Genetic characterization of a glycyl radical microcompartment used for 1,2-propanediol fermentation by uropathogenic *Escherichia coli* CFT073

Alex P. Lundin², Katie I. Stewart², Andrew M. Stewart², Taylor I. Herring², Chiranjit Chowdhury¹¹ and Thomas A. Bobik.²

²The Roy J. Carver Department of Biochemistry, Biophysics and Molecular Biology, Iowa State University, Ames, IA, USA 50011.

Running title: the Grp microcompartment of *E. coli* CFT073

²To whom correspondence should be addressed: Thomas A. Bobik, Iowa State University, Roy J. Carver Department of Biochemistry, Biophysics and Molecular Biology, 2152 Molecular Biology Building, Ames, IA, USA 50011. Tel: (515) 294-8247; Fax: (515) 294-0453; E-mail: bobik@iastate.edu.

¹¹Present Address: Amity Institute of Molecular Medicine & Stem Cell Research, Amity University Campus, Sector-125, Noida, UP-201313, INDIA

Key Words: microcompartment, carboxysome, 1,2-propanediol, *Escherichia coli* CFT073, glycyl radical diol dehydratase
Abstract

Bacterial microcompartments (MCPs) are widespread protein-based organelles composed of metabolic enzymes encapsulated within a protein shell. The function of MCPs is to optimize metabolic pathways by confining toxic and/or volatile pathway intermediates. A major class of MCPs, known as glycyl radical MCPs, has only been partially characterized. Here, we show that uropathogenic *Escherichia coli* CFT073 uses a glycyl radical MCP for 1,2-propanediol (1,2-PD) fermentation. Bioinformatic analyses identified a large gene cluster (named grp for glycyl radical propanediol) that encodes homologs of a glycyl radical diol dehydratase, other 1,2-PD catabolic enzymes and MCP shell proteins. Growth studies showed that *E. coli* CFT073 grows on 1,2-PD under anaerobic conditions but not under aerobic conditions. All nineteen grp genes were individually deleted and 8/19 were required for 1,2-PD fermentation. Electron microscopy and genetic studies showed that a bacterial MCP is involved. Bioinformatics combined with genetic analyses support a proposed pathway of 1,2-PD degradation and suggest that enzymatic cofactors are recycled internally within the Grp MCP. A two-component system (grpP and grpQ) is shown to mediate induction of the grp locus by 1,2-PD. Tests of the ECOR collection indicate that >10% of *E. coli* strains ferment 1,2-PD using a glycyl radical MCP. In contrast to other MCP systems, individual deletions of MCP shell genes (grpE, grpH and grpI) eliminated 1,2-PD catabolism suggesting significant functional differences with known MCPs. Overall, the studies presented here are the first comprehensive genetic analysis of a Grp-type MCP.

Importance.

Bacterial MCPs have a number of potential biotechnology applications and have been linked to bacterial pathogenesis, cancer and heart disease. Glycyl radical MCPs are a large but understudied class of bacterial MCPs. Here we show that uropathogenic *E. coli* CFT073 uses a glycyl radical MCP for 1,2-PD.
fermentation and conduct a comprehensive genetic analysis of the genes involved. Studies suggest significant functional differences between the glycyl radical MCP of *E. coli* CFT073 and better-studied MCPs. They also provide a foundation for building a deeper general understanding of glycyl radical MCPs in an organism where sophisticated genetic methods are available.

**Introduction**

Bacterial microcompartments (MCPs) are proteinaceous organelles produced by hundreds of species of bacteria (1-6). They are polyhedral in shape, about 100 to 150 nm in diameter and are built from tens of thousands of protein subunits of 10 to 20 different types. In the cases that have been best studied, MCPs consist of a selectively permeable protein shell that encapsulates sequentially-acting enzymes together with their substrates. The proposed function of MCPs is to optimize metabolic pathways that have toxic or volatile intermediates. Such intermediates are produced within the MCP then further metabolized into compounds that can safely exit into the cytoplasm of the cell. In the cytoplasm, MCP products are metabolized to provide the cell with energy, carbon, or carbon and energy for growth. Overall, compartmentalization of specific pathways within MCPs mitigates damage to DNA and other cytoplasmic components by toxic metabolic intermediates (7-11). It also minimizes the loss of small nonpolar intermediates (such as short chain aldehydes) that readily diffuse across the cell envelope but are retained by the protein shells of MCPs (6, 8, 12). Moreover, MCPs concentrate enzymes together with their substrates increasing reaction rates as exemplified by the carboxysome MCP which enhances CO₂ fixation (6).

Bacterial MCPs are functionally diverse and have potential importance in human health and biotechnology. Studies have indicated that MCPs are involved in ten or more metabolic processes including carbon dioxide fixation as well as the catabolism of 1,2-propanediol, ethanolamine, choline,
glycerol, rhamnose, fucose, fucoidan and aminoacetone (12-21). MCPs are associated with pathogenesis in *Salmonella* and *Listeria* (22-25) and might be linked to heart disease and cancer in humans due to their metabolic roles in the gut environment (26-29). In addition, MCPs components are being developed as nanomaterials for the production of protein-based containers for use in renewable chemicals production, drug delivery and the expression of toxic proteins (30-40).

In most cases, the genes for MCPs are found in large operons/gene clusters that are sufficient for MCP formation (1-6, 41, 42). The enzymes encoded by these clusters vary according to the MCP substrate metabolized while the proteins used to build the shells are generally conserved. The vertices of MCP shells are formed by bacterial microcompartment vertex (BMV) proteins (43). The facets of the shells are assembled from a family of conserved proteins known as bacterial microcompartment (BMC) domain proteins (44). MCP shell facets are usually composed of 3-8 different types of BMC domain proteins (5, 41). Diverse BMC domain proteins share a conserved flat hexagonal shape and tessellate into the mixed sheets that form the facets of the MCP. Functional diversification among BMC proteins is thought to determine the selective permeability of the shell and optimize MCP function in diverse host contexts.

