Organization and Sequence of the Genes Coding for the Proton-translocating ATPase of *Bacillus megaterium*

William S. A. Brusilow†, Marco A. Scarpetta, C. Ann Hawthorne, and William P. Clark

From the Department of Chemistry and Biochemistry, University of Maryland, College Park, Maryland 20742

We have cloned and sequenced the genes for the subunits of the proton-translocating ATP synthase of *Bacillus megaterium* QM B1551. The arrangement of the genes is identical to the arrangement of the same genes (the unc operon) in *Escherichia coli*. The genes for the F_{0} subunits immediately precede the genes for the F_{1} subunits and are themselves preceded by an open reading frame which codes for a protein similar to the *E. coli* i protein. In contrast to the *E. coli* ATPase genes, the transcript for these ATPase genes does not include this open reading frame.

Proton-translocating ATP synthases (ATPases) catalyze the interconversion of a transmembrane proton gradient with cellular ATP levels. These enzymes are found in bacteria, plants, and animals, and thus synthesize most of the ATP in living systems. The structures of all ATP synthases are very similar. They consist of two large sectors. The membrane, the cytoplasmic membrane of *Bacillus megaterium* contains the catalytic sites for the synthesis or hydrolysis of ATP.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EMBL Data Bank with accession number(s) J04455.

The sequences of those genes, and thus the gene order, gene-polypeptide relationships, and the protein sequences of all the subunits, was determined in a series of studies from several laboratories (3–13).

We cloned the genes for several of the F_{1} subunits of the Gram-positive, sporulating obligate aerobe *Bacillus megaterium* QM B1551 (14). We demonstrated that cloned genes for the α or β subunits of the *B. megaterium* ATPase could complement *E. coli* α or β mutants. Therefore, the *B. megaterium* subunits are capable of being synthesized by *E. coli* and being assembled into functional hybrid *E. coli-B. megaterium* ATPase complexes. In order to characterize subunit-subunit interactions in such complexes, we have sequenced the ATPase genes from this organism. We previously reported the sequences of the genes for the β and ε subunits and also demonstrated that our original cloned DNA contained the genes for only four of the F_{1} subunits, α, γ, δ, and ε (15). In this paper we report the cloning of the F_{0} genes and the gene for the δ subunit, the nucleotide sequences of the F_{0} genes and the first three F_{1} genes, and the deduced amino acid sequences of the ATPase subunits coded for by those genes.

**MATERIALS AND METHODS**

Restriction endonucleases, T4 DNA ligase, and exonuclease III were purchased from Bethesda Research Laboratories. S1 nuclease and alkaline phosphatase (calf intestine) were purchased from Boehringer Mannheim. 32P-TP (3000 Ci/mmol) was purchased from Du Pont-New England Nuclear.

DNA sequencing was carried out by the method of Sanger et al. (16) as modified by Bethesda Research Laboratories (BRL Sequencing Manual). DNA fragments generated by digestion with restriction enzymes were cloned into M13mp18 or M13mp19 (17). In addition, overlapping deletion subclones were prepared by treatment of cloned DNA with exonuclease III and S1 as described by Henikoff (18).

The mapping of the 5′ end of the RNA transcript was performed as described by Davis et al. (19). mRNA was isolated from exponentially growing *B. megaterium* as described by Ulmanen et al. (20). Using primer extension, a single-stranded sense-strand DNA probe was synthesized from a template of single-stranded anti-sense DNA cloned in M13mp18. The resultant double-stranded DNA was digested with EcoRI, and the probe was separated from the vector on a denaturing agarose gel. The probe contained 159 bases of the *atpB* gene, the complete *atpI* gene, and the sequence between the *atpI* gene and the EcoRI site at the left end of the cloned ATPase genes (Fig. 1). This probe was used to test for transcriptional starts within *atpI*. The probe was annealed to either *B. megaterium* mRNA or a tRNA control, then treated with S1 nuclease, and electrophoresed on a sequencing gel alongside molecular weight standards which consisted of a set of sequencing reactions. A second labeled probe containing 144 bases of *atpI* and the sequence between the *atpI* gene and the EcoRI site at the left end of the cloned DNA was also hybridized to both *B. megaterium* and *E. coli* mRNA to test for transcriptional start sites preceding *atpI*.

