4322–4334 *Nucleic Acids Research, 2005, Vol. 33, No. 13*
doi:10.1093/nar/gki708

Determination of protein–DNA binding constants and specificities from statistical analyses of single molecules: MutS–DNA interactions

Yong Yang[1], Lauryn E. Sass[1], Chunwei Du[3], Peggy Hsieh[3] and Dorothy A. Erie[1,2,*]

[1]Department of Chemistry and [2]Curriculum in Applied and Materials Sciences, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-3290, USA and [3]Genetics and Biochemistry Branch, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892, USA

Received May 4, 2005; Revised June 13, 2005; Accepted June 29, 2005

ABSTRACT

Atomic force microscopy (AFM) is a powerful technique for examining the conformations of protein–DNA complexes and determining the stoichiometries and affinities of protein–protein complexes. We extend the capabilities of AFM to the determination of protein–DNA binding constants and specificities. The distribution of positions of the protein on the DNA fragments provides a direct measure of specificity and requires no knowledge of the absolute binding constants. The fractional occupancies of the protein at a given position in conjunction with the protein and DNA concentrations permit the determination of the absolute binding constants. We present the theoretical basis for this analysis and demonstrate its utility by characterizing the interaction of MutS with DNA fragments containing either no mismatch or a single mismatch. We show that MutS has significantly higher specificities for mismatches than was previously suggested from bulk studies and that the apparent low specificities are the result of high affinity binding to DNA ends. These results resolve the puzzle of the apparent low binding specificity of MutS with the expected high repair specificities. In conclusion, from a single set of AFM experiments, it is possible to determine the binding affinity, specificity and stoichiometry, as well as the conformational properties of the protein–DNA complexes.

INTRODUCTION

Understanding protein–DNA interactions is fundamentally important for dissecting the molecular mechanisms underlying many biological processes. Association constants and specificities of protein binding to DNA are the primary thermodynamic properties for understanding protein–DNA interactions. Many methods, such as electrophoretic mobility shift assays (EMSA), filter binding assays, surface plasmon resonance (SPR) and calorimetric assays are used to investigate the thermodynamic equilibrium constants of protein–DNA interactions (1–5). Although these methods are very powerful, they all have two significant limitations. First, all are bulk measurements; therefore, the observed affinities are the weighted sum of all interactions occurring between the protein and the DNA (Figure 1a) (6). For example, if a protein has a significant binding affinity for the ends of the DNA, the apparent binding constant may represent this preference, especially for nonspecific binding. Second, in all of these assays, the measurement of binding is indirect, and it is generally assumed that the signal, such as heat in calorimetry or refractive index in SPR, is linearly proportional to the binding (Figure 1a). While this situation is often the case, there are many cases when this assumption is not valid (2).

A single molecule method to determine protein–DNA binding constants can overcome these limitations. Accordingly, we have developed a single molecule method using atomic force microscopy (AFM) to determine protein–DNA binding constants and specificities directly at the level of DNA-binding sites (DNA_i, Figure 1b). Using AFM, it is possible not only to determine the extent of binding of a protein to a DNA fragment but also to determine where on the DNA the protein is bound, (i.e. fractional occupancies at the specific site, the ends, or nonspecific sites, Figure 1b). In the sections that follow, we present the theory that demonstrates how binding affinities, as well as binding specificities, can be determined from the analysis of AFM images of protein–DNA complexes. We then demonstrate its utility and accuracy by analyzing complexes of the mismatch repair protein MutS with DNA and comparing our results with those of bulk measurements.

The MutS family of proteins is highly conserved and the sole known factor for the recognition of DNA mismatches and

*To whom correspondence should be addressed. Tel: +1 919 962 6370; Fax: +1 919 966 3675; Email: derie@unc.edu

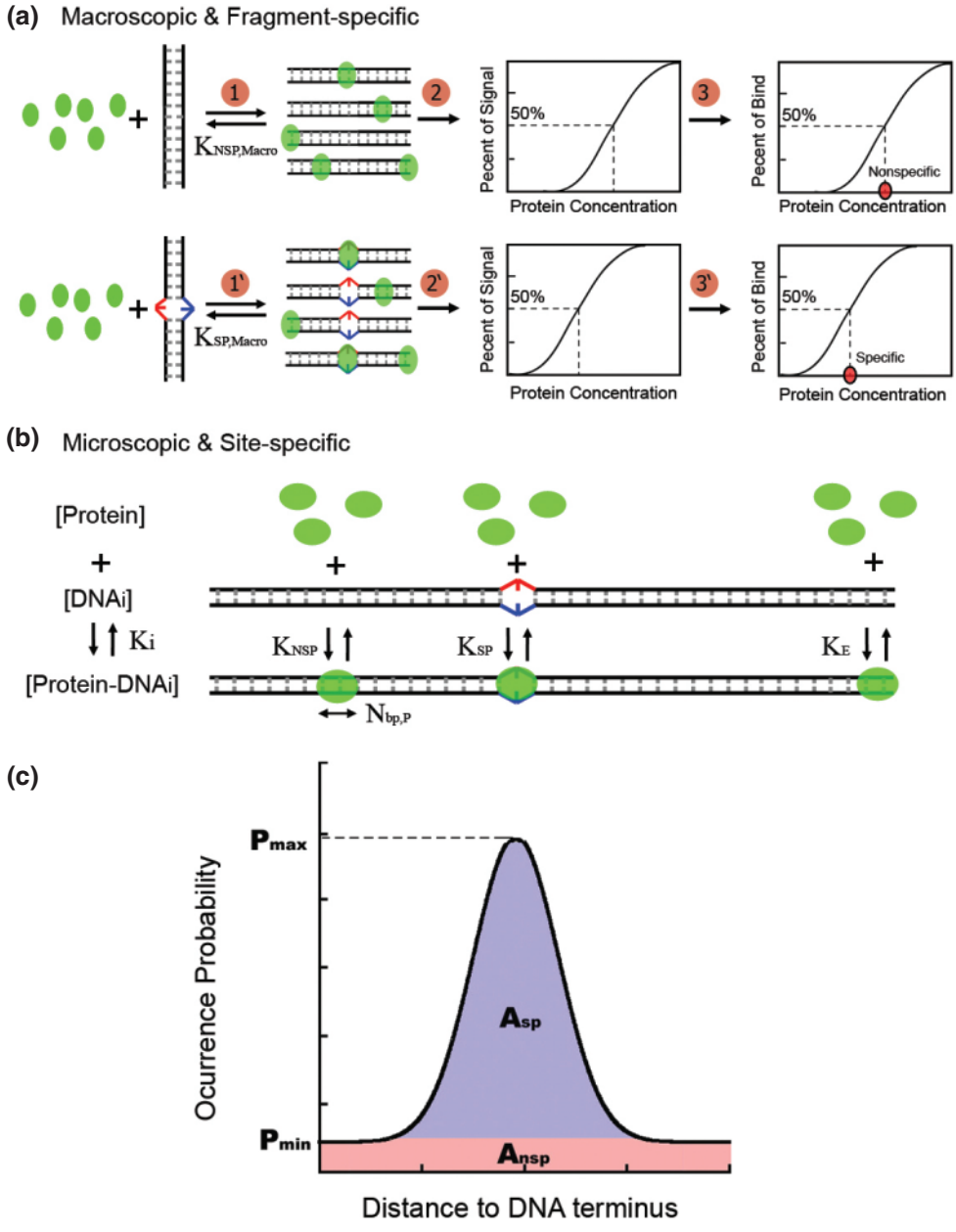

Figure 1. Illustration of the differences in determining protein–DNA binding constants and specificities by bulk methods (a) and single molecule methods (b). (**a**) In bulk assays, binding constants and specificities are determined by measuring the extent of protein binding to DNA fragments with and without a specific site. (1 and 1′). Specific site denoted as colored distortion in the DNA. In general, bulk methods cannot distinguish different types of binding interactions, such as specific, nonspecific and end binding; therefore, the binding constants will represent a weighted sum of all types of binding (2 and 2′). In addition, the extents of binding are determined indirectly from measuring changes in some signals, such as heat, absorbance, etc., and assuming that the change in signal is directly proportional to the extent of binding (3 and 3′). (**b**) Microscopic binding constants and specificities determined by AFM. Binding constants for a given DNA-binding site (DNA_i) are determined directly from the concentrations of the free protein ([protein]), of the protein bound at the given site ([Protein–DNA_i]) and of the unoccupied given site ([DNA_i]), i.e. K_i = [Protein–DNA_i]/([DNA_i] × [Protein]). The number of DNA base pairs covered by a protein ($N_{bp,P}$), which is required for the analysis (see Theory), can be obtained from the crystal structure or footprint of protein–DNA complexes, or it can be estimated from the size of complexes in AFM images. Specificities can be obtained either by comparing the binding constant at the specific site to that at a nonspecific site (i.e. $S = K_{SP}/K_{NSP}$) or by analyzing the position distribution without any knowledge of absolute binding constants [See below in (c)]. (**c**) Illustration of the position distribution of protein binding along the DNA fragment with a single specific site in the middle. The areas under the position distribution of protein–DNA complexes, A_{nsp} (red) and A_{sp} (blue), are illustrated. The observed minimum and maximum occurrence probabilities (P_{min} and P_{max}) are labeled. The specificity to the specific site can be obtained by analyzing this distribution (see Equation 8).

small insertion/deletion loops in the DNA mismatch repair (MMR) pathway (7,8). MutS–DNA interactions are crucial in the regulation of the initiation of MMR (9,10). Studies on *Escherichia coli* MutS and eukaryotic MutS homologs using traditional bulk techniques show that the binding specificities to various mismatches are very low (~30 or less) (11,12). This relatively low binding specificity to mismatches versus much higher expected MMR specificity is one of central puzzles in MMR (13,14). Interestingly, EMSA studies of *Taq* MutS binding to the single T-bulge, however, suggest a

much higher binding specificity (>1000), although the specificities for other mismatches are similarly low (11). In this paper, we present a detailed analysis of MutS–DNA interactions using AFM. Our results indicate that the binding specificities of MutS are greatly underestimated in the previous studies and suggest that this underestimation is due, in part, to a high affinity of MutS to DNA ends.