Despite their widespread occurrence and diversity, only three types of MCPs have been studied in any detail. These include the 1,2-propanediol utilization (Pdu) MCP, the ethanolamine utilization (Eut) MCP, and the carboxysome. The Pdu and Eut MCPs mediate catabolism of 1,2-PD and ethanolamine by pathways that require coenzyme B<sub>12</sub> (4, 5, 41). The carboxysome enhances autotrophic CO<sub>2</sub> fixation by optimizing RuBisCO activity (6). A widespread but understudied group of MCPs (associated with glycyl-radical enzymes) is termed glycyl radical MCPs (3, 41). Glycyl radical MCPs associated with choline TMA lyase mediate anaerobic choline degradation in varied contexts.
organisms (45-47). Glycyl radical MCPs associated with glycyl radical diol dehydratases (GR-DDH) mediate the degradation of 1,2-PD, fucose, rhamnose and possibly glycerol by pathways that are independent of coenzyme B₁₂ (19, 48-50). Recent studies partially characterized the type 3 glycyl radical MCP loci of *Rhodopseudomonas palustris* and *Rhodobacter capsulatus* (48, 49). Results indicated that these loci mediate 1,2-PD degradation (48, 49); however, relatively limited genetic studies were performed and MCP formation was not demonstrated (48, 49). Here, we establish that *E. coli* CFT073 uses a MCP for fermentation of 1,2-PD and perform a comprehensive genetic analysis to gain insights into its functions and mechanisms. Electron microscopy, genetic and physiological studies establish that a MCP is used for 1,2-PD fermentation. Genetic and bioinformatic studies support a proposed pathway for 1,2-PD degradation, define the regulation of the system, address the distribution of glycyl radical MCPs among *E. coli*, and suggest significant functional differences between the glycyl radical MCP of *E. coli* CFT073 and better-studied MCPs. In accordance with prior work, we refer to the genes used for formation of this MCP as *grp* (glycyl radical propanediol) and the MCP as the Grp MCP (3).

### Results

**The *E. coli* CFT073 grp gene cluster.**

Bioinformatic studies conducted here and previously (41) identified a large gene cluster in *E. coli* CFT073 that appears to encode a type 3 glycyl radical MCP used for the degradation of 1,2-PD or a related substrate(s) (Fig. 1). This gene cluster (the *grp* cluster) contains 21 genes transcribed in the same direction. Sequence analyses indicate that *grp* genes encode enzymes for the metabolism of 1,2-
PD to propionate and 1-propanol (including GR-DDH), five MCP shell proteins and a number of proteins of unknown function. These analyses suggest that *E. coli* CFT073 uses a MCP to metabolize 1,2-PD as shown in Fig. 2.

Further sequence analyses showed that the *grp* cluster contains three relatively large internal intergenic regions that are ≥284 bp in length (IGR2, IGR3 and IGR4) (Fig1). Since large IGRs are uncommon in bacterial operons/gene clusters, we examined these IGRs for ORFs that might have been overlooked by prior annotation or were perhaps missed due to sequencing errors. IGR2, IGR3 and IGR4 were amplified by PCR and resequenced. The sequences of all three IGRs matched the genome sequence of *E. coli* CFT073 available through NCBI (Accession: NC_004431). To search for ORFs that may have been missed, the internal IGRs and their flanking sequences were analyzed using BlastX (51), PSI-Blast (52), GeneMark.Hmm (53) and the ExPASy translate tool. IGR2 contains a small ORF (ORFA) of 171 base pairs (56 amino acids) but this ORF is not identified as a gene by GeneMark.Hmm trained with the *E. coli* CFT073 sequences and PSI-Blast found no homologs of this ORF in the NCBI non-redundant database. Thus, bioinformatic analyses suggest the small ORF located between the *grpG* and *grpH* genes is not translated. IGR3 also contains a small ORF (ORFB) (201 base pairs, 67 amino acids) transcribed in the direction opposite to the *grp* genes. This ORF is not found by GeneMark.Hmm, but it is conserved in many strains of *E. coli* so translation of this ORF is an open question. Lastly, IGR4 contains three small ORFs. Two are unlikely to be translated based on the criteria described above (ORFC and ORFD). The third (ORFE) is adjacent to *mrtB* (hypothetical maturase/reverse transcriptase gene) and encodes a protein with homology to MrtB. However, ORFE has no start codon and resequencing verified the “UAG” stop codon of *mrtB*. This raises the possibility that the UAG stop codon of *mrtB* might be translated at some frequency or that the *mrtB* gene is nonfunctional and degenerating. These ORF searches are summarized in supplementary Table S1.
As mentioned above the *grp* gene cluster includes three IGRs ≥284 bp (Fig. 1). In addition, there are IGRs of 108, 136 and 129 bp just upstream of the *grpN*, *grpO* and *grpR* genes. These intergenic sequences are long enough to include elements that influence gene expression raising the possibility that the *grp* cluster includes more than one operon and/or is regulated in a complex manner. We also analyzed the *grp* gene cluster with RegRNA 2.0 (54) and found three predicted intrinsic terminators, two flank the *mrtA-mrtB* genes (GCCAGCAGTGCTCGCTGGC and CGCCAGTAACCTGGC) and one is located just past the end of the *grp* locus (TGCCCGGTAACATTTCTTTATCGCGCA) (Fig. 1).

We note that bioinformatic analyses showed that the *grp* locus contains two genes with homology to reverse transcriptase maturase genes (*mrtA* and *mrtB*). These genes have no known role in 1,2-PD metabolism and may be remnants of horizontal gene transfer as further considered in the discussion section. Overall, the bioinformatic analyses described above suggest that the *grp* gene cluster of *E. coli* CFT073 encodes a glycyl radical MCP used for 1,2-PD degradation by the pathway shown in Fig 2. These analyses also suggest that the *grp* locus might have complex regulation, and that it might contain genes associated with horizontal gene transfer (*mrtAB*).

