The Cloning of a Putative Regulatory Gene and the *sol* Region from *Clostridium beijerinckii*

Rui Hong

Thesis submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Master of Science
in
Biochemistry

Jiann-Shin Chen, Chairman
Dennis R. Dean
Timothy J. Larson
Ann M. Stevens

August, 1999
Blacksburg, Virginia

The Cloning of a Putative Regulatory Gene and the *sol* Region from *Clostridium beijerinckii*

Rui Hong

(Abstract)

The solvent-producing clostridia are well known for their ability to produce acetone, butanol and isopropanol in industrial fermentation. Production of these compounds occurs in cells that have completed a metabolic switch under specific growth conditions. Knowledge of the regulation of the metabolic switch will make the industrial process more reliable. From an isopropanol-producing strain *Clostridium beijerinckii* NRRL B593, a gene which encodes a putative NtrC-like regulatory protein was cloned and sequenced. The gene codes for a polypeptide of 632 amino acids and has been designated the *stc* gene. Expression of the *stc* gene was confirmed by RT-PCR. The co-presence of the *stc* gene with the *adh* gene which encodes a primary/secondary alcohol dehydrogenase in isopropanol-producing clostridia suggests that the *stc* gene may be functionally related to isopropanol production.

From *C. beijerinckii* NRRL B592, a region which encompassed the solvent-production genes *ald* (aldehyde dehydrogenase), *ctfA* and *ctfB* (acetoacetate: butyrate/acetate CoA-transferase) and part of *adc* (acetoacetate decarboxylase) was cloned and sequenced. The organization of these genes was similar to that in *C. beijerinckii* NRRL B593. Northern analysis indicated that these four genes were co-transcribed on the same messenger RNA in *C. beijerinckii* NRRL B593. Therefore, in *C. beijerinckii*, the *sol* operon consists of the *ald -ctfA-ctfB-adc* genes, which differs from the *sol* operon in *Clostridium acetobutylicum*.

Acknowledgements

I wish to sincerely express my appreciation to my advisor, Dr. Jiann-Shin Chen, for his concerned mentorship and enormous support, both within and outside the research field.

I wish to thank the other members of my graduate committee, Dr. Dennis R. Dean, Dr. Timothy J. Larson, and Dr. Ann M. Stevens for making their time and expertise available to me. I also hope that my colleagues understand how grateful I am to their friendship and help from the first moment I entered this group. They are Dr. Gary Li, Dr. Adnan A. Ismaiel, Julianna Toth and Murat Kasap.

Thanks also to those friends and colleagues from other groups or even other departments for their support and enlightenment; their names are far more than can reasonably be mentioned here.

List of Abbreviations

ADC	acetoacetate decarboxylase
Aad/AdhE	aldehyde/alcohol dehydrogenase
ADH	alcohol dehydrogenase
ALDH	aldehyde dehydrogenase
CtfA, CtfB	subunits of acetoacetate: butyrate/acetate CoA-transferase
MW	molecular weight
Stc	the predicted polypeptide encoded by the putative *stc* gene
adc	gene encoding acetoacetate decarboxylase
aad/adhE	gene encoding aldehyde/alcohol dehydrogenase
adh	gene encoding alcohol dehydrogenase
ald	gene encoding aldehyde dehydrogenase
ctfA, ctfB	genes encoding subunits of acetoacetate : butyrate/acetate CoA-transferase
stc	the putative gene for a transcriptional regulatory protein

*** Due to the software I am using, the names of the genes in figures are not in italics. I apologize for the inconvenience caused.**

Table of Contents

Chapter I: Literature Review

Chapter II: The cloning and sequence analysis of the *stc* gene from *C. beijerinckii* NRRL B593

Chapter III: The cloning and sequence analysis of the *sol* operon from *C. beijerinckii* NRRL B592

Chapter IV: Concluding Remarks....55

List of Figures

Chapter I

Chapter II

Chapter III

List of Tables

Chapter I

Chapter II

Chapter III

Chapter I

Literature Review

Introduction

The genus *Clostridium* is a diverse group of obligately anaerobic, usually Gram-positive, rod-shaped, endospore-forming, and non-sulfate-reducing bacteria. By these five criteria, over 100 species have been placed in this group and many of them are genetically poorly characterized. This highly heterogeneous group comprises organisms with considerable variation in genome size (2.9 to 6.7 Mb) and DNA G+C content (22 mol% for *C. botulinum* type E and 55 mol% for *C. thermoautotrophicum*)[1]. Historically, the major research topics concerning *Clostridium* included pathogenicity, food poisoning, nitrogen fixation, acetone-butanol fermentation and polysaccharide hydrolysis. *C. botulinum, C. tetani, C. perfringens, C. acetobutylicum, C. pasteurianum,* and *C. thermocellum* are the better known species within the genus.

Acetone, butanol and isopropanol (solvent) production by the clostridia is the focus of my research. It is well known that these solvent-producing clostridia undergo a metabolic switch during growth. At the first stage of growth, the bacteria metabolize the sugary substrate to produce the volatile fatty acids, acetate and butyrate. These products accumulate in the medium and cause a pH drop. Then, during the second stage of growth, bacteria re-assimilate the acidic products and metabolize them into neutral solvents

acetone and butanol. At this time, the carbon and the electron sources are directed to the solvent products instead of acetate and butyrate. This clostridial fermentation has been described as the AB (acetone and butanol) fermentation.

The AB fermentation played a very important role in the production of bulk chemical feedstock. The successful industrial process was first developed by Chaim Weizmann in the early 1910s (later, he became the first president of the state of Israel.), and it quickly expanded into a large-scale industrial operation. During the First World War, acetone produced by the AB fermentation was utilized for the manufacture of cordite. After the war, the other fermentation product, butanol was used as a source of the solvent for lacquers in the automobile industry. Until now, clostridial fermentation has been one of the largest biotechnological processes known[2,3].

After the 1950s, this clostridial fermentation declined dramatically due to the competition of the petrochemical industry and the increased cost for the starting materials. Most of the fermentation facilities were closed or converted for ethanol fermemtation[3]. However, in the past two decades, the clostridia have redrawn attention for several reasons. The availability of petrochemicals is limited and their exploitation has evoked concern towards the environment. The clostridia show the capability to grow on a wide variety of substrates such as molasses, corn mash and whey; some of which would otherwise be disposed of as waste. With an increased understanding of the physiology of the fermentation and the tools of genetic engineering available, it is now possible to alter the substrate specificity or improve the efficiency of solvent production. A re-establishment of a clostridial fermentation plant is underway[3]. Also, the versatility of the biochemical pathways in this group of bacteria makes them a good source for useful biological products such as organic acids and enzymes. In addition, clostridia have proved to be a good system to understand basic biological questions such as the regulation of branched metabolic pathways. A better understanding of the clostridia will expand the utilization of this group of bacteria and may also facilitate the control of the pathogenic species.

Chapter I

Illustration of the Solvent Producing Pathway and Characterization of Relevant Enzymes

The solvent-producing pathway (Figure 1.1) in the clostridia has been thoroughly studied. From acetyl-CoA, the clostridia develop branched pathways to generate different products. At the acid-producing stage, acetyl-CoA is converted into a phosphorylated form, acetyl phosphate by phosphotransacetylase and then to acetate by acetate kinase, with ATP generated. Acetyl-CoA is also used to synthesize longer carbon chain molecules through four steps catalyzed by thiolase, 3-hydroxybutyryl-CoA dehydrogenase, crotonase, and butyryl-CoA dehydrogenase, forming butyryl-CoA. Butyryl-CoA becomes butyryl phosphate and then butyrate through a pathway parallel to the formation of acetate from acetyl-CoA. At the solvent-producing stage, acetyl-CoA is reduced to acetaldehyde, then to ethanol. Similarly, butyryl-CoA is reduced to butyraldehyde, and finally butanol. Acetyl-CoA can also be utilized to produce acetone through acetoacetyl-CoA and acetoacetate. If there is a secondary alcohol dehydrogenase present, acetone can be further reduced into isopropanol. Besides the biochemical reactions, the whole pathway can also be viewed as a distribution of electron and carbon. At the acid-producing stage, H_2 gas is produced to act as an electron sink; at the solvent-producing stage, electrons are directed to form alcohol and less H_2 is produced. At the end of exponential growth stage, cells begin to sporulate. The solventogenic stage comes to an end when sporulation progresses to the point at which active metabolism ceases. These morphological and physiological switches are triggered by as yet unknown signals.

Enzymes for the acid and solvent producing pathways have been purified from different clostridial species and characterized (Table 1.1). The production of acetone is metabolically coupled to the production of butanol because of the activity of the enzyme acetoacetate: acetate/butyrate CoA-transferase. When acetoacetyl-CoA is the CoA donor and butyrate is the CoA acceptor, the products acetoacetate and butyryl-CoA are precursors for, respectively, acetone and butanol. The CoA-transferase hence functions

as both an acid-assimilating enzyme and a solvent-producing enzyme. For the formation of aldehydes and alcohols, multiple enzymes are present, displaying different coenzyme specificity (particularly, NADH or NADPH specificity), substrate specificity and pH dependence. For example, multiple forms of alcohol dehydrogenase (ADH) are present as NAD^+-dependent or $NADP^+$-dependent primary alcohol dehydrogenase, primary/secondary alcohol dehydrogenase or aldehyde/alcohol dehydrogenase (Aad/AdhE)[4]. The different forms of the enzyme not only provide the organisms with a metabolic versatility, but also provide a very good source for the comparative study of the relationship between structure and function. For example, more than 200 sequences of aldehyde dehydrogenase (ALDH) from a wide variety of organisms have been collected and aligned for this purpose[5]. Furthermore, with information of crystal structures of some of the enzymes available, e.g., monomeric iron-containing hydrogenase (CpI) from *C. pasteurianum*[6] and alcohol dehydrogenase from *C. beijerinckii*[7], the determination of the structure-function relationship becomes more feasible than ever. Comparative studies can also be used to deduce the evolution of protein molecules[7].

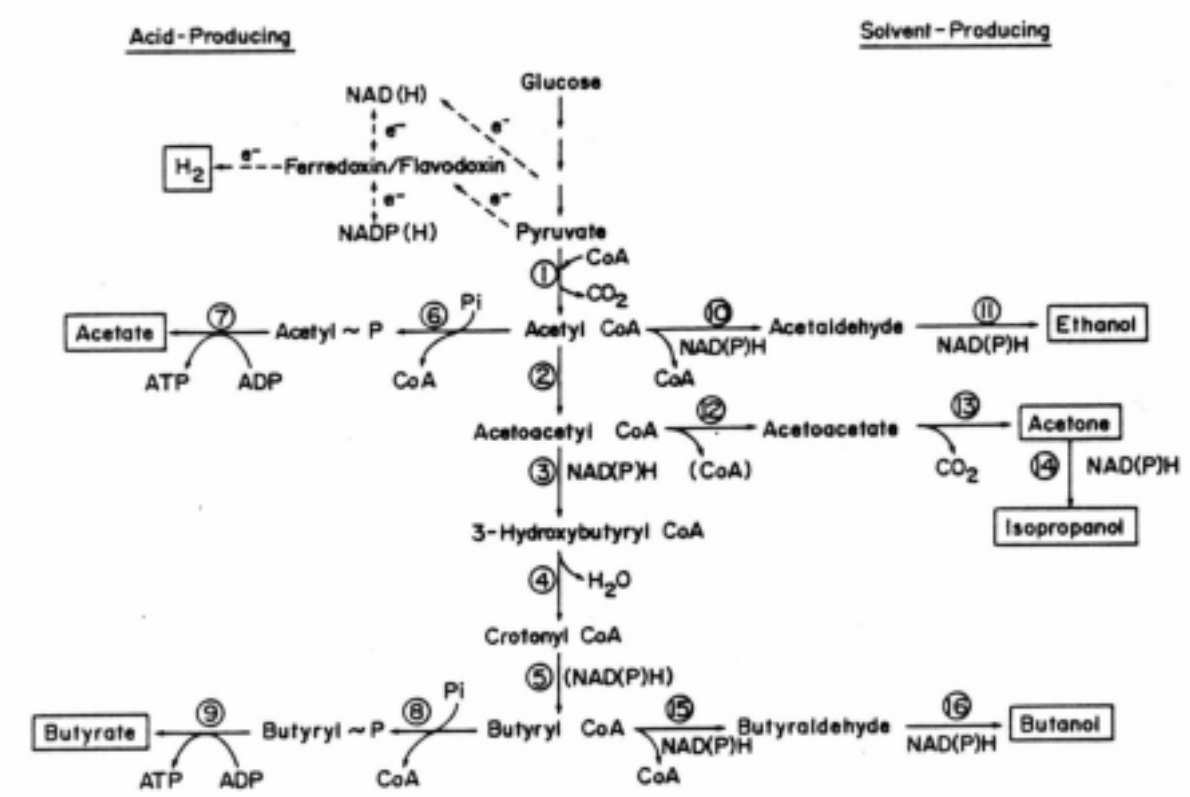


Figure 1.1: **Metabolic pathway of acid and solvent production in clostridia[8].** Enzymes catalyzing the numbered reactions are pyruvate: ferredoxin oxidoreductase (1), thiolase (2), 3-hydroxybutyryl-CoA dehydrogenase (3), crotonase (4), butyryl-CoA dehydrogenase (5), phosphotransacetylase (6), acetate kinase (7), phosphotransbutyrylase (8), butyrate kinase (9), acetoacetate: butyrate/acetate CoA-transferase (12), acetoacetate decarboxylase (13), aldehyde dehydrogenase (10, 15), and alcohol dehydrogenase (11, 14, 16).

Table 1.1: Purified enzymes or cloned genes for solvent production[4]. The name of the relevant gene is indicated in the column "Cloned".

Enzymes	Organisms	Purified	Cloned
Pyruvate:ferredoxin oxidoreductase	*C .acetobutylicum*	+ [9]	–
Thiolase	*C. acetobutylicum*	+ [10]	+ *(atoB)* [11,12]
3-Hydroxybutyryl-CoA dehydrogenase	*C. acetobutylicum* *C. beijerinckii* *Clostridium* sp. NCP 262	– +[13] –	+ *(hbd)*[14] + *(hbd)*[15] + (*hbd*) [16]
Crotonase	*C. acetobutylicum*	+[17]	+ *(crt)*[14]
Butyryl-CoA dehydrogenase	*C. acetobutylicum* *C. beijerinckii*	–	+ *(bcd)*[14] + *(bcd)*[15]
Phosphotransacetylase	*C. acetobutylicum*	Partially purified	+*(pta)*[18]
Acetate kinase	*C. acetobutylicum* *Clostridium* sp. NCP 262	+[19] +[20]	+*(ack)*[18] –
Phosphotransbutyrylase	*C. acetobutylicum* *C. beijerinckii*	+ [21] +[22]	+ *(ptb)*[23] + *(ptb)*[24]
Butyrate kinase	*C. acetobutylicum* *C. beijerinckii*	+ [25]	+ *(buk)*[23] + *(buk)*[24]
Coenzyme-A transferase	*C. acetobutylicum* *C. beijerinckii*	+[26] +	+ *(ctfA, B)*[27] +*(ctfA, B)*[15]
Acetoacetate decarboxylase	*C. acetobutylicum* *C. beijerinckii*	+ [28] +	+ *(adc)*[28, 29,30] + *(adc)*[15]
Aldehyde dehydrogenase	*C. acetobutylicum* *C. beijerinckii*	+ +[31]	– + *(ald)*[15]
Aldehyde/alcohol dehydrogenase	*C. acetobutylicum*	+[32]	+ *(aad/adhE)*[30,32]
Alcohol dehydrogenase 1, 2, 3[4]	*C. beijerinckii*	+	–
Butanol dehydrogenase I, II	*C. acetobutylicum*	+[33]	+*(bdhA* and *bdhB)*[34,35]
Primary/secondary alcohol dehydrogenase	*C. beijerinckii*	+[36]	+[37]

Chapter I

Genetics and Molecular Biology of the Solvent-Producing Clostridia

The most striking feature of the solvent-producing clostridia is the low GC content of their genome. Due to this, there is a strong bias of codon usage among these organisms: the codons dominated by A or U occur more frequently, e.g., AGA for farginine and UUA for leucine, which are rarely used in *E. coli*. Biased codon usage makes the tRNA a potential limiting factor for gene expression in a heterologous host such as *E. coli*. It has been shown experimentally that small proteins of MW 30,000 to 50,000 can be expressed in *E. coli*, while the level of expression for larger clostridial proteins in *E. coli* is low[38]. The purified RNA polymerase from *C. acetobutylicum* is analogous to α, β, and β' subunits from *E. coli* and the σ factor can cross-react with antibodies raised against the σ^{70} of *E. coli*[39]. The known promoter regions of *C. acetobutylicum* also show a high similarity to the consensus -35 and -10 element of the vegetative promoters in *E. coli* and *B. subtilis*. Other genetic elements such as phage and plasmid have also been found in the clostridia. In *C. acetobutylicum*, it was reported that CAK1 is a filamentous, single-stranded DNA phage with the size of 6.6 kb. CAK1 shows homology to the M13 phage of *E. coli*[40] and is released in the absence of cell lysis. Although plasmids conferring drug resistance have not been found in solvent-producing clostridia as in pathogenic clostridia, there is a plasmid with the size of 210 kb (later confirmed as 192 kb based on DNA sequence) in *C. acetobutylicum*[38, 41]. Some important solvent-producing genes such as *aad/adhE*, *ctfA*, *ctfB* and *adc* are located on this pSOL1 plasmid. The loss of this pSOL1 abolishes the organism's ability to produce solvents[41].

