

# Non-independence of Mnt repressor–operator interaction determined by a new quantitative multiple fluorescence relative affinity (QuMFRA) assay

Tsz-Kwong Man and Gary D. Stormo\*

Department of Genetics, Washington University Medical School, 660 S. Euclid, Box 8232, St Louis, MO 63110, USA

Received March 20, 2001; Revised and Accepted May 1, 2001

## ABSTRACT

***Salmonella* bacteriophage repressor Mnt belongs to the ribbon–helix–helix class of transcription factors. Previous SELEX results suggested that interactions of Mnt with positions 16 and 17 of the operator DNA are not independent. Using a newly developed high-throughput quantitative multiple fluorescence relative affinity (QuMFRA) assay, we directly quantified the relative equilibrium binding constants ( $K_{ref}$ ) of Mnt to operators carrying all the possible dinucleotide combinations at these two positions. Results show that Mnt prefers binding to C, instead of wild-type A, at position 16 when wild-type C at position 17 is changed to other bases. The measured  $K_{ref}$  values of double mutants were also higher than the values predicted from single mutants, demonstrating the non-independence of these two positions. The ability to produce a large number of quantitative binding data simultaneously and the potential to scale up makes QuMFRA a valuable tool for the large-scale study of macromolecular interaction.**

## INTRODUCTION

Mnt is an 82 amino acid repressor from the *Salmonella* phage P22 (1). It belongs to the ribbon–helix–helix family of DNA-binding proteins, in which the DNA contacts in the major groove are made by an antiparallel  $\beta$ -ribbon from the N-termini of two Mnt monomers (2). Mnt exists as a tetramer in solution, with each dimer binding to a half-site of a nearly symmetric 21 bp operator (3,4):

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21  
a t a g g t c c a c g g t g g a c c t g t

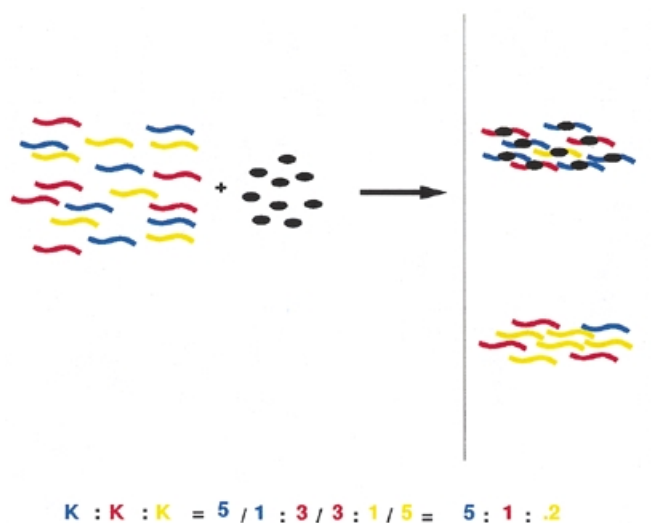
Only the inner 17 bp, positions 3–19, contribute significantly to sequence specific binding (5,6). Extensive biochemical and genetic analyses have identified the amino acid–base pair contacts that are most important for sequence-specific binding (7–9). Comparison to the crystal structure of the homologous Arc repressor–DNA complex has led to further refinement of the model for Mnt–DNA interaction (2,10). Fields *et al.* (6) measured the change in binding free energy to all possible single base changes in the mnt operator half-site. Those results

are mostly consistent with the interaction model and add quantitative information about the contribution of each base pair to the specificity of binding. Recently the structure of the tetramerization domain of Mnt has been determined (11) and a single-chain version of the Mnt dimer has also been constructed (12). The latter accomplishment, in particular, has helped to resolve the interactions of Arg2 residue with the operator DNA. So, despite the lack of a crystal structure for the Mnt–DNA complex, a considerable amount of information about the protein–DNA interaction has allowed for a detailed model of the complex (2,10,12,13).

Mnt has also been used to study how amino acid changes in the DNA-binding domain alter the specificity of the protein. It was the first DNA-binding protein for which mutants were selected that had changed their specificity from the wild-type operator to a different sequence (14). In that experiment, G at operator position 5 was changed to A, and the symmetrically related C at position 17 changed to a T. Those changes created two sites for methylation by the *Escherichia coli* *dam* methylase. Mnt proteins were selected that would bind specifically to the altered operators, and the only mutant isolated had changed His6 to Pro. It was later shown that methylation of the DNA was required for the high affinity binding by the H6P mutant Mnt (15). Raumann *et al.* (13) also showed the importance of His6 in determining the specificity of Mnt. A protein that is a hybrid between Mnt and the homologous protein Arc, but with the Mnt amino acids at the DNA contacting positions, interacts specifically with the Mnt operator. However, if His6 (in the Mnt numbering, which corresponds to position 9 in Arc) is replaced by the Arc amino acid, Gln, the specificity changes such that the protein now binds with high affinity to both the Mnt operator and the Arc operator. It is an interesting finding because the two operators differ at nearly every position (7,13). Position 6 of Mnt was termed a ‘master’ residue that not only makes direct contact with the DNA but could also affect the interactions of other residues with DNA.

