; Feng, 1990). In those mutant cells, there is no L-SD1 activity *in vivo*, but the activity can be seen *in vitro* once activated by adding iron and DTT. Protein folding was suggested to be involved in the process of post-translational activation of L-SD1 (Moniakis, 1993).

The active form of L-SD1 is very unstable *in vitro*, while the inactive form of L-SD is much more stable (Newman et al., 1985a). It was found that the inactive form of the protein can be activated with both β-mercaptoethanol and DTT (Moniakis, 1993). It was also found that the active form of L-SD can assume a second inactive form that can only be activated with DTT at a 10-times higher level than under normal activation conditions (Moniakis, 1993). Iron and DTT are known to cleave proteins (Kim et al., 1985). It had been suggested that the activation of L-SD involved a cleavage at this protein, possibly at a serine-serine bond (Su et al., 1989). Alternatively, a conformational change may be involved in the activation of SdaA protein (Moniakis, 1993).

**Part 4:** *sdaB,* **the gene which codes for L-SD2**

### 4.1. Existence of a second L-SD

Several facts suggested that there was another gene, besides *sdaA,* in *E.coli* which coded for an L-serine deaminase. First, when cloned *sdaA* was used as a probe to hybridize to chromosomal DNA from the *sdaA* insertion mutant, a minor band was observed in addition to those bands corresponding to the disrupted *sdaA* gene (Su et al., 1989), which implied that there is another gene which is very similar in nucleotide sequence to the *sdaA* gene. Second, there is still a considerable amount of L-serine deaminase activity in LB-grown cells carrying a null mutation in the *sdaA* gene (Su & Newman, 1991), which suggested that there is another L-serine deaminase (L-SD2) which is expressed in LB medium, but not in minimal medium.

As summarized below, L-SD2 is probably coded by the gene, called *sdaB,* which is very similar to the *sdaA* gene.

### 4.2. A mutant with altered *sdaB* expression

If there really were an *sdaB* gene, which codes for L-SD2 and is expressed in LB medium, one would expect that this gene may be turned on in glucose-minimal medium by changing its regulation. This was proven to be true.

H.Su isolated a mutant which grew in minimal medium on L-serine in the presence of L-leucine and glycine from a strain

carrying a insertion in *sdaA* gene (Su & Newman, 1991). This mutation, which resulted in expression of L-SD2 in minimal medium, could be in the upstream regulatory region of the *sdaB* gene or in a regulatory gene for *sdaB*. This mutation, named *sdaX*, was mapped at 60.1 min (Su & Newman, 1991).

## 4.3. Isolation of an *sdaB* mutant and cloning of the *sdaB* gene

An *sdaB* mutant was isolated by λplacMu9 insertion into a strain carrying both *sdaA*::*Cm* and *sdaX* mutations and selection of cells which lost the ability to grow on L-serine (Su & Newman, 1991). This triple mutant, *sdaA*, *sdaB*, and *sdaX*, has no L-serine deaminase activity whether in glucose-minimal medium or in LB medium (Su & Newman, 1991). Similarly, it showed no L-SD activity either *in vivo* or *in vitro* whether cells were grown in minimal medium or in LB (Su & Newman, 1991). The insertion has been mapped at the same position as the *sdaX* mutation, suggesting that *sdaB* and *sdaX* might be the same gene (Su & Newman, 1991).

The *sdaB* gene was then cloned from the strain carrying *sdaA*::*Cm* and *sdaX* mutations by selecting on L-serine-minimal medium with L-leucine and glycine after transforming the *sdaA*::*Cm* *sdaB*::λplacMu9 double mutant cells (Su & Newman, 1991). The cloned *sdaB* gene allowed the *sdaA* *sdaB* double mutant to grow on L-serine in the presence of L-leucine and glycine, and even on L-serine alone (Su & Newman, 1991). An L-SD assay showed that the activity of the *sdaA*, *sdaB* double

20

mutants carrying pBR322 *sdaB* gene was 5-6 fold higher than that seen in wildtype cells when grown in glucose-minimal medium. The activity was even higher in LB medium (Su & Newman, 1991).

Hybridization to the chromosomal DNA using the cloned *sdaB* gene as probe showed that this gene is indeed disrupted by λplacMu9 insertion in the *sdaB* mutant cells (Su & Newman, 1991). The fact that both *sdaB* and *sdaX* were mapped at 60.1 min and that the cloned *sdaB* gene showed activity in both minimal medium and LB medium suggested that the cloned DNA fragment might carry a mutated *sdaB* gene or both the *sdaB* gene and a mutated *sdaX* gene.

## 4.4. L-SD2 is similar in biochemical properties to L-SD1

The results of a preliminary characterization of L-SD2 showed that this enzyme is extremely similar in biochemical properties to that of L-SD2. Both enzymes showed no activity at low pH (below pH6), but showed activity from pH6 up to pH10 (Su, 1991). An unusually high substrate concentration is required to demonstrate both L-SD1 and L-SD2 activity (Su, 1993).

As is the case for L-SD1, iron and DTT are needed for activation of L-SD2 *in vitro* and probably the same *in vivo* activation system functions on both enzymes (Su & Newman, 1991).

## Part 5. Other proteins which deaminate L-serine

Besides the two enzymes mentioned above, L-SD1 and L-SD2, at least three other enzymes have been shown to deaminate L-serine. Among them, are two enzymes, each known as L-threonine deaminase, coded by the *ilvA* gene and the *tdcB* gene respectively (Lawther et al., 1987; Datta et al., 1987).

The product of the *ilvA* gene, the biosynthetic L-threonine deaminase (biosynthetic L-TD), catalyses the first step in the L-isoleucine synthesis pathway (Calhoun et al., 1973). An *ilvA* mutant can not produce $\alpha$-ketobutyrate for the synthesis of L-isoleucine by deaminating L-threonine and therefore needs exogenous L-isoleucine to grow (Umbarger, 1987). The activity of the biosynthetic L-TD is inhibited by the end product in the pathway, so that it can not work as a biodegradative enzyme. This enzyme can also deaminate L-serine, at least *in vitro* (Umbarger, 1957).

The *tdcB* gene product, the biodegradative L-threonine deaminase (biodegradative L-TD), is known by habit and custom as an L-threonine deaminase. This enzyme degrades both L-serine and L-threonine (Goss & Datta, 1984). The expression of the *tdcB* gene is induced under anaerobic culture conditions in rich medium or in minimal medium containing threonine, serine, valine and isoleucine, together with cyclic AMP and fumarate (Datta et al., 1987). The *tdcB* gene has been cloned and sequenced, and has been shown to be part of a polycistronic

22

operon containing three genes, *tdcABC* (Goss et al., 1988). The *tdcA* gene product was thought to be a trans-acting positive activator for autogenous regulation of the *tdc* operon, while the *tdcC* gene product was demonstrated to be a membrane-associated threonine-serine permease (Sumantran et al., 1990). The product of the *tdcR* gene, located immediately upstream of the *tdcABC* operon but in the opposite transcriptional direction, is needed for efficient expression of the operon (Schweizer & Datta, 1989).

Moreover, *E.coli* makes a similarly regulated enzyme, D-serine deaminase, which shows *in vitro* at least some activity against L-serine (Federiuk et al., 1983). D-serine deaminase is encoded by the *dsdA* gene, mapped at 51 min. The *dsdC* gene, which codes for an activator protein for *dsdA*, is located immediately upstream of *dsdA* but transcribed divergently (Bornstein-Frost et al., 1987). The activation by DsdC protein of the *dsdA* gene was shown to involve a *cis*-acting mechanism (McFall, 1986).

## Part 6. Catabolite repression

The presence in the medium of glucose, a rapidly metabolizable carbon source, will inhibit the transcription of genes coding for other carbon catabolic enzyme systems and related proteins (Magasanik, 1970). This phenomena is called catabolite repression.

23

Catabolite repression is achieved in *E.coli* through the cyclic AMP/cyclic AMP receptor protein (CRP) system (Emmer et al., 1970). Crp protein has been shown to be the positive regulatory protein for many carbon catabolic enzyme systems, such as the lactose, galactose and arabinose operons. Activation by Crp protein needs cyclic AMP as coeffector. cAMP synthesis is inhibited in the presence of glucose (Makman et al., 1965).

Crp has been demonstrated to be a DNA binding protein, which binds to the promoter region of the genes it regulates. The consensus sequence for Crp binding has been identified (de Crombrugghe et al., 1984). In order to bind to DNA, the Crp protein undergoes a conformational change, which is achieved by cAMP binding to Crp (Krakow, 1975).

One of the functions suggested for L-serine deaminase in cells is that L-serine will be used as carbon source when it is plentiful (Su, 1991). Therefore, the expression of *sdaA* and *sdaB* might be subject to catabolite repression.

**Part 7. Genes with high similarity in *E.coli***

**7.1. Some examples of genes with high similarity**

There are several sites on the *E.coli* chromosome for *rrn* loci which code for rRNA and tRNA. Since these show very high degree of similarity, one might expect frequent exchanges among the *rrn* loci. In fact, the organizations of genes are

very stable and rearrangements are rare (Riley & Krawiec, 1987). This is also seen for genes coding for proteins. That pairs or groups of structural genes, which are mapped at different sites on the chromosome, are very similar in the nucleotide sequence or in the amino acid sequence of their products is in fact quite common in *E.coli*. Table 1 lists some examples.

In some cases, the similarity among these genes or their products is extremely high through the whole sequence. The two EF-Tu proteins, coded by *tufA* and *tufB* respectively, vary by only one amino acid residue (An & Friesen, 1980). In some other cases, the similarity between related genes or their products can be very high in some regions, but much lower in other regions. An example of this is seen in the case of the four *E.coli* genes, *tar*, *tap*, *tsr* and *trg*. The level of similarity between these proteins is low (10 to 60%) at the amino end and high (60 to 100%) at the carboxyl end (Bollinger et al., 1984).

Levels of similarity among related structural genes coding for enzymes vary significantly. The amino acid sequences are usually more conserved than the nucleotide sequences. The two lysyl-tRNA synthetases, coded by *lysU* and *lysS* respectively, showed 70% identity of nucleotide sequences and 85% identity of amino acid sequences (Clark & Neidhardt, 1990).

TABLE 1. Examples of closely related genes

| Genes | | Reference(s) |
|---|---|---|
| *tufA*, *tufB* | ----------------- | An & Freisen, 1980 |
| *narZYWV*, *narGHJI* | ----------------- | Blasco et al., 1990 |
| *lysU*, *lysS* | ----------------- | Clark & Neidhardt, 1990 |
| *argF*, *argI*, *pyrB* | ----------------- | Houghton et al., 1984 |
| | | Van Vliet et al., 1984 |
| *aroF*, *aroG* | ----------------- | Davies & Davidson, 1982 |
| *hisI*, *hisT* | ----------------- | Higgins et al., 1982 |
| *metB*, *metC* | ----------------- | Belfaiza et al., 1986 |
| *trp(G)D*, *pabA* | ----------------- | Horowitz et al., 1982 |
| | | Kaplan & Nichols, 1983 |
| *trpE*, *pabB* | ----------------- | Goncharoff & Nichols, 1984 |
| | | Nichols et al., 1981 |
| *tar*, *tap*, *tsr*, *trg* | ----------------- | Bollinger et al., 1984 |

*sdaA* and *sdaB* were shown to be similar by hybridization (Su & Newman, 1991), suggesting that they might be two genes derived from the same ancestral gene. Direct nucleotide sequence comparison of *sdaA* and *sdaB* would tell us what is the exact level of similarity between the two genes and where the conserved regions are. These comparisons might help to answer some questions on the structure of L-SD.

## 7.2. Advantages to the cell of proteins with similar or identical function

Normally, the products of genes which are very similar in their nucleotide sequences, have similar functions, e.g. the products of *sdaA* and *sdaB*. To have two or more genes coding for proteins which have the same or similar function will give cells several advantages in adapting to a variable environment.

First, there may be not enough product from one copy of the gene to meet the needs of cells in certain circumstances, in which case multiple copies of the gene might be needed. Second, it provides another way to regulate the expression of some proteins. That is, cells might have several differently regulated genes each coding for a protein of similar function. This is true for many pairs or groups of genes that are similar only inside the open reading frames. Their upstream regulatory regions are very different and have been shown to lead to different regulation of expression of the genes.

27

## 7.3. Origin of genes with homology

It has been believed that genes with homology arose by duplication and subsequent mutation of an ancestral gene (Riley & Krawiec, 1987). Except for some multicopy genes which are almost identical, the level of similarity among the related genes will reflect the time when they arose (Riley & Krawiec, 1987).

The products of the duplicated genes may still have the same function as the ancestral gene, or they may have different functions. In this way, duplication provides cells an economical way to produce new genes.

It also should be noted that similar function among proteins does not necessary signify nucleotide similarities in the corresponding genes.

## MATERIALS AND METHODS

### 1. Strains, bacteriophages and plasmids

The bacterial strains used in this study, all derivatives of E.coli K-12, are described in Table 2, as are bacteriophages, and plasmids.

### 2. Media and growth conditions

#### 2.1. Glucose minimal medium

0.527% $KH_2PO_4$, 1.500% $K_2HPO_4$, 0.2% $(NH_4)_2SO_4$, 0.020% $MgSO_4$, 0.001% $CaCl_2$. 0.8% gelrite was added for solid medium.

0.2% sterile D-glucose was added after autoclaving. 50 $\mu$g/ml L-isoleucine and L-valine were added for strain MEW1 and its derivatives, since these strains carry a deletion in the ilvA gene.

#### 2.2. NSIV medium

Same medium as described above, but 0.2% L-serine was added to replace D-glucose as carbon source.

#### 2.3. LB medium

1.0% Bacto-tryptone, 0.5% Bacto-yeast extract, 0.5% NaCl. 1.8% Bacto-agar was added for solid medium.

**Table 2.**

**Bacterial strains, phages and Plasmids**

| strains | Genotype and/or relevant characteristics | Source or reference |
|---|---|---|
| A401 | *polA* | Russel & Holmgrem 1988 |
| A402 | A401 Δ*lac* | J.Bosso |
| CU1008 | *ilvA* | L.S.Williams |
| MEW1 | CU1008 Δ*lac* | Newman et al., 1985a |
| KEC9 | CU1008 *ssd* | Newman et al., 1982a |
| MEW26 | MEW1 *lrp*::Tn*10* | Lin et al., 1990 |
| MEW22 | MEW1 *sdaA*::λplacMu9 | Su et al., 1989 |
| MEW28 | MEW1 *sdaA*::*Cm* | Su et al., 1989 |
| MEW51 | MEW28 *sdaB*::λplacMu9 | Su & Newman, 1991 |

Cont.

| | | |
|---|---|---|
| MEW60 | MEW1 *sdaX sdaB*::$\lambda$placMu9 | Su & Newman, 1991 |
| SA2777 | *crp*::*Cm* | S.Adhya |
| CAG18477 | *metF*159 *zij*-501::Tn*10* | Singer et al., 1989 |
| CAG18079 | *fuc*-3072::Tn*10* | Singer et al., 1989 |
| CJ236 | *dut ung* /pCJ105(*Cm*) | Bio-Rad |
| MV1190 | *recA* (F' *lacI*$^Q$) | Bio-Rad |
| XL1-Blue | *recA$^-$ lac$^-$ hsdR$^-$* (F' *proAB lacI*$^Q$ *lacZ* M15, Tn*10*) | Stratagene Co. |
| MEW86 | MEW1 *ssd zij*-501::Tn*10* | This work |
| MEW87 | MEW1 *sdaB*::$\lambda$placMu9 | This work |
| MEW88 | MEW1 *sdaB*::*lacZ* | This work |
| MEW89 | MEW28 *lrp*::Tn*10* | This work |
| MEW90 | MEW86 *sdaB*::$\lambda$placMu9 | This work |
| MEW91 | MEWE87 *lrp*::Tn*10* | This work |
| MEW92 | MEW87 *crp*::*Cm* | This work |
| MEW93 | MEW22 *crp*::*Cm* | This work |
| MEW94 | MEW86 *sdaA*::*Cm* | This work |
| CP41 | MEW1 *sdaC*::$\lambda$placMu9 | Lin et al., 1992 |

Cont.

| | | |
|---|---|---|
| CP52 | MEW1 *sdaC*::λplacMu9 | Lin et al., 1992 |
| MEW95 | CP41 *lrp*::Tn*10* | This work |
| MEW96 | CP52 *lrp*::Tn*10* | This work |
| MEW97 | CP41 *sdaA*::*Cm* | This work |
| MEW98 | CP52 *sdaA*::*Cm* | This work |
| MEW99 | CP41 *crp*::*Cm* | This work |
| MEW100 | CP52 *crp*::*Cm* | This work |

**Phages**

| | | |
|---|---|---|
| M13KO7 | | Pharmacia |
| λ 8C5 | λ Kohara phage 457 | Kohara |
| P1 *vir* | | E.McFall |

**Plasmids**

| | | |
|---|---|---|
| pBR322 | | Bolivar 1977 |
| Bluescript | KS- and KS+ | Stratagene Co. |
| pMC1871 | a truncated *lacZ* gene carried on pBR322 | R.K.Storms |

32

Cont.

| | | |
|---|---|---|
| pMES41 | pBR322 with an 8.4-kbp *Pst*I insert from the *sdaX* mutant | Su et al., 1991 |
| pMES60 | pBluescript with 7.2-kbp *Hind*III insert from pMES41 | This work |
| pMES61 | pBluescript with 4.2-kbp *Hind*III-*Pst*I insert from pMES60 | This work |
| pWTB6 | pBluescript with 4.2-kbp *Hind*III-*Pst*I insert from Kohara phage 457 | This work |
| pWTB7 | pBR322 with 4.2-kbp *Bam*HI-*Hind*III insert from pWTB6 | This work |
| pMES63 | pBR322 with 6.8-kbp *Sal*I insert from pMES41 | This work |
| pMES64-66 | pBR322 carrying various lengths of the 6.8-kbp *Sal*I insert from pMES63. | This work |
| pMES67 | pBluescript with 8-kbp insert of pMES64 | This work |
| pMES68-69 | pBluescript carrying various lengths of the 8-kbp insert from pMES67 | This work |

Cont.

| pMES70 | pBluescript with 1.1-kbp HindIII insert from pMES41 | This work |
| pMES71 | pBluescript carrying in frame fusion of the *sdaB::lacZ* gene | This work |

**2.4. 2X YT medium**

1.6% Bacto-tryptone; 1.0% Bacto-yeast extract; 0.5% NaCl.

**2.5. Mix medium for plasmid isolation**

Minimal medium as described above; 0.2% glycerol instead of D-glucose; 0.5% Bacto-tryptone; 0.25% Bacto-yeast extract; 0.5% NaCl.

**2.6. Medium for growth of P1 phage**

1.0% Bacto-tryptone, 0.1% Bacto-yeast extract, 0.8% NaCl.

1.7% Bacto-agar for plate and 0.6% Bacto-agar for top agar. $CaCl_2$ and glucose were added to the top agar to the concentration of 2 mM and 0.1% respectively before use.

**2.7. Other additions to the media**

Ampicillin was added to a final concentration of 100 $\mu$g/ml; tetracycline, 15 $\mu$g/ml; kanamycin, 80 $\mu$g/ml and chloramphenicol, 25 $\mu$g/ml.

**3. Buffers and solutions**

MC buffer: 0.100 M $MgSO_4$, 0.500 M $CaCl_2$

Phosphate buffer for L-SD assay: 50 mM, pH7.5

Phosphate buffer for sequencing gel: 1 M, pH8.0

Glycylglycine buffer: 50 mM glycylglycine, pH8.0

TE buffer: 10 mM Tris.HCl (pH8.0), 1 mM EDTA (pH8.0)

TES buffer: 100 mM Tris.HCl (pH7.5), 5 mM EDTA and 0.1% SDS.

TAE buffer: 0.04 M Tris-acetate, 0.001 M EDTA (pH8.0)

TBE buffer: 0.09 M Tris-borate, 0.002 M EDTA (pH8.0)

Cracking buffer: 50 mM Tris.HCl (pH8.0), 0.8 mM EDTA (pH8.0),

         1% SDS, 13.7% Sucrose, 0.01% Bromophenol blue

Z buffer (per liter): 16.1 g $Na_2HPO_4.7H_2O$; 5.5 g $NaH_2PO_4.H_2O$;

         0.75 g KCl; 0.246 g $MgSO_4.7H_2O$, $\beta$-

         mercaptoethanol. pH7.0.

Triton mix buffer: 0.4% Triton100, 100mM Tris pH8.0, 100mM

         EDTA

X-gal solution: 20 mg/ml 5-bromo-4-chloro-3-indolyl-$\beta$-D-

         galactoside in N-N-dimethyl-formamide

ONPG solution: 4mg/ml o-nitrophenyl-$\beta$-D-galactoside in 0.1 M

         phosphate buffer, pH7.0

DNPH solution: 0.025% 2,4-dinitrophenylhydrazine dissolved in

         4.1% HCl.

Acrylamide stock solution: 38% acrylamide,

         2% N,N-methylenebisacrylamide.

10 M urea: 10 M urea dissolved in 1.4X TBE buffer.

20% PEG solution for single stranded DNA isolation:

         20% polyethylene glycol, 3.5 M ammonium

         acetate, pH7.5

# 4. Enzyme assays

## 4.1. L-serine deaminase assay

L-serine deaminase (L-SD) was assayed in toluene-treated cells as described previously (Pardee & Prestidge, 1955). One unit of L-SD is defined as the amount of enzyme which catalyzed the formation of 1 umol of pyruvate in 35 min.

## 4.2. $\beta$-galactosidase assay

$\beta$-galactosidase was assayed by the method described by Miller (Miller, 1972) and is expressed in Miller units.

# 5. Strain constructions

## 5.1. Construction of *ssd* derivatives

In order to transduce the *ssd* mutation, a Tn*10* insertion was placed near the *ssd* locus. This was done by first transducing a *Zij-501*::Tn*10* insertion from strain CAG18477 to strain KEC9, and then transducing the *ssd* mutation to strain MEW1 by cotransduction with the tetracycline resistance marker. The presence of the *ssd* mutation at each s. ep was verified by testing for ability to grow on L-serine. One *ssd* mutant made in this way is known as MEW86.

All other *ssd* derivatives were then made in the same way from MEW86.

## 5.2. Construction of *lrp* derivatives

*lrp* derivatives were created by transducing the mutant gene from strain MEW26 *lrp*::Tn*10*, selecting tetracycline-resistant transductants, and verifying the phenotype by growth on L-serine as described previously (Lin, et al., 1990).

## 5.3. Construction of *crp* derivatives

*crp* derivatives were made by transducing a mutant *crp* gene from strain SA2777, which carries a *crp*::*Cm*$^r$ insertion, and selecting chloramphenicol-resistant transductants. The phenotype was verified by plating on MacConkey lactose arabinose plates, on which *crp* derivatives produced white colonies.

## 5.4. Construction of *sdaA*, *sdaB* and *sdaC* derivatives

*sdaA* derivatives were made by transducing *sdaA*::*Cm* from strain MEW28 and selecting for chloramphenicol resistance.

*sdaB* derivatives with mutant promoters were made by transducing from MEW51 *sdaB*::λplacMu9 and selecting for kanamycin resistance.

*sdaC* derivatives were constructed by transducing from CP41 or CP52 *sdaC*::λplacMu9 and selecting for kanamycin resistance.

## 6. Transduction and transformation

P1-mediated transduction was carried out by the method described by Miller (Miller, 1972).

Transformation was performed according to the method described by Maniatis (Maniatis et al., 1982).

## 7. DNA isolation and restriction digestions

### 7.1. Plasmid DNA isolation

"Miniprep" of plasmids was carried out by the method of Maniatis (Maniatis et al., 1982).

Larger amounts of plasmid, for use in subcloning, deletion and sequencing, were isolated by the method used by Lang (Lang, personal communication). 30 ml of overnight culture in "mix medium" were harvested by centrifugation and resuspended in 10 ml Triton mix buffer. 2 ml of 10mg/ml lysozyme were added and the suspension was boiled for 40 seconds. After centifugation, the supernatant was treated with RNase, proteinase K, and phenol, chloroform. Plasmid DNA was finally precipitated with ethanol and desalted with a Sephadex G-50 column.

### 7.2. Bacteriophage lambda DNA isolation

Bacteriophage λ DNA was isolated by the method described by Silhavy (Silhavy et al., 1984).

## 7.3. Restriction digestions

Restriction enzyme digestions were carried out as described by Maniatis (Maniatis et al., 1982)

## 7.4. DNA recovery from agarose gels using the Geneclean kit (BioCan Scientific)

About 400 mg of agarose gel containing DNA was collected in an Eppendorf tube and 800 $\mu$l of NaI solution added. The agarose gel was melted at 50°C and 10 $\mu$l of glassmilk was added. The tube was placed on ice for 10 min and then centrifuged for 10 seconds. The pellet was washed 3 times with "new wash" solution and suspended in 10 $\mu$l of TE buffer. DNA was recovered by heat shock at 50°C for 5 min and then centrifuged for 2 min. The supernatant containing the DNA was transferred into another tube.

## 8. Deletions using Exonuclease III

DNA deletion by Exonuclease III (ExoIII) was carried out as described in the Erase-a-Base system from Promega Biotec, Madison, Wisconsin. Plasmids were digested with two enzymes, generating one 3' protruding end and one 3' recessive end, the 3' recessive end being subject to ExoIII deletion. Where no 3' protruding end could be generated by a restriction enzyme, the 3' recessive end was protected with $\alpha$-phosphorothloate nucleotides, and then another 3' recessive end was generated.

Plasmid DNA, digested as above, was incubated with ExoIII at 37°C and time points were taken at 30 second to 60 second intervals until the required deletion has been achieved. The DNA was then incubated with SI nuclease and then with Klenow enzyme and dNTPs to generate blunt ends, and ligated with T4 DNA ligase.

The resulting plasmids of mixed size from each time point were transformed into strain XL1-Blue. 3-5 transformants from each time point were grown on LB plates and lysed with cracking buffer. Plasmid size was determined on agarose gels.

## 9. Isolation of single-stranded DNA

Single-stranded DNA was isolated by the method of Su (personal communication). 100 $\mu$l of exponential phase cells carrying the appropriate plasmids were infected with 1 $\mu$l of helper phage M13K07 (titre of $10^{13}$ pfu/ml) and incubated for 1 hour at 37°C with shaking. Then 2ml of 2X YT medium containing kanamycin were added and the culture incubated for another 14-20 hours.

1.5 ml of this culture were centrifuged for 5 minutes. 1.2 ml of the supernatant were transferred to another Eppendorf tube and 300 $\mu$l of 20% PEG solution (pH7.5) were added. This mixture was left at room temperature for 10 min and then centrifuged for 10 min. The pellet was dissolved in 300 $\mu$l TES buffer, 100 ug proteinase K were added, and the mixture incubated at 37°C for at least 60 min. The DNA was

41

then extracted with phenol and chloroform and precipitated with ethanol. The precipitated single-stranded DNA was resuspended in 20 $\mu$l TE buffer.

## 10. DNA sequencing reaction

### 10.1. Single-stranded DNA sequencing

DNA sequences were determined by the dideoxy-chain termination method described by Sanger et al (Sanger et al., 1977). The reaction was carried out by the protocol indicated in the Sequenase™ kit from United States Biochemical Corporation, Cleveland, Ohio. Both DNA strands were sequenced with dGTP and one with dITP. Where there appeared to be strong secondary structures, the second strand was also sequenced with dITP.

### 10.2. Double-stranded DNA sequencing

Denaturation of about 3 $\mu$g of double-stranded DNA was carried out in 25 $\mu$l of 0.2 M NaOH 20 mM EDTA at 37°C for 10 min. Denatured DNA was recoverd by ethanol precipitation with 0.3 M sodium acetate and rinsed with 70% ethanol.

The following sequencing reaction was performed as described above.

## 11. DNA sequencing gel

The sequencing gel system used in this study was developed by Lang (Lang & Burger, 1990).

Fig.2  Subcloning of the *sdaB* gene

A 7.2-kbp *Hind*III fragment from pMES41 was inserted into the *Hind*III site of pBluescript, forming pMES60. A 4.2-kbp *Pst*I-*Hind*III fragment from pMES60 was then inserted into pBluescript, forming pMES61.

