Journal of
Proteomics & Bioinformatics

Trung et al., J Proteomics Bioinform 2014, 7:3
DOI: 10.4172/0974-276X.1000306

Research Article Open Access

SILAC-Based Quantitative Proteomics Approach to Identify Transcription Factors Interacting with a Novel *Cis*-Regulatory Element

Ngo Tat Trung[1,2], Rudolf Engelke[1,3] and Gerhard Mittler[1]*

[1]*Proteomics and Mass Spectrometry, Max Planck Institute of Immunobiology and Epignetics, Freiburg, Germany*
[2]*Department of Molecular Biology, Tran Hung Dao University Hospital, Hanoi, Vietnam*
[3]*Weill Cornell Medical College in Qatar, Doha, State of Qatar*

Abstract

The motif TMTCGCGANR (M being C or A, R being A or G, and N any nucleotide) called M8 was discovered as a putative *cis*-regulatory element present in 368 human gene promoters. Of these, 236 (64%) are conserved within promoter sequences of four related organisms: human, mouse, rat and dog. However, transcription factors (TFs) interacting with the M8 motif have not yet been described. We previously reported the use of quantitative proteomics coupled to one-step DNA affinity purification as a means of screening for TFs associated with given functional DNA elements. The procedure is performed in-vitro employing SILAC-labeled nuclear extracts and making use of well-characterized *cis*-regulatory motifs. Building on that, in this study we have combined our method with statistical analysis to filter out false positive hits from the one-step DNA affinity pull-down experiments. This resulted in the identification of zinc finger BED domain-containing protein 1 (ZBED1), alpha globin transcription factor CP2 (TFCP2), upstream binding protein 1 (UBP) and transcription factor CP2 like 1(TFCP2L1), as specific M8 interacting factors. We validated our screen demonstrating the *in vivo* binding of alpha globin transcription factor TFCP2 to selected genes harboring M8-containing promoters using ChIP (chromatin immuno-precipitation) assays. This not only implicates a functional role of the above proteins in regulating M8 motif containing genes, but also suggests the potential use of our approach to decipher protein-DNA interactions occurring in living cells.

Keywords: M8 motif; SILAC; Quanitative proteomics

Introduction

Since the sequencing of the human genome is almost fully sequenced allowing the computational prediction of genes and their open reading frames, the progress in identifying the gene regulatory *cis*-elements and their associated transcription factors is lagging behind. These elements are known to spatially and temporally control several cellular processes, including cellular differentiation programs, responses to environmental cues and initiation and progression of malignancies [1].

Theoretically, the transcription factor occupancy of gene regulatory elements can be determined by a computationally-based prediction approach. This method harnesses prior knowledge of all characterized DNA binding sites for a known TF (TF binding site position weight matrix) to seek a matching pattern within the genome, resulting in a consensus matrix of acceptable nucleotide strings at various genomics positions [2]. Unfortunately, the aforementioned strategy has several shortcomings. First, a subset of real binding sites is not found (false negatives) due to limitations in the model that do not account for the interdependency of nucleotide changes at two or more positions of a motif. Similarly, TF binding events can be partially or fully (piggy-back mechanism) dependent on the concurrent binding of other TFs [3]. Second, current models are unable to assess DNA context-dependent affinity variation of motifs (leading to false positive and negative predictions), which can be substantial [4]. Finally, *in silico* approaches cannot judge which TF protein isoforms bind to a given *cis*-element [5-7]. Strikingly, the situation is further complicated by the fact that certain TFs are able to specifically interact with two totally different DNA sequence motifs [8,9]. An alternative, more unbiased computational strategy, termed "phylogenetic footprinting" excels in determining putative TF binding motifs without requiring a position weight matrix (PWM). This is done by comparing the DNA flanking loci of orthologous genes with similar functions from related organisms. Recently, genomic foot-printing analysis of human gene orthologues became feasible as whole genome sequences of many mammalian and other vertebrate organisms are now readily available. Therefore, by enlarging or narrowing the complexity and number of input genomes, sub-clusters of putative *cis*-regulatory motifs can be revealed.

