F. DISCUSSION

1. Pathogenicity island

The recently described *cag* pathogenicity island of *H. pylori* is unique in that it appears not to code for a specific transcriptional regulator that directs the expression of the over 30 *cag* genes in a coordinate manner. Moreover, only few and often contradictory data are available regarding the transcriptional response of certain genes of the pathogenicity island to changes in environmental conditions. To gain insight into the basic mechanisms of transcriptional regulation of the *cag* pathogenicity island, the functional DNA elements required for transcription of the *cagA* and *cagB* genes were determined. The *cagA* gene codes for an immunodominant antigen, expressed in a subset of strains, associated with severe diseases (Covacci *et al*., 1993; Weel *et al*., 1996), whereas the *cagB* gene codes for a small protein of unknown function which is transcribed in the opposite orientation to *cagA*. The promoters of these genes were expected to be divergently located within the 436 bp intergenic region (Censini *et al*., 1996).

Nucleotide sequence analysis of the DNA region upstream of the transcription start points of the *cagA* and *cagB* genes revealed the presence of -10 hexamers with striking similarities to the consensus sequence recognised by $E\sigma^{70}$, the major RNA polymerase in *E. coli.* For the P_1 promoter no obvious -35 box could be found, while the P_2 and P_3 promoters exhibited -35 hexamers with some similarity (50%) to the $E\sigma^{70}$ consensus sequence (Fig. 3B). Promoters with highly conserved -10 regions but poorly conserved or missing -35 regions have been reported for other pathogens (Bashyam *et al*., 1996; Sarkari *et al*., 1994). For example, nucleotide sequence alignment of 14 promoters from *Mycobacterium smegmatis* and 10 promoters from *M. tubercolosis* revealed that the sequences in the -35 regions do not bear any homology with the *E. coli* promoters, and that apparently no single sequence is conserved in the -35 regions of these promoters (Bashyam *et al*., 1996). Expression of the Opc outer membrane protein in *Neisseria meningitidis* is regulated by a putative promoter with a -10 region similar to the *E. coli* promoters and a -35 region substituted by 11-14 cytidine residues (Sarkari *et al*., 1994). The features of the *H. pylori* $P_{1\text{-}3}$ promoters resemble the above described promoters. However, a closer inspection of the sequences immediately upstream of the -10 hexamers of the *H. pylori* P_2 promoter revealed the presence of a TGn motif (Kumar *et al*., 1993). This motif has been previously described as an extended -10 region which, in the case of *E. coli* promoters exhibiting the -10 consensus sequence but lacking a functional -35 region, is necessary to promote transcription by $E\sigma^{70}$ (Kumar *et al*., 1993). The P_1 promoter sequence shows a variant motif, TGnTG, located immediately upstream of the -10 region. Although not yet supported by direct experimental data, it is hypothesised that in the promoters of the *H.*

pylori pathogenicity island these extended -10 motifs compensate for the missing (P_1) or poorly matching (P_2) -35 region. Moreover, although with different specificity, all three promoters could be activated in *E. coli* and by $E\sigma^{70}$ *in vitro* (Fig. 4), indicating that their nucleotide sequences bear information for the major RNA polymerases from *H. pylori* and *E. coli*. The major subunit of the *H. pylori* RNA polymerase has recently been cloned and a detailed functional analysis has shown that it shares sequence recognition specificities with its *E. coli* counterpart (Beier *et al*., 1998). Furthermore, *in vitro* analyses have demonstrated that this RNA polymerase is able to transcribe efficiently from the P_1 promoter of *cagA*. Promoters of pathogenicity islands may have been selected during evolution in order to be recognised by many prokaryotic major RNA polymerases. This selection may represent an advantage if a pathogenicity island is lost by one bacterium and acquired by another one.

It has been demonstrated that the activity of the P_1 promoter is not influenced by the presence or absence of sequences located upstream of position -80 (Fig.s 8 and 9). Therefore, transcription from this promoter is not affected by the presence of the upstream divergently transcribing P_2 promoter (Fig.s 9A and 8A, mutant cag2). In contrast, full activation of the P_2 promoter requires sequences that overlap with the -50 region of the P_1 promoter (Fig. 9B and 8B, mutant cagX1). Intriguingly, deletion of the -50 region of P_1 exerted a positive effect on P_3 (Fig.s 9B and 8B). Therefore, activity of the P_2 and P_3 promoters is influenced negatively and positively, respectively, by sequences strictly belonging to, and essential for, full activation of the P_1 promoter. This suggests a complex molecular mechanism of transcriptional regulation of the three promoters.

