Genome-wide Transcriptional Response during the Shift to N₂-fixing Conditions in Heliobacterium modesticaldum
Daniel Sheehy1,2,3, Yih-Kuang Lu1, Fawsia Osman1, Zana Alattar1, Catalina Flores1, Hallie Sussman1, Sabha Zaare1, Maria Dooling1, Amir Meraban1, Patricia Baker1,3, Jeffrey W. Touchman2 and Kevin E. Redding1,3 *

1School of Molecular Sciences, Arizona State University, Tempe, AZ, USA
2School of Life Sciences, Arizona State University, Tempe, AZ, USA
3Center for Bioenergy & Photosynthesis, Arizona State University, Tempe, AZ, USA
4School of Arts and Sciences, Boston University, Boston, MA, USA

Abstract
Heliobacteria are the only known phototrophic Firmicute; all known members of this group appear to be capable of N₂ fixation but incapable of CO₂ fixation. They are anaerogenic and possess the simplest photosynthetic apparatus known. The sequence of the 3.1-Mb genome of Heliobacterium modesticaldum, a moderate thermophile within the family Heliobacteriaceae, is publicly available. The focus of this study is to understand how this organism operates at a fundamental level by examining changes in its transcriptome during a shift from ammonium-containing medium to N₂-fixing conditions. RNA was purified from cells grown with pyruvate as the carbon source and ammonia or N₂ as the nitrogen source. After RNA depletion, the RNA pool was sequenced using the Ion Torrent PGM platform. We found that the nitrogenase gene cluster was only expressed under N₂-fixing conditions, concomitant with increased expression of the high-affinity ammonium transporter. Most genes were down-regulated in N₂-fixing conditions by a factor of at least three. A drastic down-regulation of the highly expressed genes encoding proteins involved in the cyclic electron transport chain also occurred. The photosynthetic psha transcript also decreased more than 100-fold but subsequent photochemical analysis demonstrated no large drop in the concentration of the reaction center protein complex. This indicates that there is a role for substantial translational regulation in some genes. The transcriptomic analyses revealed a network of differentially expressed genes in H. modesticaldum. This study represents the first step in the creation of a quantitative genome-scale metabolic model establishing H. modesticaldum as a model organism for the Heliobacteriaceae family.

Keywords: Heliobacteria; Next generation sequencing; Transcriptome

Introduction
Originally discovered in 1983 [1], heliobacteria have been a source of great interest, as they are the only known phototrophic group within the Firmicutes [2]. Heliobacteria also possess the simplest known photosynthetic apparatus. Comprised of a type I homodimeric reaction center (RC) in the cytoplasmic membrane, this photosynthetic apparatus uses a unique pigment, bacteriochlorophyll g (BChl g) [3]. Unlike other phototrophs, heliobacteria lack a peripheral antenna system. They are not autotrophic and must obtain carbon from organic molecules [4,5], but are capable of robust nitrogen fixation [6]. Like many other bacteria in the order Clostridiales, heliobacteria can grow fermentatively in the dark using simple organic acids, such as pyruvate [7]. In the light, they use the same carbon sources and grow at a significantly faster rate (“phototertrophic growth”). These characteristics identify heliobacteria as distinct among the seven photosynthetic bacterial lineages. Several heliobacterial species have been discovered, but only the genome sequence of H. modesticaldum has been reported [8]. H. modesticaldum is an obligate anaerobe and is moderately thermophilic with an optimal growth temperature of 50-52°C [9]. The genome of H. modesticaldum is a single 3.1-Mb circular chromosome possessing 3,138 putative open reading frames (ORFs) [8]. The photosynthetic gene cluster is contained in a single operon, encoding the HbRC core polypeptide (PshA) and pigment biosynthetic enzymes (e.g. BchBNL). The order of these genes is the same as in the cluster in Heliobacillus mobilis [10]. While many enzymes involved in the central pathway of carbohydrate metabolism were predicted, it was recently reported that H. modesticaldum has an incomplete reductive TCA cycle lacking the key enzyme ATP-citrate lyase [11], as was also observed through 13C-labeling [7]. Like all heliobacteria, H. modesticaldum is capable of fixing nitrogen, even at the elevated temperatures under which it grows [9]. Nitrogen fixation is catalyzed by the nitrogenase enzyme complex, and results in the reduction of atmospheric dinitrogen (N₂) to ammonium (NH₄⁺) and the production of molecular hydrogen [12]. This process requires large amounts of chemical energy (16 ATP) and reducing power (8 Fe(II)) to convert one N₂ to two molecules of NH₄⁺ [13], which is then assimilated into many biomolecules. Sequence similarity predicts the use of a Mo-Fe group I nitrogenase consisting of a homodimer of NiFe/D polypeptides [14]. The primary pathway for NH₄⁺ assimilation in heliobacteria is the glutamine synthetase/glutamate synthase pathway [15]. This pathway is essential for growth, because glutamine is the primary intracellular nitrogen donor for purine and pyrimidine synthesis. Both ATP and reducing power are required in carbon metabolism, nitrogen assimilation, and hydrogen production, inextricably linking these pathways. In H. modesticaldum, the high-energy demand required for nitrogen fixation during diazotrophic growth has resulted in strict regulation of the nif genes encoding for nitrogenase. Thus, the addition of NH₄⁺ to cultures
results in repression of nitrogenase activity [15,16]. However, the effects of N₂-fixing versus non-fixing conditions on genome-wide expression and metabolism in Heliobacterium remained to be determined. The sequencing of the genome of H. modesticaldum, along with recent proteomic studies, have provided insights into the energy metabolism of this phototroph. However, metabolic studies have been correlated to only a few genes related to energy and carbon metabolism at the transcriptomic level [17]. To enhance our understanding of the energy metabolism of H. modesticaldum, it is necessary to explore the entire mRNA link between genome and proteome. We report here a single-nucleotide resolution map of the H. modesticaldum Ice1 transcriptome under N₂-fixing and non-fixing conditions. In general, we observed low-level repression of transcription genome-wide upon a shift to N₂-fixing conditions. On the contrary, a few genes, such as the genes involved in N₂-fixation and ammonium scavenging, were upregulated. There were also several cases of even more drastic down-regulation, including the core genes of cyclic photophosphorylation.

Materials and Methods

Growth of H. modesticaldum

Isolated colonies of H. modesticaldum strain Ice1 were cultured in gel-rite media modified by Lin and Casida (1984) for thermophilic application [18] inside an anaerobic Corry glove box at 52°C under infrared lights at 780 nm. Cells were inoculated in Pyruvate-Yeast Extract (PYE) growth media [16], which contains 1 g/L of NH₄SO₄ as the source of nitrogen and "vitamin levels" of yeast extract (0.02%). The PYE-NH₄⁺ medium was made by elimination of NH₄SO₄ and increasing the amount of Na₂S₂O₃•5H₂O from 0.2 g to 0.4 g. The pH was adjusted to 6.8 with H₂SO₄ prior to autoclaving. Growth was monitored spectroscopically at an Optical Density (OD) at 625-nm, as minimal photosynthetic pigments absorb at this wavelength in heliobacteria [3,17]. Cells were grown under anaerobic conditions at 52°C. From PYE or PYE-NH₄⁺ conditions, 3 ml of cells in late exponential growth phase were inoculated into 300 ml of similar media. This ensured no traces of ammonia were carried over to the next generation of cells. Extraction was performed when cells were in mid-log phase of growth at an OD of 0.303 (PYE-NH₄⁺) and 0.410 (PYE). A biological replicate for each condition was also prepared.

Isolation and purification of mRNA

Total RNA was isolated using the Purelink total RNA isolation kit (Invitrogen, USA). Cells were anaerobically extracted and lysed using the needle homogenization method offered in the kit protocol. Once total RNA was extracted the solutions were treated with DNasel via the Ambion RiboPure Bacteria procedure to remove all genomic DNA. Depletion of the 16S and 23S rRNAs was performed via the Ambion MICROBEXpress kit through subtractive hybridization with capture oligonucleotides. The resulting solution contained tRNAs, 55 rRNAs, and enriched mRNA.

Library preparation and Ion Torrent sequencing

Library preparation was performed using the Ion Total RNA-Seq Kit (Ambion, USA). Both sample and WT control RNAs (1 µg/µL Hela total RNA) were fragmented using RNase III at 37°C for 10 minutes. The fragmented RNA was purified via Ambion’s RiboMinus™ Concentration Module. The resulting mRNA yield was measured on the Bioanalyzer (Agilent, USA). These enriched mRNA samples were hybridized and ligated with Ion Torrent adaptors. Strand specificity was retained using the Ion Adaptor Mix containing oligonucleotides with a single-stranded degenerate sequence at the 3’ end and a defined sequence at the 5’ end. This effectively constrains the RNA orientation, with sequencing only performed from the 5’ end of the sense strand. Reverse transcription was performed and the cDNA was purified and size selected using Agencourt’s AMPure® XP reagent for an optimal fragment size between 30-200 bp. The cDNA samples were amplified using the provided Ion 5’ and 3’ PCR Primers for 16 cycles. Fragment size was verified on the Bioanalyzer (Figure S1). The prepared cDNA libraries were loaded on the chip and sequenced using the Ion PGM sequencer per the manufacturer’s instructions (Life Technologies, USA).

Bioinformatic analysis of RNA-seq data

The sequenced reads captured on the Ion Torrent platform were analyzed by following procedure described here. Initially, the sequencing reads were subject to filtering out of those with the poor quality and more than 50% of Ns in reads using PRINSEQ [19]. Bowtie2 [20] with default parameters was used to align filtered RNA-seq reads against H. modesticaldum chromosome as the reference genome. The resulting sequence alignment files were imported into Partek Genomics Suite (Partek Inc., St. Louis, MO) to compute raw and fragments per kilobase of exon model per million mapped (RPKM) reads for the normalized expression values of each transcript. A stringent filtering criterion with RPKM value of 1.0 [21] was used to obtain expressed transcripts. The RPKM values of filtered transcripts were log-transformed using log2 (RPKM + offset) with an offset value of 1.0, and fold changes in transcript expression, differential expression, and p-values were generated from these using the Partek software with default settings. Metabolic pathways that were significantly enriched between +/-NH₄⁺ were identified using Cytoscape [22] with the reference pathways of H. modesticaldum from the Kyoto Encyclopedia of Genes and Genomes (KEGG) database [23].

Real-time qPCR

A 1 µg concentration of non-fragmented mRNA was added to an iScript cDNA Synthesis Kit (Bio-Rad, USA) and reverse transcription was performed. Primers were designed and synthesized for qPCR using the SsoFast EvaGreen Supermix kit (Bio-Rad, USA). Enzyme activation was performed at 95°C for 1 minute, followed by 40 cycles of denaturation at 95°C for 15 seconds and annealing/extension at 57°C for 30 seconds. The Ct values were determined based on the PCR cycle number that crossed an arbitrarily defined threshold. Gene expression was determined using the 2⁻ΔΔCt method [24]. Three technical replicates were performed and the mean values were reported (Table S1). Primers used are listed in Table S2.

