754–763 Nucleic Acids Research, 2013, Vol. 41, No. 2 Published online 24 November 2012
doi:10.1093/nar/gks1085

STAT1:DNA sequence-dependent binding modulation by phosphorylation, protein:protein interactions and small-molecule inhibition

Andrew J. Bonham[1], Nikola Wenta[2], Leah M. Osslund[1], Aaron J. Prussin II[1], Uwe Vinkemeier[2] and Norbert O. Reich[1,*]

[1]Department of Chemistry & Biochemistry, University of California, Santa Barbara, CA 93106, USA and [2]School of Biomedical Science, University of Nottingham, Nottingham NG7 2UH, UK

Received July 3, 2012; Revised October 14, 2012; Accepted October 16, 2012

ABSTRACT

The DNA-binding specificity and affinity of the dimeric human transcription factor (TF) STAT1, were assessed by total internal reflectance fluorescence protein-binding microarrays (TIRF-PBM) to evaluate the effects of protein phosphorylation, higher-order polymerization and small-molecule inhibition. Active, phosphorylated STAT1 showed binding preferences consistent with prior characterization, whereas unphosphorylated STAT1 showed a weak-binding preference for one-half of the GAS consensus site, consistent with recent models of STAT1 structure and function in response to phosphorylation. This altered-binding preference was further tested by use of the inhibitor LLL3, which we show to disrupt STAT1 binding in a sequence-dependent fashion. To determine if this sequence-dependence is specific to STAT1 and not a general feature of human TF biology, the TF Myc/Max was analysed and tested with the inhibitor Mycro3. Myc/Max inhibition by Mycro3 is sequence independent, suggesting that the sequence-dependent inhibition of STAT1 may be specific to this system and a useful target for future inhibitor design.

INTRODUCTION

Transcriptional regulation in eukaryotes is complex (1,2) and regulated by processes as diverse as the translocation of transcription factors (TFs) into the nucleus (3) and expansion of compacted DNA by chromatin remodeling factors. TFs play an essential role by directing RNA polymerase complexes to gene targets. Understanding the combinatorial association of TFs with preferred DNA sequences, the cistrome (4) of the cell, is an ongoing challenge for molecular biology. Strategies such as chromatin immunoprecipitation coupled to microarray (ChIP-chip) (5) or high-throughput sequencing (ChIP-seq) (6) have provided novel insights into genome-wide association profiles. Similarly, the binding preferences of large numbers of TFs have been identified using protein-binding microarrays (PBMs) (4,7,8). However, the next generation of such studies will need to embrace the distinction that TFs rarely act in isolation *in vivo*. Indeed, it is generally thought that the combinatorial assembly of TFs provides much of the complexity inherent to genetic networks (9–11), and studies of multi-protein complexes, such as the basal transcription initiation complex (10) and the interferon enhanceosome (12), reveal the functional diversity controlled by protein:protein interactions.

Here we use our recently developed, microarray-based tool TIRF-PBM (total internal reflectance fluorescence protein-binding microarray) (13) to study the modulation of DNA-binding specificity of STAT1, a member of the STAT (signal activator and transducer) family of human TFs, as well as that of Myc/Max, another well-characterized human TF. Recently, a paradigm shift in the understanding of STAT biochemistry has caused a necessary re-evaluation of the role of phosphorylation, dimerization and polymerization in the DNA-binding preference of the STATs (14). STAT1 exists as a functional dimer (15), and the STATs are heavily implicated in oncogenesis when aberrantly upregulated (16,17), often associated with leukemia, lymphoma and myeloma, although STAT1 is rarely the primary STAT active in these cases (17–19). However, the STATs also serve important roles when constitutively active in normal cells, complicating strategies designed to interfere with their function (3,16,20). TIRF-PBM is well suited to address

*To whom correspondence should be addressed. Tel: +1 805 893 8368; Fax: +1 805 893 4120; Email: reich@chem.ucsb.edu
Present address:
Andrew J. Bonham, Department of Chemistry, Metropolitan State University of Denver, Denver, CO 80217, USA.

fundamental questions about the multi-protein complexes involved in transcriptional regulation.

We have evaluated several critical aspects of STAT1 biology, both of the activated, phosphorylated form, as well as the less active, unphosphorylated form. STAT1 is known to form larger assemblies than dimers when bound to DNA, including tetramers and extended polymers, which may significantly influence *in vivo* binding preferences (14). We evaluated the effect on DNA binding with or without the presence of the N-terminal domain, required for STAT1 polymerization. Due to their critical roles in tumorigenesis, there has been great interest in finding ways to regulate TF function in ways that are specific to individual proteins (16). In this study, we evaluated the efficacy of several small molecule inhibitory compounds (21) to reduce DNA-binding affinity and to investigate the possibility of sequence-dependent effects in STAT1 or Myc/Max binding, which would serve as ideal targets for future drug discovery.

MATERIALS AND METHODS

DNA array preparation

Ninety-six DNA sequences with known interactions with Myc/Max and STAT proteins *in vivo* and *in vitro* (22–25) or from promoter regions associated with the proteins in ChIP-chip assays (26–29) were selected, along with non-binding sequences as controls. dsDNA sequences were generated by primer extension of 5′ amino terminated, 51-mer template strands as previously described (13). Full DNA sequences are available in Supplementary Table S1. dsDNA-containing polyacrylamide-epoxide hydrogels were generated as previously described (13). The printed hydrogel spot morphology was evaluated in the fully hydrated and dry states. Swelled hydrogels with DyLight-649 and DyLight-549 labeled DNA controls were observed using phase contrast microscopy (Olympus ITX 70) and fluorescent confocal microscopy (Olympus Fluoview 500). Dry hydrogel spots were examined using scanning electron microscopy (SEM) with a JELO-X40 microscope at beam size 3, beam energy of 3–7 kV. Hydrogel samples were prepared for SEM imaging by Hummer 6.2 gold sputtering (Technics). Hydrogel characterization available in Supplementary Figure S1.

