



A Novel CRP-dependent Regulon Controls Expression of Competence Genes in *Haemophilus influenzae*

Rosemary J. Redfield^{1*}, Andrew D. S. Cameron², Qing Qian¹, J. Hinds³
Tahir R. Ali⁴, J. Simon Kroll⁴ and Paul R. Langford⁴

¹Department of Zoology
University of British Columbia
Vancouver, Canada V6T 1Z4

²Department of Microbiology
and Immunology, University of
British Columbia, Vancouver
Canada V6T 1Z3

³Bacterial Microarray Group
St George's Hospital Medical
School SW17 0RE, London
UK

⁴Department of Paediatrics
Imperial College London
St. Mary's Hospital Campus
London W2 1NY, UK

Natural competence for DNA uptake is common among bacteria but its evolutionary function is controversial. Resolving the dispute requires a detailed understanding of both how cells decide to take up DNA and how the DNA is processed during and after uptake. We have used whole-genome microarrays to follow changes in gene expression during competence development in wild-type *Haemophilus influenzae* cells, and to characterize dependence of competence-induced transcription on known regulatory factors. This analysis confirmed the existence of a postulated competence regulon, characterized by a promoter-associated 22 bp competence regulatory element (CRE) closely related to the cAMP receptor protein (CRP) binding consensus. This CRE regulon contains 25 genes in 13 transcription units, only about half of which have been previously associated with competence. The new CRE genes encode a periplasmic ATP-dependent DNA ligase, homologs of SSB, RadC and the *Bacillus subtilis* DNA uptake protein ComEA, and eight genes of unknown function. Competence-induced transcription of genes in the CRE regulon is strongly dependent on cAMP, consistent with the known role of catabolite regulation in competence. Electrophoretic mobility-shift assays confirmed that CRE sequences are a new class of CRP-binding site. The essential competence gene *sxy* is induced early in competence development and is required for competence-induced transcription of CRE-regulon genes but not other CRP-regulated genes, suggesting that Sxy may act as an accessory factor directing CRP to CRE sites. Natural selection has united these 25 genes under a common regulatory mechanism. Elucidating this mechanism, and the functions of the genes, will provide a valuable window into the evolutionary function of natural competence.

© 2005 Elsevier Ltd. All rights reserved.

Keywords: natural transformation; DNA uptake; cyclic AMP; genetic exchange; recombination

*Corresponding author

Introduction

Natural competence, the ability to actively take up DNA from the environment, is widely distributed among both Gram-positive and Gram-negative bacteria,^{1–4} although its mechanism remains unresolved and its significance controversial. Induction of the competent state enables a bacterium to bind free DNA fragments at the cell surface and transport them across the cell

membrane or membranes and the cell wall into the cytoplasm. DNA that escapes degradation may cause a genetic change (transformation) by recombining with the cell's chromosome.

DNA uptake mechanisms must be able to transport stiff hydrophilic DNA molecules across the cell's hydrophobic envelope. The mechanical force is thought to be provided by retraction of type IV pili or pseudo-pili, as pilin and pilus assembly proteins have been implicated in a number of different bacteria.⁵ The first requirement for understanding any mechanism is identifying all the components, and we have used microarray analysis of competence to identify all of the genes induced when *Haemophilus influenzae* cells are becoming competent.

Abbreviations used: CRE, competence regulatory element; CRP, cAMP receptor protein; TF, transformation frequency; ORF, open reading frame.

E-mail address of the corresponding author:
redfield@interchange.ubc.ca

In *H. influenzae*, genes with important roles in competence have been identified primarily by screens for mutants defective in transformation, and have been tentatively assigned regulatory or mechanistic roles based on mutant phenotypes and homology to competence genes in other bacteria. Most of the known *H. influenzae* competence genes are thought to encode proteins with direct roles in DNA uptake or in assembly of the uptake machinery. These include proteins acting at the outer membrane (*comE*, *pilA*), in the periplasm or at the inner membrane (*comC*, *comF* (= *com101A*), *rec-2*), and cytoplasmically (*comA*, *dprA*, *comM*).^{6–11} There have been several non-exhaustive screens for transformation-defective mutants,^{12–16} each has identified some new candidate genes but missed known genes, suggesting that some regulatory and DNA uptake genes may not yet be identified.

Transformation is competence's best-known consequence but it is not the usual outcome of DNA uptake; even DNA perfectly homologous to chromosomal DNA is usually degraded rather than recombined with the chromosome.¹⁷ Re-use of nucleotides from DNA degraded in the cytoplasm may be more significant than any genetic benefits, at least in the short term.¹⁸ Because regulatory mechanisms are products of natural selection, the tight regulation of competence in almost all bacteria can provide important clues to the function of DNA uptake. However, the processes regulating competence development appear diverse, and array-based investigations in Gram-positive bacteria have revealed an unexpected complexity of regulated and regulatory genes.^{19–21} *H. influenzae* is the Gram-negative bacterium whose competence regulation is best understood, and microarray analysis of regulatory mutants allowed us to investigate this regulation. In particular we wished to find out whether competence in *H. influenzae* results from a single unified response to environmental triggers or a coincidence of differently regulated processes.

Competence in *H. influenzae* is efficiently induced by transfer of exponentially growing cells to a defined starvation medium called MIV ("M-4"),²² and analysis of mRNA levels or expression of *lacZ*-fusions has shown that several of the DNA uptake genes are induced by this transfer (*comA*, *comF*, *rec-2*, *dprA*).^{7,10,14} A lower level of competence is also seen as cultures approach stationary phase in rich medium.²³

Two regulatory proteins, CRP and Sxy, are thought to activate transcription of *H. influenzae* competence genes. CRP is the cAMP regulatory protein, best characterized in *Escherichia coli*; it activates transcription of the carbon-energy regulon when rising cAMP levels signal that preferred sugars are unavailable. Like *E. coli*, *H. influenzae* has a phosphotransferase system that regulates cAMP levels and thus gene activation by CRP.^{24–27} In *H. influenzae* cAMP and CRP also regulate competence; *crp* and *cya* (adenylate cyclase) mutants are unable to become competent^{28,29} or to induce expression of the competence gene *comA*.¹⁴

Regulation of competence by such energy-supply signals is consistent with its proposed role in nutrient acquisition. In *H. influenzae* as in *E. coli*, CRP regulates diverse genes involved in nutrient acquisition or use,³⁰ and most genes regulated by CRP are predicted to be subject to additional function-specific regulation. Addition of cAMP to exponentially growing wild-type cells does not induce full competence, but only the 100-fold lower competence also seen at the onset of stationary phase,³¹ suggesting that competence genes may be subject to regulation by a second competence-specific regulator.

The Sxy protein may be the predicted second regulator. Strains over-expressing Sxy have elevated competence, and a *sxy* knockout mutant, like a *crp* mutant, is unable to become competent and fails to induce expression of *lacZ* fusions to the *comF* and *dprA* genes.^{10,32,33} However, we understand neither how Sxy regulates competence genes nor how Sxy itself is regulated. Although Sxy lacks the structural features typical of DNA binding proteins, it has been postulated to activate transcription by binding to DNA at the competence regulatory element sequences associated with promoters of DNA uptake genes (see below).¹⁰ The existence of Sxy homologs in non-competent as well as competent bacteria suggests that it may have a role outside of competence. A role for mRNA secondary structure in regulating *sxy* expression is likely: point mutations weakening base-pairing in a 5' stem of *sxy* mRNA dramatically increase *sxy* mRNA expression and competence; mutations strengthening the stem eliminate both.^{32,34}

The hypothesis that competence regulatory element (CRE) sequences in the promoters of DNA uptake genes are responsible for competence-specific regulation of transcription was strengthened by Gwinn *et al.*'s demonstration that *comM*, a gene initially identified only by its possession of a CRE sequence, is competence-regulated and essential for transformation.^{10,11} Only two symmetrical base-pairs distinguish the core CRE sequence TGCGA(N₆)TCGCA from the core CRP binding sequence TGTGA(N₆)TCACA, suggesting that CRP may bind to CRE sites and activate transcription directly, rather than regulating competence indirectly by regulating *sxy* transcription.³³ Under this model, the presence of cAMP and Sxy allows CRP to bind at CRE sites and activate transcription of DNA uptake genes.³⁵ Consistent with this, both CRP and Sxy are required for transcription of genes in the putative *comA-F* operon.^{27,33}

We have generated microarrays containing all 1738 genes of the sequenced *H. influenzae* strain KW20, and have used them to characterize (1) the changes in expression of the 1738 *H. influenzae* genes in response to transfer from rich medium to MIV, and (2) the extent to which these changes depended on regulation by Sxy and by CRP. This enabled us to show that the subset of starvation-induced genes that possess CRE sequences are also