**E. coli** CFT073 degrades 1,2-PD by fermentation.

The bioinformatic studies described above suggested that the *grp* cluster of *E. coli* CFT073 encodes a glycyl radical MCP used for the degradation of 1,2-PD. To test this, growth studies were conducted. *E. coli* CFT073 was unable to use 1,2-PD as a sole carbon and energy source under aerobic or anaerobic conditions with or without vitamin B$_{12}$ supplementation (GR-DDH enzymes do not require coenzyme B$_{12}$ but other pathways of 1,2-PD degradation are B$_{12}$-dependent). In contrast, under anaerobic conditions, 1,2-PD stimulated growth of *E. coli* CFT073 in medium supplemented with fumarate or small amounts of
yeast extract (0.2%) (Fig. 3). This indicated that *E. coli* CFT073 fermented 1,2-PD producing ATP that stimulated growth. It is likely that the yeast extract provided biosynthetic precursors since there are no known pathways in *E. coli* that convert intermediates of 1,2-PD degradation to biosynthetic building blocks. The fumarate used in some experiments may have provided biosynthetic precursors or served as terminal electron acceptor or both. The studies done here do not distinguish among these possibilities.

Further growth studies showed that *E. coli* CFT073 was unable to use 1,2-PD for anaerobic respiration with DMSO, TMAO or nitrate. Glycerol was used as a positive control and *E. coli* CFT073 grew well by anaerobic respiration of glycerol with nitrate, TMAO or DMSO as terminal electron acceptors. We also found that glycerol did not stimulate growth of *E. coli* CFT073 on medium supplemented with 0.2% yeast extract. Thus, overall, growth studies indicated that *E. coli* CFT073 degrades 1,2-PD by fermentation but does not ferment glycerol (a possible substrate for GR-DDH).

1,2-PD metabolism in the ECOR collection.

The ECOR collection is a set of 72 natural isolates of *E. coli* intended to represent the genetic diversity of the species (55). This collection was tested for strains capable of 1,2-PD degradation using MacConkey-1,2-PD indicator medium. On this medium, positive strains are red due to acid production from 1,2-PD. Results indicated that nine ECOR strains (12.5%) were able to degrade 1,2-PD. ECOR 24, 49, 50, 59, 61 and 62 gave clearly positive results while ECOR 35, 36 and 55 gave weak positive results. These ECOR strains did not require added vitamin B₁₂ for 1,2-PD degradation indicating that 1,2-PD degradation was B₁₂-independent (56). Prior studies found that about 20% of the ECOR collection could degrade 1,2-PD in a B₁₂-independent manner (including all the strains reported as 1,2-PD positive here) (56). Despite
numerous retests on MacConkey-1,2-PD plates, we were unable to replicate those studies. The reason why prior tests found more 1,2-PD positive strains will require further study.

The nine 1,2-PD-positive ECOR strains identified here were tested for the presence of the grp gene cluster by PCR. This was done by using a PCR primer pair that amplified across the grpKLM genes which encode three proteins characteristic of the grp gene cluster (a MCP shell protein, a protease-like protein and GR-DDH). For eight/nine 1,2-PD positive ECOR strains, an amplicon was obtained that had the expected size and was >98% identical in DNA sequence to the corresponding region from *E. coli* CFT073. The exception was ECOR 55 for which no PCR product was obtained. Thus, it is likely that ECOR strains 24, 49, 50, 59, 61, 62, 35 and 36 (and possibly 55) have a grp gene cluster that allows the B$_{12}$- independent degradation of 1,2-PD.

We also note that MacConkey-1,2-PD agar provides a simple test for bacterial 1,2-PD degradation that should prove helpful to varied studies of 1,2-PD metabolism. Even though 1,2-PD metabolism occurs only in the absence of oxygen, the MacConkey-1,2-PD plates were incubated under aerobic conditions because as colonies expand their centers become anaerobic.

*E. coli* CFT073 produces bacterial microcompartments in the presence of 1,2-PD

The presence of five MCP shell genes in the grp gene cluster of *E. coli* CFT073 suggested that a bacterial MCP might be involved in 1,2-PD degradation in this organism. To test this, *E. coli* CFT073 was grown anaerobically on LB medium with and without 1,2-PD and thin sections were prepared and viewed by transmission electron microscopy. Large protein complexes similar in size and appearance to bacterial MCPs were observed in cells grown in the presence of 1,2-PD but similar complexes were not seen in cells grown without 1,2-PD supplementation (Fig. 4). We also found that a grpP deletion did not form MCPs under conditions where MCPs were seen in wild-type *E. coli* CFT073. As described below
grpP is a histidine kinase required for induction of the grp operon. This indicated that grp gene(s) were required for MCP formation.

Genes in the grp cluster are required for 1,2-PD degradation by E. coli CFT073.

Each gene of the grp cluster was individually deleted by linear recombination of PCR products (57). Deletions were designed to leave the ribosome binding region (~20 bp) of the downstream gene intact (57). The target gene was initially replaced with a chloramphenicol or kanamycin resistance marker which was removed with the flp recombinase leaving behind an frt site as described (57). This method is designed to produce nonpolar deletions (57).

For each mutant, minimal medium and MacConkey-1,2-PD plates were used to test for 1,2-PD degradation. Growth curves on minimal medium are shown in Fig. S1. Results are summarized in Table 1. Deletions of, grpE, grpH, grpI, grpJ, grpM, grpN, grpP and grpQ genes eliminated 1,2-PD utilization in anaerobic liquid cultures and on MacConkey-1,2-PD indicator plates. Individual deletions of seven additional grp genes (grpA, grpB, grpC, grpG, grpK, grpL and grpS) partially impaired 1,2-PD utilization in anaerobic liquid cultures and on MacConkey-1,2-PD indicator plates. In contrast, individual deletions of four grp genes (grpD, grpF, grpO and grpR) had no detectable effect on 1,2-PD degradation in liquid cultures or on MacConkey plates. Similarly, a double deletion of both the mrtA are mrtB genes (which are flanked by grp genes) (Fig. 1) had no discernable effects on 1,2-PD degradation by E. coli CFT073.