THEORY

Site-specific binding constant

From the lattice binding model of protein–DNA interactions (5,15), a protein interacts at another DNA-binding site whenever it moves 1 bp or more away from the current binding position. In other words, the number of binding sites (N) on a linear DNA fragment is $N_{bp} - N_{bp,P} + 1$, where N_{bp} is the length of the fragment in base pairs and $N_{bp,P}$ is the number of base pairs occupied by protein (5,15). (If end binding is a different mode than nonspecific binding, then the total number binding sites on the DNA fragment will increase by 2, i.e. $N = N_{bp} - N_{bp,P} + 3$.) The number of binding sites on a circular DNA fragment is equal to the number of base pairs (i.e. $N = N_{bp}$). Assuming that all sites are independent, the binding constant, K_i, to a given site i is:

$$K_i = \frac{[\text{Protein–DNA}_i]}{[\text{DNA}_i]\times[\text{Protein}]} = \frac{O_i}{\left(1 - O_i - \sum_{j=\text{Max}\{1,\, i-(N_{bp,P}-1)\}}^{\text{Min}\{N,\, i+(N_{bp,P}-1)\}} O_{j\neq i}\right)\times\left([P]-[D]\times O_{\text{Fragment}}\right)}, \quad \mathbf{1}$$

where $[\text{Protein–DNA}_i]$ is the concentration of protein bound to site i, $[\text{DNA}_i]$ and [Protein] are the concentrations of free site i and free protein, respectively, O_i is the fractional occupancy of DNA site i by protein ($O_i = [\text{Protein–DNA}_i]/[\text{DNA}_i]_{\text{Total}}$, where $[\text{DNA}_i]_{\text{Total}}$ is the total concentration of site i), O_{Fragment} is the average number of proteins bound per DNA fragment ($O_{\text{Fragment}} = \sum_{j=1}^{N} O_j$), and [P] and [D] are the total concentrations of protein and DNA fragments, respectively. The right hand side of Equation 1 is derived by expressing the concentrations in terms of occupancies. Specifically, $[\text{Protein–DNA}_i] = O_i \times [\text{DNA}_i]_{\text{Total}}$, $[\text{Protein}] = [\text{P}] - [\text{D}] \times O_{\text{Fragment}}$ (i.e. the free protein concentration equals the total protein concentration minus the concentration of protein bound to the DNA), and

$$[\text{DNA}_i] = \left(1 - O_i - \sum_{j=\text{Max}\{1,\, i-(N_{bp,P}-1)\}}^{\text{Min}\{N,\, i+(N_{bp,P}-1)\}} O_{j\neq i}\right)\times[\text{DNA}_i]_{\text{Total}}.$$

The term $\sum_{j=\text{Max}\{1,\, i-(N_{bp,P}-1)\}}^{\text{Min}\{N,\, i+(N_{bp,P}-1)\}} O_{j\neq i}$ is included because the protein-binding site size ($N_{bp,P}$) is >1 bp, and therefore, the concentration of free site i will depend not only on those proteins bound at i, but also on those bound at sites that are within $N_{bp,P} - 1$ bp of i (assuming that $N_{bp,P}$ is independent of position).

Specific and nonspecific binding constants

Normally, only one or a few binding sites on a DNA fragment are specific and all other binding sites are nonspecific with similar protein binding affinities. Consequently, the binding constant for the specific site (K_{SP}) and the average binding constant for nonspecific binding sites (K_{NSP}) are of interest. In addition, for a linear DNA fragment, proteins may have significant binding affinities for DNA ends (K_E). A linear fragment containing N_{SP} specific sites and a total of N binding sites has $N - N_{SP} - 2$ nonspecific sites and two end binding sites. Using Equation 1, the binding constant for a given specific site (K_{SP}), and the average binding constants for a DNA end (K_E) and for a nonspecific site (K_{NSP}) can be expressed in terms of fractional occupancies of the sites. Specifically, for a linear DNA fragment containing N_{SP} specific sites that are separated from each other and from the DNA ends by at least $N_{bp,P}$ nonspecific sites:

$$K_{SP} = \frac{O_{SP}}{\left(1 - O_{SP} - \left(2N_{bp,P} - 2\right)O_{NSP}\right)\times\left([P]-[D]\times O_{\text{Fragment}}\right)} \quad \mathbf{2}$$

$$K_E = \frac{O_E}{\left(1 - O_E - \left(N_{bp,P} - 1\right)O_{NSP}\right)\times\left([P]-[D]\times O_{\text{Fragment}}\right)} \quad \mathbf{3}$$

$$K_{NSP} = (O_{NSP}) \Bigg/ \left(\left[1 - O_{NSP} - (2N_{bp,P} - 2) \times \left(O_{NSP} + \frac{\sum_{j=1}^{N_{SP}}\left(O_{SP,j} - O_{NSP}\right)}{N_{NSP}} + \frac{O_E - \left(N_{bp,P}/2\right)\times O_{NSP}}{N_{NSP}}\right)\right] \times \left([P]-[D]\times O_{\text{Fragment}}\right)\right) \quad \mathbf{4}$$

where O_{SP} is the fractional occupancy of the specific site, O_E the average fractional occupancy of a DNA end, O_{NSP} is the average fractional occupancy of a nonspecific site, and N_{SP} and N_{NSP} are the number of specific sites and nonspecific sites, respectively (the total number of sites $N = N_{NSP} + N_{SP} + 2$). For simplicity, the binding site size ($N_{bp,P}$) of the protein has been assumed to be the same for all sites, including DNA ends. The terms $[\sum_{j=1}^{N_{SP}}(O_{SP,j} - O_{NSP})]/N_{NSP}$ and $[O_E - \left(N_{bp,P}/2\right)\times O_{NSP}]/N_{NSP}$ are included in Equation 4 to account for the occlusion of the nonspecific sites by protein binding at or near the specific sites and the DNA ends, respectively.

Under conditions of low occupancy [$O_{NSP} \ll 1$ and $(O_{SP} + O_E)/N_{NSP} \ll 1$, i.e. DNA is sufficiently long], the binding constants for a linear fragment containing a single specific site can be approximated within experimental error as:

$$K_{SP} \approx \frac{O_{SP}}{(1 - O_{SP})\times\left([P]-[D]\times O_{\text{Fragment}}\right)} \quad \mathbf{5}$$

$$K_E \approx \frac{O_E}{(1 - O_E)\times\left([P]-[D]\times O_{\text{Fragment}}\right)} \quad \mathbf{6}$$

$$K_{NSP} \approx \frac{O_{NSP}}{\left([P]-[D]\times O_{\text{Fragment}}\right)} \quad \mathbf{7}$$

Binding specificity

Specificity is the relative affinity of a protein binding to a specific site versus a nonspecific site, i.e. $S = K_{SP}/K_{NSP}$. It is also the relative occurrence probability of protein binding at different sites on the same DNA fragment. Consequently, it is

not necessary to know the absolute binding constants to determine the specificity, and AFM provides a straightforward method for estimating binding specificity (16–18). If a piece of DNA contains no specific sites, a uniform distribution of protein is seen on the DNA; however, if the DNA contains a single specific site, a Gaussian distribution of protein will be observed centered at the specific site. The Gaussian distribution is a result of the error in the determination of the position of the protein. All of the proteins in the Gaussian distribution (A_{sp}, Figure 1c) result from the presence of the specific site. Accordingly, the specificity, S, can be determined by integrating the areas under the Gaussian curve of the protein distribution (A_{sp} and A_{nsp}, Figure 1c). The binding specificity can be given by the following mathematically different but physically consistent forms (see derivation in Supplementary Material):

$$S = \frac{P_{sp}}{P_{nsp}} = N \times \frac{A_{sp}}{A_{nsp}} + 1 = \frac{A_{sp}}{P_{min}} + 1 = N \times \left(\frac{P_{avg}}{P_{min}} - 1\right) + 1 \quad \textbf{8}$$

where P_{min} is the average occurrence probability of nonspecific sites (Figure 1c), and P_{avg} is the average occurrence probability for all N binding sites. The first portion of the equation defines the binding specificity as the probability of protein binding to one specific site (P_{sp}) divided by the probability binding to one nonspecific site (P_{nsp}). The second portion shows that the determination of specificity does not depend on whether position distributions are plotted as absolute positions or as relative positions, because A_{sp}/A_{nsp} is independent of the scaling of the x-coordinate. The third portion discloses that the accuracy of specificity is governed by the accuracy of A_{sp} and P_{min}. Consequently, it is not necessary to obtain the position distribution for the full length of a DNA fragment as long as the P_{min} determined by the nonspecific binding is well determined (i.e. as long as the nonspecific sites are sufficiently occupied). The last form was used in this study to facilitate the calculation of specificities using the software Kaleidagraph.

