JOURNAL OF BACTERIOLOGY, Nov. 1993, p. 6959–6969
0021-9193/93/216959-11$02.00/0

Vol. 175, No. 21

Cloning, Sequencing, and Molecular Analysis of the *sol* Operon of *Clostridium acetobutylicum*, a Chromosomal Locus Involved in Solventogenesis†

RALF J. FISCHER, JAN HELMS, AND PETER DÜRRE*

Institut für Mikrobiologie, Georg-August-Universität Göttingen, Grisebachstrasse 8, D-37077 Göttingen, Germany

Received 27 May 1993/Accepted 24 August 1993

A DNA region of *Clostridium acetobutylicum* contiguous with the *adc* operon has been cloned and sequenced. Structural genes encoding the acetoacetyl coenzyme A:acetate/butyrate:coenzyme A transferase (*ctfB* and *ctfA*) and an alcohol/aldehyde dehydrogenase (*adhE*) could be identified. These three genes together with a small open reading frame (ORF) of unknown function (upstream of *adhE*) formed an operon (*sol* operon), as shown by mRNA analyses. The complete *sol* operon was transcriptionally induced or derepressed before the onset of solventogenesis, thus confirming earlier results of Northern hybridizations with a *ctfB* gene probe (U. Gerischer and P. Dürre, J. Bacteriol. 174:426–433, 1992). Upstream of the *sol* operon, we identified two putative promoters that were located in regions with possible stem-loop structures formed by several inverted repeats. The distal promoter P1 showed only minor transcription initiation in solventogenic *C. acetobutylicum* cells but was recognized in *Escherichia coli*, presumably because of its high similarity to the σ^{70} consensus sequence. The *adhE*-proximal promoter P2 directed the major transcription start point in solventogenic *C. acetobutylicum* but was not recognized in *E. coli*. The clostridial AdhE showed high similarity to a novel family (type III) of alcohol dehydrogenases. Two other ORFs (ORF 5 and ORF 6) were found on the cloned DNA region that showed no significant similarity to sequences in various available data bases. mRNA studies revealed that ORF 5 formed a monocistronic operon and showed increased expression before onset of solventogenesis.

The formation of acetone and butanol by *Clostridium acetobutylicum* is induced at the end of the exponential growth phase, during which the organism produces considerable amounts of acids in a typical butyric acid fermentation (for reviews, see references 2, 22, and 38). The primary underlying molecular mechanisms responsible for this metabolic shift remain unknown. Significant progress has been made in the last few years with regard to the structure and organization of genes which encode the solvent-forming enzymes. Acetone production is catalyzed by acetoacetyl coenzyme A:acetate/butyrate:coenzyme A transferase (CoA transferase) and acetoacetate decarboxylase. The respective genes of *C. acetobutylicum* have been cloned and sequenced (9, 17, 35, 36). Alcohol formation is catalyzed by several different aldehyde and alcohol dehydrogenases. An NADPH-dependent alcohol dehydrogenase of *C. acetobutylicum* that is responsible for ethanol production in vivo (6) has been cloned and sequenced (52, 53), as have two NADH-dependent butanol dehydrogenase isozymes (37, 47).

Genes for acetoacetate decarboxylase, NADPH-dependent ethanol dehydrogenase, butanol dehydrogenase A, and butanol dehydrogenase B are each arranged in monocistronic operons in *C. acetobutylicum* (17, 18, 25, 36, 47). However, *ctfB*, which encodes the large subunit of CoA transferase, was found to be expressed from an approximately 4.1-kb transcript (18). Since the two CoA transferase genes (*ctfA* and *ctfB*) are located next to each other and cover roughly 1.4 kbp, approximately 2.7 kbp of coding sequence remained unidentified (9, 36). In this report, we describe the cloning, sequencing, and nucleotide sequence analysis of this DNA region from *C. acetobutylicum* which is coexpressed with the *ctfA* and *ctfB* genes.

(A preliminary account of the data presented here was given at the Annual Meeting of the Vereinigung für Allgemeine und Angewandte Mikrobiologie, Leipzig, Germany, 14 to 17 March 1993.)

MATERIALS AND METHODS

Bacterial strains, plasmids, and growth conditions. *C. acetobutylicum* DSM 792 and DSM 1731 were obtained from the Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH, Braunschweig, Germany. *Escherichia coli* XL1-B (7) and SURE (Stratagene GmbH, Heidelberg, Germany) were used as hosts and pEcoR252 (kindly supplied by D. R. Woods, University of Cape Town, Cape Town, South Africa) was used as the vector for the cloning experiments. pEcoR252 (3,349 bp) is a derivative of pEcoR251, a positive selection vector containing the *Eco*RI gene under the control of a temperature-regulated λ rightward promoter. At 37°C, the *Eco*RI gene product is produced in high amounts and becomes lethal unless insertionally inactivated (54). In pEcoR252, a point mutation at position 2597 has destroyed the *Pst*I site within the β-lactamase gene, such that the plasmid contains a unique *Pst*I cloning site in the *Eco*RI gene (in addition to the commonly used *Bgl*II and *Hin*dIII sites).

In batch culture, *C. acetobutylicum* was grown under strictly anaerobic conditions at 37°C in clostridial basal medium (32). Continuous culture experiments were performed as described by Bahl et al. (1). *C. acetobutylicum* DSM 1731 was used in these experiments because it can be reliably induced to produce solvents, whereas the type strain DSM 792 is more sensitive to the lower pH values used for solventogenic initia-

* Corresponding author.

† We are deeply distraught by the sudden death of Jan Helms during the course of this work. We are indebted to him for his important contributions to this study.

tion (18). It is interesting to note that so far, no differences at the DNA level have been found between these two strains (18, 43). Solventogenesis was induced by lowering the pH control set point from 6.0 to 4.3. Acid production resulted in a decrease in the pH value of the culture medium to 4.3, at which time solvent formation was established (18). *E. coli* was routinely grown at either 30°C (for pEcoR252 maintenance) or 37°C in LB medium or M9 medium (40). These media were supplemented with ampicillin (100 μg/ml) when necessary.

Nucleic acid isolation and manipulation. Chromosomal DNA of *C. acetobutylicum* was isolated by the method of Marmur (28) as modified by Bertram and Dürre (5). Plasmid isolation from *E. coli* was performed with the Quiagen Midi kit (Diagen GmbH, Düsseldorf, Germany). Total RNA of *C. acetobutylicum* was isolated by the hot phenol-chloroform procedure of Oelmüller et al. (33) as modified by Gerischer and Dürre (18).

DNA was manipulated by standard methods (40). Restriction enzymes, T4 DNA ligase, and *Taq* DNA polymerase were obtained from GIBCO/BRL GmbH (Eggenstein, Germany) and used according to the manufacturer's instructions.

Construction of a genomic library of *C. acetobutylicum*. Chromosomal DNA of strain DSM 792 was partially digested with *Sau*3AI (40). The resulting DNA fragments were separated by sucrose density centrifugation (10 to 40% [wt/vol]). Fractions containing fragments of the desired size were dialyzed against TE buffer (10 mM Tris hydrochloride [pH 8.0], 1 mM EDTA), ethanol precipitated, and ligated into *Bgl*II-digested pEcoR252. Transformation of competent *E. coli* cells with microdialyzed ligation mixtures (29) was performed by electroporation with a Gene Pulser (Bio-Rad Laboratories GmbH, München, Germany) as described by Dower et al. (15). Colony hybridization using Hybond N nylon filters (Amersham Buchler GmbH, Braunschweig, Germany) was used for screening the genomic library. The filters were prepared as described by Buluwela et al. (8), whereas hybridization and washing were carried out according to the protocol given below.

Hybridization. Total chromosomal DNA of *C. acetobutylicum* or plasmid DNA from *E. coli* was digested to completion with the appropriate restriction enzymes, and the obtained fragments were separated on agarose gels. Southern blots on nylon membranes (GeneScreen Plus; Dupont, NEN Research Products, Dreieich, Germany) were prepared according to the manufacturer's instructions. Membranes were prehybridized in 0.15% (wt/vol) polyvinylpyrrolidone, 0.15% (wt/vol) bovine serum albumin, 0.15% (wt/vol) Ficoll, 0.9 M NaCl, 10% (wt/vol) dextran sulfate, 1% (wt/vol) sodium dodecyl sulfate (SDS), 6 mM EDTA, 90 mM Tris hydrochloride (pH 7.5), and 100 μg of denatured salmon sperm DNA per ml (80 μl/cm^2 of membrane) for 1 to 3 h at 60°C. DNA probes were generated by the polymerase chain reaction (PCR) with a thermocycler (Trio-Thermoblock, Biometra biomedizinische Analytik GmbH, Göttingen, Germany), using synthetic oligonucleotides (17-mers) prepared for the sequencing reactions described below. Oligonucleotides (0.5 μM each), chromosomal or plasmid DNA (20 to 500 ng), a deoxynucleoside triphosphate mixture (200 μM), and *Taq* DNA polymerase (2.5 U) in reaction buffer (final volume, 100 μl) underwent 35 temperature cycles of 94°C for 1 min, 37°C for 30 s, and 72°C for 45 s. Resulting DNA fragments were phenol extracted and ethanol precipitated (40). Radiolabeling with [α-^{32}P]dATP (DuPont, NEN Research Products) was done by using a random primers DNA labeling kit (GIBCO/BRL), and the labeled probes were purified by using Sephadex G-25 columns (NAP-5; Pharmacia LKB GmbH, Freiburg, Germany). The appropriate radiolabeled probe (0.2 to 0.5 μCi/ml) was added to the prehybridization solution and incubated at 60°C for 10 to 15 h. Membranes were then washed twice in 2× SSC for 15 min at room temperature (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate), washed once in 2× SSC plus 1% (wt/vol) SDS for 5 min at room temperature, and subjected to autoradiography.

