Immobilization of Escherichia coli RNA Polymerase and Location of Binding Sites by Use of Chromatin Immunoprecipitation and Microarrays†

Christopher D. Herring,1,2,4 Marni Raffaele,2 Timothy E. Allen,1 Elenita I. Kanin,2 Robert Landick,4 Aseem Z. Ansari,2,3* and Bernhard Ø. Palsson1*

Department of Bioengineering, University of California San Diego, San Diego, California1; Department of Biochemistry, University of Wisconsin Madison, Madison, Wisconsin2; Genome Center of Wisconsin, Madison, Wisconsin3; and Department of Bacteriology, University of Wisconsin Madison, Madison, Wisconsin4

Received 20 April 2005/Accepted 14 June 2005

The genome-wide location of RNA polymerase binding sites was determined in Escherichia coli using chromatin immunoprecipitation and microarrays (chIP-chip). Cross-linked chromatin was isolated in triplicate from rifampin-treated cells, and DNA bound to RNA polymerase was precipitated with an antibody specific for the β′ subunit. The DNA was amplified and hybridized to “tiled” oligonucleotide microarrays representing the whole genome at 25-bp resolution. A total of 1,139 binding sites were detected and evaluated by comparison to gene expression data from identical conditions and to 961 promoters previously identified by established methods. Of the detected binding sites, 418 were located within 1,000 bp of a known promoter, leaving 721 previously unknown RNA polymerase binding sites. Within 200 bp, we were able to detect 51% (189/368) of the known σ70-specific promoters occurring upstream of an expressed open reading frame and 74% (273/368) within 1,000 bp. Conversely, many known promoters were not detected by chIP-chip, leading to an estimated 26% negative-detection rate. Most of the detected binding sites could be associated with expressed transcription units, but 299 binding sites occurred near inactive transcription units. This map of RNA polymerase binding sites represents a foundation for studies of transcription factors in E. coli and an important evaluation of the chIP-chip technique.

RNA polymerase (RNAP) in Escherichia coli is a key factor in gene expression and catalyzes the transcription of DNA to mRNA for all genes (reviewed in references 33 and 37). The core enzyme is composed of four subunits: two α, one β, one β′, and one ω. Core RNAP becomes transcriptionally active holoenzyme with the addition of a σ factor (Fig. 1A). E. coli possesses seven interchangeable σ factors, each with specificity for different promoters. Sigma factors function as global regulators of gene expression and mediate the transcriptional response to conditions under which a large number of genes need to be turned on or off, such as stationary phase or heat shock. Except for σ70, sigma factors do not bind to DNA except as part of the holoenzyme complex (9).

To map promoters in bacteria, we sought a way to force RNAP to reside only at promoters so that identifying DNA fragments bound to RNAP in vivo would report promoter locations. A variety of small-molecule inhibitors of RNA polymerase were evaluated for the immobilization of RNAP, and rifampin was found to work best (M. Raffaele, E. Kanin, J. Vogt, R. R. Burgess, A. Z. Ansari, unpublished data). The antibiotic rifampin inhibits bacterial growth by binding the β subunit of RNAP just upstream of the active site, blocking the synthesis of RNAs longer than 2 to 3 nucleotides (nt) (11) (Fig. 1A). Rifampin has no effect on RNAP promoter binding to form closed complexes or on open complex formation (30) and has no effect on RNAP in vitro when added after elongating RNAP has cleared the promoter (12, 38). Upon addition of rifampin to a growing culture, the rifampin stops growth without killing cells by diffusion across bacterial membranes and tight binding to RNAP not engaged in transcription. RNAP holoenzyme complexed with rifampin can still bind to promoters but is trapped in an abortive cycle and unable to extend RNAs beyond 2 to 3 nt (Fig. 1A). However, RNAP molecules in elongation complexes with RNA and DNA are resistant to rifampin binding. Once elongating RNAPs terminate transcription and release RNA and DNA, they become susceptible to rifampin binding, which traps them in newly formed open complexes.

A promoter is a DNA sequence to which RNAP binds and initiates the transcription of RNA. Knowledge of promoter locations is the first step in the elucidation of the transcriptional regulatory network. It allows the identification of which genes are cotranscribed and, from the DNA sequence, identification of regulatory motifs associated with different regulators. It also provides a basis for the interpretation of the binding site locations of various transcription factors. Promoters have been identified on an individual basis by primer extension and DNA footprinting studies, by which 961 promoters have been identified (www.cifn.unam.mx/Computational_Genomics/regulondb/) (36). In addition, DNA binding motifs have been used to predict the locations of 4,641 promoters throughout
New experimental evidence of promoter locations may validate or refute these predictions and will assist in understanding the function of uncharacterized genes and operons.