Reaction center measurements

Cells were grown in PYE and PYE-NH₄⁺ to mid-log phase of growth. Aliquots from both cultures were diluted to OD₅₅₀ = 0.30 to equalize cell density between the two conditions. Cells were transferred into a 1 cm anaerobic cuvette and brought to room temperature before analysis with a JTS-10 LED spectrometer (Bio-Logic; Claiix, France). Absorption from P₇₃₀, was monitored using 10 µs pulses of light from a 810 nm LED; 780 nm high-pass filters were placed in front of the detectors to block actinic light provided by a saturating flash (31 mJ) from a frequency-doubled Nd/YAG laser [25]. The bleaching values were measured six times per sample with a 90-s dark adaptation and cuvette inversion between measurements. After the photobleaching measurements, an acetone extraction was performed on both samples anaerobically; the supernatant was measured for BChl concentration and the pellet was resuspended in 50 µl of 50 mM MOPS (pH 7). A bicinchoninic acid assay was performed on the pellet and the
revealed several interesting features regarding The inherent strand specificity of the Ion Torrent methodology an average coverage of 24- and 22-fold per nucleotide, respectively. sets resulted in a single-nucleotide resolution transcriptome with H. modesticaldum were mapped onto the genome of ammonium-replete and ammonium-deplete cultures, respectively, subject to the alignment. Reads totaling 74 and 69 Mbps from the coverage). Further, we filtered the sequencing reads with poor quality resulted in 3.71 million reads and 345 Mbp (>112-fold genome as described above, the sequenced reads of the PYE-NH₄ culture resulted in 3.71 million reads and 345 Mbp (>112-fold genome coverage). Further, we filtered the sequencing reads with poor quality or more than 50% ambiguous bases, and the remaining reads were subject to the alignment. Reads totaling 74 and 69 Mbps from the ammonium-replete and ammonium-deplete cultures, respectively, were mapped onto the genome of H. modesticaldum. Both data sets resulted in a single-nucleotide resolution transcriptome with an average coverage of 24- and 22-fold per nucleotide, respectively. The inherent strand specificity of the Ion Torrent methodology revealed several interesting features regarding H. modesticaldum’s transcriptome. First, one of the genome strands encodes two thirds of the protein coding sequences while the opposite strand encodes the remaining coding sequences (Figure S2A). This observation is consistent with the location of open reading frames on the sequenced genome [8].

Additionally, the strand-specific sequencing revealed non-coding RNAs, which appear to include cis-encoded antisense RNAs. These are RNAs that both overlap and are complementary to the sense strand. Examples of antisense RNAs in H. modesticaldum’s transcriptome can be seen throughout the NiFe hydrogenase cluster as both cis (red) and complementary to the sense strand (green) (Figure S3). The function of antisense RNAs in bacteria is increasingly being recognized as an important regulator of metabolic and physiological processes [27]. In the cyanobacterium Synechocystis sp. PCC 6803, there are antisense RNAs for approximately 10% of the total protein-coding genes, indicating that regulation by antisense RNA may be a significant form of gene regulation in phototrophic bacteria [28]. We have annotated 775 potential antisense transcripts using an arbitrary cut-off of ≥ 10 in coverage (Data Set S1), indicating that the use of antisense RNA could be as high (or even higher) in heliobacteria.

Another feature of the baseline transcriptome is several regions of high intergenic transcription, which may indicate non-annotated genes. For instance, the region between 1352.5-1358.5 Kbps contains a high level of intergenic transcription, and nBLAST attributes a 99 percent identity to 16S rRNA for both the non-annotated transcripts and the intervening “hypothetical gene” HM1_3154 (Figure S2B). Table S1 lists 53 transcripts that map to intergenic regions (i.e. not overlapping with genes on either strand) and have coverage ≥ 10 and minimal length of 50 bases. BLAST searches were performed on all peptides potentially encoded by this region, but the vast majority did not have significant matches. Three of them are potential pseudogenes, with significant matches to peptides 50-55 residues long. One transcript, however, has 58% identity (73% similarity) to the ParA chromosome partitioning protein of Paenibacillus polymyxa over the entire length of 282 residues, and thus very likely represents a gene that was missed in the primary genome annotation (Table S1).

The ammonium-deplete transcriptome

With the general features of H. modesticaldum’s transcriptome established, it becomes possible to quantify the expression of the genome at the RNA level and how this changes in response to an environmental change. We repeated the RNA-seq experiment with cells deprived of ammonium, thus requiring them to fix N₂, as seen in Figure 1, and cells were harvested for RNA extraction at an optical density (OD) of 625 nm at 0.410 (PYE) and 0.303 (PYE-NH₄). In both cases, cells were grown phototrophically using pyruvate as the carbon source. We determined that the cells were dependent upon the N₂ in the headspace for their nitrogen needs by also inoculating a PYE-NH₄ culture in a bottle that had been purged with argon to remove all N₂. This culture stopped growing very soon after inoculation and did not attain a density more than 25% that of the PYE-NH₄ culture that had a headspace of 97% N₂ (data not shown).

Figure 1: Growth of H. modesticaldum with and without ammonium. Cells grown in pyruvate yeast extract with NH₄⁺ (black) and without NH₄⁺ (red). The Optical Density (OD) at 625 nm was monitored and RNA was extracted from cells once the cells reached an OD of 0.4 (PYE) or 0.3 (PYE-NH₄⁺), as indicated by the arrows.

Results

The baseline transcriptome

In order to be able to correlate the transcriptomic data with previous proteomic studies, it was important to use comparable growth conditions. Therefore, we chose phototrophic growth in Pyruvate Yeast Extract (PYE) medium as the baseline condition, as it is the typical growth mode used in most studies on this organism. The presence of yeast extract in this media was shown to have no significant effect on the expression levels of genes involved in metabolism and other cellular functions [17]. Following cellular extraction during mid-log phase growth, depletion of 16S and 23S rRNA was performed. This procedure reduces the sequencing bias towards highly expressed genes, as it was shown that 98% of the transcriptome consisted of RNA in three different bacterial species when total RNA was used for transcriptome analysis [26]. These highly expressed transcripts reduce the coverage of mRNA reads and undermine transcriptome analysis accuracy. Electrophoretic analysis of the RNA pools indicated that the depletion was largely successful (Figure S1 of Supplemental Material). After rRNA depletion, cDNA libraries of three biological replicates were generated and sequenced on the Ion Torrent. The sequencing reads of the PYE culture were pooled together, resulting in 3.75 million reads and 256 Mbp (>83-fold sequencing depth of the genome). Following the same procedure as described above, the sequenced reads of the PYE-NH₄ culture resulted in 3.71 million reads and 345 Mbp (>112-fold genome coverage). For instance, the region between 1352.5-1358.5 Kbps contains a high level of intergenic transcription, and nBLAST attributes a 99 percent identity to 16S rRNA for both the non-annotated transcripts and the intervening “hypothetical gene” HM1_3154 (Figure S2B). Table S1 lists 53 transcripts that map to intergenic regions (i.e. not overlapping with genes on either strand) and have coverage ≥ 10 and minimal length of 50 bases. BLAST searches were performed on all peptides potentially encoded by this region, but the vast majority did not have significant matches. Three of them are potential pseudogenes, with significant matches to peptides 50-55 residues long. One transcript, however, has 58% identity (73% similarity) to the ParA chromosome partitioning protein of Paenibacillus polymyxa over the entire length of 282 residues, and thus very likely represents a gene that was missed in the primary genome annotation (Table S1).
Discussion

In order to put into useful context, the changes in expression of the 2757 genes for which we possess data, we have made use of the annotated metabolic pathways in the Kyoto Encyclopedia of Genes and Genomes (KEGG). All of the pathways in the KEGG for the *H. modesticaldum* genome were analyzed and a table was constructed for each pathway by incorporating the expression data for all genes encoding products involved in the pathway. These tables are named according to the KEGG numbering system (e.g. hmo00010 is the glycolysis/gluconeogenesis pathway). Each can be found as a pathway in one of the Data Sets described in the Supplemental Materials. Each Data Set contains a set of pathways that are similar in nature, and the organization of the Data Sets mirrors the discussion that follows. The results from all of these pathways are summarized in Table 2. The statistical significance of the overall expression changes of the genes in each pathway were assessed by paired Student’s t-test (two-tailed) of the RPKM values for each gene in the two conditions (plus and minus ammonium). Overall expression changes with P-values lower than 0.05 were deemed significant. By this measure, expression changes in 39 of the 62 analyzable KEGG pathways were significant. To correct for the multiple comparisons problem, we applied the Bonferroni correction (i.e. α=0.05/62). After this procedure, only 7 of the KEGG pathways were deemed statistically significant (Table 2).

Nitrogen fixation, uptake, and incorporation

Providing nitrogen in the form of ammonium eliminates the need to fix N₂ via the energy-expensive nitrogenase system (encoded by the *nif* operon), which requires 16 ATP and 8 Fd for each N₂ molecule fixed [13]. Not only is nitrogenase activity absent in cells grown in medium with ammonium, but the activity is lost very quickly after addition of ammonium to N₂-fixing cultures, in this species [9] and

---

**Figure 2:** Proposed connections between fundamental carbon metabolism and nitrogen assimilation. The major pathways of carbon metabolism when cells are grown with pyruvate as the sole carbon and electron source are shown, along with the major nitrogen incorporations reactions. Ammonia is transported into the cell when available, or is made by nitrogenase by reduction of N₂ when unavailable. The ammonia is first incorporated into an organic molecule by Gln synthetase. Gln is used to make Glu, which can be used to make Asp and the rest of the amino acids via transaminasins. Glu, Glu, and Asp are the nitrogen donors for almost every other N-containing biomolecule, including amino acids (aa) and nucleobases (nb). For simplicity, the consumption/generation of reducing equivalents and ATP is largely left off this figure. Since the direction of the TCA cycle between succinate and 2-OG is unclear at this time, the arrows of the SCS and 2OGS reactions are dotted. Abbreviations: 2OGS, 2-oxoglutarate synthetase; Acon, aconitase; Amt, ammonium transporter; CS, citrate synthase; FR, fumarate reductase; GlnS, glutamine synthetase; GluS, glutamate synthase; IDH, isocitrate dehydrogenase; MDH, malate dehydrogenase; N₄ase, nitrogenase; PEPCK, phosphoenolpyruvate carboxykinase; PEP, phosphoenolpyruvate; PFOR, pyruvate:ferrodoxin oxidoreductase; PPDK, pyruvate phosphate dikinase; SCS, succinyl-CoA synthetase.
Table 1: Description: Summary of effect of ammonium depletion upon gene expression.