Deletions of sequences upstream of position -13 of the P_1 promoter in the *H. pylori* chromosome induced a defined change of the transcriptional initiation to position +8 (Fig. 9A, mutants cagX1-3). By nucleotide sequence analysis of these promoter mutants, the presence of DNA elements creating a new promoter sequence overlapping P_1 can be excluded. Remarkably, when transcription from these promoter mutants was assayed in *E. coli* and *in vitro*, RNA synthesis initiated at position +1, suggesting that transcriptional initiation is affected by the *H. pylori* promoter DNA context. The sequence between positions -13 and -20 may contribute to correct initiation by providing a suitable DNA context, thus preventing interference between possible upstream DNA structures and the RNA polymerase. Based on the observation that the DNA under study has a high A/T content (~67%), and that nucleoid-associated proteins bind preferentially to A/T rich DNAs, it is tempting to speculate that promoter context could depend on DNA-nucleoid protein interactions. This speculation could be sustained by studies that conclusively established that a class of global regulators of transcription modulates gene expression by binding to A/T rich DNAs with defined structures such as bends, or by inducing dramatic changes into the DNA structure upon binding (reviewed by Werner and Burley, 1997). Examples of these architectural proteins include the H-NS protein which binds strongly to A/T rich and preferentially includes bent DNA (Atlung and Ingmer, 1997) and the IHF protein which binds to DNA inducing strong bends of 160°-180° (Werner and Burley, 1997, and references therein). Sequencing of the *H. pylori* genome revealed only one putative nucleoprotein (designated HP835 in Tomb *et al*., 1997) with similarities

to both HU (34% identity in a 91 aa overlap) and IHF (36% identity in a 74 aa overlap) from *E. coli*. Although direct experimental evidence is missing it seems likely that this protein participates in regulation of *cagAB* transcription by providing a suitable DNA context for RNA polymerase binding.

In agreement with the deletion analyses carried out in the *H. pylori* chromosome, DNase I footprinting experiments have shown that the RNA polymerase of *E. coli* binds the DNA region from +14 to -60 of the P_1 promoter. This promoter region includes a stretch with similarities to promoter UP elements, which are constituted by A/T rich regions of 20 bp spanning from -40 to -60 and containing a possible non-perfect inverted repeat. It has been demonstrated that these UP elements can bind the C-terminal domain of the α subunit of RNA polymerase, and thereby enhance transcription severalfold (Attey *et al*., 1994; Landini and Volkert 1995; Ross *et al*., 1993). In the case of the *rrnB* P1 promoter of *E. coli*, interaction of α with the UP element enhances promoter activity by a factor of about 30 (Ross *et al*., 1993). It has been demonstrated that the C-terminal region of the α subunit binds the *cagA* P_1 UP-like element, protecting DNA positions -14 to -70 from DNaseI digestion (Fig. 10). This protected region is larger as compared to other protected UP elements and could be due to the particularly high A/T content of this promoter. However, by combining these data and the results obtained *in vivo* by deletion analysis, it can be concluded that full activation of the P_1 promoter requires the interaction of the α subunit with the UP-like element. In addition, transcription from the $P_{1\text{-}3}$ promoters may be regulated by global regulatory factors such as the HU/IHF-like protein that are likely to bind within the intergenic region, thus altering local DNA structures.

2. Flagella

Motility has been shown to be a key factor for the ability of *H. pylori* to colonise the gastric mucosa (Eaton *et al*., 1996). While a few structural components of the flagellar apparatus (Leying *et al*., 1992; Suerbaum *et al*., 1993; O'Toole *et al*., 1994; Schmitz *et al*., 1997; Porwollik *et al*., 1999) and the flagellar motor switch protein CheY (Beier *et al*., 1997) have been characterised in some detail, little is known about factors that regulate expression of genes involved in motility and chemotaxis.

