

Research Plan

A. Specific Aims

B. Backgroundp. 19
C. Preliminary Studiesp. 28
D. Research Design and Methods ..p. 32

Our broad goal is to understand the evolutionary causes of genetic exchange in microorganisms. Central to it is this project's goal: understanding the evolution of the uptake signal sequences (USS) of naturally competent bacterial pathogens. Unlike most other competent bacteria, *Haemophilus* and *Neisseria* preferentially take up homologous DNA, by recognizing short highly-repeated sequences, the USS. USS-dependent uptake has been best characterized in *Haemophilus influenzae*, where the complete genome sequence has revealed 1465 evenly distributed copies of a 9 bp USS, a substantial fraction of an otherwise streamlined genome. Furthermore, most USS are in coding regions and are likely to constrain the sequences of the corresponding proteins.

Little attention has been given to the evolutionary function of USS. Because competence has been assumed to have evolved for genetic exchange, USS and their uptake system have been viewed as a mechanism to limit or prevent uptake of unrelated DNAs, thus increasing the efficiency or safety of genetic exchange. However, this hypothesis presents several underappreciated difficulties. The USS may instead have an immediate function in the genome, independent of transformation, but this fails to explain the bias of the DNA uptake mechanism.

Importance of USS: 1. USS-dependent transformation is the primary pathway of genetic exchange in these bacteria, and has transferred virulence genes from one pathogen to another. Thus they are extremely relevant to the evolution of infectious diseases. 2. Viewed from a broader perspective, USS-dependent DNA uptake is the only strong evidence that mechanisms for exchange of chromosomal genes have been selected in bacteria. If USS are found to have another function, the existing paradigms will need drastic revision. 3. Once we understand their function, USS may also provide targets for intervention in infections by naturally-transformable organisms—either as a drug-delivery mechanism or as a part of a process to be inhibited by new drugs.

The problem of USS evolution is wide open; we know of no other laboratory working on it. We have chosen the following questions as most worth investigating.

Modeling:

	Page
I. Nucleotide acquisition is an inevitable benefit of DNA uptake. Is it consistently much larger than the other potential benefits?	32
II. A biased DNA-uptake receptor can increase the frequency of its preferred sequence in the genome. Will receptor bias inevitably do so if cells take up homologous DNA?	33
III. The <i>H. influenzae</i> USS are not randomly distributed around the genome. Can the even spacing be explained by receptor-bias-driven accumulation?.....	33

Genomic analysis:

IV. The <i>H. influenzae</i> USS are not randomly spaced around the genome. Are the same patterns seen in other genomes with USS?	34
V. Most <i>H. influenzae</i> USS are in coding sequences. How do they constrain protein function?	34
VI. Some related species with overlapping environments share a common USS. Has this led to substantial genetic exchange between them?	35

Experimental work:

VII. Uptake of heterologous DNA can kill the Rd strain of <i>H. influenzae</i> by inducing a resident prophage. Does this lethality occur in other strains?	35
VIII. Respiratory tract mucus contains high concentrations of DNA. Can DNA be taken up by <i>H. influenzae</i> growing in or under mucus?	36
IX. The strength of the <i>H. influenzae</i> uptake bias is poorly defined. How strongly does DNA uptake depend on the USS?	36
X. The USS-binding structure on the surface of competent <i>H. influenzae</i> cells has not been identified. Can the USS be cross-linked to a protein on the surface of competent cells?	37
XI. USS abundance may be primarily due to an intracellular function. Is there a cytoplasmic or inner membrane protein that specifically binds the <i>H. influenzae</i> USS?	38

B. Background and Significance	Page
Natural competence, DNA uptake and transformation	19
What is competence?	19
Phylogenetic distribution of naturally competent bacteria.	19
Recombination and transformation	20
Competence in natural environments	20
Consequences of competence for population structure	21
Uptake specificity	21
Genome analysis	21
Evolution of genetic exchange	22
The problem of the evolution of sexual reproduction	22
Recombination in bacteria	22
Evolution of natural competence	23
Possible functions of competence	23
If transformation is an accident, why have USS?	26
Possible resolution	26
Can USS-based uptake systems have another function?	27
Significance of the USS for pathogenicity	27

Natural competence, DNA uptake and transformation

What is competence?

Competence is the ability to take up DNA. In many bacteria it develops under natural conditions, controlled genetically programmed developmental pathways (reviewed by Goodgal 1982; Biswas et al. 1989; Dubnau 1991; Lorenz et al. 1994). This natural competence differs from the artificial cell-permeabilization procedures used to introduce plasmids into laboratory cultures. Most naturally competent bacteria become competent only in response to certain environmental conditions, and can then efficiently bind double-stranded DNA fragments and take one or both strands into the cell. If the sequence of an incoming DNA strand is sufficiently similar to a sequence in the chromosome, the former may displace the latter by homologous recombination, giving the cell a new (transformed) genotype; otherwise the DNA is usually rapidly degraded in the cytoplasm.

Phylogenetic distribution of naturally competent bacteria

Natural competence is not confined to particular lineages on bacterial phylogenetic trees, but is sporadically distributed on branches that also have many non-transformable groups. This sporadic distribution may result from multiple independent inventions of the ability to take up DNA, and/or to frequent loss of that ability.

Natural competence has been well characterized in only a few groups, primarily the gram-negative *H. influenzae* and *Neisseria gonorrhoeae*, and the gram-positive *Bacillus subtilis* and *Streptococcus pneumoniae*. Below is a brief overview of the basic steps in competence in *H. influenzae*. This is followed by more detailed discussion of the nature of the USS and its role in DNA uptake, and by consideration of the evolution of bacterial genetic exchange mechanisms in general, and of the evolution of the USS and its uptake system.

Induction of competence:

H. influenzae cells are not competent in exponential growth. About 1% of the cells in a culture or colony develop competence at the onset of stationary phase, and all become competent if an exponentially-growing culture is transferred to a starvation medium. Competence absolutely requires the transcription factor CRP(CAP) and its cofactor cyclic AMP (Chandler 1992; Dorocicz et al. 1993). The *sxy* gene product is also required; this may be another transcription factor, regulating the genes responsible for DNA uptake (Williams et al. 1994; Karudapuram et al. 1997).

Figure 1 here

Competence is differently regulated in other bacteria. Laboratory cultures of *N. gonorrhoeae* are competent at all stages. Both *B. subtilis* and *S. pneumoniae* use secreted factors to induce competence; *B. subtilis* also uses a complex network of nutritional and sporulation signals (Dubnau 1991). The regulation of competence has implications for its evolutionary function, and will be discussed in more detail below.

DNA binding: Bacterial cells only bind DNA when they are competent to take it up. *H. influenzae* cells preferentially bind DNA fragments containing the 9-bp uptake signal sequence AAGTGCGGT (discussed in a separate section below). The number of binding sites on each cell has not been resolved; one report estimated that about 4 large fragments, and up to 40 small fragments can be bound (Barouki et al. 1986). In lab cultures DNA uptake saturates at about 200ng of DNA per 10⁹ cells (about 200kb/cell), which may indicate that each site can bring in only one DNA fragment. Single-stranded DNA is bound and taken up only under low-pH conditions unlikely to be relevant in the natural environment (Postel et al. 1966). *N. gonorrhoeae* and *N. meningitidis* also have sequence-specific DNA uptake (Jyssum et al. 1970; Graves et al. 1982); its USS is the 10bp sequence GCCGTCTGAA (Goodman et al. 1988). In contrast, DNA binding and uptake by naturally competent Gram positive bacteria such as *B. subtilis* and *S. pneumoniae* is indiscriminate (no USS recognition).

DNA uptake: In *H. influenzae*, DNA uptake proceeds by two stages, usually called 'uptake' and 'translocation', and possibly corresponding to transport across the outer and inner cell membranes (Goodgal 1982). The DNA is first taken up into a structurally-undefined compartment where it is inaccessible to both external nucleases and intracellular restriction enzymes. This compartment was initially thought to correspond to vesicular structures seen on the surfaces of competent cells ('transformasomes') (Kahn et al. 1983), but may simply be the periplasm. Both circular and linear molecules can be taken into this compartment, and no size limits have been identified. Genes specific to the DNA binding and uptake steps have not been identified.

DNA translocation: After the first 'uptake' step the DNA is translocated into the cytoplasm. A free end is required (circular molecules remain in the unidentified compartment) (Barany et al. 1983). The time required for translocation is not well defined; in most studies the entire process was complete within a few minutes, but in some reports intact DNA was recovered after 20 minutes or more. Two genes are known to be required for translocation; *rec-2* and *com101*, and a third (*dprA*) may also be involved (Barouki et al. 1985; Larson et al. 1991; Karudapuram et al. 1997).

Naturally competent Gram positive bacteria degrade one DNA strand at the cell surface, and the other is translocated into the cytoplasm. This process may be homologous to the translocation step of gram-negative bacteria; some homologous genes are involved (Dreiseikelmann 1994). The *B. subtilis* competence gene *comE1* has a *H. influenzae* homolog, but its role in *H. influenzae* competence has not been investigated.

Recombination and transformation:

Once one end of the DNA fragment enters the cytoplasm it begins to undergo exonucleolytic degradation, rapidly from the incoming 5' end and more slowly from the incoming 3' end (Pifer et al. 1985). The degradation of the incoming 5' end may be concomitant with its translocation into the cytoplasm. The strand with the 3'-leading end may undergo recombination with a homologous sequence in the chromosome (mediated by the Rec-1 protein, a RecA homolog). However, even when incoming DNA is perfectly homologous, only 10-20% of it recombines and escapes degradation (Barany et al. 1983). In cells made competent by the MIV-starvation procedure, the nucleotides released by this degradation are rapidly reused for new DNA synthesis. The *dprA* gene product is involved in DNA processing (Karudapuram et al. 1997), but the RecBCD complex plays no role (Wilcox et al. 1975, R. Myers and R. Redfield unpublished). Even large insertions and deletions can be readily recombined if they are flanked by at least 1kb of homology (Stuy et al. 1981).

Are pathogens competent in their natural environments?

Transformation has been observed in soil and aquatic environments, but the only direct demonstration of transformation of a pathogen was Griffith's transformation of *S. pneumoniae* (Griffith 1928; Lorenz et al. 1994). Laboratory co-cultivation experiments have demonstrated genetic exchange in mixed populations of *Neisseria* and of *Haemophilus* (Stuy 1985; Frosch et al. 1992). Sequence comparisons of genes associated with pathogenesis clearly demonstrate that genetic exchange has occurred, and has, for example, been responsible for the rise of penicillin resistance in *N. gonorrhoeae* and *N. meningitidis* (Spratt et al. 1992), and for changes in capsulation type in *H. influenzae* (Kroll 1992). *N. meningitidis* has recently been discovered to have acquired several genes from a *Haemophilus* species, apparently by transformation (Kroll et al. 1998). It should also be noted that, typically, only some isolates of a 'competent' species are actually found to be transformable.

Consequences of competence for population structure:

For *H. influenzae* there is no evidence that the genetic exchange permitted by transformation actually randomises gene combinations between different lineages. Rather, *H. influenzae* isolates show a clonal population structure (linkage disequilibrium) (Musser et al. 1988), possibly suggesting that most of the available *H. influenzae* DNA comes from sibling cells in clonally-infected hosts. However, the genetic structures of some *Neisseria* and *Bacillus* species suggest that exchange of chromosomal genes is frequent enough to disrupt linkage disequilibrium, and *H. pylori* has recently been reported to show free recombination (Caugant et al. 1987; Maynard Smith et al. 1993; Cohan 1996; Suerbaum et al. 1998). In all of these species the frequency of genetic exchange is far lower than in sexual eukaryotes.

Uptake specificity

Discovery of uptake specificity:

Although competent cells of gram-positive bacteria will bind and take up all double-stranded DNAs equally well, competent *H. influenzae* and *N. gonorrhoeae* cells efficiently bind only DNAs from their own or a closely related species. The original evidence for specific uptake came from uptake-competition experiments, where addition of unrelated DNAs did not reduce uptake of homologous DNA by either *H. influenzae* or *H. parainfluenzae* competent cells. (Goodgal 1982; Graves et al. 1982).

H. influenzae DNA is highly enriched for a short sequence preferentially bound by competent cells. This sequence (the USS) was initially identified as the 11-mer AAGTGCGGTCA present on four cloned DNA fragments taken up by competent cells, and its role in DNA binding was confirmed by footprinting experiments (Danner et al. 1980). Additional uptake and DNA consensus analysis identified the 9 bp sequence AAGTGCGGT as the core USS, and implicated AT-rich flanking sequences in binding (Danner et al. 1982; Goodgal et al. 1990). The importance of flanking sequences was confirmed by testing of synthetic USS with defined flanking sequences, and by the demonstration that ethylation of both the core and flanking sequences would prevent uptake (Danner et al. 1980).

