Quantifying DNA–protein binding specificities by using oligonucleotide mass tags and mass spectroscopy

Lingang Zhang*[†], Simon Kasif[†‡§], and Charles R. Cantor*[†¶‖]

*Center for Advanced Biotechnology, [†]Department of Biomedical Engineering, [‡]Bioinformatics Program, and [§]Center for Advanced Genomic Technology, Boston University, Boston, MA 02215; and [¶]SEQUENOM, Inc., 3595 John Hopkins Court, San Diego, CA 92121

The ability to determine the relative binding affinity of different transcription-factors (TF) to their DNA binding sites is fundamentally important for a comprehensive understanding of gene regulation. Here we present a general approach for multiplex quantification of DNA-TF binding specificities *in vitro* using oligonucleotide mass tag (OMT) labeling and mass spectroscopic quantification. An OMT is a short nucleic acid sequence with a distinct mass that can be resolved by a mass spectrometer. Each putative binding sequence is labeled with a unique OMT, and PCR amplification of OMTs is performed after removing nonbound DNA. Subsequently, a primer extension reaction is carried out, and the extension products are quantified by MALDI-TOF mass spectroscopy. Using the TF NF-κB P50, we have quantified the binding specificities of up to 15 binding sequences in a single assay. The results from the multiplex assay are consistent with data from the traditional gel shift assay. The approach allows the competitive binding of multiple DNA sequences to the given protein in a homogeneous reaction. By using the commercially available homogeneous MassEXTEND platform (SEQUENOM), it is scalable for high-throughput DNA-TF binding applications, including genome-wide TF binding site mapping and analyses of SNPs in promoter regions.

polymerase chain reaction | transcription factor

Understanding transcriptional regulation of gene expression is a great challenge in modern molecular biology. Eukaryotic transcriptional networks regulate and control cellular processes, including cell growth, differentiation, apoptosis, and responses to stress and other external stimuli, largely through interactions between transcription factors (TFs) and their genomic binding sites, commonly located near transcription start sites. The sophisticated interactions of TFs and genomic DNA modulate spatial and temporal differences in gene expression. Aberrant interaction of TFs and their targets may lead to a multitude of diseases. In addition, DNA variation in the binding site is known to affect TF binding, for example the NF-κB and Oct-1 sites in the tumor necrosis factor promoter (1, 2), the AP-1-binding site in the matrix γ-carboxyglutamic acid protein promoter (3), and the SP-1 site in the matrix metalloproteinase-2 promoter (4). Much progress has been made in recent years; however, it remains difficult to predict when mutations in genomic DNA might affect gene expression and subsequently affect important processes and cell phenotypes.

The study of gene regulation requires detailed knowledge of the interaction between TFs and genomic binding sites. With the constant growth of genomic data, numerous databases and software (e.g., the TRANSFAC database and accompanying software) have been developed to characterize TF binding sites (5). Although widely used to predict genomic binding sites and their binding affinities, the majority of these existing tools provide a relatively low level of both sensitivity and specificity (6, 7). In addition to the basic limitations of modeling methodologies (6, 8, 9), a major obstacle to improving the current situation is the lack of quantitative binding data. Most existing databases, including TRANSFAC, are based on a small amount of data from the published results of nonquantitative binding assays and thus are subject to sampling biases (5, 10).

As a result, they do not represent a comprehensive definition of the DNA sequence-recognition properties of TFs. Therefore, position weight matrices from existing databases are largely unable to fully capture the impact of DNA variations on the binding affinity or predict all binding sites (6, 7). Recently, improved computational models for a small number of TFs were developed based on *in vitro* DNA–protein binding data (7, 11, 12). For instance, using nonbiased, quantitative binding data from gel shift assays, Udalova *et al.* (7) developed a principal coordinate model that allowed the prediction of the effects of DNA variations within genomic binding sites on DNA–protein interactions with higher accuracy than traditional profile models.

Given the large amount of genomic data, there is a clear need for a scalable and flexible experimental method to screen the binding specificities of a large number of TFs. One straightforward approach is ChIP DNA analyzed on oligonucleotide microarrays (ChIP-chip), which determines the entire spectrum of *in vivo* DNA binding sites for a given TF (13, 14). However, ChIP-chip can only map probable TF–DNA interaction loci to within 1- to 2-kb resolution, and statistical analysis of the enrichment of genomic fragments and experiment verification are needed to locate the actual binding sites (15, 16). In addition, condition-specific protein binding may lead to ChIP-chip experiments without significant enrichment of bound genomic fragments. In addition, yeast or bacterial one-hybrid methods have been developed to identify potential DNA–protein interactions (17–19). These methods define the binding specificity of a TF in a single round of selection and are amenable to high-throughput analysis. However, in their current formats, these methods cannot detect the potential cooperative and synergistic activity of TFs.

