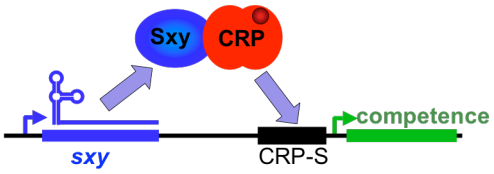
Introduction

This proposal addresses two regulatory phenomena discovered by our work on natural competence in *Haemophilus influenzae*. The first is the regulation of the new transcription factor Sxy, and the second is the action of Sxy at CRP promoters. A more detailed version of the adjacent figure is provided as Fig. 1.



Background

The following sections first introduce H. influenzae and its natural competence system. I then describe in detail the regulation of competence, emphasizing the post-transcriptional control of sxy expression and the action of CRP and Sxy at CRP-S promoters.

Until now our studies have been largely focused on *H. influenzae*, where natural competence has provided both motivation and a tool for investigation. Thus the background information we provide emphasizes regulatory processes in *H. influenzae*. Our understanding of these processes has benefited greatly from the large body of information available for *Escherichia coli*'s CRP. Our recent demonstration that *E. coli* and *H. influenzae* share the CRP-S regulon allows us to now take advantage of the more powerful methods available in *E. coli* (**#3**) **[***References to our papers will be starred; # indicates provided papers*]. Studies in both organisms will generate an enriched understanding of both this novel regulatory mechanism and the signals that induce DNA uptake in both organisms.

Global significance of H. influenzae and E. coli

H. influenzae's only niche is the human body. This small gram-negative bacterium is usually commensal in the human upper respiratory tract but is also a common cause of childhood ear infections and of respiratory disease in infants, the elderly and people with cystic fibrosis and AIDS. Serotype b strains have historically been the major cause of meningitis in infants and small children, with a 6% mortality rate and residual damage to hearing or intellect in about 50% of cases [1-3]

As a leading cause of infant diarrhea in the developing world, *E. coli* infections result in nearly a million deaths each year [4]. They also continue to be a major health concern in the first world. The mosaic genomes of *E. coli* strains arise largely by the acquisition of laterally transferred genes. Thus, gene transfer and recombination facilitates the evolution of distinct pathotypes that exploit various infectious niches [5].

H. influenzae is the model system for studies of natural competence in the γ-proteobacteria. *E. coli*, in contrast, is the model system for studies of gene regulation, especially of the archetypal transcriptional activator CRP. Recently *E. coli* has been shown to also be naturally competent, and to use many of competence genes to enhance survival in long term culture by taking up DNA, raising the possibility that uptake of environmental DNA promotes its survival on food crops, in contaminated water and in the lower intestine.

Natural competence: control and controversy

Most naturally competent bacteria express DNA uptake genes only in response to certain environmental conditions [6], and can then efficiently bind double-stranded DNA fragments and take one strand into the cell. If the sequence of this DNA strand is sufficiently similar to a sequence in the chromosome, homologous recombination may give the cell a new (transformed) genotype; otherwise the DNA is usually rapidly degraded in the cytoplasm (reviewed in **#1;**). Natural competence thus differs in both

regulation and mechanism from the artificial procedures used to introduce plasmids into cells. The regulation of competence and mechanism of DNA uptake have been well studied in the gram-negative *H. influenzae* and *Neisseria gonorrhoeae*, (and more recently *Acinetobacter*, *Helicobacter* and *Pseudomonas* species), and in the gram-positive *Bacillus subtilis* and *Streptococcus pneumoniae*.

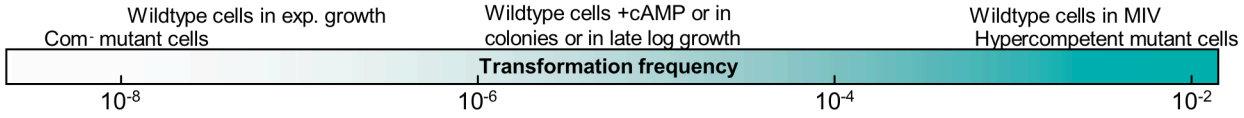
Competence is widespread, and almost all bacteria take up DNA by related mechanisms that depend on proteins of the type IV pilus/type II secretion system family [7]. However, regulation of competence in these bacteria uses a variety of apparently unrelated mechanisms. As described below, the γ−proteobacteria are an exception, as homologs of the genes of the *H. influenzae* competence regulon are regulated by the same mechanism in other *Pasteurellaceae* and in the *Enterobacteriaceae* and *Vibrionaceae* (**#3**).

One broad goal of my research is to clarify the evolutionary function of DNA uptake, which remains controversial even though the consequences of DNA uptake *per se* are not in question (**#1**). A cell that takes up DNA inevitably obtains the constituent nucleotides, reducing the demands on its biosynthetic or salvage pathways. Because DNA is abundant in natural environments and nucleotides are very expensive to synthesize (**#1**), the nucleotide benefit may be sufficient to compensate for the physiological and genetic costs of DNA uptake and thus to explain why competence exists. However another consequence of DNA uptake has dominated thinking about its function. Recombination of incoming DNA with the cell's chromosome may change its genotype, increasing or decreasing its ability to survive and reproduce. A better understanding of the control of competence will illuminate the role of genetic exchange in bacterial evolution and, in consequence, the still-controversial function of sex in eukaryotes (**#1**). I think that the best way to understand the evolution of competence is to understand its regulation, which has been shaped by natural selection to optimize the benefits. Because the genes that regulate competence evolved in the natural environment, understanding the signals they respond to will tell us which consequences of competence have been most beneficial.

Competence and transformation have medical relevance at several levels, because antibiotic resistance genes, virulence determinants and capsular serotype genes are spread by transformation. The *H. influenzae* genes responsible for DNA uptake are specifically induced in the human airway, and the DNA abundantly present in respiratory mucus is likely to be an important nutrient for these bacteria [8]. To understand how *H. influenzae* and other pathogens exploit their specific environments we must understand the regulation of competence.