Abbreviations: H, *Hind*III; P, *Pst*I; pBS, pBluescript.
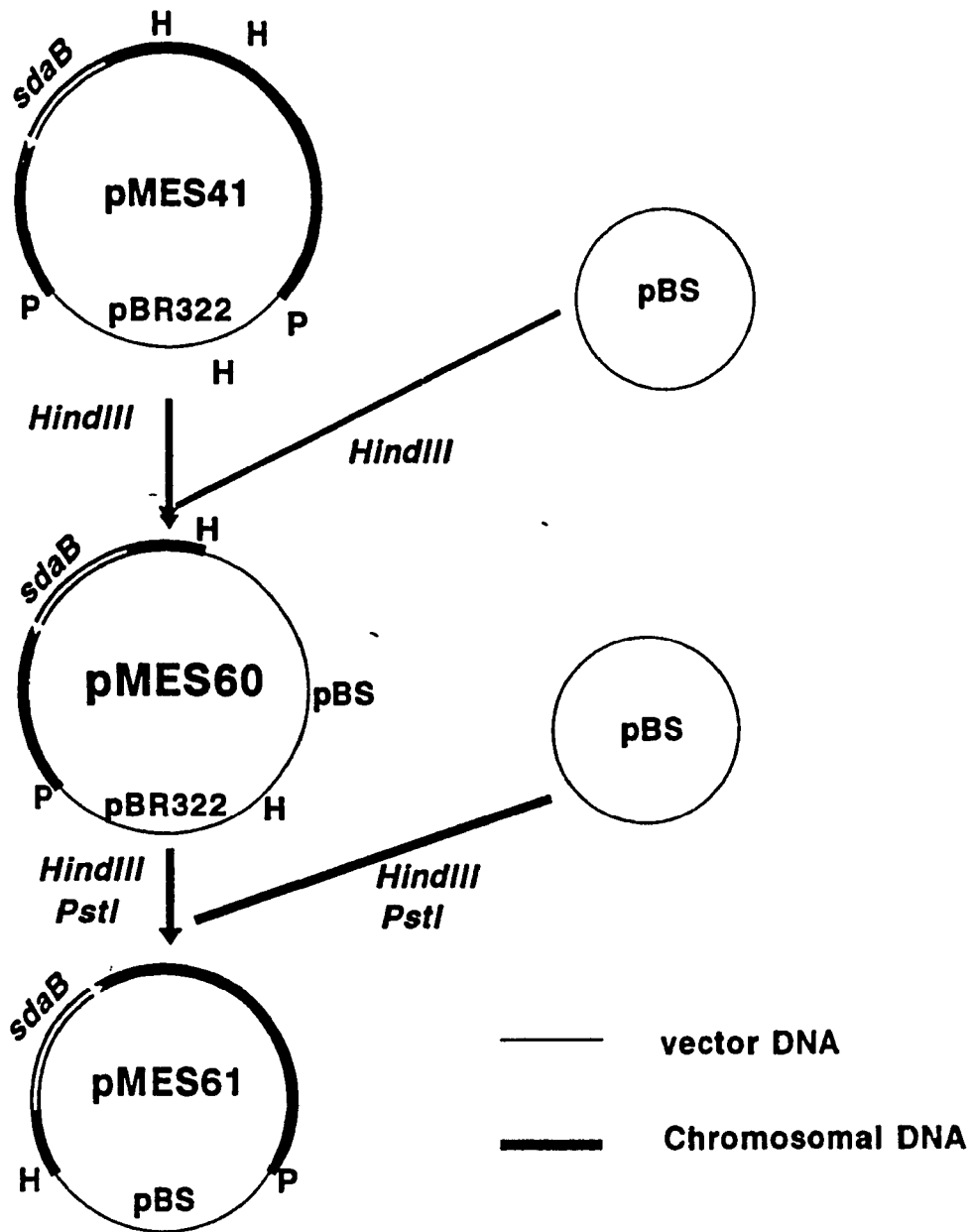
Fig.2  Subcloning of the *sdaB* gene

## 12. Plasmid constructions

### 12.1. Subcloning of the *sdaB* gene from *sdaB sdaX-*

The cloning of *sdaB* with a mini-Mu vector and subcloning of an 8.4-kbp DNA fragment into pBR322 as plasmid pMES41 was described earlier (Su & Newman, 1991).

To further subclone the *sdaB* gene for sequencing, pMES41 was cut with *Hind*III and ligated to pBluescript cut with the same enzyme, and transformants were selected on NSIV plates with ampicillin. The resulting plasmid, pMES60, contained 4.2-kbp of cloned DNA and about 3-kbp DNA from pBR322. To remove the pBR322 fragment, pMES60 was cut with *Pst*I and *Hind*III and the 4.2-kbp fragment containing the *sdaB* gene was isolated from an agarose gel with Geneclean. This fragment was then ligated into pBluescript cut with *Pst*I and *Hind*III, forming pMES61 (Fig.2). The 4.2-kbp insert on pMES61 was sequenced.

The 4.2-kbp *Bam*HI-*Hind*III fragment on pMES61 was sub-cloned back to pBR322 cut with the same enzymes, forming pMES62.

### 12.2. Subcloning of the *sdaB* gene from Kohara phage #457

DNA was isolated from Kohara phage #457 (Kohara et al., 1987) and cut with *Pst*I. An 8.4-kbp fragment was isolated from an agarose gel. This fragment had been demonstrated to hybridize with a clone of the *sdaB* gene (Su & Newman, 1991) and therefore was likely to be the same fragment cloned on

45

Fig.3 Subcloning of *sdaB* from Kohara phage #457

A 8.4-kbp *Pst*I fragment was first isolated from Kohara phage #457 DNA. This DNA fragment was cut with *Hin*dIII and a 4.2-kbp fragment was then isolated. This *Pst*I-*Hin*dIII fragment was inserted into pBluescript cut with *Pst*I and *Hin*dIII, forming pWTB6. A 4.2-kbp *Hin*dIII-*Bam*HI fragment from pWTB6 was then inserted into pBR3ᴢ2, forming pWTB7.

Abbreviations: H, *Hin*dIII; P, *Pst*I; B, *Bam*HI; pBS, pBluescript.
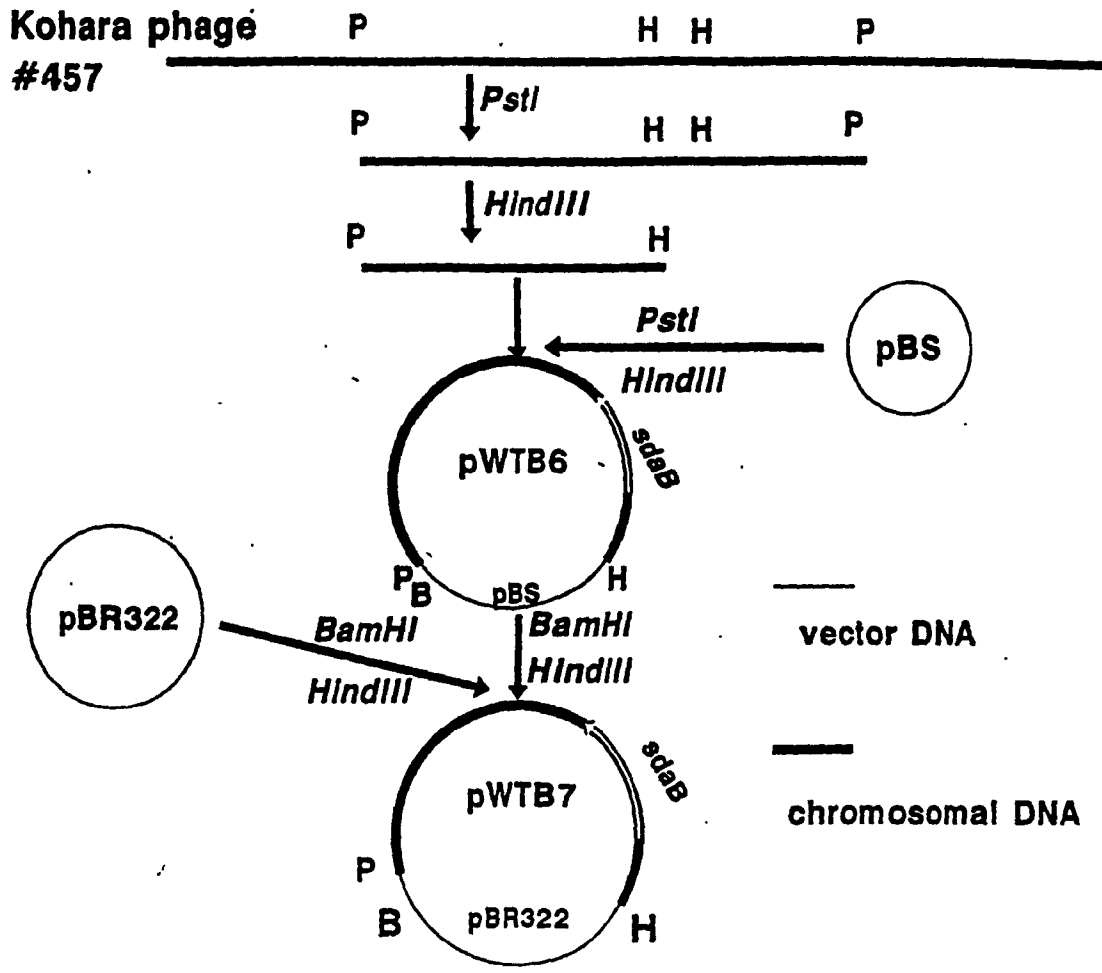
Fig.3 Subcloning of *sdaB* from Kohara phage #457

pMES41 but originating from a wild-type strain background.

The 8.4-kbp *Pst*I fragment was cut further with *Hind*III.
A 4.2-kbp *Hind*III-*Pst*I fragment was isolated from an agarose
gel and subcloned into pBluescript, forming plasmid pWTB6. The
wildtype *sdaB* gene on pWTB6 was then subcloned into pBR322
(Fig.3).

About 300 nucleotides upstream of the *sdaB* ATG
translation start site were sequenced from pWTB6.


## 12.3. Deletion analysis of the *sdaCB* operon

pMES41 was cut with *Sal*I and a 6.8-kbp fragment was
isolated from an agarose gel. This fragment was ligated to
pBR322 cut with the same enzyme and transformed. Plasmids were
isolated and digested with *Eco*RI to check the orientation of
insertions.

pMES63, the plasmid with an 6.8-kbp *Sal*I insert, oriented
such that the *Eco*RI site on the fragment was located very
close to the one on the vector, was used to further deletion
analysis (Fig.4). This plasmid was first cut with *Bam*HI and
treated with Klenow and α-phosphorothioate nucleotides, and
cut again with *Bgl*II. ExoIII deletion on this plasmid was
carried out as described above. Three resulting plasmids,
pMES64, pMES65 and pMES66, still support growth on L-serine,
were much smaller than the starting plasmid pMES63. The insert
on pMES64 is about 3.7-kbp, while the one on pMES65 is about
3.2-kbp and on pMES66 is about 2.7-kbp (Fig.5).

Fig.4   Subcloning of the 6.8-kbp *Sal*I fragment

A 6.8-kbp *Sal*I fragment was isolated from pMES41. This DNA fragment was then inserted into the *Sal*I site of pBR322, forming pMES63.

Abbreviations: H, *Hind*III; E, *Eco*RI, S, *Sal*I; P, *Pst*I.
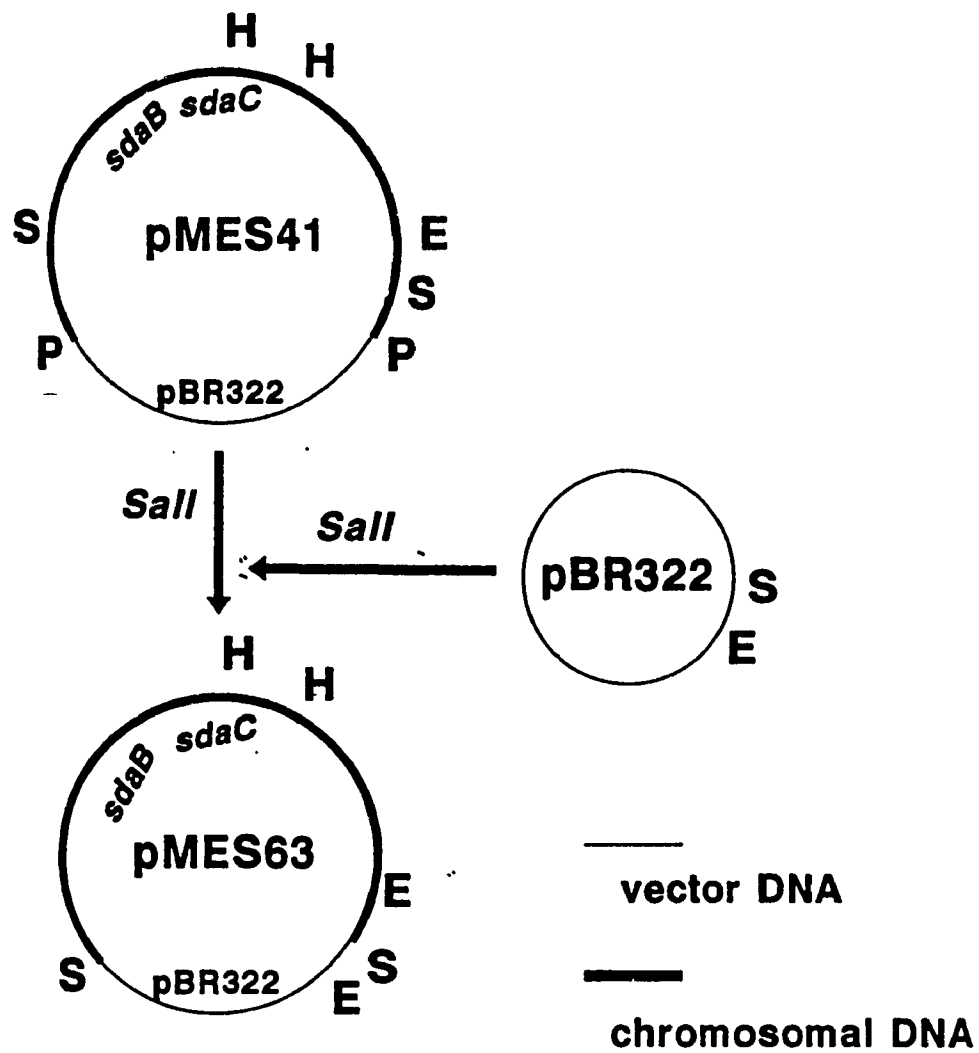
Fig.4 Subcloning of the 6.8-kbp *Sal*I fragment

Fig.5  Deletion analysis of the *sdaCB* operon

Plasmid pMES63 was cut with *Bam*HI. The resulting linear DNA fragment was treated with Klenow enzyme and α-phosphorothioate nucleotides. This DNA fragment was then cut again with *Bgl*II and deletion by exonuclease III was performed at 37°C. Partially deleted plasmids were ligated and transformed. Three smaller plasmids produced in this way, pMES64, pMES65 and pMES66, were isolated.

Abbreviations: H, *Hind*III; E, *Eco*RI; S, *Sal*I; P, *Pst*I; Bg, *Bgl*II; B, *Bam*HI; Exo III, exonuclease III.
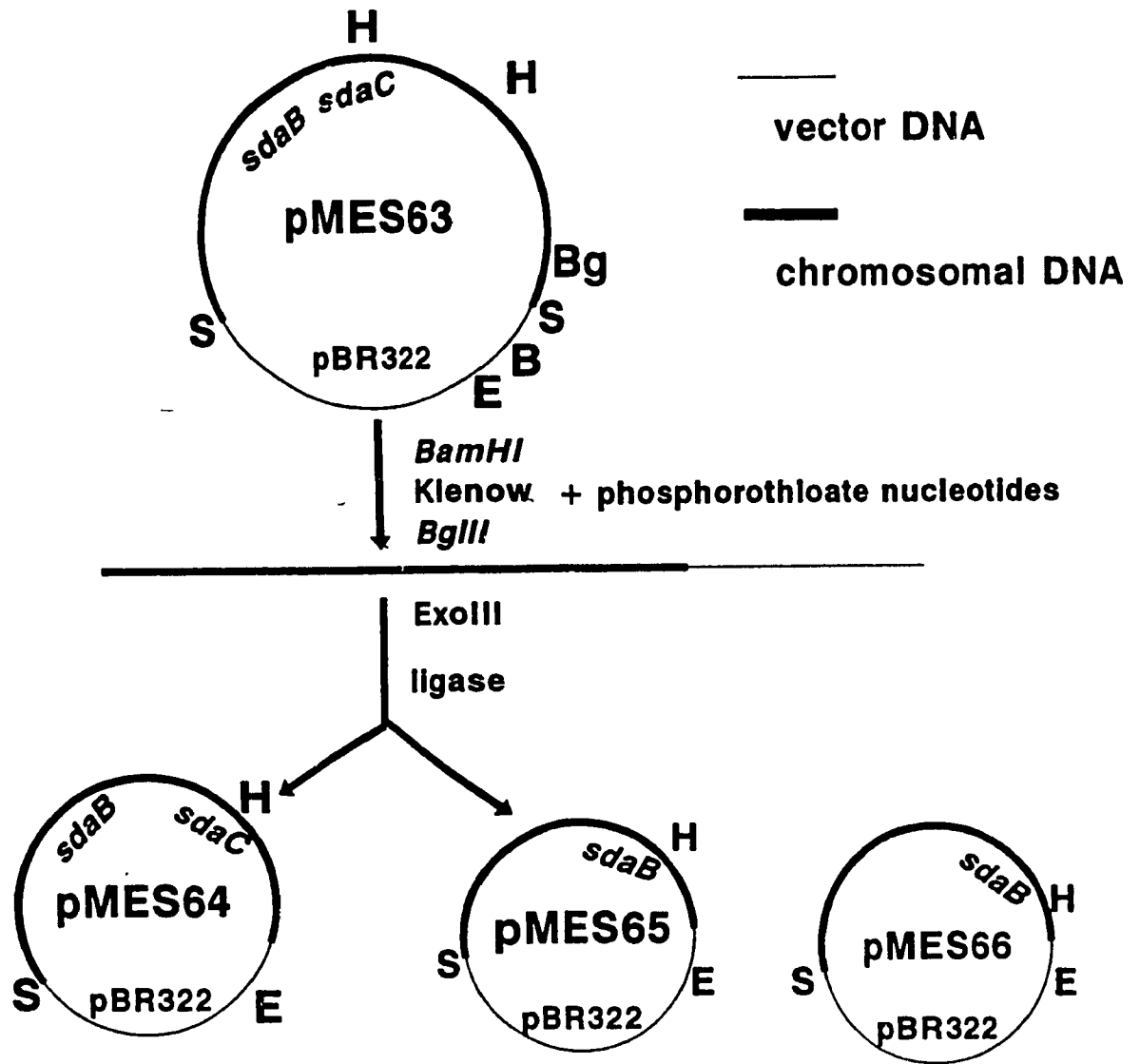
Fig. 5 Deletion analysis of the *sdaCB* operon

## 12.4. Construction of *sdaB* or *sdaC* deletion on plasmid

To test whether there is an activator gene upstream of *sdaC*, plasmids with a deleted *sdaB* gene were constructed. This was done by ExoIII deletion, since no suitable restriction enzyme sites were available.

In order to do ExoIII deletions and subsequent sequencing, the plasmid, pMES64, was first inserted into the polycloning site of pBluescript. To do that, pMES64 was cut with *Sal*I and ligated to pBluescript cut with the same enzyme. The orientation of the insert in one resulting plasmid, pMES67, was checked by restriction digestion.

ExoIII deletion was carried out on plasmid pMES67 digested with *Sac*I and *Xba*I. Two smaller plasmids that did not support growth on L-serine were isolated. One plasmid, pMES68, was sequenced using an M13 primer and was found to have 138 bases (46 amino acids) deleted from the end of the *sdaB* gene (Fig.6). Another plasmid, pMES69, was found to have the entire *sdaB* gene and 170 bases from the end of the *sdaC* gene deleted.

## 12.5. Subcloning of the 1.1-kbp *Hind*III fragment

The plasmid pMES41 was cut with *Hind*III and a 1.1-kbp fragment was isolated from an agarose gel. This fragment was then ligated to pBluescript cut with *Hind*III. One of the resulting plasmids was known as pMES70 (Fig.7).

Deletion of plasmid pMES70 and sequencing of the 1.1-kbp fragment were carried out as described above.

Fig.6  Construction of a plasmid carrying a *sdaCB* deletion

Plasmid pMES64 was cut with *Sal*I and inserted into the *Sal*I site of pBluescript, forming pMES67. pMES67 was then digested by *Sac*I and *Xba*I. The digested plasmid DNA was partially deleted by exonuclease III and two resulting plasmids were isolated. In the case of plasmid pMES68, 138 nucleotides from 3′ end of *sdaB* were deleted on pMES68 and entire *sdaB*. For plasmid pMES69, 170 nucleotides from 3′ end of *sdaC* were deleted.

Abbreviations: E, *Eco*RI; H, *Hind*III; S, *Sal*I; X, *Xba*I; pBS, pbluescript.
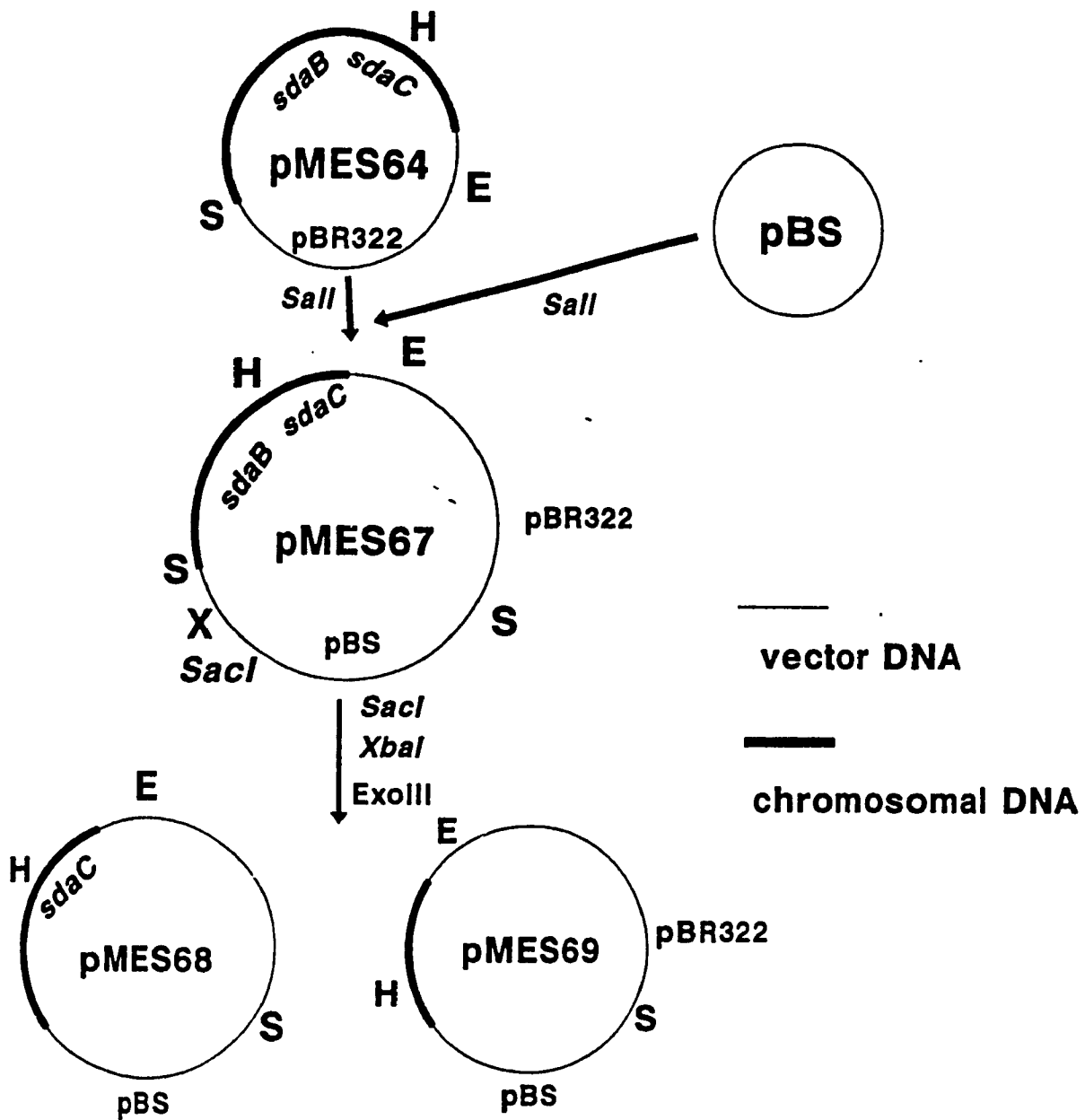
Fig.6  Construction of a plasmid carrying a *sdaCB* deletion

Fig.7  Subcloning of the 1.1-kbp *Hind*III fragment

Plasmid pMES41 was cut with *Hind*III and the 1.1-kbp *Hind*III fragment was isolated. This fragment was then inserted into the *Hind*III site of pBluescript.

Abbreviations: H, *Hind*III; P, *Pst*I; pBS, pBluescript.
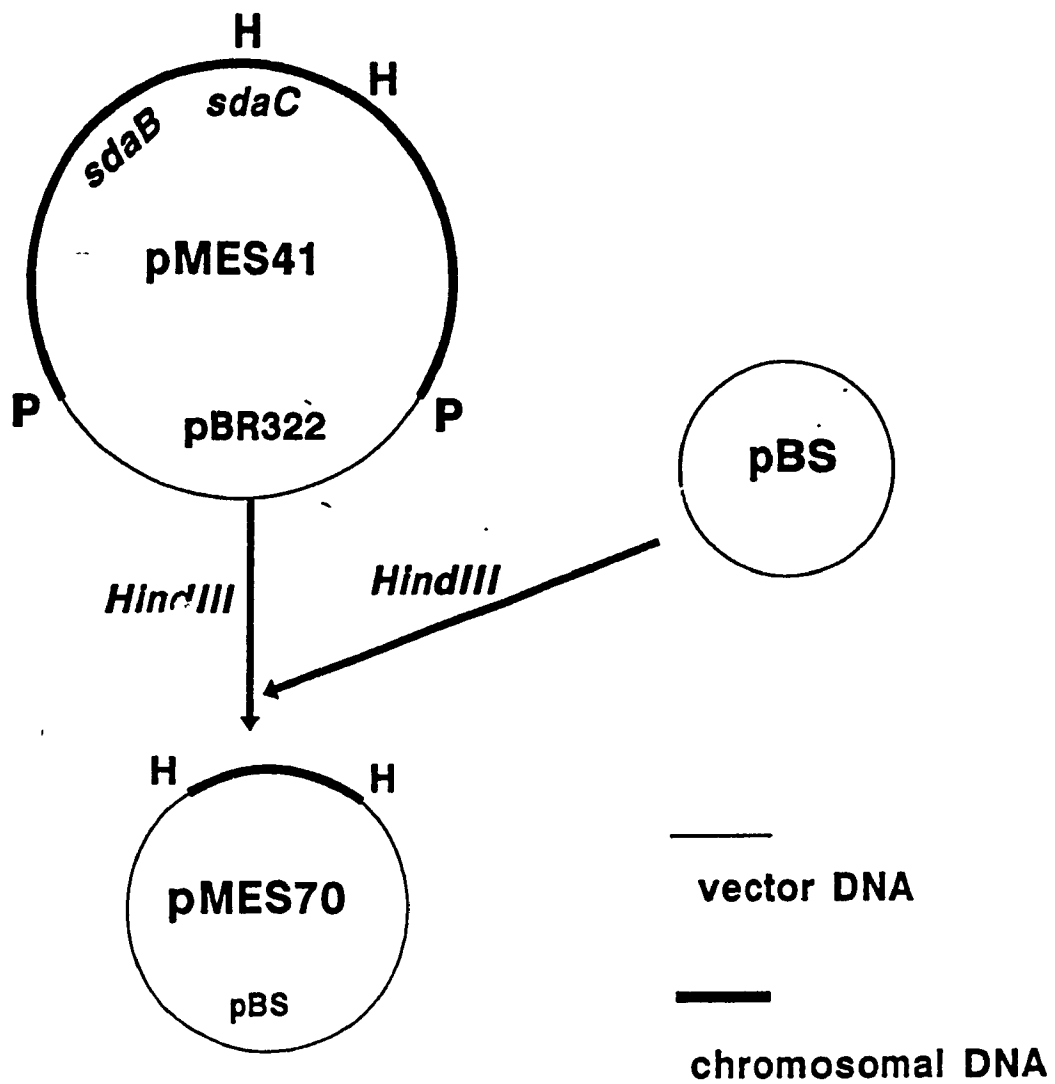
Fig. 7 Subcloning of the 1.1-kb *Hind*III fragment

## 13. Construction of a fusion of lacZ to the wildtype sdaB gene

Plasmid pMC1871 was digested with *Sal*I and the fragment containing the truncated *lacZ* gene was isolated from an agarose gel. This fragment was then treated with Klenow enzyme and dNTP mix to make blunt ends. This *lacZ*-containing fragment was ligated to an *Eco*RV fragment of pWTB6, which contains the wildtype *sdaB* gene cloned from Kohara phage #457. The ligation mix was transformed into strain MEW1 and plated on LB plate with X-gal and ampicillin. Plasmids from blue colonies on X-gal plates were isolated and checked by restriction enzyme digestion.

This resulted in an in-frame fusion of *lacZ* at codon 63 of *sdaB* controlled by the wildtype promoter. One of the plasmids was known as pMES71 (Fig.8).

## 14. Transfer of the sdaB::lacZ fusion to the E.coli chromosome

Plasmid pMES71 was used to transform a *polA⁻* strain A402, a *lac⁻* derivitive of A401, and plated on LB plates with X-gal and ampicillin. Since the plasmid could not replicate in a *polA⁻* background, colonies grown on the selection plate should be cells with plasmids inserted into chromosome by homologous recombination (Winans et al., 1985). Transformants, which are ampicillin resistant and blue on X-gal plates, were infected with a P1 phage lysate made on strain CAG12078 (Singer et al., 1989), which carries a Tn*10* at 60.5 min. P1 phage lysates were then made on a tetracycline, ampicillin

Fig. 8 *In vitro* construction of the *sdaB::lacZ* fusion

A *SalI* fragment carrying a truncated *lacZ* gene was isolated from pMC1871. This fragment was treated with Klenow enzyme and dNTP mix. The treated DNA fragment was then inserted into pWTB6 cut with *EcoRV*, forming pMES71.