Studies from ChIP-chip experiments have shown that the *in vitro* affinity of TFs binding to DNA sequences often reflects the relative occupancy of these sequences *in vivo* (20, 21). This observation suggests that, for a given TF, the knowledge of its sequence-recognition profile, measured *in vitro*, can be highly instructive in characterizing binding sites in the genome. Several experimental approaches have been developed to characterize DNA–protein interactions *in vitro*. Traditional methods, such as gel shift analysis, are laborious, nonscalable, and inaccurate when dissociation rates are fast relative to the gel migration time scale (22). Systematic evolution of ligands by exponential enrichment (SELEX) provides a method for the isolation of nucleic acids that bind to a target molecule with high affinity (23, 24). This assay, although highly
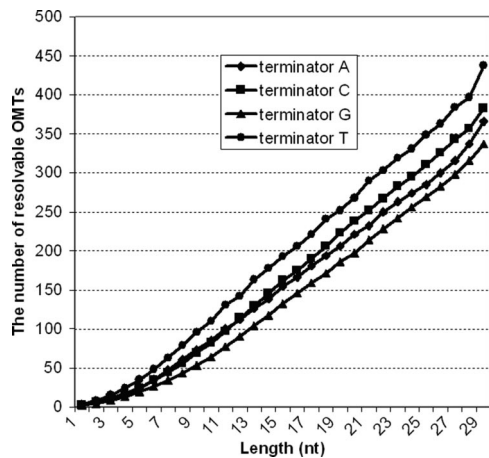

BIOCHEMISTRY

Fig. 1. The number of resolvable OMTs using four different terminators. All OMTs are designed with mass differences of 16 Da or more to allow accurate resolution by MALDI-TOF MS.

informative, identifies only the best binding sequences, whereas less optimal but often biologically relevant sequences are frequently missed. In addition, SELEX generally involves significant cloning and sequencing efforts. Additionally, microarray-based assays, in which duplex DNA molecules are immobilized on solid surfaces and protein binding is detected by surface plasmon resonance or fluorescence, provide a scalable platform for generating *in vitro* quantitative DNA binding data (11, 25–29). However, despite the demonstration of feasibility, complex processes are required to fabricate duplex-DNA microarrays. Additionally, the accessibility of target proteins to the duplex probes immobilized on the surface is still a concern, especially in studying a complex of cooperatively binding factors. These technical challenges have hindered the general application of these array platforms.

Here, we develop a *in vitro* method for sensitive, multiplex quantification of DNA–protein binding specificities by using oligonucleotide mass tags (OMTs) and mass spectroscopy (MS). Using a distinct OMT to label each protein-binding sequence, our method allows the competitive binding of multiple target sequences to a protein in a homogeneous reaction and quantification of the binding specificity of each sequence simultaneously in a single assay.

Relative binding affinities measured by our multiplex assay using TF NF-κB p50 are highly reproducible and agree with data from the traditional gel-shift assay. This method is promising for high-throughput DNA–TF binding applications, including fine-mapping of ChIP-chip results and for analyzing the impact of promoter SNPs on gene regulation.

Results

OMT Labeling. Each OMT is a short nucleic acid sequence with a unique mass but may be associated with different sequences. For instance, distinct oligonucleotides sequences 5′-ACGAAT-3′, 5′-CAAGAT-3′, and 5′-GAAACT-3′ are the same OMT because they have the same mass. All OMTs are designed to have one specific base, a 3′ terminator, and three other bases in other positions to facilitate the subsequent primer extension reaction. Different OMTs are designed to be sufficiently distinct in mass so that they are resolvable by a mass spectroscopy.

The number of resolvable OMTs was estimated, assuming a minimal 16-Da mass resolution in optimal MALDI-TOF mass spectroscopy. For a given length and terminator, every possible OMT up to the length and terminator was represented as a node in a graph. Any two nodes (OMTs) that have a difference in mass of <16 Da were connected by an edge. The most connected node (OMT) and connected edges were removed from the graph, and the process was repeated until no edge was left. Then the number of remaining OMTs was counted. This estimation shows that there are a large number of resolvable OMTs within a relatively short length (Fig. 1). For instance, within a length of 10 nt the number of resolvable OMTs can be >90. The terminator thymine generally provides the largest number of OMTs for a given maximal OMT length and was used as the terminator in our design of protein-binding probes.

DNA–Protein Binding Assay Using OMTs. Distinct OMTs are used to label different protein-binding sequences for multiplex quantification of binding specificity. The experiment is shown schematically in Fig. 2. For each candidate protein-binding sequence, we designed a DNA probe that comprises the binding site sequence (sequence C indicated in Fig. 2), a distinct OMT (sequence B), and constant sequences at the 5′ and 3′ termini (sequences A and D) that will be recognized by PCR primers. The target protein is incubated with a mixture of different duplex probes in equal molar concentration and an antibody against the target protein (first antibody). Then
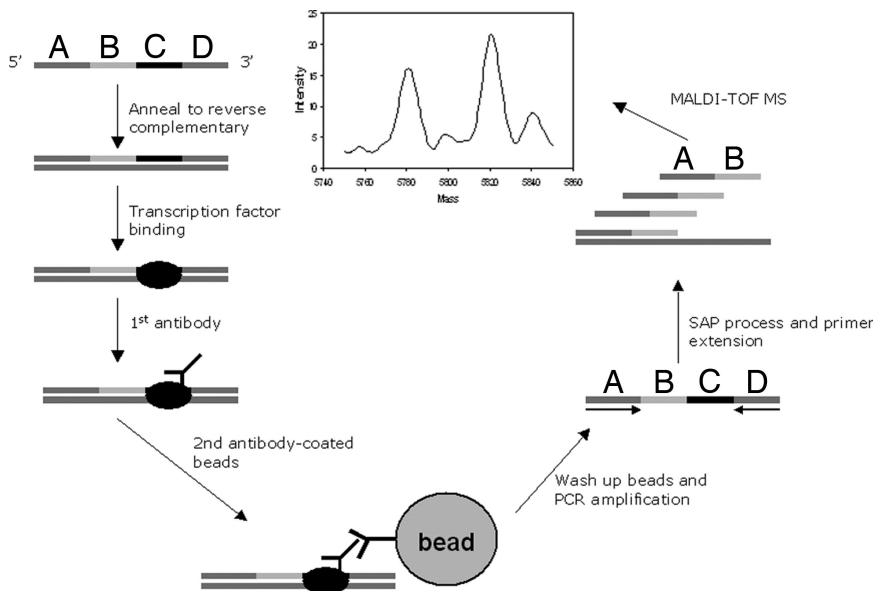

Fig. 2. A scheme for multiplexed quantification of DNA–protein binding specificity using OMTs. Each probe comprises four regions: A, a 5′ constant region recognized by a PCR primer; B, an OMT sequence; C, a 10-bp consensus binding site; and D, a 3′ constant sequence recognized by another PCR primer.

