doi:10.1093/nar/gkn477

A high-throughput percentage-of-binding strategy to measure binding energies in DNA–protein interactions: application to genome-scale site discovery

Xiaohu Wang[1], Haichun Gao[2,3], Yufeng Shen[4], George M. Weinstock[4], Jizhong Zhou[2] and Timothy Palzkill[1,*]

[1]Department of Molecular Virology & Microbiology, Baylor College of Medicine, Houston, TX 77030, [2]Institute for Environmental Genomics, Department of Botany and Microbiology, University of Oklahoma, Norman, OK 73019, USA, [3]College of Life Sciences, Zhejiang University, Hangzhou, Zhejiang 310029, P. R. China and [4]Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX 77030, USA

Received May 8, 2008; Revised June 11, 2008; Accepted July 8, 2008

ABSTRACT

Quantifying the binding energy in DNA–protein interactions is of critical importance to understand transcriptional regulation. Based on a simple computational model, this study describes a high-throughput percentage-of-binding strategy to measure the binding energy in DNA–protein interactions between the *Shewanella oneidensis* ArcA two-component transcription factor protein and a systematic set of mutants in an ArcA-P (phosphorylated ArcA) binding site. The binding energies corresponding to each of the 4 nt at each position in the 15-bp binding site were used to construct a position-specific energy matrix (PEM) that allowed a reliable prediction of ArcA-P binding sites not only in *Shewanella* but also in related bacterial genomes.

INTRODUCTION

Transcription factor proteins can recognize and bind a collection of similar DNA sequences with various affinities (1). This degenerate binding ability renders cells capable of controlling thousands of genes with relatively few regulatory proteins (2). Degenerate binding, however, poses a significant challenge for understanding the mechanism of transcriptional regulation, especially in terms of identifying new binding sites (i.e. site discovery).

During the past two decades, numerous computational site-discovery methods have been developed. However, it is still a challenge to predict transcription factor binding sites (3–5). One explanation for the difficulty is that computational predictions are usually based on sequence conservation of transcription factor binding sites rather than thermodynamic parameters that govern DNA–protein interactions. Experimental data, such as that obtained from footprinting assays and transcriptional profiling, can greatly increase the accuracy of computational predictions (6). Obtaining sufficient high quality experimental data, however, is a work-intensive task. For high-throughput experimental methods such as various ChIP-based approaches (7–10), a high-quality antibody and multiple experimental steps are usually necessary.

Although the *in vivo* occupancy of *cis*-regulatory elements may be affected by many factors (11), the occurrence and strength of a DNA–protein interaction is ultimately determined by whether it is a thermodynamically favored reaction. Thus, measuring a DNA–protein binding constant and thereby the binding energy of an interaction represents a crucial step towards understanding transcriptional regulation. Although experimental approaches for measuring these thermodynamic parameters are well established, high-throughput methods have not yet been extensively developed. The few available medium- or high-throughput experimental methods, including SPR (surface plasmon resonance) (12), microwell-based assays (13), displacement of DNA binding dye (14), MITOMI (mechanically induced trapping of molecular interactions) (15,16) and competition assays (17–20) are limited by various factors, such as cost, sensitivity and special protein/DNA constructs. To date, the most commonly used experimental approach remains the time consuming curve-fitting method.

The goal of this study was to develop an effective general approach for a rapid and accurate genome-scale prediction of transcription factor binding sites using ArcA as a model system. The ArcA transcription factor belongs to the canonical ArcA/B two-component system in which ArcB is a membrane associated histidine kinase and

*To whom correspondence should be addressed. Tel: +1 713 798 5609; Fax: +1 713 798 7375; Email: timothyp@bcm.tmc.edu

ArcA is a downstream transcription response regulator (21). As a major oxygen response regulator, the ArcA protein is well conserved in many Gram-negative bacteria including *Shewanella oneidensis* MR-1, which is a model organism for bioremediation studies (22). Recent studies, however, indicate the ArcA protein may regulate a different set of genes in *S. oneidensis* than those regulated in *Escherichia coli* (23,24). In order to determine the sequence requirements for ArcA-P binding, systematic mutagenesis of an ArcA-P binding site and subsequent quantitation of binding energy of each mutant was performed. Mathematical modeling indicated that, in principle, the binding energy in DNA–protein interactions can be determined using a simple percentage-of-binding approach instead of curve fitting. By applying this method to the traditional electrophoretic mobility shift assay (EMSA) and a recently developed protein binding microarray (PBM) technology (25), the DNA sequence requirements and associated binding energies for the ArcA-P protein were systematically determined by a simple one-step binding assay and this experimental information was used to construct a position-specific energy matrix (PEM) for genome-scale prediction of binding sites.

MATERIALS AND METHODS

Computational model of DNA–protein interactions

In a DNA–protein interaction: [L] + [P]↔[LP], where [L], [P] and [LP] represent the concentration of free (or unbound) DNA ligand, free protein and the DNA–protein complex, respectively. If it is assumed that the pressure and temperature do not change in the binding reaction, the Gibbs binding energy ΔG can be then determined by Equation (1) and the dissociation constant K_d by Equation (2) when the binding reaction reaches equilibrium, in which R, T, x and $1-x$ represent the gas constant, the absolute temperature, the fraction of DNA ligand bound to protein and the fraction of free DNA ligand, respectively (26,27). By combining equations, the ΔG can be calculated according to Equation (3).

$$\Delta G = -\mathrm{R}T \bullet \ln K_{\mathrm{eq}} = -\mathrm{R}T \bullet \ln(1/K_{\mathrm{d}}) = \mathrm{R}T \bullet \ln K_{\mathrm{d}} \qquad 1$$

$$K_{\mathrm{d}} = [\mathrm{L}] \bullet [\mathrm{P}]/[\mathrm{LP}] = [\mathrm{P}] \bullet (1-x)/x \qquad 2$$

$$\begin{aligned}\Delta G &= \mathrm{R}T \bullet \ln\{[\mathrm{P}] \bullet (1-x)/x\}\\ &= \mathrm{R}T \bullet \{\ln[(1-x)/x] + \ln[\mathrm{P}]\}\end{aligned} \qquad 3$$

$$\Delta G_{\mathrm{Ref}} = \mathrm{R}T \bullet \{\ln[(1-x_{\mathrm{Ref}})/x_{\mathrm{Ref}}] + \ln[\mathrm{P}]_{\mathrm{Ref}}\} \qquad 4$$

$$\Delta G_1 = \mathrm{R}T \bullet \{\ln[(1-x_1)/x_1] + \ln[\mathrm{P}]_1\} \qquad 5$$

$$\begin{aligned}\Delta\Delta G &= \Delta G_1 - \Delta G_{\mathrm{Ref}}\\ &= \mathrm{R}T \bullet \{\ln[(1-x_1)/x_1] - \ln[(1-x_{\mathrm{Ref}})/x_{\mathrm{Ref}}]\}\\ &\quad + \mathrm{R}T \bullet \{\ln[\mathrm{P}]_1 - \ln[\mathrm{P}]_{\mathrm{Ref}}\}\\ &= \mathrm{R}T \bullet \Delta\ln[(1-x)/x] + \mathrm{RT} \bullet \Delta\ln[\mathrm{P}]\end{aligned} \qquad 6$$

$$\begin{aligned}\Delta\Delta G &= \mathrm{R}T \bullet \{\ln[(1-x_1)/x_1] - \ln[(1-x_{\mathrm{Ref}})/x_{\mathrm{Ref}}]\\ &= \mathrm{R}T \bullet \Delta\ln[(1-x)/x]\end{aligned} \qquad 7$$

$$f(x) = \frac{\mathrm{d}\{[(1-x)/x]\}}{\mathrm{d}\{\ln[\mathrm{P}]\}} = \frac{[P]}{x(x-1)} = \frac{[\mathrm{P}]_0 - x[\mathrm{L}]_0}{x(x-1)} \qquad 8$$

Based on Equation (3), the binding energy of any DNA–protein interaction, such as the ΔG_1 and ΔG_{Ref} reactions, can be determined using Equations (4) and (5), respectively. The relative binding energy ($\Delta\Delta G$) between ΔG_1 and ΔG_{Ref} reactions can then be calculated using Equation (6). If the free protein concentration is kept constant in these two binding reactions (i.e. $[\mathrm{P}]_1 = [\mathrm{P}]_{\mathrm{Ref}}$), Equation (6) can be simplified as Equation (7), where $\Delta\Delta G$ is determined by the percentage of DNA ligand bound. In actual binding reactions, it is very difficult to keep the free protein concentration constant. It is possible, however, to achieve an approximately constant protein concentration by performing assays with a protein concentration that is much higher than the DNA ligand concentration.