In chapter E.2 of the present thesis the flagellar regulon of *H. pylori* is identified, whose transcription is under the control of the alternative sigma factor σ^{54} and the transcriptional activator FlgR. Evidence demonstrating that FlgR and σ^{54} regulate the basal-body and hook genes lies in the finding that σ^{54} is unable to activate transcription of the regulon in a $FlgR^-$ background. Conversely, *flaA*, the gene encoding the major flagellin, is regulated by the alternative sigma factor σ^{28} (Leying *et al*., 1992). Interestingly, transcription from the *flaA* promoter (P_{601}) is enhanced in the $FlgR^-$ background, suggesting a negative feedback exerted by FlgR on transcription of this σ^{28}-regulated promoter (Fig. 17). This, in turn, implies that in the context of flagellar gene expression, the

efficiency of σ^{28}-dependent transcription is dependent on σ^{54} holoenzyme transcription. Since transcription of the two flagellin genes *flaA* and *flaB* is regulated by two different σ factors (σ^{28} and σ^{54}, respectively), these genes may be differentially expressed, depending on environmental conditions. It is likely that differential expression of distinct flagellin subunits within the flagellum may enable *H. pylori* to produce flagella, which are particularly suited for motility within a given environment (high-viscosity mucus, low or high osmolarity or pH etc.).

Transcription from the *flaA* promoter was unchanged by treatment of the cells with novobiocin, an inhibitor of bacterial Gyrases, while transcription from the *flaB* promoter changed with time after drug treatment (Fig. 18). Following addition of novobiocin to the culture medium transcription of the *flaB* gene was first enhanced and subsequently reduced, thus suggesting that a perturbation of the local density of DNA supercoil may modulate transcription of the *flaB* gene. This is in agreement with the features highlighted by studies conducted on the NtrC protein from *E. coli* (Brahms *et al*., 1995). In this case, it has been reported that the binding site of NtrC can be substituted by a DNA element possessing an intrinsic supercoil structure. Consequently, expression of flagellar genes may depend on the supercoil state of some promoters which in turn may depend upon environmental conditions. In this context it should be noted that *flgR* is cotranscribed with *gyrA*, the gene encoding subunit A of DNA Gyrase, and that both genes are coregulated by a single promoter (Fig. 12, panels A, C). Therefore, it is tempting to speculate that changes in the expression of genes involved in regulation of DNA supercoiling might coordinately change expression of flagellar genes and vice versa. In support of such a hypothesis is also the finding that the FlgR-regulated *flaB* promoter overlaps the divergently transcribing promoter of *topA*, the gene encoding Topoisomerase I. This may suggest a coordinate regulation of both genes in response to the same environmental stimuli. Recently, *in vitro* phosphorylation studies with a series of purified signal-transducing histidine-kinase sensor proteins and putative response regulators of *H. pylori* have shown that FlgR is efficiently phosphorylated by the histidine-kinase encoded by *hp244* (Table 1; D. Beier, personal communication). Interestingly, unlike most other sensor-kinases, that reside in the cytoplasmic membrane, the protein encoded by *hp244* is predicted to be located in the cytoplasm. Possibly, this FlgR-specific kinase responds to the same cytoplasmic signals that alter also the supercoiling state of DNA, coordinating in this way the pathways of flagellar synthesis and DNA topology.

In *E. coli*, *Salmonella* spp., and *Caulobacter crescentus*, the regulation of motility and chemotaxis has been extensively studied (reviewed in Amsler and Matsumura, 1995; Eisenbach, 1996; Macnab, 1995; Wu and Newton, 1997). In these organisms, the genes required for flagellar biosynthesis are sequentially expressed according to a hierarchical pathway (Fig. 27). In enterobacteriaceae, three classes of flagellar biosynthetic genes have been identified. Environmental signals trigger expression of class I (early) genes, which encode two master regulatory proteins (FlhC/FlhD). These regulatory proteins are required for $E\sigma^{70}$-dependent expression of class II (middle) genes, which encode the components of the basal body and the hook. Assembly of the basal body-hook complex in turn acts as a signal for $E\sigma^{28}$-regulated expression of class III (late)

genes, including the genes for chemoreception and the flagellin gene, by allowing the release of the anti-sigma factor FlgM (Wu and Newton, 1997). In contrast, four classes of flagellar biosynthetic genes have been identified in *Caulobacter* (Wu and Newton, 1997). An early signal in the cell cycle is assumed to trigger activation of the class I gene product, a transcriptional regulator (CtrA), which in turn activates σ^{70}-dependent transcription of class II genes encoding structural components of the MS ring-switch complex. Among class II genes are also the alternative sigma factor σ^{54} and its cognate transcriptional regulator FlbD which activate expression of class III and class IV genes, encoding structural proteins of the basal body-hook complex and flagellar filament, respectively. In contrast to the situation in enterobacteriaceae, two different checkpoints in the assembly of the flagellar structure control activation of these late flagellar genes. Completion of the MS ring-switch complex acts as a signal for transcription of class III genes, while completion of the basal body-hook complex is required for activation of class IV genes.

