E. RESULTS

1. Transcriptional analysis of the divergent *cagAB* genes encoded by the *H. pylori* pathogenicity island

1.1 Determination of transcription start sites

The structural organisation of the *cag* pathogenicity island is schematically depicted in Fig. 3A. *cagA,* the gene encoding the cytotoxin associated antigen CagA, is located at the 3' end of the pathogenicity island and is transcribed in the opposite direction from the upstream *cagB* gene. The intergenic region between the two ATG start codons - outlined in nucleotide sequence in Fig. 3B - should therefore contain at least two promoters that regulate transcription of these two genes.

In order to define the start points of transcription of *cagA* and *cagB,* primer extension assays were performed with total RNA from *H. pylori* G27 and specific oligonucleotides as indicated in Fig. 3B (Table 4). The urea/acrylamide gel in Fig. 4A shows the result of a primer extension experiment carried out by hybridising oligonucleotide cagN (Fig. 3B; Table 4) to *H. pylori* RNA. Reverse transcription of the RNA yielded one major band which defines the P_1 start site of transcription at 105 nucleotides upstream of the AUG start codon of the *cagA* gene (lane 1). To determine the transcriptional start point of the *cagB* gene oligonucleotide orfX (Fig. 3B; Table 4) was hybridised to total *H. pylori* RNA and elongated with reverse transcriptase. Fig. 4B (lane 1) shows two major bands, that place the P_2 and P_3 start sites of RNA transcription at 181 and 58 nucleotides upstream of the AUG start codon of the *cagB* gene, respectively.

The intergenic region between *cagA* and *cagB* thus contains three promoters, P_1, P_2, and P_3. The P_1 promoter regulates transcripiton of the *cagA* gene while the P_2 and P_3 promoters are responsible for transcription of the *cagB* gene. All three promoters show putative conserved -10 regions with similarities to the consensus sequence (TATAAT) recognised by the major RNA polymerase $E\sigma^{70}$ from *E. coli* (Fig. 3B). Surprisingly, no obvious -35 conserved regions (TTGACA) can be detected (Fig. 3B).

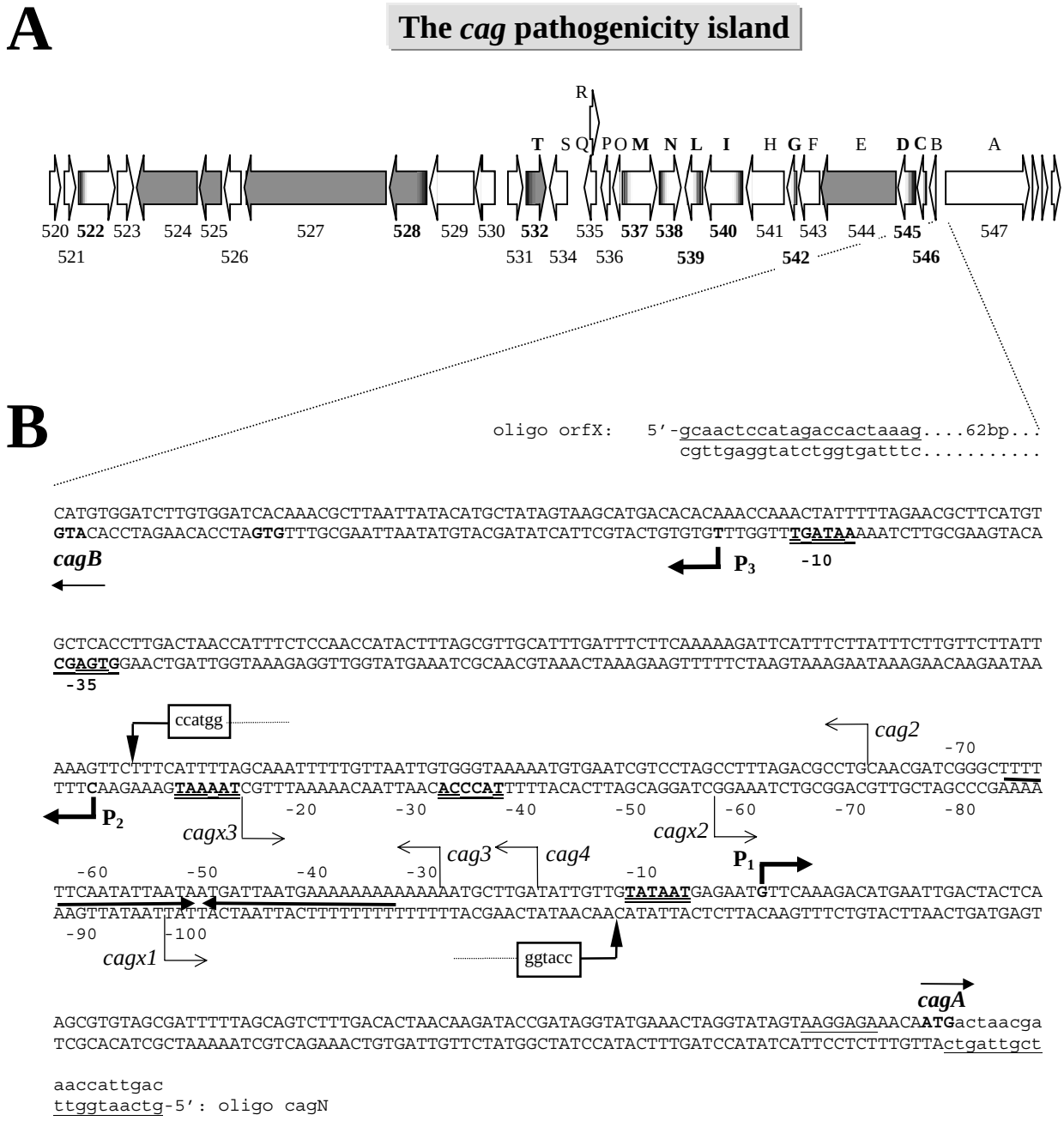


Fig. 3: A) Schematic representation of the genes contained in the *H. pylori* pathogenicity island. Numbers refer to the annotated genome sequence published by Tomb *et al*. (1997), letters to the nomenclature suggested by Censini *et al.* (1996). Other symbols are as in Fig. 2. **B) Nucleotide sequence of the 436 bp intergenic region between *cagA* and *cagB*.** The direction of transcription of promoters $P_{1\text{-}3}$ is indicated by thick arrows. The transcriptional start points are in bold letters, nucleotides matching the -10 and -35 promoter regions are double underlined, mismatches are underlined. The *cagA* ribosomal binding site is underlined. Nucleotide sequences corresponding to oligonucleotides orfX and cagN (Table 4), used in primer extension experiments, are in small letters and underlined. The translational start sites and the direction of translation of the *cagAB* genes are indicated by bold letters and arrows, respectively. An alternative GTG codon that could function as translational start codon of the *cagB* gene is indicated by bold letters. Boxed sequences with down- or up-pointing arrows indicate positions where deletions started. Left- or right-pointing arrows indicate the end-point of the deletions and are marked by the corresponding strain names (see graphic representation in Fig. 5). An inverted repeat is marked by convergent arrows.

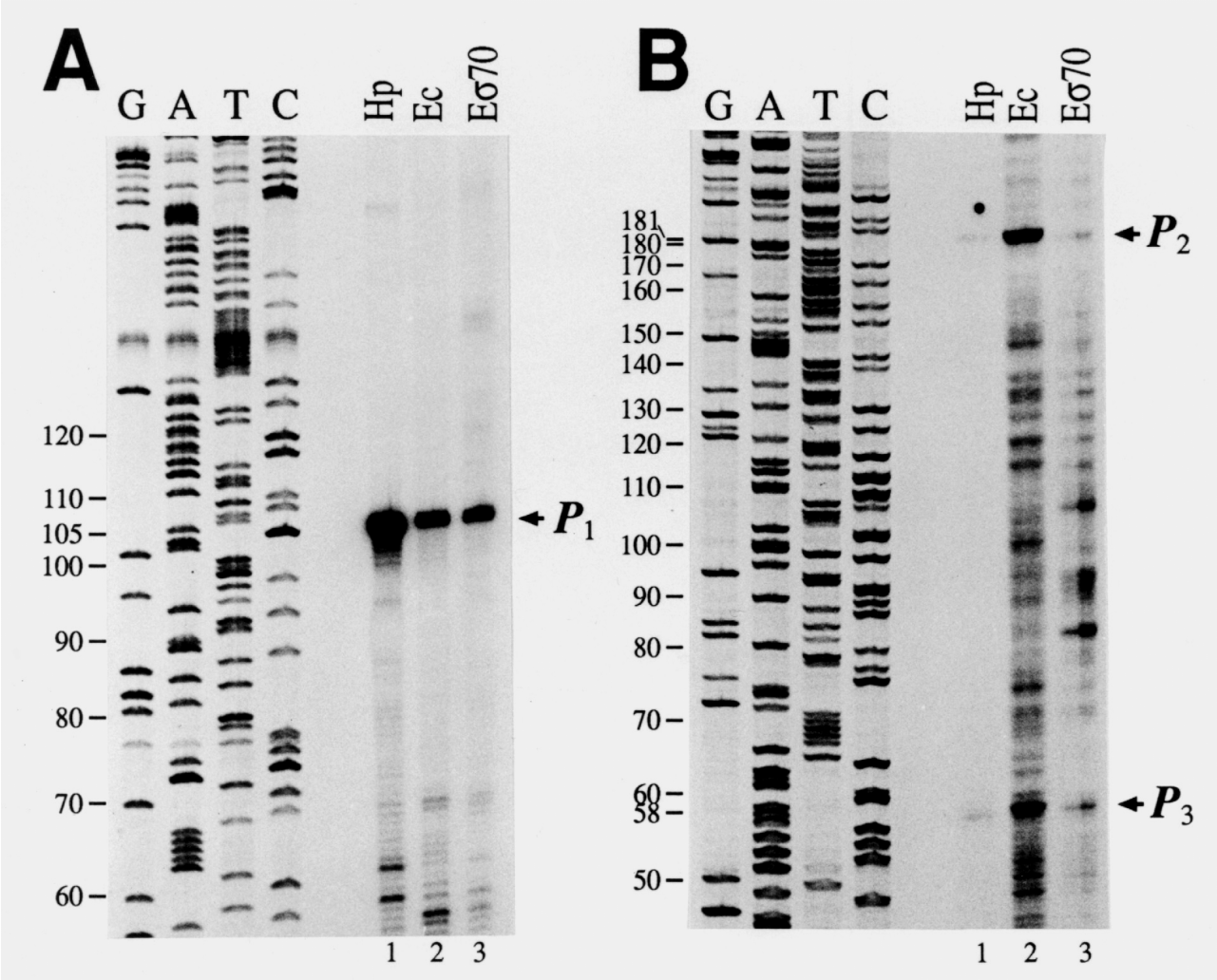


Fig. 4: Determination of transcriptional start sites in the intergenic region between *cagA* and *cagB*. Total RNA isolated from *H. pylori* CCUG17874 (lanes 1) and *E. coli* DH5α harboring plasmid pSK11.1A (lanes 2; Table 3) as well as RNA generated *in vitro* (lanes 3) using the *E. coli* RNA polymerase and plasmid pSK11.1A as template were hybridised to the radiolabeled oligonucleotides cagN (panel **A**) or orfX (panel **B**) (Table 4). Transcriptional start sites (P_1, P_2, P_3) are marked with arrows on the right side of each panel. Sequence reactions (G, A, T, C) were performed with the same oligonucleotides as in the respective primer extension experiments and plasmid pSK11.1A as a template. Numbers in the left side of each panel indicate the base pair distance from the translational start sites. Symbols: Hp, total RNA from *H. pylori* cells; Ec, total RNA from *E. coli* transformed with plasmid pSK11.1A; $E\sigma^{70}$, *in vitro* transcripts obtained with *E. coli* RNA polymerase-σ^{70} and plasmid pSK11.1A.

1.2 Specificity of promoter recognition

Primer extension assays were then performed with RNAs extracted from *E. coli* transformed with plasmid pSKA11.1 (Table 3), which carries the *cagAB* promoter region, and with RNAs generated *in vitro* by using the same plasmid DNA as template and purified $E\sigma^{70}$ RNA polymerase from *E. coli*. The experiment in Fig. 4A (lane 2) shows that the P_1 promoter is activated in *E. coli*, suggesting that the transcriptional machinery of *E. coli* is sufficient to promote transcription from this promoter. In addition, an *in vitro* transcription experiment demonstrates that promoter P_1 is

activated by the $E\sigma^{70}$ RNA polymerase (lane 3). A similar experiment carried out on the P_2 and P_3 promoters shows that transcription from these promoters in either *E. coli* and *in vitro* starts at the same site as observed in *H. pylori*, as well as at several upstream and downstream sites (Fig. 4B, lanes 2, 3). The multiple initiations of P_2 and P_3 transcription suggest that specificity of initiation requires conditions attained only in *H. pylori*.

Taken together these results strongly suggest that transcription from the P_1, P_2, and P_3 promoters is regulated by RNA polymerase containing σ^{80}, the *H. pylori* homologue of the vegetative sigma factor σ^{70} from *E. coli*. This sigma factor has been cloned and a detailed analysis including the construction of hybrid σ^{70}/σ^{80} proteins has revealed that it shares not only sequence homology but also promoter recognition and binding specificity with its *E. coli* counterpart (Beier *et al*., 1998). In fact, a hybrid RNA polymerase containing α, β, and β' from *E. coli* and σ^{80} from *H. pylori* was able to bind to and transcribe from the *cagA* P_1 promoter (Beier *et al*., 1998).

1.3 Deletion analysis of the P_1 and P_2 promoter region

In order to determine which sequences are essential for transcription of *cagA* and *cagB*, *P_1::lacZ* fusions were constructed in which internal parts of the promoter sequence between P_1 and P_2 were deleted. For the introduction of these different fusions into the *H. pylori* chromosome a two step allelic exchange procedure was chosen (Fig. 5). In the first step the proximal part of the *cagA* coding sequence and half of the *cagAB* intergenic region were substituted with a kanamycin cassette. In the second step the kanamycin cassette was then replaced with the different *P_1::lacZ* fusions and a chloramphenicol resistance gene. The advantage of this procedure as compared to a one step allelic exchange approach consists in the avoidance of undesired recombination events that could occur between the 5' untranslated regions of the *P_1::lacZ* constructs and the chromosome that would inevitably result in reconstitution of the wild type promoter sequence in the mutant strain.

For construction of the *cagA::km* derivative strain a plasmid containing a kanamycin cassette flanked by the *cagB* gene and part of the *cagAB* intergenic region on one side and a distal part of the *cagA* coding region on the other side (pGEM3(*cagA::km*), Table 3; Fig. 5) was transformed into the naturally competent *H. pylori* strain G27. One out of several hundred kanamycin resistant colonies was tested for correct replacement of the wild type sequence with the antibiotic resistance gene. For that purpose chromosomal DNA of the mutant strain was digested with *Dra*I, *Hae*III, and *Pvu*I restriction enzymes, and hybridised in a Southern Blot experiment to the labeled orf-5'/Δorf-3' PCR fragment (Table 4) as described in Fig. 6. The mutant strain had the first 2884 bp of the *cagA* gene and 254 bp of the *cagAB* intergenic region substituted by the kanamycin resistance gene (Fig. 6).