A powerful method combining chromatin immunoprecipitation and microarray analysis called chIP-chip (or genome-wide location analysis) has been developed for the global identification of transcription factor binding sites (10, 23, 34). In *E. coli* it has been used to identify the locations of MelR binding sites (16), and in yeast it has been used to identify the targets of over 200 transcription factors (18). Furthermore, chIP-chip was used to identify the targets of yeast RNA polymerase III (Pol III) and to investigate the roles of the accessory factors TFIIIB and TFIIIC (19, 31, 35).

These previous studies used microarrays spotted with PCR products of intergenic DNA and sometimes coding sequence as well. The resolution of the method depends on the size of the spotted PCR products in addition to the size of DNA fragments used in chromatin immunoprecipitation. Oligonucleotide microarrays have also been used (13) and hold the promise of greater resolution and the statistical confidence endowed by multiple signals from each region of the genome. However, the distribution of oligonucleotide probes is highly skewed in *E. coli* Affymetrix arrays and many regions of the genome are devoid of probes. Therefore, in this study we used a custom-designed oligonucleotide array from Nimblegen containing an evenly spaced “tiled” set of probes. Custom Nimblegen arrays have been used previously for chIP-chip of human transcription factors (25), but to our knowledge, this is the first publication describing the use of whole-genome, uniformly tiled arrays.

The accuracy and resolution of chIP-chip have been evaluated in a number of previous studies. For yeast RNA Pol III, essentially all known targets were correctly identified (19, 31, 35). A comparison of known transcription factor interactions to a large set of chIP-chip data revealed a false-negative rate of 80% (41). In the present study, we used rifampin to trap RNAP at promoters and compared promoter locations determined by chIP-chip to those determined by established methods. The existence of hundreds of known promoter locations in *E. coli* provides a novel opportunity for the evaluation of this technique.

FIG. 1. RNAP transcription cycle and the effect of rifampin. (A) Mechanisms of transcription initiation and rifampin binding. RNAP (blue) binds DNA to form a closed complex with affinities that differ among promoters by at least a factor of 100. The closed complex isomerizes to an open complex at rates that differ among promoters by at least a factor of 100. The open complex is stable at some promoters; at others it collapses back to the closed complex unless it binds initiating nucleoside triphosphate substrates. The competition between abortive initiation and promoter escape (initiation) is highly variable among promoters. Rifampin can bind to all species of RNAP except the elongation complex and traps RNAP in open complexes by forcing the release of 2- to 3-nt abortive transcripts. (B) Effect of rifampin on growth and RNAP location. *E. coli* cells were cultured, and the absorbance was determined at given time intervals. Rifampin (150 μg/ml) was added at 200 min to one culture (■), while the other was allowed to continue growing as a control (●). After 20 min, the culture with rifampin added stopped growing, while the control continued in exponential growth. The inset shows the results of quantitative PCR using primers designed to amplify in the promoter, 16S, or 23S subunit genes from all seven rRNA operons. Cells were treated with rifampin for 0, 20, or 40 min followed by immunoprecipitation using antibody for RNAP β' core subunit. The results of PCR performed on genomic DNA before immunoprecipitation are shown on the right end of the gel image, labeled “input.” The quantified relative intensity of each 16S and 23S band is given below the gel image.
MATERIALS AND METHODS

Chromatin immunoprecipitation. In preparatory studies performed at least three times, *E. coli* K-12 MG1655 cells were grown in 100 ml LB medium in a shaking flask at 37°C. Chromatin immunoprecipitation was performed similarly to the method described in detail below. Endpoint quantitative PCR was performed for 30 cycles using primers designed to amplify from all seven rrn operons. For 16S the primers were GGAGAATACGCGTGCAGAGG and GC GGGAGTCCCCAGAAAGCCACGCTC. For 23S they were CCAGTAGTG TTGCTTAGAAGCGACC and ACAGAAGGCTTCCCCATACCAAC. For the promoter, the reverse-direction primer was GTCTGATTAAATGGTTAAAGAG CAGT, while the forward-direction primers were a mixture of TCCCTTATA AGGCGCAACACTG, TCCCTTATAAGTGCGAATTCC, and ATCCCTAT AATGCGCCTCCGT.

For chIP-chip, three biological replicates were performed (three individual cultures). For each culture, two cell pellets of 50 ml culture were processed—one experimental sample as described below and a second control sample identical to the first but with no antibody added to the immunoprecipitation reaction. The chIP DNA of the first biological replicate was analyzed by real-time PCR using primers in the promoter regions of 5Sd and 5Fd, which were measured to be present at levels 25- and 7-fold higher, respectively, than the 16S and 23S rRNA genes. The second biological replicate was amplified using random primers and control PCR-amplified samples were labeled with Cy3 and Cy5 by Nimblegen, swapping the cyanine dyes in one biological replicate. The microarrays consisted of 371,034 oligonucleotides from the *E. coli* genome (release m58 accession no. NC_000913) plus controls. The oligonucleotides were 50 nt long and spaced every 25 nt on both top and bottom strands.

