Clustering of Genes Necessary for Hydrogen Oxidation in *Rhodobacter capsulatus*†

HONG-WU XU‡ and JUDY D. WALL*

Biochemistry Department, 117 Schweitzer Hall, University of Missouri—Columbia, Columbia, Missouri 65211

Received 10 September 1990/Accepted 30 January 1991

Three cosmids previously shown to contain information necessary for the expression of uptake of hydrogenase in *Rhodobacter capsulatus* were found to be present in a cluster on the chromosome. Earlier genetic experiments suggested the presence of at least six genes essential for hydrogenase activity that are now shown to be in a region of approximately 18 kb that includes the structural genes for the enzyme. A potential response regulator gene was sequenced as a part of the *hup* gene region.

Certain microbial enzymes, hydrogenases, catalyze H₂ evolution for the removal of excess reductant in anaerobes or H₂ oxidation as the first step in the use of H₂ as an energy source. In the photosynthetic bacterium *Rhodobacter capsulatus*, H₂ evolution appears to occur only as an obligate by-product of nitrogenase activity (21). In contrast, H₂ consumption is mediated by a membrane-bound hydrogenase (5).

*R. capsulatus* appears to express a single hydrogenase (19), with subunits of 67 and 31 kDa, that is immunologically related to the dimeric hydrogenases from several aerobic diazotrophs (15). Physiological studies of *R. capsulatus* suggested that the enzyme was a nickel-metallo protein (20), and nickel was later shown to be present in the purified enzyme (5).

The genes encoding the two subunits of hydrogenase have been identified by homology with a probe containing the *Bradyrhizobium japonicum* structural genes (9) and by complementation of photoautotrophic mutants lacking hydrogenase activity (23). Information concerning the number and location of other genes essential for hydrogen oxidation by *R. capsulatus* is still fragmentary. Three cosmids were isolated by Vignais and coworkers that appear to complement most of their Hupᵐ⁻ mutants (3, 9). Recently, these cosmids have been shown to contain sequences nearly contiguous on the chromosome (4), and it was suggested that a stretch of about 15 kb of DNA is necessary for the expression of hydrogenase genes in *R. capsulatus*.

Independently, we isolated three small cosmids that had unique restriction patterns and complemented different sets of Hupᵐ⁻ mutations (23). To confirm the arrangement of *hup* genes on the chromosome and to relate our cosmids to those reported by others, we sought to determine the chromosomal arrangement of the DNA found in our cosmids. We found these cosmids to be nearly contiguous on the chromosome, forming a cluster of approximately 18 kb. In contrast to the conclusion drawn by Colbeau et al. (4) that one of the cosmids used in this study may represent DNA unique from that isolated in their work, we find the data support a single cluster that includes all our cosmids. The exact content of hydrogenase-specific information in this region and the functions encoded have yet to be determined. To begin to identify Hup functions, a 3.1-kb DNA fragment believed to encode at least two essential genes has been sequenced. The presence of sequences with homology with regulator genes of two component signal transduction systems of procaryotes was found.

All analyses were of DNA from *R. capsulatus* B100, grown as described previously (23). The cosmids pRHP4, pRHP8, and pRHP20 were isolated by complementation of *hup* mutations (23) and are pLAFR1 with *R. capsulatus* Hup-specific DNA inserts. The cosmids were maintained in and prepared from *Escherichia coli* HB101. Individual *EcoRI* fragments used as hybridization probes were subcloned in pUC18 (24) and were maintained and prepared from *E. coli* DH5α (8). Cosmid A163 was prepared from a partial *Sau3A* genomic digest of *R. capsulatus* DNA in a *Lorist* 2 vector (7) and was a generous gift from J. Williams. Chromosomal and plasmid DNA were prepared as previously described (23). Probes for restriction mapping by Southern analysis were subcloned *EcoRI* DNA fragments from the cosmids or, alternatively, were DNA fragments cut from agarose gels and purified by a second electrophoresis into low-melting-point agarose (Fisher BioTech, Fair Lawn, N.J.). Probes were labeled according to the manufacturer’s instructions included with the oligo-labeling kits (United States Biochemical Corp., Cleveland, Ohio, and Pharmacia LKB Biotechnology, Piscataway, N.J.) Restriction endonuclease digestions of chromosomal and plasmid DNA were performed as instructed by the suppliers of the enzymes. For Southern analysis, DNA restriction fragments separated by electrophoresis were transferred to Zeta-Probe nylon membranes by the alkaline blotting technique described in the BioRad Technical Bulletin 1233 (BioRad Laboratories, Richmond, Calif.). Hybridization was carried out by the alternative protocol in the BioRad Zeta-Probe blotting membrane instruction manual.