Theoretically, the distribution of protein-binding positions is definite if the number of binding sites (N) and the binding specificity (S) are given. Consequently, another way to determine specificity or to compare specificities of two different binding sites is given by the following correlation of protein binding on any two DNA fragments that contain up to one specific site in each fragment (see derivation in Supplementary Material):

$$S^{II} = \frac{P^{I}_{min}}{P^{II}_{min}}\left(S^{I} + N^{I} - 1\right) - N^{II} + 1 \quad \textbf{9}$$

where the superscripts indicate interactions with two different DNA fragments. In particular, if there is no specific site in the first fragment (i.e. $S^{I} = 1$ and $P^{I}_{min} = 1/N^{I}$), determining the specificity in the second fragment simply requires determining the probability of protein binding at nonspecific sites (P^{II}_{min}), because $S^{II} = (1/P^{II}_{min}) - N^{II} + 1$. Consequently, this correlation provides a simpler way to determine the binding specificity as long as the binding probability on nonspecific sites (P_{min}) can be determined. This method is especially useful under the conditions where determining the integration of A_{sp} is difficult, e.g. if the position distribution is not fit well by a Gaussian distribution.

Error of measurements

The error induced by the uncertainty in O_i and $O_{Fragment}$ can be determined from multiple AFM experiments and the relative errors of the binding constant and specificity can be calculated by applying the theory of error propagation (19) for Equations 5–7 and $S = K_{SP}/K_{NSP}$. The relative errors are:

$$\frac{\sigma_{K_i}}{\overline{K_i}} \approx \sqrt{\left(\frac{\sigma_{O_i}}{\overline{O_i}}\right)^2 + \left(\frac{\sigma_{O_i}}{1-\overline{O_i}}\right)^2 + \left(\frac{[D] \times \sigma_{O_{Fragment}}}{[P]-[D] \times \overline{O_{Fragment}}}\right)^2}, \quad i = SP \text{ or } E \quad \textbf{10}$$

$$\frac{\sigma_{K_{NSP}}}{\overline{K_{NSP}}} \approx \sqrt{\left(\frac{\sigma_{O_{NSP}}}{\overline{O_{NSP}}}\right)^2 + \left(\frac{[D] \times \sigma_{O_{Fragment}}}{[P]-[D] \times \overline{O_{Fragment}}}\right)^2} \quad \textbf{11}$$

$$\frac{\sigma_S}{\overline{S}} \approx \sqrt{\left(\frac{\sigma_{O_{SP}}}{\overline{O_{SP}}}\right)^2 + \left(\frac{\sigma_{O_{SP}}}{1-\overline{O_{SP}}}\right)^2 + \left(\frac{\sigma_{O_{NSP}}}{\overline{O_{NSP}}}\right)^2} \quad \textbf{12}$$

where 'σ' represents the standard error of physical variables and the upper bar '$\bar{}$' represents the mean of physical variables in multiple AFM experiments. Inspection of these relations shows that the relative errors of binding constants and specificities are close to the relative errors of the DNA fractional occupancy, given conditions of low occupancy (O_{SP}, $O_E < 0.5$ and $O_{NSP} \ll 1$) and $[P] > 2 \times [D] \times \overline{O_{Fragment}}$ (i.e. at least 50% proteins are unbound), which are fulfilled in this study of MutS.

Relationship between site-specific and fragment-specific binding constants and specificities

The macroscopic DNA-binding constant determined by bulk solution measurements is a measure of the protein-binding affinity to the entire DNA fragment (Figure 1a). Consequently, to compare results from AFM, which provides a measure of binding affinity to individual sites (Figure 1b), it is necessary to calculate the binding constant to the entire DNA fragment from AFM microscopic constants. For a linear DNA fragment, it can be shown, based on the lattice binding model of protein–DNA interactions, that the binding constants to DNA fragments with N binding sites that contain either a single specific site ($K_{SP,Macro}$) or no specific sites ($K_{NSP,Macro}$) are:

$$K_{SP,\,Macro} = (N-3) \times K_{NSP} + 2 \times K_E + K_{SP} \quad \textbf{13}$$

$$K_{NSP,\,Macro} = (N-2) \times K_{NSP} + 2 \times K_E \quad \textbf{14}$$

Similarly, the bulk specificities are determined from the ratio of binding affinities to two DNA fragments: one with and one without the specific site (Figure 1a); whereas, AFM yields a direct measure of the relative affinities for different sites (Equation 8). The macroscopic specificity can be calculated from the ratio of $K_{SP,Macro}/K_{NSP,Macro}$ (Equations 13 and 14). Inspection of Equations 13 and 14 shows that both specificities and binding constants determined by bulk measurements may be skewed if specificities are low such that nonspecific binding makes a significant contribution to the overall binding or if

there is significant affinity for DNA ends. On the other hand, for proteins that have high specificities and low end binding affinities, the macroscopic and microscopic binding constants and specificities should be similar. EcoRI, which has a very high specificity ($\sim 10^6$), provides an example of the latter case. Specifically, a recent single-molecule study, which used an optical trap to unwind a DNA double helix with EcoRI bound, estimated the binding constant of EcoRI to its specific site and found it to be very similar to the bulk measurements (20). Finally, in comparing bulk measurements with AFM, it is important to keep in mind the limitations of the methods for determining different types of binding. For example, protein–DNA interactions that are very dynamic (rapid on and off rates) are often not detected in EMSA or filter binding assays (21). In contrast, with AFM, such interactions should be detected approximately as well as those with slow dissociation rates, because deposition of both proteins and DNA on mica is diffusion-limited, and both bind to mica irreversibly over the time scale of the deposition (seconds to minutes) (22–25). Consequently, AFM detection of protein–DNA interactions is normally not affected by the non-linear response; whereas it is a common problem for bulk methods (2).

Limitations of AFM

The primary assumption in using AFM to determine protein–DNA association constants is that the populations of bound and free DNA on the surface are the same as those in solution, i.e. that deposition on the surface does not alter the populations (23). If the free DNA deposits more efficiently onto the surface than the protein–DNA complexes, or if the surface causes the protein to dissociate from the DNA, the apparent binding constant determined by AFM will be less than the actual constant. One case where we have encountered this problem is for proteins that induce a 3D topology in the DNA such that the protein–DNA complex must be distorted to lie flat on the surface (M. Guthold, O.K. Wong, J. Gelles and D.A. Erie, unpublished data). Another case is that rinsing the AFM surface during sample preparations may cause the dissociation of protein–DNA complexes, especially for small proteins that have limited interaction with the surface when bound to DNA (M. Guthold, O.K. Wong, J. Gelles and D.A. Erie, unpublished data). In this latter case, the protein is often seen on the surface near the DNA (M. Guthold, O.K. Wong, J. Gelles and D.A. Erie, unpublished data), because the protein generally binds to the surface after dissociating from the DNA. In contrast, it is possible that the protein–DNA complex may deposit more efficiently than the free DNA, in which case the apparent binding constant would be overestimated. This latter case may be a problem if there is a very high occupancy of protein on the DNA; however, for long DNAs with low occupancy of protein, it is unlikely that the protein will change the efficiency of deposition of the DNA. For example, DNA fragments with streptavidin–horseradish peroxidase fusion protein bound to both ends exhibit the same rate and efficiency of deposition as the unbound fragments and no significant change in conformation (23,24). The binding constants could also be overestimated if the amount of the protein deposited on the surface is so high that there is a high probability that the protein coincidentally lands on the DNA (see below). This latter case may become a problem if the binding interaction is weak and a high concentration of protein is required to observe binding. In general, AFM should be a good method for determining binding constants as long as the occupancy of protein on the DNA is low and the protein does not fold the DNA into a 3D structure. Notably, our method does not depend on the relative populations of free protein and protein–DNA complexes, but only on that of free DNA and protein–DNA complexes. In addition, although both selective and systemic alterations in the DNA occupancy by the surface can affect absolute constants, only biased alterations between the occupancy on the specific site and that on nonspecific sites will affect specificities. Such bias could occur if the conformations of the specific and nonspecific protein–DNA complexes are significantly different.

Estimate of maximal contribution of random landing events to observed occupancies

The probability that a protein lands close enough to the DNA to be counted as a complex can be estimated from the size of the protein and the surface area covered by the DNA. To estimate the maximum possible contribution of a protein randomly landing on the DNA to the total observed complexes, we use a simple model. If we model the DNA as a cylinder and the protein as a sphere, we can define an area on the surface around the DNA in which a protein would be defined as bound. For a cylinder of width w and length L and a sphere of diameter d, the area is approximately $(d + w) \times (d + L)$. Accordingly, the probability of a single protein randomly landing within this area is

$$P_{\mathrm{P}} = \frac{(d+w) \times (d+L)}{A_{\mathrm{TOT}}} \qquad \textbf{15}$$

where A_{TOT} is the total area of the image. For a given surface area (A_{TOT}) containing n_{P} proteins, the total probability that a protein randomly lands on a given DNA fragment on the surface is $n_{\mathrm{P}}P_{\mathrm{P}}$. The fractional contribution of a protein randomly landing on the DNA to the observed occupancy of a fragment, O_{Fragment}, is $n_{\mathrm{P}}P_{\mathrm{p}}/O_{\mathrm{Fragment}}$. The choice of d and w can be based on two criteria in a real analysis. First, it is reasonable to define w as the diameter of DNA (2.5 nm) and d as the diameter of the protein. Alternatively, $d + w$ can be experimentally determined from images by measuring the maximal distance from the center of the DNA that a protein can be counted as a bound protein [i.e. $(d + w)/2 =$ interaction distance].

MATERIALS AND METHODS

MutS protein and DNA substrates

Taq MutS was expressed and purified as described previously (11,26). Three linear DNA fragments with blunt DNA ends, named 782Homo, 783TBulge and 982GT, were generated and purified as described previously (13), where the number represents the length of DNA in base pairs followed by the text representing homoduplex DNA or the single mismatch at a dedicated position on the DNA. For 783TBulge, an extra T is 213 bp (27%) away from its closest DNA terminus. For 982GT, a GT mismatch is 412 bp (42%) away from its closest terminus. Two other linear homoduplex DNA fragments with 3′-overhang DNA ends, named 817Puc18 and 1869Puc18,

were obtained by digesting Puc18 circular plasmid with the restriction enzyme, DrdI, where 817 and 1869 are the length of the resulted linear fragments. Two fragments were purified together with GFX DNA purification kit (Amersham Pharmacia Biotech), and incubated together with MutS, but they were identified and analyzed separately by their different lengths in AFM images.