How H. influenzae becomes competent

Development of competence: *H. influenzae* cells growing exponentially in rich medium do not take up DNA: competence genes are not expressed, DNA binding cannot be detected, and transformation frequencies with genetically marked chromosomal DNA are below the usual detection threshold of 10^{-8} ([9]*, **#2**). Levels of the competence activator protein Sxy and of cAMP (cofactor of the catabolite regulatory protein CRP), rise as cultures approach stationary phase, and rise more sharply when cells are transferred abruptly from rich medium to a starvation medium called MIV ("M-4"), which is commonly used to induce maximal competence. CRP and Sxy induce expression of a single competence regulon consisting of 25 genes in 13 transcription units, under the control of a novel CRP site (CRP-S site; details below and in **#2**, **#3** and **#5**). Properties of genes relevant to this proposal are given in Appendix Table 1. Most cells in MIV-induced cultures become competent within 60-100 minutes, giving transformation frequencies (TF) of 10^{-3}-10^{-2} as usually measured with chromosomal DNA containing

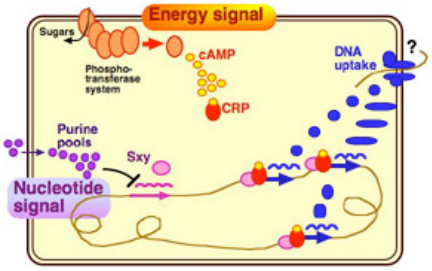


antibiotic resistance alleles. Cells also become competent when growing in colonies on agar or approaching stationary phase in broth also become competent, with transformation frequencies of about 10^{-5}-10^{-3}. Addition of cAMP to log-phase cultures induces a similar level of competence. Mutations affecting competence have a wide range of effects, from complete non-competence even in MIV ($TF<10^{-8}$) to extreme hypercompetence in exponential growth ($TF>10^{-3}$), as summarized in the diagram below.

DNA binding, uptake and translocation: Competent *H. influenzae* cells bind DNA at the cell surface, preferentially binding DNAs containing an uptake signal sequence (USS) with the 9-bp core AAGTGCGGT, present in about 2000 copies in the *H. influenzae* genome [10-12]*. DNA uptake proceeds by two stages, corresponding to transport across the outer and inner cell membranes [13]. Cells readily take up several hundred kb of DNA in a few minutes.

DNA degradation and recombination: The relative rates of degradation and recombination determine whether incoming DNA will change the cells' genotype. Degradation of one strand occurs concomitantly with translocation; the nucleotides released are rapidly reused for new DNA synthesis [14]. If sequence similarity permits, any undegraded strands may recombine with a homologous sequence in the chromosome; otherwise it will be rapidly degraded. Large insertions and deletions can be readily recombined if they are flanked by at least 1kb of homology [15], facilitating strain construction.

Key events in competence induction



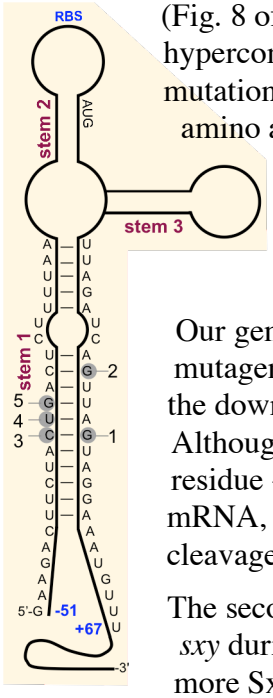
1. **Rising cAMP activates CRP.**
2. **CRP stimulates transcription of *sxy*.**
3. **Sxy is translated.**
4. **Sxy and CRP activate CRP-S regulon transcription.**
5. **PurR repression of *rec2* is released.**

1. Rising cAMP activates CRP: Transfer of exponentially growing cells to MIV causes an immediate rise in cytoplasmic cAMP. In *H. influenzae* as in most γ-proteobacteria, rising cAMP is a global signal of carbon and energy shortage [16,17]*; CRP transduces this signal by stimulating transcription of genes for use of alternative carbon or energy sources, or for sparing the wasteful use of the preferred sources.

CRP is very well characterized in *E. coli* and appears to act very similarly in *H. influenzae*; the DNA-recognition residues of the protein sequences are identical (Fig. 8 in **#3**, Fig. 2 in **#5**). It is usually abundant in the cell but is active only when cAMP levels are high [18]. CRP binds as a homodimer, specifically to symmetrical 22bp DNA sites with the consensus half site 5'-$A_1A_2A_3\mathbf{T_4G_5T_6G_7A_8}T_9C_{10}T_{11}$. The protein makes direct contact with bases G_5, G_7, and A_8 to induce a localized kink of 43° between positions 6 and 7, thus strengthening the association by wrapping the DNA around itself [19]. Because CRP binds only to very distinct DNA sites, contacts between it and RNA polymerase recruit polymerase to specific promoters.

2. CRP stimulates transcription of *sxy*: Until the competence-specific inducer Sxy has been expressed, CRP acts only at the canonical 'CRP-N' sites; these control over 100 *H. influenzae* genes, of which only *sxy* is directly relevant to competence (**#2**). Transcription of the *sxy* gene is strongly induced by active CRP (Fig. 9 of **#4**), but levels of Sxy protein are limited by one or more other regulatory factors acting through the secondary structure of *sxy* mRNA (**#4**).

3. Sxy is translated: The 5' end of *sxy* mRNA folds into a stem and complex loop; mutations that change the stability of the stem affect competence by changing the translatability of the mRNA



(Fig. 8 of **#4**). The *sxy* gene was first identified as the site of a mutation that causes hypercompetence (*sxy-1*) [9,20]*. We have now identified 4 additional hypercompetence mutations in *sxy*. All are single-nucleotide substitutions; only *sxy-1* changes the Sxy amino acid sequence but all weaken the basal stem in the 5' end of *sxy* mRNA (shown in the adjacent figure). We have confirmed most features of the secondary structure by *in vitro* RNase mapping, and confirmed its *in vivo* significance by showing that compensatory mutations that increase Stem 1's stability reduce Sxy expression and dramatically reduce or eliminate competence (**#4**).

Our genetic evidence for 2° structure is limited to a small region in Stem 1, and the *sxy* mutagenesis experiments we plan will expand this to cover the whole structure, including the downstream sequences that could contribute to regulation by alternative pairing. Although RNA structure prediction software did not show pairing of the region beyond residue +60, RNase analysis shows modest protection (*i.e.* base pairing) in wildtype *sxy* mRNA, and the *sxy-1* mutation not only destabilizes Stem 1A but also allows strong cleavage position +64 (Fig. 5 of **#4**).

The secondary structure must serve a genuinely regulatory function, as it limits translation of *sxy* during growth in rich medium but not in MIV. In other words, cells in MIV produce more Sxy protein from each mRNA (Fig. 8 in **#4**). Specific Aim I is to find out what this structure is sensing, and how the sensing occurs.

3. Depletion of purine nucleotides stimulates *sxy* translation: Nucleotide availability is known to affect competence induction [21]*. MIV medium lacks nucleotide precursors, and supplementation of MIV with purine nucleotides or nucleosides prevents induction of competence genes. As described in Section 5 below, the effect of purine nucleotides on competence induction is largely independent of the purine repressor PurR. We have now shown that purine nucleotides reduce both *sxy* mRNA and Sxy protein levels (Fig. 2A), and manuscript in preparation), and we hypothesize that nucleotides limit competence mainly by reducing *sxy* translation.