Table 1. Sequence of probes NC1, P1, P2, P3, and P4

Probe	Probe	Primer extension products (Da)
NC1	5′-taggcacctgaaaTT<u>ACAGGAAAAC</u>ctgtaggcaccat-3′	5′-ATCGTAGGCACCTGAAA-ddT-3′ (5,491.6)
P1	5′-taggcacctgaaaCTT<u>GGGATACCCC</u>ctgtaggcaccat-3′	5′-ATCGTAGGCACCTGAAAC-ddT-3′ (5,780.8)
P2	5′-taggcacctgaaaGTT<u>GGGATATCCC</u>ctgtaggcaccat-3′	5′-ATCGTAGGCACCTGAAAG-ddT-3′ (5,820.9)
P3	5′-taggcacctgaaaCCTT<u>GGGGCTTCCC</u>ctgtaggcaccat-3′	5′-ATCGTAGGCACCTGAAACC-ddT-3′ (6,070)
P4	5′-taggcacctgaaaCGTT<u>GGGGCTCCCC</u>ctgtaggcaccat-3′	5′-ATCGTAGGCACCTGAAACG-ddT-3′ (6,110)

The lowercase sequences at the 3′ and 5′ ends represent the constant sequences recognized by PCR primers. The sequence in uppercase letters is the OMT, and the 10-bp binding site is underlined and in uppercase letters. Molecular masses in daltons are shown in parentheses with the primer extension products.

magnetic beads precoated with an antibody specific to the first antibody (second antibody) are added and incubated. After washing the beads, all probes bound to the beads are PCR-amplified with a common pair of primers. Subsequently, after inactivation of unused dNTPs with shrimp alkaline phosphatase (SAP), a primer extension reaction is carried out with an extension primer and an extension mix comprising ddTTP, dATP, dGTP, and dCTP. The extension mix ensures that the extension reaction terminates at the 3′-end of sequence B, and the extension product comprises the sequence 5′-A-B-3′. Because of their resolvable masses, all of the extension products can be quantified simultaneously by using MALDI-TOF MS.

Relative Quantification of a Known Probe Mixture. Two protein-binding probes, P1 and P2 (Table 1), were used to evaluate the relative quantification. Their concentrations were measured by absorbance at 260 nm. The probes were mixed at different concentration ratios (10:1, 5:1, 2:1, 1:1, 1:2, 1:5, and 1:10) with a constant total concentration of 5×10^{-6} μg/μl. The probe mixtures were amplified by PCR, and SAP treatment and primer extension were subsequently performed. The two primer extension products, 5′-TAGGCACCTGAAAC-ddT-3′ and 5′-TAGGCACCT-GAAAG-ddT-3′, were resolved by using MALDI-TOF MS and quantified by calculating the corresponding peak areas by using the ALLELOTYPING software package (SEQUENOM). The concentration ratios measured by MS show an excellent consistency with those measured by absorbance, as shown in Fig. 3. Such linearity in relative quantification indicates that the PCR amplification is uniform for both probes, which is expected because the sizes of both probes are short (≈50 bp) and close to each other ,and the PCR efficiency largely depends on the kinetics of annealing of primers, which were designed to be the same for both probes.

Relative Quantification of NF-κB p50 Binding Specificity. We tested the quantification of DNA–protein binding specificity by using the TF NF-κB p50. NF-κB, a member of the Rel family of TFs, plays a key role in the regulation of inflammatory response, apoptosis, or tumorigenesis by binding to DNA through the Rel homology domain. NF-κB p50 homodimers are known to bind a 10-bp motif of consensus, 5′-GGRRNNYYCC-3′. However, a previous profile model, TRANSFAC, was shown to be a poor predictor of quantitative binding (7). Five probes, NC1, P1, P2, P3, and P4 were designed as shown in Table 1.