AFM imaging and analysis

The MutS–DNA reactions were carried out by incubating 12–25 nM *Taq* MutS (dimer) with 0.4–3 nM DNA substrates (double strand) for 1–3 min at room temperature (23°C) or at 65°C in a binding buffer of 20 mM HEPES, pH 7.8, 50 mM NaCl and 5 mM $MgCl_2$. The different incubation times did not affect the measured DNA occupancies, suggesting that equilibrium was established prior to deposition. The reaction mixture was deposited onto freshly cleaved mica surface (Spruce Pine Mica Company) at 23 or 65°C, incubated for <1 min on the mica, rinsed with deionized water, dried under a gentle flow of nitrogen and imaged as described previously (13). The extent of rinsing did not have significant effect on the population or conformation of the complexes on the surface. Specifically, comparison of images where the surface has been rinsed only a couple of times with those that have been rinsed dozens of times does not result in any significant differences in coverage. It is possible that some complexes were lost in the initial rinse; however, analyses of protein and DNA binding to mica indicate that they bind irreversibly over the time scale of our depositions (22–25). All images were captured with a Nanoscope IIIa microscope (Digital Instruments) in oscillating mode. Pointprobe® oscillating mode silicon probes (Molecular Imaging Cooperation) with spring constants of $\sim$50 Nm^{-1} and resonance frequencies of $\sim$170 kHz were used. All images were collected at the scan size of 1 μm × 1 μm, scan speed of 3 Hz and resolution of 512 × 512 pixels.

Generation of position distributions and determination of binding specificities

The MutS–DNA complexes were selected for the position measurement independent of whether or not multiple proteins were bound on a single DNA fragment. Only MutS molecules that completely overlapped with the DNA were counted as complexes. Only complexes in which DNA contour lengths were within the standard deviation of the DNA length were used in the position histograms. The distance from the center of complex '*i*' to its closest DNA terminus (d_i) and the contour length of DNA '*i*' (L_i) were measured in the program Nanoscope 5.12r3 (Digital Instruments). The reproducibility of this measurement was confirmed by conducting measurements in triplicate, which indicated that the error in determining the length of a given fragment is $\sim$1%. The position of the complex '*i*' is defined by $X_i = d_i/L_i$. From a large number of complexes, the position histograms were plotted between the positions from 10 to 50% away from the closest end, because two DNA ends are not distinguishable in AFM images and the shortest distance to ends was used to define the position. The histograms were presented as occurrence probability $\left(P_i = n_i/\{N_{\mathrm{bp,bin}} \times \sum n_i\}\right)$ versus position, where '*i*' represents the position of the individual bins, $n = \sum n_i$ is the total number of binding occurrences observed within the position range (10–50% DNA full length away from the closest DNA end), and $N_{\mathrm{bp,bin}}$ is the number of DNA base pairs in each position bin. Only complexes that were ⩾10% from the end were counted because of the larger error in determining the absolute positions for the complexes close to the ends. This procedure makes the assay of specificities more efficient and has little effect on the characterization of specificities (see the discussion under Equation 8). The program Kaleidagraph (Synergy Software) was used to fit the position histograms into the position distributions. For DNA fragments containing a mismatch site (the specific site), the equation $P = m_1 + m_2 \times \exp\{-[(X - m_3)/m_4)]^2\}$ was used for the fitting, where m_1–m_4 were the fitting variables. This equation represents a weighted sum of one uniform distribution ($P = m_1$) and one Gaussian distribution ($P = m_2 \times \exp\{-[(X - m_3)/m_4)]^2$). This statistical analysis is reasonable because MutS binding to mismatch-containing DNA fragments can be viewed as the sum of the binding to the homoduplex DNA and to a single mismatch. From position distributions, binding specificities were determined using Equation 8.

Determination of DNA fractional occupancies and binding constants

The occurrence probability in position histograms was not directly used for deducing the actual occupancies of individual DNA sites, because the occurrence probability at a given site will be affected by the occurrence probability at nearby sites due to the error in distance measurements. Instead, we have used the following counting of protein–DNA complexes and DNA fragments. The total number of DNA fragments (n_{Fragment}), the total number of MutS–DNA complexes at internal DNA contours ($n_{\mathrm{Complex,Int}}$) and the total number of DNA termini bound by MutS ($n_{\mathrm{Complex,Ter}}$) were counted from a set of AFM images for each MutS–DNA reaction. The DNA fragments on the edge of images, partial DNA fragments and overlapping DNA fragments were excluded from the counting. The fractional occupancy of the DNA fragment (O_{Fragment}), the apparent fractional occupancy of DNA termini (O_{Ter}) and the average fractional occupancy of internal binding sites ($O_{\overline{N}}$) were determined by $O_{\mathrm{Fragment}} = \left(n_{\mathrm{Complex,\,Int}} + n_{\mathrm{Complex,\,Ter}}\right)/n_{\mathrm{Fragment}}$, $O_{\mathrm{Ter}} = n_{\mathrm{Complex,\,Ter}}/\left(n_{\mathrm{Fragment}} \times 2\right)$ and $O_{\overline{N}} = n_{\mathrm{Complex,\,Int}}/\left(n_{\mathrm{Fragment}} \times N_{\mathrm{int}}\right)$. The number of binding sites at internal DNA contours (N_{int}) was assumed to be $N_{\mathrm{bp}}-50$, because 25 bp at the vicinity of each DNA terminus were allocated as the range of end binding based on the size of MutS in AFM images, due to the resolution limitation of AFM. To deduce the fractional occupancies of specific, nonspecific and end binding sites from O_{Ter} and $O_{\overline{N}}$, DNA-binding sites of MutS were categorized into the internal mismatch site (the specific site), internal homoduplex sites (nonspecific sites) and DNA ends. For MutS interacting with homoduplex DNA, the nonspecific fractional occupancy is $O_{\mathrm{NSP}} = O_{\overline{N}}$, whereas for MutS interacting with the mismatched DNA, the nonspecific and specific fractional occupancies can be partitioned from $O_{\overline{N}}$ based on the specificity and the number internal binding sites (N_{int}):

$$O_{\mathrm{NSP}} = \left(\sqrt{\left(N_{\mathrm{int}} \times S \times O_{\overline{N}} - N_{\mathrm{int}} - S\right)^2 + 4{N_{\mathrm{int}}}^2 \times S \times O_{\overline{N}}} + N_{\mathrm{int}} \times S \times O_{\overline{N}} - N_{\mathrm{int}} - S\right) \Big/ \left(2 \times N_{\mathrm{int}} \times S\right)$$

and

$$O_{SP} = \frac{S \times O_{NSP}}{S \times O_{NSP} + 1}$$

(see Supplementary Material for the derivation). For all the cases, the fractional occupancy of DNA ends is given by $O_E = \left(n_{Ter} - n_{Fragment} \times 2 \times 25 \times O_{NSP}\right)/\left(n_{Fragment} \times 2\right) = O_{Ter} - 25 \times O_{NSP}$. It is necessary to subtract $25 \times O_{NSP}$ from O_{Ter} to calculate the occupancies at the DNA ends (O_E) because nonspecific binding of MutS at up to 25 sites away from the ends were likely counted as end binding, due to the size of MutS in AFM images. The relative standard errors of the occupancies were determined by at least three independent measurements.

To estimate the maximum possible contribution of proteins randomly landing on DNA to the observed occupancies on the DNA fragments, we used Equation 15, the diameter of DNA ($w = 2.5$ nm) and the average diameter of MutS based on the crystal structure ($d = 7$ nm) and the length (L) of the DNA used. For the highest concentration of protein used for the 23°C depositions (20 nM), an average of ~100 proteins are seen in a 1 μm × 1 μm image ($n_P = 100$ and $A_{TOT} = 1\ \mu m^2$). For the GT-containing DNA, L is 320 nm. Accordingly, $n_P P_P / O_{Fragment} = 0.22$ ($O_{Fragment} = 1.4$, Table A in Supplementary Material).

As can be seen from the inspection of Equations 5–7, to calculate site-specific binding affinities, it is necessary to know the binding site size of the protein ($N_{bp,P}$) and the total concentrations of protein, [P], and DNA, [D], in addition to the site occupancies. $N_{bp,P}$ was set to be 15 based on the inspection of the crystal structures of MutS–DNA complexes (27–29). [P] was calculated as the dimer concentration because *Taq* MutS exists primarily as a dimer in the absence and presence of DNA at the concentration used in this study (13,26,30). The relative errors of the binding constant and specificity were calculated based on the error of occupancies using Equations 10–12, from which the standard errors were obtained by multiplying the relative error with the mean. The binding constants determined at different concentrations were within error of one another and the reported constants and standard deviations are the average of the constants determined from different concentrations.

DNA substrates and procedures for fluorescence measurements

DNA substrates for fluorescence measurements were purchased high-performance liquid chromatography (HPLC)-purified from MWG Biotech, Inc. and Integrated DNA Technologies, Inc. TAMRA-labeled ssDNA (5′-TACCTCATCTCGAGCGTGCCGATA-TAMRA-3′) was annealed with complementary strands to create T-bulge dsDNA (5′-TATCGGCACGTCTCGAGATGAGGTA-3′), GT mismatch dsDNA (5′-TATCGGCACGTTCGAGATGAGGTA-3′) and homoduplex dsDNA (5′-TATCGGCACGCTCGAGATGAGGTA-3′). The oligonucleotides were annealed in buffer containing 50 mM HEPES, pH 7.8, 100 mM NaCl and 5 mM $MgCl_2$ in a 1:1 ratio at 55°C for 20 min then slowly cooled to room temperature. Binding reactions were performed at 23°C in the same binding buffer as used for the AFM experiments. DNA concentrations used were between 5 and 100 nM. *Taq* MutS was incubated with DNA for 5 min prior to measurement acquisition.