Translation is known to be blocked by extensive regions of double-stranded RNA at or near Shine-Dalgarno sites and start codons [22]. Because the mRNA surrounding the *sxy* SD site and start codon is predicted to remain largely unstructured until more than 100nt have exited the polymerase (**#4**), we hypothesize that the rate of polymerase progress directly influences the balance between *sxy* mRNA folding and initiation of translation. Precedents are known in the coupling of transcription and translation in genes for nucleotide biosynthesis, where translation occurs efficiently only when depleted nucleotide pools cause transcription to slow or pause (reviewed in [23]). Consistent with a role for Stem 1 in sensing nucleotide pools, expression of *sxy* in the *sxy* hypercompetence mutants is insensitive to the effect of added nucleotides (Fig. 2B). We will test whether *sxy* mRNA folding and translatability is mediated by RNA polymerase stalling when nucleotide pools are depleted *in vivo* and *in vitro*. In addition, new mutants will enable us to determine each stem and loop's role in uncoupling transcription and translation

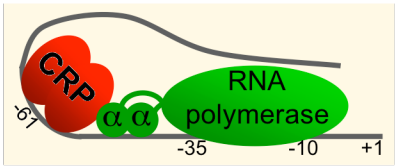
Such transcription-translation coupling is only the simplest model for *sxy* regulation. The *sxy* transcript may have specific sequences that promote polymerase pausing or stalling, either when specific nucleotides are unavailable or in response to another signal [24]. The mRNA may have an alternate folding, as in typical transcriptional attenuators [25]; this could explain the effect of the *sxy-1* mutation on downstream positions. Although the situation is ideal for regulation by a purine riboswitch, the *sxy* mRNA 2° structure has no resemblance to the well-conserved structures of known purine riboswitches (or of any other riboswitches) [26], and our unpublished RNase analysis has shown that the 2° structure is not sensitive to purines ribonucleotides (Fig. 3). Involvement of a small regulatory or antisense RNA

has not been ruled out, although the genome contains no sequences with significant complementarity to *sxy* mRNA. All *Pasteurellaceae sxy* genes are preceded by long intergenic regions that could contribute to regulation (*H. influenzae*'s is shortest at 319bp), but the sequences are quite diverged.

```
CRP-N:  wwwTGTGAnnnnnnTCACAwww
CRP-S:  TTTTGCGATnnnnATCGCAAAA
```

4. Sxy and CRP activate transcription of CRP-S regulon genes: We have recently shown that the sites at which CRP activates competence genes are distinct from those controlling other functions, both in sequence and in dependence on Sxy (**#2**). These 'CRP-S' sites differ from the canonical CRP-N sites at positions T_6 and A_{17} (red in the figure) where CRP bends DNA (**#2**, **#3**, **#5,** [27]). In *E. coli*, replacing T_6 and A_{17} with C and G respectively causes an 80-fold reduction in CRP affinity [19], and our experiments show an even stronger effect in *H. influenzae* (Fig. 3 in **#5**).

Although Sxy stimulates CRP-dependent transcription at promoters with CRP-S sites by as much as 500-fold, we have almost no information about how it does this. Sxy is a small cytoplasmic protein with no detectable DNA-binding or dimerization motifs, and our *sxy*-knockout microarrays showed that it only acts at CRP-dependent promoters (**#2**). In *E. coli* a number of proteins are known to enhance transcription by physically contacting CRP, but these all differ from Sxy in having distinct DNA-binding sites adjacent to the CRP sites (Fig. 9 of **#3**). CRP-S sites are the only recognizable motifs in the CRP-S promoters of the γ-proteobacterial genomes we examined (**#3**), suggesting that Sxy action does not involve a separate DNA-binding site. An FNR knockout has normal competence so FNR does not act at CRP-S sites (**#2**).



The simplest hypothesis for Sxy's action is that it physically contacts CRP, enabling CRP to activate transcription at what are otherwise very unfavourable CRP sites. The locations of CRP-S sites relative to their transcription start sites predict that CRP acts at them by a Class I mechanism, making direct contact with RNA polymerase through its activating region 1 (AR1) as shown in the figure above. Sxy may enhance this interaction directly by making contacts with both CRP and RNA polymerase, or by modifying CRP's ability to bend DNA. Our bandshift experiments have not detected any binding of Sxy at CRP-S sites, with or without CRP. However these are often unable to detect multi-protein complexes

```
PurR consensus:  GCAAACGTTTGC
 PurR in rec-2:  GCAACCGTTTGC
```

5. PurR repression of *rec2* is released: Our first candidate for mediating the effect of nucleotides on competence was the PurR repressor, which blocks purine-biosynthesis genes in the presence of guanine or hypoxanthine; the PurR-regulated genes are induced when cells are transferred to MIV (**#2**). However knocking out the *purR* gene did not increase competence and did not eliminate the effect of purine nucleotides on induction. Microarray comparison of wildtype and *purR* mutant cells under conditions that do not normally induce competence (exponential growth in rich medium) confirmed that the knockout relieved repression of all of the PurR-regulon genes predicted from *E. coli* homologs and PurR sites, but did not increase expression of *sxy* or most CRP-S genes.

The promoter of the *rec2* gene has a good PurR site in addition to its CRP-S site, and we have recently confirmed that PurR does repress *rec2*. Rec2 is not needed for DNA binding or uptake, but in its absence DNA cannot cross the inner membrane, so *rec2* mutants are competent for uptake but not transformable [28]. Like other CRP-S genes, the *rec2* gene is partially induced when Sxy and cAMP levels rise in late log, despite repression by PurR (**#2**). From an evolutionary perspective, repression of *rec2* by PurR adds to the evidence that natural selection has favoured cells that induce DNA uptake when nucleotide pools are depleted.

The competent state: The end result of the regulatory events described above is expression of the genes of the CRP-S regulon. What do these genes do? Most are known to contribute to DNA uptake,

either at the inner membrane or as part of the type 4 pilus-related machinery that transports DNA across the outer membrane. CRP-S regulon genes with no known role in DNA uptake are all predicted to be cytoplasmic. Some of these limit the action of nucleases on unprotected DNA, and others are predicted to affect DNA metabolism at replication forks (discussed in **#2** and **#3**). Studies in *E. coli* have highlighted the importance of these genes for protection of stalled replication forks, and suggest that the CRP-S regulon unites genes whose functions alleviate problems arising from depleted nucleotide pools. Under this model, competence proteins scavenge extracellular DNA while cytoplasmic proteins protect and resolve stalled replication forks (**#3**). Genes for both these functions are under CRP-S regulation in all *Pasteurellaceae*, *Enterobacteriaceae* and *Vibrionaceae* genomes (**#3**).

Sxy homologs are also found in all of the sequenced *Pasteurellaceae*, *Enterobacteriaceae* and *Vibrionaceae* genomes. The *Actinobacillus pleuropneumoniae sxy* gene complements a *H. influenzae sxy* knockout (J. Bosse, personal communication), but nothing is known about its regulation. Nor is anything known about how *E. coli sxy* is regulated [29]. One of the two *V. cholerae sxy* orthologs is induced when cells are cultured in the presence of chitin; this could be either a direct effect of chitin or mediated by CRP (*V. cholerae* competence is repressed by glucose) [30].

