

Video Article

DNA-affinity-purified Chip (DAP-chip) Method to Determine Gene Targets for Bacterial Two component Regulatory Systems

Lara Rajeev¹, Eric G. Luning¹, Aindrita Mukhopadhyay¹

¹Physical Biosciences Division, Lawrence Berkeley National Laboratory

Correspondence to: Lara Rajeev at lrajeev@lbl.gov

URL: <https://www.jove.com/video/51715>

DOI: [doi:10.3791/51715](https://doi.org/10.3791/51715)

Keywords: Genetics, Issue 89, DNA-Affinity-Purified-chip, response regulator, transcription factor binding site, two component system, signal transduction, *Desulfovibrio*, lactate utilization regulator, ChIP-chip

Date Published: 7/21/2014

Citation: Rajeev, L., Luning, E.G., Mukhopadhyay, A. DNA-affinity-purified Chip (DAP-chip) Method to Determine Gene Targets for Bacterial Two component Regulatory Systems. *J. Vis. Exp.* (89), e51715, doi:10.3791/51715 (2014).

Abstract

In vivo methods such as ChIP-chip are well-established techniques used to determine global gene targets for transcription factors. However, they are of limited use in exploring bacterial two component regulatory systems with uncharacterized activation conditions. Such systems regulate transcription only when activated in the presence of unique signals. Since these signals are often unknown, the *in vitro* microarray based method described in this video article can be used to determine gene targets and binding sites for response regulators. This DNA-affinity-purified-chip method may be used for any purified regulator in any organism with a sequenced genome. The protocol involves allowing the purified tagged protein to bind to sheared genomic DNA and then affinity purifying the protein-bound DNA, followed by fluorescent labeling of the DNA and hybridization to a custom tiling array. Preceding steps that may be used to optimize the assay for specific regulators are also described. The peaks generated by the array data analysis are used to predict binding site motifs, which are then experimentally validated. The motif predictions can be further used to determine gene targets of orthologous response regulators in closely related species. We demonstrate the applicability of this method by determining the gene targets and binding site motifs and thus predicting the function for a sigma54-dependent response regulator DVU3023 in the environmental bacterium *Desulfovibrio vulgaris* Hildenborough.

Video Link

The video component of this article can be found at <https://www.jove.com/video/51715/>

Introduction

The ability of bacteria to survive and thrive is critically dependent on how well they are able to perceive and respond to perturbations in their environments, and this in turn is dependent on their signal transduction systems. The number of signaling systems a bacterium encodes has been called its "microbial IQ" and can be an indication of both variability of its environment and its ability to sense multiple signals and fine tune its response¹. Two component signal transduction systems (TCS) are the most prevalent signaling systems used by bacteria, and they consist of a histidine kinase (HK) that senses the external signal and transmits via phosphorylation to an effector response regulator (RR)². RRs can have a variety of output domains and thus different effector modes, but the most common response is transcriptional regulation via a DNA binding domain³. The signals sensed and the corresponding functions of the vast majority of TCSs remain unknown.

Although *in vivo* methods such as ChIP-chip are routinely used for determination of genomic binding sites of transcription factors³, they can only be used for bacterial two component system RRs if the activating conditions or signals are known. Often the environmental cues that activate a TCS are harder to determine than their gene targets. The *in vitro* microarray based assay described here can be used to effectively and rapidly determine the gene targets and predict functions of TCSs. This assay takes advantage of the fact that RRs can be phosphorylated and thus activated *in vitro* using small molecule donors like acetyl phosphate⁴.

In this method, named DAP-chip for DNA-affinity-purified-chip (**Figure 1**), the RR gene of interest is cloned with a His-tag in *E. coli*, and a subsequently purified tagged protein is allowed to bind to sheared genomic DNA. The protein-bound DNA is then enriched by affinity-purification, the enriched and input DNA are amplified, fluorescently labeled, pooled together and hybridized to a tiling array that is custom made to the organism of interest (**Figure 1**). Microarray experiments are subject to artifacts and therefore additional steps are employed to optimize the assay. One such step is to attempt to determine one target for the RR under study using electrophoretic mobility shift assays (EMSA) (see workflow in **Figure 2**). Then, following binding to genomic DNA and the DAP steps, the protein-bound and input DNA are examined by qPCR to see if the positive target is enriched in the protein-bound fraction relative to the input fraction, thus confirming optimal binding conditions for the RR (**Figure 2**). After array hybridization, the data are analyzed to find peaks of higher intensity signal indicating genomic loci where the protein had bound. Functions may be predicted for the RR based on the gene targets obtained. The target genomic loci are used to predict binding site motifs, which are then experimentally validated using EMSAs (**Figure 2**). The functional predictions and gene targets for the RR may then be extended to closely related species that encode orthologous RRs by scanning those genomes for similar binding motifs (**Figure 2**). The DAP-chip method can provide a wealth of information for a TCS where previously there was none. The method can also be used for any

transcriptional regulator if the protein can be purified and DNA binding conditions can be determined, and for any organism of interest with a genome sequence available.

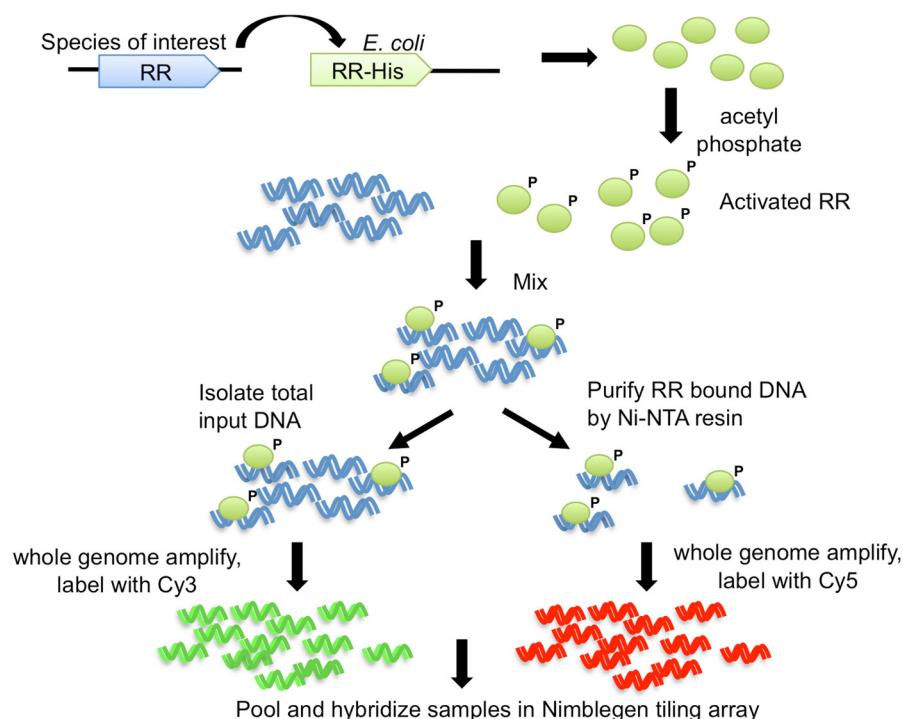


Figure 1. The DNA-affinity-purified-chip (DAP-chip) strategy⁷. The RR gene from the organism of interest is cloned with a carboxy-terminal His-tag into an *E. coli* expression strain. Purified His-tagged protein is activated by phosphorylation with acetyl phosphate, and mixed with sheared genomic DNA. An aliquot of the binding reaction is saved as input DNA, while the rest is subjected to affinity purification using Ni-NTA resin. The input and the RR-bound DNA are whole genome amplified, and labeled with Cy3 and Cy5, respectively. The labeled DNA is pooled together and hybridized to a tiling array, which is then analyzed to determine the gene targets. Figure modified and reprinted using the creative commons license from⁷.

