The gene encoding cAMP receptor protein is required for competence development in Haemophilus influenzae Rd

\((crp/catabolite\ gene\ activator\ protein/bacterial\ transformation)\)

MARK S. CHANDLER

Department of Molecular Biology and Genetics, Johns Hopkins University School of Medicine, Baltimore, MD 21205

Communicated by Hamilton O. Smith, November 18, 1991 (received for review September 30, 1991)

ABSTRACT  The Haemophilus influenzae Rd strain JG87 contains a single mini-Tn10kan insertion that causes a deficiency in the development of competence for genetic transformation. The DNA fragment containing this insertion mutation, as well as the wild-type locus, was cloned, mapped, and sequenced. The sequence contained an open reading frame that encoded a protein with a predicted amino acid sequence similar to the Escherichia coli cAMP receptor protein. The E. coli crp gene cloned on a multicopy plasmid was shown to fully complement the competence-deficient phenotype of the mutant strain; thus, the H. influenzae gene was named crp. These results suggest that H. influenzae cAMP-cAMP receptor protein complex functions to regulate one or more promoters essential for the development of competence in H. influenzae Rd. Features of a gene upstream of H. influenzae crp that is homologous to the E. coli tkt gene are also described.

In Haemophilus influenzae Rd competence—the ability to bind, take up, and integrate exogenous DNA—is a complex and highly regulated state. Competence develops during the transition from exponential to stationary phase growth or in response to environmental stimuli, such as a nutritional shift down and/or a reduction in oxygen tension (1, 2). Competent cells undergo a number of changes, including the synthesis of several specific inner and outer membrane proteins (3, 4) and the development of efficient mechanisms for DNA processing and recombination. Although the factors involved in initiating competence development are largely unknown, the phenomenon is at least partially regulated by cAMP (5, 6). Simple addition of cAMP to exponentially growing cells stimulates competence to levels within 1% of maximum competence (100- to 10,000-fold increase) (5). In Escherichia coli and other Gram-negative bacteria cAMP functions to regulate transcription in conjunction with the cAMP receptor protein (CRP, also called catabolite gene activator protein, CAP). The cAMP-CRP complex binds to specific DNA sequences at or near various promoters, stimulating transcription initiation from some and repressing it from others (for review, see ref. 7). Until now, however, the molecule that mediates the cAMP effect in H. influenzae had not been identified.

Tomb et al. (8) recently reported the isolation and characterization of a competence (Com) mutant of H. influenzae named JG87. The mutation is the result of a single mini-Tn10kan insertion in the chromosome. The strain is completely deficient in competence-induced DNA binding, uptake, and recombination (8, 9); thus, transformation is undetectable in this strain. To identify the gene responsible for this competence defect, the mutant and corresponding wild-type loci were cloned, sequenced, and analyzed by complementation studies. The results showed that the mini-Tn10kan insertion identified a gene\(^*\), now named \textit{crp}, which produces a protein highly homologous to E. coli CRP and that E. coli CRP could function to replace the \textit{Haemophilus} protein in the development of competence.

MATERIALS AND METHODS

Bacterial Strains, Plasmids, and Culture Methods. The H. influenzae Rd strains used were KW20 wild type (10), DB117 (rec-l) (11), MAP7 (Str\(^R\), Na\(^{+}\), Kan\(^R\), Nov\(^R\), Str\(^R\), Spc\(^R\), Vio\(^R\)), and JG287. A limiting amount of chromosomal DNA from the competence mutant JG87 was used to transform KW20 to produce the backcross strain JG287 (crp::mini-Tn10kan) (8). E. coli DH5\(\alpha\) served as plasmid host.

The plasmid vectors used in this study were the E. coli–H. influenzae shuttle vectors pGJB103 (Tet\(^R\), Amp\(^R\)) (8) and pSU2718 (Cam\(^R\), lacZa) (12, 13). pSU2718 is a derivative of pACYC184 and carries the P15A origin of replication. P15A-derived plasmids replicate in \textit{H. influenzae} Rd, and their use as vectors for \textit{H. influenzae} is described elsewhere (13). pSH3 contains the \textit{E. coli} crp structural gene controlled by its native promoter (14) and was provided by M. Freundlich (State University of New York, Stony Brook). The construction of crp-containing plasmids is summarized below.