In this context, the study of Eric Lander and colleagues from 2005 demarcates a milestone in the field [10]. They were the first team to create a systematic catalogue of short, common regulatory motifs, typically 6–12 bases long, present in a window comprising less than 2500 bps upstream and 500 bps downstream of the transcription start sites (TSSs) of annotated genes. These putative *cis*-elements were further ranked (M1 to M176) by virtue of their evolutionary conservation across related human, mouse, rat, and dog genomes [10]. This conserved DNA motif catalogue is considered to contain biological function as many of the identified motifs are known to be bound by given transcription factors (e.g. Sp1, Lef-1, Myc etc.). However, one sequence of the highly-ranked motifs called M8 TMTCGCGANR (M being C or A, R being A or G, and N any nucleotide) has not yet been described as a regulatory element, despite the fact that more than 300 human gene promoters harbor this motif. This implies that the M8 *cis*-element might be recognized by one or several transcription factors.

***Corresponding author:** Gerhard Mittler, Proteomics and Mass Spectrometry, Max Planck Institute of Immunobiology and Epignetics, Freiburg, Germany, Tel: +49 761 5108 712; Fax +49 761 5108 799; E-mail: mittler@ie-freiburg.mpg.de

Received October 01, 2013; **Accepted** March 17, 2014; **Published** March 20, 2014

Citation: Trung NT, Engelke R, Mittler G (2014) SILAC-Based Quantitative Proteomics Approach to Identify Transcription Factors Interacting with a Novel *Cis*-Regulatory Element. J Proteomics Bioinform 7: 082-087. doi:10.4172/0974-276X.1000306

Furthermore, the data of Xie et al. [10] suggest that genes harboring M8 exhibit increased expression levels in some leukemia or lymphoma cells, indicating a potential role of M8 motif for hematopoietic cells [10].

To identify transcription factors that interact with the M8 motif, one could perform classical biochemical fractionation coupled with electrophoretic mobility shift assays (EMSA), or follow the proteomics of the isolated chromatin segment (PiCh) protocol [11]. EMSA-based approach is slow, cumbersome and requires large amounts of nuclear extracts thus making it impractical to pursue on the proteomics scale. Likewise, PiCh can only be used for highly repetitive elements and is still challenged by the need for an enormous number of cells (~10^{11}) that are required to provide sufficient amounts of protein-DNA complexes for mass spectrometry analysis.

Recently, we described the powerful combination of quantitative proteomics employing *S*table *I*sotope *L*abeling by *A*mino acid in Cell culture (SILAC) [12] with one-step DNA affinity pull-down experiments in order to confirm the binding of well characterized transcription factors to their cognate regulatory sequences *in-vitro* [3].

In this paper, we report that biological replicates of the SILAC-based DNA affinity pull-down assay followed by statistical analysis significantly improve data quality. Consequently, we describe ZBED1 and three related but distinct proteins (TFCP2, TFCP2L1 and UBP1) as potential binding partners of the novel, highly conserved M8 motif. Notably, our chromatin immunoprecipitation data reveal the binding of TFCP2 to certain M8 motif bearing promoters *in vivo*. Therefore this study may represent a basis for functional follow-up studies in order to dissect the role of proposed M8 binders in regulating M8 bearing genes.

Materials and Methods

M8 motif bearing DNA ligand preparation

100 nM sense 5′-biotinylated oligonucleotides (synthesized by Eurofins MWG Operon) bearing either wild-type or mutant M8 motif was combined with respective 110 nM anti-sense oligonucleotide in 200 µl 20 mM Tris-HCl, pH 8.0, 200 mM NaCl, 0.5 mM EDTA, 0.03% NP40. The reaction mixture was heated at 95°C for 5 minutes and then cooled down in a stepwise manner to 65°C over 15 minutes, kept at 65°C for 5 minutes followed by cooling to room temperature (RT). At this point, the annealed double stranded oligonucleotide can either be stored at -20°C for long term use, or used in the following step.