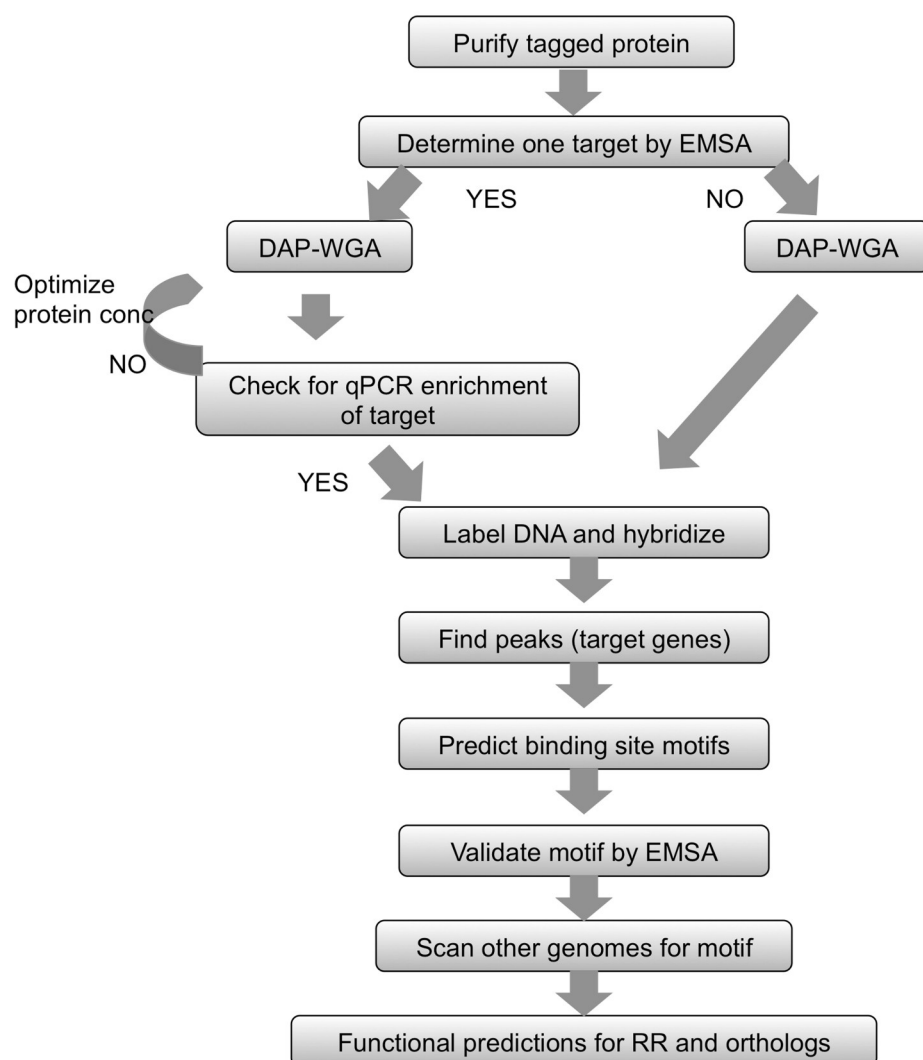


Figure 2. Summary of workflow. For any purified tagged protein, begin by determining a target using EMSA. Allow protein to bind genomic DNA and then DNA-affinity-purify (DAP) and whole genome amplify (WGA) the enriched and input DNA. If a gene target is known, use qPCR to ensure that the known target is enriched in the protein-bound fraction. If no target could be determined, proceed directly to DNA labeling and array hybridization. If enrichment by qPCR could not be observed, then repeat the protein-gDNA binding and DAP-WGA steps using different protein amounts. Use array analysis to find peaks and map them to target genes. Use the upstream regions of target genes to predict binding site motifs. Validate the motifs experimentally using EMSAs. Use the motif to scan the genomes of related species encoding orthologs of the RR under study, and predict genes targeted in those species as well. Based on the gene targets obtained, the physiological function of the RR and its orthologs may be predicted. Figure modified and reprinted using the creative commons license from⁷.

Protocol

Note: The protocol below is tailored for determination of gene targets of the RR DVU3023 from the bacterium *Desulfovibrio vulgaris* Hildenborough. It can be adapted to any other transcriptional regulator of interest.

1. Clone and Purify RR

1. Clone the RR gene, specifically DVU3023, from *D. vulgaris* Hildenborough into an *Escherichia coli* expression vector such that the gene is C-terminally His-tagged and expression is under the control of a T7 promoter.
Note: Several cloning methods may be used and is determined by the researcher. Alternate affinity tags may also be used.
2. Transform the expression construct into *E. coli* BL21 (DE3) expression strain.
3. Grow up 1 L of the BL21 expression strain at 37 °C. At mid-log phase, add IPTG to 0.5 mM to induce protein expression. Continue growth at room temperature for 24 hr.

4. Centrifuge cells at 5,000 x g for 10 min at 4 °C. Resuspend cells in lysis buffer (20 mM sodium phosphate, pH 7.4, 500 mM NaCl, 40 mM imidazole, 1 mg/ml lysozyme, 1x benzonase nuclease). Lyse cells using a French press at 4 °C.
 5. Centrifuge lysate at 15,000 x g for 30 min at 4 °C. Filter using a 0.45 µm syringe filter.
 6. Wash a 1 ml Ni Sepharose column on an FPLC instrument using 10 ml of wash buffer (20 mM sodium phosphate, pH 7.4, 500 mM NaCl, 40 mM imidazole).
 7. Load lysate on the column. Wash column with 20 ml of wash buffer.
 8. Elute DVU3023 using a gradient of 0-100% elution buffer (20 mM sodium phosphate, pH 7.4, 500 mM NaCl, 500 mM imidazole).
 9. Load eluted fractions on a desalting column, and wash with a desalting buffer (20 mM sodium phosphate pH 7.4, 100 mM NaCl).
 10. Centrifuge protein sample in a high molecular weight cutoff centrifuge filter to concentrate the protein. Add DTT to 0.1 mM and glycerol to 50% and store protein at -20 °C.
- Note: Purification methods will need to be optimized individually for each protein under study.

