A 96-well DNase I footprinting screen for drug–DNA interactions

Tom Ellis[1], David A. Evans[1], Christopher R. H. Martin[1] and John A. Hartley[1,2,*]

[1]Spirogen Ltd, London Bioscience Innovation Centre, 2 Royal College Street, London, NW1 0NH and [2]Cancer Research UK Drug-DNA Interactions Research Group, Department of Oncology, University College London, 91 Riding House Street, London W1W 7BS, UK

ABSTRACT

The established protocol for DNase I footprinting has been modified to allow multiple parallel reactions to be rapidly performed in 96-well microtitre plates. By scrutinizing every aspect of the traditional method and making appropriate modifications it has been possible to considerably reduce the time, risk of sample loss and complexity of footprinting, whilst dramatically increasing the yield of data (30-fold). A semi-automated analysis system has also been developed to present footprinting data as an estimate of the binding affinity of each tested compound to any base pair in the assessed DNA sequence, giving an intuitive 'one compound–one line' scheme. Here, we demonstrate the screening capabilities of the 96-well assay and the subsequent data analysis using a series of six pyrrolobenzodiazepine-polypyrrole compounds and human Topoisomerase II alpha promoter DNA. The dramatic increase in throughput, quantified data and decreased handling time allow, for the first time, DNase I footprinting to be used as a screening tool to assess DNA-binding agents.

INTRODUCTION

DNase I footprinting is a powerful *in vitro* technique used to identify ligand–DNA interactions at specific DNA sequences (1,2). For the past two decades it has been the fundamental assay used to determine the sequence-selectivity for both proteins and DNA-binding compounds (3,4). Through manipulations of quantified data, it has become possible to further utilize DNase I footprinting to indirectly measure thermodynamic and kinetic properties of interactions (5,6). Typically it is possible, from one or two footprinting experiments, to locate a ligand's preferential binding sites on a desired short (100–200 bp) target DNA sequence, and characterize the ligand by calculating the binding affinities at such sites (7).

Unfortunately, as many researchers will testify, DNase I footprinting is a labour-intensive procedure which may take several days to prepare and carry out (2,7,8). Furthermore, it traditionally involves the undesirable use of radioisotopes and requires tricky ethanol precipitation steps. Modifications of the DNase I footprinting protocol have been reported, and these have sought to improve the assay through the removal of radioisotope use and/or ethanol precipitation steps (9–13). However, use of these modified assays has not become commonplace, possibly as all describe single-tube methods that ultimately still require significant labour in order to yield data.

For many researchers, including us, the greatest desire is to decrease the data turnover time of DNase I footprinting to maximize productivity. In the current age of drug discovery and 'omics', the need for medium/high-throughput assays is evident, with the same technique performed repeatedly, in order to satisfy large sample sets (14). This is achieved in most experiments by carrying-out reactions in parallel, with the 96-well microtitre plate as the industry standard. Working in the field of drug design and discovery, it was our own wish to use DNase I footprinting as a screening tool, increasing the throughput such that it could be used to assess libraries of compounds produced by combinatorial methods.

Therefore, through scrutiny of every aspect of the standard protocol for DNase I footprinting, we developed a rapid assay that utilizes the 96-well format and gives increased throughput along with decreased risk, error and processing times. In the process of developing the method, it also became clear that an equally expedient analysis procedure would be required to handle the large yield of data. To this end, a semi-automated analysis protocol was also designed, tying together gel quantification software with custom-designed programs to produce a single 'footprinting profile' for each ligand across several hundred base pairs of sequence. The resultant protocol

---

provides a quick and simple microtitre-based DNase I footprinting assay that is suitable for use as a quantitative screening-tool.

MATERIALS AND METHODS

DNA cloning and plasmid purification

The proximal 705 bp of the human Topoisomerase II alpha (TOPOIIα) promoter was amplified by PCR from human genomic DNA (Promega) using primers TIIA-F (CACCGCACACAGCCTAC) and TIIA-R (TGGTGACGGTCGTGAAG) (Supplementary Figure 1). The resulting fragment was ligated into the pGEM-T Easy plasmid vector using a commercial kit (Promega). After transforming *Escherichia coli* XL-10 Gold strain (Stratagene), plasmid containing the TOPOIIα fragment was amplified and purified using a midiprep purification system kit (Qiagen). The generated template plasmid was verified by DNA sequencing (UCL Services, London, UK).

Preparation of end-labelled DNA molecules

IR700 dye (LI-COR Biosciences) end-labelled DNA was produced in $48 \times 25 \mu l$ reactions by standard *Taq* (Sigma) PCR amplification from the plasmid template. Labelled and non-labelled PCR primers (Thermo) directed to the T7 and SP6 promoter sequences of pGEM-T Easy plasmid were used, with the IR700 dye attached to the 5′-end of the T7 primer when producing forward-labelled target, and attached to the 5′-end of the SP6 primer when producing reverse-labelled target. Products were purified by PCR clean-up column with 12 reactions per column (Qiagen) and finally eluted into a total of 300 μl of 0.1 mM Tris-HCl pH 8.5. The IR700 end-labelled DNA concentration of both targets was estimated by measuring optical density both at 260 nm wavelength (for DNA) and at 685 nm wavelength (for IR700 dye). DNA product was then diluted with 0.1 mM Tris-HCl pH 8.5 to give a 100 nM stock and this was stored at −20°C. In both cases, the DNA fragment consisted of the 705 bp TOPOIIα promoter region plus 177 bp of flanking pGEM-T Easy sequence.