A library of 4–5–kb1 EcoRI fragments from the *B. megaterium* chromosome was constructed in plasmid pACYC184 (21). The chromosomal DNA was digested with EcoRI, the fragments were separated

---

*This work was supported by Grant-in-Aid 87-1297 from the American Heart Association Maryland Affiliate, Inc. Additional support was provided by Biomedical Research Support Grant RR-0742 to the University of Maryland from the National Institutes of Health (NIH). Computer resources used to carry out our studies were provided by the NIH sponsored BIOMET™ National Computer Resource for Molecular Biology, Grant 1 U41 RR-01685. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The abbreviation used is: kbp, kilobase pairs.
on an agarose gel, and those in the 4-5-kbp range were isolated as described previously (22). Those DNA fragments were ligated into pACYC184 which had been digested with EcoRI and treated with alkaline phosphatase. Recombinant plasmid DNA isolated from individual transformants was digested with EcoRI and BanHI (which cut the 4.0-kbp vector in half but did not digest the desired insert), electrophoresed on an agarose gel, then screened for hybridization to a probe containing previously cloned B. megaterium atp genes. One such clone contained the desired insert, and the plasmid consisting of the 4.3-kbp EcoRI fragment in pACYC184 was designated pWPC208. Plasmid pCAH1.3 was constructed by treating pCAH1 (14) with Sphi and ligating the resultant fragments, producing a smaller plasmid than pCAH1 but which still carried the last four genes for the F, subunits. Plasmid pWSB100 was constructed by cloning the EcoRI fragment from pWPC208 into pCAH1.3 which had been digested with EcoRI. The resultant plasmid contained all nine genes for which we have now determined the sequences.

RESULTS

Cloning the Genes for the F, Subunits—Our original cloned B. megaterium DNA in plasmid pCAH1 contained only the genes for the α, β, γ, and ε subunits. We used part of that DNA as a probe to clone the chromosomal DNA adjacent to the end carrying the gene for the α subunit. Hybridization analysis of chromosomal DNA digested with different enzymes showed that there was an EcoRI site approximately 4.5 kbp upstream of the EcoRI site present in our cloned DNA. We created a library of 4-5-kbp EcoRI fragments cloned into the EcoRI site of pACYC184 and screened that library for sequences which would hybridize to one end (the α end) of our cloned B. megaterium ATPase DNA. Sequence analysis of one of our hybridization-positive clones showed that we had cloned the desired fragment.

Sequencing the Genes for the δ, α, and γ Subunits—Our sequencing strategy involved cloning the 4.3-kbp EcoRI fragment into the EcoRI site of M13mp18, in both orientations, then constructing sequential overlapping deletions with ExoII, as described by Henikoff (18). We also constructed various well-defined deletions using restriction enzyme recognition sites which were discovered during the sequencing. We therefore completely sequenced both strands of the B. megaterium ATPase DNA.

Fig. 1 shows the restriction map, gene order, and gene-polypeptide relationships for the ATPase genes of B. megaterium, designated atp genes. Our criteria for determining the locations and identities of each gene were based on homology to the same genes and subunits of the E. coli ATPase, and we have given each gene the same designation as its equivalent gene in the E. coli unc operon. The sequence itself is shown in Fig. 2. An additional criterion for determining where each gene started was the existence of a translational initiation region containing a ribosome-binding region with the same sequence as the E. coli (23) or Bacillus subtilis (24) ribosome-binding regions (Shine and Dalgarno regions). The deduced translational initiation regions for each gene are underlined in Fig. 2. Table I presents the deduced molecular weights of each subunit, as well as the number of identical amino acids found in the same positions of the same subunit from E. coli and B. megaterium.

Determining the Transcriptional Start Site—We determined the 5’ end of the B. megaterium atp gene transcript using standard procedures. The results (Fig. 3) show that a promoter exists near the 3’ end of the atpC gene. Fig. 3 shows the major start site, at base pair 574 ± 2 and the potential −10 and −35 regions. Inspection of the sequence in this region reveals more than one set of possible −10 and −35 regions. Determination of the exact promoter-binding sites awaits detailed footprinting experiments. It is clear, however, that atpC is not necessarily part of the same transcript as atpB. Further analysis using a different probe (Fig. 3, probe A) demonstrated that there is no promoter in the 270 bases preceding the atpC gene (data not shown).