2. Determine Gene Target for RR Using Electrophoretic Mobility Shift Assay (EMSA)

1. PCR amplify 400 bp region upstream of the candidate target gene DVU3025, using *D. vulgaris* Hildenborough genomic DNA as the template, and a forward unlabeled primer and a reverse 5'-biotin-labeled primer.
Note: Tips to select a candidate target gene: Often RRs bind the upstream regions of their own gene/operon, or they may regulate proximally encoded genes. If the RR has orthologs in other species, look for genes that are conserved in the neighborhood. Alternate methods include choosing candidate genes based on regulon predictions (e.g., RegPrecise⁵), or predicting sigma54-dependent promoters⁶ for RRs that are themselves sigma54-dependent.
2. Run the PCR product on a 1% agarose gel, cut out the 400 bp sized product, and purify the DNA using a gel extraction clean up kit.
3. Mix DVU3023 protein (0.5 pmol) with 100 fmol of biotinylated DNA substrate in 10 mM Tris HCl, pH 7.5, 50 mM KCl, 5 mM MgCl₂, 1 mM DTT, 25% glycerol and 1 µg/ml poly dI.dC (non-specific competitor DNA) in a total volume of 20 µl. Also set up a reaction without any protein as a control. Incubate the reactions at room temperature on the bench for 20 min.
4. Note: These are standard conditions listed and other reaction components may be added depending on the RR of interest. If activation is desired, add 50 mM acetyl phosphate to the reaction (often RRs will bind DNA *in vitro* without activation).
5. Pre-run a precast mini 6% polyacrylamide-0.5x TBE gel in 0.5x TBE buffer at 100 V for 30 min. Add 5 µl of 5x loading buffer (0.1% bromophenol blue, 0.1% xylene cyanol, 30% glycerol in 1x TBE) to the binding reactions and load 18 µl of the reactions on to the gel. Run at 100 V for 2 hr.
Note: In order to avoid overheating, which can lead to disassociation of the protein-DNA complexes, fill the outer buffer chamber in the electrophoresis system with running buffer such that the majority of the gel is insulated by buffer. Alternately, the gel may be run at 4 °C.
6. Cut a charged nylon membrane to the size of the gel and soak it in 0.5x TBE for at least ten minutes. Remove the gel from the cassette. Cut any ridges off the gel and sandwich the gel and membrane between two thick filter papers soaked in 0.5x TBE, and place inside a semi-dry blotting apparatus and run at 20 V for 30 min.
7. Place the membrane inside a commercial UV light crosslinker instrument and set the time to 3 min.
Note: The membrane may now be stored dry at room temperature for several days before proceeding to the next step.
8. Use a commercially available chemiluminescent detection kit that employs a streptavidin-horse radish peroxidase conjugate to develop the blot as per manufacturer's instructions.
9. Image the blot using a computer hooked up to a CCD equipped camera and look for a shift in the DNA substrate mobility in the presence of the RR, which indicates that the RR binds the DNA being tested.
Note: The specificity of the protein-DNA shift can be tested by including in the reaction an excess of unlabeled competitor DNA, which should eliminate or decrease the shift seen.

3. Verify Target Enrichment after Genomic DNA-protein Binding

1. Genomic DNA-protein binding reaction
 1. Shear genomic DNA to an average size of 500 bp by sonicating 100 µl genomic DNA (at concentrations 100-200 ng/µl) in a 1.5 ml microfuge tube using a microtip with 9 low amplitude pulses of 1 sec, with 2 sec gap in between each.
Note: If the DNA splatters to the sides of the tube, spin down the contents after every 3 pulses. The exact conditions used will vary with the sonicator instrument used. The sheared DNA fragments may range from 100-1000 bp, but the average size should be within 400-600 bp. Too much shearing may result in fewer intact binding sites, and either too much or too little shearing will affect library preparation during the downstream whole genome amplification steps.
 2. Mix 2-3 µg of sheared genomic DNA with RR protein (0.5 pmol of DVU3023) in 10 mM Tris-HCl pH 7.5, 1 mM DTT, 50 mM KCl, 5 mM MgCl₂, 25% glycerol, and 50 mM acetyl phosphate. Incubate reactions at 25 °C for 30 min. Transfer 10 µl of this reaction to a 1.5 ml tube and label as input DNA.
Note: Acetyl phosphate is added to activate the protein by *in vitro* phosphorylation. Phosphorylation stimulates DNA binding, but many RRs also bind DNA *in vitro* without activation⁷. The amount of protein added to the reaction will depend on activity of the protein prep. The EMSA may be used to optimize this amount.
2. Affinity purify the protein-bound DNA
 1. Add 30 µl of Ni-NTA agarose resin to a 0.6 ml microfuge tube. Centrifuge at 100 x g for 1 min to collect the resin at the bottom, and remove the supernatant. Add 100 µl of wash buffer (10 mM Tris-HCl, pH 7.5, 5 mM MgCl₂, 50 mM KCl, 25% glycerol), flick the tube to mix, and centrifuge at 100 x g for 2 min. Remove supernatant.
 2. Add the remaining 90 µl of the binding reaction to the washed Ni-NTA resin and incubate in a rotary shaker for 30 min.
 3. Centrifuge at 100 x g for 2 min, and remove the supernatant (unbound DNA). Add 100 µl of wash buffer, flick the tube to mix contents, and centrifuge at 100 x g for 2 min. Remove supernatant. Repeat the wash step twice more.
 4. Add 35 µl of elution buffer (20 mM sodium phosphate buffer pH 8, 500 mM NaCl, 500 mM imidazole) to the resin and mix by vortexing. Incubate on the bench at RT for 5 min. Centrifuge at 100 x g for 2 min. Transfer the supernatant to a new 1.5 ml tube and label as the protein-bound DNA fraction.

5. Also add 35 μ l of the elution buffer to the input DNA.
 6. Purify the input and the protein-bound DNA fractions using a PCR purification kit.
3. Whole genome amplify the input and protein-bound DNA samples.
 1. Add 10 μ l of the input and protein-bound DNA to separate PCR tubes. Add 1 μ l of 10x Fragmentation buffer, 2 μ l of Library Preparation buffer, and 1 μ l of Library Stabilization solution to each sample. Mix well by vortexing. Heat in a thermocycler at 95 °C for 2 min, and chill on ice.
Note: The starting amount of DNA should be at least 10 ng in order to avoid introducing amplification bias.
 2. Add 1 μ l of Library Preparation Enzyme, mix by pipetting, and incubate in thermal cycler at 16 °C/20 min, 24 °C/20 min, 37 °C for 20 min, 75 °C/5 min, and hold at 4 °C.
 3. To each tube, add 47.5 μ l of water, 7.5 μ l of 10x amplification master mix, and 5 μ l of polymerase. Mix well and heat at 95 °C/3 min, followed by 20 cycles of 94 °C/15 sec, 65 °C/5 min. Hold reactions at 4 °C.
 4. Purify the amplified DNA samples using a PCR purification kit and measure DNA concentration spectrophotometrically.
 4. Verify target enrichment in the protein-bound DNA using qPCR
 1. Design qPCR primers to amplify 200 bp of the upstream region of the EMSA-verified target gene (DVU3025) using any freely available primer design software.
 2. Set up triplicate qPCR reactions for each DNA template with each primer set. Prepare a master mix for each primer set. 1x master mix contains 10 μ l of 2x Sybr Green qPCR mix, 0.5 μ M of each primer and water to a total of 18 μ l. Aliquot 18 μ l of the master mix per well of a 96-well PCR plate.
 3. Dilute the amplified and purified input and protein-bound DNA samples to 5 ng/ μ l with water and use as DNA template. Add 2 μ l of DNA template to the wells.
 4. Seal the plate with an ultra clear qPCR sealing film, spin down plate in a centrifuge at 200 x g/1 min. Place plate in a real time qPCR machine. Cycle as follows using the associated qPCR software: 95 °C/1 min, and 40 cycles of 95 °C/10 sec, 59 °C/15 sec, and 70 °C/35 sec.
 5. Also set up triplicate qPCR reactions with primers to amplify upstream regions of an unrelated gene (negative control).
 6. Calculate ΔC_T by subtracting C_T value of protein-bound DNA from that of the input DNA. Calculate fold enrichment of target gene in the protein-bound DNA as $2^{\Delta C_T}$.
Note: If the target upstream region was enriched in the protein-bound DNA vs. the input DNA, it indicates that the binding reactions and affinity purification were successful. Proceed to step 4. If enrichment of the target upstream region was not observed, repeat the binding reactions (step 3.1) with different amount of protein.