EMSA with systematically mutated SO1661 promoters

The purification of His-tagged *Shewanella* ArcA and *E. coli* ArcB$^{78\text{-}778}$ as well as the ArcA phosphorylation were performed as described previously (24,28,29). ArcA protein is labeled as ArcA-P$^{\mathrm{C}}$ or ArcA-P$^{\mathrm{E}}$ when phosphorylated by carbamoyl phosphate or *E. coli* ArcB$^{78\text{-}778}$, respectively. In this study, a total of 46 oligonucleotides of 48 bp each were synthesized. Forty-five of these primers contained a single mutation in the 15-bp ArcA-P binding site region in the SO1661 promoter (Figure 1A and Table 1). Radiolabeled promoters (144 bp each) were generated from these 46 primers by PCR amplification with a common P^{33} 5′-end labeled SO1661-328 primer (5'-CCACACCATACCGATAAAGAAGC). The interaction of each radiolabeled DNA with ArcA-P (phosphorylated ArcA) was then tested using EMSAs containing ~100–250 fmol (~10–20 nM) labeled probe and 1.5 µM ArcA-P$^{\mathrm{E}}$ as previously described (24) in which the active amount of ArcA-P was estimated to be 100 nM (Supplementary Method 1). These binding assays were repeated three times and the fraction of promoter DNA bound to the ArcA-P protein was quantified by measuring the density of both shifted and nonshifted DNA bands using ImageQuant TL (GE Healthcare Life Sciences, Piscataway, NJ, USA).

Protein binding microarray

For protein binding microarray studies, a series of 48-bp mutant promoters were constructed by synthesis of oligonucleotide pairs that were annealed rather than using PCR for second strand synthesis. The annealed promoter DNAs each contained a single mutation in the conserved 15-bp ArcA-P site (Figure 1A and Table 1). In addition, an amine group was also synthesized onto the 5′-end of one of the paired oligonucleotides. These promoter DNAs were printed and covalently immobilized onto Codelink activated slides in 50 mM PBS (pH 8.5) at 50 pmol/µl,

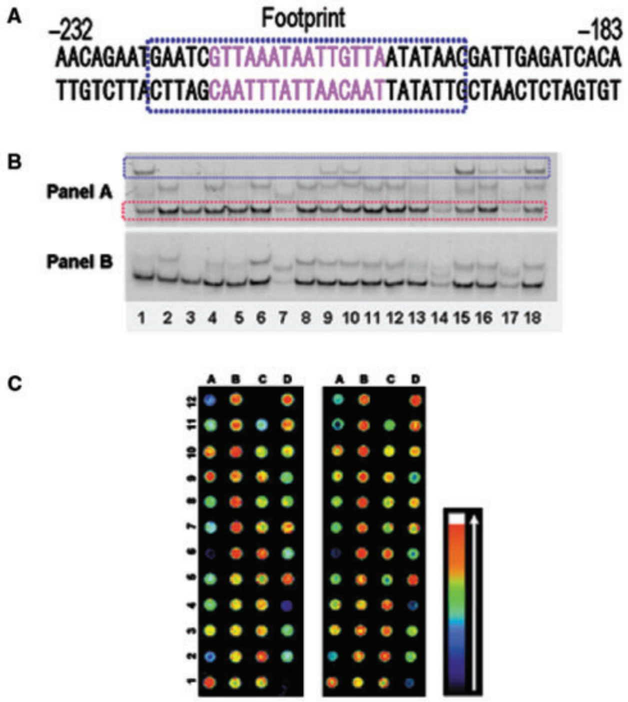


Figure 1. (**A**) SO1661 promoter DNA. The blue dotted square indicates the ArcA-P binding site as identified by DNA footprinting. The conserved 15-bp motif is shown as pink letters. (**B**) Representative EMSAs with SO1661 mutant promoters. The EMSAs were performed with 10–20 nM radiolabeled DNA ligand and 100 nM of the active form of ArcA-P^E (A). DNA ligand binding reactions lacking ArcA-P^E protein were analyzed as a negative control (B). Lane 1: SO1661^{wt} promoter. Lanes 2–18: SO1661-1 to SO1661-17 promoters, respectively (Table 1). The blue and red dotted rectangles indicate the shifted DNA–protein complexes and free unshifted DNA ligands, respectively. (**C**) PBM results with ArcA-P^C. The results of two replica PBM experiments are shown. Different colors represent relative binding strength, which is indicated by the white arrow. Red or blue indicates a strong or weak binding signal, respectively. White indicates the binding signal exceeds the upper detection limit of the method used in this study. Spot A1: SO1661^{wt} promoter. Spots A2–D12: SO1661-1 to SO1661-47 promoters, respectively.

(Amersham Biosciences, Piscataway, NJ, USA) according to the manufacturer's instructions.

The PBM experiment was performed by first incubating the microarray slides with blocking buffer (10 mM Tris/HCl, 150 mM NaCl, 5 mM $MgCl_2$, 0.05% Tween-20, 5% milk, pH 7.4) for 30 min at room temperature, and then rinsing briefly with 1 × TBS (10 mM Tris/HCl, 150 mM NaCl, 5 mM $MgCl_2$, pH 7.4). The on-slide DNA–protein interaction was initiated by covering the slides with 300 µl binding solution consisting of 100 mM Tris/HCl (pH 7.4), 20 mM KCl, 10 mM $MgCl_2$, 2 mM DTT, 10% glycerol, 0.1 µg/µl poly(dI·dC) and 2 µM carbamyol phosphate phosphorylated ArcA protein [895 nM active protein concentration (Supplementary Method 1)]. The binding reaction was performed at room temperature for 1 h. After washing off unbound protein, the microarray slides were incubated with an anti-His-tag antibody for 1 h at room temperature. The anti-His-tag antibody was purchased from Qiagen, Valencia, CA, USA (Cat# 34660) and diluted to 1:1200 in blocking buffer. The unbound anti-His-tag antibody was then washed off and the slides were incubated with Cy5-conjugated secondary antibody for 1 h at room temperature. The secondary antibody was purchased from Chemicon, Temecula, CA, USA (Cat# AP160s) and diluted 1:1200 in blocking buffer. After washing away unbound secondary antibody, the slides were dried and quantified using a microarray scanner. To reduce errors in data analysis, the signal intensity of each spot was normalized to the average signal intensity of all the 48 spots. In total, 27 binding replicas were performed with the promoter DNAs.

