

CrossMark  
click for updatesCite this: *Chem. Sci.*, 2015, 6, 745

## Photoaffinity labeling of transcription factors by DNA-templated crosslinking†

Ying Liu,<sup>a</sup> Wenlu Zheng,<sup>b</sup> Wan Zhang,<sup>b</sup> Nan Chen,<sup>a</sup> Yang Liu,<sup>a</sup> Li Chen,<sup>a</sup> Xiaozhou Zhou,<sup>a</sup> Xingshuo Chen,<sup>a</sup> Haifeng Zheng<sup>a</sup> and Xiaoyu Li<sup>\*ab</sup>

Characterization of transcription factor-DNA interaction is of high importance in elucidating the molecular mechanisms of gene transcriptions. DNA-based affinity probes were developed to capture and identify transcription factors by covalent crosslinking; however, the requirement of a crosslinker on the affinity probe remains a disadvantage, as the crosslinker itself often interferes with the protein-DNA interactions. We report a dual-probe method able to capture DNA-binding transcription factors with unmodified protein-binding sites in scenarios where conventional probes have failed. We have also shown the method's converse application in selecting specific transcription factor-binding DNA sequences from a probe library and its extension to studying proteins recognizing epigenetic marks. This study may provide a new tool for exploring DNA-binding proteins in biology.

Received 1st July 2014  
Accepted 30th September 2014

DOI: 10.1039/c4sc01953a

www.rsc.org/chemicalscience

### Introduction

Transcription factor (TF) is the major class of DNA-binding proteins that recognize and bind to specific double strand DNA sequences.<sup>1-3</sup> By binding to DNA, transcription factors modulate transcription levels of target genes and play central roles in many fundamental biological processes, usually in response to various exogenous and endogenous cellular signals in both healthy and disease states.<sup>1,4-7</sup> Consequently, transcription factors have been intensively pursued as drug targets in pharmaceutical research.<sup>8-11</sup>

Characterization of TF-DNA interactions is instrumental in elucidating transcription factors' regulatory mechanisms. Previously, many methods have been developed to identify known transcription factors' binding DNA sequences,<sup>12</sup> such as footprinting,<sup>13,14</sup> electrophoresis mobility shift assay (EMSA),<sup>15,16</sup> chromatin immunoprecipitation (ChIP),<sup>17</sup> chemiluminescent pull-down assay,<sup>18,19</sup> protein binding microarray,<sup>20</sup> and HT-SELEX.<sup>21</sup> Once a transcription factor's binding DNA sequence is known, it can be embedded into various probes for sensitive detection, such as the bimolecular proximity assay,<sup>22-25</sup> proximity-ligation assay,<sup>26,27</sup> nuclease protection assay,<sup>28</sup>

transcription factor beacon,<sup>29</sup> fluorescence recovery assay,<sup>30</sup> and enzyme amplification assay.<sup>31</sup>

On the other hand, characterization of unknown transcription factors that bind to specific DNA sequences is also highly important.<sup>32,33</sup> As many transcription factors bind to DNA transiently with low affinity,<sup>7</sup> the interactions are often lost during typical affinity purification; therefore covalent affinity probes equipped with chemical and photo-crosslinkers were developed, serving as a powerful tool to study protein-DNA interactions.<sup>14,34-51</sup> However, since the crosslinker is usually located in the protein-binding site of the affinity probe, it often contributes to or interferes with protein binding (Fig. 1a). The probe's performance strongly depends on the nature and position of the crosslinker.<sup>38,41,42</sup> Considerable efforts were

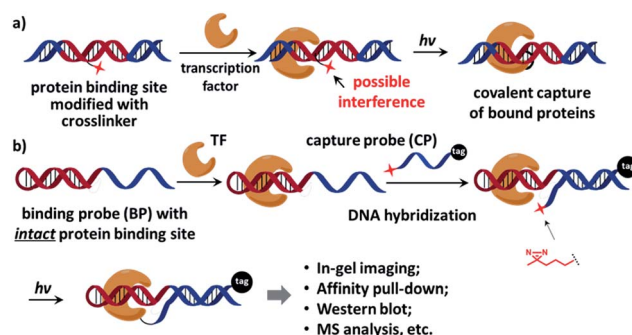


Fig. 1 (a) The crosslinker (shown as a red star) inside the protein-binding site of the affinity probe may interfere with protein binding. (b) The dual-probe method: after the transcription factor binds the binding probe (BP), the capture probe (CP) hybridizes to the BP DNA, photo-crosslinks BP-bound proteins under light irradiation, and also tags the protein. See the ESI† for probe structures and synthesis details.

<sup>a</sup>Key Laboratory of Bioorganic Chemistry and Molecular Engineering of the Ministry of Education, Beijing National Laboratory of Molecular Sciences, College of Chemistry and Molecular Engineering, Peking University, Beijing, China 100871. E-mail: xiaoyuli@pku.edu.cn

<sup>b</sup>Key Laboratory of Chemical Genomics, School of Chemical Biology and Biotechnology, Peking University Shenzhen Graduate School, Shenzhen, China 518055

† Electronic supplementary information (ESI) available: Experimental details, characterization data, DNA sequences, and detailed selection procedure. See DOI: 10.1039/c4sc01953a



undertaken to minimize the crosslinker's impact by screening for the optimal position,<sup>41,43,49,52</sup> adjusting the crosslinker's orientation,<sup>46,47,52–54</sup> and using smaller crosslinking groups.<sup>47</sup> In a recent report, Famulok and co-workers conjugated the crosslinker at the end of the aptamer so that binding interference was avoided.<sup>40</sup> Indeed, ideally an affinity probe should contain a protein-binding site free of modifications, but is still able to specifically deliver the crosslinker to the proximity of the target protein for effective labeling.