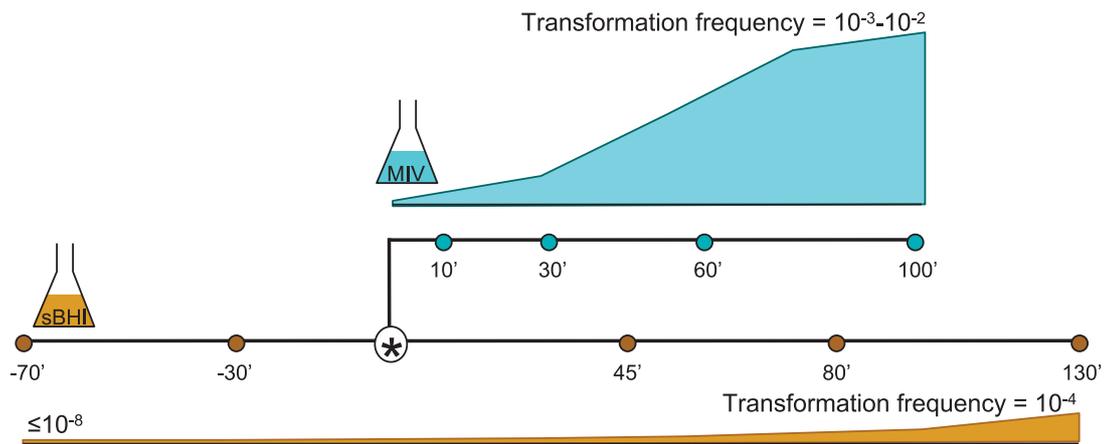


Figure 1. The sampling protocol for the competence time-courses. See Materials and Methods for details.

united by their requirement for both CRP and Sxy. This CRE regulon includes most of the identified components of the DNA uptake machinery, in addition to a number of new genes not previously associated with competence.

Results

Figure 1 illustrates the split time-course analysis used to characterize gene expression changes during competence development. Cells were sampled both during growth in the rich medium sBHI and after transfer to the starvation medium MIV. As Figure 1 indicates, cells reach peak competence after 100 minutes in MIV (transformation frequency (TF) with MAP7 DNA about 3×10^{-3}), and also become moderately competent (TF about 10^{-4}) when growth in sBHI first slows; this “late log” competence occurs before the complete cessation of growth.

The complete time-course was done twice, using RNA preparations from independent cultures. For seven of the nine time points from the two replicate time-course experiments, the replicate measurements of expression levels of 89–92% of the genes were within twofold. The exceptions were the $t=10$ minute and $t=30$ minute time points for cells in MIV, which had 81–82% of their values within twofold. Most differences greater than twofold were due to minor differences in the timing of competence development between the replicates, rather than to random variation. For 24 of the 27 microarrays used, the transcripts of less than 2% of the 1738 genes produced “No Data” reports. The other three were less than 4%. This indicated that the majority of the genes on the microarray slide were transcribed regardless of the various culture conditions and cell types used. In addition, it also demonstrated that the current microarray methodology and analysis system were sensitive enough to detect even transcripts of low abundance.

Typical microarray data from one of these time-courses is shown in Figure 2. Figure 2(a) shows

relative expression of all the 1738 genes in the *H. influenzae* genome during exponential growth ($t=0$) and at 10, 30, 60 and 100 minutes after transfer to MIV. A total of 151 genes showed reproducible >fourfold increases in mRNA after transfer to MIV. Although many genes showed modest decreases in expression after transfer to MIV, only 44 decreased by at least fourfold. (Lists of these genes are provided as Supplementary Data.) Below, we focus mainly on the genes likely to play roles in competence.

Identification of competence-induced genes

Known competence genes

We first examined starvation-induced changes in expression of the known competence genes *comA*, *comC*, *comE*, *comF* (all in the putative *comABCDEF* operon), *rec-2*, *dprA*, *comM*, and *pilA*; all five promoters contain previously identified CRE sequences.¹⁰ All genes except *comF* were induced strongly (45–450-fold) but more slowly than the majority of induced genes, with maximum expression usually seen at the $t=60$ minute sample (Figures 2(b) and 3). Expression levels of *comF* were low but quantitative PCR showed that it is induced about 40-fold in MIV, confirming previous reports.^{7,33} As expected, *comB* and *comD* were co-induced with the rest of the *comABCDEF* operon, and *pilB*, *pilC* and *pilD* were coinduced with *pilA*. However, *dprB* and *dprC* showed little induction and did not appear to be coordinately expressed with *dprA*.

Other CRE-regulated genes

Four uncharacterized genes had also been identified as having promoters with putative CRE sequences.^{10,35} Two of these were induced with the same kinetics as the genes described above: *comE1* (HI1008, 270-fold) and *ssb* (HI0250, 3.4-fold). The low but consistent induction of *ssb* was confirmed by quantitative PCR. Genes of unknown function downstream from the two other CRE

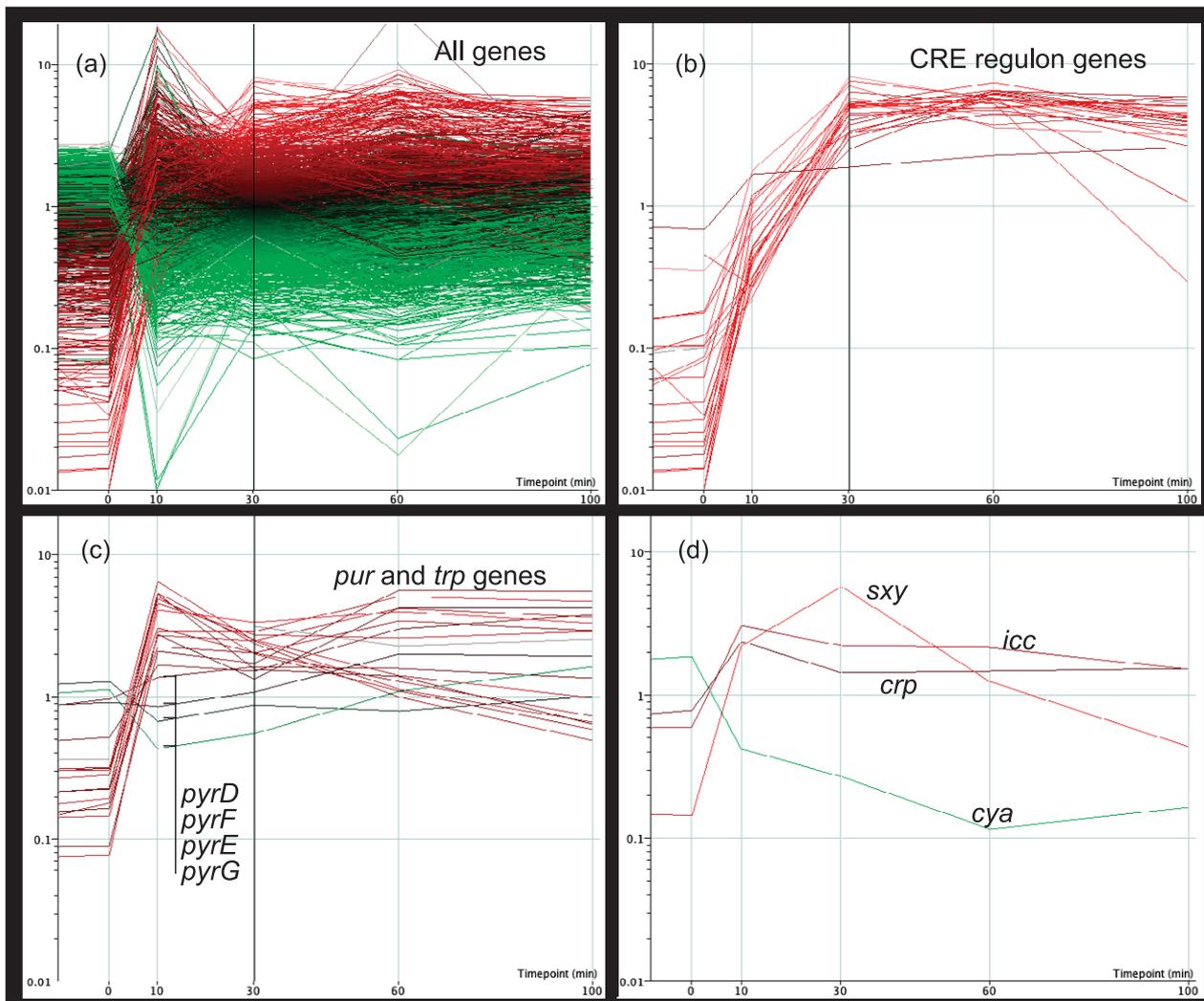


Figure 2. Data from microarray analysis of a competence time-course. The $t=0$ minute data is the mean of $t=-30$ and $t=-70$ minute samples. (a) All genes on the array. (b) All genes in the CRE regulon (listed in Figure 3). (c) *purB*, *purC*, *purD*, *purE*, *purF*, *purH*, *purK*, *purL*, *purM*, *purN*, *trpA*, *trpB*, *trpC*, *trpD*, *trpE*, *trpG*, *pyrD*, *pyrE*, *pyrF*, *pyrG*. (d) *sxy*, *crp*, *cya*, *icc*.

sequences were also induced (HI0365, ninefold, and HI1182, 50-fold); however, these CRE sequences had originally been incorrectly assigned to the divergently transcribed genes HI0364 and HI1181, respectively. Like the known competence genes, induction of HI1008, HI0250, HI0365 and HI1182 was relatively slow, with expression peaking at 30–60 minutes after transfer (Figure 2(b)).

