INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI
A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor MI 48106-1346 USA
313/761-4700  800/521-0600
RICE UNIVERSITY

CLONING AND SEQUENCING OF GENES INVOLVED IN GLYCOLYSIS FROM CLOSTRIDIUM ACETOBUTYLICUM

by

ED BELOUSKI

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE MASTER OF ARTS

APPROVED, THESIS COMMITTEE

George N. Bennett, Professor
Biochemistry and Cell Biology

Richard Gomer, Associate Professor
Biochemistry and Cell Biology

James W. Campbell, Professor
Biochemistry and Cell Biology

Susan Gibson, Assistant Professor
Biochemistry and Cell Biology

Ronald J. Parry, Professor
Chemistry

Houston, Texas
April, 1996
Abstract

Cloning and Sequencing of Genes Involved in Glycolysis from *Clostridium acetobutylicum*

by

Ed Belouski

Complementation studies were carried out with a *Clostridium acetobutylicum* ATCC 824 plasmid library and *Escherichia coli* glycolytic mutants. Degenerate primers were designed from homologous regions in known glycolysis genes, and used to screen the selected plasmids. The gene encoding phosphofructokinase and the surrounding DNA was sequenced. The gene encoding pyruvate kinase was discovered downstream of the phosphofructokinase gene. It appears that phosphofructokinase and pyruvate kinase form an operon. The gene encoding glyceraldehyde-3-phosphate dehydrogenase was also found in a cosmid library, but only partially sequenced.

Phosphofructokinase from *C. acetobutylicum* has high homology to the phosphofructokinases from *Bacillus stearothermophilus*, *Bacillus macquarensis*, *Escherichia coli* and other microbes. Conserved residues involved in the binding of substrates, cofactors and allosteric effectors are found in the *C. acetobutylicum* homologue. The pyruvate kinase/phosphofructokinase operon may act as a "regulatory operon" for glycolysis in *C. acetobutylicum*. 
Acknowledgements

I would like to thank George Bennett for taking me into his lab, and for the help thereafter. I would also like to thank everyone in the lab for their help and advice. I'd especially like to thank them for the fraternization in and out of the lab. I would like to thank my committee and any other faculty in the department that I have had to deal with. In particular, I would like to thank Bruce Cooper for all the morning training sessions. There are many others I would like to thank, but I will do that some other time, in some other way.
Oligonucleotides and their Design .............................................. 29
PCR .................................................................................................. 32
Agarose Gel and Polyacrylamide Gel Electrophoresis ................. 34
Recovery of DNA Fragments from Agarose Gels ....................... 34
DNA Ligation ............................................................................... 35
DNA Sequencing ........................................................................... 36
Computer Resources .................................................................... 36

Results

Escherichia coli Glycolysis Mutants ............................................. 38
Restrictive Media Screening ......................................................... 38
Polymerase Chain Reaction ......................................................... 39
Subcloning pAM1e ...................................................................... 40
Sequencing pBCpfk .................................................................... 43

Discussion

Homology of Clostridium acetobutylicum PFK with PFK from other Organisms, Particularly Bacillus stearothermophilus ... 49
ORFpyk ....................................................................................... 54

References .................................................................................. 58

Appendices

pRUNC51 ...................................................................................... 69
The F<sub>O</sub> Complex of the Proton-translocating F-type ATPase from Clostridium acetobutylicum .................................................. 83
List of Abbreviations

Ap  ampicillin
bp  base pair
BSA bovine serum albumin
Cm  chloramphenicol
dH₂O deionized water
ds  double stranded
EDTA ethylenediamine tetraacetate
ENO enolase
EtBr ethidium bromide
FBA fructose-bisphosphate aldolase
GAP glyceraldehyde 3-phosphate dehydrogenase
GMM glycolysis minimal medium
kb  kilobase
kDa kilo-Dalton
LB Luria broth
M  molar
NaOAc sodium acetate
Neo neomycin
PCR polymerase chain reaction
PFK 6-phosphofructokinase
PGI phosphoglucoisomerase
PGK phosphoglycerate kinase
psi pounds per square inch
TSB Transformation and Storage Buffer
TE 10 mMTris-Cl, 1 mM EDTA
Tet tetracycline
UV ultraviolet light
X-gal 5-bromo-4-chloro-3-indolyl-β-D-thiogalactosidase
List of Tables

Table 1. Glycolysis mutants and carbon sources ........................................... 8
Table 2. Plasmids and Cosmids used in this study ........................................ 26
Table 3. List of primers used in this study .................................................... 33
Table 4. Homology of PFK with PFK from other organisms ..................... 52
Table 5. Homology of PYK with PYK from other organisms .................... 55
List of Figures

Figure 1. A simple schematic of glycolysis ........................................ 5
Figure 2. Regulation of glycolysis .................................................. 16
Figure 3. Drawings of the subunit and tetramer of PFK ............... 18
Figure 4. Homologous regions used for primer design ................. 30
Figure 5. Map of pAM1e .................................................................. 41
Figure 6. Cloning strategy to produce pBCpfk .............................. 42
Figure 7. Sequencing strategy and restriction map of pBCpfk 44
Figure 8. Complete sequence of the pBCpfk insert ...................... 45
Figure 9. Amino acid sequence alignment of PFK ......................... 50
Figure 10. ORF map of the pBCpfk insert ....................................... 55
Figure 11. Complete sequence of the pRUNC51 insert ................. 71
Figure 12. ORF map of the pRUNC51 insert .................................... 78
Figure 13. Amino acid sequence alignment of DEF .................... 80
Figure 14. Amino acid sequence alignment of FMT ..................... 81
Figure 15. Partial sequence of the pTEC011 insert ...................... 84
Figure 16. Amino acid sequence alignment of ATPa .................. 90
Figure 17. Amino acid sequence alignment of ATPc .................. 92
Figure 18. Amino acid sequence alignment of ATPb .................. 93
Figure 19. Hydropathy profile and topology of ATPa ............... 98
Figure 20. Hydropathy profiles of ATPb and ATPc .................. 99
Introduction

*Clostridium acetobutylicum*

The genus *Clostridium* includes a diverse array of over 80 gram-positive species, making it one of the largest among prokaryotes. This genus is classified on four criteria: endospore formation; restriction to anaerobic energy metabolism; inability to carry out dissimilatory reduction of sulfate; and a gram-positive cell wall, although it may sometimes react gram-negative (Anderseen et al., 1989). This last criteria has become somewhat ambiguous as some species only exist as gram-positive organisms in young cultures. Other species are entirely gram-negative (Cato & Stackebrandt, 1989). Another example of this genera's heterogeneity is found with regard to the G+C content of the DNA. A first group, which includes *Clostridium acetobutylicum*, has a G+C content in the upper 20% range. While a second group may have a G+C content as high as 55%. Consideration has been given to dividing the genus on this basis (Cato et al., 1986). *Clostridia* have been found in soil, lake sediments, well water, clam gut, bovine feces and human feces (Smith & Hobbs, 1974).

Even within the *Clostridium acetobutylicum* species, there is a complex variety of strains. This study used the *C. acetobutylicum* type (reference) strain, which is American Type Culture Collection (ATCC) 824; a very similar or even identical strain to Deutsche Sammlung von Mikroorganismen (DSM) 792 (Cato et al., 1986; Wilkinson et al., 1995). The former reference strain, National
Collection of Industrial and Marine Bacteria (NCIMB) 8052, is now thought to be a *C. beijerinckii* (originally called *C. rubrum* or *C. butylicum*). This reclassification occurred after sequence analysis was carried out on the genes encoding 16S rRNA (Hutson *et al.*, 1993), and because it lacks the active restriction system found in ATCC 824 (Wilkinson & Young, 1995; Wilkinson *et al.*, 1995). Strains differ genotypically and phenotypically with regard to sporulation, metabolism, etc. Gene sequences may show high homology, but not always identity (Wilkinson *et al.*, 1995; Woolley and Morris, 1990). Finally, the reported sizes of their genomes vary drastically: ATCC 824 has a 4.0 Mb genome; P262, a 2.9 Mb genome; DSM 1731, a 3.5 Mb genome; NI-4081, a 6.5 Mb genome; and the aforementioned NCIMB 8052 has a 6.5 Mb genome (Wilkinson and Young, 1993). In comparison, the related *Bacillus subtilis* has a 4.2 Mb genome (Moszer *et al.*, 1995), and *Escherichia coli* has a 4.7 Mb genome (Medigue *et al.*, 1993). It has been suggested that some of these sizes may have been miscalculated, but it could also be that there are additional misclassifications of some of these strains.

Interest in *C. acetobutylicum* stems from its ability to produce high levels of acetate/butyrate and acetone/butanol. Researchers in the early part of this century studied ways to produce synthetic rubber from butanol. One area of investigation was the industrial use of microbes to produce the alcohols. In 1914, Chaim Weizmann isolated an alcohol producing strain he called BY. This strain would later be renamed *C. acetobutylicum* (McCoy *et al.*, 1926). Greater study was devoted to microbial fermentation due to the demand for acetone brought on by World War I munitions production. In 1916
London, Weizmann assisted in the development of the first microbial fermentation plants to produce acetone (Jones & Wood, 1986). Plants continued to open during World War I, and had spread worldwide by the 1930's. World War II made the industry even stronger as fermentation products were in high demand. It was not until the 1950's that the competition from petroleum based solvent production caused the decline, and ultimately the end of the industry.

Since then, research has continued in the field and microbial production of solvents is considered a possible alternative to chemical synthesis. Arguments for microbial production include: the fluctuations in the prices and availability of petroleum products; the manipulation and discovery of strains producing higher solvent concentrations and with an increased tolerance for solvents; and improved bioreactor designs which allow for more cost-efficient production and recovery of products. Clostridia fares well in comparisons to other fermentative microbes. Anaerobic fermentation is cheaper and easier to design, because oxygen exchange is not necessary. Clostridia can use "waste products" from industrial and agricultural processes as carbon sources. C. acetobutylicum in particular, produces large surpluses of reductive power that serve to drive high levels of solvent production (Jones & Woods, 1986). Furthermore, genetic manipulation is now feasible for C. acetobutylicum (Mermelstein et al., 1994; Green & Bennett, 1996). This metabolic engineering allows for gene enhancement or deletion, which may channel intermediates through specific pathways while eliminating other pathways. Finally, the combination of broad carbon
utilization, its anaerobic growth and unusual reductive capabilities makes *C. acetobutylicum* an attractive candidate for biotransformation and biocatalysis (Morris, 1993).

Glycolysis or the Embden-Meyerhof-Parnas pathway

The pathway commonly called glycolysis, was determined by Embden, Meyerhof and Parnas (Figure 1). For this reason, it is sometimes called the Embden-Meyerhof, or EMP pathway. However, this terminology may be misleading, because bacteria can utilize four different pathways to breakdown glucose to pyruvate. Besides the EMP pathway, there are also: the Warburg-Dickens or hexose monophosphate pathway; the Entner-Duodoroff pathway; and the phosphoketolase pathway. While the EMP pathway is anaerobic, this does not restrict it to anaerobes, nor does it restrict anaerobes to only using it. The anaerobic *C. acetobutylicum* uses the EMP pathway, and unless otherwise noted, "glycolysis" will be used in place of "the EMP pathway" for this paper.

Glycolytic Enzymes and Mutants

The first of 10 enzymes involved in the process of breaking down glucose to pyruvate, is hexokinase (ATP:D-hexose 6-phosphotransferase, EC 2.7.1.1). Hexokinase requires ATP to catalyze the phosphorylation of glucose, producing glucose 6-phosphate. Besides ATP, kinases also require a metal cofactor, usually Mg$^{2+}$. By definition, a hexokinase transfers the terminal phosphate of ATP to
Figure 1. A simple schematic of glycolysis.
any -OH group of a six carbon monosaccharide. However, it appears that only the terminal hydroxyl group of the sugar is involved (Crane, 1962). Originally discovered in yeast, prokaryotic hexokinases have a higher substrate specificity than those found in eukaryotic systems (Dixon & Webb, 1964).

Conversion of glucose 6-phosphate to fructose 6-phosphate is accomplished by phosphoglucose isomerase (D-glucose-6-phosphate ketol-isomerase, PGI, EC 5.3.1.9). Phosphoglucose isomerase is
encoded by *pgi* in *Escherichia coli* (*E. coli*). An *E. coli* mutant in *pgi* (DF40), grows about one-third as fast on glucose as its parent strain (K-10) (Table 1). The mutant also shows slower growth on galactose, maltose and lactose. However, growth is near normal on fructose and gluconate. Growth on glucose mainly proceeds via the hexose monophosphate shunt, but there may be some contribution as well by the Entner-Duodoroff pathway (Fraenkel & Levisohn, 1967).

Another kinase, phosphofructokinase (ATP:D-fructose-6-phosphate 1-phosphotransferase, PFK, EC 2.7.1.11) is responsible for the formation of fructose 1,6-bisphosphate. Again, ATP and Mg$^{2+}$ are required for this reaction. This enzyme exists as a homotetramer of 34 kDa subunits in *Bacillus stearothermophilus* (B. *stearothermophilus*) (Evans & Hudson, 1979) and 35 kDa subunits in *Escherichia coli* (*E. coli*) (Evans *et al.*, 1981).

An *E. coli* mutant in *pfk* (AM1) was discovered by Morrissey and Fraenkel (1968) (Table 1). The mutant has 5-10% the normal PFK activity. It grows slower than wild-type on glucose, galactose,
Table 1. Glycolysis mutants and carbon sources that restrict or allow growth.