Recently we reported an affinity labeling method for identifying small molecule's target proteins, in which the functions of target recognition and covalent crosslinking are separated into two probes.<sup>55,56</sup> We reason that this strategy may be employed in studying transcription factor-DNA interactions to circumvent the requirement for a crosslinker within the affinity probe. Our design is shown in Fig. 1b, a native, modification-free hairpin DNA containing the bait sequence (shown in red) is used as the "binding probe" (BP). Another DNA modified with a photo-reactive 3'-diazirine group serves as the "capture probe" (CP), which also bears a 5'-tag customizable for subsequent analysis (e.g. a fluorophore for in-gel imaging or a biotin group for affinity pull-down). Diazirine has been widely used as the crosslinker in numerous biological applications for its small size, high reactivity, and biocompatibility;<sup>47,57–60</sup> it also exhibits very low non-specific protein crosslinking with moderately elevated salt concentration.<sup>55,56</sup> After the transcription factor binds to BP, CP hybridizes to the binding probe DNA and then photo-crosslinks BP-bound protein under light irradiation. BP is free of any modification so that the original protein-DNA interaction is maintained, while CP is able to deliver the crosslinker close to the target protein for efficient crosslinking.

## Results and discussion

We initiated the study with a model transcription factor p50, a subunit of nuclear factor-kappa B (NF- $\kappa$ B) transcription factor,<sup>61</sup> which plays key roles in cell's immune responses to stimuli<sup>62</sup> and is implicated in many diseases.<sup>63</sup> First, a binding probe embedded with the p50-binding sequence (p50-BP) and a sequence-complementary capture probe with a 5'-fluorescein tag (FAM-CP) were prepared (Fig. 2a and see details in Fig. S1†). The mixture of p50-BP, FAM-CP, and p50 was irradiated under 365 nm before denaturing SDS-PAGE analysis (Fig. 2b). Results show that p50 can be specifically labelled by the dual probe (lane 4). Two product bands were observed in lane 4: by comparing with the standard samples in lane 2 and 3, they are considered to be the p50-CP conjugate (p50-CP, lower band) and the DNA duplex formed by the p50-CP and BP DNA (p50-CP/BP, upper band), as a DNA duplex may partially renature in SDS-PAGE. We have observed and experimentally confirmed this phenomenon previously.<sup>55</sup> Furthermore, little non-specific labeling was observed when additional BSA was added along with the p50 protein (1 eq. in lane 8 of Fig. 2c; 10 eq. in lane 8 and 9 of Fig. S5†). Other negative controls (without p50-BP, with a CREB-1-binding BP, and without light irradiation; lane 5, 6, and 9) also did not give noticeable p50 capture. An FAM-CP with mismatched DNA sequence for the p50-BP showed some low

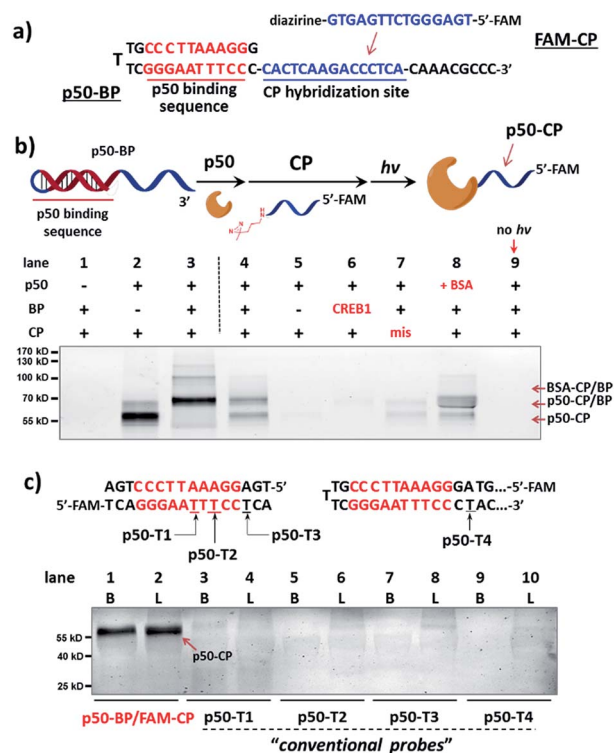


Fig. 2 (a) Sequences of p50-BP and FAM-CP. (b) Reaction scheme and denaturing SDS-PAGE analysis, monitored by FAM fluorescence. p50-BP, FAM-CP, p50, and BSA: 2  $\mu$ M each; hv: 365 nm, 15 min, 0  $^{\circ}$ C. Buffer: 10 mM MgCl<sub>2</sub>, 50 mM KCl, 10 mM EDTA, 25 mM DTT, 1 $\times$  PBS. Lane 1: BP/CP only; lane 2: standard sample of the p50-CP conjugate; lane 3: standard sample of the DNA duplex formed by p50-CP and BP; lane 4: p50 capture by p50-BP/FAM-CP after irradiation; lane 5–9: same as lane 4 but without p50-BP, with a non-p50-binding (CREB1-binding) BP, with a sequence-mismatched FAM-CP, with 1 eq. additional BSA, and without irradiation. p50-conjugated CP and p50-BP may partially renature in gel, resulting in two fluorescent bands: the p50-CP conjugate and the p50-CP/BP duplex (marked by arrows).<sup>55</sup> (c) Comparison of p50 capture by the dual-probe method with probes having directly conjugated diazirine (p50-T1/2/3/4). Experimental conditions are the same as in (b). Diazirine sites are underlined. B: reactions in buffer; L: reactions in HeLa lysate (4.8 mg mL<sup>-1</sup>, spiked with 2  $\mu$ M p50). CREB1: cAMP response element-binding protein 1.

level of labeling, possibly resulting from the BP-CP duplex partially formed at the incubation temperature (0  $^{\circ}$ C). In addition, similar labeling specificity was also observed with a 5'-biotin-tagged CP (Fig. S5†). Collectively, these results have demonstrated that the observed p50 labeling requires both specific protein-DNA interaction and photo-crosslinking mediated by a complementary capture DNA probe.