4. DNA Labeling and Array Hybridization

1. Label Input DNA with Cy3 and Enriched DNA with Cy5
Note: Cy3 and Cy5 dyes are light sensitive and care should be taken to keep light exposure to a minimum.
 1. Mix 1 μ g DNA with 40 μ l Cy3/Cy5-labeled 9-mers and adjust volume to 80 μ l with water.
 2. Heat denature at 98 °C for 10 min in the dark (in a thermocycler). Quick chill on ice for 2 min.
 3. Add 2 μ l Klenow polymerase (3'-5'-exo⁻, 50,000 U/ml), 5 mM dNTPs, and 8 μ l water to each reaction, mix well, and incubate at 37 °C for 2 hr in the dark (in a thermocycler).
 4. Add EDTA to 50 mM to stop the reaction and NaCl solution to 0.5 M.
 5. Transfer samples to 1.5 ml tubes containing 0.9 volume of isopropanol, incubate in dark for 10 min, and centrifuge at 12,000 x g for 10 min. The pellet should be pink for Cy3-labeled DNA and blue for Cy5-labeled DNA.
 6. Wash pellet with 80% ethanol (500 μ l) and centrifuge at 12,000 x g for 2 min. Air-dry pellets for 5-10 min in dark.
Note: Pellets may be stored at -20 °C.
 7. Resuspend pellet in 25 μ l water. Measure DNA concentration spectrophotometrically.
 8. Pool together 6 μ g each of the Cy3- and Cy5-labeled DNA in a 1.5 ml tube and vacuum dry in a centrifuge on low heat in the dark (cover the centrifuge lid if it is transparent).
Note: Pellets may be stored at -20 °C until ready for hybridization.
2. Microarray Hybridization
 1. Turn the hybridization system on 3-4 hr prior to use and set the temperature to 42 °C.
 2. Prepare a hybridization solution master mix such that 1x solution contains 11.8 μ l 2x Hybridization buffer, 4.7 μ l Hybridization Component A, and 0.5 μ l alignment oligo.
 3. Resuspend the pellets in 5 μ l water. Add 13 μ l of this mix to the sample. Vortex for 15 sec, incubate at 95 °C in a dry bath for 5 min. Keep samples at 42 °C in the hybridization system until ready for loading.
 4. Prepare the microarray slide-mixer assembly, according to manufacturer's protocol.
 5. Place the mixer-slide assembly within the hybridization system. Load 16 μ l of sample into the fill port, seal the ports with an adhesive film, turn mixing on in the system, and hybridize for 16-20 hr at 42 °C.
 6. Prepare 1x wash buffers I (250 ml), II (50 ml) and III (50 ml) by diluting the 10x commercially available buffers in water, and to each add DTT to 1 mM. Warm Buffer I to 42 °C.
 7. Slide the mixer-slide into a disassembly tool, and place inside a dish containing warm Buffer I. Peel the mixer off, while vigorously shaking the disassembly tool by hand.
 8. Place the slide into a container with 50 ml of Wash buffer I and shake vigorously for 2 min by hand.
 9. Transfer slide into a second container with 50 ml Wash buffer II and shake vigorously for 1 min by hand.
 10. Transfer slide into a third container with 50 ml Wash buffer III, and shake vigorously for 15 sec by hand.
 11. Quickly blot the edges of the slide on a paper towel, and place in a slide rack. Centrifuge at 200 x g for 2 min to dry the slide. Place within slide case, wrap it with foil, and store in a desiccator.

3. Scanning the Array
 1. Place the slide within the microarray scanner according to the instrument's instructions.
 2. Use a scanner software to set the wavelengths as 532 nm = Cy3 and 635 nm = Cy5, and the initial photomultiplier gains between 350 and 400.
 3. Preview the slide to locate the array on the slide. Select the array region for scanning.
 4. Scan the array and adjust the photomultiplier settings such that the array features are mostly yellow. The histogram should show the red and green curves superimposed or as close to each other as possible. The curves should end above the 10^{-5} normalized counts at saturation.
 5. Save both 532 and 635 nm images separately.
4. Array Data Analysis
 1. Import the images into an array analysis software. Using the software, create two pair reports (.pair) for each array (for Cy3 and Cy5 images). Create scaled \log_2 ratio files using the pair files. Use \log_2 ratio files to search for peaks using a sliding window of 500 bp. Map the peak loci to the upstream regions of gene targets.
 2. Check if the positive target (determined using EMSA and confirmed by qPCR, DVU3025 in this example) appears within the top peaks. Note: If the positive target is not among the top hits, it is possible that the RR under consideration has several gene targets and replicate DAP-chips can be conducted and peaks common to all replicates can be used to generate a target list.

5. Binding Site Motif Prediction and Validation

1. Retrieve sequences for upstream regions (400 bp) for the top gene targets as generated by DAP-chip analysis. Apply MEME on these sequences through Microbes Online (meme.nbcr.net) to predict motifs. Note: Enhancer binding proteins such as sigma54-dependent regulators can bind several hundred base pairs upstream of the start site. For other transcriptional regulators, a shorter region such as 200 bp upstream will be sufficient.
2. Design top and bottom strand DNA oligomers that contain the predicted binding site motif flanked by 10 bases on either end. Order the top strand 5' biotinylated.
3. Mix the top and bottom strand oligos in 1:1.5 ratio in 10 mM Tris HCl, pH 8.0, 1 mM EDTA, and 50 mM NaCl in a total reaction volume of 20 μ l in PCR tubes. Heat to 95 °C for 5 min in a thermal cycler, followed by slow cooling to 25 °C. Dilute ten fold and use as wild type dsDNA substrate for EMSA.
4. Prepare modified substrates using the same steps as above, but design oligomers to carry 4-6 substitutions in the conserved bases of the binding motif.
5. Set up binding reactions with the RR and the wild type or modified substrates and examine using EMSA as described in step 2.
6. Note: The predicted motif is validated if the RR binds the wild type substrate but not the modified one.

6. Conservation of Motif in Other Related Bacterial Species

1. Select genomes of interest that contain orthologs of DVU3023. Obtain sequence and annotation files for the genomes from the NCBI site.
2. Use a programming language such as Perl to write scripts that will use the motif sequences from the DAP-chip targets to build a position weighted matrix and use the matrix to score similar motifs present in other sequenced genomes.

Representative Results

The above method was applied to determine the global gene targets of the RRs in the model sulfate reducing bacterium *Desulfovibrio vulgaris* Hildenborough⁷. This organism has a large number of TCSs represented by over 70 RRs, indicating the wide variety of possible signals that it senses and responds to. *In vivo* analyses on the functions of these signaling systems are hard to perform since their signals and thus their activating conditions are unknown. Here the DAP-chip method was used to determine the gene targets and thus predict possible functions for a representative RR DVU3023.

DVU3023 is a sigma-54-dependent RR encoded in an operon with its cognate HK (**Figure 3A**). The C-terminal His-tagged gene was cloned into and purified from *E. coli*. For initial target determination, the purified RR was tested for binding to its downstream operon which is a ten gene-operon (DVU3025-3035) consisting of lactate uptake and oxidation genes. RR DVU3023 shifted the upstream region of DVU3025 (**Figure 3B**). Next RR DVU3023 was allowed to bind to sheared *D. vulgaris* Hildenborough genomic DNA. Although phosphorylation was not required for binding to the promoter region of DVU3025, acetyl phosphate was added to the reaction in case it was required for binding to other promoters. Following affinity purification of the protein-bound DNA fraction, qPCR was used to show that the upstream region of DVU3025 is enriched (8.45 fold) in the protein-bound fraction ($C_T = 6.9$) relative to the input DNA ($C_T = 9.98$) (**Figure 3C**), thus indicating that the binding conditions used were appropriate for RR DVU3023. Lack of enrichment of the promoter region of a randomly chosen gene (DVU0013) was used as a negative control.