RNA for Northern (RNA) blots was separated in denaturing formaldehyde gels and transferred to nylon membranes (Gene Screen Plus) according to the manufacturer's instructions. An RNA ladder (0.24, 1.4, 2.4, 4.4, 7.5, and 9.5 kb; GIBCO/BRL) was included as standard for size determination. Hybridization and washing were performed according to the protocol described above for the DNA hybridization procedure.

DNA sequencing. Double-stranded plasmid DNA was sequenced by the dideoxy-chain termination method (41), using [^{35}S]dATP (DuPont, NEN Research Products) and a T7 Sequencing kit from Pharmacia LKB GmbH according to the protocol given by the manufacturer. The entire sequence (6,352 nucleotides) of the *C. acetobutylicum* DNA insert of pK9 was determined for both strands. Sequencing started within the vector, upstream and downstream of the *Bgl*II cloning site of pEcoR252, using the newly designed pEcoR *Bgl*II sequencing forward primer (pEcoR*Bgl*II-USP, 5′-TGGG AAAAGAGGAGATCA-3′) and reverse primer (pEcoR*Bgl*II-RSP, 5′-AAGGAAAGTGGCTCTCA-3′). Further sequencing was carried out by using sequentially synthesized oligonucleotides (17-mers) fitting to the ends of the already determined DNA sequences (primer walking). Oligonucleotides were synthesized on a Gene Assembler Plus (Pharmacia LKB GmbH) according to the manufacturer's instructions. The dideoxy-terminated fragments were separated on 55-cm wedge-shaped thickness gradient gels (0.2 to 0.4 mm, 6% [wt/vol] polyacrylamide) with a Macrophor sequencing unit (Pharmacia LKB GmbH) as recommended by the manufacturer.

Primer extension analysis. Primer extension analysis was carried out as recently published (18), with the modifications of Sauer and Dürre (43), using oligonucleotides of 17 to 20 nucleotides which were complementary to the 5′ ends of the respective primary transcripts. The length of each primer extension product was calculated by running a sequencing reaction with the identical primer on the same gel. The oligonucleotides J14 (5′-TCCAAAACAGCTGCTTT-3′), J14PE (5′-CTTCTTTAATTACCTTGAGT-3′), and LORF1 (5′-AGATTTGGAGATAGATATTG-3′) were used for determination of transcription start points of the *sol* operon; J21 (5′-TTCTATCACCCTTATTA-3′) and J21PE (5′-TTATA CATATCGCAATAGGT-3′) were used for determination of the transcription start point of open reading frame 5 (ORF 5); GORF1 (5′-TTCCGGTTCTGAGGTTACTC-3′) and GORF2 (5′-AGAATGTTGGTCCAAAACAT-3′) were used for analysis of the antiparallel ORFs with divergent direction of transcription within *adhE*.

Computer analysis. The DNA sequence data and the deduced amino acid sequences were analyzed by using the DNA Strider (27) and GeneWorks (IntelliGenetics, Inc., Mountain View, Calif.) programs on a Macintosh IIsi computer (Apple Computer, Inc., Cupertino, Calif.). Further sequence analyses were carried out on a VAX 9000 computer, using the Genetics Computer Group Inc. sequence analysis software package (version 6.0) (13).

Nucleotide sequence accession number. The sequence data shown in Fig. 2 were submitted to the EMBL data base and assigned accession number X72831.

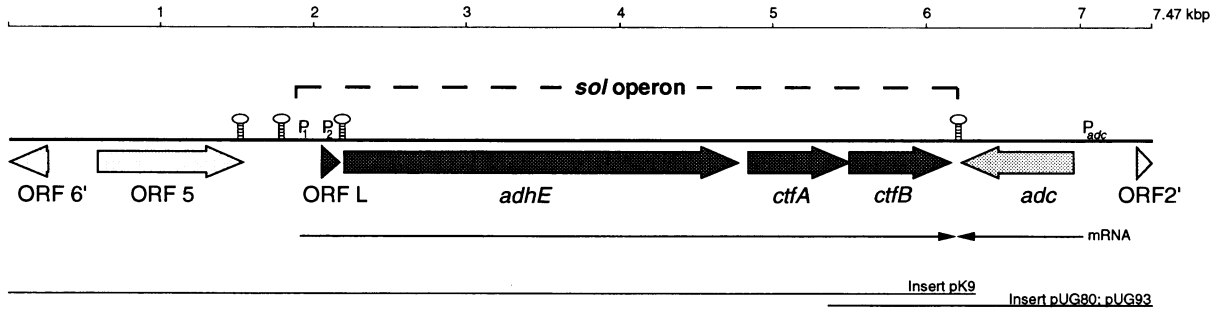

FIG. 1. Schematic ORF map of the chromosomal DNA region from *C. acetobutylicum* DSM 792 containing the *sol* operon. The 6.352-kbp insert of pK9 is shown, as well as the overlapping inserts of pUG80 or pUG93 (17). Arrows and arrowheads represent lengths, locations, and orientations of ORFs; lines with arrows indicate primary mRNA transcripts of genes induced before the onset of solventogenesis (*adhE*, *ctfA*, *ctfB*, *adc*). Dark-shaded ORFs belong to the *sol* operon, which is marked by the dashed line. A ′ at the end of an ORF indicates that the ORF is truncated. Putative promoter sequences are indicated by P_1, P_2, and P_{adc}; possible stem-loop structures are indicated by hairpin symbols.

RESULTS

Cloning and sequencing of a DNA region of *C. acetobutylicum* involved in solventogenesis. Partially *Sau*3AI-digested chromosomal DNA fragments (5 to 20 kbp) of *C. acetobutylicum* DSM 792 were ligated into *Bgl*II-cut pEcoR252 and transformed into *E. coli* SURE. The resulting library of 1,052 clones was screened by hybridization with a radiolabeled *ctfB* gene probe (18). Three of the clones on recombinant plasmids (pK9, pK16, and pK17) showed a positive hybridization signal. The recombinant plasmids pK16 and pK17 both contained the same (ca. 5.5-kbp) *C. acetobutylicum* DNA insert in opposite orientations. This insert covered the previously described *adc* gene region (17), extending only 1.4 kbp downstream of *ctfB*. However, restriction analysis of pK9 showed that it carried a larger contiguous DNA region, and this clone was therefore chosen for further analysis. Sequencing revealed that pK9 contained a 6,352-bp clostridial DNA insert (Fig. 1) that overlapped with 1,022 bp of clones pUG80 and pUG93. This fragment encoded seven ORFs, three of which represented previously cloned genes. With respect to the sequence data presented, the 3′ end of the insert contained the end of the *adc* gene, which is known to form a monocistronic operon. Furthermore, the *adc* gene has been shown to be contiguous to the *ctfB* gene (17, 18, 36) used as a probe. The *ctfB* (663-bp) and *ctfA* (654-bp) genes are located directly upstream of *adc*, with a convergent direction of transcription.

Separated by a 63-bp intergenic region, a large ORF of 2,586 bp was found upstream of the *ctf* genes. The deduced amino acid sequence from this ORF showed a high degree of similarity (76%) to the AdhE protein of *E. coli*. The *E. coli adhE* gene encodes for both alcohol and aldehyde dehydrogenase domains (19). The clostridial gene was accordingly designated *adhE*. The combined lengths of *adhE*, *ctfA*, and *ctfB* covered 3,903 bp, matching the expected size of ca. 4.1 kb (taking into account the space necessary for intergenic regions and sequences for transcription initiation and termination) which had been found earlier for a transcript containing *ctfB* (18). Since this newly identified *C. acetobutylicum adhE* gene has been previously shown to be transcribed together with the *ctf* genes induced before the onset of solventogenesis (18), we designate the complete transcription unit of *adhE*, *ctfA*, and *ctfB* as the *sol* (for solvents) operon. This designation is meant to emphasize the involvement of this operon in solventogenesis; however, it does not imply that it is the only DNA region encoding genes responsible for acetone and butanol production. Directly upstream of the *adhE* gene is a small ORF of only 108 bp which overlaps with three possible stem-loop structures formed by inverted repeats. This potential ORF (designated ORF L) showed no significant homology to any sequence in the EMBL, GenBank, and Swissprot data bases. The DNA sequence of ORF L has a much lower G+C content (22.2 mol%) than has been reported for the total genomic DNA of *C. acetobutylicum* (28 mol%) (12). Upstream of ORF L and separated by a large intergenic region (524 bp) with two typical rho-independent transcription terminators (ΔG = −90.9 and −93 kJ/mol) (46), another open reading frame (ORF 5) was identified. ORF 5 is 954 bp in length and has the same direction of transcription as the *sol* operon. A truncated ORF (ORF 6) found upstream of ORF 5 has a divergent direction of transcription. Like ORF 5, ORF 6 showed no significant similarity to any sequence in the EMBL, GenBank, and Swissprot data bases.

Since the *C. acetobutylicum* DNA insert of pK9 contained nine internal *Sau*3AI sites, it was necessary to prove that the original location of the fragments was indeed as shown in Fig. 1. This arrangement could be verified by restriction analysis of chromosomal clostridial DNA with *Hha*I, *Hin*cII, *Hin*dIII, *Nco*I, *Sau*3AI, and *Sca*I, yielding fragments of the expected sizes after hybridization with a radioactively labeled *ctfB* gene probe (data not shown). The 5.59-kbp *Sca*I fragment was found to cover almost the whole DNA insert of pK9 and to contain all of the internal *Sau*3AI sites in the expected locations. PCR studies with chromosomal DNA and various primers used for sequencing also unequivocally confirmed the sequence shown in Fig. 1 (data not shown).

