In Silico Prediction and Functional Validation of σ^{28}-Regulated Genes in Chlamydia and Escherichia coli

Hilda Hiu Yin Yu,1,2 Dennis Kibler,1,4 and Ming Tan1,2,3,*

Institute for Genomics and Bioinformatics,1 Department of Microbiology and Molecular Genetics,2 Department of Medicine,3 School of Medicine, and Department of Computer Science, Donald Bren School of Information and Computer Science,4 University of California, Irvine, California 92697-4025

Received 21 July 2006/Accepted 6 September 2006

σ^{28} RNA polymerase is an alternative RNA polymerase that has been proposed to have a role in late developmental gene regulation in Chlamydia, but only a single target gene has been identified. To discover additional σ^{28}-dependent genes in the Chlamydia trachomatis genome, we applied bioinformatic methods using a probability weight matrix based on known σ^{28} promoters in other bacteria and a second matrix based on a functional analysis of the σ^{28} promoter. We tested 16 candidate σ^{28} promoters predicted with these algorithms and found that 5 were active in a chlamydial σ^{28} in vitro transcription assay. hctB, the known σ^{28}-regulated gene, is only expressed late in the chlamydial developmental cycle only, and two of the newly identified σ^{28} target genes (tsp and dycC_1) also have late expression profiles, providing support for σ^{28} as a regulator of late gene expression. One of the other novel σ^{28}-regulated genes is dnaK, a known heat shock-responsive gene, suggesting that σ^{28} RNA polymerase may be involved in the response to cellular stress. Our σ^{28} prediction algorithm can be applied to other bacteria, and by performing a similar analysis on the Escherichia coli genome, we have predicted and functionally identified five previously unknown σ^{28}-regulated genes in E. coli.

MATERIALS AND METHODS

Development of a program for extracting sequences. We developed a program called SequenceExtractor to extract user-defined DNA sequences from a genome. The program requires two input files, consisting of a genome sequence file and a file containing the start and stop coordinates for each sequence within the genome sequence being examined. We applied this program to extract two files in fasta format from each of the genomes of C. trachomatis serovar D, E. coli K-12, and Salmonella enterica serovar Typhimurium using sequences obtained from TIGR (http://www.tigr.org). For each organism, the first output file contained up to 200 bp of sequence upstream for each gene (“200 bp upstream”). The second output file was more restrictive and contained up to 200 bp of upstream sequence for each gene, provided that these sequences were in the intergenic region and not within the coding region of the nearest upstream gene (“200-bp nonoverlap”).

Development of a program for predicting promoters. We also developed a program, called PromoterMatcher, that uses a probability weight matrix to predict promoters in a genome. We generated two probability weight matrices, each based on complementary information about σ^{28} promoter structure, and applied them to an input file consisting of extracted sequences in fasta format. The frequency matrix was based on a set of 21 known bacterial σ^{28} promoters and takes into account the frequency of occurrence of each nucleotide at each position within this promoter set. The activity matrix used functional data in the form of the relative promoter strength attributable to each nucleotide at each promoter position and was derived from a comprehensive mutational analysis of a σ^{28} promoter (25). For each promoter position, the algorithm assigned a probability value to the four possible nucleotides, with a total probability of 1. Both matrices also contained probability-weighted models for the length of the spacer between the two promoter elements based on either nucleotide frequency or relative promoter activity. The final score for each candidate promoter was determined by summing the log of the probability at each position (which is the mathematical equivalent of multiplying the probabilities). Only the highest-scoring promoter was recorded per upstream region. The predicted promoters were sorted by score, from best to worst.

Generation of the sequence logo. All sequence logos were derived using SEQLOGO, which is available online at http://ep.cbi.ac.uk/EP/SEQLOGO/. The format for data input into this site is a series of numbers representing either

Genome sequencing has indicated that all Chlamydia species encode two alternative sigma factors, suggesting a role for alternative forms of RNA polymerase in chlamydial gene regulation. We have demonstrated that one of these alternative RNA polymerases, σ^{28} RNA polymerase, transcribes hctB (24), a gene whose transcript is detectable only at late time points in the chlamydial developmental cycle (6, 16). hctB encodes Hc2, one of two histone-like proteins in Chlamydia that have been shown to be responsible for the condensation of DNA during conversion of the metabolically active form of chlamydiae, known as a reticulate body, to the infectious extracellular form, the elementary body (8). To date, hctB is the only σ^{28}-regulated gene that has been identified in Chlamydia, and it is not known whether the role of σ^{28} RNA polymerase is confined to the regulation of late gene expression in the developmental cycle.