Several complementary strategies were used to search for additional genes in the CRE regulon, and to exclude others from it. First, the MIV time-course data were examined for other genes induced with the same kinetics as the nine identified above, both by eye and by using the “find similar” function of GeneSpring. Such kinetics immediately identified a four-gene operon induced very strongly in MIV (HI0938-41, 600-fold). Examination of sequences upstream of its promoter revealed a CRE sequence that had been missed by previous searches because one base was specified only as K (G or T);

re-sequencing showed this to be an A, giving a perfect match to the CRE core consensus.

In parallel, the nine confirmed CRE sequences (excluding HI0938) were used to refine the CRE consensus, and the program RSA-tools³⁶ was then used to search the non-coding genome sequences for additional elements fitting it. A search using a stringent cutoff score of 20 returned eight of the nine input CREs, three additional CREs, and six sequences strongly matching the CRP consensus. (The *comA* CRE was missed by this search because it overlaps an upstream coding region.) Examination of array data showed that the three genes with previously unrecognized CREs (HI0659/0660, HI0952 and HI1631) were induced by MIV-starvation in the same manner as the confirmed CRE transcription units. Reducing the RSA-tools stringency cutoff to 10 and extending the search into upstream coding sequences produced many more CRP sites but only two more genes with candidate

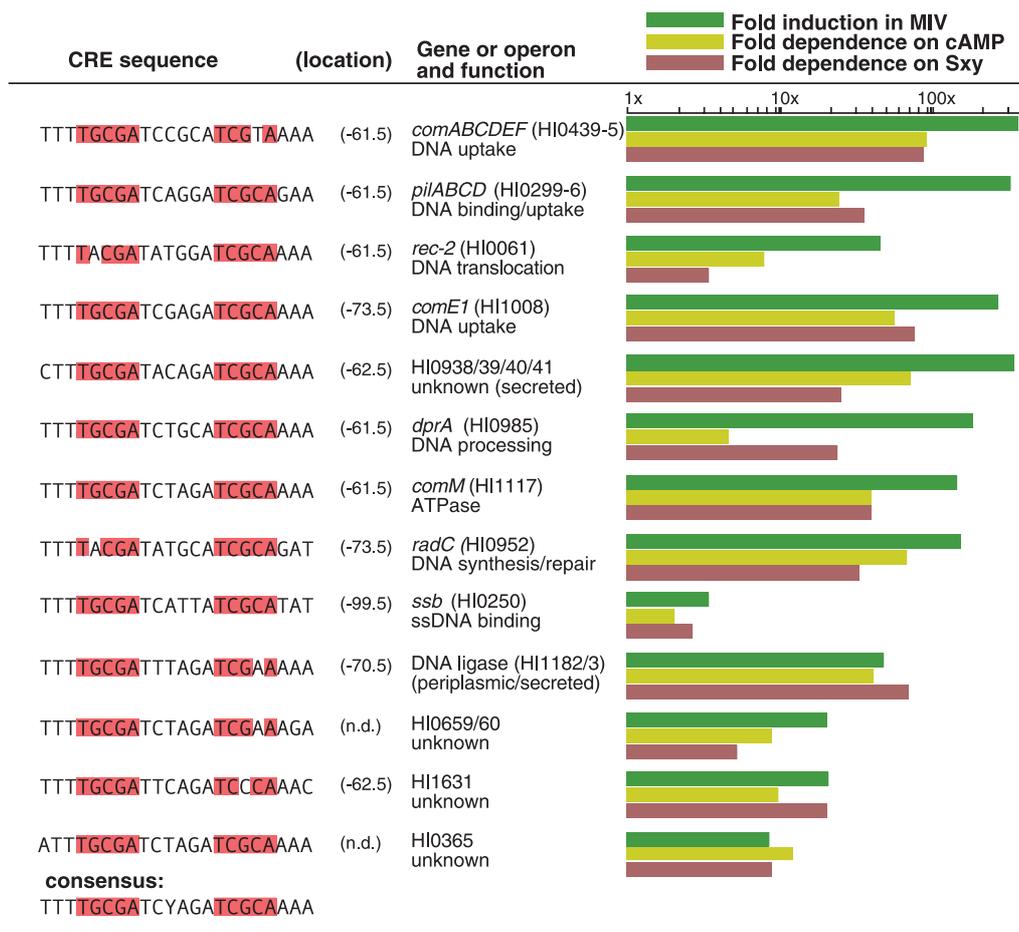


Figure 3. The CRE regulon. The CRE location is distance from the putative 5' end of the transcript to the center of the CRE; n.d., promoter not determined. Fold induction is the ratio of maximum expression in MIV expression level to $t=0$ expression. Fold dependence is the ratio of expression with cAMP or Sxy to expression without, after 100 minutes incubation in MIV.

CREs (*mobB* and *gyrB*). Both these genes lack competence-specific regulation (induction in MIV or dependence on cAMP or Sxy; see below), and neither CRE is expected to have a strong influence on its gene's transcription, the *mobB* CRE because it is more than 300 bp upstream of the transcription start site, and the *gyrB* CRE because it is a poor match to the consensus. Expression levels of genes transcribed convergently with the CRE-regulon genes were also checked to ensure that strong signals were not created by antisense transcription extending from convergently transcribed genes.

Properties of the complete CRE regulon are summarized in Figure 3. The green bars on the right show how strongly each gene is induced, with the length of each bar indicating the ratio of maximum expression in MIV to $t=0$ expression in sBHI (for multi-gene operons the ratio shown is for the first gene).

On the left of Figure 3 are the 13 CRE sequences and the distance of each from its likely promoter. The consensus of the 13 CRE sequences is shorter than but otherwise identical to that originally proposed.¹⁰ The sequence logos shown in Figure 4 reveal that CRE sites differ most strongly from the

consensus of *H. influenzae* CRP sites in having G and C rather than T and A at the highly conserved symmetric positions 6 and 17; these are the positions where CRP bends DNA.³⁷ CRE sequences are also less variable than CRP sequences, especially in the strings of T and A bases that flank the core. Consistent with the hypothesis that CRP binds to CRE sites, promoter-CRE spacing obeys the same constraints as promoter-CRP spacing in *E. coli*.³⁸ A detailed analysis of bacterial CRE sites is in preparation. The sequences upstream of each of the CRE-regulon genes were examined for additional motifs, using the programs Consensus,³⁹ Gibbs recursive sampler⁴⁰ and Bio-pro prospector.⁴¹ No new patterns were found, suggesting that CRE sites are likely to be the only sites where competence-regulatory factors bind DNA.

CRP and cAMP regulate expression of the CRE regulon

CRP is the best candidate for the factor that binds CRE sites, as it and cAMP are known to regulate transcription of *comA* and *dprA* and to be essential

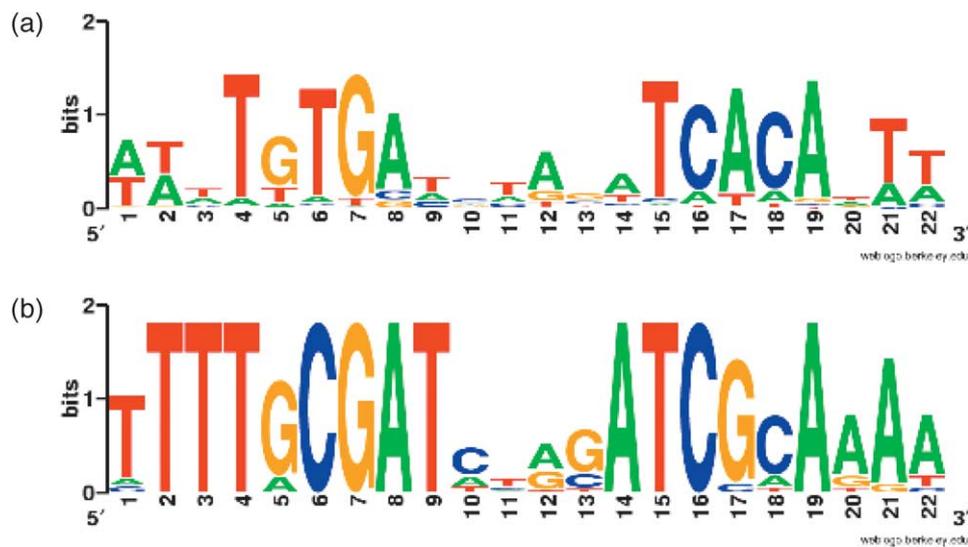


Figure 4. Sequence logo comparison of *H. influenzae* CRP and CRE consensus sequences. (a) CRP logo generated from the 45 CRP sites regulating the MIV-induced genes that are regulated by CRP but not Sxy. (b) CRE logo from the 13 CRE sites regulating the genes of the CRE regulon.

for competence. Do CRP and cAMP also regulate transcription of all the other CRE-regulon genes? This was examined using arrays where RNA from MIV-induced *cya* cells was competed with RNA from a parallel culture incubated in cAMP-supplemented MIV. The *cya* mutant does not develop competence in unsupplemented MIV but becomes fully competent if cAMP is provided.²⁹ The yellow bars in Figure 3 show the degree to which expression of the first gene in each transcription unit depended on cAMP, calculated as the ratio of expression with cAMP to expression without cAMP. The cAMP dependence of all the CRE genes' expression was roughly proportional to their levels of induction in MIV (green bars and data not shown). However, changes in *ssb* expression were not reproducibly more than twofold.