Shortly thereafter, Goodman and Scocca used cloned fragments to identify the *N. gonorrhoeae* USS as a 10-mer with the sequence GCCGTCTGAA (Goodman et al. 1988). The first isolates of this sequence occurred in inverted-repeat pairs downstream of open reading frames, where they were thought to function as rho-independent transcriptional terminators. Some *H. influenzae* SS are also found in this configuration (see below).

Biased uptake in organisms not known to have USS:

Campylobacter jejuni is naturally competent and preferentially takes up its own DNA. However, fragments cloned into *E. coli* are not preferentially taken up (D. Taylor, pers. comm), suggesting that the uptake may depend on a species-specific DNA modification rather than a species-specific sequence. Several other naturally competent bacteria are known to preferentially take up conspecific DNA, but no USS has been identified; these include *Helicobacter pylori*, *Pseudomonas stutzeri* and *Azotobacter vinlandii* (Lorenz et al. 1994). In others the specificity of uptake has not been investigated.

Genome sequence analysis

The Rd strain of *H. influenzae* is the only completely sequenced genome with USS (Fleischmann et al. 1995). Detailed analysis of these have been reported by Smith et al. (1995) and Karlin et al. (1996), provided as Appendix items B and C. Smith et al. (B, Fig. 2) found 1465 perfect matches to the previously-identified core sequence, and an additional 764 matches differing at single positions from the 9bp core sequence. Only 8 perfect copies would be expected if the core USS sequence occurred randomly in a genome of this base composition (38% G+C). Alignments of both perfect and imperfect USS revealed strong flanking-sequence consensus on the 3' side of the + strand, which agreed well with the positions previously identified by ethylation footprinting (Smith et al. Fig. 2). Karlin et al. (C, Table 1) determined that most of the imperfect USS were not over-represented in the genome. This is somewhat surprising, as it suggests that mutant versions of the USS are rapidly eliminated (to be addressed in Question 1).

Distribution and spacing of USS:

Karlin et al. (C) found that USS were more evenly distributed than expected of a randomly-located sequence. This applied both to the relative locations of USS in + and - orientations, and to the spacings of elements in each orientation. They suggested that the USS might have a structural role in evenly packaging the chromosome in the nucleoid. The distribution is illustrated in Fig. 1 of Smith et al. (B). Gaps occur only in a 30 kb region containing the genome of a Mu-like phage, whose anomalous base composition suggests is not native to *H. influenzae*, and in a cluster of ribosomal protein genes. The only conspicuous cluster of USS is in a four-fold

repeated section of gene HI1685 (2 USS in each repeat); this may be purely coincidental, as the homologous gene of *E. coli* has a similar five-fold repeat with no USS.

Non-coding regions are enriched in USS:

35% of USS are in the 14% of genome that is non-coding. One third of these occur in inverted-repeat pairs separated by <35 bp (there is only one direct-repeat pair). Discussion of these pairs has focussed almost entirely on their potential role as transcription terminators, but this has not been critically examined (Smith et al., Fig. 3). There is no evidence of a specific function for the other two thirds of the intergenic USS.

Most USS are in coding regions:

Sixty five percent of the *H. influenzae* USS (934) are in the 86% of the genome coding for proteins. The USS in each orientation have a preferred reading frame; most in the + orientation are translated as SAL, and most in the - orientation as TAV. Karlin et al. point out that these tripeptides are unlikely to seriously distort protein structure. There are no potential stop codons internal to USS in either orientation.

Conclusion to this part:

The detailed information about the abundance and distribution of USS in the *H. influenzae* genome has prompted some speculation about their evolutionary function, but to date there have been no attempts to investigate it. We propose to do so. In the next section I build up the evolutionary foundation that must underly any serious attempt to determine their true function.

Evolution of genetic exchange

To understand whether USS could have been selected to facilitate genetic exchange, we first must critically examine the other processes responsible for genetic exchange in bacteria.

The problem of the evolution of sexual reproduction:

I begin the discussion with true sexual reproduction, as misunderstandings about it have commonly provided the frame of reference for discussions of genetic exchange in bacteria.

Sexual reproduction is ubiquitous in eukaryotes. For most plants and animals, this cyclic alternation of the reductive divisions of meiosis with cell and nuclear fusion is the only way to produce the next generation. Researchers agree that sex must have substantial and reliable benefits to compensate for its costs, which include the physiological and metabolic costs of mating, the transmission of venereal diseases and meiotic drive elements, and the evolutionary cost of failing to find a mate. However they do not agree on what these benefits are (most feel that adequate benefits have yet to be identified).

The search has focused on the potential benefits of recombination. In 1896 Weissman proposed that recombination exists because, by increasing variation, it allows species to avoid extinction when their environment changed. Although this explanation is still entrenched in textbooks, it is known to be seriously inadequate because it relies on long term benefits to the species which are very weak and unreliable relative to the short term evolutionary forces acting on the individual. Recent approaches by population geneticists have instead concentrated on the more immediate benefits of randomising the distribution of deleterious mutations, and of varying the genome faster than parasites and pathogens can evolve. The issue remains unresolved; the strengths and weaknesses of the competing theories are summarised by Kondrashov (1993).

Recombination in bacteria:

Bacteria have no processes directly comparable to meiotic sex. Traditionally, bacterial processes capable of causing genetic exchange (transduction, conjugation and transformation) have been viewed as 'parasexual processes' functionally equivalent to sex. Like sex, these processes have been very useful for strain construction in the laboratory. However, horizontal transfer of chromosomal genes in bacteria is rare, fragmentary, and non-reciprocal, and closer examination of these processes provides little evidence that they evolved for the occasional benefits of genetic exchange. With the possible exception of natural transformation, genetic exchange appears to occur only as a side effect of processes selected for other functions.

Production of novel genotypes by genetic exchange (genetic recombination) in bacteria depends on both the gene transfer processes that move DNA fragments or strands from one cell to another, and the physical recombination machinery that allows incoming strands to replace homologous chromosomal sequences. With the possible exception of transformation by naturally competent cells, there is little evidence that selection for genetic exchange has played any role in their evolution.

Physical recombination: From an evolutionary perspective, the role of recombination in the formation of new genotypes is almost certainly unselected and accidental. For decades research was guided by the 'recombination pathways' paradigm, which assumed that the need to create new genotypes was paramount, and that cells had evolved multiple recombination pathways to ensure that no opportunity was lost (Clark 1971). However it is now generally accepted that these 'pathways' are a human construct, and that the enzymes required for homologous recombination exist because of their vital roles in DNA replication and repair (Morel et al. 1997). Their ability to promote recombination of free DNA fragments with homologous sequences in the chromosome is best explained as largely-unavoidable side effects of their selected functions.

Transduction: In bacteria lacking natural competence, most gene transfer is thought to occur by transduction (Milkman et al. 1990), when DNA from the chromosome of a phage-infected donor cell is injected into the recipient by a phage particle that has packaged a host DNA fragment instead of the usual phage genome. There is no evidence that selection for recombination has acted to promote transduction, and there is substantial evidence that selection has acted on the phage to promote accurate packaging of phage genomes and thus to prevent transduction, and on the host to prevent infection.

Conjugation: The accidental nature of chromosomal gene transfer by conjugation is less obvious, largely because its discoverers, believing they had discovered the bacterial equivalent of meiotic sexual reproduction, named the events and participants accordingly ('conjugation', 'fertility factor', etc.) (Lederberg et al. 1946). The clearest perspective is obtained by considering other issues (Ippen-Ihler et al. 1986). (1) The usual outcome of conjugation is not transfer of chromosomal genes, but transfer of a plasmid into a formerly plasmid-free host. (2) The genes controlling this transfer are located not on the donor or recipient chromosome but on the plasmid that is transferred. (3) The processes leading to transfer of chromosomal genes are unregulated and apparently random recombination events between the plasmid and the donor chromosome. (4) Linkage of the plasmid to chromosomal genes reduces its chance of being successfully transferred to the new host. From this perspective we see that conjugation, like transduction, is best viewed as an infectious process controlled by an infectious agent, the plasmid, and that transfer of chromosomal genes is its rare and unselected side effect.

Evolution of natural competence (discussed in Redfield 1993b, provided as Appendix item D)

The only process of gene transfer that is not obviously accidental is natural competence. No infectious agents are involved, and the genes controlling DNA uptake are in the chromosome of the recipient. Environmental DNA comes from dead cells, so the DNA 'donors' are not simply passive, but have passed away.

Costs of competence: Competent cells must undergo the expense of synthesizing the DNA uptake machinery. In the natural environment most recombination events are likely to be either selectively neutral or harmful, and most available DNA is of poor genetic quality. In particular, uptake of phage or prophage DNA can kill cells (Boling et al. 1972), and the *H. influenzae* phage HP1 has a high density of the preferred uptake signal sequences (Fitzmaurice et al. 1984). Furthermore, DNA uptake (especially of heterologous DNA) often induces the SOS response, which in the *H. influenzae* strain Rd causes cell death by inducing a prophage resident in its genome (Setlow et al. 1973; Albritton et al. 1986).

Possible functions of competence:

If we accept that natural competence is not an accident but an adaptation, there are three functions it might serve: production of recombinant progeny, recombinational repair of DNA lesions, and acquisition of nucleotides.

Could competence have evolved to cause recombination?

Although many of the costs and possible advantages of sex also apply to transformational recombination, genes responsible for transformation face two additional obstacles to success.

1. The cost of the excess mutations in DNA from dead cells: Transforming bacteria recombine their chromosomes with DNA fragments from the environment. Because we have no evidence that bacteria can remain viable while releasing copies of their genomes, we must assume that the available DNA usually comes from cells that have died and lysed. At least some of the cells suffering such a fate are likely to have carried more deleterious mutations than their surviving relatives, so any transforming cell that replaces part of its genome with homologous DNA from the environment risks increasing its burden of mutations (D)

2. The risk that an allele needed for competence will be lost by transformation: When a mutation causing transformation (we can call it a *com*⁺ allele) first arises in a population, any homologous DNA available in the environment will carry the *com*⁻ allele. Under these circumstances, each time the *com*⁺ cell takes up DNA it risks bringing in a fragment that will convert its *com*⁺ allele back to *com*⁻. This is a fate worse than

R. Redfield's NIH grant proposal (Funded Jan. 1 2000)

death: the *com*⁺ allele will not only eliminate itself, but resuscitate a competing allele. The *com*⁻ allele faces no such risk. The disadvantage to the *com*⁺ allele will be highest when it is rare, and will be proportional to the amount of transformation it causes. Because of this frequency-dependent cost of transformation, the *com*⁺ allele may fail to become established in the population even if transformation otherwise increases fitness. (Redfield et al. 1997, provided as Appendix item A)

In principle both of these barriers can be overcome if the advantages of recombination are sufficiently high. However, our models (A, and Redfield 1988, provided as Appendix item E) have found that the benefits are very unlikely to be that high. These models were constructed, like many models of sexual reproduction, so that recombination could reduce accumulation of recurrent deleterious mutations. As in the sexual models, the benefits of recombination were absolutely dependent on synergistic epistasis between mutations (i.e. a double mutant must be less fit than predicted from the product of the fitnesses of the single mutants). However, recent experiments examined epistasis between mutations in *E. coli*. Although some combinations of mutations showed synergistic epistasis, most did not, and on average there was no epistasis at all (Elena et al. 1997).

Our simulations did suggest that cells might be able to optimize the cost/benefit ratio by regulating their competence according to the potential for benefits, and that such regulation might even create benefits in the absence of synergistic epistasis. However, this has not yet been evaluated in more sophisticated models.

Independent evidence against recombination being the primary function of competence is the strong linkage disequilibrium observed for several transformable species, which implies that recombination between different lineages is very rare (Musser et al. 1988; Caugant, 1987 #28).

Could competence have evolved for DNA repair?

Double-strand damage to DNA is lethal unless an intact homologous strand can be recombined across the damaged site (Kushner 1987). Michod and coworkers have proposed that the main function of competence is to obtain, from the environment, the intact DNA strands needed for this recombinational repair. They found that when competent *B. subtilis* cultures were provided with homologous DNA after the cells' own DNA had been damaged by UV irradiation, the proportion of transformants is increased among the survivors (Michod et al. 1988). This suggested that the cells that took up DNA may have survived damage better than the cells that did not. Subsequent experiments confirmed and extended the observations (Wojciechowski et al. 1989; Hoelzer et al. 1991), but the inference remains indirect, and the data are not compelling.