Next, for comparison, a series of "conventional probes" were prepared with the diazirine crosslinker directly conjugated at the major groove side of the DNA duplex (Fig. S2†), either inside the p50-binding site (p50-T1, T2), immediately next to it (p50-T3), or 1-base away from the binding site (p50-T4; Fig. 2c). These probes were subjected to the same p50 labeling procedures as in Fig. 2b. However, in contrast to the dual probe, none of these affinity probes was able to effectively capture p50, either in buffer, in cell lysates, or in nuclear extracts (Fig. 2c & S6†).



Intrigued by this result, we further tested more transcription factors: TATA-binding protein (TBP),<sup>64</sup> Myc-associated factor X (MAX),<sup>65</sup> and CREB1.<sup>66</sup> Matching pairs of TF-BP/FAM-CP and several series of “conventional probes” were prepared for each transcription factor respectively (Fig. 3). These probes were subjected to the same labeling procedures as in Fig. 2 and their performances were compared. First, all pairs of BP/CPs can capture their respective protein targets (Fig. 3; lane 1 and 2) and also showed specificity similar to the p50 probes (Fig. S7†). Interestingly, although TBP is known to primarily interact with DNA's minor groove,<sup>67</sup> none of the “conventional probes” (with the crosslinker in the major groove) showed detectable labeling (Fig. 3a; lane 3–8). However, MAX-T2, which has the diazirine crosslinker immediately next to the binding site, was able to capture the MAX protein (Fig. 3b), and MAX-T1 and T3, with the diazirine inside and away from the binding site respectively, showed very little MAX capture. Although MAX and CREB1 are

both leucine zipper family proteins and they bind DNA's major groove very similarly,<sup>68,69</sup> all CREB1 probes can capture the CREB1 protein.

We reason that there may be two possible underlying reasons for these observations: (i) the diazirine crosslinker may have sterically hindered the protein binding, as suggested by several crystal structures of TF-DNA complexes;<sup>70,71</sup> (ii) the specific structure and conformation of the “conventional probes” do not allow for a productive crosslinking (*e.g.*, the linker connecting the diazirine to DNA may be too short or lack sufficient flexibility).<sup>43,52</sup> With the dual-probe method, the crosslinker may have better flexibility and its spatial position can be feasibly varied to access the protein target without having to be part of the binding probe. In order to test this, we compared the labeling of p50, MAX, and TBP with BP/CP pairs having different “*n* values” (*n* represents the number of protruding or recessing nucleobases after BP/CP hybridization; Fig. 4a). Results show that, in general, capture probes with positive *n* values gave higher yields than the ones with negative ones, possibly because protruding bases provide better protein access for the crosslinker (*e.g.*: similar to a long and flexible linker). *n* = 0 appeared to be optimal in most cases (Fig. 4b).

Collectively, these results have demonstrated that the effectiveness of probes with directly conjugated crosslinkers indeed depends on the specific probe structure and the specific protein–DNA interaction, while the dual-probe strategy is more generally applicable, and it has the advantage of having a separate, tuneable, and target-binding independent probe that can effectively capture and label the protein target.

Furthermore, we tested our method with endogenously expressed proteins. Taking advantage of the method's modularity, we used a 5'-biotin-tagged capture probe to pair with the existing p50-BP so that any p50-BP-binding proteins can be isolated by affinity pull-down. After incubation of these probes in p50-overexpressed HEK293T cell lysate, light irradiation at

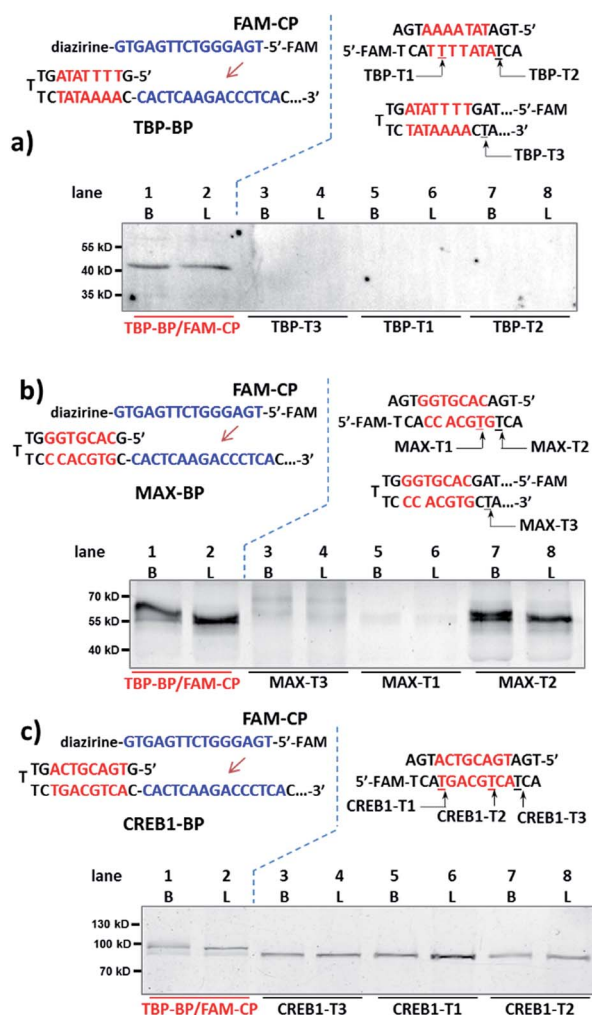


Fig. 3 Binding probe and capture probe sequences and protein labeling results analysed by denaturing SDS-PAGE for (a) TBP, (b) MAX and (c) CREB1. Reaction procedure and conditions are the same as in Fig. 2b and c, except that no BSA was added. Diazirine sites are underlined. B: reactions in buffer; L: reactions in HeLa lysate (4.8 mg mL<sup>-1</sup>, spiked with 2 μM transcription factor protein).