DISCUSSION

The sequences of ATPase genes from a variety of bacterial sources have been determined. The sequences of the F1 genes were determined for the purple nonsulfur photosynthetic bacteria Rhodopseudomonas japonica and Rhodospirillum rubrum (25, 26). In those species, the F1 genes are not adjacent to the F0 genes, as they are in E. coli. For the cyanobacterial species Synechococcus sp. strain PCC 6301 and Anabaena sp. strain PCC 7120, both the F0 and F1 genes were sequenced, but the genes in these two species are not all contiguous (27–29). The F0 genes are adjacent to the genes for the δ, α, and γ subunits, but the genes for the β and ε subunits are not. Additionally, the gene for the β subunit of the thermophilic bacterium PS3 has been cloned and sequenced (30).

The arrangement of ATPase genes in B. megaterium is virtually identical to the arrangement of ATPase genes in E. coli and represents a second example of all the ATPase genes being located together in a bacterial chromosome. The significance of this research lies both in the comparison of nucleotide and amino acid sequences and in the potential for using combinations of B. megaterium and E. coli ATPase genes to construct and study functional hybrid ATPase molecules. We have demonstrated previously that B. megaterium α or β subunits can be assembled into hybrid E. coli-B. megaterium ATPase complexes (14). The α and β subunits therefore contain the individual amino acids and protein domains necessary for both assembly and function in hybrid E. coli-B. megaterium F1 ATPase complexes. A comparison of sequences of these subunits from the two sources will therefore presumably reveal information about which amino acids and which regions of each subunit are required for assembly and activity.

Knowing the exact nucleotide and amino acid sequences will
FIG. 2. Nucleotide and deduced amino acid sequences of *B. megaterium* ATPase genes and subunits. Base 1 is the G of the EcoRI site at the left end of our cloned DNA (see Fig. 1). There is a partial unidentified reading frame preceding the first gene, designated as ‘URF.' The putative ribosome binding sites (Shine and Dalgarno sequences) are underlined. Neither *atpH* nor *atpA* are preceded by any 4-base Shine and Dalgarno sequences, but each is preceded by two 3-base sequences, both of which are reading frame preceding the first gene, designated as ‘URF.'
now permit the construction of hybrid subunits to test individual regions of each subunit for their involvement in assembly. Cross-species complementation studies have not yet been undertaken with the other genes.

The sequence of these genes reveals a number of interesting characteristics. Two of the genes start with non-AUG initiation codons. The gene for the α subunit (atpB) begins with a UUG. The gene for the β subunit (atpG) also begins with a UUG, although it is possible that the initiation codon is a GUG. Without amino acid sequence information, we cannot be sure of the start codon, but comparison to the N-terminal sequences of other β subunits (31), and the positioning of the UUG relative to the ribosome-binding region argues for the UUG start codon. It has been shown in E. coli that the UUG initiation codon results in significantly decreased expression compared to the AUG or GUG initiation codons (32). Since the α and β subunits are each present in a single copy/ATPase complex, the presence of a UUG initiation codon may be one mechanism for insuring that neither subunit is overproduced.

The spaces between genes are larger for the B. megaterium genes than for the E. coli genes. There are approximately 100-base pair intergenic regions on either side of the atpE gene, which codes for the dicyclohexylcarbodiimide-binding c subunit of the F_0. The space between atpA and atpG, which code for the α and β subunits respectively, is also about 100 base pairs. In E. coli, each of these three intergenic regions is about 50 base pairs long.