To identify additional σ^{28}-regulated genes in Chlamydia, we have combined the use of bioinformatics, to predict σ^{28}-regulated promoters in the chlamydial genome, with testing of promoter activity in a chlamydial σ^{28} in vitro transcription assay. We used two in silico approaches, identifying candidate promoters on the basis of sequences that either resemble the consensus bacterial σ^{28} promoter (9, 10) or are predicted to be highly transcribed by σ^{28} RNA polymerase based on functional studies (25). Using information from both approaches, we have developed a computer algorithm to identify candidate σ^{28} promoters in the chlamydial genome and have shown that five promoters are transcribed by chlamydial σ^{28} RNA polymerase. This method can be applied to other bacterial genomes, and we have also identified five new σ^{28}-regulated genes in Escherichia coli.
nucleotide frequency or relative promoter activity. The resulting sequence logo consists of stacks of letters at each position. The height of the stack indicates the importance of a particular position for promoter activity. The height of an individual letter within a stack indicates the relative preference for that nucleotide based on transcriptional activity or frequency (with a maximum height defined as 2 bits).

Cloning of transcription plasmids. Each candidate \( \sigma^{28} \)-regulated promoter to be tested was cloned into a plasmid so that promoter activity could be measured with a \( \sigma^{28} \) in vitro transcription assay. The promoter insert, consisting of sequence from approximately 300 bp upstream of the transcription start site to the +5 position, was amplified by PCR using either \( C. trachomatis \) serovar D or \( E. coli \) K-12 genomic DNA and respective primers (see Table S1 in supplemental material). This PCR insert was cloned upstream of a synthetic G-less cassette template in plasmid pMT1125 (23). Transcription from the predicted promoter by \( \sigma^{28} \) RNA polymerase was expected to produce a 130-nt transcript.

Overexpression and purification of \( \sigma^{28} \). \( C. trachomatis \) serovar L2 His\(_6\)-\( \sigma^{28} \) and \( E. coli \) His\(_6\)-\( \sigma^{28} \) were individually overexpressed in \( E. coli \) BL21(DE3) and purified, as previously described (24, 25), to a concentration of 35.7 \( \mu \)g/ml and 115.8 \( \mu \)g/ml, respectively.

In vitro transcription reactions. Transcription reactions were performed as previously described (24, 25). \( C. trachomatis \) \( \sigma^{28} \) RNA polymerase was reconstituted by mixing 1 \( \mu \)l \( C. trachomatis \) recombinant His\(_6\)-\( \sigma^{28} \) with 1 \( \mu \)l heparinagarose-purified \( C. trachomatis \) RNA polymerase at 4°C for 15 min, immediately prior to the transcription reaction. \( E. coli \) \( \sigma^{28} \) RNA polymerase was reconstituted from 1 \( \mu \)l \( E. coli \) recombinant His\(_6\)-\( \sigma^{28} \) and 0.03 units \( E. coli \) core enzyme (Epigenetics, Madison, Wis.). For antibody inhibition reactions, 8 \( \mu \)l of rabbit polyclonal antichlamydial \( \sigma^{28} \) antibodies (24) were preincubated with the RNA polymerase at 4°C for 15 min, immediately prior to the transcription reaction.

Purification of \( C. trachomatis \) RNA polymerase from chlamydial grown in tissue culture. \( C. trachomatis \) serovar L2 was grown in mouse L929 cells and harvested at 18 h postinfection (hpi). RNA polymerase was partially purified by heparin-agarose chromatography as previously described (21).

Purification of reticulate body RNA. L929 cells grown in suspension and infected with \( C. trachomatis \) serovar D were recovered by centrifugation and lysed by Dounce homogenization as previously described, with slight modifications (21). A second centrifugation step separated chlamydiae from host cellular debris. Chlamydial RNA was extracted using RNA STAT-60 (Teltest, Inc., Friendswood, TX).