In *E. coli* lacking other sources of carbon, uptake of DNA from themedium is sufficient to permit population growth [31]. As discussed above, we have long argued that cells take up DNA primarily as a source of nutrients, not genetic information ([32]*, **#1**). **The regulation of competence nucleotide starvation is critical evidence for this plausible hypothesis, which must now be backed up by a more detailed understanding of the molecular mechanisms.**

Specific Aims

We will answer the following questions:

I. How is sxy regulated in H. influenzae?

I-A: Does the kinetics of transcription regulate *sxy* expression?
I-B: Which base interactions are functionally important for *sxy* expression?
I-C: How do new *sxy* mutations affect competence?

II. How is sxy regulated in E. coli?

II-A: What induces *sxy* transcription?
II-B: Where does *sxy* transcription initiate?
II-C: Can *sxy* mutations increase sxy expression?
II-D: Is *sxy* translation regulated?

III. How does Sxy activate transcription in H. influenzae and E. coli?

III-A: What are Sxy's effects on transcription?
III-B: Do *E. coli* and *H. influenzae* Sxy proteins reciprocally complement?
III-C: Do CRP and Sxy physically contact each other *in vivo*?
III-D: Does Sxy replace CRP-RNAP interactions?
III-E: Can CRP or RNAP mutations bypass the need for Sxy?

Planned experiments

General *H. influenzae* methods, tools and resources: Standard methods for *H. influenzae* are described in Poje and Redfield [33,34]*. Briefly, cells are usually grown in brain-heart infusion broth with added hemin and NAD (sBHI); they double every 30' and reach a maximum density of 10^{10}cfu/ml. Transformation with antibiotic-resistance genes is the most sensitive assay for competence and for DNA uptake; it easily detects differences over 6 orders of magnitude and is the best initial screen for mutants.

We have a good collection of *H. influenzae* competence mutants (the relevant ones are listed in Table 1 of the Appendix). Additional mutants are easily constructed by chromosomal transformation as needed; the only limit is the need for about 1kp of sequence flanking any large insertion or deletion. To allow use of shorter sequences we are adapting for *H. influenzae* the recombineering system developed in *E. coli* [35,36].

I. How is sxy regulated in H. influenzae?

We hypothesize that depletion of nucleotide pools slows transcription and prevents formation of the sxy mRNA 2° structure that otherwise limits translation.

I-A: Does the kinetics of transcription regulate *sxy* expression?

I-A-1: Does slowing elongation by RNA polymerase *in vivo* increase *sxy* expression? Streptolydigin and rifampicin have been shown to inhibit the elongation steps of RNA polymerase, and we will use them to test whether slowing transcription enhances *sxy* translation [37-40]. The first test is whether partial inhibition of transcription by streptolydigin or rifampicin increases the competence of cells growing in rich medium + cAMP. If results are positive, effects on *sxy* mRNA and protein levels will also be examined, looking for a change in the ratio of protein to mRNA (as in Fig. 8 of **#4**). As these assays are very straightforward, we will test a wide range of antibiotic concentrations. This investigation meshes nicely with our collaboration with Julian Davies (UBC Dept. of Microbiology and Immunology) examining the genome-wide effects of sub-inhibitory concentrations of rifampicin.

In parallel, we will use a large-scale screen to determine whether *rif*^{R} or *std*^{R} mutations (in *rpoB*) alter the regulation of competence. We have already isolated several Rif^R mutants as part of our work with Julian Davies. Starting with wildtype *H. influenzae* cells, we will select cells that are resistant to different concentrations of antibiotic. Rather than isolating and characterizing individual mutants, we will pool all the colonies that grow up on the antibiotic plates and select hypercompetent ones by transforming the pooled cells with *nov*^{R} DNA while they are in log-phase growth in sBHI, as was done to isolate our present hypercompetence mutants. Because these methods use transformation assays they are very sensitive to modest effects on competence gene expression.

I-A-2: Do mutations that change purine pools affect *sxy* expression? The effect on intracellular nucleotide pools of supplementing MIV with nucleotides is difficult to predict because salvage pathways are complex and because *H. influenzae* can synthesize its own purines (and its pyrimidines if citrulline is provided). Intracellular pools can be more directly manipulated by mutations in nucleotide biosynthesis genes. A knockout of the purine repressor gene *purR* will cause constitutive synthesis of purines. This may not dramatically change purine nucleotide pools during growth in sBHI, but should reduce the depletion of pools caused by transfer to MIV; this mutation is predicted to reduce induction of all CRP-S genes (except perhaps *rec2*). In contrast, deletion of one of the purine biosynthetic genes will make purine pools entirely dependent on the medium, and is expected to increase *sxy* production in a defined medium with limiting purines [33]*. *H. influenzae* lacks a homologue of *E. coli*'s CytR repressor or other proteins that could sense the depletion of pyrimidine pools, but limiting the amount of citrulline in defined medium will test whether pyrimidine limitation also contributes to *sxy* regulation.

I-A-3: Does pausing by RNA polymerase contribute to regulation? Our hypothesis predicts that the kinetics of *sxy* transcript elongation will be sensitive to the state of nucleotide pools. Because transcription rates are heterogeneous both between and within single transcription events [41], such regulatory effects are most likely to be seen as sites where RNAP pauses. Pausing will be initially examined using *E. coli* RNA polymerase (Sigma) in an *in vitro* assay like that used by Donahue and Turnbough to detect pausing in the *E. coli pyrBI* leader [24]. End-labeling is accomplished by priming transcription with a ^{32}P-labelled oligo complementary to the *sxy* transcription start site. Limitation of

each nucleotide will be tested separately. Assays will also be done using partially purified *H. influenzae* RNA polymerase [42].

I-B. Which base interactions are functionally important for *sxy* expression?

I-B-1. What new mutations in *sxy* mRNA increase *sxy* expression? The regulatory importance of the basal stem in *sxy* mRNA was revealed by mutations that had been selected for derepressed competence ([9,20]*, **#4**). We will combine this selection for transformation under non-inducing conditions with saturation mutagenesis to identify all other positions in *sxy* where mutations increase expression, thus solidifying our understanding of the relevant pairing interactions.

Pools of mutations in *sxy* will be created using uniformly degenerate oligos and the Stratagene QuikChange PCR mutagenesis kit; we have had excellent experience with such kits and chose this one because mutagenesis needs only one mutagenic primer). We have arranged with Invitrogen for synthesis of sets of degenerate oligos; this experiment will use oligonucleotides covering the 5' 200nt of *sxy* mRNA. Each 25nt oligo will be prepared to uniform degeneracy, with a 3% probability of each incorrect base at each position. Mutagenesis using these oligos will produce a population of plasmids that on average will have about 2.5 substitutions in the mutagenized 25bp segment. Each plasmid population will be transformed into wildtype *H. influenzae* under conditions that give chromosomal recombination (replacement of the homologous sequence) and then selected for hypercompetence.