<table>
<thead>
<tr>
<th>Strain</th>
<th>Impaired Growth On</th>
<th>Normal Growth On</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>DF40</strong> (CGSC #4871) phosphoglucone isomerase</td>
<td>glucose, galactose, maltose, lactose, acetate</td>
<td>fructose, gluconate, glycerol, arabinose, succinate</td>
</tr>
<tr>
<td>dfB10, fhuA22, ompF627, fadL701, relA1, pit-10, spoT1, pgi-2, phoM510</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>DF263</strong> (CGSC #5781) phosphoglycerate kinase</td>
<td>glucose, glycerol, succinate, gluconate</td>
<td>glycerol+succinate, glycerol+malate</td>
</tr>
<tr>
<td>dfB10, fhuA22, ompF627, fadL701, relA1, pgk-2, pit-10, spoT1, phoM510, mcrB1</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>AM1</strong> (CGSC #4881) 6-phosphofructokinase I</td>
<td>glucose, galactose, mannitol, xylose, arabinose</td>
<td>fructose, gluconate, glycerol</td>
</tr>
<tr>
<td>tonA22, ompF627(T2), relA1, pit-10, spoT1, pfkA1</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>JM683</strong> (CGSC #5191) fructose-bisP aldolase</td>
<td>glucose, galactose, mannitol, fructose</td>
<td>glycerol</td>
</tr>
<tr>
<td>dfB10, fhuA22, ompF627 Δ(his-gnd)32, fba-1, relA1, pit-10, spoT1 Δ(aceA-pgi)33, mcrB1, phoM510</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>DR220</strong> (CGSC #5523) glyceraldehyde-3-P DH</td>
<td>glucose, glycerol, malate</td>
<td>glycerate, glycerol+malate</td>
</tr>
<tr>
<td>tonA22, ompF627(T2), gap-1, relA1, pit-10, spoT1</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>DF575</strong> (CGSC #5965) enolase</td>
<td>glucose, glycerol, succinate, gluconate</td>
<td>glycerol+succinate, glycerol+malate</td>
</tr>
<tr>
<td>f-, eno-2, rpsL104, relA1, apoT1, thi-1, pit-10, xyl-7, malA1(lambdak), mtl-2, tonA22, lambda-, supE44</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
maltose and mannitol. It grows normally on fructose, glycerol and other gluconeogenic substances. Likely containing a single gene lesion, strain AM1 frequently, spontaneously reverts back to wild-type levels of PFK activity. (Morrissey & Fraenkel, 1968).

Phosphofructokinase catalyzes the rate-limiting step of glycolysis. In addition, PFK has a large number of regulatory effectors, which make it an attractive candidate for the crucial regulatory point within the pathway. This will be discussed later.

Fructose-1,6-bisphosphate aldolase (fructose-1,6-diphosphate:D-glyceraldehyde-3-phosphate-lyase, FBA, EC 4.1.2.13) is another important enzyme in glycolysis. Acting upon fructose-1,6-bisphosphate, FBA cleaves the hexose into two triose phosphates (dihydroxyacetone phosphate and glyceraldehyde-3-phosphate). Differences again arise between single-celled and multicellular organisms. Richards and Rutter went as far as to develop two classes for aldolases (1961). Investigating the properties of muscle and yeast
aldolase, they found the latter to be inhibited by chelating agents. This inhibition could be overcome with the addition of metal ions. Because of this difference and others, they dubbed muscle aldolase a metal ion-independent (Class I) aldolase, and yeast a metal ion-independent (Class II) aldolase. Purification of a number of other aldolases continued to demonstrate a difference between the two classes (Bard & Gansalus, 1950; Bastarracea et al., 1961; Gary et al., 1955; McDonald & Mallavia, 1971; Smith, 1960).

A number of _E. coli_ mutants in _fba_ have been discovered (Irani & Maitra, 1977; Singer et al., 1991). These mutants grow slower upon glucose, galactose, mannitol and fructose. Growth is normal when compared to wild-type upon glycerol.

If it were not for the action of triosephosphate isomerase (D-glyceraldehyde-3-phosphate ketol-isomerase, TPI, EC 5.3.1.1), there would be a build up of dihydroxyacetone phosphate in the cell. Triosephosphate isomerase catalyzes the conversion between the two products formed by FBA. Due to the action of glycolytic enzymes downstream of TPI, the reaction's equilibrium lies on the side of glyceraldehyde-3-phosphate.

The following step requires both an oxidation and a phosphorylation as glyceraldehyde-3-phosphate dehydrogenase [D-
glyceraldehyde-3-phosphate:NAD oxidoreductase (phosphorylating), GAP, EC 1.2.1.12] produces 1,3-bisphosphoglycerate from the precursor glyceraldehyde-3-phosphate. To carry out this reaction,

\[
\begin{align*}
\text{CHO} & \quad \text{NAD}^+ \quad \text{NADH} + H^+ \\
\text{H} & \quad \text{OH} \\
\text{H}_2\text{COPO}_3^{2-} \quad \text{glyceraldehyde-3-phosphate} & \quad \text{glyceraldehyde-3-phosphate dehydrogenase} & \quad \text{COPO}_3^{2-} \\
\text{H} & \quad \text{OH} \\
\text{H}_2\text{COPO}_3^{2-} \quad 1,3\text{-bisphosphoglycerate}
\end{align*}
\]

GAP (sometimes abbreviated GAPDH) requires both the coenzyme NAD\(^+\) and inorganic phosphate. All known GAPs exist as 145,000 molecular weight tetrameric proteins, with each subunit possessing a site to bind the coenzyme NAD (Biesecker et al., 1977). Despite chemical identity between the subunits (Harris & Perham, 1965), as each NAD binds, it lowers the affinity for the coenzyme. This "negative cooperativity" results in the first two molecules of NAD being bound tightly, while the next two are bound only loosely (Conway & Koshland, 1968; Seydoux et al, 1973). In fact, the stable form of the enzyme appears to be an \(E-(\text{NAD})_3\) complex. The lowest affinity NAD binding site is the most reactive in the dehydrogenase and transferase catalytic activity (de Vijlinder et al., 1969). There are also four thiol (SH) groups, with one SH being very reactive and participating in the catalysis (Krimsky & Racker, 1955). The NAD's importance is not its normal redox activity. It actually lowers the pK of the SH group as it binds to the enzyme; thus greatly increasing the thiol group's reactivity (Trentham, 1968).
Hillman and Fraenkel successfully isolated an *E. coli* mutant (DF220) that expresses less than 5% of parental levels of GAP activity *in vitro* (1975) (Table 1). This mutant and others exhibit restricted growth upon glucose, glycerol, gluconate, glucuronate, fructose, galactose, ribose, glycerate, lactate, acetate, succinate, and malate. However, providing the mutants with a carbon source, such as glycerol, and an oxidative compound, such as malate does allow growth.

After the high energy compound 1,3-bisphosphoglycerate is formed, it can now be used to create ATP. This reaction is catalyzed by phosphoglycerate kinase (ATP:3-phospho-D-glycerate-1-phosphotransferase, PGK, EC 2.7.2.3). This is the first kinase in

\[
\text{COO}^\cdot \text{PO}_3^{2-} \quad \text{ADP} \quad \text{ATP} \quad \text{COO}^\cdot \text{PO}_3^{2-} \\
\text{H-} \quad \text{OH} \quad \text{phosphoglycerate kinase} \quad \text{H-} \quad \text{OH} \\
\text{H}_2\text{COPO}_3^{2-} \quad 1,3\text{-bisphosphoglycerate} \quad \text{3-phosphoglycerate}
\]

glycolysis that produces a molecule of ATP from ADP. It requires the cofactor Mg\(^{2+}\). PGK is the only kinase to catalyze a reversible reaction in glycolysis, and is the only glycolytic enzyme to exist as a monomer. Phosphoglycerate kinase is found to be one of the most abundant enzymes in the cytoplasm of yeast and bacteria undergoing fermentative metabolism (Scopes, 1973).

Hillman successfully isolated the *pgk* mutant, DF263 (Thomson *et al.*, 1979) (Table 1). As was the case with GAP mutants, PGK
mutants require a biosynthetic precursor and an oxidizable substrate. Glycerate merits special note, as it allows growth of GAP and PGK mutants, but not enolase mutants.

Phosphoglycerate mutase (2,3-diphospho-D-glycerate:2-phospho-D-glycerate phosphotransferase, PGM, EC 2.7.5.3) catalyzes the conversion of 3-phosphoglycerate to 2-phosphoglycerate. Enzymes from yeast and higher animals require the assistance of 2,3-bisphosphoglycerate in this reaction. However, higher plants' and *Bacillus subtilis'* phosphoglycerate mutases do not require it (Towne et al., 1957). *Bacillus subtilis* PGM is the only one reported to require a metal cofactor (Mn$^{2+}$) (Oh & Freese, 1976; Rose, 1980).

Enolase (2-phospho-D-glyceroate hydro-lyase, ENO, EC 4.2.1.11) catalyzes the formation of a second high energy intermediate. Enolase purified from *E. coli*, along with yeast and higher animals has a molecular weight of approximately 90,000. While there are some unique properties to individual enolases, there are only minor differences in the catalytic site. The enzyme is a homodimer, with each subunit possessing a single active site. Enolase is also dependent on a divalent metal cofactor (Mg$^{2+}$, Mn$^{2+}$ or Zn$^{2+}$) for catalytic activity (Spring & Wold, 1971).
An "intramolecular oxidation-reduction reaction" occurs during this step, resulting in a concentration of electrons around the C2 of phosphoenolpyruvate. The double bond is then formed, and increases the internal tension on the molecule. All of this is responsible for phosphoenolpyruvate being a high-energy compound.

Finally, a second molecule of ATP is produced, as pyruvate kinase (ATP:pyruvate phosphotransferase, PYK, EC 2.7.1.40) transfers the phosphoryl group from phosphoenolpyruvate to ADP. Yet another kinase reaction, PYK requires the assistance of ADP and Mg^{2+}. This is also the third irreversible glycolytic reaction, since normal physiological conditions makes the contrary reaction almost impossible (Anderson & Wood, 1969).
Metabolic Regulation of Glycolysis, Particularly at Phosphofructokinase

The glycolytic pathway possesses a number of regulatory mechanisms. Each glycolytic enzyme is subject to substrate control by saturation kinetics (Michaelis-Menten type behavior). Another basic control is the availability of cofactors, as illustrated in the structurally simple enzyme phosphoglycerate kinase. PGK can be regulated by the availability of Mg$^{2+}$ or ADP. However, these simple regulatory controls are not sufficient to control an entire pathway. More complicated controls are necessary: inhibitors (competitive, non-competitive, uncompetitive or mixed); allosteric effectors (positive and negative); protein or subunit association, disassociation and aggregation; chemical interconversion (phosphorylation); the level of enzyme synthesis and degradation. All these factors may operate alone or in concert with each other, allowing a subtle yet powerful control of the flux throughout the entire pathway. More complex controls of individual enzymes like phosphofructokinase or pyruvate kinase are available. For example, both exhibit allosteric control by intermediates in the pathway. This more complex mechanism of regulation allows control of the entire glycolytic pathway, as well as some control over other pathways (Figure 2) (Boiteux & Hess, 1981).

Unlike eukaryotic PFKs, bacterial enzymes are neither affected by the allosteric activator AMP, nor the allosteric inhibitors ATP and citrate. Accordingly, their regulation is much simpler than their eukaryotic brethren; however, there is still enough regulation to allow
Figure 2. Regulation of glycolysis through the allosteric effectors of phosphofructokinase (PFK) and pyruvate kinase (PK). Arrows represent positive allosteric effectors and bars represent negative allosteric effectors.
metabolic control of the pathway. The allosteric effectors are ADP activating PFK and phosphoenolpyruvate (PEP) inhibiting PFK. An abundance of ADP in a cell signals a shortage of energy (i.e. ATP) and a need for higher glycolytic levels. An abundance of PEP indicates that glycolysis has been slowed downstream of PFK (likely at PYK), and PFK activity is no longer necessary.

This allosteric mechanism found in *B. stearothermophilus* (Evans & Hudson, 1979), was first characterized with PFK from *E. coli* (Blangy *et al.*, 1968). An active R-state and a less-active T-state exist, dependent upon the allosteric control. The difference in activity is the result of differing affinities for F6P, as the *kcat* for each state is constant. This dual-state system is found in many PFKs, so it may be assumed that this is also the case in *C. acetobutylicum*.

To further explain this system, it helps to understand the structural characteristics of PFK. Both the substrate cooperativity and allosteric control are mediated by the ligand bridges found between subunits. Evans and Hudson resolved the structure of *B. stearothermophilus* PFK to 2.4 Å (Figure 3) (Evans & Hudson, 1979). Each subunit has two domains. Domain 1 consists of seven strands of β-sheets pocketed between five α-helices. Domain 2 only has four strands of β-sheets between 5 α-helices. The two sheets of this second domain point towards a deep cleft which forms the active site. Each subunit in the tetramer forms a close contact with only two of the others. Between the two subunits that do not form a close contact is a solvent filled hole approximately 7 Å in diameter.
Figure 3. A Drawing of a PFK subunit. Arrows represent β-sheets (A-K), cylinders represent α-helices (1-13). B Drawing of the tetramer of PFK. Subunits have Domains 1 and 2 labeled. The cylinder represents the 7 Å solvent-filled hole. Some of the A, B and C binding sites are labeled. (This figure modified from Evans & Hudson, 1979).
Each subunit has three binding sites designated A, B and C. Sites A and B form the catalytic zone. F6P binds to site A, and fructose-1,6-bisphosphate (FBP) should as well, but only weak binding of FBP has been reported (Evans & Hudson, 1979). ADP binds to site B, along with the metal cofactor (Mg$^{2+}$ or Mn$^{2+}$). Site C binds the allosteric effectors ADP and phosphoenolpyruvate. The presence of ligand bridges becomes clear upon closer inspection.

The 6-phosphoryl group of F6P is bound by histidine (His) 249 of one subunit, and arginine (Arg) 162 plus Arg 243 from the neighboring subunit (the residue numbering has changed since Evans & Hudson determined the structure in 1979; the numbering used here is based on French & Chang, 1987, GenEmbl accession number M15643). The phosphoryl group binding at site A occurs across the z-dyad axis. The groups binding the sugar ring of F6P [Arg 252, methionine (Met) 169, glutamic acid (Glu) 222 and aspartic acid (Asp) 127] all reside in domain 2. The 1-OH of F6P is in close proximity to the γ-P of the ATP bound to site B. This hydroxyl group may hydrogen-bond with Asp 127, which would increase its nucleophilicity and enable it to act as a base catalyst. The binding of adenosine diphosphate (ADP) and Mg$^{2+}$ at site B is somewhat uncertain. However, it appears that Arg 171 is associated with the nucleotide and Asp 103 with the metal cofactor. The allosteric effectors are bound across the cleft along the x-dyad axis. With ADP, the terminal phosphate group is again bound by a single group (Arg 154) from one subunit and two from a neighboring subunit (Arg 21 and Arg 25). As the phosphate group of PEP binds in the same manner, site C would best be considered a phosphate, rather than a nucleotide binding site.
Here is structural evidence for the allosteric control of PFK. This evidence is also collaborated by structural studies done with PFK from *Escherichia coli* (Shirakihara & Evans, 1985; Rypneiwski & Evans, 1989) and limited mutational studies done in both organisms (Lau *et al.*, 1987; Valdez *et al.*, 1989). This clearly demonstrates the action of glycolytic intermediates functioning as a feedback mechanism. Sequence similarity, particularly identity with the aforementioned residues, would indicate that the same mechanism may exist in the PFK of *C. acetobutylicum*. Could there be other levels of regulation and what might they be? The next section explains a situation where glycolysis (Entner-Duodoroff) is regulated at multiple levels, including coenzyme limitation and the level of enzyme expression.