A more direct test of the DNA repair hypothesis is possible in *H. influenzae*, because all the cells in *H. influenzae* cultures become competent (Herriott et al. 1970). To find out whether transforming DNA helps cells repair DNA damage, I exposed competent *H. influenzae* cultures to various doses of DNA-damaging agents and allowed them to repair the damage in the presence of either *H. influenzae* DNA or nonhomologous (control) DNA. Contrary to the predictions of the repair hypothesis, cells given *H. influenzae* DNA survived no better than cells given control DNA. (D). This negative result doesn't mean that the repair hypothesis is wrong, only that any advantage of transformational repair must be smaller than the sensitivity of the experiment. This reflects a more general problem with attempts to demonstrate natural selection in lab cultures; these experiments are often much less sensitive than natural selection can be. We are using studies of regulation to circumvent this problem.

Regulation of competence: Because regulatory mechanisms evolved by selection for adaptive expression of the traits they control, we have been using studies of regulation to answer questions about function. This is an especially valuable approach in microbiology, as it is very difficult for microbiologists, living on a scale of meters, to comprehend the natural environments of bacteria, living on a scale of microns. We do not know the significant features of the natural environments of any bacteria, even *E. coli*. But studies of regulation, in effect, ask the bacterium what features have been important to it. Thus the mechanism of regulation of competence gives us a window on what the function of competence has been in the natural environment, even though we must study regulation in the laboratory. Another benefit of studying regulation is that, because selection for appropriate regulation has integrated selective benefits over evolutionary time, the regulatory mechanism can reflect small selective advantages that would be undetectable in the short time available to lab experiments.

Regulation by DNA damage? The DNA repair hypothesis suggests the obvious question 'Is competence regulated by DNA damage?' The required components of such regulation are already in place: competence is tightly regulated in most naturally transformable bacteria, although the actual inducing signals have not usually been identified, and these bacteria all possess SOS-like pathways which regulate expression of many genes in response to the presence of DNA damage. So we examined the relationship between DNA damage and the induction of competence, and found that, contrary to the repair hypothesis's prediction, DNA damage does not

R. Redfield's NIH grant proposal (Funded Jan. 1 2000)

induce or enhance competence in either *H. influenzae* or *B. subtilis* (Redfield 1993a, provided as Appendix item F). Thus we think it unlikely that the primary function of transformation in these organisms can be DNA repair.

Could competence have evolved for obtaining nucleotides?

Nutrient acquisition is potentially a much stronger selective force than either recombination or DNA repair, and nucleotides of DNA are in great demand, as are the constituent sugars, phosphates, and nitrogenous bases. Data from *E. coli* indicates that a cell growing on unlimited glucose expends 15.4% of its energy budget on nucleotide synthesis, compared to 3.7% on amino acid synthesis (Stouthamer 1979)). But is DNA likely to be available in significant quantities in the environments of bacteria? The answer is yes for a large proportion of transformable bacteria, especially pathogens.

The naturally-competent bacteria *H. influenzae*, *S. pneumoniae*, *N. gonorrhoeae* and *N. meningitidis* live on respiratory tract or other mucosal surfaces, where double-stranded DNA is often present in extremely high amounts. DNA concentrations of 300µg/ml are typical for the respiratory tract mucus of healthy individuals, and the sputum from cystic fibrosis patients contains an average of 4mg DNA/ml (Matthews et al. 1963). These are clearly nutritionally significant amounts, especially as mucus contains little else but glycoproteins. And because the DNA is derived almost entirely from host cells (Lethem et al. 1990), it could not be used for recombination or repair. However it could provide a substantial fraction of the precursors for bacterial DNA and RNA synthesis. In support of this, we know that competent *H. influenzae* efficiently recycle into their own DNA synthesis the nucleotides released by degradation of the DNA they take up (Barany et al. 1983). In fact, *H. influenzae* is unable to synthesize its own pyrimidines, and so must obtain precursors from its environment (Fleischmann et al. 1995). Other respiratory tract bacteria are similarly fastidious.

The obvious objection to this argument is that the USS-specific uptake system should prevent uptake of host DNA. Discussion of this is deferred to the next section.

Regulation by metabolic needs: Regulation of competence in most bacteria is more consistent with a metabolic function than with one of DNA repair. Many such bacteria become spontaneously competent in stationary phase, or in the period of unbalanced growth immediately preceding it. Details of how competence is regulated provide limited (though not conclusive) support for a nucleotide-acquisition hypothesis.

We have been investigating the regulation of competence in *H. influenzae* (research funded by the Medical Research Council of Canada). We have found that competence is controlled primarily by a metabolic signal, the level of cAMP, in turn responding to changes in the availability of preferred energy sources, signalled by the phosphotransferase system (Dorocicz et al. 1993; Macfadyen et al. 1996a, 1996b (provided as Appendix item G)). We have also found (MacFadyen and Redfield, unpublished) that competence induction is greatly reduced by nucleotide supplementation of the starvation medium normally used for competence induction.

Our results suggest the following simple model for the regulation of competence in *H. influenzae*. The first step is a rise in intracellular cAMP, caused at least in part by PTS-dependent activation of adenylate cyclase when fructose is absent. In the presence of cAMP, CRP allows initiation of *sxy* transcription, but Sxy protein will not be expressed unless a second signal is also present. This specific signal is depletion of nucleotide pools, which, by limiting or preventing the formation of secondary structure in *sxy* mRNA, allows transcription to produce functional mRNA. The Sxy protein, in turn, activates transcription of DNA uptake genes, and the ensuing DNA uptake relieves pressure on nucleotide pools.

Regulation by the need for recombination: Although the regulation of competence is consistent with a solely nutritional function for transformation, it is also not inconsistent with selection for benefits of recombination. The biggest potential benefit of recombination is decreasing the load of deleterious mutations (Kondrashov 1993). Our analysis suggests that the average benefit is very small, but some cells have more to gain from recombination than others. Cells with many mutations have much to gain, because they may replace mutant genes with functional copies from other cells. But mutation-free cells can only lose. Our simulations showed cells could maximize the benefit of recombination by regulating competence according to their mutation load (Redfield 1988). How could this work? Cells have a number of stress responses, which can alter gene expression in response to various emergencies (Collado-Vides et al. 1991). Consider a cell that induced competence whenever these responses failed and intracellular levels of many key metabolites fell below a critical threshold. Because many deleterious mutations will have their main impact on these metabolites, the cell would often be responding to mutations. In this way, selection for the benefits of recombination might result in competence being induced as the ultimate global stress response.

Above I have argued that because competence is not induced by DNA damage it is unlikely to have evolved for DNA repair, and that the potential short-term benefits of recombination are so small that they will be insignificant relative to the nucleotides obtained by DNA uptake. However neither the recombination theory nor the argument from regulation are conclusive, and the true function of competence remains unresolved.

Importance of recombination: Mechanisms of recombination may exist only for DNA replication and repair, but the recombination they cause has nevertheless been extremely important for bacterial evolution. Every sequenced bacterial genome contains many horizontally-transferred segments, evidence of recurrent selective sweeps by recombinant ancestors (Lawrence et al. 1998). Despite this, there is no evidence that such selection has had any effect on the processes that produce recombinants. Two factors probably account for this. First, most random recombination events will reduce fitness rather than increase it, so recombination may be a net cost rather than a benefit. Second, beneficial recombinants arise so rarely that they may be unable to influence the evolution of the genes that produce them, because these genes are under constant strong selection for their immediate functions. This perspective on the evolution of genetic exchange in bacteria parallels our present understanding of the evolution of mutation rates. Although without mutation there would be no evolutionary change at all, selection on the processes that generate mutations appears to have acted to prevent mutations rather than to facilitate them, no doubt because almost all non-neutral mutations are deleterious.

If transformation is an accidental consequence of DNA uptake, why have USS?

If competence is a nucleotide-acquisition mechanism, the uptake specificities of competence systems seem counterproductive. On the other hand, if DNA uptake evolved to produce new genotypes by homologous recombination, sequence-specific uptake may be an adaptation ensuring that competent cells do not waste efforts taking up DNA fragments from unrelated genomes, or take up dangerous foreign DNAs. This explanation has been widely accepted, but never critically examined, and it has substantial problems.

1. The nature of the selection: For such a system to evolve, there must be selection for a USS-specific DNA-binding protein (the DNA receptor on the cell surface) and for a high frequency of USS in the genome. Given sufficiently strong selection for recombinational benefits of DNA uptake, evolution of a biased receptor could be straightforward. However, selection for more USS in the genome is problematic, because USS only become useful after the genome they reside in has been released by lysis of the cell. Thus selection for genomes with many USS might require biologically-improbable group selection to support the extreme altruism of cell death. The clonality of some bacterial populations does not strengthen group selection here, as genetic exchange within clonal populations is unlikely to confer any benefit at all unless mutational differences accumulate within the clone.

2. USS are not good markers of sequence homology: An unexpected finding some years ago was that the genomes of some other members of the family *Pasteurellaceae* (genera *Haemophilus*, *Pasteurella* and *Actinobacillus*) behaved in uptake-competition experiments as if they too had at least several hundred copies of the USS (Albritton et al. 1986). Thus the USS system causes *H. influenzae* to efficiently take up DNAs with which it cannot recombine, and which often cause substantial lethality by inducing the SOS response (Setlow et al. 1973).

3. USS may promote uptake of genetic parasites: The well-studied *H. influenzae* phage HP1 has almost as high a density of USS in its genome as does the host chromosome into which it integrates (Esposito et al. 1996). Similarly, the *Neisseria* USS is present on a *Neisseria* plasmid (Graves et al. 1982). The phage and plasmid may have evolved USS to exploit the uptake systems of their host, entering the cell by the competence pathway. An alternative explanation might be that phage and plasmid genomes residing in the host cells have been subject to the same forces that cause USS abundance in the chromosome, and that these forces act on both beneficial and harmful DNAs.

Possible resolution:

The problem of how USS abundance could be selected in living cells can be resolved by having USS be maintained by the bias of the receptor. A biased receptor will directly enrich the genome with the USS-containing fragments it takes up, because these sometimes replace chromosomal sequences (Redfield 1991) (provided).

Under this model, there was originally no USS. Because bias is an intrinsic characteristic of DNA-binding proteins, an initially-arbitrary bias of a DNA uptake protein would create a kind of molecular drive that imposed its preferred sequence on the genome. Mutations creating the preferred sequence in 'donor' DNAs would spread by transformation, and mutations destroying the sequence in the recipient chromosome would be restored by

transformation with DNA from wildtype donors. Selection for more efficient DNA uptake (for whatever benefit) could then have refined the receptor bias and further increased the USS abundance. Even though this model allows the benefit of DNA to be independent of recombination, it does require that a significant fraction of the DNA taken up be capable of recombination with the chromosome.

Along this line, Smith et al. (Smith et al. 1995) (provided) suggested that the frequency of singly-mismatched USS reflects a balance between mutation away from the preferred sequence and restoration of perfect sequences by transformation. The frequency of singly-mismatched USS is quite low relative to the perfect USS, implying that transformation would have to have been both frequent and highly specific for the perfect USS.

Could USS-based uptake systems have another function?

There is no direct evidence that the *H. influenzae* or *N. gonorrhoeae* USS affect any process other than DNA binding by competent cells. It has not been experimentally possible to determine whether DNA processing or recombination is influenced by the presence of a USS. Nor is there direct evidence for any role of the USS in non-competent cells. Several functions have been suggested by the analysis of the *H. influenzae* Rd genome sequence (discussed in detail below), but there has been no critical evaluation of these ideas.

Terminator function: A role in termination was first suggested by the finding of *N. gonorrhoeae* USS in inverted-repeat (dyad) pairs downstream from coding regions (Goodman et al. 1988). Smith et al. found 127 intergenic dyad pairs in the *H. influenzae* genome, substantially more than predicted by chance. However there has been no attempt to determine how many of these are actually in positions where they could act as terminators (i.e. immediately downstream from a transcription unit, and followed by a string of Ts). As 90% of USS occur singly, and most (65%) are within open reading frames, any role involving the regulation of transcription is likely to be secondary.

Structural function: Karlin et al. hypothesised that the significantly even spacing of USS around the genome reflected a role in DNA replication or repair, or in the physical structuring of the chromosome. However USS show no orientation bias around the chromosome, as might be expected for a sequence that interacted with DNA replication machinery. No intracellular USS-binding protein is known.