The protein-bound and input DNA samples were then labeled fluorescently and hybridized to a *D. vulgaris* Hildenborough tiling array that had a high probe density in the intergenic regions. The top four peaks were chosen as the most likely targets for DVU3023 (**Figure 3D**). These four peaks were followed by several others, which were also identified in DAP-chip analyses for several other RRs and hence appear to be sticky DNA (**Table 1**). **Figure 3E** is a schematic representation of the genes regulated by DVU3023. The positive target DVU3025 was the first peak obtained with the highest score. Two gene targets are two other singly encoded lactate permeases (DVU2451 and DVU3284). The fourth gene target does not lie in an upstream region, but in the intergenic region between two convergently transcribed genes/operons (DVU0652 and DVU0653). This is a large intergenic region, and additionally also encodes a predicted sigma54-dependent promoter. It is possible that there is an undiscovered sRNA encoded in this region that is regulated by RR DVU3023.

Using the upstream regions of the targets obtained by DAP-chip, MEME was used to predict a binding site motif (**Figure 3F**, **Table 2**). EMSA substrates carrying the specific motif upstream of DVU3025 were designed to confirm that RR DVU3023 recognizes and binds the predicted motif. The motif was further validated by making substitutions in the conserved bases within the motif which eliminated the binding shift (**Figure 3G**). This validated motif was then used in Perl-generated scripts to scan the genome sequences of other closely related sulfate reducing bacteria that had orthologs for DVU3023. Loci were chosen as possible gene targets when the motif was located in upstream regions of open reading frames (**Table 3**). Using the motif sequences predicted for the orthologous RRs, a consensus binding site motif was generated (**Figure 3F**) which closely resembled the one obtained for *D. vulgaris* Hildenborough alone.

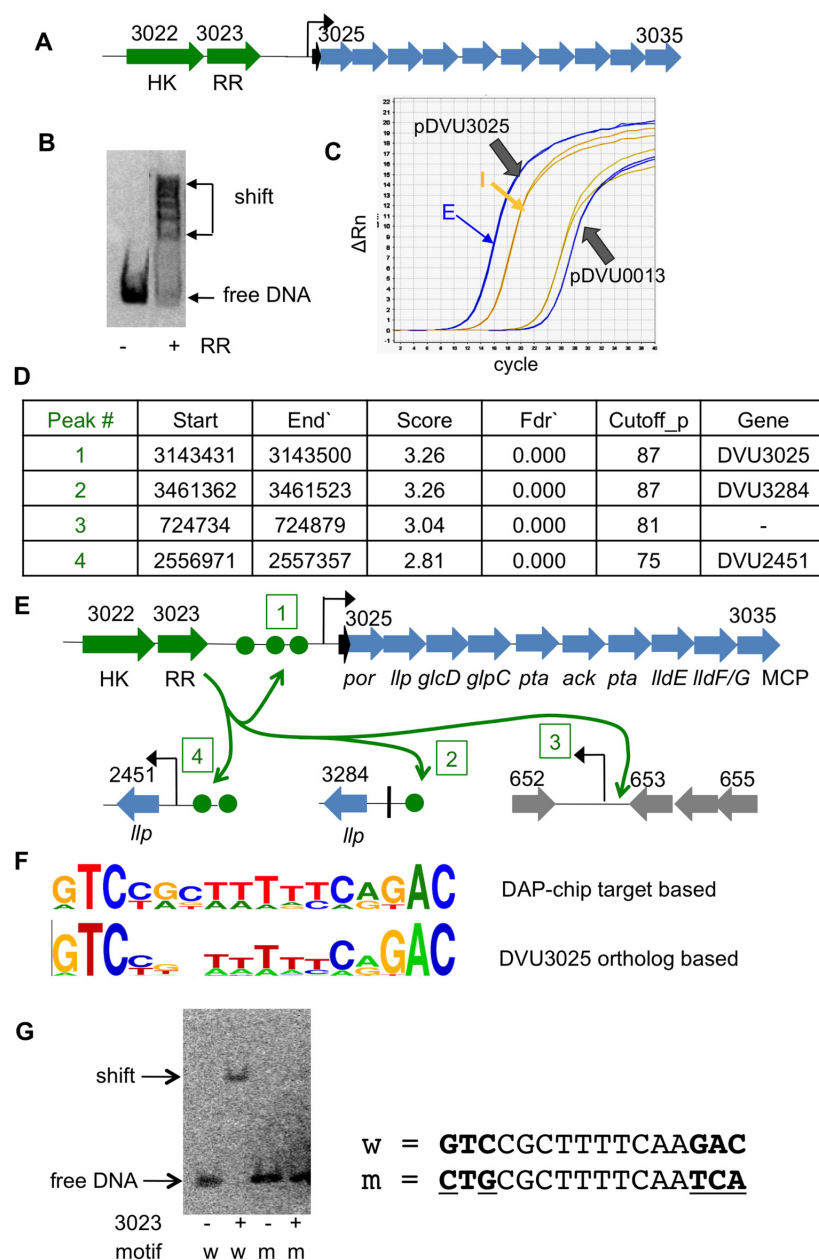


Figure 3. Determining genomic targets for *D. vulgaris* response regulator DVU3023. **A.** DVU3023 is encoded in an operon with its cognate HK. The downstream operon has a sigma54-dependent promoter (bent black arrow) and was used as a candidate target gene. **B.** Purified RR DVU3023 bound and shifted the upstream region of DVU3025⁷. **C.** q-PCR of upstream regions of DVU3025 (positive target) and DVU0013 (chosen as a negative control). **E** is protein-bound enriched DNA fraction, I is input DNA. **D.** Top four peaks obtained after DAP-chip analysis. Start and End refer to DNA coordinates at the start and end of the peaks; score refers to the log₂R ratio of the fourth highest probe in the peak; Fdr = false discovery rate; cutoff_p is the cutoff percentage at which the peak was identified. **E.** Schematic representation of the gene targets for DVU3023 based on the DAP-chip peaks. Numbers in boxes indicate the peak number in D. The HK-RR genes are in green, target genes in blue and other genes in grey. Black bent arrows are sigma54-dependent promoters, green filled circles are predicted binding site motifs. Gene names are as follows: *por* – pyruvate ferredoxin oxidoreductase; *llp* – lactate permease; *glcD* – glycolate oxidase; *glpC* – Fe-S cluster binding protein; *pta* – phosphotransacetylase; *ack* – acetate kinase; *lldE* – lactate oxidase subunit; *lldF/G* – lactate oxidase subunit; MCP – methyl-accepting chemotaxis protein. **F.** Weblogo⁸ images of the predicted binding site motif. Top – derived from DAP-chip targets; Bottom – derived from binding sites present in genomes with orthologs of DVU3023⁷. **G.** Validation of predicted binding site motif using EMSA. DVU3023 shifted the wild type motif (w) but not the modified motif (m). Sequences for the w and m motifs are shown on the right⁷. Figure modified and reprinted using the creative commons license from Rajeev *et al*⁷.