Immobilization of biotinylated oligonucleotides to the beads

100 µl suspension of Dyanabeads® Myone Streptavidin beads (Dynal Biotech ASA, Oslo, Norway) was combined with 13 µl annealed M8 motif bearing DNA ligand in 400 µl 20 mM Tris-HCl, pH 8.0, 200 mM NaCl, 0.5 mM EDTA, 0.03% NP40 and incubated for 3 hours at RT followed by incubation at 4°C overnight. Afterwards, the beads were washed by three volumes of 800 µl 20 mM Tris-HCl, pH 8.0, 200 mM NaCl, 0.5 mM EDTA, 0.03% NP40 buffer, and re-suspended in 100 µl of the same buffer. At this point, the concentration of DNA ligand on bead suspension was approximately 2.73 nM and beads could be stored at 4°C for later use.

One step-DNA affinity chromatography

The human T lymphoblastic leukemia cell line Molt 4 was metabolically labeled in a RPMI medium supplemented with either light or heavy isotopes of two essential amino acids (^{1}H4 lysine and ^{12}C6 arginine) or (^{2}H4 lysine and ^{13}C6 arginine) as described before [12]. After more than five doubling cycles, the cells show an isotope incorporation of more than 98%. Two separate populations of isotope-encoded cells were equally expanded to 5 liters of 1×10^{6} cells/ml and used for nuclear extract preparation, essentially as described in Dignam et al. [13]. The nuclear extract was then dialyzed at 4°C against a buffer (20 mM Tris-HCl, pH 7.3 containing 100 mM KCl, 20% glycerol, 0.5 mM fresh phenylmethylsulfonyl fluoride (PMSF), 1 mM fresh DTT, 0.1 mM sodium pyrophosphate, 0.1 mM sodium orthovanadate and 0.1 mM sodium fluoride) and frozen in liquid nitrogen for further use.

Three volumes of Molt4 nuclear extract from 1.5×10^{8} cells were then incubated in three vials with 40 nM immobilized M8 motif DNA, respectively. This was performed in a physiological buffer (20 mM Tris-HCl, pH 7.3 containing 100 mM KCl, 10 mM potassium glutamate, 10% glycerol, 0.5 mM fresh phenylmethylsulfonyl fluoride (PMSF), 1 mM fresh DTT, 0.1 mM sodium pyrophosphate, 0.1 mM sodium orthovanadate, 0.1 mM sodium fluoride, 0.2 mM fresh benzamidine, 1 µg/ml fresh leupeptin and 1 ug/ml fresh aprotinin) for 4 hrs on a rotating wheel at 4°C. After washing away the unbound material, protein-DNA complexes were finally liberated by PstI endonuclease restriction digestion.

Liquid chromatography mass spectrometry

Proteins eluted from the DNA affinity columns were resolved through 4-12% Bis-Tris SDS gels (NuPAGE, Invitrogen) and stained with colloidal Coomassie. The gels were sliced into 10 equally sized gel pieces and subjected to tryptic in-gel digestion [14]. Prior to LC-MS analysis, tryptic peptide mixtures were desalted using STAGE tips (Empore high performance extraction disks C18- IVA Analysentechnik e. K Meerbusch, Germany) as described previously [15] and the pooled eluates from the C18 tips were subsequently analyzed by nanoscale-LC (Agilent 1200 nanoflow system), coupled to a Hybrid LTQ OrbitrapXL+ETD mass spectrometer (Thermo Fisher Scientific Inc. Bremen, Germany).