Mapping transcription start sites by primer extension. The primer was prepared from 100 ng of a DNA oligonucleotide that was labeled with 50 \( \mu \)Ci of \( [\gamma^{32}P] \)ATP in the presence of T4 polynucleotide kinase at 37°C for 1 h. Unincorporated free nucleotides were removed with a DNA mini-Quick Spin DNA column (Roche Diagnostics, Indianapolis, Ind.). Radioactive samples were counted with a scintillation counter. Fifty micrograms of reticulate body RNA and 5 \( \times \) 10\(^5\) cpm labeled primer were preincubated at 65°C for 10 min and chilled on ice. cDNA was synthesized by adding Superscript II reverse transcriptase (Invitrogen, Carlsbad, Calif.) and 10 mM deoxynucleoside triphosphates, followed by incubation at 42°C for 50 min. The reaction was stopped by the addition of distilled water and a 1/10 volume of 5 M sodium acetate to a total volume of 100 \( \mu \)l, followed by phenol-chloroform extraction and chloroform extraction. cDNA was recovered by ethanol precipitation, dried, and resuspended in 9 \( \mu \)l formamide stop solution (95% formamide, 20 mM EDTA, 0.05% bromophenol blue, 0.05% xylene cyanol). The primer extension products were electrophoresed on a 6% acrylamide-urea sequencing gel together with a single-stranded M13mp18 DNA sequence ladder and exposed to X-ray film.

FIG. 1. Flow chart showing the use of computer algorithms for promoter prediction. The SequenceExtractor algorithm was used to extract two sequence files for each open reading frame (ORF) that were then analyzed with a promoter prediction algorithm (Promoter-Matcher) using either of two probability weight matrices (see the text). In total, this scheme produced four lists of predicted promoters ranked by scores. Details are provided in Materials and Methods and Results.

FIG. 2. Probability weight matrices and sequence logos for predicting promoters in the chlamydial genome. (A) Frequency matrix for the 16 positions in the –35 and –10 promoter elements and the four possible nucleotides at each position (see Materials and Methods). Each value in the matrix indicates how many of the 21 known \( \sigma^{28} \) promoters in the training set contained the given nucleotide at that promoter position. A sequence logo depicting the relative nucleotide preference at each position in the promoter is shown below the matrix (25). (B) Activity matrix for the –35 and –10 promoter elements with values derived from a mutational analysis of the \( C. trachomatis \) hctB promoter as described in the text (25). At each promoter position, the values are proportional to the relative promoter activity attributable to that nucleotide for a total of 100%. The sequence logo for the promoter is shown below the matrix. Details of the sequence logo format are presented in Materials and Methods and Results. All sequence logos were derived using SEQLOGO, which is available online at http://ep.ebi.ac.uk/EP/SEQLOGO/.
TABLE 1. Predicted chlamydial $\sigma^{28}$ promoters tested for in vitro activity

| Gene ID | Gene name       | Function                        | Location from ORF (nt) | $-35$ sequence | Spacer length (nt) | $-10$ sequence | Rankings$^a$ | $\sigma^{28}$ activity |
|---------|-----------------|---------------------------------|------------------------|----------------|-------------------|----------------|-------------|------------------------|
| CT047   | Possible outer membrane protein I | -96 TTTTGATAT 11 GTCGAAAT | 36, 51, 24 No          |
| CT099   | NADPH thioredoxin reductase | -104 TTAGTTTT 11 GTCGAAAC | 32, 6, 3 No          |
| CT018   | Hypothetical protein | -193 TAAGT12 CAGGAAAT* | 17, 8 No          |
| CT249   | Hypothetical protein | -188 TCAAC12 ATCGAAAT* | 9, 42 No          |
| CT256   | Hypothetical protein, possible hemolysin | -104 TACAGTGT 11 GCCGAAAGA | 4, 3, 45, 20 Yes |
| CT359   | Probable biotin synthase | -48 TAAAGGCC 12 GTCGATTTC* | 8, 4 Yes          |
| CT396   | Dnak protein | -63 TAAAGGA11 AACGGAAGA | 28, 27, * Yes      |
| CT415   | ABC transporter | -89 TAAAGAA11 GAAGAAAA | 14, 9, 14, 6 Weak |
| CT418   | GTP-binding protein | -196 TAAACGT12 GACGATAC | 10, 4, * Weak |
| CT425   | Hypothetical protein | -42 TAAAGGAC10 CTCGAACT* | 54, 28 Weak |
| CT441   | Probable tail-specific protease | -61 TAAAGGAC 11 GTCGAAGA | 3, 2, 2, 2 Yes   |
| CT553   | FMU/SUN-related methyltransferase | -37 TAAAGGAC13 CTCGAAAG | *3, 23, 12 No |
| CT631   | Hypothetical protein | -41 TCTCTTTT12 ATCGAAAT* | 41, 18 No          |
| CT693   | Phosphoglycerate kinase | -56 TGAAGTTT11 GCCTATAA | 13, 8, * Yes      |
| CT789   | Hypothetical protein | -45 TTAATATT12 AACGAAAG | 21, 12, 21, 10 No |
| CT798   | Glycogen synthase | -41 TAAAGGAC12 AACGAAAG | *28, 13 No         |

