Dual RpoH sigma factors and transcriptional plasticity in a symbiotic bacterium

RUNNING TITLE

Dual RpoH sigma factors in Sinorhizobium meliloti

Melanie J. Barnett, Alycia N. Bittner, Carol J. Toman, Valerie Oke and Sharon R. Long

Department of Biology, Stanford University, Stanford, CA 94305-5020; and Department of Biological Sciences, University of Pittsburgh, Pittsburgh, Pennsylvania 15260

Present address: Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030

Corresponding author: srl@stanford.edu
ABSTRACT

*Sinorhizobium meliloti* can live as a soil saprophyte and can engage in a nitrogen fixing symbiosis with plant roots. To succeed in such diverse environments, the bacteria must continually adjust gene expression. Transcriptional plasticity in eubacteria is often mediated by alternative sigma (σ) factors interacting with core RNA polymerase. The *S. meliloti* genome encodes 14 of these alternative σ factors, including two putative RpoH ("heat shock") σ factors. We used custom Affymetrix Symbiosis Chips to characterize the global transcriptional response of *S. meliloti rpoH1*, *rpoH2*, and *rpoH1 rpoH2* mutants during heat shock and stationary phase growth. Under these conditions, expression of over 300 genes is dependent on *rpoH1* and *rpoH2*. We mapped transcript start sites of 69 *rpoH*-dependent genes using 5' RACE (Rapid Amplification of cDNA Ends), which allowed us to determine putative RpoH1-dependent, RpoH2-dependent, and dual (RpoH1- and RpoH2-dependent) promoter consensus sequences that were each used to search the genome for other potential direct targets of RpoH. The inferred *S. meliloti* RpoH promoter consensuses share features of *E. coli* RpoH promoters, but lack extended -10 motifs.
INTRODUCTION

*Sinorhizobium meliloti* is a soil-dwelling α-proteobacterium that forms nitrogen-fixing root nodules on plants including *Medicago sativa* (alfalfa) and a model legume, *Medicago truncatula*. In the earliest stage of the symbiosis, the bacteria associate with plant roots and exchange chemical signals (13, 16). Subsequently, the actively dividing bacteria invade the root cortex and developing nodule via a plant-synthesized infection thread (28). Bacteria are released into the cytoplasm of nodule cells, where they terminally differentiate into nitrogen fixing bacteroids that convert dinitrogen into ammonia, which the plant can use as a nitrogen source for growth (15, 20, 37, 67).

Given this complicated lifestyle, *S. meliloti* must have phenotypic flexibility to adapt to a series of differing environments: unpredictable soil conditions, plant defense mechanisms, plant signals, and the internal plant milieu. Abiotic stresses may include extremes of pH, salinity, nutrient availability, and temperature, as well as toxic metals, reactive molecules, and other deleterious compounds (57). Effective adaptation usually includes major changes in gene expression; determining what, when, and how these changes occur in *S. meliloti* will increase our understanding of soil dynamics and of symbiotic nitrogen fixation (2).

Eubacterial transcription is mediated by RNA polymerase (RNAP), and the sigma (σ) factor subunit is required for promoter recognition and transcription initiation. All eubacterial genomes encode an essential housekeeping σ factor and most have at least one alternative σ factor. Housekeeping σ factors recognize a large set of promoters, while alternative σ factors recognize smaller groups of promoters for genes with a shared function. In general, σ factors recognize promoter motifs located approximately -35 and -10 nt upstream of the transcription start site (23). Within a given species, the sequence and spacing of these motifs differ depending
on which σ factor interacts with core RNAP. Since alternative σ factors compete with the
housekeeping σ factor for RNAP core enzyme, large shifts in gene expression can occur by
controlling expression, activity, and availability of alternative σ factors (49). A high number of
alternative σ factors in a species appears to correlate with a diverse lifestyle (23).

The *S. meliloti* genome encodes 14 alternative sigma factors: RpoN, essential for the
transcription of nitrogen fixation genes; 11 ECF or extracytoplasmic function-type σ factors
(RpoE1-10, FecI); and two RpoH/heat shock-type σ factors (RpoH1 and RpoH2; 19). Multiple
RpoH σ factors are common in α-proteobacterial genomes (11, 22, 33, 43). In *S. meliloti*,
RpoH1 and RpoH2 share 44% sequence identity and are 38% and 40% identical, respectively, to
the *E. coli* RpoH heat shock σ factor. In *E. coli*, the main role of RpoH is to maintain protein-
foling homeostasis under high temperature and other conditions that denature proteins, as well
as in normal conditions (24). *S. meliloti rpoH1* and *rpoH2* are partially functionally equivalent
to *E. coli rpoH*, as introduction of either gene on a plasmid restores viability of an *E. coli rpoH*
null mutant at 30°C, although not 37°C (47, 48).

Unlike wild type *S. meliloti*, *rpoH1* mutants are severely impaired for growth at 37°C,
fail to grow at 40°C, and are sensitive to acid pH, deoxycholate, sodium dodecyl sulfate, and
crystal violet (9, 40, 47, 48). In addition, *rpoH1* mutants have severe symbiotic defects: they
initiate nodule formation, invade plant roots, and are released into plant cells, but show poor
colonization and survival in nodule cells and do not fix nitrogen (40, 47). In contrast to *rpoH1*
mutation, *rpoH2* mutation has little effect on *S. meliloti* growth, stress adaptation or symbiosis
(40, 47, 48). However, the *rpoH1 rpoH2* double mutant has a more severe symbiotic phenotype
than an *rpoH1* mutant: nodules are rarely formed, and those that do form contain bacteria
carrying suppressor mutations (7).
S. meliloti rpoH1 and rpoH2 are expressed differently under both free-living and symbiotic conditions. During free-living growth, rpoH1 expression appears mostly constitutive (3, 12, 26, 36, 52), although its expression during exponential phase growth is higher in minimal medium compared to rich TY medium (3). While its expression increased with the onset of stationary phase growth in rich medium (47), other work showed either no change (52) or a decrease (8) during stationary phase growth in minimal medium. rpoH1 expression decreases during nitrogen starvation in a relA-dependent manner (31). rpoH1 is expressed strongly in M. sativa (47) and M. truncatula (3) nodules. In contrast, rpoH2 expression is only detectible in minimal medium, but not rich, after growth to stationary phase (47). rpoH2 expression is induced during heat shock in minimal medium (47), probably due to increased activity of RpoE2 (52); it also increases during osmotic stress (12). In M. sativa and M. truncatula nodules, rpoH2 is expressed at low levels (3, 47).

In E. coli, RpoH function is regulated mainly at the level of translation, protein stability, and protein activity (24). In S. meliloti, nothing is known about post-transcriptional regulation of RpoH1 or RpoH2, but since rpoH1 expression is largely constitutive, it is likely that post-transcriptional regulation plays a significant role.

To determine how these dual RpoH σ factors contribute to gene expression during S. meliloti stress adaptation, we identified putative RpoH1 and RpoH2 targets by global transcription profiling of rpoH1, rpoH2, and rpoH1 rpoH2 mutants under conditions of heat shock or stationary phase growth. We performed 5' RACE mapping on selected target genes to identify possible RpoH promoter elements. Our work shows that RpoH1 and RpoH2 directly or indirectly control hundreds of S. meliloti genes and that the putative promoters of many of these genes have overlapping, yet distinct features.
MATERIALS AND METHODS

Bacterial Strains and Culture Conditions. *S. meliloti* strains used in this study were Rm1021 (wild type), VO3128 (*rpoH1::aadA*), AB3 (*rpoH2::aacC1*), and AB9 (*rpoH1::aadA rpoH2::aacC1*) (7, 35, 47). Bacterial cultures were grown at 30°C in LB/MC medium (21) or M9 minimal medium containing 0.2% sucrose, 0.5 μg ml⁻¹ biotin, 1 mM MgSO₄, 0.25 mM CaCl₂, and 500 μg ml⁻¹ streptomycin.

For heat shock, cells were grown overnight in LB/MC medium, diluted to OD₅₉₅=0.05 the next day, and allowed to grow overnight. Growth of three replicates was staggered by one hour. Cultures were diluted to OD₅₉₅=0.05 in 65 ml LB/MC and allowed to grow to mid-exponential phase (OD₅₉₅=0.5-0.7). Each wild-type culture was split such that 30 ml remained at 30°C for 15 minutes as a control, and 30 ml was heat-shocked for 15 minutes at 42°C. Cells were harvested by mixing cultures with 1/10 volume of ice-cold stop solution (5% buffer-equilibrated phenol in ethanol) and centrifuging at 4°C. The supernatant was removed, and cells were frozen in liquid nitrogen and stored at -80°C.

For the stationary phase growth experiment, three replicates were started on the same day in identical medium. Overnight cultures were grown in LB/MC medium, diluted to OD₅₉₅ = 0.05 the next day, and allowed to grow overnight to ensure even growth. Cells (8 ml) were washed twice and diluted to OD₅₉₅=0.05 in 300 ml M9 minimal medium. Growth of the three replicates was staggered by 1.5 hours. Cultures were incubated with shaking for 48 hours, until late stationary phase. Cells were harvested in the same manner as described above for the heat shock experiments.