<table>
<thead>
<tr>
<th>Fold change</th>
<th># genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;20</td>
<td>71</td>
</tr>
<tr>
<td>&lt;10</td>
<td>224</td>
</tr>
<tr>
<td>&lt;5</td>
<td>629</td>
</tr>
<tr>
<td>&lt;2.5</td>
<td>1278</td>
</tr>
<tr>
<td>-1.25 to -2.5</td>
<td>557</td>
</tr>
<tr>
<td>-1.25 to +1.25</td>
<td>252</td>
</tr>
<tr>
<td>+1.25 to +2.5</td>
<td>233</td>
</tr>
<tr>
<td>&gt;2.5</td>
<td>110</td>
</tr>
<tr>
<td>&gt;5</td>
<td>27</td>
</tr>
<tr>
<td>&gt;10</td>
<td>10</td>
</tr>
<tr>
<td>&gt;20</td>
<td>9</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>KEGG pathway (common name)</th>
<th>Pathway Description</th>
<th>Average fold change (n)</th>
<th>P-value of RPKM differences</th>
<th>Summary of findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>hmo00010 Glycolysis/ Gluconeogenesis</td>
<td>While <em>H. modesticaldum</em> does not metabolize glucose, the enzymes of the glycolysis pathway are likely used for gluconeogenesis, which generates glucose from non-carbohydrate sources (e.g. pyruvate), and for central carbon metabolism.</td>
<td>-8.0 ± 4.9 (14)</td>
<td>P-value = 0.00284</td>
<td>The genes involved in the gluconeogenesis pathway were all downregulated in the absence of ammonium by a factor of ~8.</td>
</tr>
<tr>
<td>hmo00020 Tricarboxylic acid (TCA) Cycle</td>
<td><em>H. modesticaldum</em> cannot use oxygen as an electron acceptor, since it is a strict anaerobe, so this pathway is not used for respiration. The TCA cycle is likely used to produce precursors for other biosynthetic pathways (e.g. succinate, 2-oxoglutarate, and citrate).</td>
<td>-3.6 ± 3.5 (17)</td>
<td>P-value = 0.00644</td>
<td>Expression of the genes in the TCA pathway were reduced by a factor of ~4 on average, except for the <em>pckA</em> gene encoding phosphoenoxyglyceroxyvanine kinase (PEPCK), which was downregulated by approximately 16-fold.</td>
</tr>
<tr>
<td>hmo00030 Pentose Phosphate Pathway</td>
<td>This pathway generates NADPH and 5-carbon sugars through the oxidation of glucose and subsequent isomerization of pentoses. Primarily anabolic, the pentose phosphate pathway serves as an alternative to glycolysis and as a mechanism to generate pentoses and NADPH. The lack of the oxidative enzymes (glucose-6-phosphate dehydrogenase and glucuronate-6-phosphate dehydrogenase) indicates that this pathway does not exist, and the pentose isomerases are used to interconvert carbohydrates. (A reductive pentose phosphate pathway also does not exist, due to the lack of Ribulose.)</td>
<td>-7.5 ± 4.2 (13)</td>
<td>P-value = 0.0171</td>
<td>The genes involved in the pentose phosphate pathway were all downregulated in the absence of ammonium by approximately 7.5-fold, on average, with individual fold changes ranging from 2.8-fold to 14.7-fold reduction.</td>
</tr>
<tr>
<td>hmo00051 Fructose/Mannose metabolism</td>
<td>The purpose of the enzymes in the fructose and mannose pathways is to convert fructose and mannose to glycolysis precursors. Mannose is converted to fructose, which is then eventually converted to glyceraldehyde, where it can then enter the gluconeogenesis pathway.</td>
<td>-5.4 ± 5.3 (12)</td>
<td>P-value = 0.0772</td>
<td>Expression of genes involved in the fructose and mannose metabolism was not significantly affected by the absence of ammonium.</td>
</tr>
<tr>
<td>hmo00052 Galactose Metabolism</td>
<td>This is part of the Leloir pathway, UDP-galactose is first converted to UDP-glucose, then to glucose-1-phosphate. The major enzymes are hexokinase, UTP-glucose-1-phosphate uridylyltransferase, and UDP-glucose 4-epimerase.</td>
<td>-2.5 ± 0.8 (4)</td>
<td>P-value = 0.25</td>
<td>Expression of genes involved in galactose metabolism was not significantly affected by the absence of ammonium.</td>
</tr>
<tr>
<td>hmo00061 Fatty Acid Biosynthesis</td>
<td>Fatty acid biosynthesis relies on the fatty acid synthases and acetyl CoA carboxylases. Fatty acid synthases are large multi-enzyme complexes that create fatty acids from acetyl-CoA and malonyl-CoA. Fatty acids are long aliphatic chains with a terminal carboxylate, and are used in many applications, from energy storage to membrane structure. Acetyl-CoA carboxylase plays an essential role in regulating fatty acid synthesis; it catalyzes the committed step in fatty acid synthesis, the production of malonyl-CoA.</td>
<td>-1.7 ± 2.8 (10)</td>
<td>P-value = 0.35</td>
<td>Expression of genes involved in fatty acid biosynthesis was not significantly affected by the absence of ammonium.</td>
</tr>
<tr>
<td>hmo00071 Fatty Acid Degradation</td>
<td><em>H. modesticaldum</em> has never been reported to use fatty acids as carbon sources or energy sources; its three defined carbon sources are pyruvate, lactate and acetate. Inspection of related bacteria (<em>Desulfotobacterium hafniense</em> Y51 and <em>Desulfotomaculum reducens</em>) revealed that the same enzymes were missing (dehydrogenase, hydratase, CPT2 and CPT1). Therefore, it is unlikely that β-oxidation of fatty acids takes place in <em>H. modesticaldum</em>.</td>
<td>-6.4 ± 7.9 (6)</td>
<td>P-value = 0.065</td>
<td>The only two genes with significant expression in the fatty acid degradation pathway were HM1_0073 and aas, which were downregulated in the absence of ammonium by about 6 fold and 2 fold, respectively. HM1_0073 codes for 3-hydroxyacyl-CoA dehydrogenase while aas codes for 2-acylglycerophospho-ethanolamine acyltransferase, both of which are used in fatty acid biosynthesis.</td>
</tr>
</tbody>
</table>
This organism appears to make use of the alternate pathway in Gram-positive bacteria that proceeds via fumaraline. Some genes (mnpB and mnpD) appear to be missing, but these activities may be supplied by homologs (i.e. MenC/D) and 2 other genes in the cluster containing other menaquinone synthesis genes. Methyl group addition is catalyzed by MenG, while decarboxylation and isoprenylation appear to be catalyzed by UbD/X and UbA, respectively.

Because *H. modesticaldum* is strictly anaerobic, it does not use oxygen as an electron acceptor. However, four of the five key enzyme complexes of the general oxidative phosphorylation pathway are retained (Complexes I to III and ATP synthase), which generate most of the organism's ATP during phototrophic growth. The cytochrome bd complex is also present, most likely as a way to scavenge $O_2$.

The table was created to include all of the components lacking in *hmo00190* in order to complete the presumed cyclic electron flow pathway. This includes the cyt C$_{bb}$, the HbRC subunits, and ferredoxins, as well as FNR (ferredoxin-NADP+ oxidoreductase).

The [NiFe] hydrogenase is likely used to oxidize H$_2$ and reduce the quinone pool. There are 2 genes for [FeFe] hydrogenases, which typically oxidize ferredoxin and reduce protons to H$_2$, or the reverse, but it is unclear if they are active, given the apparent absence of maturation genes.

Purine metabolism is an anabolic process that begins with ribose-5-phosphate, resulting in the synthesis of nucleic acids and nucleotides. Specifically, ribose-5-phosphate is the precursor of IMP, which can then be used to synthesize ATP and GTP. GTP and ATP can then be used for RNA synthesis or converted to dATP and dGTP and used in DNA synthesis. Ribose-5-phosphate may also be converted to 5-aminomimidazole ribonucleotide and feed into thiamine metabolism.

This pathway focuses on the synthesis of the amino acids alanine, aspartate, and glutamate. The reactions involved are generally anabolic as they synthesize the compounds from smaller precursors. The primary branch utilizes the enzymes glutamate synthase, glutamine synthetase, glucosamine-fructose-6-phosphate aminotransferase, amidophosphoribosyltransferase, and carbamoyl-phosphate synthase. The pathway connects to glycolysis/glutamine metabolism and a number of sidefeeder pathways.

The purpose of this pathway is to synthesize the pyrimidines uracil, cytosine, thymine, and their derivatives. The reactions leading up to the end products are anabolic in nature. Some of the key enzymes of the main branch (from Asp and carbamoyl phosphate to UMP) are aspartate carbamoyltransferase, and orotate phosphoribosyltransferase. The pyrimidine metabolic pathway also connects with metabolic pathways of pentose phosphate and several amino acids.

The purpose of this pathway is to synthesize the amino acids glycine, serine, and threonine. The reactions are generally catabolic as more complex molecules yield simpler ones. More branches yield glycine and serine than threonine, since they can generally be reused in a broader range of molecules; some of the key enzymes are serine hydroxymethyltransferase, phosphor-glyceromutase, and phosphoserine aminotransferase. The pathway leads into a number of other pathways, such as glycolysis and glyoxylate metabolism.