The data presented here suggest that the transcriptional hierarchy that governs flagellar synthesis in *H. pylori* has similarities to both systems. As in *Caulobacter*, σ^{54} and a cognate transcriptional activator (FlgR) are required for transcription of genes coding for structural proteins of the basal body and hook. In contrast, expression of the major flagellin (FlaA) appears to be regulated by σ^{28}, thus reflecting the situation observed in enterobacteriaceae. In contrast to both systems, no checkpoints in flagellar assembly that regulate transcription of flagellin genes are obvious in *H. pylori*. In fact, disruption of the *flgR* gene, and the resulting defect in expression of basal body and hook genes, does not prevent FlaA synthesis. Thus, it seems unlikely that an anti-sigma factor protein blocks transcription of σ^{28}-dependent genes in the absence of a functional basal body-hook complex in *H. pylori*. In support of this hypothesis, analysis of the complete genome sequence has revealed no genes coding for proteins with significant homology to an anti-sigma factor protein (Tomb *et al*., 1997). Nevertheless, FlgR exerts a slightly similar effect by down-modulating transcription of the *flaA* gene (Fig. 17).

Interestingly, transcription of the *flgR* gene is under the control of a σ^{80}-dependent promoter (Fig. 12, panels A,C). In fact, the structure of this promoter resembles the structure of the P_1 promoter of *cagA* (Fig. 3B), which is recognised by σ^{80} and regulated through the interaction of the AT rich upstream element with the α subunit of RNA polymerase (see E.1). Searching for such types of promoters upstream of the *H. pylori* flagellar genes revealed the presence of σ^{70} consensus sequences upstream of nine genes (Table 5). For one of these genes, the *flgG* homologue (*hp1585* in Tomb *et al*., 1997), a putative σ^{54} promoter was also detected, suggesting a possible coordinate regulation of this gene by means of two alternative overlapping promoters. A putative σ^{54} promoter was also detected upstream of the *flaB'* gene encoding a homologue of the FlaB flagellin (HP295 in Tomb *et al*., 1997). Based on these observations and taking into account the data presented, a model

