The *Escherichia coli* RutR transcription factor binds at targets within genes as well as intergenic regions

Tomohiro Shimada¹, Akira Ishihama¹,², Stephen J. W. Busby³ and David C. Grainger³,*

¹Department of Frontier Bioscience and Micro-Nano Technology Research Centre, Hosei University, Koganei, Tokyo 184-8584, ²Nippon Institute for Biological Science, Ome, Tokyo 198-0024, Japan and ³School of Biosciences, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK

Received April 1, 2008; Revised May 6, 2008; Accepted May 11, 2008

ABSTRACT

The *Escherichia coli* RutR protein is the master regulator of genes involved in pyrimidine catabolism. Here we have used chromatin immunoprecipitation in combination with DNA microarrays to measure the binding of RutR across the chromosome of exponentially growing *E. coli* cells. Twenty RutR-binding targets were identified and analysis of these targets generated a DNA consensus logo for RutR binding. Complementary in vitro binding assays showed high-affinity RutR binding to 16 of the 20 targets, with the four low-affinity RutR targets lacking predicted key binding determinants. Surprisingly, most of the DNA targets for RutR are located within coding segments of the genome and appear to have little or no effect on transcript levels in the conditions tested. This contrasts sharply with other *E. coli* transcription factors whose binding sites are primarily located in intergenic regions. We suggest that either RutR has yet undiscovered function or that evolution has been slow to eliminate non-functional DNA sites for RutR because they do not have an adverse effect on cell fitness.

INTRODUCTION

A cohort of over 250 transcription factors controls gene expression in *Escherichia coli* in response to specific environmental cues. Most factors are operon specific and regulate the transcription of a small number of genes, while a small number of ‘global’ regulators coordinate transcription from a large number of promoters (1–3). Newly developed whole genome technologies now enable us to catalogue binding targets for each factor. These studies confirm that most regulators bind to intergenic DNA sequences near the 5′ end of a gene to regulate transcription.

The *E. coli rutABCDEFG* operon, that encodes genes for the catabolism of pyrimidines, is regulated by RutR, a TetR family factor whose DNA binding is modulated by uracil (4,5). RutR is transcribed divergently from *rutA* and, using genomic SELEX, Shimada and co-workers (5) identified six DNA targets for RutR, including the *rut-rutABCDEFG* intergenic region. Interestingly, this study reported that two of the targets are located within open reading frames (ORFs) and failed to detect any RutR-dependent modulation of transcription at one of the targets (*ygiF/glnE*). To investigate RutR further, here we have used chromatin immunoprecipitation (ChIP), in combination with DNA microarrays (ChIP-chip), to measure the chromosome-wide DNA-binding profile of RutR *in vivo*. We compared the binding profile of RutR in cells growing exponentially in media either with or without uracil, which inhibits DNA binding by RutR. Our study identifies 20 different binding targets for RutR. Surprisingly, 14 of these targets are located within genes and, for all but one of these targets, we were unable to measure any RutR-dependent effects on RNA levels.

MATERIALS AND METHODS

*E. coli* strains and oligonucleotides

Bacterial strains and synthetic oligodeoxynucleotides used in this work are listed in Supplementary Table 1. In all experiments we used *E. coli* strain BW25113 (6) or the ΔrutR derivative JW0998 (7). BW25113 expresses normal levels of RutR from a chromosomal copy of the *rutR* gene. Other than the *rutR* mutation, BW25113 and JW098 are isogenic. Cells were grown in M9 minimal medium supplemented with 0.4% glucose in either the presence or absence of 0.1 mM uracil. For experiments with exponentially growing cells, overnight cultures of *E. coli* strain BW25113 or JW0998 were diluted 1:100 into fresh medium either with or without uracil, and grown for ~4 hours to an OD₆₅₀ of 0.3–0.4.