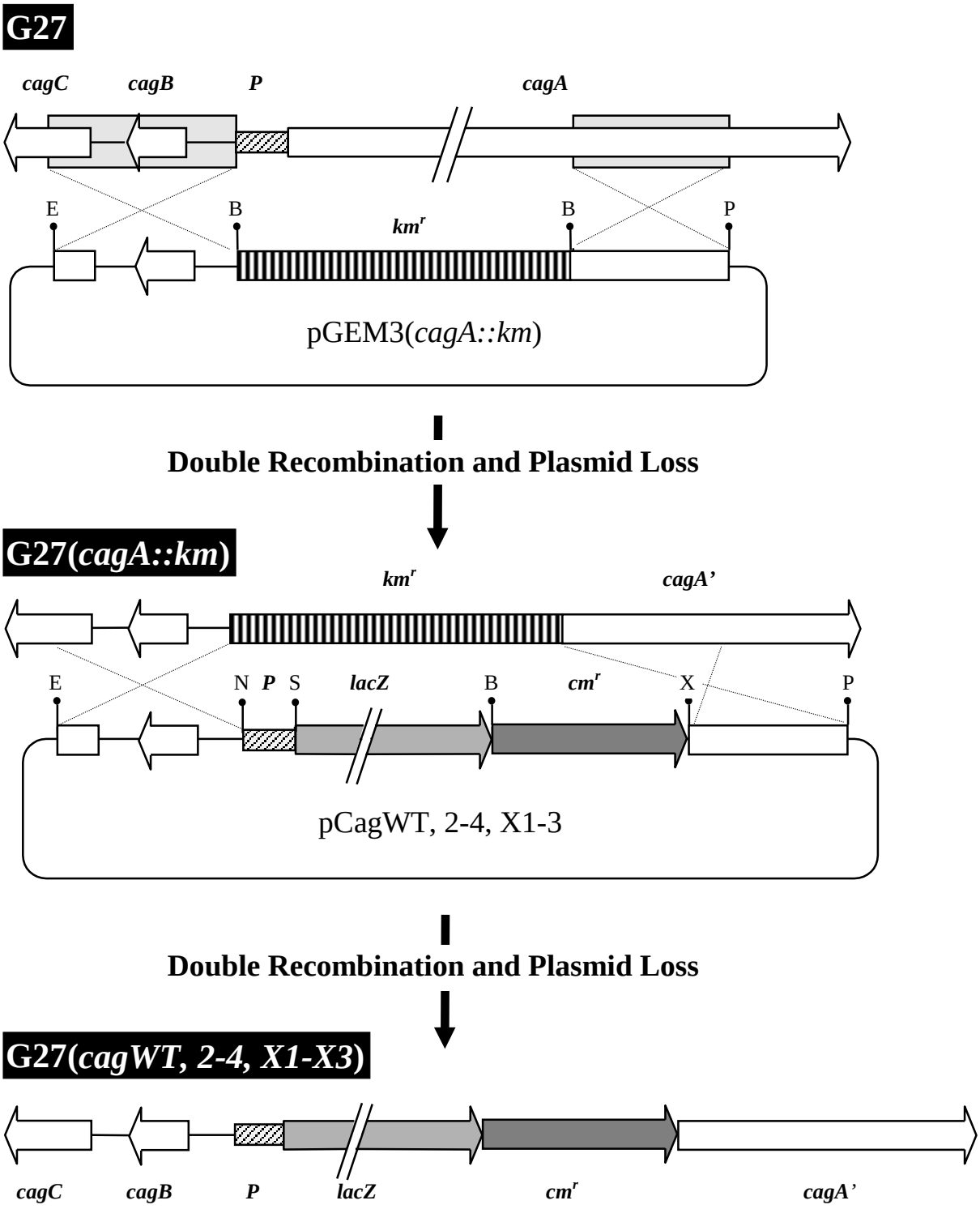


Fig. 5: Steps in the construction of *H. pylori* G27 mutants. Arrows point to the direction of transcription of the indicated genes. The gray boxes on the left and right sides mark DNA regions cloned into the pGEM3 vector (Table 3) and used for allelic exchanges. A schematic representation of the pGEM3(*cagA::km*) plasmid (Table 3) used to construct a strain carrying the kanamycin cassette integrated into the *cagA* gene is reported. Main features of *H. pylori* constructed mutant strains are indicated (see Table 3 for details). Symbols: E, *Eco*RI; B, *Bam*HI; P, *Pst*I; N, *Nco*I; S, *Sma*I; X, *Xba*I; *km*^{r}, kanamycin resistance gene; *cm*^{r}, chloramphenicol resistance gene; *lacZ*, β-galactosidase gene. *P*, promoter.

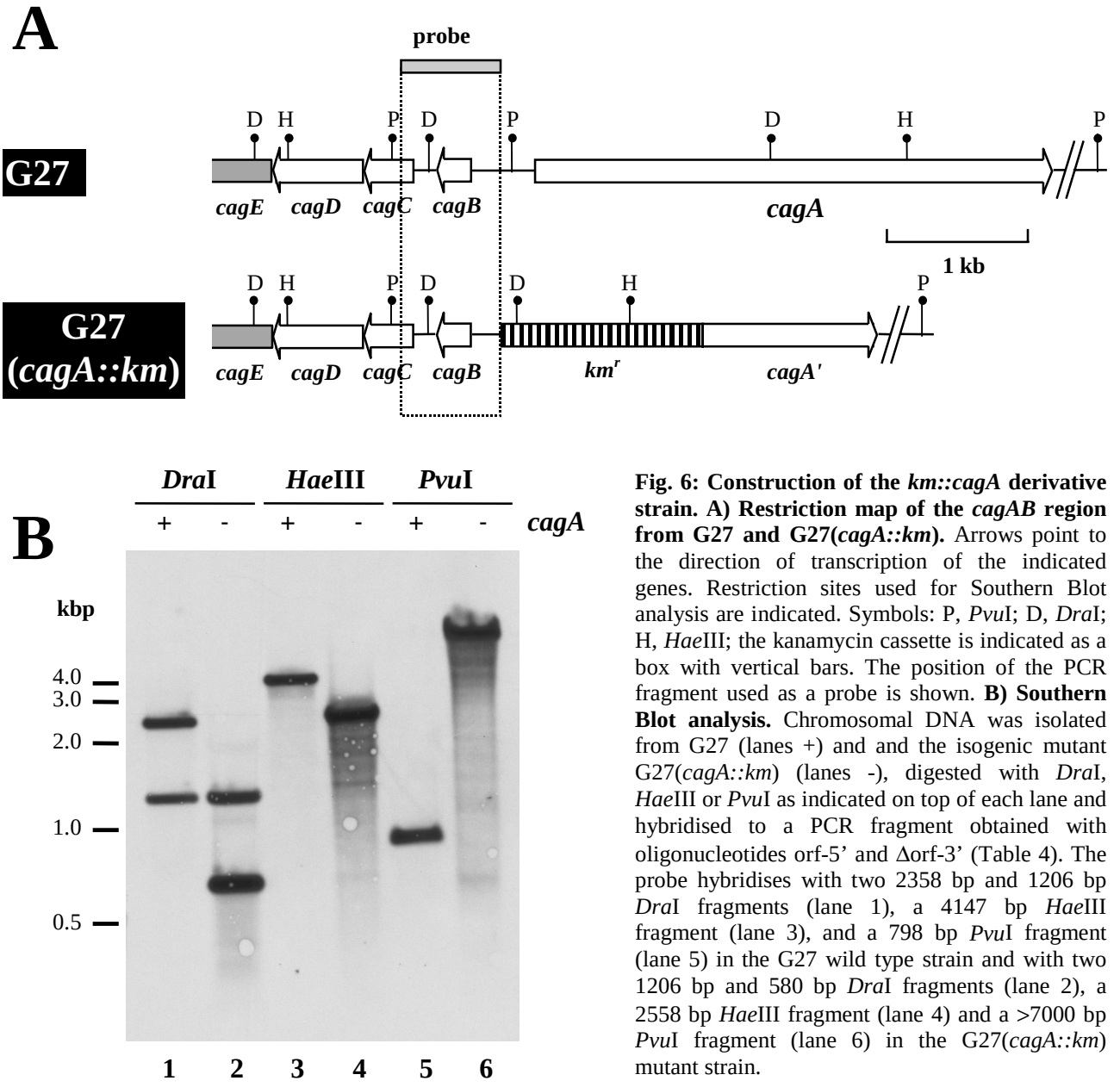
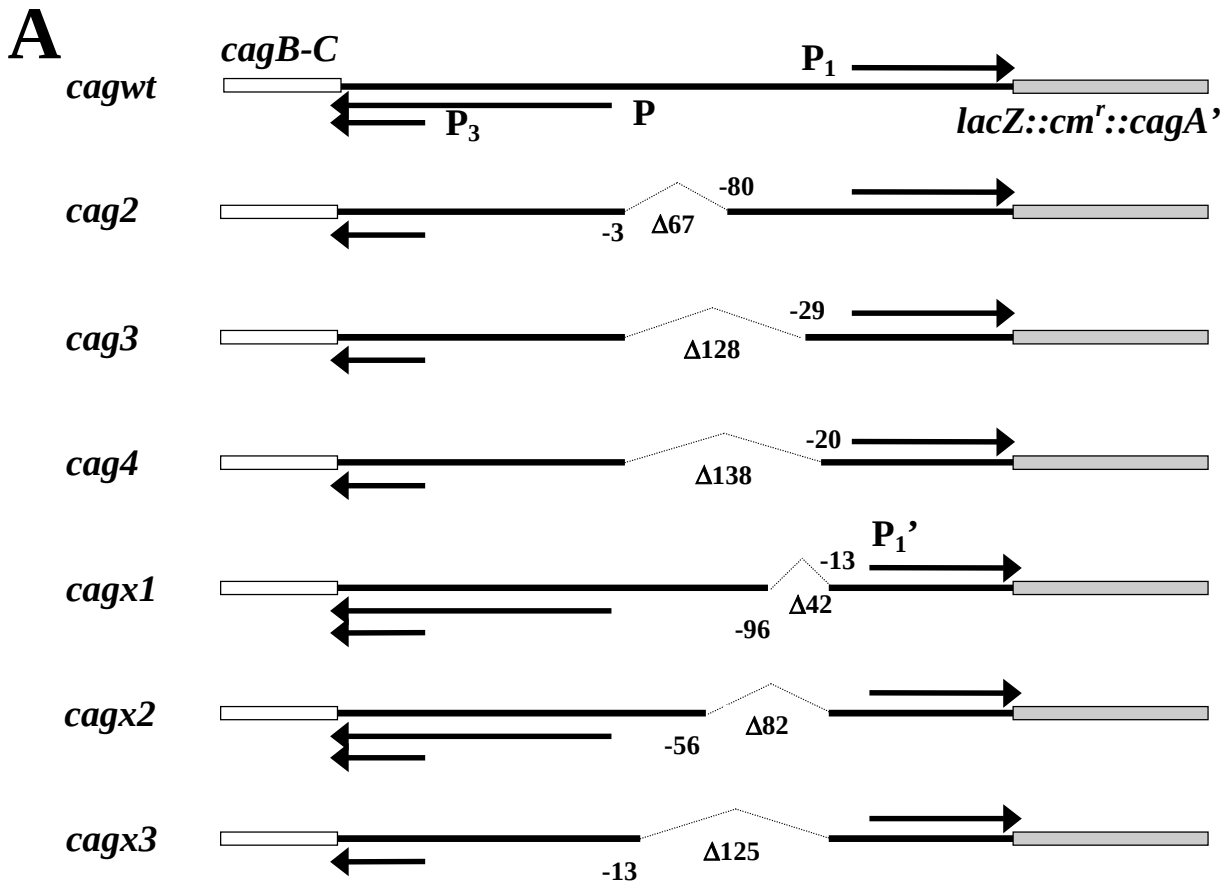


Fig. 6: Construction of the *km::cagA* derivative strain. A) Restriction map of the *cagAB* region from G27 and G27(*cagA::km*). Arrows point to the direction of transcription of the indicated genes. Restriction sites used for Southern Blot analysis are indicated. Symbols: P, *Pvu*I; D, *Dra*I; H, *Hae*III; the kanamycin cassette is indicated as a box with vertical bars. The position of the PCR fragment used as a probe is shown. **B) Southern Blot analysis.** Chromosomal DNA was isolated from G27 (lanes +) and and the isogenic mutant G27(*cagA::km*) (lanes -), digested with *Dra*I, *Hae*III or *Pvu*I as indicated on top of each lane and hybridised to a PCR fragment obtained with oligonucleotides orf-5' and Δorf-3' (Table 4). The probe hybridises with two 2358 bp and 1206 bp *Dra*I fragments (lane 1), a 4147 bp *Hae*III fragment (lane 3), and a 798 bp *Pvu*I fragment (lane 5) in the G27 wild type strain and with two 1206 bp and 580 bp *Dra*I fragments (lane 2), a 2558 bp *Hae*III fragment (lane 4) and a >7000 bp *Pvu*I fragment (lane 6) in the G27(*cagA::km*) mutant strain.

For replacement of the kanamycin cassette with the promoter fusions a series of 7 plasmids was then constructed which contained the different P_1*::lacZ* fusions together with a chloramphenicol resistance gene and suitable *H. pylori* derived fragments for homologous recombination. Fig. 7A shows an overview of these plasmids and Table 3 illustrates the details of their construction. All plasmids were introduced into the G27(*cagA::km*) mutant strain by natural transformation and transformants were screened for loss of kanamycin resistance and acquisition of chloramphenicol resistance. Correct replacement of the kanamycin cassette with the promoter constructs was verified by means of PCR using oligonucleotides orf-5' and lac (Table 4). Fig. 7B shows that all mutants contained the expected promoter deletions.



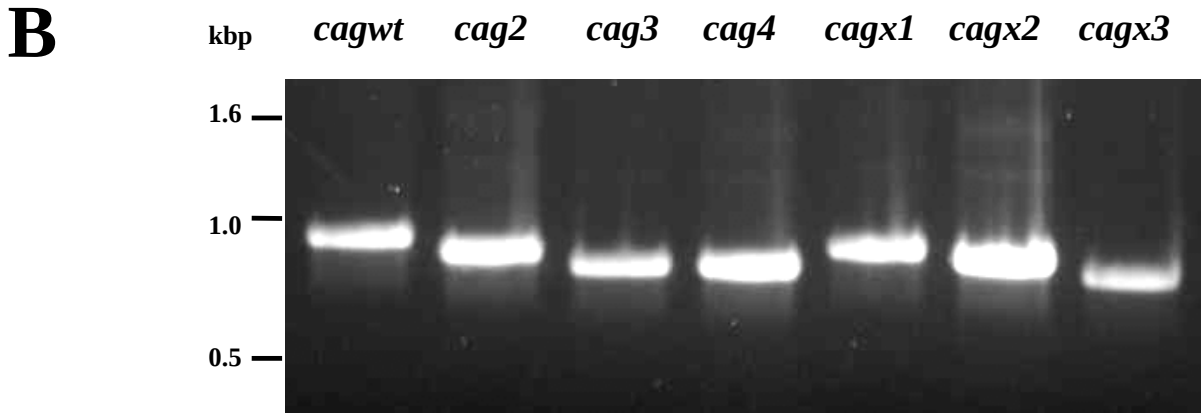


Fig. 7: P_1*::lacZ* fusions. **A) Schematic representation of plasmid DNA constructs**. Each of the plasmids contains an upstream *cagBC* DNA fragment (open box), a promoter region (black bars) and a downstream segment containing the *lacZ* gene ligated to a chloramphenicol resistance cassette and a 3' distal region of *cagA* (grey boxes). Directions of transcription are indicated by arrows, promoters by P and deleted regions by Δ. Positions of sequences retained in each promoter mutant are indicated with respect to the transcriptional start sites of P_1 (upper numbers) and P_2 (lower numbers). **B) Introduction of P_1*::lacZ* fusions into the *H. pylori* chromosome**. Promoter regions were PCR-amplified from chromosomal DNA of the mutant strains with oligonucleotides orf-5' and lac (Table 4). The sizes of the obtained fragments (972, 905, 844, 834, 930, 890, 847 bp) correspond to the internal promoter deletions of the respective mutant strains (cagwt, cag2, cag3, cag4, cagx1, cagx2, cagx3).