Identification of the Solvent-Producing Pathway Genes

The first cloned gene involved in the solvent and acid producing pathway was reported in 1988[42]. Now, almost all the genes have been identified (Table 1.1). Some of the genes were identified by complementation of *E. coli* mutant, such as the *ptb* and *buk*

genes in *C. acetobutylicum*[23]. Some other genes were identified by screening the genomic library using a DNA probe. The probe was usually generated by PCR with synthetic oligonucleotides designed according to either the N-terminal amino acid sequence of the purified protein, e.g., the *ctfA* and *ctfB* genes from *C. acetobutylicum*[27] or, the corresponding gene from other organisms, e.g., the *ack* gene from *C. acetobutylicum*[18]. Besides the DNA probe, antibodies were also used to screen the expression library to identify the corresponding clone, e.g., the *atoB* gene from *C. acetobutylicum*. Many of the genes have been expressed in *E. coli* either under a clostridial promoter or the vector promoter, e.g., the *ptb*, *buk*, *adc*, *ctfA* and *ctfB*, *bdh*A and *bdh*B genes from *C. acetobutylicum*[42]. However, some genes were poorly expressed in *E. coli*[38].

The solvent-producing genes are organized in the chromosome. In *C. acetobutylicum*, the *crt, bcd, etfB, etfA* (encoding homologue of the β and α subunits of the electron transfer flavoprotein, respectively) and *hbd* genes are located in a single operon called *bcs* (butyryl-CoA synthesis)[14.] The *aad/adhE, ctfA* and *ctf B* genes consist of the *sol* operon, which is located on the pSOL1 plasmid[41]. Some other genes are arranged in monocistronic operons, such as *adc, bdhA* and *bdhB*. Among them, the *adc* is located directly downstream of the *sol* operon in the opposite orientation; these *sol* locus genes encode enzymes necessary for acetone formation[30]. The *bdhA* and *bdhB* are next to each other on the chromosome[34, 35].

Genome Analysis

With the cloning of the individual genes continuing, physical maps of chromosome have been established for both *C. acetobutylicum* ATCC 824[43] and *C. beijerinckii* NCIMB 8052[44]. For *C. acetobutylicum* ATCC 824, the size of the chromosome is 3.9 Mb and the plasmid pSOL1 is 192 kb. Besides 11 *rrn* operons, the physical map contained 40 other genes[43]. For *C. beijerinckii* NCIMB 8052, the size of the genome is 6.7 Mb; 14 *rrn* operons and 40 other genes were assigned positions[44]. In

both cases, most of the solvent production related genes or operons were mapped. On pSOL1, the *aad/adhE*, *ctfA* and *ctfB* are organized into one operon, with the *adc* gene directly downstream in the other orientation. The genome of *C. acetobutylicum* has further been completely sequenced and become searchable on the NCBI microbial genomes BLAST server. A total of 4927 ORFs were identified on the chromosome based on homology to known genes[45].

Genetic Manipulations

The progress in the genetic studies of the clostridia largely depends on the availability of genetic tools. These tools are critical for future application of clostridia as well. For instance, in order to perform genetic engineering, it is necessary to be able to introduce and express endogenous or exogenous genes in these organisms. Two mechanisms have been utilized to introduce DNA into the solvent-producing clostridial cells, conjugation or transformation. For conjugation to occur, the plasmid needs to contain *tra* genes encoding proteins with transfer functions or at least a mobilizable element to get transferred by conjugative plasmids. Conjugation also needs cell-to-cell contact and its efficiency depends largely on the donor and the recipient bacterial strains used. Three broad host-range MLS (macrolide-lincosamide-streptogramin B resistance) plasmids, pAMβ1, pIP501 and pJH4, were transferred by a conjugation-like process from *Streptococcus faecalis* to *Clostridium acetobutylicum*. These plasmids were stably maintained and most of the antibiotic resistance genes were fully expressed in *C. acetobutylicum*[46]. Tn1545, a conjugative transposon originally discovered in *Streptococcus pneumoniae*, has been transferred from *Enterococcus faecalis* or *Bacillus subtilis* to *C. beijerinckii* NCIMB 8052[47]. Transfer of another conjugative transposon, Tn916, from *E. faecalis* to *C. beijerinckii* NCIMB 8052 was also achieved[47].

Meanwhile, transformation makes any bacterial strain potentially amenable to genetic alteration by exogenous DNA. Protoplast has been used to facilitate DNA uptake, followed by the regeneration of rod-shaped cells[48]. For vegetative cells, a

powerful method for transformation is electroporation. Cells are suspended with plasmid DNA and pulsed at a high voltage (5.0kV/cm). Transformants are selected after a recovery period. Transformation has been successfully performed in *C. acetobutylicum* and *C. beijerinckii* although the efficiency is not high[49,50,51,52].

The construction of cloning vectors also made remarkable progress. First, shuttle plasmids containing selection markers and origins of replication for *E. coli* and *Clostridium* were generated. Among them, pCBU2 was derived from *E. coli* and *C. butyricum*[53], and pTY10 was derived from *E. coli* and *C. acetobutylicum*[54]. Then a shuttle vector using replicons from *E. coli* and Gram-positive *B. subtilis* (pIM13) was constructed[55]. These plasmids were introduced into the organisms by transformation. However, for *C. acetobutylicum* strain ATCC 824, the efficiency of transformation was extremely low. Later, it was discovered that this was due to a strong restriction system present in *C. acetobutylicum* ATCC 824. The restriction endonuclease was named *Cac* 824I. It can recognize and cut at the sequence 5'-GCNGC -3', which occurs at a high frequency in the ColE1 plasmid. To overcome this restriction barrier, an in vivo methylation system was developed to modify the recognition sequence with a phage methyltransferase from *B. subtilis*[56]. Meanwhile, shuttle vectors derived from *B. subtilis* and *Clostridium*, e.g., pFNK family, were constructed to avoid the ColE1 origin, and hence were not affected by *Cac* 824I[57]. With pFNK, the genes *adc* and *ptb* were successfully transformed into *C. acetobutylicum* and expressed[58].

Among the techniques of genetic manipulations, mutagenesis is another important one. Different methods are available to achieve it. The most commonly used ones in the clostridia are chemical mutagenesis, transposon mutagenesis and more recently, homologous recombination[59]. Mutagenic chemicals include methyl methanesulfonate (MMS), ethyl methane-sulfonate (EMS), and N-methyl-N'-nitro-N-nitrosoguanidine (MNNG). A large variety of mutants have been isolated from *C. acetobutylicum* by using these chemicals[59]. No indigenous transposon has been found in the solvent-producing clostridia, yet several enterococcal conjugative transposons Tn916, Tn925, Tn1545 were widely used in transposon-mediated mutagenesis [38, 60, 61, 62]. Mutants deficient in both

structural genes of fermentation or regulatory genes which are involved in solventogenesis, sporulation or granulose formation were selected[60]. Homologous recombination occurs by a Campbell-like mechanism through a non-replicative plasmid. In *C. acetobutylicum*, either *pta* or *buk* gene was inactivated by homologous recombination. Inactivation of the *pta* gene decreased acetate production and inactivation of *buk* gene decreased butyrate production but increased butanol production[63]. Another gene which encodes a putative transcriptional repressor of solventogenesis, *solR* (see the section of "Regulation of Solventogenesis") was also inactivated by homologous recombination. The resulting mutant exhibited de-regulated solvent production[64]. Besides the mutagenesis, the modulation of gene expression, such as the newly developed antisense RNA strategy, has emerged as a useful tool to observe mutant phenotypes. Although no genetic change made, antisense RNA strategy may have a few advantages over the regular mutagenesis. It can be more rapidly implemented and by using inducible promoters to control expression of the antisense RNA, conditional mutant phenotypes can be generated to avoid lethal mutations[65].

Taxonomy of Solvent-Producing Clostridia

The genus *clostridium* is highly heterogeneous. Among them, a number of species are capable of producing solvents. As early as in the 1910s, the commercial application of clostridial fermentation was investigated. Investigator eagerly isolated and developed high-yield strains and many patents were issued[66]. The species names of these strains were mainly based on the fermentation substrates and products, which are as diverse as the genus itself. Many of these strains were grouped in the *C. acetobutylicum* after the 1950s. With the active research on solvent-producing clostridia persisting in the past 20 years, scientists realized that the data reported for presumably identical strains were very different, including those for the type strain of *C. acetobutylicum* preserved in different culture collections (ATCC 824, DSM 792 and NCIMB 8052)[3]. Establishing a correct taxonomic relationship among the solvent-producing clostridia was urgent. Yet, few usable phenotypic differences exist among these strains and the solvent production is

not a stable feature to function as a taxonomic criterion. The problem finally was solved with the molecular approaches of comparing 16S rRNA sequence and DNA relatedness. The determination of the sequence of 16S rRNA is very convenient, but its high conservation makes its use more suitable for the classification above the species level. Meanwhile, the DNA-DNA re-association study would constitute a superior method to distinguish strains with a 16S rRNA identity over 97%[67]. The results of the DNA re-association studies unequivocally classified 39 solvent-producing clostridial cultures into four species: within the same species, the degree of DNA similarity is over 70% while between species, the similarity does not exceed 32%[68].

Regulation of Solventogenesis

Solventogenesis may be regarded as a stress response in which a cell overcomes environmental stress due to the presence of accumulated toxic acidic products. Evidence shows factors such as pH, nutrition limitation, ATP or NAD(P)H availability and the concentration of undissociated butyrate are all involved in the induction of solventogenesis[2,69]. During this switch, it was observed that several heat shock proteins were also expressed[70]. In *C. acetobutylicum*, the induced genes or operons include *adc*, the *sol* operon, *bdhA*, *bdhB*, *bldh*, and several heat shock genes[70, 71]. Yet, the direct or the indispensable signal(s) that triggers this metabolic switch is still unknown.

Mutation studies show that some mutants deficient in either solventogenesis or sporulation are actually impaired in both capabilities, which suggests that solventogenesis and sporulation share some global regulatory feature(s)[61]. In bacilli, sporulation is well known to utilize a specific family of different σ factors and each of the σ factors recognizes a specific feature in the promoter region of certain gene(s)[72]. The homologous σ factors have been found in *Clostridium* spp. through DNA transcript mapping (σ^A, σ^H), heterologous DNA hybridization (σ^D, σ^H) and immuno characterization of purified proteins (σ^A)[73, 74]. However, in the solvent-producing clostridia, most of the solvent-producing genes identified to date have promoters similar to the σ^A-dependent

promoter[24, 45, 75, 76, 77]. Mutants deficient for only sporulation or solventogenesis are available[61]. Therefore, it is unlikely that solventogenesis and sporulation utilize the same control mechanism.

A homologue of the *spo0A* gene, which encodes a positive transcription factor for sporulation in *B. subtilis*, has been found in *Clostridium* spp. In *B. subtilis*, the Spo0A has two domains: an N-terminal domain serves as the acceptor for phosphorylation and a C-terminal domain serves as the effector with a helix-turn-helix motif. Single mutations in *spo0A* simultaneously impair the cells in all of the stationary phase phenotypes. Spo0A thus is considered a global regulator in *B. subtilis*[72]. The predicted Spo0A-like gene product in *Clostridium* conserves the effector domain of the Spo0A factor, particularly, the H-T-H DNA binding motif. The "*spo0A* box" with the sequence 5'-**TGNCGAA-3'**, recognized by Spo0A in *B. subtilis*, is also present in a number of solvent producing genes such as the *sol* cluster, the *bdhA* and *bdhB* genes in *C. acetobutylicum* and also the *ptb-buk* operon in *C. beijerinckii*. However, some other genes do not appear to be associated with the *spo0A* box, e.g., *atoB* (thiolase) gene and the *ptb-buk* operon in *C. acetobutylicum*[45]. The true function of the Spo0A homologue in *Clostridium* needs to be further elucidated.

In *C. acetobutylicum*, the most recent research results show that the *sol* locus genes are regulated by an upstream gene *solR* encoding a transcriptional repressor SolR[64]. The inferred amino acid sequence of SolR indicated that it contains a potential helix-turn-helix DNA-binding motif. Over-expression of *solR* gave a solvent-negative phenotype. Inactivation of *solR* gave a mutant with a very high level of solvent production. No correlation between sporulation and *solR* has been reported.

References

[1] Collins, M.D., Lawson, P.A., Willems, A., Cordoba, J.J., Fernandez-Garayzabal, J., Garcia, P., Cai, J., Hippe, H., and Farrow, J.A. 1994. The phylogeny of the genus *Clostridium*: proposal of five new genera and eleven new species combinations. Int. J. Syst. Bacteriol. 44, 812-826.

[2] Girbal, L. and Soucaille, P. 1998. Regulation of solvent production in *Clostridium acetobutylicum*. Trends Biotechnol. 16, 11-16.

[3] Durre, P. 1998. New insights and novel developments in clostridial acetone/butanol/isopropanol fermentation. Appl. Microbiol. Biotechnol. 69, 639-648.

[4] Chen, J.-S. 1995. Alcohol dehydrogenase: multiplicity and relatedness in the solvent-producing clostridia. FEMS Microbiol. Rev. 17, 263-273.

[5] http://www.psc.edu/biomed/pages/research/Col_HBN_ALDH.html Collaborative Research Project: Structure Function Relationships among Aldehyde Dehydrogenases.

[6] Peters, J.W., Lanzilotta, W.N., Lemon, B.J., and Seefeldt, L.C. 1998. X-ray crystal structure of the Fe-only hydrogenase (CpI) from *Clostridium pasteurianum* to 1.8 Angstrom resolution. Science 282, 1853-1858.

[7] Korkhin, Y., Kalb, A.J., Peretz, M., Bogin, O., Burstein Y., and Frolow, F. 1998. NADP-dependent bacterial alcohol dehydrogenases: crystal structure, cofactor-binding and cofactor specificity of the ADHs of *Clostridium beijerinckii* and *Thermoanaerobacter brockii*. J. Mol. Biol. 278, 967-981.

[8] Chen, J.-S. 1993. Properties of acid- and solvent-forming enzymes of clostridia, p. 51-76. *In* Woods, D.R. (ed.), The clostridia and biotechnology. Butterworth-Heinemann, Boston, MA.

[9] Meinecke, B., Bertram, J., and Gottschalk, G. 1989. Purification and characterization of the pyruvate-ferredoxin oxidoreductase from *Clostridium acetobutylicum*. Arch Microbiol. 152, 244-250.