Abbreviations: S, *SalI*; H, *HindIII*; P, *PstI*; Ev, *EcoRV*.
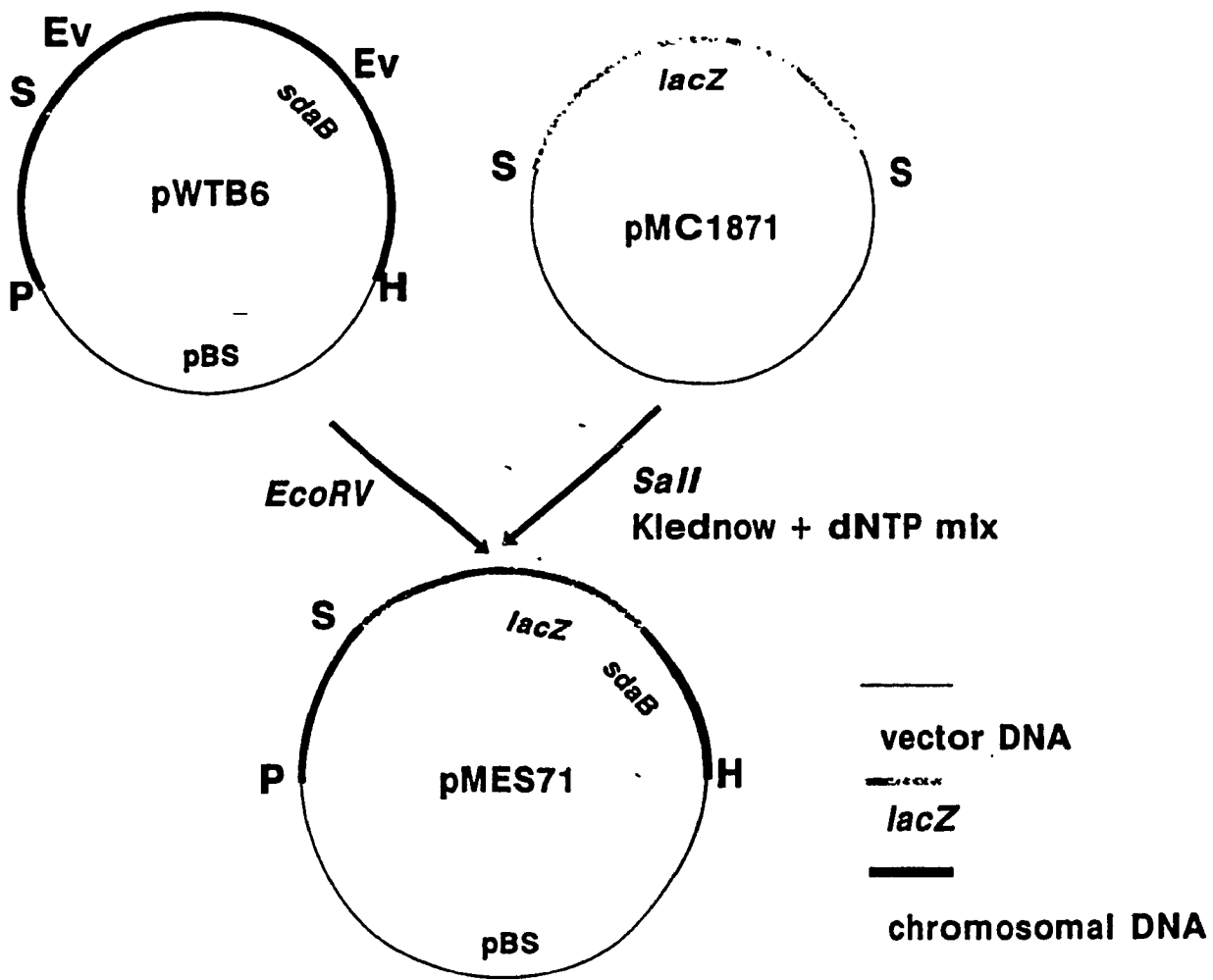
Fig. 8   Construction of the *sdaB::lacZ* fusion *in vitro*

resistant *lac*⁺ strain. This lysate was used to transduce *lac*⁺ and *tet*-resistance to strain MEW1, and an ampicillin-sensitive transductant, MEW88, was selected for further study.

## 15. Site-directed mutagenesis

Plasmid pWTB6 contains a 4.2-kb *PstI-HindIII* fragment cloned from Kohara phage #457. A synthesized oligonucleotide was used to convert the cytosine at 13 bases upstream of *sdaB* on this plasmid to an adenine.

To do this, strain CJ236 with plasmid pWTB6 was infected by helper phage M13K07. Infected cells were grown overnight in LB medium containing chloramphenicol and ampicillin. The uracil-containing single-stranded DNA was isolated from the supernatant of the culture by the method described above (Section 9). This ssDNA was annealed with the synthesized oligo-TTTTAACG*AGATGTATTTC, which corresponds to the sequence at *sdaB* ribosome-binding site region. *A indicates the mutation that would be introduced into the wildtype clone of the *sdaB* gene, which was seen in the clone obtained from the *sdaX* mutant strain. The site-directed mutagenesis process was carried out as described by Kunkel (Kunkel, 1985), using a Muta-Gene mutagenesis kit (Bio-Rad).

Ligated plasmids after mutagenesis were transformed into strain MV1190 and plasmids were isolated from the transformants. Eight such plasmids were sequenced using a synthesized primer, ATCGAAATCAACAAGC, which corresponds to the

61

sequence 305 bases upstream of ATG translation start site of *sdaB* gene. Among the eight plasmids sequenced, two plasmids were found carrying the expected mutation.

## 16. Computation on sequence analysis

Computation to search for similar sequences was performed at NCBI using the BLAST network service.

Sequence comparisons were carried out using FASTA program (Pearson et al., 1989).

Searching for possible open-reading frames in the nucleotide sequences was carried out by the method developed by Kolaskar and Reddy (Kolaskar & Reddy; 1985).

Analysis of transmembrane helices in the protein was performed using the methods of Rao and Argos (Rao & Argos; 1986) and Eisenberg et al. (Eisenberg et al., 1984).

## 17. List of abbreviations used in this thesis

| | |
|---|---|
| L-SD: | L-serine deaminase |
| DTT: | Dithiothreitol |
| C1: | one-carbon |
| cAMP: | cyclic AMP |
| Lrp: | Leucine-responsive regulatory protein |
| Crp: | cyclic AMP receptor protein |
| Kbp: | Kilobase-pair |
| 16S rRNA: | 16S ribosomal RNA (S, Sedimentation coefficient) |

## RESULTS

*E.coli* makes two L-serine deaminases (L-SD). The *sdaA* gene has been demonstrated to be the structural gene for the first L-serine deaminase (L-SD1) (Su, et al., 1991), which is made by cells grown in both glucose-minimal and LB medium. An *sdaA* null mutant shows no L-SD activity in glucose-minimal medium, but still has a considerable amount of L-SD activity in LB medium (Su & Newman, 1991), suggesting that in rich medium *E.coli* makes another L-SD.

The experiments in this thesis concern the second L-SD gene, *sdaB*. In earlier work from this laboratory, the *sdaB* gene was cloned. To do this, a mutant was isolated in which the second L-SD (L-SD2) was expressed in glucose-minimal medium due to a regulatory mutation known as *sdaX*. The *sdaB* gene was then cloned from this mutant strain (Su & Newman, 1991).

The results section begins with the determination of the sequence of the *sdaB* gene, as cloned from the *sdaX* mutant. The *sdaB* gene is shown to be very similar to *sdaA*. The precise map position of *sdaB* in the *E.coli* chromosome is determined by the discovery that the *sdaB* clone contains some of the fucose operon, which locates the fucose operon immediately downstream of the *sdaB* gene, with a possible open-reading frame, *orfX*, between *sdaB* and *fucO*.

63

The next experiments show that the regulatory mutation which established SdaB synthesis in glucose-minimal medium is due to a change in the *sdaB* ribosome-binding site. This was determined by comparing the upstream sequences of the *sdaB* gene from wildtype and mutant clones, and was further comfirmed by constructing this mutation from the wildtype *sdaB* clone by site-directed mutagenesis and showing that this is sufficient to reproduce the mutant phenotype.

Further experiments demonstrate that the expression of the two genes coding for L-SD are regulated in very different fashion. The expression of the *sdaB* gene is shown to be regulated by cyclic AMP and cAMP receptor protein (Crp). *sdaB* differs from *sdaA* also in that the expression of *sdaB* is not regulated by the *ssd* gene product and the product of the *lrp* gene has no, or very little, effect on the expression of the *sdaB* gene. Moreover, the expression of the *sdaB* gene is not affected by a variety of environmental factors influencing *sdaA* transcription.

I will then show that the *sdaB* gene is the 3′ gene in an operon which consists of two genes, with the newly recognized *sdaC* gene situated upstream of the *sdaB* gene. As judged by its sequence, the *sdaC* gene likely codes for a transmembrane protein and may share a common ancestor in early evolution with the *tdcC* gene. The SdaC protein likely functions as a transport protein, probably for L-serine.

**Part 1: Subcloning and sequencing of the *sdaB* gene.**

**1.1. Subcloning of the *sdaB* gene.**

The *sdaB* gene was cloned from the *sdaX* mutant strain using the miniMu *in vivo* cloning method (Su & Newman, 1991). An 8.4-kbp *Pst*I fragment was subcloned into pBR322 from the miniMu r ₁licon. As would be expected for DNA carrying the *sdaB* gene, this 8.4-kbp fragment conferred the ability to synthesize L-SD in LB medium. However, the L-SD activity was also seen in minimal medium, although the level was much lower. This suggested that *sdaB* and *sdaX* are on the same fragment, and that *sdaX* may be a mutation within the *sdaB* gene.

The *sdaB* plasmid hybridizes well with the *sdaA* gene, further suggesting that it carries a gene homologous to *sdaA*, i.e. that the plasmid carries the *sdaB* gene which is similar enough to the *sdaA* gene that it likely codes for the second L-serine deaminase (Su & Newman, 1991). It therefore seemed worthwhile to sequence the *sdaB* gene i.e. part of the 8.4-kbp fragment.

In order to sequence the *sdaB* gene, I first subcloned from pMES41 the 7.2-kbp *Hind*III fragment, which consists of the 4.2-kbp *Hind*III-*Pst*I fragment from the 8.4-kbp fragment and a 3-kbp fragment from pBR322, to a phagemid pBluescript. Since this subclone conferred the ability to make L-SD, the *sdaB* gene should therefore be located on this 4.2-*Y*bp *Hind*III-

65

*Pst*I fragment. This fragment was then subcloned into pBluescript, forming pMES61 (for details, see Material and Methods).

## 1.2. The sequence of the *sdaB* gene.

The 4.2-kbp *Hind*III-*Pst*I fragment on pMES61 was sequenced from both DNA strands after deletion by Exonuclease III. This sequence- a total of 4236 nucleotides in length- is presented in figure 9 in the direction from *Hind*III to *Pst*I.

Analysis of this sequence revealed two complete open-reading frames. The first open-reading frame starts with an ATG at nucleotide 971 and ends with a TAA at nucleotide 2338, coding for 455 amino acids. The second open-reading frame starts with a GTG at nucleotide 2449 and ends with a TAA at nucleotide 3204, coding for 251 amino acids.

I searched for sequences similar to the sequence presented in figure 9 using the Blast network service of NCBI in August 1992. It was found that the sequence around the first open-reading frame in the 4.2-kbp fragment is extremely similar to the *sdaA* sequence. I therefore propose that this open-reading frame codes for the structure of the SdaB protein.

There is no obvious similarity found between the sequences in the gene bank and the 855 bp open-reading frame, *orfX*. This therefore remains unidentified.

66

Fig. 9   DNA sequence of the *sdaB* gene and surrounding area.

This DNA sequence is derived from the 4.2-kbp DNA fragment inserted in pMES61. The possible ribosome-binding sites and transcriptional terminators are underlined. The sequence from *Pvu*II to *Pst*I is identical to that of *fucO* which has been published (Chen et al., 1989).

```
                    CTTCATGTCTCACCAGCTGGGTATGACGCCACCGCCGCGTGCGATTCTGTCGCTGATCCTGATCGTGGGTATGATGACC    79

ATCGTTCGCTTCGGTGAGCAGATGATCGTTAAAGCGATGAGTATTCTGGTATTCCCGTTTGTTGGCGTACTGATGCTGCTGGCTCTGTACCTGATCCCG    78

CAGTGGAACGGCGCTGCACTGGAAACGCTGTCTCTGGACACTGCATCTGCAACCGGAAACGGTCTGTGGATGACCCTGTGGCTGGCAATTCCGGTAATG   _77

GTGTTCTCGTTCAACCACTCTCCGATCATCTCTTCTTTCGCCGTTGCGAAGCGTGAAGAGTACGGCGATATGGCAGAACAGAAATGCTCGAAGATCCTG   376

GCATTCGCACACATCATGATGGTGCTGACCGTAATGTTCTTCGTCTTCAGCTGTGTACTGAGCCTGACTCCGGCAGACCTGGCTGCGGCTAAAGAGCAG   475

AACATCTCGATTCTGTCTTACCTGGCTAACCACTTTAACGCACCGGTTATCGCGTGGATGGCTCCGATTATCGCGATTATCGCTATCACCAAATCCTTC   574

CTCGGTCACTACCTGGGCGCACGTGAAGGCTTCAACGGTATGGTGATTAAATCTCTGCGTGGTAAAGGTAAGTCTATCGAAATCAACAAGCTGAACCGT   673

ATCACTGCGCTGTTCATGCTGGTAACGACCTGGATTGTTGCCAC^CTGAACCCGAGCATCCTGGGTATGATTGAAACCCTGGGCGGTCCAATCATCGCG   772

ATGATCCTGTTCCTGATGCCGATGTACGCAATTCAGAAAGTACCGGCAATGCGTAAGTACAGCGGTCACATCAGCAACGTATTCGTTGTCGTGATGGGT   871

CTGATTGCAATCTCCGCAATCTTCTACTCTCTGTTCAGCTAAGTCCTTTCGCGCCGCTTTCGGGCGGCGCTTCCTCCGTTTTAACGAGATGTATTTCCT   970
start of sdaB
  M  I  S  V  P  D  I  P  K  I  G  I  G  P  S  S  S  H  T  V  G  P  H  K  A  G  K  Q  P  T  D  D  L
ATGATTAGCGTATTCGATATTTTCAAAATCGGCATTGGCCCTTCCAGTTCTCATACCGTTGGACCAATGAAAGCGGGTAAACAATTTACCGACGATCTG   1069
  I  A  R  N  L  L  K  D  V  T  R  V  V  V  V  D  V  Y  G  S  L  S  L  S  L  T  G  K  G  H  H  T  D  I  A  I
ATTGCCCGTAACCTGCTTAAAGACGTGACCCGCGTGGTGGTTGACGTGTACGGCTCGCTCTCTCTGACCGGTAAAGGCCACCACACTGATATCGCCATT   1168
  I  M  G  L  A  G  N  L  P  D  T  V  D  I  D  S  I  P  S  F  I  Q  D  V  N  T  H  G  R  L  M  L  A
ATTATGGGCCTGGCGGGTAACCTGCCGGATACCGTGGATATCGATTCCATCCCCAGTTTTATTCAGGATGTGAATACTCATGGTCGCCTGATGCTGGCA   1267
  N  G  Q  H  V  F  P  V  D  Q  C  H  N  F  H  A  D  N  L  S  L  H  E  N  G  H  R  I  T  A  V
AACGGTCAGCATGAAGTGGAGTTCCCGGTTGATCAGTGCATGAACTTCCACGCCGA_AACCTTTCTCTGCATGAAAACGGTATGCGCATTACCGCCGTG   1366
  A  G  D  K  V  V  Y  S  Q  T  Y  Y  S  I  G  G  G  P  I  V  D  E  E  H  F  G  Q  Q  D  S  A  P  V
GCGGCCGATAAAGTCGTTTACAGCCAGACTTACTACTCTATTGGCGGTGGCTTTATCGTTGATGAAGAGCATTTTGGCCAGCAGGATAGCGCCACCGGTT   1465
  E  V  P  Y  P  Y  S  S  A  A  D  L  Q  H  C  Q  E  T  G  L  S  L  S  L  G  L  H  H  K  N  D  V  A
GAAGTTCCTTATCCGTACAGTTCAGCAGCCGATCTGCAAAAACATTGTCAGGAAACCGGGCTGTCACTCTCTGGCCTGATGATGAAAAACGACGTGGCG   1564
  L  H  S  K  E  E  L  E  Q  H  L  A  N  V  W  E  V  H  R  G  G  I  E  R  G  I  S  T  E  G  V  L  P
CTGCACAGCAAAGAAGAGCTGGAACAGCACCTGGCGAACGTCTGGGAAGTCATGCGCGGCGGTATTGAGCGCGGTATTTCCACCGAAGGCGTGTTGCCT   1663
  G  K  L  R  V  P  R  R  A  A  A  L  R  R  M  L  V  S  Q  D  K  T  T  T  D  P  H  A  V  V  D  W  I
GGCAAACTGCGTTCCACGCCGTGCGCTGCGGCCACTACGCCGGATGCTGGTCAGCCAGGATAAAACCACCACTGACCCGATGGCGGTTGTTGACTGGATC   1762
  N  H  F  A  L  A  V  N  E  E  N  A  A  G  G  R  V  V  T  A  P  T  N  G  `  C  G  I  I  P  A  V  L
AACATGTTTGCACTGGCAGTGAACGAAGAGAACGCTGCTGGCGGTCGCGTGGTGACTGCGCCGACTAACGGTC  TGCGGGATTATCCCGGCAGTTCTG   1861
  A  Y  Y  D  K  F  I  R  E  V  N  A  N  S  L  A  R  Y  L  L  V  A  S  A  I  G  S  L  Y  K  H  N  A
GCGTACTACGACAAGTTTATCCGCGAAGTGAACGCTAACTCACTGGCTCGTTACCTGCTGGTAGCCAGCGCCATTGGTTCTCTTTATAAGATGAACGCG   1960
  S  I  S  G  A  E  V  G  C  Q  G  E  V  G  V  A  C  S  H  A  A  A  G  L  A  E  L  L  G  G  N  R  A
TCGATTTCTGGTGCTGAAGTGGGTTGCCAGGGTGAAGTTGGCGTGGCCGTGCTCAATGGCGGCGGCTGGTCTGGCAGAACTATTAGGCGGCAACCGGGCG   2059
  Q  V  C  I  A  A  E  I  G  H  E  H  N  L  G  L  T  C  D  P  V  A  G  Q  V  Q  V  P  C  I  E  R  N
CAGGTGTGCATCGCGGCGGAAATCGGCCATGGAGCACAACCTCGGTCTGACGTGTGACCCGGTCGCCGGACAGGTACAGGTGCCATGCATCGAGCGTAAC   2158
  A  I  A  A  V  K  A  V  N  A  A  R  M  A  L  R  R  T  S  E  P  R  V  C  L  D  K  V  I  E  T  H  Y
GCCATTGCGGCAGTAAAAGCGGTGAACGCCGCACGTATGGCGCTGCGCCGTACCAGCGAGCCGCGCGTCTGCCTCGATAAAGTTATCGAAACCATGTAC   2257
  E  T  G  K  D  M  N  A  K  Y  R  E  T  S  R  G  G  L  A  H  K  I  V  A  C  D  *
GAAACAGGTAAAGATATGAACGCCAAGTACCGCGAAACCTCTCGCGGCGGCCTGGCAATGAAGATCGTTGCCTGCGATTAATCTGCTCCAAAGGCCTCGT   2357
                                                                       start of orfX
                                                                          V  A  V
TTTGCGAGGCCTCTTCCCGATTTCTCATCCAGCCGTAGCCTGTTCCGGCCATCGAATGTTACCCTTATCGCCTGATCTTTAAGGGGGTTATCGTGGCTGTT   2457
  H  L  L  I  V  D  A  L  N  L  I  R  R  I  H  A  V  Q  G  S  P  C  V  E  T  C  Q  H  A  L  D  Q  L
CATTTGCTTATTGTCGATGCACTGAATCTTATTCGTCGCATTCATGCCGTTCAGGGGTCGCCCTGTGTCGAAACCTGCCAGCATGCGCTCGATCAGCTC   2556
  I  M  H  S  Q  P  T  H  A  V  A  V  P  D  D  E  N  R  S  S  G  W  R  H  Q  R  L  P  D  Y  K  A  G
ATTATGCACAGCCAGCCAACCCACGCGGTCGCCGTTTTTGATGATGAAAACGCAGTAGCGGCTGGCGTCATCAGCGTTTACCAGATTACAAAGCGGGT   2655
  R  P  P  M  P  E  E  L  H  D  E  H  P  A  L  R  A  A  F  E  Q  R  G  V  P  C  W  S  T  S  G  N  E
CGACCGCCAATGCCGGAAGAGTTGCACGACGAGATGCCTGCATTACGCGCCGCCTTTGAGCAACGCGGCGTCCCGTGCTGGTCAACCAGCGGCAACGAA   2754
  A  D  D  L  A  A  T  L  A  V  K  V  T  Q  A  G  H  Q  A  T  I  V  S  T  D  K  G  Y  C  Q  L  L  S
GCCGATGACTTAGCCGCCACGCTGGCGGTCAAAGTGACACAGGCCGGGCATCAGGCAACGATTGTTTCGACAGATAAAGGCTACTGTCAGTTACTTTCA   2853   _
  P  T  L  R  I  R  D  Y  P  Q  K  R  W  L  D  A  P  P  I  D  K  E  F  G  V  Q  P  Q  Q  L  P  D  Y
CCGACATTACGTATTCGTGATTACTTCCAGAAACGTTGGCTGGATGCGCCATTTATCGATAAAGAATTTGGCGTTCAACCGCAGCAGTTGCCCGATTAC   2952   _
                                             PvuII
  W  G  L  A  G  I  S  S  S  K  V  P  G  V  A  G  I  G  P  K  S  A  T  Q  L  L  V  E  P  Q  S  L  E
TGGGGACTTGCGGGGATCAGCAGTTCAAAGGTACCGGGTGTTGCGGGGAATCGGACCAAAAAGCGCCACGCAGCTGCTGGTCGAGTTTCAGAGTCTGGAA   3051   .
  G  I  Y  E  N  L  D  A  V  A  E  K  W  R  K  K  L  E  T  H  K  E  H  A  F  L  C  R  D  I  A  R  L
GGGATATATGAGAATCTGGATGCGGGTTGCCGAAAAGTGGCGCAAAAAAATTAGAAACCCATAAAGAGATGGCGTTTCTGTGCCGCGATATTGCCCGCTTA   3150   _
  Q  T  D  L  H  I  D  G  N  L  Q  Q  L  R  L  V  R  *
CAAACCGATTTGCATATCGACGGCAATTTACAGCAATTGCGGTTGGTACGGTAACGGCGAGCCGGATACGCCGCAAACGTCGTATCCGGCATTATCACA   3249   _
```

63

```
         *  W  A  T  H  Y  L  E  V  I  D  E  L  T  A  E  R  P  N  G  G  T  C  V  D  D  L  A  A
TCAGCGCAT    TTACCAGGCGGTATGGTAAAGCTCTACAATATCCTCAAGCGTTGCTTCACGCGGGTTGCCACCGGTACAAACATCATCCAGTGCCGC  3348
   Q  A  L  A  P  I  D  E  K  R  V  G  V  D  R  L  E  P  P  I  G  V  D  R  N  L  A  P  V  A  E  V  A
CTGCGCCAGTGCCGGAATGTCTTCCTTGCGTACACCAACATCACGCAAATGTGGCGGAATACCGACATCACGGTTGAGAGCAAACACCGCTTCAACAGC  3447
   A  N  R  A  E  E  L  S  N  G  E  V  K  V  G  M  V  R  A  I  D  R  Y  K  E  G  T  P  D  A  N  Y  R
GGCATTACGCGCCTCTTCCAGGCTCATACCTTCCACTTTCACGCCCATAACGCGCGCGATATCGCGGTACTTCTCACCGGTAAAGTCAGCGTTATAACG  3546
   M  V  E  P  L  L  I  A  N  A  V  G  E  P  T  T  Y  P  A  G  L  P  E  A  M  G  E  V  L  G  L  G  V
CATGACATGCGGTAACAGGATGGCGTTCGCAACACCGTGTGGAGTGGTATAAAACGCGCCCAGTGGATGCGCCATACCATGCACCAACCCTAACCCAAC  3645
   N  S  P  G  M  G  A  V  Y  Q  G  L  A  M  E  E  G  A  D  K  D  G  A  V  S  G  R  L  A  G  A  I  I
ATTCGAGAAGCCCATACCCGCAACATACTGCCCGAGCGCCATTTCTTCTCCGGCATCCTTATCACCAGCAACCGATCCTCGCAGCGCCCCAGCAATGAT  3744
   E  I  A  K  I  H  L  A  D  T  L  A  W  A  G  R  T  I  Y  G  E  I  A  E  T  L  A  D  V  G  T  A  A
TTCAATCGCTTTAATGTGCAGTGCATCGGTTAGCGCCCACGCGCCACGGGTAATATACCCCTCAATAGCATGAGTGAGCGCATCGACACCCGTCGCAGC  3843
   K  L  A  P  P  M  G  D  M  M  D  A  D  I  P  A  V  Q  P  I  D  H  P  D  V  C  V  P  K  R  R  K  E
TTTCAGCGCTGGAGGCATACCATCCATCATGTCAGCGTCAATAAACGCCACCTGCGGGATATCATGCGGATCAACGCAAACAAACTTGCGCCGTTTCTC  3942
   E  D  T  I  V  Y  N  I  T  V  E  A  A  T  G  A  T  T  P  I  A  L  I  P  V  S  P  K  N  T  P  S  L
TTCGTCAGTGATCACGTAGTTAATGGTCACTTCTGCCGCAGTACCTGCTGTGGTAGGAATTGCCAGAATCGGTACACTGGGTTTATTGGTCGGGGAAAG  4041
   G  E  L  S  R  V  D  A  P  E  P  N  N  S  I  I  G  I  A  K  C  T  D  Q  P  S  G  G  G  I  A  I  L
CCCTTCCAGGCTACGCACATCGGCAAACTCCGGGTTGTTGCTGATAATGCCAATCGCTTTACAAGTATCCTGTGGAGAACCACCACCAATAGCGATCAG  4140
   Y  D  A  G  S  N  Q  P  V  G  L  G  E  K  V  V  T  I  T  P  N  P  V  V  G  D  Y  I  A  W  A  L  G
GTAATCCGCGCCGCTATTCTGGAATACACCGAGCCCTTCTTTGACGACAGTAATGGTTGGGTTGGGCACTACGCCGTCGTAAATCGCCCATGCCAGCCC  4231
PstI
   A
TGCAG  4236
```

Fig. 9 DNA sequence of the *sdaB* gene and surrounding area.

69

## 1.3. A DNA sequence comparison of *sdaA* and *sdaB*

The *sdaA* gene was originally considered to start from an ATG followed by an open reading frame encoding 448 amino acids (Su et al., 1989). This sequence was slightly revised later using a GTG start codon which is further upstream than the ATG that had been originally predicted to be the start codon of *sdaA*. The revised start site of *sdaA* also provides for an appropriately placed consensus Shine-Dalgarno sequence in its ribosome-binding site. The new start site is supported by the results of an Edman degradation of the N-terminus of L-SD1 (Su et al., 1993).

I, therefore, compared the sequence of the *sdaB* open-reading frame of the 4.2-kbp fragment with that of the revised *sdaA* sequence using the FASTA program (Pearson, 1988), and found 73% identical nucleotides within the 1368 nucleotides open-reading frame and the *sdaA* gene. The high similarity between *sdaA* and *sdaB* is consistent with the fact that the *sdaB* plasmid hybridized with the *sdaA* gene (Su & Newmen, 1991). Since *sdaA* has been proven to be the gene coding for L-SD1 (Su et al., 1993), it seems clear that the *sdaB* sequence reported here codes for L-SD2.

A comparison of the upstream sequence of the two genes indicated no significant similarity (Fig. 10). This was expected from the fact that the regulation of the two genes is very different, while the enzyme activity coded by the *sdaA* and *sdaB* genes is very similar (Su & Newman, 1991). The

70

```
sdaA    TAGTTAAGCCAGTCGCCGCGTTCCCTCTTACACTATGCGCTGTTATTAGTTCGTTA
        :         :         :    ::        :             :       ::
sdaB    ATCTCCGCAATCTTCTACTCTCTGTTCAGCTAAGTCCTTTCGCGCCGCCTTTCGGG

sdaA    CTGGAAGTCCAGTCACCTTGTCAGGAGTATTATC-GTGATTAGTCTATTCGACATG
        : :       :::        :       ::  :::::  :    :::::::  ::::::: ::
sdaB    CGGCGCTTCCTCCGTTTTAACGAGATGTATTTCCTATGATTAGCGTATTCGATATT

sdaA    TTTAAGGTGGGGATTGGTCCCTCATCTTCCCATACCGTAGGGCCTATGAAGGCAGG
        :: ::  : ::  ::::: :: ::    ::: ::::::::: :: ::  ::::: :: ::
sdaB    TTCAAAATCGGCATTGGCCCTTCCAGTTCTCATACCGTTGGACCAATGAAAGCGGG
```

A. A comparison of sequences upstream of the *sdaA* and

   *sdaB* gene coding regions. The translation start sites

   are underlined. Identical residues indicated with a :

   and gaps indicated with a - in this and subsequent

   figures.


```
sdaA    GAAACCGGTAAGGACATGAACGCCAAATACCGCGAAACCTCACGCGGTGGTCTGGC
        :::::  :::::  ::  :::::::::::::  :::::::::::::::  :::::  ::  :::::
sdaB    GAAACAGGTAAAGATATGAACGCCAAGTACCGCGAAACCTCTCGCGGCGGCCTGGC

sdaA    AATCAAAGTCCAGTGTGACTAATACTTCTTACTCGCCCATCTGCAACGGATGGGCG
        :::  ::  ::       :  : ::  :  :    :         ::       :    :::
sdaB    AATGAAGATCGTTGCCTGCGATTAATCCGTCTCCAAAGGCCTCGTTTTGCGAGGCC

sdaA    AATTTATACCCGCTTTCTCGTCTGCTGTAATATTCCCCACTACACTTCCACTGTTG
        ::         ::       :  : :  :        : :  ::      :  :   :   :
sdaB    TCTTCCCGATTTCTCATCCAGCCGTAGCCTGTTCCGGCATCGAATGTTACCCTTAT
```

B. A comparison of sequences downstream of the *sdaA* and

   *sdaB* gene coding regions. The translation stop sites

   are underlined.