To evaluate the transcriptional activity of the promoter mutants, bacteria were grown in liquid medium and their β-galactosidase activities were measured. Fig. 8A shows the results of these measurements. Mutant cag2, which retains 80 bp of the P_1 upstream sequence and lacks the -10 box of the P_2 promoter, showed a β-galactosidase activity similar to that of the wild type promoter sequence (cagWT), thus suggesting that activation of promoter P_2 does not interfere with transcription from promoter P_1. Further deletions extending to position -29 (cag3) or -20 (cag4) of P_1 resulted in a threefold reduction in enzymatic activity. Deletions extending to –13, and thus retaining only the -10 sequence of P_1 (cagX1-3), reduced β-galactosidase activity to 15-20% of the wild type activity. Taken together these results indicate that the DNA region between -29 and -80 is important for full activation of the P_1 promoter.

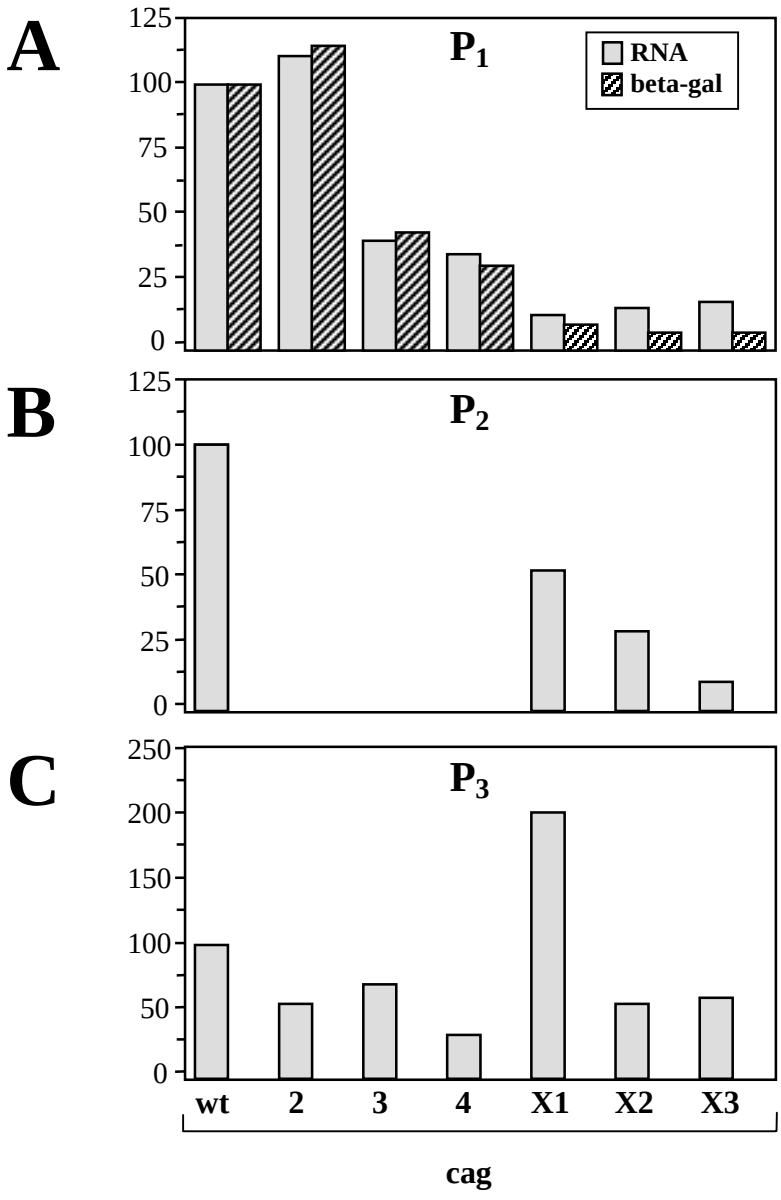


Fig. 8: Transcriptional activities of the $P_{1\text{-}3}$ promoter mutants. The abscissa reports the name of each promoter mutant (see Fig. 7A). Activities of the β-galactosidase enzyme (hatched black bars) expressed under the control of the P_1 promoter were measured as described in D.16. The value obtained in the wild type construct (cagwt) was set to 100 (arbitrary units in the ordinate axes). All assays were performed at least three times and the errors of the results were less than 20% in all cases. RNA accumulation (gray bars) at the $P_{1\text{-}3}$ promoter mutants was evaluated by exposing the quantitative primer extension analyses shown in Fig. 9 to a PhosphorImager. Values of the wild type constructs were set to 100.

To study the regulation of the P_1 promoter in more detail, total RNA from the different *H. pylori* P_1 promoter mutants was then extracted and the respective transcriptional start points and mRNA amounts were determined by primer extension. Accumulation of mRNA at the P_1 promoter was evaluated by exposing the gels shown in Fig. 9A to a PhosphorImager to measure the relative intensities of the bands. As shown in Fig. 8A the amounts of mRNA that could be detected in the different promoter mutants are in agreement with the respective β-galactosidase acitivities obtained with the *P_1::lacZ* fusions. Fig. 9A shows that the cag2-4 P_1 promoter mutants retained the same transcriptional start site as the wild type promoter (cagWT), while the cagX1-3 mutants showed a different initiation site. These latter mutants, which carry deletions from position -13 of the P_1 promoter to positions -55, -95 or -138, respectively, initiated mRNA synthesis at position +8 (P_1') with respect to the wild type transcript.

The mRNA analysis therefore confirms that the DNA region between positions -29 and -80 is important for full transcriptional activation of the P_1 promoter. In addition the DNA element spanning from position -20 to position -13 is recognised to be crucial for correct initiation of mRNA synthesis at this promoter.

To gain further information on the mechanism of promoter recognition and transcription initiation at the P_1 promoter, total RNA from *E. coli* DH5α harbouring plasmids pCag2-4 or pCagX1-3 was then extracted and the initiation sites were determined by primer extension analyses. As expected, the P_1 promoter initiated RNA transcription at the same +1 site as in *H. pylori*. Surprisingly, RNA extracted from cells with plasmids pCagX1-3 showed initiation of transcription at position +1 instead of +8 as in *H. pylori* (data not shown). The same results were confirmed by *in vitro* transcription experiments using the *E. coli* RNA polymerase and plasmids pCag2-4 and pCagX1-3 as templates. It is concluded that the shift of 7 nt in the initiation of P_1 transcription observed in *H. pylori* cagX1-3 mutants is restricted to the *H. pylori* system. Possibly, it represents repositioning of the RNA polymerase on the promoter.

1.4 Transcriptional activity of the P_2 and P_3 promoters in the different mutants

To analyse the transcriptional response of the P_2 and P_3 promoters to the internal deletions of DNA between *cagA* and *cagB* (Fig. 7A), the RNA preparations used in Fig. 9A were hybridised to oligonucleotide orfX (Fig. 3B; Table 4). The results of primer extension analysis are shown in Fig. 9B. As in the case of the P_1 specific primer extensions, accumulation of mRNA in the different mutants was evaluated by exposing the same gel to a PhosphorImager (Fig. 8). A deletion of 42 bp upstream from position -96 of the P_2 promoter and including sequences between -13 and -54 of the P_1 promoter (cagX1), determined a two-fold reduction in the amount of P_2-associated mRNA (Fig.

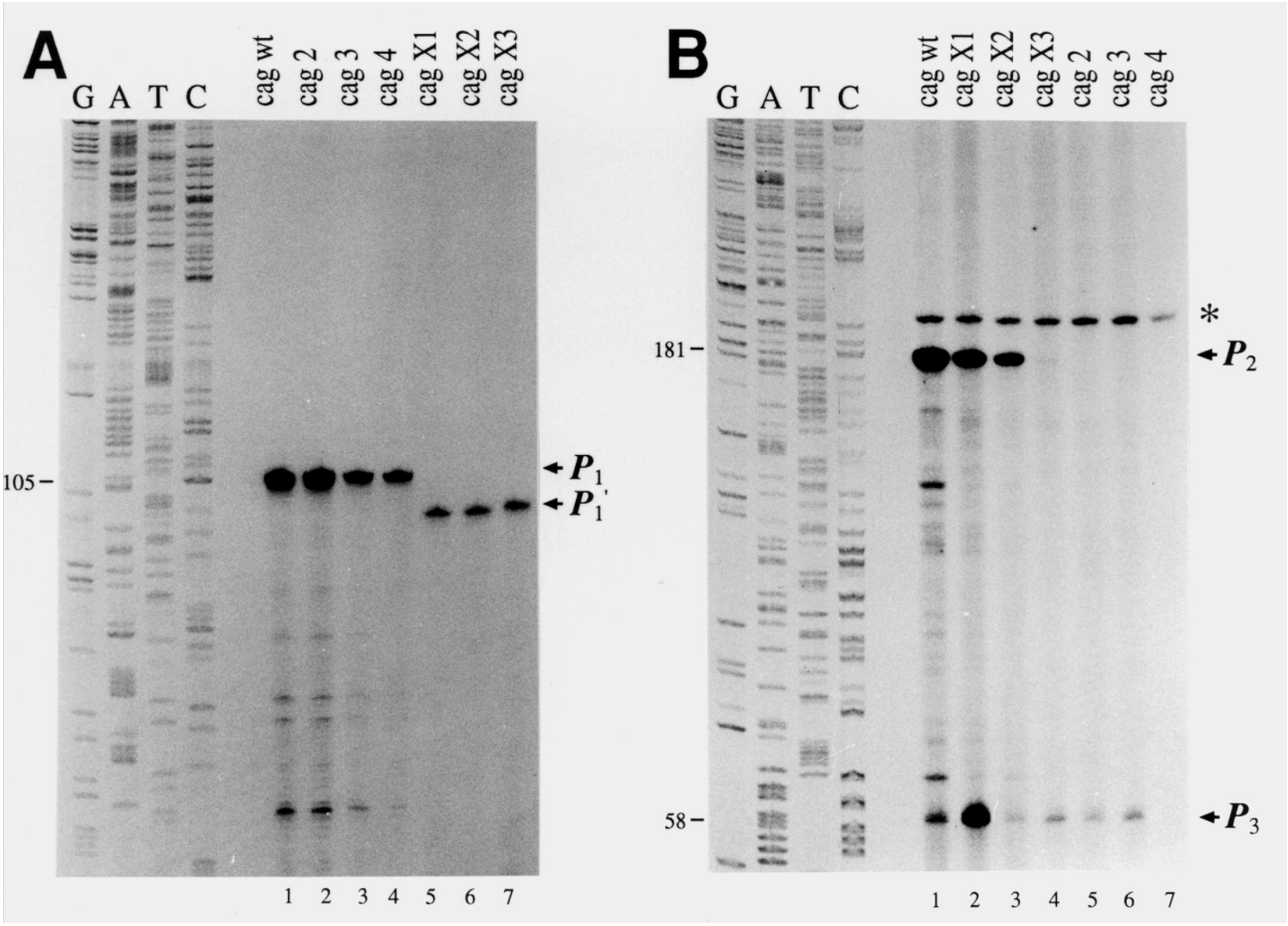


Fig. 9: Primer extension analysis of the *H. pylori* promoter mutants. **A)** Total RNA extracted from *H. pylori* cells end extended with oligonucleotide lac (Table 4). The name of the mutant strains are indicated on top of each lane. The P_1 and P_1' transcriptional start sites are indicated. GATC, nucleotide sequence of plasmid pCagwt (Table 3) with oligonucleotide lac used as a size marker. **B)** The P_2 and P_3 transcriptional start sites are indicated. RNAs are as in panel A). The star (*) indicates a cross-reactive band which is independent of the promoter activities and served as reference band. It is worth mentioning that this band was also detected in a longer exposure of the gel shown in Fig. 4B.

9B, lane 2; Fig. 8B). Conversely, in this mutant strain, the amount of RNA at the P_3 promoter was increased two-fold (Fig 9B, lane 2; Fig. 8C). A deletion extending to position -56 (cagX2), determined a further decrease in the amount of the P_2 mRNA, while the amount of the P_3 mRNA was reduced to wild type levels. In the deletion mutant extending to position -13 (cagX3), the RNA band corresponding to the P_2 mRNA was dramatically reduced and, as expected, it was absent in deletion mutants extending to position -3 (cag2-4).

It is concluded that a DNA region extending upstream from position -96 is required for full activation of the P_2 promoter. Therefore, full activation of this promoter depends on sequences belonging to the P_1 promoter. The amount of the P_3-associated mRNA was evaluated to be six-fold less than the amount of the P_2 mRNA. Very likely, the transcriptional activity of this promoter is weak and its regulation depends on sequences lying between the P_1 and P_2 promoters.

1.5 Binding of the α subunit of RNA polymerase to sequences upstream of the P_1 promoter

Nucleotide sequence analysis of the DNA sequence upstream of the P_1 promoter shows a stretch of 14 adenine residues located between positions -28 and -41 (Fig. 3B). Interestingly, the most distal part of this sequence belongs to a 33 bp non-perfect inverted repeat centered at position -52, spanning from position -36 to position -68. DNA structure and base content of this inverted repeat suggest that it may act as a promoter UP element which, for other promoters (Ross *et al*., 1993), has been shown to be the target site of the C-terminal domain of the α subunit of RNA polymerase. In a first attempt to assess whether the C-terminal domain of α can interact with this putative UP element a DNA fragment containing the wild type *P_1::lacZ* fusion was cloned into the vector pMMB208 and the resulting plasmid pMMB208cagwt (Table 3) was complemented in *E. coli* DH5α with plasmids pLAW2 and pLAW2Δ256 (Table 3), expressing the wild type α subunit and its C-terminal tuncated form, respectively. Coexpression of pMMB208cagwt with pLAW2 resulted in two-fold enhancement of β-galactosidase activity with respect to the non-complemented control while coexpression with pLAW2Δ256 yielded a 2 fold decrease in β-galactosidase activity. This suggested that the C-terminal end of the α subunit might interact specifically with the putative UP element of the P_1 promoter. To confirm this hypothesis, DNase I footprinting assays were then carried out. Fig. 10 shows a footprint experiment performed on a DNA fragment specific for the P_1 promoter using either the $E\sigma^{70}$ RNA polymerase from *E. coli* or different amounts of purified wild type α subunit or its C-terminal truncated form. A DNase I protection by the RNA polymerase could be observed from position +14 to position -60 (lane 3). Furthermore, a DNase I hypersensitive site was detected at position -51, almost in the center of the inverted repeat (see Fig. 3B). When purified wild type α subunit was added to a concentration of 2 μM, a protected region spanning from positions -17 to -70 was clearly distinguishable (lane 5). This suggested that the A/T rich inverted repeat from -32 to -68 represents a binding site for this transcriptional factor. No such protection was observed when corresponding molar amounts of the C-terminal truncated form were used (lanes 6 and 7), indicating that the C-terminal domain of α is responsible for the binding observed with the wild type subunit. The pattern of protection by α was slightly different from that obtained by RNA polymerase. Probably, this reflects conformational differences in the α subunit when it is part of the holoenzyme (Ross *et al*., 1993).