[10] Wiesenborn, D.P., Rudolph, F.B., and Papoutsakis, E.T. 1988. Thiolase from *Clostridium acetobutylicum* ATCC 824 and its role in the synthesis of acids and solvents. Appl. Environ. Microbiol. 54, 2717-2722.

[11] Petersen, D.J., Bennett, G.N. 1991. Cloning of the *Clostridium acetobutylicum* ATCC 824 acetyl coenzyme A acetyltransferase (Thiolase; EC 2.3.1.9) gene. Appl. Environ. Microbiol. 57, 2735-2741.

[12] Stim-Herndon, K.P., Petersen, D.J., and Bennett, G.N. 1995. Characterization of an acetyl-CoA C-acetyltransferase (thiolase) gene from *Clostridium acetobutylicum* ATCC 824. Gene 154, 81-85.

[13] Colby, G.D., Chen, J.-S. 1992. Purification and properties of 3-hydroxybutyryl-coenzyme A dehydrogenase from *Clostridium beijerinckii* ("*Clostridium butylicum*") NRRL B593. Appl. Environ. Microbiol. 58, 3297-3302.

[14] Boynton, Z. L., Bennett, G.N., and Rudolph, F.B. 1996. Cloning, sequencing, and expression of clustered genes encoding β-hydroxybutyryl-coenzyme A (CoA) dehydrogenase, crotonase, and butyryl-CoA dehydrogenase from *Clostridium acetobutylicum* ATCC 824. J. Bacteriol. 178, 3015-3024.

[15]Toth, J. 1999. Unpublished results of our lab.

Toth, J. and Chen, J.-S. Cloning and sequence analysis of the *ctfA* and *ctfB* genes encoding the acetoacetate: butyrate/acetate coenzyme A-transferase from *Clostridium beijerinckii* NRRL B593. 1997. Abstract for the General Meeting of the American Society for Microbiology. H-8, 286.

Toth, J., Ismaiel, A.A., and Chen, J.-S. 1999. The *ald* gene encoding a coenzyme A-acylating aldehyde dehydrogenase distinguishes *Clostridium beijerinckii* and two other solvent-producing clostridia from *Clostridium acetobutylicum*. . Appl. Environ. Microbiol. (submitted).

Toth, J. and Chen, J.-S. Organization of the acetone-butanol production genes in *Clostridium beijerinckii* NRRL B593. 1998. Abstract for the General Meeting of the American Society for Microbiology. O-39, 399.

[16] Youngleson, J.-S., Jones, D.T., Woods, D.R. 1989. Homology between hydroxybutyryl and hydroxyacyl coenzyme A dehydrogenase enzymes from *Clostridium acetobutylicum* fermentation and vertebrate fatty acid beta-oxidation pathways. J. Bacteriol. 171, 6800-6807.

[17] Waterson, R.M., Castellino, F.J., Hass, G.M., and Hill, R.L. Purification and characterization of crotonase from *Clostridium acetobutylicum*. J. Biol. Chem. 247, 5266-5271.

[18] Boynton, Z. L., Bennett, G.N., and Rudolph, F.B. 1996. Cloning, sequencing and expression of genes encoding phosphotransacetylase and acetate kinase from *Clostridium acetobutylicum* ATCC 824. Appl. Environ. Microbiol. 62, 2758-2766.

[19] Winzer, K., Lorenz K., and Durre P. 1997. Acetate kinase from *Clostridium acetobutylicum*: a highly specific enzyme that is actively transcribed during acidogenesis and solventogenesis. Microbiology 143, 3279-3286.

[20] Diez-Gonzalez, F., Russell, J.B., Hunter, J.B. 1996. The acetate kinase of *Clostridium acetobutylicum* strain P262. Arch. Microbiol.166:418-20.

[21] Wiesenborn, D.P., Rudolph, F.B., Papoutsakis, E.T. 1989. Phosphotransbutyrylase from *Clostridium acetobutylicum* ATCC 824 and its role in acidogenesis. Appl. Environ. Microbiol. 55, 317-322.

[22] Thompson, D.K., Chen, J.-S. 1990. Purification and properties of an acetoacetyl coenzyme A-reacting phosphotransbutyrylase from *Clostridium beijerinckii* ("*Clostridium butylicum*") NRRL B593. Appl. Environ. Microbiol. 56, 607-613.

[23] Cary, J.W., Petersen, D.J., Papoutsakis, E.T., and Bennett, G.N. 1988. Cloning and expression of *Clostridium acetobutylicum* phosphotransbutyrylase and butyrate kinase genes in *Escherichia coli*. J. Bacteriol. 170, 4613-4618.

[24] Oultram, J.D., Burr, I.D., Elmore, M.J., and Minton, N.P. 1993. Cloning and sequence analysis of the genes encoding phosphotransbutyrylase and butyrate from *Clostridium acetobutylicum* NCIMB 8052. Gene 1331, 107-112.

[25] Hartmanis, M.G. 1987. Butyrate kinase from *Clostridium acetobutylicum*. J. Biol. Chem. 262, 617-621.

[26] Welch, R.W., Rudolph, F.B., and Papoutsakis, E.T. 1990. Coenzyme A transferase from *Clostridium acetobutylicum* ATCC 824 and its role in the uptake of acids. Appl. Environ. Microbiol. 55, 323-329.

[27] Cary, J.W., Petersen, D.J., Papoutsakis, E.T., and Bennett, G.N. 1990. Cloning and expression of *Clostridium acetobutylicum* ATCC 824 acetoacetyl-coenzyme A: acetate/butyrate: coenzyme A-transferase in *Escherichia coli*. Appl. Environ. Microbiol. 56, 1576-1583.

[28] Petersen, and D.J., Bennett, G.N. 1990. Purification of acetoacetate decarboxylase from *Clostridium acetobutylicum* ATCC 824 and cloning of the acetoacetate decarboxylase gene in *Escherichia coli*. Appl. Environ. Microbiol. 56, 3491-3498.

[29] Gerischer, U. and Durre, P. 1990. Cloning, sequencing, and molecular analysis of the acetoacetate decarboxylase gene region from *Clostridium acetobutylicum*. J. Bacteriol. 172, 6907-6918.

[30] Fischer, R.J., Helms, J., and Durre, P. 1993. Cloning, sequencing and molecular analysis of the *sol* operon of *Clostridium acetobutylicum*, a chromosomal locus involved in solventogenesis. J. Bacteriol. 175, 6959-6969.

[31] Yan, R.T., Chen, J.S. 1990. Coenzyme A-acylating aldehyde dehydrogenase from *Clostridium beijerinckii* NRRL B592. Appl. Environ. Microbiol. 56, 2591-2599.

[32] Nair, R.V., Bennett, G.N., and Papoutsakis, E.T. 1994. Molecular characterization of an aldehyde/alcohol dehydrogenase gene from *Clostridium acetobutylicum* ATCC 824. J. Bacteriol. 176, 871-885.

[33] Welch, R.W., Rudolph, F.B. and Papoutsakis, E.T. 1989. Purification and characterization of the NADH-dependent butanol dehydrogenase from *Clostridium acetobutylicum* (ATCC 824). Arch. Biochem. Biophys. 273, 309-318.

[34] Petersen, D.J., Welch, R.W., Rudolph, F.B., and Bennett, G.N. 1991. Molecular cloning of an alcohol (butanol) dehydrogenase gene cluster from *Clostridium acetobutylicum* ATCC 824. J. Bacteriol. 173, 1831-1834.

[35] Walter, K.A., Bennett, G.N., and Papoutsakis, E.T. 1992 Molecular Characterization of two *Clostridium acetobutylicum* ATCC 824 Butanol Dehydrogenase Isozyme genes. J. Bacteriol. 174, 7149-7158.

[36] Ismaiel, A.A. and Chen, J.-S. 1993. Purification and characterization of a primary-secondary alcohol dehydrogenase from two strains of *Clostridium beijerinckii*. J. Bacteriol. 175, 5097-5105.

[37] Peretz, M., Bogin, O., Tel-Or, S., Cohen, A., Li, G., Chen, J.-S., and Burstein, Y. 1997. Molecular cloning, nucleotide sequencing, and expression of genes encoding alcohol dehydrogenases from the thermophile *Thermoanaerobacter brockii* and the mesophile *Clostridium beijerinckii*. Anaerobe 3, 259-270.

[38] Young, M. and Cole, S.T. 1993. *Clostridium*. In: *Bacillus subtilis* and other Gram-positive bacteria. Biochemistry, Physiology, and Molecular Genetics (Sonenshein, A.L., Hoch, J.A. and Losick, R., Eds.), pp35-42, American Society for Microbiology, Washington, D.C.

[39] Pich, A. and Bahl, H. 1991. Purification and characterization of the DNA-dependent RNA polymerase from *Clostridium acetobutylicum*. J. Bacteriol. 173, 2120-2124.

[40] Kim, A.Y. and Blaschek, H.P. 1991. Isolation and characterization of a filamentous virus like particle from *Clostridium acetobutylicum* NCIB 6444. J. Bacteriol. 173, 530-535.

[41] Cornillot, E., Nair, R.V., Papoutsakis, E.T., and Soucaille, P. 1997. The genes for butanol and acetone formation on *Clostridium acetobutylicum* ATCC 824 reside on a large plasmid whose loss leads to degeneration of the strain. J. Bacteriol. 179, 5442-5447.

[42] Papoutsakis, E.T. and Bennett, G.N. 1993. Cloning, structure and expression of acid and solvent pathway genes of *Clostridium acetobutylicum*, p157-195. *In* Woods, D.R. (ed.), The clostridia and biotechnology. Butterworth-Heinemann, Boston, MA.

[43] Cornillot, E., Croux, C., and Soucaille, P. 1997. Physical and genetic map of the *Clostridium acetobutylicum* ATCC 824 Chromosome. J. Bacteriol. 179, 7426-7434.

[44] Wilkinson, S.R., Young, D.I., Morris, J.G., and Young, M. 1995. Molecular genetics and the initiation of solventogenesis in *Clostridium beijerinckii* (formerly *Clostridium acetobutylicum*) NCIMB 8052. FEMS Microbiol. Rev. 17, 275-285.

[45] http://www.genomecorp.com/genesequences/clostridium/clospage.html

[46] Reysset, G. and Sebald, M. 1985. Conjugal transfer of plasmid-mediated antibiotic resistance from streptococci to *Clostridium acetobutylicum***.** Ann. Inst. Pasteur Microbiol.136B(3), 275-282.

[47]Woolley, R.C., Pennock, A., Ashton, R.J., Davies, A., and Young, M. 1989. Transfer of Tn1545 and Tn916 to *Clostridium acetobutylicum*. Plasmid 22, 169-174.

[48]Reysset, G. and Sebald, M. Transformation /Electrotransformation of Clostridia, p151-155. *In* Woods, D.R. (ed.), The clostridia and biotechnology. Butterworth-Heinemann, Boston, MA.

[49] Nakotte, S., Schaffer, S., Bohringer, M., and Durre, P. 1998. Electroporation of, plasmid isolation from and plasmid conservation in *Clostridium acetobutylicum* DSM 792. Appl. Microbiol. Biotechnol. 50, 564-567.

[50] Oultram, J.D., Loughlin, M., Swinfield, T.-J., Brehm, J.K., Thompson, D.E., and Minton, N.P. 1988. Introduction of plasmids into whole cells of *Clostridium acetobutylicum* by electroporation. FEMS Microbiol. Lett. 56, 83-88.

[51] Lee, S.Y., Mermelstein, L.D., Bennett, G.N., and Papoutsakis, E.T. Vector construction, transformation, and gene amplification in *Clostridium acetobutylicum* ATCC 824. Ann. N Y Acad. Sci. 665, 39-51.

[52] Li, G. 1998. Development of a reporter system for the study of gene expression to solvent production in *Clostridium beijerinckii* NRRL B592 and *Clostridium acetobutylicum* ATCC 824. Ph.D. thesis. Virginia Polytechnic Institute and State University, Blacksburg, VA, USA.

[53] Minton, N.P., Brehm, J.K., Swinfield, T.-J., Whelan, S.M., Mauchline, M.L., Bodsworth, N., and Oultram, J.D. 1993. Clostridial cloning vectors. *In*: Woods, D.R.(ed), The clostridia and biotechnology. Butterworth-Heinemann, Stoneham. pp 119-150.

[54] Yoshino, S., Yoshino, T., Hara, S., Ogata, S., and Hayashida, S. 1990. Construction of shuttle vector plasmid between *Clostridium acetobutylicum* and *Escherichia coli*. Agric. Biol. Chem. 54, 437-441.

[55] Azeddoug, H., Hubert, J., and Reysset, G. 1992. Stable inheritance of shuttle vectors based on plasmid pIM13 in a mutant strain of *Clostridium acetobutylicum.* J. Gen. Microbiol. 138, 1371-1378.

[56] Mermelstein, L.D.,Papoutsakis, E.T. 1993. In vivo methylation in *Escherichia coli* by the *Bacillus subtilis* phage F3Ti methyltransferase to protect plasmids from restriction upon transformation of *Clostridium acetobutylicum* ATCC 824. Appl. Environ. Microbiol. 59, 1077-1081.

[57] Truffaut, N., Hubert, J. and Reysset, G. 1989. Construction of shuttle vectors useful for transforming *Clostridium acetobutylicum.* FEMS Microbiol. Lett. 58, 15-20.

[58] Mermelstein, L.D., Welker, N.E., Bennett, G.N., and Papoutsakis, E.T. 1992. Expression of cloned homologous fermentative genes in *Clostridium acetobutylicum* ATCC 824. Biotechnology 10, 190-195.

[59] Jones, D.T. 1993. Mutagenesis and its application in Biotechnology, p 77-97. *In* Woods, D.R. (ed.), The clostridia and biotechnology. Butterworth-Heinemann, Boston, MA.

[60] Durre, P. 1993. Transposons in Clostridia, p227-243. *In* Woods, D.R. (ed.), The clostridia and biotechnology. Butterworth-Heinemann, Boston, MA.

[61] Mattsson, D.M. and Rogers, P. 1994. Analysis of Tn916-induced mutants of *Clostridium acetobutylicum* altered in solventogenesis and sporulation. J. Ind. Microbiol. 13, 258-268.

[62] Woolley, R.C., Pennock, A, Ashton, R.J., Davies, A., and Young, M. 1989. Transfer of Tn1545 and Tn916 to *Clostridium acetobutylicum*. Plasmid 22,169-174.

[63] Green, E.M., Boynton, Z.L., Harris, L.M., Rudolph, F.B., Papoutsakis, E.T., and Bennett, G.N. 1996. Genetic manipulation of acid formation pathways by gene inactivation in *Clostridium acetobutylicum* ATCC 824. Microbiology 142, 2079-2086.

[64] Nair, R.V., Green, E.M., Watson, D.E., Bennett, G.N., and Papoutsakis, E.T. 1999. Regulation of the *sol* locus genes for butanol and acetone formation in *Clostridium acetobutylicum* ATCC 824 by a putative transcriptional repressor. J. Bacteriol. 181, 319-330.

[65] Desal, R.P. and Papoutsakis, E.T. 1999. Antisense RNA strategies for metabolic engineering of *Clostridium acetobutylicum*. Appl. Environ. Microbiol. 65, 936-945.

[66] Jones, D.T., Keis, S. 1995. Origins and relations of industrial solvent-producing clostridial strains. FEMS Microbiol. Rev. 17, 223-232.

[67] Stackebrandt, E. and Goebel, B.M. 1994. Taxonomic Note: A place for DNA-DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology. Int. J. Syst. Bacteriol. 44, 846-849.

[68] Johnson, J.L and Chen, J.-S. 1995. Taxonomic relationships among strains of *Clostridium acetobutylicum* and other phenotypically similar organism. FEMS Microbiol. Rev. 17, 233-240.

[69] Girbal, L., Croux, C., Vasconcelos, I., and Soucaille, P. 1995. Regulation of metabolic shifts in *Clostridium acetobutylicum* ATCC 824. FEMS Microbiol. Rev. 17, 287-297.