*In vitro* selection of operators with high affinity to Mnt showed some context-dependent effects, where mutations at one position influence the interactions at other positions (6,16). This was most evident at positions 5 and 17, which interact directly with His6, and the neighboring positions 6 and 16 (see Table 2C). Additivity is often invoked to extrapolate from the effects of single mutations to the expected effects of multiple mutations. If changes in binding free energy were additive, then measuring the effects of all single mutations would

\*To whom correspondence should be addressed. Tel: +1 314 747 5534; Fax: +1 314 362 7855; Email: stormo@ural.wustl.edu



**Figure 1.** The principle of QuMFRA assay to simultaneously determine multiple relative equilibrium binding constants ( $K_{ref}$ ). A mixture of three different DNA sequences (red, yellow and blue) is incubated with a DNA binding protein (black). After reaching equilibrium, the bound and unbound fractions of the mixture are separated by electrophoretic mobility shift assay. The  $K_{ref}$  (blue:red:yellow = 5:1:0.2) are determined by simply measuring the ratios of each DNA in the bound and unbound fractions (see text for details).

suffice to know the binding energy to all potential binding sites. Additivity clearly fails when the affinity decreases to the point of non-specific binding (17). But for many proteins, including RNA polymerase as well as various transcription factors, additivity seems to be approximately valid over several logs of decrease in affinity (18–22). It has even been shown on Mnt that additivity holds for mutations at two positions that are in separate half-sites (23). Therefore, the systematic evolution of ligands by exponential enrichment (SELEX) results indicating non-additivity between the neighboring positions 5/6 (and 16/17) are worth investigating carefully (6).

In order to obtain accurate information about changes in free energies of binding to different sequences, and to do it efficiently enough that we could obtain results for many different sequences rapidly, we developed the new method QuMFRA described in this paper (Fig. 1). Briefly, the method utilizes a competition assay where several different fluorescently labeled DNAs compete for binding to the same pool of protein. The bound and unbound fractions of the DNA are separated using an electrophoretic mobility shift assay (EMSA). The amount of each DNA in each fraction is determined fluorescently by mathematically deconvoluting the overlap of the different fluorescent signals at the different emission wavelengths. In this paper we determined the change in binding energy for all 16 possible base combinations at positions 5 and 6, using only seven lanes of a gel. Using only four fluorescent labels the method returns three relative affinity measurements per lane, which makes the method efficient enough to collect significant amounts of quantitative data on the binding effects of many different mutant sequences. For example, a SELEX procedure that returns 60 different binding sites for a protein could be turned into quantitative effects on binding energy in

only 20 lanes of a gel. Using more fluorescent labels would make the assay even more efficient. Such data could be used to develop more complicated models of protein–DNA interactions that significantly improve predictions over the current simple models.

## MATERIALS AND METHODS

### Materials and reagents

Chemicals and reagents were purchased from Fisher Scientific (Pittsburgh, PA) and Sigma Chemical Co. (St Louis, MO) unless specified. Klenow fragment, dNTPs and extension buffer for primer extension were from Promega (Madison, WI). Fluorophore-labeled SK-1 oligos (FAM-SK1, HEX-SK1, TAMRA-SK1 and ROX-SK1) and unlabeled oligos used in primer extension were purchased from Integrated DNA Technologies (Coralville, IA). SK-1 oligo is 5'-gtggcggccgctctaact. The oligo containing the wild-type mnt operator sequence is 5'-tcgaggtcgcagctatcATAGGTCCACCGTGGACCTATtcactagt-tctagagcggccgccac. The flanking sequences (lower case) outside the mnt operator site (upper case) are SK-1 and KS-1 primer sites. The sequence of mutant oligo is the same as the wild-type oligo except for the mutation(s) indicated at positions 16 and 17 (underlined) in the mnt operator sequence.

### QuMFRA assay

Double strand labeled DNA molecules were synthesized by primer extension using 1  $\mu$ M of unlabeled oligos containing the mnt operator site, 1  $\mu$ M of FAM, HEX, TAMRA, or ROX labeled SK-1 primer, and 0.5 U Klenow fragment in a 100  $\mu$ l reaction. The reactions were extended at 37°C for 1 h. Fluorophore-labeled DNAs (1 pmol) were mixed with ~50 nM of active Mnt protein (gift of R.T.Sauer) in 1 $\times$  binding buffer (10 mM Tris pH 7.5, 200 mM KCl, 10 mM MgCl<sub>2</sub>, 0.1 mg ml<sup>-1</sup> BSA and 0.1% NP-40) at room temperature for 1 h. The binding reactions were then analyzed in a 10% polyacrylamide (0.5 $\times$  TBE) gel at 10 Vcm<sup>-1</sup> for 1 h at room temperature. The gel was scanned by Typhoon Variable Scanner (Molecular Dynamics, Sunnyvale, CA) using excitation laser at 532 nm and the following settings: fluorophore FAM, output voltage 600 V, emission filter 526 nm; fluorophore HEX, output voltage 475 V, emission filter 550 nm; fluorophore TAMRA, output voltage 475 V, emission filter 580 nm; fluorophore ROX, output voltage 475 V, emission filter 610 nm. The fluorescent intensities of the bound and unbound fractions on each lane at each emission wavelength were quantified by using volume analysis of the ImageQuant software (Molecular Dynamics, Sunnyvale, CA). The background fluorescent intensity was subtracted using object average method of ImageQuant software (Molecular Dynamics). The resultant fluorescence intensities after background subtraction at all four emission wavelengths are the output emission vector X.

The binding constant ( $K_b$ ) of a DNA species 1 is equal to:

$$K_b(D_1) = [P \cdot D_1] / [P][D_1]$$

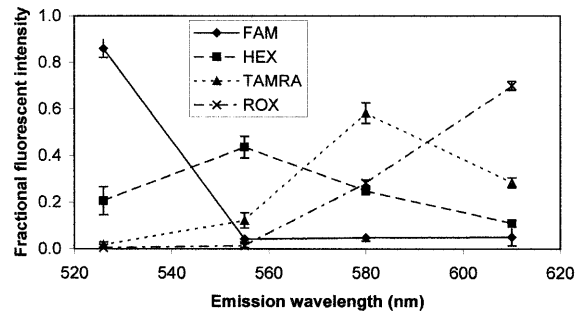
where  $[P \cdot D_1]$  is the amount of protein–DNA complex and  $[P]$  and  $[D_1]$  are the amounts of free protein and DNA, respectively. In a competition assay where two or more DNAs are binding to the same pool of protein, the relative affinities can be determined without knowing the free protein concentration.