RNA purification, cDNA Synthesis, Labeling, and Hybridization. Cell pellets were resuspended in 1 mg ml⁻¹ lysozyme in TE and buffer RLT (Qiagen RNeasy Kit). Cells grown to
stationary phase were additionally lysed by bead-beating with 0.09-0.135 mm glass beads (Thomas Scientific) at 4°C with three 30-second pulses and a one-minute incubation on ice between each pulse. Total RNA was isolated as described, but with an RNeasy Midi Kit (3). This protocol includes both on-column and off-column DNase digests to remove all contaminating chromosomal DNA. Absence of chromosomal DNA was confirmed by PCR amplification using primers to an intergenic region, and RNA integrity was validated on a 1.2% agarose formaldehyde gel.

cDNA synthesis, labeling and hybridization to Affymetrix Symbiosis Chips were performed as described previously (3), except for the stationary phase experiment, where 12 µg (instead of 4 µg) of fragmented, labeled cDNA was hybridized to each chip.

**Affymetrix Data Analysis.** Design of the *S. meliloti/M. truncatula* dual-genome Symbiosis Chip was described previously (3); the *S. meliloti* sequences on the GeneChip array correspond to the original genome annotation reported in (19). The Symbiosis Chip also contains probe sets corresponding to *S. meliloti* intergenic regions (IGR) ≥ 150 nt and to ~10,000 *M. truncatula* EST sequences. As with other array platforms, our Symbiosis Chip measures mRNA abundance, of which both transcription and mRNA turnover are components. For brevity in this report, we use the term "expression," to include the sum of all factors affecting mRNA abundance. Data were processed using GeneChip Operating Software and Data Mining Tool (Affymetrix). Chips were scaled to a target signal intensity of 500 by using the global scaling option and each experimental array was compared with a baseline array. Thus, an experiment with three control and three experimental arrays yielded nine pairwise comparisons. We deemed an increase or decrease of average signal log ratio (SLR) ≥ 0.96 to be significant if either eight or nine of the nine pairwise comparisons were evaluated by the software as
significantly changed (P ≤ 0.05). The Affymetrix data described in this publication are available in the National Center for Biotechnology Information’s Gene Expression Omnibus (GEO; 14) and are accessible through GEO Series accession number GSE36186.

**Transcription start site determination.** 5’ RACE (Rapid Amplification of cDNA Ends) was performed on a subset of *rpoH*-dependent genes essentially as described (45, 51) with minor modifications. In brief, 14 μg of RNA isolated from heat-shocked or stationary phase cells (the same RNA samples used for Affymetrix Symbiosis Chips) was treated with TAP (tobacco acid pyrophosphatase; Epicentre Technologies), ligated to an RNA primer (5’-GAGGACUCGAGCUCAGUC-3’) with T4 RNA ligase (Epicentre Technologies), and reverse transcribed with SuperScript II or III reverse transcriptase (Invitrogen). For reverse transcription, a cocktail of gene specific primers was used (Supplemental Table 1). cDNA was amplified using a primer specific to the 5’ end of each transcript (RACE primer: 5’-GAGGACTCGAGCTCAGTC-3’) and a gene-specific primer (Supplemental Table 1).

Amplified fragments were separated by polyacrylamide gel electrophoresis (7.5%) and stained with ethidium bromide or SYBR Gold (Invitrogen). Fragments that displayed decreased intensity in *rpoH* cells compared to wild-type cells were excised and subjected to a second round of PCR amplification. PCR samples were purified with USB ExoSAP-IT (Affymetrix) or QIAquick columns (Qiagen) and sequenced. Altogether, we determined the transcription start sites of 69 *rpoH*-dependent genes that were chosen to represent a range of gene expression fold changes, as well as genes orthologous to those in the *E. coli* RpoH regulon (45). All but two of these genes were significantly decreased in expression by Affymetrix GeneChip analysis, in one or more of the *rpoH* mutants. Values for SMc02703 and SMc01280 were slightly below our cutoff, but we confirmed their *rpoH*-dependence by 5’ RACE mapping (Table 1).
Promoter consensus determination and in silico genome-wide predictions. To identify putative promoter consensuses specific for RpoH1 and/or RpoH2, sequences upstream of mapped transcription start sites of protein coding genes were sorted into four sets (Table 1) based on their expression pattern: 1) 23 genes dependent only on rpoH1, only during heat shock; 2) 14 genes dependent only on rpoH2, only in stationary phase; 3) 16 dual promoter genes dependent on rpoH1 in heat shock and rpoH2 in stationary phase; and 4) 16 genes that were rpoH-dependent, but with different expression patterns. Each of the first three sets was used as input for MEME (Multiple Em for Motif Ellicitation; 1). Initially, we used 20 nt search windows for each motif (-35 and -10) and then decreased the window size in subsequent iterations to obtain the final putative RpoH-dependent promoter consensus. Eight 5' RACE sequences could not be fit to their respective consensus; these may include genes whose expression is indirectly rpoH-dependent (SMa0136, SMc00030, SMc00048, SMc00814, SMc00969, SMc01329, SMc02863, and SMc04310 in Table 1). Hence, the final three consensus sequences incorporate data from 45 mapped promoters, distributed among the three sets as follows: 1) 20 genes for the RpoH1 consensus; 2) 11 genes for the RpoH2 consensus; and 3) 14 genes for the dual promoter consensus (Table 1).

To identify putative RpoH-dependent promoters in the S. meliloti genome, we used a matrix-based search method, RSA-tools-matrix-scan (59), to search upstream of each S. meliloti ORF with position-specific scoring matrices designed for each of the three consensus gene sets. To decrease noise and prevent bias in estimation of P-values, we prohibited overlap with upstream coding regions. Only those putative promoters with both a positive weight score and a P-value ≤ 1x10⁻⁴ were considered significant. After eliminating those genes whose expression was not decreased in one or more rpoH mutants in our Affymetrix experiments (i.e.,
those not in the group of "rpoH-dependent genes"), we obtained a set of 75 putative rpoH-dependent promoters representing 100 rpoH-dependent genes. Of the eight genes listed above whose RACE sequence could not be fit to a consensus, two were identified in the genome-wide search for putative RpoH1 promoters (SMc02863 and SMc04310).
RESULTS

The *S. meliloti* transcriptome changes dramatically in response to heat shock. To identify rpoH-dependent genes, we used custom Affymetrix Symbiosis Chips (3) for transcription profiling of wild type *S. meliloti* Rm1021 cells and strains mutated in rpoH1 (VO3128), rpoH2 (AB3), and rpoH1 rpoH2 (AB9). Since rpoH1 mutants are deficient in heat shock response, we compared these strains under heat shock conditions. Exponential phase cultures grown in LB/MC medium were subjected to heat shock for 15 minutes at 42°C (Materials and Methods). We chose this early time point to favor discovery of direct RpoH1 targets and because RT-qPCR experiments showed maximal induction of a known RpoH1 target, groESL5, at 15 minutes (data not shown). Wild type *S. meliloti* cells cultured for 15 minutes at 30°C served as a normal temperature control. We performed three biological replicates for each condition and defined changes in mRNA abundance of more than twofold as significant.

Although we were interested in rpoH1- or rpoH2-dependent genes, we started with a comparison of wild type heat shock-treated *S. meliloti* cells to those cultured at 30°C to look at the full transcriptional response to heat shock. Expression of 997 genes increased and expression of 1015 genes decreased; these changes represent about 40% of the genome. (Supplemental Data Set). As expected, many of the genes whose expression increased with heat shock have annotated functions in adaptation to heat and other stress responses. Expression of genes encoding enzymes for exopolysaccharide synthesis was also increased, as observed in other *S. meliloti* stress responses (26). The *S. meliloti* genome contains three replicons: a 3.65 Mb chromosome, and two megaplasmids (1.35 and 1.68 Mb) that carry genes necessary for symbiosis. Expression of most genes involved in replication and conjugal transfer of the megaplasmids (traA1CDG, traA2, repABC-1, repABC-2, virB1-5, virB7, virB9) increased with
heat shock. Genes with decreased expression in heat shock encode housekeeping, motility, chemotaxis, and pilus biogenesis functions. Expression of biosynthesis genes decreased, especially those for amino acid and nucleotide biosynthesis. An apparent decrease in expression during heat shock of the fixNOQP1 operon was due to higher than normal expression of this operon in the wild type, normal temperature control (data not shown). A few other genes previously shown to be induced with microoxia were similarly affected during this experimental trial, suggesting that it is an artifact. Genes whose expression increased with heat shock were more likely to lack a predicted function than those whose expression decreased (472 vs. 306). Expression of numerous genes encoding regulators changed during heat shock, including that of seven sigma factors: rpoE2, rpoE5, rpoE9, rpoH2, and rpoN expression increased, while rpoE1 and rpoE4 expression decreased. In sum, our results show that 15 minutes of heat shock results in extensive transcriptional remodeling.