**Glycolytic Flux in *Zymomonas mobilis* During Fermentation**

*Zymomonas mobilis* is an obligately fermentative organism producing ethanol and CO₂ from the breakdown of glucose (Entner-Duodoroff pathway). While differing from the Embden-Meyerhof-Parnas (EMP) pathway, the two share similar enzymes from glyceraldehyde-3-phosphate dehydrogenase to pyruvate kinase (Mountenecourt, 1985). *Zymomonas mobilis* sugar conversion rates have been reported to be several times higher than those of *Saccharomyces cerevisiae* (Rogers *et al.*, 1982). The research done on glycolytic flux in *Z. mobilis* during batch fermentation reveals that efficient glucose breakdown is necessary to produce high levels of ethanol. Like *Z. mobilis*, *C. acetobutylicum* produces an alcohol
(butanol) and other solvents at high levels. Therefore, *Z. mobilis* could serve as a model for *C. acetobutylicum* with regard to the need for understanding glycolytic flux to maximize fermentative efficiency.

As the concentration of ethanol increases, *Z. mobilis* loses efficiency in its conversion of glucose to ethanol and CO₂ (Millar et al., 1982), a phenomenon also seen in yeast (Moulin et al., 1984). This occurs in two stages during *Z. mobilis* batch fermentation. There is an initial decline after glycolytic flux reaches its peak at the 18 hour mark. At 30 hours, there is a sharp decline in ethanol production (Osman et al., 1987). The exact nature of the mechanism behind this ethanol inhibition is still unclear, although there have been many suggestions (Casey et al., 1984; Hopper & Doelle, 1983; Osman & Ingram, 1985).

Osman et al. shed some light on this matter with a thorough study of glycolytic flux during batch fermentation (1987). In this study, they determined that the initial decline was caused by: a lowering in concentration of the coenzymes NAD⁺ and ADP; a reduction in intracellular pH; and a reduction in the specific activity of two glycolytic enzymes. The decrease in specific activity of glyceraldehyde-3-phosphate dehydrogenase (GAP) and phosphoglycerate kinase (PGK) was greater than would be expected due to the decrease in the intracellular pH and the limited supply of coenzymes. Another type of regulation must be acting upon the two enzymes. The authors presumed that inhibition was also occurring at the level of enzyme synthesis. Since GAP and PGK were serving to regulate ED glycolytic flux in *Z. mobilis*, they also predicted that it would be convenient if gap and pgk were to share an operon. This
would allow very efficient control of both enzymes, and therefore control of the pathway, at the level of the two enzymes' synthesis.

Conway and Ingram determined that pgk did indeed share an operon with gap (1988). The operon is under the control of a group of promoters that provides very high expression. Because PGK has a catalytic rate 4 times faster than GAP, GAP has 4-8 times higher expression than PGK (Eddy et al., 1989). This required appropriately proportioned, individual expression levels for the two genes, despite sharing a single operon. Eddy et al. found that gap transcriptional messages are present in a much higher number than the bicistronic messages (1991). They determined that the 3' pgk segment is processed for inactivation, leaving the more stable 5' gap segment. Mejia et al. extended the study of message stability to 13 genes encoding glycolytic and fermentative enzymes (1992). They found that all of these genes had significantly more stable mRNA than the mRNA from biosynthetic genes, whose levels do not need to be as high in the cell. This provides Z. mobilis a greater number of certain enzymes, while maintaining lower levels for others. Message stability has also been used to provide differential expression between operons, as well as within a single operon. Overall, message stability is one of the subtle, yet powerful methods organisms use to regulate enzymes.

The genes encoding the two proteins (GAP and PGK) most responsible for ED glycolytic control are found in a single operon in Z. mobilis. This arrangement is sensible from a metabolic viewpoint, and may be the case in organisms using the EMP glycolytic pathway. Do PFK and PYK could share an operon in organisms using the EMP
pathway? Could this be the case in \textit{C. acetobutylicum}? If so, could there be a similar increased stability in PFK, because PYK has a catalytic rate approximately 3 times as fast? Are \textit{C. acetobutylicum} glycolytic genes' mRNA more stable than biosynthetic enzymes and other enzymes that are not needed at high levels?

**Objectives**

A great deal of research and writing has been dedicated to the acid and solvent production of \textit{Clostridia}. However, very little is known about \textit{C. acetobutylicum} glycolysis. It has already been shown how glycolytic flux plays a large role in yeast and \textit{Z. mobilis} fermentation. Understanding the role \textit{C. acetobutylicum}'s glycolytic flux plays in its solventogenic stage may be equally important.

To begin exploration of glycolysis in \textit{C. acetobutylicum}, we set out to clone genes encoding enzymes involved in the pathway. This would allow further studies to be carried out on the regulation and metabolic flux in the pathway. Complementation studies were carried out with a \textit{C. acetobutylicum} ATCC 824 plasmid library and \textit{Escherichia coli} glycolytic mutants. A PCR screen was done with degenerate primers designed from homologous regions in the \textit{gap}, \textit{pgk} and \textit{pfk} genes of related organisms. Putative \textit{pfk}, \textit{pyk} and \textit{gap} genes have been cloned and the entire sequence of \textit{pfk} has been sequenced, along with the 5' end of \textit{pyk}. As was the case in \textit{Z. mobilis}, it appears that the potential regulators of the glycolytic pathway are sharing a single operon. This thesis will serve as a basis
for further work with the glycolytic pathway of *C. acetobutylicum*, as many questions remain unanswered.
Materials & Methods

Bacterial Strains and Plasmids

A list of the glycolytic mutants can be found in Table 1. *E. coli* DH5α [ΔlacU169 (φ80 lacZΔM15), recA1, relA1] was also used as a host for preparation of plasmids (Hanahan, 1983). Epicurian Coli® XL1-Blue MRF' [Δ(mcrA)183, Δ(mcrCB-hsdSMR-mrr)173, endA1, supE44, thi-1, recA1, gyrA96, relA1, lac [F' proAB, lacIqZΔM15, Tn10 (tetR)]] (Stratagene) was used as a host for electroporation of cosmids. The plasmids used in this study are listed in Table 2.

Antibiotics, Chemicals and Enzymes

Sigma Chemical Company supplied both ampicillin (Ap) and chloramphenicol (Cm). These antibiotics were necessary for plasmid selection and maintenance. Ap was added to a final concentration of 50 μg/ml and Cm to 32 μg/ml. Sigma also supplied the majority of chemicals (any exceptions will be noted), and the glycolytic enzymes: α-glycerophosphate dehydrogenase-triose phosphate isomerase (Type III: from rabbit muscle); aldolase (Type IV: from rabbit muscle); and fructose-6-phosphate kinase (Type III: from rabbit muscle). Endonucleases and T4 DNA ligase were supplied by the Promega Corporation.
Table 2. Plasmids and cosmids used in this study

<table>
<thead>
<tr>
<th>Plasmid</th>
<th>Size (in kb)</th>
<th>Properties</th>
<th>Vector Origins</th>
</tr>
</thead>
<tbody>
<tr>
<td>cosmid92&lt;sup&gt;a&lt;/sup&gt;</td>
<td>~35</td>
<td>Ap&lt;sup&gt;r&lt;/sup&gt;, Neo&lt;sup&gt;r&lt;/sup&gt;, C. acetobutylicum library, gap</td>
<td>pWE15</td>
</tr>
<tr>
<td>pAM1&lt;sub&gt;e&lt;/sub&gt;</td>
<td>10</td>
<td>Ap&lt;sup&gt;r&lt;/sup&gt;, C. acetobutylicum library, pfk</td>
<td>pBR322</td>
</tr>
<tr>
<td>pBCpfk&lt;sup&gt;c&lt;/sup&gt;</td>
<td>5.4</td>
<td>Cm&lt;sup&gt;r&lt;/sup&gt;, colE1 ori, 2 kb EcoRV fragment from pAM1&lt;sub&gt;e&lt;/sub&gt;</td>
<td>pBC SK+</td>
</tr>
<tr>
<td>pBC SK+&lt;sup&gt;d&lt;/sup&gt;</td>
<td>3.4</td>
<td>Cm&lt;sup&gt;r&lt;/sup&gt;, colE1 ori, lacZ, F1 ori</td>
<td>pUC19, pBluescript</td>
</tr>
<tr>
<td>pBR322&lt;sup&gt;e&lt;/sup&gt;</td>
<td>4.4</td>
<td>Tet&lt;sup&gt;r&lt;/sup&gt;, Ap&lt;sup&gt;r&lt;/sup&gt;, colE1 ori</td>
<td></td>
</tr>
<tr>
<td>pWE15&lt;sup&gt;f&lt;/sup&gt;</td>
<td>8.2</td>
<td>Ap&lt;sup&gt;r&lt;/sup&gt;, Neo&lt;sup&gt;r&lt;/sup&gt;</td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup>This cosmid library was prepared by Petersen & Cary in this lab.  
<sup>b</sup>This plasmid library was prepared by Petersen & Cary in this lab.  
<sup>c</sup>This study.  
<sup>d</sup>Stratagene Cloning Systems.  
<sup>e</sup>Watson, 1988.  
<sup>f</sup>Wahl <i>et al.</i>, 1987.
Maintenance of Strains

DH5α was grown in LB (Luria-Bertani: 10 g NaCl, 10 g/L Bacto®-tryptone, 5 g/L Bacto® yeast extract, 30 mg/L L-cysteine). LB plates required 15 g/L of agar (Difco Laboratories). The glycolytic mutants were grown in Glycolysis Minimal Medium (GMM was made with:
13.6 g/L KH₂PO₄, 2 g (NH₄)₂SO₄, 0.2 g/L MgSO₄·7H₂O, 0.56 mg/L FeSO₄·7H₂O; KOH was added to pH 7, the solution was then autoclaved; 1 mg/L thiamine HCl and 1 g/L Casamino acids were added after autoclaving: Fraenkel & Levisohn, 1967). Agar was added to GMM at 20 g/L. Epicurean Coli® were grown in SOC (20 g/L Bacto®-tryptone, 5 g/L Bacto® yeast extract 0.58 g/L NaCl, 93 mg/L KCl, KOH to pH 6.9, autoclave, 20 ml/L 1 M MgCl₂·6H₂O, 20 ml/L 1 M MgSO₄, 10 ml/L 2 M glucose. All strains were grown at 37° C. Cells were placed in 50% glycerol for storage at -20° C.

Bacterial Transformation

DH5α was initially transformed with the Hanahan protocol (1983), and later with a RbCl (Raleigh, 1994) protocol. The glycolysis mutants were transformed with Chung & Miller's (1988) protocol. It is important to note that each strain's appropriate carbon source was added to the Transfer and Storage Buffer (TSB), rather than the glucose supplement of Chung & Miller (1988).

Epicurian Coli® were transformed with cosmid 92 in a Gene Pulser (Bio-Rad) electroporator. The cells were electroporated in a chilled 0.1 cm electroporation cuvette. The Gene Pulser apparatus
was set at 25 μF and 1.7 kV. The Pulse Controller was set at 200 Ω. A pulse was applied, producing a time constant of 4-5 msec. Immediately after the pulse, 0.9 ml of SOC was added. Cells were plated after shaking for 1 hour at 37° C.

*C. acetobutylicum* Plasmid and Cosmid Libraries

A *C. acetobutylicum* genomic library in pBR322 had been previously prepared in this laboratory by Daniel Petersen and Jeff Cary. They ligated an incomplete *Sau3AI* digest of *C. acetobutylicum* genomic DNA to *BamHI* digested pBR322. The *Sau3AI* fragments were approximately 5-15 kb. Petersen and Cary also prepared the cosmid library. Again, genomic DNA (ATCC 824) was partially digested with *Sau3AI*. DNA fragments between 30 and 45 kb were ligated into *BamHI* digested pWE15.

Plasmid Preparations

Promega's Wizard™ midi and mini DNA purification systems were used for plasmid preparations; particularly those used for sequencing. Qiagen's Midi Kit (100) was also used, but with less success in automated sequencing. For small-scale preparations, an alkaline lysis protocol was also used. 1.5 ml of an overnight culture was spun down in a microcentrifuge tube for 2 minutes. The pellet was resuspended in 100 μl of 10 mM EDTA. 200 μl of an alkaline solution (0.15 N NaOH, 1% SDS) was added, and then placed on ice for 5 minutes. 150 μl of freshly prepared 7.5 M ammonium acetate was
added next, and the tubes were incubated on ice for another 10 minutes to allow precipitation of cell debris. Spinning down the tube for 5 minutes at maximum speed in a microcentrifuge allowed the supernatant to be removed without contaminants (a small amount of supernatant was left behind to allow a cleaner sample). Isopropanol (0.6 volume) was then added and mixed, leaving the sample at room temperature for 10 minutes to precipitate the DNA. The DNA was pelleted in a microcentrifuge for 15 minutes at maximum speed. The supernatant was then removed, and the pellet dried in a Savant Speed Vac Concentrator. Finally, the pellet was resuspended in 40 µl of dH₂O and 250 µg/ml RNase A.

Oligonucleotides and their Design

PCR primers were designed to bind domains conserved amongst several bacterial species. Figure 4 displays these domains in PFK, PGK, and GAP. Bias was given towards those strains more closely related to *C. acetobutylicum*, and for *C. acetobutylicum*'s codon bias (Walters, 1993).

Most of the sequencing and PCR primers were purchased from Genosys Biotechnologies, Inc. Genosys provides the oligonucleotide deprotected, desalted, and dried. Some oligos were synthesized on a Biosearch 8600 DNA synthesizer by E. Singleton in the Department of Biochemistry and Cell Biology at Rice University. These oligos were placed in a 55° C H₂O bath for five hours; then uncapped and left overnight in a constant-flow hood. After deprotection, the samples
Figure 4. Homologous regions used for glycolysis primers' design.

Phosphofructokinase: B. stearo is Bacillus stearothermophilus. L. del is Lactobacillus delbrueckii. B. mac is Bacillus macquarensis. E. coli is Escherichia coli.