$^a$ The rankings for each of four promoter prediction lists are shown in the following order: 200-bp upstream frequency matrix, 200-bp nonoverlap frequency matrix, 200-bp upstream activity matrix, and 200-bp nonoverlap activity matrix. *排名 is not in the top 50.

RESULTS

Development of computer algorithms to identify $\sigma^{28}$ promoters. We developed two computer algorithms, which we used in parallel to identify candidate $\sigma^{28}$ promoters within a genome (Fig. 1). The first program, called SequenceExtractor, selects sequences from a genome for analysis by the second program, PromoterMatcher, which makes predictions on the genome (Fig. 1). The first program, called SequenceExtractor, was used to select sequences up to each predicted gene where promoters were most likely to be present. SequenceExtractor was used to select sequences up to 200 bp upstream of each gene, provided that they were not in a coding region (200-bp nonoverlap region). In Chlamydia, however, many intergenic regions are short, and promoters have been located within the upstream gene (13). Thus, we also separately examined all sequences in the region 200 bp upstream of each gene even if they were beyond the intergenic region (200-bp upstream region).

To identify candidate $\sigma^{28}$ promoters within these upstream sequences, we used the PromoterMatcher algorithm to apply a weighted matrix and assign probability scores for the 16 promoter positions and the spacer length. To increase the likelihood of identifying $\sigma^{28}$ promoters, we used two weighted matrices, each based on a different measure of the contribution of sequence to promoter activity. The first matrix, called the frequency matrix, was based on the occurrence of a given nucleotide at each promoter position within a compilation of 21 known $\sigma^{28}$ promoters, including 20 promoters from E. coli and Salmonella, and the C. trachomatis hctB promoter. For example, at the seventh position in the −10 element (Fig. 2A), an A was present in 19 of the 21 promoters, and the remaining two promoters had a G at this position. Accordingly, an A was assigned a strong weighting of 19/21, while the weighting for a G was 2/21. As all known $\sigma^{28}$ promoters, with the exception of the chlamydial promoter, have a spacing of 11 nt, this spacer length was also heavily weighted.

A second weighted matrix, called the activity matrix, assigned a weighting to each of the four possible nucleotides for every position based on the promoter activity attributed to that nucleotide in a mutational analysis of the hctB promoter (25). For example, at the seventh position of the −10 element, hctB promoter activity with C. trachomatis $\sigma^{28}$ RNA polymerase was greatest when an A was present and was reduced by 2.6-, 3.4-, and 19.3-fold with a G, T, or C, respectively (25). We thus assigned probability scores for A (58/100), G (22/100), T (17/100), or C (3/100) that were proportional to these promoter activities (Fig. 2B). The probability weighting for the spacer length was based on the measured effect of a spacer length from 9 to 13 nt on hctB promoter activity (25).

By applying these two weighted matrices to the two sets of upstream sequences, we generated four lists of candidate $\sigma^{28}$
promoters using PromoterMatcher. hctB, the known C. trachomatis $\sigma^{28}$-regulated promoter, was the highest-scoring promoter sequence in all lists. The top 30 predictions for each list are shown in Table S2 (frequency matrix), and Table S3 (activity matrix) in the supplemental material.

Five candidate chlamydial promoters were transcribed by $\sigma^{28}$ RNA polymerase. We chose 16 of the top candidate promoters (Table 1) for functional testing with our chlamydial $\sigma^{28}$ in vitro transcription assay. In general, these promoters were among the top-50-scoring promoters in at least two of the four prediction lists. Since the source of our core enzyme contains chlamydial $\sigma^{28}$ RNA polymerase activity (24), we tested for transcription in the absence and presence of recombinant chlamydial $\sigma^{28}$ as a measure of $\sigma^{28}$-specific and $\sigma^{28}$-specific activity, respectively. We also assayed for $\sigma^{28}$-dependent activity by testing for inhibition of transcription by anti-$\sigma^{28}$-antibodies.