Nucleotide sequence comparisons. The complete nucleotide sequence of the *C. acetobutylicum* DNA insert of pK9 is given in Fig. 2. Putative ribosome binding sites with reasonable homology to both *E. coli* (44) and *C. acetobutylicum* ribosome binding sites (18, 30, 31, 43, 47, 51) could be identified in front of all ORFs except ORF L. The *adhE* sequence also contained two long internal antiparallel ORFs that both showed a divergent direction of transcription with respect to *adhE*. The first covered 447 bp (nucleotides 3097 to 3543) and would encode a protein of 149 amino acids; the second covered 378 bp (nucleotides 3676 to 4053), corresponding to a putative protein of 126 amino acids. Several inverted repeats that partially overlapped with the possible promoter regions and with each other were found upstream of *adhE* and ORF L (Fig. 2). The sequence of *ctfA* showed only minor differences from the recently published sequence of the same gene from *C. acetobutylicum* ATCC 824 (36). Most important is an addi-

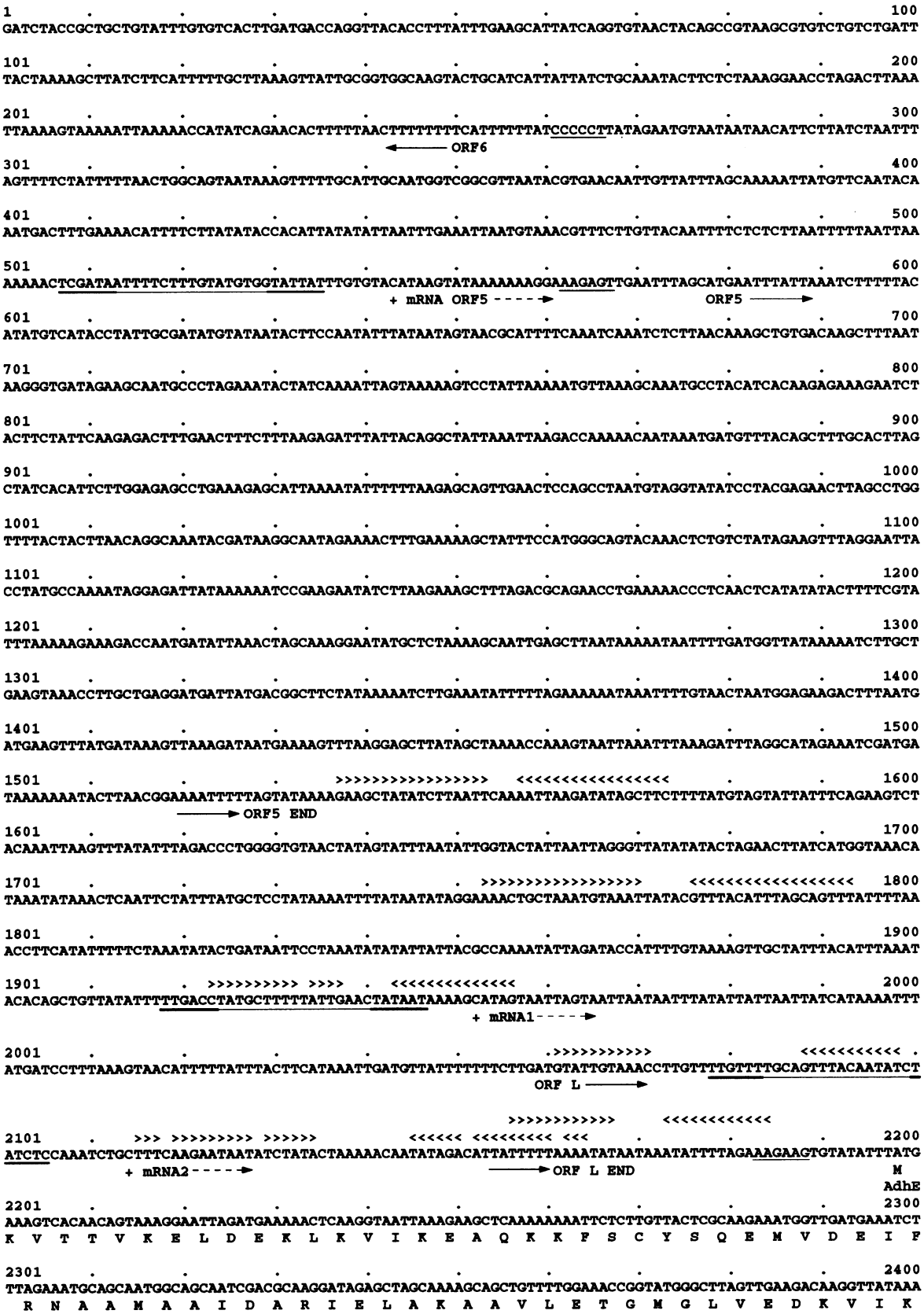

FIG. 2. Nucleotide sequence of the 6,352-bp insert of pK9 representing the region of chromosomal DNA from *C. acetobutylicum* that contains the *sol* operon genes *adhE*, *ctfA*, and *ctfB*. Only one strand is shown. The *sol* operon genes have been translated by using the one-letter amino acid code; amino acid symbols are written below the first nucleotide of the corresponding codon. Putative ribosome binding sites are underlined; potential promoter regions are marked by thick solid bars below the −10 and −35 regions joined by a line; putative secondary structures are marked by open arrows indicating the lengths and orientations of the stems. Determined mRNA starting points are marked by a + below the first base of the mRNA transcript.

```
2401 . . . . . 2500
AAATCATTTTGCAGGCGAATACATCTATAACAAATATAAGGATGAAAAAACCTGCGGTATAATTGAACGAAATGAACCCTACGGAATTACAAAAATAGCA
 N  H  F  A  G  E  Y  I  Y  N  K  Y  K  D  E  K  T  C  G  I  I  E  R  N  E  P  Y  G  I  T  K  I  A

2501 . . . . . 2600
GAACCTATAGGAGTTGTAGCTGCTATAATCCCTGTAACAAACCCCCACATCAACAACAATATTTAAATCCTTAATATCCCTTAAAACTAGAAATGGAATTT
E  P  I  G  V  V  A  A  I  I  P  V  T  N  P  T  S  T  T  I  F  K  S  L  I  S  L  K  T  R  N  G  I  F

2601 . . . . . 2700
TCTTTTCGCCTCACCCAAGGGCAAAAAAATCCACAATACTAGCAGCTAAAACAATACTTGATGCAGCCGTTAAGAGTGGTGCCCCGGAAAATATAATAGG
  F  S  P  H  P  R  A  K  K  S  T  I  L  A  A  K  T  I  L  D  A  A  V  K  S  G  A  P  E  N  I  I  G

2701 . . . . . 2800
TTGGATAGATGAACCTTCAATTGAACTAACTCAATATTTAATGCAAAAAGCAGATATAACCCTTGCAACTGGTGGTCCCTCACTAGTTAAATCTGCTTAT
 W  I  D  E  P  S  I  E  L  T  Q  Y  L  M  Q  K  A  D  I  T  L  A  T  G  G  P  S  L  V  K  S  A  Y

2801 . . . . . 2900
TCTTCCGGAAAACCAGCAATAGGTGTTGGTCCGGGTAACACCCCAGTAATAATTGATGAATCTGCTCATATAAAAATGGCAGTAAGTTCAATTATATTAT
S  S  G  K  P  A  I  G  V  G  P  G  N  T  P  V  I  I  D  E  S  A  H  I  K  M  A  V  S  S  I  I  L  S

2901 . . . . . 3000
CCAAAACCTATGATAATGGTGTTATATGTGCTTCTGAACAATCTGTAATAGTCTTAAAATCCATATATAACAAGGTAAAAGATGAGTTCCAAGAAAGAGG
  K  T  Y  D  N  G  V  I  C  A  S  E  Q  S  V  I  V  L  K  S  I  Y  N  K  V  K  D  E  F  Q  E  R  G

3001 . . . . . 3100
AGCTTATATAATAAAGAAAAACGAATTGGATAAAGTCCGTGAAGTGATTTTTAAAGATGGATCCGTAAACCCTAAAATAGTCGGACAGTCAGCTTATACT
 A  Y  I  I  K  K  N  E  L  D  K  V  R  E  V  I  F  K  D  G  S  V  N  P  K  I  V  G  Q  S  A  Y  T

3101 . . . . . 3200
ATAGCAGCTATGGCTGGCATAAAAGTACCTAAAACCACAAGAATATTAATAGGAGAAGTTACCTCCTTAGGTGAAGAAGAACCTTTTGCCCACGAAAAAC
I  A  A  M  A  G  I  K  V  P  K  T  T  R  I  L  I  G  E  V  T  S  L  G  E  E  E  P  F  A  H  E  K  L

3201 . . . . . 3300
TATCTCCTGTTTTGGCTATGTATGAGGCTGACAATTTTGATGATGCTTTAAAAAAAGCAGTAACTCTAATAAACTTAGGAGGCCTCGGCCATACCTCAGG
  S  P  V  L  A  M  Y  E  A  D  N  F  D  D  A  L  K  K  A  V  T  L  I  N  L  G  G  L  G  H  T  S  G

3301 . . . . . 3400
AATATATGCAGATGAAATAAAAGCACGAGATAAAATAGATAGATTTAGTAGTGCCATGAAAACCGTAAGAACCTTTGTAAATATCCCAACCTCACAAGGT
 I  Y  A  D  E  I  K  A  R  D  K  I  D  R  F  S  S  A  M  K  T  V  R  T  F  V  N  I  P  T  S  Q  G

3401 . . . . . 3500
GCAAGTGGAGATCTATATAATTTTAGAATACCACCTTCTTTCACGCTTGGCTGCGGATTTTGGGGAGGAAATTCTGTTTCCGAGAATGTTGGTCCAAAAC
A  S  G  D  L  Y  N  F  R  I  P  P  S  F  T  L  G  C  G  F  W  G  G  N  S  V  S  E  N  V  G  P  K  H

3501 . . . . . 3600
ATCTTTTGAATATTAAAACCGTAGCTGAAAGGAGAGAAAACATGCTTTGGTTTAGAGTTCCACATAAAGTATATTTTAAGTTCGGTTGTCTTCAATTTGC
  L  L  N  I  K  T  V  A  E  R  R  E  N  M  L  W  F  R  V  P  H  K  V  Y  F  K  F  G  C  L  Q  F  A

3601 . . . . . 3700
TTTAAAAGATTTAAAAGATCTAAAGAAAAAAAGAGCCTTTATAGTTACTGATAGTGACCCCTATAATTTAAACTATGTTGATTCAATAATAAAAATACTT
 L  K  D  L  K  D  L  K  K  K  R  A  F  I  V  T  D  S  D  P  Y  N  L  N  Y  V  D  S  I  I  K  I  L

3701 . . . . . 3800
GAGCACCTAGATATTGATTTTAAAGTATTTAATAAGGTTGGAAGAGAAGCTGATCTTAAAACCATAAAAAAAGCAACTGAAGAAATGTCCTCCTTTATGC
E  H  L  D  I  D  F  K  V  F  N  K  V  G  R  E  A  D  L  K  T  I  K  K  A  T  E  E  M  S  S  F  M  P

3801 . . . . . 3900
CAGACACTATAATAGCTTTAGGTGGTACCCCTGAAATGAGCTCTGCAAAGCTAATGTGGGTACTATATGAACATCCAGAAGTAAAATTTGAAGATCTTGC
  D  T  I  I  A  L  G  G  T  P  E  M  S  S  A  K  L  M  W  V  L  Y  E  H  P  E  V  K  F  E  D  L  A

3901 . . . . . 4000
AATAAAATTTATGGACATAAGAAAGAGAATATATACTTTCCCAAAACTCGGTAAAAAGGCTATGTTAGTTGCAATTACAACTTCTGCTGGTTCCGGTTCT
 I  K  F  M  D  I  R  K  R  I  Y  T  F  P  K  L  G  K  K  A  M  L  V  A  I  T  T  S  A  G  S  G  S

4001 . . . . . 4100
GAGGTTACTCCTTTTGCTTTAGTAACTGACAATAACACTGGAAATAAGTACATGTTAGCAGATTATGAAATGACACCAAATATGGCAATTGTAGATGCAG
E  V  T  P  F  A  L  V  T  D  N  N  T  G  N  K  Y  M  L  A  D  Y  E  M  T  P  N  M  A  I  V  D  A  E

4101 . . . . . 4200
AACTTATGATGAAAATGCCAAAGGGATTAACCGCTTATTCAGGTATAGATGCACTAGTAAATAGTATAGAAGCATACACATCCGTATATGCTTCAGAATA
  L  M  M  K  M  P  K  G  L  T  A  Y  S  G  I  D  A  L  V  N  S  I  E  A  Y  T  S  V  Y  A  S  E  Y

4201 . . . . . 4300
CACAAACGGACTAGCACTAGAGGCAATACGATTAATATTTAAATATTTGCCTGAGGCTTACAAAAACGGAAGAACCAATGAAAAAGCAAGAGAGAAAATG
 T  N  G  L  A  L  E  A  I  R  L  I  F  K  Y  L  P  E  A  Y  K  N  G  R  T  N  E  K  A  R  E  K  M

4301 . . . . . 4400
GCTCACGCTTCAACTATGGCAGGTATGGCATCCGCTAATGCATTTCTAGGTCTATGTCATTCCATGGCAATAAAATTAAGTTCAGAACACAATATTCCTA
A  H  A  S  T  M  A  G  M  A  S  A  N  A  F  L  G  L  C  H  S  M  A  I  K  L  S  S  E  H  N  I  P  S
```