   Fig.10   Sequence comparison of the regions upstream and

            downstream of the *sdaA* and *sdaB* gene coding region


71

sequences downstream of the two genes are different as well (Fig.10).

## 1.4. A comparison of the amino acid sequence of SdaA and SdaB

A comparison of the amino acid sequence of L-SD1 and L-SD2 deduced from their DNA sequences, using the FASTA program, showed that there are 76.5% identical amino acids between the two proteins (Fig.11). The similarity extends through the whole length of both proteins, suggesting that the entire structures of L-SD1 and L-SD2 are very similar.

The *sdaA* gene codes for 454 amino acids, while the *sdaB* gene codes for 455. The added amino acid is very close to the C-terminus in an area which seems to be relatively unimportant for L-SD function as judged by characteristics of fusion proteins (Su et al., 1993).

## 1.5. Map position of the *sdaB* gene

The *sdaB* gene was previously mapped at 60.1 min by conjugation and P1 phage mediated transduction (Su & Newman, 1991). This map position was further supported by the fact that the *sdaB*-carrying plasmid hybridizes with Kohara phage 457 (Kohara et al., 1987), which carries a fragment of the *E.coli* chromosome from the 60.1 min area (Su & Newman, 1991).

Examination of the sequence downstream of *sdaB* (Fig. 9) allowed a more precise localization of *sdaB*. In that sequence, the putative stop site of the *sdaB* gene is followed by *ρ*

```
sdaB  MISVFDIFKIGIGPSSSHTVGPMKAGKQFTDDLIARNLLKDVTRVVVDVYGSLSLTGKGH
      ::  ::  ::  :::::::::::::::::::: :::   ::   :::: :::::::::::::
sdaA  VISLFDMFKVGIGPSSSHTVGPMKAGKQFVDDLVEKGLLDSVTRVAVDVYGSLSLTGKGH

sdaB  HTDIAIIMGLAGNLPDTVDIDSIPSFIQDVNTHGRLMLANGQHEVEFFVDQCMNFHADNL
      ::::::::::::::: : :::::::::: :: ::    :: :: : ::: :: : ::  ::
sdaA  HTDIAIIMGLAGNEPATVDIDSIPGFIRDVEERERLLLAQGRHEVDFPRDNGMRFHNGNL

sdaB  SLHENGMRITAVAGDKVVYSQTYYSIGGGFIVDEEHFGQQDSAPVEVPYPYSSAADLQKH
      ::::::: : :  :: :::: ::::::::::::::::::::::  : :::: ::  :
sdaA  PLHENGMQIHAYNGDEVVYSKTYYSIGGGFIVDEEHFGQDAANEVSVPYPFKSATELLAY

sdaB  CQETGLSLSGLMMKNDVALHSKEELEQHLANVWEVMRGGIERGISTEGVLPGKLRVPRRA
      : ::: ::::: : :  ::::: :      : :: :   : :: ::::::: :: ::::
sdaA  CNETGYSLSGLAMQNELALHSKKEIDEYFAHVWQTMQACIDRGMNTEGVLPGPLRLPRRA

sdaB  AALRRMLVSQDKTTTDPMAVVDWINMFALAVNEENAAGGRVVTAPTNGACGIIPAVLAYY
      :::::::::: ::   ::: : :: :::::::::::::::::::::::::::: :::::::
sdaA  SALRRMLVSSDKLSNDPMNVIDWVNMFALAVNEENAAGGRVVTAPTNGACGIVPAVLAYY

sdaB  DKFIREVNANSLARYLLVASAIGSLYKMNASISGAEVGCQGEVGVACSMAAAGLAELLGG
      : ::  :     ::   : ::: :::::::::::::::::::::::::::::::::::::::
sdaA  DHFIESVSPDIYTRYFMAAGAIGALYKMNASISGAEVGCQGEVGVACSMAAAGLAELLGG

sdaB  NRAQVCIAAEIGMEHNLGLTCDPVAGQVQVPCIERNAIAAVKAVNAARMALRRTSEPRVC
      :::  ::::::::::::::::::::::::::::::::::::: ::: ::::::::::: :::
sdaA  SPEQVCVAAEIGMEHNLGLTCDPVAGQVQVPCIERNAIASVKAINAARMALRRTSAPRVS

sdaB  LDKVIETMYETGKDMNAKYRETSRGGLAMKIVACD
      :::::::::::::::::::::::::::::: : : ::
sdaA  LDKVIETMYETGKDMNAKYRETSRGGLAIK-VQCD
```

Fig.11  A comparison of the amino acid sequence of the SdaB

and SdaA proteins.

This comparison was performed with the FASTA program

(Pearson, 1988).

73

possible open-reading frame, *orfX*. Further downstream I found the coding region of the fucose operon, with the *fucO* gene reading in the opposite direction from the *sdaB* gene and *orfX* (Fig.9, Chen et al., 1989).

The fucose operon has been mapped at 60.2 min by conventional methods (Bachmann, 1990; Chakrabarti et al., 1984). Our sequence thus locates *sdaB* exactly with respect to the *fuc* operon, and is in accord with the map position of both the fucose operon and the *sdaB* gene as determined by conventional mapping.

The EMBL and Genbank databases regularly release a compilation of *E.coli* sequences containing updated information on map locations of these sequences. According to the release of 13 November 1992, the *sdaB* gene must be located between *metZ* at 60.2 min and *fucO* at 60.3 min. On this map, 0.01 min corresponds to 472 base pairs, so that there must be approximately 5 kb between *metZ* and *fucO*. Whereas the sequence of *sdaB* has now been precisely linked to the *fucO* gene, the sequence between *metZ* and *sdaB* still has a gap of 2000 bp (or less). These relationships are indicated in figure 12 which presents the restriction map and gene organization in this area of the *E.coli* chromosome.

Fig.12   Organization of genes in the 60.3 minute region of

the *E.coli* chromosome.

The map location of *fucO* gene and *metZ* gene is specified
according to the information released by EMBL and GenBank
databases in November 1992.

Abbreviations: P, *Pst*I; H, *Hind*III; S, *Sal*I; Pv, *Pvu*II;
B, *Bam*HI; E, *Eco*RI.
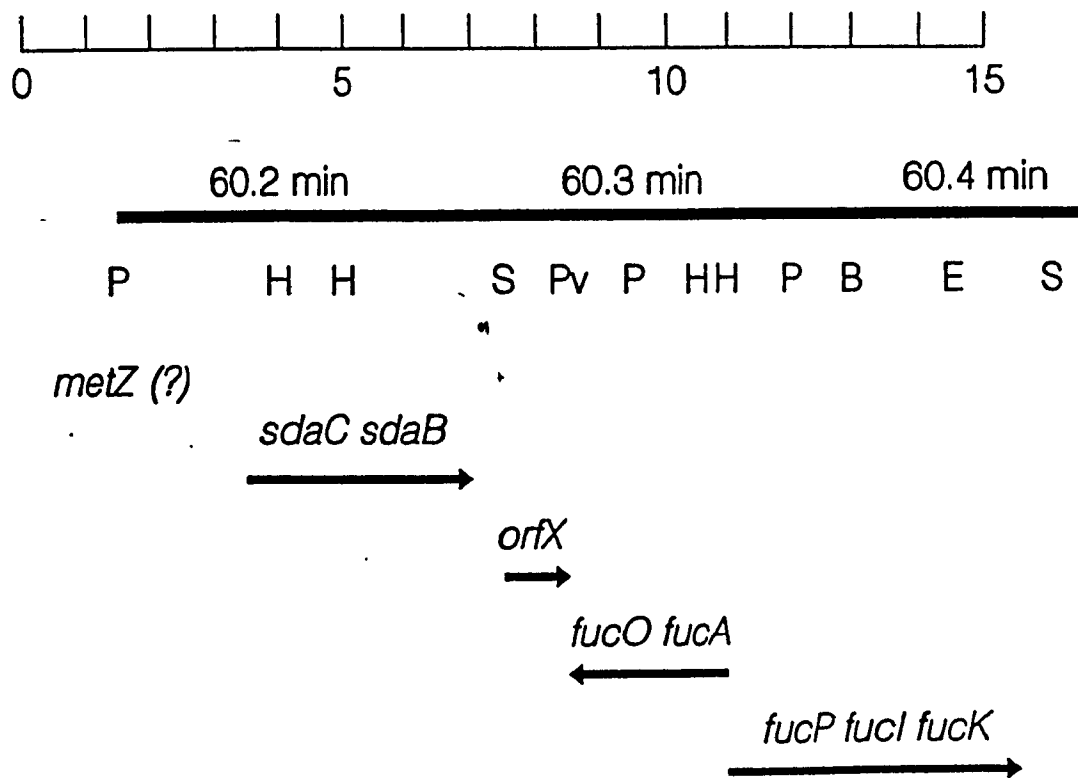
Fig. 12  Organization of genes in the 60.3 minute region
the *E.coli* chromosome

## Part 2. Characterization of the *sdaX* mutation

### 2.1. Cloning of the wildtype *sdaB* gene

The *sdaB* gene on pMES61 was cloned from a mutant strain in which *sdaB* is expressed in glucose-minimal medium, whereas wildtype *sdaB* is only expressed in LB. The mutation, which results in the expression of *sdaB* in minimal medium, may occur in a separate regulatory gene or in the regulatory region of the *sdaB* gene. The fact that the regulatory mutation, known as the *sdaX* mutation, and *sdaB* had been mapped at same position made it likely that *sdaX* is in fact a mutation in *sdaB*.

The cloned *sdaB* gene also showed L-SD activity in minimal medium (as well as LB), further suggesting that the sequence present in figure 9 might carry some alteration corresponding to *sdaX*, and probably in the upstream regulatory region of *sdaB*. To confirm this, I wished to clone and sequence this upstream regulatory region from wild-type *E.coli* and compare it with the same region from the *sdaX* mutant.

To do this, I took advantage of the finding of H. Su that the cloned *sdaB* gene hybridizes specifically with Kohara phage 457, which carries a fragment at 60.1 min of *E.coli* chromosome (Kohara et al., 1987). This indicates that Kohara phage 457 carries the *sdaB* gene.

Therefore, I first isolated a 8.4-kbp *Pst*I fragment from Kohara phage 457 DNA. This fragment was then cut with *Hind*III and a 4.2-kbp *Pst*I and *Hind*III fragment was cloned into

77

pBluescript, forming pWTB6 (for details, see Materials and Methods). The insert on pWTB6 shows exactly the same restriction map as the 4.2-kbp fragment from sdaX that has been sequenced, confirming that the sdaB gene has been cloned from a strain which should be wild-type for sdaB.

## 2.2. Sequence of the wildtype sdaB upstream region

Deletions by exonuclease III were carried out after cutting the plasmid pWTB6 with KpnI and HindIII. Two of the resulting plasmids, in which part of upstream of the sdaB gene has been deleted, were sequenced. The sequence of 300-bp upstream and 100-bp downstream of the ATG start site of the sdaB gene cloned from Kohara phage 457 has been obtained in this way.

There is only a single base difference between the wild-type upstream sequence and the sdaX sequence. This differe ce is seen at the 13th base upstream of ATG translation start site of the sdaB gene (Fig.13). This area, 7-13 bases upstream of the translation start site, is considered to be the ribosome-binding site of the gene. Its sequence closely matches the 3' sequence of 16s rRNA (Shine & Dalgarno, 1974). The nature of the Shine-Dalgarno sequence is one of the major factors influencing the extent to which the message from a gene is translated into protein. A gene with a poor ribosome binding site will generally be much less well translated than one with a more efficient sequence.

78

sequence from *sdaX* mutant     CGTTTTAACGAGATGTATTTCCT<u>ATG</u>

                                          *

sequence from Kohara phage     CGTTTTAACGCGATGTATTTCCT<u>ATG</u>


Fig.13   Location of a mutation in *sdaX*.

       The different nucleotide between the two

       sequences is indicated by *. The translational

       start of the *sdaB* gene site is underlined.

The sequence of the ribosome-binding site of the wildtype *sdaB* gene is significantly different from the consensus Shine-Dalgarno sequence, and one would expect it to be much less efficiently translated.

The *sdaX* mutation lies in this area and converts a cytosine at the 13th base upstream of ATG to an adenine (Fig. 13). The fact that the mutation is in the Shine-Dalgarno region of the *sdaB* gene, suggests that the *sdaX* mutation increases L-SD2 expression by improving translation from the *sdaB* mRNA. Indeed the *sdaX* sequence shows a better match between the mRNA of *sdaB* and 16s rRNA (Fig.14) and therefore converts a poor ribosome-binding site to a considerably better one.

Actually, the expression of the *sdaB* gene in the *sdaX* mutant strain is also increased in LB medium (Su & Newman, 1991), indicating that this mutation results in increased translation of the *sdaB* mRNA in LB as well.

## 2.3. Construction of the *sdaX* mutation by site-directed mutagenesis

While a change in the ribosome-binding site would explain the phenotype of the *sdaX* mutant, this data can not exclude the possibility that there is a further change outside the 300-bp upstream region which has been sequenced in both the wildtype and mutant *sdaB* genes. Since only part of the upstream region of the wildtype was sequenced, one can not be

```
                    3' AUUCCUCCACUAGGUU  5'   16S rRNA

                       ::  :  :  ::

Wildtype sdaB    UUUAACGCGAUGUGUA


                    3' AUUCCUCCACUAGGUU 5'    16S rRNA

                       ::  :::  ::

Mutant    sdaB    UUUAACGAGAUGUGUA
```

Fig.14  Improved match between the ribosome-binding
        site and 16S rRNA in the *sdaX* mutant. The
        mutated base is underlined.

certain that the change at the 13th nucleotide is the only difference between the two sequences, and thus sufficient in itself to cause the increased enzyme activity.

Moreover, the mutant *sdaB* gene and wildtype *sdaB* gene were cloned from different strain backgrounds. It is possible, therefore, that the difference between the two sequences shown in figure 13 is not the only- and perhaps not the critical- change accounting for the *sdaX* mutation and resulting in the expression of *sdaB* in glucose-minimal medium.

To test whether a change at the ribosome-binding site of *sdaB* gene would indeed produce the *sdaX* phenotype, I decided to create a gene which had only that change. To examine this, I changed the 13th base in the wildtype *sdaB* gene cloned on pWTB6 from Kohara phage gene bank from a cytosine to adenine.

This was done by site-directed mutagenesis using a synthesized oligonucleotide corresponding to the sequence of the ribosome-binding site of mutant *sdaB*. After the mutagenesis, plasmids from eight different transformants were isolated. The ribosome-binding site area of these plasmids was then sequenced, using another oligonucleotide which corresponds to the sequence further upstream of ATG site as primer. Two of eight sequences showed an adenine at the 13th base upstream of ATG instead of cytosine (for details, see Materials and Methods).

A 4.2-kbp *Bam*HI and *Hind*III fragment, from the plasmid containing the introduced mutation at the ribosome-binding

site of the *sdaB* gene, was then subcloned onto pBR322 from pBluescript, forming pWTB7 (Fig.3).

L-SD assay of cells carrying plasmid pWTB7 showed that the change at the 13th nucleotide is indeed sufficient to produce the *sdaX* phenotype. Strain MEW51 with the plasmid carrying the mutated *sdaB* gene showed 75 mU of L-SD activity when grown in glucose-minimal medium. This is the same level as is seen with a plasmid carrying *sdaB* cloned from the *sdaX* mutant, and much more than that seen with a plasmid carrying the wildtype gene cloned from Kohara phage, 2 mU. The mutagenized plasmid also allowed the cells to use L-serine as carbon source, as did the *sdaX* plasmid, but not the wildtype one.

I conclude that the mutation previously referred as being in *sdaX* (Su & Newman, 1991) is in fact not in a separate gene, but in the *sdaB* upstream region- specifically in the ribosome-binding site.

## Part 3: The expression and regulation of the *sdaB* gene

### 3.1. Construction of an *sdaB::lacZ* fusion with a wildtype promoter *in vitro*

Strain MEW60 carries λplacMu9 inserted in the *sdaB* gene in an *sdaX* background. I have demonstrated that the *sdaX* mutation is actually a mutation at the ribosome-binding site of the *sdaB* gene. If the ribosome-binding site was altered,

but the remainder of the upstream sequence was not, one would expect that the level of enzyme in the mutant should be higher than that in the wildtype strain, but its response to regulators should be unchanged.

To examine the nature of regulation of the wildtype gene, I needed a strain carrying *lacZ* gene fused to the *sdaB* gene in a wild-type background. To construct this, I made an *in vitro* insertion of *lacZ* gene in frame into the *sdaB* clone derived from the Kohara phage. A *Sal*I fragment containing the truncated *lacZ* from pMC1871 was blunt ended with Klenow enzyme and then ligated to an *Eco*RV fragment from pWBT6, which should result in an in-frame fusion of *lacZ* at codon 63 of the Kohara-phage-derived clone of *sdaB* (for details, see Materials and Methods).

## 3.2. Regulation of expression from the wildtype *sdaB* gene

The regulation of expression of the wildtype *sdaB* gene was studied by measuring β-galactosidase activity in strain MEW88 (WT *sdaB::lacZ*) grown in different medium. It was found that the β-galactosidase activity is very low in glucose minimal medium and is not, or very slightly, induced by L-leucine. An 8-fold induction of the wildtype *sdaB* gene expression was seen in LB medium (Table 3).

Expression from the mutant promoter is thus much higher than expression from the wild-type (Table 3). However, the regulation pattern of both promoters is very similar,

Table 3   Regulation of expression of the wildtype
and mutant *sdaB* gene

| Growth conditions | β-galactosidase activity | |
|---|---|---|
| | MEW88 | MEW60 |
| | WT *sdaB::lacZ* | *sdaB::*λplacMu9 |
| MM | 12 | 30 |
| MM + leucine | 16 | 80 |
| LB | 98 | 680 |
| LB + glucose | 50 | 400 |

Strain MEW88 (with a wildtype ribosome-binding site of *sdaB*) and MEW60 (with an *sdaX* mutation) were grown, subcultured at 37°C and assayed in exponential phase. L-leucine was added 300 μg/ml and glucose concentration in LB is 0.2%. β-galactosidase is reported as Miller units (Miller, 1972). All the data represent the average of at least three experiments.
Abbreviations: MM, minimal medium; LB, LB medium;
WT, wildtype.

suggesting that the *sdaX* mutation causes elevated expression, but does not change the overall regulation.

This is consistent with the fact that the *sdaX* mutation improves the ribosome-binding site and should therefore increase the translation of the *sdaB* mRNA.

### 3.3. The expression of *sdaB* is not affected by ultraviolet irradiation

Treatment of cells with UV irradiation during growth increases L-SD activity 6-7 fold (Newman et al., 1982 ). This increase of L-SD activity is mainly, if not all, due to increased L-SD1 activity, according to the finding of H.Su that transcription of the *sdaA* gene is greatly increased by UV irradiation (Su, 1991).

I examined whether synthesis of the SdaB product is also affected by this treatment. To do this, strain MEW28, which carries an insertion in *sdaA*, was grown in LB overnight and then subcultured into the same medium. After 1.5 hours subculture, cells were exposed to UV light for 30 seconds and then incubated for another 1.5 hours, after which L-SD was assayed. The results showed that the activity of L-SD2 is not affected by UV irradiation (Table 4).

The same question was addressed by assaying $\beta$-galactosidase activity in strains carrying *lacZ* fused to the *sdaB* gene promoters. UV irradiation had no effect on the expression of either the wildtype or mutant *sdaB* gene (Table 4).

Table 4   The effects of environmental factors

on *sdaB* gene expression

| Growth conditions | L-SD activity MEW28 | β-galactosidase activity | |
| --- | --- | --- | --- |
| | | MEW88 | MEW60 |
| | *sdaA::Cm* | WT *sdaB::lacZ* | MT *sdaB::lacZ* |
| LB | 65 | 98 | 680 |
| LB-UV irradiated | 62 | 102 | 693 |
| LB-42°C | 63 | 105 | 525 |
| LB-Anaerobic | 55 | 85 | 426 |

Strain MEW88 (with a wild-type *sdaB* promoter), MEW60 (with a mutant *sdaB* ribosome-binding site) and MEW28 (*sdaA::Cm*) were grown in a variety of conditions known to induce L-SD #1, and β-galactosidase and L-SD assayed. All cultures were grown at 37°C unless otherwise noted. Specific growth conditions were as described previously (Newman, et al. 1982b). The data presented are the average of at least 3 determinations. β-galactosidase is reported as Miller units and L-SD is expressed as μmoles pyruvate synthesized by 0.1 ml of a 100 Klett unit suspension of cells in 35 minutes.

Abbreviations: LB, LB medium; UV, ultraviolet; WT, wildtype; MT, mutant.

## 3.4. The expression of *sdaB* is not affected by growth at high temperature

*E.coli* cells have higher L-SD activity when grown at high temperature in glucose-minimal medium (Newman et al., 1982a). This regulation seems to occur at a post-transcriptional level, since transcription from the *sdaA* promoter is not affected by increase of temperature (Su, 1991).

Does *E.coli* cell also increase its L-SD2 activity when grown at high temperaure? To answer this question, strain MEW28 (*sdaA::Cm*) was grown at 37°C in LB overnight and then subcultured into the same medium, but incubated at 42°C for 3 hours. L-SD assay showed that activity of L-SD2, unlike L-SD1, is not changed by growing at 42°C (Table 4).

Similarly, the β-galactosidase activity from both MEW88 and MEW60 strains, carrying a wildtype *sdaB::lacZ* fusion and mutant *sdaB::lacZ* fusion respectively, is not altered at the higher temperature (Table 4).

## 3.5. *sdaB* expression is not affected by anaerobic growth

The transcription of the *sdaA* gene had been shown to be induced in anaerobic condition (Su, 1991). This increased transcription from *sdaA* likely accounts for most of, if not all, the increased L-SD activity in the wildtype cell in anaerobic conditions.

To test whether the transcription of the *sdaB* gene or the activity of L-SD2 is affected by anaerobiosis, MEW88 and MEW60

88

were grown in LB anaerobically overnight and subcultured in the same conditions for about 4 hours, after which β-galactosidase was assayed. *sdaB* was expressed anaerobically. However unlike *sdaA*, the expression of the *sdaB* gene from either mutant or wild-type promoters was not induced by anaerobic growth. If anything, the expression of *sdaB* was slightly lower during anaerobic growth (Table 4). Direct measurement of the *sdaB* gene product, L-SD #2, led to the same conclusion (Table 4).

### 3.6. Synthesis of L-SD2 in LB is inhibited by glucose

The *sdaA* mutant shows no L-SD activity when grown in glucose-minimal medium, indicating that the *sdaB* gene is not expressed in this medium. L-leucine and glycine, which increase the expression of *sdaA* significantly, have little effect on the expression of *sdaB*. However a large increase in *sdaB* expression was seen in cells grown in LB (Su & Newman, 1991; Table. 4).

To test which component in rich medium induces *sdaB* expression, I first tried all amino acids and some other compounds which may be present in LB medium to see if any of them affect the expression of *sdaB* in glucose-minimal medium, individually or in combination of several ones. None of them showed a significant effect (data not shown). Because LB contains a high concentration of NaCl, I tested whether NaCl might affect *sdaB* expression but it did not (data not shown).

89

I then tried to see if glucose inhibits the expression of the *sdaB* gene during growth in LB, since the difference between glucose-minimal medium and LB may lie in the extent of catabolite repression due to glucose, rather than in higher concentration of exogenous nutrients. This seems reasonable, particularly since both L-SD1 and L-SD2 are thought to be catabolic enzymes.

I tested this idea by growing cells in LB to which 0.2% of glucose was added, this being the glucose concentration used in the minimal medium. Strain MEW60, which carries an *sdaB*::λplacMu9 fusion, showed a 33% decrease in β-galactosidase activity (Table 5). However, expression from the *sdaA* promoter, as judged by the β-galactosidase activity from strain MEW22, was not affected by glucose at all.

### 3.7. Catabolite repression of the *sdaB* gene

Clearly the major factors influencing transcription of the *sdaB* gene were growth in LB and addition of glucose. This suggests that *sdaB* might be under catabolite control, and thus regulated by cAMP and cAMP receptor protein (Crp).

A *crp* null mutation (*crp::Cm*) was kindly provided by Dr.S.Adhya. To test the expression of the *sdaB* gene (and incidentally, also of the *sdaA* gene) in a *crp* mutant background, I transduced this *crp* null mutation into strains MEW60 and MEW22, forming MEW92 (*sdaB*::λplacMu9, *crp⁻*) and MEW93 (*sdaA*::λplacMu9, *crp⁻*) respectively. β-galactosidase

90

## Table 5

Effect of a *crp* mutation on expression of *sdaA* and *sdaB*

| strain | | β-galactosidase | | | L-SD | | |
|---|---|---|---|---|---|---|---|
| | | Min Glu | LB | LB Glu | Min Glu | LB | LBGlu |
| MEW60 | *sdaB::lacZ* *sdaA⁺* | 30 | 680 | 400 | 12 | 54 | 50 |
| MEW92 | *sdaB::lacZ* *sdaA⁺ crp* | − | 19 | 22 | 7 | 56 | 55 |
| MEW22 | *sdaA::lacZ* *sdaB⁺* | 65 | 630 | 550 | 2 | 64 | 7 |
| MEW93 | *sdaA::lacZ* *sdaB⁺ crp* | − | 560 | 516 | 1 | 2 | 3 |

Cultures were grown in the media noted, and assayed in mid-exponential phase. The media used were glucose minimal medium (min glu) LB, and LB with added glucose (LBglu). The data presented are the average of at least 3 determinations. β-galactosidase is reported as Miller units and L-SD is expressed as μmoles pyruvate synthesized by 0.1 ml of a 100 Klett unit suspension of cells in 35 minutes.

activity was then measured in appropriate media (Table 5).

The results showed that activity from the sdaB promoter was decreased about 30-fold by the crp mutation, which confirmed that the expression of the sdaB gene is indeed under the control of catabolite repression.

The expression of the sdaA gene is also induced in LB medium. However, activity from the sdaA promoter was essentially unaffected by the crp mutation. Therefore, I conclude that the major control on sdaB transcription is mediated by cAMP/Crp, while it seems that catabolite repression is not involved in the regulation of the sdaA gene expression.

The β-galactosidase activity from strain MEW60 is not increased when it was grown in minimal medium with other carbon sources instead of glucose, e.g. glycerol (data not shown). This indicated that cAMP/Crp is needed, but not enough, for activating the sdaB gene expression. L-serine does not induce sdaB gene expression either. Another factor, existing in LB medium but not identified yet, must be involved in the activation of sdaB.

## 3.8. The expression of the sdaB gene in an ssd mutant background

L-SD activity is dramatically increased in the ssd mutant strain and this mutant cell grows well on L-serine (Newman et al., 1982b). The transcription of sdaA had been shown to

increased in the *ssd* mutant cells (Su et al., 1989), which would explain at least part of the reason for the increased L-SD activity. It would be interesting to know if the expression of the *sdaB* gene is also affected by this pleiotropic mutation.

To investigate this, an *sdaB*::λplacMu9 *ssd* double mutant, strain MEW90, was constructed. The β-galatosidase assay showed that the expression of the *sdaB* gene is not increased in the double mutant strain in glucose-minimal medium (Table 6), suggesting that the *sdaB* gene expression is not affected by the *ssd* mutation.

I verified this further by constructing an *sdaA*::*Cm ssd* double mutant by transducing *sdaA*::*Cm* into strain MEW86 and measured the L-SD activity of this double mutant, MEW94. Again, L-SD activity of this strain is not affected by the *ssd* mutation whether cells were grown in glucose-minimal medium or in LB (Table 6).

It is clear then that the *ssd* mutation, which greatly increases the expression of the *sdaA* gene, has no effect on the transcription of the *sdaB* gene or on the activity of L-SD2.

### 3.9. The expression of the *sdaB* gene in an *lrp* mutant background

The Lrp protein had been demonstrated to be a DNA binding protein which regulates expression of the numerous genes which

93

Table 6   Effect of the *lrp* and *ssd* mutations on
expression of the *sdaB* gene

| | L-SD | | β-galactosidase | | |
|---|---|---|---|---|---|
| strains | MM | LB | Strains | MM | LB |
| MEW28 | 4 | 55 | MEW60 | 30 | 680 |
| *sdaA::Cm* | | | *sdaB::lacZ* | | |
| MEW94 | 5 | 52 | MEW90 | 47 | ND |
| *sdaA::Cm* | | | *sdaB::lacZ* | | |
| *ssd* | | | *ssd* | | |
| MEW89 | 6 | 50 | MEW91 | 28 | ND |
| *sdaA::Cm* | | | *sdaB::lacZ* | | |
| *lrp* | | | *lrp* | | |

Cultures were grown and assayed as in the preceding table. In strains MEW28, MEW94 and MEW89, L-SD is synthesized from the *sdaB* promoter. In strains MEW60, MEW90, and MEW91, β-galactosidase is synthesized from the *sdaB* promoter. β-galactosidase is reported as Miller units and L-SD is expressed as μmoles pyruvate synthesized by 0.1 ml of a 100 Klett unit suspension of cells in 35 minutes.

Abbreviations: MM, minimal medium; LB, LB medium.

