Molecular insights into the metabolism and physiology of the lactic acid bacterium *Lactobacillus delbrueckii* subsp. *lactis*

Inauguraldissertation

zur

Erlangung der Würde eines Doktors der Philosophie

vorgelegt der

Philosophisch-Naturwissenschaftlichen Fakultät

der Universität Basel

von

Aude A. Bourniquel

Aus Lorient (56), France

Basel 2000

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät

auf Antrag von Professor T. A. Bickle, Dr. B. Mollet, Prof. W.M. de Vos und Prof. C. Dehio.

Basel, den 13.11.2000

Prof. Dr. Andreas Zuberbühler

Dekan

Summary

In the course of this doctorate work, we gained insights into the energy metabolism of *Lactobacillus delbrueckii* subsp. *lactis* by purifying and characterising 3-phosphoglycerate kinase (PGK) and pyruvate kinase (PYK). These two enzymes are the two enzymes of the Embden-Meyerhof pathway able to regenerate ATP as part of glycolysis. Since *L. lactis* relies on substrate phosphorylation for energy production, the role of PGK and PYK is particularly important in this bacterium. 3-phosphoglycerate kinase is a 45 kDa–monomer and a Michaelis-Menten type enzyme with K_m of 2.6 mM for 3-phosphoglycerate and 0.7 mM for ATP. As the enzymatic assays used for the characterisation of this enzyme go in the reverse direction of glycolysis, no further study was performed. Pyruvate kinase, on the other hand, was characterised in more depth. PYK is an homotetramer with a subunit MW= 68 kDa, and an allosteric enzyme. It is activated by the glycolysis precursors, fructose 1,6-diphosphate, fructose 6-phosphate and glucose 6-phosphate, and inhibited by high concentrations of PEP, inorganic phosphate and ATP. An ATP binding motif present on a 112 residue–C-terminal extension of the enzyme was purported to be implicated in the regulation of PYK by ATP, thus reinforcing the role of pyruvate kinase as a key modulator of *L. lactis* glyolysis and energy metabolism.

In a second stage of this project, we concentrated on the study of restriction-modification (R-M) systems in *L. lactis*. The first indications of the presence of R-M systems came from the analysis of two plasmids, pN42 and pJBL2, isolated from *L. lactis* strains NCC88 and JBL2, respectively, and were two-fold: (i) evidence of modification catalysed by an N^6-methyladenine methyltransferase that may be part of a R-M system, (ii) presence of an *hsdS* gene on each of the two *L. lactis* plasmids, suggesting the presence of a complete type I *hsd* cluster on the chromosome. Purification attempts yielded a semi-purified enzyme fraction displaying activity characteristic of a type I restriction enzyme. A combination of PCR and inverted PCR techniques was implemented to isolate type I *hsd* clusters from two *L. lactis* strains NCC88 and NCC82. These *hsd* clusters encode genes that are highly conserved (>97% identity) except for the *hsdS* genes that have different specificities. The *hsd* clusters could be divided into two parts. The genes *hsdR*, *hsdM*, and *hsdS* organised in two transcriptional units, the single gene *hsdR* regulon and the *hsdMS* operon constitute the first part of *L. lactis hsd* clusters and what we referred to as the '*hsd* cluster proper' since they are the units constituting *hsd* clusters isolated and characterised in *E. coli*. *L. lactis hsd* clusters comprise a second part located downstream of the *hsd* cluster proper encoding and integrase (*int*) gene as well as a second N-terminally truncated *hsdS* gene. Comparison of the *hsdS* genes belonging to the

clusters or the *L. lactis* plasmids with *hsdS* genes isolated from four other *L. delbrueckii* subsp. strains revealed evidence of (i) horizontal transfer of plasmid-borne *hsdS* genes in-between different *L. lactis* strains, (ii) recombination of the target recognition domain-encoding sequences between *hsdS* genes. This provides an insight into a highly flexible mechanism allowing *L. lactis* to generate R-M enzymes with novel specificities, which would constitute a powerful defence tool against phage infections and may explain the unusually high phage resistance displayed by this lactic acid bacterium.

Abbreviations

ADP	adenosine diphosphate
AMP	adenosine monophosphate
approx.	approximately
ATP	adenosine triphosphate
bp	base pair
C-terminus	carboxy-terminus
Da	dalton
DNA	deoxiribonucleic acid
EDTA	ethylenediamine tetraacetic acid
FDP	fructose 1,6-diphosphate
Fru6P	fructose 6-phosphate
Glu6P	glucose 6-phosphate
hsd	host specificity for DNA
IPTG	isopropyl-β-D-thiogalactopyranoside
kb	kilo base pair
kDa	kilo dalton
LAB	lactic acid bacteria
L. lactis	*Lactobacillus delbrueckii* subsp. *lactis*
Lc. lactis	*Lactococcus lactis*
min.	minute
MRS	de Man, Rogosa and Sharpe
MRS-S	MRS medium supplemented with 2% glucose and 17% sucrose
MW	molecular weight
NAD^+	nicotinamide adenine dinucleotide (oxidised form)
NADH	nicotinamide adenine dinucleotide (reduced form)
N-terminus	amino-terminus
o/n	overnight
ORF	open reading frame
PB	protoplasting buffer
PCR	polymerase chain reaction
PEP	phosphoenolpyruvate
PGK	3-phosphoglycerate kinase
P_i	inorganic phosphate

PTS	phosphoenolpyruvate-dependent sugar phosphotransferase system
PYK	pyruvate kinase
RBS	ribosome-binding site
R-M	restriction-modification
RnaseA	ribonuclease A
RT	room temperature
SAM	S-adenosyl-methionine
sdH_2O	sterile distilled water
sec.	second
SDS	sodium dodecyl sulfate
TCA	tricarboxylic acid
TE	Tris-EDTA
TetR	tetracycline resistance repressor
Tris	tris(hydroxymethyl)aminomethane
TRD	target recognition domain
U	unit
vol.	volume
X-Gal	5-bromo-4-chloro-indoxyl-β-D-galactose

Abbreviations for unities are according to the
Système International d'Unités (SI)

TABLE OF CONTENTS

LIST OF FIGURES

Chapter VI

LIST OF TABLES

CHAPTER I

Introduction

Food conservation has been a crucial concern throughout the history of mankind. In times past, the ability to preserve food against the rigours of the lean season literally represented the difference between life and death for the population. The ability to transform easily spoiled raw materials into products that could be stored to outlast their natural and seasonal availability opened opportunities for travel and commerce thus constituting a basis for the development of increasingly complex societies. Historically, food was conserved via simple processes such as drying, smoking or salting. These methods though efficient did not alter the basic texture of the raw material. One of the most ancient food transformation processes, dating back to antiquity, is fermentation. Through the ages, many raw materials were found to respond well to this transformation process (Table 1). Fermentations were empirical and fortuitous processes based on the activities of the natural microbial flora contaminating the raw material. Nowadays, most fermentations are done on an industrial scale requiring inoculation of the raw material with well-defined starter cultures to maintain a high and constant quality in the final products. Dairy fermentations catalysed by lactic acid bacteria (LAB) represents a large percentage of the modern food industry. The mass production of well-defined starter cultures requires definite knowledge of the physiology and metabolism, as well as the culture properties of these bacteria.

Table 1. Overview of fermentations in traditional food processing.

Raw material	Main microorganisms involved	Final product
Alcoholic beverage		
Barley/ Hobs	Baker's yeast: *Saccharomyces cerevisiae*	Beer
Grapes	Yeasts, *Oenococcus oeni**	Wine
Apple	Yeasts, *Oenococcus oeni**	Cider
Rice	*Lactobacillus sakei**	Sake
Fermented food		
Wheat or cereals	Baker's yeast: *Saccharomyces cerevisiae*	Bread
Meat	*Carnobacterium* sp.*	Sausages
Dairy products		
Milk	*Lactobacillus delbrueckii* ssp. *lactis**	Swiss-type hard cheeses e.g. Comté, Emmental
Milk	*Lactococcus lactis**	Cheese e.g. Cheddar
Milk	*Lactobacillus delbrueckii* ssp. *bulgaricus** *Streptococcus thermophilus**	Yoghurt

* the asterisk indicate that this microorganism is a lactic acid bacterium

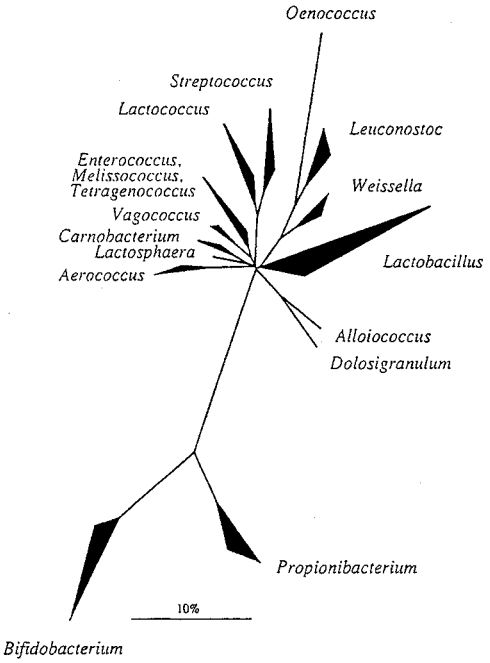


Fig. 1 Phylogenetic relationship of lactic acid bacteria. The phylogenetic distances between the bacteria were based on 16S rRNA nuceotide sequences. The consensus tree is based on distance matrix, maximum parsimony and maximum likelihood. The bar indicates 10% estimated sequence divergence.

LAB are gram-positive, microaerophilic, non-sporulating, and generally catalase-negative microorganisms. Schleifer and Ludwig (1996) described a classification of lactic acid bacteria based not on their morphology or fermentation type but on their phylogenetic relatedness (Fig. 1). Two groups were defined. The low DNA G+C content (< 54 mol % G+C) genera comprise *Aerococcus*, *Lactosphera*, *Carnobacterium*, *Vagococcus*, *Tetragenococcus*, *Melissococcus*, *Enterococcus*, *Lactococcus*, *Streptococcus*, *Oenococcus*, *Leuconostoc*, *Weissella*, *Lactobacillus*, *Alloiococcus*, and *Donosigranulum*. The high G+C content (> 55 mol % G+C) genera include *Propionibacterium*, *Bifidobacterium*, *Brevibactrium* and the microbacteria.

The genus *Lactobacillus* comprises about 50 species that are found in diverse habitats including dairy products, grain products, meat and fish products, water, sewage, beer, wine, fruits and fruit juices, pickled vegetables, sauerkraut, silage, sour dough, and mash. They are a part of the normal flora in the mouth, intestinal tract, and vagina of many homothermic animals including man (Kandler and Weiss, 1986). The species *Lactobacillus delbrueckii* comprises three subspecies *delbrueckii*, *bulgaricus*, and *lactis*. Although all three subspecies are classified as lactic acid bacteria, only two of them *L. bulgaricus* and *L. lactis* are widely used in the dairy industry. *L. bulgaricus* is used mainly in combination with the dairy bacterium *Streptococcus thermophilus* for the production of yoghurt, whereas *L. lactis* is part

of the inoculum for the manufacture of Swiss-type hard cheeses, e.g. Gruyère, Emmental, Comté, or the Italian cheese Grana.

The main dairy starter bacteria, *Lactococcus lactis*, *S. thermophilus*, and *L. delbrueckii* ssp. are classified as obligately homofermentative bacteria, i.e. they ferment glucose via the Embden-Meyerhof (glycolytic) pathway (Fig. 2), and cannot utilise pentoses or gluconates as substrates. Under normal fermentation conditions of an excess of sugar and a limited access to oxygen, pyruvate is reduced almost exclusively to lactic acid in a metabolic process referred to as homolactic fermentation. The disaccharide lactose, the sugar present in milk, is the natural substrate of dairy bacteria, in which two active lactose transport systems have been observed. *Lc. lactis* (Lawrence *et al.*, 1979; Thompson, 1979) and *L. casei* (Chassy and Alpert, 1989) contain a lactose:phosphoenolpyruvate phosphotransferase system (Lac-PTS) that catalyses the transport of lactose from the external medium and releases it into the cytoplasm as lactose-phosphate, which is cleaved by phospho-β-D-galactosidase (P-β–gal) to yield glucose and galactose-6-phosphate. Galactose-6-phosphate is metabolised through the tagatose-6-phosphate pathway (Fig. 2). A second type of lactose transport system has been observed in *S. thermophilus* and *L. delbrueckii* spp. and consists in a permease that transports lactose into the cytoplasm where it is cleaved by β-galactosidase (β–gal) to yield glucose and galactose (Premi *et al.*, 1972; Fox *et al.*, 1990). In those so-called "thermophilic" LAB, the galactose is not metabolised but excreted into the medium (Hickey *et al.*, 1986; Hutkins and Morris, 1987). In all cases, the glucose moiety is phosphorylated by glucokinase and metabolised via the glycolytic pathway to produce lactic acid and ATP. Theorically, homolactic fermentation of glucose results in 2 moles of lactic acid and a net gain of 2 moles ATP per mole glucose consumed (Axelson, 1993).

Lactic acid bacteria, lacking the TCA cycle and electron transfer chain, rely on substrate level phosphorylation for the production of ATP. LABs have been shown to possess proton-translocating ATPases but these enzymes are used for the regulation of the cytoplasmic pH rather than for ATP generation as is usual in non-lactic acid bacteria (Kobayashi, 1985; Nannen and Hutkins, 1991; Hutkins and Nannen, 1993). ATP is therefore regenerated almost exclusively by two enzymes of the Embden-Meyerhof pathway, 3-phosphoglycerate kinase (PGK) and pyruvate kinase (PYK) (Fig. 2). *Lc. lactis* is considered as the model organism for LAB and its pyruvate kinase has been extensively studied. *Lc. lactis* PYK is regulated by fructose 1,6-diphosphate (FDP) and inorganic phosphate (Pi) and plays an important role in the regulation of glycolysis (for a review, see Konings *et al.*, 1989). No reports exists,

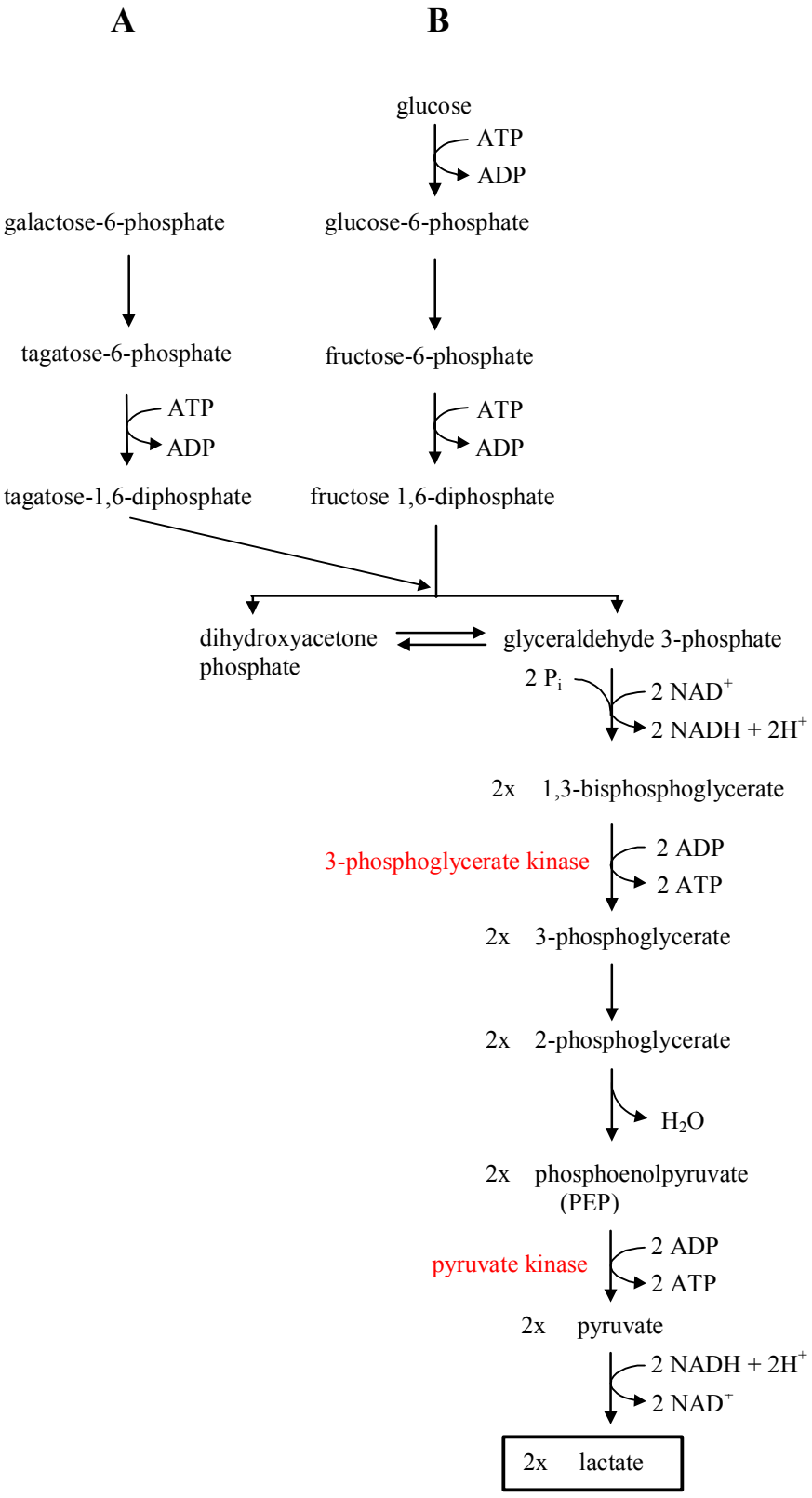


Fig. 2. Major homolactic fermentation pathways of glucose and galactose in homofermentative lactic acid bacteria. (A) tagatose-6-phosphate pathway for the assimilation of galactose-6-phosphate; (B) Embden-Meyerhof pathway of glycolysis.

however, on the 3-phosphoglycerate kinases of LAB. In the first part of this work, we describe the purification and characterisation of the the pyruvate kinase and the 3-phosphoglycerate kinase of *Lactobacillus delbrueckii lactis* and the comparison of their properties with those of enzymes from non-dairy bacteria in order to determine whether their excessively important role in the energy metabolism of *L. lactis* has induced significant mutations in their sequence or structure.

One of the major problems encountered in dairy fermentations has always been bacteriophage (phage) attacks. This problem has been compounded in the last 20 to 30 years by the current trend to use fewer, more finely tuned starter cultures at an industrial scale leading to the evolution of novel virulent phages able to infect these starter strains. Phage attacks have thus become a major concern economically as well as ecologically for the dairy industry. Infection can lead to a partial lysis of the starter cultures in the vat causing slow fermentation or even a complete failure of the starter with subsequent loss of the product (Peitersen, 1991) and requirement to process the infected product before its elimination. As *L. bulgaricus* is routinely incubated with *S. thermophilus* for the production of yoghurt and as *S. thermophilus* is a strain highly susceptible to phage infections (Bruttin *et al.*, 1997), a similarly high number of phages could have been expected to adapt to *L. bulgaricus* by horizontal transfer. In fact very few phages are known to be able to target *L. delbrueckii* ssp. Since transformation of these strains is also a highly inefficient process, it implies that *L. delbrueckii* ssp. possess a very active and reliable endogenous defence mechanism.

One of the most efficient and versatile defence mechanisms developped by bacteria against bacteriophages are the DNA restriction-modification (R-M) systems. These systems have the dual function of: (i) protecting the host DNA against restriction by methylating the DNA within specific target sites, and (ii) 'restricting' i.e. degrading any unmodified piece of DNA that may enter the cell. Restriction-modification systems thus protect the bacterium against invading phage genomes or plasmids. R-M systems may be of three types. Type I enzymes are constituted of three subunits encoded by the genes *hsdR*, *hsdM* and *hsdS* (hsd for host specificity for DNA). Methylation is catalysed by the M_2S form of the enzyme in the presence of S-adenosyl methionine (SAM), whereas the form R_2M_2S (or R_1M_2S, Janscak *et al.*, 1996) and the presence of the three cofactors SAM, Mg^{2+} and ATP is required for DNA restriction. The target site of type I R-M systems is asymetric, split into two recognition components separated by 6-8 unspecified nucleotides. Each half of the recognition site is recognised by one of the two so-called 'variable' or target recognition domains (TRDs) of the HsdS subunit, which is responsible for the specific binding of the enzyme to the DNA. Methylation occurs

within the recognition site whereas restriction is random and may occur as far as 7,000 base pairs from the closest site (for reviews, see Bickle, 1987, and Bickle and Krüger, 1993). Type II R-M systems comprise separate restriction and modification enzymes that act independently from each other and have simple cofactor requirements: restriction depends on the presence of Mg^{2+}, and modification requires SAM. Both methylation and restriction occur at fixed positions within a 4-8 nt palindromic recognition sequence. Type III enzymes are hetero-oligomeric proteins catalysing both the restriction and modification reactions. Cofactors are SAM for modification, Mg^{2+} and ATP for restriction (for a review, see Wilson and Murray, 1991). Type IIS and type III enzymes recognise asymetrical target sites and cut at a fixed distance from this site. For type III enzymes, this distance may be 25 to 30 base pairs.

In the '50s, the first reports of host-controlled modification and restriction of DNA followed experiments done with *E. coli* phages in *E. coli* bacteria (Bertani and Weigle, 1953). Since then, research concerning R-M systems and especially type I and type III enzymes has concentrated on enzymes isolated from *E. coli* strains (Bickle, 1987) or enterobacteriacea (e.g. *Salmonella*, Nagaraja *et al.*, 1985). For commercial and economic reasons, type II enzymes have been isolated from a much larger range of bacterial genera, but advanced biochemical characterisations have still been mostly restricted to *E. coli* enzymes, e.g. *Eco*RI or *Eco*RV (Stahl *et al.*, 1998; Sam and Perona, 1999; Jeltsch *et al.*, 1999). The first reported incidence of restriction-modification systems in LAB dates back to 1956 (Collins) and to the same time period as the first *E. coli* reports. However, no further research on this subject was published in lactic acid bacteria for the next 25 years, till the early '80s, when interest in the demonstration of the presence and operation of R-M systems in dairy lactococci and their role in phage resistance reawakened (Boussemaer *et al.*, 1980; Sanders and Klaenhammer, 1981; Chopin *et al.*, 1984). At that time, biochemical characterisation of R-M systems in LAB remained limited to a single type II restriction endonuclease (ENase), *Scr*FI in *Lc. lactis* (formerly, *Streptococcus lactis*, Fitzgerald *et al.*, 1982). Since then, many more R-M enzymes have been identified in LAB (see Table 2). All three types of restriction enzymes have been discovered and characterised in lactococci, whereas, though many type II ENases are been found in *Streptococcus thermophilus* strains, report of a type I enzyme is limited to the unpublished, putative enzyme translated from the sequence data posted in GenBank by Solow and Somkuti (GenBank accession number: AF177167). Although Auad *et al.* (1998) established the presence of a R-M system in *L. lactis* CNRZ 326 using the classical phage assay, no restriction enzymes have as yet been identified in *Lactobacillus delbrueckii* ssp.

We present in this work our investigation of the presence of DNA restriction and modification systems in *L. delbrueckii lactis*. This investigation was conducted in three principal stages. The first stage was the isolation and analysis of the pN42 plasmid from the *L. lactis* strain NCC88, which provided evidence of the presence of an endogneous site specific N^6-adenine methyltransferase. pN42 also encodes a gene with a structure characteristic of an *hsdS* gene suggesting the presence of an *hsd* cluster on *L. lactis* chromosome. The second stage of this project was the purification of restriction enzymes from *L. lactis* strain NCC88. This required the optimisation of the crude protein extract preparation protocol since neither the sonication, the glass beads nor the French press procedures were well-adapted for the handling of large *L. lactis* culture volumes. A type I restriction-enzyme was semi-purified and characterised. In the last stage of this project, the *hsd* clusters of two *L. lactis* strains (NCC88 and NCC82) as well as *hsdS* genes from four additional *L. delbrueckii* ssp. strains (NCC39, NCC73, NCC627, and NCC786) were sequenced and analysed providing circumstancial evidence of a mechanism combining horizontal transfer of plasmid-borne *hsdS* genes and genetic recombination of the variable regions for the production of novel *hsdS* genes with new specificities. It is worth noting that this thesis reports the first identification of restriction-modification systems in the lactic acid bacterium *L. delbrueckii* ssp.

References

- Auad, L., Azcarate Peril, M.A., de Ruiz Holgado, A.A.P., and Raya, R.R. (1998) Evidence of a restriction/modification system in *Lactobacillus delbrueckii* subsp. *lactis* CNRZ 326. *Curr. Microbiol.* **36:** 271–273.
- Axelson, L. (1993) Classification and physiology, pp. 1–72. *In:* Lactic Acid Bacteria, Microbiology and Functional Aspects, 2nd edition,. Salminen, S., and von Wright, A. (eds), Marcel Dekker Inc.
- Bertani, G., and Weigle, J.J. (1953) Host-controlled variation in bacterial viruses. *J. Bacteriol.* **65:** 113–121.
- Bickle, T.A. (1987) DNA restriction and modification systems, pp. 692–696. *In: Escherichia coli* and *Salmonella thyphimurium*: cellular and molecular biology, Ingraham, J.L., Low, K.B., Magasanik, B., Neidhardt, F.C., Schaechter, M., and Umbarger, H.E. (eds), American Society for Microbiology, Washington DC.
- Bickle, T.A., and Krüger, D.H. (1993) Biology of DNA restriction. *Microbiol. Rev.* **57:** 434–450.

- Boussemaer, J.P., Schrauwen, P.P., Sourrouille, J.L., and Guy, P. (1980) Multiple modification/restriction systems in lactic streptococci and their significance in defining a phage-typing system. *J. Dairy Res.* **47:** 401–409.
- Bruttin, A., Desiere, F., d'Amico, N., Guerin, J.P., Sidoti, J., Huni, B., Lucchini, S., and Brüssow, H. (1997) Molecular ecology of *Streptococcus thermophilus* bacteriophages infection in a cheese factory. *Appl. Environ. Microbiol.* **63:** 3144–3150.
- Chassy, B.M., and Alpert, C.–A. (1989) Molecular characterization of the plasmid-encoded lactose-PTS of *Lactobacillus casei*. *FEMS Microbiol. Rev.* **63:** 157–166.
- Chopin, A., Chopin, M.-C., Moillo-Batt, A., and Langella, P. (1984) Two plasmid-determined restriction and modification systems in *Streptococcus lactis*. *Plasmid* **11:** 206–263.
- Collins, E.B. (1956) Host-controlled variations in bacteriophages active against lactic streptococci. *Virology* **2:** 261–271.
- Fitzgerald, G.F., Daly, C., Brown, L.R., and Gingeras, T.R. (1982) *Scr*FI: a new sequence-specific endonuclease from *Streptococcus cremoris*. *Nucleic Acids Res.* **10:** 8171–8179.
- Fox, P.F., Lucey, J.A., and Cogan, T.M. (1990) Gylcolysis and related reactions during cheese manufacture and ripening. *Crit. Rev. Food Sci. Nutr.* **29:** 237–253.
- Hickey, M.W., Hillier, A.J., and Jago, G.R. (1986) Transport and metabolism of lactose, glucose and galactose in homofermentative lactobacilli. *Appl. Environ. Microbiol.* **51:** 825–831.
- Hutkins, R.W., and Morris, H.A. (1987) Carbohydrate metabolism in *Streptococcus thermophilus*: a review. *J. Food Prot.* **50:** 876–884.
- Hutkins, R.W., and Nannen, N.L. (1993) pH homeostasis in lactic acid bacteria. *J. Dairy Sci.* **76:** 2354–2365.
- Janscak, P., Abadjieva, A., and Firman, K. (1996) The type I restriction endonuclease R.*Eco*R124I: Over-production and biochemical properties. *J. Mol. Biol.* **257:** 977–991.
- Jeltsch, A., Christ, F., Fatemi, M., and Roth, M. (1999) On the substrate specificity of DNA methyltransferases. Adenine-N^6 DNA methyl transferases also modify cytosine residues at position N^4. *J. Biol. Chem.* **274:** 19538-19544.
- Kandler, O., and Weiss, N. (1986) Regular, nonsporing gram-positive rods, pp. 1208–1234. *In:* Bergey's manual of systematic bacteriology, Vol. 2, 9th ed., Sneath, P.H.A., Mair, N.S., Sharpe, M.E., and Holt, J.G. (eds), Williams and Wilkins, Baltimore.
- Kobayashi, H. (1985) A proton-translocating ATPase regulates pH of the bacterial cytoplasm. *J. Biol. Chem.* **260:** 72–76.

- Konings, W.N., Poolman, B., and Driessen, A.J.M. (1989) Bioenergetics and solute transport in lactococci. *Crit. Rev. Microbiol.* **16:** 419–476.
- Nagaraja, V., Shepherd, J.C.W., Pripfl, T., and Bickle, T.A. (1985) Two type I restriction enzymes from *Salmonella* species. Purification and DNA recognition sequences. *J. Mol. Biol.* **182:** 579–587.
- Nannen, N.L., and Hutkins, R.W. (1991) Proton-translocating adenosine triphosphatase activity in lactic acid bacteria. *J. Dairy Sci.* **74:** 747–751.
- Peitersen, N. (1991) Practical phage control. *Bull. Int. Dairy Fed.* **263:** 1–43.
- Premi, L., Sandine, W.E., and Elliker, P.R. (1972) Lactose-hydrolyzing enzymes of *Lactobacillus* species. *Appl. Microbiol.* **24:** 51–57.
- Roberts, R.J., and Macelis, D. (2000) REBASE- restriction enzymes and methylases. *Nucleic Acids Res.* **28:** 306–307. Official REBASE web site: http://rebase.neb.com
- Sam, M.D., and Perona, J.J. (1999) Catalytic roles of divalent metal ions in phosphoryl transfer by *Eco*RV endonuclease. *Biochemistry* **38:** 6576–6586.
- Sanders, M.E., and Klaenhammer, T.R. (1981) Evidence for plasmid linkage of restriction and modification in *Streptococcus cremoris* KH. *Appl. Environ. Microbiol.* **42:** 944–950.
- Schleifer, K.-H., and Ludwig, W. (1996) Phylogeny of the *Lactobacillus* and related genera. *System. Appl. Microbiol.* **18:** 461–467.
- Stahl, F., wende, W., Wenz, C., Jetsch, A., and Pingoud, A. (1998) Intra- vs intersubunit communication in the homodimeric restriction enzyme *Eco*RV: Thr 37 and Lys 38 involved in indirect readout are only important for the catalytic activity of their own subunit. *Biochemistry* **37:** 5682–5688.
- Thompson, J. (1979) Lactose metabolism in *Streptococcus lactis*: Phosphorylation of galactose and glucose moieties *in vivo*. *J. Bacteriol.* **140:** 774–785.
- Wilson, G.G., and Murray, N.E. (1991) Restriction and modification systems. *Annu. Rev. Genet.* **25:** 585–627.

Table 2. DNA restriction-modification systems identified in the main dairy starter bacteria.

Enzyme Name	Type	Recognition Site	Microorganism	GenBank accession number	References
*Lla*82I	I	ND	Lactococcus lactis (pAH82)	AF228680	
*Lla*2614I	I	ND	Lactococcus lactis (pIL2614)	U90222	
*Lld*IP*	I	ND	Lactococcus lactis bv. diacetylactis (pND861)	AF034786	
*Lla*1403I	I	ND	Lactococcus lactis IL1403	AF013165	Schouler *et al.*, 1998
*Sth*ER35IP*	I	ND	Streptococcus thermophilus (pER35)	AF177167	
S.*Lde*I	I.hsdS	ND	Lactobacillus delbrueckii subsp. lactis (pLL1212)	AF109691	
S.*Lla*103I	I.hsdS	ND	Lactobacillus lactis (pIL103)	AF013595	
S.*Lla*7I	I.hsdS	ND	Lactobacillus lactis (pIL7)	AF013596	
S.*Lla*NZ4000IP*	I.hsdS	ND	Lactococcus lactis (pNZ4000)	AF036485	van Kranenburg *et al.*, 2000
S.*Lla*33IP*	I.hsdS	ND	Lactococcus lactis DPC220	AF207855	
S.*Lla*CIS3I	I.hsdS	ND	Lactococcus lactis subsp. cremoris (pCIS3)	AF153414	Seegers *et al.*, 2000
S.*Sth*CI65IP*	I.hsdS	ND	Streptococcus thermophilus NDI-6 (pCI65st)	AF027167	O'Sullivan *et al.*, 1999
*Lla*I	IIS (?)	ND	Lactococcus lactis ssp. lactis (pTR2030)	U17233	O'Sullivan *et al.*, 1995
LlaKR2I	II	5'–GATC–3'	Lactococcus lactis ssp. lactis KR2	AF051563	Twomey *et al.*, 1998
*Lla*497I	II	5'–CCWGG–3'	Lactococcus lactis ssp. lactis NCDO 497	-	Mayo *et al.*, 1991
*Lla*DCHI	II	5'–^GATC–3'	Lactococcus lactis subsp. cremoris DCH-4	U16027	Moineau *et al.*, 1995
*Scr*FI	II	5'–CC^NGG–3'	Lactococcus lactis subsp. cremoris UC503	U89998	Fitzgerald *et al.*, 1982 Davis *et al.*, 1993
*Lla*BI	II	5'–C^TRYAG–3'	Lactococcus lactis subsp. cremoris W56	X97363	Nyengaard *et al.*, 1993

Table 2. DNA restriction-modification systems identified in the main dairy starter bacteria.

Enzyme Name	Type	Recognition Site	Microorganism	GenBank accession number	References
*Lla*BIIP*	II	ND	Lactococcus lactis subsp. cremoris W56	Y12736	Nellemann *et al.*, 1997
*Lla*AI	II	5'–^GATC–3'	Lactococcus lactis subsp. cremoris W9	-	Nyengaard *et al.*, 1993, 1995
*Lla*E1	II	ND	Lactococcus lactis W12	-	Nyengaard *et al.*, 1993
*Lla*CI	II	5'–A^AGCTT–3'	Lactococcus lactis W15	AJ002064	Josephsen *et al.*, 1998 Madsen & Josephsen, 1998 Madsen *et al.*, 1997
*Lla*DI	II	ND	Lactococcus lactis W39	-	Nyengaard *et al.*, 1995
*Lla*DII	II	5'–GCNGC–3'	Lactococcus lactis W39		Nyengaard *et al.*, 1995
*Sth*455I	II	5'–CCWGG–3'	Streptococcus thermophilus CNRZ 455	-	Guimont *et al.*, 1993
*Sth*St0IP*	II	ND	Streptococcus thermophilus St0	AJ242480	
*Sth*117I	II	5'–CC^WGG–3'	Streptococcus thermophilus ST117	-	Solaiman & Somkuti, 1991
*Sth*St8IP*	II	ND	Streptococcus thermophilus St8	AJ239049	
*Sth*134I	II	5'–C^CGG–3'	Streptococcus thermophilus strain 134	-	Solaiman & Somkuti, 1990
*Sth*132I	IIS	5'–CCCGNNNN^NNNN –3' 3'–GGGCNNNN NNNN^–3'	Streptococcus thermophilus	-	Poch *et al.*, 1997
*Lla*FI	III	ND	Lactococcus lactis LL42-1	AF054600	Su *et al.*, 1999

ND: not determined

* the suffix P indicates that those are putative enzymes predicted from sequences appearing in the GenBank database. This nomenclature follows the recommendations of the Restriction Enzyme Database (REBASE, Roberts and Macelis, 2000).

CHAPTER II

Optimisation of crude protein extract preparation for *L. lactis* cultures.

Lactic acid bacteria (LAB) are gram-positive bacteria used in the dairy industry for large-scale fermentation of milk into cheese and yogurt. Because of their economic importance, most dairy starter bacteria, such as the lactococci and streptococci, have been intensely studied on a physiological and genetic level. This is not the case however for *Lactobacillus delbrueckii* ssp. For that two major factors are responsible: (i) it has been impossible up to now to efficiently transform or conjugate these bacteria, (ii) the bacteria have a very resistant cell wall preventing cell lysis. The resistance of the cell wall is one of the reasons why the few proteins that have been purified from *L. delbrueckii* spp. were purified by heterologous overexpression (Stucky *et al.*, 1995; Klein *et al.*, 1997). That method is very efficient but requires the prior isolation and sequencing of the gene from the *L. delbrueckii* spp. host. If the gene proves difficult to isolate, the only solution is to purify the protein directly from the wild-type host. Such purifications imply dealing with large masses of cells, which renders their treatment by sonication, glass beads or even French press cumbersome. In this work, we develop for *L. lactis* bacteria a crude protein extract preparation protocol requiring only minimal hardware, and based on the formation and lysis of protoplasts so as to monitor cell-wall degradation and the totality of cell lysis.

MATERIAL AND METHODS

Bacterial strain and culture conditions. *Lactobacillus delbrueckii* subsp. *lactis* strain NCC88 was obtained from the Nestlé Culture Collection. *L. lactis* was routinely grown in MRS broth (Difco Laboratories, Detroit, MI, USA) supplemented with 2% glucose at 45°C without aeration. Several known cell-wall weakeners were added to the growth medium at different concentations and their impact on *L. lactis* cell-wall thickness and resistance studied: D,L-threonine (40 mM), glycine (1%, 2% and 3%), cysteine (0.05%) and sucrose (0.5 M). MRS broth supplemented with 2% glucose and 17% sucrose (0.5 M) was called MRS-S broth.

Optimised crude protein extract preparation. *L. lactis* was inoculated in 3 liters of MRS-S broth and incubated overnight at 45°C without aeration. The cells were harvested by centrifugation (16,000 × *g*, 10 min, 4°C). The cell pellet was washed twice with ice-cold deionised water (dH_2O) then resuspended in protoplasting buffer (PB: 50mM Tris.Cl, pH7.4; 50 mM EDTA, pH 8.0; 25% sucrose) containing 1 mg/ml lysozyme, and incubated at 37°C. After 2 hours incubation, 170 U/ml of mutanolysin were added

to the suspension. The incubation at 37°C was continued till examination under optical microscope showed that roughly 99% of the cells were protoplasted (30 min to 1 hour). 10 mM $MgCl_2$ was added to the suspension to allow a stabilization of the pellet during centrifugation. The protoplast pellet was washed with PB to eliminate lysozyme and mutanolysin as well as the proteins released during the degradation of the cell wall. Finally, the protoplast pellet was resuspended in lysis buffer (20 mM Tris.Cl, pH7.4; 50 mM KCl; 1 mM EDTA; 7 mM β-mercaptoethanol) and frozen at –20°C to insure total lysis. The suspension was thawed, then clarified by centrifugation (16,000 × *g*, 30 min, 4°C) and the supernatant i.e. the crude protein extract recovered.

RESULTS AND DISCUSSION

Influence of cell-wall weakeners on cell susceptibility to form protoplasts. Several procedures can be applied to weaken the bacterial cell wall. Treatment by muralytic enzymes such as lysozyme and/or mutanolysin is usually recommended. However, when applied to *Lactobacillus delbrueckii* ssp., that treatment is not sufficient, and Delley *et al.* (1990) recommended treating the cells with proteinase A and pronase E prior to the addition of SDS that induces the actual cell lysis. The use of proteases or SDS is however strictly prohibited during the preparation of crude protein extract for purification purposes since the proteins must be kept as best as possible in their native form and active. An alternative is the addition of cell-wall weakeners to the bacterial growth medium. In this work, a range of such compounds were tested.

The presence of 40 mM D,L-threonine in the growth medium did not affect the morphology of *L. lactis* cells (Fig. 1B) nor enhance their ability to form protoplasts. While addition of 3% glycine totally inhibits the growth of *L. lactis*, lower concentrations (1%, 2.5%) induced a clear morphological change from the long, thin chain-forming bacilli that is the classical form of *L. lactis* (Fig. 1A) to a short, compact bacillus found mostly in pair or alone (Fig. 1C). Rather than increasing *L. lactis* cells susceptibility to muralytic

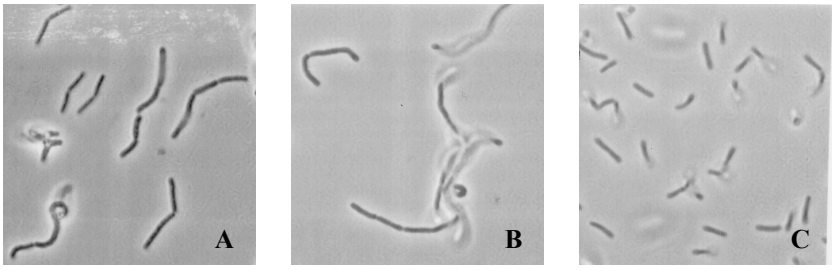


Fig. 1Observation under light microscopy of the morphology of *L. lactis* grown in different media. (A) Lactobacilli grown in MRS broth, (B) lactobacilli grown in MRS + 40 mM D,L-trheonine, and (C) lactobacilli grown in MRS + 2.5% glycine (× 100).

enzymes, however, this morphological change enhances their stability. In contrast to positive results obtained for other LABs such as *Lactococcus lactis* (Holo and Nes, 1989), *Lactobacillus helveticus* (Bhowmik and Steele, 1993), *Lactobacillus acidophilus* A1 (Walker *et al.*, 1996) or meat lactobacilli (Aymerich *et al.*, 1993), neither D,L-threonine nor glycine constitute an efficient cell-wall weakener for *L. lactis*. An attempt was then made to modify the structure of the cell wall by growing the cells in a high ionic stength medium as recommended by Argnani *et al.* (1996). *L. lactis* cells grown in MRS-S broth did not exhibit a clear morphological change (Fig. 2A-D), although the form of the cells appeared more homogeneous throughout the culture. They did however easily form protoplasts when incubated with lysozyme and mutanolysin. The addition of 0.5% cysteine to the MRS-S broth had no influence on morphology or protoplasts formation (data not shown).

Optimization of the protoplasting buffer composition. Joseph and Shockman (1974), working on autoplasts formation in *Streptococcus faecalis*, recommended the use of ammonium acetate buffer (40 mM, pH 7.0) rather than Tris or phosphate buffers to provide osmotic stability during prolonged incubations. The presence of 40 mM ammonium acetate in PB, however, prevented the formation of protoplasts from *L. lactis* cells. A Tris-based buffer was therefore chosen since it is the buffer used in many purification protocols and the incubation time was kept as short as possible (3 hours at most). After comparison of the relative efficiency of different organic compounds as osmotic stabilizers, Corner and Marquis (1969) reported that the larger sugars (raffinose, sucrose) were more efficient than the smaller sugars (glucose, ribose) or selected amino acids (proline, glycine). To maintain hypertonicity in PB, we chosed to use sucrose rather than the slightly more effective but very expensive raffinose, or glycerol which induces a swelling of the protoplasts. A concentration of 25% sucrose was found to maintain protoplast integrity whereas 17% (0.5 M) still allowed a partial lysis. Magnesium chloride ions ($MgCl_2$) are supposed to stabilize the cytoplasmic membrane and prevent the clumping of protoplasts. We found here that the addition of 10 mM $MgCl_2$ to PB enhanced cell clumping and slightly decreased the velocity of protoplast formation. The addition of EDTA, while not affecting velocity, totally inhibits clumping. EDTA also insures that protoplasts lyse immediately upon transfer to an isotonic buffer suggesting that EDTA, complexing Mg^{2+} ions, does destabilize the protoplast membrane but not to such an extent that the protoplasts will lyse when maintained in the hypertonic buffer. The final composition of the protoplasting buffer (PB) is therefore: 50 mM Tris.Cl, pH 7.4; 50 mM EDTA, pH 8.0; 25 % sucrose.

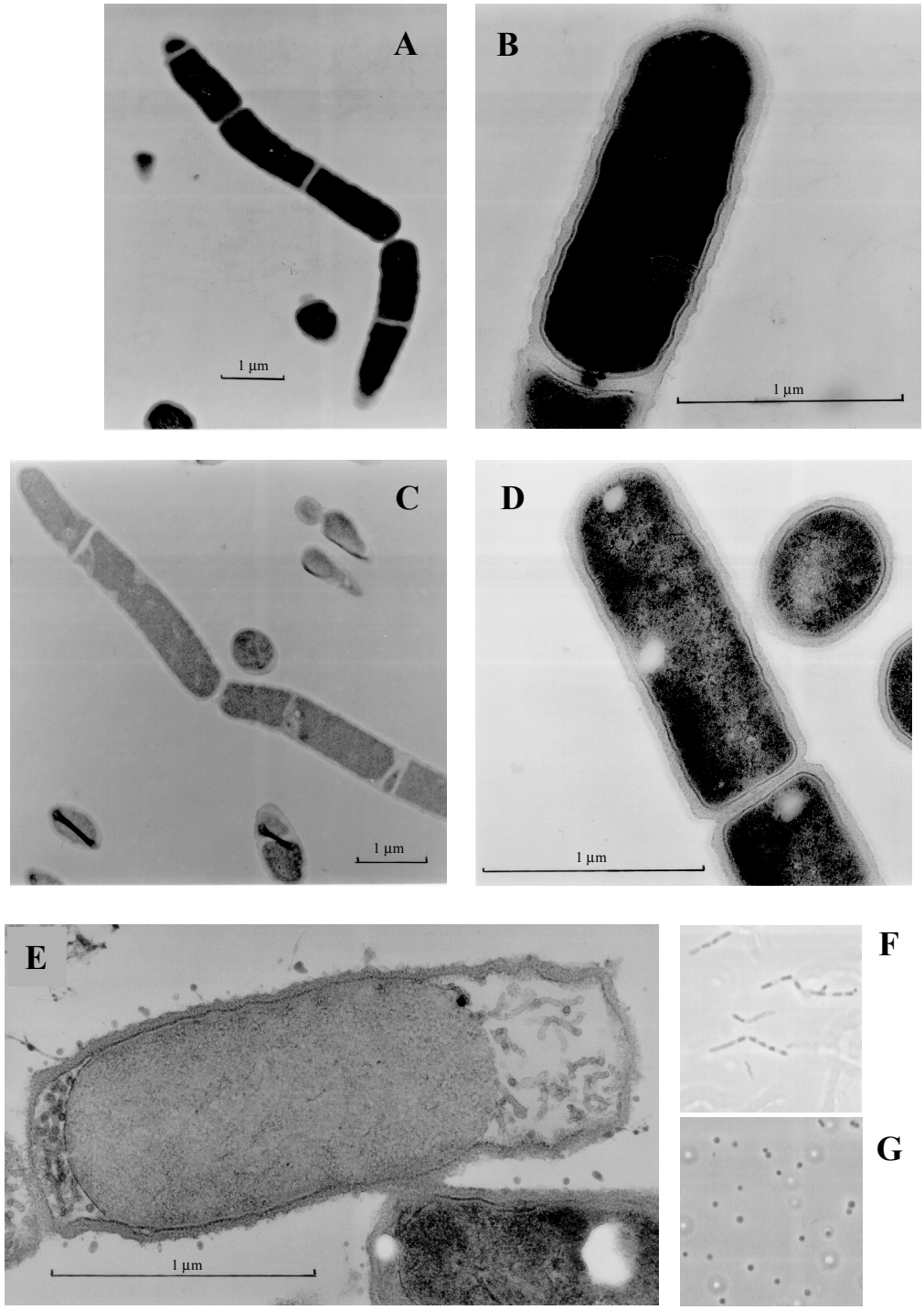


Fig.2 Transmission electron micrographs of *L. lactis* cells. (A) and (B) are views of lactobacilli grown in MRS broth with different resolutions (× 18 400 and × 64 400, respectively). (C) and (D) display lactobacilli grown in MRS-S broth (× 18 400 and × 64 400, respectively). (E) is a view of a lactobacilli cell grown in MRS-S broth and treated for 2 hours with 1 mg lysozyme per ml PB (× 64 400). (F) and (G) are light microscopy views of lactobacilli incubated with lysozyme (1 mg/ml) and lysozyme followed by mutanolysin (170 U/ml), rspectively (× 100).