Phosphoglycerate kinase: B. meg is Bacillus megaterium. T. mar is Thermotoga maritima. X. flavus is Xanthobacter flavus. E. coli is Escherichia coli. C. glut is Corynebacterium glutanicum.

Glyceraldehyde-3-phosphate dehydrogenase: A. var is Anabaena variabilis. E. coli is Escherichia coli. C. glut is Corynebacterium glutanicum. B. meg is Bacillus megaterium. C. past is Clostridia pasteurianum.

Arrows represent primer orientation and origin.
Phosphofructokinase

1
B. stearo ..MKRIGVLT SQCSSAPGMNA AIRSVHRKAI YHGVEVGHY HGYAGLIAGN
L. del ..MKRIGVLT SQCSSAPGMNA AVRAVTVRAI ANGVEFGIR YGFAVIGACD
B. mac MKTRIVAVLT SQCSSAPGMNA AVRAVRSGL FYGVEVGYIQ RGYQGILIND
E. coli ..MKRIGVLT SQCSSAPGMNA AIRGVRVSAL TEQLEVWGIY DCYLGFLYER

50
PKI

150
B. stearo VIGDGSYVG AKLITEHGFPI CLSVEGIDN DCKIDFTIG FDATHLTVID
L. del VIGDGSYVG ALQTJRHGN SKLCHIDN DCKIDDPAI VDPTDATTG VPTAATMD
B. mac VIGDGSYVG ANKLSKGIN TMLGTVIDN DCKSYTFTIG FDTSVSTVID
E. coli VIGDGSYVG AMRLEHGFPI CQILCHIDN DCKIDTDIG FTPASLSTWE

Phosphoglycerate kinase

290
PKI

340
B. meg DKKGLDAPK TEREIAYVGI NEKLVISWNAP MOVFFELDAFA NGKAVAVAEK
T. mar GAMSGLDICTIE TELPKKQKLAR ACTWVGNAP MOVPEIDDAFA EGTVQVALAIR
X. flavus DEHHSLGAPK QTVITKQKLAR VQKTVVNPAP FGAPENPDFAA AKATVAVY
E. coli DEQIIDGAPA SQGELAEILK RKTTLWNAP VGVEFPPFAF KGETETVIANAI
C. glut GAMSGLDICLP SVKNFGVLSR TKITFWNPAP MGVPFSAAFS EGTRASPRSS

390
PGKII

Glyceraldehyde-3-phosphate dehydrogenase

1
GAPI

50
A. var VLAKLVGIP GFRGKVLVL RA..GINNF NTEFGVINNL VP.FDNLAY
E. coli ..MTIKVGN GFRGKTVFR RA..AQKRS DIETVAINDL LD.ADHAY
C. glut ..MTIKVGN GFRGKTVNFRAV..ER.NG DLEWAVINND TD.NKTGSLT
B. meg ..MAVKGST GPFRGKTVFR RAAL..KND NVVEWAVIND TD.ANMLHA
C. past ..MIKVAV AGFRGKTVFR PRIL.E..VP GLEWAVINNL TD.AKMLHA

151
GAPII

200
A. var SKDVISNAS CTINCIAPFA KVINDIPGGLT BGLMTVHMAH TATQPIVDP
E. coli ..QDVSNAS CTINCIAPFA KVINDFEGII BGLMTVHMAH TATQKIDDCP
C. glut ENHNVISGAS CTINCIAPMA KVINDKGEGIE NGLMTVHAY TCQRLHDP
B. meg ANHVISNAS CTINCIAPFA KVINDKPSLKL ENMMTVHSTY TNQQIQLDP
C. past T.EIVISGAS CTINCIAPMA KVINDKPSLKL KGMFTTHAY TNQQIQLDP
were dried and resuspended in 200 μl dH₂O to determine the OD₂₆₀. They were then diluted to a concentration of ~2.5 mg/ml. Purification columns were prepared by adding sterile glass wool to a 1 ml syringe and then adding Sephadex G25-150 to 0.8 ml (after spinning off excess TE). The columns were washed with 3x100 μl dH₂O, and samples spun through. Finally, the oligos are dried and resuspended to the desired concentration in distilled H₂O (dH₂O). A list of the primers, their sequences and their sources is available in Table 3.

PCR

Perkin-Elmer Cetus’ GeneAmp PCR Core Reagent Kit provided deoxynucleotides triphosphates (dNTPs), buffer, MgCl₂ and AmpliTaq DNA polymerase. The dNTPs’ concentrations needed to be adjusted from the manufacturer’s suggestion due to C. acetobutylicum’s low G/C content (Cato et al., 1986). Both dATP and dTTP were added to 250 μM, with dGTP and dCTP at 150 μM. A PCR Optimizer™ Kit (Invitrogen) was used when different pH or MgCl₂ concentrations were needed.

The PCR was carried out in a GeneAmp PCR System 9600 (Perkin Elmer Cetus). With degenerate primers, Touchdown PCR (Roux, 1994) proved a very effective protocol, resulting in a reproducible, clean PCR product. Touchdown PCR begins with melting the DNA at 94° C for 1 minute. This is followed by consecutive cycles of decreasing annealing temperatures. That is, 94° C and 72° C are used for melting and extension respectively; but the annealing
Table 3. List of primers used in this study.

<table>
<thead>
<tr>
<th>Name</th>
<th>Gene/Sequence&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>phosphoglycerate kinase</strong></td>
<td></td>
</tr>
<tr>
<td>PGKI</td>
<td>GCWAAARACWRTWDHTGGGAAYGGWCC</td>
<td>Genosys</td>
</tr>
<tr>
<td>PGKII</td>
<td>ARRAWAYTCWBARWBWGCWCWCCWCCWCC</td>
<td>Genosys</td>
</tr>
<tr>
<td></td>
<td><strong>phosphofructokinase</strong></td>
<td></td>
</tr>
<tr>
<td>PFKI</td>
<td>ATWGSSWGTTYTWACWWSWGGWGWG</td>
<td>Genosys</td>
</tr>
<tr>
<td>PFKII</td>
<td>TCRTRTCWWATWGWCCWGGWARWCC</td>
<td>Genosys</td>
</tr>
<tr>
<td>PFKIII</td>
<td>CAGATACACCATAAGGGTTTGTAC</td>
<td>Rice</td>
</tr>
<tr>
<td>PFKIV</td>
<td>TGGTTCTTTTTATGGGAGCAC</td>
<td>Genosys</td>
</tr>
<tr>
<td>PFKV</td>
<td>GGGGWAGAGAGGTGGTGGAGGWGC</td>
<td>Genosys</td>
</tr>
<tr>
<td>PFKVI</td>
<td>GCTCTGTTTTCTGGAAATAAGCTGG</td>
<td>Genosys</td>
</tr>
<tr>
<td>pBST3b</td>
<td>GCCCTCCCATATTGATTCTTCTC</td>
<td>Genosys</td>
</tr>
<tr>
<td></td>
<td><strong>glyceraldehyde-3-P-dehydrogenase</strong></td>
<td></td>
</tr>
<tr>
<td>GAPI</td>
<td>ATWAAAYGGWTTYGGGAGWATWGG</td>
<td>Genosys</td>
</tr>
<tr>
<td>GAPII</td>
<td>GWGCWARRCARAAGWGATWGGTRC</td>
<td>Genosys</td>
</tr>
<tr>
<td>GAPIII</td>
<td>AAGCTCACGTAAGAGCAGG</td>
<td>Genosys</td>
</tr>
</tbody>
</table>

<sup>a</sup> Key to Symbols: R=A+G, S=G+C, W=A+T, Y=C+T, D=G+A+T, H=A+T+C
temperature may be dropped from 55, to 54, to 53 on to 45°C over 10 cycles. Finally, after dropping through the cycles with varying annealing temperatures, an additional 20 cycles were performed at the median annealing temperature (50°C for the above example).

**Agarose Gel and Polyacrylamide Gel Electrophoresis**

The agarose used in gel electrophoresis was purchased from Amresco®. Gels were normally 0.8% agarose and ran in chilled Tris-borate/EDTA (TBE) electrophoresis buffer. After the run, gels were stained in an ethidium bromide solution and the DNA illuminated with UV. Polyacrylamide gel electrophoresis was conducted according to the protocol of Maniatis et al. (1982).

**Recovery of DNA Fragments from Low Melting Point Gels**

If the size of the DNA fragment to be isolated was less than 2 kb, the DNA was simply separated by agarose gel electrophoresis, stained and the desired band excised with a razor blade. Long wavelength UV was used to visualize the band to lessen the damage to the DNA. The gel was then placed in a 0.5 ml microcentrifuge tube, upon a bundle of Sigmacote (Sigma) treated glass wool covering a hole in its bottom. This in turn, was placed in a 1.5 ml microcentrifuge tube and spun down at 1300 x g for 4 minutes. The TBE and DNA are collected in the 1.5 ml microcentrifuge tube, to which is added a 1/10 volume of 3 M NaOAc. Two volumes of cold, 95% ethanol was also added. This was then placed at -70°C for half an hour. The DNA was
pelleted in a refrigerated (4° C) microcentrifuge at full speed for 15 minutes. The supernatant was then decanted, and the pellet dried in a Savant Speed Vac Concentrator. Finally, the pellet was resuspended in dH₂O.

For larger sized DNA fragments, Epicentre Technologies' GELase™ was used. The DNA was separated on a 1% low-melting point (LMP) NuSieve® agarose (FMC Bioproducts) gel. EtBr staining and long wavelength UV allowed the desired band to be visualized and excised. The gel slice was weighed, and soaked in 3 volumes of 1X Gelase Buffer. After removing any excess buffer, the gel was completely melted at 70° C for at least 10 minutes. Next, the molten agarose was equilibrated to 45° C. Addition of GELase followed (1 U/600 mg of 1% LMP). The reaction was incubated at 45° C for 1 hour. Finally, the DNA was precipitated using ethanol, as described above.

DNA Ligation

T4 DNA ligase and its buffer were available through Promega. Ligations were performed overnight in a 10 µl reaction mixture at 12-16° C. The molar concentration of the insert was ten fold that of the vector. The ligation reaction was directly transformed into DH5α. 5-bromo-4-chloro-3-indolyl-β-D-thiogalactosidase (X-gal) was added to the plates, before plating out the cells, if the cloning vector being used allowed for blue/white screening. IPTG was not needed, as DH5α is lacI−.
DNA Sequencing

All sequencing was done at the Molecular Genetics Core Facility of UT-Houston's Medical School, on a Taq DyeDeoxy Terminator Cycle Sequencer. Primers may be designed in the same manner as PCR primers. MacVector Ver. 4.1 (International Biotechnologies, Inc.) designed primers when sequence data were available. Each sequencing reaction requires 3 µl of primer at a concentration of 1 pmole/µl. Template DNA can be single-stranded (ss), double-stranded (ds) or PCR products. Larger templates, such as cosmids and lambda clones are more difficult. Even large, low copy-number plasmids may be troublesome (as was the case with pAM1e). Promega's Wizard Midi and Mini-purification kits for plasmids seemed most successful. Each sequencing reaction requires 5 µl of ds template at a concentration of 200 ng/µl.

For sequencing of PCR products, the sample needs to be purified. Gel electrophoresis should be used when there are multiple, i.e. false products. However, if there is only a single product, ultrafiltration with a Microcon™ 100 microconcentrator is adequate. After purification, 25-50 ng per 100 bp are necessary for products up to 1 kb. Beyond 1 kb, the sample should be handled the same as ds samples.

Computer Resources

Searching GenBank and GenEmbl, aligning sequences, mapping and determining the presence of ORFs (open reading frames) were all
accomplished with the aid of GCG Package, Unix Version 7.3.1 and Version 8 (Genetics Computer Group, 1994). MacVector 4.1.1 (International Biotechnologies, Inc.) was used as previously mentioned, to design sequencing primers, and it was also used to locate ORFs. Database searches were also performed with the aid of the BLAST program, which scans all non-redundant sequences in the databases (Altschul, 1990). Transmembrane domain predictions were performed with the TopPredII program (Claros & Heijne, 1994).
Results

*Escherichia coli* Glycolysis Mutants

We received six strains of *E. coli*, each having a mutation in one of the genes encoding the glycolytic enzymes. Table 1 highlights the characteristic mutation of each strain. All of these strains were highly susceptible to reversion. Reversion was particularly likely to occur if the strains were grown in the presence of glucose, because it selected for the revertants. For that reason, these strains were maintained on alternative carbon sources that allowed growth.

JM683 grew quickly on growth-limiting media (glucose, galactose and fructose) the first time it was plated. It had likely arrived as a revertant. Two other strains were difficult to make competent using Chung & Miller's transformation protocol (1988). Competent DF575 cells would grow slowly upon plating, and eventually develop a solid lawn of growth. DF220 grew properly, but had such low transformation efficiency that it was impractical to screen the library with it. However, AM1, DF263 and DF40 were all capable of good transformation and used for the library screening.

Restrictive Media Screening

The restrictive media only slowed the growth of colonies, rather than entirely inhibiting it. Therefore, the screen was actually a comparison of colony size. Further complications arose, as it was discovered that different restrictive media produced a variety of
colony sizes. Testing each carbon source showed that arabinose seemed most stringent for AM1, acetate for DF40 and glycerol for DF263.

Over 13,000 AM1, 8000 DF40 and 7000 DF263 colonies were screened on restrictive media. Larger colonies were restreaked against a pBR322 control to double check the phenotype. Plasmid DNA was then isolated and used to retransform the original host strain and the other two mutant strains. Retransforming the plasmid verified that the phenotype was not just the result of an artifact or a reversion. Transforming the other strains was done in an attempt to find a plasmid that enhances the growth of multiple mutations. These could possibly be plasmids containing genes for a number C. acetobutylicum glycolytic enzymes. Ultimately, 44 plasmids were selected with AM1, 21 with DF263 and 10 with DF40 for PCR screening.

Polymerase Chain Reaction

Primers were designed as described in Materials & Methods, and are illustrated in Table 3. PCR reactions were initially conducted with the degenerate primers and genomic C. acetobutylicum ATCC 824 DNA as template. All three primer pairs gave the appropriately sized product (PKF=375 bp, PGK=207 bp, GAP=465 bp). Subsequent runs always included genomic positive controls.

Each of the plasmids selected from the earlier media screening were used as templates. Buffers A, B, C, D, F, J and N from Invitrogen's Optimization Kit were used with each plasmid. The
different buffers increase the possibility of having the proper environment for each plasmid template. With the genomic controls, all the primer pairs ran best with Buffer D. From the selected plasmids, only one gave a positive result. With the PFK primers, pAM1e produced a single band at approximately 375 bp in length. Figure 5 illustrates the characteristics of pAM1e.