The method used for chromatin immunoprecipitation was adapted from Lin and Grossman (28). *E. coli* MG1655, obtained from the *E. coli* Genetic Stock Center (CGSC no. 7740), was inoculated from frozen stocks into M9 minimal medium—0.2% glucose and grown overnight at 37°C. It was then diluted 1:50 into 200 ml M9-0.2% glucose in a 500 ml flask, and the mouth of the flask was sealed and stirred for 20 min. Cultures were monitored by OD to verify the inhibitory effects of rifampin. A slight increase in OD was observed but with a greatly reduced growth rate. Samples (50 ml) were transferred to tubes containing formaldehyde and sodium phosphate (pH 7.6) such that the final concentrations were 1% and 10 mM, respectively. They were incubated at room temperature for 10 min with gentle agitation, and then glycine was added to 100 mM. They were gently agitated at 4°C for 30 min. Cells were centrifuged at 3,500 rpm and washed twice in chilled phosphate-buffered saline. The cells were transferred to 1.5 ml tubes, centrifuged, and stored at −80°C.

Cells were resuspended in 0.5 ml of a solution containing 10 mM Tris (pH 8), 50 mM NaCl, 10 mM EDTA, and 20% sucrose, and then 0.2 μl Ready-labeled lysosome (Epicerin) was added. They were incubated for 30 min at 37°C, and then 0.5 ml of a solution containing 200 mM Tris (pH 8), 600 mM NaCl, 4% Triton X-100, 1 mM phenylmethylsulfonyl fluoride, and 2 μg/ml RNase A was added. They were incubated at 37°C for 10 min and then chilled on ice. The lysates were sonicated four times for 20 s each time using a Sonics and Materials VC50 with a 3 mm microtip. Cell debris was removed by centrifugation, and a sample was removed and extracted with phenol and phenol/chloroform for DNA size analysis. DNAs ranged from 100 to 1,200 bp, with the greatest intensity at ~600 bp. Pan Mouse immunoglobulin G magnetic beads (Dynal catalog no. 110.22) (40 μl) were resuspended and washed per the manufacturer’s instructions and then added to the lysates and rotated at 4°C for 3 h. The beads were removed using a magnet and discarded. A total of 2 μl of antibody for RNA polymerase β’ (Neclon® catalog no. W0001) was added, and the samples were rotated overnight at 4°C. Magnetic beads (50 μl) were resuspended and washed as before and then added to the samples and rotated 1 h at 4°C. The beads were recovered and then washed once in 250 mM LiCl-100 mM Tris pH 8-2% Triton X-100, twice in 100 mM Tris (pH 8)-600 mM NaCl-2% Triton X-100, twice in 100 mM Tris (pH 8)-300 mM NaCl-2% Triton X-100, and twice in TE (10 mM Tris-HCl, 1 mM EDTA). DNA and protein were eluted from the beads by resuspending in 50 mM Tris (pH 8)-10 mM EDTA-1% sodium dodecyl sulfate and then incubated at 65°C for 20 min. Beads were removed, cross-links were reversed by incubating at 65°C for 4 h, and then DNA was purified using Qiagen (QIAGEN).

ChIP DNA was amplified based on the Random DNA Amplification method of June 2001 from http://www.microarrays.org/protocols.html. The following primers were designed specifically for *E. coli*: PF43 (TGGAAATCCGAGTGAGT) and PF44 (TGGAAATCCGAGTGAGT). Round B amplification using cTaq (Takara) was performed for 35 to 40 cycles. Round C amplification was not performed. PCR products were purified with Qiagen and then ethanol precipitated. Pellets were resuspended in water, and a sample was diluted in 1 mM Tris (pH 8) to measure the DNA concentration. Experimental samples (4 μg) and control samples (3 μg) were hybridized to microarrays by Nimblegen.