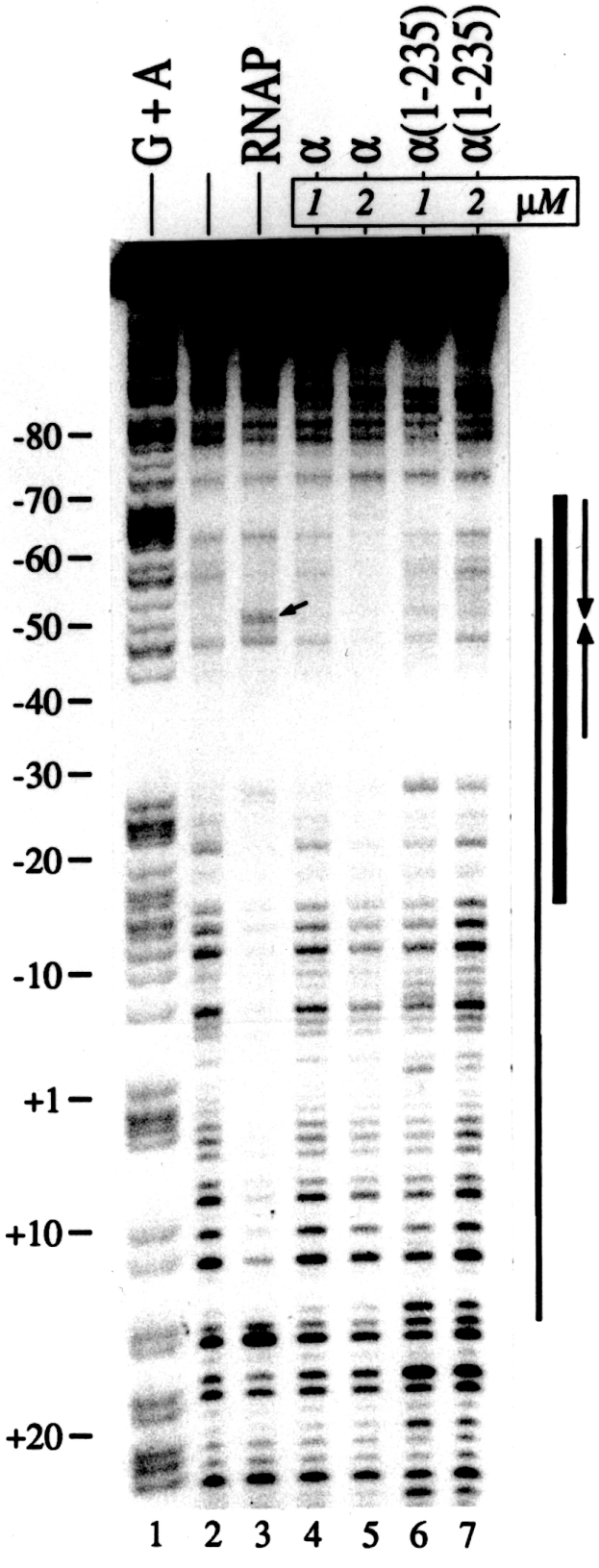


Fig. 10: DNase I footprinting analysis of the P_1 promoter. A 182 bp *Nco*I-*Sma*I DNA fragment obtained from plasmid pGEMT-P_1 (Table 3), that comprises the P_1 promoter region was labeled at its *Sma*I site and digested with DNase I (0.002 U) either in the absence of protein or in the presence of 1 Unit *E. coli* RNA polymerase (RNAP) (lane 3), or the indicated amounts of wild type α (lanes 4, 5) or C-terminally truncated α(1-235) (lanes 6, 7). The region of DNase I protection by RNAP is indicated by a vertical thin line on the right side of the figure. The DNA region protected by α is indicated by a vertical black bar, and the position of the 33 bp non perfect inverted repeat is indicated by arrows. The DNase I hypersensitive site is marked by an arrow (lane 3). Numbers on the left side of the figure indicate nucleotide positions with respect to the transcriptional start site (+1). G+A indicates a Maxam and Gilbert sequence reaction.

2. Transcriptional regulation of flagellar genes

2.1 Identification of flagellar genes transcribed by σ^{54} promoters

Analysis of the complete *H. pylori* genome sequence has revealed the presence of more than 40 putative proteins that are involved in expression, secretion and assembly of the complex flagellar apparatus (Tomb *et al.*, 1997). About 20 of these proteins constitute the structural components of the flagellar basal body, hook and filament. To gain insight into the regulation of the genes coding for these flagellar structural proteins the genome sequence was analysed in detail regarding the probable transcriptional organisation of these genes and the putative promoter sequences that might be responsible for their expression.

gene/operon	number[a)]	putative function	promoter [b)]
O-*fliF*-*fliG*-*fliH*-Δ-Δ[c)]-O[d)]	350-356	M-ring protein, motor switch, export protein	σ^{80}
fliA-*fliM*-*fliY*-O-O	1032-1028	Sigma F, motor switch, ?	?
flgB-*flgC*	1559-1558	proximal rod proteins	σ^{54}
O-*flgD*-*flgE'*-O-Δ-Δ	906-911	hook assembly, hook protein homolog	σ^{54}
flgH-Δ-*flaG*	325-327	L-ring protein, flagellar protein G	σ^{80}
flaG-*fliD*	751-752	polar flagellin, flagellar hook associated protein 2	σ^{80}
O-Δ-*motA*-*motB*-O	813-819	motor rotation proteins	σ^{80}
Δ-*flhB*-Δ-Δ	1574-1577	FlhB protein	σ^{80}
flgI-O-Δ	246-244	P-ring protein	σ^{80}
O-*flgK*	1120-1119	hook associated protein (HAP1)	σ^{54}
flaB	115	flagellin B	σ^{54}
flaB'	295	flagellin B homolog	σ^{54}
fliE	1557	basal body protein	σ^{80}
flgE	870	hook protein	σ^{54}
flgG	1585	basal body protein	σ^{80}, σ^{54}
fliN	584	motor switch	σ^{80}
flaA	601	flagellin A	σ^{28}
flgG	1092	basal body rod protein	σ^{80}
fliS-O	753-754	flagellar protein	?

Table 5: Putative flagellar genes and operons and associated promoter sequences.

a) numbers refer to the annotated sequence published by Tomb *et al*. (1997).

b) assignment to the three different sigma factors of *H. pylori* was according to similaritites to the consensus sequences recognised by $E\sigma^{70}$ (TTGACA-N_{17}-TATAAT), $E\sigma^{54}$ (GG-N_{10}-GC), and $E\sigma^{28}$ (TAAA-N_{15}-GCCGATAA).

c) the "Δ"symbol indicates putative genes with functions unrelated to flagellar biosynthesis.

d) the "O"symbol indicates open reading frames with unknown functions.

Table 5 shows the result of this analysis. Unlike the situation in *Salmonella* spp. and other enteric bacteria, where the flagellar structural genes are organised in large operons, the *H. pylori* genes are scattered throughout the chromosome. Some of them seem to be expressed as single transcriptional

units while others appear to be contained in operon-like structures. Most of the genes and operons possess sequence motifs similar to the consensus sequence recognised by $E\sigma^{70}$ in their 5' untranslated regions. These genes are likely to be transcribed by RNA polymerase containing σ^{80}, the *H. pylori* homologue of the vegetative sigma factor σ^{70} from *E. coli* (Beier *et al*., 1998; and E.1). Some of the flagellar structural genes and operons however lack σ^{70} recognition sequences and show instead GG-N_{10}-GC motifs in their 5' untranslated regions, suggesting that these genes might be regulated by RNA polymerase containing the alternative sigma factor σ^{54}. Fig. 11 shows the structural organisation of five of these putative σ^{54} dependent genes and operons in detail.

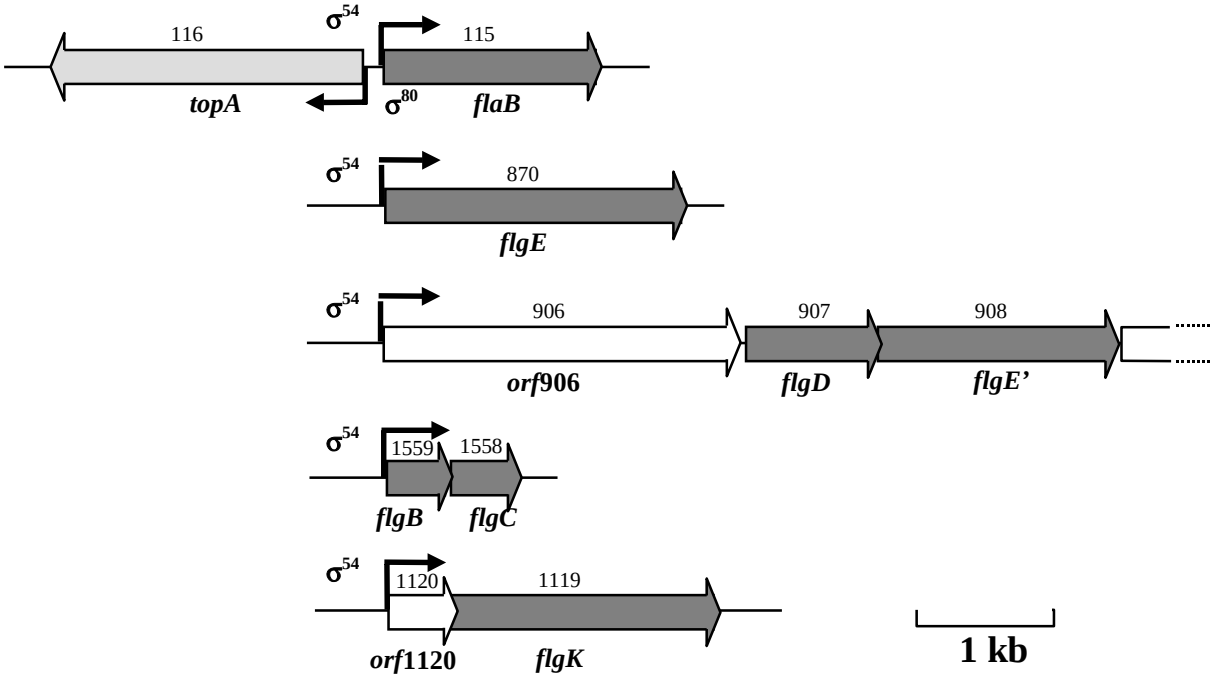


Fig. 11: **Structural organisation of σ^{54} regulated genes and operons**. Dark grey arrows indicate flagellar genes, white arrows ORFs of unknown function. The gene encoding Topoisomerase A is indicated by a light gray arrow. Numbers refer to the annotated genome sequence published by Tomb *et al*. (1997). *flgD* and *flgE'* encode proteins with 25.5% and 30.5% amino acid identity to the flagellar hook assembly protein FlgD of *Salmonella choleraesuis* (Katsukake and Doi, 1994) and the structural hook protein FlgE of *H. pylori* (O'Toole *et al*., 1994), respectively. *flgB* and *flgC* constitute an operon encoding two proteins with 31% and 46% amino acid identity to the proximal flagellar rod proteins FlgB of *S. choleraesuis* (Homma *et al*., 1990a) and FlgC of *Borrelia burgdorferi* (Ge *et al*., 1997), respectively. *flgK* encodes a protein with 27.6% amino acid identity to the flagellar hook-associated protein FlgK (HAP1) of *S. choleraesuis* (Homma *et al*., 1990b).

Besides *flaB*, for which an RNA 5' end has been mapped downstream from a σ^{54} consensus sequence (Suerbaum *et al*., 1993) and *flgE* for which a σ^{54} consensus sequence has been noted (O'Toole *et al*., 1994), three additional putative operons are represented. The first operon consists of an ORF of unknown function (*orf*906), one gene, *flgD*, encoding the flagellar hook assembly protein FlgD and one gene encoding a homologue of the structural hook protein FlgE. The second operon is constituted of two genes encoding the proximal flagellar rod proteins FlgB and FlgC. The

third operon contains one open reading frame of unknown function (*orf*1120) and the gene encoding the flagellar hook-associated protein FlgK. It is likely that these five putative σ^{54}-recognised promoters regulate expression of seven basal-body and hook genes of *H. pylori*.

2.2 Identification of the transcriptional activator of the σ^{54}-dependent genes and operons

In all cases up to now described, transcription from σ^{54}-recognised promoters is dependent on distant upstream (> 100 bp) enhancer-like sequences, which are bound by a class of transactivating proteins, usually referred to as the NtrC (NR-I) family (Kustu *et al*., 1991). Therefore, the transcriptional regulator with the highest homology to this class of proteins was selected from the *H. pylori* genome to study transcription from the putative σ^{54} promoters. Interestingly, the selected protein (named HP703 in Tomb *et al*., 1997; Table 1) shows also a high degree of similarity to proteins involved in regulation of flagellar biosynthesis. These include FlrC (43.0% identity in a 386 amino acid overlap) of *Vibrio cholerae* and FlbD (35.5% identity in a 377 amino acid overlap) of *Caulobacter crescentus*. As a preliminary test for the role of this putative regulator in transcriptional activation of flagellar genes, an HP703 isogenic mutant was constructed and assayed for bacterial motility. For that purpose a plasmid containing a kanamycin cassette flanked by the gene located upstream of *hp703* on one side and a distal part of the *hp703* coding region on the other side (pGEM3(*flgR::km*), Table 3; Fig. 12A) was transformed into the naturally competent *H. pylori* strain G27. Correct replacement of the wild type sequence with the antibiotic resistance gene was verified by PCR using oligonucleotides ntr8/ntr9 (Table 4) complementary to regions flanking the insertion site and Southern Blot analysis using *Hae*III digested chromosomal DNA of the mutant and of the wild type strains, respectively (Fig 12B). The mutant had bp 49 to 585 of the *hp703* coding sequence substituted with the kanamycin cassette. Bacterial motility was investigated by assaying the capability of the cells to spread on soft agar plates. Fig. 13 shows that the HP703 knockout mutant strain (marked G27[*flgR*$^-$], see below) grew by spreading in an area that is severely reduced with respect to the one covered by the wild type strain. In contrast, this mutant strain grew similarly to the previously characterised non-motile CheY mutant strain (Beier *et al.*, 1997; data not shown), thus showing a loss of motility functions.

Possibly, the NtrC homologous protein is the specific master activator of the hypothetical σ^{54}-regulated flagellar genes. Therefore, it is proposed to name this protein FlgR (for Flagella Regulatory protein).