Drug–DNA footprinting

In a V-bottom polypropylene 96-well plate (Eppendorf), 50 μl drug–DNA reactions were prepared against the two IR700 dye end-labelled DNA targets. In each case, six test agents were incubated with each DNA target at eight concentrations (one column) per drug: 0, 1.6, 8, 40, 200, 1000, 5000 and 25 000 nM. Serial drug dilutions were prepared beforehand in 100% DMSO at 20× final concentration allowing 2.5 μl to be added to each reaction giving a final mix of 5% DMSO, 50 mM KCl, 20 mM Tris-HCl pH 7.5, 1 mM MgCl₂, 0.5 mM DTT and 4 nM IR700 end-labelled DNA. Drug–DNA reactions were incubated for 17 h at room temperature before digestion for 8 min; initiated by adding 5 μl of 50 mM NaCl, 5 mM MgCl₂ and 5 mM MnCl₂ containing 0.02 units of DNase I (Promega). Digestion was stopped by adding 5 μl of 50 mM EDTA and digested DNA samples were purified with a 96-well vacuum filtration system (Montage SEQ96, Millipore), resuspending in 50 μl 0.1 mM Tris-HCl pH 8.5. Samples were concentrated to dryness with a Speedvac Evaporator (Thermo) and resuspended in 0.8 μl formamide containing 0.05% bromophenol blue.

Gel electrophoresis

Samples were denatured for 4 min at 95°C, and 48 samples (the first six columns) were loaded on a 52-lane sequencing gel [41 cm length, 0.25 mm thick, containing 1× TBE, 7 M urea, 10% formamide, 5% Sequagel XR (National Diagnostics, Hull, UK)]. A GA-specific chemical sequencing ladder (15) produced in advance was co-loaded in two lanes of the gel to act as markers. The gel was run for 5 h at 1500 V, 35 mA, 31.5 W, 55°C using a LI-COR model 4200 DNA Sequencer with e-Seq v2.0 Software (LI-COR Biosciences). The remaining 48 samples (the reactions with the reverse-labelled target) were subsequently loaded and run on a second gel exactly as the first 48 samples.

Footprinting gel image analysis

Raw intensity profiles (pixel intensity with regard to pixel position) for all 50 lanes were grouped in a CSV datasheet following 1D gel analysis of each 16-bit TIFF image using ImageQuant TL (GE Healthcare). With no background subtraction, band detection was then carried out in ImageQuant TL for the two marker lanes, ensuring that bands matched all the G and A positions within the analysed DNA sequence before exporting the lane measurement tables to a CSV datasheet.

Creation of differential cleavage plots

The program, 'footprint2', was used to convert the gel data into differential cleavage values. 'footprint2' is written in Perl, and the Tk toolkit is used to provide simple graphical input dialogs. The program input is CSV MS Excel datasheets for the lane intensity profiles, the band positions of one of the marker lanes (selected from the two analysed) and the analysed DNA sequence. The program proceeds in the following sequence:

(i) The raw intensity data (indexed by pixel) are sorted into bins, so that every lane has the same number of bins. This step corrects for variation in the lengths of the lanes.
(ii) The marker lane and DNA sequence input are analysed to convert the binned pixel data into data indexed by sequence position.
(iii) The intensity data are smoothed by a running average over a given number of base pairs, typically five.
(iv) Baseline correction. Each lane is separated into segments (typically ten), and the minimum in each segment found. These minima are fitted to a quadratic curve, and the curve used as the baseline to subtract.
(v) Amplitude correction. Again each lane is separated into segments (typically ten), and the range in each segment found. These ranges are fitted to a quadratic curve, and the intensity at each point is normalized by dividing by the fitted range value.

(vi) The differential cleavage value d_i for each base pair position i is then determined as $\log_{10}(I_i^l/Ii^0)$, where I_i^l is the normalized intensity of lane l at base pair i, and I_i^0 is the normalied intensity of the control 'no drug' lane at base pair i. For each drug, the output data are used to plot individual graphs of differential cleavage for all concentrations against the target DNA sequence.

The quadratic fitting was carried out using the Perl Data Language (PDL) extension to Perl (16). All gel bands were observed to have a systematic drift in baseline signal and peak amplitude that was not adequately corrected by a linear function. Polynomials of higher order than quadratic were not found to improve performance. Despite the multiple manipulations, running the program takes only about 5 s for each 50-lane gel.

Creation of footprint score graphs

Using a second Perl program, 'hitsblock2', the output file from the 'footprint2' program was combined with manually assigned differential cleavage cut-off values d_c for each drug to give a further output of a 'score' value assigned to each base pair in the sequence for each of the six agents tested per gel. The 'score' is defined as $\log_{10}c_c$, where c_c is the lowest concentration for each drug where the differential cleavage at the base pair is below the d_c value. For each drug output data were used to plot individual graphs of 'score' against the target DNA sequence.