Optimisation of protoplast formation. Lysozyme alone was not sufficient to degrade *L. lactis* cell wall. As depicted in Fig. 2E and 2F, lysozyme does not degrade *L. lactis* peptidoglycan but only separates it from the cell membrane. The cytoplasm then tended to adopt a typical spherical form limited only by the constraints exerted by the peptidoglycan still surrounding the cytoplasmic membrane. The peptidoglycan, known to constitute the "shape-maintaining structure" of the bacterium (Braun and Hantke, 1974), conserved by itself the elongated form of the intact bacterium. Addition of mutanolysin led to peptidoglycan degradation and protoplast release (Fig. 2G). Mutanolysin alone was sufficient to obtain protoplast formation from cells harvested during the exponential growth phase, but cells collected from the stationary phase required both lysozyme and mutanolysin. The successive action of lysozyme and mutanolysin was shown to be more effective than an incubation where both muralytic enzymes were present at the same time. This suggests that in the early phase of peptidoglycan degradation, the presence of lysozyme prevents the mutanolysin from accessing its substrate. In order to be applicable in all cases, the protocol developed here recommends a 2 hour-incubation at 37°C with 1 mg lysozyme per ml protoplasting buffer (PB) before adding the mutanolysin (170 U/ml).

This protocol was optimised for the preparation of crude protein extract from *L. lactis* cultures and constitute the first step required before starting a protein purification protocol with such lysis-resistant bacteria.

References

- Argnani, A., Leer, R.J., van Luijk, N., and Pouwels, P.H. (1996) A convenient and reproducible method to genetically transform bacteria of the genus *Bifidobacterium*. *Microbiology* **142:** 109–114.
- Aymerich, M.T., Hugas, M., Garriga, M., Vogel, R.F., and Monfort, J.M. (1993) Electrotransformation of meat lactobacilli. Effect of several parameters on their efficiency of transformation. *J. Appl. Bacteriol.* **75:** 320–325.
- Bhowmik, T., and Steele, J.L. (1993) Development of an electroporation procedure for gene disruption in *Lactobacillus helveticus* CNRZ 32. *J. Gen. Microbiol.* **139:** 1433–1439.
- Braun, V., and Hantke, K. (1974) Biochemistry of bacterial cells enveloppes. *Annu. Rev. Biochem.* **43:** 89–121.
- Corner, T.R., and Marquis, R.E. (1969) Why do bacterial protoplasts burst in hypotonic solutions ? *Biochim. Biophys. Acta* **183:** 544–558.

- Delley, M., Mollet, B., and Hottinger, H. (1990) DNA probe for *Lactobacillus delbrueckii*. *Appl. Environ. Microbiol.* **56:** 1967–1970.
- Holo, H., and Nes, I.F. (1989) High-frequency transformation, by elecroporation, of *Lactococcus lactis* subsp. *cremoris* grown with glycine in osmotically stabilized media. *Appl. Environ. Microbiol.* **55:** 3119–3123.
- Joseph, R., and Shockman, G.D. (1974) Autolytic formation of protoplasts (autoplasts) of *Streptococcus faecalis* 9790: release of cell wall, autolysin, and formation of stable autoplasts. *J. Bacteriol.* **118:** 735–746.
- Klein, J.R., Schick, J., Henrich, B. and Plapp, R. (1997) *Lactobacillus delbrueckii* subsp. *lactis* DSM7290 *pepG* gene encodes a novel cysteine aminopeptidase. *Microbiology* **143:** 527–537.
- Stucky, K., Klein, J.R., Schüller, A., Matern, H., Henrich, B., and Plapp, R. (1995) Cloning and DNA sequence analysis of *pepQ*, a prolidase gene from *Lactobacillus delbrueckii* subsp. *lactis* DSM7290 and partial characterization of its product. *Mol. Gen. Genet.* **247:** 494–500.
- Walker, D.C., Aoyama, K., and Klaenhammer, T.R. (1996) Electrotransformation of *Lactobacillus acidophilus* group A1. *FEMS Microbiol. Lett.* **138:** 233–237.

CHAPTER III

Purification and characterization of the 3-phosphoglycerate kinase from the moderate thermophile *Lactobacillus delbrueckii* subsp. *lactis*.

Aude A. BOURNIQUEL AND Beat MOLLET

Nestlé Research Center, Nestec Ltd., Nestec Ltd., Vers-chez-les-Blanc, P.O. Box 44, CH-1000 Lausanne 26, Switzerland

SUBMITTED TO FEMS Microbiology Letters.

Abstract

The 3-phosphoglycerate kinase (PGK) of the moderate thermophile *Lactobacillus delbrueckii* subsp. *lactis* was purified to homogeneity and found to be a monomeric enzyme with a MW of 45 kDa. PGK is a Michaelis-Menten type enzyme with a K_m=0.7 mM for ATP and a K_m=2.6 mM for 3-phosphoglycerate. The value found for T_{opt}=45°C is consistent for a moderate thermophile bacterium with an optimal thriving temperature of 45°C.

1. Introduction

3-phosphoglycerate kinase (PGK) or [ATP:3-phosphoglycerate 1-phosphotransferase, E.C. 2.7.2.3] catalyses the reversible phosphoryl transfer between 1,3-bisphosphoglycerate and ADP to form 3-phosphoglycerate and ATP, and plays a crucial role in the Embden-Meyerhoff pathway of glycolysis and in gluconeogenesis. This important role is reflected by the enzyme structure that has been highly conserved throughout evolution [1]. In most species, PGK is a monomeric enzyme with a molecular weight around 45 kDa [2,3]. The PGKs isolated from two hyperthermophilic archaebacteriae, *Pyrococcus woesei* and *Methanothermus fervidus* [4], were found to be homomeric dimers in their native state. The moderate thermophile gram-positive bacterium, *Lactobacillus delbrueckii lactis* is widely used in the food industry for the production of Swiss-type hard cheeses. As PGK is one of the key enzymes of the glycolytic pathway, we isolated and characterized the 3-phosphoglycerate kinase from *L. lactis* NCC88.

2. Material and methods

2.1. Bacterial strain and culture conditions

Lactobacillus delbrueckii subsp. *lactis* strain NCC88 was obtained from the Nestlé Culture Collection. For the purpose of protein purification, *L. lactis* was grown at 45°C without agitation in MRS broth (Difco Laboratories) supplemented with 2% glucose and 17% sucrose (MRS-S broth).

2.2. 3-phosphoglycerate kinase purification

L. lactis was grown for 12 hours at 45°C in MRS-S broth. The cells were harvested, washed twice with deionized water and suspended in 1/20th volume Protoplasting Buffer (PB: 50 mM Tris.Cl, pH 7.4; 50 mM EDTA, pH 8.0; 25 % saccharose) containing 1 mg/ml lysozyme. After 2 hours incubation at 37°C, mutanolysin was added to a final concentration of 170 U/ml and the incubation carried on until approx. 99 % of the cells were protoplasted (as observed under light microscopy). The protoplasts were washed with PB to eliminate lysozyme and mutanolysin as well as the cell wall proteins released by the muralytic enzymes. The crude protein extract was obtained by suspending the protoplast pellet in 50 ml lysis buffer (20 mM Tris.Cl, pH 7.4; 50 mM KCl; 1 mM EDTA; 7 mM β-mercaptoethanol). A freeze-thaw step at -20°C ensured total protoplast lysis. The crude protein extract was clarified by low speed centrifugation (16,000 × *g*, 30 min), followed by ultracentrifugation (100,000 × *g*, 2 hours). The concentration of NaCl in the supernatant was adjusted to 0.2 M and nucleic acids were precipitated by addition of 0.4 % polyethyleneimine (PEI). Proteins in the supernatant were precipitated by the slow addition of solid ammonium sulfate to a final saturation of 80 %. The protein pellet was suspended in 8 ml Buffer A (20 mM Tris.Cl, pH 7.4; 50 mM KCl; 10 mM $MgCl_2$; 0.1 mM EDTA; 7 mM β-mercaptoethanol) and loaded onto a DEAE-Sephacel column (Pharmacia, Uppsala, Sweden) equilibrated with Buffer A. Proteins were eluted with a 50-850 mM linear KCl gradient in Buffer A. PGK eluted between 130 and 190 mM KCl. The active fractions were pooled, concentrated in a Centriplus™ 30 (Amicon®) and loaded onto a heparin affinity column (3 x 5 ml Heparin Econo-Pac, Bio-Rad). A 50-550 mM linear KCl gradient was applied. PGK eluted between 160 and 230 mM KCl. The active fractions were pooled, concentrated, then fractionated on a Superose 6 gel filtration column (Pharmacia). The purified enzyme was stable for up to 4 months at 4°C in Buffer A.

2.3. 3-phosphoglycerate kinase activity assay

PGK activity was determined by enzymatic assay, spectrophotometrically following the oxidation of β-NADH at an optical density of 340 nm. The activity of the 3-phosphoglycerate kinase is assayed in the reverse direction of glycolysis in conditions described by Adam [5] substituting 20 mM Tris.Cl, pH 7.4, for triethanolamine and including 1mM ATP. The final volume of the assay mixture was 1.5 ml. Controls contained all additions except 3-phosphoglycerate. The baseline was monitored at OD 340 nm until constant. One unit (U) of 3-phosphoglycerate kinase is defined as the amount of enzyme which converts 1.0 μmole of 3-phosphoglycerate to 1,3-bisphosphoglycerate per minute at 45°C in the assay conditions described above, assuming that the ratio of β-NADH oxidized to 3-phosphoglycerate utilized is unity.

2.4. Protein, molecular techniques and DNA sequencing

General molecular techniques were performed as described by Sambrook *et al.* [6]. Chromosomal DNA from *Lactobacillus delbrueckii* subsp. *lactis* NCC88 was isolated as described by Delley *et al.* [7], except that the incubation step with proteinase K and pronase E was replaced by successive incubations at 37°C with lysozyme (10 mg/ml) and mutanolysin (100 μg/ml). PCR amplification was carried out as previously reported [8]. PCR products were purified using the GeneClean® kit (BIO 101, Inc., Vista, CA, USA) and used directly for sequencing. Sequencing was done according to the dideoxy-chain termination method [9]. Custom-made primers were purchased from Microsynth (Balgach, Switzerland). The sequence data were assembled and analyzed using the Wisconsin package, version 9.1 (Genetics Computer Group [GCG], Madison, WI, USA) and the ClustalW sofware (DNAStar Inc., Madison, WI). Western Blot was performed on a Mini-Protean II unit (Bio-Rad AG, Glattbuch, Switzerland) according to the instructions of the manufacturer. N-terminal sequencing of the protein was performed by Edman degradation.

3. Results and discussion

3.1. Purification of L. lactis 3-phosphoglycerate kinase

Three chromatography steps were necessary to purify *L. lactis* 3-phosphoglycerate kinase to homogeneity as determined by SDS-polyacrylamide gel electrophoresis (Fig. 1) and N-

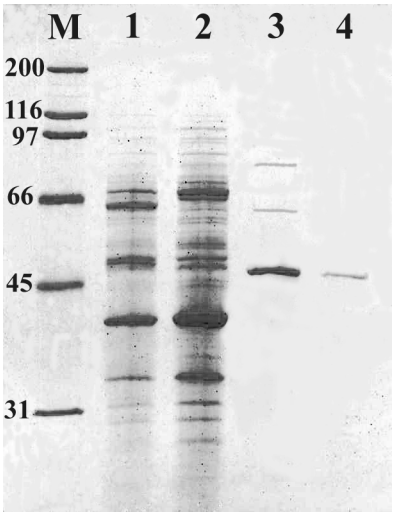


Fig. 1. Purification steps of *L. lactis* 3-phosphoglycerate kinase. Aliquots from the different purification steps were migrated on a SDS-polyacrylamide gel (10% acrylamide). Lane M: molecular weight marker; lane 1: crude extract; lane 2: DEAE-Sephacel column; lane 3: heparin column; lane 4: gel filtration column, pure protein.

terminal sequencing. *L. lactis* PGK was purified 77-fold with a recovery of 36 % of the original activity (Table 1). The size of the purified PGK was estimated on a denaturing SDS-PAGE as being approx 46 kDa (Fig. 1). The elution profile of the native enzyme through a gel filtration column corresponds to a MW of 45 kDa indicating that *L. lactis* PGK is a monomer. This is consistent with previously published results (Table 2) that indicate that bacterial 3-phosphoglycerate kinases are monomers whereas PGKs isolated from hyperthermophile archaebacteria are homodimers.

3.2. Protein and DNA sequences analysis

N-terminal sequencing of the protein by Edman degradation yielded a single 35 aa-sequence: 5'-AKLIVSDVDVKDKKVLVRVDFNVPIKDGVIGDDNR-3'. This sequence was

Table 1
Summary of the purification of *L. lactis* 3-phosphoglycerate kinase.

Step	Total protein (mg)	Total activity (U)	Sp act (U/mg)	Yield (%)	Purification (fold)
Crude extract	753	5903	7.8	100	1.00
DEAE-Sephacel	577	4606	8	78	1.02
Heparin	4.3	2356	547	40	70
Superose 6	3.5	2147	605	36	77

Table 2
Summary of the characteristics of several bacterial 3-phosphoglycerate kinases.

	Identity[a] (%)	Length (aa)	Size (kDa)	Structure	Opt. Growth Temp	GenBank Number	Ref
Bacteria							
L. lactis	100	403	45	monomer	45°C		*
L. bulgaricus	99.5	403	nd[b]	nd[b]	45°C	AJ000339	[10]
B. megaterium	50	394	42.7[c]	nd[b]	30°C	M87647	[11]
B. stearothermophilus	52	394	42.7[c]	nd[b]	55°C	X58059	[12]
T. maritima	57	390	43	monomer	80°C	X75437	[3,13]
Archaebacteria							
M. fervidus	29	410	97	**homodimer**	83°C	M55529	[4,14]
M. bryantii	30	409	39	monomer	37°C	M55530	[14]
P. woesei	30	410	102	**homodimer**	100°C	X73527	[4]

* this study
a: values given by pairwise alignment to the *L.lactis* PGK sequence using Clustal method
b: nd, not determined
c: the value given here is the M_r calculated from the nucleotide sequence

found to be 100 % identical, minus the N-terminal methionine, to the protein deduced from the *pgk* gene of *L. bulgaricus* [10]. Since these protein sequences appear to be highly conserved between the two *Lactobacillus* subspecies, the DNA sequence from *L. bulgaricus* was used to design PCR primers to amplify the *pgk* gene from the chromosomal DNA of *L. lactis*. A 1525 bp–fragment was sequenced. It encodes the entire *pgk* gene, 1211 bp, as well as the 3'-end of the *gap* gene (31 bp) situated upstream of the *pgk* and the 5'-end of the *tpi* gene (168 bp) downstream of the *pgk* gene (Appendix I). With an overall identity of 99.5 % at the nucleotide sequence level, *L. lactis* encodes a *gap* operon that is structurally identical to that described for *L. bulgaricus*. The deduced amino acid sequence of PGK shows an identity of 99.5 % to that of *L. bulgaricus*.

A BLAST search followed by multiple sequence alignment showed that the PGK structure is highly conserved between different bacteria and archaebacteria species. Pairwise sequence comparisons (Table 2) indicate slightly higher similarities with the PGK from the moderate thermophilic bacterium *B. stearothermophilus* than with the one from the mesophilic *B.megaterium*. This is coherent with the fact that *L. lactis*, growing at 45°C, might be considered as a moderate thermophile itself. However, the best identities were found not with the evolutionary close *Bacillus* sp. but with the hyperthermophilic bacteria *T. maritima*. The phylogenetic tree build from the multiple sequence alignment reinforces this observation (Fig.2).

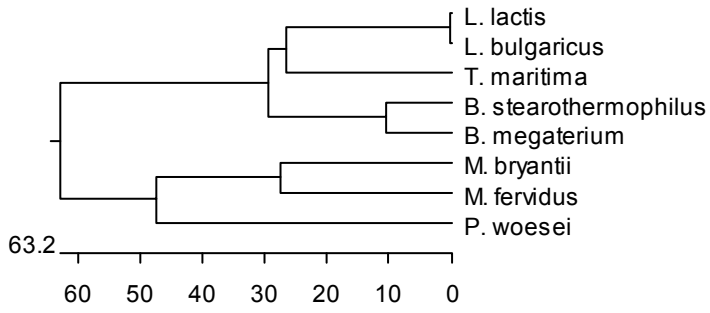


Fig. 2. Schematic representation of the phylogenetic distances between bacterial 3-phosphoglycerate kinases. The length of the branches represents the distance between pairs of sequences, while the scale at the bottom of the diagram indicates the number of substitution events. Sequences were aligned using the ClustalW software.

3.3 PGK characterization with respect to the temperature

The affinity of the 3-phosphoglycerate kinase for its substrates was determined. Since PGK is assayed in the reverse direction of glycolysis, the substrates for the enzymatic reaction are 3-phosphoglycerate and ATP. PGK is not an allosteric enzyme, but displays typical Michaelis-Menten profiles with both of the substrates of the *in vitro* assay. Curves were fitted to the Michaelis-Menten equation: $v = (V_{max} \times [S])/(K_m + [S])$. The K_m and V_{max} were determined at 45°C, the optimal growth temperature of the bacterium (Table 3).

Table 3
Kinetic parameters of *L. lactis* 3-phosphoglycerate kinase.

	V_{max}	K_m
ATP	281 U/mg	0.67 mM
3-phosphoglycerate	270 U/mg	2.64 mM

PGK was characterized with respect to the temperature. As depicted in Fig. 3, the optimal temperature for PGK activity is in the range of 40 to 50°C. These values are consistent for a 3-phosphoglycerate kinase purified from a moderate thermophilic bacterium that has an optimum growth temperature of 45°C but do not explain the close relationship observed with the hyperthermophile *T. maritima* (T_{opt}=80°C). Early and Britt [15] observed that although there is 'a definite dependence upon optimal thriving temperature', other factors such as pH and external pressure must be considered in order to explain some sequence similarities. Our results tend to support these conclusions.

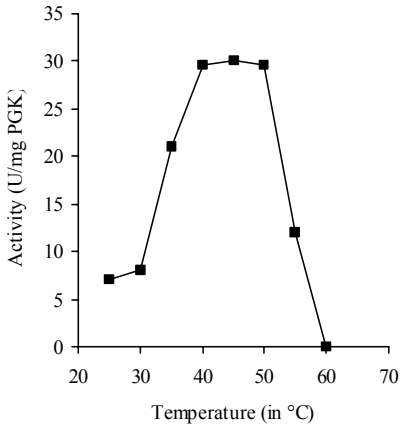


Fig. 3. Determination of the optimal temperature for PGK activity *in vitro*. This determination was done by modifying the temperature of the waterbath regulating the temperature of the cuvettes in the spectrophotometer, allowing 10 min equilibration for each temperature point.

3. Conclusion

The role of 3-phosphoglycerate kinase is crucial in the glycolytic pathway of *L. lactis*, which is an homofermentative lactic acid bacterium and has adapted to its rich growth medium, milk, by eliminating the TCA cycle. The main source of ATP for this bacterium is the degradation of substrate sugars through the Embden-Meyerhoff pathway and the regeneration of ATP through two reactions catalyzed by pyruvate kinase and 3-phosphoglycerate kinase. The pyruvate kinase of *L. lactis* is an allosteric enzyme that has a regulatory role in the physiology of the bacterium (data not shown). The 3-phosphoglycerate kinase is not an allosteric enzyme suggesting that, despite its crucial function in regenerating ATP, PGK is not one of the key enzymes involved in the regulation of *L. lactis* energy metabolism.

Acknowledgements

We are grateful to the Swiss National Science Foundation for financial support within the SPP Food Biotechnology Program no. 5002-044544.

References

[1] Mori N., Singer-Sam, J. and Riggs, A.D. (1986) Evolutionnary conservation of the substrate-binding cleft of phosphoglycerate kinases. FEBS Lett. 204, 313–317.

[2] D'Alessio, G. and Josse, J. (1971) Glyceraldehyde phosphate dehydrogenase, phosphoglycerate kinase, and phosphoglyceromutase of *Escherichia coli*. J. Biol. Chem. 246, 4319–4325.

[3] Grättinger, M., Dankesreiter, A., Schurig, H. and Jaenicke, R. (1998) Recombinant phosphoglycerate kinase from the hyperthermophilic bacterium *Thermotoga maritima*: Catalytic, spectral and thermodynamic properties. J. Mol. Biol. 280, 525–533.

[4] Hess, D., Krueger, K., Knappik, A., Palm, P. and Hensel, R. (1995) Dimeric 3-phosphoglycerate kinases from Hyperthermophilic Archaea: Cloning, sequencing and expression of the 3-phosphoglycerate kinase gene of Pyrococcus woesei in Escherichia coli and characterization of the protein, structural and functional comparison with the 3-phosphoglycerate kinase of Methanothermus fervidus. Eur. J. Biochem. 233, 227–237.

[5] Adam, H. (1965) In: Methods of enzymatic analysis (Bergmeyer, H.U., Ed.), p.539. Academic Press, New-York.

[6] Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) Molecular cloning: A laboratory manual, 2nd Edn., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

[7] Delley, M., Mollet, B. and Hottinger, H. (1990) DNA probes for *Lactobacillus delbrueckii*. Appl. Environ. Microbiol. 56, 1967–1970.

[8] Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., Mullis, K.B. and Ehlich, H.A. (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science 239, 487–491.

[9] Sanger, F., Nicklen, S. and Coulson, A.R.. (1977) DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA 74, 5463–5467.

[10] Branny, P., de la Torre, F. and Garel, J.-R. (1998) An operon encoding three glycolytic enzymes in *Lactobacillus delbrueckii* subsp. *bulgaricus*: glyceraldehyde-3-phosphate dehydrogenase, phosphoglycerate kinase and triosephosphate isomerase. Microbiology 144, 905–914.

[11] Schläpfer, B.S. and Zuber, H. (1992) Cloning and sequencing of the genes encoding glyceraldehyde-3-phosphate dehydrogenase, phosphoglycerate kinase and triosephosphate isomerase (*gap* operon) from mesophilic *Bacillus megaterium*: comparison with corresponding sequences from thermophilic *Bacillus stearothermophilus.* Gene 122, 53–62.

[12] Davies, G.J., Littlechild, J.A., Watson, H.C. and Hall, L. (1991) Sequence and expression of the gene encoding 3-phosphoglycerate kinase from *Bacillus stearothermophilus.* Gene 109, 39–45.

[13] Schurig, H., Beaucamp, N., Ostendorp, R., Jaenicke, R., Adler, E. and Knowles, J.R. (1995) Phosphoglycerate kinase and triosephosphate isomerase from the hyperthermophilic

bacterium *Thermotoga maritima* form a covalent bifunctional enzyme complex. EMBO J. 14, 442–451.

[14] Fabry, S., Heppner, P., Dietmaier, W. and Hensel, R. (1990) Cloning and sequencing the gene encoding 3-phosphoglycerate kinase from mesophilic *Methanobacterium bryantii* and thermophilic *Methanothermus fervidus*. Gene 91, 19–25.

[15] Early, C.N. and Britt, B.M. (1998) Sequence similarities of glyceraldehyde-3-phosphate dehydrogenases, phosphoglycerate kinases, and pyruvate kinases are species optimal temperature-dependent. Eur. Biophys. J. 27, 409–410.

CHAPTER IV

Purification and characterization of the pyruvate kinase of *Lactobacillus delbrueckii* subsp. *lactis*.

Aude A. Bourniquel, Frank Desiere, and Beat Mollet
Nestlé Research Centre, Nestec Ltd., Vers-chez-les-Blanc, P.O. Box 44,
CH-1000 Lausanne 26, Switzerland

SUBMITTED TO FEMS Microbiology Letters.

Abstract

The pyruvate kinase (PYK) from the lactic acid bacterium, *Lactobacillus delbrueckii* subsp. *lactis* was purified and characterised. PYK is an homotetramer and a V-type allosteric enzyme. Positive effectors of PYK are FDP, as well as fructose 6-phosphate and glucose 6-phosphate, which were found to be 1.3– and 1.6–fold more efficient than FDP, respectively. *L. lactis* pyruvate kinase is inhibited by inorganic phosphate and by ATP (EC_{50}=0.75 mM). A consensus motif characteristic of an ATP-binding domain found on a 112 amino acid C-terminal extension of the enzyme might be involved in ATP regulation of *L. lactis* pyruvate kinase.

1. Introduction

Pyruvate kinase, PYK (ATP:pyruvate-*o*-phosphotransferase, EC 2.7.1.40), is one of the key enzymes of the glycolytic pathway. PYK catalyses the transphosphorylation from phosphoenolpyruvate (PEP) to ADP: phosphoenolpyruvate + ADP ⟶ pyruvate + ATP. This reaction is essentially irreversible *in vivo* and appears to be a control point for the regulation of the glycolytic flux. Pyruvate kinases have been isolated from and characterized in many organisms from mammals, e.g. rat, cat, rabbit and bovine [1–4] to bacteria. Pyruvate kinases from various sources show a tetrameric structure consisting of four identical subunits. In mammals, isoenzymes with different kinetic properties were found in various tissues [5]. In bacteria, two main groups of pyruvate kinases have been defined based on the two isoenzymes present in *Escherichia coli*. The first group consists of pyruvate kinases similar to PYK type I from *E. coli* [6] that are activated by fructose 1,6-diphosphate (FDP). Pyruvate kinases

belonging to the second group are similar to PYK type II from *E. coli* [6] and are activated by adenosine 5'-monophosphate or ribose 5-phosphate. A few bacterial pyruvate kinases seem to diverge from these two groups. The pyruvate kinases from *Streptococcus mutans* [7] and *Lactobacillus bulgaricus* [8] are only slightly affected by the activators described above and require glucose 6-phosphate (Glu6P) for activation. PYK from *Bacillus subtilis* is activated only by its substrate PEP [9]. Eukaryotic pyruvate kinases are known to be inhibited by MgATP, whereas ATP inhibition of *E. coli* pyruvate kinase was shown to be dependent on the presence of succinyl CoA [6]. All pyruvate kinases, except the muscle enzyme [10], are allosteric enzymes.

Growth in a rich medium, milk, has allowed lactic acid bacteria (LAB) to adapt and "simplify" their carbon metabolism so that they lack the pentose phosphate pathway as well as the TCA cycle and electron transfer chain. ATP is generated by substrate level phosphorylation. Although LAB do possess an enzyme very similar to the ATP synthase, the enzyme catalyses mostly the reverse reaction, i.e. the hydrolysis of ATP with concomitant pumping of protons out of the cells [11–13]. In homofermentative LAB such as *Lactobacillus delbrueckii* subsp. *lactis*, the carbon source, mainly lactose, is therefore utilized via the Embden-Meyerhof pathway to produce lactic acid and ATP [11]. The enzymatic reaction catalyzed by pyruvate kinase is all the more important in those bacteria since it is one of only two major ATP-regenerating steps present in the cell. In this work we present the purification of the pyruvate kinase from *L. lactis* NCC88. We study its mode of regulation in the absence of both succinyl CoA and ribose-5-phosphate, and its importance in the modulation of glycolysis and cell metabolism.

2. Material and methods

2.1. Bacterial strains and culture conditions

Lactobacillus delbrueckii subsp. *lactis* strain NCC88 was obtained from the Nestlé Culture Collection. XL1-Blue *E. coli* cells (Stratagene®) were used for cloning the PCR amplified *pyk* gene. NCC88 was grown at 45°C without agitation in MRS broth (Difco Laboratories) containing 20 g/l of glucose, and 170 g/l of sucrose (MRS-S broth). *E. coli* was routinely grown in Luria-Bertani medium at 37°C with agitation. Ampicillin was used at 100 µg/ml.

2.2. Pyruvate kinase purification

L. lactis was inoculated in MRS-S broth and incubated for 16 h at 45°C. The cells were harvested by centrifugation, washed twice with deionised water and resuspended in protoplasting buffer (PB: 50mM Tris-HCl, pH 7.4; 50 mM EDTA, pH 8.0; 25% sucrose). Successive incubations at 37°C in PB with lysozyme (1 mg/ml) then mutanolysin (170 U/ml) led to the formation of protoplasts which were harvested, suspended in lysis buffer (20 mM Tris-HCl, pH 7.4; 50 mM KCl; 1 mM EDTA; 14 mM β-mercaptoethanol) then frozen at –20°C to ensure total lysis. The frozen mix was thawed at 37°C and clarified by low speed centrifugation (16,000 × g, 30 min) followed by ultracentrifugation (100,000 × g, 2 h). Nucleic acids were removed by precipitation with 0.4 % polyethyleneimine in the presence of 0.2 M NaCl. Proteins were concentrated by precipitation with ammonium sulfate (80% saturation) and resuspended in buffer A (20 mM Tris-HCl, pH 7.4; 50 mM KCl; 10 mM $MgCl_2$; 0.1 mM EDTA; 14 mM β-mercaptoethanol). The sample was loaded onto a DEAE-Sephacel column (Pharmacia Biotech). Proteins were eluted with a 50-850 mM KCl linear gradient in buffer A at a flowrate of 1 ml/min. Active fractions were pooled, concentrated, then fractionated on a Superose 6 gel filtration column (Pharmacia) at a flowrate of 0.3 ml/min. The active fraction was stored at –20°C in 50% glycerol.

2.3. Pyruvate kinase activity assay

Enzyme activity was determined by the lactate dehydrogenase-coupled enzyme assay. The assays were carried out at 45°C, the optimal growth temperature of *L. lactis*. Unless otherwise specified the reaction mixture consisted of 50 mM Tris-HCl, pH 7.4; 50 mM KCl; 7 mM $MgCl_2$; 5 mM ADP; 6 mM phosphoenolpyruvate (PEP); 10 U/ml of lactate dehydrogenase (LDH); 0.15 mM β-NADH. One unit of pyruvate kinase activity is defined as the quantity of enzyme necessary to transform 1 μmol of PEP per min under the assay conditions described above. Both K^+ and Mg^{2+} were required for pyruvate kinase activity. The ratio of free Mg^{2+} ions to ATP was kept constant. The oxidation of β-NADH was followed spectrophotometrically at 340 nm and the steady-state kinetic velocity, v, determined. All data sets were fitted to the modified version of the Hill equation as shown in Equation 1

$$v = \frac{V_{max} \quad [S]^h}{K^h_{app} + [S]^h} \qquad \text{(Eq. 1)}$$

2.4. Western blot and protein sequencing

Western blot transfer was performed as recommended by Bio-Rad (Bio-Rad Laboratories AG). N-terminal sequencing of the protein was performed by Edman degradation.

2.5. Molecular techniques and data analysis

General molecular cloning techniques were performed as described by Sambrook *et al.* [14]. Chromosomal DNA of *L. lactis* was isolated as described earlier [15]. PCR amplification was carried out as previously reported [16]. Custom-made oligonucleotide primers were purchased from Microsynth (Switzerland). The PCR products were cloned into pGEM®-T vector (Promega). The DNA sequence of the inserts was determined using the dideoxy-chain termination method [17] with the universal forward and reverse primers. The DNA sequence was confirmed by sequencing both DNA strands from at least 2 independent, cloned PCR products. Sequence data were assembled and analyzed using the Wisconsin package, version 10.0, Genetics Computer Group (GCG), Madison, WS. Protein modeling was carried out using the SWISS-MODEL [18] and the Swiss-Pdb Viewer [19] softwares available on the web.

3. Results

3.1. Purification, molecular weight and subunit structure of the pyruvate kinase

The *L. lactis* pyruvate kinase was purified using a two-step chromatography protocol. The first step consisted in a DEAE-Sephacel, weak-anion exchange column that yielded a single peak of pyruvate kinase activity. The active fraction was fractionated on a Superose 6 gel filtration column. All the PYK activity was again contained in a single peak indicating that *L. lactis* possesses only one pyruvate kinase. PYK was purified to homogeneity as observed by electrophoresis on SDS-polyacrylamide gel (Fig. 1). A single band of approx. 66-68 kDa can be detected. The elution profile of the native protein through a gel filtration column corresponds to a MW of 279 kDa. N-terminal sequencing by Edman degradation gave a unique sequence (MKKTKIVSTLGPASN) indicating that the active enzyme is a multimer of identical subunits. From the above values, we can conclude that the *L. lactis* pyruvate kinase is an homotetramer, consistent with the structure of all PYKs characterized to date.

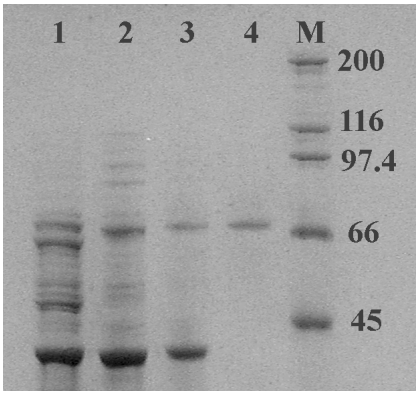


Fig. 1. Purification steps of the *L. lactis* pyruvate kinase. Aliquots from the different purification steps were migrated on a SDS-polyacryamide gel (7.5 % acryamide). Lane 1: crude extract, lane 2: ammonium sulfate precipitation, lane 3: DEAE-Sephacel column, lane 4: Superose 6 column, pure protein, lane M: marker.

3.2. Substrate kinetics of pyruvate kinase

The substrate saturation curves for both substrates, PEP and ADP, were sigmoidal which is characteristic of an allosteric enzyme. PYK was inhibited by high concentrations of PEP (data not shown). The V_{max} of the enzyme for each substrate was found to be dependent on the concentration of the second substrate (Table 1). The degree of cooperativity of the allosteric enzyme, expressed by the Hill coefficient, h, is maximal (h = 4.0) when PEP is the varied substrate, independently of the fixed concentration of ADP. However, when ADP is the varied substrate, increasing concentrations of PEP induced a decrease in the degree of cooperativity suggesting that, at PEP saturation, h will be equal to one.

3.3. Several effectors of pyruvate kinase

FDP is an heterotropic activator of *L. lactis* pyruvate kinase. As seen in Table 2, K_{app} remained constant at different FDP concentrations while h and the V_{max} values varied. V_{max} increased with increasing concentrations of FDP. This is characteristic of a V-type allosteric

Table 1
Steady-state kinetic parameters for *L. lactis* pyruvate kinase substrates.

Substrate		Hill number	$K_{0.5}$	V_{max}
Varied	Fixed	h	(mM)	(U/mg)
PEP	ADP: 5.0 mM	4.0	4.82	0.34
	ADP: 2.5 mM	4.0	5.57	0.29
	ADP: 1.0 mM	4.0	5.75	0.14
ADP	PEP: 6.5 mM	1.3	1.82	0.335
	PEP: 5.0 mM	3.1	1.81	0.14
	PEP: 3.5 mM	3.7	1.88	0.06

Table 2

Influence of different FDP concentrations on *L. lactis* pyruvate kinase.

Substrate		Effector	Hill number	K_{app}	V_{max}
Varied	Fixed		*h*	(mM)	(U/mg)
PEP	ADP: 5 mM	FDP: 0.1 mM	4.0	2.17	1.025
		FDP: 0.2 mM	3.3	2.15	1.25
		FDP: 0.5 mM	2.9	2.17	2.1

enzyme. The effect on PYK activity of two precursors of FDP during glycolysis, fructose 6-phosphate (Fru6P) and glucose 6-phosphate (Glu6P), was also determined (Fig. 2). Both compounds are heterotropic activators of PYK, and both are more effective than FDP . While FDP induced a 33-fold increase in activity, Fru6P and Glu6P induced respectively a 44- and a 54-fold increase. As none of these activators released the substrate inhibition by PEP, the increase in V_{max} is concomitant with a shift of the V_{max} towards lower PEP concentrations. AMP does not activate the PYK enzyme. Lactic acid, the end product of glycolysis in lactobacilli, has no effect on PYK activity either.

The effect on PYK activity of two potential inhibitors were evaluated (Fig. 3). Inorganic phosphate (K_2HPO_4-KH_2PO_4, pH 7.4) is a very effective inhibitor causing 50% inhibition at 0.2 mM concentration. A concentration of 0.75 mM ATP is sufficient to cause 50 % inhibition of the *L. lactis* enzyme activity. The presence of succinyl CoA is not required. Inactivation of PYK by P_i appears to be total at 2.5 mM, whereas a maximum of approx. 89% inhibition is obtained with 4 mM or more ATP.

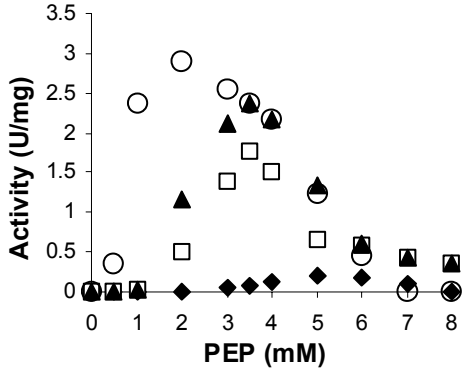


Fig. 2. Comparative effect of the activators: FDP, Fru6P, and Glu6P. The assay mixture is identical to that described in the text, except for: (□) 0.5 mM0 FDP, (▲) 0.5 mM Fru6P, and (O) 0.5 mM Glu6P. No activator (◆).

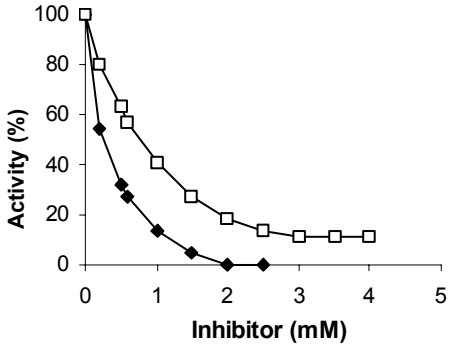


Fig. 3. Inhibition of the *L. lactis* pyruvate kinase by inorganic phosphate (◆) and ATP (□). The assay mixture was identical to that described in the text except for the presence of the inhibitors.

3.4. Analysis of the pyruvate kinase amino acid sequence

The N-terminal sequence of the *L. lactis* pyruvate kinase, obtained by Edman degradation, was compared to the sequences present in the GenBank-EMBL databases using the BLAST algorithm [20]. A very clear homology was found with the N-terminal protein sequences of bacterial pyruvate kinases. The highest homology was to *Lactobacillus bulgaricus* [8] with 14 residues identical out of 15. The sequence of the *pyk* gene from *L. bulgaricus* (GenBank Accession Number X71403) was therefore used as template to design the PCR primers needed to amplify the *pyk* gene from *L. lactis*. The gene was amplified by PCR, cloned in *E. coli*, sequenced, then translated into its amino acid sequence (Appendix II). As expected, homologies were found with the pyruvate kinases present in the protein databases. The highest score was obtained for *L. bulgaricus* PYK with 98% identity. The next highest homologies were to *Bacillus* sp. (55% identity with PYK from *B. stearothermophilus*), and *Lactococcus lactis* (54% identity), then to the type I *E. coli* pyruvate kinase (49% identity). The identity is much lower with the type II isoenzyme of *E. coli*, i.e. 39 %, suggesting that the *L. lactis* pyruvate kinase belongs to the type I group. The phylogenetic relationships represented in Fig. 4 show clearly the significant evolutionary distance existing between the type I and type II isoenzymes in gram-negative bacteria. The distinction, however, is not so definite in gram-positive microorganisms. The pyruvate kinase from *B. licheniformis* [21] known to have all the characteristics of a type II isoenzyme is found clustered with the other gram-positive enzymes. The appurtenance of the pyruvate kinase from a gram-positive bacterium to one type of isoenzyme cannot therefore be determined solely on the basis of

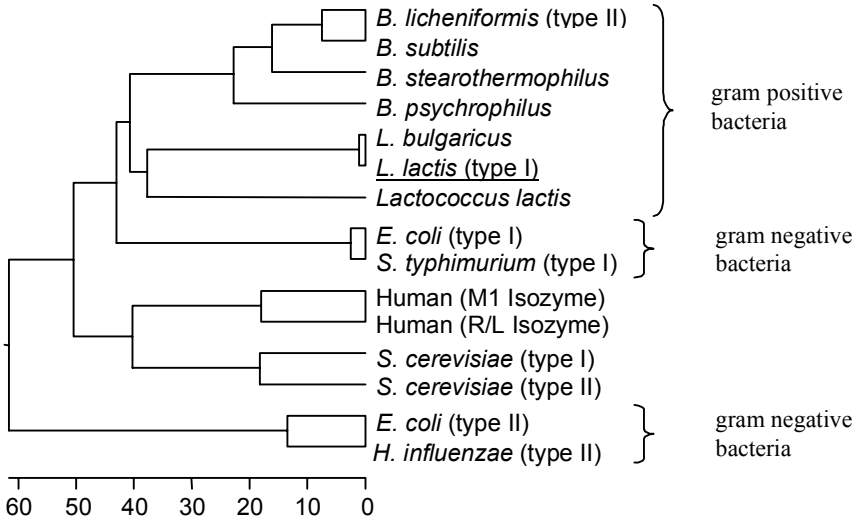


Fig. 4. Schematic representation of the phylogenetic distances between pyruvate kinases. The length of the branches represents the distance between pairs of sequence, while the scale at the bottom of the diagram indicates the number of substitution events. Sequences were aligned using the ClustalW software. The phylogenetic tree was constructed using the DNAStar software (Madison, WI).

primary sequence homologies, biochemical properties must be considered.

3.5. Functional analysis based on secondary structure and 3D-modeling

Based on the primary sequence, the *L. lactis* enzyme can be divided in three distinct components. The larger one, from residues 1 to 476, corresponds to the actual pyruvate kinase including its catalytic site and displays homology to the whole *E. coli* type I PYK. 3D-modeling of this domain to the *E. coli* 3D-structure showed clearly that although sequence identity at the amino acid level is only 49%, the structure itself is highly conserved (Fig. 5A).

The second and third components are both located within a C-terminal extension of 110 residues reported previously only in *Bacillus* sp. [22, 23] and *L. bulgaricus* [24]. The C-terminal domain composed of residues 522 to 589 encodes a motif (PROSITE #PDOC00527) characteristic of PEP-utilizing enzymes such as pyruvate, orthophosphate dikinase (PPDK, EC 2.7.9.1) found in plants, PEP synthase (EC 2.7.9.2) essential for gluconeogenesis, or phosphoenolpyruvate-protein phosphotransferase (EC 2.7.3.9) i.e. enzyme I (EI) of the phosphoenolpyruvate-dependent sugar phosphotransferase system (PTS). Modeling experiments were successful when based on the amino terminal domain of enzyme I from *E. coli*. This C-terminal component of the *L. lactis* PYK enzyme, to which we will refer as the EI-like domain, displays a structure highly similar to the C-terminus of the enzyme I from *E. coli* (Fig. 5B). The presence of this EI-like domain suggests a dual function for the *L. lactis* pyruvate kinase, where PEP can be transformed into pyruvate either by the classical pyruvate kinase activity that produces ATP, or by the EI-like activity that does not produce ATP but phosphorylates the phosphoryl carrier protein (HPr) of the PTS.

An interesting feature of this enzyme is the presence of the third component, a short, 45 residue-long domain linking the N-terminal pyruvate kinase domain to the C-terminal EI-like domain. No homology with previously determined 3D-structures could be found for this intermediary domain, nor does it display any obvious secondary structure. It encodes, however, a consensus pattern usually found in protein kinases, and known to be involved in ATP-binding (PROSITE #PS00107). The presence of this ATP-binding domain in this intermediary domain suggests the possibility of a regulatory function.