Note that DNA sequences flanking the binding site can sometimes influence the protein binding specificity of the binding site. To reduce the influence of variable OMT sequences on protein
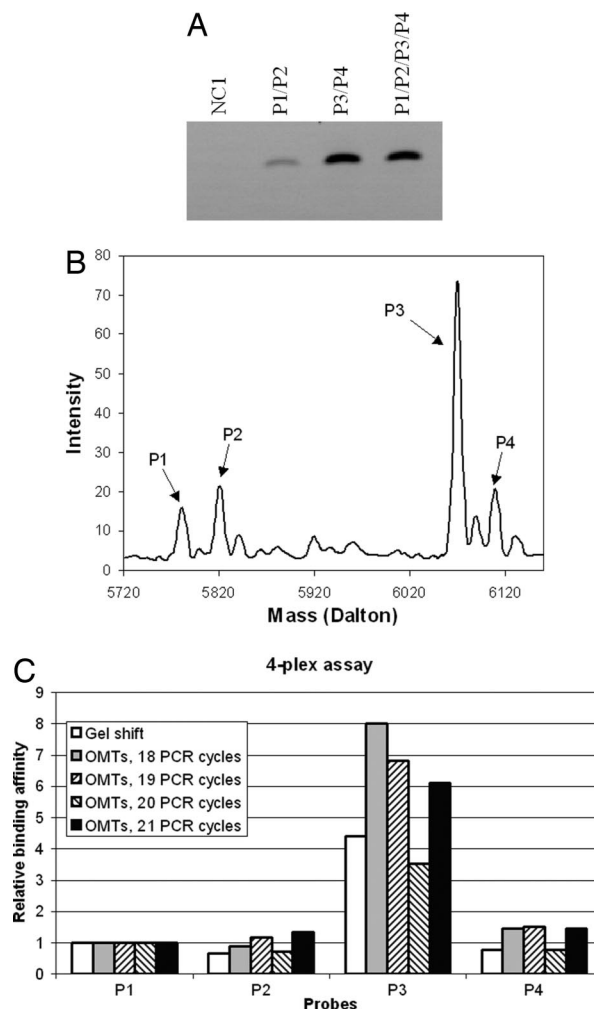
Fig. 4. PCR analysis of different probe mixtures. (*A*) Gel electrophoresis of NC1, a mixture of P1 and P2 (P1/P2), a mixture of P3 and P4 (P3/P4), and a mixture of P1–P4. (*B*) Mass spectrum of a four-plex assay comprising probes P1–P4. (*C*) Results from a four-plex assay using different PCR cycles (18–21) are consistent with the gel-shift quantification data (Pearson correlation coefficient > 0.98). The relative quantification is normalized to P1 to show the results.
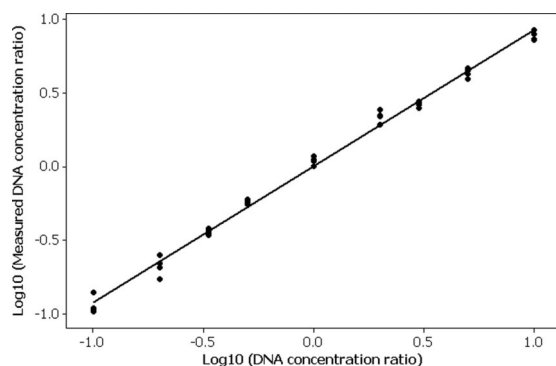
Fig. 3. Relative quantification of two probes at known concentration ratios. Two probes were mixed at different ratios (10:1, 5:1, 2:1, 1:1, 1:2, 1:5, and 1:10). After PCR, SAP processing, and primer extension, the ratio of the two primer extension products was measured by MS.
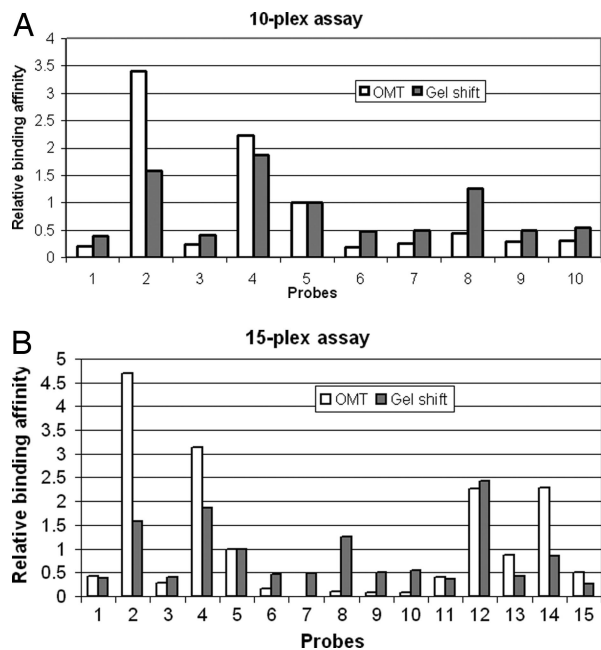
Fig. 5. Ten- and 15-plex assays. All the relative quantifications are normalized to probe 5. (*A*) Quantification from a 10-plex assay (probes 1–10) is consistent the gel-shift quantification data (Pearson correlation coefficient = 0.85). (*B*) Quantification from a 15-plex assay (probes ≈1–15) is consistent the gel-shift quantification data (Pearson correlation coefficient = 0.71).

Fig. 6. Mass spectrum of 10-plex binding assay comprising probes 1–10 using HeLa nuclear extract.

binding, we designed two terminators at the 3′ terminal of the OMTs (two continuous "T") as a spacer to isolate the OMT sequence from the binding site. As a result, all of the binding sites have the same immediate flanking sequences (within two bases) and the DNA–protein specificity is largely determined by the 10-bp sequence in the binding site. The additional terminator in the 3′ terminus of the OMTs does not change the primer extension. To further minimize the effect of variable OMTs on protein binding, one can either design a longer spacer sequence or design the binding sites containing the core binding site as well as proper flanking sequences, which are constant over all probes.

Probe NC1 does not have the consensus NF-κB p50 binding sites and was used as a negative control. Each probe was annealed to its reverse complement, and mixtures of different duplex probes at equal molar concentration were incubated with NF-κB p50 and rabbit polyclonal antibody against NF-κB p50 in the binding buffer. Then probes bound to NF-κB p50 were isolated by using magnetic beads coated with sheep anti-rabbit IgG and amplified by PCR. No PCR amplification was observed for probe NC1, as shown in Fig. 4*A*, which indicates that unbound probes were effectively removed from the PCR amplification by washing the beads, and no false-positive results are expected to have been recorded. In addition, we observed that the intensity of band with probe mixture of P3 and P4 (P3/P4) was significantly stronger than the band of the probe mixture of P1 and P2 (P1/P2), which agrees with that fact that P3 is observed to have much stronger binding affinity than the other three probes (7).