After none of the other plasmids produced any products, another PCR screen was done on a C. acetobutylicum cosmid library. The seven buffers from Invitrogen were again used. Cosmid #92 (c92) produced a single band at 465 bp when GAP primers were used as templates. Both PCR products were sequenced and each demonstrated homology to other GAP or PFK enzymes accordingly. After successfully electrotransforming c92 into Epicurean Coli® (Stratagene) electroporation-competent cells, the cosmid was placed in glycerol for storage.

Subcloning pAM1e

The cloning strategy is summarized in Figure 6. Stratagene's pBC SK+ and pAM1e were digested with EcoRV. The samples were centrifuged on Micron microconcentrators (Amicon) to remove the EcoRV, and the DNA fragments were ligated using T4 DNA ligase. The large size (~7 kb) and second origin of replication made the likelihood of the incorrect fragment ligating to pBC SK+ remote. Thus, gel purification of the 2 kb EcoRV fragment from pAM1e was not deemed necessary. The ligation reaction was used to directly transform RbCl competent DH5α. The cells were plated on LB, supplemented with
Figure 5. Map of pAM1e. Originally selected on restrictive media, this plasmid was part of a pBR322/C. acetobutylicum genomic DNA library. Note the two EcoRV sites that were used in subcloning. Ap=ampicillin resistance. pfk=phosphofructokinase. pyk=putative 5' end of pyruvate kinase.
Figure 6. Cloning strategy to produce pBCpfk by subcloning the EcoRV fragment of pAMI e into EcoRV digested pBC SK+.
chloramphenicol to select for transformants. 50 µl of 5-bromo-4-chloro-3-indolyl-β-D-thiogalactopyranoside (X-gal) (40 µg/µl) was also added to select for inserts with blue/white screening. It was not necessary to add isopropyl-β-D-thiogalactopyranoside (IPTG), because DH5α is lacI⁻. Mini-preps were performed on white colonies and the plasmids were digested with EcoRV. One plasmid was found to contain an insert of ~2 kb, and was named pBCpfk.

**Sequencing pBCpfk**

Figure 7 illustrates the sequencing strategy for the insert of pBCpfk. The sequence of pfk, and the rest of the 1.7 kb insert is provided in Figure 8. The sequence encoding pfk begins with an AUG (methionine) start codon at nucleotide 502 of the insert. It has a TAG stop codon at nucleotide 1461. There is a putative ribosomal binding site (RBS) (Shine & Dalgarno, 1974) upstream of the initiation codon for methionine, which closely resembles the canonical AGGAGG found with other *Clostridium* genes (Papoutsakis & Bennett, 1993; Young *et al.*, 1989).
Figure 7. Sequencing strategy and restriction map of pBCpfk insert. Arrows represent a single automated sequencing run.
Figure 8. Complete sequence of the pBCpik insert. The location and amino acid sequence of the 957 bp pf/C gene and the first 118 bp of a putative pfk are shown. "RBS" and underlined sequence correspond to putative ribosomal binding sites.

```
1  A0CTATTATAAAAATACGAGGAGATATTTGTCTTGTGGAGATATA
   TGGATATAATTATTTTTATGCTCTATATACAGAACTCTTACTAT

51  CCCTCTTTAAACCTCATATGAAAGAATAGCGTTTATAT
   GCACAAATTTTTAAGTAAATTTCTTTTATCCGAAATATA

101  ATGGAGACACCTCCATATAATATGTGCCAGGAGAGATAGAGTT
    TTACCTCTGAGGTITATATATAGCGCTCTCTCTTCTTTAAA

151  TCTTATGTATAAGCCTTTATGCGTCAAAAAGGAGTTGTAGTTAATTAT
    AGAATAACTAATTCCGAAATACCAAACTTTTCCTTGAACTCACATTAATT

201  TTTTAACGGGATAAAAATTTGGCGAAATAAAAATAGAGTATATATAAAA
    AAMATGCCCATTATTTMACCGCTTTTTATACATTTTCACATATTTATTT

251  GTAGGGTTTTTGCTACTAATATATATACAGGGTTAACACATAATTACTAT
    CTCGCAGAAAACAGTATTTATTATATACTTTCTTCTTCTTTGTTATATA

301  GATAAAATATACCTACTATTCTTTCTTTATTTATATAGAAATATAGCATA
    CCGATTGTTTTGGAAAATAATTATTTACCTTTTATACGTAT

351  TTTTTATATAAAAACCATGTATATTTTTACTTTTGATAAATATATAA
    AAATAATATTTGGTACTTATAAAAAATGAAAACCTAAATTTATATA

401  GGAAGAAATTGGGGTTTTGTATATATGTGATATATGTGTTC
    GCTCTTTACAACGACAATTACCATATGATACTTTGAAACAG

451  RBS

500  TATAACAAATCTGGGATATAGTGGCTCCTGGGCTATAGGCAGGTAATA
    ATATTTGTAGACCCGCTTTTTAAGGACCACAGTACCCGACTATT
```
Figure 8. (cont.)

pfk ->

501 TATGAAACAATAGCCTTATAGGATGTTAAGGATGACTACCAGATG
ATCTTTTGTTATGCAAAATGCTACCACCTTACGTTGCTACACT
MKTIALVMGSGDAPGMN

551 ATGCTGCTATAAGACGCGTAGTAAGAAGACTCCCATGAAAAGGTAATTAT
TACGACGATATTTCTCCCATCATTTCTTGACGGTACTTTTTCGCCATATTAA
AAIRAVRTRAIEKGIN

601 GTTAAGGCCATACAAAGGGGTTACGAGTGCCCCTTTAAATGGCGGAAATATT
CAATTCGGTAAATGCTCTCCCATATGTCACCGGAATTATTTACCGGCTTTTTAA
VKIGIOGGYSGLINGEIF

651 TGATATGAAATAGACGAGCCGATTACGATATAATTACAAAGGAGGGCCACTA
ACTATACATTATCTGTCGGCATAGCTATATTATATAGCTGCTCCCTGAT
DMNRHSVDIIDQRGGTI

701 TTTTAAGAACAGCGTTGTCAGAATTCTTAAAGGAAGAAGTTAGCAA
AAAAATCTTGGTGTCAGAGCGTTAAAGAATTCTTTCTCAGATCTGGTT
LRTARCPEFLKKEEVQR

751 AAAGCCAGCAATGTTTATAGGTTTTTTGGTATAGATGCTCTGCTGTAAT
TTCCGGTCTTTCTACAAMAATCCTGAAAACCATATACCTACAGAGCAACATTA
KANVLRVFVGIDGLVVVI

801 AGGTTGAAATGTTGTCGCCGCAANAAGAAGCCTTCTAAACCGCCCGAG
TCCACCTTAAAGAAATATCCCTGIGTTTTCGAAAGATTCGAAACCTC
GGNGSFMRAGAQLSKLGLGV

851 TAAAGACGTTGCTCGGAAACATGATATTGATCTCAGTATATA
ATTCTGCGAAGCGCTTTGGTACCTATATTACTAGAGATTATATTG
KTVGLPGTIDNDLPYMT
Figure 8. (cont.)

1301 TGGGAGTAAGGCTGAGAAGCTTTATTGGAAAGAAAACCTCTAGAGTT 1350
ACCCCTATTGGTACGTAGATATACCTTCTCTTCCTTTTTGAGATCTCA
GVKADEVLMEMGKTSSR

1351 ATAGGTTATTAAGGAAAATAATGTGATCAAGATATTTGATGAGGCTTT 1400
TAATTATAATTTCTCTTCCTTTTTATTAACCTATTTGCTATTGACTTCA
IGIKEGKIMDQDI Deal

1401 AGCAGTTTCCAAGAAGTTTAAATGAAAGTTATGATGATAGCAATATATGC 1450
TCTGCAGCTTCTTAAAATTTATTTCTCAATTATACATACTGTTATTACG
AVPRSFNKELYDIANML

1451 TTTCAAAAATAGCAGTAATTAGTCACAATGATTAATTTCAACACATATATA 1500
AAAGTTTTATCACTGATACATTATCAATGATATATATATT
SK Stop

1501 GATTTTTTTTCAATACAAATTTATATAAAGATGAGGAGAATGATAAC 1550
CTTTTTTTTATTTATATTATTTTCTACCTCTCTCTCTATTGT

ORFpyk ->

1551 ATGCAAAAACTAAAAAGTTTTTACATGGACAGCGAAGATCGAGACGA 1600
TACGTCTTTTATTTCAAAATGTCACCAGTGGCTCTACCTCTCTCTCTCT
MOKTMIKTVGPASETE

1601 GGAGATTTGTAACAGCTTTTGCTAAAAAGCTGGGAATGATGCTAACAGAATA 1650
CCCTACATGCTGAGGATATTTTGACCTTACTTACGAGATCTCTGAT
EIVTAGMNASRHN

1651 ACCTTTTCAACGGTGACTC 1668
TGAAAATGTTGCGGACTAG
FSHGD  End of insert
Discussion

Homology of *Clostridium acetobutylicum* PFK with PFK from other Organisms, Particularly *Bacillus stearothermophilus*

The deduced 319 amino acid sequence of *C. acetobutylicum* PFK displays very high homology with PFKs in other organisms (Figure 9 & Table 4). Not surprisingly, *C. acetobutylicum* PFK has its highest homology with two species of *Bacillus*, a closely related genus. Structural studies have been carried out on the PFK of *Bacillus stearothermophilus* (Evans & Hudson, 1979). Three binding sites and their conserved residues in *B. stearothermophilus* PFK have been elucidated. Site A and B form the catalytic site, binding the substrates fructose-6-phosphate (F6P; Site A) and adenosine triphosphate (ATP; Site B). Site B also binds the cofactor Mg$^{2+}$. Site C binds the allosteric effectors phosphoenolpyruvate (PEP; negative effector) and adenosine diphosphate (ADP; positive effector). Individual residues specifically responsible for binding are conserved throughout most of the species, including the PFK from *C. acetobutylicum*.

Residues involved in the binding of substrates and allosteric effectors are labeled in the alignment (Figure 9) with an "A", "B" or "C" designating their respective site. At site A, the 6-phosphoryl group of F6P is bound by histidine (His) 249, arginine (Arg) 162 and Arg 243. The sugar ring of F6P is bound by the residues; Arg 252, methionine (Met) 169, glutamic acid (Glu) 222 and aspartic acid (Asp) 127.
Figure 9. Alignment of deduced 319 amino acid sequence of Clostridium acetobutylicum PFK with the PFKs of other organisms. Outlined regions show identity between the different species. T. aquaticus is Thermus aquaticus. T. thermophilus is Thermus thermophilus. B. stearo is Bacillus stearothermophilus. L. delbrueckii is Lactobacillus delbrueckii. B. macqueiros is Bacillus macqueiros. C. acetobutylicum is Clostridium acetobutylicum. E. coli is Escherichia coli.

A = residues of binding site A. B = residues of binding site B. C = residues of binding site C.
Table 4. Homology of *C. acetobutylicum* PFK with PFK from other organisms. *B. stearo...* = *Bacillus stearoothermophilus*. *B. macquarensis* = *Bacillus macquarensis*. *E. coli* = *Escherichia coli*. *L. lactis* = *Lactococcus lactis*. *T. thermophilus* = *Thermus thermophilus*. *T. aquaticus* = *Thermus aquaticus*. *L. delbrueckii* = *Lactobacillus delbrueckii*. *S. citri* = *Spiroplasma citri*.

<table>
<thead>
<tr>
<th>Organism</th>
<th>% Similarity</th>
<th>% Identity</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>B. stearo...</em></td>
<td>76</td>
<td>58</td>
</tr>
<tr>
<td><em>B. macquarensis</em></td>
<td>74</td>
<td>56</td>
</tr>
<tr>
<td><em>E. coli</em></td>
<td>73</td>
<td>50</td>
</tr>
<tr>
<td><em>L. lactis</em></td>
<td>71</td>
<td>52</td>
</tr>
<tr>
<td><em>T. thermophilus</em></td>
<td>68</td>
<td>52</td>
</tr>
<tr>
<td><em>T. aquaticus</em></td>
<td>67</td>
<td>52</td>
</tr>
<tr>
<td><em>L. delbrueckii</em></td>
<td>68</td>
<td>47</td>
</tr>
<tr>
<td><em>S. citri</em></td>
<td>64</td>
<td>45</td>
</tr>
</tbody>
</table>
The binding of ADP and Mg\textsuperscript{2+} at site B in the PFK of \textit{B. stearothermophilus} was uncertain. It appeared that Arg 171 is associated with the nucleotide, while Asp 103 may be involved in binding the Mg\textsuperscript{2+}. However, in the PFK of \textit{C. acetobutylicum}, the aspartic acid has been replaced by an asparagine (Asn 103). This is a conserved substitution. The side chains are sterically similar, only Asn lacks the charge.

The key interaction between ADP and the allosteric binding site C, seems to be with the 8-phosphate. Arg 154, Arg 21 and Arg 25 all bind this group. These three arginines also bind the phosphoryl group of phosphoenolpyruvate. His 215 and threonine 158 bind the ribose of ADP. However, \textit{Thermus aquaticus} and \textit{Thermus thermophilus} lack His 215. Instead, they have a serine at this site.

Overall, the homology and conserved residues between \textit{C. acetobutylicum} PFK and \textit{B. stearothermophilus} PFK, suggests that the two share structural, catalytic and regulatory aspects. \textit{Clostridium acetobutylicum} PFK likely exists in a homotetramer possessing three binding sites between the neighboring subunits (see Figure 3). Site A and B would bind the substrates F6P and ADP accordingly. Site C would serve as the binding site for the allosteric effectors. Adenosine diphosphate would act as a positive effector upon \textit{C. acetobutylicum} PFK. This activation by ADP would be a useful regulator, as it would signal an energy-deficient state in the cell. Up regulation of PFK would increase the glycolytic flux, thus increasing the energy made available to the cell. PEP would act as a negative allosteric effector upon \textit{C. acetobutylicum} PFK. This again is an appropriate method of regulating PFK, as an excess of PEP would indicate that enzymes
downstream of PFK (potentially pyruvate kinase) have been downregulated. It would therefore be unnecessary for PFK to maintain a high level of activity.