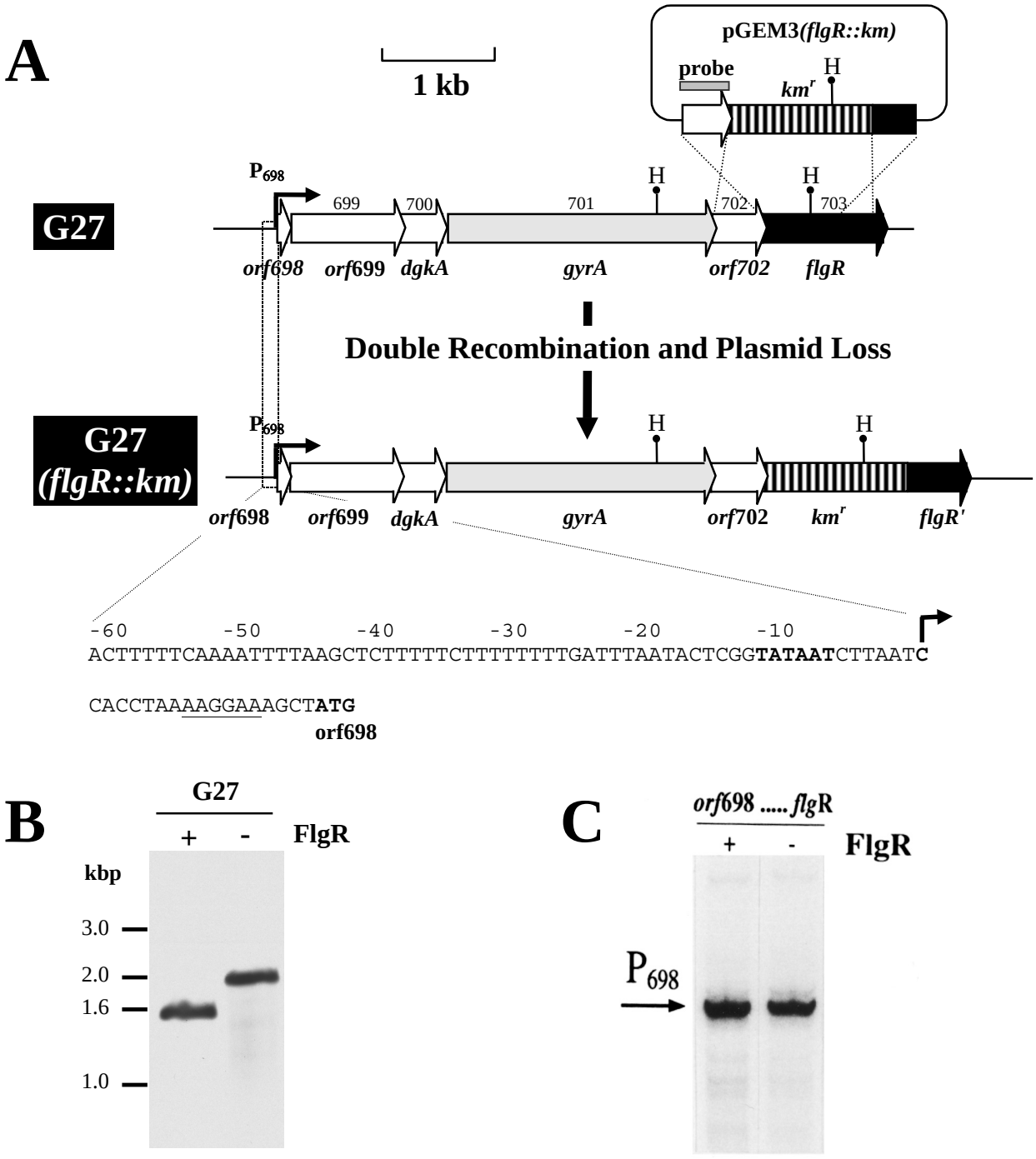
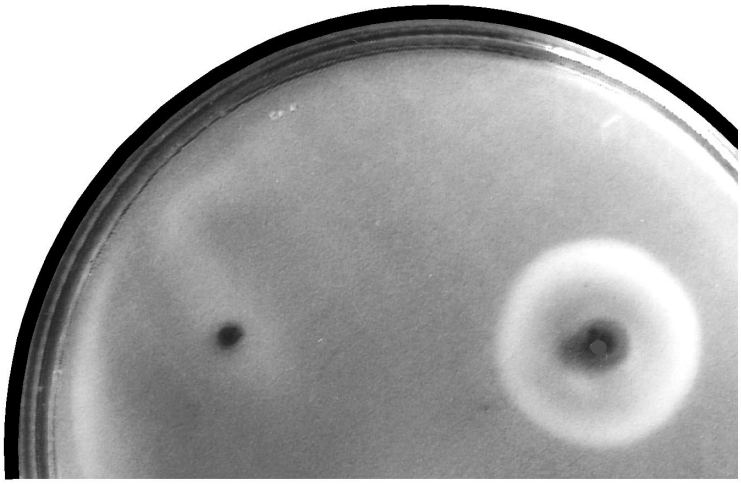


Fig. 12: A) Construction of the G27(*flgR::km*) mutant strain. Arrows point to the direction of transcription of the indicated genes. The gene coding for Gyrase A is indicated by a light grey arrow, the kanamycin cassette by boxes with vertical bars. White arrows indicate ORFs of unknown or unrelated function. *Hae*III restriction sites are labeled "H". The position of the PCR probe used for Southern Blot analysis is also indicated. Numbers refer to the annotated genome sequence published by Tomb *et al.* (1997). The putative protein encoded by *orf*703 exhibits significant amino acid sequence homology to a series of NtrC-like regulators. The highest degree of amino acid similarity (44.2% identity, 69.1% similarity) is with FleQ, an activator of mucin adhesion and flagella expression in *Pseudomonas aeruginosa* (Arora *et al.*, 1997). **B) Southern Blot analysis.** Chromosomal DNA of strains G27 (lane +) and G27(*flgR::km*) (lane -) was digested with *Hae*III and hybridised to a PCR fragment obtained with oligonucleotides ntr8 and ntr9 (Table 4). The probe hybridises with a 1560 bp fragment in the wild type and a 2040 bp fragment in the mutant strain. **C) Primer extension analysis.** Total RNA was isolated from strains G27 (lane +) and G27(*flgR::km*) (lane -), annealed to oligonucleotide gyrN (Table 4), and elongated with reverse transcriptase. The elongated primer is indicated by an arrow and labeled P_{698}. The promoter region of the FlgR encoding operon was PCR amplified with oligonucleotides gyr1/gyr2 (Table 4), cloned into pGEM3 (thus giving rise to plasmid pGyrA; see Table 3) and sequenced in parallel with the primer extension reaction.



G27 [*flg*R⁻] G27

Fig. 13: Bacterial motility assay. The indicated strains were stabbed into semi-solid agar medium and incubated at 37 °C for 72 h.

2.3 Transcriptional analysis of the *flgR* gene

Fig. 12A shows that the *flgR* gene is located at the 3' end of a putative operon which contains the gene encoding diacylglycerol kinase (*dgkA*), the gene encoding the subunit A of DNA gyrase (*gyrA*) as well as three open reading frames of unknown function (*orf*698, *orf*699, *orf*702). Unexpectedly, no cognate sensor kinase can be detected in this operon. In order to assess the transcriptional regulation of FlgR expression, a series of oligonucleotides complementary to the 5' ends of the coding sequences of *flgR*, *orf*702, *gyrA* and *orf*699 were designed (oligos ntrN, ntrY, gyrA, and gyrN, respectively; Table 4), and primer extension analysis of total RNA extracted from *H. pylori* G27 was performed. While reverse transcription of primers complementary to *flgR*, *orf*702 and *gyrA* yielded numerous bands migrating to different positions in the gel (data not shown), extension of the *orf*699 specific primer revealed one major band of high molecular weight, indicating the presence of a promoter upstream of *orf*698 (Fig. 12A). For a more precise mapping of this putative promoter, the 5' region of *orf698* was amplified by PCR from strain G27 with oligonucleotides gyr1 and gyr2, cloned into pGEM3 (giving rise to plasmid pGyrA, Table 3) and sequenced in parallel with a second primer extension experiment using the same oligonucleotide as before. This experiment placed the P_{698} transcription start site of the operon at a C nucleotide 18 nt upstream of the ATG start codon of translation (Fig. 12C). Nucleotide sequence analysis upstream of the transcription start site revealed the presence of a -10 region with a perfect match to the *E. coli* σ^{70} consensus sequence (TATAAT), an AT rich region (86% AT) spanning positions -18 to -60, and no consensus for a -35 region (Fig. 12A). This promoter thus resembles to a great extent the P_1 promoter of *cagA*, which is recognised by RNA polymerase containing σ^{80}, the vegetative sigma factor of *H. pylori* (see E.1).

In order to assess if transcription from P_{698} is affected by the FlgR protein a further primer extension experiment using the same oligonucleotide and total RNA extracted from the FlgR$^-$ mutant strain was then performed. Reverse transcription of the oligonucleotide yielded the same extension product in an amount indistinguishable from that obtained with RNA from the wild type strain (Fig. 12C), suggesting that the FlgR protein does not influence transcription of the operon.

2.4 FlgR as the master regulator of basal body and hook genes

To define the start point of transcription of the hypothetical σ^{54}-regulated *flaB, flgE, orf906-flgDE', flgBC* and *orf1120-flgK* genes, primer extension assays on total *H. pylori* RNA were then carried out. The oligonucleotides complementary to these coding sequences are shown in Table 4 and the legend to Fig. 14. To assure the correct mapping of the extension products, the respective promoter regions of these loci were cloned from the chromosomal DNA of *H. pylori* G27 and sequenced in parallel with the primer extension assays.

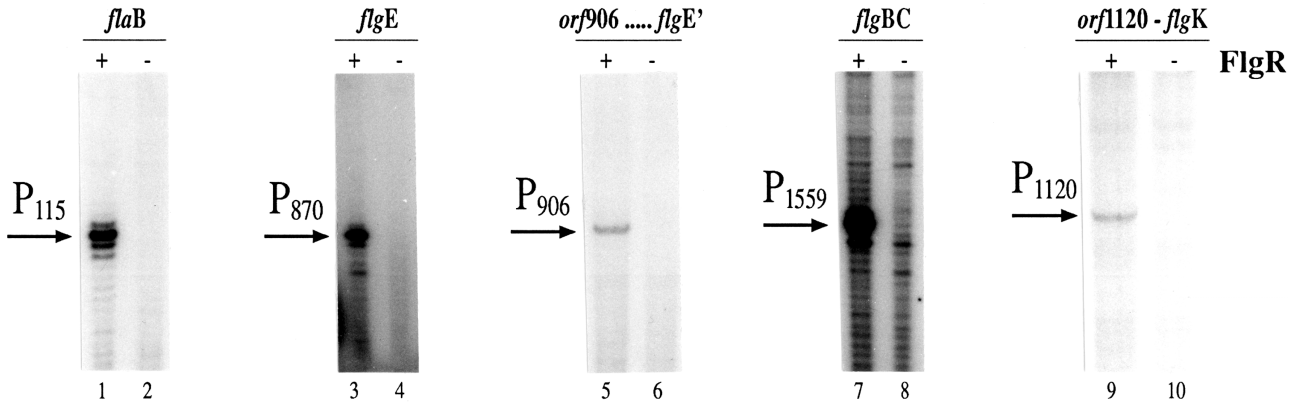


Fig. 14: Primer extension analyses of *H. pylori* RNAs. The elongated primers are indicated by arrows and marked "P". Lowerscript numbers refer to the first downstream *orf* as designated in Tomb *et al.* (1997). Total RNA was extracted from wild type *H. pylori* strain G27 (lanes +) or from the *flgR*$^-$ isogenic mutant (lanes -), annealed to specific primers and elongated with reverse transcriptase. Primers used were: flaB in lanes 1,2; flgE in lanes 3,4; flgE2 in lanes 5,6; flgB in lanes 7,8; flgK in lanes 9,10 (Table 4). Promoter fragments were PCR amplified with oligonucleotide pairs fla1/fla2, flgE5/flgE6, flgE3/flgE4, flgB1/flgB2, and flgK1/flgK2 (Table 4) and cloned into vector pGEM3, thus giving rise to plasmids plasmids pFlaB, pFlgE, pFlgD, pFlgB, and pFlgK, respectively (Table 3). The inserts of these plasmids were then sequenced in parallel with the primer extension reactions to ensure the correct mapping of the transcription start sites.

The urea/acrylamide gels in Fig. 14 show the primer extension experiments carried out by hybridising each oligo to total RNA extracted from *H. pylori*. Each reverse transcription of RNA shows a major band, which defines the indicated P start site of RNA transcription (lanes 1, 3, 5, 7, 9). In the case of *flaB*, the experiment located the start site of RNA transcription at 25 nt upstream

from the ATG start codon, which defines the P_{115} promoter. Transcription of *flgE*, directed by promoter P_{870}, started at 30 nt upstream of the ATG start codon, while *orf906-flgDE', flgBC*, and *orf1120-flgK* transcripts mapped at 33 nt, 22 nt and 24 nt upstream of the respective ATG start codons defining promoters P_{906}, P_{1559} and P_{1120}, respectively.

To demonstrate that FlgR is the transcriptional regulator of the basal-body and hook genes, total *H. pylori* RNA was extracted from liquid cultures of the G27[flgR$^-$] mutant and primer extension analyses were carried out. As shown in Fig. 14 (lanes 2, 4, 6, 8, 10) on neither of the σ^{54}-dependent flagellar genes a primer extension product could be detected, thus demonstrating the positive role of FlgR on transcription of these genes.

Nucleotide sequences of these promoters have been aligned with respect to the initiation site of transcription with results shown in Fig. 15. As expected, conserved sequences at positions -12 (GC) and -24 (GG) of the core consensus of σ^{54}-regulated promoters are found. In addition, this core consensus could be extended to conserved nucleotides both upstream and downstream from the core consensus.

```
            -60        -50        -40              -24          -12           +1
flaB        TAAAAACCATAACAGCGTTAAAAATCAAAGAGTTGGAACACCCTTTGCTTGACTAACAGCAAAT
flgE        AAATCCGCAACCAATCATTCTAAAAAGCTATTTAGGAACAACTTTGCTTTATTTTGCATAGAT
flgBC       ATCGCATTCTTTGATGAGGGAATGAGCTATTTTTGGAATGTTTTTTGCATACTAAAAATTTGCA
orf-flgDE'  CTTTAGCATTTTTAGATTTGATCACGCATCTTAAGGAACTCTTTTTGCTTAACTCCCACCATGT
orf-flgK    TTAAGAAATTGAGTTTTATAATAAAACTAAAAATGGCATAGTATTTGCTTGAATACGGCTGAAA

consensus   .W.......W...W......WW.R....W.W.WWGGNAYNNYNTTTGCWT..........
```

Fig. 15: Nucleotide sequences of *H. pylori* σ^{54} promoters. Shown is an alignment of the sequence from positions +1 to -60 of the indicated promoters and a derived consensus sequence. The σ^{54} core promoter region is shaded; identical bases are in bold face. Symbols: W = A or T; Y = C or T; R = A or G; N = G, A, T, or C.

2.5 Binding of FlgR to the *flaB* promoter

To demonstrate that FlgR regulates the σ^{54}-dependent flagellar operons by directly interacting with the respective promoter regions, the protein was purified in a recombinant form and used in *in vitro* binding studies. The coding sequence of the *flgR* gene was therefore amplified by PCR from chromosomal DNA of *H. pylori* strain G27 using oligonucleotides flgRN/flgRC (Table 4) and cloned into the expression vector pTrcHisA. The resulting plasmid, pTrcA-flgR (Table 3), which contains the entire *flgR* coding sequence fused to six histidine codons at its 5' end, was then transformed into *E. coli* strain BL21 and the recombinant [His_6]-FlgR protein was purified in its

native form as described in D.17. Purity was estimated to be 90% according to SDS acrylamide gel electrophoresis (Fig. 16A).

A 320 bp *Eco*RI-*Bam*HI fragment obtained from plasmid pFlaB (Table 3) and comprising the intergenic region between *flaB* and *topA* was then radioactively labeled and used in gel retardation experiments with the recombinant FlgR protein. As shown in Fig. 16B incubation of the probe with 2 µg protein (lane 3) yielded a complex with reduced electrophoretic mobility (marked I in Fig. 16B), thus suggesting a specific binding of FlgR to the *flaB-topA* intergenic region. As protein concentration increased (lanes 4-6) this complex became more pronounced and a second specific complex with further reduced mobility (II) appeared. Interestingly, no difference in binding specificity could be observed between non-phosphorylated (lanes 2-6) and phosphorylated (lanes 7-11) protein, indicating that binding efficiency of FlgR on the *flaB* promoter does not depend on its phosphorylation state.

FlgR therefore regulates transcription of the *flaB* gene by directly interacting with its promoter region. It is likely that also the other four σ^{54} dependent flagellar operons are regulated in a similar way by FlgR.