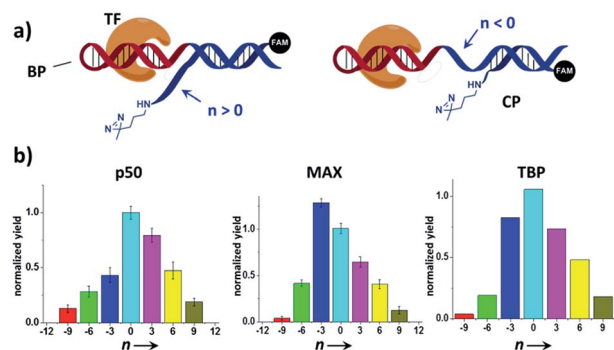


Fig. 4 (a) Varying the BP/CP hybridization site to change the diazirine position relative to the transcription factor target. “*n*” denotes the number of protruding (*n* > 0) or recessing (*n* < 0) nucleobases after probe hybridization. (b) BP/CP pairs of different *n* values were subjected to the same protein labeling procedure as in Fig. 2 for p50, MAX, and TBP. Labeling yields were determined by measuring the TF-CP fluorescence in denaturing SDS-PAGE and normalized to *n* = 0. Error bars (SD) in the p50 and MAX experiments are based on three separate experiments. The TBP experiments were performed only once.



365 nm, and then ultracentrifugation to remove free probes (MWCO: 50 kDa), the biotinylated species were captured by streptavidin beads. After elution, Western blots with anti-biotin and anti-p50 antibodies show protein bands matching the expected molecular weight of the p50-CP conjugate (Fig. 5a, lane 1; Fig. 5b, lane 2), which was not observed with a non-p50-binding negative control probe. These probes have shown excellent capture specificity in cell lysate with no significant enrichment of other proteins observed; a few protein bands appeared at high molecular weight in the anti-biotin blot, which may be from endogenous biotinylated species as they also showed up with the negative control (Fig. 5a, lane 2).

Further, we investigated whether our strategy can be used conversely to select protein-binding sequences from a “DNA-encoded probe library” for a particular transcription factor target, conceptually similar to the selection of DNA-encoded small molecule libraries against protein targets.<sup>56,72–82</sup> Our design is shown in Fig. 6a, a “DNA-encoded probe library” contains many BP/CP pairs with different sequences. The DNA sequence of the TF-binding site (S1) in BP is encoded by the DNA sequence of the CP-hybridization site (S2). Correspondingly, the hybridization site in the complementary CP (S2') is further encoded by a 3-base sequence (S3) at a distal location. In a library selection, the transcription factor target binds to the BP which contains matching S1 sequence, then BP templates target photo-crosslinking with the complementary CP to form the protein-CP conjugate. Therefore, the original target-binding S1 sequence can be decoded by reading the base sequence in the S3 site. In order to demonstrate this, first, a “probe library” composing of five equal ratio BP/CP pairs was prepared; in this library, only one BP/CP pair contains the matching p50-binding site, which is encoded by a “TTT” sequence in the S3 site (see details in Fig. S8†). This probe library was incubated with p50 and irradiated at 365 nm; the p50-CP conjugate generated was gel-purified, PCR-amplified and then sequenced. Results show that the p50-binding-encoding “TTT” was clearly enriched at the

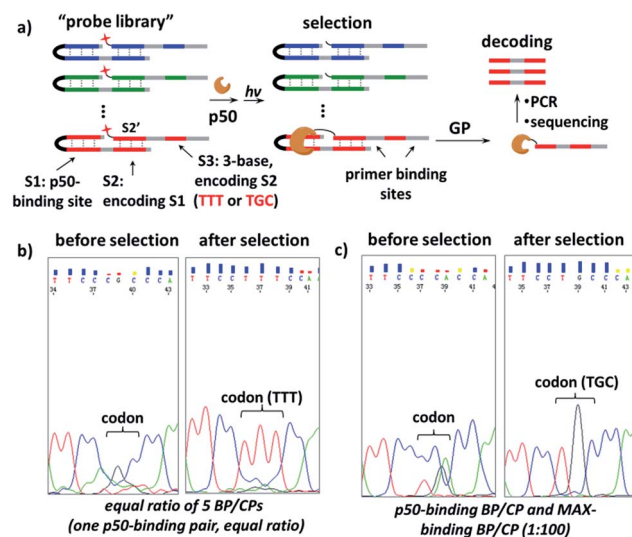


Fig. 6 (a) Scheme for the selection of a “probe library” against the p50 target. GP: gel purification. (b) and (c) Sequencing results of the encoding S3 sites before and after selection with the probe library containing (b) equal ratio of 5 different BP/CP pairs or (c) one pair of p50-binding BP/CP and one MAX-binding BP/CP at the ratio of 1 : 100. Probe library: 20  $\mu\text{M}$ ; p50 target: 4  $\mu\text{M}$ . Other conditions are the same as in Fig. 2. See Fig. S8 and S9† for details on the DNA sequences, selection and DNA sequencing; see Fig. S10 and S11† for full images of the sequencing data.