Fluorescence anisotropy was measured using a Jobin Yvon Horiba Fluorolog-3 fluorometer in T-format equipped with a Wavelength Electronics temperature control box. TAMRA-labeled dsDNA substrates were excited at 535 nm and emission was measured at 582 nm. Excitation and emission slit widths were set between 5.0 and 7.0 nm. Intensity measurements were corrected for the dark photon count, and fluorescence anisotropy was calculated using the software provided by the instrument. The fluorescence anisotropy was measured as a function of MutS concentration.

Fluorescence binding data analysis

The fluorescence anisotropy is plotted as a function of the MutS (dimer) concentration. The binding curves for the T-bulge DNA and the 10 nM GT-DNA were fit by a weighted nonlinear regression to a binding isotherm using

$$A = c \bullet \frac{K_d + P_{tot} + D_{tot} - \sqrt{\left(K_d + P_{tot} + D_{tot}\right)^2 - 4P_{tot}D_{tot}}}{2D_{tot}} + A_0 \qquad \textbf{16}$$

where A is the anisotropy, A_0 is the anisotropy of the DNA in the absence of MutS, K_d is the dissociation constant, c is the maximum value of the anisotropy, P_{tot} and D_{tot} are the total protein and DNA concentrations, respectively. A_0, c and K_d were allowed to vary in the fits. The binding curves for the GT-DNA at 30 and 50 nM were biphasic, which was clearly revealed by non-linearity in the Scatchard plots (plotting $[MutS_{Free}]/\upsilon$ versus υ, where υ is the fraction of DNA bound by MutS) (data not shown). This observation indicates that there are two binding events on the GT-DNA. Fitting the binding curves to two binding constants directly was not possible because the fits did not converge. Consequently, we used a both plots of υ versus $[MutS_{Free}]$ and Scatchard plots and to estimate the high affinity binding constant. For the Scatchard plots, the constant was estimated by fitting the linear portion of the plot to a straight line. For υ versus $[MutS_{Free}]$, the curves were fit to the sum of two binding curves $[\upsilon_1[MutS_{Free}]/([MutS_{Free}] + K_{d1}) + (\upsilon_2[MutS_{Free}]/([MutS_{Free}] + K_{d2}))]$. The reported constant is the average from both of these fits. The high affinity constants determined from these fits are consistent with the binding constants from the fits of the 10 nM data to Equation 15. For homoduplex DNA, the change in anisotropy of the DNA upon addition of MutS was too small to be able to obtain binding constants.

RESULTS AND DISCUSSION

Binding specificities and contributions of microscopic constants to macroscopic constants

Representative AFM images of free DNA and DNA in the presence of *Taq* MutS are shown in Figure 2. MutS can be seen bound to the DNA in the deposition in the presence of protein. From a large number of such images, position distributions for *Taq* MutS bound to several different DNA fragments were obtained (Figure 3). As expected, the distributions of MutS bound to DNAs that do not contain mismatches (nonspecific

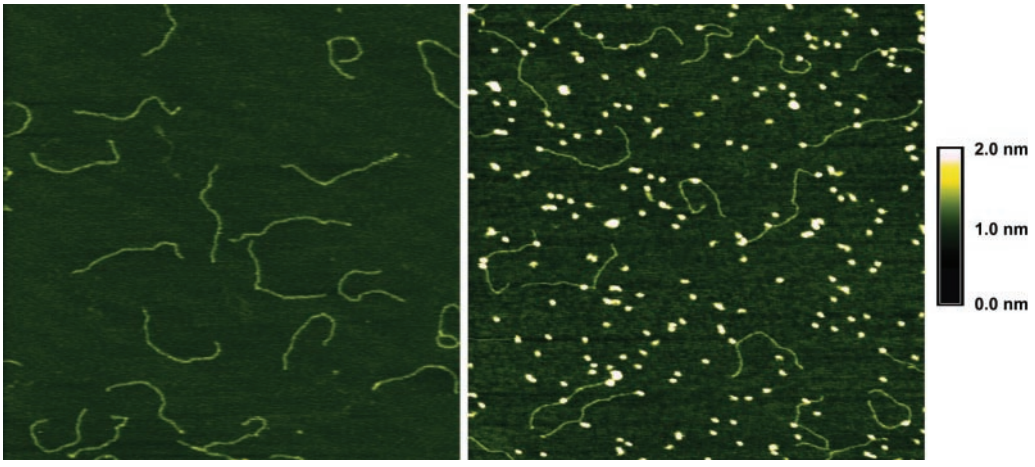

Figure 2. Representative AFM images (1 μm × 1 μm) of free DNA fragments and DNA deposited in the presence of *Taq* MutS. An aliquot of 2 nM 982GT linear DNA fragments (left) and the incubation of 2 nM 982GT with 20 nM MutS (one of the highest concentrations used; right) were deposited onto the mica surface under a similar condition. The coverage of DNA molecules on the mica surface (the number of DNA molecules per a unit of surface) in the presence and in the absence of MutS are very similar to each other based on the statistics of many AFM images.

DNA) are uniform, whereas those for MutS bound to DNAs that contain a single mismatch (specific DNA) are represented by the sum of a uniform and a Gaussian distribution. From these data, we calculated the specificities of *Taq* MutS binding to a T-bulge and a GT mismatch using Equation 8 (Table 1). In addition, we determined the standard errors of the specificity using Equation 12, and the relative accuracy is ~10% (Table 1). The binding specificity of MutS can also be calculated from the binding probability on the nonspecific sites (P_{min}) using Equation 9 or from directly estimating the number of specific and nonspecific complexes (16,17). The obtained specificities using these two alterative methods (data not shown) are consistent with those obtained using Equation 8, which integrates the areas under the Gaussian distribution.

Inspection of Table 1 reveals that *Taq* MutS has a larger binding specificity to a T-bulge than to a GT mismatch (1660 versus 300), which is consistent with data from gel shift assays (11). The specificity for a T-bulge is the same for AFM and gel shift assays (1660 for AFM versus 1700 for EMSA). The consistency between AFM and EMSA for a T-bulge suggests that both methods are suitable for determining the specificity of high affinity protein–DNA interactions. We have also used our method to analyze two other protein–DNA interactions for which position distributions have been published. In one study, binding of the human DNA damage recognition complex, XPC-HR23B, to an 800 bp DNA containing a single cholesterol moiety was investigated (31). Analyzing their raw position histogram yields a specificity of ~2600 to a cholesterol moiety, which is consistent with the biochemical studies in a similar system (32). In another recent study, human 8-oxoguanine DNA glycosylase (hOGG1) binding to a 1024 bp linear DNA containing a single oxoG was investigated (17). Analysis of their raw position histogram yields a binding specificity of 390 for hOGG1 to the single oxoG, which is the same as their estimation of 400 and is consistent with bulk measurements (17). Taken together, these results indicate that AFM is a good method for directly determining the specificities of protein–DNA interactions. In addition, these results support our suggestion that protein–DNA interactions with different dynamics are detected with similar efficiencies using AFM, because the nonspecific complexes are more dynamic than the specific ones.

Interestingly, the specificity for *Taq* MutS binding to a GT mismatch determined by AFM is significantly higher than that determined by EMSA (300 for AFM versus 12 for EMSA). AFM provides a direct measure of specificity, in that the relative probability of the protein bound to different sites on a single DNA fragment are compared, and the relative probabilities are a direct measure of the relative affinities for the different sites. In bulk studies, specificity is determined by measuring the binding affinity to two different DNA fragments: one containing a specific site and one without a specific site. Consequently, to understand the difference between the specificities determined by AFM and EMSA, it is necessary to inspect the contributions of different types of binding to macroscopic binding constants (Equations 13 and 14).

From fractional occupancies of MutS bound to different DNA sites (Supplementary Table A), we calculated the binding affinities of *Taq* MutS to specific and nonspecific sites and to DNA ends using Equations 5–7, as well as the apparent macroscopic binding affinities using Equations 13 and 14 (Table 1). Inspection of Table 1 reveals that *Taq* MutS has a weak binding affinity for nonspecific DNA sites (20–35 μM; the agreement between Puc18 and 782Homo suggests that the engineered DNA fragments are similar to plasmid DNA fragments) and that nonspecific binding makes only a small contribution to the calculated macroscopic binding constants ($K_{DNA,AFM}$) for a DNA fragment containing a T-bulge or a GT mismatch (Table 1). In contrast, *Taq* MutS has a high affinity for DNA ends (~50 nM), similar to that for a GT mismatch (77 nM). Consistent with this result, analysis of the position distribution of *E.coli* MutS on a DNA fragment containing a GT mismatch indicates that *E.coli* MutS binds to DNA ends with an affinity that is only approximately five times less than that of a GT mismatch [Supplementary Material in ref. (13); H. Wang, P. Hsieh and D.A. Erie, unpublished data]. End binding makes an increasingly important contribution to $K_{DNA,AFM}$ as specificity decreases, with $K_{DNA,AFM}$ for