FIG. 2—*Continued.*

```
4401    .       .       .       .       .       .       .       .       4500
GTGGCATTGCCAATGCATTACTAATAGAAGAAGTAATAAAATTTAACGCAGTTGATAATCCTGTAAAACAAGCCCCTTGCCCACAATATAAGTATCCAAA
  G   I   A   N   A   L   L   I   E   E   V   I   K   F   N   A   V   D   N   P   V   K   Q   A   P   C   P   Q   Y   K   Y   P   N

4501    .       .       .       .       .       .       .       .       4600
CACCATATTTAGATATGCTCGAATTGCAGATTATATAAAGCTTGGAGGAAAATACTGATGAGGAAAAGGTAGATCTCTTAATTAACAAAATACATGAACTA
 T   I   F   R   Y   A   R   I   A   D   Y   I   K   L   G   G   N   T   D   E   E   K   V   D   L   L   I   N   K   I   H   E   L

4601    .       .       .       .       .       .       .       .       4700
AAAAAAGCTTTAAATATACCAACTTCAATAAAGGATGCAGGTGTTTTGGAGGAAAACTTCTATTCCTCCCTTGATAGAATATCTGAACTTGCACTAGATG
K   K   A   L   N   I   P   T   S   I   K   D   A   G   V   L   E   E   N   F   Y   S   S   L   D   R   I   S   E   L   A   L   D   D

4701    .       .       .       .       .       .       .       .       4800
ATCAATGCACAGGCGCTAATCCTAGATTTCCTCTTACAAGTGAGATAAAAGAAATGTATATAAATTGTTTTAAAAAACAACCTTAAACCTTCATATTTCA
  Q   C   T   G   A   N   P   R   F   P   L   T   S   E   I   K   E   M   Y   I   N   C   F   K   K   Q   P  END

4801    .       .       .       .       .       .       .       .       4900
ACTACTTTTTATAATTTTAATAAAGAATTTAAAAGGAGGGATTAAAATGAACTCTAAAATAATTAGATTTGAAAATTTAAGGTCATTCTTTAAAGATGGG
                                          M   N   S   K   I   I   R   F   E   N   L   R   S   F   F   K   D   G
                                         CtfA

4901    .       .       .       .       .       .       .       .       5000
ATGACAATTATGATTGGAGGTTTTTTAAACTGTGGCACTCCAACCAAATTAATTGATTTTTTAGTTAATTTAAATATAAAGAATTTAACGATTATAAGTA
M   T   I   M   I   G   G   F   L   N   C   G   T   P   T   K   L   I   D   F   L   V   N   L   N   I   K   N   L   T   I   I   S   N

5001    .       .       .       .       .       .       .       .       5100
ATGATACATGTTATCCTAATACAGGTATTGGTAAGTTAATATCAAATAATCAAGTAAAAAAGCTTATTGCTTCATATATAGGCAGCAACCCAGATACTGG
  D   T   C   Y   P   N   T   G   I   G   K   L   I   S   N   N   Q   V   K   K   L   I   A   S   Y   I   G   S   N   P   D   T   G

5101    .       .       .       .       .       .       .       .       5200
CAAAAAACTTTTTAATAATGAACTTGAAGTAGAGCTCTCTCCCCAAGGAACTCTAGTGGAAAGAATACGTGCAGGCGGATCTGGCTTAGGTGGTGTACTA
 K   K   L   F   N   N   E   L   E   V   E   L   S   P   Q   G   T   L   V   E   R   I   R   A   G   G   S   G   L   G   G   V   L

5201    .       .       .       .       .       .       .       .       5300
ACTAAAACAGGTTTAGGAACTTTGATTGAAAAAGGAAAGAAAAAAATATCTATAAATGGAACGGAATATTTGTTAGAGCTACCTCTTACAGCCGATGTAG
T   K   T   G   L   G   T   L   I   E   K   G   K   K   K   I   S   I   N   G   T   E   Y   L   L   E   L   P   L   T   A   D   V   A

5301    .       .       .       .       .       .       .       .       5400
CATTAATTAAAGGTAGTATTGTAGATGAGGCCGGAAACACCTTCTATAAAGGTACTACTAAAAACTTTAATCCCTATATGGCAATGGCAGCTAAAACCGT
  L   I   K   G   S   I   V   D   E   A   G   N   T   F   Y   K   G   T   T   K   N   F   N   P   Y   M   A   M   A   A   K   T   V

5401    .       .       .       .       .       .       .       .       5500
AATAGTTGAAGCTGAAAATTTAGTTAGCTGTGAAAAACTAGAAAAGGAAAAAGCAATGACCCCCGGAGTTCTTATAAATTATATAGTAAAGGAGCCTGCA
 I   V   E   A   E   N   L   V   S   C   E   K   L   E   K   E   K   A   M   T   P   G   V   L   I   N   Y   I   V   K   E   P   A

5501    .       .       .       .       .       .       .       .       5600
TAAAATGATTAATGATAAAAACCTAGCGAAAGAAATAATAGCCAAAAGAGTTGCAAGAGAATTAAAAAATGGTCAACTTGTAAACTTAGGTGTAGGTCTT
END M   I   N   D   K   N   L   A   K   E   I   I   A   K   R   V   A   R   E   L   K   N   G   Q   L   V   N   L   G   V   G   L
    CtfB

5601    .       .       .       .       .       .       .       .       5700
CCTACCATGGTTGCAGATTATATACCAAAAAATTTCAAAATTACTTTCCAATCAGAAAACGGAATAGTTGGAATGGGCGCTAGTCCTAAAATAAATGAGG
P   T   M   V   A   D   Y   I   P   K   N   F   K   I   T   F   Q   S   E   N   G   I   V   G   M   G   A   S   P   K   I   N   E   A

5701    .       .       .       .       .       .       .       .       5800
CAGATAAAGATGTAGTAAATGCAGGAGGAGACTATACAACAGTACTTCCTGACGGCACATTTTTCGATAGCTCAGTTTCGTTTTCACTAATCCGTGGTGG
  D   K   D   V   V   N   A   G   G   D   Y   T   T   V   L   P   D   G   T   F   F   D   S   S   V   S   F   S   L   I   R   G   G

5801    .       .       .       .       .       .       .       .       5900
TCACGTAGATGTTACTGTTTTAGGGGCTCTCCAGGTAGATGAAAAGGGTAATATAGCCAATTGGATTGTTCCTGGAAAAATGCTCTCTGGTATGGGTGGA
 H   V   D   V   T   V   L   G   A   L   Q   V   D   E   K   G   N   I   A   N   W   I   V   P   G   K   M   L   S   G   M   G   G

5901    .       .       .       .       .       .       .       .       6000
GCTATGGATTTAGTAAATGGAGCTAAGAAAGTAATAATTGCAATGAGACATACAAATAAAGGTCAACCTAAAATTTTAAAAAAAATGTACACTTCCCCTCA
A   M   D   L   V   N   G   A   K   K   V   I   I   A   M   R   H   T   N   K   G   Q   P   K   I   L   K   K   C   T   L   P   L   T

6001    .       .       .       .       .       .       .       .       6100
CGGCAAAGTCTCAAGCAAATCTAATTGTAACAGAACTTGGAGTAATTGAGGTTATTAATGATGGTTTACTTCTCACTGAAATTAATAAAAACACAACCAT
  A   K   S   Q   A   N   L   I   V   T   E   L   G   V   I   E   V   I   N   D   G   L   L   L   T   E   I   N   K   N   T   T   I

6101    .       .       .       .       .       .       .       .       6200
TGATGAAATAAGGTCTTTAACTGCTGCAGATTTACTCATATCCAATGAACTTAGACCCATGGCTGTTTAGAAAGAAATACTATGAAACAATATTAAAAAA
 D   E   I   R   S   L   T   A   A   D   L   L   I   S   N   E   L   R   P   M   A   V  END

 >>>>>>>>>>>      <<<<<<<<<<<.       .       .       .       .       .       6300
ATAAGAGTTACCATTTAAGGTAACTCTTATTTTTATTACTTAAGATAATCATATATAACTTCAGCTCTAGGCAATATTATATCTGCAAGAATGTGAGAGC
                                  END Adc <———

6301    .       .       .       .       6352
TAGAAACAATCTCTTTTACTGGCAAATCATTAAGTGGCGCCATAGCGTGATC
```