Preparation of proteins

Phosphorylated STAT1 (P-STAT1), unphosphorylated STAT1 (U-STAT1) and truncated STAT1 (STAT1tc) were prepared as described previously (15). c-Myc and Max isoform were expressed separately in *E. coli* as recombinant, His-tagged proteins, then denatured and renatured together, as previously described (22). TATA-Binding Protein (TBP) was prepared as previously described (30). Purified proteins were fluorescently labeled with the amine-reactive dyes NHS-DyLight-649 and NHS-DyLight-549 (Pierce) and characterized as previously described for TIRF-PBM (13). Final dye-protein conjugates were evaluated for DNA-binding ability via electrophoretic mobility shift assay (EMSA) using P^{32}-labeled cognate DNA run on a 6% acrylamide gel at 4°C in 0.5× TBE for 2 h at 200 V. EMSA was used to confirm the expected binding affinity for P-STAT1 on GAS cognate DNA, with U-STAT1 displaying a >200-fold decrease in binding affinity, as well as expected binding affinity for Myc/Max and TBP (data not shown).

TIRF instrumentation

TIRF experiments were conducted using a homebuilt instrument to generate a uniform evanescent field across a plastic microscope slide printed with a microarray with temperature and flow rate control, described in previous work (13).

Reaction conditions

Care was taken to prevent non-specific association. To remove remaining reactive epoxide, a 10 mM Tris–HCL pH 8, 10 mM ethanolamine, 0.1% SDS solution was added to the flow cell for 10 min at 37°C prior to trials. To block non-specific protein contacts, the instrument was washed with PBS, 5% w/v bovine serum albumin (BSA), 1% v/v Tween-20 for 10 min at 37°C, then flushed with appropriate running buffer for the trial: Myc/Max running buffer (20 mM Tris–HCl, 60 mM KCl, 4% glycerol, 0.1 mg/ml BSA, pH 8.0), STAT1 running buffer (20 mM HEPES, 4% glycerol, 40 mM KCl, 40 mM $CaCl_2$, 2 mM DTT, 0.2 mg/ml BSA, pH 8.0), or TBP running buffer (1× PBS, 0.5% w/v BSA, 0.01% v/v Tween-20, 5 mM $MgCl_2$). Data collection used 20 s integrated exposures, at 25°C and a buffer flow rate of 50 µl/min. Running buffer with protein, or protein with inhibitor compound, was introduced and allowed to flow for >30 min. This was followed by the injection of running buffer without protein to induce dissociation of the TFs from DNA. After each trial, the flow cell was washed with 1× PBS, 1 M NaCl, 1% v/v Tween-20 to remove residual bound protein.

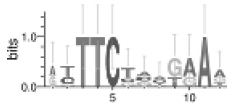
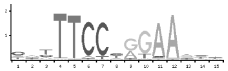
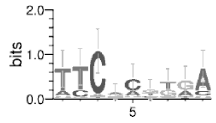
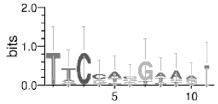
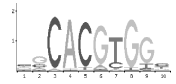
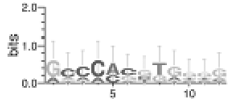
Data normalization and analysis

Each sequence was spotted on 5–10 replicate spots per array, and spot intensities for each sequence were quantified and used for all subsequent analyses of affinity and equilibrium-binding intensity, as described previously (13). Logos in Table 1 were generated as previously described (13).

Inhibitor compound preparation and analysis-

Inhibitors of STAT3 SH2 domain function (and presumably functional against closely related STAT SH2 domains) LL3 (21), FLLL1, FLLL12 and FLLL31 (Tom Li, Ohio State University) were suspended in DMSO at 5 mM, then added to TIRF-PBM trials with 50 nM P-STAT1 protein. For LLL3, final concentrations ranged from 5 µM to 200 µM. For FLLL1, FLLL12 and FLLL31, a final concentration of 20 µM was used. Mycro3, a small molecule inhibitor of Myc/Max dimerization, was suspended at 5 mM in DMSO and used at a concentration of 50 µM with Myc/Max. TIRF-PBM

Table 1. P-STAT1 and Myc/Max binding constants and logograms show the impact of dimerization and phosphorylation

TF	Consensus sequence	K_D, lit. (nM)	K_D^{app} TIRF (nM)	$k_{on,obs}$ ($10^6 M^{-1}$)	$k_{off,obs}$(10^{-3} M^{-1} s^{-1})	range of K_D^{app}	Logo TIRF	Logo JASPAR
P-STAT1	TTCnnnGAA	1–50	3.4 ±0.1	1.5 ±0.02	5.1 ±0.09	3 nM to 21 uM		
P-STAT1tc	n.d.	~1	7.9 ±1.4	1.5 ±0.2	12.4 ±0.8	2 nM to 4 uM		n.d.
U-STAT1	n.d.	>200	>400	<0.01	>40	>400 nM to >20 uM		n.d.
U-STAT1tc	n.d.	n.d.	>300	<0.07	>24	>300 nM to >30 uM		n.d.
Myc/Max	CACGTG	0.4–90.5	0.9 ±0.2	56 ±6.5	53 ±4.4	1 nM to 0.7 uM		
Myc	CACGTG	n.d.	>1000	~0.03 ± 0.6	~38.8 ±8.4	>1 uM		n.d.

n.d., not determined.

assays were conducted as above, and decreases in equilibrium binding intensity by inhibitor was assessed (Figure 2). For LLL3, equilibrium-binding intensity at different concentrations was used to derive IC_{50} values. STAT1 sequences were binned into three categories: 'High Affinity' sites, sequences with a consensus STAT1 palindromic GAS-binding site with a three nucleotide spacing between half elements; 'Half Sites', sequences with half of the palindromic GAS motif; and 'Poor Affinity' sites, sequences with either no GAS site or a GAS site with improper spacing between palindromic elements. For Myc/Max, sequences were binned into two categories: 'High affinity', sequences with consensus E-box binding DNA, and 'Low affinity', sites with mutated or absent Ebox consensus sites. Inhibition across these clusters of sequences was compared using pair-wise two-tailed Student's *t*-test with Welsh's correction.

RESULTS

Activated dimeric states are required for efficient DNA binding by STAT1 and Myc/Max

Prior efforts have shown that the P-STAT1/P-STAT1 homodimer and Myc/Max heterodimer are functionally distinct from the U-STAT1/U-STAT1 homodimer and Myc alone, respectively (31,32). For instance, phophorylation at tyrosine 701 on each STAT1 monomer allows them to orient and form the functional anti-parallel dimer (32). Recent work has shown that U-STAT1 is not a monomer, but exists in a distinct parallel dimeric orientation, with the N-terminal domains serving as the dimerization interface rather than the SH2 domains, as in the P-STAT1 dimer (32). A truncated form of STAT1 (STAT1tc) that lacks this N-terminal domain and a short C-terminal transcription activation domain (expressing amino acids 132-712) was included to probe this behavior. The binding specificity of the dimeric Myc/Max complex and the specificity of Myc alone were established to provide comparison during inhibitor trials. Our full set of proteins were individually evaluated for binding preference across the DNA sequence set, with TIRF-PBM assays used to generate equilibrium binding intensity values, using real-time measurements of binding (Figure 1b) captured to ensure that equilibrium was reached. These normalized intensity values show that each active dimer has a distinct set of highly preferred sequences (Figure 1a) which generally map to sequences that contain a full or single-base mismatch consensus site for the protein.