This pathway focuses on synthesizing the amino acids cysteine and methionine and their derivatives. The reactions are primarily anabolic in nature. Many of the branches lead to cysteine and homocysteine; important enzymes in the pathway include serine O-acetyltransferase, cysteine synthase A, and those of the S-adenosyl methionine cycle. The pathway connects with the synthesis of some other amino acids and their derivatives, as well as sulfur metabolism and methionine salvage.
<table>
<thead>
<tr>
<th>Protein</th>
<th>Degradation</th>
<th>The purpose of this pathway is to degrade the amino acids valine, leucine, and isoleucine into intermediates that can then be used in the citric acid cycle, pyrimidine metabolism, propanoate metabolism, and for the biosynthesis of more complex molecules. The lack of a branched chain α-keto acid dehydrogenase and most of the subsequent enzymes specific to this pathway, however, indicates that it does not exist in H. modesticaldum. The aminotransferase is used in the biosynthetic pathway, and the other enzymes are used in other metabolic pathways. Furthermore, this organism cannot use branched chain amino acids as a carbon source.</th>
<th>-7.0 ± 7.7 (7)</th>
<th>P-value = 0.0169</th>
<th>Expression of genes assigned to this pathway underwent a 7-fold reduction, but this pathway does not exist in this organism.</th>
</tr>
</thead>
<tbody>
<tr>
<td>hmo00290</td>
<td>Valine, Leucine and Isoleucine Biosynthesis</td>
<td>The main purpose for this pathway is to synthesize valine, leucine, and isoleucine, which contain branched aliphatic side chains. The starting point is pyruvate and 2-oxo-butanoate, which is derived from threonine or pyruvate. Acetolactate synthase is a key enzyme that catalyzes the reaction between threonine and pyruvate. Isoleucine and valine synthesis utilize the same enzymes throughout the pathway. During the synthesis of valine, 2-oxoisovalerate can be converted to 2-isopropylmalate to initiate the synthesis of leucine.</td>
<td>-7.4 ± 1.6 (8)</td>
<td>P-value = 0.00124</td>
<td>Expression of genes involved in valine, leucine and isoleucine biosynthesis decreased by a factor of 7.5.</td>
</tr>
<tr>
<td>hmo00300</td>
<td>Lysine Biosynthesis</td>
<td>The purpose of this pathway is to synthesize lysine from aspartate or homoserine. Some important enzymes in this reaction are: aspartate kinase, homoserine dehydrogenase, and diaminopimelate decarboxylase. This pathway also feeds into the pathway for peptidoglycan biosynthesis.</td>
<td>-2.5 ± 1.1 (9)</td>
<td>P-value = 0.0328</td>
<td>On average, genes involved in the biosynthesis of lysine decreased by 2.5-fold.</td>
</tr>
<tr>
<td>hmo00330</td>
<td>Arginine and Proline Metabolism</td>
<td>The purpose of this pathway is to synthesize the amino acids arginine and proline. The reaction of glutamate to ornithine is catalyzed by the enzyme cytochrome oxidase. The reaction of ornithine to arginine is catalyzed by ornithine carbamoyltransferase. The synthesis of proline from glutamate is catalyzed by glutamyl kinase.</td>
<td>-2.4 ± 2.0 (18)</td>
<td>P-value = 0.0463</td>
<td>On average, the genes involved in the metabolism of arginine and proline decreased by 2.5-fold.</td>
</tr>
<tr>
<td>hmo00340</td>
<td>Histidine Metabolism</td>
<td>The purpose of this pathway is to synthesize histidine, originating from the pentose phosphate pathway. The initial step is the reaction of PRPP with ATP phosphoribosyltransferase. Phosphoribosylformimino-AICAR-phosphate is a branchpoint metabolite; conversion to AICAR leads to purine metabolism. Conversion to imidazole-glycerol-3-phosphate by imidazoleglycerol-phosphate dehydratase will produce histidine.</td>
<td>-12.1 ± 7.4 (8)</td>
<td>P-value = 0.0000175</td>
<td>Expression of genes involved in histidine metabolism were down-regulated by 12-fold on average. (Changes are deemed statistically significant after Bonferroni correction.)</td>
</tr>
<tr>
<td>hmo00400</td>
<td>Phenylalanine, Tyrosine, &amp; Tryptophan Biosynthesis</td>
<td>The purpose of this pathway is to synthesize the aromatic amino acids: phenylalanine, tyrosine and tryptophan. D-Erythrose-4-phosphate is converted to chorismate by the shikimate pathway. Thereafter, each aromatic amino acid is synthesized by distinct pathways. If anthranilate synthase is used, the product (anthranilate) will go through a series of reactions to produce Tryptophan. Chorinate can also be converted to prephenate by chorismate mutase. This can then lead to the production of phenylalanine with the initial enzymes chorinate synthase mutase/prephenate dehydratase present or tyrosine if only prephenate dehydrogenase is used.</td>
<td>-11.4 ± 7.1 (14)</td>
<td>P-value = 0.0162</td>
<td>The expression of genes involved in phenylalanine, tryptophan and tyrosine biosynthesis decreased by an average of 14.5-fold. A strange exception was the two aroF genes: the highly expressed one decreased by 59-fold, while the lower expressed one slightly increased (1.4-fold) in the absence of ammonium.</td>
</tr>
<tr>
<td>hmo00450</td>
<td>Selenocompound Metabolism</td>
<td>This metabolism represents the process by which Selenium-based species are incorporated into biological compounds, which in turn can be incorporated into proteins. The pathways involved constitute a branched pathway. H. modesticaldum is highly expressed in this pathway and isoleucine into selenocysteine-1-IRNA and the conversion of Selenocysteine into Seleno-methionyl-tRNAMet.</td>
<td>-4.25 ± 3.95 (8)</td>
<td>P-value = 0.0277</td>
<td>Expression of genes involved in selenocompound metabolism were reduced an average of 5-fold. A notable exception to this trend – metC – showed a 4.5-fold increase in expression.</td>
</tr>
<tr>
<td>hmo00473</td>
<td>D-Alanine Metabolism</td>
<td>This pathway interconverts D-Alanine and L-Alanine and converts D-Alanine to D-Alanyl-D-Alanine, which is used in peptidoglycan metabolism.</td>
<td>-1.76 ± 0.57 (2)</td>
<td>P-value = 0.33 (not significant)</td>
<td>The highly expressed alanine racemase is reduced by ~20-fold.</td>
</tr>
<tr>
<td>hmo00500</td>
<td>Starch &amp; Glycogen metabolism</td>
<td>This pathway is used to synthesize α-(1,4)-linked glucose polymer using UDP-glucose as a monomer donor, which is synthesized by UTP-glucose-1-phosphate uridyltransferase from glucose-1-phosphate. The presence of branching and debranching enzymes suggest that the polymer present in this organism has occasional β(1,6)-linked branchpoints and thus would resemble glycogen. Glycogen phosphorylase is used to liberate glucose monomers as glucose-1-phosphate when the cell has need of the carbon source.</td>
<td>-2.9 ± 1.3 (8)</td>
<td>P-value = 0.0214</td>
<td>On average, expression of the genes in this pathway decreases ~3-fold in the absence of ammonium.</td>
</tr>
<tr>
<td>hmo00520</td>
<td>Amino/Sugar Nucleotide Synthesis</td>
<td>This pathway involves the attachment of sugars to nucleotides and the modification of these nucleotide sugars. H. modesticaldum has mechanisms to create and modify UDP and GDP sugars. These pathways interconnect with a variety of other metabolic processes. Glycolysis, glyconeogenesis, and the fructose and mannose metabolisms provide sugars to start the anabolic process.</td>
<td>-4.4 ± 3.9 (16)</td>
<td>P-value = 0.000959</td>
<td>The expression of genes involved in amino/sugar nucleotide synthesis decreased by an average of four-fold.</td>
</tr>
</tbody>
</table>
H. modesticaldum only possess the part of this pathway that produces the S-layer glycoprotein from sedoheptulose-7-phosphate. This is an anabolic process. Note that enzyme 2.7.7.71 (D-glycer-roalpha-D-manno-heptose-1-phosphate guanylyltransferase), which is required for assembly of S-layer glycoprotein in some Gram-positive bacteria, seems to be missing.

Riboflavin is an important element of several parallel pathways, utilizing glycine and tyrosine as starting points. It can then be converted to fumarate and succinate by the "left arm" of the TCA cycle. Incorporation of carbon into biomolecules then can be achieved by pyruvate carboxylase. Alternatively, it can also be converted to PEP by pyruvate phosphate dikinase. Glycerol is first phosphorylated by glycerol kinase to glycerol 3-phosphate. Glycerol-3-phosphate is then acylated at positions 1 and 2. The final product is 1,2-diacyl-glycerol-3-phosphate (phosphatidate), which serves as the biosynthetic precursor for the formation of all glycerolipids.

The expression of genes involved in riboflavin metabolism was reduced by a factor of approximately 4 on average.

Peptidoglycan is a macromolecule made of long amino-sugar polymers, cross-linked by short peptides. It forms the cell wall that encases the cell membrane, providing the cell with structural support and protection. The glycan strands are typically comprised of repeating N-acetylglucosamine (GlcNAc) and N-acetylmuramic acid (MurNAc) disaccharides. Each MurNAc is linked to a peptide of three to five amino acid residues. Disaccharide subunits are first assembled on the cytoplasmic side of the bacterial membrane on a polysaccharide anchor (lipid II). Polymerization of disaccharide subunits by transglycosylases and cross-linking of glycan strands by transpeptidases occurs on the other side of the membrane.

The expression of genes involved in peptidoglycan synthesis decreased by an average of five-fold.

Glycerolipid metabolism is an anabolic process resulting in the formation of glycerol-based phospholipids, which are a structural component of the cell membrane. Phosphatidate is the precursor. Various enzymes attach different alcohol headgroups in a phospoester linkage. In Helicobacteria, these headgroups include serine, ethanolamine, and glycerol. (They lack choline.) Pyruvate plays a vital role in the metabolism of H. modesticaldum, and serves as one of its major sources of carbon. Pyruvate can be carboxylated to OAA by pyruvate carboxylase. Alternatively, it can also be converted to PEP by pyruvate phosphate dikinase. Incorporation of carbon into biomolecules then can be achieved through PEPC, which converts PEP to OAA. In either case, OAA can then be converted to fumarate and succinate by the "left arm" of the TCA cycle.

Expression of the genes in the pyruvate metabolism pathway was reduced by a factor of approximately 6 on average.

Glycerol metabolites do not appear to exist in H. modesticaldum, as it lacks the key enzymes, isocitrate lyase and malate synthase. The role of the glcD gene product (glycocalx oxidase) is unclear at this time.

Propanoate metabolism results in the production of propionyl-CoA, which is converted to succinyl-CoA, an intermediate in the TCA cycle.

Butanoate is a four-carbon fatty acid formed by bacterial fermentation of carbohydrates. It is often used for ketone body production and fed into the citrate cycle, glycolysis, or glutamate synthesis. It does not appear that butanoate is metabolized in H. modesticaldum due to the absence of many key enzymes.

This pathway covers one-carbon molecule metabolism utilizing folate. Reduction of dihydrofolate (DHF) yields tetrahydrofolate (THF). THF can serve as an acceptor/donor of 1-carbon units in three possible oxidation states: N5-methyl-THF, N5,N10-methylene-THF, and N5-formyl-THF (and others in the same oxidation state).

Thiamine, also known as vitamin B1, is composed of pyrimidine and thiazole structures, which combine to form thiamine phosphate and are made via separate pathways. The thiamine phosphate product is phosphorylated to yield the active form of the cofactor, thiamine pyrophosphate. In this anabolic pathway, thiamine is made through this time.

Expression of genes in the carbon folate metabolism pathway was reduced by a factor of approximately 4 on average.

Riboflavin, also known as vitamin B2, is an important element of cofactors flavin mononucleotide (FMN) and flavin adenine dinucleotide (FAD). These cofactors function as redox agents for a variety of processes (i.e. the electron transport chain) due to their ability to accept/donate a hydride plus a proton (i.e. 2 electrons and 2 protons) in 1-electron or 2-electron steps. Riboflavin is synthesized from the GTP product of metabolized purines and then converted to FMN and FAD by the enzymes riboflavin kinase and FMN adenyltransferase, respectively.

Expression of genes involved in riboflavin metabolism was unaffected by the presence/absence of ammonium.
**Vitamin B6 Metabolism**

<table>
<thead>
<tr>
<th>Gene ID</th>
<th>Description</th>
<th>Change</th>
<th>P-value</th>
<th>Note</th>
</tr>
</thead>
<tbody>
<tr>
<td>hmo00750</td>
<td>Vitamin B6 Metabolism</td>
<td>-10.9 ± 9.3 (4)</td>
<td>0.0645</td>
<td>The expression of genes involved in vitamin B6 metabolism was not significantly affected by the absence of ammonium.</td>
</tr>
</tbody>
</table>

**Nicotinate and Nicotinamide Metabolism**

<table>
<thead>
<tr>
<th>Gene ID</th>
<th>Description</th>
<th>Change</th>
<th>P-value</th>
<th>Note</th>
</tr>
</thead>
<tbody>
<tr>
<td>hmo00760</td>
<td>Nicotinate (vitamin B3) and nicotinamide are precursors to NAD⁺ and NADP⁺, which are used for redox reactions. The absence of ammonium leads to an up-regulation of nicotinic acid synthetase by a factor of 2.4.</td>
<td>-2.2 ± 2.4 (6)</td>
<td>0.28</td>
<td>The expression of genes involved in nicotinate and nicotinamide metabolism was not significantly affected by the absence of ammonium.</td>
</tr>
</tbody>
</table>

**Pantothenate and CoA Biosynthesis**

<table>
<thead>
<tr>
<th>Gene ID</th>
<th>Description</th>
<th>Change</th>
<th>P-value</th>
<th>Note</th>
</tr>
</thead>
<tbody>
<tr>
<td>hmo00770</td>
<td>Pantothenate, also known as vitamin B5, is necessary for 1-carbon metabolism (see hmo01200). The expression of genes involved in pantothenate metabolism was not significantly affected by the absence of ammonium.</td>
<td>-4.0 ± 5.1 (14)</td>
<td>0.0184</td>
<td>The expression of genes involved in pantothenate and CoA biosynthesis decreased by an average of four-fold.</td>
</tr>
</tbody>
</table>

**Biotin Metabolism**

<table>
<thead>
<tr>
<th>Gene ID</th>
<th>Description</th>
<th>Change</th>
<th>P-value</th>
<th>Note</th>
</tr>
</thead>
<tbody>
<tr>
<td>hmo00780</td>
<td>Biotin, also known as vitamin B7, is a water-soluble B vitamin that is a coenzyme for carboxylases, which are involved in the synthesis of fatty acids and the amino acids isoleucine and valine. The absence of ammonium leads to an up-regulation of biotin synthetase by a factor of 1.3.</td>
<td>-1.3 ± 0.2 (4)</td>
<td>0.0719</td>
<td>The expression of genes involved in biotin metabolism was unaffected by the absence of ammonium.</td>
</tr>
</tbody>
</table>