Genome-scale ArcA-P binding site discovery method

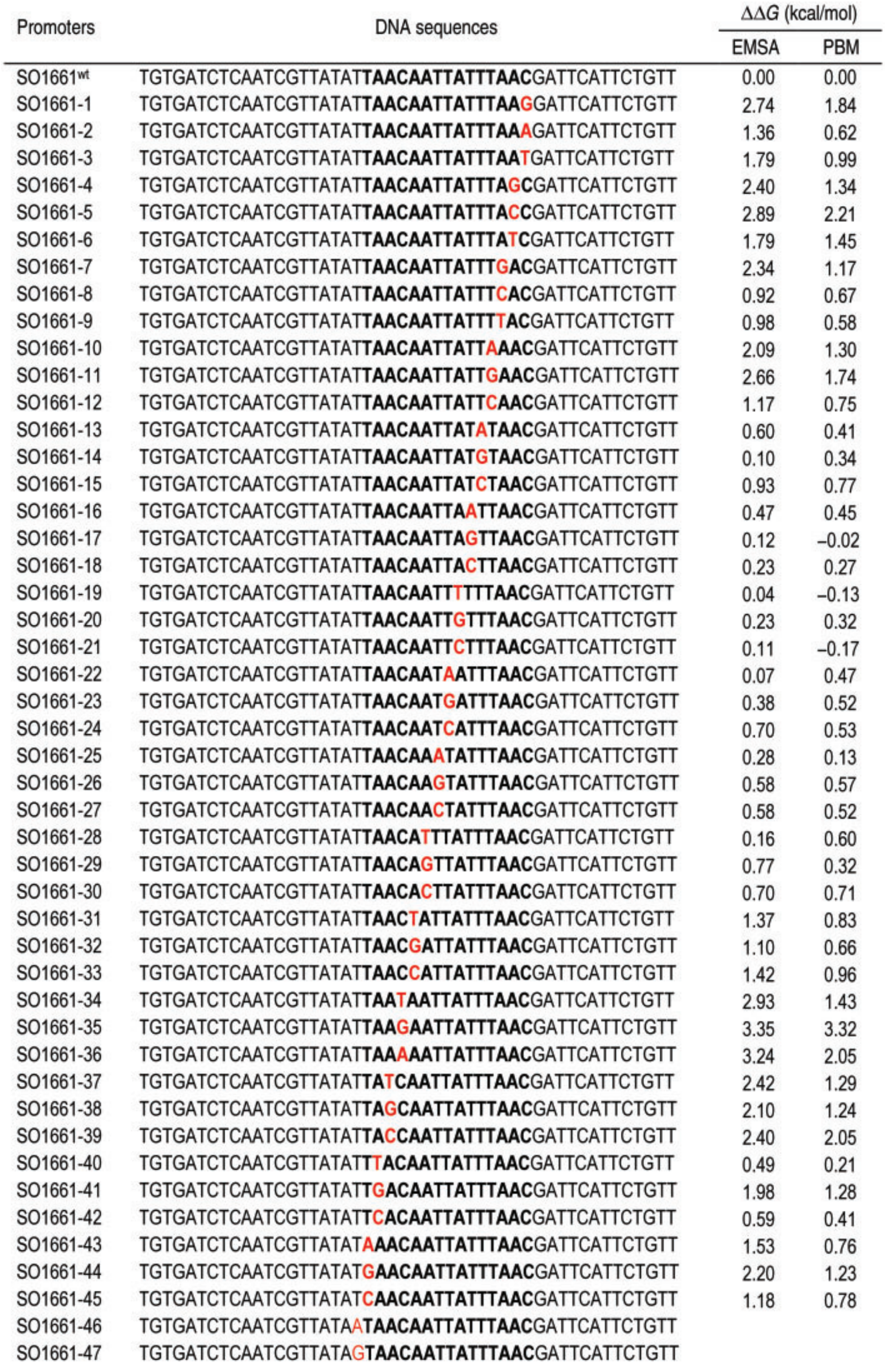
In this study, the upstream intergenic DNA for each *S. oneidensis* MR-1 ORF (open reading frame) (www.ncbi.nlm.nih.gov) including the first 100 bp of coding sequence was first obtained. These DNA sequences were then scanned with a sliding 15-bp DNA motif window, and the scores of each motif were calculated based on the energy-based ArcA-P PEM (15 × 4) (Table 2) with the assumption that each base contributes independently to the total score of the 15-bp motif (1). These scores represent the predicted binding energies for each DNA site with ArcA-P. The motif with the lowest score (most favorable binding energy) was selected for each intergenic DNA region and ranked according to their $\Delta\Delta G$ scores (Supplementary Table 1). Using the same sliding window approach with the same energy matrix, potential ArcA-P binding sites were also predicted in the promoter regions (intergenic region + the first 100-bp coding sequence) of *E. coli* K12 MG1655 and *Haemophilus influenzae* Rd KW20 genomes (Supplementary Tables 2 and 3).

RESULTS

Model for determining binding energies of mutant promoter DNA to ArcA-P

According to the computational model implemented in Equation (7) (see Materials and Methods section), the relative binding energy difference ($\Delta\Delta G$) for wild-type versus a mutant DNA with respect to a DNA–protein interaction can be determined by performing two separate binding reactions and measuring the fraction of DNA ligand bound (percentage-of-binding) for each reaction at equilibrium. This model assumes that the concentration of free active protein ([P]) remains constant in both binding reactions. In an actual experiment, [P] varies to differing extents according to the binding conditions. Based on Equation (3) or (6), the overall $\Delta\Delta G$ is determined solely by two variable factors, $\Delta\ln[(1-x)/x]$ and $\Delta\ln[\mathrm{P}]$. This suggests that the error associated with using Equation (7) to determine $\Delta\Delta G$ can be estimated according to the relative weights of $\Delta\ln[(1-x)/x]$ and $\Delta\ln[\mathrm{P}]$. The ratio of $\Delta\ln[(1-x)/x]$ versus $\Delta\ln[\mathrm{P}]$, is shown as the function $f(x)$ in Equation (8), in which [P], $[\mathrm{P}]_0$ or $[\mathrm{L}]_0$ represent the concentration of free protein, the total input protein or total input DNA ligand, respectively. The plot generated using Equation (8) is shown in Supplementary Figure 1. The results indicate that the minimal value of $f(x)$ is 9.9, 17.9 or 37.9, if $[\mathrm{P}]_0$ is 3, 5 or 10 times that of $[\mathrm{L}]_0$,

Table 1. Oligonucleotide sequences and binding energies of SO1661^{wt} and mutant promoter DNAs