P-STAT1, the functional dimer, displays strong sequence preference for consensus binding sites. Among its high-affinity binding sites, we find the highest binding from three probes (sequences 70–72) that have optimal GAS site consensus [characterized by a DNA sequence of $TTCN_{2\text{-}4}GAA$(24,33)]. The most preferred site (sequence 70) possesses an interior spacer with sequence CCG, which matches the STAT1 consensus site in the JASPAR core database (28) (see Table 1 and Discussion section below). P-STAT1 showed uniformly low binding at 'deviant GAS sites' that lack a full consensus site. STAT1 prefers binding sites that have an intra-site spacer of three bases; other STAT family member proteins prefer to bind at sites with spacers from 2 bases to 4 bases (24,34), and we find that the majority of highequilibrium binding intensity interactions occur at sites with a 3 base spacer. Noticeably, sites with a 4 base spacer (e.g. sequences 28, 32 and 36) display very low binding, reflecting a greatly reduced specificity. The variability in binding intensity of P-STAT1 among GAS sites may reflect a dependence on spacer and flanking sequence, but agrees with the consensus site as the strongest determinant of specificity.

The binding of the truncated form of P-STAT (P-STAT1tc) showed overlap of binding sites with the full-length protein, but also displayed a large number of drastic changes in binding specificity; the truncation did not appreciably bind at some consensus sites well-occupied by the full-length protein, and bound well at some sites with no apparent GAS consensus targets. These differences suggest that the N-terminal domain may have a functional role in determining site specificity of the active dimeric protein.

Similarly to STAT1, the presence of an E-box (a known consensus motif for Myc/Max binding, sequence CACGT G) is the largest determinant of binding intensity (Table 1), consistent with prior *in vitro* and cell-based motif models in the JASPAR motif core database (28). As expected, on sites with no consensus binding site, the binding of Myc/Max is greatly diminished. Two exceptions involving highly repeating DNA are evident, with high binding intensity for DNA with a 12 adenine repeat (sequence 7) as well as modest binding for DNA with a 12 guanine repeat (sequence 34) (35). These observations support the model of Myc/Max binding specificity driven primarily by interactions with the E-box.

In contrast to the binding of the active dimers, the binding of U-STAT1, U-STAT1tc or Myc show very low binding and reduced specificity for consensus DNA (Figure 1a). This does not necessarily reflect a transition from active dimer to inactive monomer, as both U-STAT1 and Myc are able to form homodimers. Historically, STAT phosphorylation was thought to be required for activation (36), but recent work suggests that even the very weak-binding properties of U-STAT1 may have functional consequence in the cell, with U-STAT1 existing in an equilibrium between monomeric, anti-parallel dimeric (DNA-binding deficient) and a minority in the parallel (DNA binding) form (32). The weak specificities displayed by Myc are likely not biologically relevant, as Myc forms homodimers only at concentrations far in excess of that found in most cells (37), although Myc displayed greatest specificity to sequences 13 and 17, both of which contain the CACGTGG preferred consensus binding site.

Quantitative affinity and specificity data for STAT1 and Myc/Max

In prior work, we established that binding intensities in TIRF-PBM are a direct indication of binding affinity (13).