**Folate Synthesis**

<table>
<thead>
<tr>
<th>Gene ID</th>
<th>Description</th>
<th>Change</th>
<th>P-value</th>
<th>Note</th>
</tr>
</thead>
<tbody>
<tr>
<td>hmo00790</td>
<td>Folate, also known as vitamin B9, is necessary for 1-carbon metabolism (see hmo01200). The expression of genes involved in folate synthesis was not significantly affected by the absence of ammonium.</td>
<td>-4.7 ± 2.3 (8)</td>
<td>0.0115</td>
<td>The expression of genes involved in folate synthesis decreased by an average of four-fold.</td>
</tr>
</tbody>
</table>

**Porphyrin and Chl Synthesis**

<table>
<thead>
<tr>
<th>Gene ID</th>
<th>Description</th>
<th>Change</th>
<th>P-value</th>
<th>Note</th>
</tr>
</thead>
<tbody>
<tr>
<td>hmo00860</td>
<td>The purpose of this pathway is to synthesize heme, (bacterial) chlorophyll, and corrin. This anaerobic pathway originates with the production of 5-aminolevulinic acid from glucose-6-phosphate. The absence of ammonium leads to an up-regulation of uroporphyrinogen III synthase by a factor of 8.6.</td>
<td>-8.6 ± 8.7 (37)</td>
<td>0.000036</td>
<td>The expression of genes involved in porphyrin and chlorophyll synthesis decreased by an average of 8.5-fold. (Changes are deemed statistically significant after Bonferroni correction.)</td>
</tr>
</tbody>
</table>

**Terpenoid Backbone Biosynthesis**

<table>
<thead>
<tr>
<th>Gene ID</th>
<th>Description</th>
<th>Change</th>
<th>P-value</th>
<th>Note</th>
</tr>
</thead>
<tbody>
<tr>
<td>hmo00900</td>
<td>The terpenoid backbone biosynthesis pathway results in the formation of natural products consisting of isoprene units. The expression of genes involved in the terpenoid backbone biosynthesis pathway was not significantly affected by the absence of ammonium.</td>
<td>-8.4 ± 5.8 (9)</td>
<td>0.0584</td>
<td>The expression of genes involved in the terpenoid backbone biosynthesis pathway was not significantly affected by the absence of ammonium.</td>
</tr>
</tbody>
</table>

**Carotenoid Biosynthesis**

<table>
<thead>
<tr>
<th>Gene ID</th>
<th>Description</th>
<th>Change</th>
<th>P-value</th>
<th>Note</th>
</tr>
</thead>
<tbody>
<tr>
<td>hmo00906</td>
<td>The purpose of this pathway is to incorporate ammonia into amino acids. First, glutamine synthetase catalyzes the reaction between glutamate and ammonium to produce glutamine. Although the organism appears to have components of the enzyme nitrite reductase (which reduces nitrite to ammonia) and 4,4'-diaponeurosporene, the typical pathway for such a C30 carotenoid requires only two gene products: CrhA and CrhB. The latter has been annotated in the genome of this organism.</td>
<td>-11.4 (1)</td>
<td>undefined</td>
<td>Expression of the crhA gene was reduced by a factor of approximately 11 in the absence of ammonium.</td>
</tr>
</tbody>
</table>

**Nitrogen Metabolism (Non-Nitrogen Fixation Genes)**

<table>
<thead>
<tr>
<th>Gene ID</th>
<th>Description</th>
<th>Change</th>
<th>P-value</th>
<th>Note</th>
</tr>
</thead>
<tbody>
<tr>
<td>hmo00910a</td>
<td>The expression of genes involved in nitrogen metabolism (but not nitrogen fixation) was not significantly affected by the absence of ammonium.</td>
<td>-4.7 ± 3.7 (5)</td>
<td>0.25</td>
<td>The expression of genes involved in nitrogen metabolism (but not nitrogen fixation) was not significantly affected by the absence of ammonium.</td>
</tr>
</tbody>
</table>

**Nitrogen Metabolism (Nitrogen Fixation Genes)**

<table>
<thead>
<tr>
<th>Gene ID</th>
<th>Description</th>
<th>Change</th>
<th>P-value</th>
<th>Note</th>
</tr>
</thead>
<tbody>
<tr>
<td>hmo00910b</td>
<td>The expression of genes involved in nitrogen fixation were found to be up-regulated by an average of ~25-fold in the absence of ammonium. If one assumes an RPMK of 1 in the ammonium-replete case for the 5 genes with no reads in that condition, then the increase in expression can be estimated to be ~23-fold (i.e. the estimation of up-regulation by the absence of ammonium is fairly robust).</td>
<td>24.9 ± 15.4 (9)</td>
<td>undefined</td>
<td>Expression of genes involved in nitrogen fixation were found to be up-regulated by an average of ~25-fold in the absence of ammonium. If one assumes an RPMK of 1 in the ammonium-replete case for the 5 genes with no reads in that condition, then the increase in expression can be estimated to be ~23-fold (i.e. the estimation of up-regulation by the absence of ammonium is fairly robust).</td>
</tr>
</tbody>
</table>
Sulfur metabolism takes place in *H. modesticaldum* via both oxidative and reductive pathways. It is present in nucleic acids, proteins, polysaccharides, phenols, etc. This organism does not seem to reduce sulfate, but rather reduces thiosulfate to sulfide, which is then used to make cysteine. Pathways that involve the incorporation of sulfur in this organism requires the input of energy. 1.4 ± 3.1 (10) P-value = 0.59 Expression of genes involved in sulfur metabolism was unaffected by the absence of ammonium. (A notable exception was one of the cysD genes, which was down-regulated about 9-fold.)

Aminoacyl-tRNAs deliver amino acids to the ribosome during translation for active transport of a wide variety of molecules. *H. modesticaldum* has ABC transporters that move mineral and organic ions, monosaccharides, phosphate and amino acids, and metal cations/iron siderophores/vitamins. -3.4 ± 2.6 (44) P-value = 0.0148 On average, the expression of genes involved in the synthesis of aminoacyl-tRNAs was reduced by a factor of 3.5 in the absence of ammonium.

*H. modesticaldum* contains at least seven such systems, presumably responding to diverse signals such as changes in temperature, pH, osmolarity, etc.

The expression of genes involved in the DNA replication pathway was reduced by an average of 5- to 6-fold. (Changes are deemed statistically significant after Bonferroni correction.)

The chemotaxis pathway of *H. modesticaldum* allow the bacterium to detect chemical gradients through its receptors and through a signal cascade regulate its flagella. Direction and speed of rotation of the flagella can also be modulated. -5.7 ± 2.9 (14) P-value = 0.000266 Expression of all genes involved in the chemotaxis pathway was reduced by an average of 5- to 6-fold. (Changes are deemed statistically significant after Bonferroni correction.)

The expression of genes involved in flagellar assembly was not significantly affected by the absence of ammonium if all genes are included. If one excludes the 6 genes that increase in expression, however, the overall drop in expression was ~7-fold (significant after Bonferroni correction).

The phosphotransferase system (PTS) is used by bacteria for the uptake of carbohydrates, especially hexoses and disaccharides. PTS also catalyzes their conversion into phosphoehopes to facilitate transportation. PTS separated into two components: enzyme I along with histidine phosphocarrier protein, and membrane-bound sugar specific permeases. However, in *H. modesticaldum*, only the PEP-protein phosphotransferase seems to be present.

The expression of genes involved in the two-component signal transduction system was reduced by a factor of ~3.5. A notable exception to this trend, csrA, displayed a 7-fold increase in expression.

Expression of all genes involved in the chemotaxis pathway was reduced by an average of 5- to 6-fold. (Changes are deemed statistically significant after Bonferroni correction.)

The ribosome is the site at which translation of mrNA into an amino acid polypeptide chain takes place. For *H. modesticaldum*, ribosomal RNAs that are involved in the construction of the ribosome include 23S/5S and 16S rRNAs to make the large 50S and small 30S subunit, respectively. -1.0 ± 3.0 (78) P-value = 0.0654 The expression of genes involved in the construction of ribosomes was unaffected by the absence of ammonium.

This pathway involves the enzymes used in the process of RNA degradation. This is a catalytic process that is used to routinely degrade messenger RNA into its component nucleotides. -1.1 ± 2.9 (9) P-value = 0.0511 Expression of genes for enzymes and proteins involved in the RNA degradation pathway is unaffected by the absence of ammonium.

RNA Polymerase is responsible for synthesis of all RNAs in the bacterium, using the genomic DNA as a template. Bacterial RNA Polymerase is made up of five subunits: β', β, two α subunits, and the ω subunit. β' is the largest subunit and encompasses some of the active site that performs RNA synthesis. The rest of the active site is located in the β subunit. The two α subunits contain determinants for interactions with DNA promoters and regulatory factors. The ω subunit is the smallest and stabilizes the formation of the protein. -1.3 ± 2.9 (5) P-value = 0.78 The expression of genes involved in RNA polymerase was unaffected by the absence of ammonium.

DNA Replication is the process by which the chromosomal DNA is copied. Proteins involved includes the subunits of the DNA polymerase III holoenzyme, and a variety of other enzymes including helicase and primase. (Helobacteria appear to missing subunits β, ϕ, and γ in the DNA polymerase III holoenzyme, but these subunits are typically not found in the Firmicutes.) -6.7 ± 7.6 (12) P-value = 0.407 The expression of genes involved in the DNA replication pathway is reduced by a factor of ~7 on average.
Table 2: For each pathway in the KEGG for *H. modesticaldum*, a table was created (present in the Supplemental Information named after the KEGG pathway; e.g. Table hmo00010 is for pathway hmo00010, which is glycolysis). Each row in Table 2 summarizes one of these tables, providing the KEGG designation, name and brief description of the pathway, as well as a summary of the RNA-seq data.