[70] Bahl, H., Muller, H., Behrens, S., Joseph, H., and Narberhaus, F. 1995. Expression of heat shock genes in *Clostridium acetobutylicum*. FEMS Microbiol. Rev. 17, 341-348.

[71] Sauer, U. and Durre, P. 1995. Differential induction of genes related to solvent formation during the shift from acidogenesis to solventogenesis in continuous culture of *Clostridium acetobutylicum*. FEMS Microbiol. Lett. 125, 115-120.

[72] Haldenwang, W.G. 1995. The sigma factors of *Bacillus subtilis*. Microbiol. Rev. 59 (1), 1-30.

[73] Sauer, U., Treuner, A., Buchholz, M., Santangelo, J.D., and Durre, P. 1994. Sporulation and primary sigma factor homologous genes in *Clostridium acetobutylicum*. J. Bacteriol. 176, 6572-6582.

[74] Sauer, U. and Durre, P. 1995. Sigma factors and sporulation genes in *Clostridium*. FEMS Microbiol. Rev. 17, 331-340.

[75] Gerischer, U. and Durre, P. 1992. mRNA analysis of the *adc* gene regulation of *Clostridium acetobutylicum* during the shift to solventogenesis. J. Bacteriol. 174, 426-433.

[76] Woods, D.R., Reid, S.J. 1995. Regulation of nitrogen metabolism, starch utilization and the β-*hbd-adh*1 gene cluster in *Clostridium acetobutylicum*. FEMS Microbiol. Rev. 17, 299-306.

[77] Walter, K.A., Bennett, G.N. 1993. Sequence and arrangement of two genes of the butyrate-synthesis pathway of *Clostridium acetobutylicum* ATCC 824. Gene 134, 107-111.

Chapter II
The cloning and sequence analysis of the *stc* gene from *C. beijerinckii* NRRL B593

Introduction

Among the solvent-forming enzymes in solvent producing clostridia, alcohol dehydrogenase (ADH) plays a key role in the production of butanol, ethanol and isopropanol. ADH catalyzes the reversible reduction of an aldehyde or a ketone to the corresponding primary or secondary alcohol:

$$\text{Aldehyde or ketone} + \text{NAD(P)H} + H^+ \longleftrightarrow \text{alcohol} + \text{NAD(P)}^+ \quad (1)$$

Consistent with the diverse properties of the species within the genus *Clostridium*, the multiple ADHs of clostridia differ in molecular size, cofactor requirement and enzymatic properties. In the past decade, with more and more ADHs purified and the taxonomic relationship of the solvent producing clostridia clarified, a better understanding of the structural and functional relationship among the different ADHs is emerging.

Among the solvent producing clostridia, three distinct groups of ADHs exist: (1) the primary ADHs, (2) the aldehyde/alcohol dehydrogenase that has primary ADH activity, and (3) the primary/secondary ADH that is able to catalyze the reactions to form both primary and secondary alcohols. The primary ADHs include BDHI and BDHII from *C. acetobutylicum* ATCC 824, ADH-1, ADH-2 and ADH-3 from *C. beijerinckii* NRRL B592. The aldehyde/alcohol dehydrogenase refers to Aad/Adh-E from *C.*

acetobutylicum ATCC 824 or DSM 792[1]. The primary/secondary ADH enables the organism, e.g., *C. beijerinckii* NRRL B593, to produce both butanol and isopropanol - the latter fermentation makes this organism distinct from other solvent producing clostridia such as the extensively studied *C. acetobutylicum* ATCC824 and DSM792 or *C. beijerinckii* NCIMB 8052. This primary/secondary ADH has been characterized[2] and its gene (*adh*) cloned from *C. beijerinckii* NRRL B593[3]. Similar to other solvent-producing genes, the *adh* gene is also induced/derepressed during the onset of solventogenesis[4]. However, the signal which triggers the expression of the *adh* gene (and the onset of solventogenesis) and how it does so are unknown. During the cloning and sequence analysis of the *adh* gene from *C. beijerinckii* NRRL B593, the partial sequence of a putative regulatory gene, which encodes an amino acid sequence related to the *ntr*C gene product, was identified upstream from the *adh* gene (Figure 2.1). This putative gene was named "*stc*" (GenBank accession number: M84723)[5]. The cloned region was later expanded in an unsuccessful attempt to clone the entire *stc* gene[4]. It is important for us to know if this putative gene is involved in the signal transduction utilizing a mechanism similar to the two component system represented by NtrB-NtrC. For my research, I completed the cloning of the remaining part of the *stc* gene and concluded an analysis of the deduced amino acid sequence and a search of other solvent-producing clostridia for the presence of the *stc* gene.

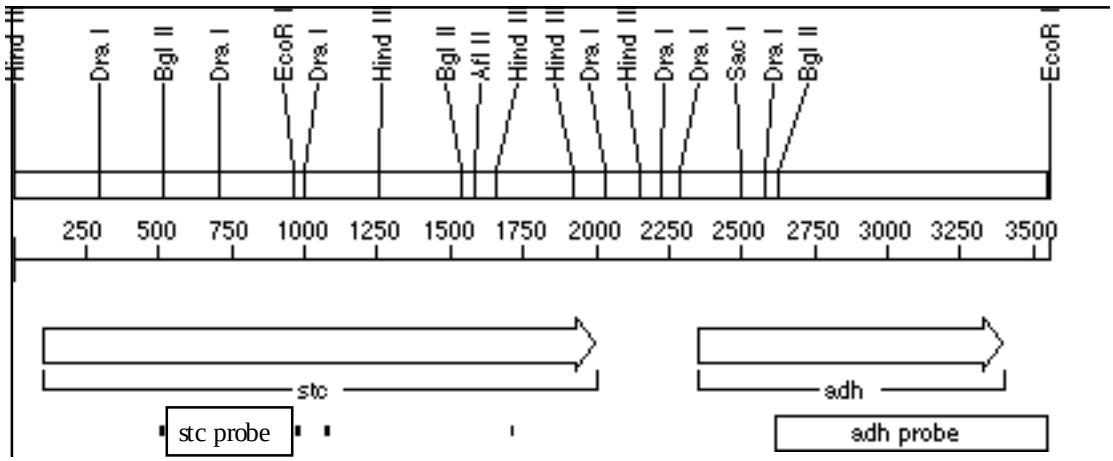


Figure 2.1: Restriction map for the *stc* and *adh* genes. Also indicated are the names and the positions of the probes for *stc* and *adh* in this study.

Chapter II

Materials and Methods

Bacterial strains and plasmids

C. beijerinckii NRRL B593 and *E. coli* DH5α were from lab stocks. Plasmid Litmus 29 was from New England BioLabs.

Bacterial growth conditions

E. coli was routinely grown in the Luria-Bertani medium aerobically at 37°C. Media were supplemented with ampicillin (100μg/ml) as needed.

The clostridial cultures were grown anaerobically under an atmosphere of 10% CO_2, 10% H_2 and 80% N_2. One ml clostridial spores were heat-shocked in the boiling water for 1.5 min and then inoculated into 45 ml CSM (Clostridial Soluble Medium[6]). After about 18 - 24 hours at 32ºC, this culture was transferred to 450 ml TYG-60 medium (60g/L glucose)[7] and incubated at 32ºC. For DNA isolation, the culture was harvested when the $O.D._{600}$ reached 2 - 3 (Beckman DU7400).

DNA isolation and manipulation

Rapid, small-scale isolation of plasmid DNA was performed by using a miniprep kit (Qiagen) according to the manufacturer's instruction. Large-scale isolation of plasmid DNA was performed with a midiprep kit (Qiagen). Genomic DNA was isolated according to a modified Marmur procedure from Johnson, J.L.[8]

Endonuclease (New England BioLabs) digestion of genomic DNA was performed, and the DNA fragments were separated on a SeaKem LE agarose (FMC BioProducts) gel (0.8%). The fragments of interest were extracted from the gel with a gel extraction kit (Qiagen). The resulting DNA fragments were ligated into Litmus29 (ligase from GIBCO BRL).

Probe generation through PCR

PCR was performed according to the procedure recommended by the *Taq* polymerase supplier (Promega Life Science) in a Genius thermal cycler (Techne).

With genomic DNA from *C. beijerinckii* NRRL B593 as the template, one region was amplified as shown in Figure 2.1. Synthetic oligonucleotides (GIBCO BRL), based on known DNA sequence in *C. beijerinckii* NRRL B593, were designed as primers for PCR. These primers are:

*stc*SB1F: 5' - TTAGATCT-GGGTTTGTTCCTCGTCGGTTGT - 3'

*stc*SB485R: 5' - TATCCCTACCATACCAGTAATATTTGAA -3'.

The PCR fragment was purified with a PCR purification kit (Qiagen) and used as the "*stc*" probe in the screening of genomic DNA for related sequences.

The probe for the *adh* gene was a subcloned fragment from the "*adh*" clone in our lab[4]. It covered the region indicated in Figure 2.1.

Southern Analysis

Probes were labeled using the ECL direct nucleic acid labeling and detection system (Amersham/Pharmacia Biotech). DNA samples were digested to completion with the proper restriction enzyme(s) and then separated by gel electrophoresis. The DNA fragments were then transferred to positively charged Nylon membrane (Boehringer Mannheim). The membrane used in Figure 2.6 was kindly provided by Julianna Toth. Prehybridization, hybridization, washing, and detection were performed at routine stringency according to the manufacturer's instructions (Amersham/Pharmacia Biotech).

SSP-PCR (Single-Specific-Primer PCR[9])

The previously sequenced region of the *stc* gene suggests that a *Hin*dIII fragment may contain the N-terminal region of this gene. The genomic DNA from *C. beijerinckii* NRRL B593 was digested with *Hin*dIII and ligated to the *Hin*dIII-digested cloning vector Litmus 29. The ligation mixture was used as the template for PCR with a generic primer based on the vector sequence and the *stc*-specific primer (*stc*SB485R). The PCR product, which was about 1.1 kb in length, was digested with *Eco*RI and *Hin*dIII and then ligated

to the *Eco*RI- and *Hin*dIII-digested Litmus 29 (Figure 2.2). This *Hin*dIII-*Eco*RI fragment contained the N-terminal region of the *stc* gene and completed the coding region of the *stc* gene.

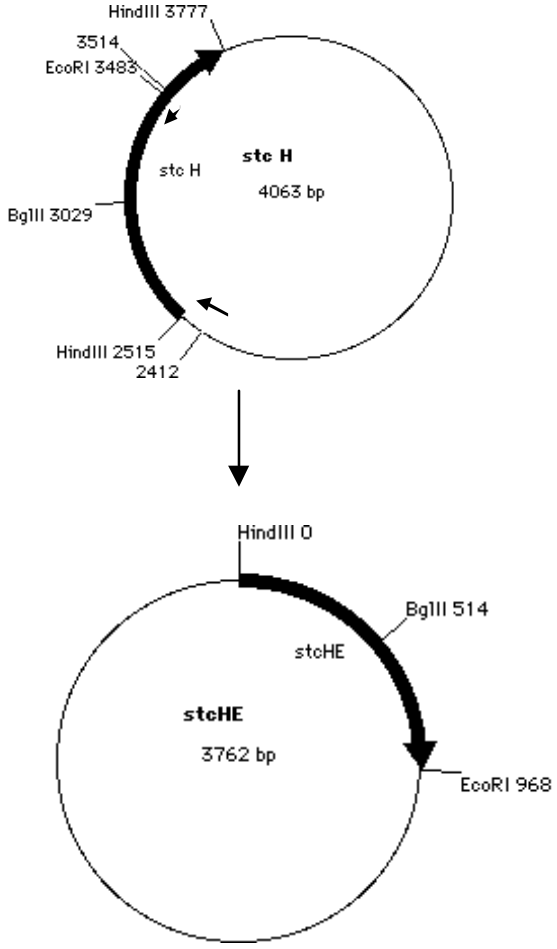


Figure 2.2: SSP-PCR. *stc* H refers to the desired ligation result which was present in the ligation mixture. *stc* HE refers to the cloned construct which contains the *Hin*dIII-*Eco*RI fragment of the *stc* gene. Small arrows in the figure of *stc* H refer to the positions of the primers used. The cloning vector used here was Litmus 29 from New England BioLab.

Cell Transformation of *E. coli*

Transformation of *E. coli* DH5α with plasmids was performed using competent cells prepared as described in **Operating Instructions and Applications Guide (Sep, 1992)** for the Bio-Rad Gene Pulser (Bio-Rad). The transformants were selected on the LB plates supplemented with ampicillin (100μg/ml), 4μM IPTG and 40μg/ml X-gal for the white-and-blue selection. The positive colonies were confirmed by Southern analysis after digestion of the plasmids with appropriate restriction enzymes.

Total RNA isolation

C. beijerinckii NRRL B593 was grown and sampled at different time points. Total RNA isolation was performed using a RNeasy Mini Kit supplemented with RNase-free DNase (Qiagen).

RT-PCR

RT-PCR with the isolated total RNA as the template was performed using the SuperScriptII RT system (GIBCO BRL). The primer for the reverse transcription to make the first strand was *stc* SB485R (about 0.9 kb downstream of the start codon of the *stc* gene). The primer pair for the following PCR was *stc*SB485R and *stc*H91 (directly upstream of the start codon):

5'-AAGTCTAGAGGGGAATATTAAATGCTCA-3' (start codon underlined) or *stc*H303 (0.2 kb downstream of the start codon):

5'-GATCTAGAAGGTTTTTCATTATATTTTACG-3'.

The two pairs of primers were expected to give products, respectively, of 0.9 and 0.7 kb, if the appropriate mRNA template was present.

DNA sequencing

Insert DNA was sequenced on either an Pharmacia ALFexpress II automated DNA sequencer of the DNA Sequencing Facility at the Center for Molecular Medicine

and Infectious Diseases, Virginia-Maryland Regional College of Veterinary Medicine or an ABI Prism 377 DNA sequencer at Davis Sequencing of Davis, California. Primers were generated based on the known sequences and synthesized by GIBCO BRL. The clone was sequenced for both strands and the ambiguous bases were clarified by repeated or overlapped sequencing.

Sequence Analysis

Sequence information was managed with the LaserGene application of DNASTAR. The homology search was carried out with BLAST 2.0 at NCBI web site[10]. The multiple sequence alignments and the secondary structure predictions were accomplished with, respectively, the protein secondary-structure prediction tool and the protein multiple sequence alignment tool CLUSTALW at the Biology WorkBench web site[11].

Results

Sequence analysis

The nucleotide sequence of the *Hin*dIII-*Eco*RI fragment (*stc*HE in Figure 2.2), which was cloned in plasmid p*stc*HE, contained an open-reading frame and it included the previously known amino acid sequence between the *Bgl*II and *Eco*RI sites[4]. A putative ribosome-binding site is (-GGGA-) present between –11 and –8 bp preceding the proposed start codon (Figure 2.3). The deduced amino acid sequence from nucleotides 103 and 517 thus represented the N-terminal region of the Stc protein. The sequence of the *stc* gene was hence complete.

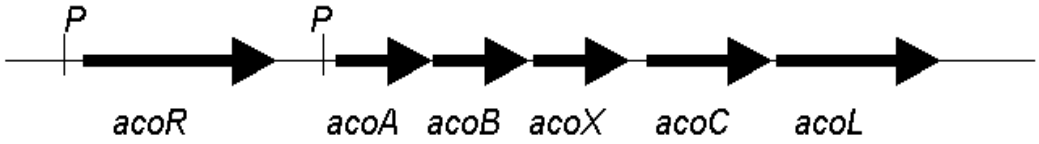
The complete nucleotide sequence (submitted to the GenBank with the accession number AF 157307) and the deduced amino acid sequence for Stc are shown in Figure 2.3, with the amino acid symbol under the second position of the codon. The complete *stc* gene (G+C mol%: 29.7) encodes a polypeptide of 632 amino acids. The calculated molecular weight (MW) is about 71,537. The calculated pI is 8.232. Table 2.1 indicates the codon usage of the *stc* gene and the amino acid composition of Stc.