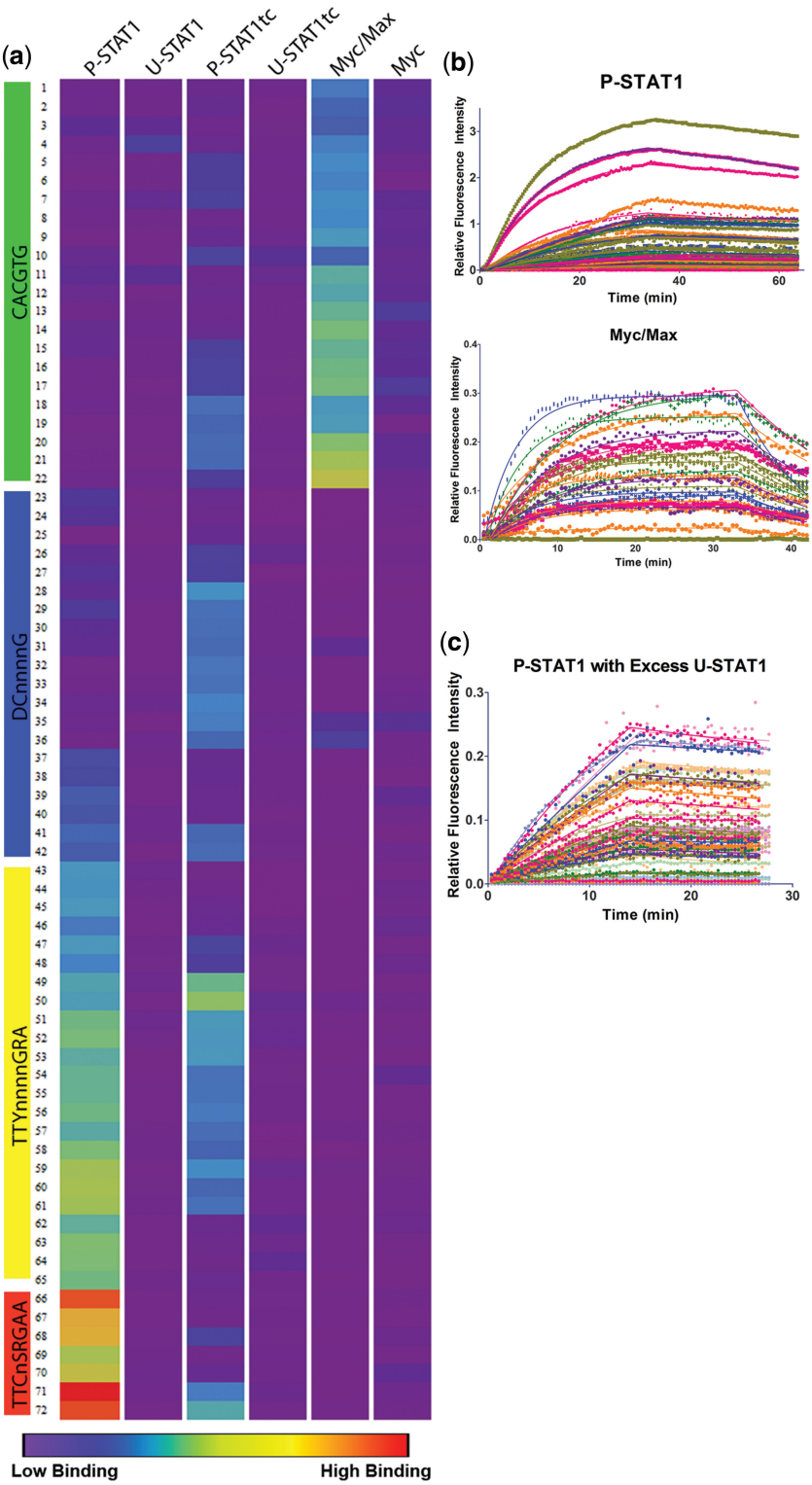


Figure 1. Analysis of the binding specificity of human TFs STAT1 and Myc/Max across DNA sequences containing cognate, mutant or no consensus binding sites in TIRF-PBM. (**a**) Equilibrium binding fluorescence intensities of DyLight 649 (used to label each protein) for each condition are shown in a spectrum from lowest (violet) to highest (red) signal (gray indicates poor quality data excluded from analysis). Colored bars on left separate sequences into regions of interest, and show consensus identities for all sequences in that region. (**b**) Insets show examples of binding equilibrium data displaying association and dissociation regimes for P-STAT1 and Myc/Max used to generate the equilibrium binding intensities. Each trace shows binding to a different sequence on the array. (**c**) The addition of a 10-fold molar excess of U-STAT1 to P-STAT1 does not change the binding preferences of P-STAT1. Representative examples shown from experiments where P-STAT1, 25 nM, with (light colors) or without (dark colors) a 20 min pre-incubation with 250 nM U-STAT1, was tested across an array of STAT recognition sequences. No significant differences in binding preference or affinity were observed.

To further probe this relationship, the real-time binding behavior and kinetics of P-STAT1 and Myc/Max across the TIRF-PBM array were measured (representative data in Figure 1b). The time-dependent binding intensity change observed is analogous to the association and dissociation traces common in surface plasmon resonance (SPR) measurements, and TIRF-PBM conditions have been optimized to avoid common pitfalls of these techniques (i.e. mass transport limitations) (13). We compared the overall observed binding affinity, K_D, on a consensus binding site in TIRF-PBM with comparable K_D measurements from the literature. P-STAT1 and Myc/Max have affinity constants that are accurate and precise compared to prior measurements (Table 1). For example, P-STAT1 in TIRF-PBM on an optimal GAS site (sequence 70) displays an affinity constant of 3.4 ± 0.1 nM, consistent with prior measurements (15), and the logogram of high-affinity binding sites (consensus motif TTCYVVGAA) reflects good agreement with known binding motifs in JASPAR (28) and TRANSFAC (*P*-value 0.00074) (38). The truncated form of STAT1, P-STAT1tc, shows a divergent binding pattern from the full-length protein (Figure 1a), but retains high-affinity (7.9 nM) binding at a GAS consensus site (sequence 51) and a logogram that is consistent with normal STAT1 binding (Table 1).

Myc/Max displayed a K_D of 0.9 ± 0.2 nM for its cognate DNA, consistent with the sub-nanomolar affinity values previously reported (37,39), and a binding logogram very similar to the motif from the JASPAR database (Table 1) (28). The JASPAR core Myc motif is derived from ChIP-seq data (40), and the tight correlation (*P*-value of 4.9e-05) between our observed motif and the ChIP derived motif suggests that PBMs are non-disruptive of biologically relevant patterns of interaction.

U-STAT1, U-STAT1tc and Myc display affinity constants for consensus DNA that are much weaker than that observed for the active forms. U-STAT1 displays weak affinity to optimal GAS site DNA (K_D of >400 nM), over a 100-fold impacted from P-STAT1 binding, consistent with prior observations (32), and reflective of the disruption of the proper SH2 domain-mediated protein:protein interface that allows the dimer's DNA-binding interface to form. Intriguingly, prior reports have shown that U-STAT1 can form a homodimer *in vitro* through a different interacting surface in the N-terminal region, leading to an anti-parallel dimeric form (32). Here, U-STAT1, despite its very low binding, displays a binding logogram showing preference for the 5′ half of the palindromic GAS consensus site, which agrees well with prior cell-based work that shows U-STAT1 can activate gene expression on only the 5′ end of the palindromic GAS site (41). To further investigate this behavior, we compared U-STAT1 binding to the binding of U-STAT1tc, which lacks the region responsible for the anti-parallel conformation and thus should be unable to form dimers. U-STAT1tc shows negligible binding in this study, but the weak binding shows a weak preference for the TTC region of half of the GAS consensus site, suggesting that monomeric STAT1 may exhibit binding to the GAS consensus site, although the affinity for these interactions is likely below the threshold of physiological importance. For Myc alone, binding affinity was also impacted compared to the Myc/Max heterodimer, with a measured K_D of >1 μM, over a 1000-fold loss in affinity. This drop in affinity is consistent with expectations (37), and reflects the poor ability for Myc/Myc homodimers to efficiently bind DNA.

These results show that TIRF-PBM accurately captures affinity constants for the human TFs STAT1 and Myc/Max, and offers a route to explore binding differences over a wide range of affinity. The range of observed affinities between highly preferred and un-preferred targets reflects the selectivity of the protein, and we find almost three orders of magnitude change in affinity for these human TFs (Table 1, range of K_D^{app} observed). This reflects the transition from weak or non-specific TF:DNA interactions to highly specific, high-affinity binding, an important range to capture, as recent reports suggest that both types of interactions may play a role *in vivo* (4).

STAT1-binding preference is modulated by polymerization

In the cell, STAT1 is in a constant flux between phosphorylated and unphosphorylated states, with significant populations of both states co-existing (32). As U-STAT1 and P-STAT1 simultaneously exist *in vivo*, we investigated the effect of dimerization and higher-order polymerization on STAT1 DNA binding.