Peptides were eluted from an analytical column by a 120 min linear gradient running from 5 to 80% (v/v) acetonitrile (in 0.5% acetic acid) with a flow rate of 250 nl/min and sprayed directly into the orifice of the mass spectrometer. The 15 cm fused silica emitter with an inner diameter of 75 µm (New Objective, USA) was packed in-house with reverse phase ReproSil-Pur C18-AQ 3 µm resin (Dr. Maisch GmbH, Germany). Data dependent acquisition of MS, MS/MS was performed: Fullscan MS spectra (m/z 350 – 1600) were acquired in the Orbitrap detector with resolution R = 60,000 at m/z = 400 with a target value of 1,000,000 ions allowing a maximum injection-time of 1200 ms. The five most intense ions were sequentially isolated with an isolation width of 2.0 and fragmented in the linear ion trap by collision induced dissociation (CID) at a target value of 10,000 (maximum fill-time = 200 ms). Target ions selected for MS/MS were dynamically excluded for 90 s. The general mass spectrometric conditions were: spray voltage, 2.3 kV; no sheath and auxiliary gas flow; capillary temperature, 160°C; normalized collision energy 35.0, ion selection thresholds were 500 counts for MS/MS acquisition, applying an activation q = 0.25 and an activation time of 30 ms.

Data analysis

The raw Orbitrap MS files were processed with MaxQuant v1.0.12.31 [16,17] a powerful computational platform for SILAC-

based quantitative proteomics, using the following settings. Enzyme specificity was set to trypsin, allowing two missed cleavages, cleavages N-terminal to proline residues, and between aspartic acid and proline. Carbamidomethylation of cysteine were set as a fixed modification, while methionine oxidation, protein N-acetylation, and loss of ammonia from glutamine and asparagine were set at variable modifications. The peak lists generated by MaxQuant were searched with an in-house Mascot 2.2 server against the human international protein index (IPI) database (v. 3.64) containing frequently observed contaminants concatenated with a decoy of the reversed sequences. For monoisotopic precursor ions, the maximum allowed mass deviation was set to 5 ppm and for MS/MS peaks to 0.5 Da. The identified proteins and peptides were further processed with MaxQuant (imported as Mascot .dat files) with a minimum required peptide length of six amino acids and a false discovery rate setting of 0.01 at the protein level. For protein identification and quantification, two unique peptides and two quantified peptides were necessary. Before statistical analysis, known contaminants and reverse hits were removed. All proteins identified with $p<0.01$ in forward or reverse experiments are shown in Table 1. Statistical significance was obtained through Benjamini-Hochberg multiple-testing correction of the MaxQuant significance A and significance B values [18] for all experiments ($p(forw) \times p(rev) < 0.05$).

Chromatin immunoprecipitation

50 ml Molt4 cell aliquots of 5×10^5 cells/ml were treated with 1% formaldehyde (final concentration) at RT for 5 minutes and then with 10 mM disuccinimidyl glutarate (final concentration) for another 2 minutes at RT. The cross-linked cells were then washed 3 times with cold PBS and resuspended in 850 µl cold-L1 lysis buffer (50 mM Tris-HCl, pH 8.0, 2 mM EDTA, 0.1% NP40, 10% glycerol, 2 mM fresh DTT, 0.5 mM fresh phenylmethylsulfonyl fluoride (PMSF), 1 mM fresh DTT, 0.1 mM sodium pyrophosphate, 0.1 mM sodium orthovanadate, 0.1 mM sodium fluoride, 0.2 mM fresh benzamidine, 1 µg/ml fresh leupeptin, 1 µg/ml fresh aprotinin and 1 mM sodium butyrate) for 10 minutes on ice. The cells were centrifuged at 800xg for 5 minutes to clarify the cytoplasmic fraction (supernatant) and pellets (cross-linked nuclei) were re-suspended in 850 µl cold-L2 lysis buffer (50 mM Tris-HCl, pH 8.0, 2 mM EDTA, 1% SDS, 2 mM fresh DTT, 0.5 mM fresh phenylmethylsulfonyl fluoride (PMSF), 1 mM fresh DTT, 0.1 mM sodium pyrophosphate, 0.1 mM sodium orthovanadate, 0.1 mM sodium fluoride, 0.2 mM fresh benzamidine, 1 µg/ml fresh leupeptin, 1 µg/ml fresh aprotinin and 1 mM sodium butyrate). At this point, the cross-linked nuclei can be snap frozen in liquid nitrogen and stored at -80°C. Otherwise, the procedure can be continued by sonicating the cross-linked nuclei, shearing the chromatin into pieces with an average length of 300-1000 bp, which can be used as input for chromatin immunoprecipitation.