FIG. 2—*Continued.*

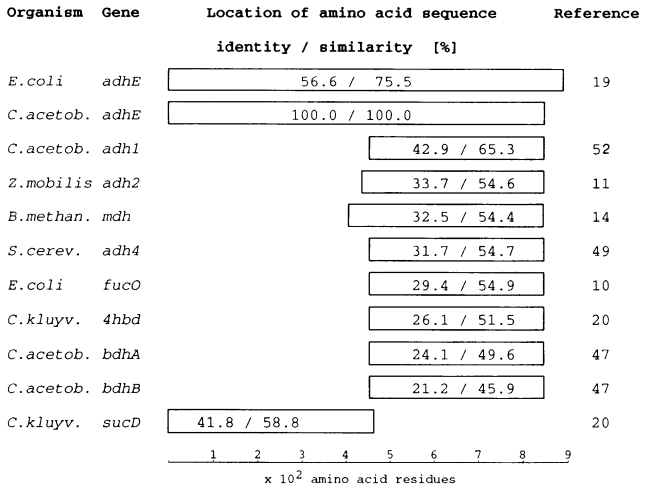

FIG. 3. Amino acid alignment (percentage of protein identities and similarities) of the *adhE* gene product compared with alcohol dehydrogenase (AdhE) of *E. coli*, alcohol dehydrogenase (Adh1) of *C. acetobutylicum*, alcohol dehydrogenase (Adh2) of *Zymomonas mobilis*, methanol dehydrogenase (Mdh) of *Bacillus methanolicus* C1, alcohol dehydrogenase (Adh4) of *Saccharomyces cerevisiae*, 1,2-propanediol oxidoreductase (FucO) of *E. coli*, 4-hydroxybutyrate dehydrogenase (4Hbd) of *C. kluyveri*, the two butanol dehydrogenase isoenzymes (BdhA and BdhB) of *C. acetobutylicum*, and succinate semialdehyde dehydrogenase (SucD) of *C. kluyveri*. The boxes indicate the domain of the *adhE* gene product of *C. acetobutylicum* where the homologies are found.

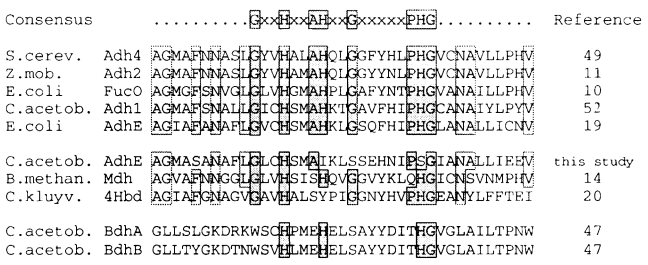

FIG. 4. Amino acid alignment of the protein regions of different alcohol dehydrogenases containing the putative iron-binding motif proposed as a typical feature of class III alcohol dehydrogenases (4). Dark-shaded amino acids indicate the core motif postulated by Bairoch (4); lighter-shaded amino acids show extended identities in at least six of the given sequences. For abbreviations, see the legend to Figure 3.

tional C found in the *ctfA* gene of strain DSM 792 at position 5465, which leads to a frameshift resulting in a 10-amino-acid longer *ctfA* gene product. Three further nucleotide exchanges were noticed: at position 5092 (A instead of T in strain ATCC 824, silent exchange), at position 5297 (G instead of A, causing the incorporation of valine instead of isoleucine), and at position 5341 (C instead of A, silent exchange). In the *ctfB* gene sequence, two differences were apparent (17, 36). At position 5564, an A was found in DSM 792 instead of a G in ATCC 824 (silent exchange); at position 5986, a G was found instead of an A. The latter represents an exchange of a cysteine for a tyrosine. However, the last three differences mentioned were found only by comparing our sequence with the data base sequence entry for these genes from strain ATCC 824 (GenBank accession number M93363). The sequence originally published (36) differed from the data bank entry at the respective positions (corresponding to nucleotides 5341, 5564, and 5986 in Fig. 2) by showing exactly the same nucleotides as given in Fig. 2 for strain DSM 792.

The gene products of *adhE*, *ctfA*, and *ctfB* as deduced from the nucleotide sequence are composed of 862, 218, and 221 amino acids, respectively, with molecular masses of 95.2, 23.6, and 23.6 kDa, respectively. The latter two values are very close to the data determined for the purified protein subunits (48). The N termini of purified CtfA and CtfB from strain ATCC 824 (9) were identical to the deduced amino acid sequences of the respective gene products from strain DSM 792.

Protein sequence comparisons. The amino acid sequence deduced from *adhE* was compared with deduced amino acid sequences from several alcohol dehydrogenase genes available in the EMBL and GenBank data bases. Figure 3 shows an alignment of AdhE from *C. acetobutylicum* to the analogous proteins from other organisms. Only the C-terminal part of the deduced protein showed significant similarity to a novel family of alcohol dehydrogenases (type III) that are distinct from the long-chain zinc-containing (type I) or short-chain zinc-lacking (type II) enzymes (14). The two recently sequenced butanol dehydrogenase isozymes of *C. acetobutylicum* are likely to form a subclass of this family (47). A 4-hydroxybutyrate dehydrogenase of *C. kluyveri* that has recently been cloned and sequenced also seems to belong to this group (20). No aldehyde dehydrogenase sequences that showed high similarity (more than 25% identity) to AdhE of *C. acetobutylicum* could be found in the data bases. However, a recently cloned and sequenced succinate semialdehyde dehydrogenase of *C. kluyveri* (20) showed 42% identity and 59% similarity to the N terminus of AdhE, thus supporting the hypothesis that the *adhE* gene product contains two different domains. The 5′ end of the gene therefore seems to encode the aldehyde dehydrogenase domain, while the 3′ end harbors the information for an alcohol dehydrogenase. Computer searches also revealed the presence of a more or less conserved iron-binding motif (4) in all mentioned type III alcohol dehydrogenases (Fig. 4). AdhE of *E. coli* has been reported to use Fe^{2+} as a cofactor for all of its catalytic activities (23). However, despite the general high similarity to this consensus sequence, AdhE of *C. acetobutylicum* lacked two of the three histidines of the iron-binding motif, leaving the question for the required metal cofactor still unanswered.

mRNA analysis. Total RNA was isolated from acidogenic and solventogenic *C. acetobutylicum* cells as well as from aerobically and anaerobically grown *E. coli*(pK9) and subsequently used for Northern blot analyses. Probes complementary to the alcohol dehydrogenase domain of *adhE* (probe ADH1, 663 bp, nucleotides 4083 to 4745), the aldehyde dehydrogenase domain of *adhE* (ACDH2, 739 bp, nucleotides 2120 to 2858), and the structural gene of ORF 5 (ORF5.1, 643 bp, nucleotides 750 to 1392) were prepared by PCR and radiolabeled as described in Materials and Methods. With both *adhE*-complementary probes, a single signal 4,100 to 4,200 nucleotides in length could be detected only in solventogenic *C. acetobutylicum* cells (Fig. 5). We were unable to show hybridization of these probes to RNA from acidogenic cells. These results are in full accordance with the results reported earlier for expression of *ctfB* (18) and provide conclusive evidence that *adhE*, *ctfA*, and *ctfB* are arranged in a common operon. RNA derived from either aerobic or anaerobic *E. coli*(pK9) cultures yielded no hybridization signals with the ADH1 and ACDH2 probes. The probe complementary to ORF 5 gave positive signals with acidogenic as well as solventogenic *C. acetobutylicum* RNA, although the hybridization