USS are not Chi sequences: We know that USS do not function as Chi sequences, interacting with the RecBCD complex and promoting production of recombinogenic strands, because *H. influenzae* has a Chi sequence unrelated to the USS (5' GNTNNTNN-3' and 5'- G(G/C)TGGAGG-3') (Sourice et al. 1998). In fact, Chi sequences may provide a paradigm, as repeats whose abundance was once thought to have been selected to promote recombination, but which are now considered to be abundant as a consequence of their degradation-blocking effect on the RecBCD complex (Myers et al. 1994). USS may arise in the same way, protected from degradation by an unselected sequence bias of an intracellular protein.

In summary: Analysis of USS in the *H. influenzae* genome has not explained their abundance. Deviations from randomness in distribution, orientation and reading frame are intriguing, but none provides strong support for any primary function. Rather, the bias for non-coding regions, frequency of dyad repeats, and preferred reading frames are probably best explained as secondary adaptations to the as-yet-unidentified forces responsible for their abundance.

Significance of USS for pathogenicity:

All of the bacteria known to contain USS are serious human pathogens. Serotype b strains of *H. influenzae* are the most common 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. An anti-type b vaccine effective in infants is available in developed countries, but there is no vaccine for other serotypes, which are a major cause of childhood ear infections and of respiratory disease in infants and elderly or immunocompromised individuals, especially people with AIDS. *Neisseria gonorrhoeae* is of course a serious sexually-transmitted disease and a major cause of infertility. *N. meningitidis* is a serious respiratory pathogen, especially in undeveloped countries, and the major cause of meningitis in young adults (Booy et al. 1998).

Competence and transformation have medical relevance at several levels. Multi-drug-resistant *H. influenzae* and *Neisseria spp.* are increasing (Doern et al. 1997; Fox et al. 1997), and antibiotic resistance genes, virulence determinants and genes specifying the different capsular serotypes can be spread by transformation (Saez-Nieto et al. 1990; Kroll 1992; Spratt et al. 1992; Kroll et al. 1998).

An improved understanding of the roles of USS may provide new targets for medical intervention. One possibility is using USS oligonucleotides to block DNA uptake - whether this is useful will depend on resolving

the function of competence. Another is using the USS as a tag to promote uptake of antibiotics. Finally, identification of a USS-dependent chromosomal process might provide a point of intervention specific to the pathogen and nontoxic to the host.

Implications for the evolution of genetic exchange:

Genetic exchange is a major driving force in bacterial evolution, but does genetic exchange, like mutation, occur only by accident? USS-dependent DNA uptake appears to be the best evidence that selection acts to promote genetic exchange. However we know of no selective processes that could have produced this system, and it is just as likely to have arisen from forces unconnected to genetic exchange.

This is an extremely important issue. Most research in this field has assumed that mechanisms producing genetic exchange have been selected for the purpose. The research described below aims to test this assumption by determining the function of the USS.

C. Preliminary Studies

	<u>Page</u>
1. Models of the evolution of USS	28
Model 1. Proportions of perfect and mismatched USS in the genome	28
Model 2. Population structure of a single USS	29
2. Analysis of USS in bacterial genes and genomes	29
a. A test of Model 2	29
b. Absence of <i>H. influenzae</i> USS in human DNA	29
c. <i>H. influenzae</i> USS in other <i>Pasteurellaceae</i>	30
d. Analysis of <i>Neisseria</i> sequences	30
e. Absence of USS-like sequences in genomes of <i>C. jejuni</i> and <i>H. pylori</i>	30
f. Constraints imposed by USS in <i>H. influenzae</i> coding sequences	30
g. Correlations between USS consensus and uptake in published data	31
3. Experimental work on interactions between the USS and its receptor	31

1. Models of the evolution of USS: (provided as Appendix item H)

Our goal is to use modeling to test hypotheses about USS evolution. The two models described below ask whether USS can be maintained by the bias of the DNA uptake system, as we have proposed (Redfield 1991, provided as Appendix item D). This is an appealing hypothesis, and relatively easy to test because it requires no assumptions about the benefits of DNA uptake or the cause of the bias.

These two preliminary models start with a population of competent cells with USS-dependent DNA uptake, and share the following assumptions: 1. Mutation both creates and destroys USS, and that recombination with incoming homologous DNA replaces genomic sequences with and without USS. 2. Changes in the USS's sequence are selectively neutral for cell functions other than DNA uptake. 3. For simplicity, the DNA fragments available for uptake (in the DNA pool) come from a population of cells identical to the population taking up the DNA (i.e. either cell death is nonselective, or all living cells spontaneously release DNA into the environment). The models are thus able to make predictions about the equilibrium state of the USS.

Model 1. Biased uptake determines the proportions of perfect and mismatched USS in the genome (H):

Figure 2 here

This simple model assumes that the ratio of perfect and mutated USS in the genome reflects an equilibrium between mutational changes to the individual USS and preferential uptake of perfect USS by a USS-biased DNA receptor. This permits derivation of an equation giving the values of the transformation frequency and uptake bias parameters needed to account for the high proportion of perfect USS in the *H. influenzae* Rd genome. These predictions are shown below. For example, if the DNA receptor is 100 times more likely to take up a fragment with a perfect USS than one with a singly-mismatched USS, cells would have to be replacing 0.0097 of their genome in each generation to account for the observed ratio of perfect and mismatched USS.

Bias (uptake of perfect USS:singly-mismatched USS)	5.0	10	20	50	100	1000
Uptake (amount of genome replaced per generation)	0.20	0.10	0.05	0.02	0.01	0.001

This is a pleasingly tidy result, as it makes good biological sense and the values are not unreasonable. Estimates of the actual bias vary, no doubt partly because they have been determined using USS with different flanking sequences, but are usually between 10 and 100. We have no good estimate of actual transformation frequencies

for *H. influenzae* in its natural environment, but if cells grow slowly, and frequently take up DNA as food then an average of 1% transformation per generation seems plausible, and even 10% not impossible.

The model is weakened by several unrealistic assumptions. It neglects USS with more than a single mismatch, and also ignores the 20bp of conserved sequences flanking the USS. A more realistic model would also include functional constraints on USS-containing sequences, and the hazards of taking up DNA from dead cells.

Model 2. Biased uptake drives the population structure of a single USS (H):

Figure 3 here

This more sophisticated model was developed by our collaborator Tony Dean. Its goal is to make testable predictions about the degree of conservation of any particular USS in different isolates of the species. Unlike the previous model, it describes a USS at one location in the genome, in all the members of a population. The frequency of the perfect USS (at equilibrium is determined by the balance between mutational changes to the USS, the effect of the various USS versions on the cell's fitness (e.g. due to coding constraints), and preferential uptake of perfect USS by a USS-biased DNA receptor.

Two versions are provided (H). In the first version, the perfect USS (USS₀) is deleterious to cell fitness, and all other versions of the USS (USS₁₋₉) are selectively neutral. The model's results allow the equilibrium frequency of each USS type to be determined, as a function of the difference between the transformation rate constant for the perfect USS and the selective cost this USS imposes on the cell. The figure on page 4 of the model shows that the equilibrium is exquisitely sensitive to this difference; 80% of the USS are perfect when the transformation rate constant is only 10⁻⁷ greater than its selective cost, but sequences unrecognizable as USS (USS₆₋₈) dominate as the difference approaches zero.

The second version of the model is similar, but assumes that all versions of the USS except one (USS_j) are deleterious. Its results are expressed differently, but show the same sensitivity to the difference between the transformation rate constant and the strength of selection against the perfect USS.

The sensitivity of the USS distribution to small changes in selection predicts that less-than-perfect USSs will not be maintained by biased uptake. Those that are not restored to perfection will be lost to additional mutation, rapidly acquiring the fittest sequence for the site.

2. Analysis of USS in bacterial genes and genomes:

a. A test of Model 2:

We have been testing the predictions of Model 2 by analyzing the conservation of USS in DNA sequences from *H. influenzae* isolates other than Rd. Our strategy was to examine all the non-Rd sequences for USS, and then compare USS-containing segments with the Rd sequence. We downloaded the 282 non-Rd sequences in GenBank (as of December 1998), discarded those with no homology to Rd, and from the remaining 333,571 nucleotides we collected 216 segments of 100bp, each centered on a perfect (118) or singly-mismatched (98) USS.

The model predicts that, when these segments are compared to their

Figure 4 here

Rd homologs, some of the USS will be more conserved than the sequences around them (Fig. 2 D). However most of the 100bp segments differ from their Rd homologues at only a few positions (Fig. 2A), an amount of variation that does not provide a strong test of the model's predictions. Some sequences had more variation, but their USS were no more often conserved or diverged than the sequences around them (Fig. 2B). The analysis is greatly weakened by the limited variation of the available sequences. To the extent that the results are significant, they do not conform to the model's predictions. This may be because the model's predictions are based on free recombination, whereas the available data indicates that the *H. influenzae* population structure is clonal. The model also includes a high rate of mutational divergence, higher than would be seen within a single bacterial species.

b. Absence of *H. influenzae* USS in human DNA:

H. influenzae's natural environment is human respiratory mucus, which contains very high concentrations of human DNA (Lethem et al. 1990). If the USS sequence is common in human DNA, the receptor's preference

for it could have evolved simply as a food-recognition adaptation. We examined 2000 kb of human sequences, and found only seven *H. influenzae* USS, the frequency expected of a random sequence of its base composition. Thus USS-preference did not evolve as an adaptation for taking up human DNA.

c. *H. influenzae* USS in *Actinobacillus actinomycetemcomitans*:

If USS exist to promote homologous recombination, we might expect that distinct species would have distinct USS. However, DNAs of several members of the family Pasteurellaceae, including *Actinobacillus actinomycetemcomitans* (*A. actino*), compete efficiently with *H. influenzae* DNA for uptake by *H. influenzae* cells, suggesting that they may contain sequences resembling its USS. Figure 3 shows the phylogenetic relationships inferred from 16S rRNAs (Dewhirst et al. 1993) and the results of several tests of DNA relatedness (Albritton et al. 1984; Albritton et al. 1986). The genome of *A. actino* is being sequenced, so we examined 133 kb of sequence (downloaded from www.genome.ou.edu/act.html) and found 52 occurrences of the *H. influenzae* USS AAGTGC GGT or its complement, a frequency of 0.4 per kb. Alignment of these revealed a flanking consensus very similar to that seen in *H. influenzae*.

Figure 5 here

The similarity of the USS cannot just be a relic of descent from a competent common ancestor, because the overall sequence divergence between the two species is considerable. Their 16S rRNA genes show only 94% identity (compare to 97% between *E. coli* and *Salmonella*), and *A. actino* DNA transforms *H. influenzae* 10,000-fold less efficiently than does *H. influenzae* DNA (Albritton et al. 1986).

Figure 6 here

This tells us that the USS have not evolved as a genetic barrier between related species (a molecular mechanism of pre-zygotic reproductive isolation). One possible explanation for the shared USS is that the USS are conserved for reasons other than a function in competence. Another is that, although the organisms' niches do overlap (*H. influenzae* is in the nasopharynx and *A. actino* in gingival mucosa and dental plaque, they encounter each other's DNA infrequently enough that there has been no selection for divergence of the USS.

d. Analysis of *Neisseria* sequences:

We tested whether the *N. gonorrhoeae* USS might also occur in *N. meningitidis* by examining *Neisseria* contigs (made available by the three *Neisseria* genome projects) for the reported *N. gonorrhoeae* USS, GCCGTCTGAA, and its reverse complement. We found this sequence to be abundant (approximately 1 copy per kb) in sequences from both species, with little evidence of any consensus in the flanking sequences.

N. gonorrhoeae and *N. meningitidis* are clearly distinct species, although more closely related than *H. influenzae* and

A. actino (their 16S rRNAs are 98% identical). Their niches overlap (about 25% of patients with genital *N. gonorrhoeae* infections have the same strain in the pharynx). As in *A. actino* and *H. influenzae*, the shared USS means that *Neisseria* cells cannot use its USS to make species-level discrimination at the level of DNA uptake.