ChIP and DNA microarray analysis

ChIP assays were used to measure the chromosome-wide DNA-binding profile of RutR in the presence and absence...
of uracil, using experimental protocols described in detail by Efromovich et al. (8). Assays were done either in tri-
plicate (−uracil) or in duplicate (+ uracil). Briefly, cultures of E. coli BW25113 and, as a control, JW0998 ΔrutR were
grown to mid-log phase at 37°C. Cells were then treated
with 1% formaldehyde and broken open by sonication
which also fragments cross-linked nucleoprotein. Cross-
l inked RutR–DNA complexes were immunoprecipitated
from cleared lysates of BW25113 using anti-RutR rabbit
polyclonal anti-serum, and parallel samples were isolated
from control JW0998 ΔrutR cells. Cross-links were then
reversed and immunoprecipitated DNA was purified.
DNA samples isolated from BW25113 cells and the control
ΔrutR cells were labelled with Cy5 and Cy3, respectively.
To identify segments of DNA specifically associated with
RutR, the two labelled samples were combined and
hybridised to a 22000 feature DNA microarray (Oxford
Gene Technology, Oxford, UK). For each probe, the Cy5/
Cy3 ratio was measured and this was plotted against the
corresponding position on the E. coli BW25113 chromo-
some, creating a profile of RutR binding (Figure 1). We
then selected ‘peaks’, formed by two or more consecutive
probes, with a Cy5/Cy3 ratio of >2.5. To increase the
stringency of our search, we discarded the small number of
peaks where the RutR-binding signal was not reduced at
least two-fold when uracil was added to cultures. The
centre of each peak is defined as the centre of the probe
within the peak that had the highest Cy5/Cy3 signal.

DNA sequence analysis

The RutR-binding motif was extracted from 500-bp DNA
sequences centred around the binding peaks using

Figure 1. Distribution of RutR binding across the E. coli chromosome. (A) The figure shows an overview of results from ChIP-chip experiments that measure the profile of RutR binding across the E. coli chromosome during exponential growth in the absence of added uracil. Binding signals (y-axis) are plotted against their location on the 4.64 Mbp E. coli chromosome (x-axis). The locations of selected signals are labelled. A complete list of the targets is presented in Table 1. Data shown in all panels are average values from replicate experiments. (B) The figure shows RutR binding to intergenic segments of the chromosome in either the absence (blue) or presence (red) of added uracil. (C) The figure shows RutR binding to coding segments of the chromosome in either the absence (blue) or presence (red) of added uracil.
BioProspector (http://ai.stanford.edu/~xsliu/BioProspector/). A DNA sequence logo describing the binding motif was then generated using WebLogo (http://weblogo.berkeley.edu/). For RutR, each of the 20 binding sites identified by ChIP-chip were aligned in PREDetector (9) and a position weight matrix (PWM) was generated to describe the information content of the binding site. Each target was then assigned a score depending on how it matched the PWM. The average score was used as a cut-off when searching genome sequences for RutR-binding sites. The same approach was used for the other transcription factors shown in Figure 4 except that we used the binding site alignments of Robison et al. (10) to generate the PWM.

RutR purification and electrophoretic mobility shift assays

His-tagged RutR was over-produced and purified exactly as described by Shimada et al. (5) using plasmid pYcdC and E. coli strain BL21(λDE3). Purified RutR was more than 95% pure as analysed by SDS PAGE. DNA fragments for EMSA experiments were generated by PCR amplification using the appropriate DNA primers and chromosomal DNA. EMSA experiments were generated by PCR amplification of RutR purification and electrophoretic mobility shift assays (EMSA). DNA fragments for labelling using [γ-32P]-ATP and polynucleotide kinase. DNA fragments were then incubated with purified RutR in buffer containing 20 mM Tris pH 7, 10 mM MgCl2, 100 μM EDTA, 120 mM KCl and 12.5 μg/ml herring sperm DNA. Reactions were loaded under tension onto a 5% polyacrylamide gel, run in 0.5x TBE at 160 V for 2–4 hours and analysed as described above.