\textit{H. influenzae} cells were grown in brain–heart infusion broth (Difco) supplemented with 2 \(\mu\)g of NAD and 10 \(\mu\)g of hemin per ml. \textit{H. influenzae} cells were made competent for chromosomal transformation by the M-IV procedure (1). Plasmid DNA was generally introduced into cells made competent by the M-IV procedure. When recombination between the plasmid insert and the chromosome was to be avoided, the glycerol-stimulated method (15) was used. Competence mutants were transformed by plasmid DNA with the CaCl\(_2\) method (16). \textit{E. coli} strains were grown in LB broth (17), and competent cells were prepared as described by Morrison (18). The Lac phenotype of cells was determined on MacConkey-lactose plates. For \textit{Haemophilus} strains tetracycline was used at 5 \(\mu\)g/ml; kanamycin was used at 20 \(\mu\)g/ml for selection of chromosomal markers and at 35 \(\mu\)g/ml for plasmid markers. Chloramphenicol was used at 2 \(\mu\)g/ml for \textit{Haemophilus} strains and at 20 \(\mu\)g/ml for \textit{E. coli} strains.

DNA Isolation and Cloning. \textit{H. influenzae} DNA was prepared by standard methods (16). Small-scale plasmid purification was done as described by Birnboim and Doly (19) with the omission of lysozyme for \textit{H. influenzae} strains. Preparative-scale plasmid isolation was carried out by alkaline extraction and ethidium bromide-CsCl gradient purification (20), also without lysozyme for \textit{H. influenzae} strains. Purified restriction fragments used in cloning experiments were iso-

Abbreviations: CRP, cAMP receptor protein; ORF, open reading frame.

\(\*)\) The sequence reported in this paper has been deposited in the GenBank data base (accession no. M77207).
lated from agarose by electrophoresis onto DEAE membranes as described (21).

To clone the mutation, DNA from JG287 was digested with Pst I and ligated to Pst I-cleaved pGJB103 DNA by using standard reaction conditions (17). The ligation mixture was used to transform DB117 by the glycerol shock method (15). Kanamycin-resistant clones were selected, and restriction analysis of their plasmids was used to confirm the presence of the mini-Tnl0kan element within the cloned H. influenzae DNA fragment. One such plasmid was retained for further study and named pXN1.

Recovery of the wild-type allele was achieved by recombinational exchange of the mutated locus in pXN1 with the wild-type chromosome in H. influenzae Rd during transformation (22). Plasmids isolated from Tet’ Kan’ clones were analyzed to check for the loss of the mini-Tnl0kan element and replacement by the wild-type H. influenzae Rd Pst I fragment. The plasmid held for further study was named pXN2 (Fig. 1).

To subclone the H. influenzae Rd crp region, the 1189-base-pair (bp) Ssp I fragment from pXN2 (Fig. 1) was gel-purified and ligated to Sma I-digested pSU2718. The ligation mix was used to transform competent E. coli DH5α, and Cam’ Lac’ clones were selected. Restriction analysis was used to identify plasmids that carried the insert in opposite orientations. The resulting plasmids were named pXN14 and pXN16 with pXN14 carrying the crp open reading frame (ORF) in the same orientation as lacZa of the vector (Fig. 1).

The E. coli crp gene was subcloned from pSH3 (14). An ≈1300-bp Ssp I fragment containing the structural gene and its native 5’ and 3’ sequences was cloned in both orientations into the Sma I site of pSU2718 (as described above). The plasmid with E. coli crp in the same orientation as lacZa of pSU2718 was named pXN13; pXN15 has crp in the opposite orientation.

DNA Sequence Determination. Sequences were determined by the dideoxyribonucleotide chain-termination method (23) using the Sequenase system (United States Biochemical) or the T7 sequencing kit (Pharmacia), dGTP or dITP labeling mixes, [α-35S]dATP, and circular plasmid or gel-purified DNA fragment templates. The reaction products were analyzed on 6% (20:1) polyacrylamide buffer-gradient gels, according to Biggin et al. (24), as described by Sambrook et al. (17). Sequencing was started at the mini-Tnl0kan insertion in pXN1 using an oligonucleotide (′5′-CCACCTTAACCTTAAAGTATT-3′) specific to the ends of the mini-Tnl0kan. Because both ends of the mini-Tnl0kan are identical for ≈400 bp, an Nsi I fragment of pXN1 containing the right end of the mini-Tnl0kan and 478 bp of H. influenzae Rd DNA (Fig. 1) was gel purified and used as template for the initial sequencing. Subsequently, specific oligonucleotide primers were designed (synthesized by Scott Morrow, Johns Hopkins University) to continue sequencing in both directions using pXN2 as the template. At intervals of ≈245 bp another primer was synthesized and used to extend the sequence. With this strategy the DNA sequence of both strands was determined. The Genetics Computer Group (Madison, WI) software package was used for computer analysis of the DNA and protein sequences.