Figure 7 here

e. Absence of a USS in genomes of *Campylobacter jejuni* and *Helicobacter pylori*:

Both *C. jejuni* and *H. pylori* are naturally competent, and *C. jejuni* preferentially takes up its own DNA. Our collaborator Paul Lee has analyzed the completely-sequenced genomes of *C. jejuni* (not yet published; available at www.sanger.ac.uk/Projects/C_jejuni/) and *H. pylori* (GenBank) and found no abundant repeats comparable to the USS of *Haemophilus* and *Neisseria*. Thus uptake specificity can exist without sequence specificity.

f. Constraints imposed by USS in *H. influenzae* coding sequences:

Two preliminary tests of the extent to which the *H. influenzae* USS influences protein-coding capacity suggest that there is surprisingly little constraint. 1. Comparison of conserved coding sequences. We have used the sequences of the USS-containing genes for three conserved proteins (*groEL* (2 USS), *trpB* and *gcp1*) to examine constraints. In each case we found that the USS-encoded amino acids of the *H. influenzae* protein did not differ significantly from the corresponding amino acids of its homologs, suggesting that the USS was not affecting protein function. 2. Availability of USS-encodable tripeptides in the proteome (the sum of all proteins

R. Redfield's NIH grant proposal (Funded Jan. 1 2000)

encoded by the genome): The USS in coding regions have preferred reading frames; SAV for USS in one orientation (354 USS) and TAL for USS in the other orientation (268 USS) (Karlin et al. 1996). This allows us to test whether the USS are constraining *H. influenzae* protein sequences, by asking whether the number of SAV and TAL tripeptides in the proteome exceeds the coding demands of the USS. We examined the amino acid sequences of 35 arbitrarily-chosen *H. influenzae* genes (genes HI0050, 0100, 0150, etc.) for these tripeptides, and found seven SAV and ten TAL out of 8575 total amino acids. Six of these tripeptides could have been encoded by USS, but were not. If this holds over the entire proteome of 525,000 amino acids, the *H. influenzae* genome could acquire more than 350 additional USS without any changes to the proteins it encodes.

Codon usage considerations: No clear patterns relating USS and codon preference were found using the detailed analysis of codon usage in *H. influenzae* (Pan et al. 1998). Karlin noted that some singly-mismatched USS occurred at a significant frequency, and suggested that they might be tolerated in coding regions only because of coding constraints. However, consideration of codon-usage patterns and the properties of the amino acids provides only partial support. Some of the significant variants specify different amino acids, but others do not change the amino acids and require use of less-favored codons

g. Correlations between USS consensus and uptake in published data:

Does the bias of the DNA uptake machinery match the extended USS identified by sequence comparisons? The original *H. influenzae* USS was identified by uptake experiments, but Smith et al.'s analysis of the genome sequence is only correlational. As a first step in remedying this, we have compared the extended USS (USS + conserved flanking sequences) with the sequences of the fragments used in published uptake experiment, especially the large data set (28 sequences) created by Goodgal's lab (1990). This analysis shows a good correspondence between a fragment's match to the extended USS and its relative uptake by competent cells, supporting the hypothesis that the observed USS consensus reflects the preference of the uptake system, and thus that its primary role is in DNA uptake. We plan to carry out a quantitative re-analysis of Goodgal's data, using the information-content matrix method described by Stormo and Hartzell (Stormo et al. 1989).

3. Experimental work on interactions between the USS and its receptor:

A new graduate student, Grant Poje, has begun a project to identify the protein or proteins that bind the USS on the surface of competent *H. influenzae* cells. A USS with best-match flanking sequences was designed according to the consensus reported by Smith et al. (B). Two complementary strands were synthesized and annealed to create a 29 bp double-stranded fragment (USS-1). In parallel, a random-sequence fragment with the same base composition as the perfect USS was created as a control (USS-R).

Although USS-1 competes with Rd DNA for uptake by competent *H. influenzae* cells (Fig. 7A), neither USS-1 nor (as expected)USS-R is efficiently taken up (Fig. 6). We have obtained much better uptake with a 50 bp synthetic fragment(USS-50, Fig. 6), equivalent to about 65 molecules per cell. This USS will be used in the receptor-tagging experiments described below (Question X).

Figure 8 here

USS consensus:	aAAGTGCGGT. rwwwww. rwwwww
USS-1:	AAAGTGCGGTAAATTTTTTAACGTATTTTT
USS-R:	TCTTGTTAGAATCTGAGTGTTATTTAAAT
USS-50	TAATGGTACCATATAAAGTGCGGTAAATTTTTACAGTATTTTTGAATTCC

Mutations in the *H. influenzae* genes *sxy-1* and *murE* greatly increase transformation frequencies (Williams et al. 1994, Ma and Redfield, in prep). The *murE* mutation affects the cell wall, and so might alter the uptake specificity and thus provide a clue to its genetic basis. However both mutants were found to have normal uptake specificity (Fig. 7B).

Figures 9 and 10 here

D. Research Design and Methods (a Table of Contents is provided on page 18)

Logic underlying the proposed research: There are three processes whose evolution needs to be explained: competence to take up DNA, a DNA-uptake system favoring a specific sequence, and the abundance of the preferred sequence in the genome. The similarity of the USS-based uptake systems of *H. influenzae* and *Neisseria* suggests that similar evolutionary forces are responsible for both systems.

The information presented in the Background and Preliminary Studies sections favors some hypotheses over others. USS do not exist to prevent uptake of DNA from related species sharing the same environment. Their primary function is not termination of transcription, as most USS are not in a position to act as terminators. However a function in nucleoid structure or chromosome packing has not been tested. The molecular drive caused by a strongly-biased DNA uptake mechanism is likely to be an important determinant of USS abundance. However, no adequate explanation of the evolution of the USS-based uptake system has appeared.

From the many unanswered questions about the evolution of USS, we have selected eleven that we think are both important and answerable, which we specifically describe below. Many other questions will also be addressed by the analysis and experiments we propose; we have not singled them out because the answers will either be difficult to interpret or will not constrain specific hypotheses.

Part I. Theory:

Question I. Nucleotide acquisition is an inevitable benefit of DNA uptake. Is it consistently much larger than the other potential benefits?

Of the three well-characterized processes that can exchange chromosomal genes, only competence may have been selected specifically for this effect. However, competence has other potential benefits, DNA repair and nucleotide acquisition, and our current models only succeed in demonstrating that recombination's benefits may be vanishingly small. The modeling we propose will continue and extend our collaboration with Tony Dean.

To address the evolution of competence we will develop a computer-simulation model that allows all of its potential benefits and costs to be examined simultaneously. The benefits of DNA repair are expected to be larger than those of recombination, because DNA damage is more frequent than mutation, and the benefits of nucleotides are expected to be larger than those of repair, because nucleotides are needed for replication more often than for repair. Recombination and repair have received the most attention, but nucleotide acquisition is also the most reliable benefit by far, as it is an inevitable consequence of DNA uptake, regardless of whether the incoming DNA is degraded or replaces a resident strand that is degraded. If the nucleotide benefit is found to be much larger than the others, models of USS evolution can be greatly simplified. The consequences would extend far beyond the evolution of USS, as this finding would imply that the genetic exchange competence permits is unselected, and thus that all genetic exchange in bacteria may be accidental.

It is less important that the models be mathematically rigorous than that they be as robust as possible, so that conclusions will not depend critically on details of the model or on its assumptions. Thus we will mainly work with computer-simulation models, because the assumptions of these are much less constrained by need for mathematical tractability, and because they allow testing of a wide range of conditions; these will be backed up by analytical work where appropriate.

For the benefits of recombination we will retain the deleterious mutation framework used by Redfield (E) and Redfield et al (A), because it is most flexible and has the best parallels with work on the evolution of sex. DNA repair can be modeled in a similar way, with lesions that can be repaired only by recombination arising according to a Poisson distribution, and corrected whenever a homologous DNA fragment is taken up. The DNA in the environment may contain damage lesions as well as mutations; the cost of these will be included in the model.

We will use experimentally determined values wherever possible. Stouthamer has determined that nucleotide biosynthesis accounts for 17% of the metabolic costs of growth for *E. coli* (Stouthamer 1979). The cost of incorporating damaged DNA can be estimated from experiments by Setlow's lab (Beattie et al. 1969) and the cost of SOS induction by non-homologous DNA from lethalities measured by Setlow and Albritton, and from experiments described below (Question VI).

We also must incorporate the regulation of competence by environmental or intracellular signals, as both common sense and our previous model (A) tell us that appropriate regulation can greatly increase benefits relative to costs, and thus can create net benefits that are absent in models without regulation. This is not difficult to simulate; it is necessary only to have cells refrain from DNA uptake whenever a certain condition is

R. Redfield's NIH grant proposal (Funded Jan. 1 2000)

satisfied. The amounts of DNA taken up will also be varied. The nucleotide benefits are likely to be linear, but a small amount of recombination is sufficient to give most of recombination's benefits.

We will use the model to simulate the growth rate of competent cells under various assumptions and conditions. We will also use it to simulate competition between competent strains with different combinations of benefits (i.e. cells that receive only the nutritional benefit with cells that receive only the recombinational benefit, which will provide more realistic estimates of the relative benefits.

This will be the first comprehensive model of the evolution of competence. With appropriate choice of parameters, its conclusions should be applicable to all naturally competent bacteria.

Question II. A biased DNA-uptake receptor can increase the frequency of its preferred sequence in the genome. Will receptor bias inevitably do so if cells take up homologous DNA? How does this depend on the strength of the bias and the amounts of DNA taken up?

Because a USS only influences DNA uptake after the cell carrying it dies and release its DNA to the environment, the abundance of the USS may be explained most readily either by an intracellular function (examined in Question XI) or as a consequence of the molecular drive we invoked in the two preliminary models described above. We now propose to develop more sophisticated models of this type, in which genomic accumulation of USS due to a biased uptake mechanism is opposed by mutational loss of the USS and by the constraints on their location in the genome.

The merit of such models is that they shift the explanation of USS abundance from selection favoring USS-rich cells to selection increasing the bias of the receptor (relatively straightforward). It is not necessary to assume that the original receptor was completely unbiased, because any DNA-binding protein is likely to exhibit some arbitrary bias in the sequences it binds. A modification of these models would assume that the initial bias did not reside in the receptor but in an intracellular DNA processing activity that preferentially spared USS-containing sequences from degradation.

Mutation is relatively simple to model as one force opposing the buildup of USS in the genome. The other force, the constraints USS place on genome function, is not understood; analysis to clarify it is proposed below (Question V). One simple interim approach would be to assume that all USS are additions to the genome, and that the cost of each USS is that of replicating and maintaining this extra DNA. Although this assumption is not consistent with the locations of most USS in coding regions, it will be useful until we have obtained better information on actual constraints.

We have begun this modeling in the absence of an adequate model of the evolutionary function of competence, but it will be much more valuable if we can build it on a solid evolutionary foundation. We expect the models to show biased uptake to be a powerful driving force, sufficient to explain the high abundance of USS in the genome. Because the bias can be experimentally determined (Question IX below), these models should also predict the frequency of transformation of these pathogens in the human host.

Question III. The *H. influenzae* USS are not randomly distributed around the genome. Can the even spacing be explained by receptor-bias-driven accumulation?

The spacing of the *H. influenzae* USS is surprisingly even, if we set aside the pairs of USS in inverted orientation. This may reflect an intracellular function, as suggested by Karlin et al. (C), (to be examined in Question XI). Alternatively, it may result from differential interactions between a biased uptake mechanism and fragments with different numbers of USS.

Specifically, if USS on the same fragment act independently, we might expect a random genomic distribution to result from biased uptake. If USS interact synergistically, so that a fragment with two close USS is much more likely to be taken up than one with widely-separated USS, we might expect an underdispersed (clumped) distribution; this could explain the large number of inverted-repeat pairs. On the other hand, if fragments with two USS are not strongly preferred over those with one USS, we might expect an overdispersed (even) distribution around the genome. We will develop a rigorous model of this intuitive explanation. If the results suggest that uptake interactions are a plausible explanation of USS distribution, we will directly measure uptake of fragments with different numbers and arrangements of USS. If not, we will focus more intensely on the search for an intracellular function of the USS.

We are not specifically proposing to model the evolution of receptor bias at this time, because as yet we know too little about both the benefits and the mechanism of DNA uptake. However the ultimate goal of our modeling

efforts is a unified model that includes the costs and benefits of DNA uptake, the evolution of a DNA-binding receptor/gene with biased uptake, and the evolution of sequences preferred by receptor.

Part 2. Genomic analysis of USS:

Bacterial genome projects provide superb resources for analyzing patterns of USS abundance and distribution. Smith et al. (B) and Karlin et al. (C) analyzed USS in the *H. influenzae* Rd genome but did not explicitly test any hypotheses about USS evolution. We will do so, using the completed and emerging genomes with USS listed in the table below.