For example, the ratio of binding constants to DNAs  $D_1$  and  $D_2$  is:

$$\begin{aligned} K_b(D_1)/K_b(D_2) &= [P \cdot D_1][P][D_2]/[P][D_1][P \cdot D_2] \\ &= [P \cdot D_1][D_2]/[P \cdot D_2][D_1] \end{aligned}$$

The ratios of bound and unbound concentrations for both DNAs can be obtained from the fluorescence measurements of the gel using the method of deconvolution described in the text (Fig. 1). Fields *et al.* (6) had previously used the technique of competition EMSA to derive the relative affinities of multiple DNA sequences, but instead of using fluorescence measurements they cut the bands from the gel and subjected them to quantitative sequencing. The multiplex nature of QuMFRA eliminates the need for sequencing and provides a more sensitive and efficient means to determine the relative affinities of multiple DNA sequences. Although the absolute protein concentration is not important in measuring relative affinities, the oligomeric state of the binding protein is still important to the results when the proteins bind to DNA as multimers. In any single binding reaction, such as in the QuMFRA assay, all of the DNAs are competing for the same oligomeric form (Mnt tetramer) of the protein. However, if one wants to compare two separate reactions, care must be taken to ensure that the protein is in the same oligomeric form in both reactions, such that the final protein concentrations of the two reactions are the same. For instance, Mnt binds DNA as a tetramer and also exists as a tetramer in solution if the concentration is high enough (24). At lower concentrations it exists in solution as a dimer and then forms tetramers upon binding DNA. Relative affinities measured in different reactions where Mnt exists in different states, tetramers versus dimers, might not be comparable. In these experiments Mnt concentrations were high enough to ensure primarily tetramers in all reactions. Fields *et al.* (6) used similar conditions and their measurements of relative affinity were generally in close agreement with those obtained by Knight and Sauer using traditional methods of determining binding constants (9). The relative affinities determined in this paper are in close agreement with those of Fields *et al.* (6) for those sequences that were in common. Therefore, this paper demonstrates that QuMFRA is a rapid and accurate method to determine relative affinities of multiple DNA sequences.

**A**



**B**

Emission Filter (nm)	FAM	HEX	TAMRA	ROX
526	0.860 ± 0.038	0.206 ± 0.060	0.013 ± 0.016	0.005 ± 0.008
555	0.042 ± 0.015	0.436 ± 0.046	0.122 ± 0.032	0.014 ± 0.008
580	0.048 ± 0.014	0.248 ± 0.011	0.582 ± 0.044	0.282 ± 0.013
610	0.050 ± 0.036	0.110 ± 0.013	0.283 ± 0.022	0.699 ± 0.019

**Figure 2.** The emission spectra (A) and the emission matrix E (B) of the four fluorophores used in QuMFRA and deconvolution. SK-1 oligos (1 μM) labeled with FAM, HEX, TAMRA or ROX were analyzed on a 10% polyacrylamide gel. The bands were then scanned and quantified by Typhoon Variable Scanner using emission filters as indicated. The values (fractional fluorescent intensities) in both emission spectrum and emission matrix E are the fraction of the fluorescence intensity of a fluorophore in one emission wavelength divided by the total fluorescence intensity of that fluorophore in all four emission wavelengths. The sum of the fractional fluorescent intensities of a fluorophore in all four emission wavelengths is equal to 1. The values shown in (B) are the means and standard deviations from five independent measurements.

## RESULTS AND DISCUSSION

### QuMFRA assay

The new QuMFRA assay can be summarized in the following steps. First, DNA sequences with different mutations are labeled with various fluorophores, such as FAM, HEX,

**Table 1.** Comparison of actual ratios of fluorophore-labeled oligos used and the ratios calculated after scanning, measurement and deconvolution

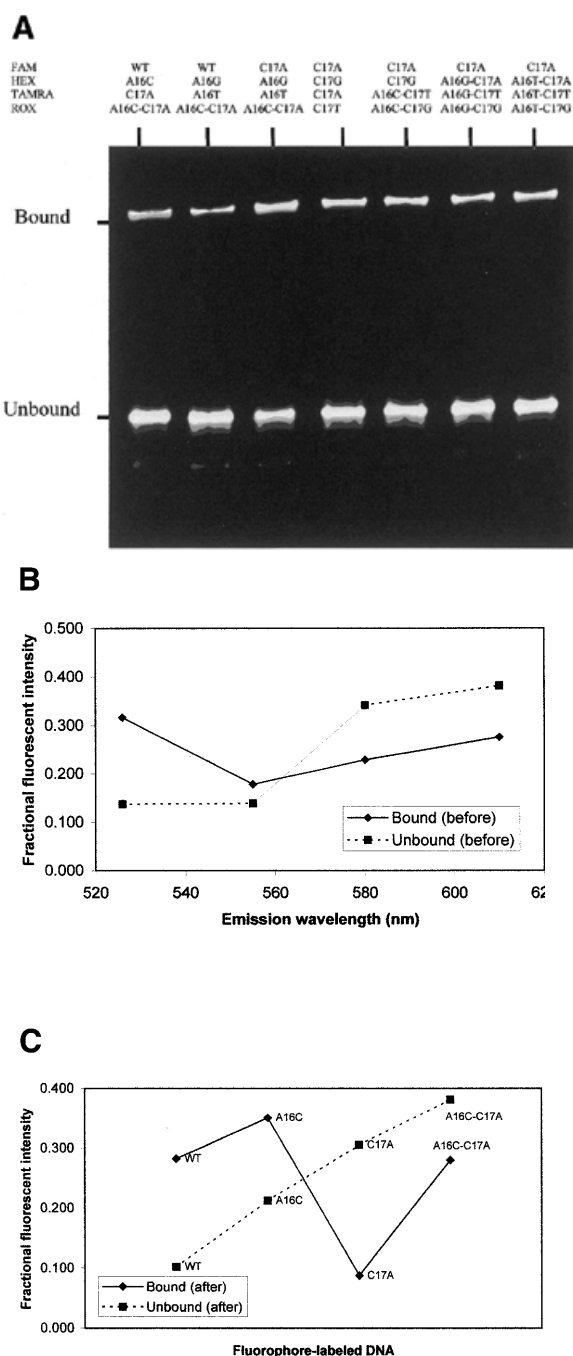
Ratio used				Ratio calculated			
FAM	HEX	TAMRA	ROX	FAM	HEX	TAMRA	ROX
0.25	0.25	0.25	0.25	0.25 ± 0.01	0.24 ± 0.01	0.24 ± 0.05	0.27 ± 0.05
0.17	0.50	0.17	0.17	0.16 ± 0.02	0.51 ± 0.03	0.16 ± 0.04	0.17 ± 0.06
0.10	0.30	0.60	0.00	0.10 ± 0.02	0.30 ± 0.02	0.60 ± 0.03	0.00 ± 0.01
0.08	0.08	0.08	0.77	0.08 ± 0.02	0.07 ± 0.02	0.08 ± 0.01	0.77 ± 0.04
0.63	0.13	0.13	0.13	0.62 ± 0.07	0.15 ± 0.04	0.10 ± 0.01	0.14 ± 0.06