Table 2.1: Codon usage of the *stc* gene and the amino acid composition of the Stc

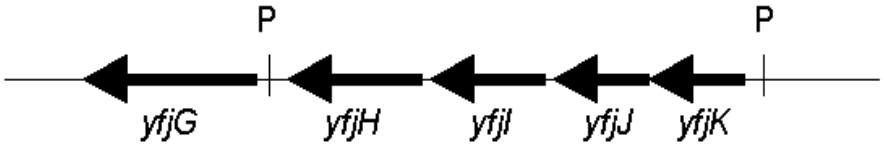
gca	Ala(A)	9	# cag	Gln(Q)	3	# uug	Leu(L)	7	# uaa	Ter(.)	0
gcc	Ala(A)	4	# ---	Gln(Q)	19	# ---	Leu(L)	50	# uag	Ter(.)	0
gcg	Ala(A)	2	# gaa	Glu(E)	38	# aaa	Lys(K)	39	# uga	Ter(.)	1
gcu	Ala(A)	10	# gag	Glu(E)	9	# aag	Lys(K)	16	# ---	Ter(.)	1
---	Ala(A)	25	# ---	Glu(E)	47	# ---	Lys(K)	55	# aca	Thr(T)	9
aga	Arg(R)	24	# gga	Gly(G)	25	# aug	Met(M)	13	# acc	Thr(T)	0
agg	Arg(R)	3	# ggc	Gly(G)	5	# ---	Met(M)	13	# acg	Thr(T)	3
cga	Arg(R)	0	# ggg	Gly(G)	8	# uuc	Phe(F)	0	# acu	Thr(T)	12
cgc	Arg(R)	1	# ggu	Gly(G)	8	# uuu	Phe(F)	21	# ---	Thr(T)	24
cgg	Arg(R)	0	# ---	Gly(G)	46	# ---	Phe(F)	21	# ugg	Trp(W)	3
cgu	Arg(R)	2	# cac	His(H)	0	# cca	Pro(P)	8	# ---	Trp(W)	3
---	Arg(R)	30	# cau	His(H)	10	# ccc	Pro(P)	2	# uac	Tyr(Y)	3
aac	Asn(N)	4	# ---	His(H)	10	# ccg	Pro(P)	1	# uau	Tyr(Y)	20
aau	Asn(N)	50	# aua	Ile(I)	39	# ccu	Pro(P)	3	# ---	Tyr(Y)	23
---	Asn(N)	54	# auc	Ile(I)	2	# ---	Pro(P)	14	# gua	Val(V)	14
gac	Asp(D)	2	# auu	Ile(I)	37	# agc	Ser(S)	5	# guc	Val(V)	2
gau	Asp(D)	29	# ---	Ile(I)	78	# agu	Ser(S)	14	# gug	Val(V)	4
---	Asp(D)	31	# cua	Leu(L)	9	# uca	Ser(S)	11	# guu	Val(V)	14
ugc	Cys(C)	5	# cuc	Leu(L)	1	# ucc	Ser(S)	2	# ---	Val(V)	34
ugu	Cys(C)	10	# cug	Leu(L)	1	# ucg	Ser(S)	2	#		
---	Cys(C)	15	# cuu	Leu(L)	13	# ucu	Ser(S)	6	#		
caa	Gln(Q)	16	# uua	Leu(L)	19	# ---	Ser(S)	40	# **TOTAL**		**632**

A BLASTP search using the complete amino acid sequence of the predicted *stc* gene product reveals that the Stc polypeptide has a sequence similarity to:

- The AcoR protein of *Clostridium magnum* (I40789) --- a positive regulatory protein for the *acoABCL* operon[12]:



- The YfjG protein of *Bacillus subtilis* (CAB12639) --- a positive transcriptional regulator for the *acoABCL* operon (map position 73° - 76°)[13]:



- A putative σ54-dependent transcriptional regulator (P54529) in the intergenic region (map position 215° - 216°) of *mmgE (*involved in *mmg* operon*)-bfmBAA* (encoding branched-chain α-ketoacid dehydrogenase E1-α) in *B. subtilis*[14].

- A putative transcriptional regulator (AAC07620) (NtrC family) in *Aquifex aeolicus.*

Further alignment of these five amino acid sequences (Table 2.2 and Figure 2.4) suggested that the Stc polypeptide may be divided into three domains: the N-terminal domain (N) of about 330 amino acids, the central domain (M) of about 240 amino acids, and the C-terminal domain (C) of about 60 amino acids. The M-domain of Stc corresponds to the highly conserved domain found in all σ-54 dependent transcriptional activators represented by NtrC. The N-domain of Stc was related only to the AcoR-family of transcriptional activators. The C-domain of Stc resembled a DNA-binding domain with a helix-turn-helix pattern (GNISKACRILGINRSTLYIKIK).

Table 2.2: Relatedness between the *C. beijerinckii* Stc sequence and members of the NtrC family

Organism and protein	Conserved region		%Identity based on the full length (632 a.a.) of Stc	Accession No.
	Range	%Identity		
Clostridium magnum AcoR	*Cb* 11-629 *Cm* 28-643	38% (243/630)	38.4%	I40789
Bacillus subtilis YfjG (AcoR)	*Cb* 11-629 *Bs* 29-601	33% (207/625)	32.8%	CAB12639
B. subtilis putative σ54-dependent regulatory protein	*Cb* 179-629 *Bs* 216-686	37% (176/473)	27.8%	P54529
Aquifex aeolicus putative regulatory protein	*Cb* 271-632 *Cb* 321-629	44% (164/369)	25.9%	AAC07620
Escherichia coli FhlA	*Cb* 321-629 *Ec* 378-686	44% (143/313)	22.6%	A36705
Klebsiella pneumoniae NifA	*Cb* 319-550 *Kp* 207-436	52% (122/232)	19.3%	CAA31682
E. coli NtrC	*Cb* 139-626 *Ec* 17-465	28% (159/565)	25.2%	P06713

A BLASTN search of the newly finished genome sequence of *C. acetobutylicum* ATCC 824 did not reveal any DNA sequence related to the *stc* gene. Also, the search did not detect any *C. acetobutylicum* sequence related to the *aco*ABCL region from *C. magnum* or *B. subtilis*. However, when the amino acid sequence of Stc or AcoR was used to do a TBLASTN search of the *C. acetobutylicum* chromosome, some common coding regions were identified. Table 2.3 compiled the statistics of the TBLASTN search. These suggest the potential presence of Stc or AcoR-like proteins in *C. acetobutylicum.*

Table 2.3: The statistics of the TBLASTN search of *C. acetobutylicum* ATCC 824 chromosome with amino acid sequence of Stc from *C. beijerinckii* NRRL B593 or AcoR from *C. magnum*. "Identity" refers to the percentage of the same amino acids. "Positive" refers to the percentage of the same and the similar amino acids. E value refers to the number of times one expects to see such a match (or better one) merely by chance in a database.

Queries/Parameters		Hit 1	Hit 2	Hit3
Stc	Aligned region of Stc	203-632	179-623	305-550
	Identities	41%	40%	28%
	Positives	61%	59%	50%
	E value	1e-79	2e-77	5e-19
	Location	2900558-2901886	189585-188224	105782-105045
AcoR	Aligned region of AcoR	190-637	330-643	258-582
	Identities	38%	47%	27%
	Positives	57%	63%	45%
	E value	8e-79	7e-72	1e-19
	Location	189579-188224	2900942-2901877	105950-104952

RT-PCR

The result of the RT-PCR is shown in Figure 2.5a. The amplified product had a size of 0.7 kb, when primers *stc*H313 and *stc*SB485R were used. The result indicated that the *stc* gene was transcribed, at least during the early stage of growth.

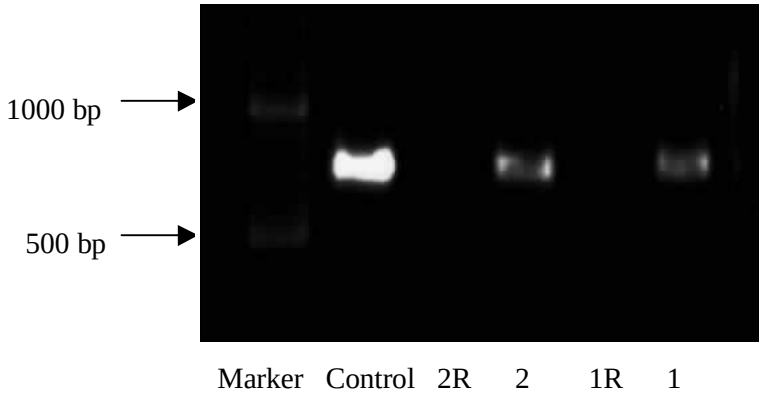


Figure 2.5a: RT-PCR for the *stc* mRNA from *C. beijerinckii* NRRL B593. Two different samples of total RNA were used as the template for RT-PCR. Sample 1 and 2 were collected, respectively, 2.5h and 4.5h after inoculation (Figure 2.5b). For each template, RNase A was (lane 1R or 2R) or was not (lane 1 or 2) added before the reverse transcription. The 100-bp DNA ladder in lane "M" was from GIBCO BRL. In the control lane, the same primers (*stc*H303 and *stc*SB485R) were used to amplify the genomic DNA from *C. beijerinckii* NRRL B593.

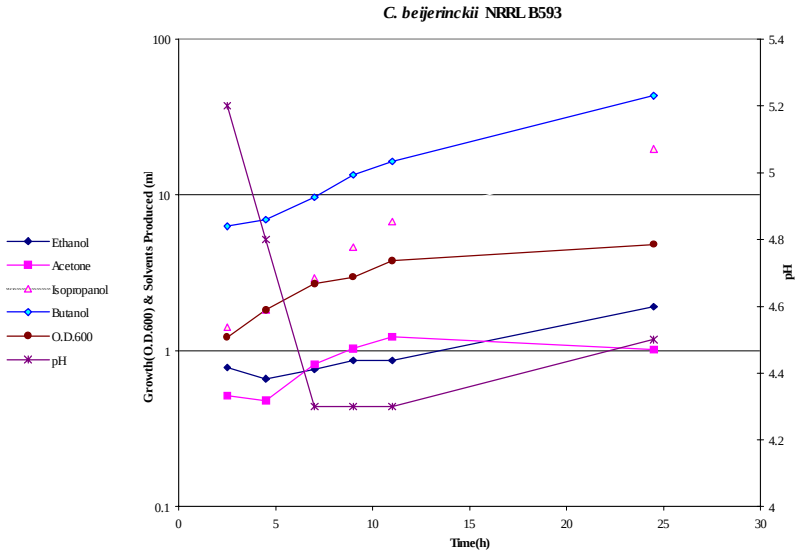


Figure 2.5b: The growth curve and solvent production of *C. beijerinckii* NRRL B593 in a TYG-60 culture. Sample 1 and 2 in Figure 2.5a were collected at the time point 2.5h and 4.5h after inoculation, respectively.

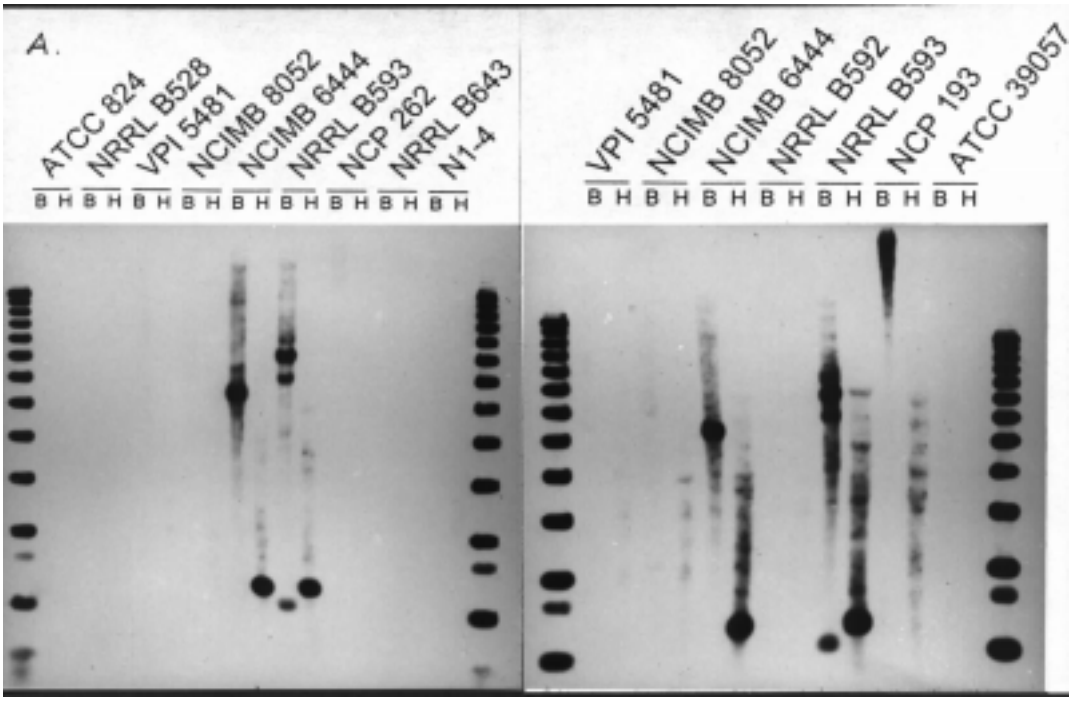
Southern Analysis

A probe for the *stc* gene (Figure 2.6 A) and a probe for the *adh* gene (Figure 2.6B) were used to screen 12 other strains of solvent-producing clostridia. The results were compiled in Table 2.4 to indicate that the *stc*-like and the *adh*-like sequences, if present, co-existed in these solvent-producing clostridia.

Table 2.4: The results of Southern analysis for 13 solvent-producing clostridial strains with the "*adh*" and "*stc*" probes.

Strains	Capable of producing secondary alcohol?	Hybridized to the *adh* probe?	Hybridized to the *stc* probe?
C. beijerinckii NRRL B593	Y	Y	Y
C. acetobutylicum ATCC 824	N	N	N
C. acetobutylicum NRRL B528	N	N	N
C. beijerinckii ATCC 39057	N	N	N
C. beijerinckii NRRL 6444	Y	Y	Y
C. beijerinckii NCIMB 8052	N	N	N
C. beijerinckii NCP 193	?	Y	Y
C. beijerinckii NRRL B592	N	N	N
C. beijerinckii VPI 5481	N	N	N
Clostridium sp. NCP 262	N	N	N
Clostridium sp. NRRL B643	N	N	N
C. saccharoperbutylacetonicum N1-4	N	N	N
C. beijerinckii NESTE 255*	Y	Y	Y

*Results not shown in Figure 2.6



B. (From Julianna Toth)

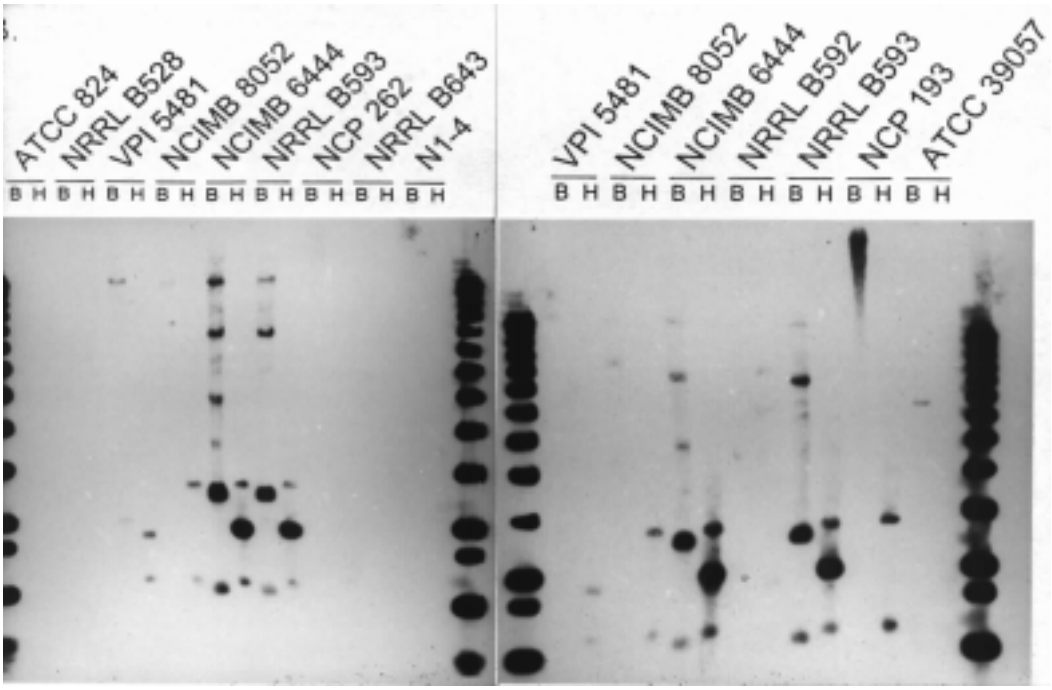


Figure 2.6: The results of Southern analysis of 12 strains of solvent-producing clostridia for the presence of *stc* and *adh*. The names of the organism were listed in Table 2.4 and abbreviated at the top of the figure. For each organism, the genomic DNA was completely digested with *Hin*dIII (represented by "H" in the figure) or *Bgl*II (represented by "B" in the figure). A, a probe for the *stc* gene was used; B, a probe for the *adh* gene was used (Figure 2.1 shows the position of the probes.).