RpoH1 plays a more significant role than RpoH2 in heat shock. We saw clear differences when wild type S. meliloti and each of the three rpoH mutant strains were subjected to heat shock. When the rpoH1 mutant was compared to wild type, 593 genes showed altered expression (Supplemental Data Set) of which 282 decreased and 311 increased. The genes whose expression decreased in the rpoH1 mutant (i.e., the set of rpoH1-dependent genes), likely include both direct as well as downstream, indirect targets. More than 1/3 of the genes whose expression was rpoH1-dependent during heat shock were not themselves significantly induced in wild type with heat shock; and about a dozen of these, including rpoE1, showed reduced expression in wild type. These may represent genes whose expression was already rpoH1-dependent at 30°C, as well as genes whose expression was induced by heat shock, but was not above the fold change cutoff in our experiments. Because not all rpoH1-dependent genes were
induced by heat in the wild type, we distinguished a subset of rpoH1-dependent, heat shock-induced genes. Figure 1 shows a Venn diagram of these relationships as they relate to a previous study of the role of rpoH1 in acid shock (described below). Among the genes that are rpoH1-dependent and heat shock-induced, annotated functions in stress response, chaperoning, macromolecular degradation, and fatty acid biosynthesis are overrepresented compared to the genome as a whole, whereas genes predicted to be involved in small molecule metabolism, including transport functions, are underrepresented (Supplemental Data Set). At least 10% of S. meliloti rpoH1-dependent genes are orthologous to members of the E. coli RpoH regulon; for example, clpA, clpB, clpP, clpX, creA, dnaK, dnaJ, ftsH, groESL, grpE, hflC, hflK, hslU, hslV, htpG, htpX, and lon (45, 63, 65, 69).

8% of rpoH1-dependent genes were previously reported to be induced during osmotic stress (12), suggesting that rpoH1 may play a role in osmotic stress resistance. It was also reported that rpoH1 is critical for response to acidic pH stress (9, see below). Although the S. meliloti global response to oxidative stress response has not been characterized, we found that about 5% of our rpoH1-dependent genes were orthologous to oxidative stress-induced genes of other bacteria: for example, methione sulfoxide reductase (msrB1, msrB2, msrA1, msrA3), thioredoxin (trxA), thioredoxin reductase (trxB), glutaredoxin (grxC), superoxide dismutase (sodB), glutathione-S-transferase (gst7, gst11), and glutathione reductase (gor).

In contrast to the rpoH1 mutant, no genes showed decreased expression in the rpoH2 mutant compared to wild type (Supplemental Data Set). In addition, a comparison of the rpoH1 rpoH2 double mutant to wild type gave essentially the same result as for the rpoH1 mutant. Moreover, directly comparing the rpoH1 strain to the rpoH1 rpoH2 strain failed to detect any
differences in expression. Thus, the impact of rpoH2 on gene expression during the early heat shock response in rich medium appears negligible, whereas rpoH1 clearly plays a major role.

**The rpoH1-dependent and rpoE2-dependent gene sets show little overlap.** *S. meliloti*

RpoE2 mediates a major transcriptional response to general stress and rpoE2 expression increases in response to various stresses (4, 52, 56). However, rpoE2 mutants show few phenotypic differences from wild type and are symbiotically normal (17, 18, 52). Sauviac et al. performed transcriptome analysis of the *S. meliloti* heat shock response in wild type and an rpoE2 mutant as part of their study on general stress response study (52). Although they used slightly different conditions to study heat shock, we still found that >75% of the changes they saw were represented in our heat shock data set. We also saw four-fold more (2012 vs. 451) expression changes, perhaps due to differences in growth and heat shock conditions or to increased detection sensitivity of our Affymetrix Symbiosis Chips.

We compared our rpoH1-dependent gene set to their rpoE2-dependent (by microarray and *in silico* prediction) gene set (52). 65 of their 89 rpoE2-dependent genes were induced in our wild type heat shock experiment, but we found that only five of these 65 were rpoH1-dependent and one was already shown to be only partially rpoE2-dependent (SMb21456; 52). Thus, RpoH1 and RpoE2 largely activate distinct sets of targets in response to heat shock.

**Many genes that were rpoH1-dependent in our study were also rpoH1-dependent during acid shock.** De Lucena et al. discovered that an rpoH1 mutant was severely impaired for growth in acidic medium (pH 5.75; 9). Their transcription profiling of wild type and rpoH1 mutant cells over a one-hour time course revealed that 68 of 124 genes whose expression was induced by acid shock also showed decreased expression in the rpoH1 mutant at one or more time points (9). Consolidating their results with our data, we conclude that 31 of these 68 genes
showed rpoH1-dependent induction in both heat and acid shock and 37 (29 + 8) genes showed 
rpoH1-dependence in acid shock only (Figure 1). Supplemental Table 2 lists rpoH-dependent, 
acid-induced genes, grouped by their rpoH1 expression pattern. We identified 259 genes whose 
expression was rpoH1-dependent in heat shock, but not acid shock: 174 (170 + 4) of these were 
significantly induced by heat shock in wild type and 85 were not (Figure 1). There may be 
additional genes whose expression is rpoH1-dependent during acid shock, but which are not 
themselves induced by acid shock in wild type; however, these were not identified by de Lucena 
(9). In summary, these comparisons demonstrated an overlapping transcriptional response to 
heat shock and acid shock in S. meliloti.

RpoH2 plays a larger role than RpoH1 during late stationary phase growth.

Previous work showed that expression of both rpoH1 and rpoH2 increases during stationary 
phase (47). To define the contributions of rpoH1 and rpoH2 in stationary phase, we compared 
transcription profiles of wild type S. meliloti Rm1021 and strains mutated in rpoH1 (VO3128), 
rpoH2 (AB3), and rpoH1 rpoH2 (AB9) grown to late stationary phase in minimal medium with 
sucrose as the carbon source and ammonium as the nitrogen source (Materials and Methods). 
We chose minimal medium because previous studies reported extremely low rpoH2 expression 
in rich medium (3, 47). Since cells from late stationary phase cultures had a low mRNA yield, 
we hybridized three times more cDNA than usual to each Affymetrix Symbiosis Chip to 
compensate (Materials and Methods).

Altogether, we identified 79 genes whose expression decreased and 12 genes whose 
expression increased in stationary phase in at least one of the mutant strains (Supplemental Data 
Set). Over half (44 genes) were dependent only on rpoH2, whereas 9 genes were dependent only 
on rpoH1 (Supplemental Figure 1). We were unable to determine rpoH1- vs. rpoH2-dependence
for 25 genes that showed decreased expression in the *rpoH1 rpoH2* double mutant, but not in either single mutant. Just one gene, SMc02900, showed decreased expression in all three mutants. 45 of the 79 genes whose expression appeared *rpoH*-dependent in stationary phase also demonstrated *rpoH1*-dependence during heat shock (see above).

About half of *rpoH2*-dependent genes lack a predicted function. Genes with predicted functions include some plausibly involved in stress responses: *ecnA*, orthologous to an *E. coli* antitoxin gene whose expression in *E. coli* is induced by osmotic stress in stationary phase (6); orthologs of *E. coli* *yagTSR*, a broad-spectrum, molybdopterin cytosine dinucleotide cofactor-containing aldehyde dehydrogenase involved in detoxification (44); *xseB*, predicted to encode a DNA repair enzyme; and SMa1158 encoding a universal stress protein that is induced in *S. meliloti* upon exposure to low oxygen and nitric oxide (5, 36). 43% of our *rpoH2*-dependent genes were induced, and none repressed, during osmotic stress in a previous study (12), suggesting that RpoH2 is important for response to osmotic stress.

Since *rpoH2* expression is induced in stationary phase, we compared our list of *rpoH2* genes with those reported in two recent studies to be induced in stationary phase (either carbon limitation serving to trigger stationary phase onset while cell densities remained low (OD$_{600}$=0.25; 52), or in cells harvested at higher cell densities (OD$_{600}$=1.2; 8). About 1/3 of our *rpoH2*-dependent genes overlapped with the combined stationary phase data, but represented less than 1% of those data. This suggests that either few stationary phase-induced genes are *rpoH2*-dependent, or that our experimental conditions failed to detect additional *rpoH2*-dependent genes.

**Expression of several *S. meliloti* sRNAs may be RpoH-dependent.** Recent studies have identified and characterized potential small, noncoding RNAs (sRNAs) in *S. meliloti*, most
of which are encoded in intergenic regions (IGR; 10, 53, 61, 62). Since our Affymetrix Symbiosis Chip can detect mRNA hybridization to IGR \( \geq 150 \) nt (3), we examined our data for IGR whose corresponding hybridization signal appeared \( \textit{rpoH} \)-dependent. The vast majority appear to represent 5'- and 3' untranslated regions (UTR) of \( \textit{rpoH} \)-dependent coding mRNAs. In addition, two IGR likely detect mRNAs encoding small ORFs that were not annotated in the original genome sequencing effort (SMc05011, SMc05020, Supplemental Table 3). We identified eight IGR (Supplemental Table 3) unlikely to represent \( \textit{rpoH} \)-dependent UTR based on the expression pattern of the adjacent genes; half of these overlap with sRNAs previously identified in a genome-wide survey (53). Of these potential \( \textit{rpoH} \)-dependent sRNAs, the trans-encoded sRNA SmelC781 is intriguing because it is adjacent to, and divergently transcribed from, \( \textit{rpoH2} \); it overlaps the predicted RpoE2-dependent promoter of \( \textit{rpoH2} \). SmelC781 expression was previously found to be induced by heat, pH, salt, and oxidative stress (53). In that study, the SmelC781 promoter (GTTGAC-N\(_{15}\).CCTAGAT) was characterized as RpoD-like (53), but we note its similarity to our RpoH-promoter consensus sequences (see below). Another sRNA candidate, SmelC456, is located upstream of \( \textit{dksA} \), a gene known to be important for environmental adaptation (31, 55), and also has a promoter (ATTGAA-N\(_{16}\).ACCAAAT) similar to our RpoH-promoter consensus sequences. We mapped the transcription start sites of both SmelC781 and SmelC456 and confirmed that their expression was indeed \( \textit{rpoH} \)-dependent (Supplemental Figure 2). SmelB130 and SmelC487 may be indirectly RpoH-regulated, as we did not identify a consensus match for these. Thus, our work identifies IGR with \( \textit{rpoH} \)-dependent expression, some of which are candidates for small regulatory RNAs involved in stress responses, and which will be objects of future research.
RpoH-dependent promoters share conserved features. As a first step in determining direct transcriptional targets of RpoH1 and RpoH2, we selected rpoH1- and rpoH2-dependent genes for 5' RACE mapping (Materials and Methods). We used RNA from heat-shocked wild type, rpoH1, rpoH2, and rpoH1 rpoH2 cells for 5' RACE, which also served as a qualitative confirmation of our Affymetrix data. As positive controls, we used groESL5, lon and clpB, which have previously been shown to be dependent on rpoH1 (40). As expected, RACE products for these control genes were less abundant in the rpoH1 and rpoH1 rpoH2 mutants, but not the rpoH2 mutant (data not shown). Moreover, the groESL5, lon and clpB transcription start sites inferred from our RACE mapping (Table 1) match those previously determined by primer extension (40). Expression of groEL4 was previously reported to be rpoH-independent (7, 40), but our Affymetrix results, which we confirmed by RACE mapping, show that its expression is dependent on rpoH1 in heat shock and rpoH2 in stationary phase, and that it is transcribed from a promoter upstream of the adjacent gene, SMc01759, encoding a protein of unknown function (Supplemental Figure 2A). Expression of dnaK was also previously reported to be rpoH-independent, although those data appear to demonstrate partial rpoH1-dependence (Figure 3d, 40). Our Affymetrix results showed a twofold decrease in dnaK expression in the rpoH1 and double mutants compared to wild type. In our RACE mapping, we observed a single, intensely staining band corresponding to the dnaK transcript, but were unable to confirm its rpoH1-dependence (Supplemental Figure 2A). We did not attempt to determine if other sigma factors act at this promoter, but note that in E. coli RpoH and RpoD initiate transcription at the same dnaK promoter in vitro (63).