The oligos used were SK-1 primer labeled with the four fluorophores as indicated. The values shown were means and standard deviations from five independent measurements.

TAMRA and ROX in this case. Labeled DNAs and the DNA binding protein of interest are mixed and allowed to attain equilibrium. The DNA–protein complexes (bound) and unbound DNAs are separated on a native polyacrylamide gel using EMSA. The gel is then scanned by a fluorescence scanner with optimized excitation and emission settings for the set of dyes used (see Materials and Methods). The fluorescent intensities of the bound and unbound fractions at each emission wavelength are quantified. Since the emission spectra of the fluorophores significantly overlap with each other (Fig. 2A), we used a deconvolution method that converts the fluorescent intensity at each emission wavelength to the corresponding amount of fluorophore-labeled DNA molecule. The actual amount of each DNA in a band is determined by using the equation,  $E \times M = X$ , where  $E$  is the emission matrix, vector  $M$  is any mixture of fluorophore-labeled DNAs and vector  $X$  is the measured output fluorescent intensities at different emission wavelengths. The emission matrix  $E$  (Fig. 2B) was determined by measuring the ratios of fluorescent intensities of SK-1 oligos labeled with FAM, HEX, TAMRA and ROX at each wavelength as described in Materials and Methods.

In a control experiment, we tested the accuracy of our deconvolution method by measuring the amounts of the four different fluorophore-labeled SK-1 oligos that were mixed at different ratios (1–10-fold) and analyzed on a native polyacrylamide gel (Table 1). The results in Table 1 show that the calculated ratios of the labeled oligos after scanning, measurement and deconvolution correspond very well with the actual ratios of labeled oligos used in the assay. The average difference between the measured and actual ratios was 5.2%, and had a standard deviation of only 17% of the measured mean.

To determine the amount of each DNA in the bound and unbound fractions of a QuMFRA assay, we measured the output emission vector  $X$  (the fluorescent intensities of each band at the four different emission wavelengths; Fig. 3A). Knowing the values in  $E$  and  $X$ , we then calculated the amount of each DNA in a band,  $M$ , using the equation above. Figure 3B shows the fractional fluorescent intensities of bound and unbound fractions of lane 1 in Figure 3A, which contained wild-type Mnt protein and four different operator sequences. The wild-type operator, and operators carrying A16C, C17A and A16C-C17A mutations were labeled with FAM, HEX, TAMRA and ROX, respectively. The fractional fluorescent intensity at each emission wavelength before the deconvolution is the sum of the fluorescent intensities of all four fluorophores at that emission wavelength (Fig. 3B). After the deconvolution, we obtained the ratio of the corresponding DNA molecules labeled with FAM, HEX, TAMRA or ROX (Fig. 3C). The relative equilibrium binding constants ( $K_{rel}$ ) are then calculated by dividing the bound to unbound ratio of each DNA sequence to that of a reference DNA sequence, such as the wild-type mnt operator (Table 2A). To test if the four fluorophores affect the Mnt binding to the operator, we performed the QuMFRA assay using the Mnt protein and four identical wild-type operator sequences, each labeled with a different fluorophore. The bound to unbound ratios of these labeled sequences were 1:1:1:1, suggesting that the fluorophore moieties do not affect the binding of protein to DNA (data not shown).



**Figure 3.** QuMFRA assay. (A) 10% polyacrylamide gel containing bound and unbound fractions of wild-type Mnt protein and operator DNA sequence containing mutations in positions 16 and 17. DNA molecules were labeled with FAM, HEX, TAMRA or ROX as indicated. Each bound and unbound fraction contained different ratios of fluorophore-labeled DNAs. Each lane contained ~50 nM of active tetrameric Mnt protein and 100 nM of the four fluorophore-labeled DNA molecules. The gel was scanned and quantified as described in Materials and Methods. The fractional fluorescent intensities of the bound and unbound fractions of the four fluorophore-labeled DNA molecules in lane 1 of (A) before (B) and after (C) deconvolution. Fractional fluorescent intensity is calculated as the intensity at one emission wavelength divided by the sum of the intensities at all four emission wavelengths. Before the deconvolution, the fluorescent intensity at each emission wavelength contains fluorescence contributions from all four fluorophores. The deconvolution method described in the text converts these mixtures of fluorescent intensities into the fluorescent intensity of each fluorophore-labeled DNA, which is proportional to the actual amount of the DNA.