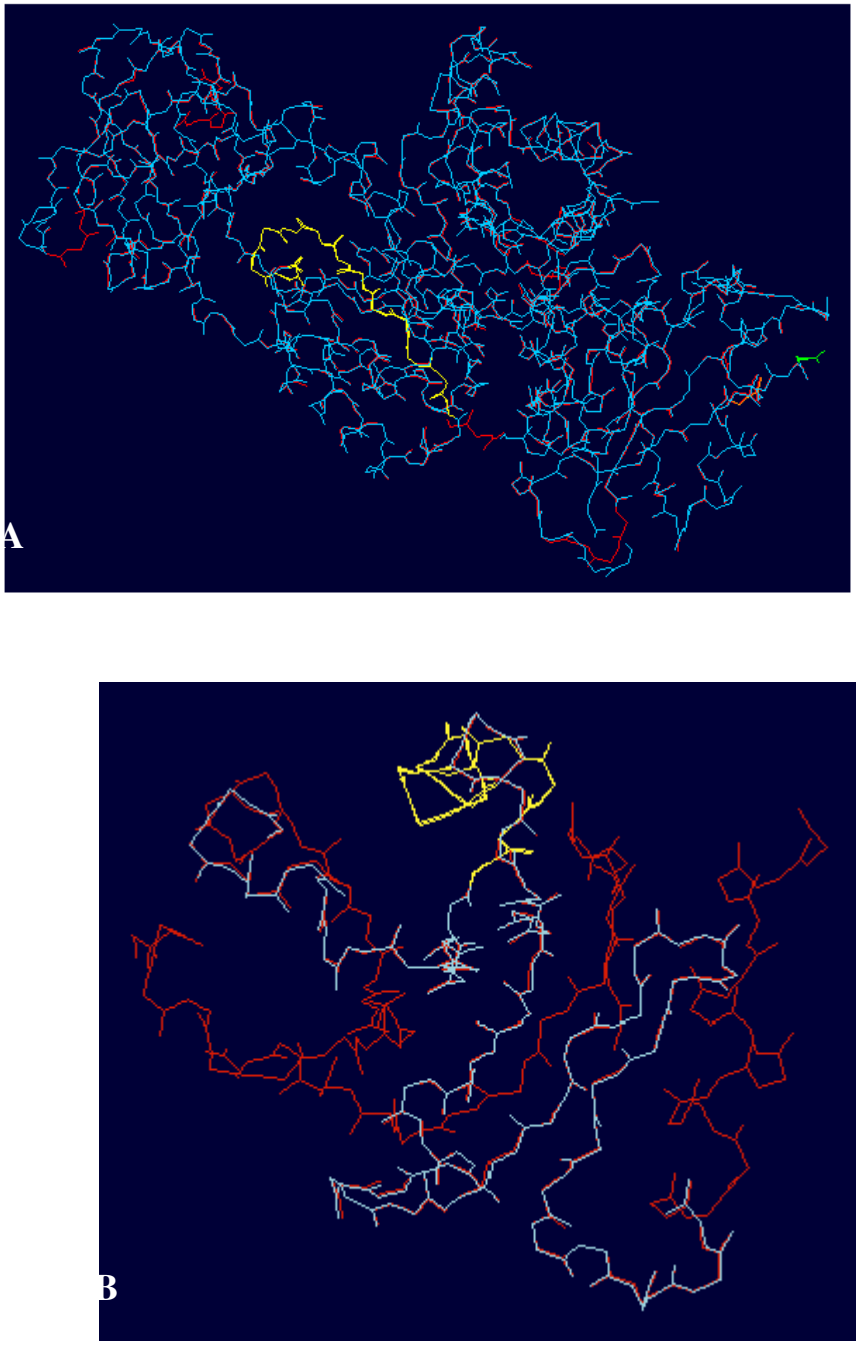


Fig. 5. 3D-modeling of *L. lactis* pyruvate kinase displays high structural homologies to A) the type I pyruvate kinase of *E. coli* (PDB # 1PKY), and B): the amino terminal domain of the *E. coli* EI enzyme (PDB # 1ZYM). The *L. lactis* enzyme is represented in blue, the *E. coli* enzymes in red. The conserved part of the structures are pink. In A) the last amino acid modelled is represented in green. The active sites are represented in yellow.

4. Discussion

The pyruvate kinase from *L. lactis* has been purified and characterized. Due to the instability of the enzyme in the following storage conditions (room tempearture, 4°C, -20°C, with or without 50% glycerol), we were unable to determine a purification rate or yield. We still managed a minimal characterization and ascertained that *L. lactis* pyruvate kinase is an homotetramer and an allosteric enzyme, as are all pyruvate kinases characterized to date.

L. lactis PYK is activated by FDP, inhibited by ATP, but not affected by AMP. These three criteria firmly place *L. lactis* pyruvate kinase within the type I group of bacterial pyruvate kinases. This appurtenance is further reinforced by the high identities, at the amino acid sequence level, between *L. lactis* PYK and the type I isoenzyme from *E. coli*, and by the biochemical properties of the enzyme. Activation by FDP induces a change in the maximal velocity that is characteristic of a V-type allosteric enzyme (nomenclature of Monod, Wyman and Changeux [25]). Type I pyruvate kinases such as those from *E. coli* and *Streptococcus lactis* [26] have been shown to be V-type enzymes, while type II PYKs such as the *B. licheniformis* enzyme [21], as well as the pyruvate kinases purified from yeast and liver, are K-type allosteric enzymes. Despite these similarities with other type I pyruvate kinases, the *L. lactis* enzyme shows several diverging properties. FDP is not the most potent activator of the enzyme. Fru6P and especially Glu6P, the precursors of FDP during glycolysis, are much more effective as activators. Thus, the *L. lactis* PYK displays some similarities with the ones from *S. mutans* [7] and *L. bulgaricus* [8] that are not affected by FDP but require Glu6P for activation. The *L. lactis* pyruvate kinase is also the only one, from all bacterial PYKs characterized to date, to show substrate inhibition by PEP. Since the *L. lactis* enzyme is a type I pyruvate kinase, it is not affected by the lack of ribose 5-phosphate in the cell, ribose 5-phosphate being an activator for type II pyruvate kinases. What is more interesting, is that the absence of succinyl CoA does not affect the *L. lactis* pyruvate kinase either. Although succinyl CoA was shown to be necessary for ATP inhibition in *E. coli* [6], the inhibition by ATP of the *L. lactis* enzyme is very efficient by itself and does not require the presence of succinyl CoA.

On a C-terminus extension common to *L. bulgaricus* [24] and *Bacillus* ssp. [22, 23], the *L. lactis* pyruvate kinase possesses an ATP-binding site purported to be involved in ATP regulation of the enzyme activity. ATP has previously been shown to regulate pyruvate kinases via a product inhibition mechanism [10, 27] which is not consistent with the partial ATP inhibition (89% inhibition) observed for the *L. lactis* enzyme. We speculate that the C-

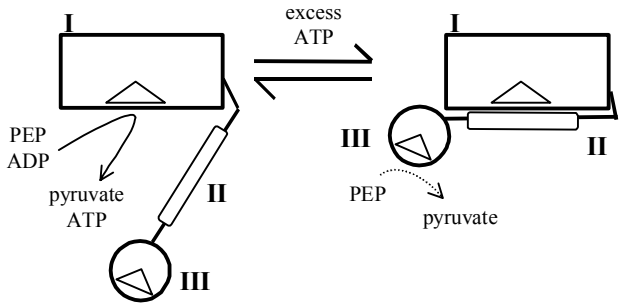


Fig. 6. Speculative model of the role of the C-terminal extension in ATP inhibition of *L. lactis* pyruvate kinase. The three component parts of the enzyme are represented by: I, the N-terminal component with a high structural homology to *E. coli* pyruvate kinase; II, the intermediate domain encoding the ATP-binding motif; III, the C-terminal component with a structural homology to the EI active domain of *E. coli*. ⌓ represents the active site in components I and III. ▭ represents the ATP-binding motif of domain II.

terminal extension of the protein act as a shunt with the ATP-binding site blocking ATP within the active site of the pyruvate kinase (Fig. 6). The entrapment of ATP in the active site would result in a reinforced product inhibition effect so that the efficiency of ATP as an inhibitor would be significantly increased. Once the pyruvate kinase active site were blocked, however, PEP might still be transformed into pyruvate by the EI-like domain, which would explain the residual activity observed in Fig. 2 during ATP inhibition.

During exponential growth of the bacteria, the levels of intracellular PEP and P_i are low, whereas the level of FDP is high, FDP appearing as the storage metabolite of the glycolytic pathway [28]. Moreover, since PYK catalyses one of only two major reactions able to produce ATP in *L. lactis*, and since the energy requirement of exponentially growing cells is high, the concentration of ATP must be low. PYK is thus activated by FDP and due to the shift of the V_{max} in the presence of the activator, the enzyme is very active at the prevalent low PEP concentrations. By the time cells reach the stationary growth phase, the lack of readily available sugars has induced a depletion in the level of glycolytic intermediates in the cell and induced a slowdown of cell metabolism. With the decrease of energy requirement, the level of ATP increases, which concomitant with a decrease in FDP concentration, induces inhibition of the pyruvate kinase activity. This in turn induces an increase in PEP concentration that reinforces the inhibition. In the model proposed by Thompson [28] for streptococci, this pool of PEP constituted during starvation was designated *PEP potential*. It has the function of providing maintenance energy to the bacterium by allowing residual PYK activity and thus production of ATP. Once a sugar becomes available, thanks to the high potential of Glu6P as an activator, PYK is able to resume a high level of activity providing ATP for glycolysis and the restoration of normal cell functions. Thus, pyruvate kinase plays a key role in the modulation of glycolysis and cell metabolism in *L. lactis* and more generally in lactic acid bacteria.

Acknowlegments

We are very grateful to R. D. Pridmore and S. Kochhar for critical reading of the manuscript. This work was supported by the Swiss National Foundation within the SPP Food Biotechnology Program no. 5002-044544.

References

[1] Riou, J.P., Claus, T.H. and Pilkis, S.J. (1978) Stimulation by glucagon of *in vivo* phosphorylation of rat hepatic pyruvate kinase. J. Biol. Chem. 253, 656–659.

[2] Muirhead, H., Clayden, D.A., Barford, D., Lorimer, C.G., Fothergill-Gilmore, L. A., Schiltz, E. and Schmitt, W. (1986) The structure of cat muscle pyruvate kinase. *EMBO J.* 5, 475–481.

[3] Larsen, T.M., Laughlin, T., Holden, H.M., Rayment, I. and Reed, G.H. (1994) Structure of rabbit muscle pyruvate kinase complexed with Mn^{2+}, K^{+}, and pyruvate. Biochemistry 33, 6301–6309.

[4] Cardenas, J.M. (1982) Pyruvate kinase from bovine muscle and liver. In: Methods in Enzymology (Wood, W.A., Ed.), pp. 140–149, Academic Press Inc, New York.

[5] Hall, E.R. and Cottam, G.L. (1978) Isozymes of pyruvate kinase in vertebrates: their physical, chemical, kinetic and immunological properties. Int. J. Biochem. 9, 785–794.

[6] Waygood, E.B. and Sanwal, B.D. (1974) The control of pyruvate kinases of *Escherichia coli*. I. Physicochemical and regulatory properties of the enzyme activated by fructose 1,6-diphosphate. J. Biol. Chem. 249, 265–274.

[7] Abbe, K. and Yamada, T. (1982) Purification and properties of pyruvate kinase from *Streptococcus mutans*. J. Bacteriol. 149, 299–305.

[8] Le Bras, G. and Garel, J.-R. (1993) Pyruvate kinase from *Lactobacillus bulgaricus*: Possible regulation by competition between strong and weak effectors. Biochimie 75, 797–802.

[9] Diesterhaft, M. and Freese, E. (1972) Pyruvate kinase of *Bacillus subtilis*. Biochim. Biophys. Acta 268, 373–380.

[10] Ainsworth, S. and Macfarlane, N. (1973) A kinetic study of rabbit muscle pyruvate kinase. Biochem. J. 131, 223–236.

[11] Axelson, L. (1993) Lactic acid Bacteria: Classification and physiology. In: Lactic Acid Bacteria, Microbiology and Functional Aspects (Salminen, S. and von Wright, A., Eds), 2nd Edn., pp. 1-72, Marcl Dekker Inc.

[12] Konings, W.N., Poolman, B. and Driessen, A.J.M. (1989) Bioenergetics and solute transport in lactococci. Crit. Rev. Microbiol. 16, 419–476.

[13] Maloney, P.C., Kashket, E.R. and Wilson, T.H. (1974) A protonmotive force drives ATP synthesis in bacteria. Proc. Natl. Acad. Sci. USA 71, 3896–3900.

[14] Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) Molecular cloning: A laboratory manual, 2nd Edn., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

[15] Delley, M., Mollet, B. and Hottinger, H. (1990) DNA probes for *Lactobacillus delbrueckii*. Appl. Environ. Microbiol. 56, 1967–1970.

[16] Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., Mullis, K.B. and Ehlich, H.A. (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science 239, 487–491.

[17] Sanger, F., Nicklen, S. and Coulson, A.R.. (1977) DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA 74, 5463–5467.

[18] http://www.expasy.ch/swissmod/SWISS-MODEL.htm

[19] Guex, N. and Peitsch, M.C. (1997) SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. Electrophoresis 18, 2714–2723. Http://www.expasy.ch/spdbv/mainpage.htm

[20] Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) Basic local alignment search tool. J. Mol. Biol. 215, 403–410.

[21] Tuominen, F.W. and Bernlohr, R.W. (1971) Pyruvate kinase of the spore-forming bacterium, *Bacillus licheniformis*. II. Kinetic properties. J. Biol. Chem. 246, 1746–1755.

[22] Sakai, H. and Ohta, T. (1993) Molecular cloning and nucleotide sequence of the gene for pyruvate kinase of *Bacillus stearothermophilus* and the production of the enzyme in *Escherichia coli*. Eur. J. Biochem. 211, 851–859.

[23] Tanaka, K., Sakai, H., Ohta, T. and Matsuzawa, H. (1995) Molecular cloning of the genes for pyruvate kinase of two bacilli, *Bacillus psychrophilus* and *Bacillus licheniformis*, and comparison of the properties of the enzymes produced in *Escherichia coli*. Biosci. Biotech. Biochem. 59, 1536–1542.

[24] Branny, P., De La Torre, F. and Garel, J.-R. (1996) The genes for phosphofructokinase and pyruvate kinase of *Lactobacillus delbrueckii* subsp. *bulgaricus* constitute an operon. J. Bacteriol. 178, 4727–4730.

[25] Monod, J., Wyman, J., and J.-P. Changeux. 1965. J. Mol. Biol. 12, 88.

[26] Collins, L.B. and Thomas, T.D. (1974) Pyruvate kinase of *Streptococcus lactis*. J. Bacteriol. 120, 52–58.

[27] Macfarlane, N. and Ainsworth, S. (1972) A kinetic study of baker's-yeast pyruvate kinase activated by fructose 1,6-diphosphate. Biochem. J. 129, 1035–1047.

[28] Thompson, J. (1987) Regulation of sugar transport and metabolism in lactic acid bacteria. FEMS Microbiol. Rev. 46, 221–231.

CHAPTER V

Lactobacillus delbrueckii subsp. *lactis* plasmids reveal evidence of an endogenous Type I restriction-modification system in their parent strains.

Aude A. Bourniquel[1], Michael G. Casey[2], Beat Mollet[1] and R. David Pridmore[1]

Nestlé Research Center[1], Nestec Ltd., Molecular Biology Group,Vers-chez-les-Blanc, CP 44, CH-1000 Lausanne 26, Switzerland; FAM[2], Swiss Federal Dairy Research Institute Liebefeld, Department of Biochemistry, Schwarzenburgstr. 161, CH-3003 Bern, Switzerland

SUBMITTED TO the Journal of Bacteriology.

The plasmids pN42 and pJBL2 were isolated from the *Lactobacillus delbrueckii* subsp. *lactis* strains NCC88 and JBL2. DNA sequence determination and bioinformatic analysis revealed a strikingly conserved genetic organization containing five major, highly conserved open reading frames. Transformation studies indicated that ORF-2 (a replicative DNA helicase), *ori* and ORF-3 constitute the minimal requirements for replication of pN42 in the heterologous host *Lactococcus lactis*. ORF-1, with the lowest overall homology (58% identity), is predicted to encode a type I restriction-modification (R-M) system HsdS specificity subunit, suggesting that these plasmids may be involved in host defense by expanding the host R-M system repertoire. We also present evidence of the presence as well as the deduced minimal consensus recognition site of an endogenous site-specific N^6-methyladenine methyltransferase that may be part of the host R-M system.

Lactobacillus delbrueckii subsp. *lactis* is a lactic acid bacterium (LAB) widely used in the dairy industry for the production of Swiss-type hard cheeses, e.g. Emmental, Gruyère, Comté. It is generally acknowledged that the current trend to use fewer, more finely tuned starter cultures at an industrial scale has promoted the evolution of novel virulent bacteriophages (phages) able to infect these strains. Phage attacks have thus become a major problem in the dairy industry where infection can lead to delays in the acidification of milk or even loss of the product (3, 25). In contrast to other LAB like *Lactococcus lactis* or *Streptococcus thermophilus*, very few phages are known to be able to target *L. lactis* strains suggesting that *L. lactis* possess a very active and reliable endogenous defense mechanism. Four distinct types of host-mediated phage defense mechanism are currently known to occur in LAB. They include adsorption inhibition, prevention of phage DNA penetration, restriction-modification

(R-M) and abortive infection. Many of these defense mechanisms can be encoded on plasmid DNA (7, 16). In this paper, we present the characterization of plasmids from two *L. lactis* strains NCC88 and JBL2, with conserved genetic organization and encoding proteins with a strong homology to type I R-M enzyme specificity subunits. The analysis of plasmid pN42 also provides evidence of the presence of an endogenous site-specific modification system in *L. lactis* NCC88.

MATERIAL AND METHODS

Bacterial strains, plasmids and culture conditions. *Lactobacillus delbrueckii lactis* NCC88 containing the plasmid pN42 was obtained from the Nestlé Culture Collection. The *Lactobacillus delbrueckii* subsp. *lactis* strains JBL1, JBL2, JCL3, JCL4, JCL5, JCL6, JCL7, JCL8, JCL9, JCL27, JCL28 and JCL30 were obtained from the Swiss Federal Dairy Research Institute, Liebefeld-Bern, Switzerland. *L. lactis* strains were grown in MRS broth (Difco, Detroit, MI) at 45°C without shaking. *Lactococcus lactis* strain MG1363 (12) was cultivated in M17 medium supplemented with 0.5% glucose at 30°C without agitation. *Escherichia coli* XL1-Blue was obtained from Stratagene (La Jolla, CA) and routinely grown in Luria-Bertani medium with aeration at 37°C. Alternatively, *E. coli* cells were grown in BHI broth (Difco, Detroit, MI) when erythromycin (Em) was added to the growth medium. Antibiotics were added at the following concentrations: ampicillin (Ap) at 100 µg/ml, Em at 150 µg/ml for *E. coli*; and Em at 4 µg/ml for lactococci. Functional β-galactosidase was detected by the addition of 300 ng of 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-gal) ml^{-1} and 60 ng of isopropyl-β-D-thiogalactopyranoside (IPTG) ml^{-1} to Luria-Bertani solid medium. The plasmids used were the *E. coli* plasmids pUC18 (35) and pJDC9 (6), and the shuttle-vector pNZ124 (26).

Maxipreparation of plasmid DNA from *L. lactis* cultures. This protocol was adapted from the protocol published by O'Sullivan and Klaenhammer (23). 100 ml MRS broth containing 2% glucose and 17% sucrose (MRS-S broth) were inoculated with a *L. lactis* preculture in MRS-S broth and incubated overnight at 45°C. The cells were harvested, washed twice with 20 ml sterile distilled water, suspended in 10 ml PB (Tris-HCl, 50 mM, pH 7.4; EDTA, 50 mM, pH 8.0; sucrose 25%) containing 1 mg/ml lysozyme, and incubated at 37°C for 2 hours. 2.5 mg/ml mutanolysin were added and the incubation continued till most of the cells appeared as protoplasts when observed by light microscopy. The protoplasts were harvested, washed with 20 ml PB then lysed by suspension in 4 ml TER (50 mM Tris.Cl, pH 7.4; 50 mM EDTA, pH 8.0; 0.5 mg/ml RNaseA) and incubated at 37°C for 15 min. Eight

milliliters of freshly prepared alkaline SDS solution (3% SDS, 0.2 N NaOH) were added and immediately mixed by inverting the tube several times. After 7 min incubation at room temperature, 6 ml of ice-cold sodium acetate (3 M, pH 4.8) were added, mixed till the solution was clear, then incubated on ice for 10-20 min and centrifuged at 5'000 rpm for 35 min at 4°C. The supernatant was transferred to a 50 ml-tube containing 13 ml isopropanol, well mixed and centrifuged at 5'000 rpm for 35 min at 4°C. All liquid was carefully removed before suspending the DNA pellet in 6.4 ml distilled water, 3.8 ml ammonium acetate 8 M, 0.2 ml ethidium bromide (10 mg/ml) and 7 ml phenol-chloroform. The mix was centrifuged at 5'000 rpm for 15 min at 4°C. The upper phase was transferred to a new tube, and then chloroform extractions were performed till all traces of ethidium bromide had been removed. 10 ml cold ethanol were then added and centrifuged at 5'000 rpm for 35 min at 4°C. The pellet was washed with 70% ethanol, air dried, and finally resuspended in 2 ml TE (10/1) containing 0.1 mg/ml RnaseA.

Molecular techniques and transformation. General molecular cloning techniques, restriction enzyme analysis and transformation of *E. coli* by $CaCl_2$-induced competence were performed as described by Sambrook *et al.* (28). Plasmid DNA was isolated from *E. coli* and *Lc. lactis* by using Qiagen columns (Basel, Switzerland); 5 mg of lysozyme ml^{-1} were added to buffer P1 to facilitate lysis of lactococci. Competent cells and electroporations of *Lc. lactis* were realized according to the method of Holo and Nes (18). PCR amplification was performed as previously reported (27).

Sequencing of plasmid DNA. The whole plasmids pN42 and pJBL2 were first cloned into pJDC9, then subcloned into pUC19. DNA sequences were determined by the dideoxy-chain termination method (29), using the universal primers or by primer walking. Sequences were assembled and analyzed using the University of Wisconsin Genetics Computer Group (GCG) computer software package (10).

Nucleotide sequence accession numbers. The pN42 and pJBL2 nucleotide sequences have been deposited in the GenBank database under nucleotide accession numbers xxxxx and xxxxx, respectively.

RESULTS

Maxi-preparations of plasmid DNA were performed for thirteen strains of *L. lactis*. Each strain yielded a single plasmid. Two strains, NCC88 and JBL2, were chosen for further study. The plasmids named pN42 and pJBL2 respectively, were subcloned and their DNA sequence determined (Fig. 1A). Analysis and comparison of the sequences revealed that both plasmids

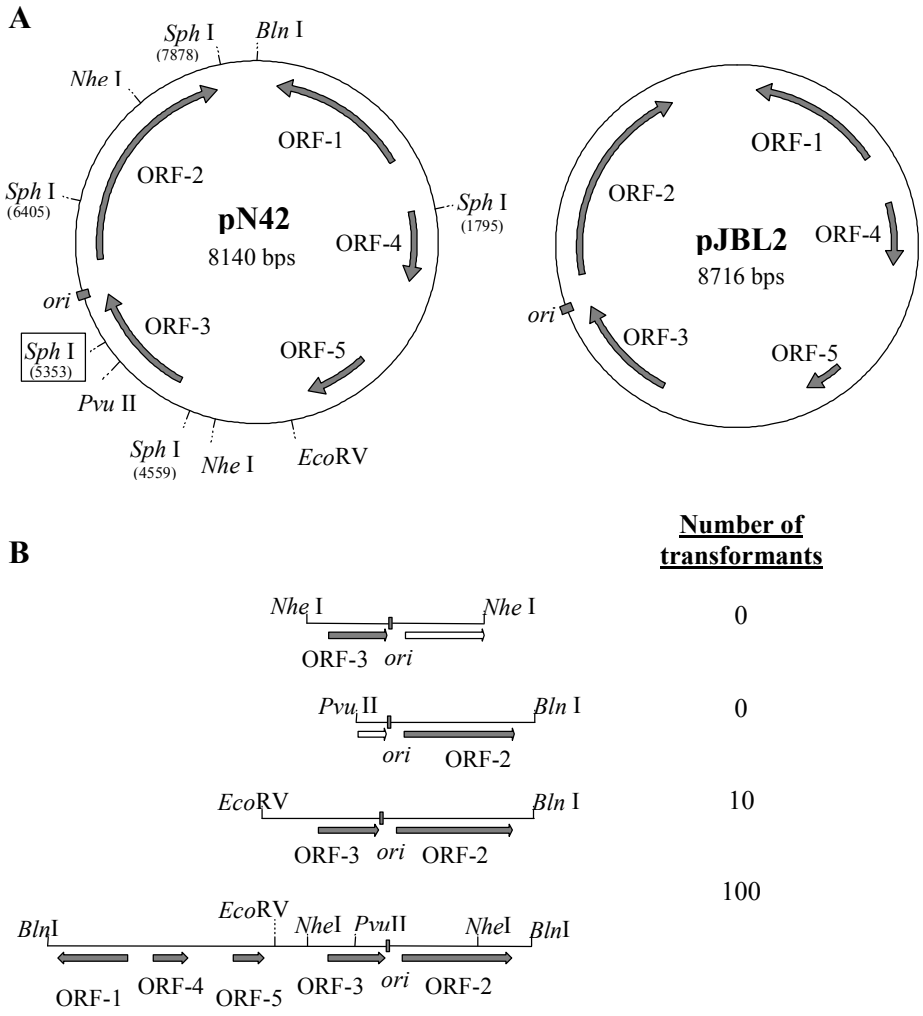


Fig. 1. Diagrams of the pN42 and pJBL2 plasmids genetic organization (A) and (B) schematic representation of the subcloning strategy for the determination of pN42 minimal replication machinery by transformation in the heterologous host *Lc. lactis*. Truncated ORFs are indicated by a clear arrow ⇨. The pN42 *Sph*I site protected against restriction in the host *L. lactis* NCC88 is boxed on pN42 plasmid map.

TABLE 1. Comparative description of pN42 and pJBL2 genetic organization.

	ORF size (in bp)		Identity*	Putative gene function
	pN42	pJBL2		
ORF-1	1179	1188	57.9 %	HsdS subunit
ORF-2	1842	1842	93.3 %	replicative DNA helicase (*dnaB*)
ORF-3	951	969	90.4 %	unknown
ORF-4	582	555	98.2 %	mobilization protein (*mobA*)
ORF-5	522	339	91.4 %	tetracyclin repressor protein (*tetR*) (in the DNA binding domain)

* percentage of identity between the nucleotide sequences of the respective ORFs from pN42 and pJBL2 determined by the GCG computer package sofware

are homologous (86.7% identity at the nucleotide level) and possess the same genetic organization containing five major open reading frames (ORFs) that are highly conserved (Table 1). Blast searches (1) were performed to assign putative functions to these ORFs based on their homologies to the annotated genes present in the GenBank-EMBL databases. As shown in Table 1, putative functions were assigned to four of the five ORFs. ORF-5 from pN42 and pJBL2 show 30% and 34% identity, respectively, to the "helix-turn-helix" DNA binding motif (PROSITE accession number: PDOC00830) of the tetracycline repressor protein (TetR) from *Salmonella ordonez* (GenBank accession number: X65876). ORF-4 from pN42 and pJBL2 show 37% and 35% identity, respectively, to the mobilization protein (MobA) of the *Bacteroides fragilis* pBIl43 plasmid (Genbank: U30316). Both ORF-4 and ORF-5 show a Leucine zipper motif (PROSITE: PDOC00029), which suggests that the proteins may interact, perhaps as a dimer. ORF-2 displays 46% similarity to the replicative DNA helicase (dnaB) from the *Borrelia burgdorferi* genome (GenBank: AE001123). The function of ORF-2 as a replication protein is further substantiated by the presence of an ATP-binding domain (Prosite database: "P-loop"; PDOC00017) characteristic of DNA and RNA helicases (13, 17, 20). ORF-3 does not display any significant homology except to a 1529 bp-long gene from *Synechocystis* sp. PCC6803 (GenBank: D90916) encoding an "hypothetical protein". Potential origins of replication for both plasmids are located between ORF-2 and ORF-3 and characterized by three direct repeats of 21 bp and 20 bp for pN42 and pJBL2, respectively.

The function of the ORF-2-encoded protein as well as the localization of the *ori* site of pN42 was investigated by subcloning pN42 fragments containing the predicted *ori*, ORF-2 and/or ORF-3 into the *E. coli* vector, pJDC9. Since pJDC9 does not replicate in the gram-positive bacterium *Lc. lactis*, transformation and replication is dependent on the addition of functional pN42 replication machinery. The interpretation of these results was limited by the low transformation frequency of pN42 based plasmids in *Lc. lactis*, where we observed a reduction of approx. 4 orders of magnitude in transformation frequency compared to shuttle vectors such as pNZ124 (10^6 transformants/μg DNA). However, though pN42 derived plasmids transformed in *Lc. lactis* were seen to be of low copy number, the constructs could be rescued intact into *E. coli*. As shown in Fig. 1B, both ORF-2 and ORF-3 plus the *ori* region were required to obtain transformants. Better transformation efficiencies were obtained when the whole plasmid was subcloned in pJDC9. This indicates that: (i) *ori* is definitely located in the 5694-5758 bp region containing the direct repeats identified by *in silico* analysis, (ii) though ORF-2 might encode a replicative DNA helicase, the presence of ORF-3 —of unknown function— is required in order to reconstitute a functional replication unit, and (iii) other plasmid elements such as ORF-4 and/or ORF-5 have a positive effect on plasmid transformation and replication.

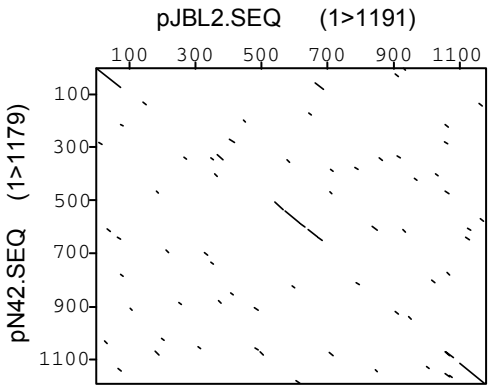


Fig. 2. The structure of ORF-1 is characteristic of type I HsdS subunits. The figure was prepared using the DotPlot program of the DNAStar software with a window of 10 base pairs and a percentage of identity of 85%.

The ORF-1 DNA sequences from plasmids pN42 and pJBL2 display the lowest overall homology (58% identity, Table 1) of the plasmids ORFs. Blast homology searches, however, indicate very high homologies to the *hsdS* (host specificity for DNA, Specificity) subunits of type I restriction-modification systems identified in *Lc. lactis*, e.g. 94% identity to the C-terminus of the *hsdS* gene encoded by the *Lc. lactis* pAH82 plasmid (Genbank: AF228680). The low overall homology observed between different *hsdS* genes derives from their structure. A DOTPLOT representation of the pairwise comparison of ORF-1 from pN42 and pJBL2 displays a structure characteristic of a *hsdS* gene with three conserved domains interspersed by two variable domains (Fig. 2). The conserved domains of ORF-1 from the two plasmids are indeed very conserved: 100% identity for the 70 bp encoding the N-terminal domain, 86% for the 139 bp encoding the central domain and 94% for the 76 bp encoding the C-terminal domain. The variable domains in contrast are much less conserved even at the amino acid level (26.4% identity or less). Homology searches with the variable domains yielded significant homologies to the variable domains, or target recognition domains (TRDs) of *E. coli* HsdS proteins. The C-terminus variable domain of pN42 ORF-1 displays 43% identity to the N-terminal TRD of the *Eco*R124 II HsdS protein (SwissProt accession number: S02167), whereas the C-terminus variable domain of pJBL2 ORF-1 shows 52% identity to the C-terminus TRD of the specificity subunit of *Eco*D restriction enzyme (GenBank: V00287). These results corroborate the identification of ORF-1 as an *hsdS* gene.

Restriction analysis of pN42 plasmid DNA extracted from its host NCC88 revealed the presence of a single *Sph*I site leading to the linearization of the plasmid. Molecular cloning and restriction enzyme analysis indicated that the plasmid was cut at position 4563. *In silico* analysis of the complete plasmid sequence revealed the presence of five *Sph*I restriction sites in pN42. Four of these sites, possessing a guanosine residue directly upstream of the recognition site (Fig. 3), and situated at positions 1799, 5357, 6409 and 7882, remained uncut

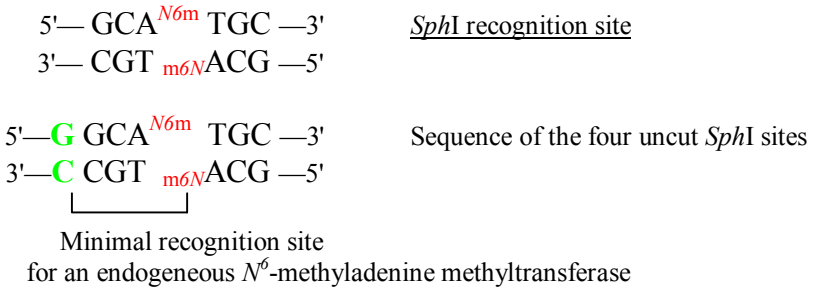


Fig. 3. Minimal recognition site for *L. lactis* NCC88 endogenous N^6-methyladenine methyltransferase.

in plasmid isolated from NCC88. Additionally, cloning of pN42 into pJDC9 and propagation in either *E. coli* or *Lc. lactis* produced plasmid DNA that was digested by *Sph*I at all five sites demonstrating that in the parent strain, the four uncut *Sph*I sites must have been modified against restriction. This indicates that *L. lactis* NCC88 methylates its DNA on the adenosine residue of the minimal conserved sequence $GGCA^{N6m}$ (Fig. 3) and suggests the presence of an endogenous N^6-methyladenine-methyltransferase which may be part of a host restriction-modification system.

DISCUSSION

In this paper, we have confirmed that the presence of plasmids in the *lactis* subspecies of *L. delbrueckii* is not such a rare occurrence (4) as in the closely related *bulgaricus* subspecies (8) for which actual reports are limited to the description of pBL10, a small cryptic plasmid of 2700 bp (5). *Lc. lactis* in contrast is known to contain easily 2 to 11 different plasmids per cell, ranging in size from 3 to > 130 kb (8). Those plasmids may encode a wide variety of metabolic functions, such as the lactose operon (9, 21), a proteolytic function (19, 33), or exopolysaccharide production (34).

The DNA sequence analysis and characterization of the pN42 and pJBL2 plasmids from *L. lactis* strains NCC88 and JBL2 respectively, showed that both plasmids had the same genetic organization and contained four major, highly conserved genes, ORF-2 to ORF-5 (90% identity or higher), and a single gene ORF-1, with a lower degree of identity (58%). Preliminary genetic analysis of the replication machinery of plasmid pN42 has revealed that although ORF-2, ORF-3 and *ori* constitute the minimal pN42 replication machinery, other plasmid encoded elements are required for efficient transformation. One possible mechanism could parallel the replication process of *Lc. lactis* single-stranded DNA plasmids (15, 32), where ORF-4 and ORF-5 might associate via the common leucine zippers to combine the DNA binding function of ORF-5 with the ability of ORF-4 to process plasmid DNA for the

synthesis of the lagging strand. Therefore the lack of ORF-4 and ORF-5 would lead to an accumulation of single-stranded DNA in the host, with a negative influence on plasmid stability. These transformation results, which were confirmed by transformation in *Lactobacillus johnsonii* (data not shown), were unfortunately limited not only by the lack of a transformation system for *L. delbruekii* subsp. *lactis* (the original host) or *bulgaricus*, but also by the lack of a suitable heterologous host for the replication analysis. pN42 is capable of replication in the heterologous hosts *L. johnsonii* and *Lc. lactis*, but the transformation frequency is very low when compared to more commonly used vectors such as pNZ124. These *L. lactis* plasmids therefore constitute the basis for the construction of shuttle and expression vectors for the genetic modification of these industrially important bacteria and future product development.

Despite a lower overall homology, ORF-1 of both plasmids encode an HsdS subunit. The HsdS protein is the subunit of type I R-M systems responsible for the recognition and binding of the enzyme to a specific target site split into two recognition components (2). Recognition of this site is facilitated by the structure of the specificity subunit (HsdS) possessing two target recognition domains (TRDs) so that each recognize one component of the target site (11, 14, 22). The low homology observed between the TRDs of the plasmid-derived HsdS subunits reflects their different target site specificity. Type I R-M systems are composed of three different subunits encoded by the genes *hsdS*, *hsdM*, and *hsdR*. In contrast to *Lc. lactis* plasmids that may encode complete *hsd* clusters (24, 30) the plasmids pN42 and pJBL2 encode only the HsdS subunits, suggesting that a complete *hsd* cluster might —or at least the *hsdM* and *hsdR* genes must— be present on the *L. lactis* chromosome. The presence of *hsdS*-encoding plasmids in a host would thereby provide new specificity subunits that could expand the range of restriction site specificities and increase biological diversity as was proposed by Schouler *et al.* in *Lc. lactis* (31).

Unlike some *Lc. lactis* plasmids, *L. lactis* plasmids do not encode phage resistance systems (16), nor do they encode complete restriction-modification systems. However, the presence of HsdS subunits on pN42 and pJBL2, combined with the demonstration of an efficient methylation system in *L. lactis* NCC88 does provide evidence of the presence of endogenous restriction-modification systems in *L. delbruekii* subsp. *lactis* and may at least in part explain the persistent lack of transformation systems for this industrially important species.

ACKNOWLEGMENT

This work was supported by the Swiss National Foundation within SPP Food Biotechnology Program no. 5002-044544. We are grateful to A.-C. Pittet for her technical assistance and transformations of *L. johnsonii*.

REFERENCES

1. **Altschul, S. F., T. L. Madden, A. A. Schäffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman.** 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. **25:**3389–3402.
2. **Bickle, T. A., and D. H. Krüger.** 1993. Biology of DNA restriction. Microbiol. Rev. **57:**434–450.
3. **Bruttin, A., F. Desiere, N. d'Amico, J. P. Guerin, J. Sidoti, B. Huni, S. Lucchini, and H. Brüssow.** 1997. Molecular ecology of *Streptococcus thermophilus* bacteriophages infection in a cheese factory. Appl. Environ. Microbiol. **63:**3144–3150.
4. **Casey, M. G., and J. Jimeno.** 1989. *Lactobacillus delbrueckii* subsp. *lactis* plasmids. Netherlands Milk Dairy J. **33:**279–286.
5. **Chagnaud, P., C. K. N. Chan Kwo Chion, R. Duran, P. Naouri, A. Arnaud, and P. Galzi.** 1992. Construction of a new shuttle vector for *Lactobacillus*. Can. J. Microbiol. **38:**69–74.
6. **Chen, J.-D., and D. A. Morrison.** 1988. Construction and properties of a new insertion vector, pJDC9, that is protected by transcriptional terminators and useful for cloning of DNA from *Streptococcus pneumoniae*. Gene **64:**155–164.
7. **Daly, C., Fitzgerald, G. F., and R. Davis.** 1996. Biotechnology of lactic acid bacteria with special reference to bacteriophage resistance. Antonie van Leeuwenhoek **70:**99–110.
8. **Davidson, B. E., N. Kordias, M. Dobos, and A. J. Hillier.** 1996. Genomic organization of lactic acid bacteria. Antonie van Leeuwenhoek **70:**161–183.
9. **Davidson, B. E., R. M. Llanos, M. R. Cancilla, N. C. Redman, and A. J. Hillier.** 1995. Current research on the genetics of lactic acid production in lactic acid bacteria. Int. Dairy J. **5:**763–784.
10. **Devereux, J., P. Haeberli, and E. Sharpe.** 1984. A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. **12:**387–395.
11. **Fuller-Pace, F. V., L. R. Bullas, H. Delius, and N. E. Murray.** 1984. Genetic recombination can generate altered restriction specificity. Proc. Natl. Acad. Sci. USA **81:**6095–6099.

12. **Gasson, M.J.** 1983. Plasmid complements of *Streptococcus lactis* NCDO712 and other streptococci after protoplast curing. J. Bacteriol. **154:**1–9.
13. **Gorbalenya, A. E., E. V. Koonin, A. P. Donchenko, and V. M. Blinov.** 1989. Two related superfamilies of putative helicases involved in replication, recombination, repair and expression of DNA and RNA genomes. Nucleic Acids Res. **17:**4713–4730.
14. **Gough, J. A., and N. E. Murray.** 1983. Sequence diversity among related genes for recognition of specific targets in DNA molecules. J. Mol. Biol. **166:**1–19.
15. **Gruss, A., and S. D. Ehrlich.** 1989. The family of highly interrelated single-stranded deoxyribonucleic acid plasmids. Microbiol. Rev. **53:**231–241.
16. **Hill, C.** 1993. Bacteriophage and bacteriophage resistance in lactic acid bacteria. FEMS Microbiol. Rev. **12:**87–108.
17. **Hodgman, T. C.** 1988. A new superfamily of replicative proteins. Nature **333:**22–23 and Nature **333:**578 (Errata).
18. **Holo, M. Y., and I. F. Nes.** 1989. High frequency transformation, by electroporation, of *Lactococcus lactis* subsp. *cremoris* grown with glycine in osmotically stabilized media. Appl. Environ. Microbiol. **55:**3119–3123.
19. **Law, J., P. Vos, F. Hayes, C. Daly, W. M. de Vos, and G. F. Fitzgerald.** 1992. Cloning and partial sequencing of the proteinase gene complex from *Lactococcus lactis* subsp. *lactis* UC317. J. Gen. Microbiol. **138:**709–718.
20. **Linder, P., P. Lasko, M. Ashburner, P. Leroy, P. J. Nielsen, K. Nishi, J. Schnier, and P. P. Slonimski.** 1989. Birth of the D-E-A-D box. Nature **337:**121–122.
21. **McKay, L. L.** 1983. Functional properties of plasmids in lactic streptococci. Antonie van Leeuwenhoek **49:**259–274.
22. **Nagaraja, V., J. C. W. Shepherd, and T. A. Bickle.** 1985. The evolution of DNA sequence specificity: a recombinant restriction enzyme has a hybrid recognition sequence. Nature **316:**371–372.
23. **O'Sullivan, D. J., and T. R. Klaenhammer.** 1993. Rapid mini-prep isolation of high-quality plasmid DNA from *Lactococcus lactis* and *Lactobacillus* spp. Appl. Environ. Microbiol. **59:**2730–2733.
24. **O'Sullivan, D., D. P. Twomey, A. Coffey, C. Hill, G. F. Fitzgerald, and R. P. Ross.** (2000). Novel type I restriction specificities through domain shuffling of HsdS subunits in *Lactococcus lactis*. Mol. Microbiol. **36:**866–875.
25. **Peitersen, N.** 1991. Practical phage control. Bull. Int. Dairy. Fed. **263:**1–43.
26. **Platteeuw, C., G. Simmons, and W. M. de Vos.** 1994. Use of the *Escherichia coli* β-glucuronidase (*gusA*) gene as a reporter gene for analyzing promoters in lactic acid bacteria. Appl. Environ. Microbiol. **60:**587–593.

27. **Saiki, R. K., D. H. Gelfand, S. Stoffel, S. J. Scharf, R. Higuchi, G. T. Horn, K. B. Mullis, and H. A. Ehlich.** 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science **239:**487–491.
28. **Sambrook, J., F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbour Laboratory, Cold Spring Harbour, N.Y.
29. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74:**5463–5467.
30. **Schouler, C., F. Clier, A. L. Lerayer, S. Dusko Ehrlich, and M.-C. Chopin.** 1998. A type IC restriction-modification system in *Lactococcus lactis.* J. Bacteriol. **180:**407–411.
31. **Schouler, C., M. Gautier, S. Dusko Ehrlich, and M.-C. Chopin.** 1998. Combinational variation of restriction modification specificities in *Lactococcus lactis*. Mol. Microbiol. **28:**169–178.
32. **Seegers, J. F. M. L., A. C. Zhao, W. J. J. Meijer, S. A. Khan, G. Venema, and S. Bron.** 1995. Structural and functional analysis of the single-strand origin of replication from the lactococcal plasmid pWV01. Mol. Gen. Genet. **249:**43–50.
33. **Tan, P. S. T., B. Poolman, and W. N. Konings.** 1993. Proteolytic enzymes of *Lactococcus lactis*. J. Dairy Res. **60:**269–286.
34. **van Kranenburg, R., H. R. Vos, I. I. van Swam, M. Kleerebezem, and W. M. de Vos.** 1997. Molecular characterization of the plasmid-encoded *eps* gene cluster essential for exopolysaccharide production in *Lactococcus lactis.* Mol. Microbiol. **24:**387–397.
35. **Yanisch-Perron, C., J. Vieira, and J. Messing.** 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene **33:**103–119.

CHAPTER VI

Lactobacillus delbrueckii subsp. *lactis* type I restriction-modification system. Evidence for horizontal transfer and exchange of *hsdS* target recognition domains.

Aude A. Bourniquel, R. David Pridmore and Beat Mollet*
Nestlé Research Centre, Nestec Ltd., Vers-chez-les-Blanc, P.O. Box 44,
CH-1000 Lausanne 26, Switzerland

Summary

Purification of *Lactobacillus delbrueckii* subsp. *lactis* NCC88 cell extracts yielded a semi-purified enzyme fraction displaying activity characteristic of a type I restriction enzyme. Type I *hsd* clusters were isolated from the chromosomal DNA of two *L. lactis* strains, NCC88 and NCC82. They encode genes that are highly conserved (>97% identity) except for the *hsdS* genes that have different specificities. The genes *hsdR*, *hsdM* and *hsdS* constitute what we referred to as the '*hsd* cluster proper' since they are the units constituting type I *hsd* clusters isolated and characterised in *E. coli*. *L. lactis hsd* clusters comprise a second part located downstream of the *hsd* cluster proper encoding an integrase (*int*) gene as well as a second N-terminally truncated *hsdS* gene. NCC88 also encodes a third *hsdS* gene, which might be transcribed independently. Comparison of the *hsdS* genes belonging to the clusters with *hsdS* genes isolated from four other *L. delbrueckii* subsp. strains revealed circumstantial evidence of (i) horizontal transfer of plasmid-borne *hsdS* genes in-between different *L. lactis* strains, (ii) recombination of the target recognition domain-encoding sequences between *hsdS* genes.

Introduction

Restriction-modification (R-M) systems are bacterial defense systems that protect the cell against bacteriophage infections. They have the dual function of: (i) protecting the host DNA against restriction by methylating the DNA within specific target sites, and (ii) 'restricting' i.e. degrading any unmodified fragment of DNA that may enter the cell. R-M systems may be of three types denoted type I, type II and type III. Very few type I restriction enzymes have been identified compared to the numerous type II endonucleases, probably because they are

the most complex of the three types of R-M systems. Type I R-M systems are encoded by the *hsd* cluster composed of the genes, *hsdR*, *hsdM* and *hsdS*, organized in two transcriptional units, the *R* regulon and the *MS* operon. These two transcriptional units are contiguous on the DNA but may be found in different positions relative to each other depending on the type I family to which they belong. Based on comparisons between the enzymes isolated from either *E. coli* or the enterobacteriacea, *Salmonella*, type I R-M systems have been distributed in three families, IA, IB and IC. R-M systems belonging to the type IA and IB families are encoded on the bacterial chromosome in the order *hsdR*, *hsdM* and *hsdS*, whereas the genes of type IC systems are encoded on plasmids and the order of the two operons are reversed so that the genes *hsdM* and *hsdS* are followed by *hsdR* (Bickle and Krüger, 1993).

Type I R-M systems are hetero-oligomeric enzymes constituted of three subunits HsdR, HsdM and HsdS encoded by the genes described above. Type I restriction enzymes are present in two different forms *in vivo*. The M_2S form catalyses methylation of the DNA in presence of the cofactor S-adenosyl-methionine (SAM), whereas the multimer R_2M_2S (or possibly R_1M_2S, Janscak *et al.*, 1996) in the presence of the three cofactors SAM, Mg^{2+} and ATP is required for DNA restriction. HsdS subunits are responsible for the recognition and the binding of the restriction enzymes to its specific target site. Type I recognition sites are asymmetrical sequences split into two components separated by 6-8 unspecified nucleotides. The S subunit characteristic structure composed of highly conserved sequences separated by two regions of non-homology, the so-called 'variable' or target recognition domains (TRDs) reflects the polypeptide specific function since the amino-terminal TRD specifies the 5' component of the recognition sequence whereas the carboxy-terminal TRD specifies the 3' component (Fuller-Pace and Murray, 1986). Consequently, the specificity of type I R-M systems is dependent on HsdS subunits and recombination events between *hsdS* genes inducing an exchange of the variable regions generate HsdS and thus type I R-M enzymes with novel specificities (Fuller-Pace, 1984; O'Sullivan *et al*., 2000).