To quantify each probe in the PCR amplification, SAP processing, primer extension, and MALDI-TOF MS were subsequently performed. It can be seen in Fig. 4*B* that the peaks of the extension products of probes P1–P4 are well resolved in the mass spectrum. Each primer extension product was quantified by measuring the corresponding peak area. The results from amplification with four different PCR cycles were compared with the gel shift data from Udalova *et al.* (7), as shown in Fig. 4*C*, to evaluate our relative quantification. The quality of our four-plex assay using different PCR cycles (≈18–21) was confirmed by the high degree of data
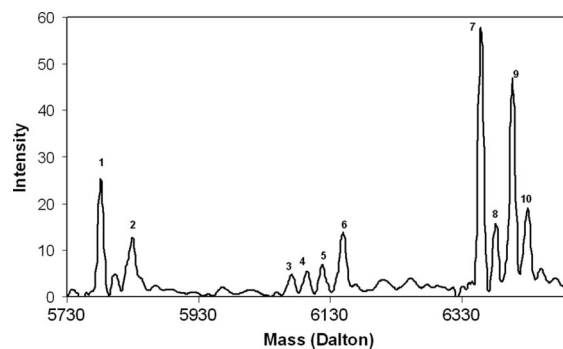
correlation with the gel-shift quantification data (Pearson correlation coefficient > 0.98).

To increase the throughput further and reduce the cost of DNA–protein binding analysis, we designed 15 probes [the sequences are available in supporting information (SI) Table 2] to test DNA–protein binding of higher multiplicity. Ten probes (1–10) or all 15 duplex probes in equal molar concentration were competitively bound with TF NF-κB p50 by incubating them together in the binding buffer. Bead isolation, PCR, SAP processing, and primer extension were subsequently performed as previously described. Experiments were carried out in quadruplicate for each multiplicity, and the binding data are available in SI Tables 3 and 4. The average binding affinity in the 10-plex (probes 1–10) assay shows good correlation (correlation coefficient = 0.85) with the gel-shift data on the same binding sequences (7), as shown in Fig. 5*A*. In addition, we observed that the results from our quadruplicate 10-plex experiments are highly reproducible (SI Table 3), with a correlation of >0.98. These data are in contrast to the gel-shift assay (7), which shows a lower correlation (0.70) between duplicated experiments (correlation was calculated on the gel shift data of the same 10 binding sites) (7). Similarly, in our 15-plex assay, the results are still fairly consistent with the gel-shift assay, with a correlation of 0.71 (Fig. 5*B*). In addition, the results from our four 15-plex experiments showed a better reproducibility, with a correlation of >0.91 (SI Table 4), than the gel-shift assay (correlation of 0.79 was calculated on the same 15 binding sites) (7).

NF-κB p50 Binding Specificity in a HeLa Nuclear Extract. We then characterized the binding specificity of NF-κB p50 in HeLa cells. Ten-plex assays using probes 1–10 were performed in quadruplicate as previously described, except that HeLa nuclear extract, rather than recombinant NF-κB p50, was used. It can be clearly seen in the mass spectrum (Fig. 6) that all 10 probes bound to NF-κB p50 in the HeLa nuclear extract. The results from four replicated experiments showed excellent correlation (Pearson correlation coefficient > 0.99) (SI Table 5). However, the binding specificity in the nuclear extract is dramatically different from that of the recombinant NF-κB p50. Notably, probes 7 and 9 showed much stronger relative binding affinity than others in the nuclear extract.

The binding specificity in the nuclear extract may be explained by the number of members of the NF-κB family and their coordinate binding to targets. The human NF-κB family contains five members, p65, p50, c-*Rel*, RelB, and p52. These subunits form homodimers and heterodimers to regulate a broad spectrum of biological processes (30). p50 lacks a transcriptional activation domain; thus, as a homodimer, it acts predominantly to repress gene expression. Importantly, however, p50 can also form heterodimers with other subunits to fulfill different functions, e.g., activation of gene expression (30). p50 heterodimers possess a binding specificity distinct from its homodimer form (7, 31). For instance, NF-κB

p50/p65 heterodimer binds the κB DNA target site of Igκ enhancer with an affinity of 5- to 15-fold higher than the p50 homodimer (31). Such a difference in binding affinity may help individual subunits fulfill both unique and shared functions with other family members by regulating different gene targets. A recent comprehensive ChIP-chip study of all of the five subunits showed that the genes bound by a single subunit are dramatically different from genes bound by two or more subunits, which suggested that p50 homodimers and heterodimers can regulate different gene targets (32). In our experiment, the antibody pulled down both the homodimers and heterodimers of NF-κB p50 and the resulted binding specificity in the nuclear extract is determined by the binding specificity of p50 homodimer and heterodimers and their relative concentration ratios.

Discussion

The interaction of TFs with genomic regulatory sites is of fundamental importance in understanding gene regulation. Although much effort has gone into identifying the binding sites, we are still far from a comprehensive genome-wide map of TF binding sites. The situation is even complicated by the fact that DNA variations, such as SNPs, in the binding sites can affect TF binding and potentially impact the normal process of gene regulation (1, 2, 4). *In vitro* DNA–protein binding assays provide a promising tool to address this challenge because quantitative binding results from these experiments may reflect the TF binding property *in vivo* (20, 21). However, technical challenges of assays to characterize *in vitro* DNA–protein interactions, including gel shift, SELEX, and protein-binding microarrays, have hindered their general application in large-scale genomic research.