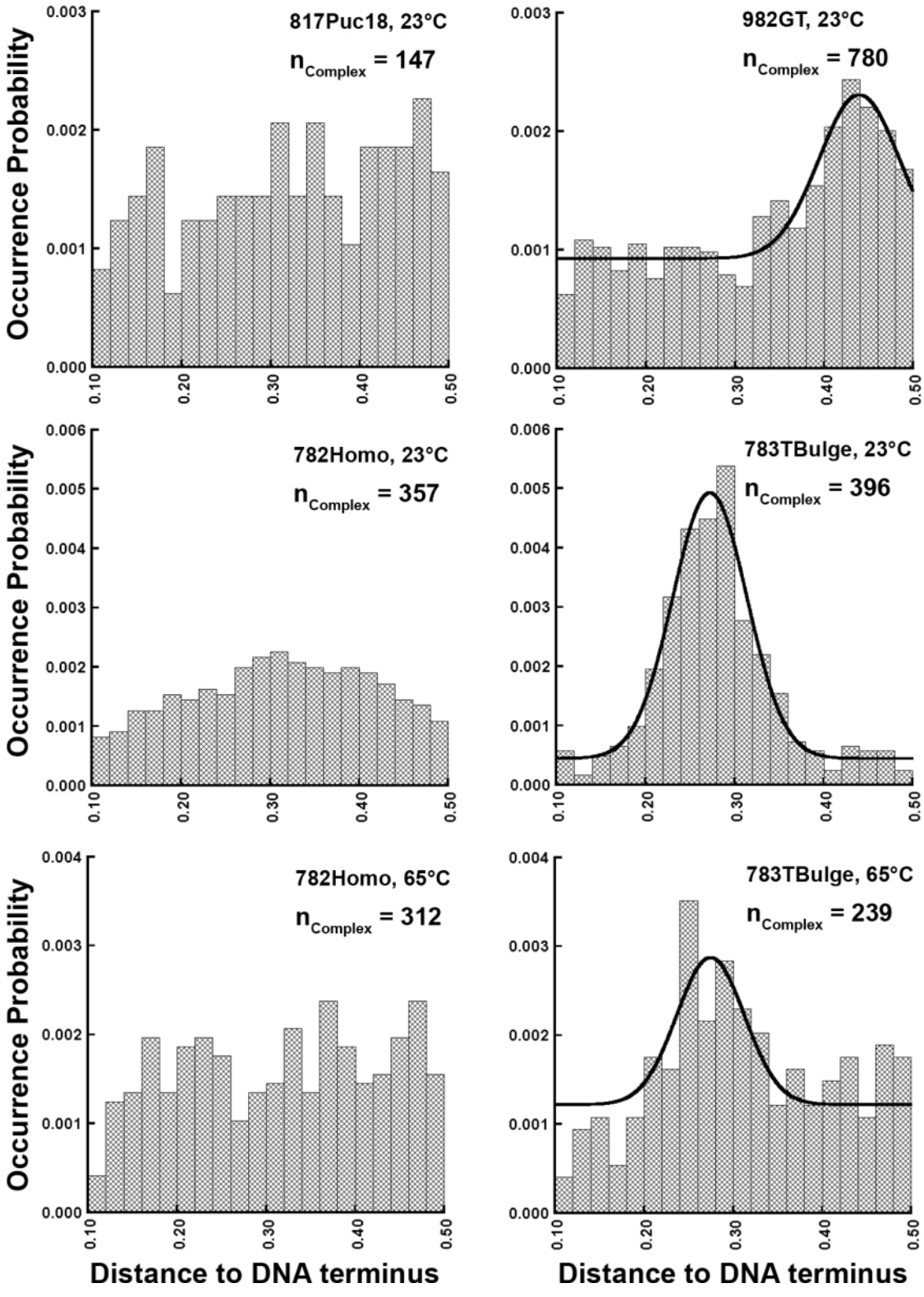

Figure 3. Position histograms of *Taq* MutS–DNA complexes along the DNA. The occurrence probability is the observed probability of MutS binding in a given range of positions, and the position is the relative position to the closest DNA end in the full length of DNA (see Materials and Methods). The DNA fragments, temperature and total number of MutS–DNA complexes are labeled on each plot. Only occupancies of MutS proteins bound at positions ⩾10% of the fragment length away from the DNA ends are included in these plots (the occupancies at the DNA ends can be found in Supplementary Table A). The histogram for 783TBulge is re-plotted from the published work [see Supplementary Material in ref. (13)]. The histogram for 1869Puc18 is similar to that on 817Puc18 (data not shown). The position histograms of MutS on homoduplex DNAs (left panels) are described well by the uniform statistics, indicating that MutS has no significant sequence-dependent DNA binding. The position histograms of MutS on mismatch DNAs (right panels) are described best by the sum of a Gaussian and a uniform distribution ($R^2 = 0.90, 0.96$ and 0.80, respectively from top to bottom), where the solid lines are the fits. Changing the number of position bins between 15 and 30 does not change the distribution.

Table 1. DNA-binding constants and specificities for *Taq* and *E.coli* MutS

		783TBulge (23°C)	982GT (23°C)	782Homo (23°C)	Puc18[a] (23°C)	783TBulge (65°C)	782Homo (65°C)
AFM site-specific constants							
$S_{Mismatch}$		1660 ± 216	300 ± 36			200 ± 18	
S_{End}		615 ± 55	460 ± 78	440 ± 80	330 ± 40	345 ± 21	320 ± 140
$1/K_{SP}$ (nM)		21 ± 2.3	77 ± 7.7			109 ± 8.7	
$1/K_{NSP}$ (nM)		34 500 ± 2400	23 300 ± 1400	20 800 ± 1000	21 300 ± 2300	21 700 ± 900	38 500 ± 12 300
$1/K_{E}$ (nM)		56 ± 3	50 ± 8	48 ± 8	64 ± 3.5	63 ± 2.5	119 ± 36
Bulk measurements							
$1/K_{DNA,FLUOR}$[b] *Taq* (nM)	24 bp fragment	5 ± 4.9	40 ± 25	ND			
$1/K_{DNA,EMSA}$[c] *Taq* (nM)	60 bp fragment	2.2 ± 2.1	310 ± 71	3800 ± 360			
$1/K_{DNA,EMSA}$[d] *E.coli* (nM)	60 bp fragment	1.4 ± 0.5	5 ± 1.7	34 ± 6.1			
Calculated AFM fragment constants							
$1/K_{DNA,AFM}$[e] *Taq* (nM)	60 bp fragment including ends	12 ± 1.4	18 ± 3.4	23 ± 4.5			
	60 bp fragment excluding ends	*20 ± 2.6*	*68 ± 6*	*495 ± 15*			

ND, not determined.
[a]Weighted average from MutS binding to 817Puc18 and 1869Puc18.
[b]Apparent macroscopic binding constant of *Taq* MutS to a 24 bp DNA measured by fluorescence anisotropy at 23°C in this study.
[c]Published apparent macroscopic binding constant of *Taq* MutS to a 60 bp DNA measured by EMSA at 21°C from ref. (11).
[d]Published apparent macroscopic binding constant of *E.coli* MutS to a 60 bp DNA measured by EMSA at 37°C from ref. (11).
[e]Compositive macroscopic binding constants of *Taq* MutS to 60 bp DNA fragments calculated using Equations 13 and 14. The constants in italic are the values ignoring the contribution from the DNA end binding by MutS.

nonspecific DNA being dominated by the contribution from end binding (Table 1). Consequently, macroscopic binding constants determined using bulk methods may vary significantly depending on whether or not end binding is detected in the assay. As discussed in the following sections, the apparent differences in specificities of *Taq* and *E.coli* MutS determined from bulk measurements likely result from differences in the extent to which end binding is being detected in bulk assays.

Comparison of AFM and bulk studies

To compare the AFM results with the existing data from bulk studies, we have calculated the macroscopic binding constants ($K_{DNA,AFM}$) of *Taq* MutS to 60 bp DNA fragments (the length of the DNA fragment in the bulk EMSA study) using Equations 13 and 14 (Table 1). Comparison of the calculated macroscopic AFM binding constants, $K_{DNA,AFM}$, with the apparent binding constants determined from gel shift assays, $K_{DNA,EMSA}$, reveals that the constants for *Taq* MutS binding to DNA containing a T-bulge, which is the highest affinity mismatch, are very similar (12 nM for AFM versus 2 nM for EMSA). Interestingly, AFM, however, yields higher binding affinities than the EMSA studies for DNA containing a GT mismatch (18 nM for AFM versus 310 nM for EMSA) and for homoduplex DNA (23 nM for AFM versus 3800 nM for EMSA). The consistency between the results from AFM and EMSA for MutS binding to a T-bulge indicates that both AFM and gel shift assays are good methods for determining this tight binding constant. The differences for homoduplex DNA and GT-containing DNA, however, indicate that either AFM is overestimating the binding constants or that the EMSA experiments are underestimating them. As discussed below, it is more likely that the gel shift assays underestimate the binding constants for the GT-mismatch and homoduplex DNA fragments.

For AFM to yield an apparent binding constant that is greater than the actual binding constant, the free DNA would have to be less efficiently deposited on the surface than the DNA with protein bound. The agreement between the AFM and EMSA data for MutS binding to the DNA containing a T-bulge suggests that there is no significant difference between the deposition efficiency of free and bound DNA for this fragment. Consequently, it is unlikely that there would be a difference for free and bound DNA containing a GT mismatch or no mismatch, because both the occupancies of MutS on the DNA (Supplementary Table A) and the overall conformations of the protein–DNA complexes are similar (13). In addition, as mentioned above the rates and efficiency of deposition of DNA fragments with protein bound on both ends are the same as those of free DNA (23,24). Another way in which AFM could overestimate the binding affinity is if the coverage of protein on the surface is too high, such that the protein coincidently lands on DNA. Because the protein coverage on the surface is low in these studies (Figure 2), this potential problem is minimized. Specifically, using a statistical analysis (see Theory), we have estimated the maximal contribution of proteins randomly landing on the DNA to the observed binding constants. This analysis indicates that for the highest protein and DNA concentrations used, random landing could result in no more than ~20% increase in the observed binding constant (see Materials and Methods), which is insufficient to explain the observed discrepancies between AFM and EMSA. Furthermore, the binding constants of MutS for homoduplex sites and for DNA ends determined from different DNA and protein concentrations are consistent with one other (within 0.5 kcal/mol) and show no trend with protein or DNA concentration (Table 1 and Supplementary Table A), indicating that the contribution from proteins randomly landing on the DNA is negligible. As discussed in Theory, it is more likely that this AFM method would underestimate a binding constant than overestimate it.