These studies demonstrated that grp genes were required for 1,2-fermentation by E. coli CFT073, supported the proposed pathway shown in Fig. 2, and suggested functional differences between the Grp MCP and other better-studied MCPs as further addressed in the discussion section.

Complementation studies.
Complementation studies were performed on the twelve grp mutants that had clear phenotypes in the growth tests (ΔgrpA, ΔgrpB, ΔgrpE, ΔgrpG, ΔgrpH, ΔgrpI, ΔgrpJ, ΔgrpK, ΔgrpL, ΔgrpM, ΔgrpN, ΔgrpP, and ΔgrpQ) (Fig. S2). For these studies, we tested whether expression of the corresponding gene from plasmid pLac22 restored growth stimulation by 1,2-PD in liquid cultures. For 10/13 mutants (ΔgrpB, ΔgrpE, ΔgrpJ, ΔgrpK, ΔgrpG, ΔgrpL, ΔgrpM, ΔgrpN, ΔgrpP, and ΔgrpQ) full complementation was observed. This indicated that the ΔgrpB, ΔgrpE, ΔgrpG, ΔgrpJ, ΔgrpK, ΔgrpL, ΔgrpM, ΔgrpN, ΔgrpP, and ΔgrpQ mutants are nonpolar and that the impaired growth of each mutant on 1,2-PD was a consequence of grp mutation being tested. On the other hand, complementation of grpA (which encodes a putative alcohol dehydrogenase) was partial (Fig. S1). Two independent mutants were tested with similar results. The DNA sequence of both mutants and the grpA gene cloned into pLac22 (the complementation vector) were confirmed. Hence, the reason for partial complementation is uncertain, but could be due to poor production of the GrpA protein. Somewhat surprisingly, the grpH and grpI deletions were not complemented by the corresponding genes expressed from pLac22. Two independent alleles of each deletion were tested. DNA sequences of clones and mutants were verified. Prior studies showed that complementation of MCP shell gene mutations can be difficult due to the effects of gene dosage or chromosomal position (58). The vector used for these studies (pLac22) allows regulated gene expression by IPTG and five different IPTG levels were tested. Thus, we tentatively suggest that the functions of grpH and grpI genes are influenced by genomic position as was shown for the pduJ MCP shell gene (59).

The grp gene cluster of E. coli CFT073 is induced by 1,2-propanediol.
RT-qPCR was used to measure transcription of the *grp* operon in wild-type *E. coli* CFT073. Results showed that the *grp* operon was induced by 1,2-PD under anaerobic conditions (Fig. 5). The various *grp* genes were induced by 1,2-PD from 20 to 500-fold with the exception of *grpP* and *grpQ* which were not induced (Fig. 5). This suggests that a 1,2-PD responsive promoters is located upstream of *grpA* in IGR1.

The finding that *grpP* and *grpQ* were not induced by 1,2-PD suggests that a transcriptional terminator is present in the intergenic region upstream of *grpP*. An intrinsic terminator was detected bioinformatically (by RegRNA 2.0) 170 bp upstream of *grpP* (CGCCAGTAACTGGCG). Presumably, this terminator accounts for why 1,2-PD induced the upstream genes of the *grp* locus but not the *grpP* and *grpQ* genes. These results also suggest that a regulated promoter is located upstream of *grpR* and *grpS* since these two genes are both induced >20-fold by 1,2-PD supplementation despite the transcriptional terminator located upstream of *grpP* (Fig 1). This promoter could possibly be located in the 128 bp IGR just upstream of the *grpR* gene.

The *grpP* and *grpQ* genes are required for induction for the *grp* gene cluster

GrpP and GrpQ have sequence similarity to histidine kinases and response regulators, respectively, suggesting a possible role in regulating transcription of *grp* genes. RT-qPCR showed that deletion of either *grpP* or *grpQ* prevented induction of *grp* genes by 1,2-PD (Fig. 5) suggesting that they are a two-component regulatory system that mediates induction of the *grp* genes in response to 1,2-PD. This idea is supported by growth and complementation studies that showed *grpP* and *grpQ* were required for 1,2-PD degradation (Table1, Fig S1 and S2) as well as electron microscopy which showed *grpP* mutants were unable to form MCPs when growing on 1,2-PD (Fig. 4).
Prior bioinformatic and structural analyses placed glycyl radical MCPs into five classes (GRM1-5) (3, 41). The Grp MCP of *E. coli* CFT073 was classified as a GRM3. Previous in vitro studies of three key enzymes encoded by a GRM3 locus in *R. palustris* indicated it was used for 1,2-PD metabolism (49) which is consistent with the studies presented here that showed the *grp* locus of *E. coli* CFT073 is used for 1,2-PD fermentation. Prior studies of a GRM3 locus of *Rb. capsulatus* showed that it mediated the metabolism of 1,2-PD during anaerobic respiration on glucose-pyruvate with DMSO (48). HPLC identified propionate, 1-propanol and propionaldehyde as metabolic products and was used to show that deletion of GR DDH eliminated 1,2-PD metabolism (48). These results are consistent with the pathway proposed in Fig. 2. However, in *Rb. capsulatus* 1,2-PD metabolism had the unexpected effect of inhibiting cell growth and it was suggested that wild-type *Rb. capsulatus* might not assemble an intact MCP shell resulting in propionaldehyde toxicity (48). Thus, the function of the GRM3 in *Rb. capsulatus* is currently enigmatic. In contrast, here we showed that a GRM3 of *E. coli* CFT073 is used for the fermentation of 1,2-PD. We also expanded on prior studies in *R. palustris* and *Rb. capsulatus* by performing a comprehensive genetic analysis of the genes involved in 1,2-PD fermentation by *E. coli* CFT073 and by establishing experimentally that a MCP was involved. These results are further discussed below.