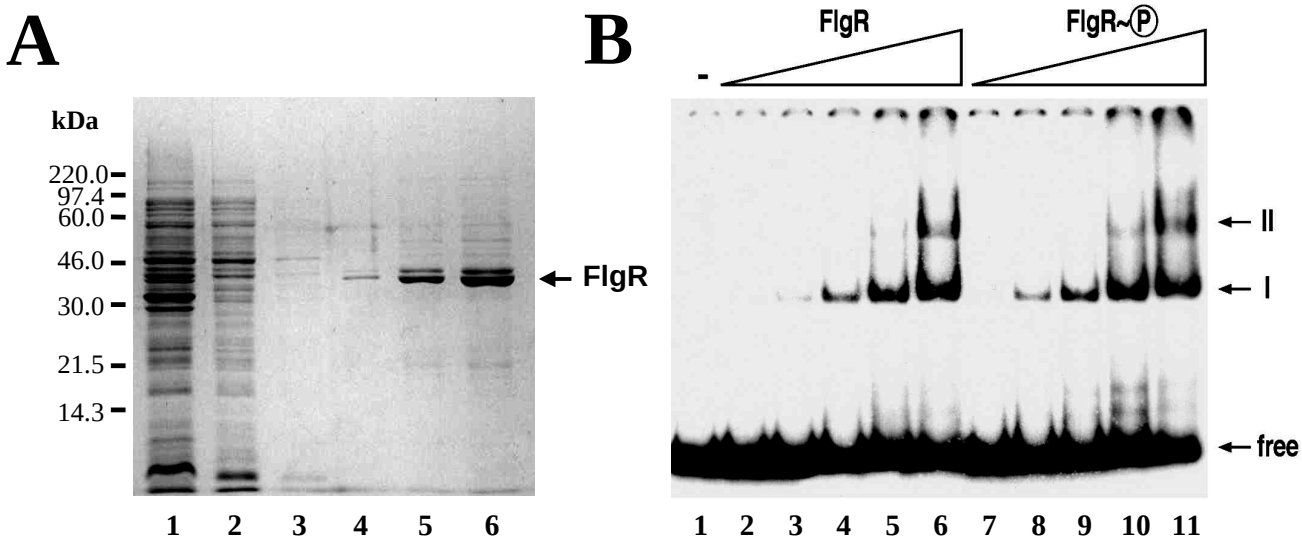


Fig. 16: A) Purification of FlgR. The [His_6] tagged FlgR protein was purified from induced cultures of DH5α pTrcA-flgR (Tables 2,3). lane 1: 50 µl cleared lysate of induced culture, lane 2: 50 µl column flow through, lane 3: 100 µl wash; lanes 4-6: 1, 5, and 10 µl eluted protein. Fractions were loaded on a 12.5% SDS-polyacrylamide gel. **B) Gel mobility shift assay with purified FlgR on the *flaB* promoter.** A 320 bp *Eco*RI-*Bam*HI fragment obtained from plasmid pFlaB (Table 3) and comprising the intergenic region between *flaB* and *topA* was incubated with increasing amounts of purified FlgR protein. Lane 1: no protein added; lanes 2-6: 1, 2, 4, 8, 16 µg purified FlgR added; lanes 7-11: like lanes 2-6 but with addition of 25 mM acetylphosphate. Protein-DNA complexes are indicated on the right side of the figure (I, II). All lanes contained 1000 ng sonicated salmon sperm DNA as non-specific competitor.

2.6 Effect of FlgR on *flaA* transcription

In order to assess the possible influence of FlgR on *flaA* transcription, which is regulated by the alternative sigma factor σ^{28} (Leying *et al*., 1992), total RNA extracted from *H. pylori* G27 and from the isogenic *flgR*$^-$ mutant strain was hybridised to an oligonucleotide specific for *flaA*. The primer extension experiment shown in Fig. 17 (lanes 1 and 2) shows a typical result. In both cases one major band (P_{601}) which corresponds to the σ^{28}-dependent *flaA* transcription start site at 50 nt upstream of the *flaA* ATG start codon reported by Leying *et al.* (1992) can be detected. Surprisingly, the intensity of the P_{601} band was higher in the mutant strain (lane 2) as compared to the wild-type strain (lane 1), suggesting that the P_{601} RNA is increased in the FlgR$^-$ mutant strain. To assess the difference in the amount of RNA, the relative intensities of the bands were evaluated by exposure of the same gel to a PhosphorImager. The evaluation revealed a two-fold increase in the amount of transcript in the mutant strain, suggesting that FlgR exerts a negative feedback on FlaA expression by specifically repressing transcription of the *flaA* gene. On the contrary, a primer extension experiment carried out on the *cagA* gene, which was demonstrated to be regulated by the σ^{80}-dependent P_1 promoter (see E.1), showed no difference in transcript size and/or amount between the wild-type and the mutant strain (Fig. 17, lanes 3, 4). Similarly, expression of the *ureAB* operon encoding the two subunits of the *H. pylori* urease enzyme was not altered in the mutant strain (Fig. 17, lanes 5, 6). Transcription of this operon was initiated at the P_{73} promoter whose transcription starts at two C nucleotides mapping 56 and 57 nt upstream of the *ureA* ATG translational start codon. Seven nucleotides upstream of the transcription start a *E. coli* Eσ^{70} -10 extended-like promoter (TGCTACAAT) with only one mismatch with respect to the *E. coli* consensus sequence (TGNTATAAT) could be detected (Kumar *et al*., 1993), indicating a σ^{80}-regulated transcription of *ureAB* rather than a σ^{54}-dependent expression as proposed previously (Labigne *et al*., 1991).

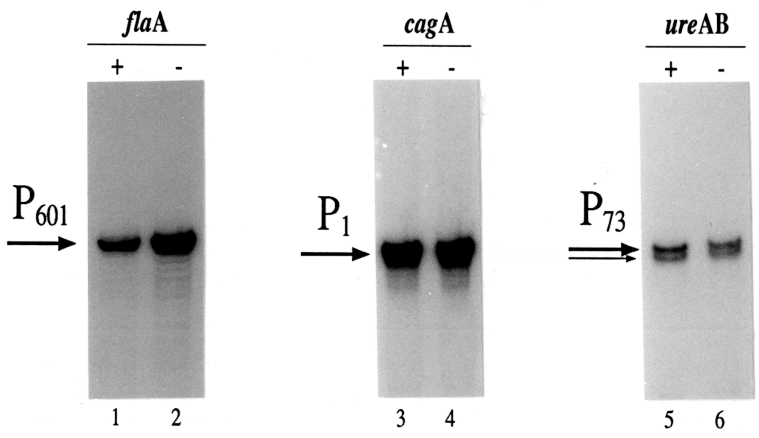


Fig. 17: **Transcriptional regulation of *flaA* by FlgR.** Primer extension experiments were performed with the same RNA preparations as in Fig 14. Primers used were: flaA in lanes 1,2; cagN in lanes 3,4; ureA in lanes 5,6 (Table 4). Symbols are as in Fig. 14.

2.7 Effect of DNA topology on *flaB* transcription

Since direction of transcription of the *flaB* gene is in the opposite direction to the *topA* gene (Fig. 11), the two promoters result in an overlapping and divergent configuration (Suerbaum *et al.*, 1998). To assess whether the FlgR regulated *flaB* promoter is sensitive to DNA topology, liquid cultures of *H. pylori* G27 were incubated with the DNA gyrase inhibitor novobiocin and the accumulation of transcripts at this promoter was assayed by primer extension analyses. Fig. 18 shows a primer extension experiment carried out with specific oligonucleotides on RNA extracted before and after addition of novobiocin to the culture medium. Following addition of novobiocin, the amount of *flaB* transcript (P_{115}) changed with time (lanes 0', 15', 30'). The P_{115} transcript increased after 15 min treatment and decreased after 30 min treatment. This suggests an immediate positive response to the perturbation followed by a negative response. In contrast, the amount of *flaA* transcript (P_{601}) was not changed by the addition of novobiocin to the culture medium (Fig. 18), thus confirming a specific response of promoter P_{115} to novobiocin treatment.

It is likely that transcription from the P_{115} (*flaB*) promoter is sensitive to changes in DNA topology, suggesting a complex mechanism of regulation of the *flaB* and *topA* genes.

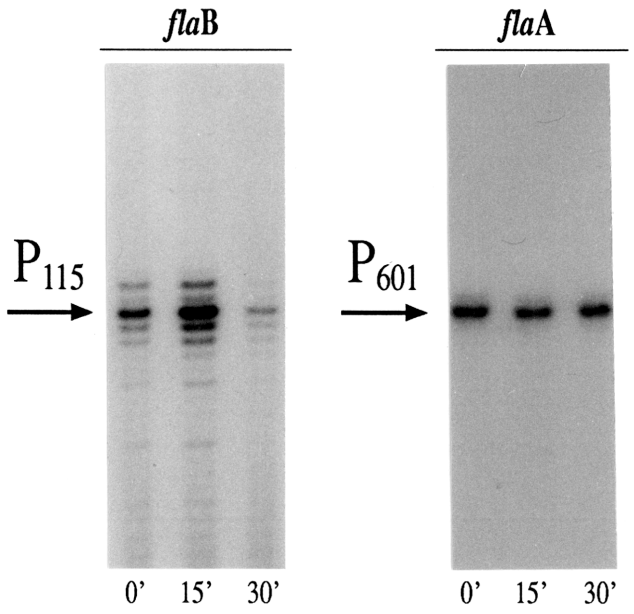


Fig. 18: Novobiocin affects transcription of *flaB* but not *flaA*. Total RNA was extracted from *H. pylori* strain G27 before addition of novobiocin (time 0') and after 15 and 30 min as indicated.

3. Transcriptional regulation of chaperone genes

3.1 Determination of transcriptional start sites

The structural organisation of the major *H. pylori* chaperone genes according to the genome sequence published by Tomb *et al.* (1997) is schematically represented in Fig. 19. Besides the *groESL* locus, which has been previously characterised (Macchia *et al.*, 1993; Suerbaum *et al*., 1994), two other putative operons are shown. One of these operons comprises a homologue of the *E. coli cbpA* gene, coding for a protein with functional similarities to DnaJ (Ueguchi *et al.*, 1994), an *orf* that codes for a protein with sequence similarities (46% identity in a 91 aa overlap) to HspR, a repressor protein of *Streptomyces coelicolor* known to be involved in negative regulation of heat shock genes (Bucca *et al*., 1995), and another *orf* that codes for a putative helicase (Tomb *et al.*, 1997). The other operon contains a homologue of the *B. subtilis hrcA* gene (Wetzstein *et al*., 1992), and the genes encoding GrpE and DnaK (Hsp70).

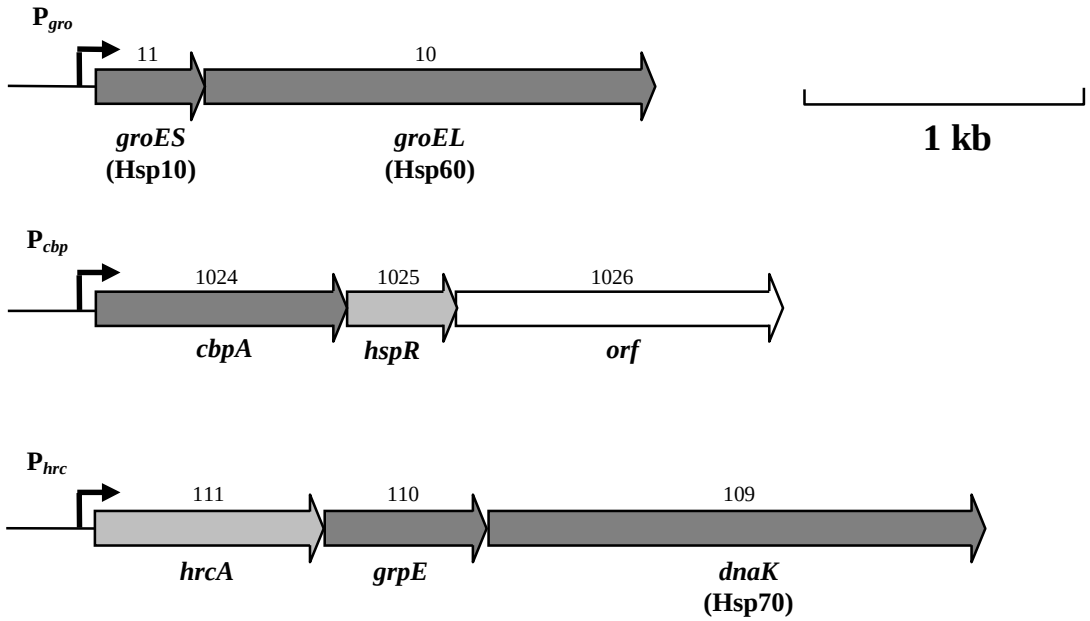


Fig. 19: **Structural organisation of *H. pylori* chaperone genes.** Dark grey arrows indicate chaperone genes, light grey arrows regulatory genes. All genes are marked according to the genome sequence published by Tomb *et al.* (1997). *groES* and *groEL* code for the HspA (Hsp10) and HspB (Hsp60) proteins (Suerbaum *et al*., 1994); *cbpA* encodes a protein with 30% amino acid identity (in a 288 aa overlap) to the co-chaperone-curved DNA binding protein CbpA from *E. coli*, a homologue of DnaJ (Ueguchi *et al*., 1994); *hspR* encodes a regulatory protein with 46% identity (in a 91 aa overlap) to HspR, the negative regulator of heat shock genes in *Streptomyces coelicolor* (Bucca *et al*., 1995); *orf* encodes a protein with 30% identity (in a 421 aa overlap) to a hypothetical helicase-like protein from *Haemophilus influenzae*; *hrcA* encodes a protein with 28% identity (in a 71 aa overlap) to the heat inducible transcriptional repressor HrcA from *Bacillus subtilis* (Wetzstein *et al*., 1992); *grpE* and *dnaK* encode the GrpE and DnaK (Hsp70) chaperones, respectively.

In order to define the start points of transcription of these operons, total RNA from *H. pylori* G27 was extracted and primer extension or S1 nuclease mapping experiments were carried out with specific primers or labeled fragments as indicated in Table 4 and the legend to Fig. 20. To ensure correct mapping of the extension or digestion products, the respective promoter regions were cloned from the chromosomal DNA of *H. pylori* G27 and sequenced in parallel.

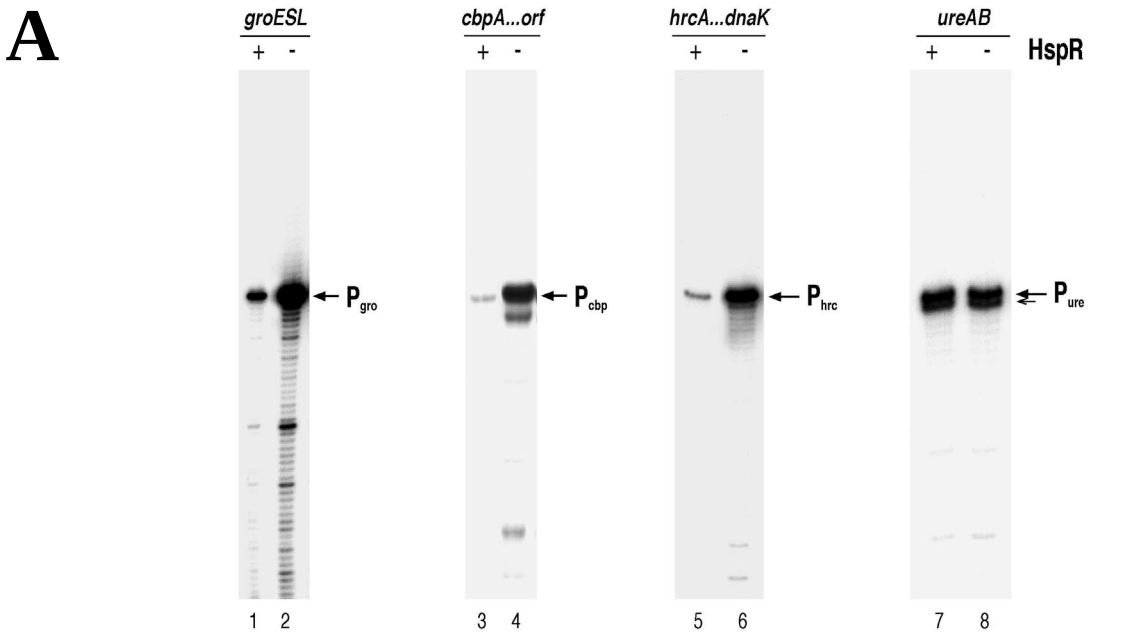


Fig. 20: A) Primer extension and S1 nuclease mapping analyses of *H. pylori* RNA. Elongated primers or protected fragments are indicated by arrows and labeled "P". Lowerscript letters refer to the first of the downstream genes (Fig. 19). Total RNA was isolated from wild type *H. pylori* G27 (lanes +) or from the isogenic *hspR*$^-$ mutant (lanes -), annealed to specific primers and elongated with reverse transcriptase or hybridised to specific DNA fragments and digested with S1 nuclease. The primer extensions in lanes (1,2), (5,6) and (7,8) were performed with oligonucleotides groS, hrcA and ureA (Table 4), respectively. For the S1 nuclease mapping in lanes (3,4) the 674 bp *Eco*RI-*Bam*HI fragment of plasmid pCbp12 (Table 3), labeled at its *Eco*RI site was used. For correct mapping of the transcription initiation sites the chaperone promoter fragments were PCR-amplified from chromosomal DNA of strain G27 with primer pairs gro1/gro2, cbp3/cbp4, and hrc1/hrc2 (Table 4), cloned into pGEM3 (thus giving rise to plasmids pGroE, pCbp34, and pHrcA; see Table 3) and sequenced in parallel with the primer extension and S1 nuclease mapping experiments. **B) Nucleotide sequences of chaperone gene promoters**. Shown is an alignment of the sequences with respect to the transcriptional start sites (+1, bold and double underlined). Putative -10 and -35 hexamers are boxed, mismatches with respect to the Eσ^{70}-recognised consensus sequence are underlined. A TG motif typical of -10 extended promoters (Kumar *et al*., 1993; see E.1, F.1) is also boxed. It should be noted that the spacing between the -10 and -35 hexamers is 17 bp in the case of the P_{gro} and P_{cbp} promoters and 19 bp in the case of the P_{hrc} promoter.