Five of the 16 candidate promoters tested showed $\sigma^{28}$-specific activity (Fig. 3), and an additional three promoters (yebL, yhbZ, and CT425) were weakly transcribed (data not shown). Four of the strongly transcribed promoters (tsp, dnaK, thyC_1, and bioY) produced a transcript only when recombinant chlamydial $\sigma^{28}$ was added, as was the case with the hctB positive control promoter. Transcription of these four promoters was also abrogated by rabbit polyclonal anti-$\sigma^{28}$-antibodies (Fig. 3, lane 3). The results were less clear-cut with the pgk promoter, because although there was a large increase in transcription when $\sigma^{28}$ was added, there was baseline transcription in the absence of $\sigma^{28}$, raising the possibility of some $\sigma^{66}$-dependent activity. Anti-$\sigma^{28}$ antibodies decreased transcription of the pgk promoter, but there was still residual transcript present. These results provide evidence that the promoters for tsp, dnaK, thyC_1, and bioY are transcribed by $\sigma^{28}$ RNA polymerase and suggest that the pgk promoter is recognized by both $\sigma^{28}$ and $\sigma^{66}$ RNA polymerases.

For in vivo validation of these results, we used primer extension to map the transcription start sites for the three strongest promoters, hctB, tsp, and pgk, to within 6 nt of the predicted $\sigma^{28}$ –10 promoter element, at a location consistent with the predicted promoter (Fig. 4). A previously mapped transcription start site for dnaK (5) was located within 8 nt of the $\sigma^{28}$ promoter that we have predicted for this gene.

**pgk is regulated by two overlapping promoters.** Analysis of the sequence in the pgk promoter revealed a possible $\sigma^{66}$ promoter overlapping the predicted $\sigma^{28}$ promoter. To confirm the presence of an active $\sigma^{28}$ promoter, without the confounding effect of a second promoter, we introduced substitutions predicted to disrupt the putative $\sigma^{66}$ promoter but not the $\sigma^{28}$ promoter (Fig. 5A). With this mutant promoter, there was no baseline $\sigma^{66}$-dependent activity, and all transcription was dependent on the addition of chlamydial $\sigma^{28}$ (Fig. 5B, lanes 1 and 2). Transcription was specifically inhibited by anti-$\sigma^{28}$ antibodies (Fig. 5B, lane 3). These results provide good experimental support for the predicted $\sigma^{28}$-dependent pgk promoter and an overlapping $\sigma^{66}$ promoter.

**Five predicted E. coli $\sigma^{28}$ promoters were transcriptionally active.** As we also had functional data for promoter recognition...
by *E. coli* σ^28^ RNA polymerase (25), we were able to apply our promoter-finding algorithm to the genomes of *E. coli* and the closely related bacterium *Salmonella*. Lists of the top 30 candidate promoters are shown in Tables S4 and S5 for *E. coli* and Tables S6 and S7 for *Salmonella* (see the supplemental material). Many of these promoters are known σ^28^ promoters in *E. coli* and *Salmonella*. We tested seven candidate σ^28^ *E. coli* promoters that had not been previously studied (Table 2) and found that five (*modA*, *ynjH*, *yecF*, *yhiL*, and *yjcS*) were functionally active in an *E. coli* in vitro σ^28^ transcription assay (Fig. 6).

**DISCUSSION**

This study demonstrates how a combination of a bioinformatic analysis and functional validation can be used to identify previously unrecognized target genes of an alternative RNA polymerase. From a genome-wide search for sequences resembling known σ^28^ promoters and sequences that have been shown to be highly transcribed by σ^28^ RNA polymerase, we identified five novel σ^28^-regulated genes in *Chlamydia* and another five new σ^28^-regulated genes in *E. coli*. Although we did not test any of the predicted *Salmonella* σ^28^ promoters, our list of top-scoring promoters includes three (STM3152, STM3216, and STM2314) of four newly identified σ^28^-regulated target genes in *S. enterica* serovar Typhimurium (7). These results demonstrate that our promoter prediction algorithm can successfully identify σ^28^-promoters, and it is likely that additional σ^28^-regulated genes remain to be discovered in many bacterial genomes.

Our results show that the combination of a frequency matrix, derived from known σ^28^-promoter sequences, and an activity matrix, based on a mutational analysis of a chlamydial σ^28^-dependent promoter, increased our chances of identifying additional σ^28^-promoters. *hctB* and *tsp* had the two strongest promoters in terms of transcriptional activity and sequence conservation with the bacterial σ^28^-consensus promoter, and both ranked equally high with the two matrices (see Tables S2 and S3 in the supplemental material). For promoters with weaker sequence conservation, such as *bioY*, the activity matrix may be a better predictor. For instance, *bioY* ranked in the top 10 using the activity matrix (Table 1) but was not in the top 50 with the frequency matrix.