<table>
<thead>
<tr>
<th>Pathway</th>
<th>KEGG Designation</th>
<th>Name</th>
<th>Brief Description</th>
<th>Expression Change</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>hmo03060</td>
<td>Protein export</td>
<td><em>H. modesticaldum</em> possesses both the Sec-SRP pathway and the Twin arginine targeting (Tat) pathway. In the former, translating or newly translated proteins with an N-terminal signal peptide (or internal integral transmembrane helix) are bound by SRP and presented to the SecYEG translocase. Working with the SecA ATPase, the unfolded polypeptide is translocated from the cytosol to the exterior, where it will fold. The Tat pathway recognizes proteins with a characteristic N-terminal signal containing 2 consecutive Arg residues and is able to transport folded proteins across the membrane.</td>
<td>-3.5 ± 7.6 (13)</td>
<td>P-value = 0.0044</td>
<td>The expression of genes involved in the bacterial secretion system was not significantly affected by the absence of ammonium.</td>
</tr>
<tr>
<td>hmo03070</td>
<td>Bacterial Secretion System</td>
<td>Helio bacteria are Gram-positive bacteria and thus lack an outer membrane. Unsurprisingly, the type I-V secretion systems are not required. Secreted proteins are transported across the membrane by both the Sec-SRP pathway and the Twin arginine targeting (Tat) pathway. Thus, all complete systems overlap entirely with the hmo03060 pathway.</td>
<td>(Since all genes in this table are also on hmo03060, we do not consider them separately here.)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>hmo03410</td>
<td>Base excision repair</td>
<td>The process of base excision repair involves the recognition and removal of damaged or mismatched bases from DNA. After excision of the damaged base by a DNA glycosylase, the AP endonuclease cleaves the phosphodiester backbone just S' of the apurinic/apyrimidinic site, allowing DNA pol I to replace the damaged strand with a newly synthesized DNA strand.</td>
<td>-4.7 ± 5.1 (6)</td>
<td>P-value = 0.0217</td>
<td>The expression of genes involved in the base excision repair pathway is reduced by a factor of ~5 on average</td>
</tr>
<tr>
<td>hmo03420</td>
<td>Nucleotide excision repair</td>
<td>Nucleotide excision repair is performed by the UvrABC endonuclease. A dimer of the UvrA ATPase/GTase binds to UvrB. The UvrA dimer functions to detect damaged sites on the DNA double helix, and UvrB attaches to the damaged site. The UvrA dimer then dissociates and UvrC attaches to UvrB, resulting in cleavage of DNA upstream and downstream of the lesion. The damaged segment is removed by PcrA/UvrD DNA helicase. UvrB prevents the reannealing of the excised segment until displaced by DNA polymerase I, which fills in the correct nucleotide sequence.</td>
<td>-4.5 ± 4.8 (7)</td>
<td>P-value = 0.0214</td>
<td>The expression of genes involved in the nucleotide excision repair pathway is reduced by a factor of about 4.5 on average.</td>
</tr>
<tr>
<td>hmo03430</td>
<td>Mismatch repair</td>
<td>DNA mismatch repair involves the recognition and excision of mismatched base pairs, and the re-synthesis of the excised DNA. Note that enzyme MutH appears to be missing from this pathway, as is generally true for all Firmicutes, which use a MutH-independent pathway.</td>
<td>-3.6 ± 3.0 (9)</td>
<td>P-value = 0.0206</td>
<td>Expression of these two genes is relatively low even in the presence ammonium, but seems to decrease about 3.6-fold in the absence of ammonium.</td>
</tr>
<tr>
<td>hmo03440</td>
<td>Homologous recombination</td>
<td>The RecFOR pathway, also called the RecF pathway, repairs DNA in bacteria via homologous recombination. The pathway typically repairs single-strand gaps, but is also capable of repairing double-strand breaks when the RecBCD pathway is inactivated or absent, as may be the case in <em>H. modesticaldum</em>. The pathway produces a Holliday junction intermediate and concludes with the branch migration and resolution of the Holliday junction.</td>
<td>-5.8 ± 3.9 (19)</td>
<td>P-value = 0.000453</td>
<td>The expression of genes involved in the homologous recombination pathway is reduced by a factor of about 6 on average. (Changes are deemed statistically significant after Bonferroni correction.)</td>
</tr>
<tr>
<td>hmo03450</td>
<td>Non-homologous end joining</td>
<td>Non-homologous end-joining is a form of double strand break repair. It is initiated when the Ku protein recognizes and binds to the broken ends of double stranded DNA. Ku then forms a bridge between the two ends, aligning them as well as preventing their degradation. DNA Ligase I is then recruited, which has polymerase activity that fills the gap. Finally, the ligase activity of LigD ligates the two ends together, thus sealing the break.</td>
<td>P-value = 0.61</td>
<td>The expression of genes involved in non-homologous end joining was extremely low under both conditions; no significant differences were detected.</td>
<td></td>
</tr>
<tr>
<td>hmo04122</td>
<td>Sulfur Relay System</td>
<td>The sulfur relay system is used to pass sulfur atoms from ubiquitin-like proteins to Molybdenum cofactors (Moco) and thiamine via separate pathways. The sulfur-containing cofactors Moco and thiamine exist in this organism and require sulfur carrier proteins to be synthesized.</td>
<td>-2.50 ± 3.42 (6)</td>
<td>P-value = 0.83</td>
<td>The expression of genes involved in sulfur transport was not significantly affected by the absence of ammonium.</td>
</tr>
</tbody>
</table>

When the bacterium was placed in N2-fixing conditions, expression of genes for nitrogen fixation was up-regulated (Figure 3A and Data Set S3, pathway hmo00910b). In particular, the *nifD* and *nifK* transcripts, encoding the core subunits of nitrogenase, increased in abundance by a factor of 25-50. A similar increase was seen for *nifH*, which encodes the Fe protein that provides electrons to the nitrogenase enzyme, but it was difficult to estimate, as no reads were detected aligning to this gene

other members of the family [6,29]. The RNA-sequencing data for all genes involved in nitrogen metabolism can be found in Data Set S3. However, due to the very different regulation seen for different sets of genes, we pulled out the genes involved in nitrogen fixation and put them into a separate table (pathway hmo00910b), leaving the rest of the genes involved in basic nitrogen metabolism in the first table (pathway hmo00910a).
in the baseline transcriptome. In cases such as this, a value of 1 RPKM in was used for the baseline value in order to estimate a minimum expression difference. If one excludes genes without reads in the baseline transcriptome, the average up-regulation of nitrogen fixation genes was ~25-fold (n=9). If one includes such genes (for nifH, nifN, fdxB, HM1_0868 and HM1_0867), using an assumed baseline value of 1 RPKM, then an average increase of 23-fold (n=14) was estimated, indicating that this is a fairly robust value. This up-regulation in structural genes, maturation factors, and regulatory protein gene transcription was expected, since all of these components are needed for the reduction of molecular N₂ and are in general not required for other functions.

The exception to this may be nifβ and HM1_1865, which encode a cysteine desulfurase and FeS cluster assembly protein, respectively, and are not found in the nif operon. Thus, the corresponding gene products likely play a role in functions required during normal growth (e.g. FeS cluster assembly) as well as assembly of the FeMoCo cofactor in dinitrogenase. Expression of these two genes is quite high in the baseline transcriptome, and only increases about 3-fold under N₂-fixing conditions (Data Set S3, pathway hmo00910b). The other exception is nifX, which is found in the nif operon, and is only up-regulated by ~3-fold as well. Its expression is quite low in both transcriptomes (3.4 and 10 RPKM), so this change may not be as reliable as for the other members of the operon.

The nif gene cluster had been previously cloned and sequenced from Helobacterium chlorum, and the gene order is completely conserved in H. modesticaldum. In both species, there are three genes at the 5‘ end of the operon that appear to be involved in regulation of N₂ fixation. Immediately upstream of nifH are the nifI1 and nifI2 genes (HM1_0868 and HM1_0867, respectively), which are members of the GlnB family of PII proteins. These PII-like proteins are typically regulated by glutamine (as a signal for nitrogen), 2-oxoglutarate (as a signal for carbon), and ATP (as a signal for energy) in a wide variety of Bacteria and Achaear. [30–32]. The nifI1/nifI2 gene pair diverges from this family and is found in the nif operons of several anaerobic diazotrophs, including Clostridium acetobutylicum and Methanococcus maripaludis [33]. Just upstream of this gene pair is a conserved protein that appears to be unique to heliobacteria, called “orf1” [34]; it was proposed to alter nitrogen expression by controlling termination of transcription after the orf1 sequence, thus attenuating expression of the downstream nif genes [33]. Under N₂-fixing conditions, expression of all 3 genes increases (Figure 3A and Data Set S3, pathway hmo00910b), as was seen previously in Helobacterium chlorum [34].

Figure 3: Ammonium regulation of nitrogenase operon and ammonium transporter. (A) The absence of ammonium induces the nif operon during N₂-fixing conditions (top) and the presence of ammonium represses expression of the nitrogenase operon, in which nifD/nifK encode the αβ subunits of the nitrogenase complex. (B) The high affinity ammonium transporter (amt) and the nitrogen regulatory protein PII (HM1_0742) are highly expressed in ammonium-deplete medium (top) and repressed in the presence of ammonium (bottom).
in ammonium-deplete media by a factor of 4-5 on average. As will be seen, this behavior is typical for most metabolic pathways. Thus, expression of genes for fundamental nitrogen metabolism behaved much as expected upon a shift from ammonium-replete to \(N_2\)-fixing conditions.

**Bioenergetics and electron transfer**

Photosynthesis is an efficient mechanism for ATP production that does not compete for the carbon of pyruvate, much of which is wasted as \(CO_2\) and small molecules (e.g. acetate, ethanol) during fermentation [7]. The heliobacterial reaction center (HbRC) uses absorbed photons to oxidize plastoquinone \(c_{553}\) and reduce cytoplasmic ferredoxin (Fd). The electrons from Fd are sent back to \(c_{553}\) via complex I and the cyt \(b_c\) complex in a cyclic electron transport (cET) process that produces a proton gradient that can be used by ATP synthase to phosphorylate ATP [38-40]. Because the KEGG does not have the heliobacterial cET pathway, we made use of the oxidative phosphorylation pathway (hmo00190) as the baseline pathway, as it contains complexes I (NADH dehydrogenase), II (succinate dehydrogenase), III (cyt \(b_c\)), and V (ATP synthase), despite the fact that fact that heliobacteria lack complex IV (cytochrome oxidase), being strict anaerobes. Other genes important in bioenergetics not represented in the KEGG pathways were added to separate tabs; hmo00190b contains the genes encoding additional proteins involved in the cET pathway (e.g. the HbRC, its electron donors and acceptors), while hmo00190c lists the genes for hydrogenases.

Given the importance of the HbRC to cellular metabolism, we expected the \(psbA\) gene encoding the HbRC core polypeptide to be highly expressed during phototrophic growth. In fact, \(psbA\) is the twelfth highest protein-encoding gene expressed (1446 RPKM; Data Set S4, pathway hmo00190b, Figure 4B). During the elucidation of the HbRC structure to 2.2 Å, we discovered a previously unknown subunit called PshX [41]. This small polypeptide consists of a single transmembrane alpha-helix with a few residues on either side. The resolution of the structure was sufficient to allow identification of sidechains and subsequent identification of the gene, which was confirmed by mass spectrometry. The gene encoding this subunit (HM1_0821), which had not been previously annotated, is not in the photosynthetic gene cluster nor does it appear to be in an operon. However, its normalized expression level (1685 RPKM; Data Set S4, pathway hmo00190b, Figure 4C) is almost exactly the same as \(psbA\). This makes sense as the stoichiometry of the subunits is 2:2 in the HbRC.

The electron donor to the HbRC is a membrane-anchored cyt \(c_{553}\) encoded by the \(cybA\) gene; it is highly expressed (490 RPKM; Data Set S4, pathway hmo00190b). A high expression level was also observed for genes \(psbB1\) and \(psbB2\) (Figure 4C), which encode the \([4Fe-4S]_{2}\) dicluster ferredoxins, now thought to be the primary electron donors and acceptors, to be drastically lowered in \(N_2\)-fixing conditions. In contrast, the mRNA levels of the other proteins in the proposed cyclic electron transfer pathway (complex I and III) as well as ATP synthase (complex V) do not follow the same pattern (Data Set S4, pathway hmo00190b). They are not as highly expressed, but they are also not nearly as strongly repressed by the absence of ammonium (average repression of 2- to 3-fold). The same is true of the genes encoding the cytochrome bd menaquinol oxidase (cyaA/B), whose primary role is likely to scavenge \(O_2\) rather than contribute to energy production.