Promoters	DNA sequences	$\Delta\Delta G$ (kcal/mol)	
		EMSA	PBM
SO1661^{wt}	TGTGATCTCAATCGTTATAT**TAACAATTATTTAAC**GATTCATTCTGTT	0.00	0.00
SO1661-1	TGTGATCTCAATCGTTATAT**TAACAATTATTTAAG**GATTCATTCTGTT	2.74	1.84
SO1661-2	TGTGATCTCAATCGTTATAT**TAACAATTATTTAAA**GATTCATTCTGTT	1.36	0.62
SO1661-3	TGTGATCTCAATCGTTATAT**TAACAATTATTTAAT**GATTCATTCTGTT	1.79	0.99
SO1661-4	TGTGATCTCAATCGTTATAT**TAACAATTATTTAGC**GATTCATTCTGTT	2.40	1.34
SO1661-5	TGTGATCTCAATCGTTATAT**TAACAATTATTTACC**GATTCATTCTGTT	2.89	2.21
SO1661-6	TGTGATCTCAATCGTTATAT**TAACAATTATTTATC**GATTCATTCTGTT	1.79	1.45
SO1661-7	TGTGATCTCAATCGTTATAT**TAACAATTATTTGAC**GATTCATTCTGTT	2.34	1.17
SO1661-8	TGTGATCTCAATCGTTATAT**TAACAATTATTTCAC**GATTCATTCTGTT	0.92	0.67
SO1661-9	TGTGATCTCAATCGTTATAT**TAACAATTATTTTAC**GATTCATTCTGTT	0.98	0.58
SO1661-10	TGTGATCTCAATCGTTATAT**TAACAATTATTAAAC**GATTCATTCTGTT	2.09	1.30
SO1661-11	TGTGATCTCAATCGTTATAT**TAACAATTATTGAAC**GATTCATTCTGTT	2.66	1.74
SO1661-12	TGTGATCTCAATCGTTATAT**TAACAATTATTCAAC**GATTCATTCTGTT	1.17	0.75
SO1661-13	TGTGATCTCAATCGTTATAT**TAACAATTATATAAC**GATTCATTCTGTT	0.60	0.41
SO1661-14	TGTGATCTCAATCGTTATAT**TAACAATTATGTAAC**GATTCATTCTGTT	0.10	0.34
SO1661-15	TGTGATCTCAATCGTTATAT**TAACAATTATCTAAC**GATTCATTCTGTT	0.93	0.77
SO1661-16	TGTGATCTCAATCGTTATAT**TAACAATTAATTAAC**GATTCATTCTGTT	0.47	0.45
SO1661-17	TGTGATCTCAATCGTTATAT**TAACAATTAGTTAAC**GATTCATTCTGTT	0.12	−0.02
SO1661-18	TGTGATCTCAATCGTTATAT**TAACAATTACTTAAC**GATTCATTCTGTT	0.23	0.27
SO1661-19	TGTGATCTCAATCGTTATAT**TAACAATTTTTTAAC**GATTCATTCTGTT	0.04	−0.13
SO1661-20	TGTGATCTCAATCGTTATAT**TAACAATTGTTTAAC**GATTCATTCTGTT	0.23	0.32
SO1661-21	TGTGATCTCAATCGTTATAT**TAACAATTCTTTAAC**GATTCATTCTGTT	0.11	−0.17
SO1661-22	TGTGATCTCAATCGTTATAT**TAACAATAATTTAAC**GATTCATTCTGTT	0.07	0.47
SO1661-23	TGTGATCTCAATCGTTATAT**TAACAATGATTTAAC**GATTCATTCTGTT	0.38	0.52
SO1661-24	TGTGATCTCAATCGTTATAT**TAACAATCATTTAAC**GATTCATTCTGTT	0.70	0.53
SO1661-25	TGTGATCTCAATCGTTATAT**TAACAAATATTTAAC**GATTCATTCTGTT	0.28	0.13
SO1661-26	TGTGATCTCAATCGTTATAT**TAACAAGTATTTAAC**GATTCATTCTGTT	0.58	0.57
SO1661-27	TGTGATCTCAATCGTTATAT**TAACAACTATTTAAC**GATTCATTCTGTT	0.58	0.52
SO1661-28	TGTGATCTCAATCGTTATAT**TAACATTTATTTAAC**GATTCATTCTGTT	0.16	0.60
SO1661-29	TGTGATCTCAATCGTTATAT**TAACAGTTATTTAAC**GATTCATTCTGTT	0.77	0.32
SO1661-30	TGTGATCTCAATCGTTATAT**TAACACTTATTTAAC**GATTCATTCTGTT	0.70	0.71
SO1661-31	TGTGATCTCAATCGTTATAT**TAACTATTATTTAAC**GATTCATTCTGTT	1.37	0.83
SO1661-32	TGTGATCTCAATCGTTATAT**TAACGATTATTTAAC**GATTCATTCTGTT	1.10	0.66
SO1661-33	TGTGATCTCAATCGTTATAT**TAACCATTATTTAAC**GATTCATTCTGTT	1.42	0.96
SO1661-34	TGTGATCTCAATCGTTATAT**TAATAATTATTTAAC**GATTCATTCTGTT	2.93	1.43
SO1661-35	TGTGATCTCAATCGTTATAT**TAAGAATTATTTAAC**GATTCATTCTGTT	3.35	3.32
SO1661-36	TGTGATCTCAATCGTTATAT**TAAAAATTATTTAAC**GATTCATTCTGTT	3.24	2.05
SO1661-37	TGTGATCTCAATCGTTATAT**TATCAATTATTTAAC**GATTCATTCTGTT	2.42	1.29
SO1661-38	TGTGATCTCAATCGTTATAT**TAGCAATTATTTAAC**GATTCATTCTGTT	2.10	1.24
SO1661-39	TGTGATCTCAATCGTTATAT**TACCAATTATTTAAC**GATTCATTCTGTT	2.40	2.05
SO1661-40	TGTGATCTCAATCGTTATAT**TTACAATTATTTAAC**GATTCATTCTGTT	0.49	0.21
SO1661-41	TGTGATCTCAATCGTTATAT**TGACAATTATTTAAC**GATTCATTCTGTT	1.98	1.28
SO1661-42	TGTGATCTCAATCGTTATAT**TCACAATTATTTAAC**GATTCATTCTGTT	0.59	0.41
SO1661-43	TGTGATCTCAATCGTTATAT**AAACAATTATTTAAC**GATTCATTCTGTT	1.53	0.76
SO1661-44	TGTGATCTCAATCGTTATAT**GAACAATTATTTAAC**GATTCATTCTGTT	2.20	1.23
SO1661-45	TGTGATCTCAATCGTTATAT**CAACAATTATTTAAC**GATTCATTCTGTT	1.18	0.78
SO1661-46	TGTGATCTCAATCGTTATAA**TAACAATTATTTAAC**GATTCATTCTGTT		
SO1661-47	TGTGATCTCAATCGTTATAG**TAACAATTATTTAAC**GATTCATTCTGTT		

The 15 bp ArcA-P binding motif is shown in bold letters and mutant nucleotides are indicated by red letters. The oligonucleotide sequences shown in this table are the complement (bottom strand) of the sequence shown in Figure 1A.

respectively. Therefore, when $[P]_0$ is 3, 5 or 10 times that of $[L]_0$, the weight of $\Delta\ln[P]$ in the estimation of total $\Delta\Delta G$ is less than 9.2%, 5.3% or 2.6%, respectively. These data suggest that $\Delta\Delta G$ can be determined accurately using Equation (7) with a ratio of $[P]_0/[L]_0$ above 5 (error $< 5.3\%$).

Relative binding energies determined using comparative EMSAs (EMSA-$\Delta\Delta G$)

The SO1661 promoter has been shown to be under the direct control of ArcA (24). Based on footprinting assays and mutational analyses (data not shown), the ArcA-P binding site within the SO1661 promoter was determined to be a 15-bp DNA motif (Figure 1A). To understand the role of ArcA in *Shewanella*, each position of the 15-bp ArcA-P binding motif within the SO1661 promoter was systematically mutated and the effect on the binding of ArcA-P^E [ArcA protein phosphorylated by *E. coli* ArcB protein is labeled ArcA-P^E (see Materials and Methods section)] was examined by EMSAs (Figure 1B). In these EMSAs, the molar ratio of active ArcA-P^E versus DNA ligand was kept above 5 (see Materials and methods section and Supplementary Method 1). The percentage of DNA bound by ArcA-P^E [equal to x in Equation (7)] of wild-type and various mutant SO1661 promoters was determined by measuring the band intensity of shifted and nonshifted DNA. The fraction DNA bound data was then used to determine the relative binding energy of the mutant promoters (i. e. $\Delta G_{\mathrm{mu}} - \Delta G_{\mathrm{wt}}$ which will be referred to as $\Delta\Delta G$ hereafter) using Equation (7). These EMSA-$\Delta\Delta G$ values (Table 1) represent the contribution relative to wild-type of each nucleotide at a given position within the 15-bp binding sequence to the total binding energy (ΔG). A PEM was generated by placing the $\Delta\Delta G$ values of the promoter DNA with each mutant nucleotide at the corresponding position within the 15-bp DNA motif (Table 2). The information in the matrix can be summarized as a sequence logo using the enoLOGOS program (30) (Figure 2). The results indicate that two repeating GTTA units are very important for binding ArcA-P (Figure 2). This pattern is similar in sequence but significantly different in position weighting to a consensus revealed by searching for a common motif in 11 *E. coli* ArcA-P interacting promoters (31). The importance of both GTTA sites may be related to the fact that the active form of ArcA-P is a dimer (32).