In general, we found that a strict pattern-matching algorithm based only on the bacterial σ^28^-consensus sequence would not be very sensitive as a means of identifying σ^28^-dependent promoters in *Chlamydia*. With the exception of the *hctB* promoter, the other chlamydial σ^28^-promoters identified in this study are not well conserved with the bacterial consensus promoter. For example, while the *dnaK* promoter (TAAAGGAA-N11-AACGAAGA) contains the signature TAAA of the σ^28^-consensus promoter, and *b0763* and *b3490*, the activity matrix ranked in the top 20 of the 28 promoters identified in this study are conserved in both the consensus promoter, and another five novel σ^28^-promoters. The rankings for each of four promoter prediction lists are shown in the following order: 200-bp upstream frequency matrix, 200-bp nonoverlap frequency matrix, 200-bp upstream activity matrix, and 200-bp nonoverlap activity matrix. * ranking is not in the top 30.

**TABLE 2. Predicted E. coli σ^28^ promoters tested for in vitro activity**

| Gene ID | Gene name | Function | Location from ORF (nt) | −35 sequence | Spacer length (nt) | −10 sequence | Rankingsa | σ^28^- activity |
|---------|-----------|----------|----------------------|--------------|-------------------|-------------|-----------|----------------|
| b0132   | yadD      | Hypothetical protein | −82               | GAA
dnaK        | 12                | 2, 5, 8, No   |
| b0763   | modA      | Molybdate ABC transporter | −129             | GTA
dnaK        | 12                | 10, 29, 18 Yes |
| b1760   | ynjH      | NADPH thioredoxin reductase, predicted protein | −62            | GTA
dnaK        | 12                | 2, 5, 8, Yes |
| b1915   | yecF      | Hypothetical protein | −134             | GAA
dnaK        | 12                | 20, 14, 11, Yes |
| b3204   | ptsN      | Enzyme IIA^1^ | −63               | GAA
dnaK        | 12                | 29, 19, 14, No |
| b3490   | yhiL      | Hypothetical protein | −118             | GAA
dnaK        | 12                | 11, 10, 12, Yes |
| b4083   | yjcS      | Hypothetical protein | −190             | GAA
dnaK        | 12                | 16, 12, 5, 5, Yes |

* The rankings for each of four promoter prediction lists are shown in the following order: 200-bp upstream frequency matrix, 200-bp nonoverlap frequency matrix, 200-bp upstream activity matrix, and 200-bp nonoverlap activity matrix. * ranking is not in the top 30.

*FIG. 6. In vitro transcription of predicted *E. coli* σ^28^-dependent promoters. Promoters were transcribed with *E. coli* core enzyme and recombinant *E. coli* σ^28^ as described in the text.*

*FIG. 7. Diagram of the *C. trachomatis* pgk gene and hrcA-gppE-dnak operon. Arrows indicate the approximate locations of transcription start sites. (A) Upstream of the pgk gene are two overlapping promoters recognized by σ^66^ and σ^28^ RNA polymerases, respectively. (B) The σ^66^ promoter for the hrcA-gppE-dnak operon is located upstream of hrcA, while a σ^28^-promoter is located upstream of dnaK within the coding region of gppE. The figure is not drawn to scale.*
polymersase in late gene expression in Chlamydia. In contrast, mRNA from pgk and bioY were detectable earlier, at 8 hpi, while the dnaK transcript was present by 3 hpi (3). It is worth noting, however, that this microarray analysis measures only steady-state transcript levels and would not be able to distinguish between the temporal activity of multiple promoters, such as transcription of pgk by both σ28 and σ66 RNA polymerases. Thus, it is entirely possible that σ28-regulated expression of these target genes may also be restricted to late time points, and as yet, there is no definitive evidence that σ28 RNA polymerase is transcriptionally active at earlier times in the chlamydial developmental cycle. In summary, there is accumulating evidence for σ28-dependent regulation of a subset of late genes in Chlamydia, distinct from the late genes transcribed by σ66 RNA polymerase (6).