Chapter II

Discussion

The putative Stc protein shows significant sequence similarity to the NtrC-like proteins, thus making it a potential transcription regulator. Although northern analysis did not detect the mRNA of the *stc* gene (result not shown), it is not contradictory to our speculation that the Stc is a regulatory protein, because it might be expressed at a low level. The results of RT-PCR showed that this gene was expressed. Also, Southern analysis revealed that *adh* and *stc*, when present, coexisted in the secondary alcohol-producing clostridia *C. beijerinckii* NRRL B593, *C. beijerinckii* NRRL 6444 and *C. beijerinckii* NESTE 255. For *C. beijerinckii* NCP 193, its ability to produce isopropanol was not conclusively demonstrated and this coincided with the weak signal seen in Southern analysis. These results suggest that besides the fact that *stc* and *adh* are located next to each other in the genome of *C. beijerinckii* NRRL B593, these two genes might be functionally related to secondary alcohol production.

The middle region of the Stc protein has high homology to NtrC-like proteins, yet only those regulatory proteins which are involved in the utilization of acetoin share homology in the N-terminal region with Stc. Whether or not Stc plays a similar role in *C. beijerinckii* is unknown. In *C. acetobutylicum*, the negative results of Southern analysis with the *stc* probe are consistent with the negative result of the BLASTN search of the chromosome of *C. acetobutylicum*. However, three related amino acid sequences were discovered when Stc from *C. beijerinckii* or AcoR from *C. magnum* was used as the query to do a TBLASTN search of the chromosome of *C. acetobutylicum*. The two searches detected three common coding regions. This suggests that the Stc or AcoR- like protein might not be unique to only secondary alcohol-producing clostridia or in other words, not unique to the signal transduction for secondary alcohol production. Further study of Stc may reveal some common regulatory features among solvent-producing clostridia. The location of the *stc*-related coding regions on the *C. acetobutylicum* chromosome did not give any clue for their function(s). In *C. acetobutylicum*, most known solvent-producing genes are located on the plasmid pSOL1 instead of on the chromosome[15]. Therefore whether or not there is an *stc*-like coding region on the

plasmid pSOL1 should be determined when a TBLASTN search can be performed with the plasmid sequence. Experimentally, the most direct approach to determine the function of the *stc* gene would be to generate null or multi-copy mutants of *stc* in *C. beijerinckii* NRRL B593 to see the phenotypic effect of the alterations.

Acknowledgements

I thank Gary Li for providing the *adh* probe and his previous work on the *stc* gene. Thanks also go to Julianna Toth for providing the membrane and helping to screen the 12 strains of solvent-producing clostridia in Figure 2.6.

References

[1] Chen, J. -S. 1995. Alcohol dehydrogenase: multiplicity and relatedness in the solvent-producing clostridia. FEMS Microbiol. Rev. 17, 263-273.

[2] Ismaiel, A. A., Zhu, C.-X., Colby, G.D., and Chen, J. -S. 1993. Purification and characterization of a primary-secondary alcohol dehydrogenase from two strains of *Clostridium beijerinckii*. J. Bacteriol. 175, 5097-5105.

[3] Peretz, M., Bogin, O., Tel-Or, S., Cohen, A., Li, G., Chen, J.-S., and Burstein, Y. 1997. Molecular cloning, nucleotide sequencing, and expression of genes encoding alcohol dehydrogenases from the thermophile *Thermoanaerobacter brockii* and the mesophile *Clostridium beijerinckii*. Anaerobe 3, 259-270.

[4] Li, G. 1998. Development of a reporter system for the study of gene expression to solvent production in *Clostridium beijerinckii* NRRL B592 and *Clostridium acetobutylicum* ATCC 824. Ph.D. thesis. Virginia Polytechnic Institute and State University, Blacksburg, VA, USA.

[5] Rifaat, M. and Chen, J.-S. 1992. A transcriptional regulator gene upstream from the *adh* gene of *Clostridium beijerinckii* NRRL B593: cloning and sequence analysis. Abstract for the General Meeting of the American Society for Microbiology. O-66, 320.

[6] Roos, J.W., Mclaughlin, J.K., and Papoutsakis, E.T. 1995. The effect of pH on nitrogen supply, cell lysis, and solvent production in fermentations of *Clostridium acetobutylicum*. Biotechnol. Bioengi. 27, 681-693.

[7] George, H.A., Johnson, J.L., Moore, W.E.C., Holdeman, L.V., and Chen, J.-S. 1983. Acetone, isopropanol, and butanol production by *Clostridium beijerinckii* (syn. *Clostridium butylicum*) and *Clostridium aurantibutyricum*. Appl. Environ. Microbiol. 45, 1160-1163.

[8] Johnson, J.L. 1994. Similarity analysis of DNAs, p.655-682. *In* P.Gerhardt, Murray, R. G. E., Wood, W.A., and Krieg, N.R. (ed.). Methods for general and molecular bacteriology. American Society for Microbiology, Washington, D.C.

[9] Shyamala, V. and Ames, G.F.-L. 1989. Genome walking by single-specific-primer polymerase chain reaction: SSP-PCR. Gene 84, 1-8.

[10] http://www.ncbi.nlm.nih.gov/BLAST

[11] http://biology.ncsa.uiuc.edu

[12] Kruger, N., Oppermann, F.B., Lorenzl, H. and Steinbuchel, A. 1994. Biochemical and molecular characterization of the *Clostridium magnum* acetoin dehydrogenase enzyme system. J. Bacteriol. 176, 3614-3630.

[13] Yamamoto, H., Uchiyama, S. and Sekiguchi, J. *(*1996*)* Cloning and sequencing of a 40.6 kb segment in the 73 °-76 ° region of the *Bacillus subtilis* chromosome containing genes for trehalose metabolism and acetoin utilization. Microbiology 142, 3057-3065.

[14] Biaudet, V., Samson, F., Anagnostopoulos, C., Ehrlich, S.D., and Bessieres, P. 1996. Computerized genetic map of *Bacillus subtilis*. Microbiology. 142, 2669-2729.

[15] Durre, P., Fischer, R.-J., Kuhn, A., Lorenz, K., Schreiber, W., Sturzenhofecker, B., Ullmann, S., Winzer, K., and Sauer, U. 1995. Solventogenic enzymes of *Clostridium acetobutylicum*: catalytic properties, genetic organization, and transcriptional regulation. FEMS Microbiol. Rev. 17, 251-262.

Chapter III
The cloning and sequence analysis of the *sol* operon from *C. beijerinckii* NRRL B592

Introduction

Solvent production requires the activity of a group of tightly regulated genes. Except for the multiple alcohol and aldehyde dehydrogenase, the other solvent-production enzymes are encoded by genes present in the *sol* locus. Figure 3.1[1] shows the organization of the solvent-production genes of the *sol* locus of *C. acetobutylicum* DSM 792. The *sol* locus consists of the *sol* operon and the *adc* operon, which are transcribed in opposite orientations. The *adhE*/*aad* gene encodes a bi-functional aldehyde/alcohol dehydrogenase (AdhE/Aad). The *ctfA* and *ctfB* genes code for the two subunits of acetoacetate: butyrate/acetate CoA- transferase. The *adc* gene constitutes a monocistronic operon, and it encodes acetoacetate decarboxylase.

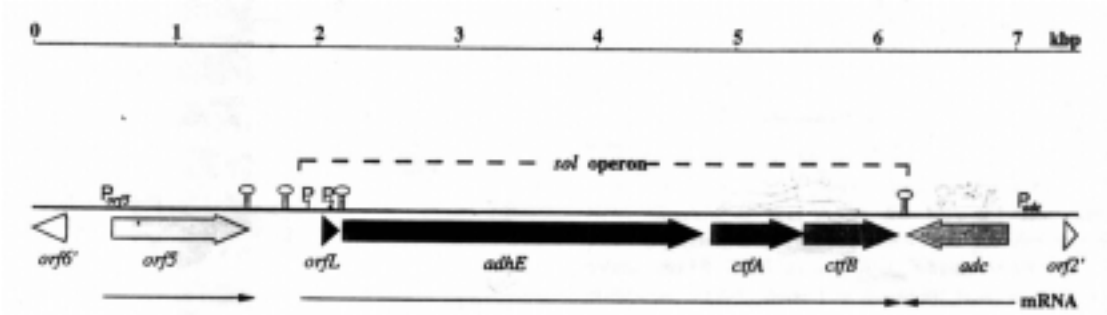


Figure 3.1: Physical map of the *sol* locus in *C. acetobutylicum* DSM 792. Putative promoters are indicated by P1, P2 and P*adc*. Hairpin symbol stands for possible stem-loop structures[1].

In *C. beijerinckii* NRRL B593, the *sol* locus (Figure 3.2) is organized differently from that in *C. acetobutylicum*[2]. Instead of the *aad/adhE* gene, an ORF upstream from the *ctfA* gene is identified as the *ald* gene (encoding aldehyde dehydrogenase). It encodes a polypeptide of 468 amino acids (MW 51,353), whose N terminal sequence matches that of the purified enzyme ALDH[3]. Also in this *sol* locus are the *ctf*A, *ctfB* and *adc* genes. The *adc* gene is in the same orientation as the other genes. A promoter region appears to be present before the *ald* gene[4].

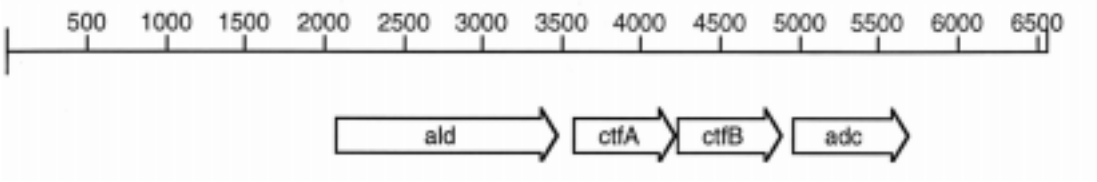


Figure 3.2: Physical map of the *sol* locus in *C. beijerinckii* NRRL B593[2].

A central question in the study of solvent fermentation is how the solventogenic switch is regulated. *C. acetobutylicum* and *C. beijerinckii* respond differently to growth conditions during the solventogenic switch[5]. The upstream region for the *sol* locus of *C. beijerinckii* NRRL B593 differed from the corresponding region preceding the *aad/adhE* gene of *C. acetobutylicum*. To understand the regulation of solventogenesis in *C. beijerinckii*, it is important to determine whether or not the organization of the solvent-production genes as found in *C. beijerinckii* NRRL B593 is unique to this strain. If this gene organization is conserved among *C. beijerinckii* strains, then the different gene organizations might reflect a general difference in the regulation of solvent production in the two species. To address this question, we chose the strain *C. beijerinckii* NRRL B592 to examine the organization of its *sol* locus.

Although belonging to the same species, *C. beijerinckii* NRRL B592 is distantly related to *C. beijerinckii* NRRL B593 because the two strains have a DNA sequence similarity of 73% as measured by DNA-DNA re-association[6]. Among other differences, *C. beijerinckii* NRRL B592 differs from *C. beijerinckii* NRRL B593 in that it does not produce isopropanol.

Chapter III

This study began with the cloning of the gene encoding the ALDH because a promoter region is likely to precede this gene. The CoA-acylating ALDH catalyzes the reaction:

$$\text{Acyl CoA} + \text{NAD(P)H} + H^+ \leftrightarrow \text{Aldehyde} + \text{CoA} + \text{NAD(P)}^+ \quad [7]$$

In solvent production, ALDH catalyzes the first specific reaction of the alcohol (ethanol or butanol) producing pathway. In *C. beijerinckii*, the activity of ALDH increases during the switch from acid production to solvent production[8]. An ALDH from *C. beijerinckii* NRRL B592 has been purified under anaerobic conditions[7]. The ALDH is a homodimer with a native molecular weight of about 100,000 and a subunit MW of 55,000. A smilar ALDH has been purified from *C. beijerinckii* NRRL B593[3]. The *C. beijerinckii* ALDH shows no alcohol dehydrogenase activity, which distinguishes it from the Aad/AdhE protein of *C. acetobutylicum*; the latter has both butyraldehyde dehydrogenase and butanol dehydrogenase activities[9]. From *C. beijerinckii* NRRL B593, the *ald* gene which encodes the ALDH has been cloned[3]. In the N-terminal 29 amino acids of the ALDHs from the two strains, only two amino acids are different and one of the two differences involves a conservative substitution (V - L) (Table 3.1). The high similarity suggests that the *ald* gene is conserved in the two strains. Therefore, a probe based on the *ald* gene of *C. beijerinckii* NRRL B593 was used for detecting the *ald* gene of *C. beijerinckii* NRRL B592. The *ald*-containing DNA fragment from *C. beijerinckii* NRRL B592 was then analyzed for the presence of other solvent-production genes.

Table 3.1: A comparison of the N-terminal amino acid sequence of ALDHs from *C. beijerinckii* NRRL B592 and *C. beijerinckii* NRRL B593.

N-terminal region of ALDH from B592 (Based on amino acid sequence of purified enzyme)[7]	MNKDTLIPTTKDLK*V*KTN*G* ENINLKNYKD…
N-terminal region of ALDH from B593 (Based on amino acid sequence of purified protein and nucleotide sequence)[3]	MNKDTLIPTTKDLK*L*KTN*V* ENINLKNYKD…

Chapter III

Materials and Methods

Bacterial strains and plasmids

C. beijerinckii NRRL B592, *C. beijerinckii* NRRL B593 and *E. coli* DH5α were from laboratory stocks. Plasmid Litmus 29 was from New England BioLabs.

Bacterial growth conditions

E. coli was routinely grown in the Luria-Bertani medium aerobically at 37°C. Media were supplemented with ampicillin (100μg/ml) for selection purposes.

The clostridial cultures were grown anaerobically under an atmosphere of 10% CO_2, 10% H_2 and 80% N_2. One ml aliquots of clostridial spores were heat-shocked in boiling water for 1.5 minutes and then inoculated into 45 ml CSM (Clostridial Soluble Medium[10]). The cultures were grown without shaking for 18-24 hours at 32 °C and then transferred into 400 ml TYG-60 medium (60g/L glucose)[11] and incubated at 32 °C with mixing.

DNA isolation, manipulation and transformation

Rapid, small-scale isolation of plasmid DNA was performed by using a miniprep kit (Qiagen). Large-scale isolation of plasmid DNA was performed with a midiprep kit (Qiagen). Genomic DNA was isolated according to a modified Marmur procedure[12].