Relative binding energies determined using PBM assays (PBM-$\Delta\Delta G$)

PBM technology is a recently developed high-throughput method to study DNA–protein interactions (25). To test if Equation (7) can also be used to determine $\Delta\Delta G$ in microarray-based DNA–protein interaction measurements, SO1661 promoter microarrays were generated using a series of synthesized 48-bp promoter DNAs (Table 1). The PBM binding reactions were performed with either ArcA-P^C [ArcA protein phosphorylated by carbamyol phosphate is labeled ArcA-P^C (see Materials and Methods section)] or ArcA-P^E. However, the results with ArcA-P^E were not as reproducible, possibly due to the low efficiency of *E. coli* ArcB$^{78\text{-}778}$ in phosphorylating the *Shewanella* ArcA protein (data not shown) and therefore the ArcA-P^E results were not used for further analysis. The PBM results indicated that ArcA-P^C exhibited varied binding affinities with different mutant promoters consistent with the results from the EMSAs (Figure 1C), while the unphosphorylated ArcA protein did not exhibit detectable DNA binding activity (data not shown).

The percentage of binding values for various mutant promoters in the PBM assays were determined indirectly by comparing their signal intensity relative to SO1661$^{\mathrm{wt}}$ promoter DNA signal intensity and the percentage of SO1661$^{\mathrm{wt}}$ DNA bound in an EMSA performed under identical conditions (Supplementary Method 2). The $\Delta\Delta G$ of different mutant promoters relative to wild-type was then calculated using Equation (7) (Table 1). With these PBM-$\Delta\Delta G$ scores, an energy-based sequence logo was created for ArcA-P^C using enoLOGOS (Figure 2). The sequence logos generated using $\Delta\Delta G$ values determined by EMS and PBM assays are similar, suggesting the percentage of promoter binding approach can be used to estimate binding energies with either assay.

Table 2. The position energy matrix (PEM) of *Shewanella* ArcA-P (unit: kcal/mol)

Position	A	C	G	T
1	1.79	2.74	0.00	1.35
2	1.79	2.40	2.89	0.00
3	0.98	2.34	0.92	0.00
4	0.00	2.65	1.17	2.09
5	0.00	0.10	0.93	0.60
6	0.00	0.12	0.23	0.47
7	0.04	0.23	0.11	0.00
8	0.00	0.38	0.70	0.07
9	0.00	0.58	0.58	0.28
10	0.16	0.77	0.70	0.00
11	1.37	1.10	1.42	0.00
12	2.93	3.35	0.00	3.24
13	2.42	2.10	2.40	0.00
14	0.49	1.98	0.59	0.00
15	0.00	2.20	1.18	1.53

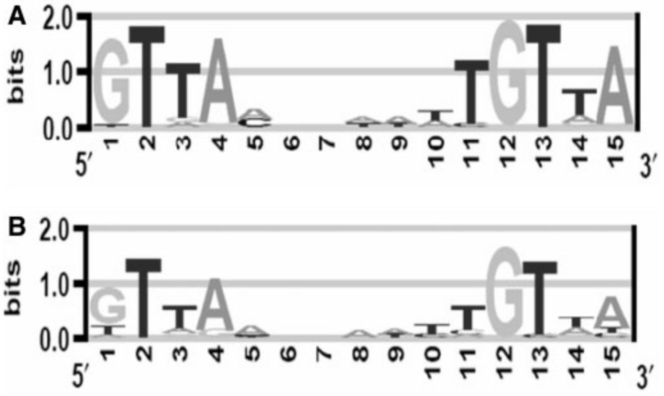


Figure 2. Sequence logo of *Shewanella* ArcA-P. Sequence logos generated by EMSAs with ArcA-P^E (**A**) or by PBM with ArcA-P^C (**B**). Both logos were created by enoLOGOS[30] based on relative binding energies ($\Delta\Delta G$) of the mutants.

Binding energies determined using a curve-fitting method (Curve-ΔG)

In order to validate the binding energy values obtained by the percentage of binding approach, several mutant SO1661 promoter DNAs (48-bp synthesized DNAs) were selected and their binding constants (K_d) with the ArcA-P protein were determined by a competitive EMSA using a curve-fitting method (33). In these assays, the input

ArcA-P was held constant and the binding of labeled promoters (radioligand) to ArcA-P was subjected to competition with various amounts of unlabeled promoter DNA (Supplementary Method 3). By fitting the EMS data to a sigmoidal dose response curve (Figure 3A and B), the IC_{50} value for $SO1661^{wt}$ was determined to be 234 ± 24 nM for ArcA-P^C and 216 ± 21 nM for ArcA-P^E, with corresponding K_d values of 154 ± 24 nM and 136 ± 21 nM, respectively (according to the formula $IC_{50} = K_d + [\text{radioligand}]$) (33) (Table 3). The K_d values of several selected mutant promoters including SO1661-15, SO1661-17, SO1661-19 and SO1661-20 were also determined using the same method with ArcA-P^C (Table 3). The binding energy (Curve fitting-ΔG) of these promoters was then determined from the K_d values according to Equation (1) (Table 3).

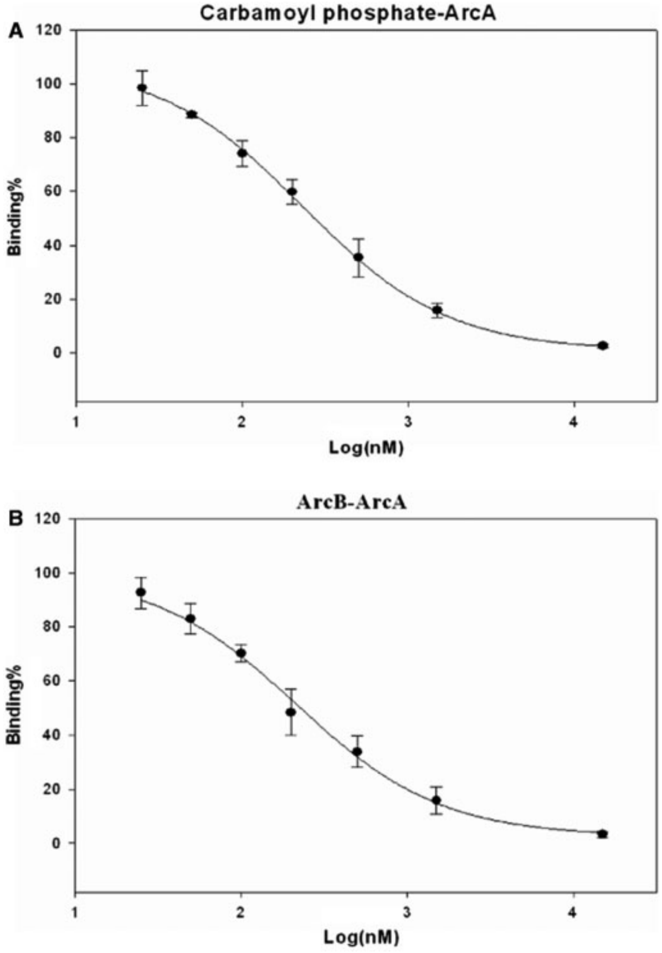


Figure 3. Competition binding curve of the $SO1661^{wt}$ promoter with (**A**) ArcA-P^C and (**B**) ArcA-P^E. The *x*-axis indicates the log concentration of the unlabeled probe in nanomolar and the *y*-axis indicates the percentage of binding relative to the maximal binding signal.