We have described a general approach for multiplex quantification of TF binding specificity by using OMT labeling and MS quantification. Traditional labeling, including fluorescence and radioactivity, can monitor only one or a very limited number of distinct targets in a single assay. We used short fragments of nucleic acid sequences of unique mass, OMTs, to label each target binding sequence. The availability of a large number of resolvable OMTs allows a multiplex binding interaction in which different targets compete for the same protein, and, hence, the relative occupancy of each target should reflect the relative thermodynamic affinity very closely. Additionally, nonspecific binding in our multiple-competition assay is less likely to compromise the result than when the different targets are measured separately. It is noted that SELEX and protein-binding microarrays also allow multiplex DNA–protein binding interactions (23, 27, 28). However, SELEX currently requires a significant effort of cloning and sequencing of a large number of binding sequences (23). Protein-binding microarrays, on the other hand, require the complicated process of fabricating a duplex probe microarray. Once a microarray is fabricated, it is very inflexible and costly to change the probe design, which is often required as biomedical research progresses. In addition, microarray suffers from the imperfect accessibility of protein to the duplex probes immobilized on a solid surface. Nonspecific surface interactions, electrostatic interactions among dense sets of DNAs, and inhomogeneous surface mixing also complicate quantitative interpretation of the results. However, in our method the binding sequences can be flexibly designed and mixed in each assay, and these probes will interact with the TF in a homogeneous reaction. Thus the method proposed in this paper provides a cost-effective technology that complements the array technology.

Another distinct advantage of OMTs is that they are amenable to amplification. Because an OMT is a nucleic acid sequence, it can be amplified by nucleic-acid amplification methods, including PCR, ligase chain reaction (33), and T-7 promoter-based RNA amplification (34). Additionally, all of the probes are uniformly amplified because of the short probe size and the common PCR primers used. The strict preservation of concentration ratios during the amplification and quantification allows for a sensitive and accurate quantification. In addition, the assay specificity, which relies on the antibody-based isolation of targets, was shown to be high in our negative control tests.

Because an OMT is a nucleic acid sequence, the labeling process is relatively easy. Whereas most labeling processes require chemical incorporation of probe labels, by using an OMT, one only needs to design a short OMT sequence in a proper position of the probe and synthesize the probe sequence by following standard methods.

Our OMT labeling has been well integrated with MS quantification. MS has been shown to be an excellent platform for the quantification of nucleic acids (35, 36). Using a mass-resolvable OMT to label each target, we expand the capacity of MS and allow multiplexing quantification assays without much optimization. The integration of the multiplex OMT-based assay with the 384-format SpectroCHIP (SEQUENOM) can generate high-throughput capacity for quantification of DNA–protein binding specificity. For instance, a large amount of valuable data from ChIP-chip experiments is available for a variety of TFs. Generally, these results map the protein–DNA binding sites at 1- to 2-kb resolution. Although statistical approaches have been developed to predict the exact genomic binding sites, experimental verification of these predictions and fine mapping of the ChIP-chip data are generally needed (15, 16). In our method, the probe sequences can be flexibly designed based on ChIP-chip data and associated statistical predictions, and the verification assay can be performed on the 384-format SpectroCHIP for high-throughput fine-mapping of binding sites at virtually a single-base-pair level. In addition, SNPs located in promoters or nonsynonymous SNPs in TF or other genes can affect gene regulation and lead to diseases (1, 2, 4, 37). Our technique also provides an approach to screening the impact of SNPs in promoter regions on TF-binding affinity and analyzing the binding specificity of proteins with nonsynonymous SNPs. In addition to proteins, our method could also be used to screen the binding specificity of sequence-specific DNA-binding small molecules, e.g., polyamides, for synthetic biology research and molecular medicine discovery (38).

One potential challenge to implementing the OMT technique in genomic study is that many TFs and DNA-binding proteins bind as complexes. Those complexes can cover binding sites significantly larger than those so far examined, and, as a result, large pieces of genomic DNA need to be incorporated as the binding sites into the probes to test the binding *in vitro*. Because of the inefficiency and technical difficulty of synthesizing long DNA, we propose to prepare the probes by PCR amplification of the corresponding genomic DNA by using PCR primers concatenated with the corresponding mass tag and constant primer-binding sequences at the 3′ ends (a diagram available in SI Fig. 7).

In addition to quantifying sequence-specific binding, OMTs can potentially be used in other applications as labels. The basic principle is to label each nucleic acid probe, which is specific to a target molecule, with a distinct OMT. After the probes recognize their target molecule, excess probes are removed, and the amount of each probe is quantified by the PCR SAP process, primer extension, and MS as described before. With minor modifications, the approach in this paper can potentially be used in multiplex quantification of protein (39, 40), mRNA, or other biological molecules.

Materials and Methods

OMTs. OMTs are short nucleic acid (usually DNA) sequences. Each OMT has a unique mass but may have many distinct sequences. All OMTs are designed to comprise a specific base, called a terminator, at the 3′ terminal and three other bases in other positions. In addition, any two OMTs are designed to be resolvable in a mass spectrum.

Protein-Binding Probes. Protein-binding DNA probes were designed for TF NF-κB p50, which has a consensus binding site that is 10 bp

long. Each protein-binding probe comprises four regions, designed in our study as 5′-taggcacctgaaa-OMT-NNNNNNNNNN-ctgtaggcaccat-3′. The 5′- and 3′-end sequences (lowercase characters) are constant in all probes and recognized by PCR primers for amplification. The OMT indicates the OMT sequence, and the middle 10 N sequences are target sites for NF-κB p50 binding. In our study, thymine was used as the terminator. All of these protein-binding probes and their reverse complementary sequences were purchased from Integrated DNA Technology (Coralville, IA).