The 3.1-kb *EcoRI* fragment from pRHP20 was subcloned into pUC18 and pTZ18U (United States Biochemicals) for DNA sequencing. Nested deletions were generated with *ExoIII* nuclease treatment, as described in the Erase-a-Base kit (Promega, Madison, Wis.). Appropriate fragments were

* Corresponding author.
† This work is journal series number 11.292 from the Missouri Agriculture Experiment Station.
‡ Present address: Genetics Institute, Andover, MA 01810-5901.
sequenced by the Sanger dideoxy-chain termination method (14) according to the instructions included with the Se-que- nase kit (United States Biochemicals). All sequencing data were generated at least twice on both strands, and dITP and 7-deaza-α-deoxy GTP were substituted for dGTP whenever necessary to resolve sequences with strong secondary structure. After manual collection of primary sequence data, analyses were performed with the University of Wisconsin Genetics Computer Group software (6) and the FASTA program (11).

To investigate the relative chromosomal arrangement of the DNA from the three cosmids previously shown to contain information essential for the expression of hydrogenase in *R. capsulatus* (23), Southern analyses were performed. EcoRI DNA fragments from each of the cosmids were used as probes for *R. capsulatus* chromosomal DNA digested with HindIII or PstI. Figure 1 shows the tentative restriction map derived from these results and the previously deduced restriction map of each cosmid (23).

To confirm the map, a Lorist cosmid, A163, containing about 35 kb of *R. capsulatus* DNA was identified by J. Williams (DuPont Chemical Co., Wilmington, Del.) that had homology to all three of our pLAFR1 cosmids. A163 was generously made available to us, and restriction digestion with EcoRI, HindIII, and PstI revealed a pattern of DNA bands compatible with the derived restriction map (data not shown). A 9.5-kb HindIII fragment and an 8.8-kb PstI fragment were generated from A163, gel purified, labeled, and used to probe EcoRI digests of each of the three cosmids containing Hup sequences. EcoRI fragments from pRHP4 and pRHP20 hybridized with both a chromosomal 9.5-kb HindIII fragment and the HindIII probe of the same size from A163. Likewise, fragments from pRHP20 and pRHP8 had an homology to an 8.8-kb PstI fragment from the chromosome and A163. The PstI probe also revealed a number of small EcoRI fragments present in A163 that were ordered by subcloning and additional restriction analysis.

Thus, the DNA fragments from A163 confirmed the order and relative chromosomal arrangement of our previously isolated *hup* cosmids. The region required for complementation of our present collection of *Hup*− mutants (23) spans approximately 17 kb of contiguous DNA.

Of the two cosmids not containing structural gene information, pRHP20 has been studied in greater detail. The 3.1-kb EcoRI fragment from this cosmid has been shown to be sufficient for the complementation of four different *Hup*− mutations (23). To explore the nature of the *hup*-specific sequences in this region, the 3.1-kb fragment was sequenced. The DNA and putative polypeptide sequences are shown in Fig. 2. The fragment was found to be 3,053 bp long.

A search for open reading frames (ORFs) preceded by a possible ribosomal binding site revealed three apparently complete ORFs in one orientation and a single large ORF in the reverse. Codon preference comparisons with a table derived from published *R. capsulatus* sequences supported the interpretation that the orientation with multiple ORFs was the one providing information for hydrogenase biosynthesis. The three complete ORFs, ORF2, -3, and -4 (Fig. 2 and 3) have the potential for encoding polypeptides of approximately 30, 12, and 35 kDa, respectively. The EcoRI sites used to generate the fragment appear to have interrupted two additional ORFs at either end of the 3.1-kb fragment. It is possible that this fragment is a part of a large operon. Additional sequencing and studies of the mRNA produced will be necessary to establish the overall genetic structure.