RESULTS

Cloning and Mapping of the Competence Mutation. A 6.5-kilobase (kb) Pst I fragment of the JG287 chromosome contained 4.6 kb of H. influenzae Rd chromosome and the 1.9-kb mini-Tnl0kan element (Fig. 1); this was cloned into pGJB103 to form pXN1. Insertion of the mini-Tnl0kan at this site has been shown responsible for the competence-deficient phenotype of this strain (8). The 4.6-kb wild-type Pst I fragment was recovered by transforming competent KW20 with pXN1. The Tet’ clones were screened for kanamycin sensitivity, and 75% (41 of 55) were found to be Kan’. Restriction and Southern analysis of the plasmid pXN2 from one of these Tet’ Kan’ clones showed that it contained the wild-type 4.6-kb Pst I fragment, having lost the mini-Tnl0kan insertion as would be expected from a recombination event (data not shown). A physical map of the 4.6-kb Pst I fragment was determined by comparison of single and multiple restriction enzyme digests of pXN1, pXN2, and pGJB103 (Fig. 1). The approximate location of the mini-Tnl0kan in pXN1 was determined from restriction analysis; the precise location was established by DNA sequence analysis (see below).

Sequence of H. influenzae Rd crp and Predicted CRP Protein Sequence. The nucleotide sequence of 1649 bp surrounding the mini-Tnl0kan insertion was determined (Fig. 2). The sequence confirmed restriction sites that had been precisely mapped physically and showed the presence of an additional Dra I site 8 bp from the other site. Examination of the sequence revealed a single large ORF, designated ORF1, extending from bp 492 to 1172. In pXN1, the mini-Tnl0kan inserted after bp 931, thus interrupting ORF1. The sequences up to 20 bp upstream of potential initiation codons (ATG, TTG, or GTG) in the region between bp 470 and 800 were examined for complementarity to the 3′ end of 16S rRNA of E. coli (3′-AUUCCUCCACUAG-3′) (25). The longest possible ribosome-binding site (TAAGGAG) at bp 499 was followed by an ATG codon at bp 501. Translation starting at this site would yield a protein of 224 amino acids and Mr 25,152. This putative protein was named CRP, based on the similarity in both sequence and function (see below) to E. coli CRP.

The 500 bp upstream of the putative CRP ORF was examined for sites similar to the bacterial promoter consensus (TgAca-17 bp-TAAtAT). The site most similar, bp 244-272, had a 17-bp hexamer spacing and 9 of 12 bases identical to the consensus (Fig. 2). An inverted repeat with potential for involvement in transcription termination of H. influenzae crp lies between bp 1196 and 1236 (Fig. 2). The predicted free energy value for this hairpin structure is ≈5 kcal/mol. The entire sequenced region was scanned for Haemophilus DNA uptake recognition sites (5′-AAGT-
GCG GTT TTA GCA CTT GTA GCA GAG CAG ATT ATT
GCAG TAT TAT CTT GTA TTA GTA

Fig. 3. Alignment of the deduced amino acid sequence of the H. influenzae CRP with those of other CRP family proteins. Proteins are as follows: Hin CRP, H. influenzae CRP; Eco CRP, E. coli CRP; Sf1 CRP, Sh. flexneri CRP; Sty CRP, Sa. typhimurium CRP; Xca CLP, X. campestris pv. campestris CRP-like protein. Amino acids identical to those in H. influenzae CRP are not shown; conservative replacements are shaded, and gaps are indicated by dashes (-). The helix-turn-helix DNA-binding motif is indicated above that sequence, and amino acids important for binding CAMP (determined for E. coli CRP) are underlined. Alignment and homologies were determined using the FASTA program of the Genetics Computer Group (GCG) software package with default settings.

FIG. 2. Nucleotide sequence and deduced amino acid sequence of the H. influenzae crp gene and protein. Putative ribosome-binding (SD) and promoter (−10, −35) sites are indicated preceding the start of the CRP ORF (bp 501). Position of the mini-Tn10 insertion (v at bp 931) and relevant restriction sites are marked. A partial ORF, ORF2, upstream of CRP is also shown (bp 1-240). Inverted repeats outside the ORFs are indicated by converging arrows.