The urea-acrylamide gels in Fig. 20A show the results of the primer extension and S1 nuclease mapping experiments carried out on *H. pylori* RNA. As shown in lane 1 the reverse

transcription of a *groES* specific oligonucleotide yielded a major band that places the P_{gro} start of transcription at 59 bases upstream of the *groES* translational start codon. Similarly the S1 mapping experiment with a *cbpA* specific DNA probe (lane 3) and the reverse transcription of a *hrcA* specific oligonucleotide (lane 5) place the P_{cbp} and P_{hrc} transcriptional start sites at 27 and 15 nucleotides upstream of the *cbpA* and *hrcA* translational start sites, respectively. Nucleotide sequence analysis of the DNA regions upstream of the three transcription start sites (Fig. 20B) revealed the presence of -10 regions with a high degree of similarity (max. one mismatch) to the the *E. coli* $E\sigma^{70}$ consensus sequence (TATAAT). Well conserved -35 regions with only one (P_{gro}) or two mismatches (P_{cbp} and P_{hrc}) with respect to the $E\sigma^{70}$ consensus (TTGACA) could also be detected.

These results taken together with the apparent lack of a σ^{32} homologue in the *H. pylori* chromosome (Tomb *et al*., 1997) strongly suggest that the promoters transcribing the chaperone genes are regulated by RNA polymerase containing σ^{80}, the *H. pylori* homologue of the vegetative sigma factor σ^{70} from *E. coli.* In fact, an *in vitro* transcription experiment on the P_{gro} promoter using purified $E\sigma^{70}$ RNA polymerase and a hybrid RNA polymerase containing α, β and β' from *E. coli* and σ^{80} from *H. pylori*, yielded the same transcriptional start site as *in vivo* (data not shown).

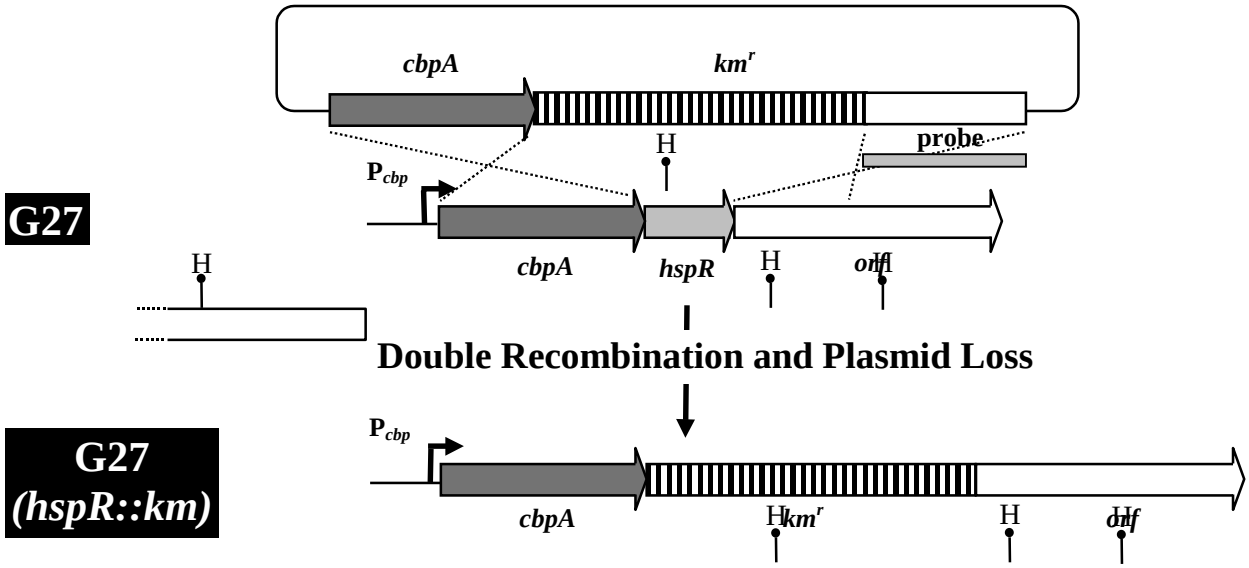
3.2 Identification of the transcriptional regulator of the chaperone genes

In order to study the regulation of the P_{gro}, P_{cbp} and P_{hrc} promoters an isogenic mutant of the *H. pylori* HspR homologue contained in the *cbpA* operon (Fig. 19) was constructed. For that purpose a plasmid containing a kanamycin cassette flanked by the *cbpA* gene on one side and a proximal part of *orf1026* on the other side (pGEM3(*hspR::km*), Table 3; Fig. 21A) was transformed into the *H. pylori* host strain G27. By homologous recombination with this plasmid most of the *hspR* coding sequence was replaced with a kanamycin resistance gene (Fig. 21A). Correct replacement of the wild type sequence with the antibiotic resistance cassette was verified by means of PCR using oligonucleotides complementary to regions flanking the insertion site (hspRN/hspRC, Table 4; data not shown) and by Southern Blot analysis using *Hae*III digested chromosomal DNA of the wild type and mutant strains, respectively, and the hsp3/hsp4 PCR fragment as a probe (Fig. 21B).

A

pGEM3*(hspR::km)*

1 kb



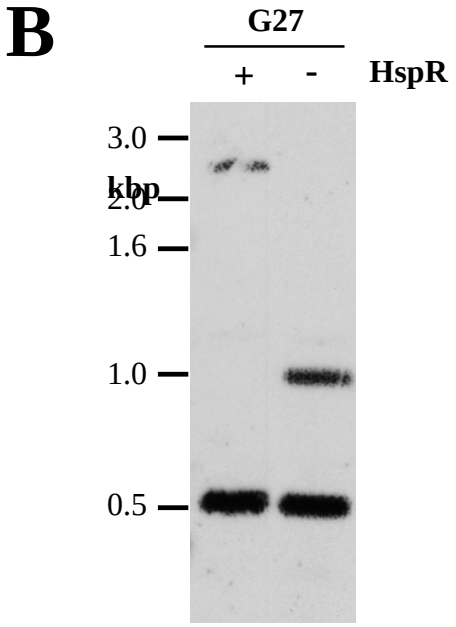


Fig. 21: A) Construction of an isogenic *hspR*⁻ mutant. Arrows point to the direction of transcription of the indicated genes. Dark grey arrows indicate chaperone genes, light grey arrows regulatory genes and open boxes ORFs with putative function. The kanamycin cassette is represented by boxes with vertical bars. Relevant *Hae*III restriction sites are labeled "H". The position of the PCR fragment used as a probe in Southern Blot is also shown. **B) Southern Blot analysis.** Chromosomal DNA was isolated from strains G27 (lane +) and G27(*hspR::km*) (lane -), digested with *Hae*III and hybridised to a PCR fragment obtained with oligonucleotides hsp3 and hsp4 (Table 4). The probe hybridises with two 502 bp and 2510 bp *Hae*III fragments in the G27 wild type strain and two 502 bp and 972 bp *Hae*III fragments in the G27(*hspR::km*) mutant strain.

Total RNA of the *hspR*⁻ mutant strain was then isolated and transcription from P_{gro}, P_{cbp} and P_{hrc} was investigated by primer extension or S1 nuclease mapping. Fig. 20A (lanes 2, 4 and 6) shows that the transcriptional start sites are unchanged in the *hspR*⁻ background while the amount of each specific transcript is increased severalfold with respect to the parental strain. Laser-scanning of the autoradiograms shown in Fig. 20A revealed that the RNAs associated with the P_{gro}, P_{cbp} and P_{hrc} promoters increased 7-, 9-, and 8-fold, respectively in the mutant strain as compared to the wild type strain. In contrast, primer extension analysis on the *ureAB* operon, which is also regulated by a σ^{80}-recognised promoter (see E.2.6) revealed no difference in the amount of RNA between wild type (lane 7) and mutant strains (lane 8). Therefore, the *H. pylori* HspR homologue acts as a specific repressor of the P_{gro}, P_{cbp} and P_{hrc} promoters, which regulate expression of the major chaperone genes in *H. pylori*.

3.3 Effect of HspR on cellular levels of GroEL and urease

To confirm the regulatory role of HspR in chaperone gene expression, Western Blot analyses on total protein extracts were carried out to compare the cellular levels of GroEL protein between wild type and isogenic *hspR*$^-$ mutant strains. Fig. 22 (lanes 1, 2) shows that antibodies specific for GroEL recognised the same protein band and that the amount of recognised protein is increased in the *hspR*$^-$ mutant. This data confirms the notion that HspR acts as a negative regulator of GroEL expression. As a control experiment, the same protein extracts were used in Western Blot analyses with antibodies raised against urease (Fig. 22, lanes 3, 4). Surprisingly, the amounts of both urease subunits (UreA and UreB) turned out to be decreased in the *hspR*$^-$ mutant with respect to the wild type strain. This decrease was not the result of transcriptional regulation (Fig. 20A, lanes 7, 8) and seemed not to be part of a possible HspR-mediated pleiotropic effect, as the levels of other *H. pylori* virulence factors, especially CagA and VacA, were unchanged in the *hspR*$^-$ mutant strain (data not shown).

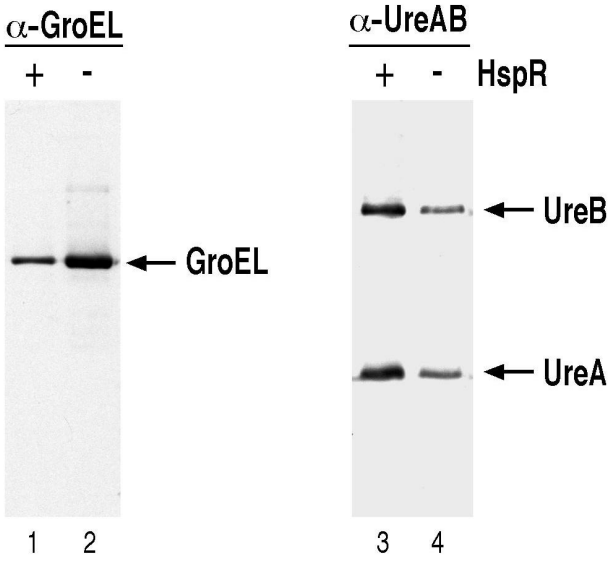


Fig. 22: Western Blots of cellular extracts from *H. pylori*. Total cell extracts were prepared from wild type *H. pylori* G27 (lanes +) and the isogenic *hspR*$^-$ mutant (lanes -) and probed with polyclonal antibodies against GroEL (lanes 1,2) or UreA/B (lanes 3,4).

3.4 Derepression of P_{gro} and P_{cbp} promoters by osmotic shock

In an attempt to identify environmental signals that lead to derepression of the HspR-dependent promoters, wild type and *hspR*$^-$ strains were grown in liquid cultures over night at 37 °C and were then incubated for 15 min either at 37 °C or at 45 °C or at 37 °C in the presence of 300 mM NaCl. Total RNA was then isolated and used in primer extension or S1 nuclease mapping experiments

with the same oligonucleotides and labeled probes that had been employed in the mapping of the transcriptional start sites. As a control, the heat shock regulated P_{sro} promoter (Beier *et al*., 1997) was also included in this analysis.

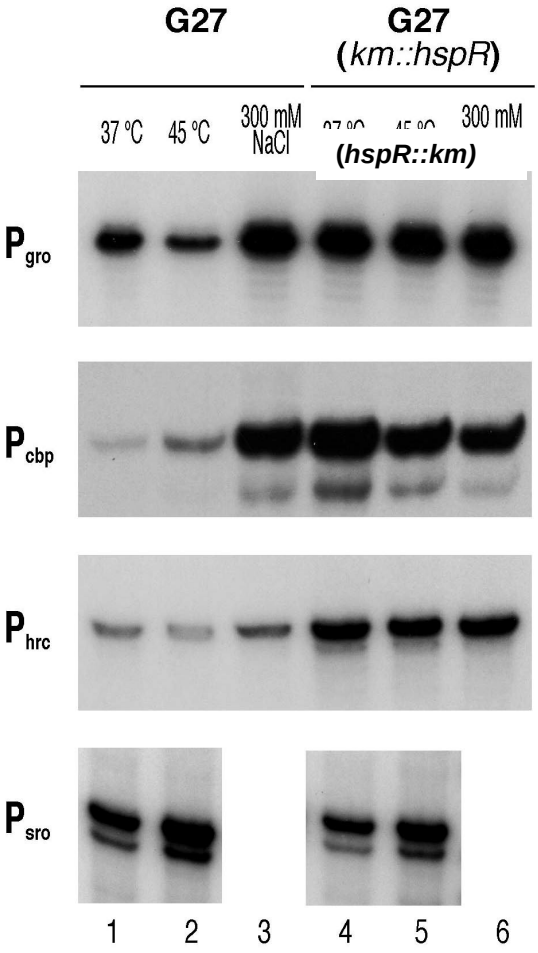


Fig. 23: Activities of *H. pylori* promoters under conditions of stress. Wild type *H. pylori* G27 and the isogenic *hspR*$^-$ mutant were grown at 37 °C and then incubated for 15 min either at 37 °C or at 45 °C or at 37 °C in the presence of 300 mM NaCl. Total RNA was isolated, hybridised to specific primers (groS for P_{gro}, hrcA for P_{hrc}, cheY for P_{sro}; Table 4) and elongated with reverse transcriptase or hybridised to specific DNA fragments (the 674 bp fragment from plasmid pCbp12 (Table 3) labeled at its *Eco*RI site for P_{cbp}) and digested with S1 nuclease. Elongated primers or protected DNA fragments are indicated with "P" and lowerscript letters that refer to the first downstream gene.