94

constitute the leucine regulon (Lin et al., 1990). R.T.Lin showed that Lrp can bind to the upstream region of the *sdaA* gene and repress its expression (Lin, 1992). This repression can be released by L-leucine which has been shown in *in vitro* experiments to reduce Lrp binding to the *sdaA* promoter. In the *lrp* mutant strain, L-SD activity is increased about 6-fold, due to the release of repression by the Lrp protein, and L-leucine no longer increased L-SD activity (Lin et al., 1990, Lin, 1992).

In order to know if the expression of the *sdaB* gene is also regulated by the *lrp* gene product, i.e. as part of the leucine regulon, I constructed the following strains and assayed L-SD and β-galactosidase.

First, the *lrp* mutation was introduced into strain MEW60 from strain MEW26, forming MEW91 *sdaB*::λplacMu9 *lrp*::Tn10. The β-galactosidase assay showed that transcription from the *sdaB* promoter is essentially unaffected by the *lrp* mutation (Table 6).

An *sdaA lrp* double mutant, MEW89, was constructed by transducing *sdaA*::Cm into MEW26 from MEW28. The L-SD activity of this strain was not affected by the *lrp* mutation (Table 6), confirming that *sdaB* is not regulated by Lrp.

Expression of the *sdaB* gene was shown to be slightly induced by L-leucine (Su & Newman, 1991; Table 3), which would usually suggest that Lrp is involved in regulation of the *sdaB* gene expression. However, since the expression of *sdaB* is

mainly seen in rich medium where Lrp synthesis is much reduced, and since direct assays showed that a mutation of lrp had no effect on *sdaB* expression, it is clear that the Lrp protein does not play an important role in regulation of *sdaB* expression.

## Part 4: Some indications as to the physiological roles of the *sda* gene products

Loss of either the *sdaA* or *sdaB* gene product has relatively little effect on the cell. Both *sdaA* and *sdaB* mutants grow reasonably normally in glucose-minimal medium, though with slightly longer doubling times (Su, 1991). As expected, the *sda* mutants, in particular the *sdaA* mutant, showed much more sensitivity to exogenous L-serine, confirming the involvement of L-SD in L-serine detoxification in the cell.

In the following sections, I investigate the effects of increasing the L-SD activity of the cells by providing an *sda* gene on a multi-copy plasmid.

## 4.1. Influence of the *sda* gene product on the ability of *E.coli* to deaminate L-serine

Wild-type *E.coli* does not use L-serine as a carbon source, but can do so when its L-SD level is increased.

96

Leucine and glycine have been known to increase L-SD activity and *E.coli* grows well on L-serine in the presence of L-leucine and glycine (Pardee & Prestidge, 1955; Newman & Walker. 1982). Pleiotropic mutations in either of two genes, *ssd* or *lrp*, make it possible for cells to use L-serine as sole carbon source. In the presence of either mutation, the L-SD level was greatly elevated (Newman et al., 1982b; Lin et al., 1990). The highest L-SD level so far assayed was in cells with plasmids carrying a cloned *sdaA* gene and, as expected, these cells grew well on L-serine (Su et al., 1989).

L-SD2 had been shown to be very similar in biochemical properties to L-SD1, so that one would expect it to be able to serve the same function as L-SD1. This is indeed true. An *sdaA::Cm*, *sdaX* double mutant, in which L-SD2 is expressed in minimal medium, grows on L-serine in the presence of L-leucine and glycine (Su & Newman, 1991). *E.coli* carrying the *sdaB* plasmid with its mutated upstream region grew well on L-serine (Su et al., 1991). However, the presence of the *sdaB* plasmid with its wildtype promoter, which does not direct synthesis of L-SD2 in minimal medium, did not support growth of *E.coli* on L-serine.

## 4.2. Influence of *sda* genes on the ability of *E.coli* to deaminate L-threonine *in vivo*.

The *tdcB* gene product, known as a threonine deaminase, degrades both L-serine and L-threonine. Whether the *sda* gene

97

products, known as L-serine deaminase, degrade L-threonine is not clear.

The *ilvA* product, L-threonine deaminase, converts threonine to α-ketobutyrate, a precursor of isoleucine. Most of the strains used for the work reported here carry an *ilvA* deletion and require isoleucine for growth. If either L-SD1 or L-SD2 could substitute for the *ilvA* product, the strains should be able to use threonine instead of isoleucine.

To test this, the *ilvA* deletion strain was inoculated into glucose-minimal medium with L-threonine, but without isoleucine. Two concentrations of L-threonine were used in this test, 100μg/ml and 2mg/ml. Cultures which grew were subcultured once again into the same medium. To make sure that they still maintained their requirement for isoleucine, I plated the subcultures on glucose-minimal medium in the absence of threonine with or without isoleucine, thus checking that they maintained their *ilvA* deficiency.

The parent strain could not grow with L-threonine as a source of isoleucine. However, both the plasmids pMES22 and pMES41, carrying the *sdaA* gene and mutated *sdaB* gene respectively, allowed growth at about half the usual rate with 2mg/ml L-threonine, and slower, but still considerable, growth at 100μg/ml L-threonine. All the cells grown maintained their requirement for isoleucine. It is clear then that the *sda* gene products both are also active *in vivo* on threonine.

## Part 5. Evidence that a sequence farther upstream is needed for efficient expression of the *sdaB* gene

Although it is clear that the *sdaB* gene is in the 4.2-kbp *Hind*III and *Pst*I fragment which has been sequenced, there is considerable evidence that this fragment may not contain all the information for *sdaB* gene expression and regulation.

### 5.1. The *sdaB* gene on pMES62 is not induced in LB medium

The plasmid pMES62 carries a 4.2-kbp *Hind*III and *Pst*I fragment, which contains the *sdaB* gene with 970-bp DNA upstream from the start codon of *sdaB* (Fig.9). A comparison of expression from this plasmid with that from plasmid pMES41 which carries another 4.2-kbp of upstream DNA showed that this upstream DNA was involved in *sdaB* regulation. The *sdaB* gene with only 970 bp of upstream DNA is expressed in both glucose-minimal medium and LB medium, but at a low level, and with the same activity in both media. On the other hand, the L-SD activity from pMES41 is higher than that from pMES62 and is greatly induced in LB medium (Table 9).

Another difference between pMES62 and pMES41 is that the expression of *sdaB* from the shortened upstream region (pMES62) is not regulated by Crp protein, while *sdaB* on pMES41 is so regulated. This is consistent with the fact that no Crp-binding site is found in the near upstream region of the *sdaB* gene.

This lack of LB induction and catabolite repression of *sdaB* on pMES62 suggested that the 4.2-kbp *Hind*III and *Pst*I fragment may not carry all the upstream regulatory elements needed for efficient expression of *sdaB* in LB medium or even that another gene which is located in the upstream region close to *sdaB* is responsible for the induction of L-SD2.

Further analysis of the sequence present in figure 9 indicated that there might be part of an open-reading frame before the *sdaB* gene. This possible open-reading frame runs through the *Hind*III site and ends at nucleotide 913, which is only 57-bp from the start codon for the *sdaB* gene. Therefore, the *sdaB* gene could be the 3' gene in an operon.

## 5.2. The map position of inserts in strain CP41 and CP52

In his studies of the Lrp-leucine regulon, R.T.Lin isolated a variety of strains, in which λplacMu9 was inserted into leucine-regulated genes (Lin et al., 1992). Among those were insertions, in strains CP41 and CP52, which seem to influence the expression of L-SD2.

The insertions in both strains CP41 and CP52 mapped very close to the Tn*10* inserted in Singer kit mapping strains CAG12079 and CAG12135, showing linkage of 78 and 94% to *fuc*::Tn*10*, and 50 and 80% to *recD*::Tn*10*, respectively (Lin, personal communication). Since the *sdaB* gene is immediately adjacent to the fucose operon on the *E.coli* chromosome (Fig.9), these results of P1 mediated transduction strongly

100

Table 7    L-SD activity in strains CP41 and CP52
and related strains

| Strains | L-SD activity | |
| | in MM | in LB |
|---|---|---|
| CP41  *sdaC::lacZ* | 11 | ND |
| CP52  *sdaC::lacZ* | 10 | ND |
| MEW26 *sdaA::Cm* | 1 | 63 |
| MEW51 *sdaA::Cm* *sdaB::lacZ* | ND | 2 |
| MEW97 *sdaC::lacZ* *sdaA::Cm* | ND | 6 |
| MEW98 *sdaC::lacZ* *sdaA::Cm* | ND | 13 |

Cultures were grown and assayed as in the preceding table. In strains CP41 and CP52, L-SD in minimal medium is synthesized from the *sdaA* gene. In strains MEW87 and MEW98, L-SD in LB medium is synthesized from the *sdaB* gene. The data presented here are the average of at least 3 determinations. L-SD is expressed as $\mu$moles pyruvate synthesized by 0.1 ml of a 100 Klett unit suspension of cells in 35 minutes.

Abbreviations:  MM,  minimal medium;  LB,  LB medium;

ND,  Not determined.

suggest that the inserts in strains CP41 and CP52 are in or near *sdaB*.

## 5.3. Lower L-SD2 expression in CP41 and CP52 strains

If λplacMu9 in CP41 and CP52 stra·ns is inserted inside the *sdaB* gene, the strains should be deficient in synthesis of L-SD2. In order to test whether CP41 and CP52 could produce L-SD2, I made double mutants by transducing *sdaA::Cm* into CP41 and CP52 from MEW28. This is because L-SD2 is only expressed in LB medium in which both L-SD1 and L-SD2 are present, and so expression from *sdaA* would confuse the results.

The two double mutants, MEW97 and MEW98, showed only 10 and 22% of wildtype L-SD activity in LB (Table 7). It is clear then that both inserts cause a reduction in expression from *sdaB* in LB-grown cells, but do not totally prevent it.

I also measured the L-SD1 made by strains CP41 and CP52 in glucose-minimal medium. The L-SD activity in minimal medium was not affected by these inserts (Table 7), suggesting that the insertions in CP41 and CP52 strains only cause a reduction in expression of L-SD2, and do not affect L-SD1 synthesis.

Therefore, the insertions in CP41 and CP52 are likely to be in the *sdaB* gene or in a 5′ gene of an operon. This 5′ gene might even produce an activator for *sdaB* expression in LB, and the CP41 and CP52 insertions might prevent synthesis of this activator. Such an activator, located just upstream of the structural gene, is seen in the *dsd* and *tdc* operons.

(Bornstein-Frost et al., 1987; Schweizer & Datta, 1989).

## 5.4. The inserts in CP41 and CP52 strains are closely linked to, but not inside the *sdaB* gene.

To test if the insertions were inside the *sdaB* gene, the 2.6-kbp *HindIII-SalI* DNA fragment from pWTB6 containing the *sdaB* gene was used as probe to hybridize to the chromosomal DNA from strain CP41 and CP52.

The results of hybridization showed that the CP41 and CP52 inserts were in fact not in the *sdaB* gene (Lin, personal communication). One band, 2.6-kbp in size, can be seen in each strain tested: MEW1 (the wildtype control strain), CP41 and CP52.

The same test was made with a DNA fragment containing more of the upstream DNA. For this, an 8.4-kbp *PstI* fragment, which carries more than 4.2-kbp DNA upstream of the *sdaB* gene and part of the fucose operon downstream of *sdaB*, was used to do the hybridization. Two hybridized bands could be seen in both CP41 and CP52, while only one band corresponding 8.4-kbp in size was seen in MEW1 (Lin, personal communication). This observation indicated that the gene(s) which carry the insertions in strain CP41 and CP52 are on the same 8.4-kbp fragment as *sdaB*, but not in the *sdaB* gene itself.

Since CP41 and CP52 grow well on fucose and deletion of *orfX* has no effect on *sdaB* gene expression from the plasmid, it is unlikely the inserts are located downstream of *sdaB*. I

103

conclude then that inserts in CP41 and CP52 strains are both in the upstream region of the *sdaB* gene. It seems that these inserts are in a gene situated close to *sdaB*, which is either a 5' gene in the operon containing *sdaB* gene or a gene which codes for an activator of *sdaB* gene expression in LB.


## 5.5. The regulation of β-galactosidase expression in strains CP41 and CP52

Both CP41 and CP52 strains were isolated by screening for genes whose expression is affected by L-leucine (Lin et al., 1992). In fact, β-galactosidase assay showed that the expression of these inserts is indeed induced about 2-3 fold by adding exogenous L-leucine (Table 8). In the *lrp* mutant background, this induction by leucine was not seen. This suggests that the induction of CP41 and CP52 insert gene expression by L-leucine is mediated by Lrp protein.

A much greater effect on CP41 and CP52 insert gene expression was seen in LB and this was not regulated by *lrp*. The β-galactosidase activity in CP41 and CP52 strains was increased greatly by growth in LB. This major induction of expression, more than 10-fold, does not depend on *lrp* function (Table 8). Adding glucose to LB significantly reduced β-galactosidase activity, suggesting that the gene carrying these inserts might be under the control of the Crp protein. This was confirmed by the finding that β-galactosidase

expression in CP41 and CP52 strains is almost completely abolished by introduction of a crp mutation. In strain MEW99 (CP41, crp) and MEW100 (CP52, crp) grown in LB medium, the β-galactosidase activity is 38 and 17 respectively.

The results shown here indicated that the gene which carries the insert in the CP41 and CP52 strains is regulated very similarly to the sdaB gene. This would further suggest that this gene is in the same operon as sdaB gene.

## Part 6. Molecular characterization of the sdaCB operon

### 6.1. Deletion analysis

The 4.2-kbp DNA fragment upstream of the sdaB gene is sufficiently large for several genes to be located in it. The existence and perhaps the nature of such gene(s) could be deduced from the sequence. In order to identify which sequence might be related to sdaB gene expression, it was necessary then to further subclone the 8.4-kbp PstI fragment to find the minimal DNA fragment which contains all the information for sdaB expression and regulation.

To do that, a 6.8-kbp SalI fragment was first subcloned into the pBR322 vector and plasmids containing insertions with both orientations were isolated. The one, pMES63, with the orientation such that the BamHI site of the vector is located upstream of the sdaB gene was used to do deletion analysis (for details, see Materials and Methods, and Fig.5).

105

Table 8. β-galactosidase activity in strains

CP41 and CP52

| Growth conditions | Strains | | | |
| --- | --- | --- | --- | --- |
| | CP41 *sdaC::lacZ* | MEW95 *sdaC::lacZ lrp* | CP52 *sdaC::lacZ* | MEW96 *sdaC::lacZ lrp* |
| MM | 161 | 157 | 35 | 37 |
| MM + L-leucine | 352 | 147 | 114 | 35 |
| LB | 1348 | 1294 | 348 | 347 |
| LB + Glucose | 659 | 789 | 252 | 187 |

Cultures were grown and assayed as in the preceding table. All the data represent the average of at least three determinations. β-galactosidase is reported as Miller units. Abbreviations: MM, minimal medium; LB, LB medium.


By subcloning the 6.8-kbp *Sal*I fragment, I eliminated the

By subcloning the 6.8-kbp *Sal*I fragment, I eliminated the 1.6-kbp *Sal*I and *Pst*I fragment downstream of the *sdaB* gene, i.e. the region which contains the *orfX* and part of the *fucO* gene. At the same time, this subcloning placed several restriction sites upstream of the gene, which would facilitate a deletion analysis. The L-SD activity from the 6.8-kbp fragment is fully induced by growing cells in LB, indicating that it contains the same regulatory sequences as the original fragment.

To make exonuclease III deletions, pMES63 was first cut by *Bam*HI and treated with Klenow enzyme and α-phosphorothioate nucleotides, which will protect the ends from deletion by exonuclease III (Promega Biotec). The linear plasmid was then cut again by *Bgl*II which cut a site at the end of the insert, upstream of *sdaB* (Fig.5). Deletion of the cut plasmid with exonuclease III would remove DNA from the upstream of the *sdaB* gene. After transforming the ligation mix of deleted plasmids and checking plasmid sizes, smaller plasmids formed in this way were isolated and transformed into an *sdaA*, *sdaB* double mutant. The transformants were checked on L-serine-minimal medium for their ability to use L-serine as carbon source, the logic being that any strain which could grow on L-serine must retain the signals for inducing L-SD.

The smallest plasmid (after deletion) which supports growth on L-serine is about 7-kbp in size. In this, the insert is only 2.7-kbp. However, L-SD activity from this plasmid,

Table 9. Characterization of the *sdaCB* operon

| plasmids | | inserts on pBR322 | | | | | Expression of *sdaB* in LB in strains | |
|---|---|---|---|---|---|---|---|---|
| | | | | | | | MEW51 | MEW93 |
| | | *sdaC* | | *B* | *orfX* | | | |
| pMES41 | 8.4-kb | PS | | H H | S | P | 215 | 68 |
| pMES63 | 6.8-kb | S | | H H | S | | 240 | 72 |
| pMES64 | 3.7-kb | | | H | S | | 228 | 62 |
| pMES65 | 3.1-kb | | | H | S | | 78 | 65 |
| pMES66 | 2.7-kb | | | H | S | | 71 | 66 |
| pMES62 | 4.2-kb | | | H | S | P | 77 | 75 |

pMES64, pMES65 and pMES66 are plasmids resulting from partial deletion of *E.coli* DNA carried on pMES63. Host cells (MEW51 *sdaA*::*Cm*, *sdaB*::λplacMu9 and MEW93 *sdaA*::λplacMu9, *crp*::*Cm*) with plasmids were grown in LB with antibiotics at 37°C. L-SD activity was assayed in exponential cells. All the data represent the average of at least three determinations.

pMES66, could not be induced in LB medium. Another plasmid, pMES65, is about 0.4-kbp bigger than pMES66 and showed the same level of L-SD activity. However its expression was not induced in LB either (Table 9).

The plasmid pMES64, which is about 1-kbp bigger than pMES66, shows higher L-SD activity in minimal medium than the two smaller plasmids. The L-SD activity from that plasmid was greatly induced in LB medium, suggesting that pMES64 carries the sequence for full induction of *sdaB* expression in LB (Table 9).

All three plasmids were also transformed into strain MEW93, which is a *crp sdaA* double mutant. The L-SD activity from pMES64 was much lower in this strain while the L-SD activity from both pMES65 and pMES66 was the same and low (Table 9).

It seems then that following conclusion could be obtained from the deletion analysis: the 1.1-kbp DNA fragment further upstream of the *sdaB* gene from the *Hind*III site contains the information for full expression of *sdaB* in LB and for the response to Crp regulation.

## 6.2. Sequence of the *sdaC* gene

Figure 15 shows the sequence of the 1.1-kbp *Hind*III fragment cloned on pMES71 and the rest of the upstream sequence of the *sdaB* gene as repeated from figure 9. The two sequences were connected by sequencing with a synthesized

Fig.15   The DNA sequence of the *sdaC* gene.

The sequence from 1 to 1182 is derived from the 1.1-kbp *Hind*III insert on pMES71. The sequence from 1183 to 2156 is taken from figure 9 (from 3 to 976).

The putative Crp binding site, ribosome-binding site and transcriptional terminator are underlined.

AAGCTTTACCCGGATCAACCTGTTGAAGTGCTGGCTGCCGACCTGGCCCGTGCCGTTCTCCGGTATTGTGGCGGGTAACGTAAAAGAAG **90**

TCGGTATTCGCGCCATTGAAGAGTTTGGTCCTTACAAAATCAACGGCGATAAAGAGATTATGCGTCGTATGGACGACCTGCTACAGGGTT **170**

TTGTTGCCCAGCATCGTATGAAGTTGCCAGGCTCAGCCTACATCCCTTGCTACGAAATCTGCACGTAATCTCCGCTCCCCCGGTGACGTT **260**

TGTTGCCGGGTTTTCTCGTTTTTGGTCACTTACTCATCAACTCATTTCATTTGTTATATGAATGTTTCTTACCACCCTCACGCGACAAAT **350**

ATCATCACAGTTAATATGTCATACAATTTATGTTGCAACGCAAACGTTTCCCTATTTTCATAAACCGTATTTTATCGCTATGAAAAAGAA **440**

ATTATCGCCATGATTAACTAAAAGTATTGATTTTTTCAGTTCAACCTACATATATTGCGCGCCCCGGAAGAAGTCAGATGTCGTTTAATG **530**

GGCAAATATTGCCCTTAAATTCTCTTTTACTTTTGATTTACAGAGTAAAGCGTTGGGATAATCTATCTTCCAAGTAGATT<u>ATTGTATTTG</u> **620**

<u>AGATCAAGATCACTGA</u>TAGATACATAACTTGTGTGTATCTTTCCGCCCTCAAATTATTACGGCGGTAAATGATTAAGCCATCGCCGATAG **710**
                                                                                                   M E

ACAGATTTCATTTTTACGGTCAGG^ACCTTCCCGGGCTGAACTGGCTAAAAGCTGAATTATTTGCATTCCTCC<u>AGGAG</u>AAATAGATGGAA **800**
T  T  Q  T  S  T  I  A  S  K  D  S  R  S  A  W  R  K  T  D  T  M  W  M  L  G  L  Y  G  T

ACGACTCAAACCAGCACGATTGCGTCGAAAGACTCTCGTAGTGCCTGGCGCAAGACAGACACCATGTGGATGCTGGGCCTTTACGGCACG **890**
A  I  G  A  G  V  L  F  L  P  I  N  A  G  V  G  G  M  I  P  L  I  I  M  A  I  L  A  F  P

GCAATCGGCGCGGGCGTGCTGTTCCTGCCAATCAACGCCGGTGTTGGCGGTATGATCCCGCTGATCATCATGGCTATCCTTGCGTTCCCG **980**
M  T  F  F  A  H  R  G  L  T  R  F  V  L  S  G  K  N  P  G  E  D  I  T  E  V  V  E  E  H

ATGACGTTTTTTGCTCACCGCGGCCTGACTCGCTTCGTACTGTCTGGTAAAAACCCGGGCGAAGACATCACCCGAGGTTGTAGAAGAACAC **1078**
F  G  I  G  A  G  K  L  I  T  L  L  Y  F  F  A  I  Y  P  I  L  L  V  Y  S  V  A  I  T  M

TTTGGTATTGGCGCAGGTAAACTGATTACCCTGCTCTACTTCTTCGCTATCTACCCGATCCTGCTGGTTTATAGCGTGGCAATCACCAAT **1168**
T  V  E  S  F  M  S  E  Q  L  G  M  T  P  P  P  R  A  I  L  S  L  I  L  I  V  G  M  M  T

ACCGTTGAAAGCTTCATGTCTCACCAGCTGGGTATGACGGCCACCGCCGCGTGCGATTCTGTCGCTGATCCTGATCGTGGGTATGATGACC **1258**
I  V  R  F  G  E  Q  M  I  V  X  A  M  S  I  L  V  F  P  F  V  G  V  L  M  L  L  A  L  Y

ATCGTTCGCTTCGGTGAGCAGATGATCGTTAAAGCGATGAGTATTCTGGTATTCCCGTTTGTTGGCGTACTGATGCTGCTGGCTCTGTAC **1348**
L  I  P  Q  W  N  G  A  A  L  E  T  L  S  L  D  T  A  S  A  T  G  N  G  L  W  M  T  L  W

CTGATCCCGCAGTGGAACGGCGCTGCACTGGAAACGGCTGTCTCTGGACACTGCATCTGCAACCGGAAACGGTCTGTGGATGACCCTGTGG **1438**
L  D  I  P  V  M  V  F  S  F  N  H  S  P  I  I  S  S  F  A  V  A  K  R  E  E  Y  G  D  M

CTGGCAATTCCGGTAATGGTGTTCTCGTTCAACCACTCTCCGATCATCTCTTCTTTCGCCGTTGCGAAGCGTGAAGAGTACGGCGATATG **1528**
A  E  Q  K  C  S  K  I  L  A  F  A  H  I  M  M  V  L  T  V  M  F  F  V  F  S  C  V  L  S

GCAGAACAGAAATGCTCGAAGATCCTGGCATTCGCACACATCATGATGGTGCTGACCGTAATGTTCTTCGTCTTCAGCTGTGTACTGAGC **1618**
L  T  P  A  D  L  A  A  A  R  E  Q  N  I  S  I  L  S  Y  L  A  N  H  F  N  A  P  V  I  A

CTGACTCCGGCAGACCTGGCTGCGGGCTAAAGAGCAGAACATCTCGATTCTGTCTTACCTGGCTAACCACTTTAACGCACCGGTTATCGCG **1708**
W  M  A  P  I  I  A  I  I  A  I  T  K  S  F  L  G  E  Y  L  G  A  R  E  G  F  N  G  M  V

TGGATGGCTCCGATTATCGCGATTATCGCTATCACCAAATCCTTCCTCGGTCACTACCTGGGCGCACGTGAAGGCTTCAACGGTATGGTG **1798**
I  K  S  L  R  G  K  G  K  S  I  E  I  N  K  L  N  R  I  T  A  L  F  M  L  V  T  T  W  F

ATTAAATCTCTGCGTGGTAAAGGTAAGTCTATCGAAATCAACAAGCTGAACCGTATCACTGCGCTGTTCATGCTGGTAACGACCTGGATT **1888**
V  A  T  L  N  P  S  I  L  G  M  I  E  T  L  G  G  P  I  I  A  M  I  L  F  L  M  P  M  Y

GTTGCCACCCTGAACCCGAGCATCCTGGGTATGATTGAAACCCTGGGCGGTCCAATCATCGCGATGATCCTGTTCCTGATGCCGATGTAC **1978**
A  I  Q  K  V  P  A  M  R  K  Y  Y  S  G  H  I  S  N  V  F  V  V  V  M  G  L  I  A  I  S  A

GCAATTCAGAAAGTACCGGCAATGCGTAAGTACAGCGGTCACATCAGCAACGTATTCGTTGTCGTGATGGGTCTGATTGCAATCTCCGCA **2068**
I  F  Y  S  L  F  S                                                         start of sdaB      M  I

ATCTTCTACTCTCTGTTCAGCTAAGTCCTTTCGCGCCGCTTTCGGGCGGCGCTTCCTCCGTTTT<u>AACGAGATG</u>TATTTCCT ATGATT **2156**

oligonucleotide corresponding to the complementary sequence of the region from nucleotide 1305 to 1323 in this figure as primer.

Analysis of this sequence revealed a large open-reading frame starting with an ATG at nucleotide 803 and ending with a TAA at 2092. A consensus Shine-Dalgarno sequence is found in the ribosome-binding site of the open-reading frame. A consensus sequence for Crp binding was also found in the promoter region of this open-reading frame.

The stop codon of this open-reading frame and the start codon for the *sdaB* gene is only 57 nucleotides apart, strongly suggesting that this gene, *sdaC*, is in the same operon as *sdaB*.

### 6.3. Analysis of the deduced SdaC protein

A striking feature of the amino acid sequence of SdaC protein, as deduced by translation from the DNA sequence, is that this protein is extremely hydrophobic, with 68.1% of all amino acid residues being hydrophobic. This suggested that SdaC protein might be a membrane protein.

Further analysis of SdaC protein using the program developed by Rao and Argos (Rao & Argos, 1986) suggested that this protein indeed contains multiple transmembrane domains (Fig. 16), which strengthens the identification of the SdaC protein as a transmembrane protein.

Fig.16  Membrane-buried helix profile of SdaC

This analysis was performed using the program developed by Rao and Argos, 1986. The x axis is the length of amino acid sequence of SdaC which is 429 amino acids. The y axis value represents the buried-helix parameter as computed by this program. ⊢——⊣ indicates a membrane-buried helix predicted according to the assumptions in this program.