**Table 2.** Non-independent effect of positions 16 and 17 in the mnt operator

Position 16	Position 17			
	A	C	G	T
<b>(A)</b>				
A	0.074 ± 0.007	1.000	0.025 ± 0.003	0.074 ± 0.012
C	0.315 ± 0.019	0.587 ± 0.099	0.109 ± 0.016	0.086 ± 0.007
G	0.028 ± 0.004	0.131 ± 0.017	0.032 ± 0.005	0.043 ± 0.005
T	0.063 ± 0.008	0.383 ± 0.027	0.027 ± 0.005	0.033 ± 0.001
<b>(B)</b>				
A	0.074	1.000	0.025	0.074
C	0.043 (7.3)	0.587	0.015 (7.5)	0.043 (2.0)
G	0.010 (2.9)	0.131	0.003 (9.9)	0.010 (4.4)
T	0.028 (2.2)	0.383	0.009 (2.9)	0.028 (1.2)
<b>(C)</b>				
A	0	93	0	0
C	3	14	2	0
G	0	2	1	0
T	0	8	0	1

(A) average measured  $K_{\text{ref}}$  and standard deviations, and (B) predicted  $K_{\text{ref}}$  of wild-type Mnt protein and all combinations of bases at positions 16 and 17 of mnt operator DNA sequence. The  $K_{\text{ref}}$  of Mnt and the wild-type operator sequence (A at position 16 and C at position 17) is equal to 1 by definition. The average values and standard deviations were calculated from four independent experiments. Predicted  $K_{\text{ref}}$  of double mutations in (B) are calculated as the products of  $K_{\text{ref}}$  of the two single mutations. The values in parentheses in (B) are the fold of differences of the measured  $K_{\text{ref}}$  in (A) and the predicted  $K_{\text{ref}}$  in (B). (C) The number of occurrences of each dinucleotide combination at positions 16 and 17 of the mnt operator DNA obtained by previously published SELEX experiments (6,16).

### Non-independent effect of positions 16 and 17 in the mnt operator

Previous results from SELEX experiments (shown in Table 2C for comparison) suggested that positions of 16 and 17 of the operator DNA of Mnt might interact non-independently (6,16). The number of occurrences of wild-type operator sequence (A at position 16 and C at position 17) was the highest (93 times out of 124). However, the Mnt protein preferentially selects the wild-type A at position 16 only if there is a wild-type C at position 17. If position 17 is not a C, the binding preference changes from A to C at position 16. Because of the small sample size (only 7 out of 124 sequences were obtained in which position 17 was not a C) and the intrinsic nature of the selection experiments (subjected to multiple rounds of selection and amplification), it is not possible to determine the quantitative effects on binding energy from the SELEX data.

To rapidly convert the results of SELEX experiments to quantitative results, we used QuMFRA to measure the  $K_{\text{ref}}$  (the wild-type mnt operator sequence is equal to 1.00 by definition) of the binding of wild-type Mnt to its operator DNA sequences containing all possible mutations (16 combinations) at positions 16 and 17 (Fig. 3A and Table 2A). Our results also confirm that C is the preferred base at position 16, not the wild-type A, when position 17 is not a C (Table 2A, row 1 and row 2). Furthermore, Mnt bound to all of the double mutants with higher measured affinity (most of them were between 2.9- and 9.9-fold) than predicted from the single mutants (Table 2B). The net effect is that Mnt binding is not as specific as predicted

from the single mutant measurements, although the total change in probability of binding to the wild-type operator sequence is not dramatic. In a binding experiment of Mnt protein with an equal mixture of Mnt operators containing all 16 possible combinations at positions 16 and 17, the wild-type sequence would represent ~33% of the total bound complex (the percentage of the  $K_{\text{ref}}$  of the wild-type operator divided by the sum of  $K_{\text{ref}}$  of all 16 operator sequences, Table 2A), whereas the prediction from the single mutants is that it would represent 41% (Table 2B). This illustrates that fairly strong deviations from additivity, in this case up to ~10-fold, may have only modest effects on the discrimination ability of the protein.

Besides using  $K_{\text{ref}}$  to infer the non-independence of position 16 and 17, we can also use the probabilities of the dinucleotides. The probability of a dinucleotide was calculated as  $K_{\text{ref}}$  of that dinucleotide divided by sum of all 16  $K_{\text{ref}}$  of dinucleotides (Table 3A). The sum of probabilities of each base at one position was also calculated as the mononucleotide probability (each column and each row of Table 3A). Using these mononucleotide probabilities, we calculated the predicted probabilities of the dinucleotide at each position (Table 3B). The ratios between the probabilities derived from dinucleotides and mononucleotides are shown in Table 3C. The dinucleotide probability of the wild-type AC is 23% higher than the predicted probability from mononucleotides, suggesting that the Mnt protein has higher affinity to wild-type AC than expected, and a deviation from independence. C at position 16

**Table 3.** Probabilities of Mnt binding to the dinucleotide at positions 16 and 17 of the mnt operator

Position 16	Position 17				SUM
	A	C	G	T	
<b>(A)</b>					
A	0.023	0.332	0.010	0.023	0.389
C	0.106	0.196	0.037	0.030	0.369
G	0.010	0.043	0.010	0.013	0.076
T	0.020	0.126	0.010	0.010	0.166
SUM	0.159	0.698	0.066	0.076	1.000
<b>(B)</b>					
A	0.062	0.271	0.026	0.030	0.389
C	0.059	0.257	0.025	0.028	0.369
G	0.012	0.053	0.005	0.006	0.076
T	0.026	0.116	0.011	0.013	0.166
SUM	0.159	0.698	0.066	0.076	1.000
<b>(C)</b>					
A	0.38	1.23	0.39	0.78	–
C	1.81	0.76	1.49	1.06	–
G	0.82	0.81	1.96	2.28	–
T	0.75	1.09	0.90	0.79	–

(A) Dinucleotide probabilities calculated from experimental  $K_{\text{ref}}$  values measured by QuMFRA assay in Table 2A. (B) Predicted dinucleotide probabilities derived from mononucleotide probabilities at positions 16 and 17 of the mnt operator, assuming independence. (C) The ratios of dinucleotide probabilities from experimental  $K_{\text{ref}}$  and predicted from mononucleotides.

has higher dinucleotide probabilities than predicted in all bases at position 17 except the wild-type C. This further indicates that C is the preferred base at position 16 if the wild-type C at position 17 is changed to other bases. The differences between the measured and predicted probabilities are much smaller than those predicted from single mutant measurements, suggesting that probabilities may be a better way to estimate the binding affinity of a protein in cases of non-independence.