We searched for consensus motifs upstream of our putative mapped transcription start sites using MEME as described in Materials and Methods. To identify differences in RpoH1-
versus RpoH2-dependent versus dual (RpoH1- and RpoH2-dependent) promoters of protein coding genes, we constructed three different consensus sequences (Figure 2). The RpoH1-specific consensus (CTTGAA-N_{15-16}-CCTATAT) comprises promoters of 20 genes dependent only on \textit{rpoH1} in heat shock (Table 1). The RpoH2-specific consensus (CTTGCC-N_{15-16}-CCTATCT) comprises promoters of 11 genes dependent only on \textit{rpoH2} during stationary phase growth (Table 1). Finally, the dual promoter consensus (CTTGAA-N_{15-16}-CCTATCT) comprises 14 promoters dependent on \textit{rpoH1} in heat shock and on \textit{rpoH2} in stationary phase (Table 1). The remaining RACE-mapped promoters that did not belong to any of the aforementioned groups (16 genes), or that could not be fit to their respective consensus sequences (8 genes), are also listed in Table 1. The vast majority of RpoH-dependent promoters have spacing of 15 or 16 nt between the end of the -35 motif and start of the -10 motif. RpoH1 consensus promoters were biased toward 16 nt spacers (70% of promoters), whereas RpoH2 consensus promoters favored 15 nt spacers (91% of promoters). Promoters responsive to both sigma factors showed a slight preference for 15 nt spacers (64% of promoters).

Our RpoH1 consensus sequence closely matches the consensus previously determined from three \textit{S. meliloti} \textit{rpoH1}-dependent promoters (\textit{groES5}, \textit{lon}, \textit{clpB}; CTTGAA-N_{17-CCANAT}; 40). All three of our consensus sequences share some similarity to the \textit{E. coli} RpoH-dependent promoter consensus (TTGAAA-N_{13-14}-CCCCATAT) identified by Nonaka et al. (45), with the RpoH1-consensus showing the most similarity. Discriminator length (distance between the end of the -10 motif and the mapped start site) for most \textit{S. meliloti} \textit{rpoH1}-dependent promoters was seven or eight nucleotides, as is the case in \textit{E. coli} (45). We searched 30 nt windows upstream of the -35 motif of our mapped \textit{S. meliloti} promoters for sequences similar to the A/T rich "UP element" enhancer found upstream of \textasciitilde40% of \textit{E. coli} RpoH-dependent promoters (45). While
about 1/3 had A/T rich stretches, we could not identify a consensus sequence among them (data not shown). Similarly, we failed to identify an extended -10 region such as that found in *E. coli* RpoH promoters. The spacing between the -35 and -10 motifs is identical between *E. coli* and *S. meliloti* once the different motif lengths are taken into account. Our RpoH-dependent promoter motifs closely match RpoH-dependent promoters from other α-proteobacterial species, such as *Rhodobacter sphaeroides* (46), *Caulobacter crescentus* (34), *Agrobacterium tumefaciens* (41), and *Bradyrhizobium japonicum* (39).

As with *E. coli*, the *S. meliloti* RpoH and RpoD -35 and -10 consensus sequences are similar: the *S. meliloti* consensus determined from 24 mapped putative RpoD-dependent promoters (32) is CTTGAC-N17,18-CTATAT, although we note that the -10 motif of the mapped RpoD-dependent promoters is much more variable than that of our mapped RpoH-dependent promoters (data not shown). Taking into account the differences in consensus motif lengths, an RpoD-dependent promoter with 17 nt spacing would have the same effective spacing as an RpoH-dependent promoter with 16 nt spacing. These similarities present challenges in predicting whether an *S. meliloti* promoter is RpoD- or RpoH-dependent; in fact, at least two promoters (*grpE* and *ibpA*) predicted to be RpoD-dependent by MacLellan et al. (32), are probably RpoH-dependent based on our array and 5' RACE mapping data. Likewise, we found two candidate RpoH-dependent promoters in our genome-wide search (see below) that were previously predicted to be RpoD-dependent (SMb20361 and SMc00043; 32).

To identify additional RpoH-dependent promoter candidates, we used each of the consensus sequences (Figure 2) to search the *S. meliloti* genome as described in Materials and Methods. This analysis predicted 75 candidate promoters, representing 100 genes that showed *rpoH1*- and/or *rpoH2*-dependent decreases in gene expression (Supplemental Table 4).
Consensus sequence matches did not strictly correlate with expression data. For example, of the 60 transcripts in Supplemental Table 4 that were identified as rpoH1-dependent during heat shock in our Affymetrix analysis, 46 were detected as significant using the RpoH1-specific promoter consensus as input; however, 29 of these were also detected with at least one of the other consensus sequences. This was not unexpected due to the similarity of the three consensus sequences (Figure 2).
DISCUSSION

In this study, we used transcription profiling to assess the roles of dual RpoH sigma factors in *S. meliloti* under two different stress conditions: heat shock and late stationary phase growth. We identified over 300 ORFs and putative sRNAs whose expression is *rpoH*-dependent. We mapped the transcription start sites of ~20% of these genes by 5' RACE and identified likely RpoH-dependent promoters. From these mapped promoters, we developed three different consensus sequences that we used to search the genome. Based on this analysis, we predict that up to half of the *rpoH*-dependent genes are direct targets of RpoH1 and/or RpoH2.

In wild-type *S. meliloti*, ~20% of genes show increased expression in heat shock compared to normal temperature. However, expression of only a quarter of these is dependent on either RpoH1 (this study) or the general stress response sigma factor, RpoE2 (52). Thus, it is possible that other alternative sigma factors play a role in the *S. meliloti* heat shock response. In support of this hypothesis, we found that, in addition to *rpoE2*, expression of *rpoE5*, *rpoH2*, *rpoE9*, and *rpoN* increased during heat shock. Moreover, it was previously reported that *rpoE5* and *rpoH2* are targets of RpoE2 (52).

Beyond the use of sigma factors, previous research suggests that regulatory mechanisms such as CIRCE/HrcA and RNA thermometers are important in *α*-proteobacterial response to heat shock (42, 64). HrcA repressor-binding to a cis-linked CIRCE element mediates negative regulation of bacterial heat shock genes (42). For instance, in *S. meliloti*, expression of *groESL1* and *groESL2* is not RpoH-dependent, but may be regulated by HrcA because upstream CIRCE elements are present (7). In *Bradyrhizobium japonicum*, *hrcA* expression itself is RpoH-dependent (38). Although *S. meliloti*, *hrcA* expression increased ~1.4 fold during heat shock, its expression was RpoH-independent in our experiments (Supplemental Data Set). Post-
transcriptional regulation of heat shock gene expression is mediated by elements such as RNA thermometers, possibly via high temperature-induced unmasking of the Shine-Dalgarno sequence required for translation (64). A given gene may have multiple regulatory inputs; for example, two RpoH targets identified in our study, which encode small heat shock proteins (SMB21295 & \textit{ibpA}), were previously predicted to possess upstream RNA thermometer elements (64). We infer that regulation of the \textit{S. meliloti} heat shock response is complex.