S3 site after selection (Fig. 6b). In a second “probe library”, a pair of p50-binding BP/CP, encoded by a “TGC” sequence at the S3 site, was mixed with 100-fold excess of MAX-binding BP/CP (see details in Fig. S9†). This library was also selected against the p50 target and again the encoding “TGC” was distinctly enriched (see the ESI for details; Fig. S10 and S11†). These selection results suggest that our strategy may be used as a selection method for the identification of target sequences for DNA-binding proteins.

Finally, we studied proteins recognizing DNAs containing 5-methyl-C (mC) and 5-hydroxymethyl-C (hmC) sequences, two important epigenetic marks implicated in gene transcriptions.<sup>83,84</sup> We prepared binding probes containing mC and hmC sites (mC-BP and hmC-BP; Fig. 7a), respectively, and a control probe without cytosine modification (C-BP).<sup>85</sup> With the capture probe (C-CP), these probes were applied to pull-down experiments in HEK293T lysate overexpressing MeCP2, a well-known protein recognizing both of these two modifications.<sup>85,86</sup> For mC-BP, Western blots showed specific enrichment of the MeCP2 protein (Fig. 7b, left and middle panels). Importantly, it was not observed with the control probe C-BP. mC-BP also specifically enriched another band at  $\sim 65$  kD, which can be blotted by the anti-MBD1 antibody, and MBD1 is known to bind mC sites on DNA.<sup>87,88</sup> Similarly, for hmC-BP, specific enrichment of MeCP2 was also observed (Fig. 7c). The band at  $\sim 40$  kD was identified as possibly to be MBD3, another protein reported that is able to bind hmC.<sup>86,89</sup> In addition, pull-down experiments in lysates without protein overexpression have identified several other mC- and hmC-binding proteins (Fig. S9 and S10†). Collectively, these results have demonstrated that our method

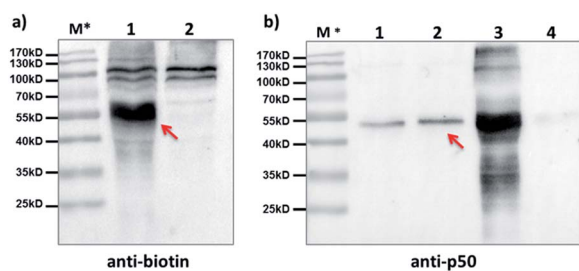


Fig. 5 After affinity pull-down with the p50-BP and 5'-biotin-CP in cell lysate, proteins captured by streptavidin beads were eluted and blotted by (a) anti-biotin antibody; lane 1: pull-down with p50-BP/5'-biotin-CP; lane 2: pull-down with a CREB1-binding control BP/5'-biotin-CP; and by (b) anti-p50 antibody; lane 1: 0.1 pmol purified p50; lane 2: pull-down with p50-BP/5'-biotin-CP; lane 3: cell lysate only; lane 4: pull-down with a CREB1-binding BP/5'-biotin-CP. BP and CP: 2  $\mu\text{M}$  each; HEK293T cell lysate: 6.8 mg mL<sup>-1</sup>, 1.2 mL used in pull-down; hv: 365 nm, 15 min, 0 °C. Elution buffer: 95% formamide, 40 mM NaOAc, 1 mM free biotin. Red arrows indicate captured p50-CP conjugates. M\*: overlaid ladder.



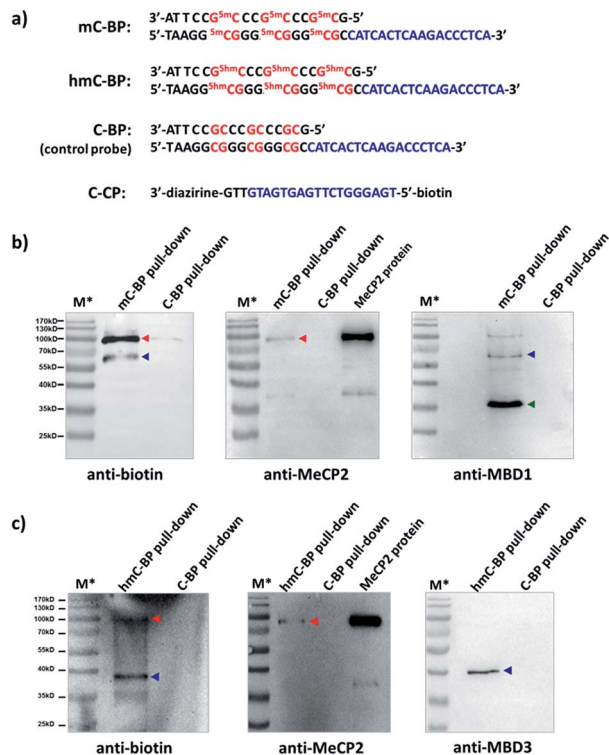


Fig. 7 (a) Sequences of mC-BP, hmC-BP, control C-BP, and the capture probe (C-CP). <sup>5mC</sup>: 5-methyl-C; <sup>5hmC</sup>: 5-hydroxymethyl-C. After affinity pull-down of DNA-binding proteins in HEK293T lysate with different probes, proteins captured by streptavidin beads were eluted and blotted by different antibodies. (b) mC-BP experiments in MeCP2-overexpressed lysate. (c) hmC-BP in MeCP2-overexpressed lysate. BP and CP: 20  $\mu$ M each; lysate: 7.44 mg mL<sup>-1</sup>, 0.2 mL used. Specific probes and antibodies used are marked. Arrows indicate captured proteins; in (b), red: MeCP2; blue: MBD1; green: possible an MBD1 degradation fragment as it did not show in the biotin blot (see Fig. S9b†); in (c), red: MeCP2; blue: MBD3. For all experiments: hv: 365 nm, 15 min, 0 °C; elution buffer: 95% formamide, 40 mM NaOAc, 1 mM free biotin. M\*: overlaid ladder. See the ESI† for experimental details.

may also be extended to study 5-methyl-C and 5-hydroxymethyl-C-binding proteins in epigenetic studies.