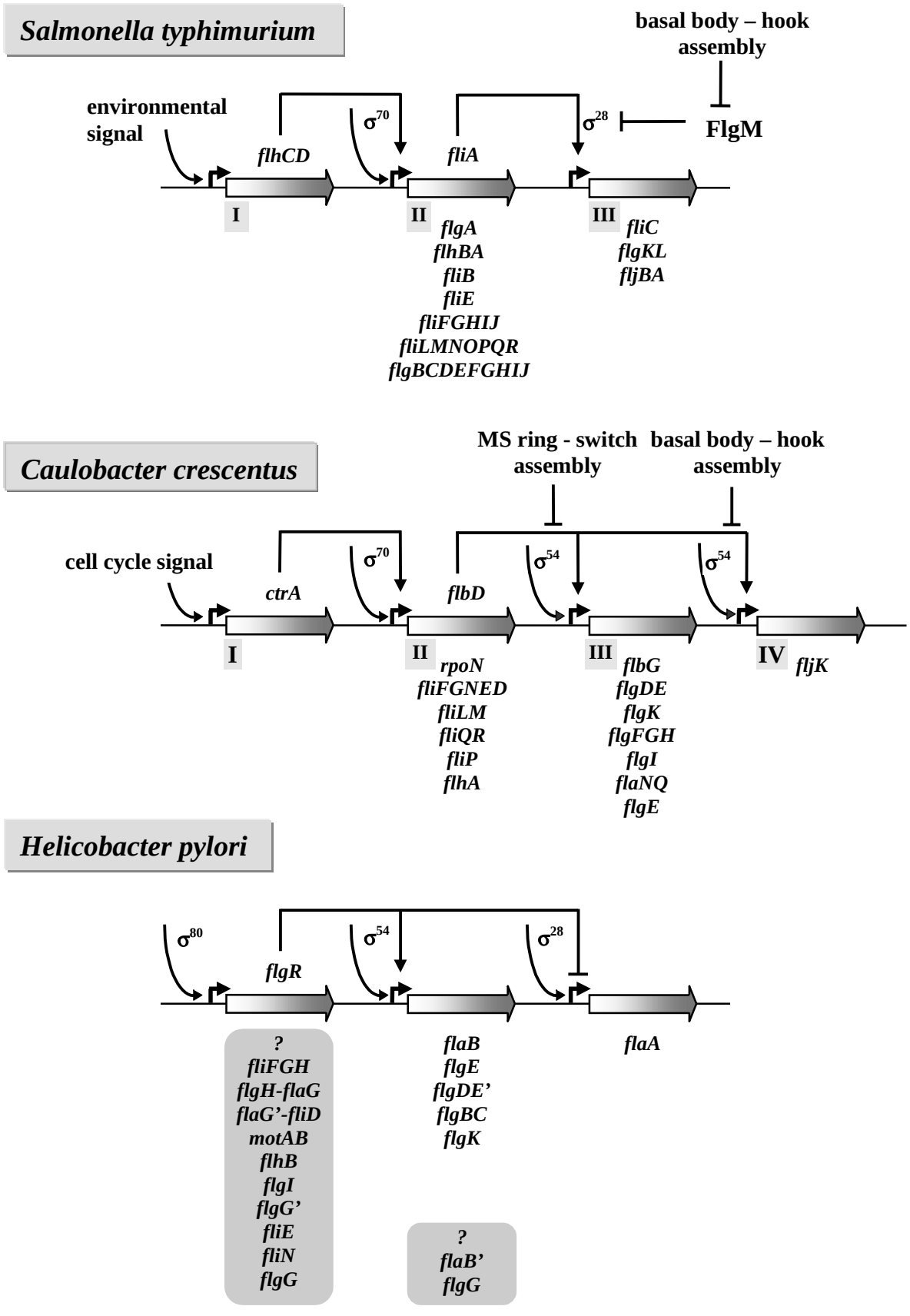

Fig. 27: Comparison of flagellar regulatory hierarchies in *Salmonella typhimurium, Caulobacter crescentus,* and *Helicobacter pylori*. Dependence of transcription on the specified sigma factors is indicated. Classes of flagellar genes are indicated by roman numbers. *H. pylori* genes and operons whose promoters have been deduced on the basis of nucleotide sequence analysis are enclosed by grey boxes. Symbols: promoter; → activation; ⊥ repression.

for the regulation of flagellar gene expression in *H. pylori* is proposed (Fig. 27). In this model, $E\sigma^{80}$ directs transcription of the master regulator FlgR and genes coding for structural components of the flagellar export apparatus, motor, and basal body. FlgR in turn activates transcription of σ^{54}-dependent genes encoding structural components of the basal body-hook complex and represses transcription of the σ^{28}-dependent gene encoding the major flagellin subunit FlaA.

3. Chaperones

The chaperones of *H. pylori* have been studied in some detail, because of their possible involvement in specific virulence mechanisms (Huesca *et al*., 1996; Macchia *et al*., 1993; Suerbaum *et al*., 1994). These studies resulted in a large amount of biochemical and immunological data concerning the specific roles of some *H. pylori* chaperones, but left analyses of transcriptional regulation of the respective genes almost untouched. In chapter E.3 of the present thesis, evidence is provided that the *groESL*, *hrcA-grpE-dnaK* and *cbpA-hspR-orf* operons, encoding the major chaperones of *H. pylori*, are transcribed by the vegetative sigma factor σ^{80}, and are regulated negatively by the transcriptional repressor HspR. Notably, the *hspR* gene is contained in one of the operons regulated by the HspR protein. Therefore, expression of HspR is autoregulated.