To test whether a CRE is in fact a CRP-binding site, we purified native CRP from *E. coli* and used electrophoretic mobility-shift assays to measure DNA-binding specificity. Band shifts were apparent with both the *mglBAC* (CRP-binding site) and *comA-F* (CRE) promoter regions in reactions

containing 5 to 500 nM CRP and were detectable with 0.5 nM CRP (Figure 5). A DNA fragment without a CRP-binding site showed no band shift even with 500 nM CRP. The relative affinity of CRP for different bait DNAs was estimated by titrating CRP into binding reactions. Comparison of lanes in which about half of the bait DNA was shifted (Figure 5; 50 nM CRP for *comA-F*, 5 nM CRP for *mglBAC*), revealed that CRP binds the *mglBAC* promoter with around tenfold greater affinity than the *comA-F* promoter. This finding is consistent with the 80-fold greater affinity of CRP for a synthetic (perfect) CRP-binding site over the same sequence with CRE-like G:C substitutions at base-pairs 6 and 17.⁴²

We also confirmed that the CRP-related transcriptional activator FNR does not play a role in regulating expression of competence genes. FNR is the only other member of the CRP family in *H. influenzae*; in *E. coli* the two proteins have very similar binding sites.⁴³ FNR's binding specificity has not been investigated in *H. influenzae*, and it could have evolved to bind to the CRP-related CRE sites. However, a *fur* knockout, constructed by

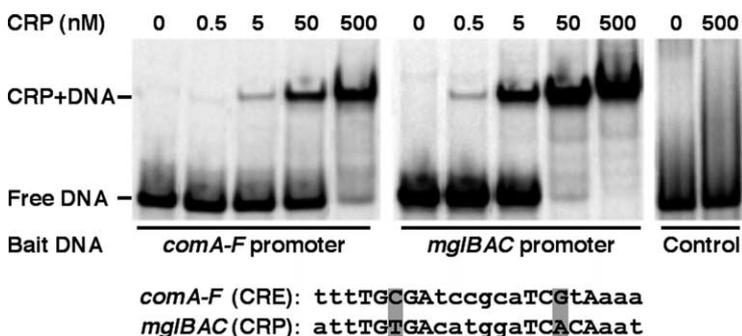


Figure 5. Electrophoretic mobility shift analysis of CRP-DNA complexes. Bait DNAs are 130 bp fragments amplified from *H. influenzae* chromosomal DNA; control DNA lacks an apparent CRP-binding site. CRP-binding site alignment: capital letters indicate agreement with the highly conserved regions of CRP-binding sites, grey boxes highlight the distinguishing bases of CRE sites.

insertion of a KanR cassette, had no effect on competence in sBHI or MIV (data not shown).

Sxy regulates expression of the CRE regulon

A *sxy* knockout mutant is unable to become competent and fails to induce expression of *comF* and *dprA lacZ*-fusions.^{10,32–34} To find out whether Sxy controls all CRE regulon genes, and whether it controls other genes, we used microarrays to compare MIV-induced gene expression in cells carrying a *sxy* knockout³² with that in wild-type cells. Except for *ssb* (reduced only two- to threefold) genes in the CRE regulon were expressed at 20–200-fold lower levels in the mutant. The purple bars in Figure 3 show the magnitude of the Sxy dependence for the first gene in each transcription unit. Although the level of *comF* transcripts was too low to measure reliably in microarrays, Zulty *et al.* have shown that Sxy is needed for *comF* expression³³ and we have now confirmed this with quantitative PCR. With one exception, expression of genes lacking CRE sites was not changed by deletion of *sxy*, indicating that Sxy's only role may be to regulate the CRE regulon genes.

The one exception was the operon containing genes HI0658-0654. These are moderately competence-induced (four- to tenfold) with the same kinetics as the CRE regulon genes, and this induction depends on cAMP and Sxy. However, no CRE sequence could be identified in the 160 bp non-coding region upstream of HI0658, and none of the genes' functions are obviously related to DNA uptake. This operon is directly downstream of the CRE-regulated HI0660-0659 operon (Figure 3), so transcription may read through from it into HI0658-0654 (the intervening 160 bp lack any obvious transcriptional terminator). Comparisons of HI0658 and HI0659 to homologs in other species confirmed the stop and start codon assignments.

Other starvation-induced genes

To ensure that no other competence genes had been overlooked, all genes at least fourfold induced in MIV were examined for function and for

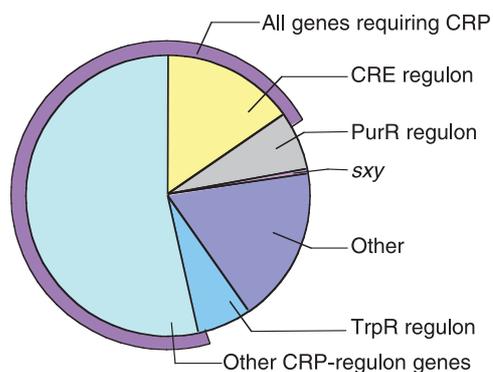


Figure 6. Starvation-induced genes. Categorization of the 151 genes induced at least fourfold on transfer to MIV.

dependence on cAMP and Sxy (Figure 6). Most of these 151 genes were found to be CRP-dependent, consistent with the evidence that cAMP levels rise during competence induction and with the large number of CRP-regulated genes postulated by Tan *et al.*³⁰ The induced genes included 23 of the 25 genes in the CRE regulon (*ssb* and *comF* did not meet the fourfold induction criterion), two genes in the PurR regulon, one in the TrpR regulon, and 81 other genes with CRP sites and cAMP-dependent, *sxy*-independent expression in MIV (CRP-regulon genes). The other CRP-regulon genes fell into several groups: 27 genes of unknown functions, 23 genes involved in sugar utilization, and 31 genes mainly with roles in nutrient uptake and central metabolism. None of these non-CRE genes has been implicated in DNA uptake.

Forty-four of the MIV-induced genes depended on neither CRP nor Sxy. In addition to *sxy* itself (discussed below), these included the PurR-regulon and TrpR-regulon genes shown in Figure 2(c). Although genes for synthesis of other amino acids were not induced, transfer to MIV caused rapid induction of genes for tryptophan biosynthesis, presumably due to the lack of tryptophan in the Casamino acid component of MIV. Supplementation of MIV with tryptophan did not affect competence development (data not shown). Genes in the purine biosynthetic pathway were also rapidly induced, confirming that transfer to MIV causes rapid depletion of purine pools. Genes for pyrimidine synthesis (*pyrD*, *E*, *F* and *G*) were expressed quite strongly during exponential growth in sBHI and were not further induced by transfer to MIV (Figure 2(c)). The final categories of MIV-induced genes comprise 15 genes whose functions have no obvious connection to competence, and 13 genes whose functions are unknown. The lack of any competence-related genes in this set suggests that the CRE regulon includes all of the genes that need to be induced for competence development.

Genes not induced in MIV

The many genes that were not induced on transfer to MIV medium include genes reported to play roles in DNA uptake or recombination. Most of these were expressed at moderate to high levels in sBHI, indicating likely roles in normal growth. These include genes likely to have general effects on transcription and translation (*topA*, *atpA*, *fis*),^{14,34,44} a gene that affects protein export (*por*),¹⁴ genes that affect peptidoglycan metabolism and cell division (*murE*, *pbp1*, *pbp2*, *rodA*, *mreB*, *pgsA*, *ftsN*),^{6,45,13} and a gene required for DNA utilization by *E. coli* (*comJ*).^{13,46} We also examined the sites of mini-Tn10kan insertions that had been isolated by Tomb *et al.*¹² but not previously mapped to specific genes. Comparison of the original restriction maps with the genome region identified by Southern blotting showed that the insertions in the JG6 cluster of mutants are all in the *trmE* gene (a tRNA

methyltransferase), the insertion in JG27 is in *gor* (glutathione reductase), and the insertion in JG49 is in the 250 bp segment between *rpoB* and *rpoC*. None of these were induced in MIV.