The overall question remains, what are the roles of the two RpoHs in \textit{S. meliloti}? All α-proteobacterial genomes sequenced to date contain at least one predicted \textit{rpoH} gene; most contain two, and some contain more. Previous work in \textit{S. meliloti} suggested largely distinct, yet partly overlapping roles for each RpoH (40, 47, 48). Transcription of \textit{rpoH1} occurs constitutively during exponential phase growth in rich medium and increases upon entry into stationary phase (47), but nothing is known about how RpoH1 activity is regulated. Transcription of \textit{rpoH2} was reported to be entirely dependent on RpoE2, suggesting that RpoH2 acts in the general stress response circuit (52). In addition, we observed increased \textit{rpoH2} expression in the \textit{rpoH2} and double mutants in both heat shock and stationary phase growth. This raises the possibility that RpoH2 either directly or indirectly regulates its own expression. An important caveat is that the insertion mutation in the middle of the \textit{rpoH2} gene might alter the stability of the resulting \textit{rpoH2} transcript or its detection by our Affymetrix Chip. Hybridization data corresponding to the region upstream of \textit{rpoH2} shows that, while abundance of the putative \textit{rpoH2} 5' untranslated region increased in heat shock as expected, it was unchanged in the \textit{rpoH2} and double mutants. Thus, additional experiments are necessary to determine whether RpoH2 negatively regulates its own expression.
With respect to heat shock response, we show that \textit{rpoH1} appears to be the main player in the early transcriptional response to heat shock in rich medium, with no detectable contribution from \textit{rpoH2}. On the other hand, with respect to late stationary phase growth in minimal medium, \textit{rpoH2} appears to play a larger role than \textit{rpoH1}. To reconcile our findings that \textit{rpoH2} expression increases during heat shock, yet \textit{rpoH2} itself appears to play a minimal role in heat shock gene expression, we suggest that RpoH2 was not sufficiently active to initiate transcription at RpoH2-dependent promoters at 15 minutes of heat shock in rich medium. We speculate that higher levels of \textit{rpoH2} transcripts and/or appropriate post-transcriptional regulation are required for full RpoH2 activity and that this serves to fine-tune the response depending on the particular stressor(s) encountered. Similarly, even though we found that \textit{rpoH1} mRNA is abundant, and that RpoH1 contributes to transcription during late stationary phase growth, RpoH1 may not be active enough to affect the same magnitude of change in gene expression as it does during heat shock.

We note that the apparent differences between \textit{rpoH1} and \textit{rpoH2} contributions to gene expression in heat shock versus late stationary phase growth may be due to other factors. For example, cells were grown in rich medium for the heat shock experiment and in minimal medium for the late stationary phase growth experiment. Osmoprotective compounds such as glycine betaine and proline are more abundant in rich (LB) medium than in minimal medium (27, 68). Therefore, if RpoH2 responds more readily to osmolality than it does to heat, it could explain the minimal contribution of \textit{rpoH2} to heat shock in rich medium. Indeed, we find that >40% of \textit{rpoH2}-dependent genes were previously shown to be induced during osmotic stress (12).
Our work points to an additional component in the *S. meliloti rpoH1/H2* regulatory circuits: small, noncoding RNAs (sRNAs). We identified eight different intergenic regions with differential expression in wild type versus one or more *rpoH* mutants (Supplemental Table 3). We used 5' RACE mapping to confirm the transcription start sites and *rpoH*-dependence for two of these, SmelC781 and SmelC456 (Supplemental Figure 2). Both of these were classified as trans-encoded sRNAs in a genome-wide survey (53). Most trans-encoded sRNAs interact with their target mRNAs via short, imperfect base-pairing and may control gene expression by a variety of mechanisms including activation or repression of translation, degradation or stabilization of mRNA, and modification of protein activity (58). Because of the short, imperfect nature of sRNA-mRNA interactions it is difficult to identify potential targets using bioinformatic approaches alone (58). Based on the *rpoH1*- and *rpoH2*-dependent expression of SmelC781 and its proximity to *rpoH2*, we speculate that SmelC781 affects expression of unknown targets involved in response to heat shock, stationary phase growth, and other stresses. The SmelC781 transcript maps to nt 3939507-3539650 of the chromosome (53), overlapping the predicted RpoE2-binding site of the *rpoH2* promoter (52). Thus, even if SmelC781 does not affect post-transcriptional regulation of *rpoH2*, its transcription may interfere with that of *rpoH2*. The putative trans-encoded sRNA SmelC456 lies immediately upstream of *dksA*; therefore, it is tempting to speculate that SmelC456 might regulate DksA expression or function. In *E. coli*, the transcription factor DksA regulates transcription initiation in concert with the nucleotide (p)pGpp, which is produced by the synthetase RelA during stress (49). Recent studies show that (p)pGpp allows many alternative sigma factors, including *E. coli* RpoH, to compete more effectively for RNAP binding (49). In *S. meliloti*, *relA* is required for root nodule formation (66), and *dksA* appears to be required for most *relA*-dependent gene regulation (31).
Interestingly, rpoH1 expression showed a relA-dependent, dksA-independent decrease during nitrogen starvation (31), a mechanism that could enhance RNAP binding to other stress response sigma factors, such as RpoE2. The role, if any, of Smel456 in the S. meliloti relA, dksA, and rpoH regulatory circuits remains to be elucidated.

In other α-proteobacteria possessing dual rpoH genes there is a division of labor. Notably, in the closely related species, Rhizobium etli, the rpoH1 and rpoH2 orthologs contributed to the oxidative stress response, while rpoH2 appeared to be the main player in the response to osmotic stress (33). R. etli rpoH1 mutants also have similar symbiotic defects as S. meliloti rpoH1 mutants (33). In Rhodobacter sphaeroides, rpoH1 is the main σ factor for responding to heat shock and rpoH2 controls the response to singlet oxygen stress (22, 46). In the intracellular pathogen Brucella melitensis, rpoH2 (orthologous to S. meliloti rpoH1) is important for response to heat shock, cold shock, and hydrogen peroxide, expression of virulence factors, and invasion and survival in mammalian cells, while rpoH1 may be important for full virulence (11). It appears that for S. meliloti and other α-proteobacteria, multiple rpoH genes provide flexibility in adapting to diverse environments.

We are especially interested in understanding why S. meliloti rpoH mutants are symbiotically impaired. De Lucena et al. suggested that decreased resistance to acid pH or oxidative stress might explain the non-fixing phenotype of rpoH1 mutants (9). Adding support to this hypothesis is our prediction that at least 5% of rpoH1-dependent genes are involved in response to oxidative stress. Direct studies of RpoH function during symbiosis are difficult given that rpoH1 rpoH2 double mutants do not form nodules and rpoH1 mutants senesce soon after their release into plant cells (7, 40). Comparative studies can provide some insights: a comparison of rpoH-dependent genes from this work and de Lucena et al. (9) to other published
data (3, 5, 8) reveals that 14% of rpoH-dependent genes are induced in nodule bacteria. Most of these genes lack a predicted function, but others are predicted to encode transcriptional regulators, small heat shock proteins, glutaredoxin, an outer membrane protein, and proteins involved in DNA repair, protein turnover, stress response, and small molecule metabolism. It is therefore possible that the primary function of RpoH1 and RpoH2 during symbiosis is to transcribe stress response genes either to repair cellular damage or for production of new proteins during symbiosis. However, the RpoH sigma factors may also be required for symbiosis-specific functions. While the RpoH regulons in several pathogens overlap with the E. coli RpoH regulon, they also include targets not found in E. coli, some of which might be involved in pathogenesis (25, 54). Here we have identified many RpoH-dependent genes that do not have obvious E. coli homologs; these could play a role in symbiosis. Targeted study of individual rpoH-dependent, nodule-induced genes may help elucidate the role of RpoH in the S. meliloti symbiosis and be generalizable to other α-proteobacteria that invade eukaryotic hosts.

The ability to recognize a broader range of promoters may be an important feature of S. meliloti RpoH1 and RpoH2. Our RpoH promoter consensus sequences are similar to RpoH promoters previously identified in S. meliloti and other α-proteobacteria (34, 39-41, 46) and to the -35 and -10 core hexamer motifs of the E. coli RpoH promoter consensus (45). However, we found that S. meliloti RpoH promoters lack the two upstream (CC) nucleotides that make up the extended -10 motif of E. coli RpoH promoters (45). This is an expected finding given that, like all known α-proteobacterial RpoH proteins, S. meliloti RpoH lacks the conserved K130 residue that recognizes the extended -10 region in E. coli (29). Further, it was demonstrated that RpoH from the α-proteobacterium Caulobacter crescentus recognizes E. coli promoters that lack the extended -10 motif (29). It has been postulated that by increasing promoter stringency, the
extended -10 motif allows for an efficient response to a restricted set of promoters, and that lack
of an extended motif may allow α-proteobacterial RpoH to transcribe larger sets of genes
involved in a variety of processes (30). This premise is supported by previous work showing
that, in addition to heat shock, *S. meliloti* rpoH1 is required for response to acid shock and
survival in root nodules (9, 40, 47), and by our current study defining a large set of rpoH-
dependent genes.

In *E.coli*, A/T-rich UP elements located upstream of the -35 motif interact with the α
subunits of RNAP to increase promoter strength (29, 50). Although about 1/3 of our mapped
RpoH-dependent promoters contained A/T-rich regions upstream of the -35 motif, we could not
define a consensus sequence among them using MEME (data not shown). However, recent work
modeling full-length *E.coli* RpoE promoters demonstrated that UP elements do not show
position-specific sequence conservation; instead, the number of overlapping 3 nt A- and T-tracts
in the -64 to -35 region is the best predictor of UP element strength (50). Several of our mapped
promoters contain overlapping A- and T-tracts; however, functional tests are necessary to
determine if the *E. coli* UP element model is applicable to *S. meliloti* RpoH promoters.