In recent years, many projects have been aimed at the study of R-M systems in lactic acid bacteria and their potential industrial application as protection against bacteriophage infections. Lactococci were found to possess all three types of R-M systems, while only type II endonucleases have been discovered in dairy streptococci. In 1998, Auad *et al.* published a report establishing the presence of a R-M system in *L. lactis* CNRZ 326 using the classical phage assay. The plasmids pN42 and pJBL2 isolated from *L. lactis* NCC88 and JBL2 respectively, each encode one *hsdS* gene suggesting the presence of a complete *hsd* cluster on the chromosomal DNA (chapter V). In this work, we demonstrate the presence of a type I R-M system in *L. lactis* NCC88 and analyze the sequence of *hsd* clusters isolated from *L. lactis* strains NCC82 and NCC88. We present circumstantial evidence of a mechanism combining

horizontal transfer of plasmid-borne *hsdS* genes and genetic recombination of the variable regions for the production of *hsdS* genes with novel specificities.

Experimental procedures

Bacterial strains and growth conditions

Lactobacillus delbrueckii subsp. *lactis* strains NCC39, NCC73, NCC82, NCC88, NCC627, and *Lactobacillus delbrueckii* subsp. *bulgaricus* strain NCC786 were obtained form the Nestlé Culture Collection. *L. delbrueckii* ssp. was routinely grown in MRS broth (Difco, Detroit, MI) supplemented with 2% glucose at 45°C without shaking. For plasmid isolation and protein purification purpose, *L. lactis* was grown in MRS broth supplemented with 2% glucose and 17% sucrose (MRS-S broth). *Escherichia coli* XL1-Blue was obtained from Stratagene (La Jolla, CA) and grown in Luria-Bertani broth with aeration at 37°C. Ampicillin (Ap) was added at at 100 µg/ml to the growth medium. Functional β-galactosidase was detected by the addition of 80 µg of 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-gal) ml^{-1} and 0.5 mM of isopropyl-β-D-thiogalactopyranoside (IPTG) to Luria-Bertani solid medium.

Plasmids

The plasmid pN42 was extracted from *L. lactis* NCC88 using the protocol described by O'Sullivan and Klaenhammer (1993) modified as indicated below. pN42 was linearised by digestion with *Bln*I and subcloned into the *E. coli* vector pJDC9 (Chen and Morrison, 1988) thus forming the plasmid pN42-sub1. The *E. coli* vector pGEM®-T (Promega, Madison, WI) was used to clone PCR products for sequencing.

Maxipreparation of plasmid DNA from L. lactis cultures

This protocol was adapted from the protocol published by O'Sullivan and Klaenhammer (1993). 100 ml MRS broth containing 2% glucose and 17% sucrose (MRS-S broth) were inoculated with a *L. lactis* preculture in MRS-S broth and incubated overnight at 45°C without aeration. The cells were harvested, washed twice with 20 ml sterile distilled water, suspended in 10 ml PB (Tris-HCl, 50 mM, pH 7.4; EDTA, 50 mM, pH 8.0; sucrose 25%) containing 1 mg/ml lysozyme, and incubated at 37°C for 2 hours. 2.5 mg/ml mutanolysin were added and the incubation continued till most of the cells appeared as protoplasts when observed by light

microscopy. The protoplasts were harvested, washed with 20 ml PB then lysed by suspension in 4 ml TER (50 mM Tris.Cl, pH 7.4; 50 mM EDTA, pH 8.0; 0.5 mg/ml RnaseA) and incubated at 37°C for 15 min. Eight milliliters of fresh alkaline SDS solution (3% SDS, 0.2 N NaOH) were added and immediately mixed by inverting the tube several times. After 7 min incubation at room temperature, 6 ml of ice-cold sodium acetate (3 M, pH 4.8) were added, mixed till the solution was clear, then incubated on ice for 10-20 min and centrifuged at 5'000 rpm for 35 min at 4°C. The supernatant was transferred to a 50 ml-tube containing 13 ml isopropanol, well mixed and centrifuged at 5'000 rpm for 35 min at 4°C. All liquid was carefully removed before suspending the DNA pellet in 6.4 ml distilled water, 3.8 ml ammonium acetate 8 M, 0.2 ml ethidium bromide (10 mg/ml) and 7 ml phenol-chloroform. The mix was centrifuged at 5'000 rpm for 15 min at 4°C. The upper phase was transferred to a new tube, and then chloroform extractions were performed till all traces of ethidium bromide had been removed. 10 ml cold ethanol were then added and centrifuged at 5'000 rpm for 35 min at 4°C. The pellet was washed with 70% ethanol, air dried, and finally resuspended in 2 ml TE (10/1) containing 0.1 mg/ml RnaseA.

Preparation and partial purification of protein extracts from L. lactis NCC88.

L. lactis NCC88 was inoculated in MRS-S broth and incubated for 16h at 45°C without aeration. The cells were harvested by centrifugation, washed twice with deionised water and resuspended in Protoplasting Buffer (PB: 50mM Tris-HCl, pH 7.4; 50 mM EDTA, pH 8.0; 25% sucrose). Successive incubations at 37°C in PB with lysozyme (1 mg/ml) then mutanolysin (170 U/ml) led to the formation of protoplasts which were harvested, suspended in lysis buffer (20 mM Tris-HCl, pH 7.4; 50 mM KCl; 1 mM EDTA; 14 mM β-mercaptoethanol) then frozen at –20°C to ensure total lysis. The frozen mix was thawed at 37°C and clarified by low speed centrifugation (16,000 × *g*, 30 min) followed by ultracentrifugation (100,000 × *g*, 2 h). Nucleic acids were removed by precipitation with 0.4% polyethyleneimine in the presence of 0.2 M NaCl. Proteins were concentrated by precipitation with ammonium sulfate (80% saturation), dialysed overnight against buffer A (20 mM Tris-HCl, pH 7.4; 50 mM KCl; 10 mM $MgCl_2$; 0.1 mM EDTA; 14 mM β-mercaptoethanol) then loaded on a heparin affinity column (5 ml-Econo-Pac column, Bio-Rad, Glattsburg, Switzerland) and eluted with a linear KCl gradient. The active fractions eluting around 80 mM KCl were pooled, concentrated and fractionated on a Superose 6 gel filtration column (Pharmacia, Uppsala, Sweden).

Endonuclease activity assay

Endonuclease activity assays were performed in 20 µl of a buffer containing 50 mM Tris-HCl (pH 8.0), 1 mM DTT, 10 mM $MgCl_2$, and 50 mM NaCl. Unless otherwise indicated, the cofactors SAM and ATP were added at concentrations of 0.2 mM and 2 mM respectively, as well as 0.5 µg of the substrate, N^6-methyladenine-free λ DNA (New England Biolabs, Beverly, MA). Reactions were started by addition of 2 µl of crude extract or semi-purified fraction and incubated for 30 minutes at 45°C, the optimal growth temperature of the bacterium. The reactions were stopped by the addition of gel loading dye, then each mixture was applied to an 0.8% agarose gel for electrophoresis.

Molecular techniques, transformation and sequencing

General molecular techniques, restriction enzyme analysis, and transformation of *E. coli* by $CaCl_2$-induced competence were performed as described by Sambrook *et al.* (1989). Plasmid DNA was isolated from *E. coli* using CONCERT™ columns (Life Technologies, Basel, Switzerland). *L. lactis* genomic DNA was obtained as described by Delley *et al.* (1990), except that the incubation step with proteinase K and pronase E was replaced by successive incubations at 37°C with lysozyme (10 mg/ml) and mutanolysin (100 µg/ml). PCR amplification was performed as previouly reported (Saiki *et al.*, 1988). Ligation of the PCR products into pGEM®-T was carried out with the kit as recommended by the manufacturer (Promega, Madison, WI). DNA sequencing was performed by the dideoxy chain termination method (Sanger *et al.*, 1977) with the universal primers or by primer walking. Custom-made primers were purchased from Microsynth (Balgach, Switzerland). The sequence data were assembled and analyzed using the Lasergene99 software package (DNAStar Inc., Madison, WI) and the Wisconsin package, version 10 (Genetics Computer group [GCG], Madison, WI).

Isolation of the hsd clusters by inverted PCR

Specific primers for the PCR amplification of *hsdS* genes from lactic acid bacteria, 5'–AATTAAGRTTTRAWGGTTTYRC–3' and 5'–CAAMAAATCYAACTTACG–3', were designed from conserved regions revealed by multiple alignment of *hsdS* sequences from *Lc. lactis* strains (GenBank accession numbers: AF034786, U90222 and AF013165) and the *L. lactis* plasmids pN42 and pJBL2 (Appendix III and IV, respectively). The PCR fragments were cloned into pGEM®-T, sequenced and used as the starting point for the isolation of the *hsd* clusters by inverted PCR. Genomic DNA (4 µg) was digested separately with the

restriction enzymes *Bam*HI, *Eco*RI, *Hin*dIII, *Pst*I and *Xma*CI, phenol-chloroform extracted, ethanol precipitated and ligated in a 200 µl volume o/n at 4°C. The ligation products were phenol-chloroform extracted, ethanol precipitated and suspended in a final volume of 20 µl TE (10/1). 1 µl of the ligated DNA was used as template for amplification with the Expand™ Long Template PCR System (Roche Diagnostics Ltd, Rotkreuz, Switzerland). The template was denatured for 5 min at 95°C, followed by 30 cycles of 30 sec. at 95°C, 30 sec. at 50°C and 68°C for 10 min with 10 sec.–increments per cycle. Reactions were concluded by an incubation of 7 min at 68°C.

Results

Type I R-M system activity in L. lactis *NCC88.*

λ DNA incubated with NCC88 cell extract formed neither a smear of fragments characteristic of type I or type III restriction enzymes nor a pattern typical of type II enzymes, but was totally degraded, independently of the presence or absence of ATP and SAM. This degradation activity is the result of the activity of nucleases able to aspecifically degrade genomic DNA.

Purification procedures were implemented to separate the aspecific nucleases from the restriction enzymes potentially present in the cell extract. Crude protein extract was prepared from *L. lactis* NCC88 and separated on a heparin affinity column. Elution fractions were tested for endonuclease activity. As depicted in Fig. 1A, cleavage λ DNA by fractions 11 to 14 resulted in a smear. Incubation of the same fractions with plasmid DNA (Fig.1B)

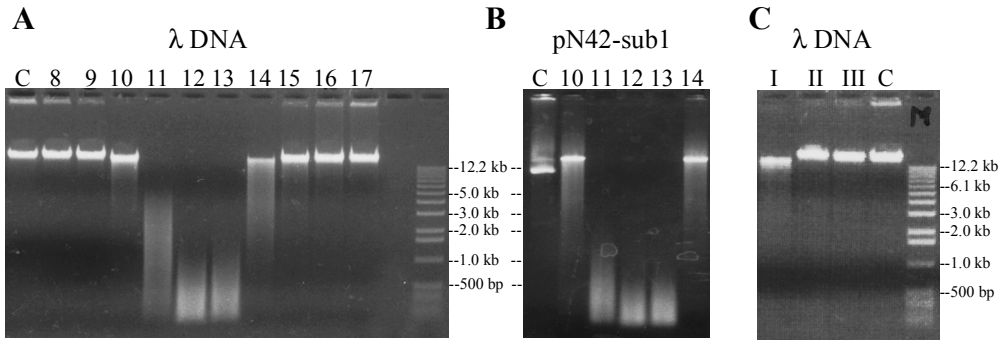


Fig. 1. Agarose gel electrophoresis of DNA digested with the partially purified enzyme preparations from *L. lactis* NCC88. (A) Screening for restriction activity in the elution fractions (8 to 17) from the heparin affinity column in the presence of the cofactors Mg^{2+}, ATP and SAM. C is the control λ DNA. (B) Testing the active fractions for endonuclease activity on plasmid DNA propagated in r^-m^- *E. coli*, in the presence of the cofactors Mg^{2+}, ATP and SAM. C is the control, undigested pN42-sub1. (C) Cofactors requirements of the partially purified fraction from the gel filtration column. Lane I, Mg2+ plus ATP plus SAM; Lane II, Mg2+ plus SAM; Lane III, Mg2+ plus ATP. C is the control λ DNA.

confirmed that this cleavage pattern was the result of endonuclease activity alone, eliminatingthe possibility of nuclease contamination. In an attempt to further purify the active enzyme, the four active fractions were pooled, concentrated and fractionated on a gel filtration column. Enzyme activity eluted in a single fraction (MW=200-280 kDa) but no peak corresponding to this activity could be observed on the elution profile, nor could the enzyme be detected on SDS-polyacryamide gels. Comparison of λ DNA degradation in Fig.1B and Fig.1C revealed a significant loss of endonuclease activity. The enzyme is very labile and could not be purified to homogeneity.

Characterisation of the partially purified enzyme (Fig.1C) revealed that cleavage and smearing required the presence of Mg^{2+} and ATP and were strictly dependent on the presence of SAM. Those properties are typical of type I restriction enzymes.

L. lactis strains NCC82 and NCC88 encode type I R-M system gene clusters

hsd gene clusters were isolated from the chromosomal DNA of two *L. lactis* strains, NCC88 and NCC82 (Appendix V and VI) and encode, respectively, five and six related open reading frames or ORFs (Fig.2).

The first ORF (3018 bp) of the clusters encodes a 1006 amino acid-protein with a predicted MW of 116 kDa and is highly conserved between the two *L. lactis* strains (98.5% identity). Blast homology searches (Altschul *et al.*, 1997) revealed 33% identity to the HsdR subunits of putative type I R-M systems encoded on plasmids pER35 and pND861 of the lactic acid bacteria *S. thermophilus* and *Lc. lactis* (Genbank accession numbers: AF177167 and AF034786, respectively). Homologies (32% identity) were also found to the HsdR subunit of the purified and extensively characterized (Price *et al.*, 1987; Price *et al.*, 1989; Taylor *et al.*,

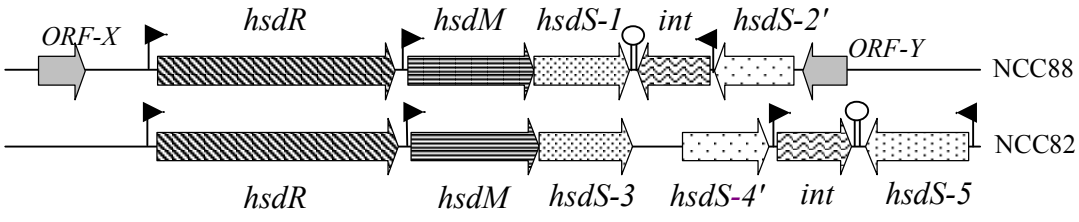


Fig. 2 Diagrams of the *hsd* clusters from *L. lactis* NCC88 and NCC82. The two ORFs indicated with a prime sign, *hsdS-2'* and *hsdS-4'* are truncated at the 5' end and lack a functional initiation codon.The function of the grey ORFs *ORF-X* and *ORF-Y* is unknown.

Table 1. Seven conserved helicase motifs identified in the *L. lactis* NCC88 and NCC82 HsdR subunits.

R-M subunits	Helicase motifs*		
	I	IA	II
NCC88 HsdR	SGYIWHT**TG**S**GKT**LTSYKVARN	IDKSIFLIDRKDLDTQT	LAFIVDECHR
NCC82 HsdR	SGYIWHT**TG**S**GKT**LTSYKVARN	IDKSIFLIDRKDLDTQT	LAFIVDECHR
*Eco*R124 HsdR	gGYIWHT**TG**S**GKT**LTSFkAARl	IDKvfFvvDRKDLDyQT	vvFIfDECHR
Consensus	**+ + tg GKT**	**+ ++++ +**	**+++++DEcHr**

	III	IV	V	VI
NCC88 HsdR	WYGF**TGT**PI	VLG**F**SI**D**NNQ	LVIV**V**NRLLT**G**F**D**SPSLSTLYI-**D**	**Q**AF**SR**TN**R**
NCC82 HsdR	WYGF**TGT**PI	VLG**F**SI**D**NNQ	LVIV**V**NRLLT**G**F**D**SPSLSTLYI-**D**	**Q**AF**SR**TN**R**
*Eco*R124 HsdR	qfGF**TGT**PI	VLk**F**KV**D**YND	LlIV**V**gmfLT**G**F**D**aPtLnTLfv-**D**	**Q**AF**SR**TN**R**
Consensus	**+ ++TGT**	**+++f d**	**++++t + g+ + +**	**q +GR R**

* Boldface type indicates amino acids which are either completely conserved or replaced by similar amino acids. Plus signs represent hydrophobic acid residues. The consensus residues indicated were determined by Gorbalenya *et al.* (1989) and Gorbalenya and Koonin (1991).

1992; Dreier *et al.*, 1996) type I enzyme encoded on the *E. coli* plasmid R124/3 (GenBank: X13145). HsdR subunit sequences are generally not conserved, except for the presence of seven motifs characterising the 'DEAD' family subdivision of the helicase superfamily II (Gorbalenya and Koonin, 1989; Gorbalenya *et al.*, 1991). Analysis of the NCC82 and NCC88 protein sequences revealed the presence of these seven domains (Table 1) confirming the identification of this first ORF of the cluster as a *hsdR* gene.

The second ORF (1596 bp) from NCC82 and NCC88 encode a 532 aa-long protein with a calculated MW of 60.2 kDa and 97.6% identity to each other down to the nucleotide level. The protein sequences displayed 53% identity to the HsdM subunit of *S. thermophilus* and *Lc. lactis* putative type I enzymes and 35% identity to the M subunit of *Eco*R124/3. The presence, at position 296 of the protein sequence, of the motif 'V-V-m-N-P-P-Y' characteristic of N^6-adenine-specific DNA methylase (Loenen *et al.*, 1987; PROSITE accession number: PDOC00087) led us to identify this second ORF as a *hsdM* gene.

The third ORF present on the both clusters encodes, respectively for NCC88 and NCC82, 387 aa– and 401 aa–long proteins (predicted MW 45.3 and 44.6 kDa, respectively) which both display structures characteristic of HsdS subunits. The low overall homology (36.4% identity) observed between the two proteins derives from this typical S subunit structure composed of three short, highly conserved domains separated by two variable domains (Kannan *et al.*, 1989). Homology observed between the conserved domains of HsdS-1 and HsdS-3 is high: 100% identity for the 21 aa–N-terminal domain, 85% for the 46 aa–central domain and 64% positive identities for the 38 aa–C-terminal domain. The variable domains in contrast are much less conserved (25% identity or less). Blast homology searches in the GenEMBL database revealed 27% identity between the C-terminus variable domain of HsdS-

1 and the N-terminal TRDs of *Salmonella potsdam* and *E. coli* K12 HsdS proteins (GenBank: M14984 and V00288, respectively) corroborating the identification of the third ORFs of the *L. lactis hsd* clusters as *hsd*S genes.

L. lactis strains NCC82 and NCC88 thus both encode classical type I restriction-modification *hsd* clusters constituted of three genes *hsdR*, *hsdM* and *hsdS*, adjacent on the DNA.

Genetic organization of L. lactis hsd clusters

Examination of the DNA sequences for transcriptional and translational regulatory sequences revealed putative promoter regions upstream of the genes *hsdR* and *hsdM* and very distinctive ribosome binding sites (RBS) upstream of all three genes (Table 2). The *hsd* cluster encoded on the chromosomal DNA of *L. lactis* strains is organized in two operons allowing the transcription of *hsdR* alone and of *hsdM* and *hsdS* as a single regulon. As the stop codon of *hsdM* overlaps the start codon of *hsdS* translation of the *hsdMS* mRNA may be coupled (Oppenheim and Yanofsky, 1980).

In contrast to *E. coli*, *L. lactis hsd* clusters may encode two (NCC88) or three (NCC82) additional related ORFs (Fig. 2) located downstream of the *hsd* cluster proper. The DNA sequences annotated *hsdS-2'*, *hsdS-4'* and *hsdS-5* display structures characteristic of *hsdS* genes (Fig. 3) and were therefore identified as such. *hsdS-5* is a complete and fully functional *hsdS* gene. As it is orientated in the opposite direction of the cluster and especially the *hsdMS* operon, it cannot be considered as being part of the *hsd* cluster proper but putative promoter

Table 2. Putative promoter regions and RBS from the *L. lactis hsd* clusters.

Genes	Sequence			Spacing
	-35	-10	RBS	-35 to -10
NCC88 *hsdR*	TTCAGT	TATAAT	GTTT **AGGGAAGGGGAAAG** CT **ATG**	23
NCC82 *hsdR*	TTCAGT	TATAAT	GTTT **AGGGAAGAGGAAAG** CT **ATG**	23
NCC88*hsdM*	TTGAAT	TTAAAT	GCAT **AGAAGGAAGAAGAA** T **ATG**	18
NCC82 *hsdM*	TTGAAT	TTAAAT	GCAT **AGAAGGAAGAAGAA** T **ATG**	18
NCC82 *hsdS-5*	TTGAAA	TTAATG	CCCC **GCTGGA** CT **ATG**	8
NCC88 *int*	TTGTTA	TAAAAT	TGCT **AAGGAGA** TTTTT **ATG**	20
NCC82 *int*	TTCATC	TAAAAT	TGCT **AAGGAGA** TTTTT **ATG**	5
NCC88 *hsdS-1*			TTTT **GAGGAGGAAA** TCCGCTA **ATG**	
NCC82 *hsdS-3*			TTTT **GAGGAGGAAA** TCCGCTA **ATG**	
Gram-positive consensus sequence	TTGAcA	TATAAT		17

regions and a potential terminator loop were identified upstream and downstream of the *hsdS-5* gene, respectively, suggesting that it might be transcribed independently. *hsdS-2'* and *hsdS-4'* were tentatively identified as *hsdS* genes due to the presence of the conserved domains and variable domains in the correct spatial proportions, but they do not encode functional proteins since both are truncated at the 5'-end and missing a functional initiation codon. Futhermore a stop codon is present at position 442–444 of the *hsdS-4'* gene in the central conserved domain. As no frameshift has been induced in the gene sequence by this stop codon, it might have resulted from a point mutation.

On the clusters from both *L. lactis* strains are 915 bp–long genes (*int* genes) encoding 305 aa-proteins that display 99% identity to each other. Blast homology searches revealed 36% and 34% identity to the integrases of the *S. thermophilus* bacteriophage Sfi21 (GenBank: AF115103) and the lactococcal bacteriophage ΦLC3 (Genbank: U38906), respectively. The respective orientation of the integrase and the truncated *hsdS* gene in the two *L. lactis* strains suggests the possibility of a flip-flop mechanism.

L. lactis strains encode a large repertoire of restriction-modification specificity proteins

hsdS genes were amplified by PCR from type I operons encoded on the genomic DNA of four different *L. delbrueckii* ssp. strains, NCC39, NCC73, NCC627 and NCC786. The PCR products were sequenced, translated and, using the GCG PILEUP software (Devereux *et al.*, 1984), compiled in a multiple alignment with the *L. lactis* HsdS subunits isolated previously, including those encoded by the two N-terminal truncated genes *hsdS-2'* and *hsdS-4'*. As can be seen on Fig.4, three conserved domains can be distinguished: (i) at the N-terminus, (ii) at the C-terminus, and (iii) in a central position of the proteins. The *L. lactis* HsdS proteins do not encode the TAEL repeats characteristic of *Eco*R124 S proteins (Gubler and Bickle, 1991), but a slightly degenerated version of the Argos repeats (Argos, 1985; Kannan *et al.*, 1989) could be identified directly before the central and C-terminal conserved domains (Fig.3). Furthermore, one common motif specific to *L. lactis* strains is present in all three conserved domain: ‘w-[E,Q]-[E,Q]-x-K-l-g’ and located at the end of the N-terminal and central domain and at the start of the C-terminal domain (Fig.3).

The size of the HsdS proteins isolated from *L. lactis* are in the range of 360 to 430 amino acids except for the 204 aa-long NCC39 polypeptide. This polypeptide is composed of a single TRD inserted between a standard N-terminal conserved domain and a C-terminal domain that comports only the first four amino acids of the central conserved domain, but encodes the Argos repeats.

```
          1                                                  50
  NCC73   ~~~~~~~~~~ ~~~~~~MKKG PKLRFKGFTD DWEQRKLGDV CEEVSGNNGN
 NCC786   ~~~~~~~~~~ ~~~~~~MKKG PKLRFKGFTD DWEQRKLGDV CEEVSGNNGN
HsdS-2'   ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
pLL1212   ~~~~~~~~~~ ~MMNDTQAKY PQLRFKGFAD PWEQRKLGDV CEPITDSIDT
 HsdS-3   ~~~~~~~~~~ ~~MKD.EKKA PKLRFKGFTD DWEQRKLGDV CEPITDSIDT
 NCC627   ~~~~~~~~~~ ~~~~~~~~~~ PKLRFKGFTD DWEQRKLGDV CEPLTDSIDT
  pJBL2   ~~~~~~~~~~ ~~MNDTQAKY PQLRFKGFAD PWEQCKLGDV AKITMGQSPN
HsdS-4'   ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~MGQSPN
 HsdS-1   ~~~~~~~~~~ ~~~MKDEKKA PKLRFKGFTD DWEQVKYGEI FQRRSKMGVS
  NCC39   ~~~~~~~~~~ ~~~MKDEKKA PKLRFKGFTD DWEQCKLGDV AN.FSKGTGY
 HsdS-5   MAWSAGKMKQ IPWSSLSARS CRNRRKGCNF SWEQCKAKEL FE.....SKS
   cons   ~~~~~~~~~~ ~~~~~~.kk. pklrfkgftd dWEQrKlgdv cep.sgs..n

                                                conserved motif
          51                                                 100
  NCC73   VKGLPILTIS AANGWMNQKD RFSQVIA... GNELKKYTLL ..EKGHLAYN
 NCC786   VKGLPILTIS AANGWMNQKD RFSQVIA... GNELKKYTLL ..EKGHLAYN
HsdS-2'   ~~~~~~~~~~ ~~~~~MNQKD RFSQVIA... GNELKKYTLL ..EKGHLAYN
pLL1212   QKYPNEVFAE YSMPAFDASM KPDIVLGSSM NS.VRK...I .ITRPCLLVN
 HsdS-3   QKYPNEVFAE YSMPAFDASM KPDIVLGSSM NS.VRK...I .ITRPCLLVN
 NCC627   QKYPNEVFAE YSMPAFDASM KPDIVLGSSM NS.VRK...I .ITRPCLLVN
  pJBL2   SKNYTDNPKD HILVQGNADM KDGQVHPRIW TTEITK...I .ADKGDLI..
HsdS-4'   SKNYTDNPKD HILVQGNADM KDGQVHPRIW TTEITK...I .ADKGDLI..
 HsdS-1   TPALPSVEYD DINPGMGTLN KE....PKSK GTSKRG...I HFNPGDVLFG
  NCC39   SKSDLKGTGS PIILYGRLYT KYETIIRNVD TFVVPKSGSV FSKGGEVIVP
 HsdS-5   KKDHSDLPVL SATQENGIVF RNDLDIDIKF DTHTALSNYK VVSPGDYIIS
   cons   .ky...v.a. yi.pamna.m k.dqvi..s. gtevrk...i ...kgdl.vn

          101                                                150
  NCC73   HGNSKLAKYG TVFVQNLYDQ ALVPRVYHSF KMKTENNPYY VEYYFATKKL
 NCC786   HGNSKLAKYG TVFVQNLYDQ ALVPRVYHSF KMKTENNPYY VEYYFATKKL
HsdS-2'   HGNSKLAKYG TVFVQNLYDQ ALVPRVYHSF KMKTENNPYY VEYYFATKKL
pLL1212   KLNVR.KK.R IWYVKKPNKN AVCSAEFIPL YSDTVDLT.F LNQVAKSETF
 HsdS-3   KLNVR.KK.R IWYVKKPNKN AVCSAEFIPL HSDTVDLT.F LNQVAKSETF
 NCC627   KLNVR.EK.R IWYVKKPNKN AVCSAEFIPL YSDTVDLT.F LNQVAKSETF
  pJBL2   .LSVR.AP.. VGDIGKTSYD VVIGRGVAAI KGN....E.F IFQLLKRMKT
HsdS-4'   .LSVR.AP.. VGDIGKTSYD VVIGRGVAAI KGN....E.F IFQLLKRMKT
 HsdS-1   KLRPY.LK.N .WLFA..CFE GVAVGDFWVL TSSKIDHG.F TYSLIQAPEF
  NCC39   GSGET.AE.D ISIASVVEPA GILLGGDLNI IYPNSDLD.P AFLAITISNG
 HsdS-5   LRSFQ.GGFE LSEKRGIISP AYTVFAFNNQ EKNKQNST.F WKYLFKRYDF
   cons   klnvr.ak.. iw.v.k.... av..r.f..l ks.t.dlt.f ..qlfk..kf

          151                                                200
  NCC73   DRELARLVTS GARMDGLLNI NKKDFFKIKF EVPTPV.EQS LISTILQKLD
 NCC786   DRELARLVTS GARMDGLLNI NKKDFFKIKF EVPTPV.EQS LISTILQKLD
HsdS-2'   DRELARLVTS GARMDGLLNI NKKDFFKIKF EVPTPV.EQS LISTILQKLD
pLL1212   TRYLENH.SS GSS.NSQKRI TPRSLMLSKL HIPT.IEEQK LIGKIFESLD
 HsdS-3   TRYLENH.SS GSS.NSQKRI TPRSLMLSKL HIPT.IEEQK LIGKIFESLD
 NCC627   TRYLENH.SS GSS.NSQKRI TPRSLMLSKL HIPT.IEEQK LIGKIFESLD
  pJBL2   VGYWTKY.ST GS...TFESI NSLEINNAVI NLPK.EHEQN KVGKILSYMD
HsdS-4'   VGYWTKY.ST GS...TFESI NSLEINNAVI NLPK.EHEQN EVGKILSYMD
 HsdS-1   .QYIANL.SS GSK...MPRS DWGLVSNART FIPTNLSEQK SISSVLFGLD
  NCC39   KPHFDMA.RR AQG.KSVVHL HNVDLKHILL KTP.NLSEQK RISKIFESLD
 HsdS-5   IESLKRL.TF GIR.DG.KAI SFNQFSTLKL SFP.NFDEQN KIATLLHELD
   cons   .ryla.l.ss gsr.d..kri n..d...ikl .iPt...EQk li.kileslD

Argos repeat                                    iP ---EQ- -I-k----LD
```

```
          201                                                       250
   NCC73  QIITLHEEKK CLLERLKSVL SQEMFANKNG YPAVRFEGFD KAWKQSKLGE
  NCC786  QIITLHEEKK CLLERLKSVL SQEMFANKNG YPAVRFEGFD KAWKQSKLGE
  HsdS-2  QIITLHEEKK CLLERLKSVL SQEMFANKNG YPAVRFEGFD KAWKQSKLGE
 pLL1212  HTITLHEEKK RQLECLKSAL LQKMFADKSG YPVVRFEGFD KAWEERKLKD
  HsdS-3  HTITLHEEKK RQLERLKSAL LQKMFADESG YPVVRFEGFS DEWEQRKLKD
  NCC627  HTITLHEEKK RQLERLKSAL LKKMFADKSG YPVVRFKEFC DEWQGTKLRK
   pJBL2  HAITLHEEKK RQLECLKSAL LQKMFADKSG YPVVRFEGFS DEWEERKLGD
 HsdS-4'  HAITLHEEKK C*LERLKSAL LQKLFADKSG YPAVRFKGFG GNWKERKFGE
  HsdS-1  TAITLHEEKK RQLERLKSAL LQKMFADKSG YPAVRFKGFD DIWDQEKLNS
   NCC39  HTITLHDQKL NLLKLVKQSL RQNMM~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
  HsdS-5  RTITLHEEKK CQLERLKSAL LQKLFADENG NPAVRFKDFS EAWESRKLFD
    cons  htITLHeeKk .qLerlKsaL lqkmfadksg ypavrfegfd .aW.qrKlg.

                                                  conserved motif

          251                                                       300
   NCC73  ITKVI..... AAKPYISTPS THGNYLVIQQ ........GD KPIAGFSNSN
  NCC786  ITKVI..... AAKPYISTPS THGNYLVIQQ ........GD KPIAGFSNSN
 HsdS-2'  ITKVI..... AAKPYISTPS THGNYLVIQQ ........GD KPIAGFSNSN
 pLL1212  VVEKQI..KG KAQLEKLAP. ........GE VEYLDTSRLN G.....GQAI
  HsdS-3  VVEKQI..KG KAQLEKLAP. ........GE VEYLDTSRLN G.....GQAI
  NCC627  .ISKIV..RG ASPRPISDPK WFDNDSNVGW LRISDVTSQD GRIHHLKQHI
   pJBL2  ..AVSI.... .SSGVTGDAT LQDGEYRLTR IESISQGTLN ..VARLG.FT
 HsdS-4'  LGTIEM.... .CKRIFKDQT TDAGEIPFYK IGTFG.GMPN AYISREI.FE
  HsdS-1  L....V..RL HRGLTYSPNN VQDSGIRI.. LRSSNILDGQ FVMTDDDIFV
   NCC39  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
  HsdS-5  VIVKIIDFRG RTPKKMGLNW SERGHLALSA LNVKDGFIDF STDAHYANEI
    cons  .tk.ii.... aak.yis.ps th.nylvi.q ....d...g. ..iagf.n.i

          301                                                       350
   NCC73  PFKNYNNITL FGDHTLSLFK PRSPFLVAS. .......... .DGIKILSPS
  NCC786  PFKNYNNITL FGDHTLSLFK PRSPFLVAS. .......... .DGIKILSPS
 HsdS-2'  PFKNYNNITL FGDHTLSLFK PRSPFLVAS. .......... .DGIKILSPS
 pLL1212  LTNGLKD..V TLDDILI.LW DGS..KAGTV YHGFE..GAL GSTLKAY..R
  HsdS-3  LTNGLKD..V TLDDILI.LW DGS..KAGTV YHGFE..GAL GSTLKAY..R
  NCC627  SKQGQAKTRV ISEPHLL.LS IAA..TVGKP LINYVKIGVH DGFLIFI..E
   pJBL2  NKKPDQKYLL NLGDILY.SN INSLSHIGKV ALVDTTGIYH GINLLRFQMR
 HsdS-4'  RYKRLYPYPT K.GDILI.ST SGSIGRV..V EYTGQEAYYQ DSNIV.WLNH
  HsdS-1  KSSVVNIPTV KDGDILITAA NGSIKLVGKH AIISGISENT AVSGGFMLVG
   NCC39  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
  HsdS-5  LYKKWMSGNE LLKGQVLFTT EAPMGNVAQV PDDSGYVLSQ .RTISFVSNN
    cons  .fk..nnit. flddili..k pgs..lvg.v .......... ...ik.ls..

          351                                                       400
   NCC73  IEMNGLFYFY ELEKYK.... .......PKS EGYKRHFTIL KKCKANFPTD
  NCC786  IEMNGLFYFY ELEKYK.... .......PKS EGYKRHFTIL KKCKANFPTD
 HsdS-2'  IEMNGLFYFY ELEKYK.... .......PKS EGYKRHFTIL KKCKANFPTD
 pLL1212  TSANSKFVYQ YLKRHQDNIY ..NNYRTPNI P..HVQKDFL NVFTISVPVS
  HsdS-3  TSANSKFVYQ YLKRHQDNIY ..NNYRTPNI P..HVQKDFL NVFTISVPVS
  NCC627  PQFDLEFMFQ WLEMFRPQ.W ..SKYGQPGS QI.NLNSDLV KNQIIYLPSR
   pJBL2  NDVDSEFLFQ RLNTTPMKNW AVSHANPAVS QA.SINQTEL SKQPISLPTI
 HsdS-4'  NDLISNSFLK YLYKT..FNW ..NGVEGSTI KR.LYNNNIL .KTKIRIPSS
  HsdS-1  SSRIPDFVNS LFDTSWYQRF IRKYVTGGNG SIGNLKKNDL DKQYVKVPTT
   NCC39  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
  HsdS-5  SEITDNFLAV VLRSSKVQRA LSTLASGGTA K..GVSQSTL SQLIMVIPIS
    cons  iemnslf.fq .Lekyk.... .......p.s e..k...til kk.ki..Pts
Argos repeats                                                 iP--
```

```
          401                                                  442
   NCC73  IDEQKLIGSL LLHIDNLITK .Q~~~~~~~~~ ~~~~~~~~~~ ~
  NCC786  IDEQKLIGSL LLHIDNLITK .Q~~~~~~~~~ ~~~~~~~~~~ ~
 HsdS-2'  IDEQKLIGSL LLHIDNLITK QQLKMDKLNE TKESLLQNMF I*
 pLL1212  .DEQEKIGSF FKQLDDTIAF HQRKLDLLKE QKKGFLQKMF VV
  HsdS-3  .DEQEKIGSF FKQLDDTIDL HQRKLDLLKE QKKGFLQKMF V*
  NCC627  .EEQNQIGIF LNQIDSLINL HQQK~~~~~~ ~~~~~~~~~~ ~~
   pJBL2  .TEQQKIGSF FKQLDKTIAL HQRKLDLLKE QKKGFLQKMF V~
 HsdS-4'  .CEQEKIGGI LSTLDHLLSH HQQKIDIVKL IKQSLLQNMF I*
  Hsds-1  .SEQERIGEF FREIDQLIIN NQIKHEKLLE LKKFLLQNMF I*
   NCC39  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~
  HsdS-5  LDEQQKIERI IRIVEELTRL YQNKLEILTE LKKSLLQKMF I*
    cons  .dEQekIgsf l.qidnlitl hQ.K.d~l.e ~Kk~.LQ.MF ..

    conserved motif

Argos      -EQ--I-k- ---LD
repeats
```

Fig. 3 Multiple alignement of HsdS polypeptides isolated from *L. lactis* strains, using the PILEUP sofware from GCG. The boldface type indicates that the amino acid residue is conserved in at least 6 of the aligned sequences. Capital letters in the consensus indicates that the residue is conserved in all sequences. The conserved motif repeated in all three conserved domain is 'w-[E,Q]-[E,Q]-x-K-l-g'.

Pairwise comparisons of these proteins revealed several significant occurences. HsdS-3 isolated from the chromosomal NCC82 *hsd* cluster was 95.9% identical to the HsdS subunit encoded by the *L. lactis* plasmid pLL1212 (GenBank: AF109691). Contrary to expectations, the few diverging amino acids were not located in the TRDs of the proteins but in the conserved domains. The N-terminal conserved domain and TRD of NCC627 S subunit also displayed 95.0% identity to HsdS-3. The percentage of identity decreased sharply towards the C-terminus of the central conserved domain. Similarly the N-terminal TRD of the pJBL2-encoded HsdS subunit showed 97% to the N-terminal TRD of the truncated S protein encoded by the chromosomal *hsdS-4'* gene. The complete HsdS subunit amplified from the NCC73 strain was 93.6% identical to the truncated S polypeptide encoded by the *hsdS-2'* gene. Interestingly, the C-terminal TRD of the NCC627 HsdS revealed 84% positive identities to the N-terminal TRD of the pN42-encoded S protein.

Discussion

L. lactis is a bacterium of industrial importance and is used daily worldwide in huge quantities for the production of cheese. In contrast to other dairy bacteria, *L. lactis* is rarely infected by bacteriophages implying that the bacterium possess a very efficient and adaptable defence mechanism. Enzymatic studies of the *L. lactis* strain NCC88 clearly showed the presence of a

type I restriction-modification enzyme in this strain. Further molecular biology analyses led to the isolation and sequencing of *hsd* gene clusters encoding type I R-M systems from the two *L. lactis* strains NCC88 and NCC82, as well as *hsdS* genes from four additional *L. lactis* strains. Analysis of the *hsd* clusters from NCC82 and NCC88 revealed a similar genetic organization comprising: (i) the *hsd* cluster proper encoding the *hsdR*, *hsdM* and *hsdS* genes necessary for the expression of a type I restriction enzyme, and (ii) a 5'-truncated *hsdS* gene and an integrase-encoding (*int*) gene, both located downstream of the *hsd* cluster proper and positioned in opposite orientations in the two *L. lactis* strains. NCC82 also encodes an additional complete *hsdS* gene on the minus strand of its chromosome and downstream relative to the *hsd* cluster proper.

The *hsdR*, *hsdM* and *hsdS* genes are organised in two transcriptional units with the single *hsdR* gene operon directly upstream of the *hsdMS* operon. This structure is similar to that of type IA and type IB *hsd* clusters (Loenen *et al.*, 1987; Murray *et al.*, 1993) but diverge significantly from type IC clusters for which the order of the two transcriptional units is reversed (Price *et al.*, 1989; Skrzypek and Piekarowicz, 1989; Tyndall *et al.*, 1994). In 1993, Murray *et al.* reported that the level of identity between *E. coli* R subunits from different type I families is so low that comparisons fall into the twilight zone defined by Doolittle *et al.* (1986) and that only well-chosen gap penalties yield a PILEUP alignment in which the helicase motifs are aligned among the R polypeptides. In this work, the best alignment for identification of the seven helicase motifs in *L. lactis* R subunits was obtained with the type IC R.*Eco*R124II. Moreover, the level of identity observed between the *L. lactis* and *Eco*R124 R and M subunits (32% and 35%, respectively) are consistent with a common ancestry for the *hsd* genes from these bacteria. The structure of the *L. lactis* S subunits composed of three conserved domains interspersed by two variable domains is reminiscent of either type IB or IC enzymes, whereas the presence of the Argos repeats directly before the conserved domains is consistent with type IC restriction enzymes (Kannan *et al.*, 1989). Despite the genetic organization and the chromosomal location of their *hsd* clusters (type IC *E. coli* systems being plasmid-encoded), *L. lactis* strains thus encode type IC DNA restriction and modification systems.

It has been argued that the presence of the Argos repeats in all *E. coli* S proteins indicate that the three type I–families are descended from a common ancestral gene, and that familial divergence and the generation of new specificities within the families are a consequence of evolution (Argos,1985; Kannan *et al.*, 1989). The presence of a slightly degenerated form of the Argos repeats in *L. lactis* S polypeptides, and the level of homologies observed between the R and M subunits of *L. lactis* and *Eco*R124, as high as that found between *E. coli* subunits from different type I families (Murray *et al.*, 1993) indicate that *L. lactis* may have acquired

its *hsd* cluster from *E. coli*. This is further supported by the fact that the *L. lactis* enzymes belong to the type IC family of restriction enzymes which are plasmid–borne in *E. coli* and thus more readily available for horizontal transfer than the chromosomally–encoded type IA and IB families. The homology observed between the TRDs of HsdS-1 and S.*Eco*K combined with the fact that TRDs from pN42– and pJBL2–encoded S polypeptides display homologies to different *E. coli* S subunits, *Eco*R124II and *Eco*DI, respectively (chapter V), suggests that the exchange of *hsdS* genes between *L. lactis* and *E. coli* was more than a one-time event. It is clear, on the other hand, from the near identity of the *hsdR* and *hsdM* genes in both *L. lactis* strains that they share common ancestral genes and that genetic variability in the *L. lactis hsd* cluster is limited to the *hsdS* genes. The *L. lactis* acquisition, probably by horizontal transfer, of a type IC plasmid–encoded *E. coli hsd* cluster must have been followed evolutionarily by futher exchanges of *hsdS* genes between these two bacterial genera.

In type I R-M systems, the HsdS subunit is responsible for the recognition and binding of the enzyme to its specific target site split into two recognition components (Bickle and Krüger, 1993). Recognition of this site is facilitated by the structure of the specificity subunit possessing two target recognition domains (TRDs) so that each recognize one component of the target site (Gough and Murray, 1983; Fuller-Pace *et al.*, 1984; Nagaraja, 1985). The adaptability of type I R-M systems thus depends on the capacity of the host to acquire new *hsdS* genes as well as its ability to implement recombination between these genes to generate new specificities. This phenomenon has been well-characterized in enteric bacteria (Gann *et al.*, 1987) and an example of domain shuffling was recently reported in the dairy bacterium *Lc. lactis* (O'Sullivan *et al.*, 2000). Comparative analysis of S polypeptides encoded by the *hsdS* genes isolated from different *L. lactis* strains highlighted the fact that such recombination events must be frequent in *L. lactis*. Near identical TRDs (94% identity or higher) were identified in HsdS subunits from different strains either in the same positions, e.g. as amino terminal domains, or in different positions, one N-terminal TRD constituting the carboxy domain of a second HsdS protein. The high number of such occurences observed for a relatively small number of sequences analysed suggests a very high frequency of domain-shuffling events. This is substantiated by the presence, downstream of the *hsd* cluster proper, of the *int* gene encoding a protein homologous to the lactococcal bacteriophage ΦLC3 integrase. ΦLC3 Int being essential for both DNA excision and integration in the genome acts as a site-specific recombinase (Lillehaug and Birkeland, 1993). The presence of the conserved motif ‘w-[E,Q]-[E,Q]-x-K-l-g’ in all three conserved domains of the *L. lactis* proteins might provide an attachment site for a recombination event catalysed by such an integrase. It may also explain the presence in the vicinity of the *int* genes of inactive, truncated *hsdS* genes that might have provided the genetic material for recombination with the *hsdS* gene belonging to

the *hsd* cluster proper. The identification of *hsdS* genes encoding near identical proteins (96% identity) on the *L. lactis* plasmid pLL1212 and NCC82 chromosome supplies circumstantial evidence that the material required for the generation of new specificities is provided by horizontal transfer and exchange of plasmid DNA between different *L. lactis* strains. The *hsdS*–encoding plasmids have thus the dual function of providing genetic variability for the chromosomally encoded type I system, as was reported in *Lc. lactis* (Schouler *et al.*, 1998) and acting as vectors for the exchange of *hsdS* genes between *L. lactis* strains by horizontal transfer.

A type I DNA restriction and modification system relying on such an efficient system for the obtention of new genetic material and the generation of novel specificities by recombination of TRDs constitute a very adaptable defense mechanism against invasive foreign DNA and may explain the difficulties to transform or conjugate this microorganism as well as the exceptional resistance of *L. lactis* to bacteriophage infections.

Acknowledgement

This work was supported by the Swiss National Foundation within SPP Food Biotechnology Program no. 5002-044544.