Transposon mapping had been used to identify the portions of the 3.1-kb EcoRI fragment from pRHP20 necessary for complementation of four *hup* mutations (23) (Fig. 3). Insertions in two regions affected complementation. Tn5 in positions 5, 6, and 7 eliminated complementation of *hup*-101, and insertions at positions 9 and 10 prevented complementation of *hup*-42 and reduced that of *hup*-36 and *hup*-43 (23). If this region is a part of a large operon, it is surprising that strong polarity is not evident, although other cases in which an operator-proximal Tn5 is not strongly polar on distal genes have been reported (22). It is also possible that multiple promoters naturally occur within the sequence or that Tn5 provides an adventitious promoter for downstream genes. Further analysis will be necessary to establish which of these alternatives may be operative.

Taken together with sequencing information, the complementation results of the insertionally interrupted DNA support the assumption that one or more of the ORFs encoded in this fragment are expressed. Attempts to establish this possibility in the T7 RNA polymerase and promoter system (18) have been unsuccessful.

A FASTA protein sequence comparison revealed no inter- pretable homologies to the putative polypeptides from ORF1−2, -3, or -4. In contrast ORF5 was shown to have 21 to 30% identity over more than 100 amino acids with the N terminus of proteins known to be the response regulator components of bacterial two-component signal transduction systems (16) (Fig. 4). In fact, the three strictly conserved residues Asp-13, Asp-57, and Lys-109 (CheY numbering [16]) are present in this peptide. The N terminus of HydG,

FIG. 1. Restriction endonuclease map of the *R. capsulatus* chromosomal region containing *hup* genes. The relative locations of the three cosmid inserts previously described (23) are indicated below the composite map. The two restriction fragments from the Lorist recombinant A163 that were used to confirm the arrangement of the three cosmids are also shown. Hup structural genes, *hup*SL, are contained almost entirely within pRHP8. EcoRI; HindIII; PstI.

FIG. 2. Nucleotide and derived polypeptide sequences of the 3.1-kb EcoRI fragment from pRHP20. The deduced amino acid sequences corresponding to the putative ORFs are shown below the DNA in standard one-letter amino acid designations. The regions that could serve as ribosome-binding sites for translation of the putative ORFs are underlined.
transcriptional activator for the labile hydrogenase genes of *E. coli* (17), and ORF5’ are 24% identical. However, the two peptides do not show any increase in similarity relative to other members of these signal proteins. From the sequence analysis and the map position, ORF5’ would appear to be the 5’ end of the *hupR* gene shown to be necessary for expression of the hydrogenase structural gene (13). A candidate for the cognate protein kinase has not yet been identified.

Visual examination of the amino acid sequences of the translation products of the remaining putative ORFs suggested that ORF4 was unusual in having a high histidine content in the amino terminus. The arrangement was not random. Beginning at nucleotide 1667 (Fig. 2), a motif of at least six histidines alternating with hydrophobic amino acids was repeated three times in close proximity, i.e., HXHXHX HXHXHX, where X is a hydrophobic residue. A FASTA search of data bases revealed a similar motif in the homeotic protein of drosophila encoded at the *Deformed* locus (12). Because this deduced homeotic protein has several conspicuous regions of monotonic amino acid sequences, the significance of the homology is questionable. It is tempting to speculate that these three regions of histidines may function in metal metabolism or ion movements. This motif is absent from the hydrogenase genes sequenced for *E. coli* (2, 10, 17) and the 24-kb nif region of *Klebsiella pneumoniae* (1). Further mutagenesis will be necessary to determine the significance, if any, of this motif for hydrogen metabolism.

In summary, the three previously described cosmid isolates by complementation of *hup* mutations of *R. capsulatus* (23) were found to be clustered, with the structural genes occurring at the right end as drawn in Fig. 4. A potential response regulator gene was found 5.5 kb downstream and is apparently transcribed in the same direction as the structural genes.

**Nucleotide sequence accession number.** The sequence reported here appears in the GenBank Nucleotide Sequence Database under accession number M-55089.

---

**Fig. 3.** Diagram of relative locations of ORFs and Tn*δ* inserts in the 3.1-kb *EcoRI* fragment from pRHP20 (23). Numbered arrows above the ORFs indicate the locations of Tn*δ* inserts determined by restriction mapping. The ORF locations were derived from the data in Fig. 2. E, *EcoRI; H, HindIII.*

---

We thank Joseph Forrester, Associate Director of the DNA Core Facility of the University of Missouri—Columbia, for patient sequence analysis, Frank Schmidt for codon preference plots, and Barbara Rapp-Giles for technical assistance.

This work was supported by grant DE-FG02-87ER13713 from the Department of Energy.

**REFERENCES**