Further information on the software programs described can be obtained from the corresponding author.

Standard ^{32}P-radiolabelled DNase I footprinting

As a comparison, compound 2 was footprinted against ^{32}P-labelled TOPOIIα promoter DNA using a previously described protocol (17) with minor modifications. A GA-specific chemical sequencing ladder (15) was used and drug was incubated with DNA at eight different concentrations; 0, 1.6, 8, 40, 200, 1000, 5000 and 25000 nM.

RESULTS

To demonstrate the DNase I footprinting screen, a series of DNA minor groove-binding compounds was assessed using the TOPOIIα promoter region previously characterized (17,18). The investigated compounds (Figure 1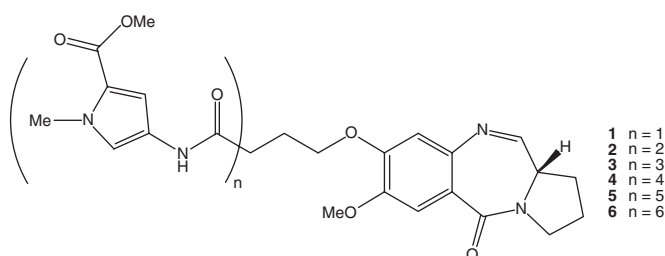) are an established series of novel methyl ester-terminated C8-linked pyrrolobenzodiazepine–poly(N-methylpyrrole) conjugates (19). This PBD-polypyrrole series increases in length from one to six consecutive pyrrole groups (1–6) held together by peptide linkers. For the experiments detailed here, a 705 bp section of the proximal promoter region of the TOPOIIα gene was investigated (Supplementary Figure 1). The DNA target was amplified by PCR to yield two footprinting fragments differing only by the strand on which they are 5′-end labelled with IR700 dye. The use of two fragments provided optimum resolution of footprints along the investigated promoter region.

DNase I footprinting in 96-well format

A layout of six compounds assessed against two footprinting fragments at eight different concentrations was found to give optimum use of the 96-well format and allowed all liquid handling steps to be carried out using standard multi-channel pipettes. Five-fold serial dilutions were chosen, in order to give a suitable 5^6 (~16000-fold) range of reaction concentrations for each compound. To circumvent any compound solubility issues, serial dilutions were performed in 100% DMSO to give $20\times$ concentration stocks in 96-well format that could be prepared in advance (and stored long-term at $-20°C$). The presence of 5% DMSO in the final reaction was tested for its effect on DNase I and found to be negligible (data not shown).

In the development of the method, the clean up of digestion products presented the only major hurdle to carrying out multiple parallel DNase I footprinting reactions in 96-well format. This step is typically achieved through ethanol precipitation of the DNA, a procedure that does not lend itself to microtitre plates. Thorough testing of many technologies including magnetic beads, dialysis and spin columns revealed the most efficient and usable DNA clean-up system to be a 96-well vacuum filter plate originally designed to purify Sanger-sequencing reaction products by membrane-based size-exclusion. With digestion products safely retained on the plate filter they can easily be washed, resuspended into low-salt buffer, concentrated to dryness and resuspended into the small volume of formamide (0.8 μl) required for loading on an automated sequencer.

Infrared DNase I footprinting

Previous research has described single-tube, non-radioisotopic DNA-protein DNase I footprinting experiments analysed on several automated sequencer systems including an automated laser fluorescent (ALF) sequencer (10), capillary sequencers (12,13) and a LI-COR infrared DNA sequencer (11). Of the three, the LI-COR system had the highest sensitivity due to its use of infrared dyes rather than traditional fluorescent dyes (e.g. fluorescein); however, the capillary sequencers were the fastest and easiest to use. Given that using low concentrations of DNA is essential for quantitative footprinting experiments (8), we chose to demonstrate our assay on the system with the highest sensitivity—the LI-COR 4200 automated DNA sequencer. This system uses the infrared

Figure 1. The investigated series of PBD-polyamide conjugate compounds.

fluorophore, IR-700 (Li-Cor), as a safe and high-sensitivity alternative to radioactive isotope labels such as ^{32}P and ^{33}P (20) and the subsequent data collection on an automated sequencer also allows for long read lengths and as many as 64 samples per gel (21). For quantifiable resolution, electrophoresis conditions were optimized, both with the TOPOIIα DNA fragments described here and with several other footprinting targets (data not shown). Although the sequencer was fully capable of resolving >800 nt in a single overnight run, it was found that beyond approximately 500 nt (5 h), individual bands could no longer be measured with a confidence suitable for the analysis. Additionally, it was found that standard denaturing polyacrylamide gels had to be supplemented with 10% formamide to increase further their denaturation strength so as to prevent the formation of DNA secondary structures (22).

The data from each run performed on the automated sequencer are displayed as a 16-bit TIFF image, with two high-resolution DNase I footprinting images produced per 96-well plate of reactions (Figure 2). Visual inspection of these gel images reveals a considerably increased yield of high quality footprinting data, unparalleled by conventional radioactive DNase I footprinting.