Evidence demonstrating that HspR acts as a specific transcriptional repressor of the three chaperone-encoding operons comes from the observation that the mRNA levels of these operons are increased by an average of 8-fold in the *hspR*$^-$ background. This effect is specific for the chaperone operons as transcription of other genes is unaffected in the mutant (Fig. 20A). Electrophoretic mobility shift assays and footprint experiments with purified HspR have revealed that the protein binds directly to the promoters of the three operons. In all cases, a large DNA region of about 75 bp is covered by HspR. However, the position of this region varies considerably between the different promoters. In the case of the P_{cbp} promoter, HspR binding occurs between positions +14 and -59, thus interfering probably directly with the binding of RNA polymerase to the -10 and -35 sequences, thereby autoregulating its own expression. For P_{gro} and P_{hrc}, HspR binds upstream of -46 and -78, respectively, in atypical positions for a repressor protein. On the P_{gro} promoter, contact between HspR and the RNA polymerase bound to the -10 and -35 regions may be possible and the negative effect of HspR on P_{gro} transcription may, therefore, be the result of inhibition of promoter clearance. However, in the case of the P_{hrc} promoter, HspR binding occurs too far upstream to allow a similar kind of interaction and other mechanisms of transcriptional repression that involve accessory DNA binding proteins and/or structural changes in DNA topology may have to be considered.

The most surprising result of this study concerns the environmental stimulus that leads to derepression of the HspR-regulated operons. Unlike the situation in other bacterial species investigated (Gross, 1996; Neidhardt and VanBogelen, 1987; Lindquist and Craig, 1988), it seems

that in *H. pylori* transcription of chaperone genes is not induced by heat but by osmotic shock. Under the experimental conditions used, no significant increase in transcription could be observed following a temperature shift from 37 °C to 45 °C, whereas incubation with 300 mM NaCl leads to a complete derepression of the HspR regulated promoters P_{gro} and P_{cbp}. Therefore, the *H. pylori* HspR homologue might not represent a "heat shock regulator" but rather an "osmotic shock regulator" that senses changes in the osmolarity of the environment and mediates a transcriptional response to these changes. A basis for these different sensor capacities is inferred by the amino acid sequence of HspR, which is highly homologous to the respective sequence of the heat shock regulator HspR from *S. coelicolor* in the N-terminal putative DNA binding part (51.5% amino acid identity within the first 70 residues), but not in the C-terminal part (15.1% amino acid identity within the last 53 residues) (Bucca *et al.*, 1995). As this latter part of the protein is supposed to account for protein-protein interactions and conformational changes in response to certain stimuli, it seems plausible that the divergent C-terminal sequence of the *H. pylori* protein represents an adaptation to a different function, perhaps rendering it insensitive to changes in temperature but, on the other hand, making it sensitive to changes in osmolarity. Support for this hypothesis comes from the fact that the HspR homologue of the extreme thermophilic bacterium *Aquifex aeolicus* (Deckert *et al.*, 1998) shares homology with its *H. pylori* counterpart not only in the N-terminal (58.1% identity) but also in the C-terminal part (24.5% identity). As heat shock is quite unlikely to occur in *A. aeolicus* which lives at 95 °C, the homology of its HspR sequence with the respective *H. pylori* sequence may indicate a common adaptation to a function other than heat shock regulation in these two organisms that are otherwise completely different in their ecological niches.

Interestingly, the P_{hrc} promoter, although clearly regulated by HspR (Fig. 20, 23, 25C), did not respond to osmotic stress with an increase in its transcriptional activity. Unlike the P_{cbp} and P_{gro} promoters that appeared to be completely derepressed under the osmotic shock conditions used, this promoter showed no change in activity after addition of salt to the medium (Fig. 23). A possible explanation for this phenomenon lies in the putative higher affinity of the P_{hrc} promoter for the HspR protein (Fig. 25C) with respect to the other two promoters, which might account for maintenance of the repressed status also in the presence of high concentrations of salt. DNA binding proteins might also be involved in repression of the P_{hrc} promoter, since the distant binding site of the HspR protein on this promoter probably does not allow for direct inhibition of transcription by interference with RNA polymerase binding. Cooperation with a second repressor that binds further downstream might be necessary to repress transcription. In support of this hypothesis is the observation that the first gene of the operon regulated by P_{hrc} encodes a homologue of the HrcA transcriptional repressor of class I heat shock genes from *B. subtilis*. Furthermore, an inverted repeat with some similarities to CIRCE, the consensus binding site of HrcA (TTAGCACTC-N9-GAGTGCTAA, underlined are unmatching bases), is present at position -44 with respect to the transcriptional start site of the operon (data not shown). The *H. pylori* HrcA homologue might bind to this inverted repeat thereby repressing transcription of the operon in cooperation with the distally binding HspR repressor. Interestingly, HrcA is predicted to be an integral transmembrane protein,

suggesting a combined sensor and transcriptional regulator capacity as described for the ToxR protein of *Vibrio cholerae* (Miller *et al*., 1987; DiRita and Mekalanos, 1991)