First, we considered that U-STAT1 and P-STAT1 may interact in the cell, giving rise to unique specificity. Estimates have placed the total concentration of STAT1 in the cell at ~50 nM, and so to ensure an observable effect we tested the binding of 25 nM P-STAT1 in the presence or absence of a 10-fold molar excess of U-STAT1. The STAT1 proteins were mixed and incubated for 1 hour, to ensure that if dimer dissociation was the limiting step, ample time to dissociate and re-associate as hemi-dimers of P-STAT1 and U-STAT1 was possible. However, even under these permissive conditions, we observed unchanged P-STAT1-binding specificity and affinity (representative measurements in Figure 1c), suggesting that STAT1 does not form hemi-dimers composed of one monomer of U-STAT1 and one monomer of P-STAT1.

Next, we tested an alternative model that homo-dimers of P-STAT1 and homo-dimers of U-STAT1 polymerize into tetramers and higher-order polymers [which have been observed for P-STAT1 alone (32)], and this polymerization may change P-STAT1-binding preference. To address this possibility, we performed an experiment where we first allowed 50 nM P-STAT1 to bind to an array and reach equilibrium. This was followed by 10-fold excess of U-STAT1, and the binding of U-STAT1 to the pre-bound P-STAT1:DNA array was monitored. However, under these conditions we observed no U-STAT1 binding to the array (data not shown). As we observed minimal binding of U-STAT1 alone, even at 400 nM (Figure 1), this result shows that P-STAT1 does not have sufficiently strong interactions with U-STAT1 dimers to cause mixed polymerization,

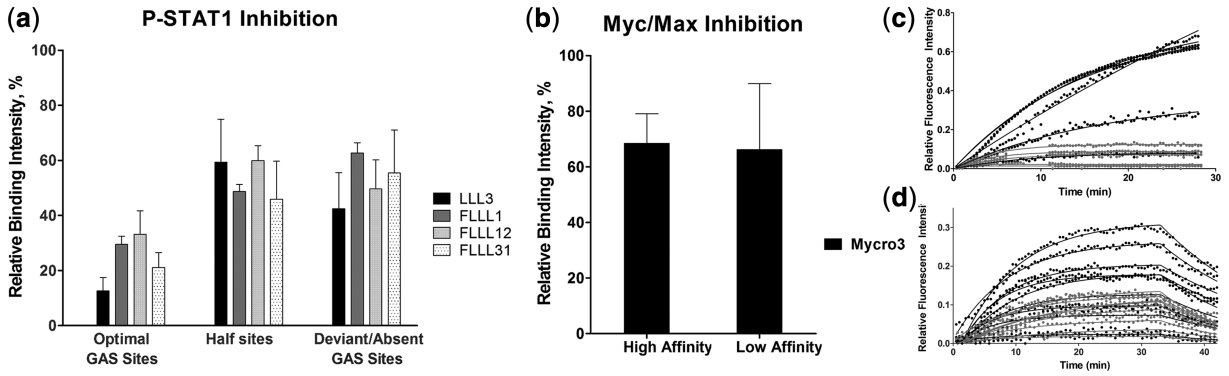


Figure 2. The inhibition of STAT1–DNA complexes but not Myc/Max–DNA complexes show sequence-dependent effects. (**a**) Inhibitor compound LLL3 and analogs were incubated with P-STAT1 and affect equilibrium binding intensity of P-STAT1 to DNA in a sequence-dependent manner (20 μM compound and 50 nM P-STAT1). IC_{50} for LLL3 was shown to vary in a sequence-dependent manner (Supplementary Figure S2) and ranges from ~1 μM to >60 μM. ***P-value < 0.001 for optimal GAS site affect to half GAS sites or deviant GAS sites by Student's t-test.(**b**) Inhibitor compound Mycro3 was incubated with Myc/Max and affects equilibrium binding intensity in a sequence-independent manner. (**c**) and (**d**) Representative data traces showing that the addition of the protein:protein inhibitors cause decreased TF binding.

consistent with the understanding that STAT1 polymerization is driven by N-terminal domain contacts, which are occluded in U-STAT1 anti-parallel dimers. These findings support a model for STAT1 behavior where P-STAT1 and U-STAT1 do not significantly interact.

However, tetramerization and higher-order polymerization of P-STAT1 may still have important effects on sequence preference. P-STAT1tc is missing the N-terminal domain predicted to be utilized in these polymerization processes (32), providing an important means by which we can investigate the role of higher-order polymerization. We find that P-STAT1tc has reduced binding compared to P-STAT1 at preferred binding sites, with an average 2.6-fold impact on binding, and thus the truncated regions appear to impart a measure of affinity and specificity to the binding, possibly through facilitating polymerization. The mechanism for the affinity imported by this presumptive polymerization remains unclear at present, but intriguingly, we find that P-STAT1tc binds at a number of sites with low sequence homology to the GAS consensus, suggesting more complex roles for polymerization.

Small molecule inhibitors of STAT1 and Myc/Max display sequence-specific effects. P-STAT1 at a concentration of 50 nM was tested with 20 μM concentrations of the compound LLL3 and several LLL3 analogs (FLLL1, FLLL12 and FLLL31) (21) (Figure 2a, c). LLL3 has been previously characterized as a STAT3 dimerization inhibitor (21), but the high conservation among proteins in the STAT family (42) strongly suggested the possibility of an effect on STAT1. LLL3 had the greatest inhibitory effect on 'High Affinity' sequence binding of the analogs tested, although all reduce equilibrium binding intensity by a significant extent. Across all 'High Affinity' sites, LLL3 shows an average adjusted binding intensity of 12.6 ± 4.9% of the no-inhibitor control, nearly an 8-fold reduction in binding. In comparison, the analogs display an average adjusted binding intensity of 27.9 ± 5.6%. In previous work with STAT3, LLL3 was shown to reduce DNA binding to ~40% of normal activity at this concentration, but those assessments were performed in cells and *in vivo* (21), so the higher activity may be attributed to the presence of other factors that could partially rescue the disrupted interaction.