That RpoH1 and RpoH2 have partly overlapping, yet distinct roles is supported by our
consensus promoter analyses. The consensus sequences for RpoH1, RpoH2, and dual (RpoH1
and RpoH2) promoters share similar core motif features, -35 TTG and -10 CTANNT (Figure 2).
However, we found that the TTG consensus motif was more strongly conserved in RpoH1
promoters than in RpoH2 or dual promoters (Figure 2). In addition, unlike the RpoH1 and dual
promoter consensuses, the -35 motif of the RpoH2 consensus is biased toward terminal C
nucleotides (CTTGA vs. CTTGCC). This is similar to *Rhodobacter sphaeroides*, where a
consensus of seven -35 motifs specific for RpoH2 (CTTGCC) was distinguishable from a
consensus of eight -35 motifs (CTTGAN) recognized by both RpoH1 and RpoH2 (46). The -10 motif also shows differences among our three consensus sequences, with the 3rd position T being more conserved in RpoH2 and dual promoters than in RpoH1 promoters (Figure 2). Perhaps the most striking difference between the three sets of *S. meliloti* promoters is their spacer length distributions. Most RpoH1-specific promoters have 16 nt spacers (70% of total), whereas most RpoH2-specific promoters have 15 nt spacers (91% of total). Spacer length of dual promoters is more evenly distributed (15 nt, 64%; 16 nt, 29%). If this trend is generalizable to the full regulon of *S. meliloti* RpoH target genes, it could have large effects on promoter selectivity. For example, in *E. coli*, the core -35 and -10 motifs of RpoD and RpoS promoters are extremely similar, yet the corresponding RpoD and RpoS RNAP holoenzymes have distinct regulons and roles in vivo (60). In this instance, a major factor in RpoD versus RpoS promoter selectivity is spacer length: RpoD holoenzyme strongly prefers promoters with 17 nt spacing, while RpoS recognizes promoters that deviate ±1-2 nucleotides from this length (60). Thus, our work suggests that differences in core promoter motifs and spacer lengths play a role in RpoH1 and RpoH2 promoter selectivity.

In sum, it is likely that for species encoding multiple RpoHs, promoter selectivity is controlled in various ways from expression of the *rpoH* genes themselves to RpoH selectivity for specific promoters. Our work here provides a foundation for further study of RpoH-RNAP-promoter interactions, regulation of *rpoH* expression, control of RpoH activity, and characterization of RpoH target genes, some of which may be important for symbiosis.
ACKNOWLEDGEMENTS

M.J.B. was supported by the Hoover Circle fund and National Institutes of Health grant #GM093628 to S.R.L. A.N.B. was supported by an Andrew Mellon Predoctoral Fellowship from the University of Pittsburgh and NRI Competitive Grants Program/USDA award 2001-35319-10902 to V.O. We thank Robert Fisher and Claus Lang for critically reading the manuscript, are grateful to Jamil Afza and Glen Barnett for assistance with data extraction and software, and appreciate the constructive suggestions of the anonymous reviewers.
REFERENCES


represents the first molybdopterin cytosine dinucleotide cofactor containing molybdo-
flavoenzyme from *Escherichia coli*. FEBS J. **276**:2762-2774.

Regulon and promoter analysis of the *E. coli* heat-shock factor, σ^{32}, reveals a 

sigma factor regulons in the response to singlet oxygen in *Rhodobacter sphaeroides*. J. 

Identification of the heat-shock sigma factor RpoH and a second RpoH-like protein in 
*Sinorhizobium meliloti*. Microbiology **147**:2399-2408.

responsible for the expression of heat shock protein genes in *Sinorhizobium meliloti*. Mol. 

49. Österberg, S., T. Del Peso-Santos, and V. Shingler. 2011. Regulation of alternative 

50. Rhodius, V. A., V. K. Mutalik, and C. A. Gross. 2012. Predicting the strength of UP-
elements and full-length *E. coli* σ^{E} promoters. Nucleic Acids Res. **40**:2907-2924.


sigma factor acts as a general stress response regulator in *Sinorhizobium meliloti*. J. 
Bacteriol. **189**:4204-4216.
survey of sRNAs in the symbiotic nitrogen-fixing alpha-proteobacterium *Sinorhizobium*
meliloti. BMC Genomics 11:245.


The third pillar of bacterial signal transduction: classification of the extracytoplasmic

Washington, DC.


RSAT to scan genome sequences for transcription factor binding sites and cis-regulatory

60. Typas, A., G. Becker, and R. Hengge. 2007. The molecular basis of selective promoter

chromosomal alpha-proteobacterial small RNAs by comparative genome analysis and
detection in *Sinorhizobium meliloti* strain 1021. BMC Genomics 8:467.


Figure 1. Intersection of rpoH1-dependent genes with heat shock- and acid shock-induced genes. Genes induced by heat shock were identified in this study, and genes induced by acid shock were determined by de Lucena et al. (9). 17 genes whose expression was significantly decreased in the rpoH1 rpoH2 double mutant during heat shock, and whose expression was also decreased in the rpoH1 mutant, but with a SLR of > -0.96, are also shown in this diagram. Genes whose expression was rpoH1-dependent during acid shock, but which are not themselves induced by acid shock in wild type, were not identified by de Lucena; therefore, this diagram shows only 85 genes that were rpoH1-dependent in heat shock, but not themselves induced by heat or acid shock. Not shown on the diagram are 12 genes whose expression increased with both heat and acid shock treatments, but was rpoH1-dependent in only one of these treatments: 4 of these were rpoH1-dependent during heat shock and 8 were rpoH1-dependent during acid shock.

Figure 2. S. meliloti rpoH-dependent promoter consensus motifs. Colored sequence logos for gene sets 1, 2, and 3 of Table 1 were generated using MEME as described in Materials and Methods. The total information content of each logo is given in bits; the height of each nucleotide in the logo represents the positional probability of that nucleotide multiplied by the information content of the logo. -35 consensus motifs are displayed on the left side of the figure, and -10 consensus motifs on the right. Non-conserved spacer regions of 15-16 nt between the motifs are not shown. 1) RpoH1-specific promoter consensus motifs (20 genes), 2) RpoH2-specific promoter consensus motifs (11 genes), and 3) Dual promoter (RpoH1 and RpoH2) consensus motifs (14 genes).
<table>
<thead>
<tr>
<th>Unique Identifier</th>
<th>Gene Description</th>
<th>Avg. SLR$^c$</th>
<th>Dist.$^c$</th>
<th>Sequence$^d$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SMa1894</strong></td>
<td>msrB2 Peptide methionine sulfoxide reductase</td>
<td>-2.9 H1, hs</td>
<td>26</td>
<td>AATTGGATATCTCGAAGGACGCGCTCCCATATTTCAGGCA</td>
</tr>
<tr>
<td><strong>SMa1896</strong></td>
<td>msrA3 Peptide methionine sulfoxide reductase</td>
<td>-2.4 H1, hs</td>
<td>n/a</td>
<td></td>
</tr>
<tr>
<td><strong>SMb20708</strong></td>
<td>Methylated-DNA-protein-Cys methyltransferase</td>
<td>-3.0 H1, hs</td>
<td>68</td>
<td>GATTGAATTTGAGGCCGTCGGACCGCCTAG</td>
</tr>
<tr>
<td><strong>SMb20707</strong></td>
<td>cyaG2 Adenylate cyclase</td>
<td>-2.7 H1, hs</td>
<td>n/a</td>
<td></td>
</tr>
<tr>
<td><strong>SMb22023</strong></td>
<td>groES5 60 kD chaperonin</td>
<td>Probe set not on chip</td>
<td>96</td>
<td>CTGGGATATGCTTTTTGACCCCAACGGATCATCGG</td>
</tr>
<tr>
<td><strong>SMb21566</strong></td>
<td>groEL5 10 kD chaperonin</td>
<td>-3.5 H1, hs</td>
<td>n/a</td>
<td></td>
</tr>
<tr>
<td><strong>SMc00109</strong></td>
<td>Transcription regulator</td>
<td>-1.1 H1, hs</td>
<td>70^*</td>
<td>CTGGATGAGGCCGATCCTGCAATTAACCTCTA</td>
</tr>
<tr>
<td><strong>SMc00117</strong></td>
<td>msrB1 Peptide methionine sulfoxide reductase</td>
<td>-2.5 H1, hs</td>
<td>18/20</td>
<td>TTGTTTGCGGACCGATCCTGCAATTTATATCCTGCA</td>
</tr>
<tr>
<td><strong>SMc00467</strong></td>
<td>Conserved hypothetical protein</td>
<td>-1.7 H1, hs</td>
<td>82</td>
<td>CTGGGCTCTGACGGGAAACATTCTATTCTCCG</td>
</tr>
<tr>
<td><strong>SMc00466</strong></td>
<td>Conserved hypothetical protein</td>
<td>-1.5 H1, hs</td>
<td>n/a</td>
<td></td>
</tr>
<tr>
<td><strong>SMc01042</strong></td>
<td>grpE Heat shock protein</td>
<td>-1.3 H1, hs</td>
<td>89</td>
<td>CTGGATTGCCGCGCCGGCAAGCTCATTGCGCTTCC</td>
</tr>
<tr>
<td><strong>SMc01224</strong></td>
<td>trxB Thioredoxin reductase</td>
<td>-3.0 H1, hs</td>
<td>66</td>
<td>CTGGGATGCGCAGCGGATGATTCAGGCA</td>
</tr>
<tr>
<td><strong>SMc01256</strong></td>
<td>sda L-serine dehydratase</td>
<td>-2.1 H1, hs</td>
<td>106</td>
<td>CTGGGATATGCTTTTTGACCCCAACGGATCATCGG</td>
</tr>
<tr>
<td><strong>SMc01441</strong></td>
<td>hflK Membrane bound protease</td>
<td>-1.3 H1, hs</td>
<td>65</td>
<td>CTGGGATGAGGCCGATCCTTAAACCTCTGCA</td>
</tr>
<tr>
<td><strong>SMc01440</strong></td>
<td>hflC Membrane bound protease</td>
<td>-1.3 H1, hs</td>
<td>n/a</td>
<td></td>
</tr>
<tr>
<td><strong>SMc01465</strong></td>
<td>creA Conserved hypothetical protein</td>
<td>-4.4 H1, hs</td>
<td>20</td>
<td>CTGGGATGAGGCCGATCCTTAAACCTCTGCA</td>
</tr>
<tr>
<td>Accession</td>
<td>Gene</td>
<td>Function</td>
<td>H1 Change</td>
<td>Hs Change</td>
</tr>
<tr>
<td>-----------</td>
<td>----------</td>
<td>-----------------------------------</td>
<td>-----------</td>
<td>-----------</td>
</tr>
<tr>
<td>SMc01905</td>
<td>lon</td>
<td>ATP-dependent protease</td>
<td>-1.5</td>
<td>hs 72</td>
</tr>
<tr>
<td>SMc02110</td>
<td>clpS1</td>
<td>Elongation factor TS</td>
<td>-1.4</td>
<td>hs 52</td>
</tr>
<tr>
<td>SMc02109</td>
<td>clpA</td>
<td>ATP-dependent protease</td>
<td>-1.5</td>
<td>n/a</td>
</tr>
<tr>
<td>SMc02380</td>
<td></td>
<td>Oxidoreductase</td>
<td>-2.9</td>
<td>hs 23</td>
</tr>
<tr>
<td>SMc02720</td>
<td>clpP2</td>
<td>ATP-dependent protease</td>
<td>-1.4</td>
<td>hs 24</td>
</tr>
<tr>
<td>SMc02882</td>
<td></td>
<td>Conserved hypothetical protein</td>
<td>-4.5</td>
<td>hs 28</td>
</tr>
<tr>
<td>SMc02886</td>
<td></td>
<td>Conserved hypothetical protein</td>
<td>-2.4</td>
<td>hs 23</td>
</tr>
<tr>
<td>SMc02885</td>
<td>mscA1</td>
<td>Peptide methionine sulfoxide</td>
<td>-2.2</td>
<td>n/a</td>
</tr>
<tr>
<td>SMc03152</td>
<td></td>
<td>Conserved hypothetical protein</td>
<td>-4.0</td>
<td>hs 68</td>
</tr>
<tr>
<td>SMc03801</td>
<td></td>
<td>Thioredoxin</td>
<td>-1.9</td>
<td>hs 41</td>
</tr>
<tr>
<td>SMc04403</td>
<td>dcp</td>
<td>Peptidyl-dipeptidase</td>
<td>-2.8</td>
<td>hs 53</td>
</tr>
</tbody>
</table>