**ORFpyk**

MacVector was used to locate the presence of other open reading frames (ORFs, Figure 10). The 118 bp ORF 90 bp downstream of pfk was translated using the Genetics Computer Group program, and its amino acid sequence analyzed by the Basic Local Alignment Search Tool (BLAST). The sequence scored many hits with pyruvate kinases (PYKs), and was named ORFpyk.

The 39 amino acids encoded by ORFpyk have 69-82% similarity, and 51-69% identity with the N-termini of other pyruvate kinases (Table 5). The best match at 82% similarity and 69% identity was with the PYK of *Clostridium perfringens*. A putative RBS is located upstream of the initiator methionine of ORFpyk. It appears that ORFpyk may be in the same operon as pfk in *C. acetobutylicum*, because of their proximity (90 bp) and the lack of a transcriptional stop site between the two genes.

The presence of pyk in the same operon as pfk is intriguing. *C. acetobutylicum* may have a "regulatory operon" much like the one in *Zymomonas mobilis*. Glyceraldehyde-3-phosphate dehydrogenase (GAP) and phosphoglycerate kinase (PGK) provide metabolic control in *Z. mobilis* at the level of regulating the Entner-Doudoroff glycolytic flux. Sharing an operon is convenient for the organism on a
Figure 10. MacVector derived open reading frame (ORF) map of the insert in pBCpfk. Rows 1, 2 and 3 read 5' to 3'. Rows 4, 5 and 6 read 3' to 5'. The ORF from 502 to 1461 corresponds to pfk. The incomplete ORF starting at 1551 corresponds to pyk.

Start/Stop Method: AA span >= 25
Genetic Code: universal

Table 5. Homology of the deduced 39 amino acid 5' terminus of C. acetobutylicum PYK with the N-termini of PYK from other organisms. C. perfringens = Clostridium perfringens. B. licheniformis = Bacillus licheniformis. B. psychrophilus = Bacillus psychrophilus. E. coli = Escherichia coli. B. stearo... = Bacillus stearothermophilus L. delbrueckii = Lactobacillus delbrueckii.

<table>
<thead>
<tr>
<th>Organism</th>
<th>% Similarity</th>
<th>% Identity</th>
</tr>
</thead>
<tbody>
<tr>
<td>C. perfringens</td>
<td>82</td>
<td>69</td>
</tr>
<tr>
<td>B. licheniformis</td>
<td>74</td>
<td>59</td>
</tr>
<tr>
<td>B. psychrophilus</td>
<td>74</td>
<td>56</td>
</tr>
<tr>
<td>E. coli</td>
<td>72</td>
<td>54</td>
</tr>
<tr>
<td>B. stearo...</td>
<td>72</td>
<td>52</td>
</tr>
<tr>
<td>L. delbrueckii</td>
<td>69</td>
<td>51</td>
</tr>
</tbody>
</table>
regulatory level, because the crucial enzymes in Z. mobilis's glycolytic pathway may have the same signal acting on both genes simultaneously. This could also be the case for pfk and pyk in C. acetobutylicum. The two enzymes most credited with regulation in the Embden-Meyerhof-Parnas pathway are PFK and PYK. The pfk/pyk operon of C. acetobutylicum may be efficiently regulated, for the same reasons as the gap/pgk operon of Z. mobilis discussed previously.

The lack of complete sequence data and the lack of any knowledge on the glycolytic genes' transcriptional regulation in C. acetobutylicum prevent the formation of any definite conclusions. It is possible that C. acetobutylicum simply has a large operon of glycolytic genes, consisting of more than pfk and pyk. In this case, transcriptional studies would still be necessary. The post-transcriptional processing of gap and pgk mRNA in Z. mobilis demonstrates that sharing an operon does not necessarily entail sharing message stability.

It could also be a coincidence that both pfk and pyk share an operon C. acetobutylicum. In other gram-positive organisms (including the mollicutes) possessing the EMP glycolytic pathway, pfk and pyk share an operon (Lactobacillus bulgaricus, Branny et al., 1993; Bacillus steatorrhophilus, Sakai & Ohta, 1993; Spiroplasma citri, Chevalier et al., 1990). The gram-negative organisms using the EMP pathway do not appear to have a "regulatory operon" (Escherichia coli, Hellinga & Evans, 1985; Thermus thermophilus, Xu et al., 1991). Other times, the operon may contain these two genes with others. This is the case in Lactococcus lactis, which has the pfk,
pyk and ldh genes in a single operon (Llanos et al., 1993). The ldh gene encodes lactate dehydrogenase, which is involved in the production of lactic acid in the EMP pathway. This last case could be an example of a larger "regulatory-operon." The anaerobic organisms like L. bulgaricus and C. acetobutylicum require a more efficient regulation of glycolytic flux than aerobic microbes. It could be this physiological need for efficient glycolytic flux that led to the formation of pfk/pyk "regulatory-operons."

Although message stability has not been determined in Lactococcus lactis, a definite codon bias has been found. This bias has been found in other glycolytic operons as well, where it serves to maximize the expression of these genes (Llanos et al., 1993; Cancilla et al., 1995). Codon bias and higher expression were first found in Escherichia coli (Guoy & Gautier, 1982) and yeast (Sharp et al., 1986). Determination of a codon bias in C. acetobutylicum glycolysis operons remains to be done.

The incomplete ORF upstream of pfk showed some homology to DNA polymerase III of Saccharopolyspora erythraea, Vibrio cholerae, Escherichia coli and Salmonella typhimurium. However, this homology was rather weak and this ORF does not appear to be in the operon with pfk and pyk. None of the other ORFs showed any conclusive homology in searches using BLAST.
References


Spring, T. G. and Wold, F. The purification and characterization of *Escherichia coli* enolase. *J. Biol. Chem.* 246:6797-6802.


pRUNC51

Plasmid pRUNC51 had previously been examined in this lab (Gui, 1995). This plasmid has ~3 kb of *C. acetobutylicum* ATCC 824 genomic DNA inserted into pUC19. A 1179 bp open reading frame (ORFfrubr) encoding a putative rubredoxin oxidoreductase was discovered within the insert. The deduced 393 amino acid sequence of this ORF has a range of similarities from 39%-58% with *Escherichia coli* thioreductase, *Pseudomonas oleovorans* rubredoxin-reductase and *Clostridium pasteurianum* rubredoxin oxidoreductase. *Escherichia coli* and *C. pasteurianum* had the highest homology with the *C. acetobutylicum* rubredoxin oxidoreductase. However, the amino acid sequence encoded by this ORF also has homology with the C-terminus of primosomal protein n' (PriA, replication factor Y). PriA actually has a range of similarities from 42%-53% with the four proteins above. The highest homology is conferred by the rubredoxin oxidoreductase of *C. acetobutylicum*

Gui proposed an evolutionary relationship between the rubredoxin oxidoreductases and PriA; however, this relationship remains unclear. The biological functions of rubredoxin oxidoreductase are still uncertain. Primosomal protein n' is involved in the replication of the bacterial chromosome (Schloimai & Kornberg, 1980). The 76 kDa PriA contains two nucleotide-binding motifs, which are found in most ATP-binding proteins and ATPases (Walker et al., 1982).
I sequenced the rest of the insert and searched for other ORFs with MacVector (Figure 11, Figure 12). Immediately downstream (22 bp) of the deduced termination codon for rubredoxin oxidoreductase is another ORF. This ORF was found to have homology with polypeptide deformylases (def) upon searching with BLAST. Homologies ranged from 51-65% similarity and 22-47% identity. Sixteen nucleotides downstream of the putative def is an incomplete second ORF. The ORF runs to the end of the insert and most likely continues beyond it in the genome. This incomplete ORF showed homology to methionyl-tRNA<sub>Met</sub> formyltransferases (fmt). Homology ranged from 40-62% similarity and 18-45% identity. None of the other ORFs showed any conclusive homology in searches using BLAST.

Polypeptide deformylase plays an essential role in the maturation of eubacterial, chloroplast and mitochondrial proteins (Mazel et al., 1994; Guillen et al., 1992). Deformylase removes the formyl group from the N-terminal methionine post-translationally, allowing aminopeptidase to act upon the peptide (Adams, 1968; Livingston & Leder, 1969; Takeda & Webster, 1968). Methionyl-tRNA transformylase is responsible for transferring the formyl group to the initiator Met-tRNA<sub>Met</sub>. Transformylase activity enhances the rate of initiation in vitro (Eisenstadt & Lengyel, 1966; Kung et al., 1979), but may be dispensable under certain growth conditions or genetic contexts (Guillon et al., 1992). Closely related on a functional level, it is not surprising to find def sharing an operon with fmt in Escherichia coli (Mazel et al., 1994), Thermus thermophilus (Meinnel & Blanquet, 1994) and C. acetobutylicum (there are neither genes coding for
Figure 11. Complete sequence of the pRUNC51 insert. The location and amino acid sequence of the 1179 nt ORFrubr, 450 nt def and the first 570 bp of a putative fmt are shown. "RBS" and underlined sequence correspond to putative ribosomal binding sites.

1 ATCTGGCCCGAAATAGCCTCAAGCAGCTATTTGATAGCTCTGGCGTAC 50
  TAGACCGGCTTTAATTCGAGTTCTGCTGATATAAATATCGACGGCATTG

51 CATCATGCAGATCTGGTGAGACCTGGAGCTCAACCGATCTGTATTTCTACT 100
  GTAGTATGCTCTCAGAACCAGCTGGAGCTCCAGTTGCTAGCACAAAAGATGA

101 GITGATTAAGGCTTTTTTATTTTTTGTTAACCCTAATCTCATTTAACACGATTAATTTTCAACATATTGAGAAMAAAAACAAATGGAATTTTAGGTATATAAT 150

151 AGGAGTTTTTGTTATAGGCAAAATTTTCATTTCTCTTAAAGATTTAAGACGA TCTCCAAAACAATACTCTGTTAAAGTAAAGAGATTTTCATAATCTGCT 200

201 TATCAGAAAGTCCATCCGCTATAGCTACTCTGTTAGGGTTTTTTAAATATTA ATAGTGTTCAGGTAGCGATATCGATGAACGATCCAAAATTATAAT 250

251 ATAGCTGTTACTTTAAGTTAAAAAAATAAAAATTCAACTACAAGATGTTATATG 300
  TATCGACAAAATGATTCTCAATTTTTTATTTTAAGTGATGTAAAATCA

301 AGCTAGTGTTGGATTTTGGCGACTTTCAACTGTTCTAAAGTTTTATTAGCTCAAGTTCAACCAACCAAGCGTGAAGTGGACAGATCCAAAATAT 350

351 ATATAGAATATCCTGCCCTTTATTGAGATGTTGCTTAAGAGGTATGGTTTTTATTTAATGTTAAATCTCGCTGCTGCCAA 400

401 CCTATGACTAATATTGCTCTCTTTTTAATGGCTCTGAGTGGTTTATTTAAC GGAATTGATGTTAAGGGAGAATAATACCCAGGACTTCTCTCAAAAATTTG 450
Figure 11. (cont.)

451  CGGCTTCTATTTGGAAATATAGATTCGGCAACTTTTTCTACCATTGATTA  500
     GGCAGAGATAAAACCTTTGATTTACATGGTTAAGAAAATGGAATCTAATT

501  GTTGGGTTPATTGTAATAGATGAAGAAGAACATGAGCCACCTTACAAATCAG  550
     CAACCCAAAATTAAGATATATCTACTCTCTCTCTGATCTCCGCTAAATGGTTTAGTC
     .  RBS  .  ORFrbr  ->  .

551  AAAGTGACTCTAGTCATACATACGAAGATTATTCTGGATAAATGAGAAAAGTCGA  600
     TTTGCACTAGGATTICATGTTACCTATCTTTCAACGGACTTTTCATTTCACTGT
     MKSA

601  ATAGAAGATTGTATTTCTTGGTTGGGATAACCTGCTTGGAATGAACAC  650
     TATCTTTGTTCTAAGAAAAACAAACCTAGGCGTTAGGGCAGCATCCTTG
     IDECILVLSATPAVET

651  CTATTATAAAACCCTTTAAAGGAAAAGATTATGTGTCTGATAACACTAAAACA  700
     GATAATATTGGGAAATTCTCTATTTCTGATACATTTGATTGATTTTTGT
     YKTLKEDYDLDITLKNR

701  GGGCTGATGGTGCTCTATTGCGAGAAGTATCAGTGGTTGATATGAGAGAA  750
     CCCGACTACCCAGAGATACGGCTGCTCTTATAGTCAACCAGCATACCTCCCTT
     ADGALMPEVSVDMDRE

751  GAATTAAGAAACAAAATCATATAATTCTGTACGAGCCCTTGTATGAAG  800
     CTATATTCTTTTGTTATATGATAATATGGTCTGGAACATACCTCCCTT
     ELRENKNKSIFFSRALYES

801  CATTATTGAGAAGCGTTGGAGAGAAACAAATTATATCCCTTTCTCAATA  850
     GTAAAAATCCTGCCAATCTTCTTTTTCTGTTTAATATGAAAGAAAGATTTAT
     ILETLEKEQIILFLNRR
Figure 11. (cont.)

<table>
<thead>
<tr>
<th>851</th>
<th>GGAGAGGATATTCAACCTTTTGCCTTAGGAAATGTGGTATGTTTTT</th>
<th>900</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>GCTCTCTCTAGTGTTGAGCTCAGCATCTTTACACCAATACAAAAA</td>
<td></td>
</tr>
<tr>
<td></td>
<td>RGYSTFVSCRKCGYVF</td>
<td></td>
</tr>
</tbody>
</table>

| 901 | AAGTGCAAATTTGGCAATGTTTCAATGACTTATGACTATAATAAGGAA |
|-----|----------------------------------------------|----|
|     | TTTACGTTTTAAAGCTACAAGTGTTGATTATTATTCGCTAT    |
|     | KCKNCDSMTYHNNKGY                               |

| 951 | TCTGATATGCGCCTACTGCGGAGACCTCAAGAATTCCCTTCAATGTTGTC |
|-----|--------------------------------------------------|--|
|     | AGAGCTTAGTGGTAAAGCCTTGGTTAGTTTCTTTAAAGGAGATGACCATCAG |
|     | LICHYC GSTQRIPSVC P                                |

| 1000 | CTAAAGTGCCGTGAGCAAGTGTTGTTTGGGAGCAGAAGCAGAGAGG |
|------|-----------------------------------------------|--|
|      | GATTCAGCCATGTTTCTCATATAATCAAAACCTCGTCTGTCTTCC |
|      | KCGSKYVKYFGAGTER                                  |

| 1050 | ATAGAAAAAGAGAGAAATTTTTTCCAAATTGCAAAGACCTGAAGAT |
|------|-----------------------------------------------|--|
|      | TATTTCTCTCTATTTTCGAAAAGGTTTACGTTTCTTGACATCTTA |
|      | IEREIKRFLPNARTVRM                                 |

| 1100 | GGATAGGGATTCTACGGTAAAAAGATCTCTAGTAAGATATAATGG |
|------|-----------------------------------------------|--|
|      | CCTATTCTATGATGGGCAATTTCCTATAATACAACTTTCTGCTA  |
|      | DRTTVKKSYERYIMA                                 |

| 1150 | CTTCAAGAACAAGGGATATGTTTGGTGATAGGTACTAAATTGATGT |
|------|-----------------------------------------------|--|
|      | GAAGTTCTTGTCTGCCTAAATGCTTAAACTACATTTCACTGAGT|
|      | SRTKGYDILIGTQMIA                                 |

| 1200 | AAGGCACCATGATTTTAAAGGATGTTTTAGTTGGAGTAGTAGGTGGAG |
|------|-----------------------------------------------|--|
|      | TTCCGTTATCTAAATCTCCATACCTAAATCAACCTCATCATCGAGCCCT |
|      | KALDFKDVTLVGVVVAAAD                           |
Figure 11. (cont.)