To verify that infrared DNase I footprinting provides reliable data equivalent to standard footprinting techniques, a 150 bp section from Figure 2A is compared to an equivalent ^{32}P-radiolabelled DNase I footprinting gel image in Supplementary Figure 2. Almost identical digestion patterns are seen at equivalent concentrations, despite the differences in label detection, reaction volumes and electrophoretic resolution.

Quantitative analysis of DNase I footprinting

Standard radioisope footprinting experiments are commonly subjected to quantitative analysis, in order to make thermodynamic and kinetic measurements of ligand–DNA interactions (6). A number of techniques have been previously described for quantitative analysis that interpret measurable changes in intensity of either a footprinted 'block' of the gel image or in peak-fitted individual bands (23,24). Furthermore, several research groups have made available software packages that carry out such analyses (24,25). By virtue of both the vast increase in quantifiable data per gel and the linear band separation afforded by the use of an automated sequencer, none of the described analytical methods was fully suited to the 96-well screening method described here. It was therefore necessary for us to make our own reassessment of footprint analysis, in order to establish a novel protocol serving the needs of the assay.

The analysis of footprint data requires two steps: (i) identifying footprints as regions of decreased band intensity in lanes where compound has been introduced, and (ii) locating the sites of these footprints by interpreting the local DNA sequence using the GA-specific chemical sequencing ladder. For our needs, we were confident that image quantification software packages (e.g. IQTL, GE Healthcare) would be adequately able to detect the necessary intensity changes; however, we were
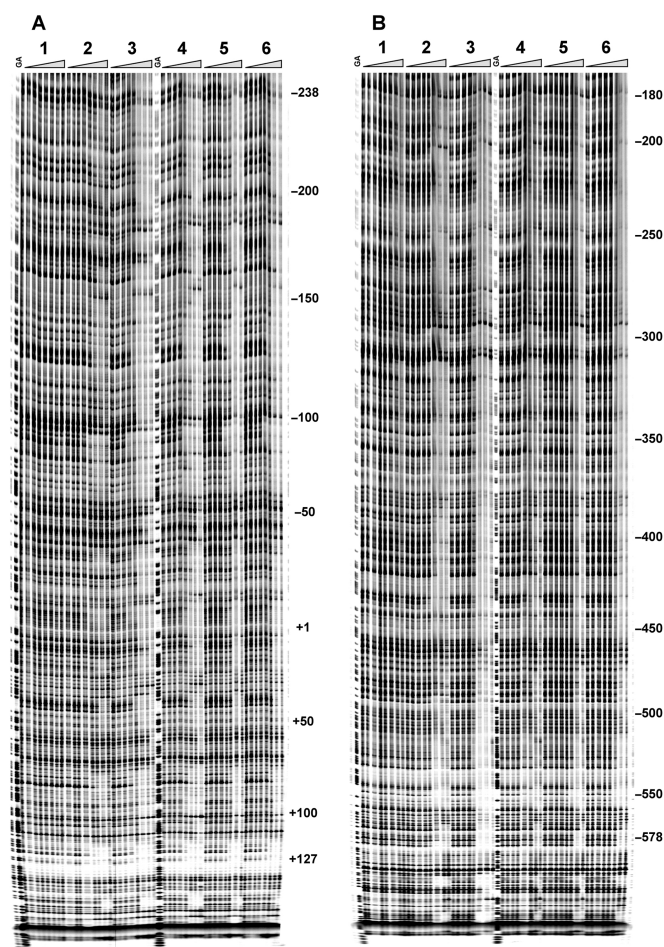

Figure 2. Infrared gel electrophoresis images of DNase I footprinting products following incubation of compounds 1–6 with TOPOIIα promoter DNA. Compounds were incubated for 17 h with TOPOIIα promoter DNA IR-labelled on the sense (**A**) or antisense (**B**) strand. Each gel consists of 50 lanes: two GA marker lanes (GA) and six sets of eight lanes for compound (l–r; 0, 0.0016, 0.008, 0.04, 0.2, 1, 5 and 25 µM). The sequence positions relating to Supplementary Figure 1 are given vertically.

required to develop a program to carry out the sequence interpretation. To this effect, a simple yet novel program was designed to perform the interpretation of data from image quantification, using the band positions in the GA-specific chemical sequencing ladder together with the sequence of analysed DNA to assign nucleotide positions to the intensity profiles of each sample lane.

With this section of the program in place, we then sought to improve on existing techniques to normalize sample data to take into account experimental inaccuracies due to pipetting error, variation in DNase I cleavage efficiency and common 'smiling' and 'warping' effects of gel electrophoresis. Traditionally, this is achieved by some form of background subtraction followed by normalization of all band intensities to selected 'invariant band intensities' (25). Selecting invariant bands/blocks was regarded as particularly undesirable given the large number of bands found on one of our infrared DNase I footprinting gel images (~25 000); thus, a variation of this was incorporated. First, when assigning nucleotide

positions to the intensity profiles of each sample lane, the variation in lane length due to gel electrophoresis effects was normalized. Also at this point, the slightly faster migration of bands in the chemical sequencing ladder was taken into account. Next, the nucleotide position-assigned intensity profiles of each sample lane were normalized for all other likely experimental inaccuracies via a series of transformations that utilized automatically selected invariants. Effectively by treating profiles as waveforms, normalization was achieved through polynomial transformations that modified the waveform amplitude, baseline and tilt so that the profile of each lane approximated that of the nearest appropriate control (no ligand) lane at all positions except those footprinted.