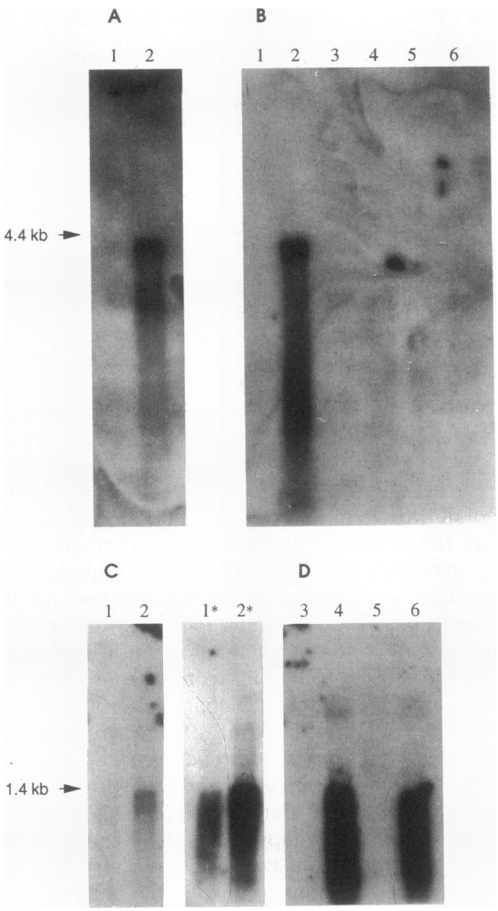

FIG. 5. Northern hybridizations using radiolabeled PCR-generated fragments which are complementary to the *adhE* gene and ORF 5. Each lane contained 10 μg of total RNA from acid-producing *C. acetobutylicum* DSM 1731 (lane 1), solvent-producing *C. acetobutylicum* DSM 1731 (lane 2), anaerobically grown *E. coli* SURE (M9 medium) (lane 3), anaerobically grown *E. coli* SURE(pK9) (M9 medium) (lane 4), aerobically grown *E. coli* SURE (LB medium) (lane 5), and aerobically grown *E. coli* SURE(pK9) (LB medium) (lane 6). Northern hybridizations were performed by using the radiolabeled probe ADH1 (663 bp, representing the region at the end of the *adhE* gene (A), the radiolabeled probe ACDH2 (739 bp, located within the first third of the *adhE* gene (B), and a PCR fragment (ORF5.1; 642 bp, complementary to ORF 5) (C and D). In panel C, lanes 1 and 2 were autoradiographed for 4 h and lanes 1* and 2* were photographed for 48 h; and in panel D, lanes 3 to 6 were autoradiographed for 4 h. All probes were generated by using sequencing primers in the PCRs. The sizes of selected marker bands (RNA ladder) are indicated at the left.

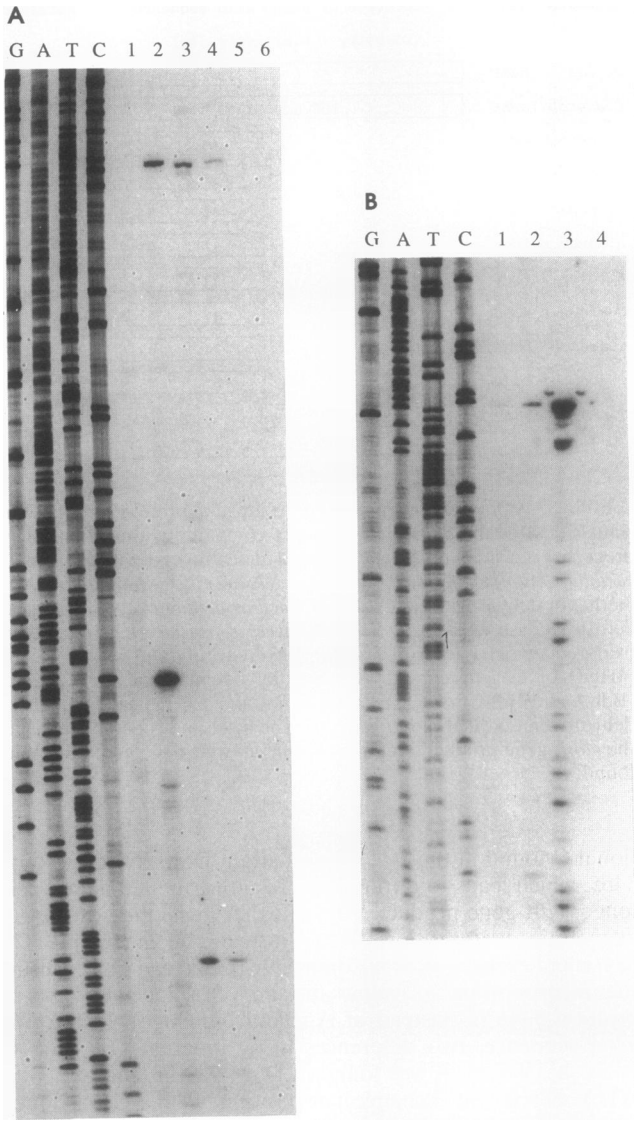

FIG. 6. Mapping of the 5′ end of the transcripts from the *sol* operon and from ORF 5 by primer extension analysis. (A) The ^{32}P-radiolabeled oligonucleotide J14PE, complementary to the mRNA of the *adhE* gene, was hybridized to 10 μg each of total RNA isolated from cells of acid-producing *C. acetobutylicum* (lane 1), solvent-producing *C. acetobutylicum* (lane 2), anaerobically (lane 3) or aerobically (lane 4) grown *E. coli* SURE(pK9), or *E. coli* SURE (lanes 5 and 6). (B) The 17-mer J21, complementary to the transcript of ORF 5, was used for primer extension analysis. The sources of RNA were acid-producing *C. acetobutylicum* cells (lane 1), solvent-producing *C. acetobutylicum* cells (lane 2), cells of anaerobically cultured *E. coli* SURE(pK9) (lane 3), and cells of anaerobically cultured *E. coli* SURE (lane 4). The sequencing ladders were obtained by using the same oligonucleotides in sequencing reactions.

was more intense with RNA from solventogenic cells, and two distinct bands were visible (Fig. 5). The lengths of these transcripts were determined to be 1.3 and 1.0 kb, which correspond to transcription termination of a monocistronic ORF 5 operon at the two stem-loop structures marked in Fig. 2. As mentioned previously, both of these structures resemble typical rho-independent transcription terminators. ORF 5 was highly expressed in the *E. coli* clone, irrespective of whether the strain was grown aerobically or anaerobically.

Determination of transcription start points. The RNA used for the Northern hybridizations mentioned above was also subjected to primer extension analysis in an effort to determine the exact start point of the transcripts. The oligonucleotides J14 and J14PE, which are complementary to sequences in the 5′ region of *adhE* (J14, nucleotides 2348 to 2364; J14PE, nucleotides 2233 to 2252) and LORF1 which is complementary to a part of ORF L (nucleotides 2093 to 2112) were used to map the start of transcription of the *sol* operon (Fig. 6). As expected from the Northern experiments, no signals could be obtained from acidogenic *C. acetobutylicum*. However, two bands of different intensity were observed with use of RNA from solventogenic cells. The major start point could be

located at nucleotide 2114, which is 84 nucleotides upstream of the ATG start codon of the *adhE* gene. A minor initiation site could be mapped at nucleotide 1952, which is 107 bp upstream of the ORF L start codon and 246 bp upstream of the beginning of *adhE*. In *E. coli*(pK9), the major signal (position 2114) could not be found. In this case, initiation started at nucleotide 1952, which has been shown to be a minor start point in *C. acetobutylicum*. A second signal could be detected in the *E. coli* clone at position 2164, directly at the base of a large stem-loop structure. It is therefore likely that this signal is an artifact caused by secondary structure inhibition of readthrough by the reverse transcriptase. We constructed two further primers that were complementary to the 5′ regions of the two antiparallel ORFs internal to *adhE* (GORF1, nucleotides 3991 to 4010; GORF2, nucleotides 3483 to 3502). No signals could be obtained in primer extension analyses with these oligonucleotides, indicating that the *adhE*-internal antiparallel ORFs are not genes, as has also been found for the *adhE* gene of *E. coli* (19).

A single start point of transcription was mapped 35 bp upstream of the ATG start codon of ORF 5, using two oligonucleotides complementary to the beginning of this ORF (J21PE, nucleotides 610 to 629; J21, nucleotides 697 to 713). As expected from the results of the Northern blots, the signal was more intense with RNA isolated from solventogenic *C. acetobutylicum* cells (Fig. 6). In *E. coli*(pK9), this position also proved to be the predominant transcription start point. Shorter transcripts of low intensity were identified, but it is likely that these represent degradation products or are the result of less specific initiation in *E. coli*. A similar phenomenon was shown for *adc* transcription initiation in *E. coli* (18).

From the thus determined transcription start points, promoter sequences could be deduced. Both putative promoters of the *sol* operon are located in regions with stem-loop structures (Fig. 2). P2, the major promoter employed during solventogenesis in *C. acetobutylicum*, showed only little homology to the consensus of gram-positive and gram-negative bacteria. A second possible structure of this promoter region is TTGTAA(21 bp)TACAAT (nucleotides 2064 to 2096, 17 bp upstream of the transcription start point). The spacing of such a promoter would be rather atypical. It would be completely located within a stem-loop structure, with the loop showing the motif TTG-TT-TTG-TT-TTG (Fig. 2). The minor transcription initiation start in *C. acetobutylicum* was controlled by the putative promoter P1. This structure showed high similarity to the consensus of gram-positive and gram-negative bacteria, thus also explaining why this site was recognized in *E. coli*(pK9). The putative promoter identified upstream of ORF 5 also showed reasonable homology to the consensus sequence (Fig. 2). This is probably a reason for the high expression of ORF 5 in *E. coli*.