Data analysis. For each array, the Cy5 signal values were multiplied by a scaling factor in Microsoft Access so that the total signals for Cy3 and Cy5 were equal. The log of the ratio of experimental signal to control signal was then calculated. The microarray was designed from both the top- and bottom-strand sequences, which means that for every probe, its reverse complement was also represented on the array. To distinguish them in the resulting data, 12 nt were added to the nominal genome position of probes from the bottom strand. Genome-scale data was visualized using Nimblegen’s SignalMap software, and most other analysis was performed with Matlab version 6.5 (see the program code in the supplemental material). For each probe, the mean log2 ratio from all three biological replicates was calculated and then smoothed using a moving average with a window size of 500 bp. Peaks were identified using a reiterative process in which the probe with the highest log2 ratio was located upon each iteration. To prevent the same peak from being reidentified each time, the probe with the highest log2 ratio and all data within 1,000 bp were set to zero. This process was reiterated 1,000 times, and the resulting peaks were manually edited using SignalMap for visualization. An additional 139 peaks with a log2 ratio greater than 0.1 were then identified by the same method and manually edited.

RESULTS AND DISCUSSION

Effects of rifampin and experimental design. We first determined the time required for rifampin to localize RNAP to promoters when added to mid-log *E. coli* cultures. The outer membrane of gram-negative bacteria like *E. coli* poses a significant barrier to rifampin, requiring much higher concentrations for inhibition of RNAP in vivo than in vitro (4). When 150 μg/ml rifampin was added, growth ceased almost immediately (Fig. 1B). However, chIP signals in the rRNA operon persisted much longer, indicating a lack of coupling between rRNA transcription and cell growth. We found that 20 min of incubation with rifampin was required for chIP signals in the rRNA operon to reach near-background levels. Based on this result, we chose 20 min as a suitable time for rifampin treatment. This amount of time should allow detection of even the weakest promoters; when the occupancy rate of a promoter by RNAP is only once every 20 min, then a chIP-chip signal will be detectable, though slightly diminished.

Binding sites for RNA polymerase in the genome of *E. coli* strain MG1655 were identified using chIP-chip. Briefly, cells were grown to mid-log phase in minimal glucose medium and then treated for 20 min with rifampin. DNA and protein were cross-linked with formaldehyde, and cells were lysed. DNA was sheared to ~600 bp with sonication, and RNAP was immunoprecipitated using an antibody specific for the β’ subunit. DNA bound by RNAP and from a no-antibody mock immunoprecipitation control was amplified using random primers and then labeled with Cy3 and Cy5 fluorescent dyes and hybridized to a custom microarray consisting of 371,034 *E. coli* oligonucleotides spaced 25 bp apart across the whole genome.

Reproducibility. This experiment was repeated three times. The reproducibility of the log2 ratios from each array is shown in Fig. 2. Higher correlation was seen when comparing just the experimental channels (coefficients of 0.71, 0.63, and 0.84) or just the controls (0.66, 0.61, and 0.87). This level of correlation is lower than typically observed in microarray analysis of gene expression, but important differences should be considered. In this experiment, the whole genome was represented with oligonucleotide probes, and yet only a fraction of the genome should be present in the DNA that was hybridized to it. Thus,
the majority of oligonucleotide probes on the array are not hybridized, leading to an increased amount of noise. The log₂ ratios of chIP-chip experiments tend to have less dynamic range and greater noise than those from expression analysis (R. Green, Nimblegen, personal communication).

Unlike a typical gene expression experiment, the absolute log₂ ratio is not as important as the shape and location of peaks within the log₂ ratio data, as visible in Fig. 3. As expected from the high density of evenly spaced probes, the log₂ ratio showed a clear pattern of peaks and valleys. When peaks were identified from each biological replicate individually, the mean standard deviation of peak location was 206 bp. The maximum log₂ ratio in each biological replicate ranged from 4.2 to 5.9, and for all further analysis, replicates were combined by calculating the mean. To control for differences in hybridization efficiency of the oligonucleotides, the log₂ ratios were smoothed with a moving average. In Fig. 3B it can be seen that smoothing preserved the shape of peaks yet eliminated much of the noise.

A large cluster of outliers is visible in the top two panels of Fig. 3. As expected from the oligonucleotides, the log₂ ratios were smoothed with a mean. To control for differences in hybridization efficiency of ratio in each biological replicate ranged from 4.2 to 5.9, and for log₂ ratio is not as important as the shape and location of peaks (R. Green, Nimblegen, personal communication).

The rifampin stock used in biological replicate 1 was filter sterilized, which probably removed some rifampin and lowered the actual concentration. For biological replicates 2 and 3 the rifampin was not filtered. It seems likely that in biological replicate 1 the rifampin concentration was lower, and RNAP was able to escape from these very strong promoters. It should be noted that two peaks were observed at all of the seven rRNA gene operons. The first peak corresponds to the P1 and P2 promoters, indistinguishable at this scale. The other peak is probably a conserved promoter of unknown significance noted previously (2).