The STAT dimerization interface contributes to the formation of the DNA-binding interface (36), and we investigated whether STAT1 dimerization inhibition would show differential effects against weaker affinity sites as well as high affinity, preferred binding targets. Unexpectedly, LLL3 and its analogs display markedly less perturbation of the low-affinity interactions (Figure 2a). At 'Half Sites', LLL3 showed 59.4 ± 15.5% adjusted binding intensity, while the analogs tested showed 51.5 ± 7.2% intensity. LLL3 thus shows a 4.7-fold greater reduction in binding at 'High Affinity' sites compared to the perturbation of binding at lower affinity 'Half Sites', with the LLL3 analogs following a similar pattern. 'Poor Affinity' sites, similarly, showed 42.4 ± 13.1% intensity when tested with LLL3, and 55.9 ± 10% average adjusted intensity with the other LLL3 analogs. The difference in inhibitory effect between 'High Affinity' sites and 'Half Sites' or 'Poor Affinity' is statistically significant, with a P-value $< 1.0 \times 10^{-3}$, suggesting that LLL3's inhibition may function in a sequence-dependent manner. It is possible that the STAT1 dimer binds to 'high affinity' sites in a specific conformation of the DNA-binding domain and the SH2 dimerization domain that is readily disrupted, while lower affinity interactions with DNA primarily depend on weakly specific interactions that are not as sensitive to protein conformational changes. To explore this behavior, LLL3 inhibition was tested from 5 μM to 200 μM, allowing generation of IC_{50} dose–response curves for the binding of P-STAT1 at each sequence (Supplementary Figure S2). Each sequence showed a unique IC_{50}, ranging for 1 μM to >60 μM, supporting the hypothesis that LLL3 disruption of STAT1 DNA binding occurs in a sequence-dependent fashion.

It was not apparent if this sequence-specific inhibition could be a general feature of dimeric TF interactions or

specific to the protein, so we pursued Myc/Max inhibition studies using Mycro3, a compound previously characterized as a selective inhibitor of Myc/Max dimerization (43). 50 μM Mycro3 [comparable to the reported IC_{50} of 40 ± 13 μM obtained with a cognate-binding site (43)] incubated with 10 nM Myc/Max showed a reduction in binding across the array (Figure 2b). Mycro3 caused an observed change in Myc/Max binding intensity at 'high affinity' sites from 100% to 68.5 ± 10.6% signal, and a statistically indistinguishable change to 66.2 ± 23.7% binding signal at weaker affinity sites. This decrease in Myc/Max binding also affected the kinetics of the reaction (Figure 2d), and increased the K_D observed for cognate site binding from 0.9 ± 0.2 nM to 3.6 ± 1.4 nM. These results establish that Mycro3′s affect on Myc/Max occurs in a sequence-independent fashion, consistent with its proposed disruption of the formation of the binding-competent heterodimer complex (44), and suggests that the interactions of Myc or Max with DNA do not significantly alter this inhibition. As this Myc/Max compound study did not reveal sequence-dependent inhibition, the sequence-specific inhibition observed with STAT1 and LLL3 appears to be specific to the TFs and compounds tested.

DISCUSSION

This TIRF-PBM study of the full length, dimeric, human TF STAT1 demonstrates the substantial variation in DNA sequence preference caused by STAT1 phosphorylation, multimerization capability, and in the presence of small molecule inhibitors of dimerization. TFs in 'active' states (i.e. the tyrosine-phosophorylated homodimer of STAT1 or the heterodimeric complex of Myc and Max) display binding specificity and affinity in agreement with prior work, although neither the binding motif in JASPAR nor our single-protein TIRF-PBM studies recapitulate the diversity of STAT1-binding sites seen in cell-based studies (29), a general trend in TF studies (4,45–47).

Probing less active states (i.e. the unphosphorylated form of STAT1 or Myc protein alone) shows that DNA-binding affinity is reduced 100- or 1000-fold with non-cognate targets, yet reveals that there is distinct binding preference even among these poorly bound sites. These observations challenge a simplistic model where individual STAT1 dimers can efficiently bind at preferred binding sites and that STAT1 polymerization serves only to increase binding at poor binding sites. The mechanisms by which higher-order polymerization orients and directs STAT1 binding remain unclear, and the biological role of this polymerization could be reflected in the ability to utilize non-ideal binding sites topologically near other sites, a possibility not addressed in this work. However, our results suggest that the N-terminal region may play an important role in regulating site specificity, and future work should investigate this possibility with studies of polymerized state in relation to binding.

More broadly, these results agree with other PBM efforts (4) that illustrate that protein:DNA interactions cannot be accurately viewed as a single, reductionist consensus site. Rather, protein binding must be viewed as a continuous scale, with motifs of high affinity (or even multiple, distinct motifs of high affinity) present among a sea of less preferred, but still possibly functional, motifs. In notable agreement, a recent wide-ranging report on the motifs for a broad set of DNA-binding domains from mouse TFs found numerous high, moderate and low-affinity sites for each TF tested (4). Communicating this wide specificity in a convenient form is an ongoing challenge; position weight matrixes generate logograms that help show the range of possible binding sites (28,48), but still fail to accurately convey information about TFs that bind to multiple distinct motifs.

The expansion of TIRF-PBM methods into the screening of small molecule inhibitors of protein:protein interaction (LLL3 for STAT1 and Mycro3 for Myc/Max) revealed the unexpected, but perhaps desirable, result that LLL3 appears to disrupt the function of STAT1 in a DNA sequence-dependent manner. As aberrant STAT activity plays a key role in many cancers (19), yet also plays a centrally important constitutive regulatory role (36), finding sequence-selective inhibitors such as LLL3 may aid in the design of chemotherapeutics with reduced off-target effects. This highlights the proposed importance of protein:protein inhibitors as therapeutic targets (16,49), and suggests that the screening of such compounds to reveal sequence-dependent variations may be of value (50).

This study supports the view that STAT1 DNA-binding affinity and specificity are modulated by the dimeric conformation of the TF, and that disruption of normal dimer interfaces by small molecule inhibition or truncation of the N-terminal domain of STAT1 will result in altered, but not abolished, DNA binding. These observations are consistent with recent models of STAT biology (14,32), and with recent findings that suggest both preferred and non-preferred binding sites are important for the majority of TFs (4,7). The modulation of the functional occupancy of these sites *in vivo* is likely to rely on the effect of various co-activators, competing TFs and the basal transcriptional machinery (13,51). As TIRF-PBM has the ability to resolve multiple components of protein complexes that bind DNA (13), there is great potential for future studies of STAT1 in concert with co-activators that may reveal further functional consequence and provide a bridge between *in vitro* and *in vivo* characterization.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Table 1 and Supplementary Figures 1 and 2.