Following successful normalization and nucleotide position assignment, we chose initially to express intensity differences as logarithmic differential cleavage and present these graphically (Figure 3B) as in previously reported analyses (19,26,27). However, although such plots achieve an attractive graphical representation of footprint sites, we believed that they could be improved upon due to two inherent disadvantages: (i) lines are required for every concentration of each compound, making graphs cluttered and making it awkward to compare multiple compounds, and (ii) the experimenter is misled into interpreting the magnitude of the differential cleavage as the strength of interactions at the footprint—although in reality the magnitude is a factor of the original intensity of the bands in the no-compound control lane. To overcome these features, we set out to write a program that could use differential cleavage values to produce, for each compound, a single 'affinity' value for every DNA base. This 'one compound–one line' scheme intuitively reports the strength of the footprint at any DNA site and facilitates a rapid comparison of DNA-binding properties when multiple compounds are assessed.

Our initial intention was to use quantitative DNase I footprint titrations to determine an association constant (K_a) for every single nucleotide by fitting background-corrected band intensities to the Hill equation (28). Although this approach was successful at many binding sites, its universal use was not feasible as band intensity decreases were insufficient in regions that were not footprinted. Additionally, the requirement to model thousands of curves required more computing power and data manipulation than was practicable. Next, we looked at a simpler approach that can be adopted when the lowest concentrations at which footprints are observed are greater than the concentration of DNA in the samples (8). In this method, the ligand dissociation constant is effectively given by the ligand concentration required to produce a 50% decrease in band intensity (29). In the screen described here, observable decreases in band intensity due to binding are almost always greater than 50% due to the 5-fold concentration difference between sample lanes. Subsequently, it was possible to design a simple program to determine, for every nucleotide position, the ligand concentration at which a significant band intensity decrease is observed compared to the control. This value, referred to as the 'score' (log concentration), gives a good approximation of the
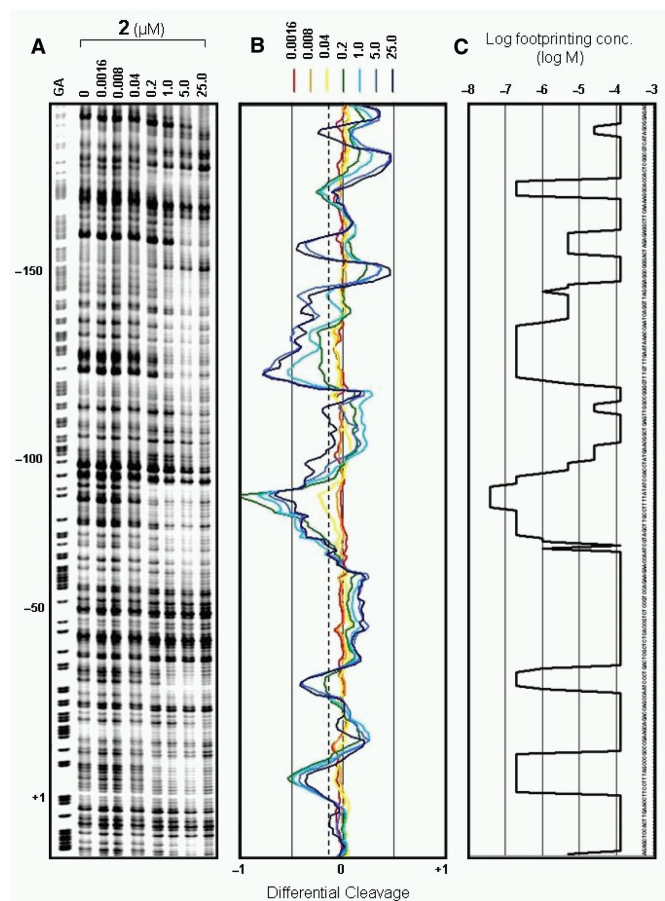

Figure 3. An example of the semi-automated analysis of DNase I footprints. A short region of the gel image shown in Figure 2A is used as an example to illustrate analysis of binding of compound 2 over 180 bp. By converting the lane images (A) into intensity profiles, a program-generated differential cleavage plot, (B) and a single 'score' profile (C) for the compound can be rapidly produced (the cut-off value is indicated by a broken line in the differential cleavage plot).