**Binding site locations.** From the combined smooth data, 1,139 peaks were identified, representing binding sites for RNAP. They may also represent promoters, if RNA is transcribed from them. Binding sites less than a few hundred base pairs apart or on opposite strands are expected to appear as a single peak. Binding and initiation by RNAP is strongly dependent on other transcription factors and the conditions tested. Presumably, the sites we detected are a subset of the total binding sites in *E. coli*. For instance, peaks were not observed at the sites of either the araBAD or lacZYA promoters, probably because those binding sites are occluded by repressors in glucose medium.

There are ~2,800 RNAP molecules in *E. coli* in minimal medium (8) and 2,428 predicted transcription units (TUs) in the genome, not including ~40 TUs containing recently annotated small regulatory RNAs (TU predictions are from www.biostat.wisc.edu/gene-regulation/) (7). Of these TUs, 1,552 contain an open reading frame (ORF) that appears to be expressed under these conditions (log₂ signal > 8.0 in gene expression data generated with Affymetrix arrays) (data from reference 14). Of these, 1,141 have 5’ ends that are more than 500 bp apart, providing an estimate of the number of binding sites that we might expect in this experiment. This number agrees remarkably well with our results. Of the 1,139 sites we detected, 958 occurred within 1,000 bp of the start site of a TU. For 659 of these, the TU contained an ORF with a log₂ signal > 8.0 in gene expression data from the same growth conditions. Interestingly, this indicates 299 binding sites that are not associated with active transcription. In these cases, RNAP may bind to the chromosome but not activate transcription until a repressor or activator is triggered in response to environmental conditions, as is generally the case for transcription from σ^54- specific promoters. It should be noted, though, that after treatment with rifampin, all active promoters may become saturated by RNAP, forcing binding to sites that are not normally bound. Additionally, some promoters may bind RNAP weakly and yet still be active by having high rates of open complex formation or promoter clearance.

Of the 1,139 binding sites, 501 occurred in intergenic regions...
and 638 occurred in ORFs. Those in ORFs can be subdivided into three groups on the basis of the distance from the site to the nearest 5' end of an ORF (median distance = 158 bp). The first group consists of binding sites that are probably located upstream of the ORF but mistakenly identified inside the ORF due to poor resolution of the chIP-chip technique. In this group are 372 sites for which the nearest 5' end of an ORF is that of the ORF in which they occur and is 500 bp away. The second group consists of sites that are probably promoters for a downstream ORF. In this group are 191 sites for which the nearest 5' end of an ORF is an ORF that is different from the one in which it occurs and is <500 bp away. The third group consists of sites that occur >500 bp away from the 5' end of any ORF (see Table S3 in the supplemental material). Of these 75 sites, many have small log2 ratios and occur in insertion elements or in genes of unknown function. The others occur either in the middle of an ORF (sites in sapA, rfbX, evgS, emrA, nlpD, pyrG, xylG, and fimD) or near the intergenic region between two convergent genes (sites in pinQ, intQ, and otsA). These sites may play a regulatory role, possibly as RNAP pause sites or promoters of antisense transcripts, or they may be cryptic promoter-like elements ordinarily masked by transcribing RNAP.

**Correlation to other factors.** There was very little correlation between the chIP-chip log2 ratio and the similarity of known promoters to a σ70 consensus binding matrix (Fig. 4A). This result is surprising given that (i) the relative efficiency of chIP, often measured as “occupancy,” has been observed to correlate with in vitro binding affinity (39) and (ii) RNAP has high affinity for promoters containing sequences close to consensus (26). A possible explanation for this lack of correlation may be the excess of RNAP molecules over the number of
Affymetrix arrays using the same growth conditions as for chIP-chip from reference 14 were generated from three biological replicates with expression of the nearest downstream ORF. Gene expression data of the genome or some other factor, is unknown (1, 24) through the meaning, whether related to the physical organization of the genome and other factors are associated with the transcription of genes for flagella biosynthesis and stress response. Gene expression measurements indicate that the genes downstream from $\sigma^{28}$-specific promoters are probably not expressed under these conditions (average log$_2$ signal = 7.13), while those downstream of $\sigma^{24}$- and $\sigma^{32}$-specific promoters probably are (average log$_2$ signals of 9.62 and 10.71, respectively). The expression of stress response genes does not necessarily indicate the presence of stress; heat-shock proteins are normally expressed at $\sim$37°C (20), and $\sigma^{32}$ is required for growth above 20°C (42). The log$_2$ ratio of chIP-chip peaks near $\sigma^{24}$- and $\sigma^{32}$-specific promoters showed poor correlation to gene expression ($r = 0.17$). Few $\sigma^{35}$- and $\sigma^{32}$-specific promoters were detected by chIP-chip. These sigma factors are associated with the transcriptional response to stationary phase and nitrogen assimilation, respectively.