Binding Assay. Each protein-binding probe was mixed with its reverse complementary sequence at equal molar concentrations in 10 mM Tris·HCl (pH 8.0), 100 mM NaCl, and 1 mM EDTA. The mixture was heated at 95°C for 5 min and then cooled down slowly to room temperature to form duplex probes.

Binding of TF and duplex probes was measured in a buffer containing 10 mM Hepes, 50 mM KCl, 0.1 mM EDTA, 1.0 mM DTT, 10% glycerol, 0.05% Nonidet P-40, and 0.05 mg/ml poly(dI-dC) (Amersham Biosciences). Human recombinant NF-κB p50 (1 μl) (Promega, Madison, WI) or 2.0 μl of HeLa nuclear extract (Promega), and 2.0 μl of rabbit polyclonal antibody against NF-κB p50 (Santa Cruz Biotechnology, Santa Cruz, CA) were preincubated in 20 μl of binding buffer at room temperature for 10 min. Then a set of different duplex probes at equal molar concentrations was added to a final concentration of 20 nM for each duplex probe. The mixture was incubated at room temperature for 40 min. Subsequently, we added 100 μl of magnetic beads coated with sheep anti-rabbit IgG (Dynal Biotech, Oslo, Norway) that were pre-washed with PBS buffer containing 0.5% BSA, resuspended, and incubated in 100 μl of binding buffer for 10 min. The mixture was incubated for 30 min at room temperature on a rotator. Finally, the magnetic beads were washed six times with PBS buffer containing 5% BSA and 0.05% Tween-20 and resuspended in 200 ml of Tris-EDTA buffer for PCR amplification.

PCR Amplification and Primer Extension. Protein-binding probes bound to the magnetic beads through the TF NF-κB p50 were amplified by PCR with the primers 5′-ATCGTAGGCACCT-GAAA-3′ and 5′-ATTGATGGTGCCTACAG-3′. Amplification of 0.04 μl of suspended magnetic beads was performed by using 1.0 μM PCR primers, 1.5 mM MgCl$_2$, 200 μM dNTP, and 0.1 unit of HotStartTaq DNA polymerase (Qiagen) in a total reaction of 5 μl with the following PCR conditions: 95°C hot start for 15 min, followed by ≈18–26 cycles of 95°C for 30 s, 55°C for 30 s, then 72°C for 20 s, with a final hold of 72°C of 4 min.

After PCR, the products were treated with 0.04 unit of SAP (SEQUENOM), which inactivates unused dNTPs from the amplification cycles, for 20 min at 37°C followed by heat inactivation at 85°C for 5 min. Then, for the primer extension cycle, a 1.2 μM final concentration of extension primer 5′-ATCGTAGGCACCT-GAAA-3′ (same as one PCR primer) and 0.6 unit of ThermoSequenase (SEQUENOM) were added to a total volume of 9 μl with the extension mix containing dATP, dCTP, dGTP, and ddTTP at 50 μM for each base. The primer extension conditions include a 94°C hold for 2 min and ≈40–70 cycles of the following conditions: 94°C for 5 s, 52°C for 5 s, and 72°C for 15 s. All reactions (PCR amplification, SAP processing, and primer extension) were carried out in a GeneAmp 9700 thermocycler (Applied Biosystems).

MALDI-TOF MS and Quantitative Analysis. Before MALDI-TOF MS analysis, salts from the reactions were removed by using Spectro-CLEAN (SEQUENOM) resin and 16 μl of water. After a quick centrifugation, ≈10 nl of reaction solution was dispensed onto a 384-format SpectroCHIP (SEQUENOM) by using a Spectro-POINT nanodispenser (SEQUENOM). A SpectroCHIP is a disposable silicon dioxide chip prespotted with an optimized MALDI matrix for DNA analysis in either 96 or 384 pad format and calibrant pads for positive control samples. Mass spectrometric data were automatically imported into the SpectroTYPER (SEQUENOM) database for processing, i.e., noise normalization and peak area analysis. Each primer extension product was quantified by measuring its associated peak area in the mass spectrum.

Note. While this manuscript was under review, we learned that Landegren and colleagues (41) developed another *in vitro* method for specific and sensitive analysis of interactions between proteins and nucleic acids by using the proximity ligation assay (PLA) (ref. 41, which is published in this issue of PNAS). In this method, the detection of protein binding also is achieved by quantification of nucleic acids. Both methods use PCR amplification for sensitive detection and antibodies for targeting specific proteins. One difference is that our OMT-based method allows multiplex assay in a homogeneous reaction, whereas the PLA-based method requires single-stranded DNA tag microarrays for multiplexing applications. We believe that the two distinct approaches are largely complementary to each other, and our OMT labeling technology could be used in the PLA-based method for multiplex analysis.