dissociation constant by effectively estimating the C_{50} value. The program uses the differential cleavage data and a user-defined differential cleavage cut-off value to determine a 'score' for every nucleotide position by reporting the lowest concentration at which the differential cleavage value at that base is lower than the cut-off value. The result is a 'score' profile across the tested concentration range that describes the apparent affinity of a compound to each position in the entire DNA sequence tested. Positions with no band intensity decrease over the tested range are assigned a 'score' equivalent to binding at 125 μM, 5-fold above the maximum tested concentration; which allows them to be easily disregarded. A plot of the 'score' profile allows primary binding sites to be distinguished from secondary and tertiary binding sites, as the primary sites are bound with a higher approximate association constant and thus receive a lower 'score' value (Figure 3C). In addition, the DNA-binding properties of multiple compounds can readily be compared when multiple 'score' profiles are displayed together (Figure 4).
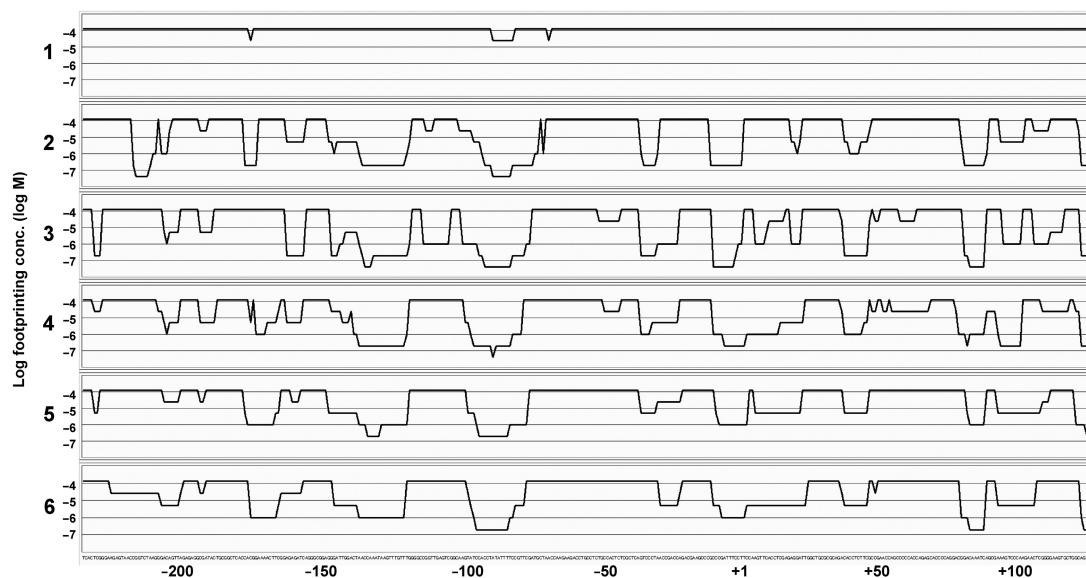
Figure 4. A comparison of footprint profiles of compounds 1–6 on the proximal region of the TOPOIIα promoter. Single footprint profiles were program-generated starting with the IR footprint gel image shown in Figure 3A. Where no footprinting is seen, the sequence has been given an arbitrary score of −3.9 (log 125 μM). The DNA sequence is given in Supplementary Figure 1.

DISCUSSION

A rapid 96-well DNase I footprinting screen was developed that gives readily comparable quantitative binding data for multiple test agents over long sequences of DNA. By increasing substantially the throughput of footprinting, we have demonstrated the suitability of this assay as a screening tool relevant in the current age of drug discovery. A typical radioisotope DNase I footprinting experiment might be able to assess up to a maximum of two ligands over 100 bp of DNA, whilst requiring approximately 3–4 days to proceed from an unlabelled DNA template to an analysable gel image (2,7,8). The 96-well format screen described herein permits the assessment of 12 ligands over 500 bp of DNA (or six ligands over 1 kb) in as little as two days, along with a greatly decreased proportion of user-time spent on difficult liquid-handling steps, which could ultimately be programmed to be carried-out by a benchtop robot system. This represents up to a 30-fold increase in data yield within a significantly shortened timescale. Indeed, within our lab, the screen is routinely used to assess 12 compounds over 500 bp every 24 h, using only one automated sequencer. The described assay has the further advantage of producing quantifiable, linearly-separated band intensities collected in real-time and does not require such unreliable steps as ethanol precipitation and the transfer of an electrophoresis gel to blotting paper.

As with other previously described modifications to DNase I footprinting (9–13), useful improvements are afforded by the introduction of an automated DNA sequencing system and the omission of radioisotope use—although in a well-equipped laboratory it would be perfectly reasonable to use the 96-well procedure with radiolabelled DNA. Despite the relative expense of automated DNA sequencing systems, their versatility allows them to be used for many other applications, notably DNA sequencing (30), AFLP analysis, (31) electrophoretic mobility shift assays (32) and *in vivo* footprinting (33). Additionally, the dye labels utilized are considerably safer, have long lifetimes and are low in cost (34). It is also common to find that such systems are available for communal use in many research environments, and although fluorescent sequencing systems are more common than the LI-COR infrared system described here, continual improvement in the fluorescent dyes and optics used by automated sequencing systems should no longer prevent such fluorophores being used as an alternative to radioisotopes in exactly the same method described here.