Such a large drop in transcript abundance after a shift to ammonium-deplete medium is not a common feature of the other abundant mRNAs. For example, \(rpoE\) (sigma factor) only decreases by 1.6-fold and \(cstA\) (carbon starvation protein) increases by 1.4-fold. In addition, the four most highly expressed genes in ammonium-replete conditions (HM1_3146, HM1_3147, HM1_3149, HM1_3151; 6200-19000 RPKM), all of which are conserved hypothetical proteins of unknown function located next to each other in the genome, exhibit very little repression in the absence of ammonium (at most 2-fold; Data Set S2). In order to make the comparison more meaningful, we compiled a list of the most highly expressed protein-encoding genes in ammonium-replete conditions (i.e., with normalized transcript levels >300 RPKM). When the log, of the expression change (-NH\(_4\) compared to +NH\(_4\)) is plotted against the log of the expression level in +NH\(_4\) medium of these 42 genes, something striking emerges: the structural genes for the HbRC (\(psbA, psbX\)), its electron donor (\(cybA\)) and acceptor (\(psbB1, psbB2\)) form a discrete group of genes that are highly repressed by the lack of ammonium (red dots in Figure 4B).

To test the effects of ammonium depletion on expression of the HbRC, we grew cells in PYE or PYE-\(NH_4\) media, took samples at a few points along the growth curve, and directly assayed functional HbRC levels by a quantitative P\(_{800}\) photobleaching experiment. Briefly, cells were excited with a saturating laser flash and the absorbance at 800 nm was monitored (Figure 4C). The primary electron donor of the HbRC (P\(_{800}\)) absorbs maximally at 800 nm and is bleached upon light-driven electron transfer to the ferredoxin acceptors. The bleaching recovers within a few milliseconds due to reduction of oxidized P\(_{800}\) by the membrane-anchored cyt \(c_{553}\) donor [43,44]. We observed only a modest drop (~30%) in cellular P\(_{800}\) levels in PYE-\(NH_4\) cultures relative to the PYE cultures (Table S4), and in both cases there was a small drop (~20%) as the cultures aged. Moreover, the kinetics of P\(_{800}\) recovery were about the same in both (t\(_{1/2}\) = 2.5 ms; Figure 4C), which is inconsistent with a large drop in cyt \(c_{553}\); the reduction of P\(_{800}\) by cyt \(c_{553}\) is a second-order reaction and is thus sensitive to the concentration of the electron donor [44]. Thus, despite the large drop in \(psbA\) and \(cybA\) mRNA levels, we find no evidence for correspondingly large drops in the levels of the encoded proteins. Thus, the transcript levels of the
The genome of *H. modesticaldum* contains two copies of the genes (hupSLC) for a membrane-bound [NiFe] uptake hydrogenase; one of these operons also contains the genes encoding the maturation factors (hupD hypABCDEF) [11]. This type of hydrogenase maturation factors were not found. It is thus unclear if this organism assembles an active [FeFe] hydrogenase. In any case, these genes are also not induced by the shift to N₂-fixing conditions.

Regulation of fundamental carbon metabolism by ammonium

As mentioned, *H. modesticaldum* cannot grow as a photoautotroph, as it lacks at least one key enzyme for each carbon fixation pathway. For example, heliobacteria seem to possess every enzyme for the reverse TCA (rTCA) cycle except for ATP-citrate lyase, which prevents them from using this cycle to fix CO₂ in an autotrophic manner [11,17]. Thus, *H. modesticaldum* is an obligate heterotroph; it uses pyruvate, lactate, or acetate as carbon sources [16]. Figure 2 summarizes central carbon metabolism and connections to nitrogen metabolism in heliobacteria.

Pyruvate was the sole carbon source during the growth of heliobacteria in these experiments. Expression of genes encoding enzymes in pyruvate metabolism was reduced by approximately 6-fold in the absence of ammonia (Data Set S5, pathway hmo00620). Pyruvate can be oxidatively decarboxylated to produce acetyl-CoA with concomitant reduction of 2 ferredoxins by the pyruvate-ferredoxin oxidoreductase (PFOR), as this organism lacks an NAD(P)-dependent pyruvate dehydrogenase. During fermentative growth, acetyl-CoA synthetase converts acetyl-CoA to acetyl-phosphate, which can be used as a phosphate donor to ADP to make ATP via acetate kinase. These latter 2 enzymes are moderately expressed (7-16 RPKM) under both conditions, while PFOR is highly expressed (30-168 RPKM). This may be because the cultures used were grown in the light and could use photosynthetic cET to produce ATP, allowing pyruvate to be used primarily as a carbon source.

This organism appears to lack pyruvate carboxylase to convert pyruvate to oxaloacetate (OAA). However, it can convert pyruvate to phosphoenolpyruvate (PEP) by pyruvate phosphate dikinase (PPDK). Incorporation of carbon into biomolecules can then be achieved through PEP carboxylase (PEPCK) [17], which carboxylates PEP to OAA. OAA can then be converted to malate, fumarate, and succinate by the “left arm” of the TCA cycle, these enzymes are highly expressed.

**Figure 4:** Ammonium regulation of the photosynthetic reaction center. (A) The absence of ammonium repressed the pshA gene encoding the HbRC core polypeptide. (B) Log-log plot of expression change vs. expression for the 42 most highly expressed protein-coding genes. This list includes all protein-encoding genes with normalized transcript levels of >300 RPKM (in the presence of NH₄⁺) that do not overlap with rRNA or tRNA genes (potentially raising their transcript counts). The x-axis is the log of the normalized transcript level (in RPKM). The y-axis is the log of the ratio of the normalized transcript levels in ammonium-deplete to ammonium-replete conditions. The names of the six genes with repression levels >16-fold are marked. The five red dots correspond to structural genes for the HbRC (pshA, pshX, ferredoxins (pshB1, pshB2), and cyt c₅₅₃ (cyhA)). (C) Photobleaching of P₈₀₀ induced by a 6-ns laser flash (at time t) was monitored in cells using 10-μs pulses of light from an 803-nm LED. The ammonium-replete and ammonium-deplete cultures had A₆₂₅s of 0.42 and 0.30, respectively, at the time of the measurement. Absorption changes at 803 nm are reported as mAU normalized to the A₆₂₅ values of the culture to provide an estimate of cellular HbRC content. The inset is the post-flash data plotted on a log₂ time scale.
in *H. modesticaldum* (Data Set S5, pathway hmo00020). The transcript level of *pckA* encoding PEPCK was reduced by ~16-fold during N$_2$-fixing conditions, which is the largest fold change for any gene in the TCA cycle, which exhibited an average ~3.5-fold down-regulation in transcript level. Malate dehydrogenase, which would reduce OAA to malate in the reverse direction, exhibited an ~11-fold reduction in transcript level, the second highest change. It may be that the organism slows the flow of carbon into the left arm of the TCA cycle when ammonium is low.

We note that there are two genes annotated to be similar to the alpha subunit of OAA decarboxylase (oadA), which are moderately expressed under both conditions (30-120 RPKM). The OAA decarboxylase complex usually consists of 3 subunits – the beta and gamma subunits are integral membrane proteins required for coupling Na$^+$ efflux to decarboxylation of OAA by the alpha subunit [39]. These latter two subunits are not annotated in the genome, so the existence of this enzyme is unclear at this time. We note, however, that the first oadA gene exhibits similarity to pyruvate carboxylase, and is preceded by HM1_0379, which is annotated as a biotin carboxyl carrier protein; both genes are moderately expressed (80-120 RPKM in +NH$_4$+) and exhibit little repression in the absence of ammonium. There is a similar story with the second oadA gene, which is followed by a gene similar to the biotin carboxylase subunit of acetyl-CoA carboxylase (accC). Thus, it is possible that at least one of these putative oadA genes is part of a functional pyruvate carboxylase, which could allow synthesis of OAA in low ammonium.

The key enzyme of the TCA cycle – citrate synthase – appears to be missing in the *H. modesticaldum* genome. There is no gene encoding the typical (S)-citrate synthase, but it has been suggested that this organism uses an (R)-citrate synthase, an enzyme evolutionarily related to homocitrate synthases [7]. There are two genes in the *H. modesticaldum* genome that show high similarity (>40% identity) to the putative (R)-citrate synthase of *Clostridium kluyveri* [46]. The most similar gene (50% identity) is *aksA* (HM1_2993), which is well expressed under both conditions (21-28 RPKM). The other is *nifV* (42% identity) and is part of the *nif* operon. Expression of this gene was increased by a factor of ~33-fold in the absence of ammonium.

We thus propose that this gene is a true homocitrate synthase, and this increase in expression is expected, as homocitrate is an essential part of the FeMoCo cofactor in dinitrogenase, serving as one of the ligands to the Mo metal [47,48]. Helio bacteria do not have the lysine biosynthetic pathway via homocitrate – they use the pathway via aspartate 4-semialdehyde (Data Set S5, pathway hmo00300) – and there is no other known use for homocitrate. Thus, it is likely that the *aksA* gene product is actually a (R)-citrate synthase, as proposed by Tang *et al.* [7]. If so, OAA could be converted to citrate by this enzyme, and thence to 2-oxoglutarate via isocitrate by the “right arm” of the TCA cycle (Figure 2). This would allow the two arms of the TCA cycle to work in the opposite directions and OAA would be a major branch point compound.

PEP would also be a major branch point, as it can be converted to OAA by PEPCK to feed into the TCA cycle, or converted to carbohydrate via the enzymes of gluconeogenesis. As *H. modesticaldum* does not utilize carbohydrates as a carbon or energy source, it is most likely that the enzymes of glycolysis are all used in the anabolic direction (i.e. gluconeogenesis) for making fructose-6-phosphate, which then can be converted into whichever hexoses are needed. The genes involved in the gluconeogenesis pathway were all down-regulated in the absence of ammonium by an average of ~8-fold (Data Set S5, pathway hmo00010). The fructose-6-phosphate generated by this pathway can also be converted to pentose-phosphates via the pentose-phosphate pathway. It should be noted that the oxidative enzymes of this pathway (glucose-6-phosphate dehydrogenase and 6-phosphogluconate dehydrogenase) are absent in this organism, so it would have to use primarily transketolase and transaldolase to convert the hexose-phosphates and triose-phosphates from gluconeogenesis into pentose-phosphates. The genes involved in the pentose-phosphate metabolism were all down-regulated in the absence of ammonium by a factor of ~7.5 (Data Set S5, pathway hmo00030). Gluconeogenesis products also feed into the fructose/mannose and galactose metabolic pathways, allowing production of any needed hexoses. The genes encoding enzymes in these pathways were not significantly affected by the absence of NH$_4^+$ (Data Set S5, pathways hmo00051 and hmo00052).

There has so far been no report of carbon storage polymers in *H. modesticaldum*, such as polyhydroxybutyrate or glycogen/starch. While the genome lacks enzymes for synthesis of the former, there are genes encoding a putative glycogen synthase (glgA) and glycogen branching enzyme (HM1_2990) in the genome. There are two genes for glycogen phosphorylase (glpP); one is highly expressed (108 -135 RPKM) and the other is barely expressed (0.3 – 0.7 RPKM). There is also a debranching enzyme (mluQ). Thus, the minimum complement of enzymes is present for glycogen/starch synthesis and breakdown; they are down-regulated only ~3-fold in the absence of ammonium (Data Set S5, pathway hmo00500). If the rate of carbon incorporation exceeds that of ammonia synthesis during N$_2$-fixing conditions, it is possible that the cells store some of the excess carbon as glycogen.

Expression of enzymes in fatty acid biosynthesis was unaffected by the absence of ammonia (Data Set S5, pathway hmo00061). *H. modesticaldum* does not appear to use fatty acids as a carbon or energy source [5,9]. The only genes that showed significant expression in this pathway were HM1_0073 and aas, which are involved in other pathways, such as fatty acid biosynthesis. The metabolic pathways of lipopolysaccharide and glycolipid synthesis were not significantly affected by the absence of ammonium, while glycerophospholipid synthesis genes collectively showed a ~6-fold decrease in gene expression (Data Set S5, pathways hmo00540, hmo00561 and hmo00564). A deviation from this trend was seen with glkK, which was up-regulated by a factor of ~3.5-fold. This gene encodes glycerol kinase, which can be involved in other pathways.