Peak attributes						Start Coordinate Mapping						End Coordinate mapping					
Location	Start	End	Score	Fdr	cutoff_p	Start	offset	geneStrand	overlapGene	DVU	Desc	End	GeneStrand	Offset	OverlapGene	DVU	Desc
GENOME	3143431	3143500	3.26	0.000	87	3143431	184	+	FALSE	DVU3024	hypothetical prc	3143500	+	115	FALSE	DVU3024	hypothetical protein (TIGR)
GENOME	3461362	3461523	3.26	0.000	87	3461362	196	-	FALSE	DVU3284	L-lactate perme	3461523	-	112	FALSE	DVU3285	hypothetical protein (TIGR)
GENOME	724734	724879	3.04	0.000	81	724734	201		FALSE			724879		201	FALSE		
GENOME	2556971	2557357	2.81	0.000	75	2556971	-76	+	TRUE	DVU2452	hypothetical prc	2557357		201	FALSE		
GENOME	1782380	1782629	2.74	0.000	73	1782380	-99	-	TRUE	DVU1705	type I restrictio	1782629	-	-81	TRUE	DVU1705	type I restriction-modification enzyme, S subunit (TIGR)
GENOME	1242717	1243131	2.59	0.000	69	1242717	-880	+	TRUE	DVU1147	alginate o-acety	1243131	+	87	FALSE	DVU1148	conserved hypothetical protein (TIGR)
GENOME	2280533	2280874	2.59	0.000	69	2280533	200	+	FALSE	DVU2182	hypothetical prc	2280874	+	11	FALSE	DVU2183	hypothetical protein (TIGR)
GENOME	2243710	2243987	2.44	0.000	65	2243710	45	-	FALSE	DVU2145	chloramphenicc	2243987		201	FALSE		
GENOME	2270761	2270902	2.44	0.000	65	2270761	-146	-	TRUE	DVU2174	hypothetical prc	2270902	-	-5	TRUE	DVU2174	hypothetical protein (TIGR)
GENOME	2088394	2089032	2.36	0.000	63	2088394	0	-	TRUE	DVU2007	nuclease, putati	2089032	+	-277	TRUE	DVU2009	conserved domain protein (TIGR)
GENOME	2944256	2944934	2.29	0.000	61	2944256	201		FALSE			2944934	-	-51	TRUE	DVU2840	conserved hypothetical protein (TIGR)
GENOME	1643740	1644119	2.25	0.000	60	1643740	-424	-	TRUE	DVU1562	HAMP domain	1644119	-	-45	TRUE	DVU1562	HAMP domain protein (TIGR)
GENOME	2080350	2080759	2.25	0.000	60	2080350	-790	+	TRUE	DVU1999	sulfate transpor	2080759	+	-1199	TRUE	DVU1999	sulfate transporter family protein (TIGR)
GENOME	231502	231941	2.21	0.000	59	231502	-811	-	TRUE	DVU0187	GGDEF domain	231941	-	-372	TRUE	DVU0187	GGDEF domain protein (TIGR)
GENOME	1582106	1582493	2.21	0.000	59	1582106	-333	-	TRUE	DVU1509	conserved hypo	1582493	-	54	FALSE	DVU1509	conserved hypothetical protein (TIGR)
GENOME	2816118	2816816	2.21	0.000	59	2816118	44	+	FALSE	DVU2709	hypothetical prc	2816816	+	43	FALSE	DVU2710	conserved hypothetical protein (TIGR)
GENOME	333493	333834	2.18	0.000	58	333493	-4	-	TRUE	DVU0287	hypothetical prc	333834	-	54	FALSE	DVU0288	hypothetical protein (TIGR)
GENOME	2694523	2694632	0.56	0.000	15	2694523	-1290	-	TRUE	Dv23SD	23S ribosomal	2694632	-	-1181	TRUE	Dv23SD	23S ribosomal RNA (NCBI)
GENOME	626560	626745	2.48	0.008	66	626560	-66	-	TRUE	DVU0559	lipoprotein, put	626745	+	-192	TRUE	DVU0560	conserved domain protein (TIGR)
GENOME	2076336	2077093	2.48	0.008	66	2076336	-80	+	TRUE	DVU1995	anti-anti-sigma	2077093	+	91	FALSE	DVU1996	quaternary ammonium compound-resistance protein QacC, putative (TIGR)

Table 1. Top 20 peaks from the DAP-chip array analysis. The table is divided into three sections. Peak attributes show details of the peaks as generated by array analysis software, where Location refers to whether the peak was found in the genome or the extrachromosomal plasmid, Start and End refer to the start and end loci for the peak, Score refers to the log₂R ratios for the fourth highest probe in the peak, Fdr refers to the false discovery rate value, and cutoff_p is the cutoff percentage at which the peak was identified. The other two sections Start coordinate mapping and End coordinate mapping map the start and end loci, respectively, of the peak to the gene. In these sections Gene strand refers to the strand coding the gene, Offset indicates distance from the start of the gene (positive values indicate loci is upstream of gene, while negative values indicate loci is within the gene), Overlap gene value is TRUE if the locus overlaps a gene, DVU refers to the DVU# of the gene that the coordinates map to, and Description indicates the gene annotation. Table modified and reprinted using the creative commons license from Rajeev *et al*⁷. [Please click here to view a larger version of this table.](#)

RR	Target	Motif	Location
DVU3023	DVU3025	GTCCGCTTTTCAAGAC	-350
		GTCCACTTTTTCAGAC	-250
		ATCCGCTTTTTCAGAC	-389
	DVU2451	GTCTGCAATGTCGGAC	-310
		GTCCATTTTTCAGAC	-401
	DVU3284	GTCTGGAAAACCGTAC	-127

Table 2. Sequences used to build the consensus DAP-chip target based motif in Figure 3. Table modified and reprinted using the creative commons license from Rajeev *et al*⁷.

Target	bp from ATG	Score	Sequence	Operon	Description
<i>D. vulgaris</i> Hildenborough					
DVU_2451	-310	6.29	GTCTGCAATGTCGGAC	DVU_2451	L-lactate permease
DVU_3025	-351	5.96	GTCCACTTTTTCAGAC	DVU_3025-3035	pyruvate ferredoxin oxidoreductase, I-lactate permease, glycolate oxidase, Fe-S cluster binding protein, phosphotransacetylase, acetate kinase, N-terminal of phosphotransacetylase, conserved hypothetical protein, Fe-S cluster binding protein. methyl accepting chemotaxis protein
DVU_3025	-251	5.96	GTCCGCTTTTTCAGAC	DVU_3025-3035	
DVU_3025	-390	4.61	ATCCGCTTTTTCAGAC	DVU_3025-3035	
DVU_2451	-401	4.01	GTCCATTTTTCAGAC	DVU_2451	L-lactate permease
DVU_3284	-127	3.75	GTCTGGAAACCGTAC	DVU_3284	L-lactate permease
<i>D. vulgaris</i> Miyazaki DvMF_1859 (78%)					
DvMF_1860	-466	5.96	GTCCGGATTTTTCGGAC	DvMF_1860-1869	pyruvate ferredoxin oxidoreductase, I-lactate permease, glycolate oxidase, Fe-S cluster binding protein, phosphotransacetylase, acetate kinase, N-terminal of phosphotransacetylase, conserved hypothetical protein, Fe-S cluster binding protein. methyl accepting chemotaxis protein
DvMF_1860	-592	*	ATCCGTTTTTTCGGAC	DvMF_1860-1869	
DvMF_1860	-631	*	GTCTACCTTTTTCGGAC	DvMF_1860-1869	
<i>D. desulfuricans</i> G20 Dde_3234 (69%)					
Dde_3237	-363	6.81	GTCCGCTTTTTCAGAC	Dde_3237-3247	pyruvate ferredoxin oxidoreductase, I-lactate permease, glycolate oxidase, Fe-S cluster binding protein, phosphotransacetylase, acetate kinase, N-terminal of phosphotransacetylase, conserved hypothetical protein, Fe-S cluster binding protein. methyl accepting chemotaxis protein
Dde_3237	-388	4.01	GTCTGTTTTTTCAGAC	Dde_3237-3247	
Dde_3237	-262	2.91	GTCTTTTTTTCAGAC	Dde_3237-3247	
<i>D. salicigena</i> DSM 2638 Desal_2575 (55%)					
Desal_2574	-229 *		GTCTTAAATTCGGAC	Desal_2574-2566	pyruvate ferredoxin oxidoreductase, I-lactate permease, glycolate oxidase, Fe-S cluster binding protein, phosphotransacetylase, acetate kinase, conserved hypothetical protein, Fe-S cluster binding protein. methyl accepting chemotaxis protein
Desal_2574	-376 *		GTCTGGTTTTTTCAGAC	Desal_2574-2567	
Desal_2574	-352 *		GTCTTTTTTTCAGAC	Desal_2574-2568	
<i>D. magneticus</i> RS1 DMR_20080 (59%)					
DMR_20070	-441	4.86	GTCCCTTTTTCAGAC	DMR_20070-20000	pyruvate ferredoxin oxidoreductase, glycolate oxidase, Fe-S cluster binding protein, CoA binding domain protein, N-terminal of phosphotransacetylase, I-lactate permease, conserved hypothetical protein, Fe-S cluster binding protein.
DMR_20070	-251	3.5	GTCTGAAAAACCGAC	DMR_20070-20000	
DMR_20070	-278	3.5	GTCTGAAAAACCGAC	DMR_20070-20000	
<i>Desulfotomobium baculatum</i> Dbac_1042 (58%)					
Dbac_1044	-459	4.6	GTCCGAAAAATCGGAC	Dbac_1044-1049	pyruvate ferredoxin oxidoreductase, I-lactate permease, glycolate oxidase, Fe-S cluster binding protein, phosphotransacetylase, acetate kinase
Dbac_1044	-188	2.91	GTCTTTTTTTCAGAC	Dbac_1044-1049	
<i>Desulfohalobium retbaense</i> Dret_0776 (63%)					
Dret_1039	-173	4.35	GTCCGCTATCAGGAC	Dret_1039-1031	I-lactate permease, glycolate oxidase, Fe-S cluster binding protein, pyruvate ferredoxin oxidoreductase, phosphotransacetylase, acetate kinase, conserved hypothetical protein, Fe-S cluster binding protein.