The above discussion leads to the suggestion that the gel shift assays may be underestimating the binding affinities of *Taq* MutS to GT-containing and homoduplex DNA fragments. One significant limitation to using gel shift assays for determining binding constants is its insensitivity to weak or dynamic protein–DNA interactions (21). Notably, the AFM and EMSA results agree for the T-bulge, which has the highest binding affinity. As discussed above, MutS binding to DNA ends makes an increasingly important contribution to the calculated macroscopic binding constants, $K_{DNA,AFM}$, as specificity decreases (Table 1). Consequently, if binding of *Taq* MutS to DNA ends is not detected in the gel shift assays, the apparent binding constants for nonspecific and lower specificity DNA fragments determined by EMSA would be significantly less than those calculated from AFM. To assess this possibility further, we calculated apparent binding constants from AFM ignoring the contribution from end binding (Table 1, italic values). Although the affinities are still higher than that from EMSA (68 nM for AFM versus 310 nM for EMSA for GT-containing fragment; 495 nM for AFM versus 3800 nM for EMSA for homoduplex DNA fragment), they are similar given the different techniques and different DNA fragments that were used in the two studies.

To further resolve this discrepancy, we have measured the binding affinity, using fluorescence anisotropy (33,34), of MutS to short (24 bp) DNA fragments that have a fluorescent label on one end. Figure 4 shows binding curves for T-bulge and GT-containing DNA. The curves for T-bulge DNA are fit well to a single binding isotherm (Figure 4). The binding constant for T-bulge DNA, determined from the average of seven measurements, is 5 ± 4.9 nM, which is in good agreement with both the AFM and EMSA results. This result indicates that both AFM and EMSA yield accurate binding constants for MutS binding to DNA containing a T-bulge. The binding curves for the GT-DNA at 10 nM are fit well to a single binding isotherm (Figure 4); however, at 30 and 50 nM GT-DNA, they are biphasic, which is clearly revealed by nonlinearity in Scatchard plots (data not shown). This result indicates that there are multiple binding events on the GT-DNA, which is consistent with the AFM data, which yields similar binding constants for DNA ends and a GT mismatch. The binding constant for GT-DNA, determined from the average of four independent measurements (see Materials and Methods), is 40 ± 25 nM. This binding affinity is in good agreement with the binding constant determined by AFM and is significantly tighter than that determined by EMSA. For homoduplex DNA, the change in anisotropy upon the addition of MutS was too small to determine a binding constant (data not shown).

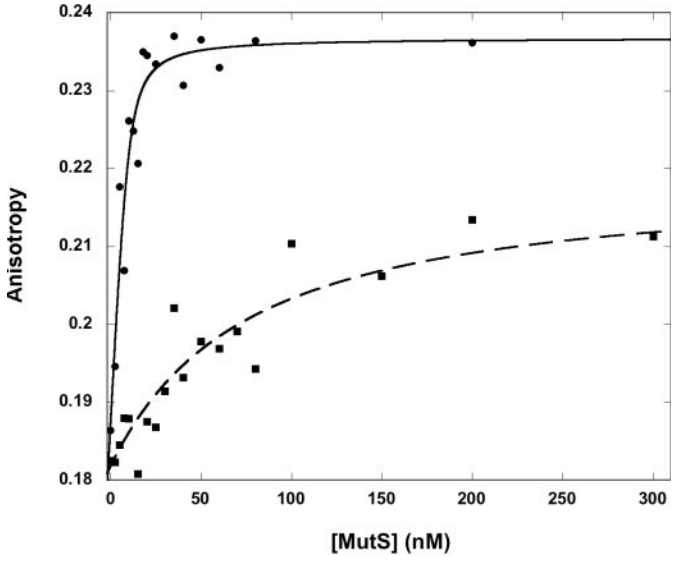

Figure 4. Binding of MutS to fluorescently labeled DNA fragments. The fluorescence anisotropy of 24 bp DNA fragments labeled on the 3′end with TAMRA is plotted as a function of added MutS protein. Typical data are shown for 10 nM TAMRA-labeled DNA fragments containing a T-bulge (circles) or a GT mismatch (squares). The curves are the best fits to a binding isotherm using Equation 16. The average binding constants from 4 to 7 determinations are given in Table 1.

Taken together, these analyses strongly suggest that end binding of *Taq* MutS was not detected by EMSA and that the binding to the weaker nonspecific sites and to a GT mismatch was only partially detected. On the other hand, AFM is expected to be significantly less sensitive to the dynamics of the protein–DNA interactions because the deposition of the complexes is rapid and irreversible over the time scale of the deposition (23,24). This suggestion is supported both by the agreement between specificities calculated from biochemical and AFM data (see previous section) and by the observation that the binding constant for *Taq* MutS to nonspecific DNA is similar at 23 and 65°C (Table 1). In summary, these results strongly suggest that the differences between AFM and EMSA are a result of the dynamic limitations of EMSA and that AFM provides an accurate measure of the binding constants.

Binding specificities of *E.coli* and *Taq* MutS

E.coli MutS presents an interesting counter example to *Taq* MutS, in that end binding by *E.coli* MutS appears to be detected by EMSA. Consistent with our AFM studies (H. Wang, P. Hsieh and D.A. Erie, unpublished data), recent biochemical studies indicate that *E.coli* MutS has a strong binding affinity to DNA ends and suggest that this end binding is stable to electrophoresis (35). This suggestion is further supported by inspection of the gels from the EMSA study, which reveals a super-shifted band on increasing *E.coli* MutS concentration, indicating that two MutS proteins are bound to the fragments at higher concentrations (11). In contrast, no super-shifted bands are seen with *Taq* MutS at a concentration of 100 nM (13), supporting our assertion that end binding of *Taq* MutS is not stable to electrophoresis. Detection by EMSA of end binding by *E.coli* MutS but not by *Taq* MutS likely results from the interactions between *E.coli* MutS and DNA ends being less dynamic than those between *Taq* MutS and DNA ends. Consistent with this suggestion, kinetic studies indicate that the dissociation rate of *Taq* MutS from a T-bulge (substrate with the highest affinity) is four times faster than that of *E.coli* MutS from a T-bulge (36).

The differences in the detection of end binding provide an explanation for the apparent differences in the specificities of *E.coli* and *Taq* MutS determined from EMSA studies. The bulk specificity for a T-bulge ($K_{T\text{-bulge,EMSA}}/K_{Homo,EMSA}$) is ~1700 for *Taq* MutS, whereas it is only ~30 for *E.coli* MutS

(Table 1). The differences in these specificities result primarily from differences in the apparent binding affinities to homoduplex DNA (30 nM for *E.coli* MutS versus 3000 nM for *Taq* MutS). The higher apparent binding affinity of *E.coli* MutS for homoduplex DNA results from stable end binding and is predicted by our analysis of the microscopic binding constants and the observations of high binding affinity of *E.coli* MutS for DNA ends. Consequently, for *E.coli* MutS, the ratio $K_{\text{T-bulge,EMSA}}/K_{\text{Homo,EMSA}}$ is not a good measure of specificity. It is likely that *E.coli* and *Taq* MutS have similar specificities for a T-bulge, given that their EMSA binding affinities are similar (Table 1). Published AFM data of the position distribution of *E.coli* MutS bound to DNA fragments containing a GT mismatch support this idea. Analysis of these data [see Supplementary Material in ref. (13)] reveals that *E.coli* MutS has a specificity of ~1000 for a GT mismatch, which is similar to that for *Taq* MutS (~300) but significantly higher than the specificity ($K_{\text{T-bulge,EMSA}}/K_{\text{Homo,EMSA}}$) determined by EMSA (~7).

Our results indicate that MutS has significantly higher specificities for mismatches than those previously determined from bulk measurements. *Taq* MutS maintains a high specificity (~200) for a T-bulge even at its physiological temperature of 65°C (Table 1). These higher specificities resolve, in part, the previous conundrum of apparent low MutS binding specificities, but highly efficient repair (13). AFM is powerful method for determining specificities, because it provides a direct measure of the relative binding affinities to different sites on the same DNA fragment and knowledge of the absolute binding affinities is not required. The only way that specificities determined from AFM studies can be skewed is if there is a difference in deposition efficiency of specific and nonspecific complexes, which is unlikely unless there is a significant difference in the overall conformation of the specific and nonspecific complexes.