Genetic and bioinformatic analysis indicated that *E. coli* CFT073 ferments 1,2-PD by the pathway shown in Fig. 2. Bioinformatic analyses performed here and previously (41) showed the *grp* locus encodes homologs of all the pathway enzymes shown in Fig. 2 including GR DDH (GrpM) and its activating enzyme (GrpN) as well as propionaldehyde dehydrogenase (GrpJ), phosphotransacetylase (GrpB), 1-propanol dehydrogenase (GrpA) and propionate kinase (GrpG). This pathway is very similar to the 1,2-PD degradative pathway of *Salmonella* except that a GR-DDH (coenzyme B$_{12}$-independent)
replaces the coenzyme B$_{12}$-dependent DDH found in *Salmonella*. The pathway proposed in Fig. 2 is further supported by analysis of deletion mutants that removed each pathway enzyme individually. Deletions of the GrpM GR-DDH, the GrpN GR-DDH activating enzyme, and GrpJ aldehyde dehydrogenase eliminated growth stimulation by 1,2-PD consistent with their roles as pathway enzymes. Deletions of the GrpA Adh and the GrpB P$\text{ATC}$ substantially impaired 1,2-PD catabolism but did not eliminate it. Similar partial phenotypes were seen when genes encoding the analogous enzymes of the *Salmonella* $B_{12}$-dependent pathway of 1,2-PD degradation (PduQ and PduL) were deleted (60, 61). In this case, the reason for partial phenotypes was that although the *Salmonella* Pdu enzymes are redundant with housekeeping enzymes (and should have little to no phenotype) they are still needed to recycle cofactors internally within the Pdu MCP (60, 61). Hence, we propose that GrpA and GrpB are pathway enzymes that play an important role recycling of cofactors internally with the Grp MCP. Results also showed that the *grpG* Prk was mildly impaired for 1,2-PD degradation. For the Pdu pathway of *Salmonella*, deletion of the gene that encodes Prk (*pduW*) has no effect on 1,2-PD degradation (62). The *Salmonella* Prk is thought to be redundant with the housekeeping acetate kinase (which uses both acetate and propionate as substrates). It seems likely to us that GrpG is a pathway enzyme that is partially redundant with the housekeeping acetate kinase. Thus, overall, the bioinformatic and genetic analyses of *grp* genes described above support the 1,2-PD degradative pathway shown in Fig. 2 and suggests internal cofactor recycling.

Several studies indicated a bacterial MCP is used for 1,2-PD degradation by *E. coli* CFT073. Electron microscopy showed that large complexes similar in size and shape to bacterial MCPs were produced during growth of *E. coli* CFT073 on 1,2-PD but not during growth on other substrates (Fig. 4). In addition, MCPs were absent from cells where induction of the *grp* operon was prevented by deletion of the GrpP histidine kinase. Bioinformatic analyses performed here and previously (3, 41) identified five...
grp genes that encoded homologs of microcompartment shell proteins (grpE, grpH, grpI, grpK and grpS).

Hence, several independent lines of evidence indicate grp locus of E. coli CFT073 encodes are glycyll radical MCP used for 1,2-PD fermentation. Based on analogy with the 1,2-PD, ethanolamine and carboxysome MCPs, possible functions of the Grp MCP are to increase pathway flux and/or sequester propionaldehyde to prevent toxicity and carbon loss (7-11); however, alternative/additional functions are suggested by studies reported here.

The genetic analyses reported here suggested significant functional differences between the Grp MCP and other MCPs that have been studied. The grp locus of E. coli CFT073 includes five genes that encode homologs of proteins that make up the shells of MCPs (grpE, grpH, grpI, grpK and grpS).

Individual deletions of the grpH, grpI or grpE genes prevented growth stimulation by 1,2-PD. This result was unexpected. For the eut and pdu MCPs, shell gene deletions lead to temporary growth arrest (due to aldehyde toxicity) or have no effect on growth depending on culture conditions (7-9). In the case of the choline utilization (Cut) MCP, disruption of the MCP shell does not impair growth on choline under standard anaerobic growth conditions (47). The Cut MCP is proposed to protect cells from DNA damage by acetaldehyde and minimize the loss of acetaldehyde and neither condition is expected to impair growth under standard anaerobic culture conditions (47). Hence, the finding that the shell of the Grp MCP is required for fermentation 1,2-PD contrasts with studies done on other MCPs and suggests that the Grp MCP has a substantially different function than previously studied MCPs. This is of particular interest since the function of a Grp-MCP has not been determined experimentally in any organism. One interesting possibility is that the shell of the Grp MCP is required for the activity or stability of its associated 1,2-PD degradative enzymes although more work will be needed to sort this out.

We note that individual deletions of two shell genes, grpS and grpK, only partially impaired growth of E. coli CFT073 on 1,2-PD. Presumably, some shell function remains when the GrpS and GrpK shell proteins are eliminated. A similar situation is seen for the Pdu and Eut MCPs where deletion of...
minor abundance shell proteins has little effect on MCP function (8, 58). In cases where shell gene deletions lack significant phenotypes their encoded proteins are thought to have specialized roles that are required under specific environmental conditions (58).