Ultimately, the optimization of the practical steps for the DNase I footprinting screen were less challenging than the development of the analysis procedure. As many features of the screen were novel, no commercially available programs were suitable for the entire analysis. The final analysis system is, in effect, a hybrid between existing image analysis software and our own programs, and as such some degree of user-input is still required, in order to transfer data between steps and to make occasional informed judgements. Nevertheless, one advantage of a semi-automated procedure over a fully automated procedure is that it allows the user to monitor the analysis at several stages and in greater depth (25). With the DNase I footprinting screen in particular, the use of a semi-automated procedure allowed us to overcome the difficulty of having a wide range of band intensities within each lane (due to the inherent non-uniform cleavage of DNA by DNase I). By allowing the user to monitor the differential cleavage outputs, it became possible to ensure that the desired 'score' data gave an accurate description of the initial gel image data without compromising on

the speed of analysis. A further advantage of the semi-automated procedure is that it introduces flexibility to the analysis. With only minor modifications, our programs can be adjusted to analyse gel images of different dimensions (length, width, and number of lanes) or even to analyse electrophoresis results from alternative sequencing systems and from conventional radioisotope footprinting. Indeed, the whole design of our 96-well DNase I footprinting screen is intentionally modular, in order to permit the user to adapt the protocol to suit their needs. For example, a researcher wanting to screen a mutant library of transcription factors on several promoters could perform the DNase I footprinting in 96-well format (incubating with proteinase K before sample clean-up, in order to remove the protein content) before resolving the gel images using standard radio-isotope electrophoresis over several days. The subsequent data could then be processed rapidly using the semi-automated procedure providing the user with readily-comparable profiles of how each mutation effects transcription factor binding.

The 30-fold increase in throughput afforded by the 96-well DNase I footprinting assay allows, for the first time, the use of DNase I footprinting as a screening tool for assessing the DNA-binding properties of ligands, DNA-binding proteins and potentially therapeutic test agents. In our lab, we have found it possible to work at a capacity that allows as many as 42 compounds to be characterized per week against a 500 bp DNA target when using the 5^6 concentration range described here. The power of the assay is illustrated by the results described here, obtained using the PBD-polypyrrole series. The full series was characterized—including full analysis over 700 bp of DNA and preparation of labelled DNA—in a total time of just three days. By comparison, our previous characterization that included analysis of the series (over 200 bp of DNA) took more than four weeks to achieve (19). By comparing the 'score' profiles of a large number of compounds, differences in binding-site size, sequence-selectivity and affinity can all be determined with ease. In research programs to design novel drugs, these data are invaluable for selecting lead compounds, directing further synthesis and in determining the nature of drug–DNA interactions.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

Conflict of interest statement. None declared.

REFERENCES

1. Galas,D.J. and Schmitz,A. (1978) DNAse footprinting: a simple method for the detection of protein-DNA binding specificity. *Nucleic Acids Res.*, **5**, 3157–3170.
2. Bailly,C., Kluza,J., Martin,C., Ellis,T. and Waring,M.J. (2005) DNase I footprinting of small molecule binding sites on DNA. *Methods Mol. Biol.*, **288**, 319–342.
3. Fox,K.R. (1997) DNase I footprinting. *Methods Mol. Biol.*, **90**, 1–22.
4. Rippe,R.A., Brenner,D.A. and Tugores,A. (2001) Techniques to measure nucleic acid-protein binding and specificity. Nuclear extract preparations, DNase I footprinting, and mobility shift assays. *Methods Mol. Biol.*, **160**, 459–479.
5. Brenowitz,M., Senear,D.F., Shea,M.A. and Ackers,G.K. (1986) Quantitative DNase footprint titration: a method for studying protein-DNA interactions. *Methods Enzymol.*, **130**, 132–181.
6. Petri,V. and Brenowitz,M. (1997) Quantitative nucleic acids footprinting: thermodynamic and kinetic approaches. *Curr. Opin. Biotechnol.*, **8**, 36–44.
7. Trauger,J.W. and Dervan,P.B. (2001) Footprinting methods for analysis of pyrrole-imidazole polyamide/DNA complexes. *Methods Enzymol.*, **340**, 450–466.
8. Fox,K.R. and Waring,M.J. (2001) High-resolution footprinting studies of drug-DNA complexes using chemical and enzymatic probes. *Methods Enzymol.*, **340**, 412–430.
9. Sandaltzopoulos,R. and Becker,P.B. (1994) Solid phase DNase I footprinting: quick and versatile. *Nucleic Acids Res.*, **22**, 1511–1512.
10. Sandaltzopoulos,R., Ansorge,W., Becker,P.B. and Voss,H. (1994) Nonradioactive, solid-phase DNase I footprints analyzed on an A.L.F. DNA sequencer. *Biotechniques*, **17**, 474, 476, 478.
11. Machida,M., Kamio,H. and Sorensen,D. (1997) Long-range and highly sensitive DNase I footprinting by an automated infrared DNA sequencer. *Biotechniques*, **23**, 300–303.
12. Yindeeyoungyeon,W. and Schelm,M.A. (2000) Footprinting with an automated capillary DNA sequencer. *Biotechniques*, **29**, 1034–1036, 1038, 1040–1031.
13. Wilson,D.O., Johnson,P. and McCord,B.R. (2001) Nonradiochemical DNase I footprinting by capillary electrophoresis. *Electrophoresis*, **22**, 1979–1986.
14. Wang,K., Gan,L., Boysen,C. and Hood,L. (1995) A microtiter plate-based high-throughput DNA purification method. *Anal. Biochem.*, **226**, 85–90.
15. Bencini,D.A., O'Donovan,G.A. and Wild,J.R. (1984) Rapid chemical degradation sequencing. *Biotechniques*, **2**, 4–5.
16. PDL – The Perl Data Language. http://pdl.perl.org.
17. Tolner,B., Hartley,J.A. and Hochhauser,D. (2001) Transcriptional regulation of topoisomerase II alpha at confluence and pharmacological modulation of expression by bis-benzimidazole drugs. *Mol. Pharmacol.*, **59**, 699–706.
18. Isaacs,R.J., Harris,A.L. and Hickson,I.D. (1996) Regulation of the human topoisomerase IIalpha gene promoter in confluence-arrested cells. *J. Biol. Chem.*, **271**, 16741–16747.
19. Wells,G., Martin,C.R., Howard,P.W., Sands,Z.A., Laughton,C.A., Tiberghien,A., Woo,C.K., Masterson,L.A., Stephenson,M.J. *et al.* (2006) Design, synthesis, and biophysical and biological evaluation of a series of pyrrolobenzodiazepine-poly(N-methylpyrrole) conjugates. *J. Med. Chem.*, **49**, 5442–5461.
20. Steffens,D.L., Jang,G.Y., Sutter,S.L., Brumbaugh,J.A., Middendorf,L.R., Muhlegger,K., Mardis,E.R., Weinstock,L.A. and Wilson,R.K. (1995) An infrared fluorescent dATP for labeling DNA. *Genome Res.*, **5**, 393–399.
21. Middendorf,L.R., Bruce,J.C., Bruce,R.C., Eckles,R.D., Grone,D.L., Roemer,S.C., Sloniker,G.D., Steffens,D.L., Sutter,S.L. *et al.* (1992) Continuous, on-line DNA sequencing using a versatile infrared laser scanner/electrophoresis apparatus. *Electrophoresis*, **13**, 487–494.
22. Humphries,S.E., Gudnason,V., Whittall,R. and Day,I.N. (1997) Single-strand conformation polymorphism analysis with high throughput modifications, and its use in mutation detection in familial hypercholesterolemia. International Federation of Clinical Chemistry Scientific Division: Committee on Molecular Biology Techniques. *Clin. Chem.*, **43**, 427–435.
23. Takamoto,K., Chance,M.R. and Brenowitz,M. (2004) Semi-automated, single-band peak-fitting analysis of hydroxyl radical