**Regulation of intermediary metabolism and cofactor synthesis**

The transcripts of genes involved in the synthesis of alanine, aspartate, glutamate, lysine, arginine, proline, glycine, serine, threonine, cysteine, methionine, and D-alanine, as well as the sulfur relay system, were all reduced by 2- to 3-fold on average in the absence of ammonium (Data Set S6). The genes involved in the nitrogen metabolism pathway—which utilizes the products formed during nitrogen fixation in order to synthesize glutamine, arginine and proline—displayed a 5-fold decrease in expression. The same was seen for genes involved in selenocompound metabolism. However, in both the cysteine/methionine and selenocompound metabolism pathways, there was an exception to both trends. The gene metC (cystathionine β-lyase) deviated significantly with an increase in expression of 4.5-fold (Data Set S6, pathway hmo00270). Expression of genes in basal sulfur metabolism was unaffected by the absence of ammonium, except for one of the cysD genes, which was repressed ~9-fold (Data Set S6, pathway hmo00920). Metabolism of the branched aliphatic amino acids (valine, leucine and isoleucine) demonstrated a 7-fold reduction in gene expression, while biosynthesis of the aromatic amino acid tryptophan was unaffected by the absence of ammonium.
acids (histidine, phenylalanine, tyrosine and tryptophan) displayed a higher level of repression in the absence of ammonium (~12-fold). An interesting exception is the two aroF genes, which encode the enzyme catalyzing the first committed step of aromatic amino acid biosynthesis (3-deoxy-7-phosphohexulonate synthase); the more highly expressed one is strongly repressed (~60-fold; Data Set S6, pathway hmm00460; see Figure 4B), while the less highly expressed one is unaffected by the absence of ammonium. Thus, the repression of genes involved in biosynthesis of most amino acids was not very different from the average, with the exception of the aromatic and (to a lesser extent) the branched aliphatic amino acids.

The purine and pyrimidine pathways of *H. modesticaldum* synthesize the ribo- and deoxyribonucleotides that are polymerized to make DNA and RNA, while the amino/sugar nucleotide pathway involves the incorporation of sugars into nucleotides. Peptidoglycan, which forms the cell wall that encases the cell membrane and provides structural support and protection, is synthesized as amino-sugar (rather than nucleotide-sugar) polymers. In the absence of ammonium, the overall expression of genes in the nucleotide and peptidoglycan synthetic pathways decreased by an average of 4-fold. All individual genes were down-regulated by a factor of less than 17-fold or up-regulated by a factor less than 3-fold. Overall, nucleotide metabolism and peptidoglycan synthesis do not seem to be greatly affected by the absence of ammonia, as one might expect (Data Set S6, pathways hmm00230 and hmm00240).

There are several pathways for the biosynthesis of molecules used as cofactors or prosthetic groups. The flux through these pathways is thus smaller than the pathways for intermediary metabolism. For most of these pathways, changes in transcript levels of the enzymes was either not significant or not very large (~6-fold, on average; Data Set S6, pathways hmm00670 to hmm00790). Synthesis of porphyrin is initiated by the conversion of glutamyl-tRNA to 5-aminolevulinate; it then diverges into two pathways: one leads to the synthesis of Vitamin B12, while the other path leads to synthesis of heme and chlorophyll. The porphyrin and chlorophyll biosynthesis pathway was one of the few pathways for which the change in transcript abundance (~9-fold) was statistically significant after the Bonferroni correction (Table 2, Data Set S6, pathway hmm00860). The thiD gene, which encodes phosphomethylpyrrimidine kinase involved in thiamine and Vitamin B6 synthesis, was an outlier, being repressed ~23-fold in the absence of ammonium (Data Set S6, pathways hmm00730 and hmm00750). Other outliers were bchG (encoding chlorophyll synthase, 28-fold), as well as HM1_2396 and cbIB (encoding cobalamin biosynthesis protein, CbiB/CobD; 20-fold and 24-fold, respectively).

**DNA, RNA, and protein metabolism**

The genes involved in replication and repair of genomic DNA, including base excision repair, nucleotide excision repair, mismatch repair, and homologous recombination, were reduced by an average of ~5-fold in the absence of ammonium (Data Set S7). The expression of genes encoding subunits of RNA polymerase and those involved in RNA degradation were unaffected by the absence of ammonium (Data Set S7, pathways hmm03020 and hmm03018).

The ribosome is a major cellular component – ribosomal RNA (rRNA) is by far the most abundant form of RNA and ribosomal proteins are also abundant proteins. On average, expression of rRNA and ribosomal proteins was not affected by the absence of ammonium (Data Set S7, pathway hmm03010). Expression of tRNAs and aminoacyl-tRNA synthetases exhibited a ~3.4-fold reduction in message level (Data Set S7, pathway hmm00970).

Like all Gram-positive bacteria, heliobacteria lack an outer membrane and periplasm. Only one membrane must be crossed to direct a protein outside the cell. Heliobacteria possess both the Sec-SEP pathway and the Twin-arginine targeting (Tat) pathways for protein export. The expression of genes involved in the bacterial secretion systems was not significantly affected by ammonium depletion (Data Set S7, pathway hmm03060).

**Regulation of active transport, flagella, and chemotaxis**

ABC (ATP binding cassette) transporters couple ATP hydrolysis to active transport of a wide variety of molecules. *H. modesticaldum* has 11 different ABC transporter systems, which import phosphate, sulfate/thiosulfate, mineral ions (zinc, nickel/cobalt, tungstate, molybdate), monosaccharides (ribose/xyllose), amino acids (glutamine, spermidine/putrescine, branch chain amino acids), and vitamins (biotin). There are also two ABC-2 type systems for export of teichoic acid (for lipopolysaccharide synthesis) and bacitracin (presumably for resistance to this antibiotic), as well as the FtsEX complex involved in cell division. Expression of genes encoding ABC transporters decreased by an average of 2.6-fold in the absence of ammonia (Data Set S8, pathway hmm02010).

The phosphotransferase system (PTS) is responsible for importing carbohydrates, such as hexoses and disaccharides. In order to prevent the acquired substrate from prematurely exiting the cell, phosphotransferase chemically modifies the substrate via phosphorylation using PEP as the phosphoryl donor. Only one gene in this pathway was identified in *H. modesticaldum*, ptsI (encoding phosphoenolpyruvate-protein phosphotransferase), and was down-regulated by 7-fold in the absence of ammonia (Data Set S8, pathway hmm02060). It is unclear if the PTS system exists in this organism.

Two-component signal transduction systems consisting of a sensor and regulator allow cells to sense changes in environment, such as carbon sources, pH, quorum signals, and osmolarity. Expression of the genes involved in these systems decreased 3.5-fold on average in the absence of ammonium (Data Set S8, pathway hmm02020). A notable exception to this trend is corA, which underwent a 7-fold increase in expression. This gene encodes the carbon storage regulator protein, an RNA-binding protein that is a pleiotropic regulator of many genes. The CorA protein facilitates the destruction of many mRNAs, presumably by binding to their Shine-Dalgarno sequences and inhibiting translation, thus removing the protection from nucleases provided by the ribosome [49-53]. It is possible that increased expression of this protein is at least partially responsible for the decrease in the steady-state level of many other mRNAs.

Bacterial chemotaxis allows *H. modesticaldum* to regulate the speed of rotation and direction of its flagella in response to perceived chemical gradients. Transcripts of genes in this pathway were reduced on average of ~6-fold in the absence of ammonium (Data Set S8, pathway hmm02030). Similarly, the expression of genes involved in flagellar assembly (i.e. flagellar components or assembly proteins) was reduced by ~7-fold (Data Set S8, pathway hmm02040). There were 6 genes in this group that exhibited a moderate increase in expression; exclusion of these 6 genes was required to calculate a change in expression that was statistically significant. The gene that demonstrated the largest increase in expression, *flgM*, encodes for a negative regulator for flagellin synthesis. It is unclear why the transcripts of some genes involved in flagellar synthesis are increasing, while those of most of the genes are decreasing in abundance after the shift to ammonium-deplete medium.
Importance

The Heliobacteria appear to be the result of a lateral gene transfer of photosynthetic genes into an ancestral bacterium likely resembling a Clostridial cell, and indeed they can grow in the dark fermentatively like their non-photosynthetic cousins. The transcriptome reported here provides insights into how such a hybrid organism operates under normal conditions, and gives us the first view at a large scale of how the genetic information in its genome can be used under different conditions, such as when it is forced to use atmospheric N₂ as a source of nitrogen. These results should prove to be foundational for future studies on this particular organism (H. modesticaldum), which is rapidly becoming a model organism for the Heliobacteria.

This study reports the first transcriptome of H. modesticaldum. We also report the genome-wide effect that the absence of NΗ₄⁺ has on expression. These findings show that the shift from ammonium-replete to N₂-fixing conditions is accompanied by a general decrease of ~4-fold in 74% of the genome. There was also increased expression of nitrogenase and the high-affinity ammonium transporter, along with a few other proteins, including some regulators (e.g. PII proteins and csrA). We also observed a significant decrease in the level of the mRNA encoding the photosynthesis RC core polypeptide and electron donors/acceptors (cyt c₅₅ and ferredoxins), which are among the most highly expressed protein-coding genes in ammonium-replete conditions. However, functional assays of the RC revealed only a modest decrease of the protein in vivo, indicating a significant role for regulation of translation or post-translational processes. Proteomic studies will be required in the future to unravel these contradictory responses. Future applications of the two H. modesticaldum transcriptomes reported will likely benefit from additional RNA-seq studies using a variety of growth conditions and comparative RNA-seq studies of other heliobacterial species.

Acknowledgement

This work was partially supported by the Division of Chemical Sciences, Geosciences, and Biosciences, Office of Basic Energy Sciences, of the U.S. Department of Energy through grant DE-SC0010575 to Kevin Redding. Daniel Sheehy was supported by funds from the Howard Hughes Medical Institute through the Undergraduate Science Education Program and from the School of Life Sciences Undergraduate Research Program (SOLUR) at Arizona State University. We thank Scott Bingham at the Genomics Core at Arizona State University for all sequencing performed on the Ion Torrent PGM. The following ASU Honors students and undergraduates were supported: Anoosha Murella, Daniel Murphy, Miranda Ngan, Geneva Nguyen, Dixit Cheung, Devin Corrigan, Paige Davis, Fredrick Frost, Melynn Grebe, Amritpal Singh, and Kiri Olsen. Amritpal Singh was supported by funds from the Howard Hughes Medical Institute through the Science Undergraduate Research Program (SOLUR) at Arizona State University. Sheehy was supported by funds from the Howard Hughes Medical Institute through the Undergraduate Science Education Program and from the School of Life Sciences Undergraduate Research Program (SOLUR) at Arizona State University. We thank Scott Bingham at the Genomics Core at Arizona State University for all sequencing performed on the Ion Torrent PGM. The following ASU Honors students and undergraduates were supported: Anoosha Murella, Daniel Murphy, Miranda Ngan, Geneva Nguyen, Kiri Olsen, Dixit Cheung, Devin Corrigan, Paige Davis, Fredrick Frost, Melynn Grebe, Amritpal Singh, and Kiri Olsen.

References