Genes down-regulated in MIV

Many of the 44 genes down-regulated at least fourfold play roles in translation. Transcription of the 29 genes in the two ribosomal protein operons (HI0776-HI-786 and HI0788-HI0803) was reduced transiently, though not all genes met the fourfold reduction cutoff. The conserved operon containing *nusA* and *infB* was also down-regulated, as was *rpoBC*. The other down-regulated genes did not fall into any evident groups; *cya* was the only down-regulated gene with a known connection to competence.

Expression of regulatory genes

Regulation of CRP and cAMP

Unlike its *E. coli* homolog, the *H. influenzae* *crp* promoter has no CRP sites; consistent with this, microarrays showed that *crp* mRNA was only weakly induced by transfer of cells to MIV (Figure 2(d)) and unaffected by mutation of *cya*. The *H. influenzae* *cya* gene, like its *E. coli* homolog, has a good CRP site overlapping its promoter, which is predicted to act as a repressor rather than an activator of transcription.²⁹ Consistent with this, *cya* mRNA was sharply decreased after transfer to MIV (Figure 2(d)) and increased about sixfold in *cya* mutant cells. The *icc* gene (cAMP phosphodiesterase) was induced five- to sixfold in MIV (Figure 2(d)) and decreased about twofold in the absence of cAMP. The induction of *icc* in MIV will increase cAMP turnover and, with decreased transcription of *cya*, limit the cell's long-term response to activation of adenylate cyclase by the PTS. The *sxy* knockout mutation had no effect on transcription of *cya*, *icc* or *crp*.

Regulation of Sxy

The regulatory gene *sxy* was induced 16–40-fold after transfer to MIV, with maximum expression in the 30 minute sample (Figure 2(d)). This induction is consistent with Sxy's role as a positive regulator of CRE-regulon genes, and with previous primer-extension analysis.³³ The *cya* mutation had no effect on *sxy* expression, contrary to a previous report that cAMP induces *sxy* transcription.^{33,34} Expression of a *lacZ* fusion to the *sxy* promoter does not depend on the presence of an intact *sxy* gene, so autoregulation is not a factor, but differences between results with transcriptional and translational fusions suggest a role for post-transcriptional regulation.³⁴

Other genes with regulatory roles were not induced by transfer to MIV. Although point mutations in *murE* are known to increase expression of the CRE-regulon genes *comA* and *rec-2*,⁴⁵

expression of *murE* and other genes involved in cell wall synthesis and recycling was not altered by any of the treatments tested. With the exception of *radC*, genes with direct roles in DNA recombination and repair were not induced by transfer to MIV and are not regulated by CRP or Sxy (*lig*, *mutT*, *mutY*, *radA*, *recA*, *recBCD*, *recF*, *recG*, *recN*, *recO*, *recR*, *ruvABC*, *ung*, *uvrABCD*). Expression of the genes for mismatch repair was neither increased nor decreased in our array experiments, suggesting that recombination of heterologous DNA may not have played a significant role in the evolution of competence.

Competence development in rich medium

Cultures become modestly competent in colonies on sBHI agar plates and when liquid sBHI cultures approach stationary phase.³² With the exception of *ssb*, all genes in the CRE regulon were also modestly induced (4–20 times) as stationary phase approached during growth in rich medium (data not shown). This suggests that the low level of competence seen at this stage is not due to failure of a particular competence function, but to a general low induction of all components. This is supported by the modest increases in expression of *sxy* and of the CRP-regulon genes induced in MIV.

Discussion

What do the CRE-regulon genes do?

DNA uptake and translocation functions

Several of the genes in the CRE regulon are known to have roles in assembly of the uptake machinery or in DNA transport. Insertions disrupting *comA* and *comC* prevent DNA binding and uptake; however, their mutant phenotypes could be due to polar effects on *comE*. ComA is predicted to be cytoplasmic and ComC to be targeted to the inner membrane. ComE is a member of the secretin family of gated pore proteins associated with type IV pili. The *pilA* gene encodes a typical pilin subunit of type IV pili; the *pilBCD* genes are homologous to genes for pilin processing and pilus assembly. As *H. influenzae* lacks visible type IV pili these genes likely produce a short pseudopilus. An insertion disrupting the inner-membrane protein Rec-2 allows DNA binding and uptake into the periplasm but the DNA cannot be translocated into the cytoplasm.⁸ Mutations in *comF* (original name *com101A*) cause a similar phenotype.⁴⁷ The ComE1 protein is homologous to the C-terminal region of the *Bacillus subtilis* DNA-uptake protein ComEA, and our preliminary data implicate ComE1 in DNA uptake by *H. influenzae* (S. Molnar & R.R., unpublished results).

Genes in the HI0938-0941 operon have not previously been associated with competence. They have good homologs only in the Pasteurellaceae,

but weak homologs occur in similar operons in many other bacteria. All are predicted to be secreted from the cytoplasm, and an insertion in HI0938 prevents DNA uptake. (S. Molnar & R.R., unpublished results). HI1182/1183 (incorrectly annotated as two ORFs due to a sequencing frameshift) belongs to a small family of ATP-dependent DNA ligases with signal sequences for secretion into the periplasm. Both the *H. influenzae* and *Neisseria gonorrhoeae* ligases have been well-characterized *in vitro*, but no periplasmic function is known.^{48,49}

Cytoplasmic functions

The CRE-regulon proteins with known cytoplasmic functions all interact with DNA but have not been implicated in DNA uptake. Insertions in *comM* and *dprA* cause DNA entering the cytoplasm to be degraded before it can recombine with the chromosome.^{9,11} ComM is a member of the YifB subfamily of AAA-ATPase proteins, its possession of a *lon* protease domain suggests it may be an ATP-dependent protease.⁵⁰ DprA (Smf) is also predicted to bind ATP; it is required for transformation in a number of bacteria and its homolog in *Streptococcus pneumoniae* also protects DNA from degradation.⁵¹ The SSB protein is ubiquitous and well characterized. Its ability to bind and stabilize single-stranded DNA is essential for DNA replication and repair, and also plays a role in homologous recombination.⁵² Perhaps because *ssb* transcripts are abundant in log-phase cells, competence-specific regulation had only modest effects. The function of *radC* is less well understood; it encodes a RecG-like protein thought to function at stalled DNA replication forks, and is also a component of the *S. pneumoniae* and *B. subtilis* competence regulons.^{20,53,54}

Other proteins

Other CRE regulon genes encode cytoplasmic proteins of unknown function. HI0365 contains a Fe-S oxidoreductase domain. HI0659 and HI0660 are short proteins that are conserved as an operon in a number of distantly related bacteria; HI0659 contains a helix-turn-helix domain in the XRE family.⁵⁵ HI1631 has no known homologs.

In *H. influenzae* the ability to take up DNA develops as a unified response to changing conditions. The CRE regulon contains all competence genes induced in MIV starvation medium and all genes regulated by Sxy, suggesting that the regulation of competence will be understood only when we understand the regulation of *sxy* expression and its role in expression of the CRE-regulon genes. Both depleted purine pools and mutant forms of *murE* can cause overexpression of CRE-regulon genes, and our preliminary results indicate that both act by increasing *sxy* expression (Q.Q. & J. Hill, unpublished results). Furthermore, *sxy* point mutations causing hyper-competence affect base-pairing in the 5'-leader of *sxy* mRNA.³⁴ This may reflect coupling of transcription to translation,

involvement of antisense RNA or action of a riboswitch.⁵⁶

This work has identified 11 new genes not previously associated with competence. The most intriguing of these is the periplasmic ATP-dependent DNA ligase encoded by HI1182/83. Similar secreted ATP-dependent DNA ligases are found in *Neisseria* and several other bacteria;^{48,49} they belong to one of two newly discovered families of bacterial ATP-dependent DNA ligases. Both the *H. influenzae* and *Neisseria* ligases are known to seal nicks but not blunt ends, and Magnet & Blanchard presciently speculated that they might function in competence.⁴⁹ Consistent with this, transformation with cloned fragments bearing restriction fragment ("sticky") ends often gives transformants containing conjoined DNA fragments, and a periplasmic ligase activity was proposed in explanation.⁵⁷ However, there is no obvious role for ligation in DNA uptake, and there is unlikely to be any ATP in the periplasm, especially because one of the MIV-induced genes is the periplasmic 5'-nucleotidase encoded by HI0206.⁵⁸

We do not know whether the remaining ten genes of the CRE regulon all contribute to competence or reflect a broader role of the CRE regulon, perhaps resolving other problems created by depletion of nucleotide pools. In *B. subtilis* and *S. pneumoniae*, what were originally thought to be competence-specific signals control many genes of diverse function.^{19,20,54} Competence regulons have not yet been identified in other Gram-negative bacteria. Characterization of the functions of the new CRE-regulon genes and of possible CRE regulons in related bacteria should help resolve this issue.