We will begin with a library of cells produced by saturation mutagenesis of the 5'portion of *sxy* mRNA, (bases -51 to +150). Hypercompetent mutants will be selected from the library by transformation under noninducing conditions (as done in [9]* and **#4**.); each colony that has become NovR will be tested for elevated competence during log phase growth in sBHI, and those passing this screen will have their *sxy* alleles sequenced for candidate hypercompetence mutations. Our present hypercompetent mutants were identified in preliminary screens of one pool of EMS-mutagenized cells (**#4**), and we predict that this targeted mutagenesis will produce many new mutations.

Our null hypothesis is that mutations that weaken any part of the candidate structure (Fig. 5A in **#4**) will cause hypercompetence. Mutations that change Stem 2 (site of the Shine-Dalgarno sequence and start codon) might be expected to have the strongest effects, but this short stem may interfere with expression less than the long basal Stem 1. We are especially interested in mutations that suggest differences from the candidate structure, and any that suggest a role for alternate pairing, either within the 110bp segment we have analyzed or distal to it. Because pairing in Stem 1 limits *sxy* expression, Stem 1 mutations that appear to increase base pairing would suggest existence of an alternate pairing structure.

I-B-2. What *sxy* mutations reduce *sxy* expression in hypercompetent mutants? Mutations that reduce *sxy* expression are intrinsically just as informative as those that increase it, but they will be harder to isolate because reduced competence cannot be selected for. We will target this analysis by mutagenizing cells carrying hypercompetence mutations, using the same degenerate oligos used in section I-A-1. Replica plating on novobiocin agar prespread with Nov^R DNA (no expression time needed) provides an efficient screen for reduced-competence mutants.

Together the locations and base pairing abilities of both hypercompetence and reduced-competence mutants will identify the relevant components of secondary structure. Regions where regulatory mutations do not arise will be equally informative, as they will identify residues that do not contribute to regulation.

I-C: How do new *sxy* mutations affect competence?

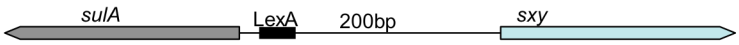
The mutations isolated in I-B could affect any or all of (i) pausing by RNAP, (ii) *sxy* mRNA 2° structure, and (iii) *sxy* mRNA translatability. After preliminary bioinformatic analysis of potential

pairing interactions involving mutant bases, we will directly create and test any mutations needed to confirm pairing relationships, as we have done for Stem 1 (**#4**). This will also be used to exclude effects due to coding changes.

Changes to pausing will be investigated in selected mutants using the *in vitro* transcription assay described in I-A-C. Changes to 2° structure will also be confirmed using the RNase sensitivity assays described in **#4**. The assays of *sxy* mRNA and Sxy protein levels shown in Fig. 8 of **#4** demonstrated unambiguously that mRNA 2° structure limits translation, and they will be used to evaluate the consequences of changed pause sites or 2° structure. The end product will be a refined model of *sxy* regulation, with the roles of base pairing, RNAP pausing, and translation initiation explicitly specified.

II. How is sxy regulated in E. coli?

II-A: What induces *sxy* transcription? Although we have shown that the *E. coli ppdD* gene can be induced by *sxy* overexpression (**#3**), an exhaustive search by Sauvonnet *et al*. did not identify any conditions that induced *ppdD* [29]. Our starting point will be the hypothesis that *sxy* is similarly regulated in *E. coli* and *H. influenzae*. Because *E. coli* has maintained a fully functional sxy gene and CRP-S regulon, we are confident that these genes are expressed under some conditions. This is supported by analysis of *E. coli* microarray data in the NCBI-GEO database, which shows *sxy* expression to be elevated under conditions of nutrient depletion (Fig. 3). Accordingly, we will first test abrupt transfer from rich medium to minimal salts, and from aerobic to anaerobic conditions, using both wildtype cells and purine auxotrophs. Fusions of *lacZ* to both *sxy* and *ppdD* will be used as reporters; the latter gene's 150-fold induction by Sxy may amplify sensitivity (see Fig. 7 in **#3**). Although the *sxy* promoter region has no recognizable sites for either activators or repressors, knockouts of possible repressors (*e.g.* Ftagged, CytR) will be tested. Once inducing conditions have been identified, knockouts of possible activators (*e.g.* CRP, FNR, Fis) will also be tested.

II-B: Where does *sxy* transcription initiate? 

Before the *sxy* promoter can be characterized the 5' end of *E. coli sxy* mRNA must be mapped by run-off transcription [43]. The *sxy* gene is divergently transcribed from the LexA-repressed *sulA* gene. Data in the NCBI-GEO array database shows that *sxy* is not regulated by LexA, but we do not know where transcription initiates in the adjacent 200bp.

II-C: Can *sxy* mutations increase *sxy* expression? Once the transcription start site has been identified, we will screen for *sxy* mutations that increase *sxy* expression, using the targeted mutagenesis procedure described above (Section I-B-1). If *E. coli sxy* expression is regulated like *H. influenzae sxy* expression, such mutations will be in positions where they alter an expression-limiting mRNA 2° structure. The assay will use the same *lacZ* fusion used for the induction studies in Section II-A, with screening by plating on minimal glycerol agar with X-Gal. If this search is unsuccessful we will also screen EMS-mutagenized cells for such mutations; this will identify any trans-regulators.

II-D: Is *sxy* translation regulated? We will raise polyclonal antibodies against *E. coli* Sxy as we did for *H. influenzae* Sxy (we have had bad experiences with commercial anti-peptide antibodies) and use these to directly measure *in vivo* protein levels. These can then be compared with mRNA levels measured by real-time PCR to give accurate measures of translatability, as described in **#4**. If inducing conditions identified in II-A implicate nucleotide pools, the effect of these on translatability will also be investigated; this will be much easier in *E. coli* than in *H. influenzae*, especially because all mutants are freely available from Genobase (http://ecoli.aist-nara.ac.jp).

Finding the conditions that induce the *E. coli* CRP-S regulon will allow us to determine whether these genes cause significant DNA uptake and possibly natural transformation, overturning decades of (unpublished) negative results. Although Steve Finkel's group showed some of these genes enable *E.*

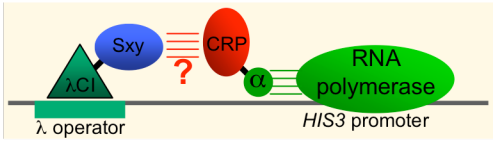
coli to use DNA as a nutrient [31], they were unable to directly detect DNA uptake or transformation; they do not plan to work on *sxy* regulation (personal communication).