pgk and dnaK are the first examples of genes in Chlamydia that can be transcribed by more than one form of RNA polymerase. With pgk, the promoters for σ28 RNA polymerase and σ66 RNA polymerase overlap and appear to have the same transcription start site (Fig. 5 and 7A), which raises the question of how promoter occupancy by the two forms of RNA polymerase is regulated. dnaK is known to be transcribed as part of the hrcA-gprE-dnak operon by σ66 RNA polymerase (21) under the control of the HrcA repressor (23). We now show that dnaK has an independent promoter that is transcribed by σ28 RNA polymerase (Fig. 4 and 7B), and we predict that this promoter is responsive to heat shock. We base this prediction on the observation that elevated temperatures have been shown to upregulate levels of the dnaK transcript by greater than 10-fold, while hrcA and gprE mRNA levels were not similarly increased (5).

While we have identified a total of six σ28-regulated genes in Chlamydia, it is not clear whether these genes belong to a specific functional group. hctB encodes Hc2, a histone-like protein that causes DNA condensation (1, 2, 4, 15, 15). Tsp is a predicted protease with similarity to CPAF, a secreted chlamydial protease that cleaves host transcription factors involved in major histocompatibility complex class I and class II antigen expression (18, 26). tycC_1 encodes a hypothetical protein, which may be involved in hemolysis (20). Of the remaining three target genes, dnaK encodes a heat shock chaperone, pgk encodes a phosphoglycerate kinase, and bioY encodes a hypothetical protein with homology to a predicted biotin synthase in 

In vitro transcription late in the developmental cycle may be triggered in response to cellular stress, such as nutrient deprivation or other conditions within the chlamydial inclusion, although the details remain to be elucidated.

Our promoter search algorithm is versatile and can be applied to predict σ28 promoters in other bacteria or promoters for other forms of RNA polymerase. σ28 promoter recognition appears to be conserved among bacteria (25) and, thus, our existing frequency- and activity-weighted matrices can be readily used for other prokaryotic genomes. With the appropriate probability weight matrix, the algorithm can also be used to identify promoters recognized by different forms of RNA polymerase. More generally, this same algorithm could be applied to any DNA sequence, such as a protein-binding site, as long as examples are available to build the weighted matrix. As our results have shown, however, an essential component of this bioinformatic approach is the validation of the in silico predictions with functional testing.

ACKNOWLEDGMENTS

We thank Eike Nichus, Johnny Akers, Elizabeth Di Russo, Allan Chen, and Narac Park for their support and G. Wesley Hatfield and Marian Waterman for critical review of the manuscript. This work was supported by a grant from the NIH (AI 44198). M.T. is supported by an NIH Independent Scientist Award (AI 057563), and H.H.Y. Y. was supported by a predoctoral training grant from the NIH (National Research Service Award I T15 LM007443 from the National Library of Medicine).