Previous results have shown that His6 of the Mnt protein interacts with the G-C base pairs at positions 5 and 17 of the Mnt operator (9,13,14). But when position 17 is changed from a C to anything else, there is then a preference for C at position 16, instead of the wild-type A. This suggests that the Mnt protein is able to partially compensate for the loss of a favorable contact at position 17 by making a new contact at position 16. One intriguing possibility is that the His6 from the other monomer is now able to interact with the DNA. An important difference in how Mnt and Arc bind to DNA is shown in figure 6 of Raumann *et al.* (13). In Mnt, His6 from only one of the monomers is thought to interact directly with DNA bases, whereas in Arc both of the homologous Gln9 residues make specific contacts with the bases (2). The other His6 residue should be positioned near to position 16, but apparently does not interact (or at least does not contribute significantly to the binding affinity) when there is the wild-type position 17 interaction. Disruption of the position 17 interaction may allow the His6 at the other monomer to interact directly with the

DNA, still preferring the C on the same strand, but moved over one position to position 16. Such an alternative means of residue 6 interacting with DNA may also alter the positioning of the B-ribbon in the major groove and so influence the interactions of the other residues with DNA, thereby explaining its 'master residue' status (13).

### From SELEX data to quantitative modeling

The SELEX results are sufficient to indicate non-additivity between positions 16 and 17 (6,16), but they do not provide quantitative values for the changes in binding affinities (Table 2C). Half of the possible dinucleotides do not occur at all in the data, and two others only occur once. A much larger sample size would be needed to obtain the frequencies for all combinations. For the dinucleotides that occur more than once, there is a remarkable correlation between the frequency of occurrence and the measured  $K_{\text{ref}}$ , with both having exactly the same order: AC (wt) > CC > TC > CA > GC, CG > all others (Table 2A and C). The order of  $K_{\text{ref}}$  predicted from the single mutants is different: AC (wt) > CC > TC > GC > AT, TA > all others (Table 2B). But while the correlation in rank is quite strong, it is not possible to infer quantitative relative affinity values from the SELEX frequencies. Even taking into account that the SELEX sites were obtained after multiple rounds of selection and amplification, which implies that they should be related to the relative affinities by the power of the number of rounds, they do not produce the correct, or even consistent, values (data not shown). Undoubtedly effects of saturation, especially during the early rounds, and variations in PCR efficiency as well as chance fluctuations contribute to the differences between observed frequencies and directly measured  $K_{\text{ref}}$ . If anything, it is surprising that the ranks are correlated to the extent they are. This suggests that QuMFRA can be used to directly and rapidly quantify the relative affinities of DNA sequences that were selected from the qualitative SELEX experiments.

### The advantages of QuMFRA assay

Studying DNA-protein interactions provides an insight into how DNA binding proteins recognize target DNA sites and regulate gene expression and other cellular processes. The large amount of DNA sequencing data that has been generated recently allows us to search for many more new DNA binding proteins. The need to predict binding sites of these new proteins, and hence to understand their functions, is increasing dramatically. During the past two decades, many methods have been used to elucidate DNA-protein recognition rules of how proteins recognize their cognate DNA targets (25–29). Most of these recognition rules were based on the X-ray crystallography of many DNA-protein complexes (25,30–36) or selection-based methods, such as SELEX and phage display (26,37–43). Although results of these methods provide an invaluable source of information on how proteins interact with their cognate DNA, they are usually qualitative and do not allow quantitative modelling of DNA-protein interaction for binding site prediction (see above).

The new QuMFRA assay has the following advantages over the existing methods that are used to study DNA-protein interactions. First, QuMFRA produces direct and quantitative results of DNA-protein interaction. As described above, the  $K_{\text{ref}}$  are much more quantitative and reliable than the sequence

occurrence frequencies obtained from SELEX experiments. These quantitative data are important in generating a model to predict binding sites of new DNA binding proteins in the same class (6,16). Besides, the ability to study multiple mutations simultaneously allows us to address questions of additivity and context-dependent effects, which are often ignored in the current algorithms for DNA binding site prediction (6,16,43). Secondly, the multiplex nature of QuMFRA eliminates the deficiencies of not being able to distinguish between different DNA sequences using the traditional radioactive labeling method, or the need to generate different fragment sizes for discrimination (44). Thirdly, QuMFRA is highly sensitive. Although we used ~1 pmol of DNA in our assay, the sensitivity of our assay can certainly be increased by increasing the output voltages of the fluorescence scanner. The lowest detection limits of FAM, HEX, TAMRA and ROX are ~2–20 fmol (at output voltage = 1000 V, data not shown), which is comparable to or higher than the radioactive- (10 fmol) or Cy5- (50 fmol) labeled EMSA (45). Fourthly, QuMFRA is a high-throughput method with a potential to scale up. By using four fluorophores in this paper, we can measure the  $K_{ref}$  of three different DNA sequences per lane. In a typical 15-lane gel, we can obtain  $K_{ref}$  of 45 ( $15 \times 3$ ) sequences. Therefore, one needs less than three gels ( $16 \times 8/45$ ) to study all the possible adjacent dinucleotide changes in the 17 bp symmetric mnt operator sequence. Using QuMFRA together with SELEX, we can further limit our quantitation to positions that are important to binding or highly biased, such as positions 16 and 17. In addition, we can certainly scale up the number of fluorophores used in the assay, provided that they give distinct emission profiles using suitable combinations of excitation lasers and emission filters. The Typhoon scanner we used contains two excitation lasers (532 and 633 nm) and can maximally accommodate up to 14 emission filters at a time, and there are hundreds of different commercially available fluorophores. If we could use the full capability of the scanner, we can obtain 195 ( $13 \times 15$ )  $K_{ref}$  per 15-lane gel. The ability to use more fluorophores is under development.