References

- Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25:** 3389–3402.
- Argos, P. (1985) Evidence for a repeating domain in type I restriction enzymes. *EMBO J* **4:** 1351–1355.
- Auad, L.,Peril, M.A.A., Holgado, A.A.P.R., and Raya, R.R. (1998) Evidence of a restriction/modification system in *Lactobacillus delbrueckii* subsp. *lactis* CNRZ 326. *Curr Microbiol* **36:** 271–273.
- Bickle, T.A., and Krüger, D.H. (1993) Biology of DNA restriction. *Microbiol Rev* **57:** 434–450.
- Chen, J.-D., and Morrison, D.A. (1988) Construction and properties of a new insertion vector, pJDC9, that is protected by transcriptional terminators and useful for cloning of DNA from *Streptococcus pneumoniae*. *Gene* **64:** 155–164.

- Delley, M., Mollet, B., and Hottinger, H. (1990) DNA probe for *Lactobacillus delbrueckii*. *Appl Environ Microbiol* **56:** 1967–1970.
- Devereux, J., Haeberli, P., and Sharpe, E. (1984) A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res* **12:** 387–395.
- Doolittle, R.F., Feng, D.F., Johnson, M.S., and McClure, M.A. (1986) Relationships of human protein sequences to those of other organisms. *Cold Spring Harbor Quant Biol* **51:** 447–455.
- Dreier, J., MacWilliams, M.P., and Bickle, T.A. (1996) DNA cleavage by type IC restriction-modification enzyme *Eco*R124II. *J Mol Biol* **264:** 722–733.
- Fuller-Pace, F.V., Bullas, L.R., Delius, H., and Murray, N.E. (1984) Genetic recombination can generate altered restriction specificity. *Proc Natl Acad Sci USA* **81:** 6095–6099.
- Fuller-Pace, F.V., and Murray, N.E. (1986) Two DNA recognition domains of the specificity polypeptides of a family of type I restriction enzymes. *Proc Natl Acad Sci USA* **83:** 9368–9372.
- Gann, A.A.F., Campbell, A.J.B., Collins, J.F., Coulson, A.F.W., and Murray, N.E. (1987) Reassortment of DNA recognition domains and the evolution of new specificities. *Mol Microbiol* **1:** 13–22.
- Gorbalenya, A.E., Koonin, E.V., Donchenko, A.P., and Blinov, V.M. (1989) Two related superfamilies of putative helicases involved in replication, recombination, repair and expression of DNA and RNA genomes. *Nucleic Acids Res* **17:** 4713–4730.
- Gorbalenya, A.E., and Koonin, E.V. (1991) Endonucleases (R) subunits of type-I and type-III restriction-modification enzymes contain a helicase-like domain. *FEBS Lett* **291:** 277–281.
- Gough, J.A., and Murray, N.E. (1983) Sequence diversity among related genes for recognition of specific targets in DNA molecules. *J Mol Biol* **166:** 1–19.
- Gubler, M., and Bickle, T.A. (1991) Increased protein flexibility leads to promiscuous protein–DNA interactions in type IC restriction-modification systems. *EMBO J* **10:** 951–957.
- Janscak, P., Abadjieva, A. and Firman, K. (1996) The type I restriction endonuclease R.EcoR124I: over-production and biochemical properties. *J Mol Biol* **257:** 977–991.
- Kannan, P., Cowan, G.M., Daniel, A.S., Gann, A.A., and Murray, N.E. (1989) Conservation of organization in the specificity polypeptides of two families of type I restriction enzymes. *J Mol Biol* **209:** 335–344.

- Lillehaug, D., and Birkeland, N.-K. (1993) Characterization of genetic elements required for site-specific integration of the temperate lactococcal bacteriophage ΦLC3 and construction of integration-negative ΦLC3 mutants. *J Bacteriol* **175:** 1745–1755.
- Loenen, W.A.M., Daniel, A.S., Braymer, H.D., and Murray, N.E. (1987) Organization and sequence of the *hsd* genes of *Escherichia coli* K-12. *J Mol Biol* **198:** 159–170.
- Murray, N.E., Daniel, A.S., Cowan, G.M., and Sharp, P.M. (1993) Conservation of motifs within the unusually variable polypeptide sequences of type I restriction and modification enzymes. *Mol Microbiol* **9:** 133–143.
- Nagaraja, V., Shepherd, J.C.W., and Bickle, T.A. (1985) The evolution of DNA sequence specificity: a recombinant restriction enzyme has a hybrid recognition sequence. *Nature* **316:** 371–372.
- O'Sullivan, D.J., and Klaenhammer, T.R. (1993) Rapid mini-prep isolation of high-quality plasmid DNA from *Lactococcus lactis* and *Lactobacillus* spp. *Appl Environ Microbiol* **59:** 2730–2733.
- O'Sullivan, D., Twomey, D.P., Coffey, A., Hill, C., Fitzgerald, G.F., and Ross, R. P. (2000) Novel type I restriction specificities through domain shuffling of HsdS subunits in *Lactococcus lactis. Mol Microbiol* **36:** 866–875.
- Price, C., Pripfl, T., and Bickle, T.A. (1987) *Eco*R124 and *Eco*R124/3: the first members of a new family of type I restriction and modification systems. *Eur J Biochem* **167:** 111–115.
- Price, C., Lingner, J., and Bickle, T.A. (1989) Basis for changes in DNA recognition by the *Eco*R124 and *Eco*R124/3 type I DNA restriction and modification enzymes. *J Mol Biol* **205:** 115–125.
- Oppenheim, D.S., and Yanofsky, C. (1980) Translational coupling during expression of the tryptophan operon of *Escherichia coli*. *Genetics* **95:** 785–795.
- Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., Mullis, K.B., and Ehlich, H.A. **(**1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* **239:** 487–491.
- Sambrook, J., Fritsch, F., and Maniatis, T. (1989) Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbour Laboratory, Cold Spring Harbour, N.Y.
- Sanger, F., Nicklen, S., and Coulson, A.R. (1977) DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA* **74:** 5463–5467.
- Schouler, C., Gautier, M., Ehrlich, S.D. and Chopin, M.C. (1998) Combinational variation of restriction modification specificities in Lactococcus lactis. *Mol Microbiol* **28:** 169-78.

- Skrzypek, E., and Piekarowicz, A. (1989) The *Eco*DXXI restriction and modification system: cloning the genes and homology to type I restriction and modification systems. *Plasmid* **21:** 195–204.
- Taylor, I., Patel, J., Firman, K., and Kneale, G. (1992) Purification and biochemical characterisation of the *Eco*R124 type I modification methylase. *Nucleic Acids Res* **20:** 179–186.
- Tyndall, C., Meister, J., and Bickle, T.A. (1994) The *Escherichia coli prr* region encodes a functional type IC DNA restriction system closely integrated with an anticodon nuclease gene. *J Mol Biol* **237:** 266–274.

CHAPTER VII

General conclusion

Lactobacillus delbrueckii subsp. *lactis* is used worldwide daily in huge quantities for the production of Swiss-type hard cheeses. In contrast to other lactic acid bacteria of industrial importance, e.g. lactococci or streptococci, that have been extensively studied, we have few informations on the metabolism or physiology of *L. lactis*. The lack of transformation or conjugation systems adapted to this bacterium combined with the difficulty of extracting proteins may explain in part this situation. One interesting feature of this lactic acid bacterium (LAB) is its unusually high resistance to bacteriophage infections suggesting that this microorganism possess very effective phage defence mechanisms. One such mechanism is the use of restriction-modification (R-M) systems to degrade invading DNA, e.g. phage genomes.

The aim of this doctorate work was to gain insights into the metabolism and physiology of *L. lactis*. In order to try and elucidate its phage resistance mechanisms, we concentrated on the study of *L. lactis* R-M systems. Protein purification methods using chromatography columns yielded a semi-purified enzyme fraction containing an endonuclease degrading λ DNA in a smear and strictly dependent on the presence of the three cofactors Mg^{2+}, ATP and S-adenosyl-methionine (SAM) for its activity. Those properties are characteristic of a type I restriction-modification system. A molecular biology approach combining PCR and inverted PCR allowed us to amplify the genes coding for type I restriction enzymes from two different *L. lactis* strains: NCC88 from which the active type I enzyme had been semi-purified, and NCC82. These genes are encoded on 8 kb- and 10 kb-clusters for NCC88 and NCC82, respectively. The clusters can be divided in two parts. The first part of the cluster referred to as the '*hsd* cluster proper' encodes the genes *hsdR*, *hsdM* and *hsdS*. These genes constitute the basic *hsd* cluster identified in well-characterised *E. coli* type I R-M systems (Murray, 2000). *L. lactis* clusters, however, contain a second part encoding an integrase (*int*) gene and a second, 5'-truncated *hsdS* gene, both located downstream of the *hsd* cluster proper, but oriented in opposite directions in the two *L. lactis* strains. NCC82 cluster also encodes a third, complete *hsdS* gene. Comparison of the clusters from both *L. lactis* strains revealed that the *int*, *hsdR* and *hsdM* genes are very highly conserved (>97% identity). The *hsdS* genes display lower overall homologies but the sequences encoding the conserved domains of HsdS subunits are just as conserved and only the variable domains encoding the target recognition domains (TRDs) are different indicating that the *hsdS* genes present in the clusters express HsdS subunits with different specificities. *hsd* clusters from different *L. lactis* strains thus share a common ancestry, whereas new specificities were generated subsequently. The

isolation of *hsdS* genes from four further *L. delbrueckii* strains and the comparison of these genes with the *hsdS* genes previously identified revealed circumstantial evidence of: (i) horizontal transfer of the plasmid-borne genes, and (ii) target recognition domain-shuffling between different *hsdS* genes.

R-M systems are the most powerful but paradoxically the most fragile of bacteriophage defence mechanism. Indeed, most incoming phage genomes are degraded by the restriction enzymes protecting the host against infection. However, in the rare instances where the phage DNA is methylated at all its recognition sites before restriction can occur, the phage becomes protected against further restriction and may replicate with impunity and infect the whole bacterial culture. The ability of a bacterium to acquire and/or generate novel restriction specificities is thus determinant for its phage resistance efficiency. The highly flexible mechanism identified in *L. lactis* for the generation of novel specificities may therefore go a long way towards explaining the unusually high phage resistance of *L. lactis* strains.

In this work, we could semi-purify and isolate the genes encoding a type I restriction-modification enzyme, but observed no evidence of the presence of either type II or type III R-M systems. This doesn't mean that the bacterium possesses none of them. As was found for *Helicobacter pylori*, a genome sequencing project reveals numerous, previously unknown R-M systems (Tomb *et al.*, 1997; Xu *et al.*, 2000). Those previously unidentified R-M systems might not be expressed in laboratory strains in the absence of the induction that phage attacks represent for the bacteria. Indeed, experiments that followed genome sequencing of the pathogen *Neisseria gonorrhoeae*, revealed that the M genes are expressed to keep the host genome modified, whereas the cognate R genes may be silent. An interesting approach to continue this work would be a genomic approach. With the total sequence of *L. lactis* genome, it would be possible to ascertain whether type II or type III R-M systems or even other *hsdS* genes might be present in the bacterium. A genomic approach would also allow us to identify other phage resistance mechanisms, e.g. adsorption inhibition, prevention of phage DNA penetration and abortive infection that have been reported in lactic acid bacteria (Daly *et al.*, 1996). Differential expression analysis of these genes in the presence or absence of phage infection would provide a clear basis for understanding the mechanisms of phage resistance in this dairy starter bacterium.

A genomic approach would also provide information on the energy metabolism of *L. lactis*. However, the information provided by differential expression analyses might not be sufficient to elucidate the regulation of the pathways. In the glycolysis pathway, for example, certain enzymes have key regulator roles that can only be identified following biochemical characterisation and identification of the various effectors acting on the enzyme. The characterisation of 3-phosphoglycerate kinase and especially of the allosteric enzyme

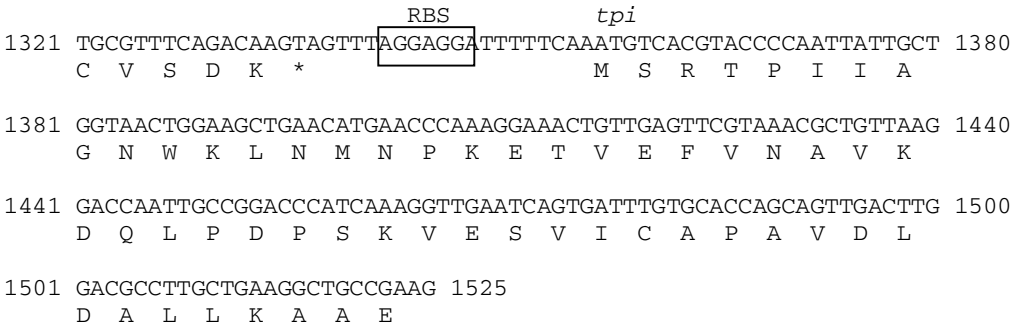
pyruvate kinase provided insights into the energy metabolism of *L. lactis*. Further studies should consider the expression of these enzymes and whether they are subjected to catabolite repression by CcpA as is the case for *Bacillus* and *Streptococcus thermophilus* (Tobish *et al.*, 1999; van der Bogaard *et al*., 2000).

References

- Daly, C., Fitzgerald, G.F. and Davis, R. (1996) Biotechnology of lactic acid bacteria with special reference to bacteriophage resistance. *Antonie van Leeuwenhoek* **70:** 99–110.
- Murray, N.E. (2000) Type I restriction systems: sophisticated molecular machines (a legacy of Bertani and Weigle). *Microbiol. Mol. Biol. Rev.* **64:** 412–434.
- van Den Bogaard, P.T., Kleerebezem, M., Kuipers, O.P. and de Vos, W.M. (2000) Control of lactose transport, β-galactosidase activity, and glycolysis by CcpA in *Streptococcus thermophilus*: Evidence for carbon catabolite repression by a non-phosphoenolpyruvate-dependent phosphotransferase system sugar. *J. Bacteriol.* **182:** 5982–5989.
- Tobisch, S., Zuhlke, D., Bernhardt, J., Stulke, J. and Hecker, M. (1999) Role of CcpA in regulation of the central pathways of carbon catabolism in Bacillus subtilis. *J. Bacteriol.* **181:** 6996–7004.
- Tomb, J.F., White, O., Kerlavage, A.R., Clayton, R.A., Sutton, G.G., Fleischmann, R.D., Ketchum, K.A., Klenk, H.P., Gill, S., Dougherty, B.A., *et al.* (1997) The complete genome sequence of the gastric pathogen Helicobacter pylori. *Nature* **388:** 539–547.
- Xu, Q., Morgan, R.D., Roberts, R.J. and Blaser, M.J. (2000) Identification of type II restriction and modification systems in *Helicobacter pylori* reveals their substancial diversity among strains. *Proc. Natl. Acad. Sci. USA* **97:** 9671–9676.

I. Nucleotide sequence of *L. lactis* NCC88 3-phosphoglycerate kinase.

```
      gap
    1 TCGTACTTTGTTGCACTTTGCTACTCTTTAATTTAAAGTTTTCTAAATTGAATTAGTAGT  60
      R  T  L  L  H  F  A  T  L  *
                                                          RBS
   61 GTGAAAGAAGGCGGAGGGAAATTTCTTCCTTCCGCCTTTTTTCGATGAATATAGGAGATT 120

          pgk
  121 TTTTCAATGGCTAAATTGATTGTTTCTGACGTAGACGTTAAGGACAAGAAGGTTTTGGTT 180
            M  A  K  L  I  V  S  D  V  D  V  K  D  K  K  V  L  V

  181 CGCGTTGACTTCAACGTGCCGATTAAGGACGGCGTTATCGGCGACGACAACCGTATCGTG 240
      R  V  D  F  N  V  P  I  K  D  G  V  I  G  D  D  N  R  I  V

  241 GCTGCTTTGCCAACTATCAAGTACATCATCGAAAACGGCGGCAAGGCTATCTTGCTTTCC 300
      A  A  L  P  T  I  K  Y  I  I  E  N  G  G  K  A  I  L  L  S

  301 CACCTTGGCCGGATCAAGAGCGATGAAGACAAGAAGAGCTTGAGCCTGGCTCCAGTTGCC 360
      H  L  G  R  I  K  S  D  E  D  K  K  S  L  S  L  A  P  V  A

  361 AAACGTTTGGGCGAATTGCTTGAAAAGCCTGTAACTTTCGTACCTTCAAACGAAGGCAAG 420
      K  R  L  G  E  L  L  E  K  P  V  T  F  V  P  S  N  E  G  K

  421 GAAGTTGAAGACGCCATCAACAACATGAAGGACGGCGACGTAGTTGTTTTGGAAAACACC 480
      E  V  E  D  A  I  N  N  M  K  D  G  D  V  V  V  L  E  N  T

  481 CGTTTCCAAGACATCGACAACGACTTCGGCAAGCGTGAATCAAAGAACGACCCTAAGCTG 540
      R  F  Q  D  I  D  N  D  F  G  K  R  E  S  K  N  D  P  K  L

  541 GGCGAATACTGGGCATCATTGGGTGACGTTTTCGTAAACGACGCGTTCGGTACTGCTCAC 600
      G  E  Y  W  A  S  L  G  D  V  F  V  N  D  A  F  G  T  A  H

  601 AGAAGCCACGCTTCAAACGTTGGTATCGCTACGGCCATGAAGGCTGCCGGTAAACCAGTA 660
      R  S  H  A  S  N  V  G  I  A  T  A  M  K  A  A  G  K  P  V

  661 GCTGCTGGTTTCCTGCTTGAAAAGGAAATCAAGTTCCTGGGCAACGCTGTTGCTAACCCA 720
      A  A  G  F  L  L  E  K  E  I  K  F  L  G  N  A  V  A  N  P

  721 GTTCACCCATTCGTAACTATTCTTGGCGGGGCTAAGGTTTCTGACAAGATCGGCGTTATT 780
      V  H  P  F  V  T  I  L  G  G  A  K  V  S  D  K  I  G  V  I

  781 ACCAACTTGATTCCAAAGGCTGACCACATCATCATCGGTGGTGGTATGGCTTACACCTTC 840
      T  N  L  I  P  K  A  D  H  I  I  I  G  G  G  M  A  Y  T  F

  841 CTTAAGGCTCAAGGCCACAATATCGGCAAGTCCCTGGTTGAAGACGACAAGGTTGAATTT 900
      L  K  A  Q  G  H  N  I  G  K  S  L  V  E  D  D  K  V  E  F

  901 GCCAAGGAATTGCTGGAAAAGGCTGGCGACAAGCTGGTTCTGCCAATCGACCACGTAGCC 960
      A  K  E  L  L  E  K  A  G  D  K  L  V  L  P  I  D  H  V  A

  961 GCAACTGAATTCAACAACGATGCTGCTTCAGAAGTTGTTGGCCAAGACATCCCAGACAAC 1020
      A  T  E  F  N  N  D  A  A  S  E  V  V  G  Q  D  I  P  D  N

 1021 GAAATGGGCTTGGACATCGGTCCTAAGACTATTGAACTCTTCAAGAAGACTCTTGAAGGT 1080
      E  M  G  L  D  I  G  P  K  T  I  E  L  F  K  K  T  L  E  G

 1081 GCCAAGACTGTTGTTTGGAACGGGCCAATGGGCGTCTTCGAAATGCCAAACTTCGCCAAG 1140
      A  K  T  V  V  W  N  G  P  M  G  V  F  E  M  P  N  F  A  K

 1141 GGTACTTTGGAAGTTGGCCGTGCTTTGGCTGACTTGCCAGACGCTACTACTATTGTCGGC 1200
      G  T  L  E  V  G  R  A  L  A  D  L  P  D  A  T  T  I  V  G

 1201 GGTGGTGACTCAACTGCTGCTGCTAAGCAATTGGGGATCGCTCCTAAGTTGACTCACATC 1260
      G  G  D  S  T  A  A  A  K  Q  L  G  I  A  P  K  L  T  H  I

 1261 TCAACTGGTGGTGGTGCATCTCTTGAATACCTTGAAGGTAAGGAATTGCCAGGTATCGCT 1320
      S  T  G  G  G  A  S  L  E  Y  L  E  G  K  E  L  P  G  I  A
```


RBS
tpi
1321 TGCGTTTCAGACAAGTAGTTTAGGAGGATTTTTCAAATGTCACGTACCCCAATTATTGCT 1380
C V S D K * M S R T P I I A
1381 GGTAACTGGAAGCTGAACATGAACCCAAAGGAAACTGTTGAGTTCGTAAACGCTGTTAAG 1440
G N W K L N M N P K E T V E F V N A V K
1441 GACCAATTGCCGGACCCATCAAAGGTTGAATCAGTGATTTGTGCACCAGCAGTTGACTTG 1500
D Q L P D P S K V E S V I C A P A V D L
1501 GACGCCTTGCTGAAGGCTGCCGAAG 1525
D A L L K A A E

II. Nucleotide sequence of *L. lactis* NCC88 pyruvate kinase.

pyk

```
   1 ATGAAAAAAACAAAGATTGTTAGTACTTTAGGACCAGCTTCAAACGATATTGAAACTATT  60
     M  K  K  T  K  I  V  S  T  L  G  P  A  S  N  D  I  E  T  I

  61 ACCAAGTTAGCCGAAGCAGGCGCAAACGTATTCCGTTTCAACTTCTCACACGGCGACCAC 120
      T  K  L  A  E  A  G  A  N  V  F  R  F  N  F  S  H  G  D  H

 121 GAAGAACACTTGGCAAGAATGAACATGGTTCGTGAAGTTGAAAAGAAGACTGGCAAGCTT 180
     E  E  H  L  A  R  M  N  M  V  R  E  V  E  K  K  T  G  K  L

 181 TTGGGCATCGCTTTGGACACCAAGGGTGCTGAAATCAGAACCACTGACCAAGAAGGCGGC 240
     L  G  I  A  L  D  T  K  G  A  E  I  R  T  T  D  Q  E  G  G

 241 AAGTTCACTATCAACACTGGTGACGAAATCCGCGTGTCAATGGACGCAACTAAGGCCGGC 300
     K  F  T  I  N  T  G  D  E  I  R  V  S  M  D  A  T  K  A  G

 301 AACAAGGACATGATCCACGTTACCTACCCAGGTCTGTTCGACGACACTCACGTAGGCGGC 360
     N  K  D  M  I  H  V  T  Y  P  G  L  F  D  D  T  H  V  G  G

 361 ACTGTATTGATCGACGACGGTGCTGTTGGTTTGACTATCAAGGCCAAGGACGAAGAAAAG 420
     T  V  L  I  D  D  G  A  V  G  L  T  I  K  A  K  D  E  E  K

 421 CGCGAATTGATTTGTGAAGCTCAAAACACTGGTGTCATCGGCTCAAAGAAGGGTGTTAAC 480
     R  E  L  I  C  E  A  Q  N  T  G  V  I  G  S  K  K  G  V  N

 481 GCTCCAGGTGTTGAAATCCGCCTCCCAGGGATTACTGAAAAGGACACTGACGACATCCGC 540
     A  P  G  V  E  I  R  L  P  G  I  T  E  K  D  T  D  D  I  R

 541 TTTGGTTTGAAGCACGGTATTAACTTCATCTTTGCTTCTTTTGTACGTAAGGCTCAAGAC 600
     F  G  L  K  H  G  I  N  F  I  F  A  S  F  V  R  K  A  Q  D

 601 GTTCTTGACATTCGCGCACTTTGCGAAGAAGCTAACGCACCATACGTTAAGATCTTCCCA 660
     V  L  D  I  R  A  L  C  E  E  A  N  A  P  Y  V  K  I  F  P

 661 AAGATTGAATCACAAGAAGGTATTGACAACATCGACGAAATCTTGCAAGTTTCAGATGGT 720
     K  I  E  S  Q  E  G  I  D  N  I  D  E  I  L  Q  V  S  D  G

 721 TTGATGGTTGCCCGTGGTGACATGGGTGTTGAAATCCCATTCATCAACGTGCCATTTGTT 780
     L  M  V  A  R  G  D  M  G  V  E  I  P  F  I  N  V  P  F  V

 781 CAAAAGACTTTGATCAAGAAGTGCAACGCTTTGGGCAAGCCAGTTATCACTGCTACTCAA 840
     Q  K  T  L  I  K  K  C  N  A  L  G  K  P  V  I  T  A  T  Q

 841 ATGCTGGACTCAATGCAAGAAAACCCACGTCCAACCCGTGCCGAAGTAACTGACGTTGCT 900
     M  L  D  S  M  Q  E  N  P  R  P  T  R  A  E  V  T  D  V  A

 901 AACGCCGTTCTTGACGGTACTGACGCAACTATGCTGTCAGGCGAATCAGCAAACGGTTTG 960
     N  A  V  L  D  G  T  D  A  T  M  L  S  G  E  S  A  N  G  L

 961 TACCCAGTAGAATCAGTTCAAGCTATGCACGACATCAATGTTCGGACTGAAAAGGAAATG 1020
     Y  P  V  E  S  V  Q  A  M  H  D  I  N  V  R  T  E  K  E  M

1021 GACACCCGGAACACTCTGGCTCTGCAACGCTTTGAAGAATACAAGGGCTCAAACGTTACT 1080
     D  T  R  N  T  L  A  L  Q  R  F  E  E  Y  K  G  S  N  V  T

1081 GAAGCTATCGGCGAATCAGTTGTCCGCACTGCTCAAGAACTGGGCGTTAAGACTATCATC 1140
     E  A  I  G  E  S  V  V  R  T  A  Q  E  L  G  V  K  T  I  I

1141 ACTGCTACTAGCTCCGGCTACACAGCTCGTATGATCTCCAAGTACCGTCCAGACGCAACC 1200
     T  A  T  S  S  G  Y  T  A  R  M  I  S  K  Y  R  P  D  A  T

1201 ATCGTTGCCTTGACTTTCGACGAAAAGATCCAACACTCATTGGGTATCGTTTGGGGCGTT 1260
     I  V  A  L  T  F  D  E  K  I  Q  H  S  L  G  I  V  W  G  V

1261 GAACCAGTTTTGGCAAAGCAACCTTCAAACACTGACGAAATGTTCGAAGAAGCTGCCCGC 1320
     E  P  V  L  A  K  Q  P  S  N  T  D  E  M  F  E  E  A  A  R
```

```
1321 GTAGCTAAGGAACACGGTTTCGTTAAGGATGGCGACCTGGTAATCATCGTTGCCGGCGTA 1380
     V  A  K  E  H  G  F  V  K  D  G  D  L  V  I  I  V  A  G  V

1381 CCATTCGGCCAATCAGGTACTACTAACTTGATGAAGCTGCAAATCATCGGCAACCAACTT 1440
     P  F  G  Q  S  G  T  T  N  L  M  K  L  Q  I  I  G  N  Q  L

1441 GCTCAAGGTTTGGGCGTAGGCACTGGCTCAGTTATCGGCAAGGCTGTTGTTGCGAACAGC 1500
     A  Q  G  L  G  V  G  T  G  S  V  I  G  K  A  V  V  A  N  S

1501 GCTGAAGAAGCCAACAATAAGGTTCACGAAGGCGACATCCTGGTAGCTAAGACTACTGAC 1560
     A  E  E  A  N  N  K  V  H  E  G  D  I  L  V  A  K  T  T  D

1561 AAGGACTACATGCCAGCTATCAAGAAGGCCAGCGGTATGATCGTTGAAGCTTCCGGCTTG 1620
     K  D  Y  M  P  A  I  K  K  A  S  G  M  I  V  E  A  S  G  L

1621 ACCAGCCACGCAGCTGTTGTCGGCGTATCACTCGGCATTCCAGTTGTTGTCGGTGCTGCT 1680
     T  S  H  A  A  V  V  G  V  S  L  G  I  P  V  V  V  G  A  A

1681 GACGCAACTTCAAAGATCGCTGACGGCTCAACTTTGACTGTTGACGCACGTCGCGGCGCA 1740
     D  A  T  S  K  I  A  D  G  S  T  L  T  V  D  A  R  R  G  A

1741 ATTTACCAAGGTGAAGTTTCAAACCTGTAATCTAGGATTGCAAGTCTGACAACCACAGAA 1800
     I  Y  Q  G  E  V  S  N  L  *

1801 GAAAGAGTTCAATCAATCGATTGAGCTCTTTTTCTTTGCCAATTTTTTCTCTGCTTTCTT 1860
          ------>      <------

1861 TAATGATTGCCTTGAATATAGCTAAGGCAA 1890
```

III. Nucleotide sequence of the *L. lactis* plasmid pN42.

```
   1  CCTAGGCTTG AAATTGACGC ATAGGCGCAA AGGGAGCGGG CGACAGGGGG TAAAGCACGA
  61  TAAATTCGTT TTTTACAGAC GTTCAGTCCA TGTTGTCATA TTTGTACTCC CGTTTTTAGG
 121  GCTGTTTTAA AAGTATTTTT AGCGGCGATT TGTTAATTAT AGCCCCTATA CAAACATCTT
 181  TTGTAAAAAG CCTTTTTTCT GTTCTTTCAA CAAATCTAAC TTACGTTGAT GAAGAGCGAT
 241  AGTGTCATCT AGCTGTTTTA AAAATGAGCC TATTTTTTTT TGTTCTTCCT GACTAGGTTT
 301  ATAGATTTTA AATGATGAAA ATTTAGAAAT CCAATGACGT TCATGACTTT GAGGTACATA
 361  TTTTATATTC TTCAATGTAT TAAACATAAA ATAGAAATTG TCAGAATTAT CATTCAAACT
 421  AAGTAATTTC ATTGCGGAGC TCTTAATTTT AAAAGGGAAA TCTACATAAT GAGAGTCAGT
 481  TGTAAAATCA TCAAATATAA CAACTGGATT TTCTACGGTA GCATTTTTAA TCCCGCTAAT
 541  TTCATCTGTA TAGCCCAATA AGAAACTCTT GCCTGCTGTT AAAACAGGGG TATTAAAATT
 601  GTCATCGTAC TCTGTAGATT TGACAATATA TTTTGTTGGT TGCTCATAGT TAAATACCTC
 661  CCCCAACTTA CACTGCTCCC ATTCGTCACT AAATCCTTCA AACCGAATAG CTGGATACCC
 721  GCTCTTATAA GCGAACATTT TCTGCAGTAA AGCGCTTTTT AAGCATTTAA GTTGCTGTTT
 781  CTTTTCCTCA TGTAAAGTGA TTGCAGTATC CAATTCAGAG AAGAAGTTAG CAATTCTTTC
 841  TTGTTCAGAC GTAGTTGGAA ACGCAACAGA CTGATTTCCG ACAATATCCG AGTTCAAATT
 901  AACCTGACTT CCCGGCTGAC CATATTTGTT CCAATATGGT TTGAACATAA GAAGCCATTG
 961  AAACATAAAT TCCTTATTAA ATGTTGGGTT GAGAAATATT AAGAATCCAT CGTGAACTCC
1021  TGTGTTAACG TAATTGATCA CTGGACTACC CACAGTAGCA GCAATACTTA ACAATAAATG
1081  TGGTTCTGTG ATAACACGCG TTTTAGATTG ACCAGCTTTT GAAATGTGTT GCGATAAGTG
1141  ATGAATGCGT CCTTTTTGTT CAGTGACATC GGATATTCTT AGCCATCCAA CATTTGAATT
1201  ATCATCGAAC CATTTGGGGT TAGAAATAGG TCTTGGACTC GCTCCACGTA CGATTTCCGC
1261  TTTGTTTTTT AACTTACACT GCTCCCAAGG ATCAGCGAAA CCTTTAAATC TTAATTGCGG
1321  ATATTTAGCT TGTGTATCAT TCATTATTTT TCCTCCGGTT TAATGTCTAA GGCCATTTTA
1381  TCAAATTAAA AATCAGCAAA ACCTATTTTG TGTCTGGTGG AACCAACAAG CGGCTAGAAA
1441  ATATGCTGCC AAACACCCTA AAGAACAAAA TATTGATAAC GAGCATACTT GGCATTAAAC
1501  GCCGTATAAG CTCATTTAAG CCGTTTTAAG TGTTATATGC ATAATTATAT TAAAACTGCT
1561  TTAAAATCGC TTAGAAGCAA GAATAGGCAG CTTGAGTGGC TGAATTGGCG ATGACTGAAC
1621  TAAGGACTAG GCCAAGAAAC TTTTGCACAG TCAACAATTC CCCGGACTAA TTCGGACTTT
1681  TTCTTTCTGG TCAGGTCTCC TAATGGTCAG TAAGGTCAGC CGCTTCAGCG GTCAATCGTG

                                    RBS                 ORF4
1741  TATAATAATA ATCAAGATTG ACAAGAGGAG GGCTGACAAT GGCAAATAGC GCTGGCATGC
                                                    M  A  N  S  A  G  M

1801  TGTCAGTAGG TCAAATAGCT AAAATGCTGA AGACCAACAG ACAGAACATT TACAACGTGC
       L  S  V  G  Q  I  A  K  M  L  K  T  N  R  Q  N  I  Y  N  V

1861  TTAAAGCTGA GCATATTAAA CCTGACGGCT TCAATGACAA GCACTATTCA CTTTACAGCC
       L  K  A  E  H  I  K  P  D  G  F  N  D  K  H  Y  S  L  Y  S
```

```
1921  CGGAAACAAT TCAAGAGATC AAGGCCGCTC TGTCTAAGAA GGCAACGCTG AGAAGTAAGA
       P  E  T   I  Q  E  I   K  A  A   L  S  K   K  A  T  L   R  S  K

1981  AGGTAGTAGC AAAAGAGCAG GCTGAAGAGA TAGCTGACTT GAAGAATCAG CTGTCAGAAC
       K  V  V   A  K  E  Q   A  E  E   I  A  D   L  K  N  Q   L  S  E

2041  AGCAGAGATT GACAACCTGG CTACAGTCTC AGCTGGTTCA ACTTCAAGTA GAGGCTGACA
       Q  Q  R   L  T  T  W   L  Q  S   Q  L  V   Q  L  Q  V   E  A  D

2101  AGCTCAGGAG TCAGAACAGC CAGTTACAGC TAGACAATGC AAAGACTCAG CTCCTTATTG
       K  L  R   S  Q  N  S   Q  L  Q   L  D  N   A  K  T  Q   L  L  I

2161  GCCAGGTTGA CCAGGAGAAG ACAACACTGA AGGCCGAGAA TGACCGACTG AGCGCTGAAA
       G  Q  V   D  Q  E  K   T  T  L   K  A  E   N  D  R  L   S  A  E

2221  ATAACAAACT AGGACAATTA ACCGATAAGG TGCTGAAGGA CGCTCAGAGA GCAGAAGAGG
       N  N  K   L  G  Q  L   T  D  K   V  L  K   D  A  Q  R   A  E  E

2281  ACGCTCAGAA GGCTAAAGCT GATCTAGATA AAGCCCAAGC CCGGCGGGCT GGCTTATGGT
       D  A  Q   K  A  K  A   D  L  D   K  A  Q   A  R  R  A   G  L  W

2341  CTAGAATCAC CAGGAATTAT TAAGAGTGGT ATAGCCGTTA TCTGACTTTG TGAAATTCCT
       S  R  I   T  R  N  Y   -

2401  TATTGGCTCT GTCAGATCAA GCGATTTTAA ACCTATACGA GTTTGTGAAT CCTAGTTTAC

2461  GGAATTGGGC GATAAGGAAG CCCGTCATTG CAAGGATAGA AGGTTAGTTC CAATAAGACA

2521  CATTATGTAA AGTTGTAAGT GGTATACCTG TAATTGATTG ACAGGAACTA TACACGGGCT

2581  AGACACTTGC CAGCATTGAC TGTAGCGGCT TTACAATGAC ACTAGATCTA CACTATAATT

2641  ACAGCGGAAA GAGAAAGGCT GAGCGGTCTC CTAATGGACA ACTACAACTG GCCAGCCCGG

2701  CAACTTTGAG AGCCGTTAAA GAGCTCTCTC AGCATGGTTA GAGTATAGAA AGAGTGCTGA

2761  ACATGGACTT TAAAAAAGGG CTGAAGGGCT TGCAAGATCA GCAGACCCGG CTTGAAGCTA

2821  AACAGGAAGT ACTGTTAGAC ATCATGGCTG AGTTCTGGCC TAAAGTAGCT AAAGAAGGCA

2881  ATGACGTTGC TGAAGCGGTC AAGGTAGAAG ACCTGGCTGA ATGGTTCGCT AAGAACAGCC

2941  GGAAAACTGT TATTTGCGTG TCAGCAAGAC AGAAGACGGC TATGACCTGG CTTTTGAACC

3001  ACAACAGCCT TCAAGAGAAT TGTTATGGTA CGATGATCTT TATTGGCGGC TGGGTAAAAC

                                                      RBS      ORF5
3061  AGCTGACCAA CTCAAAACGT AAATCTAAGG TCAAGACGCT AGAGGAAATT ATCTAATGGC
                                                                  M

3121  GGTTTACAAA GAATGGACTG ATTCAGATCA TTTAGAGTTA GTCAAAAATT GGAAATTACA
      A  V  Y  K   E  W  T   D  S  D   H  L  E  L   V  K  N   W  K  L

3181  CGGGCTGACT AACGTTGAGA TAGCTCAAAG AATAGGCATT GCTGAGAAGA CTTTGTACGT
      H  G  L  T   N  V  E   I  A  Q   R  I  G  I   A  E  K   T  L  Y

3241  ATGGTTGAAG AAGTCTCCTA AGCTGAAGAA GGCCATTAGA GGCGGCAAGG ATATTGCCAG
      V  W  L  K   K  S  P   K  L  K   K  A  I  R   G  G  K   D  I  A

3301  GGCTAGGGCT GAGAATGCAC TGTATGAGCT TGCTCTTAAT GGCGATAGGC AAGCCCTTTT
      R  A  R  A   E  N  A   L  Y  E   L  A  L  N   G  D  R   Q  A  L

3361  CTTTTGGCTC AAAAACAACT ACAGAGAACG CTACTCAGAC AAGCCGTTAA GCCCGGCTGA
      F  F  W  L   K  N  N   Y  R  E   R  Y  S  D   K  P  L   S  P  A

3421  AGCCGATTTG ATGAGTCAGA AGGCAAGGCT GGCCAAATTA CAGGCTGACC TGGCTGAGGC
      E  A  D  L   M  S  Q   K  A  R   L  A  K  L   Q  A  D   L  A  E

3481  TCAGCTGAAG GCCATTAAGG AAGACCAGGG AGACCAAGCA ACGCAATTAA ACAACCTGTT
      A  Q  L  K   A  I  K   E  D  Q   G  D  Q  A   T  Q  L   N  N  L
```

```
3541  AGACAGTCTG AAGGAAGCCG TGTTAGATGA GGGAATTAGC CCCGATAACA TCGTTCCTAC
      L  D  S  L   K  E  A   V  L  D   E  G  I  S   P  D  N   I  V  P

3601  TGGCAACGGC TTAATTATCG ATGATATTCC TGACTCTTAG GTTTACACGA CATTGACAGT
      T  G  N  G   L  I  I   D  D  I   P  D  S  -

3661  GTAAACACAA GATAGCGGAA AATCTTCTGA TTATTATATT TACAAGCACT GTATATTGTG

3721  CTATTCTAAG ATGTGCTAAA CGGATTTGGG GAATGCAACT AACTGCTGTA AGGTATCAAC

3781  TTTTTTTGTT GCGCTCTTTA ATTCTTTAGC AAAAAGCTAG ATATCAAAAA AGAGCGAGAC

3841  CGGGTATTGC TTCACGGGTT CGCTCTTATT TTTTTATCTG GCTAGTTGCC TACTGGTACT

3901  ATGCTGACAC CCTAGCGGCA TGTTTGCGGT ATTGCACTAC AGCGGCAACA ATGGTAAAAA

3961  TAATAATAGG TAACAAAAAA GCCTTTAGTA CTGGCAATAC TAGAGGCGGG CTGTGTTTAG

4021  CTCTGGCAAA GCTTAACACG GTTAGAATTA TATTCCGTAC CACATATGAT ACGTTTAAAC

4081  GTAACACTCT GTCAAGGAGA ACATATCACC TTAAGGGTAC ATATAGTAGT TTTCTTCTAA

4141  CATTATGTTG TAAAAACATA ACATTTTGTA GACAAACACT ATACTTCTAT GACTCTAACC

4201  ATGTTTAAGA CAGGCCAGGC TAACACCTAT TGGCCTGTTT TTTGTTGCCA AAATTTCAAA

4261  AGAAAGGCGG TAACAGCCGT GATTAAACAA CAAAACATTG ATGTTAGAGC GGCTATTAAA

4321  GCTTCTGGTC TGAAGCAATA TGAGGTAGCT ACTTTGATGA ATGTTTCAGC TAGCTATCTC

4381  AGCCAGCTTT TACTTCAACC ATTGTCAGAA GGCCATAAGA AGCGCATTAT GGCGGCGATT

4441  AAACAAGGCG AGTCATTGAA GGGAGAACAA GAATAATGAT GAGCTTAGAA GAACGTGAGC

4501  AAGAAATTGA AAAGGTAGTA CGCATTGCTG AAGCTGACTT CAACAACGCT TGTCAATTGC

4561  ATGCTATCAA CAAGGAAGAT GTTATTAAGA ACCATGCTTA CAAGTATGCT GAAGTGCTGA

4621  GGCTTCAGGA ATTGCTGGCA TTGAACAAGA CCATTAGGGA CGGTCTGAAC GGCATTGAAA

                                      RBS          ORF3
4681  TGTCAGTAGA TCTCATTGAG TAGCGGGGAG ACCCGCCATG AACAACAGTG AAAAAAACTC
                                            M   N  N  S   E  K  N

4741  TCTAATGGCT GAACCGTATA ACTCAGACCG CAACGCCATT GACAGACTCA GAATCAACCA
      S  L  M  A   E  P  Y   N  S  D   R  N  A  I   D  R  L   R  I  N

4801  GAAGGCCTTA CAGGCGGGCT CTGTCAAGCG TGAAGAGGGC TACAACTCAG AGGGCTTAGA
      Q  K  A  L   Q  A  G   S  V  K   R  E  E  G   Y  N  S   E  G  L

4861  AATGGTCTCC TACACGGCTT ATAAGAGCGG CATTCAGTAT GTCATTTCTT CAGAAGCTGA
      E  M  V  S   Y  T  A   Y  K  S   G  I  Q  Y   V  I  S   S  E  A

4921  AGGCGGCAAA ATGGTTATTA ACGAGACCTT CAGCAAGGTT CAACATCTAC TAATTGCCAG
      E  G  G  K   M  V  I   N  E  T   F  S  K  V   Q  H  L   L  I  A

4981  CTGGTATAGC CAGCCAGACA GAGCCAGCAA TTTCAGAATA CAGCTGACCT TTAAAGAGAT
      S  W  Y  S   Q  P  D   R  A  S   N  F  R  I   Q  L  T   F  K  E

5041  CTCAGAGGCG CTAGGAGTCA GCAGAAGCCA GGCTACAGCG CTCAGAAAGC AGCTGAGAGA
      I  S  E  A   L  G  V   S  R  S   Q  A  T  A   L  R  K   Q  L  R

5101  GCTAATTACA CAGCTAGTAC GTTGTACTTT TGTTAACAGC AATAAAGACG GCATAGACGC
      E  L  I  T   Q  L  V   R  C  T   F  V  N  S   N  K  D   G  I  D

5161  TGTCAATCTC TTTGCAGCTG GCAACTACAG TAAAGGGAAG CTGACAATGT GGTTAACTCC
      A  V  N  L   F  A  A   G  N  Y   S  K  G  K   L  T  M   W  L  T

5221  TAACATGGCT GAGCGGCTTC TGTCAGAAGA ATCATCTACG GAATATTTTC CGTTATCTTT
      P  N  M  A   E  R  L   L  S  E   E  S  S  T   E  Y  F   P  L  S

5281  ACTGAAGCTG AAAGGGACAG CCTATTATTT AGCCTTAAAG GTCATGCACA ACGCAAACAT
```

```
     L  L  K  L   K  G  T   A  Y  Y   L  A  L  K   V  M  H   N  A  N

5341 TAATGCACGC TGGCATGCTG ACAGAGTTGA CAGATTGGGC TTAGAAAACA CGCTGAAGGC
     I  N  A  R   W  H  A   D  R  V   D  R  L  G   L  E  N   T  L  K

5401 CTTGCCTACA CTCCCCGACC CGGTAAAACT CTCTAAAGGC AACAGCAGAA GCCTATACCT
     A  L  P  T   L  P  D   P  V  K   L  S  K  G   N  S  R   S  L  Y

5461 AAAAATCTTA ACTCCCCTGG CTAAAGCTAT TGAAGAGCTT GAAGCCGTCA CTGGCATTGT
     L  K  I  L   T  P  L   A  K  A   I  E  E  L   E  A  V   T  G  I

5521 CGTTAGACCT AGCCAGCCAC TAAAGGGAAT GAAGACGAAA GATCTGTCTA AAGTCACTTT
     V  V  R  P   S  Q  P   L  K  G   M  K  T  K   D  L  S   K  V  T

5581 GAATGTCATT GATTGGGGAC AGGTTGATAT AGCCGAATTG ACCAGAAATA AGAGAAAACG
     L  N  V  I   D  W  G   Q  V  D   I  A  E  L   T  R  N   K  R  K

5641 CTTGCGAAAA AATAATGTTC GTGAGGACTA AAACTATATT TGTCCTAATT CGTATGTAGG
     R  L  R  K   N  N  V   R  E  D   -
```

ori

```
5701 TAATTATGGT CGCAAATGTA GGTAATTATG GTCGCAAATG TAGGTAATTA TGGTCGCATT
```

```
5761 GTGAAATTTA GGCAAGTGCC TTGAGGCATT GAGCCAGTAA GGAGTAAGCG CATTTTTTTA

5821 AAAAGCTTCA CTTGCTAATA GTTTAATAGT ATTAAAAGCA ACGGCTCAGC TTGACGCTGG

5881 CCTTGCTTGA AAATTGAAAA AAGATGAAAC AGCCAGGGAG AGCAGAGGCT TCTACTGGCC
```