An unresolved question underlying any comparison of mechanisms or regulation is the evolutionary origin of competence. The known distribution is sporadic; bacteria from many different lineages have been shown to take up DNA, but often competence cannot be demonstrated in their close relatives.⁵⁹ Competence may be truly sporadic, arising independently many times in lineages of otherwise non-competent bacteria, in response to changing conditions, or it may have arisen once early in bacterial evolution but be frequently lost or undetectable in laboratory cultures. The diversity of regulatory mechanisms appears to support the former, but the evolutionarily lability of gene regulation may mask ancestral homology. The mechanisms of DNA uptake have components that are clearly homologous in almost all competent bacteria, and a better understanding of their function may clarify the issue.

Materials and Methods

Microarray slide preparation

The *H. influenzae* whole genome microarray was based on the annotated sequence of the Rd strain.⁶⁰ Primer3 software⁶¹ was used to design primer pairs to amplify an

internal sequence of each ORF. Software parameters dictated the annealing temperatures of approximately 55 °C and PCR product sizes between approximately 175 and 600 bp. BLAST analysis was used to minimize homology with other ORFs within the genome. All PCR products were confirmed using agarose gels. Reactions with multiple or no products were repeated at lower and higher annealing temperatures, and those that produced incorrect-sized products had their primers redesigned.

PCR products of all 1738 *H. influenzae* genes were spotted in duplicate onto poly-L-lysine-coated glass microscope slides by a MicroGridII robot (BioRobotics, UK), using the facilities of the Bacterial Microarray Group at St. George's Hospital Medical School, London. Control spots were: *H. influenzae* 5 S, 16 S and 23 S rRNA genes; human and rat actin genes; and *E. coli lacZ* and *glpD* genes. tRNA genes were not included. Slide processing prior to hybridization has been described.⁶² Quality controls used the first and last slides of each print run.

Strains

KW20 is the standard *H. influenzae* Rd strain sequenced by Fleischman *et al.*⁶⁰ The MAP7, *cya*, and *sxy* knockout strains have been described.^{28,29,32,57} The FNR knockout strain (RR838) was constructed by cloning the PCR-amplified *fnr* gene from KW20 into pGEM-T easy (Promega), inserting a Sall-cut kanamycin-resistance cassette from pWJC3⁶³ into the internal XhoI site, and transforming the disrupted gene into KW20. The structure of the mutant gene was confirmed by Southern blotting.

Culture growth and competence protocols have been described.^{57,64} MIV medium contains (all amounts in µg/ml): Arg, 21; Asp, 4032; Cys, 6; Glu, 314; Leu, 61; Lys, 35; Met, 18; Ser, 65; Tyr, 42; Ile, 33; Gly, 2.5; His, 13; Val, 35; Phe, 46; Thr, 20; Ala, 48; Pro, 50; fumarate, 1000; citrulline, 12; Tween-80, 200; NaCl, 4675; MgSO₄, 124; CaCl₂, 147; KH₂PO₄, 1740. Cultures used for the time-courses were pre-grown in sBHI at densities below 2 × 10⁸ cfu/ml for at least two hours before the first time point was taken. Sample times are specified relative to *t*=0, when cells in sBHI at a density of 8 × 10⁸ (*A*₆₀₀=0.2) were transferred to MIV. Time-course samples were taken from cells in sBHI at *t* = -70, -30, 45, 80 and 130 minutes, and from cells in MIV at *t* = 10, 30, 60 and 100 minutes. Competence of the 100 minute sample was confirmed by transformation to Novobiocin resistance with DNA of the Nov^R strain MAP7. Samples for *cya* (four replicate experiments) and *sxy* analysis (five replicate experiments) were taken after 100 minutes of incubation in MIV.

RNA methods

Aliquots of cells (usually 2 ml) were taken from liquid cultures, pelleted (one minute at 10,000g), quick-chilled and stored frozen at -80 °C. RNAs were prepared from these pellets using Qiagen RNeasy kits, and were freed of contaminating DNA with either Qiagen on-column DNase I digestion or an Ambion DNA-Free kit. RNA concentrations were determined spectrophotometrically and RNA quality was assessed by gel electrophoresis.

Microarray methods

cDNAs from signal and control RNAs were labeled with either Cy3 and Cy5 or Cy5 and Cy3, respectively, in corresponding replicate experiments to limit artifacts caused by the potential differences of Cy3 or Cy5 in

labeling efficiency. Production of cDNA probes labeled with Cy3 and Cy5 and microarray hybridization used either of two protocols, one developed by the Bacterial Microarray Group for labeling with Cy3-dUTP and Cy5-dUTP⁶⁵ and the other by TIGR for amino-allyl labeling with Cy3 and Cy5†. For analysis of time-course samples, a control RNA pool containing equal amounts of RNA from all nine samples was prepared and used as competitor for each sample. This improves the quantification of RNAs that are expressed at very low levels in some samples.²⁰

Analysis of microarray data

Microarray slides were scanned and intensity data collected from the images using Imagene software.

Time-course data

Data for virtual *t*=0 minute samples from the time-courses were created by averaging the two exponential growth samples (*t* = -70 and *t* = -30 minutes). The data were imported into GeneSpring (Silicon Genetics, v6.0) and assembled into the multi-sample experiments indicated in Figure 1 (e.g. see Figure 2). Datasets were normalized using GeneSpring's default "per spot" normalization step and a modified "per chip" normalization that restricted the measurements used in the calculation of the median to current normalized values of at least 0.01. In addition, extra background correction was applied when needed.

cAMP and *sxy* data

Replicate slides were combined into a ±cAMP dataset and a ±*sxy* dataset, and each dataset was normalized using the default per chip step and a refined per spot step that decreased the cut-off value of the control channel from 10 to 0.01 to improve the spot-detection sensitivity.

Quantitative PCR

RNAs were prepared from an independent MIV time-course (-40, and 0 minutes in sBHI, 20, 60 and 100 minutes in MIV), from wild-type and *sxy* knockout cells at 60 and 100 minutes in MIV, and from *cya* knockout cells at 60 and 100 minutes in MIV ± 1 mM cAMP. cDNA templates were generated using the iScript cDNA synthesis kit (BioRad). Reactions were carried out in duplicate in a 7000 SDS (Sequence Detection System) (Applied Biosystems) using the iTaqSYBR Green Supermix with Rox (BioRad) and primers designed with Primer Express 2.0 (Applied Biosystems) and on-line Net Primer for PCR products. The standard curves used five serial fivefold dilutions of a MAP7 genomic DNA template. Relative RNA abundance measurements were calculated by normalizing derived quantity of cDNA template (*ssb* or *comF*) to that of a control (*murG*), chosen because of its strong constant expression in the microarray time-courses.

Electrophoretic mobility-shift assays

Fragments containing the *mgIBAC* and *comA-F* promoters (each 130 bp) were PCR-amplified from *H. influenzae* genomic DNA, purified on a 5% (w/v) acrylamide gel, and internally labeled in 12.5 µl reactions containing

† <http://www.tigr.org/tdb/microarray/protocolsTIGR.shtml>

50 ng of DNA, 6 μ M of each PCR primer, 50 μ M d(C,G,T)TP mix, 2 μ M dATP, 0.8 μ M [α - 33 P]dATP (2500 Ci/mmol), 1 \times Klenow buffer, and two units of Klenow enzyme. DNA and oligonucleotides were heated to 94 $^{\circ}$ C for three minutes then placed on ice. Nucleotides, buffer, and enzyme were then added and the reaction was incubated at room temperature for two hours. The reaction was stopped by heating to 80 $^{\circ}$ C for 20 minutes, diluted in 150 μ l of TE (10 mM Tris, 1 mM EDTA (pH 7.5)), and stored at -20° C. CRP was purified from *E. coli* (DH5 α) cells carrying the plasmid pXN15, which encodes *E. coli* *crp* and its native promoter.^{28,66} Protein purity was assessed using SDS-PAGE and Coomassie staining and protein concentration was measured using the BioRad DC Protein (Lowry) assay. Binding reactions (10 μ l) contained 10 mM Tris (pH 8.0), 50 mM KCl, 0.5 mM EDTA, 10% (v/v) glycerol, 250 μ g/ml bovine serum albumin, 100 μ M cAMP, 1 mM dithiothreitol, 40 μ g/ml poly(dI-dC) DNA, 2.8 ng labeled DNA (400,000 cpm/ng), and purified CRP as indicated. Reactions were incubated at room temperature for ten minutes then loaded onto a prerun polyacrylamide gel (30:1 acrylamide/bisacrylamide; 0.2 \times TBE (89 mM Tris, 89 mM borate, 2 mM EDTA (pH 8.3)), 2% glycerol, and 200 μ M cAMP; running buffer 0.2 \times TBE and 200 μ M cAMP). Following electrophoresis for 2.5 hours at 100V, the gel was dried and exposed for one hour to a phosphor screen. Bands were visualized using a STORM 860 scanner.