REFERENCES

1. Barry, C. E., III, S. Hayes, and T. Hackstadt. 1992. Nucleoid condensation in Escherichia coli that express a chlamydial histone homolog. Science 256: 377–379.
2. Barry, C. E., III, T. J. Brrickman, and T. Hackstadt. 1993. Hc1-mediated effects on DNA structure: a potential regulator of chlamydial development. Mol. Microbiol. 9:273–283.
3. Belland, R. J., G. Zheng, D. D. Crane, D. Hogan, D. Sturdevant, J. Sharma, W. L. Beatty, and H. D. Caldwell. 2003. Genomic transcriptional profiling of the developmental cycle of Chlamydia trachomatis. Proc. Natl. Acad. Sci. USA 100:8478–8483.
4. Brickman, T. J., C. E. Barry III, and T. Hackstadt. 1993. Molecular cloning and expression of hcb encoding a strain-variant chlamydial histone-like protein with DNA-binding activity. J. Bacteriol. 175:4274–4281.
5. Engel, J. N., J. Pollack, E. Perara, and D. Ganem. 1990. Heat shock response of marine Chlamydia trachomatis. J. Bacteriol. 172:6979–6972.
6. Fahr, M. J., A. L. Douglas, W. Xia, and T. P. Hatch. 1995. Characterization of late promoters of Chlamydia trachomatis. J. Bacteriol. 177:4242–4260.
7. Frye, J. J., E. Karlinsky, H. R. Felise, B. Marzolf, N. Dowidar, M. McClelland, and K. T. Hughes. 2006. Identification of new flagellar genes of Salmonella enterica serovar Typhimurium. J. Bacteriol. 188:2233–2243.
8. Hackstadt, T. 1999. Cell biology, p. 321. In R. S. Stephens (ed.), Chlamydia: intracellular biology, pathogenesis, and immunity. ASM Press, Washington, D.C.
9. Haldenwang, W. G. 1995. The sigma factors of Bacillus subtilis. Microbiol. Rev. 59:1–30.
10. Helmann, J. D. 1991. Alternative sigma factors and the regulation of flagellar gene expression. Mol. Microbiol. 5:2875–2882.
11. Hua, L., P. S. Hefty, Y. J. Lee, Y. M. Lee, R. S. Stephens, and C. W. Price. 2006. Core of the partner switching signalling mechanism is conserved in the obligate intracellular pathogen Chlamydomonas trachomatis. Mol. Microbiol. 59: 623–636.
12. Ide, N., T. Ikebe, and K. Kutsukake. 1999. Reevaluation of the promoter structure of the class 3 flagellar operons of Escherichia coli and Salmonella. Genes Genet. Syst. 74:113–116.
13. Mathews, S., and P. Timms. In silico identification of chlamydial promoters and their role in regulation of development. In P. Bavoil and P. Wyrick (ed.), Chlamydia: genomics, pathogenesis and implications for control, in press. Horizon Scientific Press, Wymondham, United Kingdom.
14. Pedersen, L. B., S. Birkelund, and G. Christiansen. 1994. Interaction of the Chlamydomonas trachomatis histone H1-like protein (Hc1) with DNA and RNA causes repression of transcription and translation in vitro. Mol. Microbiol. 11:1085–1098.
15. Pedersen, L. B., S. Birkelund, and G. Christiansen. 1996. Purification of recombinant Chlamydia trachomatis histone H1-like protein Hc2, and comparative functional analysis of Hc2 and Hc1. Mol. Microbiol. 20:295–311.
16. Perara, E., D. Ganem, and J. Engel. 1992. A developmentally regulated chlamydial gene with apparent homology to eukaryotic histone H1. Proc. Natl. Acad. Sci. USA 89:2125–2129.
17. Schaumburg, C. S., and M. Tan. 2003. Mutational analysis of the Chlamydia trachomatis dnaK promoter defines the optimal –35 promoter element. Nucleic Acids Res. 31:551–555.
18. Shaw, A. C., B. B. Vandahl, M. R. Larsen, P. Roepstorff, K. Gevaert, J. Vandekerckhove, G. Christiansen, and S. Birkelund. 2002. Characterization of a secreted Chlamydia protease. Cell. Microbiol. 4:411–424.
19. Shen, L., M. Li, and Y. X. Zhang. 2004. Chlamydia trachomatis sigma28 recognizes the fliC promoter of Escherichia coli and responds to heat shock in chlamydiae. Microbiology 150:205–215.
20. Stephens, R. S., S. Kalman, C. Lammel, J. Fan, R. Marathe, L. Aravind, W. Mitchell, L. Olinger, R. L. Tatusov, Q. Zhao, E. V. Koonin, and R. W. Davis. 1998. Genome sequence of an obligate intracellular pathogen of humans: Chlamydia trachomatis. Science 282:754–759.
21. Tan, M., and J. N. Engel. 1996. Identification of sequences necessary for transcription in vitro from the Chlamydia trachomatis rRNA P1 promoter. J. Bacteriol. 178:6975–6982.
22. Tan, M., T. Gaul, R. L. Gourse, and J. N. Engel. 1998. Mutational analysis of the Chlamydia trachomatis rRNA P1 promoter defines four regions important for transcription in vitro. J. Bacteriol. 180:2359–2366.
23. Wilson, A. C., and M. Tan. 2002. Functional analysis of the heat shock regulator HrcA of Chlamydia trachomatis. J. Bacteriol. 184:6566–6571.
24. Yu, H., and M. Tan. 2003. Sigma 28 RNA polymerase regulates hctB, a late developmental gene in Chlamydia. Mol. Microbiol. 50:577–584.
25. Yu, H. Y., E. G. Di Russo, M. A. Rounds, and M. Tan. 2006. Mutational analysis of the promoter recognized by Chlamydia and Escherichia coli σ^70 RNA polymerase. J. Bacteriol. 188:5524–5531.
26. Zhong, G., P. Fan, H. Ji, F. Dong, and Y. Huang. 2001. Identification of a chlamydial protease-like activity factor responsible for the degradation of host transcription factors. J. Exp. Med. 193:935–942.