## Conclusions

In summary, we have developed a dual-probe method for characterizing transcription factor-DNA interactions and proteins recognizing epigenetic marks. By separating target recognition and capture, affinity probes can be feasibly designed to specifically capture and label DNA-binding proteins without affecting the original protein-DNA interactions. Binding probes are completely native DNAs which can be rapidly prepared in large quantity by automated DNA synthesis, making this method potentially suitable for high throughput identification of DNA-binding proteins in genomic studies.<sup>90,91</sup> On the other hand, chip-based large-scale *de novo* DNA synthesis<sup>92</sup> could be used to prepare probe libraries with diverse sequences, suitable for selections to identify DNA binding sequences for transcription factors and other DNA-binding

proteins. Currently our laboratory is actively exploring these opportunities.

## Acknowledgements

This work was supported by Ministry of Science and Technology Basic Research Program (2011CB809100), NSFC (21272016, 21002003, 91013003, J1030413) and the Doctoral Fund of Ministry of Education of China (20120001110083). We thank the mass spectrometry facility of National Center for Protein Sciences at Peking University and Dr Rong Meng for the kind help with mass spectroscopy analysis. We thank Prof. Huanran Tan at the Peking University Health Science Centre for the gift of the p50 plasmid.

## Notes and references

- 1 F. Spitz and E. E. Furlong, *Nat. Rev. Genet.*, 2012, **13**, 613–626.
- 2 M. M. Babu, N. M. Luscombe, L. Aravind, M. Gerstein and S. A. Teichmann, *Curr. Opin. Struct. Biol.*, 2004, **14**, 283–291.
- 3 J. M. Vaquerizas, S. K. Kummerfeld, S. A. Teichmann and N. M. Luscombe, *Nat. Rev. Genet.*, 2009, **10**, 252–263.
- 4 E. de Nadal, G. Ammerer and F. Posas, *Nat. Rev. Genet.*, 2011, **12**, 833–845.
- 5 T. Chen and S. Y. Dent, *Nat. Rev. Genet.*, 2014, **15**, 93–106.
- 6 V. M. Weake and J. L. Workman, *Nat. Rev. Genet.*, 2010, **11**, 426–437.
- 7 J. Stewart-Ornstein, C. Nelson, J. DeRisi, J. S. Weissman and H. El-Samad, *Curr. Biol.*, 2013, **23**, 2336–2345.
- 8 P. A. Konstantinopoulos and A. G. Papavassiliou, *JAMA, J. Am. Med. Assoc.*, 2011, **305**, 2349–2350.
- 9 J. E. Yeh, P. A. Toniolo and D. A. Frank, *Curr. Opin. Oncol.*, 2013, **25**, 652–658.
- 10 E. Sen, *Drug Discovery Today*, 2011, **16**, 1044–1051.
- 11 C. Yan and P. J. Higgins, *Biochim. Biophys. Acta*, 2013, **1835**, 76–85.
- 12 G. D. Stormo and Y. Zhao, *Nat. Rev. Genet.*, 2010, **11**, 751–760.
- 13 D. J. Galas and A. Schmitz, *Nucleic Acids Res.*, 1978, **5**, 3157–3170.
- 14 M. J. Storek, A. Ernst and G. L. Verdine, *Nat. Biotechnol.*, 2002, **20**, 183–186.
- 15 M. Fried and D. M. Crothers, *Nucleic Acids Res.*, 1981, **9**, 6505–6525.
- 16 M. M. Garner and A. Revzin, *Nucleic Acids Res.*, 1981, **9**, 3047–3060.
- 17 M. J. Solomon and A. Varshavsky, *Proc. Natl. Acad. Sci. U. S. A.*, 1985, **82**, 6470–6474.
- 18 T. Iwasaki, W. Miyazaki, N. Rokutanda and N. Koibuchi, *Biotechniques*, 2008, **45**, 445–448.
- 19 K. A. Vuori, J. K. Ahlskog, L. Sistonen and M. Nikinmaa, *FEBS J.*, 2009, **276**, 7366–7374.
- 20 M. F. Berger, A. A. Philippakis, A. M. Qureshi, F. S. He, P. W. Estep 3rd and M. L. Bulyk, *Nat. Biotechnol.*, 2006, **24**, 1429–1435.
- 21 A. Jolma, T. Kivioja, J. Toivonen, L. Cheng, G. Wei, M. Enge, M. Taipale, J. M. Vaquerizas, J. Yan, M. J. Sillanpaa,