III. How does Sxy activate transcription in H. influenzae and E. coli?

III-A: What are Sxy's effects on transcription? Before we can investigate the mechanism of Sxy's effects on transcription, we need to clarify what these effects are. Although we refer to CRP promoters that do not fit the CRP-S consensus as CRP-N promoters, these are diverse and we have not yet ruled out weak effects of Sxy at some of them. Furthermore, we have only examined Sxy-dependence in cells induced by transfer to MIV (**#2**), and the effects we saw may have been confounded by additional responses to this complex nutritional shift. CRP-dependent promoters are even more variable in *E. coli* and we know much less about how Sxy affects them; we have tested only 4 of the >100 candidates (**#3**).

We will now do a thorough microarray analysis in both *H. influenzae* and *E. coli*, examining expression from all promoters in the presence and absence of Sxy, and under conditions where CRP is either fully induced or only partially induced. These experiments will be done in rich medium (sBHI and LB) to eliminate confounding effects of MIV. CRP activity will be manipulated by using cells with *cya* knockouts, with cAMP provided at limiting or fully inducing concentration. Both the *H. influenzae* and *E. coli* analyses will use cells carrying *sxy* under CRP-independent promoters (*sodC* [44] and *lacUV5* respectively). These analyses will using *H. influenzae* microarrays from the Bacterial Microarray Group at the University of London and *E. coli* microarrays from the Microarray and Proteomics Facility at the University of Alberta.

III-B: Do *E. coli* and *H. influenzae* Sxy proteins reciprocally complement? The functional similarity of the *E. coli* and *H. influenzae* Sxy proteins was confirmed in **#3** by our demonstration that expression of *E. coli* CRP-S promoters is Sxy-dependent. The above microarray analysis will reveal whether their regulatory targets are the same. However the two Sxy proteins are less well conserved than CRP (30% identity, 50% similarity), and they may exert their similar effects in somewhat different ways. This will be tested by reciprocal complementation experiments, using chromosomal *sxy* knockouts (we have both) and the sxy-expression plasmids used in III-A. Complementation will be assayed in *H. influenzae* by transformability, and in *E. coli* by expression of the *ppdD::lacZ* reporter used in II-A. If complementation is poor or negligible we will test the effect of providing the homologous CRP. An increase in complementation would strongly suggest that CRP and Sxy physically interact, and we may then be able to use mutant or hybrid CRPs to pin down the sites of interaction.



III-C: Do CRP and Sxy physically contact each other *in vivo*? The simplest model for how Sxy activates CRP-S promoters is that Sxy makes physical contact with CRP, increasing its ability to kink the CRP-S site or interact with RNAP (see Discussion of **#5**). Two-hybrid assays in *E. coli* and yeast will be used to detect contact between CRP and Sxy. We will begin with assays using *E. coli* Sxy fused to the λ CI protein and CRP fused to the α-NTD of RNAP (Stratagene Biomatch system). Because evaluating transcription factors in a test system that depends on transcription can give false-positive results, the analysis will also be done in yeast.

Co-precipitation analysis will also be used to identify *in vivo* interactions between CRP, Sxy, and promoter DNA. We have His-tagged *crp* and *sxy* clones for both *H. influenzae* and *E. coli*, and have confirmed that the tagged proteins stimulate CRP-S promoters (*H. influenzae* Sxy has not yet been tested). Several E. coli CRP-S genes (*ppdD, comA, comM*) have also been cloned with His-tags and these will serve as negative controls for interactions detected using His-CRP and His-Sxy.

His-CRP from both species will be used to precipitate promoter DNA and associated proteins, using the method of Grainger *et al.* [18]. Cells will be lysed and His-CRP will be immobilized on Ni-agarose along

with all DNA and proteins to which it is bound. PCR will be used to test for promoter DNA that has been precipitated by CRP, while our anti-Sxy antibodies will be used to test for the presence of Sxy in promoter complexes. Formaldehyde fixation of cells prior to lysis can be used to improve stability and isolation of protein-DNA complexes; its reversibility will facilitate downstream experiments such as PCR [45].

To test whether Sxy interacts with CRP and/or DNA, His-Sxy will be induced in both *crp*+ and *crp*- cells. After isolation of His-Sxy, PCR will again be used to detect co-precipitated promoter DNA. Commercially available anti-CRP will can be used to test for co-precipitation of CRP by His-Sxy, and PCR allows us to quickly test for precipitation of any promoter. We have identified sequences of interest adjacent to CRP-S sites (see Fig. 6 in **#5**). Their importance will be examined by deletion and site-directed mutagenesis experiments.

III-D: Does Sxy replace CRP-RNAP interactions? In *E. coli* CRP, any of several mutations in activating region 1 (AR1) eliminate essential CRP-RNAP interactions and prevent transcription (ref?). We will test whether expression of Sxy restores transcription to these mutants by eliminating the need to direct CRP-RNAP contacts. Because the *E. coli* and *H. influenzae* CRP proteins are so similar, the same mutations will also be created in *H. influenzae* CRP and similarly tested. We will also test CRP-S activation by a *crp* mutation we have obtained from Helen Berman which changes CRP specificity so it preferentially binds to CRP-S sites over CRP-N sites [19]. However, because the mutation dramatically reduces CRP's affinity for all CRP sites the activation may be too weak to detect.

III-E: Can *crp* or RNAP mutations bypass the need for Sxy at CRP-S sites? Although such gain-of-function mutations may not be possible, we are proposing to search for them because selection in *H. influenzae* is straightforward and very powerful, and because these mutations would be very informative. The basic strategy will be to select for restored transformation in a *H. influenzae sxy* knockout strain, by incubating MIV-induced cells with *nov*^{R} DNA and selecting for Nov^R cells. Using a *crp* knockout complemented with *E. coli* CRP will allow any mutations we find to be directly mapped onto CRP's functional domains (for example mutations in the AR1 domain of CRP would imply that Sxy enhances contacts with RNA polymerase). Furthermore, we have shown that *E. coli* CRP has a higher affinity for CRP-S sites than *H. influenzae* CRP, which means the desired phenotype may be easier to obtain (**#3**). To increase the frequency of mutations in CRP or RNAP, we will begin by mutagenizing cells with pools of mutant PCR fragments created using the same kind of degenerate oligos described in Section I-B.

Significance:

We are the experts in both *H. influenzae* competence and the novel transcription factor Sxy (most of the recent papers we have cited are our own, only because few others exist). We are poised to make substantial advances, as we have both the expertise and the genetic and molecular tools to build on our discovery of the CRP-S regulon. The answers we get will apply to most of the γ-proteobacteria, a group that includes many important human pathogens in addition to *H. influenzae* and *E. coli*. Despite the importance of the questions we raise above, no other researchers are studying them.