These differences in the proportions of known promoters detected do not appear to match the relative abundances of the potential binding sites in the presence of rifampin. In excess and “trapped” at promoters, RNAP may bind to all sites equally. We also found very little correlation between the log$_2$ ratio of chIP-chip peaks and the expression of the nearest downstream ORF (Fig. 4B), which is not surprising considering that the log$_2$ ratio correlates poorly with the consensus matrix. This observation differs from studies of yeast Pol III in which the log$_2$ ratio was directly correlated with transcription (35). The discrepancy may be due to biological differences between eukaryotic and prokaryotic RNAP, such as the ratio of enzyme molecules to potential binding sites or to disparate mechanisms of gene regulation. More likely, it is due to experimental differences; in the study performed by Roberts et al. (35) rifampin was not used to trap RNAP at the promoters and occupancy was measured in a transcribed region.

We examined the distribution and periodicity of binding sites in the E. coli genome. The densities of binding sites differed throughout the genome, ranging from 13 to 36 sites per 100 kb. There was no correlation between the number of binding sites per 100 kb and the average predicted TU size in each 100-kb region (correlation coefficient = $-0.11$; TU predictions from Bockhorst et al.) (7). We did not observe any significant periodicity in the location of binding sites, but we did detect periodicity when both the location and log$_2$ ratio of the peaks were considered (Fig. 5). A wavelet transform (6, 29) revealed that the average log$_2$ ratio ranged in a region from $\sim$1,900 to 3,400 kb with periodicity of $\sim$660 kb, which is significant relative to the periodicity found in randomized data. The regions of high and low average log$_2$ ratios correlated with regions of high and low average gene expression levels (Fig. 5C). Periodicity of this sort has been observed for many different genomic parameters across the evolutionary spectrum, though the meaning, whether related to the physical organization of the genome or some other factor, is unknown (1, 24) (T. E. Allen, N. D. Price, A. R. Joyce, B. Ø. Palsson, unpublished data). It is unusual that only a fraction of the genome shows periodicity for RNAP binding sites. It is also interesting that a correlation between log$_2$ ratio and gene expression was only observed when the log$_2$ ratios were smoothed over 330 kb. On a gene-by-gene basis there may have been too much noise to see this trend.

Comparison to locations of known promoters. To validate the RNAP binding sites, we examined the ability of chIP-chip to detect 961 known promoters compiled in RegulonDB (36). For each known promoter, the nearest peak was located in the RNAP chIP-chip data (Table 1). A total of 418 RNAP binding sites were located within 1,000 bp of a known promoter, leaving 721 binding sites identified by chIP-chip that have not been previously characterized. Of the known promoters detected, most were specific to $\sigma^{34}$, but high percentages of $\sigma^{24}$-, $\sigma^{28}$-, and $\sigma^{32}$-specific promoters were also detected. These sigma factors are associated with the transcription of genes for flagella biosynthesis and stress response. Gene expression measurements indicate that the genes downstream from $\sigma^{30}$-specific promoters are probably not expressed under these conditions (average log$_2$ signal = 7.13), while those downstream of $\sigma^{24}$- and $\sigma^{32}$-specific promoters probably are (average log$_2$ signals of 9.62 and 10.71, respectively). The expression of stress response genes does not necessarily indicate the presence of stress; heat-shock proteins are normally expressed at $\sim$37°C (20), and $\sigma^{32}$ is required for growth above 20°C (42). The log$_2$ ratio of chIP-chip peaks near $\sigma^{24}$- and $\sigma^{32}$-specific promoters showed poor correlation to gene expression ($r = 0.17$). Few $\sigma^{35}$- and $\sigma^{32}$-specific promoters were detected by chIP-chip. These sigma factors are associated with the transcriptional response to stationary phase and nitrogen assimilation, respectively.
sigma factors. During logarithmic growth in LB medium, protein measurements showed that σ^{70} is predominant, followed in abundance by σ^{38} and σ^{28}, while the others were undetectable (22). Under the conditions used here for chIP-chip, the mRNAs for all of the sigma factors appeared to be expressed, as mRNAs for all of the sigma factors appeared to be expressed;

FIG. 5. Density and periodicity of RNA polymerase binding sites in glucose minimal medium as detected by chIP-chip. (A) The log2 ratios of 1,139 binding sites were averaged in each 1-kb section of the genome. Those sections lacking a binding site were assigned a value of zero. These data were examined for periodicity by wavelet analysis, which smooths the data over a range of window sizes to reveal the zero. These data were examined for periodicity by wavelet analysis, which smooths the data over a range of window sizes to reveal the zero. Those sections lacking a binding site were assigned a value of 1,139 binding sites were averaged in each 1-kb section of the genome. Those sections lacking a binding site were assigned a value of zero. These data were examined for periodicity by wavelet analysis, which smooths the data over a range of window sizes to reveal the zero.