Apart from its role in repressing transcription of chaperone genes (summarised in Fig. 28), the HspR protein might have additional functions in the regulatory network that governs protein expression in *H. pylori.* Evidence for this hypothesis comes from the observation that the *hspR*$^-$ mutant has lost motility functions, as assayed on agar motility plates (data not shown). Indication for additional HspR functions comes also from the fact that the cellular level of urease enzyme is reduced in the *hspR*$^-$ mutant. This reduction is apparently not due to direct transcriptional regulation (Fig. 20A), and is probably not a consequence of the generally altered level of chaperones in the mutant cells, as the amount of other *H. pylori* proteins, such as VacA and CagA, seems to be unaffected in the mutant. Possibly, HspR regulates a protease, that specifically degrades the subunits of the urease enzyme under certain environmental conditions. Increased expression of chaperones and decreased expression of urease, and possibly other proteins, may thus be a coordinated response of *H. pylori* to changes in the environmental conditions that are mimicked *in vitro* by high osmolarity, but may correspond *in vivo* to the achievement of a certain ecological niche. Such a niche could be represented by the mucus overlaying the epithelial cells of the stomach. In the neutral milieu of this mucus layer, urease might not be essential for survival of the bacteria any more, and might therefore be downregulated by HspR, while chaperones might be necessary for adherence of the bacteria to epithelial cells and their expression might therefore be derepressed by HspR. Thus HspR might be directly involved in the colonisation of *H. pylori* on the gastric epithelial cells and consequently in the establishment of a successful infection.

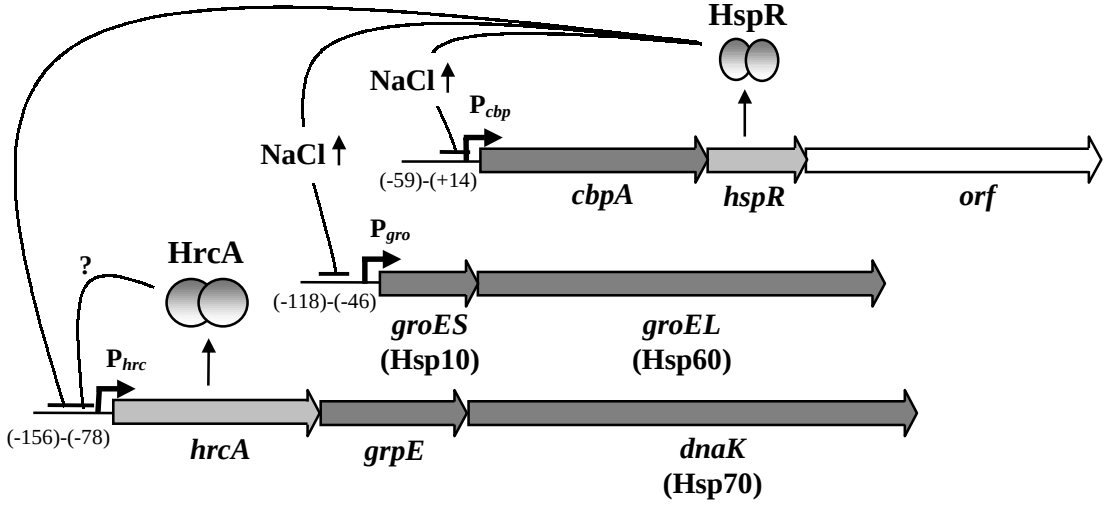

Fig. 28: Regulation of *H. pylori* chaperone genes. The HspR regulatory protein represses transcription of the chaperone encoding operons by binding to large DNA regions of about 75 bp in all three promoters (positions of the binding sites are indicated in brackets with respect to the transcriptional start sites). Transcription from the P_{cbp} and P_{gro} promoters can be derepressed by osmotic shock treatment (symbolised NaCl↑), while transcription from the P_{hrc} promoter remains unaffected under these conditions. Transcription from this promoter is likely to be corepressed by HspR and HrcA.