Work in our lab addresses all aspects of competence; its regulation, its mechanisms, and its evolutionary significance. Our goal is to understand competence regulation in its full biological context. Because the study organisms are human commensals and pathogens, this context is the human body. However our work is not 'systems biology' or '-omics'. Rather we are doing the fundamental research on which these surveys rely. Our benchwork is where the rubber meets the road.

Papers provided with this proposal:

#1. Redfield, R. J. 2001. Do bacteria have sex? Nature Reviews Genetics **2:** 634-639.

#2. Redfield, R. J., Cameron A. D, Qian, Q., Hinds, J., Ali, T. R., Kroll, J. S. and Langford, P. 2004. A novel CRP-dependent regulon controls expression of competence genes in *Haemophilus influenzae*. 2005. Journal of Molecular Biology. **347:** 735-747.

#3. Cameron, A. D. S., and Redfield, R. J. 2006. Non-canonical CRP sites control competence regulons in *Escherichia coli* and many other gamma-proteobacteria. Nucleic Acids Research **34:** 6001-6014.

#4. Cameron, A. D. S., Bannister, L. A., Volar, M. and Redfield R. J. 2007. RNA secondary structure regulates sxy expression and competence development in *Haemophilus influenzae*. (Under review at Molecular Microbiology)

#5. Cameron, A. D. S., and Redfield R. J. 2007. Sxy enhances CRP binding and transcription activation at CRP-S sites. (Submitted to J. Bacteriol.)

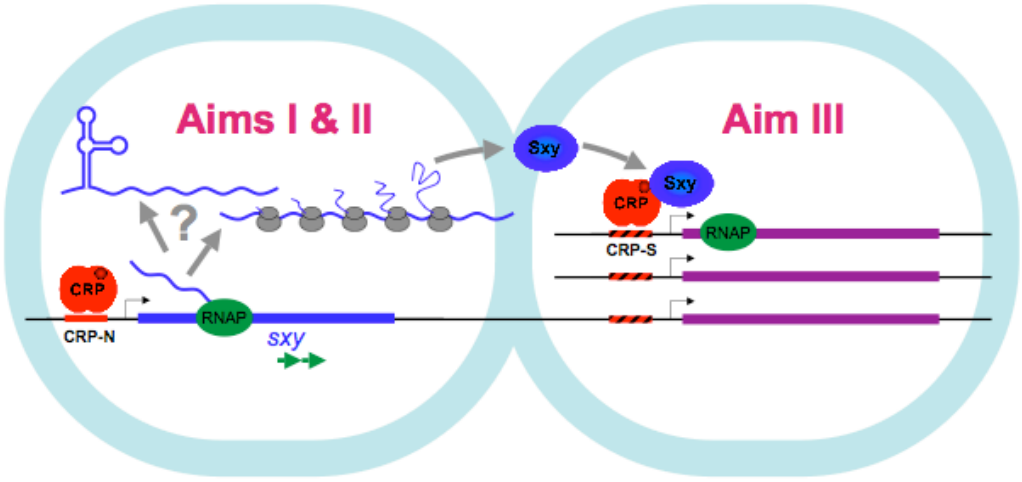
References

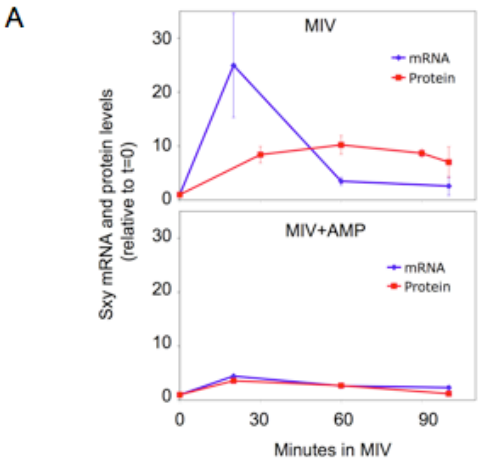
1. Saha, S. K. et al. Invasive Haemophilus influenzae type B diseases in Bangladesh, with increased resistance to antibiotics. *J Pediatr* **146**, 227-233 (2005).
2. Heath, P. T. Haemophilus influenzae type b conjugate vaccines: a review of efficacy data. *Pediatr. Infect. Dis. J.* **17**, S117-22 (1998).
3. Dworkin, M. S., Park, L. & Borchardt, S. M. The changing epidemiology of invasive Haemophilus influenzae disease, especially in persons > or = 65 years old. *Clin Infect Dis* **44**, 810-816 (2007).
4. Chen, H. D. & Frankel, G. Enteropathogenic Escherichia coli: unravelling pathogenesis. *FEMS Microbiol Rev* **29**, 83-98 (2005).
5. Kaper, J. B., Nataro, J. P. & Mobley, H. L. Pathogenic Escherichia coli. *Nat Rev Microbiol* **2**, 123-140 (2004).
6. Solomon, J. M. & Grossman, A. D. Who's competent and when: regulation of natural genetic competence in bacteria. *Trends Genet.* **12**, 150-155 (1996).
7. Chen, I. & Dubnau, D. DNA uptake during bacterial transformation. *Nat Rev Microbiol* **2**, 241-249 (2004).
8. Lethem, M. I., James, S. L., Marriott, C. & Burke, J. F. The origin of DNA associated with mucus glycoproteins in cystic fibrosis sputum. *Eur Respir J* **3**, 19-23 (1990).
9. Redfield, R. J. sxy-1, a Haemophilus influenzae mutation causing greatly enhanced spontaneous competence. *J. Bacteriol.* **173**, 5612-5618 (1991).
10. Danner, D. B., Smith, H. O. & Narang, S. A. Construction of DNA recognition sites active in Haemophilus transformation. *Proc Natl Acad Sci U S A* **79**, 2393-2397 (1982).
11. Redfield, R. J. et al. Evolution of competence and DNA uptake specificity in the Pasteurellaceae. *BMC Evol Biol* **6**, 82 (2006).