ACKNOWLEDGEMENTS

The authors wish to thank and acknowledge Ernest Martinez (UC Riverside) for his gift of Myc and Max expression constructs, Thorsten Berg (Max Planck Inst. of Biochem.) for Myc/Max inhibitor compounds, and Tom Li (Ohio State Univ.) for STAT inhibitor

compounds. We also thank Thorsten Neumann (UCSF) for TIRF technical expertise and Teisha Rowland (UCSB) for manuscript proofing.

FUNDING

Funding for open access charge: DoD grant [W81XWH-09-1-0698].

Conflict of interest statement. None declared.

REFERENCES

1. Wray,G.A., Hahn,M.W., Abouheif,E., Balhoff,J.P., Pizer,M., Rockman,M.V. and Romano,L.A. (2003) The evolution of transcriptional regulation in eukaryotes. *Mol. Biol. Evol.*, **20**, 1377–1419.
2. Carninci,P., Sandelin,A., Lenhard,B., Katayama,S., Shimokawa,K., Ponjavic,J., Semple,C.A.M., Taylor,M.S., Engstrom,P.G., Frith,M.C. *et al.* (2006) Genome-wide analysis of mammalian promoter architecture and evolution. *Nat. Genet.*, **38**, 626–635.
3. Levy,D.E. and Darnell,J.E. (2002) STATs: transcriptional control and biological impact. *Nat. Rev. Mol. Cell Biol.*, **3**, 651–662.
4. Jaeger,S.A., Chan,E.T., Berger,M.F., Stottmann,R., Hughes,T.R. and Bulyk,M.L. (2010) Conservation and regulatory associations of a wide affinity range of mouse transcription factor binding sites. *Genomics.*, **94**, 185–195.
5. Horak,C.E. and Snyder,M. (2002) ChIP-chip: a genomic approach for identifying transcription factor binding sites. *Guide Yeast Genet. Mol. Cell Biol.*, Pt B, 350, 469–483.
6. Visel,A., Blow,M.J., Li,Z., Zhang,T., Akiyama,J.A., Holt,A., Plajzer-Frick,I., Shoukry,M., Wright,C., Chen,F. *et al.* (2009) ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature*, **457**, 854–858.
7. Zhu,C., Byers,K., McCord,R., Shi,Z., Berger,M., Newburger,D., Saulrieta,K., Smith,Z., Shah,M., Radhakrishnan,M. *et al.* (2009) High-resolution DNA binding specificity analysis of yeast transcription factors. *Genome Res.*, **19**, 556–566.
8. Mukherjee,S., Berger,M.F., Jona,G., Wang,X.S., Muzzey,D., Snyder,M., Young,R.A. and Bulyk,M.L. (2004) Rapid analysis of the DNA-binding specificities of transcription factors with DNA microarrays. *Nat. Genet.*, **36**, 1331–1339.
9. Dean,A. (2006) On a chromosome far, far away: LCRs and gene expression. *Trends Genet.*, **22**, 38–45.
10. Hahn,S. (2004) Structure and mechanism of the RNA polymerase II transcription machinery. *Nat. Struct. Mol. Biol.*, **11**, 394–403.
11. Martinez,E. (2002) Multi-protein complexes in eukaryotic gene transcription. *Plant Mol. Biol.*, **50**, 925–947.
12. Panne,D. (2008) The enhanceosome. *Curr. Opin. Struct. Biol.*, **18**, 236–242.
13. Bonham,A.J., Neumann,T., Tirrell,M. and Reich,N.O. (2009) Tracking transcription factor complexes on DNA using total internal reflectance fluorescence protein binding microarrays. *Nucleic Acids Res.*, **37**, e94.
14. Sehgal,P.B. (2008) Paradigm shifts in the cell biology of STAT signaling. *Semin. Cell Dev. Biol.*, **19**, 329–340.
15. Vinkemeier,U., Cohen,S.L., Moarefi,I., Chait,B.T., Kuriyan,J. and Darnell,J.E. Jr (1996) DNA binding of in vitro activated Stat1 alpha, Stat1 beta and truncated Stat1: interaction between NH2-terminal domains stabilizes binding of two dimers to tandem DNA sites. *EMBO J.*, **15**, 5616–5626.
16. Darnell,J.E. (2002) Transcription factors as targets for cancer therapy. *Nat. Rev. Cancer*, **2**, 740–749.
17. Yu,H. and Jove,R. (2004) The stats of cancer – New molecular targets come of age. *Nat. Rev. Cancer*, **4**, 97–105.
18. Kovacic,B., Stoiber,D., Moriggl,R., Weisz,E., Ott,R.G., Kreibich,R., Levy,D.E., Beug,H., Freissmuth,M. and Sexl,V. (2006) STAT1 acts as a tumor promoter for leukemia development. *Cancer Cell*, **10**, 77–87.
19. Bowman,T., Garcia,R., Turkson,J. and Jove,R. (2000) STATs in oncogenesis. *Oncogene*, **19**, 2474–2488.
20. Siddiqui-Jain,A., Grand,C.L., Bearss,D.J. and Hurley,L.H. (2002) Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription. *Proc. Natl Acad. Sci. USA*, **99**, 11593–11598.
21. Fuh,B., Sobo,M., Cen,L., Josiah,D., Hutzen,B., Cisek,K., Bhasin,D., Regan,N., Lin,L., Chan,C. *et al.* (2009) LLL-3 inhibits STAT3 activity, suppresses glioblastoma cell growth and prolongs survival in a mouse glioblastoma model. *Br. J. Cancer*, **100**, 106–112.
22. Farina,A., Faiola,F. and Martinez,E. (2004) Reconstitution of an E box-binding Myc:Max complex with recombinant full-length proteins expressed in Escherichia coli. *Protein Expres. Purif.*, **34**, 215–222.
23. Solomon,D.L.C., Amati,B. and Land,H. (1993) Distinct DNA binding preferences for the c-Myc/Max and Max/Max dimers. *Nucleic Acids Res.*, **21**, 5372–5376.
24. Ehret,G.B., Reichenbach,P., Schindler,U., Horvath,C.M., Fritz,S., Nabholz,M. and Bucher,P. (2001) DNA binding specificity of different STAT proteins. *J. Biol. Chem.*, **276**, 6675–6688.
25. Kraus,J., Borner,C. and Hollt,V. (2003) Distinct palindromic extensions of the 5′-TTC...GAA-3′ motif allow STAT6 binding in vivo. *FASEB J.*, **17**, 304–306.
26. Fernandez,P.C., Frank,S.R., Wang,L., Schroeder,M., Liu,S., Greene,J., Cocito,A. and Amati,B. (2003) Genomic targets of the human c-Myc protein. *Genes & Development*, **17**, 1115–1129.
27. Haggerty,T.J., Zeller,K.I., Osthus,R.C., Wonsey,D.R. and Dang,C.V. (2003) A strategy for identifying transcription factor binding sites reveals two classes of genomic c-Myc target sites. *Proc. Natl Acad. Sci. USA*, **100**, 5313–5318.
28. Bryne,J.C., Valen,E., Tang,M.H.E., Marstrand,T., Winther,O., da Piedade,I., Krogh,A., Lenhard,B. and Sandelin,A. (2008) JASPAR, the open access database of transcription factor-binding profiles: new content and tools in the 2008 update. *Nucleic Acids Res.*, **36**, D102–D106.
29. Robertson,G., Hirst,M., Bainbridge,M., Bilenky,M., Zhao,Y., Zeng,T., Euskirchen,G., Bernier,B., Varhol,R., Delaney,A. *et al.* (2007) Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. *Nat. Meth.*, **4**, 651–657.
30. Bonham,A.J., Braun,G., Pavel,I., Moskovits,M. and Reich,N.O. (2007) Detection of sequence-specific protein-DNA interactions via surface enhanced resonance Raman scattering. *J. Am. Chem. Soc.*, **129**, 14572–14573.
31. Fisher,F., Crouch,D.H., Jayaraman,P.S., Clark,W., Gillespie,D.A. and Goding,C.R. (1993) Transcription activation by Myc and Max: flanking sequences target activation to a subset of CACGT G motifs in vivo. *EMBO J.*, **12**, 5075–5082.
32. Wenta,N., Strauss,H., Meyer,S. and Vinkemeier,U. (2008) Tyrosine phosphorylation regulates the partitioning of STAT1 between different dimer conformations. *Proc. Natl Acad. Sci. USA*, **105**, 9238–9243.
33. Wesoly,J., Szweykowska-Kulinska,Z. and Bluyssen,H.A. (2007) STAT activation and differential complex formation dictate selectivity of interferon responses. *Acta Biochim. Pol.*, **54**, 27–38.
34. Seidel,H.M., Milocco,L.H., Lamb,P., Darnell,J.E. Jr, Stein,R.B. and Rosen,J. (1995) Spacing of palindromic half sites as a determinant of selective STAT (signal transducers and activators of transcription) DNA binding and transcriptional activity. *Proc. Natl Acad. Sci. USA*, **92**, 3041–3045.
35. Simonsson,T. (2001) G-quadruplex DNA structures–variations on a theme. *Biol. Chem.*, **382**, 621–628.
36. Hoey,T. and Schindler,U. (1998) STAT structure and function in signaling. *Curr. Opin. Genet. Dev.*, **8**, 582–587.
37. Jung,K.C., Rhee,H.S., Park,C.H. and Yang,C.H. (2005) Determination of the dissociation constants for recombinant c-Myc, Max, and DNA complexes: the inhibitory effect of linoleic acid on the DNA-binding step. *Biochem. Biophys. Res. Commun.*, **334**, 269–275.
38. Wingender,E. (2008) The TRANSFAC project as an example of framework technology that supports the analysis of genomic regulation. *Brief. Bioinform.*, **9**, 326–332.
39. Farina,A., Faiola,F. and Martinez,E. (2004) Reconstitution of an E box-binding Myc:Max complex with recombinant full-length