Status of selected bacterial genome projects as of March 1, 1999:

<u>Organism</u>	<u>Genome size</u>	<u>Organization</u>	<u>Completeness</u>
<i>Haemophilus influenzae</i>	1.83	TIGR	Complete => 1,830,138bp
<i>Actinobacillus actinomycetemcomitans</i>	2.2	OU-ACGT	780 contigs; 2008273bp
<i>Neisseria gonorrhoeae</i>	2200 kb (1855 orfs)	OU-ACGT	Sequencing complete, assembly underway.
<i>Neisseria meningitidis</i> MC58	2157 kb (1841 orfs)	TIGR	2533 contigs => 1,406,901bp
<i>Neisseria meningitidis</i> group. A	2230 kb (1841 orfs)	Sanger	Sequencing complete, assembly underway.

Question IV. The *H. influenzae* USS are not randomly distributed around the genome. Are the same patterns seen in other genomes with USS?

We have detailed information about USS abundance and location only for *H. influenzae*. We do not know which features or patterns are unique or arbitrary, and which reflect the forces responsible for the sequence-specific uptake system. If non-random spacing is not a general feature of USS in different genomes, we will not devote much time to seeking an explanation.

There will be four genomes to be analyzed (*A. actino*, *N. gonorrhoeae*, and two *N. meningitidis* strains); the work will be done with our collaborator Paul Lee, who has developed programs to analyze the frequencies of all possible oligomers in genome sequences. Analysis must wait for the complete assembly of each genome sequence, but all are likely to be completed by the end of 1999. The first step will be to completely characterize the USS in each genome—at present we have only sampled a small fraction of the genome. We will determine the consensus of both perfect and singly-mismatched USS; the latter is of particular value as it reveals the relative importance of each of the core bases, and ensures that we have not overlooked an abundant sequence related to the expected USS.

This analysis will also address a number of other questions. The ratios of perfect to singly-mismatched USS will be used in the model of bias-driven USS accumulation (Question II). The presence or absence of the USS in prophages and horizontally transferred genes will address the role of USS in reducing the risks of taking up foreign DNA. It will also lay the groundwork for analysis of the distribution and impact of USS in coding and non-coding regions in these species.

Question V. Most *H. influenzae* USS are in coding sequences. How do they constrain protein function?

To evaluate possible explanations for the abundance of USS, we need estimates of the selective cost of maintaining them in the genome. Most USS are in sequences with other known functions. This suggests that they are best viewed not as additions that increase the genome size by 2-3%, but as constraints on the functions of a genome of unchanged size. The most significant costs are likely to derive from the 65% of USS that are in coding regions.

Coding functions: The two preliminary analyses described above (Section C-f) suggest that many USS do not seriously interfere with the coding functions of the genome. Both analyses will be extended to much larger sample sizes, and the full range of organisms known to have USS. This can be done immediately for *H. influenzae*, using the FASTA file of all coding regions available from TIGR's website, but analysis of *A. actino*,

R. Redfield's NIH grant proposal (Funded Jan. 1 2000)

N. gonorrhoeae and *N. meningitidis* will wait until complete annotated sequences have been released (expected within a year).

Some questions to be answered are: Are USS unevenly distributed between coding and non-coding sequences? Are there preferred reading frames? Do they use common codons, and encode peptides likely to be tolerated in many locations? Does the frequency of these peptides in the proteome exceed the requirements of the existing USS? The 10 bp *Neisseria* USS will require more complex analysis than the 9 bp *H. influenzae* USS, because it necessarily will constrain but not absolutely specify one amino acid. If time permits we will include the flanking consensus sequences in the analysis of constraints in *H. influenzae* and *A. actino*. This is very important for the *H. influenzae* USS, because the flanking sequences cover 20 bp, but may be difficult to do as there will be many coding possibilities. We would expect to see an inverse relationship between the complexity of the analysis (the number of coding possibilities) and the expected degree of constraint on gene function, and will try to formalize this.

We will also extend the analysis of USS in conserved genes, comparing USS-encoded residues with the variation seen in homologous residues from genes without USS.

Question VI. Some related species with overlapping environments share a common USS. Has this led to substantial genetic exchange between them?

Once the complete genome sequences of *A. actino* and the three *Neisserias* are available, we will use complete genome by genome nucleotide comparisons to search for nucleotide sequences whose similarity is anomalously high. This will be much more effective than searching each genome for segments of anomalous base composition or codon usage (Lawrence et al. 1998).

If a large number of transferred genes is found, perhaps bacteria sharing a common USS belong to a 'species complex', where genetic exchange is frequent but does not break down species identities. Although the shared USS might be the cause of such a relationship, it seems unlikely to have been selected for that function. Selection for such could be responsible for the shared USS. Strongly arisen by direct selection, so such a finding is unlikely to explain the existence of USS. If no evidence of genetic exchange is found, we can conclude that there are other barriers preventing exchange between

One potential complication is that the species with sequenced genomes are not the only possible sources of transferred genes. *H. parainfluenzae*, *H. aphrophilus* and *H. paraphrophilus* all compete for uptake with *H. influenzae*, and have been found in normal oral flora. The *Neisseria* USS may also occur in species other than *gonorrhoeae* and *meningitidis*. (a number of normally-nonpathogenic *Neisseria* species share the oropharynx).

Part 3. Experimental Work:

Question VII. Uptake of heterologous DNA can kill the Rd strain of *H. influenzae* by inducing a resident prophage. Does this lethality occur in other strains?

The rightmost column in Fig. 3 shows the survival of competent *H. influenzae* exposed to high concentrations of heterologous DNAs. This cell death is potentially a source of strong selection against competence, but we do not know if it occurs in *H. influenzae* strains other than Rd.

We have 21 independently-isolated *H. influenzae* strains in our lab stocks, and 50 more are readily available from the American Type Culture Collection. We will initially test 10 transformable strains. Each will be treated with the standard competence-induction ritual (see Methods below), and tested each for transformation to antibiotic resistance by DNA of the multiply-resistant strain MAP7, and for killing after exposure to a high concentration of *H. parainfluenzae* or *H. aphrophilus* DNA. *H. influenzae* strain BC200 is a derivative of Rd that does not express the prophage responsible for the cell killing (Setlow et al. 1973); it will be used as a negative control strain (no death expected with the foreign DNAs). DNA from strain BC200 will also be used as the negative control DNA; unlike Rd DNA there will be no concern that zygotic induction of the prophage could cause death of non-lysogenic recipients.

If no lethality is observed in non-Rd strains, we will conclude that uptake of foreign DNA does not impose substantial costs. If uptake of foreign DNA does kill other strains, we will conclude that this is potentially a very substantial cost of taking up any foreign DNA. We will use the killing caused by these divergent USS-containing DNAs as an estimate of the fitness decrease that would be caused by taking up human or other unrelated DNAs if the DNA-uptake system was not specific for the USS. The answers to Questions VI and VIII will allow estimation of the actual benefits and costs of the USS-specific uptake system. Specifically, to the extent that cells efficiently take up USS-containing DNA from overlying mucus layers on agar plates (Question

VIII), the requirement for the USS is protecting them from the potentially-lethal effects of human DNA that the need for a USS limits uptake of human DNA in the respiratory tract. The extent of the evidence for USS-promoted genetic exchange between the USS-containing species (Question VI) will indicate the actual frequency of uptake of lethal USS-containing DNAs in the respiratory tract.

If time permits, we will also test *Neisseria* strains for similar effects.

Question VIII. Respiratory tract mucus contains high concentrations of DNA. Can DNA be taken up by *H. influenzae* growing in or under mucus?

H. influenzae cells normally grow in respiratory mucus or below the mucus layer in small colonies on the surface of epithelial cells (Hendrixson et al. 1998). We do not know whether the large amount of DNA in mucus can be taken up by *H. influenzae* cells, nor the extent to which DNA is released by lysis of some cells and taken up by others. We do know that *H. influenzae* colonies on agar plates do contain many competent cells. When colonies on agar are resuspended in medium containing marked chromosomal DNA, transformation frequencies are about 10^{-5} . We have used this competence as an assay for our hypercompetent mutants (Williams et al. 1994), but we do not know at what stage of colony growth the competence develops. Stuy found little exchange within mixed colonies on agar plates (Stuy 1985), suggesting that little DNA becomes available for uptake.

We will simulate more natural conditions by culturing cells in association with DNA-containing solutions of porcine mucin (readily available from Sigma). In one set of experiments, mixtures of genetically-marked cells will be cultured in semi-solid media supplemented with mucin, and single strains in mucin media supplemented with genetically-marked *H. influenzae* DNA. This will reveal the ability of cells to take up DNA from mucus. In other experiments, cells or mixtures of cells will be grown in thin lawns on the surface of agar plates, and overlaid with DNA-containing mucin, where appropriate supplemented with genetically marked DNA. We have no experience working with *H. influenzae* *in vivo*, so we will not attempt to carry out *in vivo* transformation experiments, although we will be alert for any potential collaborator who might be interested in doing them.

Question IX. The strength of the *H. influenzae* uptake bias is poorly defined. How strongly does DNA uptake depend on the USS?

If the USS functions primarily in DNA uptake, then we expect the sequence-specificity of uptake to match the USS consensus. Preliminary analysis of published uptake data is consistent with this hypothesis, but does not strongly test it. We will break this analysis down into several subsidiary questions.

How efficiently are imperfect USSs taken up? How strongly is uptake influenced by the conserved flanking sequences? The low frequency of singly-mismatched USSs in the genome suggests that these may not be efficiently taken up. This hypothesis is supported by the finding that an 8-mer USS mismatched at the first base is taken up only 1% as well as a 9-mer (Fitzmaurice et al. 1984); (we have checked the flanking sequences; they appear equivalent). As described in Section C-g, we will also use Goodgal's data to derive a matrix relating DNA uptake to USS sequence. However we need much more solid uptake data to evaluate the role of uptake-bias in USS abundance (Question II).

We will synthesize all 27 of the singly-mismatched USSs, each with optimal flanking sequences. (Because of the concentrations needed for these experiments, synthesis of 50 bp fragments is more economical than cloning or PCR amplification.) These will be tested both for uptake by competent cells and in a competition assay, where the ability of each to limit transformation with genetically-marked DNA will be measured, using our perfect USS as a positive control. The results in Section C-3 indicate that competition is a more sensitive measure of binding ability than is uptake. The results will be related to the base frequencies found in the singly-mismatched USSs (Fig. 2b of Smith et al. (B)), to see if these could be being maintained by uptake.

The analysis of core specificity must also be extended to the sequences flanking the USS. The rightmost of the two conserved AT-rich regions has not been directly implicated in DNA uptake, but its strong conservation would suggest that it plays an important role. This must be tested, because if the flanking sequences do not act in DNA uptake, they must be conserved because of an uptake-independent function. Furthermore, to determine the constraints USS impose on genome function, we need to assess the role of flanking sequences as well as the core USS. We will again use the perfect synthetic USS as control, and construct variants with the least-common base at each of the 12 conserved flanking positions. As above, these will be tested as competitors in transformation experiments as well as for uptake.

The results with synthetic USS will be confirmed and extended to longer fragments by testing uptake of PCR-amplified fragments of the *H. influenzae* genome. We will amplify fragments of 500 bp, selected to contain

R. Redfield's NIH grant proposal (Funded Jan. 1 2000)

USS differing from the consensus sequence at positions identified as important by the previous experiment. We will also compare fragments containing two or more USS to those with a single USS (relevant to Question III), and will investigate the effect of USS location within the fragment (central or terminal).

Does DNA uptake absolutely require a USS? It is often assumed that a DNA fragment cannot be taken up unless it carries a USS, but there is conflicting evidence on whether unrelated DNAs are taken up not at all, or only less efficiently. For example, Goodgal and Mitchell reported that, given enough time competent *H. influenzae* cells could take up as much *E. coli* DNA as *H. influenzae* DNA (Goodgal et al. 1984). We do not know whether this is because the uptake system will slowly take up DNAs that have no USS-like sequences, or because it must wait for the rare randomly-occurring USS-like sequences present in unrelated DNAs. This information is needed for predictions of the benefits of specific and non-specific uptake. The evolutionary benefits of an absolute requirement of USS are not obvious - this would be an extreme endpoint of the evolution of a biased uptake system.

One difficulty in interpreting past measurements is that the DNAs used were often undefined. We will measure uptake of DNAs of defined sequence. We will use radioactively-labeled linearized plasmids containing the synthetic USS and control sequences described above (USS-1 and USS-R). Because the experiments will require longer incubation times than usual, and might be confounded by the presence of extracellular nucleases or other factors allowing competence-independent uptake of label, we will repeat the experiments with DNAs labeled in several different ways.

Question X. The USS-binding structure on the surface of competent *H. influenzae* cells has not been identified. Can the USS be cross-linked to a specific protein or proteins on the surface of competent cells? If so, micro-sequencing will be used to identify the protein(s) and the corresponding gene(s).