For these genes and the other two sets of cloned and sequenced F_0 genes, there is an open reading frame preceding the start of the gene for the α subunit of the F_0. In E. coli, this gene is uncI, and the protein it codes for is the i protein. Although the protein can be synthesized in vitro and in minicells, it has never been detected in vivo. Mutations in

### Table I

| Amino acid compositions of the predicted protein subunit products of the B. megaterium atp genes. At the bottom are listed the total number of amino acids in each subunit, the deduced molecular weights, and the number of amino acids in each B. megaterium subunit which are identical to an amino acid in the same position of the homologous E. coli subunit, as determined by the ALIGN program of Bionet |

<table>
<thead>
<tr>
<th>Amino acid</th>
<th>Subunit</th>
<th>i</th>
<th>a</th>
<th>c</th>
<th>b</th>
<th>α</th>
<th>γ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alanine</td>
<td></td>
<td>9</td>
<td>18</td>
<td>12</td>
<td>21</td>
<td>16</td>
<td>42</td>
</tr>
<tr>
<td>Cysteine</td>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Aspartic acid</td>
<td></td>
<td>3</td>
<td>6</td>
<td>0</td>
<td>6</td>
<td>7</td>
<td>26</td>
</tr>
<tr>
<td>Glutamic acid</td>
<td></td>
<td>2</td>
<td>6</td>
<td>3</td>
<td>27</td>
<td>14</td>
<td>39</td>
</tr>
<tr>
<td>Phenylalanine</td>
<td></td>
<td>7</td>
<td>18</td>
<td>2</td>
<td>5</td>
<td>6</td>
<td>15</td>
</tr>
<tr>
<td>Glycine</td>
<td></td>
<td>8</td>
<td>19</td>
<td>10</td>
<td>8</td>
<td>7</td>
<td>47</td>
</tr>
<tr>
<td>Histidine</td>
<td></td>
<td>5</td>
<td>0</td>
<td>2</td>
<td>6</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>Histonine</td>
<td></td>
<td>10</td>
<td>18</td>
<td>9</td>
<td>13</td>
<td>14</td>
<td>36</td>
</tr>
<tr>
<td>Lysine</td>
<td></td>
<td>5</td>
<td>9</td>
<td>2</td>
<td>16</td>
<td>14</td>
<td>25</td>
</tr>
<tr>
<td>Leucine</td>
<td></td>
<td>17</td>
<td>35</td>
<td>17</td>
<td>16</td>
<td>18</td>
<td>50</td>
</tr>
<tr>
<td>Methionine</td>
<td></td>
<td>3</td>
<td>12</td>
<td>4</td>
<td>7</td>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td>Asparagine</td>
<td></td>
<td>0</td>
<td>4</td>
<td>8</td>
<td>14</td>
<td>10</td>
<td>9</td>
</tr>
<tr>
<td>Proline</td>
<td></td>
<td>3</td>
<td>8</td>
<td>2</td>
<td>1</td>
<td>5</td>
<td>20</td>
</tr>
<tr>
<td>Glutamine</td>
<td></td>
<td>5</td>
<td>4</td>
<td>2</td>
<td>15</td>
<td>10</td>
<td>24</td>
</tr>
<tr>
<td>Arginine</td>
<td></td>
<td>8</td>
<td>4</td>
<td>2</td>
<td>8</td>
<td>11</td>
<td>29</td>
</tr>
<tr>
<td>Serine</td>
<td></td>
<td>6</td>
<td>15</td>
<td>3</td>
<td>8</td>
<td>14</td>
<td>31</td>
</tr>
<tr>
<td>Thrething</td>
<td></td>
<td>7</td>
<td>10</td>
<td>9</td>
<td>4</td>
<td>10</td>
<td>29</td>
</tr>
<tr>
<td>Valine</td>
<td></td>
<td>15</td>
<td>18</td>
<td>7</td>
<td>14</td>
<td>17</td>
<td>43</td>
</tr>
<tr>
<td>Tryptophan</td>
<td></td>
<td>2</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Tyrosine</td>
<td></td>
<td>8</td>
<td>8</td>
<td>0</td>
<td>1</td>
<td>4</td>
<td>14</td>
</tr>
<tr>
<td>Total amino acids</td>
<td>123</td>
<td>236</td>
<td>70</td>
<td>172</td>
<td>181</td>
<td>502</td>
<td>285</td>
</tr>
<tr>
<td>M_0</td>
<td>14,032</td>
<td>26,088</td>
<td>6,952</td>
<td>19,379</td>
<td>20,386</td>
<td>54,609</td>
<td>31,228</td>
</tr>
<tr>
<td>No. amino acids identical to E. coli subunit</td>
<td>32</td>
<td>81</td>
<td>28</td>
<td>35</td>
<td>39</td>
<td>268</td>
<td>114</td>
</tr>
</tbody>
</table>
Sequence of the B. megaterium H⁺ ATPase Genes