TABLE 1. Proximity of known promoters to RNAP binding sites detected in this study

<table>
<thead>
<tr>
<th>Sigma factor specificity</th>
<th>No. of promoters in RegulonDB</th>
<th>No. (%) of RNAP sites within 200 bp</th>
<th>No. (%) of RNAP sites within 500 bp</th>
<th>Log2 mRNA expression of the sigma factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>σ^{19}</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>8.9</td>
</tr>
<tr>
<td>σ^{24}</td>
<td>32</td>
<td>12 (38)</td>
<td>15 (47)</td>
<td>10.9</td>
</tr>
<tr>
<td>σ^{28}</td>
<td>13</td>
<td>4 (31)</td>
<td>6 (46)</td>
<td>8.2</td>
</tr>
<tr>
<td>σ^{32}</td>
<td>29</td>
<td>12 (41)</td>
<td>14 (48)</td>
<td>9.6</td>
</tr>
<tr>
<td>σ^{38}</td>
<td>62</td>
<td>4 (6)</td>
<td>9 (15)</td>
<td>11.9</td>
</tr>
<tr>
<td>σ^{54}</td>
<td>28</td>
<td>3 (11)</td>
<td>6 (21)</td>
<td>12.0</td>
</tr>
<tr>
<td>σ^{70}</td>
<td>681</td>
<td>331 (49)</td>
<td>440 (65)</td>
<td>10.3</td>
</tr>
<tr>
<td>Unknown</td>
<td>143</td>
<td>39 (27)</td>
<td>56 (39)</td>
<td></td>
</tr>
</tbody>
</table>

a To avoid counting binding sites due to a nearby σ^{70} promoter, only those promoters that occurred more than 500 bp away from a known σ^{70}-specific promoter were counted.

b The log2 of signal from Affymetrix gene expression arrays of cells grown to log phase in M9 minimal medium plus glucose (from reference 14).

RNAP to σ^{38}- and σ^{54}-specific promoters under these conditions. In any case, the detection of binding sites for the alternative sigma factors indicates that either those sigma factors are present at some level under these conditions or there is degeneracy in the binding specificity of σ^{70}-holoenzyme in vivo.

The distribution of distances between known σ^{70}-specific promoters and chIP-chip RNAP binding sites is shown in Fig. 6. We found that 49% of known σ^{70}-promoters correlated to a chIP-chip peak within 200 bp, and 73% correlated within 1,000 bp. Of 681 known σ^{70}-promoters, 532 occur upstream of an apparently expressed ORF (the nearest downstream ORF had a log2 signal > 8.0). Of these, 368 were separated by at least 500 bp, and out of 368 sites, 273 could be linked to a chIP-chip peak within 1,000 bp. This leaves 95 known expressed promoters that we were unable to detect, resulting in a 26% (95/368) false-negative rate. This compared favorably to chIP-chip experiments in yeast, for which a false-negative rate of 80% has been estimated (41). It falls short, though, of the nearly perfect detection of RNA Pol III binding sites (19, 31, 35). The false-positive rate is impossible to estimate, because binding does not always result in transcription and because of the lack of any other viable methods for measuring protein binding in vivo.

False negatives. A variety of factors could explain the false negatives. One explanation may be failure to amplify some parts of the chromosome by random PCR or low efficiency of formaldehyde cross-linking at some promoters (17). As chIP-chip data becomes increasingly common, characteristics of poorly cross-linked proteins or DNA sequences may emerge. It is also possible that the epitope for the antibody we used was occluded by other transcription factors at some promoters. Antibodies specific to other RNAP subunits are available, and it may be found that their use detects an overlapping set of RNAP binding sites. Results obtained by performing a single replicate of chIP-chip with antibodies for σ^{70} and β’ with Affymetrix arrays showed that the σ^{70} sites were a subset of the sites detected using antibody for β’, as expected (see Table S4 and Fig. S2 in the supplemental material). Of 13,499 probes with σ^{70} log2 ratios > 1, 97% had similar β’ log2 ratios (difference < 1). On the other hand, of 15,032 probes with β’ log2 ratios > 1, only 74% had similar σ^{70} log2 ratios. The log2 ratios from the β’ and σ^{70} Affymetrix chIP-chip experiments corre-
lated with a coefficient of 0.72. The log2 ratios from the Affymetrix chips were compared to those from the Nimblegen chips. For each Affymetrix probe, the nearest Nimblegen probe in the same genomic orientation was identified, correcting for the different versions of the E. coli genome sequence that were used in chip design. The correlation coefficients were 0.60 for comparisons of Affymetrix β′ to Nimblegen β′ and 0.64 for comparisons of Affymetrix σ70 to Nimblegen β′. Affymetrix arrays were not used in any further experiments because of the incomplete coverage of the E. coli genome provided by Affymetrix expression arrays (see Fig. S2 in the supplemental material).