RBS *ORF2*

```
5941 TGTTTTT**AGA AGAAGG**TATC TAGCATGAAC AATAACTTAG TTAAACCAAC AGATTTAAAG
                               M  N   N  N  L   V  K  P   T  D  L  K

6001 GGCTTGGTCT CTTTACCGGA ATACATTGCC AGCGTGGTTA GCATGGACTC TAAAGGCTTC
       G  L  V   S  L  P   E  Y  I  A   S  V  V   S  M  D   S  K  G  F

6061 TTTAGCTGTC TCAATCCGAA CCACCCGGAC AATCACCCTA GCATGTGTTT AGACCCTAAC
       F  S  C   L  N  P   N  H  P  D   N  H  P   S  M  C   L  D  P  N

6121 CACCCGCAAT ATGTTCATTG CTTCAGTTGC GGCGTGTCCT ATGATCTGTT TGATTGTTGG
       H  P  Q   Y  V  H   C  F  S  C   G  V  S   Y  D  L   F  D  C  W

6181 GCGCTGATTA ATGACGGCGT GACAGAGACC AAGAAGAATA GCGCTGGCAA GGAAAAGCCA
       A  L  I   N  D  G   V  T  E  T   K  K  N   S  A  G   K  E  K  P

6241 GTCTATAACT TCAATGCTGT AGCTTCAGAG ATTGCTGACC ATTACGGCTA TGCTCTTATT
       V  Y  N   F  N  A   V  A  S  E   I  A  D   H  Y  G   Y  A  L  I

6301 GGCGACCCGG CAAATGATCT CTATTCGGTA GAACCACCCT TGCCAGAACC ACCAGCAGAA
       G  D  P   A  N  D   L  Y  S  V   E  P  P   L  P  E   P  P  A  E

6361 CCAGCTCAGA CCAGCACCAA TTTTAGAGAG CAATTAGAAG ATTGGCATGC TAACTTGAAT
       P  A  Q   T  S  T   N  F  R  E   Q  L  E   D  W  H   A  N  L  N

6421 CAGACTGACT ATCTTCAGAA GCGGGGAATC ACTCAGACAA CAGCAGAGAT TTTCAATTTA
       Q  T  D   Y  L  Q   K  R  G  I   T  Q  T   T  A  E   I  F  N  L

6481 GGCTACTCCC CGTTGACCAA CAGCATTATT ATCCCTTACG GTCAGGACGG CTATTACGTT
       G  Y  S   P  L  T   N  S  I  I   I  P  Y   G  Q  D   G  Y  Y  V

6541 CAGAGGGCGC TGAATCCAAT TGAGAAGCGT GACCGCTACC GCTTCCCTAT TGGCCAGGCT
       Q  R  A   L  N  P   I  E  K  R   D  R  Y   R  F  P   I  G  Q  A

6601 AGAGCCTACA ACATTGAAGC ATTGGCTAAA TGCAAGACGG TATTCATCGT TGAAGGCCAG
       R  A  Y   N  I  E   A  L  A  K   C  K  T   V  F  I   V  E  G  Q

6661 TTTGACGCTC TGTCAATCAT GCAAGAATCC GATGTAGGAG CTGTAGCAAC TTCAACCAGC
       F  D  A   L  S  I   M  Q  E  S   D  V  G   A  V  A   T  S  T  S

6721 CAGACTCGGC TTATTGTCAA GGCCTTACAG AAGTTCAAAG AGCAAGACCC AACAATTAAC
       Q  T  R   L  I  V   K  A  L  Q   K  F  K   E  Q  D   P  T  I  N
```

```
6781  CCGACTATCA TTCTCAGCAT GGACAACGAC AGAGCAGGCC AGAAGGCGAA TAGAGCCCTT
        P  T  I   I  L  S   M  D  N  D   R  A  G   Q  K  A   N  R  A  L

6841  CAGAGGGACT TAGAAGCCCT GGGCTTTACT TGCTATGTCA ACCCGGTTAA CGGCGACTAC
        Q  R  D   L  E  A   L  G  F  T   C  Y  V   N  P  V   N  G  D  Y

6901  AAGGACGCTA ACGAGTTCCT GGTAAAGGAT AGAGAGGGCT TCAGACAGAA ACTTCAGCAC
        K  D  A   N  E  F   L  V  K  D   R  E  G   F  R  Q   K  L  Q  H

6961  GTCATCAATC AGCCCGACAA TTGGCTTGAC AATTACTATG CTGACATCAA AAAACGCCAT
        V  I  N   Q  P  D   N  W  L  D   N  Y  Y   A  D  I   K  K  R  H

7021  GACTACCCGG ACAATATCCC TACTGGCTTC AAGAATTTAG ATGATGAGCT TGACGGCGGT
        D  Y  P   D  N  I   P  T  G  F   K  N  L   D  D  E   L  D  G  G

7081  CTTCAGCCTA AACTGTATGT TTTAGGCGCT GTCAGTTCGC TAGGGAAAAC GACTTTTGCC
        L  Q  P   K  L  Y   V  L  G  A   V  S  S   L  G  K   T  T  F  A

7141  TTGAATATTG CTGACAACCT GGCTAAACAG GGGAGACATG TTTTCTTCTT CAGCATGGAA
        L  N  I   A  D  N   L  A  K  Q   G  R  H   V  F  F   F  S  M  E

7201  TCTAGCAAGA GAGAAGTGAC GGACAAGCTT TTAAGCCGGG CTAGCTGTCT CTCTAACGGC
        S  S  K   R  E  V   T  D  K  L   L  S  R   A  S  C   L  S  N  G

7261  CATAAATGGA CTCAGCTTCA AGTCAGCCGG GGAGAATGGT TGAACAATGC TGAGGACAAA
        H  K  W   T  Q  L   Q  V  S  R   G  E  W   L  N  N   A  E  D  K

7321  GAAGAGTTTG ACGGCCTGTT TAAAGCCTTC AGCCGTTACC AGCACTTCTT ACATATCTAT
        E  E  F   D  G  L   F  K  A  F   S  R  Y   Q  H  F   L  H  I  Y

7381  GACAATAGAG TTAAGGCAAG TCAGGTAAAA GACCTGGTCA ATAGTTGGCT TGACAACCAC
        D  N  R   V  K  A   S  Q  V  K   D  L  V   N  S  W   L  D  N  H

7441  CCGGACGAGA AGAAGCCGCT TGTAGTCGTT GACTATCTTC AGATCTTGCA AGCTGAGCAG
        P  D  E   K  K  P   L  V  V  V   D  Y  L   Q  I  L   Q  A  E  Q

7501  GACAATGTGA CAGATAAGGC GAAAGTGACG GACAGCGTGA GTGTTCTCTC AGAGCTGACT
        D  N  V   T  D  K   A  K  V  T   D  S  V   S  V  L   S  E  L  T

7561  AAACAGGCTG AAGTCCCTGT TCTGGTCATC TCATCATTGA ACCGGGCTTC CTACTGGCAA
        K  Q  A   E  V  P   V  L  V  I   S  S  L   N  R  A   S  Y  W  Q

7621  GACGTAAGTT TTGAATCCTT CAAGGAATCC GGGGAAATTG AGTACTCAGC AGACGTTATG
        D  V  S   F  E  S   F  K  E  S   G  E  I   E  Y  S   A  D  V  M

7681  TTAGGATTAG AGTTCGCTCA TCGTGAAGAA TACATTACAG TTAAGGGCAA CGGCCATGTT
        L  G  L   E  F  A   H  R  E  E   Y  I  T   V  K  G   N  G  H  V

7741  GAATTGAACA AAGAGAAGTT TGACCAGCGG AAACAGGAAG TCCTAGACGG GTTGAAATGG
        E  L  N   K  E  K   F  D  Q  R   K  Q  E   V  L  D   G  L  K  W

7801  TCATTCTGAA GAATCGAACT GGCAAGACAG GCGGTCATAT CTTCTTCAAG TACAACGCCA
        S  F  -

7861  TGTTTAACAG CTACCAGGCA TGCACTGAGC AAGAGGCGGC AATACCCAAT AACTTTAATA

7921  AGTTGTTTCA TAGCAAGGAA GTAGGCAAGC CAATTGAAGC GGCTGTGCGT GATTACACGG

7981  TAGACCCGGT AACAGGCCTG GCAACAGAGA AGAAGCCCGA TAAATAGAAC TGAAGAAGCT

8041  GGCCAGGAAT GGCTGGCTTT TGTTTTGCCT TCAGACGCTC TCAGAAGCTC ATAGAGCCCC
```

Complemented nucleotide sequence of ORF-1.

```
                                                                  RBS       ORF1
1400© TTTGCTGATT TTTAATTTGA TAAAATGGCC TTAGACATTA AACCGGAGGA AAAATAATGA
                                                                    M

1340© ATGATACACA AGCTAAATAT CCGCAATTAA GATTTAAAGG TTTCGCTGAT CCTTGGGAGC
      N  D  T   Q  A  K  Y   P  Q  L   R  F  K   G  F  A  D   P  W  E

1280© AGTGTAAGTT AAAAAACAAA GCGGAAATCG TACGTGGAGC GAGTCCAAGA CCTATTTCTA
      Q  C  K   L  K  N  K   A  E  I   V  R  G   A  S  P  R   P  I  S

1220© ACCCCAAATG GTTCGATGAT AATTCAAATG TTGGATGGCT AAGAATATCC GATGTCACTG
      N  P  K   W  F  D  D   N  S  N   V  G  W   L  R  I  S   D  V  T

1160© AACAAAAAGG ACGCATTCAT CACTTATCGC AACACATTTC AAAAGCTGGT CAATCTAAAA
      E  Q  K   G  R  I  H   H  L  S   Q  H  I   S  K  A  G   Q  S  K

1100© CGCGTGTTAT CACAGAACCA CATTTATTGT TAAGTATTGC TGCTACTGTG GGTAGTCCAG
      T  R  V   I  T  E  P   H  L  L   L  S  I   A  A  T  V   G  S  P

1040© TGATCAATTA CGTTAACACA GGAGTTCACG ATGGATTCTT AATATTTCTC AACCCAACAT
      V  I  N   Y  V  N  T   G  V  H   D  G  F   L  I  F  L   N  P  T

 980© TTAATAAGGA ATTTATGTTT CAATGGCTTC TTATGTTCAA ACCATATTGG AACAAATATG
      F  N  K   E  F  M  F   Q  W  L   L  M  F   K  P  Y  W   N  K  Y

 920© GTCAGCCGGG AAGTCAGGTT AATTTGAACT CGGATATTGT CGGAAATCAG TCTGTTGCGT
      G  Q  P   G  S  Q  V   N  L  N   S  D  I   V  G  N  Q   S  V  A

 860© TTCCAACTAC GTCTGAACAA GAAAGAATTG CTAACTTCTT CTCTGAATTG GATACTGCAA
      F  P  T   T  S  E  Q   E  R  I   A  N  F   F  S  E  L   D  T  A

 800© TCACTTTACA TGAGGAAAAG AAACAGCAAC TTAAATGCTT AAAAAGCGCT TTACTGCAGA
      I  T  L   H  E  E  K   K  Q  Q   L  K  C   L  K  S  A   L  L  Q

 740© AAATGTTCGC TTATAAGAGC GGGTATCCAG CTATTCGGTT TGAAGGATTT AGTGACGAAT
      K  M  F   A  Y  K  S   G  Y  P   A  I  R   F  E  G  F   S  D  E

 680© GGGAGCAGTG TAAGTTGGGG GAGGTATTTA ACTATGAGCA ACCAACAAAA TATATTGTCA
      W  E  Q   C  K  L  G   E  V  F   N  Y  E   Q  P  T  K   Y  I  V

 620© AATCTACAGA GTACGATGAC AATTTTAATA CCCCTGTTTT AACAGCAGGC AAGAGTTTCT
      K  S  T   E  Y  D  D   N  F  N   T  P  V   L  T  A  G   K  S  F

 560© TATTGGGCTA TACAGATGAA ATTAGCGGGA TTAAAAATGC TACCGTAGAA AATCCAGTTG
      L  L  G   Y  T  D  E   I  S  G   I  K  N   A  T  V  E   N  P  V

 500© TTATATTTGA TGATTTTACA ACTGACTCTC ATTATGTAGA TTTCCCTTTT AAAATTAAGA
      V  I  F   D  D  F  T   T  D  S   H  Y  V   D  F  P  F   K  I  K

 440© GCTCCGCAAT GAAATTACTT AGTTTGAATG ATAATTCTGA CAATTTCTAT TTTATGTTTA
      S  S  A   M  K  L  L   S  L  N   D  N  S   D  N  F  Y   F  M  F

 380© ATACATTGAA GAATATAAAA TATGTACCTC AAAGTCATGA ACGTCATTGG ATTTCTAAAT
      N  T  L   K  N  I  K   Y  V  P   Q  S  H   E  R  H  W   I  S  K

 320© TTTCATCATT TAAAATCTAT AAACCTAGTC AGGAAGAACA AAAAAAAATA GGCTCATTTT
      F  S  S   F  K  I  Y   K  P  S   Q  E  E   Q  K  K  I   G  S  F

 260© TAAAACAGCT AGATGACACT ATCGCTCTTC ATCAACGTAA GTTAGATTTG TTGAAAGAAC
      L  K  Q   L  D  D  T   I  A  L   H  Q  R   K  L  D  L   L  K  E

 200© AGAAAAAAGG CTTTTTACAA AAGATGTTTG TATAGGGGCT ATAATTAACA AATCGCCGCT
      Q  K  K   G  F  L  Q   K  M  F   V  -
```

IV. Nucleotide sequence of the *L. lactis* plasmid pJBL2.

```
   1  ATTCGTTTTT TACAATTTTG CTATACTCCC ACCTATTTAT ACGGGCTGAA CCTGCTTTAA
  61  TCAAGCCAAG CGTCTTACAG ACGTTCAGTC TATGTTGTCA TATTTGTACT CCCGTTTTTA
 121  GGGCTGTTTT AAAAGTATTT TTTAGGGGCG ATTTTATAAT TATAGACCCT AAACGAACAT
 181  CTTTTGTAAA AAGCCTTTTT TCTGTTCTTT CAACAAATCT AACTTACGTT GATGAAGAGC
 241  GATAGTTTTG TCGAGTTGTT TGAAGAATGA ACCTATTTTT TGTTGTTCAG TTATTGTTGG
 301  CAACGAAATT GGCTGTTTAC TTAATTCCGT TTGATTAATG CTTGCTTGGC TGACTGCGGG
 361  ATTCGCGTGC GAAACTGCCC AATTTTTCAT TGGCGTCGTG TTTAGTCTCT GAAATAAAAA
 421  TTCCGAATCA ACATCATTGC GCATTTGGAA TCGCAAAAGA TTAATTCCGT GATAAATCCC
 481  TGTCGTATCA ACCAAAGCCA CCTTTCCAAT GTGACTCAAT GAGTTAATAT TGCTGTAAAG
 541  AATATCGCCC AAATTTAGAA GATACTTCTG GTCTGGTTTT TTATTTGTGA ATCCTAACCT
 601  CGCAACATTC AACGTTCCTT GCGAAATACT TTCGATGCGT GTAAGTCTAT ATTCGCCATC
 661  TTGAAGTGTT GCGTCACCCG TGACTCCACT TGATATACTA ACCGCATCAC CCAACTTACG
 721  CTCTTCCCAC TCGTCACTAA ATCCTTCAAA TCTAACAACA GGATACCCAC TCTTGTCAGC
 781  GAACATCTTC TGCAATAAAG CGCTTTTCAA GCATTCAAGT TGACGTTTCT TTTCCTCATG
 841  TAAAGTGATG GCATGATCCA TATAGGAAAG TATCTTTCCA ACCTTATTTT GTTCATGCTC
 901  CTTAGGGAGA TTTATTACCG CATTATTTAT TTCTAATGAA TTAATGCTCT CAAATGTTGA
 961  ACCAGTACTA TATTTAGTCC AATAACCTAC TGTTTTCATT CGTTTAAGCA GTTGAAAAAT
1021  AAATTCATTA CCTTTGATGG CGGCTACTCC ACGACCGATA ACCACATCAT AGCTTGTTTT
1081  TCCAATATCG CCTACAGGTG CTCTAACGCT TAAGATTAAA TCACCTTTGT CTGCGATTTT
1141  TGTTATTTCC GTAGTCCAAA TTCTAGGATG TACTTGACCA TCCTTCATAT CTGCATTTCC
1201  TTGAACGAGA ATGTGATCTT TTGGATTATC AGTGTAATTT TTAGAATTTG GTGATTGACC
1261  CATGGTGATT TTTGCCACGT CCCCCAACTT ACACTGCTCC CAAGGATCAG CGAAACCTTT
1321  AAATCTTAAT TGCGGATATT TAGCTTGTGT ATCATTCATT ATTTTTTCTC CGGTTTAATG
1381  TCTAAGGCCA TTTTATCAAA TTAAGAATCA GCAAAACCTA TTTTGTGTCT GGTGGAACCA
1441  ACAAGCGGCT AGAAAATATG CTGCCAAACA CCCTAAAGAA CAAAATATTG ATAACGAGCA
1501  TACTTGGCAT TAAACGGCGT ATAAGCTCAT TTAAGCCGTT TTAAGTGTTA TATGCATAAT
1561  TATATTAAAG CTGCTTTAAA ATCGCTTAGA AGTAAGAATA GGCATCTTGA GTGGCTGAAT
1621  TGGCGATGAC TGAACTAAGG ACTAGGCCAA GAAACTTTTG CGCAGTCAAC AATTCCCCGG
1681  ACTGGTTCGG AATTTTGCTT CCTGGCCAGG CCTTCTAATG GTCGGCAAGG TCAGCTGCTT
                                                 RBS              ORF4
1741  CAGCGGTCAA TCGTGTATAA TATAGTCAAG ATTGACAAGA GGAGGGCTGA CAATGGAAAA
                                                                  M  E
1801  TAGCGCTGGC ATGCTGTCAG TTGGTCAAAT AGCTAAAATG CTGAAGACCA ACAGACAGAA
      N  S  A  G   M  L  S   V  G  Q   I  A  K  M   L  K  T   N  R  Q
1861  CATTTACAAC GTGCTTAAAG CTGAGCATAT TAAACCGGAC GGCTTCAATG ACAAGCACTA
      N  I  Y  N   V  L  K   A  E  H   I  K  P  D   G  F  N   D  K  H
```

```
1921  TTCACTTTAC AGCCCGGAAA CGATTCAAGA GATCAAGGCG GCTCTGTCTA AGAAGGCAAC
      Y  S  L  Y   S  P  E   T  I  Q   E  I  K  A   A  L  S   K  K  A

1981  GCTGAGAAGT AAGAAGGTAG TAGCAAAAGA GCAGGCTGAA GAGATAGCTG ACTTGAAGAA
      T  L  R  S   K  K  V   V  A  K   E  Q  A  E   E  I  A   D  L  K

2041  TCAGCTGTCA GAACAGCAGA GATTGACAAC CTGGCTACAG TCTCAGCTGG TTCAACTTCA
      N  Q  L  S   E  Q  Q   R  L  T   T  W  L  Q   S  Q  L   V  Q  L

2101  AGTAGAGGCT GACAAGCTAA GGAGTCAGAA CAGCCAGTTA CAGCTAGACA ACGCAAAGAC
      Q  V  E  A   D  K  L   R  S  Q   N  S  Q  L   Q  L  D   N  A  K

2161  TCAGCTCCTT ATTGGCCAGG TTGACCAGGA GAAGACAACA CTGAAGGCCG AAAATGACCG
      T  Q  L  L   I  G  Q   V  D  Q   E  K  T  T   L  K  A   E  N  D

2221  ACTGAGCGCT GAAAATGAAA AACTAGGACA ATTAACCGAT AAGGTGCTGA AGGACGCTCA
      R  L  S  A   E  N  E   K  L  G   Q  L  T  D   K  V  L   K  D  A

2281  GAGAGCAGAA GAGGACGCTC AGAAGGCTAA AGCTGATCTA GATAAAGCGG CGGGCTGGCT
      Q  R  A  E   E  D  A   Q  K  A   K  A  D  L   D  K  A   A  G  W

2341  TATGGTCTAG AATCACCAGG AATTATTAAG AGTGGTATAG CCGTTATCTG ACTTTGTGAA
      L  M  V  -

2401  ATTCCTTATT GGGTCTGTCA GATCAAGCGA TTTTAAACCT ATACGAGTTT GTGAATCCTA

2461  GTTTACGGAA TTAGGCGACA AGGAAGCCCG TCATTGCAAG GATAGAAGGT TAGTTCCAAT

2521  AAGACACATT ATGTAAAGTT GCAAGTGGTA TAAGGTCGCA AAACGCTACC TTAGCTCATA

2581  GGAGGGTAAA AAATAGACAG CCGGGAAAAG GACTTCCAAC AAGCGCTCTT ACTGGACTTG

2641  GTGGCTGGAA TTTTCGGCCG TATGAAGACC AGGACAAGCA AGACGGGGCA ACCGTTACAG

2701  TGCCATGAGC TATTCAGACT GCCTGGAAGA AATTATAAAC TGCTACTAAA AAAGACTACC

2761  CGCAATAGGT AGCCTTTTTA TGTGCCTGGA AAAACCTTAA AAAACCATAG ATTTTCGTTT

2821  ACCACGTGTG GGTAAAACCT GTCAAAACTT GACATTTATA GCAACCAGAC TTTACGCCTG

2881  TAACTGGAGT TACGGACATG GTTTTCCCTA CCTGACGCCT TAAGTTAAGC TTAGACCGCT

2941  TGCCTGCAAG GGAGCGGCGA TTCGTGGTTT CGCACAAGTG GAGAAAAAAC TGGTAAATAC

3001  GGGTATGAAG TGGAGTTTCT AAGGCATCTA TAAAAGAGAG CCTAACAAAA CCGGACATTT

3061  TGTTAACCTT AATATTTCTT AATATCTGCT GCGGGGCAAG GGAGCGGGCG GTTTGTGTCT

3121  TTCAGCAAGT GGCAAAAAAA GGTAAATACT GGTACATCTA AAAGACAGTC CCCAATAGTG

3181  GGAGCTGGAG TGAAGATGAC TCCCTAAAGC AGGGACCTCA CCGAAGACAT TCCCTTAAAA

3241  GAGTGAGCGA TTCGCTACCA CCCTTCCCTT CCTAACATGT TGGGAAAGGC GATTGGAAAG

3301  TTGTAAGTGG TATACCTGTA ATTGGTTGAC AGGTACTAGA CATAGGCTAG ACACTTGCCA

                  RBS          ORF5
3361  GCATTGACTG T**AGCGG**TTTT ACAATGACAC TAGATCTACA CTATAACTAC AGTAGAAAGG
                                M  T   L  D  L   H  Y  N  Y   S  R  K

3421  CTGAGCGGTC TCCTAATGTA CGACTGACTC AGACCGATTT AGAGCTGGTC AAGAATTGGA
       A  E  R   S  P  N  V   R  L  T   Q  T  D   L  E  L  V   K  N  W


3481  AATTACACGG GCTGACTAAC GTTGAGATAG CTCAGAAGAT AGGTATTGCT GAGAAGACCT
       K  L  H   G  L  T  N   V  E  I   A  Q  K   I  G  I  A   E  K  T

3541  TGTATGTTTG GTTGAAGAAG TCTCCTAAGC TGAAGAAGGC CATTAGAGGC GGGCAAGATA
       L  Y  V   W  L  K  K   S  P  K   L  K  K   A  I  R  G   G  Q  D

3601  TTGCCAGGGC TAGGGCTGAG AATGCACTGT ATGAGCTTGC TCTTAATGGC GATAGACAAG
       I  A  R   A  R  A  E   N  A  L   Y  E  L   A  L  N  G   D  R  Q
```

```
3661  CCCTTTTCTT CTGGCTCAAA AACAACTACA GAGAACGCTA CTCAGACAAG CCGTTAAGTT
        A  L  F   F  W  L  K   N  N  Y   R  E  R   Y  S  D  K   P  L  S

3721  AAGAGCTGAA GCCGATTTGA TGAGTCAGAA GGCAAGACTG GCCAAATTAC AGGCTGACCT
       -

3781  GGCTGAGGCT CAGCTGAAGG CCATTAAGGA AGACCAGGGA GACCAAGCAA CGAAATTAAA

3841  CAACCTGTTA GACAGTCTAA AGGAAGCCGT GTTAGGTGAG GGAATCAGCC CCAATAACAT

3901  CGTTCCTACT GGCAACGGCT TAATTATCGA TGATACTTCT GACTCTTATG TTTACACGAC

3961  ATTGACAGTG TAAACACAAG ATAGCGGAAA ATCTTCTGAT TATTATATTT ACAAGTACTG

4021  TATATTGTGT TATTCTAAGA ATGTGAAGAG GAATTAAAGA ACGGAGTTAG GGAAAAGGTA

4081  TCAACTTTTT TTGTTTCGCT CTTTAATACT CTTTAGCAAA AAGCTAGATA TCAAAAAAGA

4141  GCGAGCCGGC ATTAATTCAC GGGTTCGCTC TTATTTTTTT ATCGGTTGCG GCTTGCGTAC

4201  GGTGCTTTTG CGGTCAATAG ACGGTTTAAG TGTATTGCAT TGCACTAGCA AGCATGGTAG

4261  AAATAATAAT AGACAACAAA AAAAGCCTTT AGTGCTGGTA ACACTAGAGG CGGGCTGTGT

4321  TCAGTGCTGG TAACGCTAAA CACGGTTAAA GCCATATTTT TCTTTACAGA ATATGATACT

4381  TTTAAACTTT ACATTCTGTC AAGGGTGACG TATCACTTCT AAAGTACATA TAGTAGTTTT

4441  CTTCTAACAT TATGTTGTAA AAACATAACA TTTTGTAGAC AAACACTATA CTTCTATGAC

4501  TCTAACCATG TTTAAGACAG GCCAGGCTAA CACCCTACTG GCCTGTTTTT TGTTGCCAAA

4561  ATTTCAAAAG AAAGGCGGTA ACAACCGTGA TTAAACAACA AAACAGCGAT GTTAGAGCGG

4621  CTATTAAAGC TTCTGGCTTG AAGCAATATG AGGTAGCTAC TTTGATGCAT GTTTCAGCTA

4681  GCTATCTCAG CCAGCTTCTA CTTCAACCAT TGTCAGAAGG CCATAAGAAA CGCGTTATGG

4741  CGGCGATTAA ACAAGGTGAG TCATTGAAGG GAGAACAAGA ATAATGATGA GCTTAGAAGA

4801  GCGTGAGCGT GAGATTGAAA AGGTAGTACG CATTGCTGAA GCTGACTTCA ACAACGCTTG

4861  TCAATTGCAT GCTATCAACA AGGAAGACGT TATTAAGAAC CAGGCTTACA AGTATGCTGA

4921  AGTGCTGAAG CTTCAGGAAT TGCTGGCATT GAACAAGACC ATAAGACGGG CTGAATGGTA

                                          RBS          ORF3
4981  TTGAAATGTC AGTAGATCTC ATTGAGTAC**G** **GGGAGA**CCCG CTATGAACAA TAGTGAAAAA
                                                 M  N  N  S  E  K

5041  ACTTCTCTAA TGGCTGAGCC GTATAACTCA GACCGCAACG CCATTGACAG ACTCAGAATC
        T  S  L   M  A  E   P  Y  N  S   D  R  N   A  I  D   R  L  R  I

5101  AACCAGAAGG CCTTACAGGC GGGCTCTGTC AAGCGTGAAG AGACCTATAT AGCCGATGAA
        N  Q  K   A  L  Q   A  G  S  V   K  R  E   E  T  Y   I  A  D  E

5161  TCCGGCAACT CAGAGGCTTC AGAATGGCCT TCCTACACGG CCAATAAGAG CGGCGTTCAG
        S  G  N   S  E  A   S  E  W  P   S  Y  T   A  N  K   S  G  V  Q

5221  TACGCTATAC CAGTAGAGGC TGAGCCCGAC AAAATGGTAA TAAATGAGAC CTTCAGCAAG
        Y  A  I   P  V  E   A  E  P  D   K  M  V   I  N  E   T  F  S  K

5281  GTTCAACATC TACTAATTGC CAGCTGGTAT AGCCAGCCAG ACAGAACAAG TAACTTCAGA
        V  Q  H   L  L  I   A  S  W  Y   S  Q  P   D  R  T   S  N  F  R

5341  ATACAGCTGA CCTTCAAAGA GATCTCAGAG GCGCTAGGAT ACAGCAGAAG CCAGGCTACA
        I  Q  L   T  F  K   E  I  S  E   A  L  G   Y  S  R   S  Q  A  T


5401  GCGCTCAGAA AACAGCTGAG AACGCTGACT GTAACGCTAG TACGTTGTAC TTTTGTTAAC
        A  L  R   K  Q  L   R  T  L  T   V  T  L   V  R  C   T  F  V  N
```

```
5461  AGCAATAAAG ACGGCATAGA CGCTGTCAAT CTCTCTGCTG CTGGCAACTA CAGTAAAGGG
        S  N  K   D  G  I   D  A  V  N   L  S  A   A  G  N   Y  S  K  G

5521  AAGCTGACAA TGTGGTTAAC TCCTAACATG GCTGAGCGGC TTCTGTCAGA AGAATCATCT
        K  L  T   M  W  L   T  P  N  M   A  E  R   L  L  S   E  E  S  S

5581  ACAGAATATT TTCCGTTATC GTTACTGAAG CTGAAAGGGA CAGCCTATTA TTTAGCCTTA
        T  E  Y   F  P  L   S  L  L  K   L  K  G   T  A  Y   Y  L  A  L

5641  AAGGTCATGC ACAACGCAAA CATTAATGCA CGCTGGCATG CTGACAGAGT TGACAGATTG
        K  V  M   H  N  A   N  I  N  A   R  W  H   A  D  R   V  D  R  L

5701  GGCTTAGAAA ACACGCTGAA GGCCTTGCCT ACACTCCCCG ACCCGGTAAC CATGAGCGCC
        G  L  E   N  T  L   K  A  L  P   T  L  P   D  P  V   T  M  S  A

5761  AAAAATGGTA AACACATTCA GCAAAAAATT TTAACTCCCC TGGCTAAAGC CATTGAAGAG
        K  N  G   K  H  I   Q  Q  K  I   L  T  P   L  A  K   A  I  E  E

5821  CTTGAAGCCG TCACTGGCAT TGTCGTTAGA CCTAGCCAGC CGCTAAAGGG ACTGAGGACA
        L  E  A   V  T  G   I  V  V  R   P  S  Q   P  L  K   G  L  R  T

5881  AAAGACCTGT CTAAAGTCAC TCTGAATGTC ATTGATTGGG GACAGGTCGA CATAGCAGAA
        K  D  L   S  K  V   T  L  N  V   I  D  W   G  Q  V   D  I  A  E

5941  TTGACCAGAA AAAAGCGAAA ACGTTTGCGA AAAAATAATG TTCGTGAGGA CTAAAACTAT
        L  T  R   K  K  R   K  R  L  R   K  N  N   V  R  E   D  -

                                                        Ori
6001  ATTTTTTTAA ATTCGTATGT AGGTAATTAT GGTCACTAAT GTAGGTAATT ATGGTCACTA
                 ------------------------------>  ------------------->

6061  ATGTAGGTAA TTATGGTCAC ATTGTGAAAT TTCAGCAAGT GCCTTGAAGC CTTGAGCCAG
      ------------------->

6121  TAGGGAGTAA GCGCATTTTT TTAAAAAGTT TCACTTGTTA ATAGTTTAAT AGTATTAAAA

6181  GCAACGGCAC AGCTTGACGC TGGCCTTGCT TGAAAATTGA AAAAAGATGA AACAGCCAGG

                                          RBS              ORF2
6241  GAGAGCAGAG GCTTCTACTG GCCTGTTTTT AGAAGAAGGT ATCTAGCATG AACAATAATT
                                                          M   N  N  N

6301  TAGTTAAACC AGCAGATTTA AAGAGCTTGG TCTCTTTACC GGAATACATT GCCAGCGTGG
        L  V  K   P  A  D  L   K  S  L   V  S  L   P  E  Y  I   A  S  V

6361  TTAGCATGGA CTCTAAAGGC TTCTTTAGCT GTCTCAATCC GAACCACCCC GACAAGCACC
        V  S  M   D  S  K  G   F  F  S   C  L  N   P  N  H  P   D  K  H

6421  CTAGCATGTG TGTAGATCGT AACCACCCGC AATATGTTCA TTGCTTCAGC GCCACTTGCC
        P  S  M   C  V  D  R   N  H  P   Q  Y  V   H  C  F  S   A  T  C

6481  AGGCCAGTTA TGACCTGTTT GATTGTTGGG CTCTGATTAA TGACGGCGTG ACAGAGACCA
        Q  A  S   Y  D  L  F   D  C  W   A  L  I   N  D  G  V   T  E  T

6541  GGAAGAATAG CACTGGCAAG GAAAAGCCAG TCTATAACTT CAATGCTGTA GCTTCAGAGA
        R  K  N   S  T  G  K   E  K  P   V  Y  N   F  N  A  V   A  S  E

6601  TAGCTGACCA TTACGGGTAT ACTCTTATTG GCGACCCGGC AAATAACTTG CCAGAACCGC
        I  A  D   H  Y  G  Y   T  L  I   G  D  P   A  N  N  L   P  E  P

6661  CATTACCCGA ACCAGAGCCA GAGCCAGCTC AGACCAGCAC CAATTTTAGA GACCAATTAG
        P  L  P   E  P  E  P   E  P  A   Q  T  S   T  N  F  R   D  Q  L

6721  AAGCCTGGCA TGCTAACTTG AATCAGACTG ACTATCTTCA GAAGCGGGGG ATCACTCAGA
        E  A  W   H  A  N  L   N  Q  T   D  Y  L   Q  K  R  G   I  T  Q

6781  CAACAGCAGA GATTTTCAAT TTAGGCTACT CCCCGTTGAC CAACAGCATT ATTATCCCTT
        T  T  A   E  I  F  N   L  G  Y   S  P  L   T  N  S  I   I  I  P

6841  ACGGTCAGGA CGGCTATTAC GTTCAGAGAG CGCTTAATCC GATTGAGAAG CATGACCGCT
        Y  G  Q   D  G  Y  Y   V  Q  R   A  L  N   P  I  E  K   H  D  R
```

```
6901  ACAGATTTCC TATTGGCCAG GTGAGAGTCT ACAACGCTGA GGCGCTGAAA GAATGCAAGA
       Y  R  F   P  I  G  Q   V  R  V   Y  N  A   E  A  L  K   E  C  K

6961  CGGTATTCAT CGTTGAAGGT CAGTTTGACG CTCTGTCAAT CATGCAAGAA TCCGGTGTAG
       T  V  F   I  V  E  G   Q  F  D   A  L  S   I  M  Q  E   S  G  V

7021  GAGCTGTAGC AACTTCAGCC AGTCAGACCC AACTCATTGT CAAAACCTTA CAGAAGTTCA
       G  A  V   A  T  S  A   S  Q  T   Q  L  I   V  K  T  L   Q  K  F

7081  AAGAGCAAGA CCCAACAATT AACCCGACTA TCATACTCAG CATGGACAAC GACAGAGCAG
       K  E  Q   D  P  T  I   N  P  T   I  I  L   S  M  D  N   D  R  A

7141  GCCAGAAGGC GAATAGAGCA CTTCAGAGGG ACTTAGAAGC GCATGGCTTT ACTTGCTATG
       G  Q  K   A  N  R  A   L  Q  R   D  L  E   A  H  G  F   T  C  Y

7201  TCAACCCGGT TAACGGCGAT TACAAGGACG CTAACGAGTT TCTGGTAAAG GACAGAGAGG
       V  N  P   V  N  G  D   Y  K  D   A  N  E   F  L  V  K   D  R  E

7261  GCTTCAGACA GAAGCTTCAG CATGTCATCA ACCAGCCCGA TAATTGGCTT GACAAGTACT
       G  F  R   Q  K  L  Q   H  V  I   N  Q  P   D  N  W  L   D  K  Y

7321  ATGCTGACAT TAAGCAACGC CATGACTACC CGGACAATAT CCCTACTGGC TTCAAGAATT
       Y  A  D   I  K  Q  R   H  D  Y   P  D  N   I  P  T  G   F  K  N

7381  TAGATGATGA GCTTGACGGC GGGCTTCAGC CTAAACTCTA TGTTTTAGGT GCTGTATCAT
       L  D  D   E  L  D  G   G  L  Q   P  K  L   Y  V  L  G   A  V  S

7441  CATTGGGTAA AACGACTTTT GCCTTGAATG TTGCTGACAA CCTGGCTAAA CAGGGGAGGC
       S  L  G   K  T  T  F   A  L  N   V  A  D   N  L  A  K   Q  G  R

7501  ATGTTTTCTT CTTCAGCATG GAATCTAGCA AGAGAGAAGT GACGGACAAG CTTCTAAGCC
       H  V  F   F  F  S  M   E  S  S   K  R  E   V  T  D  K   L  L  S

7561  GGGCTAGCTG TCTCTCTAAC GGCCATAAAT GGACTCAGCT ACAAGTCAGC CGGGGAGCAT
       R  A  S   C  L  S  N   G  H  K   W  T  Q   L  Q  V  S   R  G  A

7621  GGTTAAACAA TGCTGAGGAC AAAGAAAAGT TTGACGGCCT GTTTAAAGCC TTCAGCCGTT
       W  L  N   N  A  E  D   K  E  K   F  D  G   L  F  K  A   F  S  R

7681  ACCAGCGCTT CTTACACATC TATGACAATA GAGTTAAGGC AAGTCAGGTA AAAGACCTGG
       Y  Q  R   F  L  H  I   Y  D  N   R  V  K   A  S  Q  V   K  D  L

7741  TCAATGGCTG GCTTGACAAC CACCCGGACG AAAAGAAGCC GCTTGTGGTC GTTGACTATC
       V  N  G   W  L  D  N   H  P  D   E  K  K   P  L  V  V   V  D  Y

7801  TTCAGATCTT GCAAGTTGAG CAGGACAACG TGACAGACAA GGCCAAGGTG ACGGACAGCG
       L  Q  I   L  Q  V  E   Q  D  N   V  T  D   K  A  K  V   T  D  S

7861  TGAGTGTTCT CTCAGAGCTG ACTAAACAGG CTGAAGTCCC TGTTCTGGTC ATCTCATCAT
       V  S  V   L  S  E  L   T  K  Q   A  E  V   P  V  L  V   I  S  S

7921  TGAACCGGGC TTCCTACTGG CAAGACGTAA GTTTTGAATC CTTCAAGGAA TCCGGGGAAA
       L  N  R   A  S  Y  W   Q  D  V   S  F  E   S  F  K  E   S  G  E

7981  TTGAATACTC AGCTGACGTT ATGTTAGGCT TAGAGTTCGC TCATCGTGAA GAATACATTA
       I  E  Y   S  A  D  V   M  L  G   L  E  F   A  H  R  E   E  Y  I

8041  CAGTTCAGAA AAACGGCCAT GTTGAATTGA ACAAAGAGAA GTTTGACCAG CGAAAACAGG
       T  V  Q   K  N  G  H   V  E  L   N  K  E   K  F  D  Q   R  K  Q

8101  AAGTCCTAGA CGGGTTGAAA TGGTCATTCT GAAGAATCGA ACTGGCAAGA CAGGCGGTCA
       E  V  L   D  G  L  K   W  S  F   -

8161  TATCTTCTTC AAGTACAACG CCAAGTACAA CAACTACCAG GCATGCACTG AGAAAGAGGC

8221  GGCAATAGCA GCAACTTTAA TAAGTTGTTT CATAGCAAGG AAGTAGGCAA GCCAATTGAA

8281  GCGGCCATGT CTGAATACAC GGTAGACCCG GAAACAGGCC GGTAACAGAG AAGAATCAAG

8341  ATAAATAGAG CTGAAGAAGC TGGCCAGGAA TGGCTGGCTT TTGTTTTGTC TGCTGTCCGT

8401  GTGGATATCA CCGACCGCCT GGTCCAGGTT CGCCAGGGAG TGACCGGGCT AAAAGTCTGA
```

```
8461  GCCTTTAAAG TTGACGATCA AAAATTCGAC AGTCAAAACT CAAGGCAAAC AAAAAGAGCC

8521  GCAATTCTTG CAGCTCTGAG TGCAGTTGAA TGTCATGCTC TGAGATAGCC CTAGAAGGCC

8581  CATATTTGCG TTCTGAGGCG TTTTAGCTTT GCCTGGCAGA ATTTTGCCCA TGACCGGGGC

8641  TGAGGATACC CCCCTGGGCC AACATTTGGC GAGAACGACC GTTTGGGAGC GGGCGCCAGG

8701  GGGTAAAGCA CGATAA
```

Complemented nucleotide sequence of ORF-1.

```
                                          RBS       ORF1
1400© TTTGATAAAA TGGCCTTAGA CATTAAACCG GAGAAAAAAT AATGAATGAT ACACAAGCTA
                                                    M  N  D  T  Q  A

1340© AATATCCGCA ATTAAGATTT AAAGGTTTCG CTGATCCTTG GGAGCAGTGT AAGTTGGGGG
      K  Y  P  Q  L  R  F  K  G  F  A  D  P  W  E  Q  C  K  L  G

1280© ACGTGGCAAA AATCACCATG GGTCAATCAC CAAATTCTAA AAATTACACT GATAATCCAA
      D  V  A  K  I  T  M  G  Q  S  P  N  S  K  N  Y  T  D  N  P

1220© AAGATCACAT TCTCGTTCAA GGAAATGCAG ATATGAAGGA TGGTCAAGTA CATCCTAGAA
      K  D  H  I  L  V  Q  G  N  A  D  M  K  D  G  Q  V  H  P  R

1160© TTTGGACTAC GGAAATAACA AAAATCGCAG ACAAAGGTGA TTTAATCTTA AGCGTTAGAG
      I  W  T  T  E  I  T  K  I  A  D  K  G  D  L  I  L  S  V  R

1100© CACCTGTAGG CGATATTGGA AAAACAAGCT ATGATGTGGT TATCGGTCGT GGAGTAGCCG
      A  P  V  G  D  I  G  K  T  S  Y  D  V  V  I  G  R  G  V  A

1040© CCATCAAAGG TAATGAATTT ATTTTTCAAC TGCTTAAACG AATGAAAACA GTAGGTTATT
      A  I  K  G  N  E  F  I  F  Q  L  L  K  R  M  K  T  V  G  Y

 980© GGACTAAATA TAGTACTGGT TCAACATTTG AGAGCATTAA TTCATTAGAA ATAAATAATG
      W  T  K  Y  S  T  G  S  T  F  E  S  I  N  S  L  E  I  N  N

 920© CGGTAATAAA TCTCCCTAAG GAGCATGAAC AAAATAAGGT TGGAAAGATA CTTTCCTATA
      A  V  I  N  L  P  K  E  H  E  Q  N  K  V  G  K  I  L  S  Y

 860© TGGATCATGC CATCACTTTA CATGAGGAAA AGAAACGTCA ACTTGAATGC TTGAAAAGCG
      M  D  H  A  I  T  L  H  E  E  K  K  R  Q  L  E  C  L  K  S

 800© CTTTATTGCA GAAGATGTTC GCTGACAAGA GTGGGTATCC TGTTGTTAGA TTTGAAGGAT
      A  L  L  Q  K  M  F  A  D  K  S  G  Y  P  V  V  R  F  E  G

 740© TTAGTGACGA GTGGGAAGAG CGTAAGTTGG GTGATGCGGT TAGTATATCA AGTGGAGTCA
      F  S  D  E  W  E  E  R  K  L  G  D  A  V  S  I  S  S  G  V

 680© CGGGTGACGC AACACTTCAA GATGGCGAAT ATAGACTTAC ACGCATCGAA AGTATTTCGC
      T  G  D  A  T  L  Q  D  G  E  Y  R  L  T  R  I  E  S  I  S

 620© AAGGAACGTT GAATGTTGCG AGGTTAGGAT TCACAAATAA AAAACCAGAC CAGAAGTATC
      Q  G  T  L  N  V  A  R  L  G  F  T  N  K  K  P  D  Q  K  Y

 560© TTCTAAATTT GGGCGATATT CTTTACAGCA ATATTAACTC ATTGAGTCAC ATTGGAAAGG
      L  L  N  L  G  D  I  L  Y  S  N  I  N  S  L  S  H  I  G  K

 500© TGGCTTTGGT TGATACGACA GGGATTTATC ACGGAATTAA TCTTTTGCGA TTCCAAATGC
      V  A  L  V  D  T  T  G  I  Y  H  G  I  N  L  L  R  F  Q  M

 440© GCAATGATGT TGATTCGGAA TTTTTATTTC AGAGACTAAA CACGACGCCA ATGAAAAATT
      R  N  D  V  D  S  E  F  L  F  Q  R  L  N  T  T  P  M  K  N

 380© GGGCAGTTTC GCACGCGAAT CCCGCAGTCA GCCAAGCAAG CATTAATCAA ACGGAATTAA
      W  A  V  S  H  A  N  P  A  V  S  Q  A  S  I  N  Q  T  E  L
```

```
320© GTAAACAGCC AATTTCGTTG CCAACAATAA CTGAACAACA AAAAATAGGT TCATTCTTCA
      S  K  Q   P  I  S  L   P  T  I   T  E  Q   Q  K  I  G   S  F  F

260© AACAACTCGA CAAAACTATC GCTCTTCATC AACGTAAGTT AGATTTGTTG AAAGAACAGA
      K  Q  L   D  K  T  I   A  L  H   Q  R  K   L  D  L  L   K  E  Q

200© AAAAAGGCTT TTTACAAAAG ATGTTCGTTT AGGGTCTATA ATTATAAAAT CGCCCCTAAA
      K  K  G   F  L  Q  K   M  F  V   -
```