Digestion with restriction enzymes (New England BioLabs) was performed according to the manufacturer's specifications and the DNA fragments were separated on a SeaKem LE agarose (FMC BioProducts) gel (0.8% or 1.2%). The fragments of interest were extracted from the gel with a gel purification kit (Qiagen). The resulting DNA fragments were then ligated into Litmus 29 (ligase from GIBCO BRL).

Transformation of *E. coli* with plasmids was performed with competent cells prepared as described in **Operating Instructions and Applications Guide** (September, 1992) for the Bio-Rad Gene Pulser (Bio-Rad).

Chapter III

Total RNA isolation

C. beijerinckii NRRL B593 was grown and sampled at different time points. Total RNA isolation was performed using a RNeasy Mini Kit supplemented with RNase-free DNase (Qiagen).

Probe generation through PCR

PCR was performed according to the procedure recommended by the supplier of *Taq* polymerase (Promega Life Science) in a Genius thermal cycler (Techne).

Probes for the *ald, ctfA, ctfB* and *adc* were generated using the genomic DNA of *C. beijerinckii* NRRL B593 as the template. Figure 3.3 shows the sequence and the position of the primers kindly provided by Julianna Toth.

The probe for the *ald* gene encompassed the region from 231 bp upstream to 291 bp downstream of the start codon of the *ald* gene of *C. beijerinckii* NRRL B593. The probe for the *ctf* genes encompassed the region beginning from 308 bp downstream of the *ctfA* start codon to 452 bp downstream of the *ctfB* start codon. The probe for the *adc* gene encompassed the region from 129 bp upstream of the start codon to 113 bp downstream of the stop codon of the *adc* gene.

Southern/northern analysis

The probe was labeled using the ECL direct nucleic acid labeling and detection system (Amersham/Pharmacia Biotech). Genomic DNA digested with appropriate restriction enzymes or the total RNA was separated by electrophoresis on an agarose gel. DNA fragments or RNA were transferred to Hybond-N^+ membrane (Boehringer Mannheim). Prehybridization, hybridization, washing, labeling and detection were performed according to the manufacturer's instructions (Amersham/Pharmacia Biotech).

Cloning and subcloning

The *ald* probe from *C. beijerinckii* NRRL B593 detected a 4-kb *Hin*dIII fragment from *C. beijerinckii* NRRL B592. The 4 kb *Hin*dIII-fragments of *C. beijerinckii* NRRL B592 DNA were cut out of the agarose gel and the DNA was recovered with a gel extraction kit (Qiagen). The DNA fragments were ligated into the *Hin*dIII-digested

cloning vector and then used to transform *E. coli* DH5α. The transformants were selected on LB plates supplemented with ampicillin (100μg/ml), 4μM IPTG and 40μg/ml X-gal. Based on the blue-and-white selection, white colonies were transferred to another LB plate (with ampicillin 100μg/ml) and lifted to a Hybond–N^+ filter (Amersham/Pharmacia Biotech). The membrane was screened with the *ald* probe. Secondary screening using the same probe was carried out to confirm the positive colonies. After plasmid isolation and appropriate digestion, Southern analysis with the *ald* probe was performed to check the inserts. The subclones containing the *Hin*dIII-*Eco*RI fragments were screened for the presence of the *ald, ctf* or *adc* genes by using both Southern hybridization and PCR amplification.

DNA sequencing

The insert DNA was sequenced on either a Pharmacia ALF Express II automated DNA Sequencer of the DNA Sequencing Facility at the Center for Molecular Medicine and Infectious Diseases, Virginia-Maryland Regional College of Veterinary Medicine or an ABI Prism 377 DNA sequencer of Davis Sequencing at Davis, California. Sequencing primers were generated based on the known sequences and synthesized by GIBCO BRL. The clones were sequenced for both strands and the ambiguous bases were clarified by repeated or overlapped sequencing.

Sequence Analysis

Sequence information was managed with the LaserGene application of DNASTAR. The homology search was carried out with BLAST 2.0 at the NCBI web site[13]. The multiple sequence alignments and the secondary structure predictions were accomplished with, respectively, the protein secondary-structure prediction tool and the protein multiple sequence alignment tool CLUSTALW at the Biology WorkBench web site[14].

Results

Sequence and analysis of the *sol* locus of *C. beijerinckii* NRRL B592

The *ald* gene on a 4-kb *Hin*dIII fragment of *C. beijerinckii* NRRL B592 was cloned into Litmus 29 and named as "*ald*". Later, this clone was also found to contain the *ctfA*, *ctfB* and part of the *adc* genes. To facilitate sequencing, two *Eco*RI-*Hin*dIII subclones were constructed (Figure 4.3a). The *Hin*dIII fragment was partially sequenced. The sequenced region was 3440bp in length and contains three complete and one partial open-reading frame. Figure 3.4b shows the DNA and the deduced amino acid sequences of the cloned region.

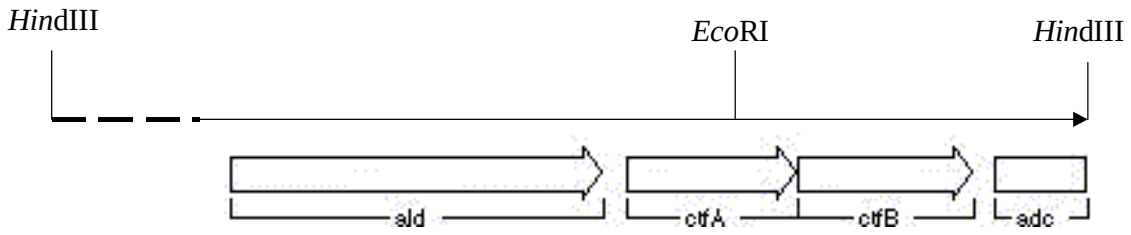


Figure 3.4a: Physical map for the cloned *sol* locus from *C. beijerinckii* NRRL B592. The solid line refers to the sequenced region and the dashed line refers to the cloned but yet unsequenced region.

The first ORF was 1,404 bp long and preceded by a putative ribosome-binding site (AGGAGG) at -14 to –9 bp before the start codon. This ribosome-binding site was also present in several solvent-producing genes in *C. acetobutylicum* such as *bdhA* and *bdhB*[1]. The complete ORF encoded a polypeptide of 468 amino acid residues, with a calculated MW of 51,316 and a predicted pI of 6.064. The N-terminal 29 amino acid residues matched the sequence determined from the purified ALDH[6]. This ORF was identified as the *ald* gene of *C. beijerinckii* NRRL B592. The coding region of the *ald* gene had a mol% G+C content of 32.5.

The ORF 91 bp downstream from the *ald* gene was the *ctfA* gene, which encoded the small subunit of the acetoacetate: acetate/butyrate CoA-transferase. The *ctfA* gene was preceded by a putative ribosome-binding site (GGAG) at –12 to –9 bp from the start codon, and there was no potential transcription termination site between the *ald* and the *ctfA* gene. The *ctfA* gene encoded a polypeptide of 217 amino acid residues, with a calculated MW of 23,285 and a predicted pI of 7.041. The N-terminal 39 amino acid residues matched the sequence determined from the purified CtfA from *C. beijerinckii* NRRL B593: MNKLVKLTDLKRIFKDGMTIMVGGFLDXGTPE NIIDMLVD[15].

The ORF directly downstream of the *ctfA* gene was the *ctfB* gene, which encoded the large subunit of the CoA transferase. The start codon for the *ctfB* gene was UUG which followed the stop-codon of the *ctfA* gene, with no other bases separating the two genes. The putative ribosome-binding site GGAG for the *ctfB* was located –13 to –10 bp from the start codon and was within the coding region of the *ctfA* gene. The arrangement of the *ctfA* and *ctfB* genes suggested the presence of translational coupling, which is a mechanism for the coordinated expression of different subunits for a protein. The UUG codon only occurred once in the coding region of the *ctfB* gene. The *ctfB* gene encoded a polypeptide of 221 amino acid residues, with a calculated MW of 23,769 and a predicted pI of 5.635. The 45 N-terminal amino acid residues had two differences from the sequence determined from the purified CtfB from *C. beijerinckii* NRRL B593: MVDKVLAKEIIAKRVAKELKK**G**QLVNLGIGLPTLVANYVPKE**Y**M[15].

The fourth ORF occurred 72 bp downstream from the *ctfB* gene and was identified as the *adc* gene, which encoded the acetoacetate decarboxylase. The *adc* gene was preceded by a putative ribosome-binding site (GAGG) at –17 to –12 bp from the start codon, and there was no potential transcription termination site between the *ctfB* and *adc* genes. Only the N-terminal 117 amino acid residues were present in the cloned region.

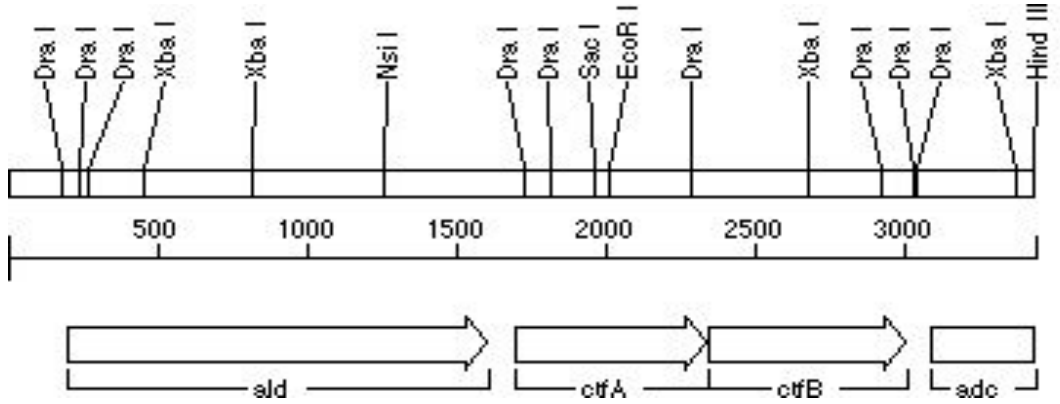
Since the identification of these genes was largely based on a similarity to the corresponding genes of *C. beijerinckii* NRRL B593, a comparison of the *sol* region from

the two strains were made. Figure 3.5 shows the restriction sites of the *sol* region from the two strains. Furthermore, an alignment of the *sol* region of *C. beijerinckii* NRRL B592 and *C. beijerinckii* NRRL B593 was performed. Figure 3.6 shows the alignments for the polypeptides of ALDH, CtfA and CtfB. Table 3.2 summarizes the data for the alignments.

Table 3.2: Comparison of the DNA and the encoded polypeptides of the *sol* locus from *C. beijerinckii* NRRL B592 and *C. beijerinckii* NRRL B593.

Organisms		*C. beijerinckii* NRRL B592	*C. beijerinckii* NRRL B593	Identity
The nucleotide sequence	Overall	3437 bp	3437 bp	97.3%
	ald	1404 bp (mol% G+C: 32.5)	1404 bp(mol% G+C: 31.9)	97.2%
	ctfA	651 bp (mol% G+C: 32.9)	651 bp (mol% G+C: 32.3)	98.5%
	ctfB	663 bp (mol% G+C: 33.2)	663 bp (mol% G+C: 34.2)	95.7%
	adc (partial)	351 bp	351 bp	98.9%
The deduced a.a. sequence	ALDH	468 a.a.	468 a.a.	97%
	CtfA	217 a.a.	217 a.a.	99.5%
	CtfB	221 a.a.	221 a.a.	97%
	ADC (partial)	117 a.a.	117 a.a.	97.4%

A.



B.

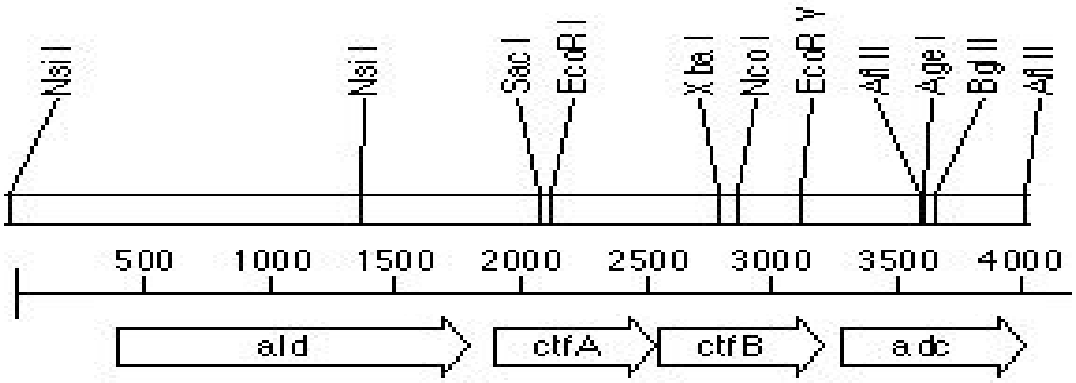


Figure 3.5: The gene and restriction maps of the *sol* cluster in *C. beijerinckii* NRRL B592 (A) and *C. beijerinckii* NRRL B593 (B).

Northern Analysis:

The probe "*ald*" or "*adc*" was used to screen the total RNA from *C. beijerinckii* NRRL B593 (Figure 3.7a). A 4kb RNA band was detected by either probe in cells harvested at 4h after inoculation. A similar band was seen using cells harvested at 9.5h. However, cells harvested at 2h or 7h did not give a band. From the gel picture (results not shown), we know that the reason that cells harvested at 7h did not give a band is due to the heavy degradation of this sample.

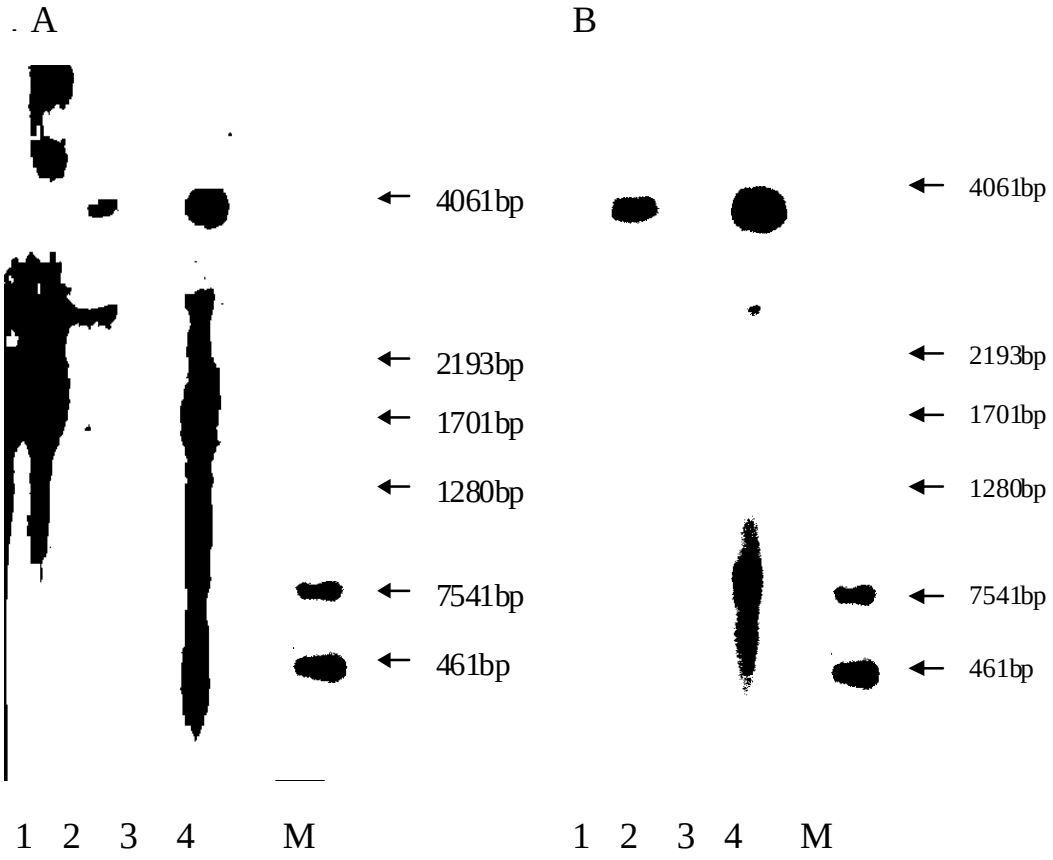


Figure 3.7a: Northern analysis of total RNA isolated from *C. beijerinckii* NRRL B593 with the "*ald*" probe (A) or the "*adc*" probe. (B). Lanes 1, 2, 3 and 4 correspond to cultures sampled at 2h, 4h, 7h and 9.5 h after inoculation as shown in Figure 3.7 b. "M" refers to the lane for the molecular weight ladder. 10μg of total RNA was loaded for each sample.