Comparison of the thermodynamic parameters ($\Delta\Delta G$, ΔG or K_d) determined using EMSA, PBM and curve-fitting methods

In order to evaluate the relationship between the $\Delta\Delta G$ values determined by EMSA versus the PBM methods, the values obtained for different mutant promoters were compared by linear regression (Figure 4, solid line). The resulting Pearson correlation coefficient is 0.97 (Figure 4), indicating the results are in close agreement in terms of the trend of the $\Delta\Delta G$ values, i.e. the methods give very similar results on the relative importance of a position. Consistent with this finding, the sequence logos generated using the EMSA and PBM data are also very similar (Figure 2). However, the EMSA-$\Delta\Delta G$ values are usually higher than the corresponding PBM-$\Delta\Delta G$ values as indicated by the position of the data points in Figure 4 relative to the dotted line that depicts a perfect correlation between the absolute values of $\Delta\Delta G$. A possible explanation is that the ratio of $[P]_0/[L]_0$ in the EMSAs was relatively low (5–10 : 1) and the larger $\Delta\Delta G$ values may be a result of neglecting the contribution of ΔlnP. In addition, the EMSA values in Table 1 were determined using $ArcA^E$ and the PBM values were determined using $ArcA^C$. To test these possibilities, several mutant promoters (48-bp synthesized DNAs) were selected (Table 3) and their interaction with ArcA-P^C was determined using the same comparative EMSA but at an increased ratio of $[P]_0/[L]_0$ (~100:1), and the resulting EMSA-$\Delta\Delta G$ values agree more closely with the corresponding PBM-$\Delta\Delta G$ values (Table 3).

Table 3. Binding affinities of selected promoter DNAs

Promoter	$\Delta\Delta G$ (kcal/mol)		K_d (nM)			ΔG (kcal/mol)		
	EMSA	PBM	Curve fitting	EMSA	PBM	Curve fitting	EMSA	PBM
$SO1661^{wt}$	0.00	0.00	154 ± 24	–	–	−8.98 ± 0.08	–	–
SO1661-12	0.87	0.75		711 ± 111	573 ± 89			
SO1661-13	0.51	0.41		376 ± 59	313 ± 49			
SO1661-14	0.29	0.34		257 ± 40	280 ± 44			
SO1661-15	0.55	0.77	390 ± 80	401 ± 63	591 ± 81	−8.45 ± 0.11	−8.43 ± 0.08	−8.21 ± 0.07
SO1661-17	0.37	−0.02	131 ± 39	291 ± 45	150 ± 23	−9.08 ± 0.13	−8.62 ± 0.08	−9.00 ± 0.08
SO1661-18	0.33	0.27		276 ± 43	248 ± 39			
SO1661-19	−0.50	−0.13	59 ± 23	64 ± 10	123 ± 19	−9.53 ± 0.19	−9.49 ± 0.08	−9.11 ± 0.08
SO1661-20	0.39	0.32	69 ± 2	303 ± 47	269 ± 42	−9.45 ± 0.02	−8.60 ± 0.08	−8.66 ± 0.08
SO1661-21	−0.18	−0.16		113 ± 18	115 ± 18			
SO1661-22	0.28	0.47		251 ± 39	347 ± 54			

The data in this table were determined using the 48 bp synthesized promoters and ArcA-P^C.

As stated earlier, the binding energy (Curve-ΔG) of the SO1661-16, SO1661-17, SO1661-19 and SO1661-20 mutant DNAs was determined by a curve-fitting method. As a comparison, the EMSA-$\Delta\Delta G$ and PBM-$\Delta\Delta G$ of these four mutant promoters was converted into binding energy (EMSA- or PBM-ΔG) according to Equation (6) ($\Delta\Delta G = \Delta G_{\mathrm{mu}} - \Delta G_{\mathrm{wt}}$) (Table 3). The average difference between the Curve-ΔG, EMSA-ΔG and PBM-ΔG of the four selected promoters is <5.7% (Table 3). Since ΔG is relatively insensitive to changes in binding affinity, these ΔG scores were then converted into binding constants (K_{d}) using Equation (1) ($\Delta G = \mathrm{RT} \bullet \ln K_{\mathrm{d}}$). The results show that the average difference of these K_{d} scores determined by different methods is within a 2.6-fold range (Table 3). Considering that it is not uncommon to observe 2- to 3-fold variations when determining K_{d} even by curve-fitting methods (1), these results suggest that ΔG and K_{d} can be reliably obtained using the comparative EMS and PBM assays described in this study.

Genome-scale prediction of ArcA-P binding sites

The EMSA results indicated that any mutant binding site with a $\Delta\Delta G$ score above 1.77 kcal/mol interacted with ArcA-P$^{\mathrm{E}}$ weakly (Table 1 and Figure 1B). This energy score, which yields an estimated binding constant of ~1 μM, was used as the cut-off $\Delta\Delta G$ value to predict the ArcA-P binding sites with a PEM based on the values in Table 2. In total, 45 ArcA-P binding sites with a $\Delta\Delta G$ score below 1.77 kcal/mol were identified within the *Shewanellla* genome (Supplementary Table 1). Because a single binding site can be situated between divergently transcribed genes, there are 61 genes directly associated with the 45 binding sites. In a previous study, several promoters with varying binding energies were selected and their interactions with the ArcA-P protein were examined by EMSAs (24). Of the 14 promoters that contain a binding site with predicted $\Delta\Delta G$ values ranging from 0.64 to 1.77 kcal/mol, 13 exhibited clear binding with ArcA-P when tested by *in vitro* EMSAs with radiolabeled PCR products (24). The promoter (SO3659) that bound weakly has a relatively high $\Delta\Delta G$ value (1.60 kcal/mol). Of the nine promoters that contain sites with predicted $\Delta\Delta G$ values above 3.09 kcal/mol, none exhibited noticeable binding with ArcA-P (24). These results suggest that the predicted binding energies are strongly associated with the ability to interact with ArcA-P. To date, microarray gene expression data is available for the genes encoded at 43 of the 45 predicted ArcA-P sites corresponding to 61 potentially regulated genes. Of these 43 sites, 27 (63%) encode genes exhibiting >2-fold regulation and 35 (81%) exhibit >1.5-fold regulation by ArcA (Supplementary Table 1). Because the microarray study only examined transcriptional regulation at a given condition and a given time, the accuracy of the site-discovery approach described in this study is likely >81% with regard to *in vivo* gene regulation. Thus, the number of false positive predictions appears to be low. False negatives, however, may be higher since a total of ~300 genes were identified as under ArcA regulation (>2-fold regulation) in the microarray study (24).

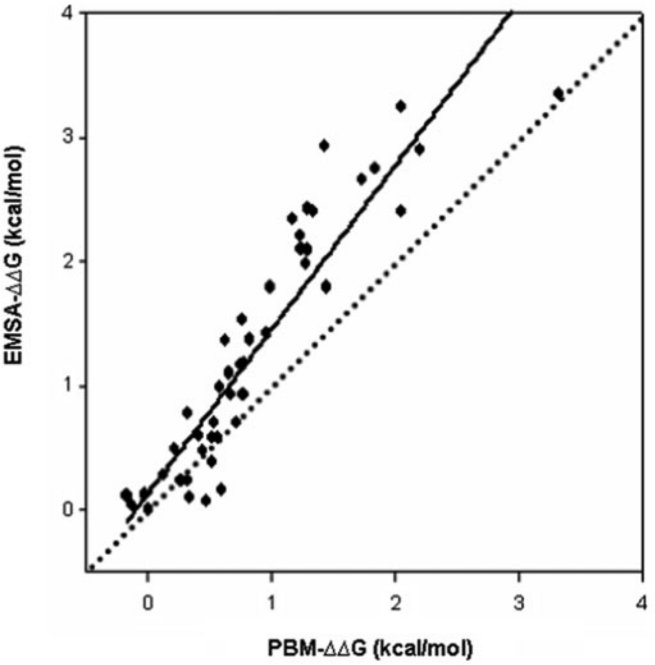


Figure 4. Correlation of relative binding energy ($\Delta\Delta G$). The *x*- and *y*-axis represent the $\Delta\Delta G$ values derived from PBM (with ArcA-P$^{\mathrm{C}}$) and EMS (with ArcA-P$^{\mathrm{E}}$) assays, respectively. The R^2-value is 0.94 (solid line). The dotted line represents the perfect correlation line for absolute values of $\Delta\Delta G$.