proteins expressed in Escherichia coli. *Protein Expres. Purif.*, **34**, 215–222.
40. Chen,X., Xu,H., Yuan,P., Fang,F., Huss,M., Vega,V.B., Wong,E., Orlov,Y.L., Zhang,W., Jiang,J. *et al.* (2008) Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell*, **133**, 1106–1117.
41. Chatterjee-Kishore,M., Wright,K.L., Ting,J.P.Y. and Stark,G.R. (2000) How Stat1 mediates constitutive gene expression: a complex of unphosphorylated Stat1 and IRF1 supports transcription of the LMP2 gene. *EMBO J.*, **19**, 4111–4122.
42. Seidel,H.M., Milocco,L.H., Lamb,P., Darnell,J.E. Jr, Stein,R.B. and Rosen,J. (1995) Spacing of palindromic half sites as a determinant of selective STAT (signal transducers and activators of transcription) DNA binding and transcriptional activity. *Proc. Natl Acad. Sci. USA*, **92**, 3041–3045.
43. Kiessling,A., Sperl,B., Hollis,A., Eick,D. and Berg,T. (2006) Selective inhibition of c-Myc/Max dimerization and DNA binding by small molecules. *Chem. Biol.*, **13**, 745–751.
44. Berg,T., Cohen,S.B., Desharnais,J., Sonderegger,C., Maslyar,D.J., Goldberg,J., Boger,D.L. and Vogt,P.K. (2002) Small-molecule antagonists of Myc/Max dimerization inhibit Myc-induced transformation of chicken embryo fibroblasts. *Proc. Natl Acad. Sci. USA*, **99**, 3830–3835.
45. Farnham,P.J. (2009) Insights from genomic profiling of transcription factors. *Nat. Rev. Genet.*, **10**, 605–616.
46. Sandelin,A., Carninci,P., Lenhard,B., Ponjavic,J., Hayashizaki,Y. and Hume,D.A. (2007) Mammalian RNA polymerase II core promoters: insights from genome-wide studies. *Nat. Rev. Genet.*, **8**, 424–436.
47. Mccord,R.P., Berger,M.F., Philippakis,A.A. and Bulyk,M.L. (2007) Inferring condition-specific transcription factor function from DNA binding and gene expression data. *Mol. Syst. Biol.*, **3**, 100.
48. Bailey,T.L., Williams,N., Misleh,C. and Li,W.W. (2006) MEME: discovering and analyzing DNA and protein sequence motifs. *Nucleic Acids Res.*, **34**, 369–373.
49. Redell,M.S. and Tweardy,D.J. (2005) Targeting transcription factors for cancer therapy. *Curr. Pharm. Design*, **11**, 2873–2887.
50. Li,X.Q., Jiang,X. and Yaoi,T. (2006) High throughput assays for analyzing transcription factors. *Assay Drug Dev. Techn.*, **4**, 333–341.
51. Naar,A.M., Lemon,B.D. and Tjian,R. (2001) Transcriptional coactivator complexes. *Ann. Rev. Biochem.*, **70**, 475–501.