1651 AAAGGCCGATATAACCTTATGAGTAGCCCTGCTTTTATAAAAATACTTAATTT
TTTCGCTATATTGAGATCTCCATCGAGGAATATTTTTTGTGATAAT
KGDIYEVACFIKKSLEY

1701 TGAATTTGTTAAAAAAACTTTATAATAATATAAGGTTTTCTAGATATA
ACTTACATAATTATTATTATTACTTCAAAGATATCTATAAT
ELLKLYNNIRVSLDIN

1751 RBS def ->

1751 ACCATTGAGTTATATTATAATTATTTTGAGCTATGTTAATGCGAATAAGT
TGGTAACCTCAAATATATATATAAAAACCCATTAAATTACCGTTATAGCA
PLSLStopMAIR

1801 AGTTATAAGGAAGATGTTGACGGAATATTATAGGGAAAAAGATGAAAAAGT
TCATATTCCTTCATACCCCTGCTAAATAATTCTCTTCTTCTTCTTCTTCTTCA
SRKYGDELRLKKSRRKV

1851 AGAAAAAGATAGATAAAGGATATAATAGCTTATTTAGATATGTTTGAAT
TCATTTCTATCTATCTCTATATAATTTAGAATATCTACTATAAACTTT
EKIDKRLLTLIDDMFET

1901 CCATGTATATGAGTGATGAGCTCTGCGATGTTGCTCTACACAGGTTTGGTATA
GTCACTATATGCTCTACCTCACTCACTCAAGACGTGGGTTGCCACCATAT
MYNADGVLAAAPQVGIC

1951 TTTAAAAAGACTTTGTAGTAATAGATATGAGGAAGGTTCTGTTGTGATAAT
AAATTTTCGACATACATATATACATCATCTTTCCTGGAGAACATGATA
LKRLLVVIDGEGPVVVLI

2001 AAATCTGTGGAAATTCATAGACAAGGTTGGAAAAGCCTGAGAAGTTAGAACGCT
TTAGGACTTTATACCTCCTGTTCACCINTTGGACATCTACATCTCTCGAG
NPENILETSGLKAVDVGC
Figure 11. (cont.)

2051 GCCCTAGTATTTCCGAAAGCACAAGGAGAAGTGGAAAGGACACATATGTT 2100
GGATTCTAAGGGTCTTTGCTTCTACCTCTCTCTGTGGTATACAA
LSUPERQGEVERPTYV

2101 AAGGCAAAAAGCTCTAAATGAAAMGGAAGMAATGTAATTGGAAGCAGA 2150
TTCCGTTTGGAGATTATTTTTCTCCTTCTTAACTTTAATTTGTGC
KAKALNEKGEIEIVIEAE

2151 AGATTTATTTGGCAAGAGCTATATGCGCTAGAAACAGAATTTAAATGGAG 2200
TTCAATTTAAGGTTCTGTATACGGACTTTGTCTAGATAATTACTCT
DLFAARACHTEDHLNGV

2201 TTTATTTTTGAGAAATTTAGGGAAGTGAGGAAATTTATTGGAAGA 2250
AAAATACATCTATTTATCGACTTTCCACTTTTAAATTACACTTT
LFVKLAESSGNNStop

fmrT ->

2251 ATCTGGTTTTATGGGAACCGCGATTTTCCAGTACCCCTCCCTGAGAAACT 2300
TAACACAAATGCTCTTGTGAGGAGGAAACTCCTTTGAG
MGTPEFSSVPSLEK

2301 TATAAGAAATTATTGATGIAAGGCTGTTTGGCCGAGCAATGACTA 2350
ATATCTTTTAACTAGACTCTTCCCGAGAAAACGTGCTGGTCTATTGGAT
IENYDVRAVLQTQPDKPK

2351 AGGAGAGCGGGAAGGAGAGCAGTCTTGATGCTGATGAAAAGATGAGGAGTT 2400
TCCCTGACACTTTTTCTGTGGTAGATAGAGACACTGAAAATTCTCTATGTC
GRGKKSMLAMSEVEKAV

2401 AAAAACAAATTTCCGAGTTTTCAAAGAGTTAAATTTAAAATTCAGTTTGA 2450
TTTGATTTAAAGAAGCAAGTTGCGTTGAAATTTATTTATCATTTAATCT
KNIPFVFPVPVLKNDIE
**Figure 11. (cont.)**

2451  AGTTATATATTACAAAGAAAAGAATAGCAAGATTTATAGATTATGATGG 2500  
      TCAATATTATTATTATTATTATCGTTGATCTATATATACGACC  
      V I N K L K E I A P D F I V V V A  

2501  CATTGCGACATTTATACAGGAAAGCTCTGATATCTCTAATGCTATGACGCA 2550  
      GTAAMCCGATTTTTATCTTTCTTGAAGACTATAATGGAATCTACGTTG  
      F G Q I L S K E V L D I P K Y A  

2551  TGTATATATCTGCAATGCACTCTCTTTTGCACAAAATATACGAGATGATCAGACC 2600  
      ACAAATTTAGAGCTACGTAGAGAAAACGTTTTATATCTCGAGAAGTG  
      C I N L H A S L L P N Y R G A A P  

2601  TATAGTGGCCATATTATATTTTAAAGAGAAGACAGGAAATATACACTA 2650  
      ATATTTAACCCTTAATATTACCTCTTTGTTCTGCTTTATATGATGAT  
      I N W A I I N G E T K T G N T T M  

2651  TGATTATGCTGAAAGATGAGTGATATGCTCTTTAAAGGATGAA 2700  
      ACTTTACGAGACTCTCTTAACCTATGCTACTATACGMAAGATTCTACTCTT  
      I M A E G L D T G D M L L K D E  

2701  GTTGATATATTAAAGAGATATAGACAGGCTCGACTTTATGATATTGTGAT 2750  
      CAACTATATTATTCTCTATGCTCTCCATGTAAGTACTATAAAMCTA  
      V D I K R D M T A G E L H D I L M  

2751  GAAATGAGGCGACAGTCTGATGATGTGTTACGCGACATTACATTACATAT 2800  
      CTTATCTCGAGCTAGCTACTACCATACGATCGCTGCTGTTATATAGCG  
      N R G A D L D D G Y A R L S N M L  

2801  TGCTGAGGCTATTACGCGCGCA 2823  

ACGAACCTCGAATAACGCGCGCTG  

LE A Y S G A  End of Insert
Figure 12. MacVector derived open reading frame (ORF) map of the insert in pRUNC51. Rows 1, 2 and 3 read 5' to 3'. Rows 4, 5 and 6 read 3' to 5'. The ORF from 589 to 1767 corresponds to ORFrubr. The ORF from 1789 to 2238 corresponds to def. The incomplete ORF starting at 2254 corresponds to fmt.

Start / Stop Method: AAspan >= 25
Genetic Code: universal
rubredoxin oxidoreductase nor genes coding for primosomal protein n' reported near def and fmt in the other organisms).

In the *C. acetobutylicum* operon, *def* encodes a 150 amino acid (16,500) polypeptide which shows high homology to DEF from other organisms (Figure 13). Deformylase has recently been found to be related to the zinc protease family Meinnel & Blanquet, 1993). The zinc proteases are characterized by the amino acid motif HEXXH (Vallee & Auld, 1990), which is found in the deformylases from *C. acetobutylicum* and several other organisms. The *fmt* initiator ATG codon is 16 nucleotides (nt) downstream of *def* in *C. acetobutylicum* (Figure 11). However, only 600 nt were sequenced before reaching the end of the insert. Approximately 300 nt remain to be sequenced. The known sequence encodes a polypeptide with good homology to FMT from other organisms (Figure 14).
Figure 13. Alignment of deduced 150 amino acid sequence of *Clostridium acetobutylicum* DEF with the DEFs of other organisms. Outlined regions show identity between the different species. H. influenzae is *Haemophilus influenzae*. E. coli is *Escherichia coli*. C. acetobutylicum is *Clostridium acetobutylicum*. T. thermophilus is *Thermus thermophilus*. 

HEXXH = zinc protease motif.

<table>
<thead>
<tr>
<th></th>
<th>H. influenzae</th>
<th>E. coli</th>
<th>C. acetobutylicum</th>
<th>T. thermophilus</th>
</tr>
</thead>
<tbody>
<tr>
<td>Seq</td>
<td>MTA LNVL YPDIL KVKCEP VTKVND AIRKIV DDMDTMY EKKGIGL AAP</td>
<td>MSLQVLHLDPL-LRKVAEPVE LT ARLQIVDDMFETMYAEGIGLAAAT</td>
<td>MAIRSKIRKYGDELRK36RRKVEKKIDKLRLXXLI/DDMFEETMYNADGVGLAAAF</td>
<td>MVYPIRLYGDPVKRKRPAVEQDFSGLRLAEDEMGLEAKEGVGLAAAP</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Seq</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Seq</td>
<td>QEGCLSIPEQ RALVPRKKEVKVTQALRDGKKEFLTDDAGGELAACIQHEID</td>
<td>QEGCLSIPEQ RALVPRKKEVKVTQALRDGKKEFLTDDAGGELAACIQHEID</td>
<td>TEGCLSIPEQER QGERPTYVKKAKNEKgeeIVLAEFLFARAIHED</td>
<td>TEGCLSIPEQGLYSEEVPRAEIRVEYQDEEGRGRVLELEGYMARVFHEID</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Seq</td>
<td>HLNGLFVDYSLP KQRKIKE</td>
<td>HLGKLFMDVLPSLQPQRIQ</td>
<td>HLNVLFVRLAASEGN</td>
<td>HLGKLFMDVLPSLKP</td>
</tr>
<tr>
<td>Seq</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 14. Alignment of deduced 200 amino acid partial sequence of *Clostridium acetobutylicum* FMT with the FMTs of other organisms. Outlined regions show identity between the different species. C. acetobutylicum is *Clostridium acetobutylicum*. R. prowazekii is *Rickettsia prowazekii*. E. coli is *Escherichia coli*. H. influenzae is *Haemophilus influenzae*. Synechocystis is *Synechocystis sp.*
The $F_0$ Complex of the Proton-translocating F-type ATPase from *Clostridium acetobutylicum*

The F-type ATP synthase ($F_1F_0$) catalyzes the production of ATP, driven by an electrochemical proton gradient formed by respiration. Conversely, the enzyme complex can generate an electrochemical gradient of protons by hydrolysis of ATP. An F-type synthase is also found in *Propionigenium modestum*, but it is a sodium pump rather than a proton pump (Laubinger & Dimroth, 1987; Laubinger & Dimroth, 1988). F-type synthases are composed of two distinct entities; a catalytic component ($F_1$) possessing 5 subunits ($\alpha, \beta, \gamma, \delta$ and $\epsilon$), and a structural component ($F_0$) possessing 3 subunits ($a, b$ and $c$). An operon containing genes encoding the latter has been cloned and sequenced from *Clostridium acetobutylicum*.

A gene encoding thiolase had been cloned and sequenced from pTEC011 by Stim-Hemdon et al. (1995). Downstream of thiolase, Stim-Hemdon sequenced most of an operon possessing three ORFs with homology to the subunits of $F_0$ (arranged $atpa$, $atpc$ and $atpb$). I finished sequencing the operon and analyzed the genes in the operon for homology with other $F_0$ subunits and for hydrophobicity.

Within the operon, the 663 nucleotide (nt) $atpa$ is followed by the 243 nt $atpc$, which is followed by the 477 nt $atpb$. Seventy-seven nucleotides separate $atpa$ and $atpc$. Fifty-two nucleotides separate $atpc$ and $atpb$ (Figure 15). This is the common gene arrangement found with F-type ATP synthase operons in non-photosynthetic bacteria (Walker et al., 1984; Brusilow et al., 1989). The five genes encoding the $F_1$ complex are usually located downstream of the Fo
Figure 15. Partial sequence of the pTECO11 insert. The location and amino acid sequences of the 663 bp atpa, the 243 bp atpc and the 477 bp atpb are shown. "RBS" and underlined sequence correspond to putative ribosomal binding sites.