Table 3. Binding site motifs for DVU3023 orthologs present in other sequenced *Desulfovibrio* and related species. For each genome scanned, the organism name is indicated, followed by the locus tag for the DVU3023 ortholog and its percent identity to DVU3023 in parentheses. Score indicates the value assigned by the Perl program based on similarity to the input sequences. Description states the gene annotations for the genes in the target operon. Table modified and reprinted using the creative commons license from Rajeev *et al*⁷. [Please click here to view a larger version of this table.](#)

Discussion

The DAP-chip method described here was successfully used to determine the gene targets for several RRs in *Desulfovibrio vulgaris* Hildenborough⁷ of which one is shown here as a representative result. For RR DVU3023, choosing a candidate gene target was straightforward. DVU3025 is located immediately downstream of the RR gene, and the RR and target genes are conserved in several *Desulfovibrio* species, and additionally DVU3025 has a predicted sigma54-dependent promoter. The EMSA provides a simple method to rapidly test the RR for binding to the candidate target gene, and also allows the assessment of the activity of the purified protein sample, as well as determine optimal protein-DNA binding conditions.

The DNA binding activity can also be tested in the presence and absence of acetyl phosphate to see if phosphorylation is necessary for DNA binding. Not all RRs are phosphorylated by small molecule donors⁹, in these cases the purified cognate sensor kinase, if available, may be used to activate the protein. There are also atypical RRs known that lack key active site residues in the receiver domain and are not activated by phosphorylation¹⁰. For the majority of RRs studied, phosphorylation stimulates DNA binding¹¹. However there are examples where phosphorylation does not affect *in vitro* DNA binding¹², and there are examples where phosphorylation is required for binding¹³, and also cases where the subset of promoters bound increases with phosphorylation¹⁴. The DAP-chip method may be performed with and without phosphorylation to determine if there are differences in the set of promoters bound by the RR. However, it should also be noted that some RRs may be purified in a functionally phosphorylated form from *E. coli*¹³.

The protocol described here is for a His-tagged protein such that the protein-bound DNA is affinity purified using Ni-NTA agarose resin, but it can easily be adapted for any kind of tagged protein by using the appropriate affinity resin for pull-down. Following the DAP-WGA steps, the qPCR step provides an additional control to ensure that the protein-gDNA binding conditions were appropriate and that the protein-bound DNA was successfully enriched by the affinity purification. Greater than 3-fold enrichment is sufficient to proceed with the hybridization step. Unlike the EMSA reaction, there is no non-specific competitor DNA like poly-dI.dC added to the gDNA binding reaction since it interferes with the WGA step. Due to this if the protein sample has non-specific DNA binding activity then clear enrichment of the confirmed target DNA will not be observed by qPCR. Thus the qPCR control step can be used to optimize the amount of protein used in the gDNA binding reaction. Commercially available kits for whole genome amplification claim to introduce no amplification bias when following their guidelines for minimum starting DNA material and low number of amplification cycles. The amplification bias may be checked using qPCR with various primer sets on amplified versus unamplified DNA. Any effect on downstream array hybridization may also be determined by differentially labeling and hybridizing pooled amplified and unamplified genomic DNA.

The list of peaks that are generated following array data analysis is usually long with several peaks having a false discovery rate value of 0. Therefore a combination of other peak attributes such as high log₂R scores and cutoff_p values (as generated by the array analysis software used in this study) are used to cull the list down to the highest confidence gene targets. The presence of the pre-determined gene target

among the top five hits greatly strengthens the confidence in the data set. Performing this assay for a number of regulatory proteins from the same organism will also help to identify “sticky” DNA sequences that appear in several data sets⁷. Additionally if a binding site motif can be predicted and validated then the presence of the motif in peaks lower down in the list may also be used to choose a conservative target gene list. Enrichment of DNA sequences other than promoter regions may be artifacts of array hybridization or may indicate the regulation of previously unidentified open reading frames or small RNAs⁷. For regulators where a gene target could not be predetermined using EMSA, the DAP-chip assay may be performed “blind”⁷. In such cases also, identification of a binding site motif will improve the selection of gene targets. Determining the protein concentration to be used in the assay will depend on the non-specific binding activity of the protein, which may be assessed by EMSA using randomly chosen DNA substrates. Lower concentrations work better for those proteins with high nonspecific binding activity. The reliability of a blind DAP-chip may be improved by performing replicate assays with different protein concentrations. For RRs with few targets, just two or even a single assay may be sufficient to generate a target list which may be validated using subsequent EMSAs. The DAP-chip data for such RRs usually show a clear jump in the log₂R scores or cutoff_p values beyond the initial few peaks. For RRs with several targets, data from three replicates may be analyzed to generate a list of common peaks, some of which may be selected for EMSA validation.

The ability to predict binding site motifs based on the DAP-chip targets adds to the value of this method and vastly increases the information gained. EMSA again provides an efficient protocol to experimentally verify the predictions. Software scripts in Perl or other programming language may be used to rapidly search through available genome sequences. The results will identify both orthologous target genes as well as target genes uniquely regulated in the genomes searched. In the representative result shown here, three of the four upstream target regions identified for DVU3023 have annotated gene functions related to lactate uptake and oxidation. Additionally since the binding site motif for DVU3023 was validated, other genomes could be searched for similar motif sites. Together the results indicate that the DVU3022-3023 TCS is well conserved in the *Desulfovibrio* and other related sulfate-reducing bacteria, and that it regulates genes for lactate transport and oxidation to acetate. Since lactate is the primary carbon and electron source used by these organisms, DVU3023 likely plays a key role in their physiology. Among the top DAP-chip peaks for RR DVU3023, there was also an intergenic region. Although a binding site motif was not found in this region, the presence of a predicted sigma-54-dependent promoter suggests that there may be an unidentified *orf* or sRNA encoded, and that it could be a true target for DVU3023. This finding highlights the value of the experimental DAP-chip approach combined with binding site predictions as opposed to determining target sites based on computational predictions alone.