GCGGT-3'). No matches were found, which is not necessarily unexpected for a sequence of this size (1.7 kb) (26, 27).

Similarity of H. influenzae CRP to Prokaryotic CRP Family Proteins. The predicted H. influenzae CRP sequence was compared to those in the protein sequence data bases. Strong similarities were found to the three previously described CRP proteins from E. coli, Shigella flexneri, and Salmonella typhimurium (Fig. 3) (28-31). A fourth protein, CLP (cAMP receptor protein-like protein) of Xanthomonas campestris
acids) had 41% identical residues and an additional 43% conservative replacements when compared with \( H. \) \( influenzae \) CRP. CRP of \( E. coli \) is a well-characterized protein. The \( E. coli \) CRP helix-turn-helix DNA-binding domain has been identified (33). This domain was conserved in \( H. influenzae \) CRP (Fig. 3). The specific sites of interaction of \( E. coli \) CRP with DNA are at Arg-181, Glu-182, Arg-186, and Lys-188 (34). The first three of these residues were identical in \( H. influenzae \) CRP (Arg-195, Glu-196, Arg-200); the fourth was changed to isoleucine (Ile-202) (Fig. 3). The amino acids of \( E. coli \) CRP postulated to interact with cAMP were also conserved in \( H. influenzae \) CRP, as shown in Fig. 3.

One of the functions of \( E. coli \) CRP is to act as a regulator of its own synthesis (14, 35, 36). Thus, the DNA sequence upstream of the \( H. influenzae \) crp gene was examined for sequences similar to the consensus \( E. coli \) CRP-binding site (AA-TGTGA -T- -TCA-AT) (37). The site most similar was at bp 318–339 (AA-CGTGA -T- -GCG-AG) with 9 of 13 bases identical to the consensus. This potential CRP-binding site was located between the putative promoter and the ribosome-binding site, the same region where in \( E. coli \) CRP binds to repress transcription of crp. Several other less conserved potential CRP-binding sites were found in this region, including sites on both strands of the DNA.

**Complementation of the Competence-Deficient (Com') Phenotype.** To identify the role, if any, of CRP in competence, subclones of \( E. coli \) and \( H. influenzae \) crp were constructed and used to test for complementation of the competence defect in strain JG287. \( H. influenzae \) crp was subcloned as an 1189-bp \( SspI \) fragment (Fig. 1). This fragment contains the crp structural gene and its putative transcriptional signals (Fig. 2). \( E. coli \) crp was also subcloned as an \( SspI \) fragment that contains the \( E. coli \) structural gene with its native transcriptional signals and \( ~200 \) bp of \( pB322 \) DNA at the 3' end of the gene. The transformability of JG287 containing these various plasmid constructs was tested. The results are shown in Table 1. Both the \( E. coli \) and \( H. influenzae \) DNA fragments complemented the competence defect in JG287, restoring approximately wild-type levels of transformation. The level of complementation with each of the four plasmids is reproducible and may reflect the amount of expression of the crp gene in each strain. These results, the homology of the \( H. influenzae \) fragment to \( E. coli \) crp and the ability of \( E. coli \) crp to restore wild-type levels of transformation in JG287, confirmed that the insertion in JG287 disrupts the \( H. influenzae \) Rd homologue of \( E. coli \) crp and that CRP is required for competence development in \( H. influenzae \) Rd.

**Neighboring ORFs.** Examination of the region sequenced revealed a second partial ORF (ORF2), from bp 1–240, with similarity to another \( E. coli \) protein. ORF2 (75 amino acids) was similar, 41% identical with another 49% conservative replacements, to the protein encoded by the ttk gene that is downstream of the \( dut \) (dUTPase) gene of \( E. coli \) (38). The ttk gene of \( E. coli \) encodes a protein of 210 amino acids with a \( M_r \) of 23,500 that is detectable from plasmids containing the \( dut \) locus (39). The ttk gene product is apparently cotranscribed with \( dut \), and its function is unknown (38, 40). Several other small ORFs (41–83 amino acids) were detected in the region sequenced, none with similarity to known proteins.