TAAATTCAGAAAATGAATATATATTTTATCAATTTCAGTCATTGAAAGA
 ATTTAAGCTTTTTACTTATATAAAATAGTTAAGTCGTAACCTTCT

TTATGAGCTAAATTCAGTACTAGGCGTAATTTGAAAAATTTTATAACTATA
 AAATACCGAATTCCGGTATGATCGGATTATTTATTAAATTTAATGATAT

GCCATAAGAAATGGGCTAAAAACGTTTTGGCTATTGAATAGAAAGACGGTAAAT
 CGCTATTCTTTTCGCGATTTTTGCAAACGCTACTTTTTCTTTGGCATTTA

ATTATAGAAAAATCTTTAAAACAGATTTTATTTTATAAAATTTAAGATA
 TATATTTTTTTAGATTTTGCTCTCAAATAAATATTTTTAAATTTCTAT

TATATTTAAATAACGGTTTAAATTAGGAGGAGTATTTTTGAACCTGA
 ATATTTAAATTTTTGAGAATTCTCTCTTCATAAATCTTAGACT

ATAATTAAAGAATGGTAAAAGGTGAACGTCTTTATATGCAATTATTGCT
 TATAATTTTTCTCTAAATTTTCAACATGGAAATACTAGTTAAATACGA

GCCATAGTTTCAGTAAATCTTTTGGCTCAATTATAGATTTTCTCTT
 CGTTATCAAAGCTATTGAAACAAGACGGATTATAATCCTTAAGCAA

AAATACGATTATAGGGATTTTTGCAAGAAAAATTTTATTATAAGATA
 TTATCATATATATCCCTAAGAAAGTGTATTAAATATATATCTCTAT

ATTAGACGGATAAATGGGAAATAGGTCACC
 TAAATGCTGACAAAGCAAACATTTTTTTACCTTTATACAGTGA

ATATTTCTTAAGTTCAAATTTTTAGAGTAATACTTGGTTTTTTTATTAGGATAT
 TATAAGAAATCAAGTAAATCTCTTATGAAAACAAATAATATCCATA
Figure 15 (cont.)

\[ \text{atpc} \rightarrow \]

1351 \text{GGAACGAAATATATGATCTCATACTTTTACTAGGAATGCAATA} 1400
\text{CCTTTGCTTACTATACACCTAAAGGTATTGACATCTTACGTTAT}
\text{M N I D S H T F L G M Q Y}

1401 \text{TCTAGGCTCGGTCTAGCAGCAATTTGGATATTGGTGAGGGTTAGGTA} 1450
\text{AGATGCCAGTTCAGATCTGTTAACACTACATAACCCACCTCCACATCCAT}
\text{L G A G L A A I G C I G G G V G I}

1451 \text{TAGGTACTGTATCGATTTAAAGATGTGGAAGACAGCACACAGAA} 1500
\text{ATCCATGACACTTACATCTT GCCATTTCGTTAACCTGTTGTGTCTT}
\text{G T V T G K A V E A I G R Q P E}

1501 \text{TCAAGTCTCATAAAGTTATGCTCACAAATGTATGGGGTTAGATCTTGGC} 1550
\text{AGTCGAAGGTTCATCAATACGATTTACACCAATGCAATAACGACT}
\text{S A S K V M P T M I M G L A F A E}

1551 \text{AGTTACATCCATTAGCGTTTTTTCATTCTTGCTCTATATTATTATTGA} 1600
\text{TCAATGATAGTATACTGGAATAATCGAAATACGATATTACACATATAATACCAAT}
\text{V T S L Y A L F V A I M L L F V K}

1601 \text{ATATATAATAGTGTATCTTAAAATAAGAATAAGCTGAGGGGAGGGCTTTTAG} 1650
\text{TTATTATCTCAATAGAATTATTCTTTATATGTTACGAACTCTCTCCGGAATACCT}
\text{Stop}
\text{atpb} \rightarrow \]

1651 \text{TGTATGGAAATTATTTTATAGTTATGCTTTCTACATGTAACATTCAT} 1700
\text{ACATACCTTAAATATAATGAATACCAATCGGAAATATGTATCATTTCAGTTAAG}
\text{M E F N L V T I G F T I V N F I}

1701 \text{ATTCCTATATGTGATACCTAAGCACCCTTCTTTTGATAAAGATGAAATAAG} 1750
\text{TTAGGAACTACATGAGCTGAAAAGAAGAATTTCTACTTTTATTTAC}
\text{I L M I L I L K H F F F D K V N K V}
| 1751 | TTATTTGATGATGAGGAAATAATGAAATGAGTACAACTTAAATAAAAAAGACTGAT      | 1800 |
|      | AATTAACACTACTATCTCTATTACTACTTCATGTAATTGTATTCTTTTTGACTA        |      |
|      | I D D R N N E V A L T I K K A D                              |      |
| 1801 | GCCACAAAAAGGAAGCTAGGCCTTTTAAAAGTTGAAAGGTAAAAGATCT            | 1850 |
|      | C G I T T T G C T T C T G A T T C G A A A A A A T T T T A C T T C    |      |
|      | A Q N E E A R L L K V E S E K N L                              |      |
| 1851 | AGAAGATTCAAAACTTCAAGGAAGAAAACATTGTGAAAAAATTACAAAGTGAA         | 1900 |
|      | T C T C T A A G G T T T T A G G T T C T T T T G G I A A C T C T T T T T A T G T T C A T T   |      |
|      | E D S K L Q G K T I V E N Y K V K                             |      |
| 1901 | AAGCAGAAAAAGTTTTCAGAGGAATAACAGCAGAAGCTAAAACGTGAAGCT          | 1950 |
|      | T T C G T C T T T T C A A A A T C C T T T C C T C T T T T T G A C T T C G A A T T T G C A C T T C G A   |      |
|      | A E K V S E E I T A E A K T E A                              |      |
| 1951 | CAGAATATATATTAGAAAGAGCCTAAAAAGAGAACCAGAAACGAAAAAGAAA         | 2000 |
|      | G I C T T A T A T A T C T T T C T C T G A T T T T C T C T T T T T T G G T T T T C T C T T T C T T T T T T   |      |
|      | Q N I L E R A K R E T Q R E K E K                             |      |
| 2001 | GGCCTGAGAGGAATCAAATAATCAGATTGTTGAACCTCTGCTATTATTATT          | 2050 |
|      | C G A C T C T C A T T T T G T T T T T A G T T T A C A A A A C G A C T G A A G C A T A A T A A A A A   |      |
|      | A E D E I K N Q V V E L A V L I S                            |      |
| 2051 | CCTCAAAGCTTTTCAAAAAATTCCATAAAATGAAAGCAGAACACAGAAAACCTT      | 2100 |
|      | G G A G T T T T C G A C T C T T T T A G T T A C T T T T T C T T G T G C T C T T C T T T T G A   |      |
|      | S K A L E N S I N E A E H R K L                                |      |
| 2101 | ATGAAGATTTTGTATCTAAGGTTAGGTTATTATAATTGATGAAATTTTACT          | 2150 |
|      | T A T C T T C A A A A C A T A G A T T C A T C A T C A T A A A T T A T T A T A C A T A C T T A A G T G A   |      |
|      | I E D F V S K V G I Stop                                      |      |
genes. This may also be the case in *C. acetobutylicum*, but the insert ends less than 20 bp past the end of *atpb*.

Subunit *a* is a 221 amino acid (24,300) polypeptide whose highest homology (66% similarity, 35% identity) is with the Na⁺ motive subunit *a* from *Propionigenium modestum* (Figure 16). The *a* subunits from chloroplasts show the next highest homology (61-64% similarity, 27-34% identity). *Escherichia coli*, *Bacillus megaterium* and *Bacillus subtilis* subunits have progressively lower homology (62%, 60% and 60% similarity; 30%, 29% and 27% identity).

Subunit *c* is an 81 amino acid (8300) polypeptide whose highest homology (71% similarity, 44% identity) is with *Synechococcus 6301*, a cyanobacterium (Figure 17). The next highest homologies (67-69% similarity, 40-42% identity) were found with the higher plants, so it appears that subunit *b* has its best homology with photosynthetic organisms. *P. modestum*, followed by *B. subtilis* and *B. megaterium* show the lowest homology (60%, 70% and 65% similarity, 42%, 33% and 28% identity).

Subunit *b* is a 159 amino acid (18,200) whose highest homologies were with *B. megaterium*, *B. subtilis* and *Streptococcus pneumoniae* (58%, 58% and 56% similarity, 34%, 32% and 30% identity) (Figure 18). *P. modestum* and *E. coli* were next (55% and 50% similarity, 27% and 24% identity), leaving the photosynthetic organisms with the lowest homologies (47-48% similarity and 23-27% identity).
Figure 16. Alignment of deduced 221 amino acid sequence of *Clostridium acetobutylicum* ATPa with the ATPa of other organisms. Outlined regions show identity between the different species. *C. acetobutylicum* is *Clostridium acetobutylicum*. *P. modestum* Na is *Propionigenium modestum* (Na transport). *E. gracilis* is *Euglena gracilis*. *Antithamnion* is *Antithamnion* sp. *N. tabacum* is *Nicotiana tabacum*. *Synechocystis* is *Synechocystis* *Sp*. *E. coli* is *Escherichia coli*. *B. subtilis* is *Bacillus subtilis*. TMH = transmembrane helix determined by TopPred II.
Figure 17. Alignment of deduced 81 amino acid sequence of Clostridium acetobutylicum ATPc with the ATPc of other organisms. Outlined regions show identity between the different species. C. acetobutylicum is Clostridium acetobutylicum. Synechococcus is Synechococcus Sp. S. oleracea is Spinacia oleracea. P. satium is Pisum satium. M. polymorpha is Marchantia polymorpha. G. sulphuraria is Galdieria sulphuraria. P. modestumNa is Propionigenium modestum (Na transporter). B. subtilis is Bacillus subtilis. E. coli is Escherichia coli. TMH = transmembrane helix determined by TopPred II for C. acetobutylicum ATPc.
**Figure 18.** Alignment of deduced 159 amino acid sequence of Clostridium acetobutylicum ATPb with the ATPb of other organisms.Outlined regions show identity. Shaded regions show homology. C. acetobutylicum is Clostridium acetobutylicum. S. pneumoniae is Streptococcus pneumoniae. B. megarerium is Bacillus megarerium. B. subtilis is Bacillus subtilis. P. modestumNa is Propionigenium modestum (Na transport). E. coli is Escherichia coli. G. sulphuraria is Gallideria sulphuraria. TMH = transmembrane helix determined by TopPred II for C. acetobutylicum ATPb.
Figure 18. (cont.)

C. acetobutylicum 85 KEVSELITEAQNIIKELERAKRETQREKEKAEDENKNQVELAVLISS 134
S. pneumoniae 89 EQSKANILADAKYVEAGRLKEKANQETIAQNAELQSYSGEVDLTYESL 138
B. megaterium 96 EQKRFEIYAAARFEQRLKAAKQETEQQDKQATAAAEREQVASELSTAS 145
B. subtilis 94 EQKREEIQAARAFESRLEKAEATIEKREQASVALEREQVASELSTAS 143
P. modestumNa 92 DERKETEGRKEANTOREKMLKSAEVEIIEKMKKEQARKELOQLTDEVLAVKLAE 141
E. coli 84 SQILDIAKAAEAEAERQRTKTVQAAQAELAEERKRAREELRKOVAILAVACAE 133
G. sulphuraria 101 QEFYNAQIQSAQKQEAQLQQIMMQFEKEKNIYSEKVEQISEQIKN 150
C. acetobutylicum 135 KALENSINEAEHRLIKLEDPSKVG 159
S. pneumoniae 139 KIISKNLDSHAKHKEIDQITDOLGEA 164
B. megaterium 146 KVIEKESELSEQDQEKILHIQEYGDRV 172
B. subtilis 144 KVIEKEDEQAOQKLIDYLKEVGESR 170
P. modestumNa 142 KMEKEVDAIKGANLIDOFITGEVGEIK 168
E. coli 134 KIIFERSVDEAANSIDVDKLYAE 156
G. sulphuraria 151 KLIISIYI 157
Based on the homology of subunit \( a \), *C. acetobutylicum* may have a sodium motive ATP synthase; however, the other two subunits' homologies do not entirely corroborate this. *P. modestum* has the first demonstrated Na\(^+\) motive F-type ATP synthase, but it very much resembles the proton motive ATP synthases. *P. modestum*’s F-type synthase can also function as a proton pump under low concentrations of sodium (Laubinger & Dimroth, 1989). This suggests the two classes are phylogenetically related, and that homology between *C. acetobutylicum* subunits and those of *P. modestum* may stem from this. However, it would be reasonable to expect the F\(_o\) subunits from *C. acetobutylicum* would show their highest homologies with the closely related *Bacillus* species' F\(_o\) subunits. The homology to higher plants was very unexpected. Further sequencing of the catalytic portion may reveal a more definitive trend with more homology to one class of synthase. Otherwise, expressing and purifying the enzyme may be the only answer.

Similar F\(_{1}\)F\(_o\) enzyme complexes are found in bacteria, mitochondria and chloroplasts (Schneider & Altendorf, 1987; Senior, 1990). The representative structures have been determined in *Escherichia coli* (Deckers-Hebestreit & Altendorf, 1992). Subunit \( a \) is characterized by its high hydrophobicity and 5-8 transmembrane domains. The common feature of subunit \( c \) is a hairpin-like structure. Two hydrophobic transmembrane helices are found on either side of a central hydrophilic loop. Subunit \( b \) is predominantly hydrophilic, but its hydrophobic N-terminus anchors it to the membrane.
The hydrophobicity of the three subunits of *C. acetobutylicum* ATP synthase were analyzed with the computer program, TopPred II (topology prediction; Claros & Heijne, 1994). Subunit *a* has five predicted transmembrane helices (Figure 19). The hydrophobicity of four of the domains was quite high, but the fifth was less definite. Analyses with Plot.A/KKD (based on mathematical algorithm devised by Klein, Kaneshi and DeLisi) and Plot.A/TMH (transmembrane helices) also predicted five transmembrane helices (data not shown). Subunit *a* of *C. acetobutylicum* most likely has five transmembrane helices. Subunit *c* was analyzed by all three programs also. The profile developed in TopPred II is displayed in Figure 20. The two transmembrane helices were also predicted with Plot.A/KKD and Plot.A/TMH (data not shown). Finally, subunit *b*'s TopPred II derived profile is given in Figure 20. Both Plot.A/KKD and Plot.A/TMH corroborated, with a single transmembrane helix at the N-terminus (data not shown).

While the homologies of *C. acetobutylicum*’s *F₀* genes are ambiguous, the hydrophobic analyses are very homologous to the subunits from other species. Cloning and sequencing the DNA downstream of this *F₀* operon will likely reveal the presence of another operon containing the five genes encoding the catalytic *F₁* portion of the enzyme complex. These five genes may even be sharing the operon with the three genes of the structural *F₀* portion. The homologies in the catalytic genes may resolve some of the ambiguity found with the structural gene homologies. Protein expression and purification of the entire *F₁F₀* enzyme complex from *C.*
acetobutylicum may also provide a clearer picture about its use as a proton or sodium translocating complex.
Figure 19. A Hydropathy profile of ATP synthase subunit a (Kyte-Doolittle) determined with the TopPred II program. B Predicted pattern of transmembrane helices within ATP synthase subunit a. (Figure is a modified output of TopPred II)
Figure 20. A Hydropathy of ATP synthase subunit b (Kyte-Doolittle) determined with the TopPred II program. B Hydropathy of ATP synthase subunit c (Kyte-Doolittle) determined with the TopPred II program. (Figure is a modified output of TopPred II)