The *Shewanella* ArcA protein shares high sequence identity with ArcA from several other Gram-negative bacteria, such as *E. coli* (81%) and *H. influenzae* (79%). The role of ArcA has been extensively studied in *E. coli*, thus providing an ideal system to examine the accuracy of the site-discovery approach described in this study. For this purpose, a genome-scale prediction ArcA-P binding sites in *E. coli* was performed using the same ArcA-P$^{\mathrm{E}}$-derived PEM as that used above for *S. oneidensis* (Table 2). Using the same 1.77 kcal/mol energy threshold, a total of 57 putative ArcA-P binding sites were identified including seven of the nine canonical ArcA-P regulated promoters (21,31) (Supplementary Table 2). Among the 57 predicted sites, 27 (47%) have been reported to encode genes exhibiting >2-fold regulation by ArcA (Supplementary Table 2). This number could increase as additional gene regulatory data becomes available. To date, footprinting assay results have been published for a total of 15 *E. coli* ArcA-P interacting DNAs including 14 promoters (Supplementary Table 4). For these ArcA-P footprinted DNAs, a total of 27 ArcA-P binding sites (up to two sites per footprinted DNA) were predicted using the *Shewanella* ArcA-P$^{\mathrm{E}}$ PEM with no preset threshold (Table 2), among which 24 sites are located exactly within the ArcA-P footprinting regions and three are within a region that was not examined by footprinting or any other assays (Supplementary Table 4). For the 14 *E. coli* ArcA-P footprinted promoters, eight contain a site with a predicted $\Delta\Delta G$ value below 1.82 kcal/mol and these promoters are all strongly regulated by ArcA-P

(>5-fold regulation), including the *lctPRD* promoter which contains an ArcA-P site with the most favorable predicted $\Delta\Delta G$ score and which also exhibits the most significant level of regulation by ArcA-P (~90- to 100-fold) (34). For the six promoters containing sites with predicted $\Delta\Delta G$ values >2.95 kcal/mol, five are weakly regulated by ArcA-P (<3-fold regulation). Taken together, these results indicate a strong correlation of predicted $\Delta\Delta G$ scores with the strength of ArcA-P binding and regulation.

Genome-scale predictions of *H. influenzae* ArcA-P binding sites were also performed using the ArcA-P^E PEM (Table 2). By using the 1.77 kcal/mol cut-off $\Delta\Delta G$ score used for the *S. oneidensis* and *E. coli* predictions, a total of 22 ArcA-P target binding sites were identified (Supplementary Table 3). The 22 binding sites is a somewhat smaller number of predictions than those for the *S. oneidensis* (45) and *E. coli* (57) genomes, but it is consistent with a recent microarray study where only 23 genes exhibited >2-fold regulation by ArcA in *H. influenzae* (35). Among these 23 genes identified in the microarray study, 12 were predicted to contain an ArcA-P binding site using the 1.77 kcal/mol threshold. Interestingly, the eight predicted ArcA-P binding sites with the most favorable energy scores ($\Delta\Delta G \leq 1.13$ kcal/mol) were all within promoter regions for genes exhibiting >2-fold regulation by ArcA-P in the microarray study (35) (Supplementary Table 3). These results, in addition to the *S. oneidensis* and *E. coli* results, suggest that false positive predictions are rare among the binding sites with favorable energy scores.

DISCUSSION

In this study, a simple model was used to examine binding energy in DNA–protein interactions using electrophoretic gel shift and PBM assays. With this approach, the importance of each position within the ArcA-P binding site was quantitatively established by characterizing the interaction between *Shewanella* ArcA-P and a series of mutant promoter DNAs, whereby each position in the binding site was systematically mutated to all possible single nucleotide changes. The results of the fine mapping were used to create a PEM that was used for a genome-scale prediction of 45 ArcA-P sites in *Shewanella*. A further examination suggests that this prediction is >81% consistent with *in vivo* gene regulation according to microarray studies and >92% (13/14) accurate in terms of published *in vitro* gel shift validation binding assays (24). In addition, this study predicted 27 ArcA-P sites for 15 published *E. coli* ArcA-P footprinted DNAs, and 24 of them were found exactly within the footprinting protected regions and the other three sites fall into the regions that were not examined by footprinting assays (Supplementary Table 4). This is the first report showing that footprinting protected regions can be effectively predicted by starting from a single known transcription factor binding site. Finally, the predicted *H. influenzae* ArcA-P sites correlate well with *in vivo* regulation determined by a microarray analysis in that the eight predicted binding sites with the most favorable $\Delta\Delta G$ scores all exhibit ArcA dependent gene regulation (Supplementary Table 3) (35).

As indicated earlier, the available validation data suggest the identification of binding sites using binding energies is highly accurate in terms of very few false positives but that false negatives clearly occur. There are several possible explanations for false negative predictions. One obvious contributor to false negatives is the $\Delta\Delta G$ threshold chosen for the scores obtained from the genome scan. False negative predictions may also occur due to cooperative protein binding to multiple weak binding sites present in a promoter region. It has been shown that ArcA protein multimerizes upon phosphorylation and that the multimeric protein can bind to multiple sites within a promoter region (36,37).

The one-step percentage-of-binding strategy described in this study provides a rapid approach to examine binding energy in DNA–protein interactions *via* systematic mutation of the DNA binding site. Since most *cis*-regulatory sites are ~6–12 bp long (38), the one-step EMSA described here provides an efficient means of generating a PEM for genome-scale site discovery. Compared with other site-discovery approaches, the method described in this study requires little previously known experimental data (only a single known binding site is necessary). Compared with the few available high-throughput methods (12–20) to measure DNA–protein binding energies, the percentage-of-binding approach represents a simple yet effective method. In addition, the application of percentage-of-binding strategy to microarray-based DNA–protein interactions could result in a low cost and high throughput genome-scale site-discovery approach for many other transcription factors.

SUPPLEMENTARY DATA

Supplementary data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Dr Hiram F. Gilbert of Baylor College of Medicine for discussions and comments on the article, and Dr Dimitris Georgellis of the Universidad Nacional Autonoma de Mexico for the gift of *E. coli* ArcB expressing plasmid pQE30-ArcB$^{78\text{-}778}$. This research was supported by the U.S. Department of Energy under the Genomics: GTL Program through *Shewanella* Federation, Office of Biological and Environmental Research, Office of Science. Funding to pay the Open Access publication charges for this article was provided by Baylor College of Medicine.

Conflict of interest statement. None declared.

REFERENCES

1. Stormo,G.D. and Fields,D.S. (1998) Specificity, free energy and information content in protein-DNA interactions. *Trends Biochem. Sci.*, **23**, 109–113.
2. Bulyk,M.L. (2003) Computational prediction of transcription-factor binding site locations. *Genome Biol.*, **5**, 201.