Fig.16  Membrane-buried helix profile of SdaC

114
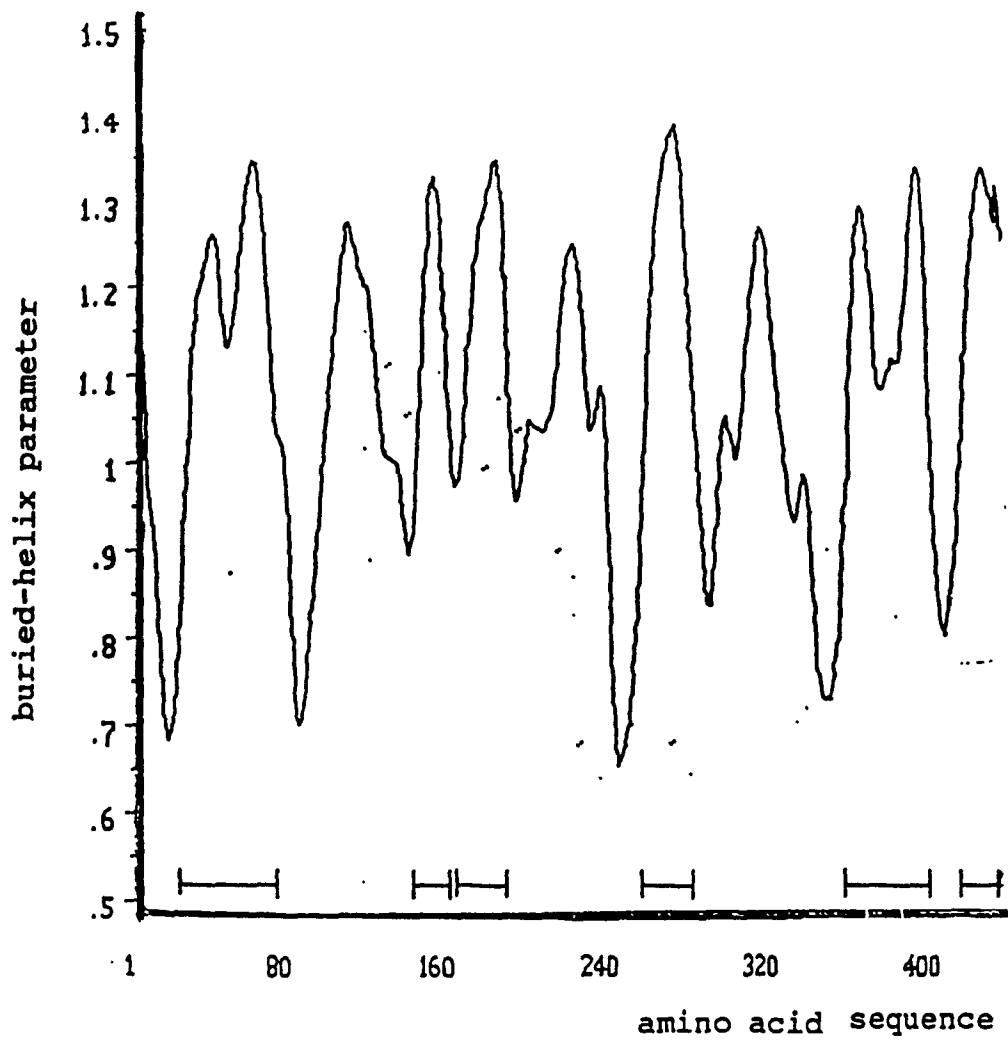
```
SdaC  METTQTSTIASKDSRSAWRKTDTMWMLGLYGTAIGAGVLFLPINAGVGGMIPLIIMA
        :  :    :   : ::: :: : ::: ::::::::::: :: :: :: ::    :
TdcC   MSTSDSIVSSQTKQSSWRKSDTTWTLGLFG1AIGAGVLFFPIRAGFGGLIPILLML

SdaC  ILAFPMTFFAHRGLTRFVLSGKNPGEDITEVVEEHFGIGAGKLITLLYFFAIYPILL
       :: :  :   :: : ::: ::   ::: :::::: :  :: :::::: : :
TdcC  VLAYPIAFYCHRA-ARLCLSGSNPSGNITETVEEHFGKTGGVVITFLYFFAICPLLW

SdaC  VYSVAITNTVESFMSHQLGMTPPPRAILSLILIVGMMTIVRFGEQMIVKAMSILVFP
       :  : :::: :  ::: : :  : :  :     ::    :: :: :: :
TdcC  IYGVTITNTFMTFWENQLGFAPLNRGFVALFLLLMAFVIWFGKDLMVKVMSYLVWP

SdaC  FVGVLMLLALYLIPQWNGAALETLSLDTASATG-NGLWMTLWLAIPVMVFSFNHSPI
       :   : :: : ::: :: :        : ::  : : :: :  :::::: :::
TdcC  FIASLVLISLSLIPYWNSAVIDQVDLGSLSLTGHDGILITVWLGISIMVFSFNFSPI

SdaC  ISSFAVAKREEYG-----DMAEQKCSKILAFAHIMMVLTVMFFVFSCVLSLTPADLA
       ::: : :::::    :  : ::: :  :    ::  :::: :::   : ::  :
TdcC  VSSFVVSKREEYEKDFGRDFTERKCSQIISRASMLMVAVVMFFAFSCLFTLSPANMA

SdaC  AAKEQNISILSYLANHFNAPVIAWMAPIIAIIAITKSFLGHYLGAREGFNGMVIKSL
       :: ::: :::::::::
TdcC  EAKAQNIPVLSYLANHFASMTGTKTTFAITLEYAASIIALVAIFKSFFGHYLGTLEG

SdaC  RGKGKSIEINKLNRITAL

TdcC  LNGLVLKFGYKGDKTKVS
```

Fig.17   A comparison of the amino acid sequence of SdaC and

TdcC

This comparison was performed with the FASTA

program (Pearson & Lipman, 1988)

## 6.4. A sequence comparison between the *sdaC* and *tdcC* gene and their products

The sequence present in figure 15 was compared with the sequences in EMBL and GenBank databases released in April 1993. A very high similarity was found between the *sdaC* gene and the *tdcC* gene in the *tdc* operon. No other significant similarity was found between the *sdaC* gene or other parts of this sequence with the sequences in the gene bank.

A detailed comparison of the *sdaC* and *tdcC* genes, using the FASTA program, revealed that there is 64.2% identity in a stretch of 840 bp. The similarity is found through most of the coding sequence, but the sequences near the 3' end of the coding region shows significant differences. The sequences immediately surrounding the two genes are not similar.

The deduced amino acid sequence of the *sdaC* gene shows 52.6% identity with the *tdcC* gene product (Fig.17). As one might expect from the DNA sequence comparison, the C-termini of the two protein show significant differences.

The discovery that the nucleotide sequences of the *sdaC* and *tdcC* genes are very similar may indicate that they are derived from a common ancestor. The TdcC protein has been demonstrated to be a permease for both L-threonine and L-serine. This implies that SdaC protein is likely to function as a transport protein.

116

## 6.5. No coding sequence is found immediately upstream of the sdaC gene

While there may be an open-reading frame starting upstream and running through the *Hind*III site, or transcribed in the opposite direction, there is no complete open-reading frame upstream of the *sdaC* gene in the sequence presented in figure 15. Since I already showed that this sequence together with the sequence in figure 9 carries all the information for *sdaB* expression, it is clear that the mutation in CP41 or CP52 strain is inside the *sdaC* gene, and that the decrease in *sdaC* transcription causes the reduction of *sdaB* expression.

I also constructed two plasmids, which carry all the sequence upstream of *sdaC* from pMES64 but do not carry the entire *sdaB*. One of these is deficient also in *sdaC* and thus carries only the region upstream of *sdaC*. This was done by deletion starting downstream of the *sdaB* gene, since there is no appropriate cutting site within the *sdaB* gene to permit direct subcloning.

To make this construct, plasmid pMES64 was cut with *Sal*I and the whole plasmid was inserted into pBluescript, forming pMES67 (for details, see Material and Methods). This plasmid showed very high L-SD activity, and the activity was further induced in LB. Exonuclease III deletion on pMES67 was carried out after double digestion with *Sac*I and *Xba*I. Plasmids which did not support growth on L-serine and were smaller than pMES67 were selected. Two of the plasmids were sequenced using

117

M13 primers to find out how many nucleotides had been deleted.

It was found that 48 amino acids from the C-terminus of the SdaB protein were deleted from one of the plasmids. This plasmid showed some L-SD activity in LB, 54 mU which is the level from single copy sdaB, when it was transformed into MEW97 or MEW98 strains. It was originally thought that this is because a gene upstream of sdaB codes for a trans-acting activator, which activates sdaB on the chromosome and therefore some L-SD activity could be synthesized.

The second plasmid, in which the deletion ends at 232 nucleotides upstream of sdaB gene, showed no L-SD activity in MEW97 or MEW98 strain. The deletion on this second plasmid actually ends inside of sdaC gene with 57 amino acids deleted from the C-terminus of the SdaC protein.

It is clear then that there is no gene coding for a trans-acting activator for the sdaB gene upstream of sdaC, and that the inserts in CP41 and CP52 are inside the sdaC gene. The fact that the plasmid, which carries the sdaB gene with 48 amino acids deleted from its C-terminus, showed L-SD activity suggests that the truncated SdaB protein is still functional, but with much lower activity.

# DISCUSSION

In this thesis, I have presented the nucleotide sequence of the *sdaCB* operon of *E.coli*. The *sdaC* gene likely codes for a transport protein for L-serine, while *sdaB* codes for the second L-serine deaminase (L-SD2). I showed that the *sdaB* gene is very similar to *sdaA*, the gene which codes for L-SD1, but the two genes are regulated differently. The nucleotide sequence confirmed that the *sdaB* gene is located upstream of the *fucO* gene with a possible open-reading frame, *orfX*, between them in the 60.2 minute region of the *E.coli* chromosome. A mutation which established synthesis of L-SD2 in glucose-minimal medium was demonstrated to cause a change at the ribosome-binding site of the *sdaB* gene.

In this discussion, I will first discuss the strategy for localizing and sequencing the *sdaB* gene. This will be followed by a consideration of the regulation of *sdaB* gene expression at both translational and transcriptional levels. I will then discuss the strategy used to demonstrate the existence of the *sdaC* gene and the possible function of the SdaC protein. Finally, the possible physiological roles of both L-SD1 and L-SD2 will be discussed.

## Part 1. Localization of the *sdaB* gene on the 8.4-kbp clone

This work began with an 8.4-kbp clone carrying *sdaB*, this clone having been isolated by H.Su (Su & Newman, 1991). L-SD2 is normally expressed in LB, and therefore one cannot clone it by complementing a deficiency. However, H.Su isolated a mutant strain, *sdaX*, in which the regulation of *sdaB* was altered such that L-SD2 is expressed in glucose-minimal medium. This allowed the strain to grow with serine as carbon source, and this characteristic was then used to clone *sdaB*. The 8.4-kbp *Pst*I fragment used in this work was thought to carry the *sdaB* gene, since it conferred high L-SD activity and allowed the cell to grow on L-serine (Su & Newman, 1991).

From the first miniMu clone, an 8.4-kbp *Pst*I fragment complementing *sdaB* deficiency, was subcloned onto pBR322 (Su & Newman, 1991). In order to do deletion and single stranded sequencing, I further subcloned a 4.2-kbp fragment into the phagemid pBluescript multicloning site. The DNA sequence of both strands of this 4.2-kbp fragment carrying the *sdaB* gene was then determined.

### 1.1. Subcloning of the *sdaB* gene

There are two *Hind*III sites inside the 8.4-kbp *Pst*I fragment, located such that *Hind*III digestion will cut the fragment into two *Hind*III-*Pst*I fragments, 3-kbp and 4.2-kbp in size respectively (Fig.2).

The first subclone on pBluescript, containing the 4.2-kbp fragment along with about 3-kbp further DNA from pBR322, showed L-serine deaminase activity, suggesting that the 4.2-kbp fragment carries the *sdaB* gene.

However, the pBR322 DNA proved to be essential for *sdaB* expression. Removal of the 3-kbp pBR322 fragment, the *Pst*I fragment on pMES60 (Fig. 2), from the clone on pBluescript abolished expression of the *sdaB* gene; that is, no L-SD activity was made from the *sdaB* gene in the 4.2-kbp *Hind*III-*Pst*I fragment when it was cloned on pBluescript. Again, if the 3-kbp pBR322 fragment was reinserted into this clone, the *sdaB* gene was expressed.

While the *sdaB* gene was not expressed from the 4.2-kbp fragment in Bluescript, it was expressed when this 4.2-kbp DNA fragment was subcloned into pBR322. The reason why the *sdaB* gene in the 4.2-kbp fragment was not expressed when it is on pBluescript is not clear. It is similarly unclear why the 3-kbp fragment from pBR322 would help to express *sdaB*.

This work has shown that *sdaB* is the downstream gene of the *sdaCB* operon. It is clear that the 4.2-kbp fragment does not contain the whole *sdaCB* operon and therefore may need a promoter from the vector to express the *sdaB* gene. However, a larger fragment (6.8-kbp *Sal*I fragment) on pBluescript containing the entire *sdaCB* operon does not show L-SD activity either. This 6.8-kbp *Sal*I fragment showed L-SD activity when it was cloned on pBR322 in either orientation.

121

On the other hand, the 3-kbp fragment from pBR322 does not provide any promoter for the expression of *sdaB*, because it is located downstream of the *sdaB* gene (inside the *fucO* gene).

The vector, pBluescript, produces many more copies per cell than does pBR322 (Instruction Manual, Stratagene Co.). It may be that too many copies of the *sdaB* gene are toxic to the cells or interfere with their own expression. However, it seems unlikely that this is the reason for nonexpression of *sdaB* on pBluescript. First, cells with pBluescript carrying the 4.2-kbp *Hind*III-*Pst*I fragment grow well, i.e. do not have toxicity problems. Second, it seems that pBluescript carrying both the *sdaB* gene and the 3-kbp fragment from pBR322 still has a high copy number in the cells, but the *sdaB* gene is expressed.

Another possibility is that there are several other promoters on pBluescript which may interfere with *sdaB* expression from its own promoter. Even if this is true, it is difficult to see how the 3-kbp downstream fragment from pBR322 could help to express the *sdaB* gene.

## 1.2. The open-reading frame for the *sdaB* gene

Analysis of the sequence of the 4.2-kbp fragment carrying the *sdaB* gene revealed that there are two complete open-reading frames on the DNA strand as read in the direction from *Hind*III to *Pst*I (Fig.9).

122

The first open-reading frame is presumed to code for *sdaB*. It starts with an ATG and ends with a TAA, coding for 455 amino acids with a predicted molecular mass of 48,655. A possible transcription terminator, CCAAAGGCCTCGTTTTG, was found immediately after the stop codon. Codon usage analysis indicated that this open-reading frame has a codon bias characteristic of *E.coli* protein-coding genes (Gouy & Gauthier, 1982). The sequence in the ribosome-binding site of the open-reading frame is close to the consensus Shine-Dalgarno sequence.

A comparison of nucleotide sequences showed that this complete open-reading frame has very high similarity with the *sdaA* gene, suggesting that they are homologous. Since I was looking for a gene whose product has a function similar to that of the SdaA protein, I concluded that this open-reading frame codes for SdaB. This is also consistent with the fact that *sdaB* hybridizes well with the *sdaA* gene.

## 1.3. Existence of an open-reading frame (*orfX*) of unknown function

The second open-reading frame, *orfX*, is likely to code for a protein. It starts with a GTG, which is 111 nucleotides distant from the stop codon for the *sdaB* gene. The stop codon of this open-reading frame, a TAA, is 54 nucleotides from the stop codon of the *fucO* gene, which is transcribed in the opposite direction from *sdaB* and *orfX*. A putative

123

transcriptional terminator, CGGCGAGGCGGATACGCCGC, is present immediately after the stop codon. A consensus Shine-Dalgarno sequence is present in the expected ribosome-binding site of this open-reading frame.

The *orfX*, coding for 251 amino acids with a predicted molecular mass of 28,102, is likely to constitute a single gene operon, since there is a transcriptional terminator after the *sdaB* gene and a relatively long sequence between *sdaB* and *orfX*.

The function for *orfX* is unknown. Moreover, I do not know if this open-reading frame is expressed or not. Since the subclone of the 6.8-kbp *Sal*I fragment, which does not carry *orfX*, showed the same level of L-SD activity as the original 8.4-kbp *Pst*I fragment, it is clear that the absence of *orfX* has no effect on L-SD2 expression.

## 1.4. Determination of the exact map position of the *sdaB* gene

The conventional mapping methods, conjugation and transduction, are still widely used to map genes on the *E.coli* chromosome. However, in recent years, information from restriction maps and sequences has shown more and more impact on locating genes precisely. By searching DNA sequence databases, the *sdaA* gene was found to be closely linked to the *pabB* gene (Su, 1991). According to the recent release of a compilation of *E.coli* sequences from the EMBL and GenBank databases, it has been estimated that about 45% of the entire

*E.coli* genome has been sequenced by 1992.

Both the *sdaCB* and fucose operons were mapped at 60.1 min by conventional mapping (Su & Newman, 1991; Chakrabarti et al., 1984). My sequence of 1.2-kbp downstream of the *sdaB* gene, from the *Pvu*II site to the *Pst*I site, was found to be identical to that of part of the fucose operon. Therefore, the *sdaB* gene and *fucAO* operon are indeed closely linked with a possible open-reading frame (*orfX*) between them (Fig.9). The *sdaB* gene and *orfX* are transcribed in a clockwise direction, while *fucAO* is transcribed in the opposite direction (Fig.12; Chen et al., 1989).

**Part 2. Comparisons of the *sdaA* and *sdaB* sequences.**

**2.1. A comparison of nucleotide sequences of *sdaA* and *sdaB***

The initial search for possible similar sequence with the sequence of 4.2-kbp fragment was performed using the NCBI blast services in August, 1992. This located an open-reading frame in the *sdaB* sequence, which was found to be highly similar to the *sdaA* gene. The similar region starts after the ribosome-binding sites of the two genes and stopped shortly after the stop codon (Fig.10). A total of 71.8% identical nucleotides were found in the stretch of 1444 nucleotides.

A more detailed comparison of the coding sequences of the *sdaA* and *sdaB* genes was conducted using the FASTA program (Pearson & Lipman, 1988). This showed 73% nucleotide identity

125

with the similar region extending through the entire coding sequences.

Since the *sdaA* gene has been demonstrated to code for the structure of L-SD1, the high similarity between the *sdaA* and *sdaB* genes makes it extremely likely that *sdaB* gene codes for the structure of L-SD2.

The start codon for *sdaB* gene is an ATG, while the one for *sdaA* is a GTG. Both of them end with a TAA. The upstream region of *sdaB* showed no significant similarity with that of *sdaA*. Since the *sdaB* gene is the downstream gene in the *sdaCB* operon, I then compared the upstream noncoding region of the *sdaCB* operon and the upstream noncoding region of the *sdaA* gene. These showed no similarity. These results indicate that the expression of the two L-serine deaminases is regulated differently. The downstream sequence of the *sdaB* gene is also different from that of the *sdaA* gene.

## 2.2. A comparison of the amino acid sequences of SdaA and SdaB

Comparison of the amino acid sequences of SdaA and SdaB as deduced from their nucleotide sequence revealed even higher level of similarity with 76.5 % identity (Fig.11). Therefore, SdaA and SdaB should have similar structures, the more so since both of them deaminate L-serine.

It was thought at the beginning of this work that it would be possible to obtain some knowledge about the relationship between the structure and the function of L-

126

serine deaminating enzymes by comparing two similar enzymes in *E.coli*. However, since these two proteins are so similar through the entire structure, it is difficult to assign the functional domains of L-SD from the results of this comparison.

As one might expect from the high similarity of the two proteins, the two Sda proteins have very similar biochemical properties. L-SD1 and L-SD2 showed similar requirements for pH and concentration of substrate (Su, 1991). Both enzymes are made in an inactive form (Su & Newman, 1991) and activated by an as yet unknown series of enzymes (Newman et al., 1985a). It has been suggested earlier that this activation might involve serinolysis as described by Snell and coworkers (Recsei et al., 1983). This is made considerably less likely by the fact that the serine-serine bond implicated in the *sdaA* gene product is not found in the *sdaB* sequence. On the other hand, the *sdaB* sequence contains a serine-serine bond not seen in *sdaA*.

## 2.3. *sdaA* and *sdaB* may share a common ancestor

There are now known in *E.coli* many pairs or groups of genes which show a very high level of similarity in their nucleotide sequences (Table 1). Those pairs or groups of genes are each believed to be derived from a common ancestor. In many cases, the products of those pairs or groups of genes have similar function, e.g., enzymes show similar catalytic

127

ability. However, as in the case of the *sdaA* and *sdaB* genes, the upstream regulatory regions of most of those genes showed great differences.

The *tufA* and *tufB* gene, coding for elongation factors (EF-Tu), are homologous (An & Friesen, 1980). Most differences in the nucleotide sequences are in the silent third codon position and the products of the two genes vary by only one amino acid residue. However, the nucleotide sequences immediately on the 5' side of the first codon of the structural genes differ completely, suggesting that the regulation of the expression of the two genes is different. The translation start codon for *tufA* is GTG, like that of the *sdaA* gene; the translation start codon for *tufB* is ATG, like that of the *sdaB* gene.

Two nitrate reductases have been reported in *E.coli*, encoded by *narZYWV* and *narGHJI* respectively (Blasco et al., 1989; Blasco et al., 1990). These two transcription units were found to be 73% identical at nucleotide sequences and it was suggested that they are descended from a common ancestor by duplication (Blasco et al., 1990). The expression of the enzymes had been demonstrated to be regulated differently. The *narGHJI* genes are positively regulated by Fnr in anaerobiosis and depend on the presence of nitrate for expression (Stewart, 1988), while *narZYWV* genes may be not controlled by Fnr and their expression is independent of nitrate (Bonnefoy et al., 1987). Indeed, Fnr and NarL boxes were found in the upstream

of *narG* gene, but were absent from the upstream of *narZ* gene (Blasco et al., 1990).

Like the examples mentioned above, the two *sda* genes are extremely similar and their products show similar structure and catalytic activity, strongly suggesting that they are derived from a common ancestor by duplication. The fact that the nucleotide sequences immediately surrounding the two genes are very different indicated that the duplication is restricted to the coding region of the two genes.

The similarity between the *sdaA* gene and *sdaB* gene spreads through the whole coding region. Thus the two proteins must be similar from the amino terminus to carboxyl terminus. It is therefore very difficult to deduce the location of the catalytic site or other important regulatory sites from the comparison of the two sequences. On the other hand, this high similarity through the whole protein between SdaA and SdaB may suggest that there is a strong selective pressure for L-SD to maintain its overall structure.

The lack of similarity between the regions immediately on the 5′ side of the *sdaA* and *sdaB* genes suggested that there are no regulatory sequences held in common by the two genes. This is consistent with the fact that the regulation of expression of the two genes is very different.

**Part 3   Translational control of *sdaB***

## 3.1. Identification of the *sdaX* mutation

Because *sdaB* is not normally expressed in minimal medium, it was cloned indirectly (see discussion, part 1). To do that, H.Su isolated a mutant which did produce L-SD2 in glucose-minimal medium and named the mutated gene *sdaX*. An *sdaB* mutant was then isolated by screening λplacMu9 insertions in the *sdaX* mutant strain and choosing those which were deficient in L-SD activity in both minimal medium and LB. The *sdaX* and *sdaB* mutations were mapped at the same positon, suggesting that they are closely linked, and might even be in the same gene (Su & Newman, 1991). The *sdaB* gene was cloned from the strain carrying the *sdaX* mutations. High L-SD activity from the cloned *sdaB* gene was seen in glucose minimal medium (Su & Newman, 1991), although the activity was much higher in LB, suggesting that the *sdaX* mutation might be carried on this clone.

To identify the *sdaX* mutation, I cloned the 4.2-kbp *Hind*III–*Pst*I DNA fragment containing the wildtype *sdaB* gene from the Kohara phage gene bank. The L-SD activity from this clone was extremely low when cells were grown in glucose-minimal medium, consistent with the fact that the cloned wild-type *sdaB* gene is not expected to be expressed in this growth condition. The wild-type clone was expressed in LB, but to a much lower level than was produced by the *sdaB* clone derived

130

from the *sdaX* strain. The results clearly showed that the 4.2-kbp clone carries both *sdaB* and a regulatory mutation.

I then sequenced part of the upstream region of the wildtype *sdaB* gene. This differed from the sequence obtained from the clone derived from the *sdaX* mutant cells by only one base pair, in that a cytosine at the 13th position upstream of the ATG translational start site was replaced by an adenine in the mutant clone (Fig.14). This point mutation might be the *sdaX* mutation which results in the expression of the *sdaB* gene in minimal medium.

## 3.2. Creation of the *sdaX* mutation by site-directed mutagenesis.

The single base change found in the clone from the mutant cell is inside the Shine-Dalgarno sequence (Shine & Dalgarno, 1974), a location which may explain the phenotype of the *sdaX* mutation. However, this difference between the ribosome-binding site of the mutant and wildtype clones may not be the only difference between the two clones, and therefore may not be responsible for the *sdaX* mutation, for the following reasons:

1. The two clones were derived from different strain backgrounds and the *sdaB* gene in different strains may not be identical as to nucleotide sequence.

2. Only 300 bases of the DNA upstream of the ATG start site of the Kohara phage-derived clone were sequenced. There

131

might be another difference between the two clones somewhere outside these 300 bases, and this might be responsible for the change in expression.

3. The promoter for the *sdaB* gene may be much further upstream, e.g. upstream of the *sdaC* gene, since the *sdaB* gene is a 3' gene in the *sdaCB* operon. This is unlikely, because different *sdaB* expression was seen from the two 4.2-kbp *HindIII-PstI* DNA fragments cloned from the *sdaX* mutant and Kohara phage gene bank respectively. This 4.2-kbp fragment carries only part of the *sdaC* gene, and not the promoter.

I wanted therefore to know whether the single mutation at base pair -13 from the ATG, which resulted in a changed ribosome-binding site, indeed produced the *sdaX* phenotype. To examine this, I replaced the same nucleotide of the wildtype clone, a cytosine 13 nucleotides upstream from the ATG codon, with an adenine as seen in the *sdaX* strain by site-directed mutagenesis. The nature of the sequence change on the plasmids was confirmed by subsequent sequencing using a synthezised oligonucleotide as primer. The plasmid carrying the introduced mutation was then transformed into cells deficient in both L-SD1 and L-SD2. By testing the ability of cells containing mutated plasmid to grow on L-serine and assaying L-SD activity, I showed that the change from the wild-type to the *sdaX* sequence at base-pair -13 indeed reproduced the *sdaX* mutant phenotype. That is, a change in the ribosome-binding site resulted in increased expression in minimal medium.

Table 10

Effect of ribosome binding site mutations

on synthesis of gene products

| Gene | | Sequence of RBS | Effect of mutation | Source |
|------|------|------|------|------|
| *sdaB* | wild type | TAACGCGATGT | Increase | this |
| | mutant | TAACGAGATGT | | work |
| *malT* | wild type | TGAAGTGATTA | Increase | Chapon |
| | mutant | TGGAGTGATTA | | 1982 |
| gene 0.3 | wild type | CACGAGGTAAC | Decrease | Dunn |
| of T4 | mutant | CACGAAGTAAC | | 1978 |
| gene H | wild type | TGAGGTGATTT | Decrease | Gillam |
| of φX174 | mutant | TGAGG-GATTT | | 1980 |

Data is presented for 2 *E. coli* genes, *sdaB* and *malT*, and two phage genes, gene 0.3 from phage T4, and gene H from φX174. Nucleotides mutated are underlined in the mutant sequence. The mutant form of gene H carries a deletion indicated by the symbol -.

Abbreviations: RBS, ribosome-binding site

## 3.3. A consideration of the effect of a mutation in the ribosome-binding site on expression of *sdaB*

Translation from mRNA can be regulated through the extent of the attachment between the mRNA and the ribosomes. It has been suggested that this type of regulation may be a common mode of regulating translation, especially where quick and subtle changes of gene expression are required (Watson et al., 1987).

The ribosome of *E.coli* distinguishes between AUG codons at the start of a gene and those coding for internal methionines by recognizing the ribosome-binding site upstream of the actual start site (Shine & Dalgarno, 1974). The sequence at the ribosome-binding site matches the sequence of the 3' end of 16S rRNA and this match is very important in determining the translational efficiency of the mRNA. Generally, the closer the match, the higher the efficiency of translation will be. Indeed, changes in the amount of protein produced as a result of ribosome-binding site mutations have been described for a variety of genes, some of which are listed in table 10. In the case of the *malT* mutation, the change in the ribosome-binding site brought the sequence closer to the consensus sequence UAAGGAGUGA (McCarthy & Gualerzi, 1990), which resulted in increased *malT* expression (Chapon, 1980). The mutations in both gene 0.3 of '4 and gene H of $\phi$X174 made the sequences less like the consensus sequence, resulting in decreased expression (Dunn et al.,

134

1978; Gillam et al., 1980).

The wildtype Shine-Dalgarno region of the *sdaB* gene, TAAcGcGaTg, differs from the consensus sequence at 4 of 10 nucleotides, and 3 of those are in the most critical central 5 bases. The "*sdaX*" mutation removed one of the mismatches in the central 5 (Fig.14), and thus should improve translation.

## 3.4. Translational regulation of the *sdaB* gene in LB medium

Although wildtype cells grown in glucose-minimal medium make very little L-SD2 (if any), they must make *sdaB* mRNA - otherwise a change in the ribosome-binding site could not, by itself, increase the level of the enzyme.

On the other hand, when they were grown in LB medium, the wildtype cells did make L-SD2. The L-SD2 made in LB could come from the same mRNA which is not translated at all in minimal medium or there might be a different promoter in LB for the *sdaB* gene. Of course, the wildtype cell must make more mRNA in LB than in minimal medium, as would be expected from the known catabolite repression on *sdaB*. It must also translate it more efficiently, since the ribosome-binding site is improved. This is indeed true that the *sdaX* mutant makes more L-SD2 in LB than wildtype strain (Su & Newman, 1991).

Why then is wildtype mRNA not translated in minimal medium, when it is translated in LB?

Besides the match between the ribosome-binding site and 16S rRNA, translational efficiency is also determined by

135

several other factors, in particular the secondary structure of the mRNA. The ribosome-binding site could be trapped in a particular secondary structure and therefore, the ribosome might not be able to initiate protein synthesis. This secondary structure of mRNA may also be affected by changes in the sequence (De Smit & Duin, 1990; McCarthy & Gualerzi, 1990).

The *uncH* gene in *E.coli*, for example, is expressed at very low level. It has been demonstrated that *uncH* mRNA is translated poorly due to the secondary structure around the Shine-Dalgarno sequence of the mRNA. Mutations which decrease the stability of the putative stem-loop structure cause an increase in expression (Pati et al., 1992).

It is likely then, that in cells grown in minimal medium, wildtype *sdaB* mRNA assumes a nontranslatable form, e.g. a secondary structure which stops the initiat. on of translation. The *sdaX* mutation then changes the ribosome-binding site and improves the match; and also changes the secondary structure of the mRNA to a more efficient translated form. Therefore, *sdaB* mRNA in the *sdaX* mutant cell is translated.

The wildtype *sdaB* mRNA is translated less efficiently in LB-grown cell than is the *sdaX* mRNA, since the *sdaX* mutation was shown to improve the translation from the *sdaB* mRNA in LB-grown cells. However, wildtype mRNA is nonetheless highly expressed, suggesting that the mRNA in LB-grown cell is in a translatable form and that improved translation by the *sdaX*

mutation results from its improving the match between the mRNA and the 16S RNA.

Cases are known in which transition from a non-translatable form to a translatable form can be achieved by a translational accessory protein. It has been suggested that interaction between mRNA and other proteins can improve translational efficiency, e.g. RNase III protein at the lambda *cIII* gene (Altuvia et al., 1987). Therefore, it may be that cells grown in LB medium make such an accessory translation factor and increase translation of wildtype *sdaB* mRNA, thus expressing *sdaB* in LB medium.