nucleic acid footprint autoradiograms for the quantitative analysis of transitions. *Nucleic Acids Res.*, **32**, 119.

24. Shadle,S.E., Allen,D.F., Guo,H., Pogozelski,W.K., Bashkin,J.S. and Tullius,T.D. (1997) Quantitative analysis of electrophoresis data: novel curve fitting methodology and its application to the determination of a protein-DNA binding constant. *Nucleic Acids Res.*, **25**, 850–860.

25. Das,R., Laederach,A., Pearlman,S.M., Herschlag,D. and Altman,R.B. (2005) SAFA: semi-automated footprinting analysis software for high-throughput quantification of nucleic acid footprinting experiments. *RNA*, **11**, 344–354.

26. Crow,S.D., Bailly,C., Garbay-Jaureguiberry,C., Roques,B., Shaw,B.R. and Waring,M.J. (2002) DNA sequence recognition by the antitumor drug ditercalinium. *Biochemistry*, **41**, 8672–8682.

27. Martin,C., Ellis,T., McGurk,C.J., Jenkins,T.C., Hartley,J.A., Waring,M.J. and Thurston,D.E. (2005) Sequence-selective interaction of the minor-groove interstrand cross-linking agent SJG-136 with naked and cellular DNA: footprinting and enzyme inhibition studies. *Biochemistry*, **44**, 4135–4147.

28. Trauger,J.W., Baird,E.E., Mrksich,M. and Dervan,P.B. (1996) Extension of sequence-specific recognition in the minor groove of DNA by pyrrole-imidazole polyamides to 9-13 base pairs. *J. Am. Chem. Soc.*, **118**, 6160–6166.

29. Dabrowiak,J.C. and Goodisman,J. (1989) In Kallenbach,N. R. (ed), *Chemistry and Physics of DNA-Ligand Interactions*, Adenine Press, New York, pp. 143–174.

30. Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y., Arakawa,T., Hara,A., Fukinishi,Y. *et al.* (2001) Functional annotation of a full-length mouse cDNA collection. *Nature*, **409**, 685–690.

31. Myburg,A.A., Remington,D.L., O'Malley,D.M., Sederoff,R.R and Whetten,R.W (2001) High-throughput AFLP analysis using infra-red dye-labeled primers and an automated DNA sequencer. *Biotechniques*, **30**, 348–352, 354, 356–347.

32. Sano,M., Ohyama,A., Takase,K., Yamamoto,M. and Machida,M. (2001) Electrophoretic mobility shift scanning using an automated infrared DNA sequencer. *Biotechniques*, **31**, 1056–1058, 1060, 1062.

33. Dai,S.M., Chen,H.H., Chang,C., Riggs,A.D. and Flanagan,S.D. (2000) Ligation-mediated PCR for quantitative in vivo footprinting. *Nature Biotechnol.*, **18**, 1108–1111.

34. McIndoe,R.A., Hood,L. and Bumgarner,R.E. (1996) An analysis of the dynamic range and linearity of an infrared DNA sequencer. *Electrophoresis*, **17**, 652–658.