DISCUSSION

In a previous study, the *ctfB* gene, encoding the large subunit of the CoA transferase, had been cloned and sequenced (17). Recent experiments showed that this gene was part of a longer operon induced or derepressed before the onset of solventogenesis (18). In this study, the remainder of this operon has been cloned and sequenced. Directly upstream of *ctfB*, the gene encoding the small subunit of CoA transferase (*ctfA*) could be found, as was expected from data reported for strain ATCC 824 (9, 36). A large gene with high similarity to an alcohol/aldehyde dehydrogenase-encoding gene of *E. coli* was detected upstream of *ctfA*. This clostridial *adhE* gene most probably encodes both alcohol and aldehyde dehydrogenase domains, as indicated by the high degree of homology to type III alcohol dehydrogenases and an aldehyde dehydrogenase from *C. kluyveri* (Fig. 3). We assume that in vivo, the *C. acetobutylicum* AdhE is active in butanol production. This hypothesis is supported by the following findings. (i) *adhE* is induced or derepressed only before the onset of butanol formation, whereas ethanol is produced constitutively throughout the fermentation (18). (ii) The *adhE* gene was found to contain an ACG codon, as did all other sequenced acetone- or butanol-forming enzymes (acetoacetate decarboxylase, CoA transferase, and butanol dehydrogenases A and B [17, 36, 47]). Expression of the tRNA (*thrA* gene product) recognizing this codon was hampered by insertion of Tn*916* close to the gene in *C. acetobutylicum* mutant AA2 (6, 42). This mutation rendered the mutant unable to produce acetone and butanol. No activities of the above-mentioned enzymes could be detected in this mutant (6), and thus ACG-containing genes were not expressed. This finding also indicates that no AdhE was synthesized; however, this mutant was still able to form ethanol. Activities of a specific acetaldehyde dehydrogenase and an NADPH-dependent alcohol dehydrogenase could be detected. The gene of the latter enzyme has been sequenced (52) and does not contain an ACG codon. (iii) Spontaneous butanol-negative mutants of *C. acetobutylicum* have been isolated in the presence of the suicide substrate allyl alcohol (16). In *E. coli*, such a selection has led to mutants defective in *adhE* (26). Surprisingly, all *C. acetobutylicum* mutants were still able to make ethanol and showed high butanol dehydrogenase activity. However, the strains were devoid of butyraldehyde dehydrogenase activity. An explanation for this phenomenon is now possible, assuming that the clostridial AdhE is active in butanol formation. A gene defect would lead to loss of both of its functions. However, the loss of butanol dehydrogenase activity could be compensated for by action of the two other butanol dehydrogenases A and B that are present in *C. acetobutylicum* (47).

A third activity of AdhE of *E. coli* is the deactivase function of pyruvate formate lyase (24). Several *Clostridium* species (*C. beijerinckii*, *C. butyricum*, and *C. kluyveri*) have been reported to perform a pyruvate formate lyase reaction that in contrast to other organisms was readily reversible (45). Inactivation and reactivation of the *C. butyricum* enzyme was similar to that of *E. coli* (45, 50). However, the metabolic function of the clostridial pyruvate formate lyase is to catalyze formate formation for C_1 unit synthesis in anabolism, in contrast to the *E. coli* enzyme, which mediates acetyl-CoA production needed for ATP synthesis in catabolism (45). It is not known whether *C. acetobutylicum* also contains a pyruvate formate lyase that eventually might be negatively regulated by the clostridial AdhE.

The results of the Northern experiments indicate that there is no processing of the *adhE* transcript. So far, no data have been reported on an alcohol or acetaldehyde dehydrogenase with a size of more than 95 kDa. The analogous protein in *E. coli* is even known to form spirosome-like high-molecular-weight complexes (21, 23). Butyraldehyde dehydrogenase has been purified from *C. acetobutylicum* NRRL B643 (34). The enzyme consists of two identical subunits with a molecular mass of 56 kDa. The aldehyde dehydrogenase-encoding domain of the clostridial *adhE* would permit formation of a protein of that size. However, we have no data indicating that AdhE from strain DSM 792 is processed into two proteins with either aldehyde or alcohol dehydrogenase activity. Experiments are in progress to purify the *adhE* gene product of *C. acetobutylicum* DSM 792. It could also be possible that in strain NRRL B643, the genes and gene products responsible

for butanol formation are different from those in strain DSM 792. Support for this idea comes from the observation that in NRRL B643, allyl alcohol-resistant mutants were able to make acetone and butanol but could not form ethanol (39). This is in contrast to the analogous mutants of strain DSM 792 mentioned above.

The organization of genes encoding three important enzymatic activities for solventogenesis (i.e., alcohol dehydrogenase, aldehyde dehydrogenase, and CoA transferase) in one transcription unit has led to its designation *sol* operon. Comparison of its regulatory region including the putative promoters P1 and P2 with that of the *adc* operon (18) did not reveal any striking sequence similarities. This is not too surprising, since it is known that the ratio of acetone and butanol formation can be dramatically altered (3), which is possible through differential expression of the respective genes. The positions of P1 and P2 in regions with stem-loop structures raise questions as to additional transcription factors. The finding that no transcription of the *sol* operon could be observed in the *E. coli* clone is most probably due to the fact that the RNA polymerase in this organism is unable to read through the secondary structures formed by the inverted repeats in front of *adhE*. However, as shown by the Northern blots, *C. acetobutylicum* is capable of some form of antitermination. Similar mechanisms might be responsible for expression of several heat shock genes found in this organism that have an upstream conserved inverted repeat structure (30, 31) which, however, shows no similarity to the stem-loop sequences upstream of *adhE*. The putative ORF L overlaps P2 as well as the three different inverted repeat structures. Translation of this ORF might lead to blocking of the hybridizing regions, thus preventing the formation of termination structures, thereby providing to an attenuation mechanism. So far, we do not have any evidence for the existence of a leader peptide in ORF L. It was surprising to find transcription initiation at P1 in the *E. coli* clone, as the Northern blot experiments did not reveal any transcription products. One explanation might be that only a limited number of transcripts were produced by the *E. coli* RNA polymerase which had to read through the various secondary structures. The difference in the methods used then led to signals in the primer extension analysis, whereas not enough transcription products were synthesized to be visible in the Northern blots.

ORF 5, whose function is still unknown, is organized in a monocistronic operon, according to the sequence data and mRNA analysis. This gene showed increased expression before the onset of solventogenesis, although a basal level was already observed in the acidogenic growth phase. Future experiments will be aimed at elucidation of the ORF 5 function, the regulation of the *sol* operon, and the reactions catalyzed by the *adhE* gene product at the enzymatic level.

ACKNOWLEDGMENTS

We are grateful to H. Bahl, G. Gottschalk, J. D. Santangelo, B. Söhling, and D. R. Woods for advice, critical reading of the manuscript, communicating results prior to publication, provision of plasmid pEcoR252, and stimulating discussions.

This work was supported by a grant from the Deutsche Forschungsgemeinschaft.