1. Udalova IA, Richardson A, Denys A, Smith C, Ackerman H, Foxwell B, Kwiatkowski D (2000) *Mol Cell Biol* 20:9113–9119.
2. Knight JC, Udalova I, Hill AV, Greenwood BM, Peshu N, Marsh K, Kwiatkowski D (1999) *Nat Genet* 22:145–150.
3. Farzaneh-Far A, Davies JD, Braam LA, Spronk HM, Proudfoot D, Chan SW, O'Shaughnessy KM, Weissberg PL, Vermeer C, Shanahan CM (2001) *J Biol Chem* 276:32466–32473.
4. Price SJ, Greaves DR, Watkins H (2001) *J Biol Chem* 276:7549–7558.
5. Matys V, Kel-Margoulis OV, Fricke E, Liebich I, Land S, Barre-Dirrie A, Reuter I, Chekmenev D, Krull M, Hornischer K, *et al.* (2006) *Nucleic Acids Res* 34:D108–10.
6. Gershenzon NI, Stormo GD, Ioshikhes IP (2005) *Nucleic Acids Res* 33:2290–2301.
7. Udalova IA, Mott R, Field D, Kwiatkowski D (2002) *Proc Natl Acad Sci USA* 99:8167–8172.
8. Bulyk ML, Johnson PL, Church GM (2002) *Nucleic Acids Res* 30:1255–1261.
9. Man TK, Yang JS, Stormo GD (2004) *Nucleic Acids Res* 33:4026–4032.
10. Pavesi G, Mauri G, Pesole G (2004) *Brief Bioinform* 5:217–236.
11. Linnell J, Mott R, Field S, Kwiatkowski DP, Ragoussis J, Udalova IA (2004) *Nucleic Acids Res* 32:e44.
12. Roulet E, Busso S, Camargo AA, Simpson AJ, Mermod N, Bucher P (2002) *Nat Biotechnol* 20:831–835.
13. Ren B, Robert F, Wyrick JJ, Aparicio O, Jennings EG, Simon I, Zeitlinger J, Schreiber J, Hannett N, Kanin E, *et al.* (2000) *Science* 290:2306–2309.
14. Iyer VR, Horak CE, Scafe CS, Botstein D, Snyder M, Brown PO (2001) *Nature* 409:533–538.
15. Buck MJ, Lieb JD (2004) *Genomics* 83:349–360.
16. Liu XS, Brutlag DL, Liu JS (2002) *Nat Biotechnol* 20:835–839.
17. Deplancke B, Dupuy D, Vidal M, Walhout AJ (2004) *Genome Res* 14:2093–2101.
18. Deplancke B, Mukhopadhyay A, Ao W, Elewa AM, Grove CA, Martinez NJ, Sequerra R, Doucette-Stamm L, Reece-Hoyes JS, Hope IA, *et al.* (2006) *Cell* 125:1193–1205.
19. Meng X, Brodsky MH, Wolfe SA (2005) *Nat Biotechnol* 23:988–994.
20. Horak CE, Mahajan MC, Luscombe NM, Gerstein M, Weissman SM, Snyder M (2002) *Proc Natl Acad Sci USA* 99:2924–2929.
21. Harbison CT, Gordon DB, Lee TI, Rinaldi NJ, Macisaac KD, Danford TW, Hannett NM, Tagne JB, Reynolds DB, Yoo J, *et al.* (2004) *Nature* 431:99–104.
22. Jansen C, Gronenborn AM, Clore GM (1987) *Biochem J* 246:227–232.
23. Beinoraviciute-Kellner R, Lipps G, Krauss G (2005) *FEBS Lett* 579:4535–4540.
24. Tuerk C, Gold L (1990) *Science* 249:505–510.
25. Wegner GJ, Lee HJ, Marriott G, Corn RM (2003) *Anal Chem* 75:4740–4746.
26. Wang JK, Li TX, Lu ZH (2005) *J Biochem Biophys Methods* 63:100–110.
27. Mukherjee S, Berger MF, Jona G, Wang XS, Muzzey D, Snyder M, Young RA, Bulyk ML (2004) *Nat Genet* 36:1331–1339.
28. Bulyk ML, Gentalen E, Lockhart DJ, Church GM (1999) *Nat Biotechnol* 17:573–577.
29. Egener T, Roulet E, Zehnder M, Bucher P, Mermod N (2005) *Nucleic Acids Res* 33:e79.
30. Hayden MS, Ghosh S (2004) *Genes Dev* 18:2195–2224.
31. Phelps CB, Sengchanthalangsy LL, Malek S, Ghosh G (2000) *J Biol Chem* 275:24392–24399.
32. Schreiber J, Jenner RG, Murray HL, Gerber GK, Gifford DK, Young RA (2006) *Proc Natl Acad Sci USA* 103:5899–5904.
33. Barany F (1991) *Proc Natl Acad Sci USA* 88:189–193.
34. Van Gelder RN, von Zastrow ME, Yool A, Dement WC, Barchas JD, Eberwine JH (1990) *Proc Natl Acad Sci USA* 87:1663–1667.
35. McCullough RM, Cantor CR, Ding C (2005) *Nucleic Acids Res* 33:e99.
36. Ding C, Cantor CR (2003) *Proc Natl Acad Sci USA* 100:3059–3064.
37. Stitziel NO, Tseng YY, Pervouchine D, Goddeau D, Kasif S, Liang J (2003) *J Mol Biol* 327:1021–1030.
38. Warren CL, Kratochvil NC, Hauschild KE, Foister S, Brezinski ML, Dervan PB, Phillips GN, Jr, Ansari AZ (2006) *Proc Natl Acad Sci USA* 103:867–872.
39. Fredriksson S, Gullberg M, Jarvius J, Olsson C, Pietras K, Gustafsdottir SM, Ostman A, Landegren U (2002) *Nat Biotechnol* 20:473–477.
40. Sano T, Smith CL, Cantor CR (1992) *Science* 258:120–122.
41. Gustafsdottir SM, Schlingemann J, Rada-Iglesias A, Schallmeiner E, Kamali-Moghaddam M, Wadelius C, Landegren U (2007) *Proc Natl Acad Sci USA* 104:3067–3072.