3. Kim,T.H. and Ren,B. (2006) Genome-wide analysis of protein-DNA interactions. *Annu Rev. Genomics Hum. Genet.*, **7**, 81–102.
4. Tompa,M., Li,N., Bailey,T.L., Church,G.M., De Moor,B., Eskin,E., Favorov,A.V., Frith,M.C., Fu,Y., Kent,W.J. *et al.* (2005) Assessing computational tools for the discovery of transcription factor binding sites. *Nat. Biotechnol.*, **23**, 137–144.
5. Li,N. and Tompa,M. (2006) Analysis of computational approaches for motif discovery. *Algorithms Mol. Biol.*, **1**, 8.
6. Tronche,F., Ringeisen,F., Blumenfeld,M., Yaniv,M. and Pontoglio,M. (1997) Analysis of the distribution of binding sites for a tissue-specific transcription factor in the vertebrate genome. *J. Mol. Biol.*, **266**, 231–245.
7. Ren,B., Robert,F., Wyrick,J.J., Aparicio,O., Jennings,E.G., Simon,I., Zeitlinger,J., Schreiber,J., Hannett,N., Kanin,E. *et al.* (2000) Genome-wide location and function of DNA binding proteins. *Science*, **290**, 2306–2309.
8. Iyer,V.R., Horak,C.E., Scafe,C.S., Botstein,D., Snyder,M. and Brown,P.O. (2001) Genomic binding sites of the yeast cell-cycle transcription factors SBF and MBF. *Nature*, **409**, 533–538.
9. Kim,J., Bhinge,A.A., Morgan,X.C. and Iyer,V.R. (2005) Mapping DNA-protein interactions in large genomes by sequence tag analysis of genomic enrichment. *Nat. Methods*, **2**, 47–53.
10. Johnson,D.S., Mortazavi,A., Myers,R.M. and Wold,B. (2007) Genome-wide mapping of *in vivo* protein-DNA interactions. *Science*, **316**, 1497–1502.
11. Audic,S. and Claverie,J.M. (1998) Visualizing the competitive recognition of TATA-boxes in vertebrate promoters. *Trends Genet.*, **14**, 10–11.
12. Brockman,J.M., Frutos,A.G. and Corn,R.M. (1999) A multistep chemical modification procedure to create DNA arrays on gold surfaces for the study of protein-DNA interactions with surface plasmon resonance imaging. *J. Am. Chem. Soc.*, **121**, 8044–8051.
13. Hallikas,O., Palin,K., Sinjushina,N., Rautiainen,R., Partanen,J., Ukkonen,E. and Taipale,J. (2006) Genome-wide prediction of mammalian enhancers based on analysis of transcription-factor binding affinity. *Cell*, **124**, 47–59.
14. Boger,D.L., Fink,B.E., Brunette,S.R., Tse,W.C. and Hedrick,M.P. (2001) A simple, high-resolution method for establishing DNA binding affinity and sequence selectivity. *J. Am. Chem. Soc.*, **123**, 5878–5891.
15. Thorsen,T., Maerkl,S.J. and Quake,S.R. (2002) Microfluidic large-scale integration. *Science*, **298**, 580–584.
16. Maerkl,S.J. and Quake,S.R. (2007) A systems approach to measuring the binding energy landscapes of transcription factors. *Science*, **315**, 233–237.
17. Zhang,L., Kasif,S. and Cantor,A.C. (2007) Quantifying DNA-protein binding specificities by using oligonucleotide mass tags and mass spectroscopy. *Proc. Natl Acad. Sci. USA*, **104**, 3061–3066.
18. Fields,D.S. and Stormo,G.D. (1994) Quantitative DNA sequencing to determine the relative protein-DNA binding constants to multiple DNA sequences. *Anal. Biochem.*, **219**, 230–239.
19. Fields,D.S., He,Y., Al-Uzri,A.Y. and Stormo,G.D. (1997) Quantitative specificity of the Mnt repressor. *J. Mol. Biol.*, **271**, 178–194.
20. Luo,B., Perry,D.J, Zhang,L., Kharat,I., Basic,M. and Fagan,J.B. (1997) Mapping sequence specific DNA-protein interactions: a versatile, quantitative method and its application to transcription factor XF1. *J. Mol. Biol.*, **266**, 479–492.
21. Lynch,A.S. and Lin,E.C.C. (1996)Responses to molecular oxygen. In Neidhardt,F.C., Curtiss,R. III, Ingraham,J. L., Lin,E.C.C., Low,K.B., Magasanik,B., Reznikoff,W.S., Riley,M., Schaechter,M. and Umbarger,H.E. (eds), *Salmonella: Cellular and Molecular Biology*, 2nd edn. American Society for Microbiology, Washington, DC, pp. 1526–1538.
22. Heidelberg,J.F., Paulsen,I.T., Nelson,K.E., Gaidos,E.J., Nelson,W.C., Read,T.D., Eisen,J.A., Seshadri,R., Ward,N., Methe,B. *et al.* (2002) Genome sequence of the dissimilatory metal ion-reducing bacterium *Shewanella oneidensis*. *Nat. Biotechnol.*, **20**, 1118–1123.
23. Gralnick,J.A., Brown,C.T. and Newman,D.K. (2005) Anaerobic regulation by an atypical Arc system in *Shewanella oneidensis*. *Mol. Microbiol.*, **56**, 1347–1357.
24. Gao,H., Wang,X., Yang,Z.K., Palzkill,T. and Zhou,J. (2008) Probing the regulation of ArcA in *Shewanella oneidensis* MR-1 by integrated genomic analyses. *BMC Genomics*, **9**, 42.
25. Mukherjee,S., Berger,M.F., Jona,G., Wang,X.S., Muzzey,D., Snyder,M., Young,R.A. and Bulyk,M.L. (2004) Rapid analysis of the DNA-binding specificities of transcription factors with DNA microarrays. *Nat. Genet.*, **36**, 1331–1339.
26. Pyle,A.M., McSwiggen,J.A. and Cech,T.R. (1990) Direct measurement of oligonucleotide substrate binding to wild-type and mutant ribozymes from Tetrahymena. *Proc. Natl Acad. Sci. USA*, **87**, 8187–8191.
27. Del Carmine,R., Molinari,P., Sbraccia,M., Ambrosio,C. and Costa,T. (2004) 'Induced-fit' mechanism for catecholamine binding to the beta2-adrenergic receptor. *Mol. Pharmacol.*, **66**, 356–363.
28. Iuchi,S. and Lin,E.C. (1992) Purification and phosphorylation of the Arc regulatory components of *Escherichia coli*. *J. Bacteriol.*, **174**, 5617–5623.
29. Georgellis,D., Lynch,A.S. and Lin,E.C. (1997) *In vitro* phosphorylation study of the *arc* two-component signal transduction system of. *Escherichia coli*. *J. Bacteriol.*, **179**, 5429–5435.
30. Workman,C.T., Yin,Y., Corcoran,D.L., Ideker,T., Stormo,G.D. and Benos,P.V. (2005) enoLOGOS: a versatile web tool for energy normalized sequence logos. *Nucleic Acids Res.*, **33**, W389–W392.
31. Liu,X. and De Wulf,P. (2004) Probing the ArcA-P modulon of *Escherichia coli* by whole genome transcriptional analysis and sequence recognition profiling. *J. Biol. Chem.*, **279**, 12588–12597.
32. Toro-Roman,A., Mack,T.R. and Stock,A.M. (2005) Structural analysis and solution studies of the activated regulatory domain of the response regulator ArcA: a symmetric dimer mediated by the alpha4-beta5-alpha5 face. *J. Mol. Biol.*, **349**, 11–26.
33. Bylund,D.B. and Murrin,L.C. (2000) Radioligand saturation binding experiments over large concentration ranges. *Life Sci.*, **67**, 2897–2911.
34. Iuchi,S. and Lin,E.C.C. (1988) *arcA* (*dye*), a global regulatory gene in *Escherichia coli* mediating repression of enzymes in aerobic pathways. *Proc. Natl Acad. Sci. USA*, **85**, 1888–1892.
35. Wong,S.M., Alugupalli,K.R., Ram,S. and Akerley,B.J. (2007) The ArcA regulon and oxidative stress resistance in *Haemophilus influenzae*. *Mol. Microbiol.*, **64**, 1375–1390.
36. Jeong,J.Y., Kim,Y.J., Cho,N., Shin,D., Nam,T.W., Ryu,S. and Seok,Y.J. (2004) Expression of *ptsG* encoding the major glucose transporter is regulated by ArcA in *Escherichia coli*. *J. Biol. Chem.*, **279**, 38513–38518.
37. Lynch,A.S. and Lin,E.C. (1996) Transcriptional control mediated by the ArcA two-component response regulator protein of *Escherichia coli*: characterization of DNA binding at target promoters. *J. Bacteriol.*, **178**, 6238–6249.
38. Berger,M.F., Philippakis,A.A., Qureshi,A.M., He,F.S., Estep,P.W. III and Bulyk,M.L. (2006) Compact, universal DNA microarrays to comprehensively determine transcription-factor binding site specificities. *Nat. Biotechnol.*, **24**, 1429–1435.