2. *RpoH2*-dependent in stationary phase, used to determine the *RpoH2* promoter consensus sequence

<table>
<thead>
<tr>
<th>Accession</th>
<th>Gene</th>
<th>Function</th>
<th>H2 Change</th>
<th>Sp Change</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>SMa1158</td>
<td></td>
<td>Conserved hypothetical protein,</td>
<td>-1.6</td>
<td>sp 34</td>
<td>ATTTCGTTTTTTATATTTGGCCTGCTATCTGCTTGGTA</td>
</tr>
<tr>
<td>SMa1364</td>
<td></td>
<td>UspA-like</td>
<td>-2.4</td>
<td>sp 61</td>
<td></td>
</tr>
<tr>
<td>SMa2301</td>
<td></td>
<td>ABC transporter, periplasmic</td>
<td>-4.1</td>
<td>sp 0*</td>
<td></td>
</tr>
<tr>
<td>SMa2349</td>
<td></td>
<td>solute-binding protein</td>
<td>-3.8</td>
<td>sp 187</td>
<td></td>
</tr>
<tr>
<td>SMa2351</td>
<td></td>
<td>Diguanylate cyclase</td>
<td>-2.7</td>
<td>sp n/a</td>
<td></td>
</tr>
<tr>
<td>SMa2353</td>
<td></td>
<td>Oxidoreductase, YagT-like</td>
<td>-2.5</td>
<td>sp n/a</td>
<td></td>
</tr>
<tr>
<td>SMc01354</td>
<td></td>
<td>Oxidoreductase, YagR-like</td>
<td>-1.0</td>
<td>sp 67</td>
<td></td>
</tr>
<tr>
<td>SMc01723</td>
<td></td>
<td>Conserved hypothetical protein</td>
<td>-3.9</td>
<td>sp 39</td>
<td></td>
</tr>
<tr>
<td>SMc01960</td>
<td></td>
<td>Oxidoreductase</td>
<td>-1.8</td>
<td>sp 42</td>
<td></td>
</tr>
<tr>
<td>SMc02703</td>
<td></td>
<td>Conserved hypothetical protein</td>
<td>-0.93</td>
<td>sp 58</td>
<td></td>
</tr>
<tr>
<td>Gene ID</td>
<td>Protein Description</td>
<td>H1 Expression</td>
<td>H2 Expression</td>
<td>Sp or hs</td>
<td></td>
</tr>
<tr>
<td>------------</td>
<td>------------------------------------</td>
<td>---------------</td>
<td>---------------</td>
<td>----------</td>
<td></td>
</tr>
<tr>
<td>SMc04146</td>
<td>Phosphoketolase</td>
<td>-3.2 H2, sp</td>
<td>58</td>
<td>ATTGCCCTGGCGCCGACCCACCGCATAGATTCTGCAGG</td>
<td></td>
</tr>
<tr>
<td>SMc04181</td>
<td>Hypothetical protein</td>
<td>-1.4 H2, sp</td>
<td>59</td>
<td>ATGTTACGCCTCTACGAGACTATATATTTACATC</td>
<td></td>
</tr>
<tr>
<td>SMc05020 ecnA</td>
<td>Entericidin A</td>
<td>-2.2 H2, sp</td>
<td>25</td>
<td>ATGATTTTCGCAACGACATACGATTTTTCTGAA</td>
<td></td>
</tr>
<tr>
<td>SMa2061</td>
<td>Conserved hypothetical protein</td>
<td>-1.8 H1, hs; -3.0 H2, sp 26</td>
<td>26</td>
<td>CTGCAAAATCATACGCGCGGATTTCTGCTCC</td>
<td></td>
</tr>
<tr>
<td>SMa2063</td>
<td>Sensor histidine kinase</td>
<td>-1.3 H1, hs; -2.8 H1H2, sp n/a</td>
<td>n/a</td>
<td>CTGCAAGACGCACGCGCATCAATGAGATCGT</td>
<td></td>
</tr>
<tr>
<td>SMb20117</td>
<td>Glycosyltransferase</td>
<td>-4.0 H1, hs; -2.9 H2, sp 70</td>
<td>70</td>
<td>CTGCCCTTTGCTTCTGCCGATATGGAAAGAGA</td>
<td></td>
</tr>
<tr>
<td>SMb20302</td>
<td>Conserved hypothetical protein</td>
<td>-1.1 H1, hs; -5.4 H2, sp 32</td>
<td>32</td>
<td>CTGCCCTTTGCTTCTGCCGATATGGAAAGAGA</td>
<td></td>
</tr>
<tr>
<td>SMc04146</td>
<td>Hypothetical protein</td>
<td>-1.6 H1, hs; -4.3 H2, sp 15'</td>
<td>15'</td>
<td>CTGCAAGACGCACGCGCATCAATGAGATCGT</td>
<td></td>
</tr>
<tr>
<td>SMc00106</td>
<td>Conserved hypothetical protein</td>
<td>-2.7 H1, hs; -3.9 H2, sp 36</td>
<td>36</td>
<td>CTGCAAGACGCACGCGCATCAATGAGATCGT</td>
<td></td>
</tr>
<tr>
<td>SMc00367</td>
<td>Conserved hypothetical protein</td>
<td>-5.3 H1, hs; -3.0 H2, sp 83</td>
<td>83</td>
<td>CTGCAATTTTCTGGCATCTACATTGATCGT</td>
<td></td>
</tr>
<tr>
<td>SMc00952</td>
<td>Thioredoxin-like protein</td>
<td>-2.3 H1, hs; -1.4 H2, sp 25</td>
<td>25</td>
<td>ATGAAAGATTTTGGCTGGGCGGCTCTAGCAGT</td>
<td></td>
</tr>
<tr>
<td>SMc01280</td>
<td>degP3 Serine protease</td>
<td>-2.0 H1, hs; -1.1 H2, sp 24</td>
<td>24</td>
<td>TTGCAATTTTGGCTGGGCGGCTCTAGCAGT</td>
<td></td>
</tr>
<tr>
<td>SMc01279</td>
<td>Replication-associated recombination protein</td>
<td>-1.3 H1, hs n/a</td>
<td>n/a</td>
<td>ATGAAAGATTTTGGCTGGGCGGCTCTAGCAGT</td>
<td></td>
</tr>
<tr>
<td>SMc01759</td>
<td>Conserved hypothetical protein</td>
<td>-3.8 H1, hs; -1.9 H2, sp 48</td>
<td>48</td>
<td>TTGCAATTTTGGCTGGGCGGCTCTAGCAGT</td>
<td></td>
</tr>
<tr>
<td>SMc01758</td>
<td>groEL4 60 KD chaperonin</td>
<td>-3.4 H1, hs; -1.8 H2, sp n/a</td>
<td>n/a</td>
<td>TTGCAATTTTGGCTGGGCGGCTCTAGCAGT</td>
<td></td>
</tr>
<tr>
<td>SMc03176</td>
<td>Conserved hypothetical protein</td>
<td>-1.5 H1, hs; -1.7 H2, sp 0</td>
<td>0</td>
<td>TTGCAATTTTGGCTGGGCGGCTCTAGCAGT</td>
<td></td>
</tr>
<tr>
<td>SMc03177</td>
<td>ATP-dependent DNA ligase</td>
<td>-1.1 H1, hs n/a</td>
<td>n/a</td>
<td>TTGCAATTTTGGCTGGGCGGCTCTAGCAGT</td>
<td></td>
</tr>
<tr>
<td>SMc03246</td>
<td>Phage-related integrase</td>
<td>-3.0 H1, hs; -1.1 H2, sp 47</td>
<td>47</td>
<td>TTGCAATTTTGGCTGGGCGGCTCTAGCAGT</td>
<td></td>
</tr>
<tr>
<td>SMc03802</td>
<td>Peptidase</td>
<td>-2.6 H1, hs; -2.6 H2, sp 52</td>
<td>52</td>
<td>TTGCAATTTTGGCTGGGCGGCTCTAGCAGT</td>
<td></td>
</tr>
<tr>
<td>SMc03803</td>
<td>Conserved hypothetical protein</td>
<td>-2.3 H1, hs; -2.6 H2, sp n/a</td>
<td>n/a</td>
<td>TTGCAATTTTGGCTGGGCGGCTCTAGCAGT</td>
<td></td>
</tr>
<tr>
<td>SMc03836</td>
<td>Arylesterase</td>
<td>-3.2 H1, hs; -2.0 H2, sp 44'</td>
<td>44'</td>
<td>TTGCAATTTTGGCTGGGCGGCTCTAGCAGT</td>
<td></td>
</tr>
<tr>
<td>SMc04202</td>
<td>Conserved hypothetical protein</td>
<td>-3.3 H1, hs; -1.3 H2, sp 220</td>
<td>220</td>
<td>TTGCAATTTTGGCTGGGCGGCTCTAGCAGT</td>
<td></td>
</tr>
</tbody>
</table>