V. Nucleotide sequence of the *hsd* cluster from *L. lactis* strain NCC88.

```
   1  GGATCCCGAC CAAATTGTCA TGGTCAACCT GTCAGGCAGA GGGGACAAGG ACGTTAATCA

  61  AGTTGCCGCC TACCTGGGTG AAGAAATTTA AGTGATTTAA TATACAAGGC CGGAAAGAGA

 121  CCCCATTCTC TTTCCGGTCT TCCTTTTGTC TGGAGATACG CAAGAAATTA TATTGTTATC

 181  ACTTTTTTCT CCTAAGAATT AGAACATTGT TTCTATAGGT AAAAACTGGT TGGTTTAGCT

 241  GTCGACTATC TGACCCGCTT TATGTCGGGG GCAAGTGCCC AAGAATCTTT CGAAATTTCG
                            >>.................ORF-Y.................>

 301  CTGCGGGGAG CAAGTAATAT TGGAGAAGAT GCTTTAGCTG CAAAGCTGCT GGCAGAAGTG
      >.............................ORF-Y.............................>

 361  AAAGGACTGG ATGATCGCTC GATTACGAAT GCCATTAAGC TGACCAGATT CGATGTCTAC
      >.............................ORF-Y.............................>

 421  TACCGGGCTG GCCTTGGCTA TAAGCCAGTT AGCGAGATTA AGCCGGATCA AGCAACGATT
      >.............................ORF-Y.............................>

 481  CAAAACGTGA GGACGATGGT TGAACGCTCA CTGCATTTTC TTGAAGTCTA TGGCCCAAAG
      >.............................ORF-Y.............................>

 541  CTGCTGGATG GATTTACTTT TGAAGGCGGC TATAAAGACA CGGTAAGCAA GGGGGATGGT
      >.............................ORF-Y.............................>

 601  GACTTAACCA CGGCTGACAC CCTCTGGGAC TTTAAGGTGT CGAAGGCGAA GGTGAAAAAA
      >.............................ORF-Y.............................>

 661  GAGTACACCC TGCAACTACT TATGTACTGG CGGATGGGCT TGCATTCGGT TCATCCGGAG
      >.............................ORF-Y.............................>

 721  TTTCAAAGCA TCAAGTATCT AGGTATCTAC AATCCCTGTC TGAATCAAGT CTATCGCATT
      >.............................ORF-Y.............................>

 781  GCCGTCACTG ACATTTCTGA AGATGTTATT AGGGAAGTCG AAGAAAAAGT AATCGGGTAT
      >.............................ORF-Y.............................>

 841  GGGATGCTGT AATGCAAAAA ATGGTGGTTC AGAACGGACA ATGTCATTAA GTTAAGGCAT
      >..ORF-Y...>>

 901  TGTGCTGGTT GAGCAGACCC ATACCTTTTA ATATACAGGG CGGGTCAAAC CGTCAGTATC

 961  TGGGTCAGCA ATCCCCAGGC TGTCCCGGCT GCCAAAAGTT GGTCGGCATT GACTGGATAG

1021  CTTCCTTCTG GGTTGCCGCC AAAGCAGAAG TCCCGGTTTC ACTTGCCGAA TTTACCTGGA

1081  ACTTCATCAT AGATGCCTAA AAGGCTGGTT TCAACCAACT GGCTGATCTT ATTTGCTTCA

1141  GCTGCCGCGA TGATATTTTT GGTTGGGCGG TTGTTACCAG CGTCGTCATG GTCAACGACA

1201  GCTGAGAAGA TCATATCACA GCCCTTAGCC GCGTCAACTA CGACTTGTTT GTTGTTGATG

1261  TCACCATCGA CCAAAGTTAC CCGGTCTGGG TAGCGGTCTT TTAAGCTTTG CAAGTGGCTG

1321  GCCCGGCGCA GGAGCAAGAC CAGGTCGATA TTCTTGAACT TGTCTTCAGT CAAAATTCTT

1381  TCTTCAACAA TGCGGGCAAT TTGACCGTTA GCTGCCAAAA TCAGTAATTT CATGTTTATT

1441  ATTTTCCTTT CTGCTTGTAG GATTCTCTTG ATGCTTTTAT GATAGCGGGG CGCAAGCTTT

1501  TTGAAAAATG TTCAATGTGC AAGATAGTAT GCAGGAAGGA CATAGTAGAT ATCAAATGCT

1561  GTAAATTCTG ATGTATATAA AAGATCATGA TTGCGGATAA CATAGGATCA GCCCTTTTTT

1621  CAGCCCTATT TTTTCGAAAA TAGGCAATAA AGGTTGATCA AAAGGAAATG CTTTCACTTC
```

```
      -35                          -10
1681  AGTAGAGAAA AAGTTAAAAT AGTATTTATA ATGTAATACA ATCAGCATCG TTAGGTAAGT

                  RBS          hsdR
1741  TGCTGTTTAG GGAAGGGGAA AGCTATGACT TTAGAATCAC AACTGGAAGA CAACCTAATC
                                 M  T  L  E  S  Q  L  E  D  N  L  I

1801  GCGCAGTTGA CGCAGGATGT CCATCAGTGG AAGTTCCGTG ATGATTTACG CACCGTAGAT
        A  Q  L  T  Q  D  V  H  Q  W  K  F  R  D  D  L  R  T  V  D

1861  CAGCTTTGGG ACAACTTTTT CCGTATCTTG GAATCAAACA ATAAGGATCA GTTAAATGAC
        Q  L  W  D  N  F  F  R  I  L  E  S  N  N  K  D  Q  L  N  D

1921  CATCCGTTGA CTCCTAACGA AAAAATGACG GTGAGAACGG CGATCGTTAA ACCGACCTTT
        H  P  L  T  P  N  E  K  M  T  V  R  T  A  I  V  K  P  T  F

1981  TACCGGGCAA CGGAGTTTAT GGTCGGGGCC AACCGGCAGG TTCGCTATCA CTTAAGAAGA
        Y  R  A  T  E  F  M  V  G  A  N  R  Q  V  R  Y  H  L  R  R

2041  GAAGATTCTT CTATTCCCGA CGCTGATCTG CTGATTTTAG ACAATACCAA CATTGCGGGT
        E  D  S  S  I  P  D  A  D  L  L  I  L  D  N  T  N  I  A  G

2101  GGAAACTCAG TTTATGAAGT TGTACACCAG GTTCAGCTAC AGAAGAAGAC CGCGCTTAAT
        G  N  S  V  Y  E  V  V  H  Q  V  Q  L  Q  K  K  T  A  L  N

2161  CAAGACCGTC GTTTCGACGT TAGTTTGTTG ATCAACGGCT TGCCGGTAAT TCACATTGAG
        Q  D  R  R  F  D  V  S  L  L  I  N  G  L  P  V  I  H  I  E

2221  CTTAAAGCTC CAAATGTTCC TTATAAGAAG GCCTTTAACC AAATTCAAAA GTATATCGAC
        L  K  A  P  N  V  P  Y  K  K  A  F  N  Q  I  Q  K  Y  I  D

2281  GAAGGACAAT TTACTGACAT TTACAGCTTC GTAGAAATGT TTGTGGTAAC TAATGGTACT
        E  G  Q  F  T  D  I  Y  S  F  V  E  M  F  V  V  T  N  G  T

2341  CAAACAAGAT ATATATCTGC TGGGCAGAAT TTGAATGCCA AGTTTTTAAC GGCCTGGGTT
        Q  T  R  Y  I  S  A  G  Q  N  L  N  A  K  F  L  T  A  W  V

2401  GATAAGAATA ATAAGCGGGT AGACAATTAT CTGAGTTTTG CAGAAGAGGT TTTGTCAATA
        D  K  N  N  K  R  V  D  N  Y  L  S  F  A  E  E  V  L  S  I

2461  CCTGCTGCTC ACCATATGAT TGCCGACTAT GTGGTTTTAG ACAGCGAAAG CAAGAGCGTT
        P  A  A  H  H  M  I  A  D  Y  V  V  L  D  S  E  S  K  S  V

2521  ATCCTGCTCC GTCCTTACCA GATACATGCG ATTCAAGCGA TTTTTAAGGC TTCTAGAGAG
        I  L  L  R  P  Y  Q  I  H  A  I  Q  A  I  F  K  A  S  R  E

2581  AGTAAGTCGG GCTATATTTG GCATACGACA GGGTCAGGTA AGACGTTAAC TTCGTACAAG
        S  K  S  G  Y  I  W  H  T  T  G  S  G  K  T  L  T  S  Y  K

2641  GTTGCCCGTA ACTTGTTGCA AATTCCGTCA ATTGATAAGT CAATCTTCCT AATCGACCGT
        V  A  R  N  L  L  Q  I  P  S  I  D  K  S  I  F  L  I  D  R

2701  AAAGACCTGG ACACGCAGAC TACAACTGCT TTCAAGATTT ACGCCAACAA CGATACGATC
        K  D  L  D  T  Q  T  T  T  A  F  K  I  Y  A  N  N  D  T  I

2761  AGCGTTAACG AAACAAATAA TAGTTATGAC CTTGCTGACC AGCTGACTGA TGGCGACCGG
        S  V  N  E  T  N  N  S  Y  D  L  A  D  Q  L  T  D  G  D  R

2821  ACTGTAGTAG TTACTACCCG CCAGAAGATG CAAAACATGT TTAAGCGGAT TGATGAATTA
        T  V  V  V  T  T  R  Q  K  M  Q  N  M  F  K  R  I  D  E  L

2881  GATCAGTTGC CTAAACGGTA TGAGAATTTA AAGAATATGC GGCTAGCCTT CATCGTCGAT
        D  Q  L  P  K  R  Y  E  N  L  K  N  M  R  L  A  F  I  V  D

2941  GAATGTCACA GAACGATTAC CCCTAGCCAA AAGCGGGAGA TTGATAAATT CTTTAGCCGC
        E  C  H  R  T  I  T  P  S  Q  K  R  E  I  D  K  F  F  S  R

3001  AAACCGCTGT GGTATGGCTT TACTGGTACG CCAATTTTTA ACGAGAATGC CCGGGCAAAG
        K  P  L  W  Y  G  F  T  G  T  P  I  F  N  E  N  A  R  A  K

3061  AATGGTCAAG ACGCGCGGAC GACTGAAGAA TTATATGGGC CAGTCCTGCA CAAGTATACG
        N  G  Q  D  A  R  T  T  E  E  L  Y  G  P  V  L  H  K  Y  T
```

```
3121  ATCGGGGATG CAATTAGAGA CAAGATGGTT TTGGGCTTCT CCATCGATAA TCAAGGTGGC
        I  G  D   A  I  R   D  K  M  V   L  G  F   S  I  D   N  Q  G  G

3181  AGCAACGAAG ATGGGAATGA AGAAGACACC AAGAAAATGG ACCAGATTTA CCGGTCCAAG
        S  N  E   D  G  N   E  E  D  T   K  K  M   D  Q  I   Y  R  S  K

3241  GCCCATATGC ATTCAGTTGC GACGGCAGTA ATTAAAGCGG CGTACCGCAA GCAGGGCCTT
        A  H  M   H  S  V   A  T  A  V   I  K  A   A  Y  R   K  Q  G  L

3301  ATTTCTGGTA AGAAATACTC AGCCATTTTG ACGACTTCGT CAATTGAGCA AGCCCAGAAG
        I  S  G   K  K  Y   S  A  I  L   T  T  S   S  I  E   Q  A  Q  K

3361  TACTACCGTA TCTTTAAGAA AATCATCGAT GGGGAAGACG AAGAATTCAA GATTCCGGAA
        Y  Y  R   I  F  K   K  I  I  D   G  E  D   E  E  F   K  I  P  E

3421  CGAATCAAGA AAGTGGCACC AGACTTCCCA AAGATCGCTA TTACTTACTC AGTTAGTGAA
        R  I  K   K  V  A   P  D  F  P   K  I  A   I  T  Y   S  V  S  E

3481  AACGAAGATG ATTCAGAATC AGTGCAAGAT GAGATGAAGC AGTCGCTTGC GGACTACAAC
        N  E  D   D  S  E   S  V  Q  D   E  M  K   Q  S  L   A  D  Y  N

3541  GCCGTTTATG GGACGAACTT CTCGATGGCT GAGCTGGATC AGTACAATCA AAACGTTAAT
        A  V  Y   G  T  N   F  S  M  A   E  L  D   Q  Y  N   Q  N  V  N

3601  GCCCGGCTTG CCCGCAAGAA GGCTCAATAC CAAGCTGACA ACCAACGCTT AGACCTAGTG
        A  R  L   A  R  K   K  A  Q  Y   Q  A  D   N  Q  R   L  D  L  V

3661  ATAGTTGTTA ACCGTTTGCT GACTGGTTTT GACTCACCAA GTTTGTCGAC TTTGTACATT
        I  V  V   N  R  L   L  T  G  F   D  S  P   S  L  S   T  L  Y  I

3721  GACCGGCCGC CAATGAGTCC ACAGGACATT ATCCAGGCTT TTTCTAGAAC CAACCGGATT
        D  R  P   P  M  S   P  Q  D  I   I  Q  A   F  S  R   T  N  R  I

3781  TTTGATAAAG ACAAGACTTG GGGACAAATC GTAACTTACC AGTATCCCAA GACCTTTAGT
        F  D  K   D  K  T   W  G  Q  I   V  T  Y   Q  Y  P   K  T  F  S

3841  GAAAAGATTG ATGATGCAAT CGTCCTATAT TCTAATGGCG GGGAGAAGTA TGCGGTTGCT
        E  K  I   D  D  A   I  V  L  Y   S  N  G   G  E  K   Y  A  V  A

3901  CCAAGCTGGG AAGAATCAAA ACAAAGCTAC ATATCAGCCC GGTCTAAAAT TGAAATGTAC
        P  S  W   E  E  S   K  Q  S  Y   I  S  A   R  S  K   I  E  M  Y

3961  AGCTTCGACG CAGATGGACC GTCAATTTAT GATGCTTCCA AGGAAGAAAA GAAGAAATTC
        S  F  D   A  D  G   P  S  I  Y   D  A  S   K  E  E   K  K  K  F

4021  GTTAAAGCTT TCCAAGAATT TGACAAGGCT TTGGCTGCAA TCAAGACCTA TGATGAATTA
        V  K  A   F  Q  E   F  D  K  A   L  A  A   I  K  T   Y  D  E  L

4081  GATACCGAAG AAGGCCTAAT GCAATTAGGT GTGAGCGACA TTACACCAGA TGATTGGGAA
        D  T  E   E  G  L   M  Q  L  G   V  S  D   I  T  P   D  D  W  E

4141  GCCATGCGGG GAGTTTATGA GGATATTCTG GACGATCTGC GGAAAGATCC AGATGAAGAT
        A  M  R   G  V  Y   E  D  I  L   D  D  L   R  K  D   P  D  E  D

4201  CCTGATCACA ATGACGACAT CGATGATGAA TACGAACTTG AGTCATTTGG ACAAAAGGAA
        P  D  H   N  D  D   I  D  D  E   Y  E  L   E  S  F   G  Q  K  E

4261  ATCGATGAGC GCTATATCAT GAACTTGATC CAGGCATTCT TGCCAGAGAG CTCTAACAAT
        I  D  E   R  Y  I   M  N  L  I   Q  A  F   L  P  E   S  S  N  N

4321  CAAGAAGAAG CTAGCAGCAA TGAAATTTCG CCTGAGACAG TCAAGGAAAT CAATGGCCAC
        Q  E  E   A  S  S   N  E  I  S   P  E  T   V  K  E   I  N  G  H

4381  ATTGATGAAT TGGCGAAGAC AAACGAGCTT TTGGCAGAGA TTATGCGGAA GCTCTGGCAG
        I  D  E   L  A  K   T  N  E  L   L  A  E   I  M  R   K  L  W  Q

4441  CAGATTTTGC AGGATCCGGC TAAATACGCT GGCAAGCAGG TCGATGAGCT CCTTGAATCC
        Q  I  L   Q  D  P   A  K  Y  A   G  K  Q   V  D  E   L  L  E  S

4501  TTGATTGACC AGGAATTGCA ATCAATCATG CGTGAGTTTG CTGATAAATA CAAGGTTGAC
        L  I  D   Q  E  L   Q  S  I  M   R  E  F   A  D  K   Y  K  V  D
```

```
4561  TACGACCAGT TCCGGTACGT TATGGCTAAC TATGACCCTA AGTTAAAGGG AAATAAGCAA
        Y  D  Q   F  R  Y   V  M  A  N   Y  D  P   K  L  K   G  N  K  Q

                                                          -35
4621  AAGGGGATGA ATGACCTTTT GGACAAGGAA CGTTTTGTCG ACTACTTGAA TGACAATCCG
        K  G  M   N  D  L   L  D  K  E   R  F  V   D  Y  L   N  D  N  P

                -10
4681  GATTCTGACT TAAATAAGCC GTACCGCTGG AAGAGTGAAG TTCGGGAACA GGCCAAGCAA
        D  S  D   L  N  K   P  Y  R  W   K  S  E   V  R  E   Q  A  K  Q

                                                        RBS
4741  TACTATGTCG ATAAGATTGG CCCGTTAATT AACAGAGAAG CATAGAAGGA AGAAGAATAT
        Y  Y  V   D  K  I   G  P  L  I   N  R  E   A  -

      hsdM
4801  GGCAGAAGAA AATTCAACAG TTAGCTTGCA GAGTGGTTTG TTTGCGGCTG CAGACGTCTT

      M  A  E  E   N  S  T   V  S  L   Q  S  G  L   F  A  A   A  D  V

4861  GCGTTCAAAG ATGGACGCCA ATGAGTATAA GAACTATCTT TTGGGGACTG TTTTCTATAA

      L  R  S  K   M  D  A   N  E  Y   K  N  Y  L   L  G  T   V  F  Y

4921  GTACCTTTCA GACCAGCAGC TTTACAAGTT GGCTGAAGAC GCCGGTGAAG ATGACGTTAC

      K  Y  L  S   D  Q  Q   L  Y  K   L  A  E  D   A  G  E   D  D  V

4981  TTTAGACAAG GCTCAGAAAA TCTATGAGGA AAACCTTGAA GAAGAAGACC TACTGGAAGA

      T  L  D  K   A  Q  K   I  Y  E   E  N  L  E   E  E  D   L  L  E

5041  GGTTAAAGAC GAGCTTGGAT ATTTGATCGA GCCGGAATAC ACCTACACCA AGATCTTAGA

      E  V  K  D   E  L  G   Y  L  I   E  P  E  Y   T  Y  T   K  I  L

5101  TAATGCCAAC GATGGCAGTT TCCAACTCAA TCAGTTAGGG GATGCCTTTA ACAAGCTAGA

      D  N  A  N   D  G  S   F  Q  L   N  Q  L  G   D  A  F   N  K  L

5161  AAGTCAAGGC AGCAGTTTTG AAGGCTTGTT TGACGACTAT GACCTGTATT CAAAGCGGCT

      E  S  Q  G   S  S  F   E  G  L   F  D  D  Y   D  L  Y   S  K  R

5221  GGGTCAAAAC TTGCAGAAGC AGACAGATAC AATTGCCGGA GTGATTAAGG CGATCGGCAA

      L  G  Q  N   L  Q  K   Q  T  D   T  I  A  G   V  I  K   A  I  G

5281  ATTAGAACTG GTTAAGACTC CTGGTGACAC CTTGGGGGAT GCTTACGAAT ACTTGATCAG

      K  L  E  L   V  K  T   P  G  D   T  L  G  D   A  Y  E   Y  L  I

5341  TCAATTTGCC TCAGAATCAG GTAAGAAGGC CGGTGAGTTC TATACTCCGC AAGAAGTGTC

      S  Q  F  A   S  E  S   G  K  K   A  G  E  F   Y  T  P   Q  E  V

5401  GGAACTTTTG GCACGGTTGA CTTTAGTCGG TAAGGATTAC TCTTCTGGGA TGAGCGTTTA

      S  E  L  L   A  R  L   T  L  V   G  K  D  Y   S  S  G   M  S  V

5461  CGACCCGGCT ATGGGGTCAG GTTCACTGCT GCTGAACTTT AGAAAGTATG TTCCAAACTC

      Y  D  P  A   M  G  S   G  S  L   L  L  N  F   R  K  Y   V  P  N

5521  ATCAAGAATT ACTTATTACG GGCAGGAAAT CAACACATCA ACCTTTAACT TGGCTAGAAT

      S  S  R  I   T  Y  Y   G  Q  E   I  N  T  S   T  F  N   L  A  R

5581  GAACATGATT TTGCACCACG TTGATCTGGC AAACCAGAAG TTGAGAAACG GGGATACGTT

      M  N  M  I   L  H  H   V  D  L   A  N  Q  K   L  R  N   G  D  T
```

```
5641  AGACGAGGAC TGGCCAGCTG AAGAAACCAC TAACTTCGAC TCAGTTGTCA TGAACCCGCC
      L  D  E  D  W  P  A  E  E  T  T  N  F  D  S  V  V  M  N  P
5701  ATATTCACTT AAATGGAGCG CGGACAAGGG CTTCTTGGAT GACCCACGTT TTTCTAAATA
      P  Y  S  L  K  W  S  A  D  K  G  F  L  D  D  P  R  F  S  K
5761  CGGTGTTTTG CCGCCAAAGT CTAAGGCGGA CTATGCTTTC TTGCTTCACG GCTTCTACCA
      Y  G  V  L  P  P  K  S  K  A  D  Y  A  F  L  L  H  G  F  Y
5821  CTTGAAGCAT AGTGGGGCAA TGGCTATCGT TCTGCCACAC GGGATTCTTT TCCGTGGTGC
      H  L  K  H  S  G  A  M  A  I  V  L  P  H  G  I  L  F  R  G
5881  AGCGGAAGGA AAGATCCGGC AAAAGTTGCT TGAAGAGGGC GCGATTGATG CAGTAATCGG
      A  A  E  G  K  I  R  Q  K  L  L  E  E  G  A  I  D  A  V  I
5941  TTTGCCTGCA AACTTGTTCT ACTCAACTGG TATCCCAACT ACTATTGTCG TTCTTAAGAA
      G  L  P  A  N  L  F  Y  S  T  G  I  P  T  T  I  V  V  L  K
6001  GGACAAGCAG GATCGGAGCG TGCTGTTTAT TGACGCATCT AAGGAGTTCG AGAAAGTTAA
      K  D  K  Q  D  R  S  V  L  F  I  D  A  S  K  E  F  E  K  V
6061  GACTCAGAAC AAGCTGCGGC AAGAAGACAT TGATAAGATC TTGAAGACCT ACGAAGAGCG
      K  T  Q  N  K  L  R  Q  E  D  I  D  K  I  L  K  T  Y  E  E
6121  GCCAGCAGAT GTTGAGAAGT ATGCCCACTT GGCAAGTTTT GACGAAATCA AAGAAAATGA
      R  P  A  D  V  E  K  Y  A  H  L  A  S  F  D  E  I  K  E  N
6181  CTTCAATTTG AACATTCCTC GTTACGTTGA TACTTTTGAG CCAGAACCAG AAATTGATCT
      D  F  N  L  N  I  P  R  Y  V  D  T  F  E  P  E  P  E  I  D
6241  GCGGGACGTG GCTAAGGAGC TGCGGGATAT TGACCAACAG ATCAATGAAA ATGAGAAAGA
      L  R  D  V  A  K  E  L  R  D  I  D  Q  Q  I  N  E  N  E  K
6301  ATTGGTAGGG ATGCTTAAGG AATTGACTTC AAGTGACGAT GATATCATGG CAGGTTTGCA
      E  L  V  G  M  L  K  E  L  T  S  S  D  D  D  I  M  A  G  L
                                RBS          hsdS-1
6361  GAGCATCATC GAAAATTTTG AGGAGGAAAT CCGCTAATGA AAGATGAAAA AAAGGCCCCT
                                                 M  K  D  E  K  K  A  P
    Q  S  I  I  E  N  F  E  E  E  I  R  -
6421  AAACTGCGGT TTAAAGGCTT TACTGACGAT TGGGAGCAAG TAAAGTATGG GGAGATATTT
        K  L  R  F  K  G  F  T  D  D  W  E  Q  V  K  Y  G  E  I  F
6481  CAACGAAGGT CAAAAATGGG TGTTAGTACA CCCGCTCTCC CAAGCGTTGA ATACGACGAC
        Q  R  R  S  K  M  G  V  S  T  P  A  L  P  S  V  E  Y  D  D
6541  ATTAATCCAG GGATGGGCAC TCTAAATAAG GAACCTAAAA GTAAAGGTAC AAGCAAGCGA
        I  N  P  G  M  G  T  L  N  K  E  P  K  S  K  G  T  S  K  R
6601  GGAATTCACT TCAATCCTGG AGATGTTCTA TTCGGGAAAT TACGTCCTTA CTTAAAGAAT
        G  I  H  F  N  P  G  D  V  L  F  G  K  L  R  P  Y  L  K  N
6661  TGGCTTTTTG CTTGTTTTGA AGGAGTGGCT GTTGGAGACT TTTGGGTTCT AACATCAAGT
        W  L  F  A  C  F  E  G  V  A  V  G  D  F  W  V  L  T  S  S
6721  AAGATTGATC ATGGATTTAC ATATAGTTTA ATTCAAGCAC CCGAATTTCA GTATATAGCT
        K  I  D  H  G  F  T  Y  S  L  I  Q  A  P  E  F  Q  Y  I  A
```

```
6781  AATTTATCAT CTGGTTCGAA AATGCCTAGA TCTGATTGGG GATTGGTCTC AAATGCAAGA
        N  L  S   S  G  S   K  M  P  R   S  D  W   G  L  V   S  N  A  R

6841  ACGTTTATCC CAACAAATCT TTCAGAACAA AAAAGTATAT CTTCTGTCTT ATTTGGCTTA
        T  F  I   P  T  N   L  S  E  Q   K  S  I   S  S  V   L  F  G  L

6901  GATACCGCAA TCACTTTACA TGAGGAAAAG AAACGCCAAC TTGAGCGCCT TAAAAGCGCT
        D  T  A   I  T  L   H  E  E  K   K  R  Q   L  E  R   L  K  S  A

6961  TTATTGCAGA AGATGTTCGC TGACAAGAGC GGGTATCCAG CTGTTCGGTT TAAAGGATTT
        L  L  Q   K  M  F   A  D  K  S   G  Y  P   A  V  R   F  K  G  F

7021  GACGACATCT GGGACCAAGA AAAATTGAAC AGTCTTGTAA GACTGCACCG GGGATTAACA
        D  D  I   W  D  Q   E  K  L  N   S  L  V   R  L  H   R  G  L  T

7081  TATTCACCAA ATAATGTTCA AGATAGTGGC ATTAGAATTC TACGATCTTC AAATATTTTA
        Y  S  P   N  N  V   Q  D  S  G   I  R  I   L  R  S   S  N  I  L

7141  GACGGCCAGT TTGTGATGAC AGATGATGAT ATATTTGTAA AATCTAGTGT CGTAAATATT
        D  G  Q   F  V  M   T  D  D  D   I  F  V   K  S  S   V  V  N  I

7201  CCTACCGTAA AGGACGGAGA TATATTGATC ACTGCAGCAA ACGGGTCTAT TAAATTGGTA
        P  T  V   K  D  G   D  I  L  I   T  A  A   N  G  S   I  K  L  V

7261  GGCAAACATG CCATAATTTC AGGAATATCG GAAAACACCG CGGTTTCCGG TGGCTTTATG
        G  K  H   A  I  I   S  G  I  S   E  N  T   A  V  S   G  G  F  M

7321  TTAGTCGGAT CATCTCGAAT ACCGGATTTC GTAAATTCAT TATTTGATAC ATCTTGGTAC
        L  V  G   S  S  R   I  P  D  F   V  N  S   L  F  D   T  S  W  Y

7381  CAAAGATTTA TTAGAAAATA TGTTACAGGT GGTAATGGGT CCATCGGAAA TCTAAAGAAA
        Q  R  F   I  R  K   Y  V  T  G   G  N  G   S  I  G   N  L  K  K

7441  AATGATCTCG ATAAACAATA CGTAAAGGTA CCAACAACAA GTGAACAAGA ACGAATTGGA
        N  D  L   D  K  Q   Y  V  K  V   P  T  T   S  E  Q   E  R  I  G

7501  GAGTTTTTTA GGGAAATAGA CCAGCTTATC ATTAACAATC AGATTAAGCA TGAAAAATTG
        E  F  F   R  E  I   D  Q  L  I   I  N  N   Q  I  K   H  E  K  L

7561  TTGGAACTCA AAAAATTCCT TCTCCAAAAT ATGTTCATAT AGTTTAAAAG GATAGCGATA
        L  E  L   K  K  F   L  L  Q  N   M  F  I   -

7621  CCAAGCCGCT ATCCTTTTCC TTACAATGCA GCCAAGTGCT GCATTATTTT GGTATTATCT
                          <<.................int..................<

7681  TTGTTCTCCA ATTCTTGAAT AATGTGTAGG TATGTTTGTT GCGTTGTTGT CATATCAGCA
      <.............................int.............................<

7741  TGTCCTAAAC GATTCGCAAC ACTTGCAATA GAAACACCGG CAAATAAAAG CAACGAAGCA
      <.............................int.............................<

7801  TGGGTGTGGC GTAGCCCGTG TACAGAGATA ACAGGGATAT TTAACTCCTT ACAGTGCTTC
      <.............................int.............................<

7861  TTCAGTAAAT TATTAACAGT TGAGTTAAAA ACTCGCTTTT GGGCAAAAAT AGGCTCTCCA
      <.............................int.............................<

7921  TTAGGCAAAT CTTGAATGAC TTGATTGAGC TGCATGGCAA GCTTCCAATC GATGGGAACT
      <.............................int.............................<

7981  TTTCTTACTG ATGACGAATT CTTTGTCGGC TTGAAACTCC CCGTATAGCT CTTGTAGTCC
      <.............................int.............................<

8041  CATGTCTTAT CTACGTTGAT CATTTGTTGC TCTAAATCTA TGTCTTCTTT AGTGAGCCCT
      <.............................int.............................<

8101  AAAGCTTCGC TAAATCGCAA GCCTGTTTTA GCAATTAAAA GGATCATCCA ATCAAAATTT
      <.............................int.............................<

8161  GGGAAAGCTG ACAGATCTAA GTGACGAAGT AGCAACTTTA ATTCGAATTC ATTCAAATAC
      <.............................int.............................<
```

```
8221  TTTGCCTTCT TTTCGGAAGG CTCCTTTCCT TTAACCACAA TTTTCCTTGT TGGATCTCTC
      <.............................int.............................<

8281  TGAATATAGC CTTCGTCATA CGCATCTAGC AACATAGCCT TCAGCTGATG GTGAAAATCC
      <.............................int.............................<

8341  ATTACTGTTG CTTTCTCATG AGTTTGAGCA AAAGTATTTA AGATTTGCTG ATATTCAAGT
      <.............................int.............................<

8401  CTTGTAGTAT CGACTAATTT TAGATTGGGC GCAATTTCTT GTAAATGGCG ATGAGATAGC
      <.............................int.............................<

8461  CAGTACTTAT CTAGAGTCCT TTCTCGAACT GCTCCATCCT TGTATACCTC AATCCATTGG
      <.............................int.............................<

8521  GCAAAATAAT CATGTAGTAA AATTTGTTTC CTACGCATAA AAATCTCCTT AGCAAATATC
      <.................int..................<<

8581  TTGCTAGGGA GATTTTAGAT TAGATGAACA TGTTTTGTAA CAATGACTCC TTTGTTTCGT
                         <<.................hsdS-2'.................<

8641  TCAATTTATC CATCTTTAGT TGCTGCTTAG TGATAAGATT ATCTATATGC AACAATAGGC
      <............................hsdS-2'............................<

8701  TGCCGATCAG TTTCTGTTCA TCAATGTCTG TTGGAAAATT TGCCTTACAT TTTTTCAAAA
      <............................hsdS-2'............................<

8761  TAGTAAAGTG CCTTTTATAG CCTTCACTTT TAGGTTTATA CTTTTCAAGT TCATAGAAGT
      <............................hsdS-2'............................<

8821  AAAACAAACC ATTCATTTCT ATTGATGGTG ACAATATTTT TATTCCATCA GAAGCAACAA
      <............................hsdS-2'............................<

8881  GAAATGGTGA CCTAGGCTTA AATAAGGATA GTGTATGGTC ACCAAATAAT GTTATATTAT
      <............................hsdS-2'............................<

8941  TATAATTCTT GAATGGATTA CTATTAGAGA AACCCGCGAT TGGTTTATCA CCTTGTTGTA
      <............................hsdS-2'............................<

9001  TCACTAAATA ATTTCCATGT GTTGAAGGTG TTGATATGTA AGGTTTAGCT GCTATTACTT
      <............................hsdS-2'............................<
9061  TAGTAATTTC TCCTAACTTA CTTTGTTTCC ATGCTTTATC AAATCCCTCA AACCGAACAG
      <............................hsdS-2'............................<

9121  CTGGATATCC GTTTTTGTTA GCAAACATTT CCTGCGATAA AACGCTTTTT AAGCGCTCAA
      <............................hsdS-2'............................<

9181  GTAGGCATTT CTTTTCCTCA TGTAAAGTGA TGATCTGATC TAGCTTCTGA AGGATAGTAC
      <............................hsdS-2'............................<

9241  TAATTAATGA TTGTTCAACA GGGGTTGGTA CTTCAAACTT GATTTTGAAG AAATCTTTCT
      <............................hsdS-2'............................<

9301  TATTAATATT GAGCAACCCA TCCATCCTAG CCCCGGATGT TACTAGTCTT GCCAATTCTC
      <............................hsdS-2'............................<

9361  TATCAAGTTT CTTGGTGGCA AAATAATACT CGACGTAATA AGGATTATTC TCTGTCTTCA
      <............................hsdS-2'............................<

9421  TTTTGAAGCT ATGATAAACA CGAGGAACTA GTGCCTGATC ATATAGATTC TGTACAAAGA
      <............................hsdS-2'............................<

9481  CAGTTCCGTA TTTTGCCAGC TTTGAATTTC CATGGTTATA TGCTAAATGA CCTTTTTCTA
      <............................hsdS-2'............................<

9541  AAAGAGTATA CTTTTTCAAT TCATTTCCGG CGATTACTTG AGAAAAGCGG TCTTTTTGAT
      <............................hsdS-2'............................<

9601  TCATCCAACC ATTTGCAGCA GATATAGTTA AAATTGGTAA ACCTTTGACA TTGCCATTAT
      <.<<
```

```
 9661  TTCCAGACAC TTCTTCACAA ACATCCCCCA ACTTACGTTG CTCCCAAGCA AAAATTACAT
                                                                  <<...<

 9721  CCCTTCCGCC TGTTCCTGCA TGACCTGGCA GAAGGCGATG ACCAGGGGAT CTGCTTCATC
       <.............................ORF-X.............................<

 9781  TTTCCGGCAG ACCAGGCCAT AGTCCAGCAG GGCAGGGTAG TTCAACTTCA CCGCCTTGAT
       <.............................ORF-X.............................<

 9841  CACATCTGCT TTTTCGATCG CCACAAAACT TGGCATCACG GTGATCCCTA GACCCGCCTT
       <.............................ORF-X.............................<

 9901  GACCATTAAG CTGACGATCT CAACATCGTT AACGTGGCTG ATGTCTAAGT CATCGACCTT
       <.............................ORF-X.............................<

 9961  CTTTCTGATC GTCTCCTGTA GCTTCAGTTG TTCTGGCGGG CACCAGTTGT TGTCCATCAG
       <.............................ORF-X.............................<

10021  CAAAATCTTT TCTCCTGCCA AATCATCCAG TTCAAGCTCC TCTCTGTCAC TTAAATGATG
       <.............................ORF-X.............................<

10081  GTCCGCTGGT ACAATTGCCA TAAAGTGTCC GCTAAAAAGA TGATAGTACT TCACCTCGGC
       <.............................ORF-X.............................<

10141  TAGATCAGTG ATATCATCAT GTGTTGTAAA AATCACATCG CTATCATGGT CAATCAAGTG
       <.............................ORF-X.............................<

10201  GTGTTTGAGC CGGTTATGAT CAAATCCCTC AAGAAAGATC TTGCAATTCG GGTATTCCCG
       <.............................ORF-X.............................<

10261  GTGAAAAGCT TTGATCATCT CTGGCAAAAT TGCTTGCTCT AGCGGGGTAT CCGTAATTCC
       <.....ORF-X......<<

10321  AATCGTTAAG TTTGCTTTGC TCCCGGTTGA AAGAATTTCT GGTCCGCTGC A
```

hsdS 2' and *int*

```
                                9800© ATGGCCTGGT CTGCCGGAAA GATGAAGCAG ATCCCCTGGT

                                                                        hsdS-2'
 9760© CATCGCCTTC TGCCAGGTCA TGCAGGAACA GGCGGAAGGG ATGTAATTTT TGCTTGGGAG
                                                                    W  E

 9700© CAACGTAAGT TGGGGGATGT TTGTGAAGAA GTGTCTGGAA ATAATGGCAA TGTCAAAGGT
        Q  R  K    L  G  D    V  C  E  E    V  S  G    N  N  G    N  V  K  G

 9640© TTACCAATTT TAACTATATC TGCTGCAAAT GGTTGGATGA ATCAAAAAGA CCGCTTTTCT
        L  P  I    L  T  I    S  A  A  N    G  W  M    N  Q  K    D  R  F  S

 9580© CAAGTAATCG CCGGAAATGA ATTGAAAAAG TATACTCTTT TAGAAAAAGG TCATTTAGCA
        Q  V  I    A  G  N    E  L  K  K    Y  T  L    L  E  K    G  H  L  A

 9520© TATAACCATG GAAATTCAAA GCTGGCAAAA TACGGAACTG TCTTTGTACA GAATCTATAT
        Y  N  H    G  N  S    K  L  A  K    Y  G  T    V  F  V    Q  N  L  Y

 9460© GATCAGGCAC TAGTTCCTCG TGTTTATCAT AGCTTCAAAA TGAAGACAGA GAATAATCCT
        D  Q  A    L  V  P    R  V  Y  H    S  F  K    M  K  T    E  N  N  P

 9400© TATTACGTCG AGTATTATTT TGCCACCAAG AAACTTGATA GAGAATTGGC AAGACTAGTA
        Y  Y  V    E  Y  Y    F  A  T  K    K  L  D    R  E  L    A  R  L  V

 9340© ACATCCGGGG CTAGGATGGA TGGGTTGCTC AATATTAATA AGAAAGATTT CTTCAAAATC
        T  S  G    A  R  M    D  G  L  L    N  I  N    K  K  D    F  F  K  I

 9280© AAGTTTGAAG TACCAACCCC TGTTGAACAA TCATTAATTA GTACTATCCT TCAGAAGCTA
        K  F  E    V  P  T    P  V  E  Q    S  L  I    S  T  I    L  Q  K  L
```

```
9220© GATCAGATCA TCACTTTACA TGAGGAAAAG AAATGCCTAC TTGAGCGCTT AAAAAGCGTT
       D  Q  I   I  T  L   H  E  E  K   K  C  L   L  E  R   L  K  S  V

9160© TTATCGCAGG AAATGTTTGC TAACAAAAAC GGATATCCAG CTGTTCGGTT TGAGGGATTT
       L  S  Q   E  M  F   A  N  K  N   G  Y  P   A  V  R   F  E  G  F

9100© GATAAAGCAT GGAAACAAAG TAAGTTAGGA GAAATTACTA AAGTAATAGC AGCTAAACCT
       D  K  A   W  K  Q   S  K  L  G   E  I  T   K  V  I   A  A  K  P

9040© TACATATCAA CACCTTCAAC ACATGGAAAT TATTTAGTGA TACAACAAGG TGATAAACCA
       Y  I  S   T  P  S   T  H  G  N   Y  L  V   I  Q  Q   G  D  K  P

8980© ATCGCGGGTT TCTCTAATAG TAATCCATTC AAGAATTATA ATAATATAAC ATTATTTGGT
       I  A  G   F  S  N   S  N  P  F   K  N  Y   N  N  I   T  L  F  G

8920© GACCATACAC TATCCTTATT TAAGCCTAGG TCACCATTTC TTGTTGCTTC TGATGGAATA
       D  H  T   L  S  L   F  K  P  R   S  P  F   L  V  A   S  D  G  I

8860© AAAATATTGT CACCATCAAT AGAAATGAAT GGTTTGTTTT ACTTCTATGA ACTTGAAAAG
       K  I  L   S  P  S   I  E  M  N   G  L  F   Y  F  Y   E  L  E  K

8800© TATAAACCTA AAAGTGAAGG CTATAAAAGG CACTTTACTA TTTTGAAAAA ATGTAAGGCA
       Y  K  P   K  S  E   G  Y  K  R   H  F  T   I  L  K   K  C  K  A

8740© AATTTTCCAA CAGACATTGA TGAACAGAAA CTGATCGGCA GCCTATTGTT GCATATAGAT
       N  F  P   T  D  I   D  E  Q  K   L  I  G   S  L  L   L  H  I  D

8680© AATCTTATCA CTAAGCAGCA ACTAAAGATG GATAAATTGA ACGAAACAAA GGAGTCATTG
       N  L  I   T  K  Q   Q  L  K  M   D  K  L   N  E  T   K  E  S  L

     -35                      -10                            RBS
8620© TTACAAAACA TGTTCATCTA ATCTAAAATC TCCCTAGCAA GATATTTGCT AAGGAGATTT
       L  Q  N   M  F  I   -

       int
8560© TTATGCGTAG GAAACAAATT TTACTACATG ATTATTTTGC CCAATGGATT GAGGTATACA
          M  R   R  K  Q   I  L  L  H   D  Y  F   A  Q  W   I  E  V  Y

8500© AGGATGGAGC AGTTCGAGAA AGGACTCTAG ATAAGTACTG GCTATCTCAT CGCCATTTAC
       K  D  G   A  V  R   E  R  T  L   D  K  Y   W  L  S   H  R  H  L

8440© AAGAAATTGC GCCCAATCTA AAATTAGTCG ATACTACAAG ACTTGAATAT CAGCAAATCT
       Q  E  I   A  P  N   L  K  L  V   D  T  T   R  L  E   Y  Q  Q  I

8380© TAAATACTTT TGCTCAAACT CATGAGAAAG CAACAGTAAT GGATTTTCAC CATCAGCTGA
       L  N  T   F  A  Q   T  H  E  K   A  T  V   M  D  F   H  H  Q  L

8320© AGGCTATGTT GCTAGATGCG TATGACGAAG GCTATATTCA GAGAGATCCA ACAAGGAAAA
       K  A  M   L  L  D   A  Y  D  E   G  Y  I   Q  R  D   P  T  R  K

8260© TTGTGGTTAA AGGAAAGGAG CCTTCCGAAA AGAAGGCAAA GTATTTGAAT GAATTCGAAT
       I  V  V   K  G  K   E  P  S  E   K  K  A   K  Y  L   N  E  F  E

8200© TAAAGTTGCT ACTTCGTCAC TTAGATCTGT CAGCTTTCCC AAATTTTGAT TGGATGATCC
       L  K  L   L  L  R   H  L  D  L   S  A  F   P  N  F   D  W  M  I

8140© TTTTAATTGC TAAAACAGGC TTGCGATTTA GCGAAGCTTT AGGGCTCACT AAAGAAGACA
       L  L  I   A  K  T   G  L  R  F   S  E  A   L  G  L   T  K  E  D

8080© TAGATTTAGA GCAACAAATG ATCAACGTAG ATAAGACATG GGACTACAAG AGCTATACGG
       I  D  L   E  Q  Q   M  I  N  V   D  K  T   W  D  Y   K  S  Y  T

8020© GGAGTTTCAA GCCGACAAAG AATTCGTCAT CAGTAAGAAA AGTTCCCATC GATTGGAAGC
       G  S  F   K  P  T   K  N  S  S   S  V  R   K  V  P   I  D  W  K

7960© TTGCCATGCA GCTCAATCAA GTCATTCAAG ATTTGCCTAA TGGAGAGCCT ATTTTTGCCC
       L  A  M   Q  L  N   Q  V  I  Q   D  L  P   N  G  E   P  I  F  A

7900© AAAAGCGAGT TTTTAACTCA ACTGTTAATA ATTTACTGAA GAAGCACTGT AAGGAGTTAA
       Q  K  R   V  F  N   S  T  V  N   N  L  L   K  K  H   C  K  E  L
```

