Modeling of Rho Dependent Transcription Termination Sites in the Bacterium *Helicobacter pylori*

Lise Petersen and Anders Krogh

Bioinformatics Centre, 15 Universitetsparken, DK-2100 Copenhagen, Denmark.
lpe@binf.ku.dk, krogh@binf.ku.dk

Keywords. *Helicobacter pylori*, hidden Markov model, transcription termination, Rho, RNA secondary structure

Helicobacter pylori is a gram negative bacterium that has been linked to numerous severe gastroduodenal diseases, including peptic ulcer and gastric cancer. The basis for these diverse clinical outcomes is not understood, but part of the explanation may be differential virulence among strains. A considerable number of strain specific genes have been revealed by comparative genomics studies of two fully sequenced genomes [1] and DNA-DNA microarray studies of clinical isolates [2]. Many of the strain specific genes are located in a plasticity zone. The expression level of those genes or whether they are expressed at all, is presently unknown, as is the function of many of them. Mapping of gene regulatory motifs and transcription units on the genomes may aid in elucidating the importance of different genes.

Rho-dependent termination in bacteria is poorly understood, and the Rho binding site is not well conserved. We investigated the occurrence of RNA secondary structure and conserved sequence motifs around transcription termination sites in the genomes of two strains of the bacterium. Earlier studies have indicated that *H. pylori* only to a very limited degree rely on mRNA hairpin structure formation for transcription termination. The gene for transcription termination factor Rho is present in both strains, and sensitivity to bicyclomycin indicates that Rho-dependent termination is essential for the bacterium.

We used a hidden Markov model (HMM) to model conserved motifs in the downstream regions of genes. We constructed a training set consisting of downstream regions of genes that are followed by a gene transcribed in the opposite direction, and therefore must have a transcription termination site. Sequences were removed from the training set if they contained inverted repeats that were followed by a stretch of T-residues, and based on length of stem, distance to stop codon, and a free energy criterion were comparable to Rho independent transcription termination sites described in *E. coli* [3]. Such sequences can form stable hairpin structures in the mRNA, thereby causing release of the RNA polymerase holoenzyme. The frequent occurrence of poly-G tracts in transcription termination regions led us to the discovery that also shorter, GC rich inverted repeats could be found that satisfied the same free energy criterion and such sequences were subsequently removed from the training set. Altogether we reduced the training set to a total of 118 sequences.

A motif resembling known characteristics of Rho-dependent termination [4] was identified using the HMM (Fig 1). The HMM was used to decode the two *H. pylori* genomes, as well as random sequence, using a non-looped model architecture and posterior decoding, as described previously [5].

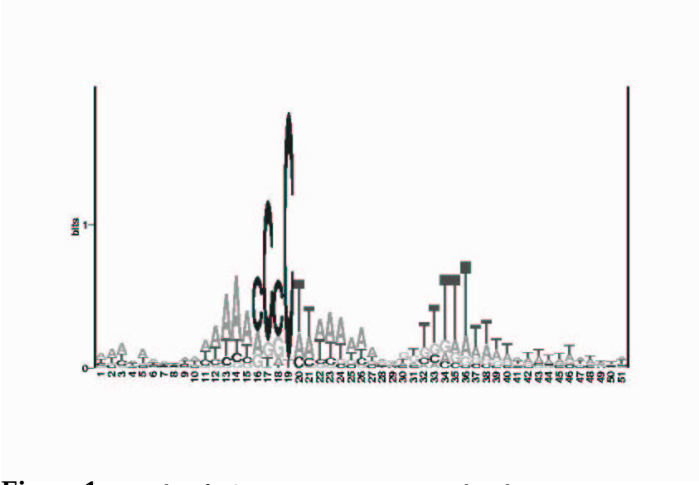

Figure 1 Logo plot of 185 putative terminators predicted in downstream regions of genes annotated in *H. pylori* J99

The H. pylori genomes were decoded with the model and the predictions showed a characteristic distribution around annotated stop and start codons (Fig. 2).

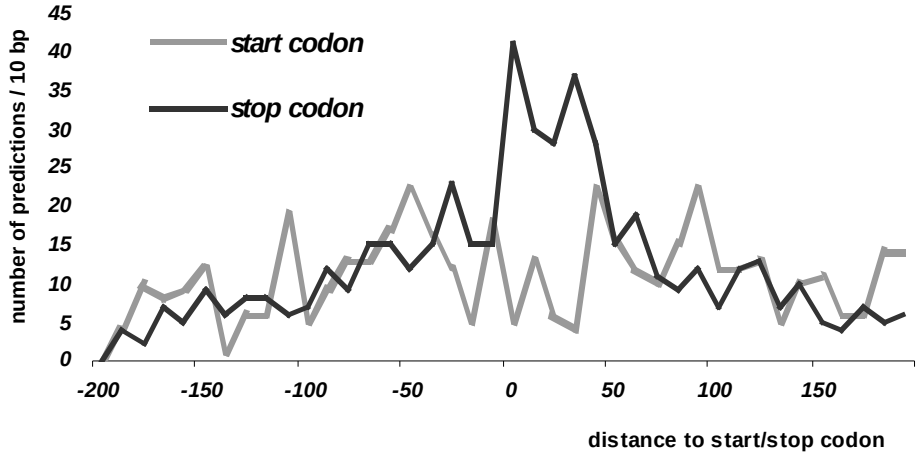

Figure 2 Distribution of predicted termination sites in relation to annotated stop and start codons in two *H. pylori* genomes

Conclusions

The HMM developed in this study identified a consensus motif consisting of a C-tract, followed by a spacer and a T-tract. Such a motif may be able to terminate transcription, in that a stretch of C-residues is known to bind Rho, and the T-tract in turn will function as a pause site for the RNA polymerase holoenzyme [4].
In both genomes predictions are relatively frequent around stop codons, and relatively rare around start codons. This observation indicates that the identified motif is indeed involved in transcription termination.
In contrast to earlier studies we found that stable RNA secondary structures (hairpins) are frequently found in transcription termination regions, suggesting that Rho-independent termination is more important in *H. pylori* than previously believed [6].

References

[] Janssen, P.J., B. Audit and C.A. Ouzounis, Strain-specific genes of *Helicobacter pylori*: distribution, function and dynamics. *Nucleic Acids Research*, 29:4395-4404, 2001.
[2] Salama, N., K. Guillemin, T.K. McDaniel, G. Sherlock, L. Tompkins and S. Falkow, A whole-genome microarray reveals genetic diversity among *Helicobacter pylori* strains. *Proceedings of the National Academy of Sciences*, 97:14668-14673, 2000.
[3] Lesnik, E.A., R. Sampath, H.B. Levene, T.J. Henderson, J.A. McNeil, and D.J. Ecker, Prediction of rho-independent transcriptional terminators in *Escherichia coli*. *Nucleic Acids Research*, 29:3583-3595, 2001.
[4]Richardson, J.P., Rho-dependent termination and ATPases in transcript termination. *Biochimica and Biophysica Acta*, 1577:251-260, 2002.
[5] Petersen, L., T.S. Larsen, D.W. Ussery, S.L.W. On, and A. Krogh, *Rpo*D promoters in *Campylobacter jejuni* exhibit a strong periodic signal instead of a –35 box, *Journal of Molecular Biology*, 326:1361-1372, 2003.
[6] Unniraman, S., R. Prakash and V. Nagaraja, Conserved economica of transcription termination in eubacteria. *Nucleic Acids Research*, 30:675-684, 2002.