RESULTS AND DISCUSSION

ChIP-chip analysis of RutR binding in mid-log phase E. coli

Our aim was to use ChIP to measure the binding of RutR across the chromosome of growing E. coli cells. Thus, strains BW25113 and the ΔrutR derivative JW0998 were grown aerobically, in M9 minimal medium supplemented with 0.4% glucose, to an OD650 of 0.3–0.4. Cells were then treated with formaldehyde, and cellular DNA was extracted and sonicated, yielding DNA fragments of ∼500–1000 bp. After immunoprecipitation with anti-RutR antibodies, DNA fragments from BW25113 or control JW0998ΔrutR cells were purified and labelled with Cy5 and Cy3, respectively, mixed and hybridised to the microarray. After washing and scanning, the Cy5/Cy3 signal intensity ratio was calculated for each probe. In parallel, the experiment was repeated using cells grown in the presence of 0.1 mM uracil. Complete data sets are shown in Supplementary Table 2. Figure 1A gives an overview of the profile for RutR binding. Most peaks for RutR binding are discrete, easily distinguishable from the background signal and sensitive to uracil (examples are shown in Figure 1B and C).

Identification and sequence analysis of RutR targets

To select peaks for RutR binding, a cut-off was applied to the dataset (see Materials and Methods section). A total of 77 probes passed this cut-off, corresponding to 20 separate peak locations (listed in Table 1). The targets we identified for RutR included four of the six RutR-binding targets reported previously (carA, rutA, hyi, and ygiF/glnE). Our failure to find all of the targets identified by Shimada et al. (5) may be due to differences in the in vivo and in vitro DNA-binding properties of RutR or the high false-negative rate of ChIP-chip analysis. Surprisingly, although RutR is bound to some targets in intergenic regions (Figure 1B) most targets are located within genes (Figure 1C).

To pinpoint the precise RutR-binding sequences, we used BioProspector (http://ai.stanford.edu/~xsliu/BioProspector/) to search for short, over-represented DNA sequences in 500-bp segments centred on each peak.

Table 1. RutR-binding sites identified by ChIP-chip analysis

<table>
<thead>
<tr>
<th>Peak centre</th>
<th>Sequence (5'-3')</th>
<th>Match to consensus</th>
<th>Position of site with respect to nearest start codon</th>
<th>ORF or intergenic</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. DNA/RNA related</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3211536</td>
<td>TTTGACTTCTGTGTCAA</td>
<td>13/14 + 804.5</td>
<td>ORF (purD)</td>
<td></td>
</tr>
<tr>
<td>3994092</td>
<td>TTTGACTGCGTCGTGCA</td>
<td>11/14 + 258.5</td>
<td>ORF (xerC)</td>
<td></td>
</tr>
<tr>
<td>B. Metabolism</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>29360</td>
<td>TTTGACCAATTTGGTCCA</td>
<td>13/14 + 284.5</td>
<td>intergenic (carA)</td>
<td></td>
</tr>
<tr>
<td>135994</td>
<td>CACACCAGTTGGTCAA</td>
<td>11/14 + 1667.5</td>
<td>ORF (ygiF)</td>
<td>intergenic (yabA)</td>
</tr>
<tr>
<td>331599</td>
<td>No consensus site</td>
<td>N/A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>553553</td>
<td>TTAATCTGTCGGTCTGG</td>
<td>9/14 + 312.5</td>
<td>ORF (bjiD)</td>
<td></td>
</tr>
<tr>
<td>1015645</td>
<td>TTTGACCAACAGGTGTCGA</td>
<td>12/14 + 117.5</td>
<td>ORF (jaa)</td>
<td>intergenic (rutA)</td>
</tr>
<tr>
<td>1073323</td>
<td>TTTGACAAAGCGGTGTCGA</td>
<td>11/14 + 114.5</td>
<td>intergenic (rutA)</td>
<td></td>
</tr>
<tr>
<td>2781517</td>
<td>TGGACACAAAGCACTGTC</td>
<td>9/14 + 373.5</td>
<td>ORF (yitQ)</td>
<td></td>
</tr>
<tr>
<td>3197797</td>
<td>CCAACACTTCGTCGTCGA</td>
<td>10/14 + 373.5</td>
<td>ORF (yigF)</td>
<td></td>
</tr>
<tr>
<td>3221219</td>
<td>TTTACACTTGTCAT</td>
<td>12/14 + 885.5</td>
<td>ORF (ehgA)</td>
<td>overlapping (yihX)</td>
</tr>
<tr>
<td>3578836</td>
<td>ATGGACATGATTTGCT</td>
<td>8/14 + 0.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C. Function unknown</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1272801</td>
<td>CTGACCAATGGTCGAC</td>
<td>11/14 + 69.5</td>
<td>intergenic (ydhM)</td>
<td></td>
</tr>
<tr>
<td>1724133</td>
<td>TGCAGCCAGCTGCCTTA</td>
<td>11/14 + 22.5</td>
<td>intergenic (ydhM)</td>
<td></td>
</tr>
<tr>
<td>1822567</td>
<td>TTTACACTTGTCGCCG</td>
<td>11/14 + 273.5</td>
<td>ORF (rev)</td>
<td></td>
</tr>
<tr>
<td>D. Drug resistance/sensitivity</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2371490</td>
<td>TTTGACACACCATTCC</td>
<td>10/14 + 9.5</td>
<td>ORF (pmrD)</td>
<td></td>
</tr>
<tr>
<td>E. Transport</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>624085</td>
<td>ATGACAAATTGTCAC</td>
<td>10/14 + 45.5</td>
<td>intergenic (fepB)</td>
<td></td>
</tr>
<tr>
<td>2510573</td>
<td>ATGGCATTGCTGGTGG</td>
<td>7/14 + 183.5</td>
<td>ORF (mntH)</td>
<td></td>
</tr>
<tr>
<td>4419024</td>
<td>TTGACCATCAGGTAAA</td>
<td>12/14 + 102.5</td>
<td>ORF (ihaB)</td>
<td></td>
</tr>
<tr>
<td>4603388</td>
<td>ATGGACATGGCGGCCAG</td>
<td>9/14 + 176.5</td>
<td>ORF (jhaF)</td>
<td></td>
</tr>
</tbody>
</table>