In summary, we quantified the non-independent effect of positions 16 and 17 of mnt operator DNA using a new high-throughput QuMFRA assay. These results challenge the independence assumption used by many algorithms for binding site prediction. By combining the results of SELEX and QuMFRA, we now have a much faster and quantitative method to measure  $K_{ref}$  of DNA sequences containing mutations at positions that are important for DNA–protein interaction. These quantitative data are essential for better understanding of DNA–protein interactions and binding site prediction of a DNA binding protein (16).

## ACKNOWLEDGEMENTS

We thank R.T.Sauer for the wild-type Mnt protein used in this experiment. We also thank F.S.Silbaq, Manisha Thakker and Ashwin Krishna for the technical support. The project is supported by grant GM28755 from the National Institutes of Health (G.D.S.).

## REFERENCES

- Sauer, R.T., Krovatin, W., DeAnda, J., Youderian, P. and Susskind, M.M. (1983) Primary structure of the imm1 immunity region of bacteriophage P22. *J. Mol. Biol.*, **168**, 699–713.
- Raumann, B.E., Rould, M.A., Pabo, C.O. and Sauer, R.T. (1994) DNA recognition by  $\beta$ -sheets in the Arc repressor–operator crystal structure. *Nature*, **367**, 754–757.
- Vershon, A.K., Youderian, P., Susskind, M.M. and Sauer, R.T. (1985) The bacteriophage P22 arc and mnt repressors. Overproduction, purification and properties. *J. Biol. Chem.*, **260**, 12124–12129.
- Vershon, A.K., Liao, S.M., McClure, W.R. and Sauer, R.T. (1987) Bacteriophage P22 Mnt repressor. DNA binding and effects on transcription *in vitro*. *J. Mol. Biol.*, **195**, 311–322.
- Knight, K.L., Bowie, J.U., Vershon, A.K., Kelley, R.D. and Sauer, R.T. (1989) The Arc and Mnt repressors. A new class of sequence-specific DNA-binding protein. *J. Biol. Chem.*, **264**, 3639–3642.
- 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.
- Knight, K.L. and Sauer, R.T. (1989) DNA binding specificity of the Arc and Mnt repressors is determined by a short region of N-terminal residues. *Proc. Natl Acad. Sci. USA*, **86**, 797–801.
- Knight, K.L. and Sauer, R.T. (1989) Identification of functionally important residues in the DNA binding region of the mnt repressor. *J. Biol. Chem.*, **264**, 13706–13710.
- Knight, K.L. and Sauer, R.T. (1992) Biochemical and genetic analysis of operator contacts made by residues within the  $\beta$ -sheet DNA binding motif of Mnt repressor. *EMBO J.*, **11**, 215–223.
- Raumann, B.E., Brown, B.M. and Sauer, R.T. (1994) Major groove DNA recognition by B-sheet: the ribbon–helix–helix family of gene regulatory proteins. *Curr. Opin. Struct. Biol.*, **4**, 36–43.
- Nooren, I.M., Kaptein, R., Sauer, R.T. and Boelens, R. (1999) The tetramerization domain of the Mnt repressor consists of two right-handed coiled coils. *Nat. Struct. Biol.*, **6**, 755–759.
- Berggrun, A. and Sauer, R.T. (2000) Interactions of Arg2 in the Mnt N-terminal arm with the central and flanking regions of the Mnt operator. *J. Mol. Biol.*, **301**, 959–973.
- Raumann, B.E., Knight, K.L. and Sauer, R.T. (1995) Dramatic changes in DNA-binding specificity caused by single residue substitutions in an Arc/Mnt hybrid repressor. *Nat. Struct. Biol.*, **2**, 1115–1122.
- Youderian, P., Vershon, A., Bouvier, S., Sauer, R.T. and Susskind, M.M. (1983) Changing the DNA-binding specificity of a repressor. *Cell*, **35**, 777–783.
- Vershon, A., Youderian, P., Weiss, M., Susskind, M. and Sauer, R. (1985) Mnt repressor–operator interactions: altered specificity requires n-6 methylation of operator DNA. In Calendar, R. and Gold, L. (eds), *Sequence Specificity in Transcription and Translation*. Alan R. Liss, New York, pp. 209–218.
- Stormo, G.D. and Fields, D.S. (1998) Specificity, free energy and information content in protein–DNA interactions. *Trends Biochem. Sci.*, **23**, 109–113.
- Frank, D.E., Saecher, R., Bond, J., Capp, M., Tsodikov, O., Melcher, S., Levandoski, M. and Record, M.J. (1997) Thermodynamics of the interactions of lac repressor with variants of the symmetric lac operator: effects of converting a consensus site to a non-specific site. *J. Mol. Biol.*, **267**, 1186–1206.
- Sarai, A. and Takeda, Y. (1989) Lambda repressor recognizes the approximately 2-fold symmetric half-operator sequences asymmetrically. *Proc. Natl Acad. Sci. USA*, **17**, 6513–6517.
- Takeda, Y., Sarai, A. and Rivera, V. (1989) Analysis of the sequence-specific interactions between Cro repressor and operator DNA by systematic base substitution experiments. *Proc. Natl Acad. Sci. USA*, **86**, 439–443.
- Mulligan, M., Hawley, D., Entriken, R. and McClure, W. (1984) *Escherichia coli* promoter sequences predict *in vitro* RNA polymerase selectivity. *Nucleic Acids Res.*, **12**, 789–800.
- Berg, O. and von Hippel, P. (1988) Selection of DNA binding sites by regulatory proteins. II. The binding specificity of cyclic AMP receptor protein to recognition sites. *J. Mol. Biol.*, **20**, 709–723.
- Berg, O. and von Hippel, P. (1987) Selection of DNA binding sites by regulatory proteins. Statistical–mechanical theory and application to operators and promoters. *J. Mol. Biol.*, **193**, 723–750.
- Stormo, G.D., Strobl, S., Yoshioka, M. and Lee, J.S. (1993) Specificity of the Mnt protein. Independent effects of mutations at different positions in the operator. *J. Mol. Biol.*, **229**, 821–826.