Sequence analysis

The program RSA-tools was used to search for sequences resembling CRE sites.³⁶ The input matrix was based on the first nine CRE sequences shown in Figure 3, using the calculations described by Macfadyen.³⁵ Sequence motifs were identified using the programs Consensus,³⁹ Gibbs recursive sampler⁴⁰ and Bioprospector.⁴¹ Sequence logos were generated using WebLogo.⁶⁷

Database deposition

Fully annotated data from these arrays have been placed in the B μ G@Sbase, accession no. E-BUGS-20 (<http://bugs.sghms.ac.uk/E-BUGS-20>) and ArrayExpress accession no. E-BUGS-20.

Acknowledgements

R.J.R. & P.R.L. thank the Canadian Institutes for Health Research and the Wellcome Trust, respectively, for financial support. We acknowledge Phil Butcher and B μ G@S (the Bacterial Microarray Group at St George's Hospital Medical School) for supply of the microarrays and advice and the Wellcome Trust for funding the multi-collaborative microbial pathogen microarray facility under its Functional Genomics Resources Initiative. The work made extensive use of the TIGR Comprehensive Microbial Resource. Robert Hancock's group (UBC) provided array analysis facilities and Michelle Brazas gave invaluable advice. Sebastian Molnar and Jessica Hill provided data prior to publication, and Leah Macfadyen provided editorial assistance.

Supplementary data

Supplementary data associated with this article can be found, in the online version, at [doi:10.1016/j.jmb.2005.01.012](https://doi.org/10.1016/j.jmb.2005.01.012)

References

- Stone, B. J. & Kwai, Y. A. (1999). Natural competence for DNA transformation by *Legionella pneumophila* and its association with expression of Type IV pili. *J. Bacteriol.* **181**, 1395–1402.
- Chang, K. C., Yeh, Y. C., Lin, T. L. & Wang, J. T. (2001). Identification of genes associated with natural competence in *Helicobacter pylori* by transposon shuttle random mutagenesis. *Biochem. Biophys. Res. Commun.* **288**, 961–968.
- Friedrich, A., Hartsch, T. & Averbach, B. (2001). Natural transformation in mesophilic and thermophilic bacteria: identification and characterization of novel, closely related competence genes in *Acinetobacter* sp. strain BD413 and *Thermus thermophilus* HB27. *Appl. Environ. Microbiol.* **67**, 3140–3148.
- Aas, F. E., Wolfgang, M., Frye, S., Dunham, S., Lovold, C. & Koomey, M. (2002). Competence for natural transformation in *Neisseria gonorrhoeae*: components of DNA binding and uptake linked to type IV pilus expression. *Mol. Microbiol.* **46**, 749–760.
- Chen, I. & Dubnau, D. (2004). DNA uptake during bacterial transformation. *Nature Rev. Microbiol.* **2**, 241–249.
- Tomb, J. F., el-Hajj, H. & Smith, H. O. (1991). Nucleotide sequence of a cluster of genes involved in the transformation of *Haemophilus influenzae* Rd. *Gene*, **104**, 1–10.
- Larson, T. G. & Goodgal, S. H. (1991). Sequence and transcriptional regulation of *com101A*, a locus required for genetic transformation in *Haemophilus influenzae*. *J. Bacteriol.* **173**, 4683–4691.
- McCarthy, D. (1989). Cloning of the *rec-2* locus of *Haemophilus influenzae*. *Gene*, **75**, 135–143.
- Karudapuram, S., Zhao, X. & Barcak, G. J. (1995). DNA sequence and characterization of *Haemophilus influenzae* *dprA+*, a gene required for chromosomal but not plasmid DNA transformation. *J. Bacteriol.* **177**, 3235–3240.
- Karudapuram, S. & Barcak, G. J. (1997). The *Haemophilus influenzae* *dprABC* genes constitute a competence-inducible operon that requires the product of the *tfoX* (*sxy*) gene for transcriptional activation. *J. Bacteriol.* **179**, 4815–4820.
- Gwinn, M. L., Ramanathan, R., Smith, H. O. & Tomb, J. F. (1998). A new transformation-deficient mutant of *Haemophilus influenzae* Rd with normal DNA uptake. *J. Bacteriol.* **180**, 746–748.
- Tomb, J. F., Barcak, G. J., Chandler, M. S., Redfield, R. J. & Smith, H. O. (1989). Transposon mutagenesis, characterization, and cloning of transformation genes of *Haemophilus influenzae* Rd. *J. Bacteriol.* **171**, 3796–3802.
- Dougherty, B. A. & Smith, H. O. (1999). Identification of *Haemophilus influenzae* Rd transformation genes using cassette mutagenesis. *Microbiology*, **145**, 401–409.
- Gwinn, M. L., Stellwagen, A. E., Craig, N. L., Tomb, J. F. & Smith, H. O. (1997). *In vitro* Tn7 mutagenesis of

- Haemophilus influenzae* Rd and characterization of the role of *atpA* in transformation. *J. Bacteriol.* **179**, 7315–7320.
15. Beattie, K. L. & Setlow, J. K. (1971). Transformation-defective strains of *Haemophilus influenzae*. *Nature New Biol.* **231**, 177–179.
 16. Caster, J. H., Postel, E. H. & Goodgal, S. H. (1970). Competence mutants: isolation of transformation deficient strains of *Haemophilus influenzae*. *Nature*, **227**, 515–517.
 17. Pifer, M. L. & Smith, H. O. (1985). Processing of donor DNA during *Haemophilus influenzae* transformation: analysis using a model plasmid system. *Proc. Natl Acad. Sci. USA*, **82**, 3731–3735.
 18. Redfield, R. J. (2001). Do bacteria have sex? *Nature Rev. Genet.* **2**, 634–639.
 19. Dagkessamanskaia, A., Moscoso, M., Henard, V., Guiral, S., Overweg, K., Reuter, M. *et al.* (2004). Interconnection of competence, stress and CiaR regulons in *Streptococcus pneumoniae*: competence triggers stationary phase autolysis of *ciaR* mutant cells. *Mol. Microbiol.* **51**, 1071–1086.
 20. Peterson, S. N., Sung, C. K., Cline, R., Desai, B. V., Snesrud, E. C., Luo, P. *et al.* (2004). Identification of competence pheromone responsive genes in *Streptococcus pneumoniae* by use of DNA microarrays. *Mol. Microbiol.* **51**, 1051–1070.
 21. Ogura, M., Yamaguchi, H., Kobayashi, K., Ogasawara, N., Fujita, Y. & Tanaka, T. (2002). Whole-genome analysis of genes regulated by the *Bacillus subtilis* competence transcription factor ComK. *J. Bacteriol.* **184**, 2344–2351.
 22. Herriott, R. M., Meyer, E. M. & Vogt, M. (1970). Defined non-growth media for stage II development of competence in *Haemophilus influenzae*. *J. Bacteriol.* **101**, 517–524.
 23. Redfield, R. J. (1991). *sxy-1*, a *Haemophilus influenzae* mutation causing greatly enhanced spontaneous competence. *J. Bacteriol.* **173**, 5612–5618.
 24. Reizer, J. & Reizer, A. (1996). A voyage along the bases: novel phosphotransferase genes revealed by *in silico* analyses of the *Escherichia coli* genome. *Res. Microbiol.* **147**, 458–471.
 25. Macfadyen, L. P., Dorocicz, I. R., Reizer, J., Saier, M. H., Jr & Redfield, R. J. (1996). Regulation of competence development and sugar utilization in *Haemophilus influenzae* Rd by a phosphoenolpyruvate: fructose phosphotransferase system. *Mol. Microbiol.* **21**, 941–952.
 26. Macfadyen, L. P. (1999). Regulation of intracellular cAMP levels and competence development in *Haemophilus influenzae* by a phosphoenolpyruvate: fructose phosphotransferase system. PhD Thesis, University of British Columbia.
 27. Gwinn, M. L., Yi, D., Smith, H. O. & Tomb, J. F. (1996). Role of the two-component signal transduction and the phosphoenolpyruvate: carbohydrate phosphotransferase systems in competence development of *Haemophilus influenzae* Rd. *J. Bacteriol.* **178**, 6366–6368.
 28. Chandler, M. S. (1992). The gene encoding cyclic AMP receptor protein is required for competence development in *Haemophilus influenzae* Rd. *Proc. Natl Acad. Sci. USA*, **89**, 1626–1630.
 29. Dorocicz, I. R., Williams, P. M. & Redfield, R. J. (1993). The *Haemophilus influenzae* adenylate cyclase gene: cloning, sequence, and essential role in competence. *J. Bacteriol.* **175**, 7142–7149.
 30. Tan, K., Moreno-Hagelsieb, G., Collado-Vides, J. & Stormo, G. D. (2001). A comparative genomics approach to prediction of new members of regulons. *Genome Res.* **11**, 566–584.
 31. Wise, E. M., Alexander, S. & Powers, M. (1973). Adenosine 3':5'-cyclic monophosphate as a regulator of bacterial transformation. *Proc. Natl Acad. Sci. USA*, **70**, 471–474.
 32. Williams, P. M., Bannister, L. A. & Redfield, R. J. (1994). The *Haemophilus influenzae sxy-1* mutation is in a newly identified gene essential for competence. *J. Bacteriol.* **176**, 6789–6794.
 33. Zully, J. J. & Barcak, G. J. (1995). Identification of a DNA transformation gene required for *com101A* + expression and supertransformer phenotype in *Haemophilus influenzae*. *Proc. Natl Acad. Sci. USA*, **92**, 3616–3620.
 34. Bannister, L. A. (1999). An RNA secondary structure regulates *sxy* expression and competence development in *Haemophilus influenzae*. PhD Thesis, University of British Columbia.
 35. Macfadyen, L. P. (2000). Regulation of competence development in *Haemophilus influenzae*. *J. Theoret. Biol.* **207**, 349–359.
 36. van Helden, J. (2003). Regulatory sequence analysis tools. *Nucl. Acids Res.* **31**, 3593–3596.
 37. Schultz, S. C., Shields, G. C. & Steitz, T. A. (1991). Crystal structure of a CAP-DNA complex: the DNA is bent by 90 degrees. *Science*, **253**, 1001–1007.
 38. Kolb, A., Busby, S., Buc, H., Garges, S. & Adhya, S. (1993). Transcriptional regulation by cAMP and its receptor protein. *Annu. Rev. Biochem.* **62**, 749–795.
 39. Hertz, G. Z. & Stormo, G. D. (1999). Identifying DNA and protein patterns with statistically significant alignments of multiple sequences. *Bioinformatics*, **15**, 563–577.
 40. Thompson, W., Rouchka, E. C. & Lawrence, C. E. (2003). Gibbs Recursive Sampler: finding transcription factor binding sites. *Nucl. Acids Res.* **31**, 3580–3585.
 41. Liu, X., Brutlag, D. L. & Liu, J. S. (2001). BioProspector: discovering conserved DNA motifs in upstream regulatory regions of co-expressed genes. *Pac. Symp. Biocomput.* 127–138.
 42. Chen, S., Gunasekera, A., Zhang, X., Kunkel, T. A., Ebright, R. H. & Berman, H. M. (2001). Indirect readout of DNA sequence at the primary-kink site in the CAP-DNA complex: alteration of DNA binding specificity through alteration of DNA kinking. *J. Mol. Biol.* **314**, 75–82.
 43. Spiro, S., Gaston, K. L., Bell, A. I., Roberts, R. E., Busby, S. J. & Guest, J. R. (1990). Interconversion of the DNA-binding specificities of two related transcription regulators, CRP and FNR. *Mol. Microbiol.* **4**, 1831–1838.
 44. Chandler, M. S. & Smith, R. A. (1996). Characterization of the *Haemophilus influenzae topA* locus: DNA topoisomerase I is required for genetic competence. *Gene*, **169**, 25–31.
 45. Ma, C. & Redfield, R. J. (2000). Point mutations in a peptidoglycan biosynthesis gene cause competence induction in *Haemophilus influenzae*. *J. Bacteriol.* **182**, 3323–3330.
 46. Finkel, S. E. & Kolter, R. (2001). DNA as a nutrient: novel role for bacterial competence gene homologs. *J. Bacteriol.* **183**, 6288–6293.
 47. Larson, T. G. & Goodgal, S. H. (1992). Donor DNA processing is blocked by a mutation in the *com101A* locus of *Haemophilus influenzae*. *J. Bacteriol.* **174**, 3392–3394.