7840© ATATCCCTGT TATCTCTGTA CACGGGCTAC GCCACACCCA TGCTTCGTTG CTTTTATTTG
N I P V I S V H G L R H T H A S L L L F

7780© CCGGTGTTTC TATTGCAAGT GTTGCGAATC GTTTAGGACA TGCTGATATG ACAACAACGC
A G V S I A S V A N R L G H A D M T T T

7720© AACAAACATA CCTACACATT ATTCAAGAAT TGGAGAACAA AGATAATACC AAAATAATGC
Q Q T Y L H I I Q E L E N K D N T K I M

7660© AGCACTTGGC TGCATTGTAA GGAAAAGGAT AGCGGCTTGG TATCGCTATC CTTTTAAACT
Q H L A A L - 

7600© ATATGAACAT ATTTTGGAGA AGGAATTTTT TGAGTTCCAA CAATTTTTCA T

VI. Nucleotide sequence of the *hsd* cluster from *L. lactis* strain NCC82.

```
   1  TAGCGGTCTT TTAAGCTTTG CAAGCGGCTG GCCCGGCGCA AAAGCAGGAC TAGGTCGACA

  61  TCCTTGAACT TGTCTTCAGT CAAAATTCTT TCTTCAACAA TGCGGGCAAT TTGACCGTTA

 121  GCTGCCAAAA TCAGTAATTT CATGTTTATT ATTTTCCTTT CTGCTTGTCG GATTCTCTTG

 181  ATGCTTTTAT GATAGCGGGG CGCAAGCTTT TTGAAAAATA CTGTTTGAGC AAGATGGTAT

 241  GCAGGAAGGA CATAGCAGAT ATCAAATACT GTAAATTCTG ATGTATATAA AAGATCATGA

 301  TTGCAGATAA CATAGGATCA GCCCTTTTTT CAGCCCTGTT TTTTTGAAAA TAGGCAATAA

                                      -35                          -10
 361  AGGTTGATCA AAAGGAAATG CTTTCACTTC AGTAGAGAAA AAGTTAAAAT AGTATTTATA

                                                        RBS          hsdR
 421  ATGCAATACA ATCAGCATCG TTAGGTAAGT TGCTGTTTAG GGAAGAGGAA AGCTATGACT
                                                                  M  T

 481  TTAGAATCAC AACTGGAAGA CAACCTAATC GCGCAGTTGA CGCAGGATGT CCATCAGTGG
        L  E  S   Q  L  E   D  N  L  I   A  Q  L   T  Q  D   V  H  Q  W

 541  AAGTTCCGTG ATGATTTACG CACCGTAGAT CAGCTTTGGG ACAACTTTTT CCGTATCTTG
        K  F  R   D  D  L   R  T  V  D   Q  L  W   D  N  F   F  R  I  L

 601  GAATCAAACA ATAAGGATCA GTTAAATGAC CATCCGTTGA CTCCTAACGA AAAAATGACG
        E  S  N   N  K  D   Q  L  N  D   H  P  L   T  P  N   E  K  M  T

 661  GTGAGAACGG CGATCGTTAA ACCGACCTTT TACCGGGCAA CGGAGTTTAT GGTCGGGGCC
        V  R  T   A  I  V   K  P  T  F   Y  R  A   T  E  F   M  V  G  A

 721  AACCGGCAGG TTCGCTATCA CTTAAGAAGA GAAGATTCTT CTATTCCCGA CGCTGATCTG
        N  R  Q   V  R  Y   H  L  R  R   E  D  S   S  I  P   D  A  D  L

 781  CTGATTTTAG ACAATACCAA CATTGCGGGT GGAAACTCAG TTTATGAAGT TGTACACCAG
        L  I  L   D  N  T   N  I  A  G   G  N  S   V  Y  E   V  V  H  Q

 841  GTTCAGCTAC AGAAGAAGAC CGCGCTTAAT CAAGACCGTC GTTTCGACGT TAGTTTGTTG
        V  Q  L   Q  K  K   T  A  L  N   Q  D  R   R  F  D   V  S  L  L

 901  ATCAACGGCT TGCCGGTAAT TCACATTGAG CTTAAAGCTC CAAATGTTTC TTATAAGAAG
        I  N  G   L  P  V   I  H  I  E   L  K  A   P  N  V   S  Y  K  K

 961  GCCTTTAACC AAATTCAAAA GTATATCGAC GAAGGACAAT TTACTGACAT TTACAGCTTC
        A  F  N   Q  I  Q   K  Y  I  D   E  G  Q   F  T  D   I  Y  S  F

1021  GTAGAAATGT TTGTGGTAAC TAATGGTGCT CAAACAAGAT ATATATCTGC TGGGCAGAAT
        V  E  M   F  V  V   T  N  G  A   Q  T  R   Y  I  S   A  G  Q  N

1081  TTGAATGCCA AGTTTTTAAC GGCCTGGGTT GATAAGAATA ATAAGCGGGT AGACAATTAT
        L  N  A   K  F  L   T  A  W  V   D  K  N   N  K  R   V  D  N  Y

1141  CTGAGTTTTG CAGAAGAGGT TTTGTCAATA CCTGCTGCTC ACCATATGAT TGCCGACTAT
        L  S  F   A  E  E   V  L  S  I   P  A  A   H  H  M   I  A  D  Y

1201  GTGGTTTTAG ACAGCGAAAG CAAGAGCGTT ATCCTGCTCC GTCCTTACCA GATACATGCG
        V  V  L   D  S  E   S  K  S  V   I  L  L   R  P  Y   Q  I  H  A

1261  ATTCAAGCGA TTTTTAAGGC TTCTAGAGAG AGTAAGTCGG GCTATATTTG GCATACGACA
        I  Q  A   I  F  K   A  S  R  E   S  K  S   G  Y  I   W  H  T  T

1321  GGGTCAGGTA AGACGTTAAC TTCGTACAAG GTTGCCCGTA ACTTGTTGCA AATTCCGTCA
        G  S  G   K  T  L   T  S  Y  K   V  A  R   N  L  L   Q  I  P  S

1381  ATTGATAAGT CAATCTTCCT AATCGACCGT AAAGACCTGG ACACGCAGAC TACAACTGCT
        I  D  K   S  I  F   L  I  D  R   K  D  L   D  T  Q   T  T  T  A
```

```
1441  TTCAAGATTT ACGCCAACAA CGATACGATC AGCGTTAACG AAACAAATAA TAGTTATGAC
        F  K  I   Y  A  N   N  D  T  I   S  V  N   E  T  N   N  S  Y  D

1501  CTTGCTGACC AGATGACTGA TGGCGACCGG ACTGTAGTAG TTACTACCCG CCAGAAGATT
        L  A  D   Q  M  T   D  G  D  R   T  V  V   V  T  T   R  Q  K  I

1561  CAAAACATGT TTAAGCGGAT TGATGAATTG GATCAGTTGC CTAAACGGTA TGAGAACTTA
        Q  N  M   F  K  R   I  D  E  L   D  Q  L   P  K  R   Y  E  N  L

1621  AAGAATATGC GGCTGGCCTT CATCGTCGAT GAATGTCACA GAACGATTAC CCCTAGCCAA
        K  N  M   R  L  A   F  I  V  D   E  C  H   R  T  I   T  P  S  Q

1681  AAGCGGGAGA TTGATAAATT CTTTAACCGC AAACCGCTGT GGTATGGCTT TACTGGTACG
        K  R  E   I  D  K   F  F  N  R   K  P  L   W  Y  G   F  T  G  T

1741  CCAATTTTTA ACGAGAATGC CCGGGCAAAG AATGGTCAAG ACGCGCGGAC GACTGAAGAA
        P  I  F   N  E  N   A  R  A  K   N  G  Q   D  A  R   T  T  E  E

1801  TTATATGGGC CAGTCCTGCA CAAGTATACG ATCGGGGATG CAATTAGAGA CAAGATGGTT
        L  Y  G   P  V  L   H  K  Y  T   I  G  D   A  I  R   D  K  M  V

1861  TTGGGCTTCT CCATCGATAA TCAAGGTGGC AGCAACGAGG ATGGAAATGA AGAAGATACC
        L  G  F   S  I  D   N  Q  G  G   S  N  E   D  G  N   E  E  D  T

1921  AAGAAAATGG ATCAGATTTA CCGGTCCAAG GCCCATATGC ATTCAGTTGC GACAGCAGTA
        K  K  M   D  Q  I   Y  R  S  K   A  H  M   H  S  V   A  T  A  V

1981  ATTAAAGCGG CGTACCGCAA GCAGGGCCTT ATTTCTGGTA AGAAATACTC AGCCATTTTT
        I  K  A   A  Y  R   K  Q  G  L   I  S  G   K  K  Y   S  A  I  F

2041  ACGACTTCGT CAATTGAGCA AGCCCAGAAG TACTACCGTA TCTTTAAGAA AATCATCGAT
        T  T  S   S  I  E   Q  A  Q  K   Y  Y  R   I  F  K   K  I  I  D

2101  GGGGAAGACG AGGAATTCAA GATTCCGGAA CGAATCAAGA AAGTGGCACC AGACTTCCCA
        G  E  D   E  E  F   K  I  P  E   R  I  K   K  V  A   P  D  F  P

2161  AAGATCGCTA TTACTTACTC AGTTAGTGAA AACGAAGATG ATTCAGAATC AGTGCAAGAT
        K  I  A   I  T  Y   S  V  S  E   N  E  D   D  S  E   S  V  Q  D

2221  GAGATGAAGC AGTCGCTTGC GGACTACAAC GCCGTTTATG GGACGAACTT CTCGATGGCT
        E  M  K   Q  S  L   A  D  Y  N   A  V  Y   G  T  N   F  S  M  A

2281  GAGCTGGATC AGTACAATCA AAACGTTAAT GCCCGGCTTG CCCGCAAGAA GGCTCAATAC
        E  L  D   Q  Y  N   Q  N  V  N   A  R  L   A  R  K   K  A  Q  Y

2341  CAAGCTGACA ACCAACGCTT AGACCTAGTG ATAGTTGTTA ACCGTTTGCT GACTGGTTTT
        Q  A  D   N  Q  R   L  D  L  V   I  V  V   N  R  L   L  T  G  F

2401  GACTCACCAA GTTTGTCGAC TTTGTACATT GACCGGCCGC CAATGAGTCC ACAGGACATT
        D  S  P   S  L  S   T  L  Y  I   D  R  P   P  M  S   P  Q  D  I

2461  ATCCAGGCTT TTTCTAGAAC CAACCGGATT TTTGATAAAG ACAAGACTTG GGGACAAATC
        I  Q  A   F  S  R   T  N  R  I   F  D  K   D  K  T   W  G  Q  I

2521  GTAACTTACC AGTATCCCAA GACCTTTAGT GAAAAGATTG ATGATGCGAT CGTCCTATAT
        V  T  Y   Q  Y  P   K  T  F  S   E  K  I   D  D  A   I  V  L  Y

2581  TCTAATGGTG GGGAGAAGTA TGCGGTTGCT CCAAGCTGGG AAGAATCAAA ACAAAGCTAC
        S  N  G   G  E  K   Y  A  V  A   P  S  W   E  E  S   K  Q  S  Y

2641  ATATCAGCCC GGTCTAAAAT TGAAATGTAC AGCTTCGACG CAGATGGACC GTCAATTTAT
        I  S  A   R  S  K   I  E  M  Y   S  F  D   A  D  G   P  S  I  Y

2701  GATGCTTCTA AGGAAGAAAA GAAGAAATTC GTTAAAGCTT TTCAAGAATT TGACAAGGCT
        D  A  S   K  E  E   K  K  K  F   V  K  A   F  Q  E   F  D  K  A

2761  TTGGCTGCAA TCAAGACCTA TGATGAATTA GATACCGAAG AAGGCCTAAT GCAATTAGGT
        L  A  A   I  K  T   Y  D  E  L   D  T  E   E  G  L   M  Q  L  G

2821  GTGAGCGACA TTACACCAGA TGATTGGGAA GCCATGCGGG GAGTTTATGA GGATATTCTG
        V  S  D   I  T  P   D  D  W  E   A  M  R   G  V  Y   E  D  I  L
```

```
2881  GACGATCTGC GGAAAGATCC AGATGAAGAT CCTGATCACA ATGACGACAT CGATGATGAA
        D  D  L   R  K  D   P  D  E  D   P  D  H   N  D  D   I  D  D  E

2941  TACGAACTTG AGTCATTTGG ACAAAAGGAA ATCGATGAGC GCTATATCAT GAATTTGATC
        Y  E  L   E  S  F   G  Q  K  E   I  D  E   R  Y  I   M  N  L  I

3001  CAGGCGTTCT TGCCAGAAAG TTCTGACAAT CAAGAAAAAG CTAGTGGCAA TGAAATCTCG
        Q  A  F   L  P  E   S  S  D  N   Q  E  K   A  S  G   N  E  I  S

3061  CCAGAAACTG TCAAGGAAAT CAATGGCTAC ATTGATGAAC TGGCGAAGAC AAACGAACTT
        P  E  T   V  K  E   I  N  G  Y   I  D  E   L  A  K   T  N  E  L

3121  TTGGCAGAAA TCATGCGTAA GCTTTGGCAA CAGATTTTGC AGGACCCGGC TAAATACGCA
        L  A  E   I  M  R   K  L  W  Q   Q  I  L   Q  D  P   A  K  Y  A

3181  GGCAAGCAGG TTGATGAGCT CCTTGAATCC TTGATTGACC AGGAATTGCA ATCAATCATG
        G  K  Q   V  D  E   L  L  E  S   L  I  D   Q  E  L   Q  S  I  M

3241  CGTGAGTTTG CTGATAAGTA CAAGGTTGAC TACGACCAGT TCCGCTACGT TATGGCTAAC
        R  E  F   A  D  K   Y  K  V  D   Y  D  Q   F  R  Y   V  M  A  N

3301  TATGACCCTA AGTTAAAGGG AAATAAGCAA AAGGGGATGA ACGACCTTTT GCACAAGGAA
        Y  D  P   K  L  K   G  N  K  Q   K  G  M   N  D  L   L  H  K  E

                        -35                         -10
3361  CGTTTTGTCG ACTACTTGAA TGACAATCCG GATTCTGACT TAAATAAGCC GTACCGCTGG
        R  F  V   D  Y  L   N  D  N  P   D  S  D   L  N  K   P  Y  R  W

3421  AAGAGTGAAG TTCGGGAACA GGCCAAGCAA TACTATGTCG ATAAGATTGG CCCGTTAATT
        K  S  E   V  R  E   Q  A  K  Q   Y  Y  V   D  K  I   G  P  L  I

                           RBS       hsdM
3481  AACAGAGAAG CATAGAAGGA AGAAGAATAT GGCAGAAGAA AATTCAACAG TTAGCTTGCA
        N  R  E   A  -
                                     M  A  E  E   N  S  T   V  S  L

3541  GAGTGGTTTG TTTGCAGCTG CAGACGTCTT GCGTTCAAAG ATGGACGCCA ATGAGTATAA

      Q  S  G   L  F  A  A   A  D  V   L  R  S  K   M  D  A   N  E  Y

3601  GAACTATCTT TTGGGGACTG TTTTCTATAA GTACCTTTCA GACCAGCAGC TTTACAAGTT

      K  N  Y   L  L  G  T   V  F  Y   K  Y  L  S   D  Q  Q   L  Y  K

3661  GGCTGAAGAC GCCGGTGAAG ATGACGTTAC TTTAGACGAA GCTCAGAAAA TCTATGAGGA

      L  A  E   D  A  G  E   D  D  V   T  L  D  E   A  Q  K   I  Y  E

3721  AAATCTTGAA GAAGAAGGCC TCCTAGATGA GGTTAAAGAC GAGCTTGGAT ATTTGATCGA

      E  N  L   E  E  E  G   L  L  D   E  V  K  D   E  L  G   Y  L  I

3781  ACCGGAGTAC ACATACACCA AGATCTTAGA TAATGCCAAC GATGGCAGTT TCCAACTCAA

      E  P  E   Y  T  Y  T   K  I  L   D  N  A  N   D  G  S   F  Q  L

3841  TCAGTTAGGG GATGCCTTTA ACAAGCTAGA AAGTCAAGGC AGCAGTTTTG AAGGCCTGTT

      N  Q  L   G  D  A  F   N  K  L   E  S  Q  G   S  S  F   E  G  L

3901  TGACGACTAT GACCTGTATT CAAAGCGGCT GGGTCAAAAC TTGCAGAAGC AGACAGATAC

      F  D  D   Y  D  L  Y   S  K  R   L  G  Q  N   L  Q  K   Q  T  D

3961  AATTGCCGGA GTGATTAAGG CGATCGGCAA ATTAGAACTG GTTAAGACTC CTGGTGACAC

      T  I  A   G  V  I  K   A  I  G   K  L  E  L   V  K  T   P  G  D

4021  CTTGGGGGAT GCTTACGAAT ACTTGATCAG TCAATTTGCC TCAGAATCAG GTAAGAAGGC

      T  L  G   D  A  Y  E   Y  L  I   S  Q  F  A   S  E  S   G  K  K
```

```
4081  CGGTGAATTC TATACTCCAC AAGAAGTTTC TGAACTTTTA GCACGGTTGA CTTTAGTCGG
      A  G  E  F  Y  T  P  Q  E  V  S  E  L  L  A  R  L  T  L  V
4141  TAAGGATTAC TCTTCTGGGA TGAGCGTTTA CGACCCGGCG ATGGGATCAG GTTCACTGCT
      G  K  D  Y  S  S  G  M  S  V  Y  D  P  A  M  G  S  G  S  L
4201  GCTGAACTTT AGAAAGTATG TTCCAAACTC ATCAAGAATT ACTTATTACG GGCAGGAAAT
      L  L  N  F  R  K  Y  V  P  N  S  S  R  I  T  Y  Y  G  Q  E
4261  CAACACGTCA ACCTTTAACT TGGCTAGAAT GAACATGATT TTGCACCACG TTGATCTGGC
      I  N  T  S  T  F  N  L  A  R  M  N  M  I  L  H  H  V  D  L
4321  AAACCAGAAG TTGAGAAACG GGGATACGTT AGACGAGGAC TGGCCCGCTG AAGAAACTAC
      A  N  Q  K  L  R  N  G  D  T  L  D  E  D  W  P  A  E  E  T
4381  CAATTTCGAC TCAGTTGTAA TGAACCCGCC ATATTCACTT AAATGGAGCG CGGACAAGGG
      T  N  F  D  S  V  V  M  N  P  P  Y  S  L  K  W  S  A  D  K
4441  CTTCTTGGAT GACCCACGTT TCTCTAAGTA TGGTGTTTTG CCGCCAAAGT CTAAGGCGGA
      G  F  L  D  D  P  R  F  S  K  Y  G  V  L  P  P  K  S  K  A
4501  CTATGCTTTC TTGCTTCACG GCTTCTACCA CTTGAAACAC AGTGGGGCAA TGGCTATCGT
      D  Y  A  F  L  L  H  G  F  Y  H  L  K  H  S  G  A  M  A  I
4561  TCTGCCACAC GGGATTCTTT TCCGTGGTGC AGCGGAAGGA AAGATCCGGC AAAAGTTGCT
      V  L  P  H  G  I  L  F  R  G  A  A  E  G  K  I  R  Q  K  L
4621  TGAAGAGGGC GCGATTGATG CAGTAATCGG TTTGCCTGCA AACTTGTTCT ACTCAACTGG
      L  E  E  G  A  I  D  A  V  I  G  L  P  A  N  L  F  Y  S  T
4681  TATCCCAACT ACTATTGTCG TTCTTAAGAA GGACAAGCAG GATCGGAACG TGCTGTTTAT
      G  I  P  T  T  I  V  V  L  K  K  D  K  Q  D  R  N  V  L  F
4741  TGACGCATCT AAGGAGTTCG AGAAAGTTAA GACTCAGAAC AAGCTGCGGC AAGAAGACAT
      I  D  A  S  K  E  F  E  K  V  K  T  Q  N  K  L  R  Q  E  D
4801  TGATAAGATC TTGAAGACCT ACGAAGAGCG GCCAGCAGAT GTTGAGAAGT ATGCCCACTT
      I  D  K  I  L  K  T  Y  E  E  R  P  A  D  V  E  K  Y  A  H
4861  GGCAAGTTTT GACGAAATCA AAGAAAATGA CTTCAACTTG AACATTTCTC GTTACGTTGA
      L  A  S  F  D  E  I  K  E  N  D  F  N  L  N  I  S  R  Y  V
4921  CACTTTTGAG CCAGAACCAG AAATTGATCT GCGAGACGTG GCTAAGGAAC TACGGGATAT
      D  T  F  E  P  E  P  E  I  D  L  R  D  V  A  K  E  L  R  D
4981  TGACCAACAG ATCAATGAAA ACGAGAAAGA ATTGGTAGGG ATGCTTAAGG AATTGACTTC
      I  D  Q  Q  I  N  E  N  E  K  E  L  V  G  M  L  K  E  L  T
                                                        RBS
5041  AAGTGACGAT GATATCATGG CAGGTTTGCA GAGCATCATC GAGAATTTTG AGGAGGAAAT
      S  S  D  D  D  I  M  A  G  L  Q  S  I  I  E  N  F  E  E  E
```

hsdS-3

```
5101  CCGCTAATGA AAGACGAAAA AAAGGCCCCT AAACTGCGGT TTAAAGGCTT TACTGACGAT
             M   K  D  E   K  K  A  P   K  L  R   F  K  G   F  T  D  D
      I  R  -

5161  TGGGAGCAAC GTAAGTTAGG GGATGTGTGT GAACCGATAA CAGATAGCAT TGATACTCAA
        W  E  Q   R  K  L   G  D  V  C   E  P  I   T  D  S   I  D  T  Q

5221  AAATACCCCA ACGAAGTATT TGCAGAATAT AGCATGCCTG CATTCGATGC ATCTATGAAG
        K  Y  P   N  E  V   F  A  E  Y   S  M  P   A  F  D   A  S  M  K

5281  CCAGATATTG TTTTAGGTTC TTCGATGAAT AGTGTTAGGA AAATAATTAC TAGACCATGC
        P  D  I   V  L  G   S  S  M  N   S  V  R   K  I  I   T  R  P  C

5341  TTATTAGTCA ACAAGCTAAA CGTGAGGAAA AAACGGATAT GGTATGTTAA AAAACCTAAT
        L  L  V   N  K  L   N  V  R  K   K  R  I   W  Y  V   K  K  P  N

5401  AAGAATGCAG TCTGTTCGGC AGAATTTATT CCTTTGCATT CAGATACCGT TGATTTAACT
        K  N  A   V  C  S   A  E  F  I   P  L  H   S  D  T   V  D  L  T

5461  TTCTTGAATC AAGTAGCCAA GAGTGAAACT TTTACAAGAT ATCTTGAAAA TCACTCATCT
        F  L  N   Q  V  A   K  S  E  T   F  T  R   Y  L  E   N  H  S  S

5521  GGCTCTTCAA ATAGCCAAAA GCGAATTACT CCACGATCAT TAATGCTTTC TAAGTTGCAT
        G  S  S   N  S  Q   K  R  I  T   P  R  S   L  M  L   S  K  L  H

5581  ATTCCAACAA TTGAGGAGCA GAAGTTAATT GGTAAGATCT TTGAATCCCT AGATCATACC
        I  P  T   I  E  E   Q  K  L  I   G  K  I   F  E  S   L  D  H  T

5641  ATCACTTTAC ATGAGGAAAA GAAACGCCAA CTTGAGCGCC TTAAAAGCGC TTTATTGCAG
        I  T  L   H  E  E   K  K  R  Q   L  E  R   L  K  S   A  L  L  Q

5701  AAGATGTTTG CTGATGAGAG CGGGTATCCT GTTGTTAGGT TTGAAGGATT TAGTGACGAG
        K  M  F   A  D  E   S  G  Y  P   V  V  R   F  E  G   F  S  D  E

5761  TGGGAACAGC GTAAGTTAAA AGATGTCGTT GAAAAGCAAA TAAAAGGTAA GGCTCAACTT
        W  E  Q   R  K  L   K  D  V  V   E  K  Q   I  K  G   K  A  Q  L

5821  GAAAAATTAG CACCGGGAGA GGTTGAATAT CTTGATACAT CAAGGTTGAA TGGTGGCCAA
        E  K  L   A  P  G   E  V  E  Y   L  D  T   S  R  L   N  G  G  Q

5881  GCAATATTAA CTAACGGTCT TAAAGATGTA ACGTTAGATG ATATTTTAAT TCTCTGGGAT
        A  I  L   T  N  G   L  K  D  V   T  L  D   D  I  L   I  L  W  D

5941  GGATCAAAAG CAGGAACTGT TTATCATGGA TTTGAAGGGG CCTTAGGCTC TACACTCAAG
        G  S  K   A  G  T   V  Y  H  G   F  E  G   A  L  G   S  T  L  K

6001  GCTTACAGAA CCTCTGCTAA TTCAAAATTT GTTTACCAAT ATTTGAAACG TCACCAAGAT
        A  Y  R   T  S  A   N  S  K  F   V  Y  Q   Y  L  K   R  H  Q  D

6061  AATATTTATA ACAATTATCG GACGCCTAAT ATACCGCACG TCCAAAAAGA CTTTCTGAAT
        N  I  Y   N  N  Y   R  T  P  N   I  P  H   V  Q  K   D  F  L  N

6121  GTATTTACGA TTAGCGTACC CGTGAGTGAT GAGCAAGAAA AAATAGGTTC ATTCTTTAAA
        V  F  T   I  S  V   P  V  S  D   E  Q  E   K  I  G   S  F  F  K

6181  CAGCTAGATG ACACTATCGA TCTTCATCAA CGTAAGTTAG ATTTGTTGAA AGAACAGAAA
        Q  L  D   D  T  I   D  L  H  Q   R  K  L   D  L  L   K  E  Q  K

6241  AAAGGCTTTT TACAAAAGAT GTTCGTTTAG GGTCTATAAT TATAAAATCG CCCCTAAAAA
        K  G  F   L  Q  K   M  F  V  -

6301  ATACTTTTAA AACAGCCCTA AAAACGGGAG TACAAATATG ACAACATAGA CTGAACGTCT

6361  GTAAGACGCT TGGCTTGATT AAAGCAGGTT CAGCCCGTAT AAATAGGTGG GAGTATAGCA

6421  AAATTGTAAA AAACGAATTT ATCGTGCTTT ACCCCCTGGC GCCCGCTCCC AAACGGTCGT

6481  TCTCGCCAAA TGTTGACCCA GGGGGGTATC CTCAGCTCCG GTCATGGGCA AAATTCTGCC

6541  AGGCAAAGCT AAAACGCCTC AGAACGCAAA TATGGGCCTT CTAGGGCTAT CTCAGAGCAT
```

```
6601  GACATTCAAC TGCACTCAGA GCTGCAAGAA TTGCGGCTCT TTTTGTTTGC CTTGAGTTTT

6661  GACTGTCGAA TTTTTGATCG TCAACTTTAA AGGCCCAGAC TTTTATCTCG GTCACTCCCT

6721  GGCGAACCTG GACCAGGCGG TCGGTGATAT CCACACGGAC AGCAGAGAAA ACAAAAGCCA

                      hsdS-4'
6781  GCCATTCACT GGCTAGCAGT GTAAGTTGGG GGATGTGGCA AAAATCACCA TGGGTCAATC
                        Q   C   K   L   G   D   V   A   K   I   T   M   G   Q

6841  ACCAAATTCT AAAAATTACA CTGATAATCC AAAAGATCAC ATTCTCGTTC AAGGAAATGC
      S   P   N   S   K   N   Y   T   D   N   P   K   D   H   I   L   V   Q   G   N

6901  AGATATGAAG GATGGTCAAG TACATCCTAG AATTTGGACT ACGGAAATAA CAAAAATCGC
      A   D   M   K   D   G   Q   V   H   P   R   I   W   T   T   E   I   T   K   I

6961  AGACAAAGGT GATTTAATCT TAAGCGTTAG AGCACCTGTA GGCGATATTG GAAAAACAAG
      A   D   K   G   D   L   I   L   S   V   R   A   P   V   G   D   I   G   K   T

7021  CTATGATGTG GTTATCGGTC GTGGAGTAGC CGCCATCAAA GGTAATGAAT TTATTTTTCA
      S   Y   D   V   V   I   G   R   G   V   A   A   I   K   G   N   E   F   I   F

7081  ACTGCTTAAA CGAATGAAAA CAGTAGGTTA TTGGACTAAA TATAGTACTG GTTCAACATT
      Q   L   L   K   R   M   K   T   V   G   Y   W   T   K   Y   S   T   G   S   T

7141  TGAGAGCATT AATTCATTAG AAATAAATAA TGCGGTAATA AATCTCCCTA AGGAGCATGA
      F   E   S   I   N   S   L   E   I   N   N   A   V   I   N   L   P   K   E   H

7201  ACAAAATGAG GTTGGAAAGA TACTTTCCTA TATGGATCAT GCCATCACTT TACATGAGGA
      E   Q   N   E   V   G   K   I   L   S   Y   M   D   H   A   I   T   L   H   E

7261  AAAGAAATGC TAACTTGAGC GCTTAAAAAG CGCTTTACTG CAGAAGTTGT TCGCTGATAA
      E   K   K   C   -   L   E   R   L   K   S   A   L   L   Q   K   L   F   A   D

7321  GAGCGGGTAT CCGGCAGTTC GGTTTAAGGG CTTTGGTGGT AACTGGAAAG AGCGAAAATT
      K   S   G   Y   P   A   V   R   F   K   G   F   G   G   N   W   K   E   R   K

7381  TGGTGAGCTT GGAACCATAG AGATGTGCAA ACGAATTTTT AAAGATCAAA CAACCGATGC
      F   G   E   L   G   T   I   E   M   C   K   R   I   F   K   D   Q   T   T   D

7441  TGGAGAAATC CCCTTCTATA AAATTGGAAC ATTTGGCGGG ATGCCTAATG CATATATATC
      A   G   E   I   P   F   Y   K   I   G   T   F   G   G   M   P   N   A   Y   I

7501  TAGAGAAATT TTTGAAAGAT ATAAAAGATT ATATCCCTAT CCTACAAAGG GGGATATCTT
      S   R   E   I   F   E   R   Y   K   R   L   Y   P   Y   P   T   K   G   D   I

7561  AATTTCTACT TCTGGTAGTA TTGGACGAGT CGTTGAATAC ACCGGCCAAG AAGCCTATTA
      L   I   S   T   S   G   S   I   G   R   V   V   E   Y   T   G   Q   E   A   Y

7621  TCAAGATTCA AATATCGTGT GGTTAAATCA CAACGATTTA ATTAGTAATT CCTTTTTAAA
      Y   Q   D   S   N   I   V   W   L   N   H   N   D   L   I   S   N   S   F   L

7681  GTATCTGTAC AAAACTTTTA ATTGGAACGG TGTTGAAGGA AGTACGATTA AACGTTTATA
      K   Y   L   Y   K   T   F   N   W   N   G   V   E   G   S   T   I   K   R   L

7741  CAATAACAAT ATATTGAAAA CAAAGATCAG AATTCCATCC AGTTGTGAAC AAGAGAAAAT
      Y   N   N   N   I   L   K   T   K   I   R   I   P   S   S   C   E   Q   E   K

7801  AGGGGGAATT TTGTCTACTC TCGATCATCT TCTTAGCCAT CATCAACAAA AAATAGATAT
      I   G   G   I   L   S   T   L   D   H   L   L   S   H   H   Q   Q   K   I   D

                                                      -35         -10
7861  AGTAAAGTTA ATCAAGCAAT CTCTACTACA AAACATGTTC ATCTAATCTA AAATCTCCCT
      I   V   K   L   I   K   Q   S   L   L   Q   N   M   F   I   -

                        RBS         int
7921  AGCAAGATAT TTGCTAAGGA GATTTTTATG CGTAGGAAAC AAATTTTACT ACATGATTAT
                                    M   R   R   K   Q   I   L   L   H   D   Y

7981  TTTGCCCAAT GGATTGAGGT ATACAAGGAT GGAGCAGTTC GAGAAAGGAC TCTAGATAAG
        F   A   Q   W   I   E   V   Y   K   D   G   A   V   R   E   R   T   L   D   K
```

```
8041  TACTGGCTAT CTCATCGCCA TTTACAAGAA ATTGCGCCCA ATCTAAAATT AGTCGATACT
        Y  W  L   S  H  R   H  L  Q  E   I  A  P   N  L  K   L  V  D  T

8101  ACAAGACTTG AATATCAGCA AATCTTAAAT ACTTTTGCTC AAACTCATGA GAAAGCAACA
        T  R  L   E  Y  Q   Q  I  L  N   T  F  A   Q  T  H   E  K  A  T

8161  GTAATGGATT TTCACCATCA GCTGAAGGCT ATGTTGCTAG ATGCGTATGA CGAAGGCTAT
        V  M  D   F  H  H   Q  L  K  A   M  L  L   D  A  Y   D  E  G  Y

8221  ATTCAGAGAG ATCCAACAAG GAAAATTGTG GTTAAAGGAA AGGAGCCTTC CGAAAAGAAG
        I  Q  R   D  P  T   R  K  I  V   V  K  G   K  E  P   S  E  K  K

8281  GCAAAGTATT TGAATGAATT CGAATTAAAG TTGCTACTTC GTCACTTAGA TCTGTCAGCT
        A  K  Y   L  N  E   F  E  L  K   L  L  L   R  H  L   D  L  S  A

8341  TTTCCAAATT TTGATTGGAT GATCCTTTTA ATTGCTAAAA CGGGCTTGCG ATTTAGCGAA
        F  P  N   F  D  W   M  I  L  L   I  A  K   T  G  L   R  F  S  E

8401  GCTTTAGGGC TCACTAAAGA AGACATAGAT TTAGAGCAAC AAATGATCAA CGTTGATAAG
        A  L  G   L  T  K   E  D  I  D   L  E  Q   Q  M  I   N  V  D  K

8461  ACATGGGACT ACAAGAGCTA TACGGGGAGT TTCAAGCCGA CAAAGAATTC GTCATCAGTA
        T  W  D   Y  K  S   Y  T  G  S   F  K  P   T  K  N   S  S  S  V

8521  AGAAAAGTTC CCATCGATTG GAAGCTTGCC ATGCAGCTCA GTCAAGTCAT TCAAGATTTG
        R  K  V   P  I  D   W  K  L  A   M  Q  L   S  Q  V   I  Q  D  L

8581  CCTAATGGAG AGTCTATTTT TGCCCAAAAG CGAGTTTTTA ACTCAACAGT TAATAATTTA
        P  N  G   E  S  I   F  A  Q  K   R  V  F   N  S  T   V  N  N  L

8641  CTGAAGAAGC ACTGTAAGGA GTTAAATATC CCTGTTATAT CTGTACACGG GCTACGCCAC
        L  K  K   H  C  K   E  L  N  I   P  V  I   S  V  H   G  L  R  H

8701  ACCCATGCTT CGTTGCTTTT ATTTGCCGGT GTTTCTATTG CAAGTGTTGC GAAGCGTTTA
        T  H  A   S  L  L   L  F  A  G   V  S  I   A  S  V   A  K  R  L

8761  GGGCATGCTG ACATGACGAC AACACAGCAA ACATACCTAC ACATTATTCA AGAATTGGAG
        G  H  A   D  M  T   T  T  Q  Q   T  Y  L   H  I  I   Q  E  L  E

8821  AACAAAGATA ATACTAAGAT TATGCAGCAC TTGGCTGCAT TGTAAGGAAA AGGATAGCGA
        N  K  D   N  T  K   I  M  Q  H   L  A  A   L  -

8881  CTTGGTGTCG CTATCCTTTT CAAACTATAT GAACATTTTT TGTAACAATG ACTTCTTGAG

8941  TTCCGTTAGA ATTTCTAGCT TATTCTGATA AAGGCGTGTA AGTTCTTCCA CTATCCTAAT

9001  AATTCTTTCT ATTTTTTGTT GTTCGTCTAA TGATATCGGA ATTACCATTA TTAGCTGTGA

9061  TAGTGTACTC TGGCTTACTC CTTTAGCCGT TCCACCACTA GCAAGTGTAG AAAGAGCTCT

9121  TTGTACTTTT GATGATCTTA AAACTACTGC TAAAAAATTA TCTGTAATTT CACTGTTATT

9181  TGAAACAAAA GAAATTGTTC TTTGACTTAA AACATATCCA CTATCATCTG GAACTTGCGC

9241  AACATTTCCC ATTGGAGCTT CTGTTGTAAA AAGTACTTGC CCCTTTAGTA ATTCATTTCC

9301  AGACATCCAT TTCTTATACA GAATTTCATT TGCGTAATGT GCATCAGTAG AAAAATCTAT

9361  GAACCCATCT TTCACATTCA AGGCAGAGAG GGCCAAATGT CCTCTTTCAC TCCAATTCAG

9421  TCCCATTTTT TTAGGAGTAC GACCTCTAAA ATCGATAATT TTTACAATTA CATCGAACAA

9481  CTTTCTGCTT TCCCATGCTT CACTAAAATC TTTAAACCGA ACAGCAGGAT TTCCATTCTC

9541  ATCAGCGAAC AACTTCTGCA GTAAAGCGCT TTTTAAGCGC TCAAGTTGGC ATTTCTTTTC

9601  CTCATGTAAA GTGATGGTCC TATCAAGCTC GTGAAGTAAT GTAGCAATTT TATTTTGTTC

9661  ATCAAAATTC GGAAAAGATA GCTTTAAAGT ACTAAACTGA TTAAAGCTGA TTGCTTTGCC

9721  ATCTCGTATT CCAAAAGTTA ATCTTTTTAG TGATTCAATG AAATCATATC GCTTAAAAAG
```

```
 9781  ATACTTCCAA AAAGTACTAT TCTGTTTATT TTTCTCTTGA TTATTAAATG CGAAGACAGT

 9841  ATACGCAGGG CTGATAATTC CTCGTTTTTC TGATAATTCA AATCCACCTT GAAATGATCT

 9901  AAGACTTATT ATGTAATCTC CGGGTGATAC AACCTTATAG TTACTCAATG CCGTGTGTGT

 9961  GTCAAATTTA ATGTCGATAT CCAAATCATT TCGAAAAACT ATGCCATTCT CTTGAGTTGC

10021  AGAAAGGACT GGCAAATCAC TGTGATCTTT CTTAGATTTA CTTTCAAATA ATTCCTTTGC

10081  TTTACACTGC TCCCAAGAAA AATTACATCC CTTCCGCCTG TTCCTGCATG ACCTGGCAGA

10141  AAGCGAGGAC CAAGGGATCT GCTTCATCTT TCCGGCAGAC CAGGCCATAG TCCAGCGGGG

10201  CGGGGTAGTT CAACTTTACC GCCTTGATCA CATCTGCTTT TTCGATCGCC ACAAAACTTG

10261  GCATCACGGT GATCCCTAGA CCCGCCTTGA CCATTAAGCT GACGATTTCA ACATCGTTAA

10321  CGTAGTTGAT GTCCAAGTCA TCGACCTTCT TTCTGATCGT CTCCTGGAGC TTCAGTTGTT

10381  CTGGCGGGCA CCAGTTGTTG TCCATCAGCA AAATCTTTTC TCCCGCCAAA TCATCCAGGG

10441  CAAGTTCCTC TCTGTCACTT AAATGATGGT CTGCAGGTAC AATTGCCACA AAGTGTCCGC

10501  TAAAAAGATC ATAGTACTTC ACCTCGGCTA GATCAGTGAT ATCATCATGG GTCGTAAAAA

10561  TCACATCGCT GTCGTGGTCA ATCAAGTGGT GCTTGAGCCG GTTATGATCA AATCCCTCAG

10621  AAAAATCTTG CAATTCGGGT ATTCCCGGTG AAAAGCTTTG ATCATCTCTG GCAGAATTGC

10681  TTGCTCTAGC GGGGTATCCG TAATACCAAT CGTTAAGTTG CTTTGCTCCC GGTTGAAAGA

10741  ATTTCTGGTT CGCTGCAGGG CCTTATTATA GCTGTTGATC AAGCTCTTGC TATCATCGCA

10801  AAACAGCTTC CCATTCTTAG TCAGCTGTAC TTCCCGGTGG TTACGGTATG ATGCCAACTT

10861  CGCCTTCTAT CCGCTGAATG ATCCGGCTAA CGGCCGGTTG AGAAAGGTGC AGGTTCTCCG

10921  CGGTCTTTGA ATAGTTCAAA CTCTCCGCCA AATTGATAAA AACTTCCAAC CGCCGATTGT

10981  AAAATTAAAC TGAACACTGT TCCGATCCAG TAAGAAAAAT CGTTTTCACT GGATTAGAGC

11041  ACAAAAAATC CATTTCAACC ATCTGGTAGA ATATGAATTA CCACAAACAT TTCTAG
```

Complemented nucleotide sequence of the *hsdS*-5 gene.

```
                -35             -10
10320© TTAACGATGT TGAAATCGTC AGCTTAATGG TCAAGGCGGG TCTAGGGATC ACCGTGATGC

10260© CAAGTTTTGT GGCGATCGAA AAAGCAGATG TGATCAAGGC GGTAAAGTTG AACTACCCCG

            RBS    hsdS-5
10200© CCCCGCTGGA CTATGGCCTG GTCTGCCGGA AAGATGAAGC AGATCCCTTG GTCCTCGCTT
                    M  A   W  S  A  G   K  M  K   Q  I  P   W  S  S  L

10140© TCTGCCAGGT CATGCAGGAA CAGGCGGAAG GGATGTAATT TTTCTTGGGA GCAGTGTAAA
       S  A  R   S  C  R   N  R  R  K   G  C  N   F  S  W   E  Q  C  K

10080© GCAAAGGAAT TATTTGAAAG TAAATCTAAG AAAGATCACA GTGATTTGCC AGTCCTTTCT
       A  K  E   L  F  E   S  K  S  K   K  D  H   S  D  L   P  V  L  S

10020© GCAACTCAAG AGAATGGCAT AGTTTTTCGA AATGATTTGG ATATCGACAT TAAATTTGAC
       A  T  Q   E  N  G   I  V  F  R   N  D  L   D  I  D   I  K  F  D

 9960© ACACACACGG CATTGAGTAA CTATAAGGTT GTATCACCCG GAGATTACAT AATAAGTCTT
       T  H  T   A  L  S   N  Y  K  V   V  S  P   G  D  Y   I  I  S  L

 9900© AGATCATTTC AAGGTGGATT TGAATTATCA GAAAAACGAG GAATTATCAG CCCTGCGTAT
       R  S  F   Q  G  G   F  E  L  S   E  K  R   G  I  I   S  P  A  Y
```

```
9840© ACTGTCTTCG CATTTAATAA TCAAGAGAAA AATAAACAGA ATAGTACTTT TTGGAAGTAT
       T  V  F   A  F  N   N  Q  E  K   N  K  Q   N  S  T   F  W  K  Y

9780© CTTTTTAAGC GATATGATTT CATTGAATCA CTAAAAAGAT TAACTTTTGG AATACGAGAT
       L  F  K   R  Y  D   F  I  E  S   L  K  R   L  T  F   G  I  R  D

9720© GGCAAAGCAA TCAGCTTTAA TCAGTTTAGT ACTTTAAAGC TATCTTTTCC GAATTTTGAT
       G  K  A   I  S  F   N  Q  F  S   T  L  K   L  S  F   P  N  F  D

9660© GAACAAAATA AAATTGCTAC ATTACTTCAC GAGCTTGATA GGACCATCAC TTTACATGAG
       E  Q  N   K  I  A   T  L  L  H   E  L  D   R  T  I   T  L  H  E

9600© GAAAAGAAAT GCCAACTTGA GCGCTTAAAA AGCGCTTTAC TGCAGAAGTT GTTCGCTGAT
       E  K  K   C  Q  L   E  R  L  K   S  A  L   L  Q  K   L  F  A  D

9540© GAGAATGGAA ATCCTGCTGT TCGGTTTAAA GATTTTAGTG AAGCATGGGA AAGCAGAAAG
       E  N  G   N  P  A   V  R  F  K   D  F  S   E  A  W   E  S  R  K

9480© TTGTTCGATG TAATTGTAAA AATTATCGAT TTTAGAGGTC GTACTCCTAA AAAAATGGGA
       L  F  D   V  I  V   K  I  I  D   F  R  G   R  T  P   K  K  M  G

9420© CTGAATTGGA GTGAAAGAGG ACATTTGGCC CTCTCTGCCT TGAATGTGAA AGATGGGTTC
       L  N  W   S  E  R   G  H  L  A   L  S  A   L  N  V   K  D  G  F

9360© ATAGATTTTT CTACTGATGC ACATTACGCA AATGAAATTC TGTATAAGAA ATGGATGTCT
       I  D  F   S  T  D   A  H  Y  A   N  E  I   L  Y  K   K  W  M  S

9300© GGAAATGAAT TACTAAAGGG GCAAGTACTT TTTACAACAG AAGCTCCAAT GGGAAATGTT
       G  N  E   L  L  K   G  Q  V  L   F  T  T   E  A  P   M  G  N  V

9240© GCGCAAGTTC CAGATGATAG TGGATATGTT TTAAGTCAAA GAACAATTTC TTTTGTTTCA
       A  Q  V   P  D  D   S  G  Y  V   L  S  Q   R  T  I   S  F  V  S

9180© AATAACAGTG AAATTACAGA TAATTTTTTA GCAGTAGTTT TAAGATCATC AAAAGTACAA
       N  N  S   E  I  T   D  N  F  L   A  V  V   L  R  S   S  K  V  Q

9120© AGAGCTCTTT CTACACTTGC TAGTGGTGGA ACGGCTAAAG GAGTAAGCCA GAGTACACTA
       R  A  L   S  T  L   A  S  G  G   T  A  K   G  V  S   Q  S  T  L

9060© TCACAGCTAA TAATGGTAAT TCCGATATCA TTAGACGAAC AACAAAAAAT AGAAAGAATT
       S  Q  L   I  M  V   I  P  I  S   L  D  E   Q  Q  K   I  E  R  I

9000© ATTAGGATAG TGGAAGAACT TACACGCCTT TATCAGAATA AGCTAGAAAT TCTAACGGAA
       I  R  I   V  E  E   L  T  R  L   Y  Q  N   K  L  E   I  L  T  E

8940© CTCAAGAAGT CATTGTTACA AAAAATGTTC ATATAGTTTG AAAAGGATAG CGACACCAAG
       L  K  K   S  L  L   Q  K  M  F   I  -
                                        ------------------------>    <

8880© TCGCTATCCT TTTCCTTACA ATGCAGCCAA GTGCTGCATA ATCTTAGTAT TATCTTTGTT
     ----------

8820© CTCCAATTCT TGAATAATGT G
```

This study was performed between November of 1996 and October 2000 at the Nestlé Research Center, Nestec Ltd., Lausanne, Switzerland. It was promoted by Univ.-Prof. Dr. T. A. Bickle, Dept. of Microbiology, Biozentrum, Universität Basel, Basel, Switzerland.

I would like to express my gratitude to:

- Univ.-Prof. T. A. Bickle, who accepted to promote this thesis and act as a reviewer. I would also like to thank him for his help with the organisational aspects associated with its realisation.

- Univ.-Prof. W. M. de Vos for immediately accepting to act as a reviewer and for the scientific discussions, which were a great help during those four years.

- Dr. Beat Mollet, Dr Andrea Pfeiffer and Dr Werner Bauer for giving me the opportunity to conduct this work within the excellent conditions of the Molecular Biology group of the Nestlé Research Center.

- Dr. Beat Mollet and R. David Pridmore for their supervision of my work and for interesting scientific discussions and helpful suggestions. I would like to thank R. David Pridmore for his great help and assistance with the molecular biology side of this project.

- Dr. Elaine Vaughan (now at Wageningen University, NL) for her interest in my work and very interesting and lively discussions.

- Dr. Nadji Rehkif (now PTC Orbe) and Dr. Frederic Chavagnat for their help and technical support with the protein work, i.e. column chromatography and SDS-PAGE.

- Dr. Michael Casey for his collaboration and his help in the study of the *L. lactis* plasmids.

- Dr. Marie-Lise Dillmann for her microscopy work and the very nice micrographs of lactobacilli cells she produced for me.

- Michèle Delley and Hengameh van der Kaaij for their lively interest in the evolution of my work, their technical assistance and especially, their constant encouragement and moral support.

- To everybody in the 'Molecular Biology' group, the 'Enzymology' group and the 'Fermentation' group for the great working atmosphere and the close camaraderie.

- Nestlé and the Swiss National Fund for financial support of this work.

Finally, I particularly want to thank my parents for their support and encouragements. I could never have achieved this doctorate without them. Papa, Maman, merci.

CURRICULUM VITAE

Name:	Aude Amanda BOURNIQUEL
Date of birth:	20.06.1972
Place of birth:	Lorient (56), France
Nationality:	French
Personal address:	36, rte de Trébale F-44600 St-Nazaire France

EDUCATION:

1978-1983	Primary school in St-Nazaire (44), France
1983-1987	Secondary school in St-Nazaire (44), France
1987-1990	High school in St-Nazaire (44), France
1990-1992	Diploma (DUT) in Biology Applied to the Food Industry at the polytechnique (IUT) of Montpellier (34), France
1992-1993	BSc. (Hons) in Applied Biochemistry at the University of Central Lancashire (Preston, UK)
1993-1996	Masters degree studies ('ingénieur en Biotechnologie') at the European College of Biotechnology, Strasbourg (67), France.
Since 1996	Ph.D. thesis at the Nestlé Research Center (Lausanne, CH) under the supervision of Dr. Beat Mollet. Thesis title: " Molecular insights into the metabolism and physiology of the lactic acid bacterium *Lactobacillus delbrueckii* subsp. *lactis.*"

WORK EXPERIENCE:

1992	3-month work experience as a microbiologist at the WHITBREAD BEER COMPANY in Salmesbury (Lancs., UK).
1993	6-week stage in the laboratory of Prof. T.A. Bickle at the Biozentrum (Basel, CH).
1996	8-month diploma work at the firma Hoffmann-LaRoche (Basel, CH) in the department of Vitamins and Fine Chemicals (Molecular Biology).