uncI do not affect the activity of the ATPase, so the function of this gene and the protein it codes for are unknown (33, 34). There is little amino acid or nucleotide sequence homology between the I genes of E. coli, Synechococcus, or B. megaterium, although the I proteins of B. megaterium and E. coli have a higher percentage of identical amino acids than the more poorly conserved b or a subunits (Table I). Cozens and Walker (27) analyzed the hydropathy characteristics of the E. coli and Synechococcus proteins and found a strong similarity in the hydrophobicity plots of the two. We carried out the same analysis on the B. megaterium protein and found that it resembles the E. coli protein very closely. (Fig. 4). Therefore, even though the sequences are only slightly similar, they appear to code for proteins with similar characteristics.

FIG. 3. Transcriptional start site for the B. megaterium atp genes. The left side of the figure shows the first 1000 bases of our sequenced DNA and the limits of the atpI and atpB coded for by that DNA. Horizontal lines beneath the genes indicate the lengths and limits of the two labeled DNA probes which were hybridized to B. megaterium mRNA followed by treatment with S1-nuclease. The bottom line indicates the amount of probe B protected by the mRNA. The right side of the figure shows the result of one such experiment with probe B. Lane 1 is untreated probe B. Lane 2 is probe B treated with E. coli tRNA followed by S1 digestion. Lane 3 shows probe B annealed to B. megaterium mRNA followed by S1 digestion. To the right of lane 3 is a sequencing ladder used to determine the length of the protected fragment. The lengths of the DNA represented by the bands in the sequencing ladder are indicated on the right. The equivalent experiments with probe A are not shown. At the bottom of the figure is the sequence of the promoter region. The transcriptional start site ± 2 bases is indicated by the arrow, and possible -10 and -35 regions are also indicated.

FIG. 4. Hydropathy profiles of the I protein of E. coli and the predicted product of the B. megaterium atpI gene. Values were calculated using the PEP program in Bionet with a span of 6. The scales on the abscissae represent amino acid residues. Values above the horizontal line represent hydrophilic regions and values below the line represent hydrophobic regions.
Even though mutations in this gene in *E. coli* do not appear to affect ATPase function, the conservation of this gene indicates that it must serve some function.

Our analysis of transcriptional start sites in our cloned DNA add to the confusion about the atpI gene. In *B. megaterium*, the transcriptional start for at least atpB, if not for all the subunit-encoding genes, is within the atpI gene. This gene itself appears to be part of a different transcriptional unit. These results strengthen the conclusions from mutational studies that the I gene is not involved in ATPase synthesis, assembly, or activity. Its presence as part of the *E. coli* unc operon may be a conservation of location (e.g. in close proximity to the origin of replication) rather than an indication of a relationship to the ATPase genes.

Studying the synthesis and assembly of the *E. coli* ATPase is complicated by the fact that the genes are differentially translated from the same transcript (35–37). McCarthy et al. (36–38) have demonstrated that the intergenic region preceding uncE (c subunit of the F) contains a sequence which substantially increases the translation of any gene it precedes. The mechanism for this translational enhancement is not understood. One of the sequence features of this region is a string of Us followed by a region of interrupted As. The same region of the *B. megaterium* ATPase genes contains two such sequences, which are overlined in Fig. 2. We have not yet tested the ability of these sequences to enhance translation of the *B. megaterium* atpI gene.

One of the more interesting features of the sequences of these genes is that the genes for the b subunit of the F, and the δ subunit of the F, overlap by four bases. An identical overlap has been reported for the b and δ genes of the ATPase of *Anabaena* sp. strain PCC 7120 (28). One model for the structure of the *E. coli* ATPase hypothesizes an interaction between the b and δ subunits which is important in binding the F1 to the F0 (39). The overlapping of these two genes may facilitate either the translational coupling of the expression of these genes or some essential assembly step involving these two subunits.

REFERENCES