- M. Bonke, K. Palin, S. Talukder, T. R. Hughes, N. M. Luscombe, E. Ukkonen and J. Taipale, *Genome Res.*, 2010, **20**, 861–873.
- 22 T. Heyduk and E. Heyduk, *Nat. Biotechnol.*, 2002, **20**, 171–176.
- 23 E. Knoll and T. Heyduk, *Anal. Chem.*, 2004, **76**, 1156–1164.
- 24 T. Krusinski, A. Ozyhar and P. Dobryszkycki, *Nucleic Acids Res.*, 2010, **38**, e108.
- 25 E. Heyduk, B. Dummit, Y. H. Chang and T. Heyduk, *Anal. Chem.*, 2008, **80**, 5152–5159.
- 26 I. Weibrecht, M. Gavrilovic, L. Lindbom, U. Landegren, C. Wahlby and O. Soderberg, *New Biotechnol.*, 2012, **29**, 589–598.
- 27 S. M. Gustafsdottir, J. Schlingemann, A. Rada-Iglesias, E. Schallmeiner, M. Kamali-Moghaddam, C. Wadelius and U. Landegren, *Proc. Natl. Acad. Sci. U. S. A.*, 2007, **104**, 3067–3072.
- 28 J. Wang, T. Li, X. Guo and Z. Lu, *Nucleic Acids Res.*, 2005, **33**, e23.
- 29 A. Vallee-Belisle, A. J. Bonham, N. O. Reich, F. Ricci and K. W. Plaxco, *J. Am. Chem. Soc.*, 2011, **133**, 13836–13839.
- 30 X. Xu, Z. Zhao, L. Qin, W. Wei, J. E. Levine and C. A. Mirkin, *Anal. Chem.*, 2008, **80**, 5616–5621.
- 31 A. Cao and C. Y. Zhang, *Anal. Chem.*, 2013, **85**, 2543–2547.
- 32 J. T. Kadonaga and R. Tjian, *Proc. Natl. Acad. Sci. U. S. A.*, 1986, **83**, 5889–5893.
- 33 M. F. Carey, C. L. Peterson and S. T. Smale, *Cold Spring Harb. Protoc.*, 2012, **2012**, 18–33.
- 34 G. L. Verdine and D. P. Norman, *Annu. Rev. Biochem.*, 2003, **72**, 337–366.
- 35 C. Yi, G. Jia, G. Hou, Q. Dai, W. Zhang, G. Zheng, X. Jian, C. G. Yang, Q. Cui and C. He, *Nature*, 2010, **468**, 330–333.
- 36 S. K. Hota, M. L. Dechassa, P. Prasad and B. Bartholomew, *Methods Mol. Biol.*, 2012, **809**, 381–409.
- 37 M. J. Schmidt and D. Summerer, *Angew. Chem., Int. Ed.*, 2013, **52**, 4690–4693.
- 38 K. M. Meisenheimer and T. H. Koch, *Crit. Rev. Biochem. Mol. Biol.*, 1997, **32**, 101–140.
- 39 G. W. Preston and A. J. Wilson, *Chem. Soc. Rev.*, 2013, **42**, 3289–3301.
- 40 J. L. Vinkenburg, G. Mayer and M. Famulok, *Angew. Chem., Int. Ed.*, 2012, **51**, 9176–9180.
- 41 M. Winnacker, S. Breger, R. Strasser and T. Carell, *ChemBioChem*, 2009, **10**, 109–118.
- 42 M. Winnacker, V. Welzmler, R. Strasser and T. Carell, *ChemBioChem*, 2010, **11**, 1345–1349.
- 43 D. Forget, M. F. Langelier, C. Therien, V. Trinh and B. Coulombe, *Mol. Cell. Biol.*, 2004, **24**, 1122–1131.
- 44 L. Lercher, J. F. McGouran, B. M. Kessler, C. J. Schofield and B. G. Davis, *Angew. Chem., Int. Ed.*, 2013, **52**, 10553–10558.
- 45 J. T. Reardon and A. Sancar, *Mol. Cell. Biol.*, 2002, **22**, 5938–5945.
- 46 M. Liebmann, F. Di Pasquale and A. Marx, *ChemBioChem*, 2006, **7**, 1965–1969.
- 47 U. K. Shigdel, J. Zhang and C. He, *Angew. Chem., Int. Ed.*, 2008, **47**, 90–93.
- 48 J. J. Tate, J. Persinger and B. Bartholomew, *Nucleic Acids Res.*, 1998, **26**, 1421–1426.
- 49 P. Mallikaratchy, Z. Tang, S. Kwame, L. Meng, D. Shangguan and W. Tan, *Mol. Cell. Proteomics*, 2007, **6**, 2230–2238.
- 50 S. D. Bruner, D. P. Norman and G. L. Verdine, *Nature*, 2000, **403**, 859–866.
- 51 J. C. Fromme and G. L. Verdine, *Nat. Struct. Biol.*, 2002, **9**, 544–552.
- 52 M. B. Renfrow, N. Naryshkin, L. M. Lewis, H. T. Chen, R. H. Ebright and R. A. Scott, *J. Biol. Chem.*, 2004, **279**, 2825–2831.
- 53 Z. Qiu, L. Lu, X. Jian and C. He, *J. Am. Chem. Soc.*, 2008, **130**, 14398–14399.
- 54 B. J. Lannutti, J. Persinger and B. Bartholomew, *Biochemistry*, 1996, **35**, 9821–9831.
- 55 G. Li, Y. Liu, L. Chen, S. Wu and X. Li, *Angew. Chem., Int. Ed.*, 2013, **52**, 9544–9549.
- 56 G. Li, Y. Liu, X. Yu and X. Li, *Bioconjugate Chem.*, 2014, **25**, 1172–1180.
- 57 L. Dubinsky, B. P. Krom and M. M. Meijler, *Bioorg. Med. Chem.*, 2012, **20**, 554–570.
- 58 A. Blencowe and W. Hayes, *Soft Matter*, 2005, **1**, 28.
- 59 M. Hashimoto and Y. Hatanaka, *Eur. J. Org. Chem.*, 2008, 2513–2523.
- 60 Z. Li, P. Hao, L. Li, C. Y. Tan, X. Cheng, G. Y. Chen, S. K. Sze, H. M. Shen and S. Q. Yao, *Angew. Chem., Int. Ed.*, 2013, **52**, 8551–8556.
- 61 R. Sen and D. Baltimore, *Cell*, 1986, **46**, 705–716.
- 62 D. Baltimore, *Nat. Immunol.*, 2011, **12**, 683–685.
- 63 M. Karin, Y. Yamamoto and Q. M. Wang, *Nat. Rev. Drug Discovery*, 2004, **3**, 17–26.
- 64 L. Comai, N. Tanese and R. Tjian, *Cell*, 1992, **68**, 965–976.
- 65 C. Grandori, S. M. Cowley, L. P. James and R. N. Eisenman, *Annu. Rev. Cell Dev. Biol.*, 2000, **16**, 653–699.
- 66 M. R. Montminy and L. M. Bilezikjian, *Nature*, 1987, **328**, 175–178.
- 67 J. L. Kim, D. B. Nikolov and S. K. Burley, *Nature*, 1993, **365**, 520–527.
- 68 A. R. Ferre-D'Amare, G. C. Prendergast, E. B. Ziff and S. K. Burley, *Nature*, 1993, **363**, 38–45.
- 69 M. A. Schumacher, R. H. Goodman and R. G. Brennan, *J. Biol. Chem.*, 2000, **275**, 35242–35247.
- 70 C. W. Muller, F. A. Rey, M. Sodeoka, G. L. Verdine and S. C. Harrison, *Nature*, 1995, **373**, 311–317.
- 71 G. Ghosh, G. van Duyne, S. Ghosh and P. B. Sigler, *Nature*, 1995, **373**, 303–310.
- 72 R. E. Kleiner, C. E. Dumelin and D. R. Liu, *Chem. Soc. Rev.*, 2011, **40**, 5707–5717.
- 73 L. Mannocci, M. Leimbacher, M. Wichert, J. Scheuermann and D. Neri, *Chem. Commun.*, 2011, **47**, 12747–12753.
- 74 M. A. Clark, R. A. Acharya, C. C. Arico-Muendel, S. L. Belyanskaya, D. R. Benjamin, N. R. Carlson, P. A. Centrella, C. H. Chiu, S. P. Creaser, J. W. Cuzzo, C. P. Davie, Y. Ding, G. J. Franklin, K. D. Franzen, M. L. Gefter, S. P. Hale, N. J. V. Hansen, D. I. Israel, J. W. Jiang, M. J. Kavarana, M. S. Kelley, C. S. Kollmann, F. Li, K. Lind, S. Mataruse, P. F. Medeiros, J. A. Messer, P. Myers, H. O'Keefe, M. C. Oliff, C. E. Rise, A. L. Satz, S. R. Skinner, J. L. Svendsen, L. J. Tang, K. van Vloten,