REFERENCES

1. **Bahl, H., W. Andersch, and G. Gottschalk.** 1982. Continuous production of acetone and butanol by *Clostridium acetobutylicum* in a two-stage phosphate limited chemostat. Eur. J. Appl. Microbiol. Biotechnol. **15:**201–205. (Erratum, **17:**73, 1983.)
2. **Bahl, H., and G. Gottschalk.** 1988. Microbial production of butanol/acetone, p. 1–30. *In* H. J. Rehm and G. Reed (ed.), Biotechnology, vol. 6b. VCH Verlagsgesellschaft mbH, Weinheim, Germany.
3. **Bahl, H., M. Gottwald, A. Kuhn, V. Rale, W. Andersch, and G. Gottschalk.** 1986. Nutritional factors affecting the ratio of solvents produced by *Clostridium acetobutylicum*. Appl. Environ. Microbiol. **52:**169–172.
4. **Bairoch, A.** 1991. PROSITE: a dictionary of sites and patterns in proteins. Nucleic Acids Res. **19**(Suppl.)**:**2241–2245.
5. **Bertram, J., and P. Dürre.** 1989. Conjugal transfer and expression of streptococcal transposons in *Clostridium acetobutylicum*. Arch. Microbiol. **151:**551–557.
6. **Bertram, J., A. Kuhn, and P. Dürre.** 1990. Tn*916*-induced mutants of *Clostridium acetobutylicum* defective in regulation of solvent formation. Arch. Microbiol. **153:**373–377.
7. **Bullock, W. O., J. M. Fernandez, and J. M. Short.** 1987. XL1-Blue: a high efficiency plasmid transforming *recA E. coli* strain with beta-galactosidase selection. BioTechniques **5:**376–378.
8. **Buluwela, L., A. Forster, T. Boehm, and T. H. Rabbitts.** 1989. A rapid procedure for colony screening using nylon filters. Nucleic Acids Res. **17:**452.
9. **Cary, J. W., D. J. Petersen, E. T. Papoutsakis, and G. N. Bennett.** 1990. Cloning and expression of *Clostridium acetobutylicum* ATCC 824 acetoacetyl-coenzyme A:acetate/butyrate:coenzyme A-transferase in *Escherichia coli*. Appl. Environ. Microbiol. **56:**1576–1583.
10. **Conway, T., and L. O. Ingram.** 1989. Similarity of *Escherichia coli* propanediol oxidoreductase (*fucO* product) and an unusual alcohol dehydrogenase from *Zymomonas mobilis* and *Saccharomyces cerevisiae*. J. Bacteriol. **171:**3754–3759.
11. **Conway, T., G. W. Sewell, Y. A. Osman, and L. O. Ingram.** 1987. Cloning and sequencing of the alcohol dehydrogenase II gene from *Zymomonas mobilis*. J. Bacteriol. **169:**2591–2597.
12. **Cummins, C. S., and J. L. Johnson.** 1971. Taxonomy of clostridia: wall composition and DNA homologies in *Clostridium butyricum* and other butyric acid-producing clostridia. J. Gen. Microbiol. **67:**33–46.
13. **Devereux, J., P. Haeberli, and O. Smithies.** 1984. A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. **12:**387–395.
14. **De Vries, G. E., N. Arfman, P. Terpstra, and L. Dijkhuizen.** 1992. Cloning, expression, and sequence analysis of the *Bacillus methanolicus* C1 methanol dehydrogenase gene. J. Bacteriol. **174:**5346–5353.
15. **Dower, W. J., J. F. Miller, and C. W. Ragsdale.** 1988. High efficiency transformation of *E. coli* by high voltage electroporation. Nucleic Acids Res. **16:**6127–6145.
16. **Dürre, P., A. Kuhn, and G. Gottschalk.** 1986. Treatment with allyl alcohol selects specifically for mutants of *Clostridium acetobutylicum* defective in butanol synthesis. FEMS Microbiol. Lett. **36:**77–81.
17. **Gerischer, U., and P. Dürre.** 1990. Cloning, sequencing, and molecular analysis of the acetoacetate decarboxylase gene region from *Clostridium acetobutylicum*. J. Bacteriol. **172:**6907–6918.
18. **Gerischer, U., and P. Dürre.** 1992. mRNA analysis of the *adc* gene region of *Clostridium acetobutylicum* during the shift to solventogenesis. J. Bacteriol. **174:**426–433.
19. **Goodlove, P. E., P. R. Cunningham, J. Parker, and D. P. Clark.** 1989. Cloning and sequence analysis of the fermentative alcohol-dehydrogenase-encoding gene of *Escherichia coli*. Gene **85:**209–214.
20. **Gottschalk, G., and B. Söhling (University of Göttingen, Göttingen, Germany).** Personal communication.
21. **Imamura, R., H. Niki, M. Kitaoka, K. Yamanaka, T. Ogura, and S. Hiraga.** 1992. Characterization of high molecular weights of complexes and polymers of cytoplasmic proteins in *Escherichia coli*. Res. Microbiol. **143:**743–753.
22. **Jones, D. T., and D. R. Woods.** 1986. Acetone-butanol fermentation revisited. Microbiol. Rev. **50:**484–524.
23. **Kessler, D., W. Herth, and J. Knappe.** 1992. Ultrastructure and pyruvate formate-lyase radical quenching property of the multienzymic AdhE protein of *Escherichia coli*. J. Biol. Chem. **267:**18073–18079.

24. **Kessler, D., I. Leibrecht, and J. Knappe.** 1991. Pyruvate-formate-lyase-deactivase and acetyl-CoA reductase activities of *Escherichia coli* reside on a polymeric protein particle encoded by *adhE*. FEBS Lett. **281:**59–63.
25. **Lin, F.-P.** 1992. Molecular characterization of three linked genes, *fixB*, β*hbd* and *adh1*, from *Clostridium acetobutylicum* P262. Ph.D. thesis. University of Cape Town, Cape Town, South Africa.
26. **Lorowitz, W., and D. Clark.** 1982. *Escherichia coli* mutants with a temperature-sensitive alcohol dehydrogenase. J. Bacteriol. **152:** 935–938.
27. **Marck, C.** 1988. 'DNA Strider': a 'C' program for the fast analysis of DNA and protein sequences on the Apple Macintosh family of computers. Nucleic Acids Res. **16:**1829–1836.
28. **Marmur, J.** 1961. A procedure for the isolation of deoxyribonucleic acid from microorganisms. J. Mol. Biol. **3:**208–218.
29. **Marusyk, R., and A. Sergeant.** 1980. A simple method for dialysis of small-volume samples. Anal. Biochem. **105:**403–404.
30. **Narberhaus, F., and H. Bahl.** 1992. Cloning, sequencing, and molecular analysis of the *groESL* operon of *Clostridium acetobutylicum*. J. Bacteriol. **174:**3282–3289.
31. **Narberhaus, F., K. Giebeler, and H. Bahl.** 1992. Molecular characterization of the *dnaK* gene region of *Clostridium acetobutylicum*, including *grpE*, *dnaJ*, and a new heat shock gene. J. Bacteriol. **174:**3290–3299.
32. **O'Brien, R. W., and J. G. Morris.** 1971. Oxygen and the growth and metabolism of *Clostridium acetobutylicum*. J. Gen. Microbiol. **68:**307–318.
33. **Oelmüller, U., N. Krüger, A. Steinbüchel, and C. G. Friedrich.** 1990. Isolation of prokaryotic RNA and detection of specific mRNA with biotinylated probes. J. Microbiol. Methods **11:**73–84.
34. **Palosaari, N. R., and P. Rogers.** 1988. Purification and properties of the inducible coenzyme A-linked butyraldehyde dehydrogenase from *Clostridium acetobutylicum*. J. Bacteriol. **170:**2971–2976.
35. **Petersen, D. J., and G. N. Bennett.** 1990. Purification of acetoacetate decarboxylase from *Clostridium acetobutylicum* ATCC 824 and cloning of the acetoacetate decarboxylase gene in *Escherichia coli*. Appl. Environ. Microbiol. **56:**3491–3498.
36. **Petersen, D. J., J. W. Cary, J. Vanderleyden, and G. N. Bennett.** 1993. Sequence and arrangement of genes encoding enzymes of the acetone-production pathway of *Clostridium acetobutylicum* ATCC 824. Gene **123:**93–97.
37. **Petersen, D. J., R. W. Welch, F. B. Rudolph, and G. N. Bennett.** 1991. Molecular cloning of an alcohol (butanol) dehydrogenase gene cluster from *Clostridium acetobutylicum* ATCC 824. J. Bacteriol. **173:**1831–1834.
38. **Rogers, P.** 1986. Genetics and biochemistry of *Clostridium* relevant to development of fermentation processes. Adv. Appl. Microbiol. **31:**1–60.
39. **Rogers, P., and N. Palosaari.** 1987. *Clostridium acetobutylicum* mutants that produce butyraldehyde and altered quantities of solvents. Appl. Environ. Microbiol. **53:**2761–2766.
40. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
41. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with the chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74:**5463–5467.
42. **Sauer, U., and P. Dürre.** 1992. Possible function of $tRNA^{Thr1}_{ACG}$ in regulation of solvent formation in *Clostridium acetobutylicum*. FEMS Microbiol. Lett. **100:**147–154.
43. **Sauer, U., and P. Dürre.** 1993. Sequence and molecular characterization of a DNA region encoding a small heat shock gene of *Clostridium acetobutylicum*. J. Bacteriol. **175:**3394–3400.
44. **Shine, J., and L. Dalgarno.** 1974. The 3′-terminal sequence of *Escherichia coli* 16S ribosomal RNA: complementarity to nonsense triplets and ribosome binding sites. Proc. Natl. Acad. Sci. USA **71:**1342–1346.
45. **Thauer, R. K., F. H. Kirchniawy, and K. A. Jungermann.** 1972. Properties and function of the pyruvate-formate-lyase reaction in *Clostridiae*. Eur. J. Biochem. **27:**282–290.
46. **Tinoco, I., Jr., P. N. Borer, B. Dengler, M. D. Levine, O. C. Uhlenbeck, D. M. Crothers, and J. Gralla.** 1973. Improved estimation of secondary structure in ribonucleic acid. Nature (London) New Biol. **246:**40–41.
47. **Walter, K. A., G. N. Bennett, and E. T. Papoutsakis.** 1992. Molecular characterization of two *Clostridium acetobutylicum* ATCC 824 butanol dehydrogenase isozyme genes. J. Bacteriol. **174:**7149–7158.
48. **Wiesenborn, D. P., F. B. Rudolph, and E. T. Papoutsakis.** 1989. Coenzyme A transferase from *Clostridium acetobutylicum* ATCC 824 and its role in the uptake of acids. Appl. Environ. Microbiol. **55:**323–329.
49. **Williamson, V. M., and C. E. Paquin.** 1987. Homology of *Saccharomyces cerevisiae* ADH4 to an iron-activated alcohol dehydrogenase from *Zymomonas mobilis*. Mol. Gen. Genet. **209:**374–381.
50. **Wood, N. P., and K. Jungermann.** 1972. Inactivation of the pyruvate formate lyase of *Clostridium butyricum*. FEBS Lett. **27:**49–52.
51. **Young, M., N. P. Minton, and W. L. Staudenbauer.** 1989. Recent advances in the genetics of the clostridia. FEMS Microbiol. Rev. **63:**301–326.
52. **Youngleson, J. S., W. A. Jones, D. T. Jones, and D. R. Woods.** 1989. Molecular analysis and nucleotide sequence of the *adh1* gene encoding an NADPH-dependent butanol dehydrogenase in the Gram-positive anaerobe *Clostridium acetobutylicum*. Gene **78:** 355–364.
53. **Youngleson, J. S., J. D. Santangelo, D. T. Jones, and D. R. Woods.** 1988. Cloning and expression of a *Clostridium acetobutylicum* alcohol dehydrogenase gene in *Escherichia coli*. Appl. Environ. Microbiol. **54:**676–682.
54. **Zappe, H., D. T. Jones, and D. R. Woods.** 1986. Cloning and expression of *Clostridium acetobutylicum* endoglucanase, cellobiase and amino acid biosynthesis genes in *Escherichia coli*. J. Gen. Microbiol. **132:**1367–1372.