## Part 4. Transcriptional control of *sdaB* gene expression

### 4.1. Different transcriptional regulation of the *sdaB* and *sdaA* genes.

The high degree of similarity of both nucleotide and amino acid sequences suggested that L-SD1 and L-SD2 have very similar structure and catalytic activity. However, lack of homology in the upstream regions between the two genes indicated that the expression of the two genes is regulated differently.

The *sdaA* gene is transcribed in both glucose-minimal medium and LB. Lrp protein was shown to bind to the upstream region of the *sdaA* gene and repress its expression. This repression by Lrp protein can be released by adding exogenous

137

L-leucine (Lin, 1992). Another global regulatory protein, the product of the *ssd* gene, also regulates transcription from *sdaA* gene expression (Su et al., 1989) by an unknown mechanism. Increased SdaA activity is induced by a wide variety of environmental factors. Some factors, e.g. UV irradiation and anaerobic condition, were shown to induce the transcription of the *sdaA* gene (Su, 1991), while some other factors, e.g. high temperature, may increase L-SD1 expression at a post-transcriptional level (Su, 1991).

On the other hand, *sdaB* is only expressed in LB medium. No amino acids, not even L-serine, has been shown to increase *sdaB* expression in minimal medium. UV irradiation, anaerobic growth condition and elevated temperature also were not inducing. Similarly the *ssd* gene product does not affect the expression of *sdaB* in either minimal medium or LB.

L-leucine induced the β-galactosidase activity from an *sdaX sdaB*::*lacZ* fusion and from the wildtype *sdaB*::*lacZ* fusion, although very slightly (Table 3). This indicates that Lrp protein may be involved in the regulation of *sdaB* gene expression. This however may be a result of regulation of the putative *sdaC* gene. β-galactosidase activity from CP41 and CP52 strains containing the *sdaC*::λplacMu9 fusion was shown to be affected by L-leucine and this induction is mediated by Lrp protein (Table 8). The expression of the *sdaB* gene should be then regulated by Lrp protein as well, since it is in the same operon as *sdaC* and downstream from it.

138

It is surprising, then, that the activity of L-SD2 in the *lrp* mutant background was not changed significantly in either minimal medium or in LB (Table 6). In fact, the L-SD2 activity is barely detectable in glucose-minimal medium with or without exogenous L-leucine.

The *sdaCB* mRNA should be made in minimal medium and probably more such mRNA is made in the presence of L-leucine. That is why the *sdaC* gene is expressed in minimal medium and is induced by L-leucine. However, it is another story for the *sdaB* gene. A consensus Shine-Dalgarno sequence is present in the ribosome-binding site for *sdaC*, while the ribosome-binding site for the *sdaB* gene functions poorly, particularly in glucose-minimal medium. As discussed above, the *sdaB* mRNA is likely in a nontranslatable form and therefore is not translated at all in minimal medium.

In LB medium, I showed that the Lrp protein is not involved in the induction of β-galactosidase expression in CP41 and CP52 strains (Table 8). Therefore, it is unlikely that Lrp protein would influence the high expression of L-SD2 in LB medium. This is consistent with the fact that the expression of Lrp protein itself was highly repressed in LB medium (Lin, 1992).

The fact that the expression of the *sdaA* and *sdaB* genes is regulated completely differently suggests that L-SD1 and L-SD2 might play different physiological roles in the cell.

## 4.2. Catabolite repression of sdaB gene expression

One of the possible functions suggested for L-SD is that it is a catabolic enzyme (Su, 1991). By deaminating L-serine, this reaction would provide the cell with a carbon and energy source. If this is true, one would expect that the expression of L-SD might be controlled by catabolite repression.

While the signal which activates the expression of the sdaB gene in LB medium is not identified, I did find that the expression of the sdaB gene is repressed significantly by adding glucose to LB medium, which is consistent with regulation by catabolite repression.

Catabolite repression is mediated by the cyclic AMP/cAMP receptor protein (Crp) system in E.coli. To test whether this is involved in sdaB regulation, a crp null mutation, crp::Cm, was introduced into our strains. As expected, the expression of sdaB is almost completely abolished in the crp mutant background (Table 6). It is clear then that the major control of transcription of the sdaB gene is catabolite repression. This is consistent with the idea that L-SD2 is an essentially catabolic enzyme, with the function of degrading L-serine when amino acids are plentiful and alternative carbon sources are not.

Growing cells in minimal medium with other carbon sources does not induce the expression of sdaB, indicating that cAMP and Crp alone could not activate sdaB expression. Addition of exogenous L-serine did not activate the sdaB expression

either. Therefore, there must be a signal in LB medium, which together with Crp protein, starts expression of sdaB.

Although sdaB is clearly shown to be regulated by catabolite repression, sequence analysis showed that there is no potential Crp binding site found immediately upstream of the sdaB gene. There are two possible explanations for the absence of a Crp binding site upstream of sdaB. The first one is that sdaB may be the 3' gene in an operon, with the Crp binding site located at the beginning of the operon. The second possibility is that the regulation by cAMP and Crp protein on sdaB is indirect, e.g. through another protein. Synthesis of this protein would be regulated by Crp protein, and this protein would then regulate the expression of sdaB.

I demonstrated in this work that the sdaB gene is indeed a 3' gene in the sdaCB operon and a consensus sequence typical of a Crp binding site was found in the promoter region of the sdaCB operon. This is consistent with the data showing that Crp protein is involved in the regulation of L-SD2 expression.

The sdaA gene is expressed in glucose-minimal medium and its expression is increased in LB as well. However, this is not due to catabolite repression. Expression of the sdaA gene in LB is not affected by glucose. Similarly, introduction of a crp::Cm mutation did not affect expression of the sdaA gene.

This result further indicates that the L-SD1 and L-SD2 must play a different role in the cell and suggests that the major function of L-SD1 is not purely catabolic.

141

### 4.3. A possible mechanism for induction in LB

The expression of both the *sdaA* and *sdaB* genes is greatly induced in LB. However, the mechanism by which this occurs is not clear. The case of the *tdcB* gene, also expressed in rich medium, is similar (Umbarger, 1978). L-serine does not induce *sdaA* or *sdaB* expression, while L-threonine does not, at least alone, induce *tdcB* gene expression (Hobert & Datta, 1983).

It seems then that there might be a common signal to activate those genes, whose products are involved in degradation of amino acids and perhaps other compounds present in rich medium. This common signal may tell the cell that the overall level of amino acids and other nutrients is very high and therefore all the degrading genes should be activated. In this way the cell would not have to sense individual amino acid levels, but would respond to the overall level.

However, even if this is true, it cannot apply to all the genes whose products are involved in the degradation of amino acids. The expression of the *dsdA* gene, coding for D-serine deaminase, has been shown to be induced by its substrate (McFall, 1987), as is the expression of the *putA* gene, coding for proline dehydrogenase (Wood, 1981).


### Part 5    Characterization of the *sdaCB* operon

### 5.1. Characterization of insertions near the *sdaB* gene

In order to screen genes regulated by L-leucine, λplacMu9 was inserted ramdonly into the chromosome of *E.coli* cells. Two

142

such insertions, in strains CP41 and CP52, were mapped by P1 mediated transduction in the same area in which the *sdaB* gene was located (Lin et al., 1992; Lin, 1992). This result led to the suggestion that $\lambda$placMu9 in both strains CP41 and CP52 might be inserted in a gene very close to the *sdaB* gene, or even inside it.

The fact that L-SD activity was much reduced in both strains suggested that they both might be in *sdaB*. Moreover, the regulation of $\beta$-galactosidase expression in CP41 and CP52 strains was found to be extremely similar to that of the *sdaB* gene. Expression of the insert *lacZ* was very low in minimal medium. They differed from *sdaB* in being induced by exogenous L-leucine, in a mechanism demonstrated to be mediated by Lrp protein. However, like *sdaB*, CP41 and CP52 showed the greatest induction during growth in LB (Table 8; Lin et al., 1992) and the *lrp* mutation had no effect on this induction.

Nonetheless the CP41 and CP52 inserts were not in *sdaB*. R.T.Lin showed that the gene inserted by $\lambda$placMu9 in strain CP41 and strain CP52 is actually a gene upstream of *sdaB*, called *sdaC*. This was demontrated by hybridization to chromosomal DNA isolated from CP41 and CP52 strains, using a cloned DNA fragment around *sdaB* gene as probe. The results of hybridization showed that, in strain CP41 and strain CP52, $\lambda$placMu9 was inserted not in the *sdaB* gene, but in the upstream region very close to *sdaB*.

The map position, and the similarity of regulation of the inserts in CP41 and CP52 strains and in the *sdaB* gene, strongly suggested that the gene carrying the CP41 and CP52 inserts is a 5' gene in the same operon as *sdaB*. The level of β-galactosidase activity is different in strains CP41 and CP52, and in strain MEW60 in which λplacMu9 is inserted in the *sdaB* gene of an *sdaX* mutant. In strain CP41, the β-galactosidase activity is much higher than that in the other two strains (Table 8). However, this could happen if the truncated *lacZ* gene were inserted in different positions within the gene or operon. If there is indeed an operon with λplacMu9 in CP41 and CP52 inserted in the gene upstream of *sdaB*, for instance, it might then be that in strain CP41, the λplacMu9 is inserted at the beginning of the gene so that the amino acids fused to the β-galactosidase have little effect on the activity of the protein. In strain CP52, on the other hand, the amino acids fused to β-galactosidase might decrease the activity of the protein significantly. Things are quite different in MEW60 strain in which the β-galactosidase is fused to a downstream gene of the operon. The transcription of the fusion gene, as well as the translation, may simply be weaker than that of the gene upstream of *sdaB*. The fused amino acids may also affect the β-galactosidase activity of the MEW60 strain.

It would be nice to know the exact location of the λplacMu9 insertion in the CP41 and CP52 strains. I actually

144

attempted to do that by sequencing the junction of λplacMu9 and the gene carrying the insert in both CP41 and CP52 strains. Two primers were used, one corresponding to the sequence of the MuS end, which connects the target gene and the truncated *lacZ* gene, and another corresponding to the sequence inside the *lacZ* gene. The template was digested chromosomal DNA isolated from CP41 and CP52 strains and sequencing was carried out by the method described in the double stranded DNA cycle sequencing system (Gibco BRL, Life Tech.Inc.). Though this did not work when I attempted it, it should be possible to determine it in this way.

## 5.2. Characterization of the *sdaCB* operon carried on a plasmid

Expression of *sdaB* is greatly dependent on how much upstream DNA accompanies it. The plasmid with the 4.2-kbp *Hind*III-*Pst*I fragment carrying the *sdaB* gene conferred L-SD activity in glucose-minimal medium, because of the improved ribosome-binding site due to the *sdaX* mutation. The 8.4-kbp *Pst*I fragment, carrying 4.2-kbp more DNA upstream of the *sdaB* gene, produced a higher L-SD activity than the 4.2-kbp fragment.

This difference was accentuated in LB. The expression of *sdaB* from the 4.2-kbp fragment is not induced in LB, while the L-SD activity from the 8.4-kbp fragment is greatly induced by growing the cells in LB.

Another major difference between the 4.2-kbp fragment and the 8.4-kbp fragment is the response of the *sdaB* gene to catabolite repression. The L-SD activity from the 4.2-kbp fragment remains the same in the *sdaA* mutant cells and in the *sdaA crp* double mutant cells. The L-SD activity from the 8.4-kbp fragment, however, is much lower in the *sdaA crp* double mutant than in the *sdaA* mutant cells. That is, the level of *sdaB* expression from the 4.2-kbp clone is very low in LB medium, and Crp protein does not activate its expression. This result clearly suggested that the 4.2-kbp DNA fragment which was originally sequenced does not contain all the information for *sdaB* expression and regulation. This is consistent with the results of the characterization of expression from the insertions in CP41 and CP52.

The 8.4-kbp *Pst*I fragment, from which the L-SD activity is induced in LB and repressed in the *crp* mutant background, carries more than 4.2-kbp of the DNA upstream of *sdaB*. This is a big enough fragment to carry several genes. I therefore further subcloned the DNA fragment so as to locate the minimum sequence that contains all the information for *sdaB* gene expression.

The 6.8-kbp *Sal*I fragment is missing about 1.6 kbp of DNA downstream of *sdaB* including the possible open-reading frame downstream of *sdaB* (*orfX*) and some of the *fucO* sequence (Fig.12). From that construct, the upstream DNA which is not related to *sdaB* gene expression was then deleted using

146

exonuclease III. The 3.7-kbp remaining contained all the information needed for expressing *sdaB* and responding to LB induction and catabolite repression.

## 5.3. The complete sequence of the *sdaC* gene

Of the 3.7-kbp DNA fragment, 2.6-kbp *Hind*III-*Sal*I fragment had already been sequenced during the characterization of *sdaB* itself, the sequence from *Hind*III to *Sal*I site in figure 9. To sequence the rest of the fragment, I subcloned a 1.1-kbp *Hind*III fragment from pMES41 into pBluescript. The sequence surrounding the *Hind*III site was confirmed by sequencing the junction using a synthesized oligonucleotide, corresponding to the sequence downstream of the *Hind*III site, as primer.

Analysis of the sequence indicated that there is indeed an open-reading frame, here known as the *sdaC* gene, upstream of the *sdaB* gene. The *sdaC* gene starts with an ATG and ends with a TAA. A consensus Shine-Dalgarno sequence is present at the site expected for a ribosome-binding site of this open-reading frame. Moreover, a consensus Crp binding site, TATTTGAGATCAAGATCACTGA, was found in the promoter region of this gene, as would be expected from the fact that expression in CP41 and CP52 strains is regulated by Crp protein.

The putative stop codon of the *sdaC* gene is only 57-bp from the start codon of *sdaB* which strongly suggested that the *sdaC* and *sdaB* genes are in the same operon. This is consistent

147

with the fact that the insertion in *sdaC* decreased the expression of the *sdaB* gene. It also explains why I was unable to find a Crp binding site just upstream of the *sdaB* gene, even though expression of the *sdaB* gene was shown to be regulated by Crp.

Although the *sdaB* gene is a 3′ gene in this operon, its expression is unlikely to be totally dependent on the promoter upstream of the *sdaC* gene. First, the insertion in the *sdaC* gene results in reduction of *sdaB* expression, but does not eliminate *sdaB* expression completely. Second, L-SD activity was produced from the subclone carrying the 4.2-kbp fragment, which contains only part of the *sdaC* gene. This seems to indicate that the *sdaB* gene has its own promoter, which may lie inside the coding region of the *sdaC* gene. However, this promoter for the *sdaB* gene is a minor one and does not account for full expression of the *sdaB* gene, since the level of expression from this promoter is very low and is not induced in LB medium. The regulation by LB and the catabolite repression of *sdaB* are mediated by the major promoter in the upstream of the *sdaC* gene.

This arrangement - that the downstream gene(s) in a operon have their own promoters which can be located inside of the coding region of the upstream gene- is very common. For instance, in the $\beta$ operon for RNA polymerase $\beta$ and $\beta'$ subunits, the promoter of the gene for $\beta'$ subunit is at the end of the gene coding for the $\beta$ subunit (Burton et al.,

1983). In a more closely related case, there is also a minor promoter for *tdcB* in the *tdcABC* operon which is located inside the open-reading frame for the *tdcA* gene (Goss & Datta, 1988).

## Part 6  The *sdaC* gene product, a previously undescribed possible transport protein for L-serine

### 6.1. SdaC, a hydrophobic protein

The *sdaC* open-reading frame codes for 429 amino acids with a predicted molecular mass of 46,906. The SdaC protein is highly hydrophobic, with 68.1% of all amino acid residues hydrophobic in nature (Table 11), suggesting that it might be a membrane protein. The hydropathy profile of the amino acid sequence revealed that SdaC protein has multi-transmembrane domains (Fig.16).

### 6.2. A comparison of the *sdaC* and *tdcC* genes and their products

The *sdaB* gene likely codes for the structure of L-SD2, since it is likely homologous to the *sdaA* gene which has been shown to code for the structure of L-SD1. A sequence comparison reveals that *sdaC* is very similar in nucleotide sequence to the *tdcC* gene, a gene known to code for a membrane protein. This result indicated that the *sdaC* and *tdcC* genes might be derived from the same ancestor gene.

**Table 11  Amino acid composition of the SdaC protein**

Amino acid residues

| polar | | % | nonpolar | | % |
|---|---|---|---|---|---|
| Arg(R) | 11 | 2.5% | Phe(F) | 27 | 6.2% |
| Lys(K) | 16 | 3.7% | Leu(L) | 49 | 11.4% |
| Asp(D) | 6 | 1.3% | Ile(I) | 48 | 11.1% |
| Asn(N) | 14 | 3.2% | Val(V) | 30 | 6.9% |
| Glu(E) | 15 | 3.4% | Trp(W) | 7 | 1.6% |
| Gln(Q) | 7 | 1.6% | Ala(A) | 42 | 9.7% |
| His(H) | 8 | 1.8% | Tyr(Y) | 11 | 2.5% |
| Thr(T) | 27 | 6.2% | Gly(G) | 32 | 7.4% |
| Ser(S) | 30 | 6.9% | Met(M) | 28 | 6.5% |
| | | | Cys(C) | 2 | 0.4% |
| | | | Pro(P) | 19 | 4.4% |
| Subtotal | 134 | 28.1% | Subtotal | 295 | 68.1 |

The deduced amino acid sequence of the SdaC protein is also very similar to that of the TdcC protein (Fig.17). The SdaC protein has 429 amino acids, while the TdcC protein is two amino acids longer than the SdaC protein. 49.5% amino acids of SdaC protein are found identical to those of TdcC protein (Fig.17).

TdcC protein appeared to be an integral membrane protein with several membrane-spanning domains exhibiting a striking similarity with other bacterial permeases (Sumantran et al., 1990). Analysis of SdaC protein, using the program developped by Eisenberg et al. (Eisenberg et al., 1984), revealed that it has 11 membrane associated α-helices (Table 12), while 10 such α-helices were predicted for the TdcC protein. Those membrane associated α-helices between SdaC and TdcC proteins are remarkably similar (Table 12), suggesting that their structure is similar and therefore they might serve similar functions in the cell.

Although both sdaB and tdcB are associated in an operon with a gene coding for a membrane protein, the two operons differ in that the sdaC gene is the first gene in the sdaCB operon, while the tdcC gene is the last gene in the tdcABC operon (Goss et al., 1988). A positive regulatory gene was found upstream of the tdcABC operon and transcribed divergently (Schweizer & Datta, 1989), but no such gene was found in the upstream of sdaCB operon. The integration host factor was required for positive regulation of the tdc operon

Table 12    Comparison of membrane associated α-helices

between SdaC and TdcC

| Proteins | | Positions | | α-helices |
|---|---|---|---|---|
| TdcC | (1) | 22 to | 42 | TTWTLGLFGTAIGAGVLFFPI |
| | | | | : : ::: :::::::::: :: |
| SdaC | (1) | 23 to | 43 | TMWMLGLYGTAIGAGVLFLPI |
| TdcC | (2) | 44 to | 64 | AGFGGLIPILLMLVLAYPIAF |
| | | | | :: ::: : :: : : |
| SdaC | (2) | 47 to | 67 | VGGMIPLIIMAILAFPMTFFA |
| TdcC | (3) | 96 to | 116 | GVVITFLYFFAICPLLWIYGV |
| | | | | :: ::::::: : : : : |
| SdaC | (3) | 100 to | 120 | LITLLYFFAIYPILLVYSVAI |
| TdcC | (4) | 139 to | 159 | FVALFLLLLMAFVIWFGKDLM |
| | | | | : : : |
| SdaC | (4) | 141 to | 161 | ILSLILIVGMMTIVRFGEQMI |
| TdcC | (5) | 162 to | 182 | VMSYLVWPFIASLVLISLSLI |
| | | | | :: :: :: : : : :: |
| SdaC | (5) | 164 to | 184 | AMSILVFPFVGVLMLLALYLI |
| TdcC | (6) | 206 to | 226 | ILITVWLGISIMVFSFNFSPI |
| | | | | : :: : ::::: |
| SdaC | (6) | 202 to | 222 | ATGNGLWMTLWLAIPVMVFSF |
| TdcC | (7) | 260 to | 280 | MLMVAVVMFFAFSCLFTLSPA |
| | | | | :: :::: ::: |
| SdaC | (7) | 250 to | 270 | ILAFAHIMMVLTVMFFVFSCV |
| TdcC | (8) | 311 to | 331 | AITLEYAASIIALVAIFKSFF |
| | | | | : : ::: :: ::: |
| SdaC | (8) | 298 to | 318 | APVIAWMAPIIAIIAITKSFL |
| TdcC | (9) | 365 to | 385 | ISMIFIMGSTWVVAYANPNIL |
| | | | | : : :: :: :: |
| SdaC | (9) | 348 to | 368 | LNRITALFMLVTTWIVATLNP |
| TdcC | (10) | 388 to | 408 | IEAMGAPIIASLLCLLPMYAI |
| | | | | :: : :::: : : : |
| SdaC | (10) | 370 to | 390 | ILGMIETLGGPIIAMILFLMP |
| SdaC | (11) | 407 to | 427 | ISNVFVVVMGLIAISAIFYSL |

152

(Wu & Datta, 1992), but its involvement in *sdaCB* expression has not been examined.

## 6.3. A possible function for the SdaC protein

The TdcC protein has been shown to be a transport protein for both L-serine and L-threonine, and also transports L-leucine to a small degree. The expression of this L-threonine and L-serine permease is only seen anaerobically, especially in rich medium (Sumantran et al., 1990).

The high similarity between the TdcC and SdaC proteins makes it very likely that the product of *sdaC* gene is also a transport protein. Further, its genetic association with the *sdaB* gene suggests that *sdaC* codes for a new L-serine permease. The *sdaC* gene is in the same operon as *sdaB* gene and therefore regulated in the same manner. One would therefore expect that this L-serine permease is mainly expressed in LB medium and is under catabolite repression.

It is not a surprise that the *sdaB* gene is associated with a gene that codes for a transport protein. In fact, it is common in *E.coli* that a catabolic enzyme is associated with a permease for the substrate of the catabolic enzyme. In the *lac* operon, the *lacZ* gene is associated with *lacY*, which codes for a lactose permease (Beckwith, 1987); in the *fuc* operon, the *fucA* gene is associated with the *fucP* gene, which codes for a permease for fucose (Chen et al., 1987).

153

At least three amino acid utilization systems have been shown to be organized in the same way. As demonstrated in this work, the *sdaB* gene is associated with *sdaC* gene for L-serine degradation. The *tdcB* gene and *tdcC* genes are in the same operon as well (Schweizer & Datta; 1990). Another well-established system, the proline utilization system, also consists of a structural gene for a catabolic enzyme and a gene for a transport protein (Wood, 1981). The *putA* gene, coding for a proline dehydrogenease, is associated with the *putP* gene, coding for the proline carrier (Nakao et al., 1987). However, these two genes, unlike *sdaB* and *sdaC*, are transcribed divergently.

## 6.4. The transport systems for L-serine in *E.coli*

The possible L-serine transport protein, the product of the *sdaC* gene, as elucidated in this work, is induced by L-leucine via ɪ ₋p protein, and is further induced during growth in LB. The degree of specificity of this system for L-serine is not known, but it is very common that a transport system is used by several amino acids.

L-serine transport is very complex in *E.coli*, with at least four other systems already characterized. The first system transports L-serine and L-threonine, and has been demonstrated in both membrane vesicles (Lombardi & Kaback, 1972) and whole cells (Templeton & Savageau, 1974). This is the major system for l-serine uptake and has been demonstrated

154

to be an Na$^+$-coupled cotransport system (Hama et al., 1987).

The second system mediates transport of L-serine, L-threonine, L-alanine and L-leucine (Robbins & Oxender, 1973). This system is sensitive to osmotic shock and is repressed by L-leucine. Therefore it was believed that it is a same system as one of the LIV (leucine-isoleucine-valine) system and it is not an ion-coupled cotransport system.

The third system is very specific for L-serine taken up by a mechanism of H$^+$-serine cotransport, and it is induced by L-leucine. Some other amino acids may also be involved in full induction of this system (Hama et al., 1988). It is also suggested that the higher level of L-serine transport activity in LB medium is due to this H$^+$/L-serine system (Kayahara et a., 1992).

The newly described TdcC system, which transports both L-threonine and L-serine, has been demonstrated to be different from all the other systems (Sumantran et al., 1990).

The LIV-1 and LIV-2 systems were mapped at 74-76 minutes region and 74-78 minutes region respectively (Antonucci & Oxender, 1986). The tdcC gene was mapped at 68 minutes (Schweizer & Datta, 1990). It is clear then that the possible L-serine transport system described here is different from these two, since sdaC is mapped at 60.1 minute (Su & Newman, 1991). Moreover, the two LIV systems are repressed by L-leucine, while the expression of SdaC protein is induced by L-leucine.

It is not clear, however, if the SdaC system is different from the other two systems. The system described by Lombardi and Kaback (Lombardi & Kaback, 1972) was found in minimal medium, where the expression of SdaC is very low. This might indicate that they are different.

The permease described by Hama et al. (Hama et al., 1988) is very similar to the system described in this work. Both of them are induced by L-leucine, likely via Lrp protein, and further induced in rich medium (Kayahara et al., 1993). However, the Hama permease did not decrease in medium to which glucose is added (Hama et al., 1988), which suggests that it is not regulated by catabolite repression. This is not entirely convincing because the effect of glucose on *sdaC* is not seen under the conditions they used. For instance, the expression of SdaC protein is not significantly induced in minimal medium with other carbon sources which are less efficient in catabolite repression than glucose. Therefore, the catabolite repression on SdaC expression is only seen in rich medium. Even in LB medium, only partial repression was detected by glucose on SdaC expression, though the effect of a *crp* mutation was profound.

## Part 7   The possible physiological roles of L-SD1 and L-SD2

The catalytic activity of the products of the *sdaA* and *sdaB* genes has been well established. However, their physiological role is not completely understood. Both the *sdaA*

and *sdaB* mutants grow normally in minimal medium and in LB, suggesting that they are not necessary for cell metabolism at least in these conditions. On the other hand, however, the existence of two L-serine deaminases and the complicated regulation of their expression indicates that their function must be very important for the cell. They may not be necessary for cell survival, but their function may help the cell to achieve an optimum and well-regulated growth. The slightly longer doubling time of the *sda* mutants suggests a subtle role for L-SD (Su, 1991; Zhang, personal communication).

The extremely high similarity between L-SD1 and L-SD2 suggested that these two proteins have the same enzymatic function. The possible physiological significance of L-SD has been discussed in detail by Su (Su, 1991). In this section, I will focus on the possibility that L-SD1 and L-SD2 may play different roles in the cell. I suggest that L-SD1 is likely the major enzyme involved in L-serine detoxification, while L-SD2 functions as a catabolic enzyme.

L-serine has been shown to be toxic to *E.coli* (Cosloy & McFall, 1970; Isenberg & Newman, 1974; Uzan & Danchin, 1978), although it is a necessary amino acid for protein synthesis. The inhibition of growth by a low level of L-serine could be released by adding L-isoleucine, L-threonine, 2-keto-butyric acid, or homoserine (Hama et al., 1990). However, the inhibition by a higher level of L-serine cannot be reversed by L-isoleucine and the mechanism of inhibition by the high level

157

of L-serine is not understood (Su, 1991). In any case, it is clearly necessary for the cell to control its L-serine concentration.

The *sdaA* mutant is much more sensitive to exogenous L-serine than the wildtype cell (Su, 1991), suggesting that L-SD1 is indeed involved in L-serine detoxification. Moreover, cells carrying *sdaA* on a plasmid can tolerate much higher L-serine. On the other hand, the *sdaB* mutant showed almost the same level of sensitivity to L-serine as the wild-type (Zhang, personal communication), indicating that L-SD2 is not important in L-serine detoxification. However it should be noted that these tests were made in minimal medium where *sdaB* is not normally expressed.

On the other hand, a plasmid carrying both *sdaB* and *sdaX* allows the cell to grow well in the presence of a high level of L-serine, just as the *sdaA* plasmid did. This is consistent with the idea that the two enzymes have the same catalytic function.

The *sdaA* gene is expressed in both minimal medium and its expression is largely induced in LB, while the *sdaB* gene is only expressed in LB. Therefore, one might expect that L-SD2 is also involved in L-serine detoxification in LB, and simply acts as a supplement to L-SD1. This may not be true, since the expression of *sdaB* is not only controlled by the signal in LB, but also by the cAMP/Crp system. In other words, the expression of L-SD2 is regulated as a catabolic enzyme,

and not according to the need for controlling the L-serine pool in the cell. The fact that the gene coding for L-SD2 is associated with a gene coding for a transport protein further suggested that the major fuction of L-SD2 is as a catabolic enzyme.

## Part 8 Summary

The main focus of this work was the molecular characterization of the second L-serine deaminase (L-SD2).

L-SD2 is encoded by the *sdaB* gene, which is only expressed in LB medium. In earlier studies, the *sdaB* gene was cloned from an *sdaX* mutant background, in which L-SD2 is expressed in glucose-minimal medium. Since the *sdaX* mutation and the *sdaB* mutation were mapped in the same position, and the cloned *sdaB* gene is expressed in minimal medium, it is likely then that this clone carries a mutant *sdaB* gene.

I first sequenced 4.2-kbp *Hind*III-*Pst*I fragment of cloned DNA, which conferred L-SD activity when it is cloned on pBR322, but not on pBluescript. The sequence of an open-reading frame in the 4.2-kbp fragment was found to be homologous with that of the *sdaA* gene, which codes for first L-serine deaminase (L-SD1), suggesting that this open-reading frame, *sdaB*, codes for L-SD2. This is consistent with the fact that the *sdaB* clone hybridized well with the *sdaA* gene.