The grp operon encodes a number of proteins of unknown/uncertain function including GrpC (putative eutJ-like chaperone), GrpD (possible flavoprotein), GrpF (heme-binding protein), GrpL (protease), GrpO (permease) MrtA and MrtB (maturase-reverse transcriptase) and GrpR (methionine adenosyltransferase). Deletion of grpC, grpD, grpF, grpO, grpR or mrtAB individually had little effect on growth of E. coli CFT073 on 1,2-PD. However, in conjunction with prior work, we can make inferences about their possible functions and suggest reasons why deletion mutants did not have apparent phenotypes. GrpR, a putative methionine adenosyltransferase, might be used to augment the cellular supply of SAM which is needed for activation/reactivation of GR-DDH. However, GrpR might only be needed when GR-DDH inactivation rates are high such as under microaerobic conditions. Similarly, the GrpD flavoprotein could be an electron donor for the GrpN GR-DDH activase/reactivase and perhaps only needed when the rate of inactivation of GR-DDH is high. Presumably, the GrpO permease is used to facilitate diffusion of 1,2-PD across the cytoplasmic membrane. However, given the structural similarities between 1,2-PD and glycerol, this protein might be redundant with the GlpF glycerol facilitator encoded outside the grp locus. In the case of the GrpC EutJ-like chaperone and the GrpF heme-binding proteins, homologs are widely found in MCP gene clusters suggesting a role in MCP assembly or function. However, deletion of the grpC and grpF genes did not result in substantial phenotypes suggesting their functions may depend on environmental conditions. Deletion of the mrtA and mrtB genes which encodes a maturase reverse transcriptase homolog (MrtAB) also had no effect on the growth of E. coli CFT073 on 1,2-PD. Since genes related to mrtAB mediate the movement of mobile genetic elements (63), the mrtAB genes may have had a role in horizontal gene transfer into E. coli CFT073 rather than a direct role in 1,2-PD metabolism. Deletion of one gene of unknown function (grpL)
significantly impaired growth of *E. coli* CFT073 on 1,2-PD but did not eliminate growth. The GrpL protein has homology to proteases. By analogy with viral capsid assembly (64), the GrpL protease might be needed for maturation of the MCP shell. This would be unprecedented among bacterial MCPs.

Studies also showed that the *grpP* and *grpQ* genes are essential for growth of *E. coli* CFT073 on 1,2-PD, and are required for induction of the *grp* locus by 1,2-PD. GrpP and GrpQ have homology to histidine kinases and response regulators, respectively. The N-terminal region of the GrpP histidine kinase is related in sequence to the PocR protein of *Salmonella* which induces the Pdu MCP in response to 1,2-PD (65, 66). In addition, recent studies showed a homologous two-component system regulates the GRM3 locus of *R. capsulatus* (48). Thus, it is likely that GrpPQ are a two-component system that induces *grp* genes in response to 1,2-PD.

Finally, to better evaluate the distribution of Grp MCPs among *E. coli* strains, we screened the ECOR collection using MacConkey agar supplemented with 1,2-PD. The ECOR collection consists of 72 *E. coli* strains that are representative of the genetic variation of the species as a whole (55). Nine of 72 ECOR strains were found to degrade 1,2-PD without added vitamin B12. PCR followed by DNA sequencing showed that eight of nine 1,2-PD positive ECOR strains carried genes characteristic of the *grp* locus including *grpM* (encodes GR-DDH), *grpK* (encodes a hexameric shell protein) and *grpL* (encodes a putative protease). Two of the 1,2-PD positive ECOR strains were isolated from patients with urinary tract infections (UTI), and seven were from feces of healthy individuals. *E. coli* CFT073 which was studied here is a uropathogenic strain. Although, 1,2-PD degradation by the vitamin B12-dependent pathway has been linked to pathogenesis in *Salmonella* and *Listeria* (22-25), to our knowledge, a link between 1,2-PD metabolism via the Grp MCP has not yet been found.
Material and Methods

Chemicals and reagents

Antibiotics were from Sigma Chemical Company (St. Louis, MO). Choice Taq Blue Master Mix was from Denville Scientific (Holliston, MA). KOD Hot Start Master Mix was from EMD Millipore (Billerica, MA). Isopropyl-β-D-1-thiogalactopyranoside (IPTG) was from Diagnostic Chemicals Limited (Charlottesville, PEI, Canada). Restriction enzymes HiFi DNA Assembly Master Mix and T4 DNA ligase were from New England Biolabs (Beverly, MA).

Bacterial strains and growth conditions

The bacterial strains used in this study are listed in Table S2. BSL2 precautions were used for handling all strains. The rich media used were lysogeny broth (LB), also known as Luria-Bertani medium (Becton, Dickinson and Company, Franklin Lakes, NJ) (67) and Terrific Broth (MP Biomedicals, Solon, OH). MacConkey-1,2-PD indicator plates were prepared using Difco MacConkey Agar Base supplemented with 1% 1,2-PD. The liquid medium used for growth curves and complementation studies was no-carbon-E (NCE) medium (68) supplemented with 1 mM MgSO₄, 50 μM ferric citrate, 125 mM NaCl, 1x BME vitamins (Sigma B6891) and one or more of the following at the concentrations indicated in the text: yeast extract, 1,2-PD, disodium fumarate (pH 7.0) and glycerol. The growth studies shown in Fig. 2 were performed using 18 by 150-mm serum tubes sealed with butyl rubber stoppers as described (47). All other grow curves were determined using a Synergy HT Microplate reader (BioTek, Winooski, VT) as described (69) with the modification that the microplate reader was placed inside an anaerobic chamber (Coy Laboratory Products, Grass Lake, MI) having an atmosphere of 90% N₂, 5% CO₂ and 5% H₂.

For growth and complementation studies, microplate cultures were inoculated to 2% with an aerobic LB overnight culture supplemented with 50 μg/mL ampicillin as appropriate.
Construction of chromosomal mutations

Chromosomal deletions were made by linear recombination of PCR products as described (57, 69). Kanamycin resistance was selected and the resistance markers were removed using the Flp recombinase (57). This leaves behind an 82-nucleotide scar (an frt site). All mutations were verified by PCR amplification across the scar followed by DNA sequencing.