Similar *in vitro* methods such as the one described here have only been used in a few cases¹⁵⁻¹⁷ and very rarely for a previously unstudied regulator. The optimization EMSA and qPCR steps prior to the array hybridization will substantially aid in analyzing the results when the method is to be used for a novel regulator. If array printing becomes an impediment, the DAP steps may be combined with next generation sequencing to obtain binding sites^{18,19}. As more array-based methods become substituted with sequencing based strategies, such future adaptation of the method circumvents the need to design custom tiling microarrays for the organism of interest. ChIP-Seq technologies result in greater sensitivity and specificity of detection of peaks when compared to ChIP-chip methods, and are also rapidly becoming more cost-effective²⁰. Although this article focused on two component response regulators, this method can be used for any prokaryotic or eukaryotic transcription factor¹⁵. Nucleosome occupancies often have an effect on targets that are regulated and comparing the target binding sites for eukaryotic transcription factors by *in vivo* and *in vitro* analyses can reveal these effects²¹.

Disclosures

The authors have no conflict of interest to disclose.

Acknowledgements

We thank Amy Chen for her help in preparing for the video shoot and for demonstrating the technique. This work conducted by ENIGMA: Ecosystems and Networks Integrated with Genes and Molecular Assemblies (<http://enigma.lbl.gov>), a Scientific Focus Area Program at Lawrence Berkeley National Laboratory, was supported by the Office of Science, Office of Biological and Environmental Research, of the U. S. Department of Energy under Contract No. DE-AC02-05CH11231.

References

- Galperin, M. Y. Diversity of structure and function of response regulator output domains. *Curr Opin Microbiol.* **13**, 150-159, doi:S1369-5274(10)00010-X [pii]10.1016/j.mib.2010.01.005 (2010).
- Casino, P., Rubio, V., & Marina, A. The mechanism of signal transduction by two-component systems. *Curr Opin Struct Biol.* **20**, 763-771, doi:S0959-440X(10)00145-4 [pii]10.1016/j.sbi.2010.09.010 (2010).
- Aparicio, O., Geisberg, J. V., & Struhl, K. Chromatin immunoprecipitation for determining the association of proteins with specific genomic sequences *in vivo*. *Curr Protoc Cell Biol.* **Chapter 17**, Unit 17 17, doi:10.1002/0471143030.cb1707s23 (2004).
- McCleary, W. R., & Stock, J. B. Acetyl phosphate and the activation of two-component response regulators. *J Biol Chem.* **269**, 31567-31572 (1994).
- Novichkov, P. S. *et al.* RegPrecise web services interface: programmatic access to the transcriptional regulatory interactions in bacteria reconstructed by comparative genomics. *Nucleic Acids Res.* **40**, W604-608, doi:gks562 [pii]10.1093/nar/gks562 (2012).
- Studholme, D. J., Buck, M., & Nixon, T. Identification of potential sigma(N)-dependent promoters in bacterial genomes. *Microbiology.* **146 Pt 12**, 3021-3023 (2000).
- Rajeev, L. *et al.* Systematic mapping of two component response regulators to gene targets in a model sulfate reducing bacterium. *Genome Biol.* **12**, R99, doi:gb-2011-12-10-r99 [pii]10.1186/gb-2011-12-10-r99 (2011).
- Crooks, G. E., Hon, G., Chandonia, J. M., & Brenner, S. E. WebLogo: a sequence logo generator. *Genome Res.* **14**, 1188-1190, doi:10.1101/gr.849004 [doi]14/6/1188 [pii] (2004).
- Barbieri, C. M., Mack, T. R., Robinson, V. L., Miller, M. T., & Stock, A. M. Regulation of response regulator autophosphorylation through interdomain contacts. *J Biol Chem.* **285**, 32325-32335, doi:M110.157164 [pii]10.1074/jbc.M110.157164 (2010).

10. Bourret, R. B. Receiver domain structure and function in response regulator proteins. *Curr Opin Microbiol.* **13**, 142-149, doi:S1369-5274(10)00020-2 [pii]10.1016/j.mib.2010.01.015 (2010).
11. Gao, R., & Stock, A. M. Molecular strategies for phosphorylation-mediated regulation of response regulator activity. *Curr Opin Microbiol.* **13**, 160-167, doi:S1369-5274(09)00188-X [pii]10.1016/j.mib.2009.12.009 (2010).
12. Hung, D. C. *et al.* Oligomerization of the response regulator ComE from *Streptococcus mutans* is affected by phosphorylation. *J Bacteriol.* **194**, 1127-1135, doi:JB.06565-11 [pii]10.1128/JB.06565-11 (2012).
13. Ladds, J. C. *et al.* The response regulator Spo0A from *Bacillus subtilis* is efficiently phosphorylated in *Escherichia coli*. *FEMS Microbiol Lett.* **223**, 153-157, doi:S0378109703003215 [pii] (2003).
14. Wen, Y., Feng, J., Scott, D. R., Marcus, E. A., & Sachs, G. Involvement of the HP0165-HP0166 two-component system in expression of some acidic-pH-upregulated genes of *Helicobacter pylori*. *J Bacteriol.* **188**, 1750-1761, doi:188/5/1750 [pii]10.1128/JB.188.5.1750-1761.2006 (2006).
15. Liu, X., Noll, D. M., Lieb, J. D., & Clarke, N. D. DIP-chip: rapid and accurate determination of DNA-binding specificity. *Genome Res.* **15**, 421-427, doi:gr.3256505 [pii]10.1101/gr.3256505 (2005).
16. Gossett, A. J., & Lieb, J. D. DNA Immunoprecipitation (DIP) for the Determination of DNA-Binding Specificity. *CSH Protoc.* **2008**, pdb prot4972 (2008).
17. Mascher, T. *et al.* The *Streptococcus pneumoniae* cia regulon: CiaR target sites and transcription profile analysis. *J Bacteriol.* **185**, 60-70 (2003).
18. Park, P. J. ChIP-seq: advantages and challenges of a maturing technology. *Nat Rev Genet.* **10**, 669-680, doi:nrg2641 [pii]10.1038/nrg2641 (2009).
19. Lefrancois, P., Zheng, W., & Snyder, M. ChIP-Seq using high-throughput DNA sequencing for genome-wide identification of transcription factor binding sites. *Methods Enzymol.* **470**, 77-104, doi:S0076-6879(10)70004-5 [pii]10.1016/S0076-6879(10)70004-5 (2010).
20. Ho, J. W. *et al.* ChIP-chip versus ChIP-seq: lessons for experimental design and data analysis. *BMC Genomics.* **12**, 134, doi:1471-2164-12-134 [pii]10.1186/1471-2164-12-134 (2011).
21. Liu, X., Lee, C. K., Granek, J. A., Clarke, N. D., & Lieb, J. D. Whole-genome comparison of Leu3 binding *in vitro* and *in vivo* reveals the importance of nucleosome occupancy in target site selection. *Genome Res.* **16**, 1517-1528, doi:gr.5655606 [pii]10.1101/gr.5655606 (2006).