159

The amino acid sequence of L-SD2 is very similar to that of L-SD1 as well, suggesting that these two proteins are similar in structure.

The wildtype *sdaB* gene was cloned from the Kohara phage gene bank. By sequencing the upstream region of the wildtype *sdaB* gene, I showed that the *sdaX* mutation is the result of a change in the ribosome-binding site of the *sdaB* gene. This was further confirmed by reconstructing the mutation from a wildtype *sdaB* gene by site-directed mutagenesis and showing that this reproduces the mutant phenotype.

By constructing a wildtype *sdaB*::*lacZ* fusion, I showed that the expression of wildtype *sdaB* and mutant *sdaB* gene are regulated in the same pattern, though the mutant *sdaB* always has higher expression.

The expression of the *sdaA* and *sdaB* genes was found to be regulated very differently. Neither the *ssd* nor the *lrp* mutation has any significant effect on *sdaB* expression. The environmental factors, which have been shown to induce *sdaA* gene expression, do not increase L-SD2 activity. On the other hand, the expression of the *sdaB* gene was regulated by catabolite repression, whereas a *crp* mutation has no effect on *sdaA* gene expression.

Characterization of mutants with insertions upstream of the *sdaB* gene and characterization of plasmids carrying *sdaB* with a long upstream sequence suggested that *sdaB* is a 3′ gene in a operon. The *sdaC* gene was then found to be the first gene

160

in the *sdaCB* operon by sequencing further upsteam of the *sdaB* gene. Analysis of the deduced SdaC protein suggested that it is a transmembrane protein. The high similarity between SdaC and TdcC proteins indicated that SdaC might function as a transport protein. Because of its genetic location next to the gene coding for L-SD2, I propose that the *sdaC* gene codes for an L-serine permease.

# REFERENCE

Altuvia, S., Locker-Giladi, H., Koby, S., Ben-Nun, O. and Oppeneheim, A.B. 1987. RNaseIII stimulates the translation of the cIII gene of bacteriophage lambda. *Proc. Natl. Acad. Sci.* USA.**84**: 6511-6515.

Ambartsoumian,G., R.D'Ari, R.T.Lin, and E.B.Newman. 1993. Alter l amino acid metabolism in *lrp* mutants of *E.coli* K-12 and their derivatives. Submitt, J.

An, G., and Friesen, J.D. (1980). The nucleotide sequence of tufB and four nearby tRNA structural genes of *Escherichia coli*. *Gene*. **12**:33-39.

Antonucci,T., and Oxender,D. 1986. The molecular biology of amino-acid transport in bacteria. *Adv.Microb.Physiol*. **28**:145-180.

Bachmann, B.J., 1990. Linkage map of *Escherichia coli* K-12, edition 8. *Microbiol. Rev.* *54*:130-197.

Beckwith,J. 1987. The lactose operon, p.1444-1452. In F.C.Neidhardt, J.L.Ingraham, K.B.Low, B.Magasanik, M.Schaechter, and H.E.Umbarger(ed.), *Escherichia coli* and

*Salmonella typhimurium*: cellular and molecular biology, vol.2. American Society for Microbiology, Washington,D.C.

Belfaiza,J.N., C.R.Parsot, A.Martel, C.B.de Tour, D.Margarita, G.Cohen, and I.Saint-Girons. 1986. Evolution in biosynthetic pathways: two enzymes catalyzing consecutive steps in methionine biosynthesis originate from a common ancestor and possess a similar regulatory region. *Proc.Natl.Acad.Sci.* USA **83**:867-871.

Blakley,R.L. 1969. The biochemistry of folic acid and related pteridines. Elsevier/North Holland Publishing Co. Amsterdam.

Blasco,F., C.Iobbi, G.Giordano, M.Chippaux, and V.Bonnefoy. 1989. Nitrate reductase of Escherichia coli: completion of the nucleotide sequence of the *nar* operon and reassessement of the role of the $\alpha$ and $\beta$ subunits in iron binding and electron transfer. *Mol Gen.Genet.* **218**:249-256.

Blasco, F., Iobbi, C., Ratouchniak, J., Bonnefoy, V., and Chippaux. (1990). Nitrate reductases of *Escherichia coli*: Sequence of the second nitrate reductase and comparison with that encoded by the *narGHJI* operon. *Mol Gen Genet.* **222**:104-111.

Bolivar, F., R.L. Rodriguez, P.J. Greene, M.C. Betlach, H.L.

Heyneker, H.W. Boyer, J.H. Cross, and S. Falkow. 1977. Construction and characterization of new cloning vehicles. II. A multi-purpose cloning system. *Gene* **2**:95-113.

Bollinger,J., C.Park, S.Harayama, and G.L.Hzelbauer. 1984. Structure of Trg protein: Homologies with and differences from other sensory transducers of *Escherichia coli*. *Proc.Natl.Acad. Sci*. USA **81**:3287-3291.

Bonnefoy,V., J.F.Burini, G.Giordano, M.C.Pascal, and M.Chippaux. 1987. Presence in the "silent" terminus region of the *Escherichia coli* K-12 chromosome of cryptic gene(s) encoding a new nitrate reductase. *Mol. Microbiol*. **1**:143-150.

Bornstein-Frost,S., E.McFall, and S.Palchaudhuri. 1987. *In vivo* D-serine deaminase transcription start sites in wildtype *Escherichia coli* and in *dsdA* promoter mutants. *J.Bacteriol*. **169**:1056-1060.

Burton,Z.F., C.Gross, K.Watanabe, and R.Burgess. 1983. the operon that encodes the sigma subnuit of RNA polymerase also encodes ribosomal protein S21 and DNA primase in *E.coli* K-12. **Cell**. 32:335-349.

Calhoun, D.H., R.A. Rimerman, and G.W. Hatfield. 1973. Threonine deaminase from *Escherichia coli*. I. Purification and

properties. *J. Biol. Chem. 248*:3511-2516.

Chakrabarti, T., Chen, Y., and Lin, E.C.C. (1984). Clustering of genes for L-fucose dissimilation by *Escherichia coli*. *J. Bacteriol*. **157**:984-986.

Chan, T.T.K., and E.B. Newman. 1981. Threonine as a carbon source for *Escherichia coli*. *J. Bacteriol*. **145**:1150-1153.

Chapon, C. (1982) Expression of *malT*, the regulator gene of the maltose regulon in *Escherichia coli*, is limited both at transcription and translation. *The EMBO J*. **1**:369-374.

Chen, Y., Lu, Z., and Lin, E.C.C. (1989) Constitutive activation of the fucAO operon and silencing of the divergently transcribed fucPIΛ operon by an IS5 element in *Escherichia coli* mutants selected for growth on L-1,2-propanediol. *J Bacteriol* **171**:6097-6105

Clark, R.L. and F.C. Neidhardt. 1990. Roles of the two lysyl-tRNA synthetases of *Escherichia coli*: Analysis of nucleotide sequeι e and mutant behaviour. *J. Bacteriol. 172*:3237-3243.

Cosloy, S.D. and E. McFall. 1970. L-Serine-sensitive mutants of *Escherichia coli* K-12. *J. Bacteriol*. **103**:840-841.

165

Datta, P., T.J. Goss, J.R. Omnaas, and R.V.Patil. 1987. Covalent structure of biodegradative threonine dehydratase of *Escherichia coli*: homology with other dehydratases. *Proc. Natl. Acad. Sci.* USA **84**:393-397.

Davies,W.D., and B.E.Davidson. 1982. The nucleotide sequence of *aroG*, the gene for 3-deoxy-D-arabinoheptulosonate-7-phosphate synthetase (phe) in *Escherichia coli* K-12. *Nucleic Acids Res.* **10**:4045-4058.

de Crombrugghe,B., S.Busby, and H.Buc. 1984. Cyclic AMP receptor protein: role in transcription activation. *Science* **224**:831-838.

De Smit, M.H., and Duin, J.Van. (1990) Secondary structure of the ribosome binding site determines translational efficiency: A quantitative analysis. *Proc. Natl. Acad. Sci.* USA. **87**:7668-7672.

Dunn, J.J., Buzash-Pollert, E., and Studier, F.W. (1978) Mutations of bacteriophage T7 that affect initiation of synthesis of the gene 0.3 protein. *Proc. Natl. Acad. Sci.* USA. **75**:2741-2745.

Eisenberg,D., E.Schwarz, M.Komaromy, and R.Wall. 1984. Analysis of membrane and surface protein sequences with the

hydrophobic moment plot. *J.Mol.Biol.* **179**:125-142.

Emmer,M., B.de Crombrugghe, I.Pastan, and R.Perlman. 1970. Cyclic AMP receptor protein of *E.coli*: its role in the synthesis of inducible enzymes. *Proc.Natl.Acad.Sci.* USA **72**:2300-2304.

Federiuk, C.S., Bayer, R., and Shafer, J.A. (1983). Characterization of the catalytic pathway for D-serine dehydratase: evidence for variation of the rate-determining step with substrate structure. *J. Biol. Chem.* **258**: 5379-5385.

Feng,X.P. 1990. Study of MEW84-A mutant related to the post-translational modification of L-serine deaminase in *Escherichia coli* K-12. M.Sc Thesis at Concordia Univeristy.

Fraser, J., and E.B. Newman. 1975. Derivation of glycine from threonine in *Escherichia coli* K-12 mutants. J. Bacteriol. **122**:810-817.

Gaston, K., A. Bell, A. Kolb, H. Buc, and S. Busby. 1990. Stringent spacing requirements for transcription activation by CRP. *Cell* **62**: 733-743.

Gillam, S., Astell, C.R., and Smith, M. (1980) Site-specific mutagenesis using oligodeoribonucleotides: isolation of a

phenotypically silent ϕX174 mutant, with a specific nucleotide deletion at very high efficiency. *Gene.* **12**:129-137.

Goncharoff, P., and B.P. Nichols. 1988. Evolution of Aminobenzoate Synthases: Nucleotide sequence of *Salmonella typhimurium* and *Klebsiella aerogenes pabB*. *Mol. Biol. Evol.* **5**:531-548.

Goncharoff, P., and B.P. Nichols. 1984. Nucleotide sequence of *Escherichia coli pabB* indicates a common evolutionary origin of p-aminobenzoate synthetase and anthranilate synthetase. *J. Bacteriol.* **159**:57-62.

Goss, T.J. and P. Datta. 1984. *Escherichia coli* K-12 mutation that inactivates biodegradative threonine dehydrogenase by transposon Tn5 insertion. *J. Bacteriol.* **158**:826-831.

Goss, T.J., H.P. Schweizer, and P. Datta. 1988. Molecular characterization of the *tdc* operon of *Escherichia coli* K-12. *J. Bacteriol.* **170**:5352-5359.

Gouy, M., and C. Gauthier. 1982. Codon usage in bacteria: correlation with gene expressivity. *Nucleic. Acids Res.* **10**:7055-7074.

Hama,H., T.Shimamoto, M.Tsuda, and T.Tsuchiya. 1987.

Properties of a Na$^+$-coupled serine-threonine transport sustem in *Escherichia coli. Biochim.Biophys.Acta.* **905**:231-239.

Hama,H., T.shimamoto, M.Tsuda, and T.Tsuchiya. 1988. Characterization of a novel L-serine transport system in *Escherichia coli. J.Bacteriol.* **170**:2236-2239.

Hama, H., Y. Sumita, Y. Katutagi, M. Tsuda, and T. Tsuchiya. 1990. Target of serine inhibition in *Escherichia coli. Biochem. Biophys. Res. Comm.* **168**:1211-1216.

Higgins,C.F., G.F.Ames, W.M.Barnes, J.-M.Clement, and M.Hofnung. 1982. A novel intercistronic regulatory element in procaroytic operons. *Nature*(London) **298**:760-762.

Hobert,E.H., and P.Datta. 1983. Synthesis of biodegradative threonine dehydratase of *Escherichia coli*: role of amino acids, eletron acceptors, and certain intermediary metabolites. *J.Bacteriol.* **155**:586-592.

Horowitz,H., G.E.Christie, and T.Platt. 1982. Nucleotide sequence of *trpD* gene encoding anthranilate synthetase component II of *E.coli. J.Mol.Biol.* **156**: 245-256.

Houghton,J.E., D.E.Bencine, G.A.O'Donovan, and J.R.Wild. 1984. Protein differentiation: a comparison of aspartate

transcarbamoylase and ornithine transcarbamoylase from *Escherichia coli* K-12. *Proc.Natl.Acad.Sci.* USA **81**:4864-4868.

Isenberg, S., and E.B. Newman. 1974. Studies on L-serine deaminase in *Escherichia coli* K-12. *J. Bacteriol.* **118**:53-58.

Kaplan, J.B., and B.P.Nichols. 1983. Nucleotide sequence of *Escherichia coli pabA* gene and its evolutionary relationship to *trpG(D)*. *J.Mol.Biol.* **168**: 451-468.

Kayahara,T., P.Thelen, W.Ogawa, K.Inaba, M.Tsuda, E.B.Goldberg, and T.Tsuchiya. 1992. Properties of recombinant cells capable of growing on serine without NhaB Na$^+$/H$^+$ antiporter in *Escherichia coli*. *J.Bacteriol.* **174**:7482-7485.

Kim, K., S.G. Rhee, and E.R. Stadtman, 1985. Nonenzymatic cleavage of proteins by reactive oxygen species generated by dithiothreitol and iron. *J. Biol. Chem.* **260**:15394-15387.

Kohara, Y., K. Akiyama, and K. Isono. 1987. The physical map of the whole *Escherichia coli* chromosome: application of a new strategy for rapid analysis and sorting of a large genomic library. *Cell* **50**:495-508.

Kolaskar,A.S., and Reddy,B.V.B. 1985. A method to locate protein coding sequences in DNA of prokaryotic systems.

*Nucleic Acid Res.* **13**:185-194.

Krakow,J.S. 1975. Cyclic adenosine monophosphate receptor: effect of cyclic AMP analogues on DNA binding and proteolytic inactivation. *Biochim.Biophys.Acta.* **383**:345-350.

Kredich, N.M., and G.M. Tomkins. 1966. The enzymatic synthesis of L-cysteine in *Escherichia coli* and *Salmonella typhimurium.* *J. Biol. Chem.* **241**:4955-4965.

Kunkel. 1985. Rapid and efficient site-specific mutagenesis without phenotypic selection. *Proc. Natl. Acad. Sci.* USA **82**:488-492.

Lang, B.F., and G. Burger. 1990. A rapid, high resolution DNA sequencing gel system. *Anal. Biochem.* **188**:176-180.

Lawther, R.P., R.C. Wek, J.M. Lopes,R. Pereira, B.E. Taillon, and G.W. Hatfield. 1987. The complete nucleotide sequence of the *ilvGMEDA* operon of *Escherichia coli* K-12. *Nucleic Acids Res.* **15**:2137-2155.

Lin, R.T. Characterization of the leucine/Lrp regulon in *Escherichia coli* K-12. PhD thesis at Concordia University.

Lin, Rongtuan, R. D`Ari, and E.B. Newman. 1990. The leucine

171

regulon of *Escherichia coli*: a mutation in *rblA* alters expression of L-leucine-dependent metabolic operons. *J. Bacteriol.* **172**:4529-4535.

Lin, R.T., D'Ari, R.D., and Newman, E.B. (1992). *λplacMu* insertions in genes of the leucine regulon: Extension of the regulon to genes not regulated by leucine. *J. Bacteriol.* **174**:1948-1955.

Lombardi,F.J., and Kaback,H.R. 1972. Mechanisms of active transport in isolated bacterial membrane vesicles. VIII. The transport of amino acids by membranes prepared from *Escherichia coli*. *J.Biol.Chem.* **247**:7844-7857.

Lowry, O.H., J. Rosebrough, A.L. Farr, and R. J. Randall. 1951. Protein measurement with Folin phenol reagent. *J. Biol. Chem.* **193**·265-275.

Lu, Z., and E.C.C. Lin. 1989. The nucleotide sequence of *Escherichia coli* genes for L-fucose dissimilation. *Nucleic. Acids Res.* **17**:4883-4884.

Magasanik,B. 1970. Glucose effects: inducer exclusion and repression. In: The lactose Operon. J.R.Beckwith and D.Zipser, eds. Cold Spring harbour laboratory, Cold Spring Habour, New York. pp. 189-219.

Makman,R.S., and E.Q.Sutherland. 1965. Adenosine 3',5'-phosphate in *Escherichia coli*. *J.Biol.Chem.* **240**:1309-1314.

Moniakis,J. 1993. A characterization of L-SD1, the gene product of *sdaA* in *Escherichia coli* K-12. M.Sc thesis at Concordia University.

Maniatis, T., E.F.Fritsch, and J. Sambrook. 1982. In Molecular Cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

McCarthy, J.E.G. and Gualerzi, C. 1990. Translational control of prokaryotic gene expression. *TIG* **6**: 78-85.

McFall, E. 1973. Role of adenosine 3'5'-cyclic monophosphate and its specific binding protein in the regulation of D-serine deaminase synthesis. *J. Bacteriol.* **113**:781-785.

McFall, E. 1987. The D-serine deaminase operon. pp.1520-1526. In *Escherichia coli* and *Salmonella typhimurium* Cellular and Molecular Biology. Ed. F. C. Neidhardt. American Society for Microbiology, Washington.

McKitrick,J.C., and L.I.Pizer. 1980. Regulation of phosphoglycerate dehydrogenase levels and effect on serine synthesis in *Escherichia coli* K-12. *J. Bacteriol.* **141**:235-

245.

Miller, J.H. (ed). 1972. Experiments in molecular genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

Miller,B.A., and E.B.Newman. 1974. Control of serine transhydroxymethylase synthesis in *Escherichia coli* K-12. *Can.J.Microbiol.* **20**:41-47.

Morris,J.F., and E.B. Newman. 1980. Map locatio., of the *ssd* mutation in *Escherichia coli* K-12. *J. Bacteriol.* **143**:1504-1505.

Mudd, S.H., and G.L. Cantoni. 1964. Biological transmethylation, methyl-group neogenesis and other "one-carbon" metabolic reactions dependent upon tetrahydrofolic acid, p. 1-47. In M. Florkin and E.H. Stotz(ed.), Comprehensive biochemistry, Vol. 15. Elsevier, Amsterdam.

Nakao,T., Yamto,I., and Anraku,Y. 1987. Nucleotide sequence of putP, the proline carrier gene of *Escherichia coli* K-12. *Mol.Gen.Genet.* **208**:70-75.

Newman, E.B., D. Ahmad, and C. Walker. 1982a. L-Serine deaminase activity is induced by exposure of *Escherichia coli* K-12 to DNA-damaging agents. *J. Bacteriol.* **152**:702-705.

Newman, E.B., D. Dumont, and C. Walker. 1985a. *In vitro* and *in vivo* activation of L-serine deaminase in *Escherichia coli* K-12. *J. Bacteriol.* **162**:1270-1275.

Newman, E.B., and V. Kapoor. 1980. *In vitro* studies on L-serine deaminase activity of *Escherichia coli* K-12. *Can. J. Biochem.* **58**:1292-1297.

Newman,E.B., V.Kapoor, and R.Potter. 1976. Role of L-threonine dehydrogenase in the metabolism of threonine and synthesis of glycine by *Escherichia coli*. *J.Bacteriol.* **126**:1245-1249.

Newman, E.B., and B. Magasanik. 1963. The relation of serine-glycine metabolism to the formation of single carbon units. *Biochim. Biophys. Acta* **78**:437-448.

Newman,E.B., B.Miller, and V.Kapoor. 1974. Biosynthesis of single-carbon units in *Escherichia coli* K-12. *Biochim.Biophys.Acta* **338**:529-539.

Newman, E.B., Malik, N., and C. Walker. 1982b. L-Serine degradation in *Escherichia coli* K-12: Directly isolated *ssd* mutants and their intragenic revertants. *J. Bacteriol.* **150**:710-715.

Newman, E.B., B. Miller, L.D. Colebrook, and C. Walker. 1985b.

A mutation in *Escherichia coli* K-12 results in a requirement for thiamine and a decrease in L-serine deaminase activity. *J. Bacteriol.* **161**:272-276.

Newman, E.B., J.F. Morris, C. Walker, and V. Kapoor. 1981. A mutation affecting L-serine and energy metabolism in *Escherichia coli* K-12. *Mol. Gen. Genet.* **182**:143-147.

Newman, E.B., and C. Walker. 1982. L-Serine degradation in *Escherichia coli* K-12: a combination of L-serine, glycine and leucine used as a source of carbon. *J. Bacteriol.* **151**:777-782.

Newman, E.B., C. Walker, and K. Ziegler-Skylakakis. 1990. A possible mechanism for the *in vitro* activation of L-serine deaminase activity in *Escherichia coli* K-12. *Biochem. Cell Biol.* **68**:123-128.

Nichols,B.P., M.van Cleemput, and C.Yanofsky. 1981. Nucleotide sequence of *Escherichia coli* trpE anthranilate synthetase component I contains no tryptophan residues. *J.Mol.Biol.* **146**: 45-54.

Pardee, A.B., and L.S. Prestidge. 1955. Induced formation of serine and threonine deaminase by *Escherichia coli*. *J. Bacteriol.* **70**:677-674.

Pati,S., DiSilvestre,D., and Brusilow,W.S.A. 1992. Regulation of the *Escherichia coli uncH* gene by mRNA secondary structure and translational coupling. *Mol.Microbiol.* **6**:3559:3566.

Pearson, W.R., and Lipman, D.J. (1988). Improved tools for biological sequence comparison. *Proc. Natl. Acad. Sci.* USA **85**:2444-2448.

Pizer, L.I. 1963. The pathway and control of serine biosynthesis in *Escherichia coli.* *J. Biol. Chem.* **238**:3934-3944.

Pizer,L.I. 1965. Glycine synthesis and metabolism in *Escherichia coli.* *J.Bacteriol.* **89**:1145-1150.

Pizer, L.I., and M.L. Potochny. 1964. Nutritional and regulatory aspects of serine metabolism in *Escherichia coli.* *J. Bacteriol.* **88**:611-619.

Rainwater, S., and P.M. Silverman. 1990. The Cpx proteins of *Escherichia coli* K-12: evidence that *cpxA, ecfB, ssd,* and *eup* mutations all identify the same gene. *J. Bacteriol.* **172**:2456-2461.

Rao,J.K.M., and Argos,P. 1986. A conformational preference parameter to predict helices in intergral membrane proteins.

177

*Biochim.Biophys.Acta. 869*:197-214.


Recsei, P.A., Q.K. Huynh, and E.E. Snell. 1983. Conversion of prohistidine decarboxylase to histidine decarboxylase: peptide chain cleavage by nonhydrolytic serinolysis.
*Proc. Natl. Acad. Sci.* USA **80**:973-983.


Riley,M., and S.Krawic. 1987. Genome organization. pp.967-981. In *Escherichia coli* and *Salmonella typhimurium* Cellular and Molecular Biology. ed. F.C.Neidhardt. American Society for Microbiology, Washington.


Robbins,J.C., and D.L.Oxender. 1973. Transport system for alanine, serine and glycine in *Escherichia coli* K-12. *J.Bacteriol.* **116**:12-18.


Roberts,R.B., P.H.Abelson, D.B.Cowie, E.T.Bolton, and R.J.Britten. 1955. Studies of biosynthesis in *Escherichia coli*. Carnegie Inst.Wash.Publ. 607.


Russel, M., and A. Holmgrem. 1988. Construction and characterization of glutaredoxin-negative mutants of *Escherichia coli. Proc. Natl. Acad. Sci.* USA. **85**:990-994.


Sanger, F., S. Nicklen, and A.R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci.* USA.

74:5463-5467.


Schiltz.E., and D.Schnackerz. 1976. Sequence studies on D-serine dehydratase of *Escherichia coli*. Primary structure of the tryptic phosphopyridoxyl peptide and of the N-terminus. *Eur.J.Biochem*. 71:109-116.


Schweizer, H.P., and P. Datta. 1989. The complete nucleotide sequence of the *tdc* region of *Escherichia coli*. *Nucleic. Acids Res*. 17:3994.


Schweizer, H.P., and P. Datta. 1990. Physical map location of the *tdc* operon of *Escherichia coli*. *J. Bacteriol*. 172:2825.


Shine, J., and Dalgarno, L. (1974) The 3'-terminal sequence of *Escherichia coli* 16S ribosomal RNA: complementarity to nonsense triplets and ribosome binding sites. *Proc. Natl. Acad. Sci*. USA. 71:1342-1346.


Silhavy, T.J., M.L. Berman, and L.W. Enquist. 1984. Experiments with gene fusions. Cold Spring Harbor laboratory. Cold Spring Harbor. N.Y.


Singer, M., T.A. Baker, J.Schnitzler, S.M. Deichel, M. Goel, W. Dove, K.J. Jaack, A.D. Grossman, J.W. Frickson, and C.A. Gross. 1989. A collection of strains containing genetically

linked alternating antibiotic resistance for genetic mapping of *Escherichia coli*. *Microbiol. Rev.* **53**:1-24.

Stauffer, G.V. 1987. Biosynthesis of serine and glycine. In *Escherichia coli* and *Salmonella typhimurium* Cellular and Molecular Biology. Ed. Frederick C. Neidhardt. American Society for Microbiology, Washington.

Stauffer,G.V., M.D.Plamann, and L.T.Stauffer. 1981. Construction and expression of hybrid plasmids containing the *Escherichia coli glyA* gene. *Gene* **14**:63-72.

Stauffer,G.V., and J.E.Brenchley. 1974. Evidence for the involvement of serine transhydroxymethylase in serine and glycine interconversions in *Salmonella typhimurium*. *Genetics* **77**:185-198.

Stewart,V. 1988. Nitrate respiration in relation to facultative metabolism in Enterobacteriaceae. *Microbiol.Rev.* **52**:190-232.

Stratagene Inc. Bluescript DNA sequencing system. Instruction Manual.

Su,H. 1991. Molecular studies of the *sdaA* and *sdaB* genes and their products in *Escherichia coli* K-12. PhD thesis at

Concordia University.

Su, H., Lang, B. F., and Newman, E.B. (1989) L-Serine degradation in *Escherichia coli* K-12. Cloning and sequencing of the *sdaA* gene. *J Bacteriol* **171**:5095-5102.

Su,H., J.Moniakis, and E.B.Newman, 1993. Use of gene fusion of the structural gene *sdaA* to purify L-serine deaminase 1 from *Escherichia coli* K-12. *Eur.J.Biochem.* **211**:521-527.

Su, H., and Newman, E.B. (1991) A novel L-serine deaminase activity in *Escherichia coli* K-12. *J Bacteriol* **173**:2473-2480.

Sumantran, V.N., H.P. Schweizer, and P. Datta. 1990. A novel membrane-associated threonine permease encoded by the *tdcC* gene of *Escherichia coli*. *J. Bacteriol.* **172**:4288-4294.

Templeton,B.A., and Savageau,M.A. 1974. Transport of biosynthetic intermediates: homoserine and threonine uptake in *Escherichia coli*. *J.Bacteriol.* **117**:1002:1009.

Tran, V.S., E. Schaeffer, O. Bertrand, R. Mariuzza, and P. Ferrara. 1983. Purification, molecular weight and N-terminal sequence of cystathionine-r-synthase of *Escherichia coli*. *J. Biol. Chem.* **258**:14872-14873.

Umbarger, H.E., and Brown, B. (1957). Threonine deamination in *Escherichia coli* II. evidence for two L-threonine deamnases. *J. Bacteriol.* **73**:105-112.

Umbarger, H.E., M.A. Umbarger, and P.L.M. Siu. 1963. Biosynthesis of serine in *Escherichia coli* and *Salmonella typhimurium*. *J. Bacteriol.* **85**:1431-1439.

Umbarger,H.E. 1978. Amino acid biosynthesis and its regulation. *Annu.Rev.Biochem.* **47**:533-606.

Umbarger, H.E. 1987. Biosynthesis of the branched-chain amino acids. In *Escherichia coli* and *Salmonella typhimurium* Cellular and Molecular Biology. Ed. Frederick C. Neidhardt. American Society for Microbiology, Washington.

Uzan, M., and A. Danchin. 1978. Correlation between the serine sensitivity and the derepressibility of the *ilv* genes in *Escherichia coli relA⁻* mutants. *Mol. Gen. Gent.* **165**:21-30.

van Vliet,F., A.Jacobs, J.Piette, D.Gigot, M.Lauwreys, A.Pierard, and N.Glansdorff. 1984. Evolutionary divergence of genes for ornithine and aspartate carbamoyl-transferases - complete sequence and mode of regulation of the *Escherichia coli argF* gene; comparison of *argF* with *argI* and *pyrB*. *Nucleic Acids Res.* **12**:6277-6289.

Watson,J.D., Hopkins,N.H., Robert,J.W., Steitz,J.A., and Weiner,A.M. 1987. In Molecular Biology of the Gene. The Benjamin/Cummings Publishing Company.

Weber, R.F., and P.M. Silverman. 1988. The Cpx proteins of *Escherichia coli* K-12: structure of the CpxA polypeptide as an inner membrane component. *J. Mol. Biol.* **203**:467-478.

Winans, S.C., S.J.Elledge, J.H.Krueger, and G.C.Walker. 1985. Site-directed insertion and deletion mutagenesis with cloned fragments in *Escherichia coli*. *J.Bacteriol.* **161**:1219-1221.

Wood, J.M. 1981. Genetics of L-proline utilization in *Escherichia coli*. *J. Bacteriol.* **146**:895-901.

Wu,Y., and P.Datta. 1992. Integration host factor is required for positive regulation of the *tdc* operon of *Escherichia coli*. *J.Bcateriol.* **174**:233-240.

Yanofsky, C. 1960. The tryptophan synthetase system. *Bacteriol. Rev.* **24**:221-245.