**DISCUSSION**

The mini-Tn10kan insertion in strain JG287 has identified a gene, crp, required for the development of competence in \( H. influenzae \). The similarity of \( H. influenzae \) CRP to the \( E. coli \) cAMP receptor protein suggests that \( H. influenzae \) CRP functions by forming a complex with cAMP to regulate the expression of one or more genes involved in the induction of competence. Comparison of the \( H. influenzae \) and \( E. coli \) CRP proteins shows that the regions of \( E. coli \) CRP required for function—that is, the helix-turn-helix DNA-binding motif and the specific amino acids required for binding cAMP are conserved in the \( H. influenzae \) protein. In addition, \( E. coli \) CRP can apparently substitute for \( H. influenzae \) CRP, as shown by the complementation experiments. There are differences between the \( E. coli \) and \( H. influenzae \) CRP proteins. \( H. influenzae \) CRP has an additional 14 amino acids at the N terminus of the protein. The significance of this is not clear; however, evidence from the study of the structure of \( E. coli \) CRP suggests that the N-terminal end is not involved in dimer formation, cAMP binding, or DNA binding (33). Finally, the results presented here further demonstrate the utility of several genetic tools recently described for \( H. influenzae \). (i) Homologous exchange in transformation provides a reliable mechanism for introducing specific insertions or transposition mutations into the \( H. influenzae \) chromosome, as was described for isolation of this competence mutant (8). (ii) P15A-derived plasmids are useful vectors for genetic analysis in \( H. influenzae \) (13).

The specific genes in \( H. influenzae \) regulated by the cAMP–CRP complex are not known. However, transcription of the \( E. coli \) crp gene is known to be negatively regulated by the cAMP–CRP complex (36). This suggests the possibility that cAMP–CRP may also negatively regulate the \( H. influenzae \) crp gene. The presence of consensus CRP-binding sites downstream of the putative transcription start site in \( H. influenzae \) crp supports this hypothesis. Negative autoregulation may partially explain the observation that addition of cAMP to exponential-phase \( H. influenzae \) cultures induces competence but gives transformation frequencies \( ~100 \)-fold lower than that seen with M-IV medium (5, 6). Another possible explanation for the lower transformation frequency seen with the addition of cAMP, suggested by Zoon et al. (6), is that multiple regulatory events are required for full induction of competence. Additional mutations that effect competence in genes other than crp have been identified (8), and these would be candidates for other regulatory genes. Also, Redfield has recently isolated a competence-enhancing mutation \( sxy-I \) (9). During exponential growth cells carrying this mutation transform 100 times more efficiently than wild-type cells (9). The gene identified by this mutation could also function as another regulator of competence. Regarding this, JG287 pXN14 was tested for its ability to transform during exponential growth, and these cells were found to transform at the same low level as wild-type cells (data not shown).

The CRP protein appears highly conserved among species (30) and yet acts to control many different functions (7). \( H. influenzae \) has apparently adapted it to control competence development; however, the role of CRP in other transformable species is uncertain. The cloning, sequencing, and

<table>
<thead>
<tr>
<th>Table 1. Complementation of the competence defect (Com') in JG287 by plasmids containing crp</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Strain</strong></td>
</tr>
<tr>
<td>--------------------------</td>
</tr>
<tr>
<td>KW20 wild type</td>
</tr>
<tr>
<td>JG287 Com-</td>
</tr>
<tr>
<td>JG287 pSU2718</td>
</tr>
<tr>
<td>JG287 pXN13 (Ecpp)&gt;</td>
</tr>
<tr>
<td>JG287 pXN15 (Ecpl&gt;</td>
</tr>
<tr>
<td>JG287 pXN14 (Hcpp)&gt;</td>
</tr>
<tr>
<td>JG287 pXN16 (Hcpp)&lt;</td>
</tr>
</tbody>
</table>

*cfu. Colony-forming units.
*DNA fragment and orientation present on plasmid are in parentheses; arrow pointing right represents the same orientation as that of lacZa of the vector, and arrow pointing left is the opposite.
*IV competent cells were incubated with 1 μg of strain MAP7 chromosomal DNA per ml for 30 min, diluted, plated in agar, and overlaid with agar containing streptomycin.
identification of crp as a gene regulating competence induction in H. influenzae provides a firm basis for further physiological and biochemical studies of the development of competence in H. influenzae.

I am grateful to Hamilton Smith in whose laboratory the work was done. I thank Martin Freundlich for his gift of strains and Jean-Francois Tomb for critical review of the manuscript. This work was supported by Grants 5-T32-CA091349 and 1-RO1-AI27783 from the National Institutes of Health and by Grant MV-517 from the American Cancer Society.