Another likely explanation for failure to detect RNAP at known σ70 or other sigma factor promoters is the high variability of the rates of open complex formation and promoter escape and of the affinity of promoters for initial binding of RNAP holoenzyme (33). These same parameters may explain why it takes so long to clear RNAP from the rRNA operons. At some E. coli promoters, open complexes are so unstable that they initiate immediately upon binding nucleoside triphosphate without the slow promoter escape and high probability of abortive initiation observed at many promoters. The rm P1 promoter is a classic example of such a promoter. It exhibits no abortive initiation and has the added features of highly efficient initial RNAP binding and the ability to sequester up to 80% of transcribing RNAP into elongation complexes on the seven rm operons in E. coli (15, 32). RNAP released at a terminator may rebind an rRNA promoter and reform an rm elongation complex more rapidly than rifampin can bind to it (σ70 can associate with RNAP during termination) (3). Thus, RNAP may become trapped at rRNA promoters only after multiple rounds of transcription and when the rifampin concentration inside the cell accumulates to higher concentration, accounting for the slow clearance of RNAP from the rm operon (Fig. 3A). Other promoters, which form unstable open complexes but lack the avid initial RNAP binding and rapid promoter escape properties of rm promoters, may never accumulate significant RNAP, because RNAP will equilibrate away to promoters able to form stable open complexes. Thus, certain promoters that form unstable open complexes may escape detection. Consideration of the kinetic properties of promoters will be an important issue in the development of promoter mapping by the rifampin-trapping chIP-chip method.

It should also be noted that many of the sites we identified are likely to be specific to the minimal glucose conditions used here. Expansion of this experiment to other culture conditions will likely reduce the number of undetected known promoters and reveal differences in RNAP binding resulting from the activity of the regulatory network. Gene expression data also show changing patterns under different conditions but are often difficult to interpret. By using chIP-chip, direct regulatory interactions can be distinguished from indirect regulatory interactions (27, 34), but care should be practiced given the false-negative rate of 26% that we observed.

Closing remarks. Transcriptional regulation can be seen as a combinatorially based molecular “computation.” In this analogy, the binding and interaction of various transcription factors with RNAP are the input variables and transcription level is the output. The identification of transcription factors involved in each “computation” is a fundamental step in understanding observations of gene expression. As each new factor is identified by chIP-chip or some other technique, its role will be defined by its interaction and colocation with RNAP. The work presented here lays the foundation for future work in elucidating the regulatory network of E. coli. The location data can also be used for validation of computational predictions of promoters, and the log2 values may provide insight into some characteristic of RNAP binding or activity. Whether the observed false-negative detection of known promoters reveals an inherent limitation of the technique or is merely a peculiarity of the experimental conditions is yet to be seen. A comparison of experiments with or without rifampin added, under different growth conditions, and with antibodies for various sigma factors will shed light on this question.

ACKNOWLEDGMENTS

The analysis of known promoters would not have been possible without the data provided by Araceli Huerta-Moreno, Heladia Salgado-Osorio, and Julio Collado-Vides. We thank B. Ren, B. K. Cho, and A. Joyce as well as H. Holster, K. Nuwaysir, and M. Hogan at Nimblegen for their help.

This work was initiated at the University of Wisconsin Madison with the support of University of Wisconsin I&EDR funds, Hatch McIntyre Stennis (U.S. Department of Agriculture), and University of Wisconsin trust funds (to A.Z.A.) and a grant from the National Institutes of Health (GM38660 to R.L.) and completed at the University of California San Diego with support from the National Institutes of Health (GM62791 to B.O.P.). C.D.H. was supported in part by NLM T15LM007359.

B.O.P. serves on the Scientific Advisory Board of Genomatica Inc., a spin-off company from the University of California San Diego.

ADDENDUM IN PROOF

After this article was accepted, a genome-wide map of human RNA polymerase II preinitiation complexes made using chIP-chip and Nimblegen microarrays was published (T. H. Kim, L. O. Barrera, M. Zheng, C. Qu, M. A. Singer, T. A. Richmond, Y. Wu, R. D. Green, and B. Ren. Nature doi: 10.1038/nature03877 [29 June 2005]). A large number of false negatives were encountered in that study as well, which the authors attribute to lack of sensitivity in microarray detection of immunoprecipitated DNA.