Fig. 23 (lanes 1, 2) shows that in the wild type strain a 15 min incubation at 45 °C lead to a slight increase in the amount of P_{cbp} associated RNA, while the amount of RNA transcribed from the P_{gro} and P_{hrc} promoters remained unaffected. As expected, following heat shock treatment the amount of RNA at the P_{sro} promoter was increased. In the *hspR*$^-$ mutant strain the three chaperone gene transcribing promoters P_{gro}, P_{cbp} and P_{hrc} showed the same amount of RNA at both 37 °C and 45 °C (Fig. 23, lanes 4, 5), while the RNA level at the P_{sro} promoter was higher at 45 °C than at 37 °C. This suggests that transcription of the chaperone genes is not significantly influenced by a 15 min heat shock treatment. In contrast, addition of 300 mM NaCl to cultures of *H. pylori* strain G27 yielded an increase in the amount of P_{gro} and P_{cbp} associated RNA as compared to untreated cells (Fig. 23, compare lane 3 to lane 1), thus indicating that osmotic shock induces transcription from these promoters. Analysis of the RNA extracted from the *hspR*$^-$ strain cultured at 37 °C or after heat shock at 45 °C or following salt treatment showed the same amount of RNA at the P_{gro} and P_{cbp} promoters as in the wild type cells treated with 300 mM NaCl (compare lanes 4-6 to lane 3).

Surprisingly, following salt treatment the amount of RNA at the P_{hrc} promoter showed no significant increase in the wild type strain (compare lane 3 to lane 1), indicating that this promoter is not induced by osmotic shock. However, independently of the growth conditions, the amount of RNA at this promoter was higher in the *hspR*$^-$ mutant as compared to the wild type strain (lanes 4-6), thus confirming the negative regulation exerted by HspR on transcription from this promoter.

These results suggest that HspR can mediate the transcriptional response to changes in osmolarity, whereas it seems not to be involved in regulation of heat shock inducible promoters of *H. pylori*.

3.5 Binding of HspR to the P_{gro} promoter

In a first attempt to determine if HspR binds directly to the promoters of the chaperone operons, a recombinant HspR protein was purified from *E. coli* and used in gel retardation experiments. The coding sequence of the *hspR* gene was therefore amplified by PCR from chromosomal DNA of *H. pylori* strain G27 using oligonucleotides hspRN/hspRC (Table 4) and cloned into the expression vector pTE22b$^+$ (Table 3). The resulting plasmid, pTE22b-hspR (Table 3), which contains the entire *hspR* coding sequence fused to six histidine codons at its 3' end, was then transformed into *E. coli* strain BL21 and the recombinant HspR-[His_6]-protein was purified in its native form as described in D.17. According to coomassie blue staining of SDS-polyacrylamide gel, purity of the protein was estimated to be about 85-90% (Fig 24A). Impurity was mainly due to a smear of faint bands and a distinct band migrating to a position of 70 kDa. N-terminal microsequencing showed that this contaminant band corresponded to a putative transformylase from *E. coli* (accession No AAC75315.1). The recombinant HspR protein was then assayed in band shift experiments with a 427 bp probe obtained from plasmid pGroE (Table 3) which comprises the entire *groESL* promoter region. As shown in Fig. 24B, incubation of the probe with an amount of approximately 50 ng of purified HspR yielded a specific complex with dramatically reduced electrophoretic mobility (lane 3). In the presence of 250 ng of HspR the labeled fragment was completely retarded (lane 5), suggesting a specific binding of the protein to the *groESL* promoter region. In addition to the specific complex, fuzzy bands migrating at lower positions in the gel were detected (indicated with a bar in Fig. 24B). These bands may indicate the formation of intermediate unstable protein-DNA complexes.

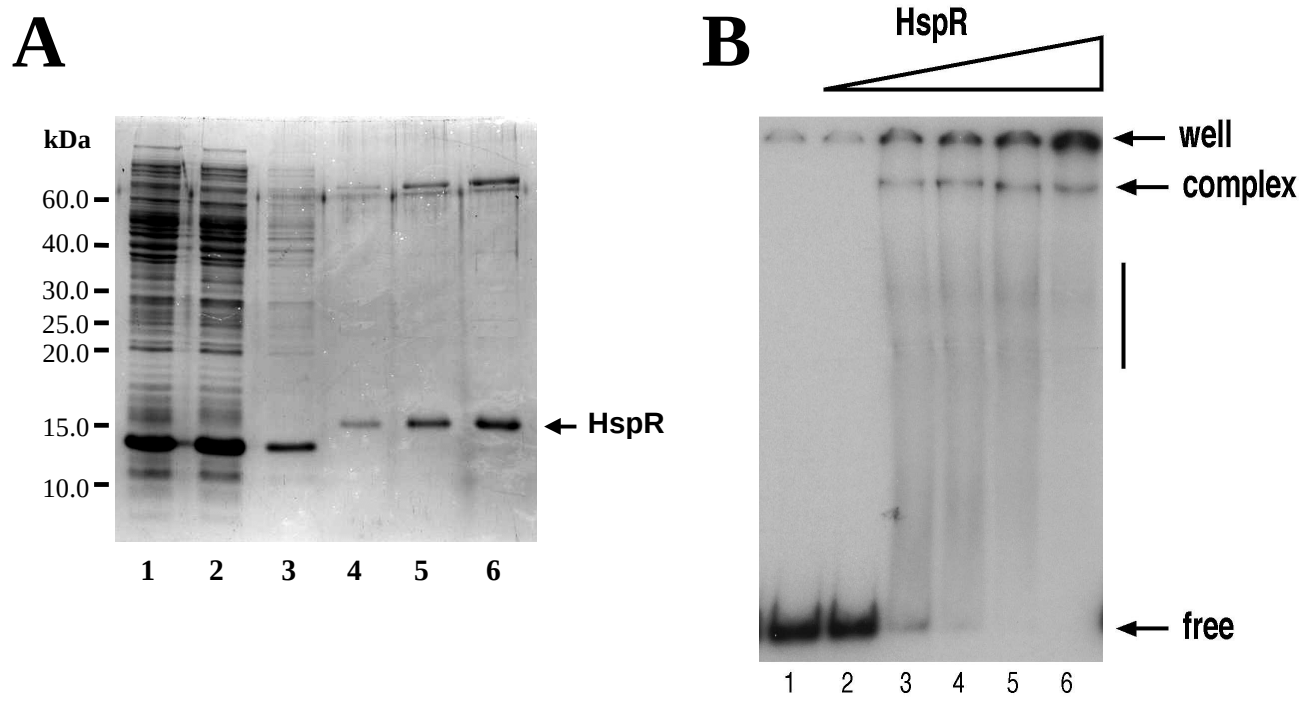


Fig. 24: A) Purification of HspR. The [His_6]-tagged HspR protein was purified from induced cultures of BL21 pTE22b-hspR (Tables 2,3). lane 1: 50 µl cleared lysate of induced culture, lane 2: 50 µl column flow through, lane 3: 100 µl wash; lanes 4-6: 1, 5, and 10 µl eluted protein. Fractions were loaded on a 15% SDS-polyacrylamide gel. **B) Gel mobility shift assay with purified HspR on the *groESL* promoter.** A 427 bp *Eco*RI-*Bam*HI-fragment obtained from plasmid pGroE (Table 3) and comprising positions +148 to –279 of the P_{gro} promoter was incubated with increasing amounts of purified HspR protein. Lane 1: no protein added; lanes 2-6: 10, 50, 100, 250, 500 ng purified HspR added. All lanes contained 1000 ng sonicated salmon sperm DNA as non-specific competitor. Unbound (free) and bound (complex) DNAs are indicated on the right side of the figure. Unstable lower molecular weight complexes are indicated with a bar.

3.6 Differential binding of HspR to the P_{gro}, P_{cbp} and P_{hrc} promoters

In order to localise in more detail the binding site(s) of HspR DNase I footprinting assays were then performed on the different promoters. Fig. 25A shows the result of a footprinting experiment carried out on a DNA fragment specific for the P_{gro} promoter. With the addition of 2 µg of purified HspR, a large protected region appeared that became more distinct after addition of 4 or 8 µg of purified protein. The region extended from -46 to -118 with respect to the transcriptional start site of the P_{gro} promoter and showed six bands of enhanced DNase I sensitivity at both adjacent and internal sites. The same experiment carried out on the P_{cbp} promoter (Fig. 25B) showed a protected area of similar size between positions +14 and -59, and revealed a similar pattern of enhanced DNase I sensitivity. As in the case of the P_{gro} promoter, the protected area could be detected after addition of 2 µg protein and became more pronounced as protein concentration increased. Finally, a footprint on a P_{hrc} specific probe (Fig. 25C) revealed a protected region between -78 and -156 with respect to the transcriptional start site. Similarly to the other two promoters, six DNase I hypersensitive sites mapping to different positions internal and adjacent to the protected region could be detected. Notably, unlike the P_{gro} and P_{cbp} specific fragments,

protection of the P_{hrc} specific probe could be detected with as little as 500 ng of purified HspR protein (Fig. 25C, lane 3).

It is concluded that HspR binds specifically to the three chaperone promoters, covering a large DNA region of about 70-80 bp that overlaps the transcriptional start site in the case of the P_{cbp} promoter, but maps upstream of the -35 box in the P_{gro} and P_{hrc} promoters. The P_{gro} and P_{cbp} promoters seem to have comparable affinity for HspR while the P_{hrc} promoter seems to have a slightly enhanced affinity for the protein with respect to these two promoters.

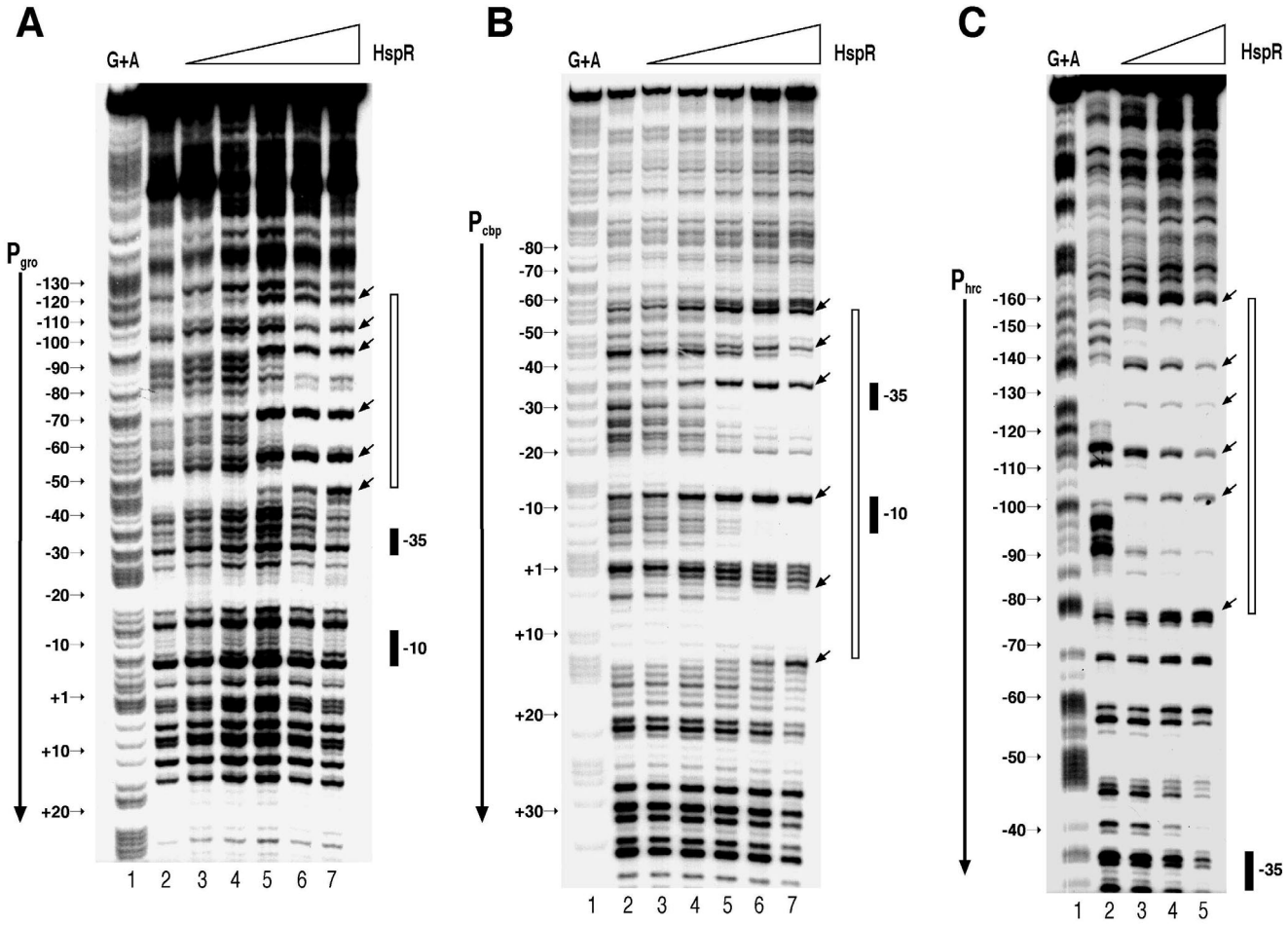


Fig. 25: DNase I footprinting of HspR on the P_{gro}, P_{cbp}, and P_{hrc} promoters. Specific end-labeled fragments were incubated with increasing amounts of purified HspR protein and digested with 0.02 U DNase I. Lanes 2-7: 0, 0.5, 1.0, 2.0, 4.0, and 8.0 µg HspR added, respectively. One thousand ng sonicated salmon sperm DNA was added in all lanes as non-specific competitor. The vertical arrows on the left side of each panel indicate the direction of transcription; numbers refer to nucleotide positions with respect to the transcriptional start sites. The vertical open bars on the right side of each panel indicate the areas of DNase I protection, small arrows refer to DNase I hypersensitive sites. The positions of the -35 and/or -10 regions of each promoter are marked by a vertical black bar. The G+A lane was a sequence reaction on the DNA probe used as a size marker. DNA probes used in the footprinting assays were: **A)** a 427 bp *Eco*RI-*Bam*HI fragment obtained from plasmid pGroE (Table 3), labeled at its *Bam*HI site. **B)** a 272 bp *Eco*RI-*Bam*HI fragment obtained from plasmid pCbp34 (Table 3), labeled at its *Eco*RI site. **C)** a 293 bp *Eco*RI-*Bam*HI fragment obtained from plasmid pHrcA (Table 3), labeled at its *Bam*HI site.

Fig. 26 shows the sequences of the DNA regions that are covered by HspR and the positions of the DNase I hypersensitive sites on the three promoters. Interestingly, the P_{cbp} and P_{gro} promoters show an identical pattern of DNase I hypersensitive sites. In fact, in the middle of both binding sites, a sequence with similarity to the HAIR consensus sequence (CTTGAGT-N_7-ACTCAAG) proposed by Grandvalet *et al.* (1999) can be detected. In contrast, the P_{hrc} promoter shows a pattern of DNase I hypersensitivity that is different from the one observed on the other two promoters. In addition, unlike the P_{gro} and P_{cbp} promoters, a sequence with similarity to the HAIR consensus is barely detectable within the protected DNA region. The mechanism of HspR binding to the P_{hrc} promoter might therefore be different from that of the other two promoters.

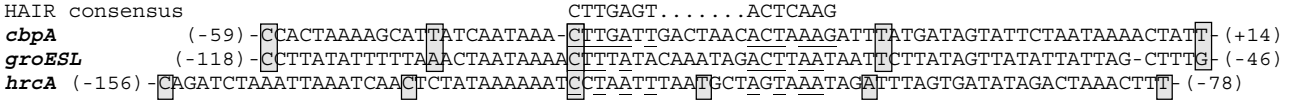


Fig. 26: **Sequences of the HspR binding sites of the P_{gro}, P_{cbp} and P_{hrc} promoters**. DNase I hypersensitive sites are boxed, nucleotide positions are indicated with respect to the transcriptional start sites. Nucleotides matching the HAIR consensus sequence proposed by Grandvalet *et al.* (1999) are underlined.