CONCLUSION

Intrinsic limitations of bulk techniques on the quantification of protein–DNA binding constants and specificities have been long recognized, especially for proteins with low binding specificity and/or a second specific site, such as DNA ends. Only apparent binding constants to the entire DNA fragment can be determined from these methods. Furthermore, many of these bulk techniques are sensitive to the dynamics of the interactions. We have established a single-molecule method for determining protein–DNA binding constants and specificities in a site-specific fashion. This method provides direct measure of the binding affinity to a particular site because the occurrence of a protein binding to each site on the DNA is directly observed. Similarly, the binding specificities from AFM are site-to-site (microscopic) comparisons of binding constants instead of fragment-to-fragment comparisons. In addition, this method is straightforward because it relies only on counting and 1D-distance measurements of hundreds of complexes, which are sufficient for the assay. Finally, from a single set of AFM experiments, it is possible to determine the binding affinity, specificity and stoichiometry, as well as the conformational properties of the protein–DNA complexes (23,37). Application of this method to MutS–DNA interactions has revealed that the apparent differences between *Taq* and *E.coli* MutS DNA binding properties are due to the strong end-binding affinity of MutS, as well as to a problem in detecting end binding and nonspecific DNA binding by *Taq* MutS in the EMSA measurements. The significantly higher DNA-binding specificities of MutS obtained in this study are obviously important for achieving efficient DNA repair in the cell. The biological significance of strong binding affinity for DNA ends is unclear; however, it may be functionally important in other biological processes in which MutS is involved, such as DNA double-strand break repair and recombination, because DNA ends are critical intermediates in these pathways.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at NAR Online.

ACKNOWLEDGEMENTS

The authors are grateful for helpful discussions with Xiaolei Zhou. The authors thank Dr Ashutosh Tripathy, Director, UNC Macromolecular Interactions Facility, for use of fluorescence instrumentation. This work was supported by the National Institutes of Health grant GM R01-54316 and the American Cancer Society (DAE).

Conflict of interest statement. None declared.

REFERENCES

1. Carey,J. (1991) Gel retardation. *Methods Enzymol*., **208**, 103–117.
2. Lohman,T.M. and Bujalowski,W. (1991) Thermodynamic methods for model-independent determination of equilibrium binding isotherms for protein–DNA interactions: spectroscopic approaches to monitor binding. *Methods Enzymol*., **208**, 258–290.
3. Myszka,D.G. (2000) Kinetic, equilibrium, and thermodynamic analysis of macromolecular interactions with BIACORE. *Methods Enzymol*., **323**, 325–340.
4. Oda,M., Furukawa,K., Ogata,K., Sarai,A. and Nakamura,H. (1998) Thermodynamics of specific and non-specific DNA binding by the c-Myb DNA-binding domain. *J. Mol. Biol*., **276**, 571–590.
5. Winzor,D.J. and Sawyer,W.H. (1995) *Quantitative Characterization of Ligand Binding*. Wiley-Liss, Inc., New York, NY.
6. Record,M.T.,Jr, , Ha,J.H. and Fisher,M.A. (1991) Analysis of equilibrium and kinetic measurements to determine thermodynamic origins of stability and specificity and mechanism of formation of site-specific complexes between proteins and helical DNA. *Methods Enzymol*., **208**, 291–343.
7. Modrich,P. (1989) Methyl-directed DNA mismatch correction. *J. Biol. Chem*., **264**, 6597–6600.
8. Modrich,P. and Lahue,R. (1996) Mismatch repair in replication fidelity, genetic recombination, and cancer biology. *Annu. Rev. Biochem*., **65**, 101–133.
9. Hsieh,P. (2001) Molecular mechanisms of DNA mismatch repair. *Mutat. Res*., **486**, 71–87.
10. Schofield,M.J. and Hsieh,P. (2003) DNA mismatch repair: molecular mechanisms and biological function. *Annu. Rev. Microbiol*., **57**, 579–608.
11. Schofield,M.J., Brownewell,F.E., Nayak,S., Du,C., Kool,E.T. and Hsieh,P. (2001) The Phe-X-Glu DNA binding motif of MutS. The role of hydrogen bonding in mismatch recognition. *J. Biol. Chem*., **276**, 45505–45508.
12. Gradia,S., Acharya,S. and Fishel,R. (2000) The role of mismatched nucleotides in activating the hMSH2–hMSH6 molecular switch. *J. Biol. Chem*., **275**, 3922–3930.
13. Wang,H., Yang,Y., Schofield,M.J., Du,C., Fridman,Y., Lee,S.D., Larson,E.D., Drummond,J.T., Alani,E., Hsieh,P. *et al*. (2003) DNA bending and unbending by MutS govern mismatch recognition and specificity. *Proc. Natl Acad. Sci. USA*, **100**, 14822–14827.

14. Wang,H. and Hays,J.B. (2002) Mismatch repair in human nuclear extracts. Quantitative analyses of excision of nicked circular mismatched DNA substrates, constructed by a new technique employing synthetic oligonucleotides. *J. Biol. Chem.*, **277**, 26136–26142.
15. McGhee,J.D. and von Hippel,P.H. (1974) Theoretical aspects of DNA–protein interactions: co-operative and non-co-operative binding of large ligands to a one-dimensional homogeneous lattice. *J. Mol. Biol.*, **86**, 469–489.
16. de Jager,M., Wyman,C., van Gent,D.C. and Kanaar,R. (2002) DNA end-binding specificity of human Rad50/Mre11 is influenced by ATP. *Nucleic Acids Res.*, **30**, 4425–4431.
17. Chen,L., Haushalter,K.A., Lieber,C.M. and Verdine,G.L. (2002) Direct visualization of a DNA glycosylase searching for damage. *Chem. Biol.*, **9**, 345–350.
18. Solis,F.J., Bash,R., Yodh,J., Lindsay,S.M. and Lohr,D. (2004) A statistical thermodynamic model applied to experimental AFM population and location data is able to quantify DNA–histone binding strength and internucleosomal interaction differences between acetylated and unacetylated nucleosomal arrays. *Biophys. J.*, **87**, 3372–3387.
19. Taylor,J.R. (1997) *An Introduction to Error Analysis: The Study of Uncertainties in Physical Measurements*, 2nd edn. University Science Books, Sausalito, CA.
20. Koch,S.J., Shundrovsky,A., Jantzen,B.C. and Wang,M.D. (2002) Probing protein–DNA interactions by unzipping a single DNA double helix. *Biophys. J.*, **83**, 1098–1105.
21. Lim,W.A., Sauer,R.T. and Lander,A.D. (1991) Analysis of DNA–protein interactions by affinity coelectrophoresis. *Methods Enzymol.*, **208**, 196–210.
22. Lee,C.S. and Belfort,G. (1989) Changing activity of ribonuclease A during adsorption: a molecular explanation. *Proc. Natl Acad. Sci. USA*, **86**, 8392–8396.
23. Ratcliff,G.C. and Erie,D.A. (2001) A novel single-molecule study to determine protein–protein association constants. *J. Am. Chem. Soc.*, **123**, 5632–5635.
24. Rivetti,C., Guthold,M. and Bustamante,C. (1996) Scanning force microscopy of DNA deposited onto mica: equilibration versus kinetic trapping studied by statistical polymer chain analysis. *J. Mol. Biol.*, **264**, 919–932.
25. Gettens,R.T., Bai,Z. and Gilbert,J.L. (2005) Quantification of the kinetics and thermodynamics of protein adsorption using atomic force microscopy. *J. Biomed. Mater. Res. A*, **72A**, 246–257.
26. Biswas,I., Ban,C., Fleming,K.G., Qin,J., Lary,J.W., Yphantis,D.A., Yang,W. and Hsieh,P. (1999) Oligomerization of a MutS mismatch repair protein from *Thermus aquaticus*. *J. Biol. Chem.*, **274**, 23673–23678.
27. Obmolova,G., Ban,C., Hsieh,P. and Yang,W. (2000) Crystal structures of mismatch repair protein MutS and its complex with a substrate DNA. *Nature*, **407**, 703–710.
28. Lamers,M.H., Perrakis,A., Enzlin,J.H., Winterwerp,H.H., de Wind,N. and Sixma,T.K. (2000) The crystal structure of DNA mismatch repair protein MutS binding to a G × T mismatch. *Nature*, **407**, 711–717.
29. Natrajan,G., Lamers,M.H., Enzlin,J.H., Winterwerp,H.H., Perrakis,A. and Sixma,T.K. (2003) Structures of *Escherichia coli* DNA mismatch repair enzyme MutS in complex with different mismatches: a common recognition mode for diverse substrates. *Nucleic Acids Res.*, **31**, 4814–4821.
30. Bjornson,K.P., Allen,D.J. and Modrich,P. (2000) Modulation of MutS ATP hydrolysis by DNA cofactors. *Biochemistry*, **39**, 3176–3183.
31. Janicijevic,A., Sugasawa,K., Shimizu,Y., Hanaoka,F., Wijgers,N., Djurica,M., Hoeijmakers,J.H. and Wyman,C. (2003) DNA bending by the human damage recognition complex XPC-HR23B. *DNA Repair (Amst.)*, **2**, 325–336.
32. Hey,T., Lipps,G., Sugasawa,K., Iwai,S., Hanaoka,F. and Krauss,G. (2002) The XPC–HR23B complex displays high affinity and specificity for damaged DNA in a true-equilibrium fluorescence assay. *Biochemistry*, **41**, 6583–6587.
33. Jabonski,A. (1960) *Bull Acad Pol Sci Ser A*, **8**, 259–264.
34. Lakowicz,J.R. (1999) *Topics in Fluorescence Spectroscopy*. Chapter 10, Kluwer Academic/Plenum Publishers, New York, pp. 291–318.
35. Acharya,S., Foster,P.L., Brooks,P. and Fishel,R. (2003) The coordinated functions of the *E. coli* MutS and MutL proteins in mismatch repair. *Mol. Cell*, **12**, 233–246.
36. Schofield,M.J., Nayak,S., Scott,T.H., Du,C. and Hsieh,P. (2001) Interaction of *Escherichia coli* MutS and MutL at a DNA mismatch. *J. Biol. Chem.*, **276**, 28291–28299.
37. Yang,Y., Wang,H. and Erie,D.A. (2003) Quantitative characterization of biomolecular assemblies and interactions using atomic force microscopy. *Methods*, **29**, 175–187.