24. Waldburger, C.D. and Sauer, R.T. (1995) Domains of Mnt repressor: roles in tetramer formation, protein stability and operator DNA binding. *Biochemistry*, **34**, 13109–13116.
25. Brennan, R.G. and Matthews, B.W. (1989) Structural basis of DNA–protein recognition. *Trends Biochem. Sci.*, **14**, 286–290.
26. Choo, Y. and Klug, A. (1997) Physical basis of a protein–DNA recognition code. *Curr. Opin. Struct. Biol.*, **7**, 117–125.
27. Desjarlais, J.R. and Berg, J.M. (1992) Toward rules relating zinc finger protein sequences and DNA binding site preferences. *Proc. Natl Acad. Sci. USA*, **89**, 7345–7349.
28. Harris, L.F., Sullivan, M.R. and Hickok, D.F. (1993) Conservation of genetic information: a code for site-specific DNA recognition. *Proc. Natl Acad. Sci. USA*, **90**, 5534–5538.
29. Suzuki, M., Brenner, S.E., Gerstein, M. and Yagi, N. (1995) DNA recognition code of transcription factors. *Protein Eng.*, **8**, 319–328.
30. Elrod-Erickson, M., Rould, M.A., Nekludova, L. and Pabo, C.O. (1996) Zif268 protein–DNA complex refined at 1.6 Å: a model system for understanding zinc finger–DNA interactions. *Structure*, **4**, 1171–1180.
31. Elrod-Erickson, M., Benson, T.E. and Pabo, C.O. (1998) High-resolution structures of variant Zif268–DNA complexes: implications for understanding zinc finger–DNA recognition. *Structure*, **6**, 451–464.
32. Mandel-Gutfreund, Y., Schueler, O. and Margalit, H. (1995) Comprehensive analysis of hydrogen bonds in regulatory protein–DNA complexes: in search of common principles. *J. Mol. Biol.*, **253**, 370–382.
33. Mandel-Gutfreund, Y. and Margalit, H. (1998) Quantitative parameters for amino acid–base interaction: implications for prediction of protein–DNA binding sites. *Nucleic Acids Res.*, **26**, 2306–2312.
34. Matthews, B.W. (1988) Protein–DNA interaction. No code for recognition. *Nature*, **335**, 294–295.
35. Suzuki, M. and Yagi, N. (1994) DNA recognition code of transcription factors in the helix–turn–helix, probe helix, hormone receptor and zinc finger families. *Proc. Natl Acad. Sci. USA*, **91**, 12357–12361.
36. Suzuki, M. (1994) A framework for the DNA–protein recognition code of the probe helix in transcription factors: the chemical and stereochemical rules. *Structure*, **2**, 317–326.
37. Choo, Y. and Klug, A. (1994) Toward a code for the interactions of zinc fingers with DNA: selection of randomized fingers displayed on phage [published erratum appears in *Proc. Natl Acad. Sci. USA* (1995) **92**, 646]. *Proc. Natl Acad. Sci. USA*, **91**, 11163–11167.
38. Choo, Y. and Klug, A. (1994) Selection of DNA binding sites for zinc fingers using rationally randomized DNA reveals coded interactions. *Proc. Natl Acad. Sci. USA*, **91**, 11168–11172.
39. Choo, Y. and Klug, A. (1995) Designing DNA-binding proteins on the surface of filamentous phage. *Curr. Opin. Biotechnol.*, **6**, 431–436.
40. Greisman, H.A. and Pabo, C.O. (1997) A general strategy for selecting high-affinity zinc finger proteins for diverse DNA target sites. *Science*, **275**, 657–661.
41. Irvine, D., Tuerk, C. and Gold, L. (1991) SELEXION. Systematic evolution of ligands by exponential enrichment with integrated optimization by non-linear analysis. *J. Mol. Biol.*, **222**, 739–761.
42. Jamieson, A.C., Wang, H. and Kim, S.H. (1996) A zinc finger directory for high-affinity DNA recognition. *Proc. Natl Acad. Sci. USA*, **93**, 12834–12839.
43. Wolfe, S.A., Greisman, H.A., Ramm, E.I. and Pabo, C.O. (1999) Analysis of zinc fingers optimized via phage display: evaluating the utility of a recognition code. *J. Mol. Biol.*, **285**, 1917–1934.
44. Desjarlais, J.R. and Berg, J.M. (1994) Length-encoded multiplex binding site determination: application to zinc finger proteins. *Proc. Natl Acad. Sci. USA*, **91**, 11099–11103.
45. Ruscher, K., Reuter, M., Kupper, D., Trendelenburg, G., Dirnagl, U. and Meisel, A. (2000) A fluorescence based non-radioactive electrophoretic mobility shift assay. *J. Biotechnol.*, **78**, 163–170.