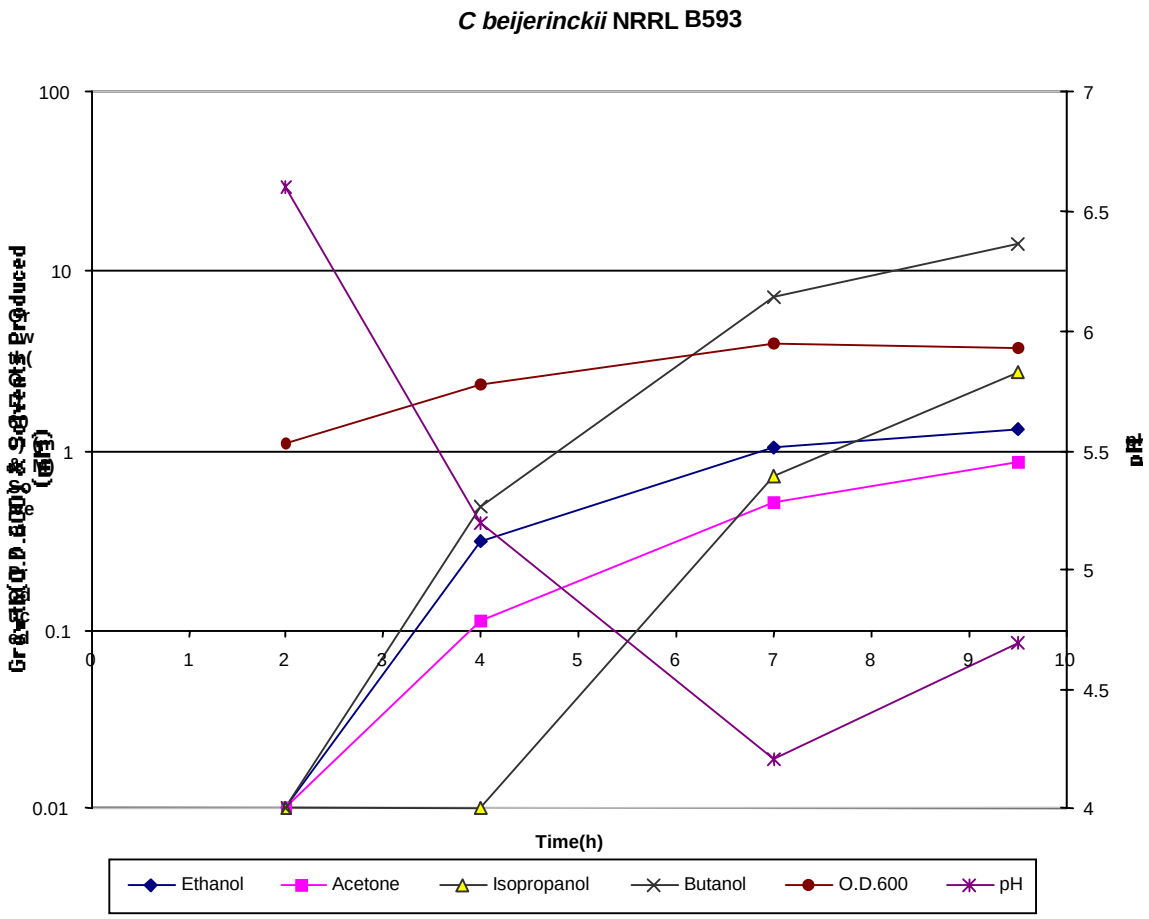


Figure 3.7b: Growth and solvents produced for a TYG-60 culture of *C. beijerinckii* NRRL B593 used for the northern analysis.

Chapter III

Discussion

The differences in the organization of the *sol* locus in *C. acetobutylicum* and *C. beijerinckii* suggest that solvent production may be regulated differently in the two species. In *C. acetobutylicum*, there are two promoters for the *sol* operon P1 and P2 (Figure 3.1). P1 is more similar to the vegetative promoter. P2 and P*adc* each has a heptamer motif (-ACAATAT-) located between the –10 and –35 regions of the promoters and forms part of a stem-loop structure that engulfs the –35 region. The P2 and P*adc* initiate high level transcription at the onset of solventogenesis[1]. In *C. beijerinckii*, with the *adc* gene included in the *sol* operon, this operon is responsible for the production of acetone as well as aldehydes. The ALDH encoded by the *ald* gene in the *sol* operon is the predominant ALDH in *C. beijerinckii* NRRL B592 and *C. beijerinckii* NRRL B593[3, 7]. A potential promoter motif (TTGATA --N_{15}-- TATATT) is present –157 bp to –131 upstream of the *ald* start codon. In northern analysis, besides the 4-kb band, the presence of ambiguous bands with the sizes of around 1.5 kb and 2.9 kb may be due to the degradation of the RNA or unspecific hybridization with abundant 16S rRNA (about 1.5 kb) and 23S rRNA (about 2.9 kb). The latter was confirmed when other probes such as *stc* or *hbd* (gene encoding 3-hydroxybutyryl CoA dehydrogenase) were used (results not shown). The different integrity of the RNA samples also caused the loading of uneven amount of total RNA in each lane in spite of the fact that data of UV measurements were used as a reference for loading. Since no internal control is available, it is not possible to ascertain the pattern of the expression of the *sol* operon at this point even though the northern results themselves display certain on-and-off pattern. Further studies of the location and features of the P*sol* in *C. beijerinckii* would help us to understand the regulation of this promoter.

Southern analysis showed that there is no *ald*-like DNA sequence present in *C. acetobutylicum*. But it was present in three other species of solvent-producing clostridia[3]. A TBLASTN search indicated that in the chromosome of *C. acetobutylicum*, there were two amino acid sequences that were related to ALDH from *C. beijerinckii*[3]. It would be

interesting to know whether or not an ALDH is encoded by one of the ORFs identified by the TBLASTN search.

ALDH is present in all kind of organisms. Besides the bacterial CoA-dependent ALDHs which catalyze the reversible conversion between acyl-CoA and the corresponding aldehydes, another group of ALDHs, the CoA-independent ALDHs, are present in both prokaryotes and eukaryotes. The CoA-independent ALDHs act in detoxifying a wide variety of organic compounds and toxins[3]. Recently, more than 200 ALDH sequences which represent a wide range of substrate specificity and organisms have been aligned[16]. With the availability of X-ray crystal structures for ALDH, this alignment has proved useful in predicting the structure of other ALDHs and in interpreting site mutagenesis results[16]. The ALDH from both *C. beijerinckii* NRRL B592 and B593 has the three invariant amino acid residues for the CoA-dependent ALDHs: Gly (227), Cys (280) and Glu (376), but not the other three conserved amino acid residues for CoA-independent ALDHs. The six residues: Thr (89), Met (91), Thr (148), Gly (242), His (398), and Gly (449), which likely occupy positions lining the catalytic pocket are conserved in these two ALDHs[3]. Further structural study would determine whether or not some of these residues have a role in CoA-binding and also what residues are involved in the coenzyme specificity.

Coenzyme A-transferases are a family of enzymes with a diverse substrate specificity and subunit composition. Members of this group of enzymes are found in anaerobic fermenting bacteria, aerobic bacteria and in the mitochondria of humans and other mammals. All CoA-transferases share a common mechanism in which the CoA moiety is transferred from a donor (e.g., acetoacetyl-CoA) to an acceptor, acetate, whereby acetyl-CoA is formed. The transfer has been described by a ping-pong mechanism in which CoA is bound to the active-site residue of the enzyme as a covalent thiol ester intermediate [15, 17].

As expected, the CoA transferase (both subunits) are shown conserved in a wide range of bacteria (BLAST search, result not shown). Among them are *Clostridium acetobutylicum, Haemophilus influenzae Rd, Bacillus subtilis, Escherichia coli,* and

Helicobacter pylori. CtfB is also conserved in *Streptomyces sp.* 2065, *Sphingomonas* sp., and *Mycobacterium tuberculosis*. Currently, the only crystal structure known for the CoA-transferase is glutaconate CoA transferase from *Acidaminococcus fermentans* (GenBank Accession No. 2981809)[17]. No significant similarity was found between the sequence of this glutaconate CoA transferase and the CtfA or CtfB sequence from *C. beijerinckii* (BLAST 2sequence alignment, result not shown). It remains to be shown whether or not the different amino acid sequences can fold into a similar protein with a similar reaction mechanism.

Conclusion

The sequence data revealed that the *sol* locus is highly conserved in *C. beijerinckii* NRRL B592 and *C. beijerinckii* NRRL B593. Northern analysis indicated that the four ORFs in the *sol* region could be transcribed on the same messenger RNA with a size of about 4 kb. Thus, this region can be described as a *sol* operon with the organization of the genes *ald-ctfA-ctfB-adc*.

Acknowledgements

I thank Julianna Toth for providing the sequence information and the primers for the *sol* region of *C. beijerinckii* NRRL B593. Thanks also go to Adnan A. Ismaiel and Murat Kasap for the enzyme information and their assistance in culturing the bacteria.

References

[1] Durre, P., Fischer, R.-J., Kuhn, A., Lorenz, K., Schreiber, W., Sturzenhofecker, B., Ullmann, S., Winzer, K., and Sauer, U. 1995. Solventogenic enzymes of *Clostridium acetobutylicum*: catalytic properties, genetic organization, and transcriptional regulation. FEMS Microbiol. Rev. 17, 251-262.

[2] Toth, J. and Chen, J.-S. Cloning and sequence analysis of the *ctfA* and *ctfB* genes encoding the acetoacetate: butyrate/acetate coenzyme A-transferase from *Clostridium beijerinckii* NRRL B593. 1997. Abstract for the General Meeting of the American Society for Microbiology. H-8, 286.

Toth, J. and Chen, J.-S. Organization of the acetone-butanol production genes in *Clostridium beijerinckii* NRRL B593. 1998. Abstract for the General Meeting of the American Society for Microbiology. O-39, 399.

[3] Toth, J., Ismaiel, A.A., and Chen, J.-S. 1999. The *ald* gene encoding a coenzyme A-acylating aldehyde dehydrogenase distinguishes *Clostridium beijerinckii* and two other solvent-producing clostridia from *Clostridium acetobutylicum.* . Appl. Environ. Microbiol. (submitted)

[4] Li, G. Unpublished results, 1998.

[5] Johnson, J.L and Chen, J.-S. 1995. Taxonomic relationships among strains of *Clostridium acetobutylicum* and other phenotypically similar organism. FEMS Microbiol. Rev. 17, 233-240.

[6] Johnson, J.L., Toth, J., Santiwatanakul, S., and Chen, J.-S. 1997. Cultures of "*Clostridium acetobutylicum*"from various collections comprise *Clostridium acetobutylicum*, *Clostridium beijerinckii*, and two other distinct types based on DNA-DNA reassociation. Int. J. Syst. Bacteriol. 47, 420-424.

[7] Yan, R.T., Chen, J.S. 1990. Coenzyme A-acylating aldehyde dehydrogenase from *Clostridium beijerinckii* NRRL B592. Appl. Environ. Microbiol. 56, 2591-2599.

[8] Yan, R.T., Zhu, C.-X., Golemboski, C., and Chen, J.S. 1988. Expression of solvent-forming enzymes and onset of solvent production in batch cultures of *Clostridium beijerinckii.* Appl. Environ. Microbiol.54, 642-648.

[9] Ismaiel, A.A. and Chen, J.-S. 1998. Purification of the aldehyde-alcohol dehydrogenase encoded by the *aad* gene from *Clostridium acetobutylicum* ATCC 824. Abstract for the general Meeting of the American Society for Microbiology. O-40, 400.

[10] Roos, J.W., Mclaughlin, J.K., and Papoutsakis, E.T. 1985. The effect of pH on nitrogen supply, cell lysis, and solvent production in fermentations of *Clostridium acetobutylicum.* Biotechnol. Bioengi. 27, 681-693.

[11] George, H.A., Johnson, J.L., Moore, W.E.C., Holdeman, L.V., and Chen, J.-S. 1983. Acetone, isopropanol, and butanol production by *Clostridium beijerinckii* (syn. *Clostridium butylicum*) and *Clostridium aurantibutyricum.* Appl. Environ. Microbiol. 45, 1160-1163

[12] Johnson, J.L. 1994. Similarity analysis of DNAs, p.655-682. *In* P.Gerhardt, Murray, R. G. E., Wood, W.A., and Krieg, N.R. (ed.). Methods for general and molecular bacteriology. American Society for Microbiology, Washington, D.C.

[13] http://www.ncbi.nlm.nih.gov/BLAST

[14] http://biology.ncsa.uiuc.edu

[15] Colby, G.D. 1993. CoA-transferase and 3-hydroxybutyryl-CoA dehydrogenase: acetoacetyl-CoA-reacting enzymes from *Clostridium beijerinckii* NRRL B593. Ph.D. thesis. Virginia Polytechnic Institute and State University, Blacksburg, VA, USA.

[16] http://www.psc.edu/biomed/pages/research/Col_HBN_ALDH.html

[17] Jacob, U., Mack, M., Clausen, T., Huber, R., Buckel, W. and Messerschmidt, A. 1997. Glutaconate CoA-transferase from *Acidaminococcus fermentans*: the crystal structure reveals homology with other CoA-transferases. Structure 5, 415-426.

Chapter IV

Concluding Remarks

With more and more genetic tools available for the solvent-producing clostridia, a comprehensive study of these organisms is now feasible. Based on the results reported in this thesis, areas of future investigation could include the following:

The *stc* gene was expressed in *C beijerinckii* NRRL B593. Primer extension analysis would be useful for further characterization of its promoter. The potential promoter sequence can be compared with other known promoter sequence in clostridia to help elucidate the possible regulation mechanism. In order to test our presumption about the function of the Stc protein, a direct approach is to make an *stc* knockout mutant or to add multiple copies of *stc* to see the phenotypic effect(s). The antisense RNA method, although not mutagenesis, would be a promising approach to generate Stc^- phenotype. Since we do not know if Stc is essential for the organism, an inducible promoter can be utilized in the antisense RNA construct to control the transcription of the antisense RNA.

The *sol* region encodes the enzymes crucial to solvent production. Genetic and biochemical analysis of its promoter might reveal features of the regulation for solventogenesis. The construction of a reporter system to study the function of the promoter of the *sol* operon and other solvent-production related promoters will help us to understand what triggers and how it triggers the onset of solventogenesis.

Vita

Rui Hong

Rui Hong was born in Kunming, China --- a beautiful *spring (season) city* in Southwest China on April 25, 1974. She finished her pre-college education there and went to Fudan University in Shanghai, China, in September 1992. There, she spent one year in military training, then a four-year study in biological sciences. In June 1997, she graduated from Fudan as an honor student and a multi-scholarship recipient with the Bachelor of Science degree in Biophysics and Physiology. In August 1997, she was enrolled in the Department of Biochemistry at Virginia Tech and later, joined Dr. Jiann-Shin Chen's group. At Virginia Tech, she held the graduate assistantship supported by the department and Dr. Chen. In early 1999, she received the B.M. Anderson Award for her performance in the first year of graduate study. She graduated with the Master of Science degree in August. Rui went on to the Department of Pharmacology, University of Pennsylvania to pursue her study for the Ph.D. degree.