Electron microscopy

For each strain, an aerobic overnight culture was prepared on LB medium supplemented with 10 mM MgSO₄. The overnight culture was placed in the anaerobic chamber for ~1 h then 30 μL was used to inoculate 3 mL of double strength LB supplemented with 10 mM MgSO₄, 50 μM Fe Citrate and 1% 1,2-PD as appropriate. The resulting culture was incubated for 8 hours at 37 °C, statically, in the anaerobic chamber. 200 μL was used to inoculate 20 mL of double strength LB supplemented with 10 mM MgSO₄, 50 μM Fe Citrate and 1% 1,2-PD as needed. These cultures were incubated under anaerobic conditions, statically at 37 °C to an OD600 of 1.2 to 1.5 (~16 hours). Cultures were pipetted into 50 mL polycarbonate centrifuge tubes and tightly sealed with screw caps having rubber o-rings. Cells were centrifuged at 5,000 x g for 5 min then returned to the anaerobic chamber. The supernatant was decanted to a waste container. 1 mL 2% glutaraldehyde was added and the suspension was mixed by pipetting up and down. Cells were transferred to a 1.5 mL microcentrifuge tube, removed from the anaerobic chamber and incubated at room temperature for 1 h. Cells were pelleted by centrifugation for 5 min at 5,000 x g. The supernatant was decanted. 1 mL of 2% glutaraldehyde was added followed by pipetting up and down to mix. The suspension was incubated at room temperature overnight then washed by pelleting three times and resuspending in 0.1 M cacodylate buffer, pH 7.2. From this point, imbedding, sectioning and electron microscopy were carried out as described (69).
Molecular Biology Methods.

Agarose gel electrophoresis, plasmid purification, PCR, restriction digests, ligation reactions and electroporation were carried out using standard protocols as described (13, 70). Taq or Phusion DNA polymerase (New England Biolabs) were used for amplification of chromosomal DNA and colony PCR. KOD DNA polymerase was used for amplification of plasmid templates. Plasmid DNA was purified using Qiagen products (Qiagen, Chatsworth, CA) according to the manufacturer’s instructions. Following restriction digestion or PCR amplification, DNA was purified using Promega Wizard PCR Preps (Madison, WI) or Qiagen gel extraction kits. For ligation of DNA fragments, T4 DNA ligase or NEBuilder HiFi DNA Assembly Master Mix was used according to the manufacturer’s instructions (New England Biolabs).

Complementation studies

Genes used for complementation were cloned into plasmid pLac22 which allows for tight regulation of cloned genes by IPTG (71). Chromosomal genes were amplified using Phusion or Taq DNA polymerase and cloned between the BglII and HinDIII sites of pLac22 by Gibson Assembly. PCR primers with 20 bp of homology to the vector were used to make inserts for Gibson Assembly which was done with retention of the BglII site to provide correct placement of the ribosome binding site. Primers for Gibson Assembly were designed using SnapGene software (GSL biotech). Positive clones were identified by colony PCR and the DNA sequence of all clones was verified. Growth curves for testing complementation were done in 48-well microplates as described above using IPTG at 0, 10, 20, 50, 100, and 200μM.

Amplification and sequencing of the grpKLM genes of ECOR strains.
The *grpKLM* genes of selected ECOR strains were amplified by colony PCR using Choice Taq Blue Master Mix and primers TGCCGGTTCCGGATCTCTAT and TAATCGCGGGCAGTTGTTCT which were also used for sequencing of the PCR products.

**RT-qPCR for measuring grp expression**

Cells from *E. coli* strain CFT073 were grown overnight in LB media anaerobically in the presence or absence of 1,2-PD. A fresh 20 mL flask of LB was inoculated at 1% and grown anaerobically for 3.5 h with 1,2-PD at 37 °C then harvested. Total RNA was extracted using the RNeasy Protect Bacteria Mini Kit (Qiagen), and converted to cDNA with EcoDry Premix (Takara) following manufacturer’s instructions. Genomic DNA was cleaned up using SYBR Green PCR Master Mix (Applied Biosystems). Samples of 50 µL volume were prepared in MicroAmp Optical Strips (Applied Biosystems) with three replicates per gene of interest. Three replicates of housekeeping gene GapA were collected in parallel as an internal standard. Samples were analyzed using a StepOnePlus RT-PCR instrument (ThermoFisher) using the Comparative ΔΔCt experimental protocol, which was comprised of 40 cycles of 15 s at 95 °C followed by 60 s at 60 °C, with a final melt curve step performed under the same conditions. An identical procedure was utilized for ΔgrpP and ΔgrpQ strains. Data was analyzed using the ΔΔCt method and standard error was reported for each gene of interest.
Acknowledgements

This work was supported by grant AI081146 from the National Institutes of Health to TAB. We thank Dr. Harry L.T. Mobley and Stephanie Himpsl for kindly providing *E. coli* CFT073. We also thank the ISU DNA Sequencing and Synthesis Facility for assistance with DNA analyses and the ISU Microscopy and Nanoimaging facility for help with electron microscopy.
### Table 1. Growth of grp deletions on 1,2-PD.