No mutations directly affecting DNA binding or the first stage of uptake have been identified; the only known mutants unable to bind DNA have regulatory defects. Although there have been a number of reports of competence-specific proteins, most of the early work has not proved reproducible, and recent efforts to identify USS-binding proteins have not been successful. In 1989, J.-F. Tomb characterized proteins in the membranes of competent and non-competent cells, but found no USS-binding activity (Tomb, personal communication). H. O. Smith spent 1991 attempting to isolate the USS-binding protein from competent cells, with no success (Smith, personal communication). Consequently we are trying a completely different approach, one which exploits the USS consensus and the sequence of all *H. influenzae* proteins revealed by the genome sequence.

We are developing a high-affinity USS as a 'fishing lure', to crosslink to the USS-binding protein or proteins on the cell surface (see Section C-3). The USS-binding activity has not been reproducibly detected in cell lysates, so crosslinking will be done using intact competent cells. Competent cells will be incubated briefly with the synthetic USS, and exposed to a pulse of UV irradiation, causing direct crosslinks between DNA and protein. Dr. Michel Roberge of the Dept. of Biochemistry has experience with this technique; he is providing assistance and the use of his single-pulse UV Nd:YAG laser (Ho et al. 94).

Initially we are crosslinking a radiolabelled USS to cells, and using SDS-PAGE to determine whether a single protein or multiple proteins become crosslinked to the USS. If a single protein, we will recover the crosslinked protein by heating the cell lysate to separate the USS strands, and annealing the mixture with magnetic beads linked to a single USS strand. An alternative approach will be to use a USS with a single-stranded poly-A tail, and oligo-dT-linked beads. If the USS has become crosslinked to more than one protein, we will separate the protein fractions by preparative SDS-PAGE before annealing them to beads.

Microsequencing of the USS-linked protein or proteins will be used to obtain peptide sequence of sufficient length to unambiguously identify the corresponding gene or genes from the genome sequence. Mass-spectrometry-based sequencing will be most sensitive, and is available at UBC through the Nucleic Acid and Protein Sequencing (NAPS) Unit.

Finding the gene for the USS receptor would be a major breakthrough with relevance beyond the USS and pathogenesis, as the mechanism of DNA uptake remains a mystery. If the receptor gene is identified, it will be cloned and inactivated by cassette mutagenesis. Cells will be transformed with the inactivated gene, and their ability to take up DNA measured. If the receptor functions only in DNA uptake, the mutant cells should grow normally. Any effects of the mutation on non-competent cells would signal that the receptor has a function other than DNA uptake. We will also check the distribution of USS in the neighborhood of this gene; any anomalies will have implications for the role of the receptor in driving USS abundance and location.

Question XI. USS abundance may be primarily due to an intracellular function. Is there a cytoplasmic or inner membrane protein that specifically binds the *H. influenzae* USS?

If the USS has an intracellular function (i.e. a function other than DNA uptake), there must be an intracellular protein that binds it. The protein should be reasonably abundant if its role is structural (i.e. if most or all the USS are usually bound to it), and must be extremely sequence-specific, given the tight conservation of the perfect USS core relative to its mutated variants. There may also be cooperative binding with the DNA-binding protein HU; in *E. coli* this preferentially binds to sequences containing curved AAAAA stretches like those flanking the USS (Tanaka et al. 1993).

We are presently using Southwestern blotting to look for proteins capable of binding the USS (see Section C-3). We will also do bandshift assays, using our high-affinity synthetic USS. If no specific binding is seen in these assays, and if time permits, we will make a phage display library and screen it with the USS probe (Zwick et al. 1998). Jamie Scott of the nearby Simon Fraser University Institute of Molecular Biology has offered assistance.

We will also test whether USS in nucleoids are in a structure that protects them from digestion by added nuclease. Cells will be gently permeabilized in the presence of spermidine to preserve nucleoid structure, and incubated with sufficient DNase to digest the bulk of the DNA into fragments smaller than 10 bp (Murphy et al. 1997). The surviving DNA will be purified, and end-labeled. It will then be denatured and incubated with magnetic beads linked to the + and - USS strands, and the amount of bound label measured. Control DNA will be total genomic DNA digested with a combination of restriction enzymes to an average size of 30-50 bp.

Identification of an intracellular protein that binds the USS would provide strong evidence that the USS does not function solely in DNA uptake. If a protein is not homologous to any proteins of known function, our next step would be to clone the corresponding gene, inactivate it by cassette mutagenesis, and determine its role in the cell. One particularly interesting possibility is that the same protein could function within the cell and in DNA uptake, being exposed on the cell surface only when cells are competent. A negative result would be less conclusive, indicating only that any USS-binding protein must be of low abundance or low affinity.

Why no laboratory-based selection experiments are proposed:

We seek evidence of how selection has acted in the real world. Laboratory selection experiments can easily be misleading, especially for microorganisms, because culture conditions for pathogens necessarily differ so much from natural conditions, and because we rarely even know what the relevant natural growth and survival conditions are, at the microscopic scale that these organisms live. Furthermore, the conditions used for selection experiments are often contrived to match a theoretical model - in effect using the organism to simulate the model rather than exploring the roles of natural factors. Thus we feel that real sequences and real mechanisms are our best evidence of how selection *has* acted (as opposed to lab experiments, which show how selection *can* act). Laboratory selection experiments are often further compromised by use of the standard laboratory culture conditions rather than those that might approximate natural ones, and even our best attempts to mimic natural conditions are unlikely to be adequate, simply because we have so little understanding of which factors are relevant.

Other questions to be addressed if time and resources permit

The research questions proposed above have been limited to those most likely to be productive, and will most likely require the full five years requested. Many other questions have not been explicitly targeted, and will be considered as appropriate. Examples are: Does the USS have a direct function in processing or recombination of incoming DNA? Have the *Haemophilus* and *Neisseria* systems evolved independently? Are they maintained by the same forces? Are USS common on genetic parasites of *Haemophilus* and *Neisseria*, or is the *H. influenzae* phage HP1 an exception? Does DNA methylation play any role in uptake specificity?

Future directions

Because there has been no previous exploration of USS evolution, it is difficult to predict which investigations will be most productive. The future work will be determined by the answers we find to the questions posed above. For example, if modeling shows nucleotide acquisition to be the only large and reliable benefit of DNA uptake, we will use both modelling and experiments to explore conditions favoring biased uptake; if many different strains are harmed by taking up foreign DNA, we will determine whether a significant amount of homologous DNA is available in colonies; if an intercellular USS-binding protein is found, its relationship to the biased uptake system will be explored. This project is only the beginning.

Methods:

Computer simulations will be written in Pascal or C++, and run on Macintosh computers. We also have Mathematica, Matlab and Maple software. Genomic analysis will use MacVector software and GCG; the latest release of MacVector interacts directly with GenBank via the Entrez interface, and will export data in GCG format. Both Francis Ouellette (formerly with the Microbial Genomes division of NCBI, and now at UBC's Centre for Molecular Medicine and Therapeutics) and Christoph Sensen (Director of the *Sulfolobus* genome project) are available to guide us; letters of support are attached. Genetic methods used for *H. influenzae* are described in Redfield (1999), provided as Appendix item J. An older but more detailed general reference for *H. influenzae* culture and genetic analysis is (Barcak et al. 1991) *Haemophilus influenzae* strains: We have a substantial collection of *H. influenzae* competence mutants and some clinical isolates. Additional strains will be obtained from the American Type Culture Collection.

References:

- Albritton, W., Setlow, J., Thomas, M. and Sottnek, F. (1986). Relatedness within the family *Pasteurellaceae* as determined by genetic transformation. *Int. J. Syst. Bact.* **36**: 103-106.
- Albritton, W. L., Setlow, J. K., Thomas, M., Sottnek, F. and Steigerwalt, A. G. (1984). Heterospecific transformation in the genus *Haemophilus*. *Mol. Gen. Genet.* **193**: 358-63.
- Barany, F., Kahn, M. E. and Smith, H. O. (1983). Directional transport and integration of donor DNA in *Haemophilus influenzae* transformation. *PNAS* **80**: 7274-8.
- Barcak, G. J., Chandler, M. S., Redfield, R. J. and Tomb, J.-F. (1991). Genetic systems in *Haemophilus influenzae*. *Methods in Enzymol.* **204**: 321-342.
- Barouki, R. and Smith, H. O. (1985). Reexamination of phenotypic defects in *rec-1* and *rec-2* mutants of *Haemophilus influenzae* Rd. *J. Bact.* **163**: 629-34.
- Barouki, R. and Smith, H. O. (1986). Initial steps in *Haemophilus influenzae* transformation. Donor DNA binding in the com10 mutant. *J. Biological Chemistry* **261**: 8617-23.
- Beattie, K. L. and Setlow, J. K. (1969). Killing of *Haemophilus influenzae* cells by integrated ultraviolet-induced lesions from transforming deoxyribonucleic acid. *J. Bact.* **100**: 1284-8.
- Biswas, G. D., Thompson, S. A. and Sparling, P. F. (1989). Gene transfer in *Neisseria gonorrhoeae*. *Clinical Microbiology Reviews* :
- Boling, M. E., Setlow, J. K. and Allison, D. P. (1972). Bacteriophage of *Haemophilus influenzae*. I. Differences between infection by whole phage, extracted phage DNA and prophage DNA extracted from lysogenic cells. *J. Molecular Biology* **63**: 335-48.
- Booy, R. and Kroll, J. (1998). Bacterial meningitis and meningococcal infection. *Curr Opin Pediatr* **10**: 13-8.
- Caugant, D. A., Mocca, L. F., Frasch, C. E. et al. (1987). Genetic structure of *Neisseria meningitidis* populations in relation to serogroup, serotype, and outer membrane protein pattern. *J. Bact.* **169**: 2781-92.
- 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.
- Clark, A. J. (1971). Toward a metabolic interpretation of genetic recombination of *E. coli* and its phages. *Ann Rev Microbiol* **25**: 437-64.
- Cohan, F. M. (1996). The role of genetic exchange in bacterial evolution. *ASM News* **62**: 631-636.
- Collado-Vides, J., Magasanik, B. and Gralla, J. D. (1991). Control site location and transcriptional regulation in *Escherichia coli*. *Microbiol. Rev.* **55**: 371-394.
- Danner, D., Smith, H. O. and Narang, S. (1982). Construction of DNA recognition sites active in *Haemophilus transformation*. *PNAS* **79**: 2393-2397.
- Danner, D. B., Deich, R. A., Sisco, K. L. and Smith, H. O. (1980). An eleven-base-pair sequence determines the specificity of DNA uptake in *Haemophilus transformation*. *Gene* **11**: 311-8.
- Dewhirst, F., Paster, B., Olsen, I. and Fraser, G. (1993). Phylogeny of the Pasteurellaceae as determined by comparison of 16S ribosomal ribonucleic acid sequences. *Int J Med Micro. Virol Parasi. Inf. Dis* **279**: 35-44.
- Doern, G. V., Brueggemann, A. B., Pierce, G., Holley, H. P., Jr. and Rauch, A. (1997). Antibiotic resistance among clinical isolates of *Haemophilus influenzae* in the United States in 1994 and 1995 and detection of beta-lactamase- positive strains resistant to amoxicillin-clavulanate: results of a national multicenter surveillance study. *Antimicrob Agents Chemother* **41**: 292-7.