12. Wilson, L. A. & Redfield, R. J. Defining the Haemophilus influenzae uptake signal sequence. *Genome Res.* submitted
13. Goodgal, S. H. DNA uptake in Haemophilus transformation. *Annu. Rev. Genet.* **16**, 169-192 (1982).
14. Barany, F., Kahn, M. E. & Smith, H. O. Directional transport and integration of donor DNA in Haemophilus influenzae transformation. *Proc Natl Acad Sci U S A* **80**, 7274-7278 (1983).
15. Stuy, J. H. & Walter, R. B. Addition, deletion, and substitution of long nonhomologous deoxyribonucleic acid segments by genetic transformation of Haemophilus influenzae. *J. Bacteriol.* **148**, 565-571 (1981).
16. Macfadyen, L. P., Dorocicz, I. R., Reizer, J., Saier, M. H. J. & Redfield, R. J. Regulation of competence development and sugar utilization in Haemophilus influenzae Rd by a phosphoenolpyruvate:fructose phosphotransferase system. *Mol. Microbiol.* **21**, 941-952 (1996).
17. Barabote, R. D. & Saier, M. H. J. Comparative genomic analyses of the bacterial phosphotransferase system. *Microbiol Mol Biol Rev* **69**, 608-634 (2005).
18. Grainger, D. C., Hurd, D., Harrison, M., Holdstock, J. & Busby, S. J. Studies of the distribution of Escherichia coli cAMP-receptor protein and RNA polymerase along the E. coli chromosome. *Proc Natl Acad Sci U S A* **102**, 17693-17698 (2005).
19. Chen, S. et al. 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. (2001).
20. Williams, P. M., Bannister, L. A. & Redfield, R. J. The Haemophilus influenzae sxy-1 mutation is in a newly identified gene essential for competence. *J. Bacteriol.* **176**, 6789-694. (1994).
21. MacFadyen, L. P. et al. Competence development by Haemophilus influenzae is regulated by the availability of nucleic acid precursors. *Mol. Microbiol.* **40**, 700-77. (2001).
22. de Smit, M. H. & van Duin, J. Translational standby sites: how ribosomes may deal with the rapid folding kinetics of mRNA. *J. Mol. Biol.* **331**, 737-743 (2003).
23. Uptain, S. M., Kane, C. M. & Chamberlin, M. J. Basic mechanisms of transcript elongation and its regulation. *Annu. Rev. Biochem.* **66**, 117-172 (1997).
24. Donahue, J. P. & Turnbough, C. L. J. Nucleotide-specific transcriptional pausing in the pyrBI leader region of Escherichia coli K-12. *J. Biol. Chem.* **269**, 18185-18191 (1994).
25. Yanofsky, C. Transcription attenuation: once viewed as a novel regulatory strategy. *J. Bacteriol.* **182**, 1-8 (2000).
26. Mandal, M. & Breaker, R. R. Gene regulation by riboswitches. *Nat Rev Mol Cell Biol* **5**, 451-463 (2004).
27. Schultz, S. C., Shields, G. C. & Steitz, T. A. Crystal structure of a CAP-DNA complex: the DNA is bent by 90 degrees. *Science* **253**, 1001-107. (1991).
28. McCarthy, D. Cloning of the rec-2 locus of Haemophilus influenzae. *Gene* **75**, 135-143 (1989).
29. Sauvonnet, N., Gounon, P. & Pugsley, A. P. PpdD type IV pilin of Escherichia coli K-12 can Be assembled into pili in Pseudomonas aeruginosa. *J. Bacteriol.* **182**, 848-854 (2000).
30. Meibom, K. L., Blokesch, M., Dolganov, N. A., Wu, C. Y. & Schoolnik, G. K. Chitin induces natural competence in Vibrio cholerae. *Science* **310**, 1824-1827 (2005).
31. Palchevskiy, V. & Finkel, S. E. Escherichia coli competence gene homologs are essential for competitive fitness and the use of DNA as a nutrient. *J. Bacteriol.* **188**, 3902-3910 (2006).

32. Redfield, R. J. Genes for breakfast: the have-your-cake-and-eat-it-too of bacterial transformation. *J Hered* **84**, 400-404 (1993).
33. Poje, G. & Redfield, R. J. General methods for culturing Haemophilus influenzae. *Methods Mol Med* **71**, 51-56 (2003).
34. Poje, G. & Redfield, R. J. Transformation of Haemophilus influenzae. *Methods Mol Med* **71**, 57-70 (2003).
35. Yu, D. et al. An efficient recombination system for chromosome engineering in Escherichia coli. *Proc Natl Acad Sci U S A* **97**, 5978-5983 (2000).
36. Datsenko, K. A. & Wanner, B. L. One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proc Natl Acad Sci U S A* **97**, 6640-6645 (2000).
37. Kyzer, S., Zhang, J. & Landick, R. Inhibition of RNA polymerase by streptolydigin: no cycling allowed. *Cell* **122**, 494-496 (2005).
38. Stuy, J. H. Transfer of genetic information within a colony of Haemophilus influenzae. *J. Bacteriol.* **162**, 1-4 (1985).
39. Walter, R. B. & Stuy, J. H. Isolation and characterization of a UV-sensitive mutator (mutB1) mutant of Haemophilus influenzae. *J. Bacteriol.* **170**, 2537-2542 (1988).
40. Jin, D. J. & Gross, C. A. RpoB8, a rifampicin-resistant termination-proficient RNA polymerase, has an increased Km for purine nucleotides during transcription elongation. *J. Biol. Chem.* **266**, 14478-14485 (1991).
41. Davenport, R. J., Wuite, G. J., Landick, R. & Bustamante, C. Single-molecule study of transcriptional pausing and arrest by E. coli RNA polymerase. *Science* **287**, 2497-2500 (2000).
42. Wang, D. et al. Discontinuous movements of DNA and RNA in RNA polymerase accompany formation of a paused transcription complex. *Cell* **81**, 341-350 (1995).
43. Zulty, J. J. & Barcak, G. J. Identification of a DNA transformation gene required for com101A+ expression and supertransformer phenotype in Haemophilus influenzae. *Proc Natl Acad Sci U S A* **92**, 3616-3620 (1995).
44. Bosse, J. T., Nash, J. H., Kroll, J. S. & Langford, P. R. Harnessing natural transformation in Actinobacillus pleuropneumoniae: a simple method for allelic replacements. *FEMS Microbiol Lett* **233**, 277-281 (2004).
45. Pugh, B. F. & Gilmour, D. S. Genome-wide analysis of protein-DNA interactions in living cells. *Genome Biol* **2**, REVIEWS1013 (2001).

Figure 1





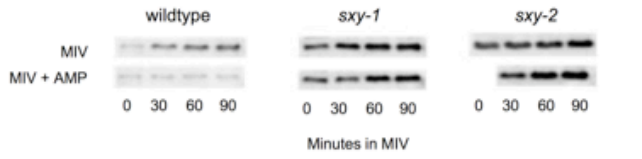


Figure 2. Transcription and translation of sxy in MIV ± AMP. A. mRNA and protein levels in wildtype cells. **B.** Protein levels in wildtype and hypercompetence mutants.

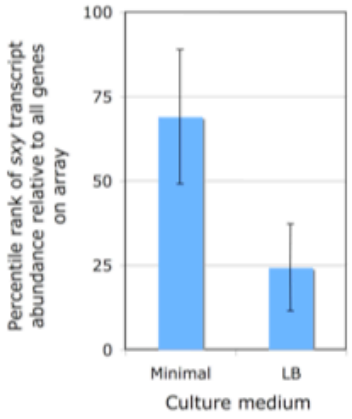


Figure 3. Average and standard deviation of *sxy* transcript levels in *E. coli* cultured in minimal (M9 or Davis) or Lauria-Bertani (LB) media. Number of samples: Minimal medium (24), LB medium (36).