3. RpoH1-dependent in heat shock and RpoH2-dependent in stationary phase, used to determine the dual RpoH1/RpoH2 promoter consensus sequence:

- Conserved hypothetical protein
- Sensor histidine kinase
- Glycosyltransferase
- Conserved hypothetical protein
- Thioredoxin-like protein
- degP3 Serine protease
- Replication-associated recombination protein
- groEL4 60 KD chaperonin
- Conserved hypothetical protein
- ATP-dependent DNA ligase
- Phage-related integrase
- Peptidase
- Conserved hypothetical protein
- Arylesterase
- Conserved hypothetical protein

43
<table>
<thead>
<tr>
<th>Gene</th>
<th>Description</th>
<th>Promoters</th>
</tr>
</thead>
<tbody>
<tr>
<td>SMa0136</td>
<td>Conserved hypothetical protein</td>
<td>-2.4 H2, sp 60</td>
</tr>
<tr>
<td>SMb20303</td>
<td>Hypothetical protein</td>
<td>-4.3 H1, hs; -1.2 H1, sp 38</td>
</tr>
<tr>
<td>SMb21295</td>
<td>Heat shock protein, hsp20 family</td>
<td>-4.6 H1, hs; -2.7 H1H2, sp 87</td>
</tr>
<tr>
<td>SMb21294</td>
<td>Heat shock protein, hsp20 family</td>
<td>-5.3 H1, hs n/a</td>
</tr>
<tr>
<td>SMb21296</td>
<td>Hypothetical protein</td>
<td>-5.5 H1, hs; -2.2 H1H2, sp 119</td>
</tr>
<tr>
<td>SMc00030</td>
<td>Hypothetical protein</td>
<td>-1.9 H2, sp 30</td>
</tr>
<tr>
<td>SMc00031</td>
<td>Hypothetical protein</td>
<td>-1.3 H2, sp n/a</td>
</tr>
<tr>
<td>SMc00048</td>
<td>Conserved hypothetical protein, Usp-like</td>
<td>-1.1 H2, sp 197</td>
</tr>
<tr>
<td>SMc00049</td>
<td>Conserved hypothetical protein</td>
<td>-1.2 H2, sp n/a</td>
</tr>
<tr>
<td>SMc00814</td>
<td>Hypothetical protein</td>
<td>-1.9 H1, hs; -2.9 H2, sp 46</td>
</tr>
<tr>
<td>SMc00949</td>
<td>HspQ-like protein</td>
<td>-3.8 H1, hs; -1.6 H1, sp 17</td>
</tr>
<tr>
<td>SMc00969</td>
<td>Amidohydrolase</td>
<td>1.1 H1, hs; -2.1 H2, sp 0</td>
</tr>
<tr>
<td>SMc00970</td>
<td>Exodeoxyribonuclease VII</td>
<td>-1.4 H2, sp n/a</td>
</tr>
<tr>
<td>SMc01180</td>
<td>Conserved hypothetical protein</td>
<td>-4.7 H1, hs; -1.3 H1H2, sp 19</td>
</tr>
<tr>
<td>SMc01329</td>
<td>azoR</td>
<td>FMN-dep. NADH-azoreductase</td>
</tr>
<tr>
<td>SMc02382</td>
<td>Conserved hypothetical protein</td>
<td>-3.9 H1, hs; -2.5 H1H2, sp 30</td>
</tr>
<tr>
<td>SMc02433</td>
<td>clpB</td>
<td>ATP-dependent protease</td>
</tr>
<tr>
<td>SMc02575</td>
<td>hslV</td>
<td>ATP-dependent protease</td>
</tr>
<tr>
<td>SMc02576</td>
<td>GCN5-related N-acetyltransferase</td>
<td>-4.4 H1, hs n/a</td>
</tr>
<tr>
<td>SMc02577</td>
<td>hslU</td>
<td>ATP-dependent protease</td>
</tr>
<tr>
<td>SMc02857</td>
<td>dnaK</td>
<td>Chaperone</td>
</tr>
<tr>
<td>SMc02769</td>
<td>Conserved hypothetical protein</td>
<td>-4.6 H1, hs; -3.2 H1, sp 44</td>
</tr>
<tr>
<td>Identifier</td>
<td>Description</td>
<td>SLR</td>
</tr>
<tr>
<td>----------------</td>
<td>------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>SMc02863</td>
<td>recF: DNA repair protein</td>
<td>-2.2</td>
</tr>
<tr>
<td>SMc02900</td>
<td>Conserved hypothetical protein</td>
<td>-1.7</td>
</tr>
<tr>
<td>SMc03857</td>
<td>ffb: Signal recognition particle protein</td>
<td>-1.1</td>
</tr>
<tr>
<td>SMc03858</td>
<td>pheAa: Chorismate mutase</td>
<td>-1.1</td>
</tr>
<tr>
<td>SMc03859</td>
<td>rpsP: 3OS ribosomal protein S16</td>
<td>-1.4</td>
</tr>
<tr>
<td>SMc04040</td>
<td>ibpA: Heat shock protein, hsp20 family</td>
<td>-5.8</td>
</tr>
<tr>
<td>SMc04310</td>
<td>Hypothetical protein</td>
<td>-3.6</td>
</tr>
<tr>
<td>SMc04406</td>
<td>Conserved hypothetical protein</td>
<td>-4.3</td>
</tr>
<tr>
<td>SmelC456</td>
<td>Trans-encoded sRNA</td>
<td>-2.4</td>
</tr>
<tr>
<td>SmelC781</td>
<td>Trans-encoded sRNA</td>
<td>-2.5</td>
</tr>
</tbody>
</table>

* Identifier in italics indicates that the gene is predicted to be in an operon with the gene listed directly above it.

b Average SLR change of all pairwise comparisons for rpoH mutants compared to wild type. A negative value means expression was decreased in the mutant. Abbreviations: H1, rpoH1 mutant; H2, rpoH2 mutant; hs, heat shock; sp, stationary phase. SLR values for the double mutant (H1/H2) are shown only when there was no change for a single mutant under that growth condition. SLR is expressed as the log2 ratio of the change, i.e., a SLR of 1 equals a 2-fold change. Values in bold were slightly below our cutoff for SLR (SMc02703) or pairwise comparisons (SMc01280), but the rpoH-dependence of the respective genes was confirmed by RACE mapping.

c Distance between the transcription start site (+1) and the first nucleotide of the translational start codon. An asterisk next to the value indicates that the value was adjusted because the annotated start codon is likely incorrect.

d Putative RpoH-dependent promoters determined by RACE mapping. Transcription start site is in bold. Putative -35 and -10 motifs are underlined.
Heat shock
997 total
Acid shock
124 total
RpoH1
327 total
26
758
170
29
31
85
1) RpoH1

2) RpoH2

3) RpoH1/H2