R. Redfield's NIH grant proposal (Funded Jan. 1 2000)

- Dorocicz, I., Williams, P. and Redfield, R. J. (1993). The *Haemophilus influenzae* adenylate cyclase gene: cloning, sequence and essential role in competence. *J. Bact.* **175**: 7142-7149.
- Dreiseikelmann, b. (1994). Translocation of DNA across bacterial membranes. *Micro. Rev.* **58**: 293-316.
- Dubnau, D. (1991). Genetic competence in *Bacillus subtilis*. *Microbiol. Rev.* **55**: 395-424.
- Elena, S. F. and Lenski, R. E. (1997). Test of synergistic interactions among deleterious mutations in bacteria. *Nature* **390**: 395-8.
- Esposito, D., Fitzmaurice, W. P., Benjamin, R. C. et al. (1996). The complete nucleotide sequence of bacteriophage HP1 DNA. *Nucleic Acids Res* **24**: 2360-8.
- Fitzmaurice, W., Benjamin, R., Huang, P. & Scocca, J. (1984). Characterization of recognition sites on bacteriophage HP1c1 DNA which interact with the DNA uptake system of *H. influenzae*. *Gene* **31**: 187-96.
- Fleischmann, R. D., Adams, M. D., White, O. et al. (1995). Whole-Genome Random Sequencing and Assembly of *Haemophilus influenzae* Rd. *Science* **269**: 496-512.
- Fox, K., Knapp, J., Holmes, K. et al. (1997). Antimicrobial resistance in *Neisseria gonorrhoeae* in the USA, 1988-1994: the emergence of decreased susceptibility to the fluoroquinolones. *J Infect Dis* **175**: 1396-403.
- Frosch, M. and Meyer, T. F. (1992). Transformation-mediated exchange of virulence determinants by co-cultivation of pathogenic *Neisseriae*. *Fems Micro. Letters* **79**: 345-9.
- Goodgal, S. H. (1982). DNA uptake in *Haemophilus* transformation. *Ann. Rev. Genet.* **16**: 169-92.
- Goodgal, S. and Mitchell, M. (1984). Uptake of heterologous DNA by *H. influenzae*. *J. Bact.* **157**: 785-8.
- Goodgal, S. H. and Mitchell, M. A. (1990). Sequence and uptake specificity of cloned sonicated fragments of *Haemophilus influenzae* DNA. *J. Bact.* **172**: 5924-8.
- Goodman, S. D. and Scocca, J. J. (1988). Identification and arrangement of the DNA sequence recognized in specific transformation of *Neisseria gonorrhoeae*. *PNAS* **85**: 6982-6.
- Graves, J. F., Biswas, G. D. and Sparling, P. F. (1982). Sequence-specific DNA uptake in transformation of *Neisseria gonorrhoeae*. *J. Bact.* **152**: 1071-7.
- Griffith, F. (1928). The significance of pneumococcal types. *J. Hygiene* **27**: 113-159.
- Hendrixson, D. R. and St Geme, J. W., 3rd (1998). The *Haemophilus influenzae* Hap serine protease promotes adherence and microcolony formation, potentiated by a soluble host protein. *Mol Cell* **2**: 841-50.
- Herriott, R. M., Meyer, E. M. and Vogt, M. (1970). Defined nongrowth media for stage II development of competence in *Haemophilus influenzae*. *J. Bact.* **101**: 517-24.
- Ho, D., Sauve, D., and M. Roberge (1994) Detection and isolation of DNA-binding proteins using single-pulse ultraviolet laser crosslinking. *Anal. Biochem.* **218**: 248-254.
- Hoelzer, M. and Michod, R. (1991). DNA repair and the evolution of transformation in *Bacillus subtilis*: III. Sex with damaged DNA. *Genetics* **128**: 215-223.
- Ippen-Ihler, K. A. and Minkley, E. G., Jr. (1986). The conjugation system of F, the fertility factor of *Escherichia coli*. *Ann. Rev. Genet* **20**: 593-624.
- Jyssum, S. and Jyssum, K. (1970). Specific uptake of homologous DNA accompanying transformation in *Neisseria meningitidis*. *Acta Pathol Microbiol Scand [B] Microbiol Immunol* **78**: 140-8.
- Kahn, M. E., Barany, F. and Smith, H. O. (1983). Transformasomes: specialized membranous structures that protect DNA during *Haemophilus* transformation. *PNAS* **80**: 6927-31.
- Karlin, S., Mrazek, J. and Campbell, A. M. (1996). Frequent oligonucleotides and peptides of the *Haemophilus influenzae* genome. *Nucleic Acids Res* **24**: 4263-72. (provided as Appendix item C)
- Karudapuram, S. and Barcak, G. (1997). The *H. influenzae* *dprABC* genes constitute a competence- inducible operon that requires the product of the *tfoX* (*sxy*) gene for transcriptional activation. *J Bact.* **179**: 4815-20.
- Kondrashov, A. (1993). Classification of hypotheses on the advantage of amphimixis. *J. Hered.* **84**: 372-387.
- Kroll, J. S. (1992). The genetics of encapsulation in *Haemophilus influenzae*. *J. Inf. Dis.* **165**: S93-6.
- Kroll, J. S., Wilks, K. E., Farrant, J. L. and Langford, P. R. (1998). Natural genetic exchange between *Haemophilus* and *Neisseria*: intergeneric transfer of chromosomal genes between major human pathogens. *PNAS* **95**: 12381-5.
- Kushner, S. R. (1987). DNA repair. pp. 1044-1053 in *Escherichia coli* and *Salmonella typhimurium* Ed. F. Neidhardt. Washington DC, American Society for Microbiology..

R. Redfield's NIH grant proposal (Funded Jan. 1 2000)

- Larson, T. G. and Goodgal, S. H. (1991). Sequence and transcriptional regulation of com101A, a locus required for genetic transformation in *Haemophilus influenzae*. *J Bacteriol* **173**: 4683-91.
- Lawrence, J. & Ochman, H. (1998). Molecular archaeology of the *Escherichia coli* genome. *PNAS* **95**: 9413-7.
- Lederberg, J. and Tatum, E. L. (1946). Gene recombination in *Escherichia coli*. *Nature* **158**: 558.
- Lethem, M. I., James, S. L., Marriott, C. and Burke, J. F. (1990). The origin of DNA associated with mucus glycoproteins in cystic fibrosis sputum. *European Respiratory Journal* **3**: 19-23.
- Lorenz, M. G. and Wackernagel, W. (1994). Bacterial gene transfer by natural genetic transformation in the environment. *Micro. Rev.* **58**: 563-602.
- Macfadyen, L. P., Dorocicz, I. R., Reizer, J., Saier, M. H., Jr. and Redfield, R. J. (1996a). Regulation of competence development and sugar utilization in *Haemophilus influenzae* Rd by a phosphoenolpyruvate:fructose phosphotransferase system. *Mol Microbiol* **21**: 941-52.
- Macfadyen, L. P., and Redfield, R. J. (1996b). Life in mucus: sugar metabolism in *Haemophilus influenzae*. *Res. Microbiol.* **147**: 541-51 (provided as Appendix item G)
- Matthews, L., Spector, S., Lemm, J. and Potter, J. (1963). Studies on pulmonary secretions. 1. The overall chemical composition of pulmonary secretions from patients with cystic fibrosis, bronchiectasis and laryngectomy. *Am. Rev. Respir. Dis.* **88**: 199-204.
- Maynard Smith, J., Smith, N., O'Rourke, M. and Spratt, B. (1993). How clonal are bacteria? *PNAS* **90**: 4384-4388.
- Michod, R. E., Wojciechowski, M. and Hoelzer, M. (1988). DNA repair and the evolution of transformation in the bacterium *Bacillus subtilis*. *Genetics* **118**: 31-39.
- Milkman, R. and Bridges, M. (1990). Molecular evolution of the *Escherichia coli* chromosome. III. Clonal frames. *Genetics* **126**: 505-517.
- Morel, P., Cherny, D., Ehrlich, S. D. and Cassuto, E. (1997). Recombination-dependent repair of DNA double-strand breaks with purified proteins from *Escherichia coli*. *J Biol Chem* **272**: 17091-6.
- Murphy, L. D., and Zimmerman, S. B. (1997) Stabilization of compact spermidine nucleoids from *Escherichia coli* under crowded conditions: implications for in vivo nucleoid structure. *J. Struct. Biol.* **119**:336-46
- Musser, J. M., Kroll, J. S., Moxon, E. R. and Selander, R. K. (1988). Clonal population structure of encapsulated *Haemophilus influenzae*. *Infection & Immunity* **56**: 1837-45.
- Myers, R. and Stahl, F. (1994). Chi and the RecBC D enzyme of *Escherichia coli*. *Ann.Rev Genet* **28**: 49-70.
- Pan, A., Dutta, C. and Das, J. (1998). Codon usage in highly expressed genes of *Haemophilus influenzae* and *Mycobacterium tuberculosis*: translational selection versus mutational bias. *Gene* **215**: 405-13.
- Pifer, M. L. and Smith, H. O. (1985). Processing of donor DNA during *Haemophilus influenzae* transformation: analysis using a model plasmid system. *PNAS* **82**: 3731-5.
- Postel, E. H. and Goodgal, S. H. (1966). Uptake of "single-stranded" DNA in *Haemophilus influenzae* and its ability to transform. *J. Molecular Biology* **16**: 317-27.
- Redfield, R. J. (1988). Evolution of bacterial transformation: is sex with dead cells ever better than no sex at all? *Genetics* **119**: 213-21. (provided as Appendix item E)
- Redfield, R. J. (1991). Bacterial Mating Preferences. *Nature* **352**: 25-26. (provided as Appendix item I)
- Redfield, R. J. (1993a). Evolution of natural transformation: testing the DNA repair hypothesis in *Bacillus subtilis* and *Haemophilus influenzae*. *Genetics* **133**: 755-761. (provided as Appendix item F)
- Redfield, R. J. (1993b). Genes for breakfast: the have your cake and eat it too of transformation. *J. Hered.* **84**: 400-404. (provided as Appendix item D)
- Redfield, R. J., Schrag, M. and Dean, A. (1997). The evolution of bacterial transformation: sex with poor relations. *Genetics* **146**: 27-38. (provided as Appendix item A)
- Redfield, R. J. (1999). *Haemophilus influenzae* Genetics. To appear in Encyclopedia of Microbiology, 2nd Ed. (provided as Appendix item J)
- Saez-Nieto, J. A., Lujan, R., Martinez-Suarez, J. V. et al. (1990). *Neisseria lactamica* and *Neisseria polysaccharea* as possible sources of meningococcal beta-lactam resistance by genetic transformation. *Antimicrob Agents Chemother* **34**: 2269-72.
- Setlow, J. K., Boling, M. E., Allison, D. P. and Beattie, K. L. (1973). Relationship between prophage induction and transformation in *Haemophilus influenzae*. *J. Bact.* **115**: 153-61.

R. Redfield's NIH grant proposal (Funded Jan. 1 2000)

- Smith, H. O., Tomb, J.-F., Dougherty, B., Fleischmann, R. and Venter, J. (1995). Frequency and distribution of DNA uptake signal sequences in the *Haemophilus influenzae* Rd genome. *Science* **269**: 538-540.
(provided as Appendix item B)
- Sourice, S., Biaudet, V., El Karoui, M., Ehrlich, S. D. and Gruss, A. (1998). Identification of the Chi site of *Haemophilus influenzae* as several sequences related to the *Escherichia coli* Chi site. *Mol Micro.***27**: 1021-9.
- Spratt, B. G., Bowler, L. D., Zhang, Q. Y., Zhou, J. and Smith, J. M. (1992). Role of interspecies transfer of chromosomal genes in the evolution of penicillin resistance in pathogenic and commensal *Neisseria* species. *J. Molecular Evolution* **34**: 115-25.
- Stormo, G. and Hartzell, G. (1989). Identifying protein-binding sites from unaligned DNA fragments. *PNAS* **86**: 1183-7.
- Stouthamer, A. H. (1979). The search for correlation between theoretical and experimental growth yields. *Int. Rev. Biochem.* **21**: 1-47.
- Stuy, J. H. (1985). Transfer of genetic information within a colony of *H. influenzae*. *J. Bact.* **162**: 1-4.
- Stuy, J. H. and Walter, R. B. (1981). Addition, deletion, and substitution of long nonhomologous deoxyribonucleic acid segments by genetic transformation of *Haemophilus influenzae*. *J. Bact.* **148**: 565-71.
- Suerbaum, S., Smith, J. M., Bapumia, K. et al. (1998). Free recombination within *Helicobacter pylori*. *Proc Natl Acad Sci U S A* **95**: 12619-24.
- Tanaka, H., Goshima, N., Kohno, K., Kano, Y. and Imamoto, F. (1993). Properties of DNA-binding of HU heterotypic and homotypic dimers from *Escherichia coli*. *J Biochem (Tokyo)* **113**: 568-72.
- Wilcox, K. W. and Smith, H. O. (1975). Isolation and characterization of mutants of *Haemophilus influenzae* deficient in an adenosine 5'-triphosphate-dependent deoxyribonuclease activity. *J. Bact.* **122**: 443-53.
- Williams, P. M., Bannister, L. A. and Redfield, R. J. (1994). The *Haemophilus influenzae* *sxy-1* Mmutation is in a newly identified gene essential for competence. *J. Bact.* **176**: 6789-6794.
- Wojciechowski, M., Hoelzer, M. and Michod, R. E. (1989). DNA repair and the evolution of transformation in *Bacillus subtilis*. II. Role of inducible repair. *Genetics* **121**: 411-422.
- Zwick, M., Shen, J., and Scott, J. (1998) Phage-displayed peptide libraries. *Curr. Opin. Biotechnol.* **9**:427-36