- R. W. Wagner, G. Yao, B. G. Zhao and B. A. Morgan, *Nat. Chem. Biol.*, 2009, **5**, 647–654.
- 75 M. C. Jean Pierre Daguier, S. Alvarez, S. Barluenga and N. Winssinger, *Chem. Sci.*, 2011, **2**, 9.
- 76 Z. J. Gartner, B. N. Tse, R. Grubina, J. B. Doyon, T. M. Snyder and D. R. Liu, *Science*, 2004, **305**, 1601–1605.
- 77 J. B. Doyon, T. M. Snyder and D. R. Liu, *J. Am. Chem. Soc.*, 2003, **125**, 12372–12373.
- 78 S. Melkko, J. Scheuermann, C. E. Dumelin and D. Neri, *Nat. Biotechnol.*, 2004, **22**, 568–574.
- 79 S. Melkko, Y. Zhang, C. E. Dumelin, J. Scheuermann and D. Neri, *Angew. Chem., Int. Ed.*, 2007, **46**, 4671–4674.
- 80 S. J. Wrenn, R. M. Weisinger, D. R. Halpin and P. B. Harbury, *J. Am. Chem. Soc.*, 2007, **129**, 13137–13143.
- 81 F. Buller, M. Steiner, K. Frey, D. Mirsof, J. Scheuermann, M. Kalisch, P. Buhlmann, C. T. Supuran and D. Neri, *ACS Chem. Biol.*, 2011, **6**, 336–344.
- 82 M. H. Hansen, P. Blakskjaer, L. K. Petersen, T. H. Hansen, J. W. Hojfeldt, K. V. Gothelf and N. J. V. Hansen, *J. Am. Chem. Soc.*, 2009, **131**, 1322–1327.
- 83 A. Bird, *Cell*, 1992, **70**, 5–8.
- 84 R. J. Klose and A. P. Bird, *Trends Biochem. Sci.*, 2006, **31**, 89–97.
- 85 C. G. Spruijt, F. Gnerlich, A. H. Smits, T. Pfaffeneder, P. W. Jansen, C. Bauer, M. Munzel, M. Wagner, M. Muller, F. Khan, H. C. Eberl, A. Mensinga, A. B. Brinkman, K. Lephikov, U. Muller, J. Walter, R. Boelens, H. van Ingen, H. Leonhardt, T. Carell and M. Vermeulen, *Cell*, 2013, **152**, 1146–1159.
- 86 M. Mellen, P. Ayata, S. Dewell, S. Kriaucionis and N. Heintz, *Cell*, 2012, **151**, 1417–1430.
- 87 S. H. Cross, R. R. Meehan, X. Nan and A. Bird, *Nat. Genet.*, 1997, **16**, 256–259.
- 88 B. Hendrich and A. Bird, *Mol. Cell. Biol.*, 1998, **18**, 6538–6547.
- 89 O. Yildirim, R. Li, J. H. Hung, P. B. Chen, X. Dong, L. S. Ee, Z. Weng, O. J. Rando and T. G. Fazzio, *Cell*, 2011, **147**, 1498–1510.
- 90 W. Z. Lin, J. A. Fang, X. Xiao and K. C. Chou, *PLoS One*, 2011, **6**, e24756.
- 91 J. Wang, J. Lu, G. Gu and Y. Liu, *J. Endocrinol.*, 2011, **210**, 15–27.
- 92 S. Kosuri and G. M. Church, *Nat. Methods*, 2014, **11**, 499–507.