*Previously identified targets. RutR binding sites are shown in bold and bases that match the consensus RutR binding sequence are underlined.
This identified the 16-bp sequence motif shown in Figure 2, which matches the previously proposed consensus RutR-binding sequence, 5'-TTGACCAnnTGGT-CAAn-3', and predicts that positions 4 and 13 of the binding site are the most important for RutR binding. Following this, we located RutR-binding sites at 19 of the 20 targets identified by our ChIP-chip analysis (no motif was found at the yahA promoter). For the previously identified RutR targets, carA, rutA, hyi and ygiF/glnE (5), the motif that we identified corresponds exactly to the location of the known RutR-binding site (Table 1).

In vitro analysis of RutR binding

DNA fragments covering each of the 20 RutR targets were amplified, end-labelled and incubated with purified RutR protein in in vitro EMSA assays to measure the binding of RutR to each target (Figure 3). For 16 of the 20 targets, purified RutR clearly retarded the migration of purified DNA fragments (Figure 3A), while for four of the targets (fepB, yahA, mntH and fhuF), addition of purified RutR resulted in little or no retardation, indicating that RutR has a low affinity for these targets (Figure 3B). To understand this, we consulted the RutR-binding motifs shown in Table 1 and the RutR-binding site sequence logo shown in Figure 2. Interestingly, all 16 of the high-affinity RutR-binding sites matched the consensus at both of the strongly conserved positions 4 and 13. In contrast, the low-affinity fepB, mntH and fhuF targets all had non-consensus sequences at either position 4 or 13 and we could find no match to the RutR-binding site at the yahA target. In this latter case, RutR-binding specificity may be more relaxed in vivo or RutR might bind cooperatively with some other factor.

Figure 2. The RutR-binding site DNA sequence logo. The sequence logo was generated by aligning binding sites identified by ChIP-chip.

Figure 3. Binding of RutR to its DNA targets in vitro. The figure shows the results of electrophoretic mobility shift assays in which the binding of RutR (10, 25 or 50 nM) to purified end-labelled PCR products was measured in vitro. Free DNA fragments (F) and RutR–DNA complexes (C) are labelled. High-affinity targets are shown in (A) and low-affinity targets are shown in (B).
We thank Tomoya Baba for providing E. coli strain BW25113 and the rutR derivative JW0998. This work was supported by a Wellcome Trust programme grant awarded to S.J.W.B. Funding to pay the Open Access publication charges for this article was provided by Wellcome Trust.

Conflict of interest statement. None declared.
REFERENCES