48. Cheng, C. & Shuman, S. (1997). Characterization of an ATP-dependent DNA ligase encoded by *Haemophilus influenzae*. *Nucl. Acids Res.* **25**, 1369–1374.
49. Magnet, S. & Blanchard, J. S. (2004). Mechanistic and kinetic study of the ATP-dependent DNA ligase of *Neisseria meningitidis*. *Biochemistry*, **43**, 710–717.
50. Iyer, L. M., Leipe, D. D., Koonin, E. V. & Aravind, L. (2004). Evolutionary history and higher order classification of AAA+ ATPases. *J. Struct. Biol.* **146**, 11–31.
51. Berge, M., Mortier-Barriere, I., Martin, B. & Claverys, J. P. (2003). Transformation of *Streptococcus pneumoniae* relies on DprA- and RecA-dependent protection of incoming DNA single strands. *Mol. Microbiol.* **50**, 527–536.
52. Raghunathan, S., Kozlov, A. G., Lohman, T. M. & Waksman, G. (2000). Structure of the DNA binding domain of *E. coli* SSB bound to ssDNA. *Nature Struct. Biol.* **7**, 648–652.
53. Saveson, C. J. & Lovett, S. T. (1999). Tandem repeat recombination induced by replication fork defects in *Escherichia coli* requires a novel factor, RadC. *Genetics*, **152**, 5–13.
54. Berka, R. M., Hahn, J., Albano, M., Draskovic, I., Persuh, M., Cui, X. *et al.* (2002). Microarray analysis of the *Bacillus subtilis* K-state: genome-wide expression changes dependent on ComK. *Mol. Microbiol.* **43**, 1331–1345.
55. Wood, H. E., Devine, K. M. & McConnell, D. J. (1990). Characterisation of a repressor gene (*xre*) and a temperature-sensitive allele from the *Bacillus subtilis* prophage, PBSX. *Gene*, **96**, 83–88.
56. Nudler, E. & Mironov, A. S. (2004). The riboswitch control of bacterial metabolism. *Trends Biochem. Sci.* **29**, 11–17.
57. Poje, G. & Redfield, R. J. (2003). Transformation of *Haemophilus influenzae*. *Methods Mol. Med.* **71**, 57–70.
58. Zagursky, R. J., Ooi, P., Jones, K. F., Fiske, M. J., Smith, R. P. & Green, B. A. (2000). Identification of a *Haemophilus influenzae* 5'-nucleotidase protein: cloning of the *nucA* gene and immunogenicity and characterization of the NucA protein. *Infect. Immun.* **68**, 2525–2534.
59. Sikorski, J., Teschner, N. & Wackernagel, W. (2002). Highly different levels of natural transformation are associated with genomic subgroups within a local population of *Pseudomonas stutzeri* from soil. *Appl. Environ. Microbiol.* **68**, 865–873.
60. Fleischmann, R. D., Adams, M. D., White, O., Clayton, R. A., Kirkness, E. F., Kerlavage, A. *et al.* (1995). Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science*, **269**, 496–512.
61. Rozen, S. & Skaletsky, H. J. (2000). Primer3 on the WWW for general users and for biologist programmers. In *Bioinformatics Methods and Protocols: Methods in Molecular Biology* (Krawetz, S. & Misenes, S., eds), pp. 365–386, Humana Press, Totowa, NJ.
62. DeRisi, J. L., Iyer, V. R. & Brown, P. O. (1997). Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science*, **278**, 680–686.
63. Chen, W., Gross, L., Joho, K. & McAllister, W. T. (1992). A modified kanamycin-resistance cassette to facilitate two-codon insertion mutagenesis. *Gene*, **111**, 143–144.
64. Poje, G. & Redfield, R. J. (2003). General methods for culturing *Haemophilus influenzae*. *Methods Mol. Med.* **71**, 51–56.
65. Bacon, J., James, B. W., Wernisch, L., Williams, A., Morley, K. A., Hatch, G. J. *et al.* (2004). The influence of reduced oxygen availability on pathogenicity and gene expression in *Mycobacterium tuberculosis*. *Tuberculosis (Edinb)*, **84**, 205–217.
66. Zhang, X. P., Gunasekera, A., Ebright, Y. W. & Ebright, R. H. (1991). Derivatives of CAP having no solvent-accessible cysteine residues, or having a unique solvent-accessible cysteine residue at amino acid 2 of the helix-turn-helix motif. *J. Biomol. Struct. Dynam.* **9**, 463–473.
67. Crooks, G. E., Hon, G., Chandonia, J. M. & Brenner, S. E. (2004). WebLogo: a sequence logo generator. *Genome Res.* **14**, 1188–1190.

Edited by J. Karn

(Received 12 October 2004; received in revised form 23 December 2004; accepted 5 January 2005)

Note added in proof: While this paper was under review, Van Waggoner *et al.* also identified the CRE sites in the HI0665, HI0938 and HI1182 gene promoters and used real-time PCR to confirm competence-induced expression (Van Waggoner *et al.* (2004). *J. Bacteriol.* **186**, 6409–6421). Gene knockouts showed that HI0366 (but not HI0365) and HI0939 are required for DNA uptake, and that HI1182 plays a modest role in transformation but is not required for DNA uptake.