<table>
<thead>
<tr>
<th>gene</th>
<th>Closest homolog of deleted gene</th>
<th>$t_1^{1,2}$ PD utilization</th>
<th>protein product (accession)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ΔgrpA</td>
<td>Alcohol dehydrogenase</td>
<td>+--</td>
<td>WP_000135343.1</td>
</tr>
<tr>
<td>ΔgrpB</td>
<td>Phosphotransacylase</td>
<td>+--</td>
<td>WP_001109982.1</td>
</tr>
<tr>
<td>ΔgrpC</td>
<td>EutJ-like chaperone of unknown function</td>
<td>++-</td>
<td>WP_001332070.1</td>
</tr>
<tr>
<td>ΔgrpD</td>
<td>Hypothetical flavoprotein</td>
<td>+++</td>
<td>WP_001329683.1</td>
</tr>
<tr>
<td>ΔgrpE</td>
<td>EutN MCP vertex protein</td>
<td>-</td>
<td>WP_000280292.1</td>
</tr>
<tr>
<td>ΔgrpF</td>
<td>Heme-binding protein of unknown function</td>
<td>+++</td>
<td>WP_001301167.1</td>
</tr>
<tr>
<td>ΔgrpG</td>
<td>Propionate kinase</td>
<td>++-</td>
<td>WP_001362987.1</td>
</tr>
<tr>
<td>ΔgrpH</td>
<td>Shell hexamer</td>
<td>-</td>
<td>WP_0013011171.1</td>
</tr>
<tr>
<td>ΔgrpI</td>
<td>Shell trimer (stacked/gated)</td>
<td>-</td>
<td>WP_001205499.1</td>
</tr>
<tr>
<td>ΔgrpJ</td>
<td>Aldehyde dehydrogenase</td>
<td>-</td>
<td>WP_000997839.1</td>
</tr>
<tr>
<td>ΔgrpK</td>
<td>Shell hexamer with FeS cluster</td>
<td>+/-</td>
<td>WP_000746009.1</td>
</tr>
<tr>
<td>ΔgrpL</td>
<td>Putative protease of unknown function</td>
<td>+/-</td>
<td>WP_000722279.1</td>
</tr>
<tr>
<td>ΔgrpM</td>
<td>glycyl radical diol dehydratase (GR-DDH)</td>
<td>-</td>
<td>WP_000890295.1</td>
</tr>
<tr>
<td>ΔgrpN</td>
<td>grpM activating enzyme</td>
<td>-</td>
<td>WP_001044978.1</td>
</tr>
<tr>
<td>ΔgrpO</td>
<td>Permease</td>
<td>+++</td>
<td>WP_001086421.1</td>
</tr>
<tr>
<td>ΔmtrA</td>
<td>Maturase, reverse transcriptase</td>
<td>+++</td>
<td>AAN82974.1</td>
</tr>
<tr>
<td>ΔmtrB</td>
<td>Maturase, reverse transcriptase</td>
<td>+++</td>
<td>WP_000237741.1</td>
</tr>
<tr>
<td>ΔgrpP</td>
<td>Histidine kinase</td>
<td>-</td>
<td>AAN82979.1</td>
</tr>
<tr>
<td>ΔgrpQ</td>
<td>DNA-binding response regulator</td>
<td>-</td>
<td>WP_000494233.1</td>
</tr>
<tr>
<td>ΔgrpR</td>
<td>Methionine adenosyltransferase</td>
<td>+++</td>
<td>WP_096489982.1</td>
</tr>
<tr>
<td>ΔgrpS</td>
<td>Shell hexamer with C-terminal extension</td>
<td>++</td>
<td>WP_000015571.1</td>
</tr>
<tr>
<td>ΔmrtAB</td>
<td>maturease/reverse transcriptase</td>
<td>+++</td>
<td></td>
</tr>
</tbody>
</table>

†WT = Wild type E. coli CFT073

‡Qualitative estimate of growth based on the growth curves show in Fig. S1
Figure 1. The **grp** gene cluster of *E. coli* CFT073. The **grp** gene cluster of *E. coli* CFT073 includes 21 genes transcribed in the same direction. General gene functions are indicated by color. The green colored genes encode all the enzymes needed for the conversion 1,2-PD to 1-propanol and propionate as shown in Fig. 2 including a glycyl radical diol dehydratase (**grpM**). The numbers beneath genes indicate the size of intergenic regions (IGRs) in base pairs. The largest internal IGRs are numbered 2-4. The graphical stem-loop structures indicate the locations of possible intrinsic terminators predicted by RegRNA 2.0 software.

Figure 2. Model for 1,2-PD degradation by the Grp-MCP of *E. coli* CFT073. The Grp MCP allows *E. coli* CFT073 to grow by 1,2-PD fermentation. 1,2-PD diffuses through the protein shell of the MCP and enters the lumen where it is converted to propionaldehyde by a glycyl radical diol dehydratase (**grpM**) supported by its activating enzyme (**grpN**). Propionaldehyde is further metabolized to 1-propanol and propionate by the **grpJ** aldehyde dehydrogenase, the **grpA** alcohol dehydrogenase, the **grpB** phosphotransacetylase, and the **grpG** propionate kinase. This pathway generates ATP used to support growth. Based on analogy with other MCPs, its likely function is to sequester propionaldehyde to prevent toxicity and/or diffusive loss through the cell envelope. Also based on analogy, the **grpJ** aldehyde dehydrogenase and the **grpA** 1-propanol dehydrogenase are likely required for recycling HS-CoA and NAD⁺, respectively, internally within the MCP.
Figure 3. Growth response of *E. coli* CFT073 to 1,2-propanediol. 1,2-propanediol stimulated growth of *E. coli* CFT073 under anaerobic conditions on minimal medium containing (A) 0.2% yeast extract or (B) 20 mM fumarate. Cells were grown anaerobically in sealed serum tubes. Cell growth was measured by monitoring absorbance at 600 nm (OD600 nm) over time.

Figure 4. Electron microscopy of *E. coli* CFT073 grown in the presence and absence of 1,2-PD. (A) *E. coli* CFT073 grown in the absence of 1,2-PD. (B) Grown in the presence of 1,2-PD. (C) Grown in the presence of 1,2-PD (higher magnification). (D) A grpP mutant (unable to express grp genes) grown in the presence of 1,2-PD. White triangles point to bacterial MCPs.

Figure 5: Regulation of the grp genes of *E. coli* CFT073. RT-qPCR was used to measure the relative fold induction of each grp gene by 1,2-PD in wild-type CFT073 *E. coli* and in mutants where grpP or grpQ were deleted. Results shown are the average of three replicates with induction calculated using the ΔΔCt method with the gapA gene as the standard. The error bars represent one standard deviation. The growth conditions and RT-qPCR methods are described in the experimental procedures section.


736