Sheared chromatin corresponding to one aliquot of cross-linked nuclei (2.5×10^6 cells) was then diluted 10 times into 1.5 ml DB buffer (50 mM Tris-HCl, pH 8, 5 mM EDTA, 200 mM NaCl, 0.5% NP40, 2 mM fresh DTT, 0.5 mM fresh phenylmethylsulfonyl fluoride (PMSF), 1 mM fresh DTT, 0.1 mM sodium pyrophosphate, 0.1 mM sodium orthovanadate, 0.1 mM sodium fluoride, 0.2 mM fresh benzamidine, 1 µg/ml fresh leupeptin, 1 µg/ml fresh aprotinin and 1 mM sodium butyrate) and incubated for 5 hrs with 5.0 µg rabbit anti-TFCP2 antibody (Abcam, Ab80445). The control assay was carried out at the same time using 5.0 µg rabbit pre-immune-serum and equal amounts of input material.

The DNA-protein complexes were precipitated by adding 30-40 µl salmon sperm DNA saturated protein G. The incubation time was extended for 45 minutes and the beads were then washed three times with a NaCl washing buffer (20 mM Tris-HCl, pH 8, 2 mM EDTA, 500 mM NaCl, 1% NP40, 1% SDS) and once with a LiCl washing buffer (20 mM Tris-HCl, pH 8, 2 mM EDTA, 500 mM LiCl, 1% NP40, 1% SDS). The protein G bound material was released by incubating with 2 volumes of 150 µl 20 mM Tris-HCl buffer, pH 8 containing 2% SDS and 0.5 mM EDTA for 5 minutes at 65°C. The released immunoprecipitated (IP) DNA was finally purified using a Qiagen PCR cleaning kit and stored in 250 µl 20 mM Tris-HCl, pH 8, 0.5 mM EDTA at -20°C for realtime qPCR analysis.

Real-time qPCR analysis

2 µl of IP or input DNA served as a template for real-time-qPCR reactions using sybr green and primer pairs as indicated in Table 1. Enrichment ratios of individual M8 motif bearing DNA promoter regions were calculated by normalizing the qPCR signal from IP DNA templates against the corresponding signal from the input genomic DNA.

Result and Discussion

We recently published the combination of SILAC coupled to liquid chromatography mass spectrometric analysis in screening for protein-DNA interactions using well-characterized TF binding sites [3]. Here, we apply this methodology to identify transcription factors that associate with a specific, *cis*-regulatory element that has not been previously characterized. We chose a ten nucleotide-length DNA motif (TMTCGCGANR, termed M8) as our current study model for the following reasons. First, the M8 motif was co-discovered in a bioinformatic genome-wide phylogenetic footprinting analysis in conjunction with other well-characterized TF binding sites [10], strongly arguing for its role as a *bona fide cis*-regulatory element. Second, the motif M8 occurs 368 times in human promoters, of which 236 (64%) are conserved across four related mammalian species (human, mouse, rat, and dog). Since the seminal study of Xie et al. [10] proposes a potential role of M8 motif bearing genes in human leukemia cells we used the T lymphoblastic cell line Molt4 throughout this study.

Primer name	Primer sequence (5 prime to 3 prime)	Primer position
RBP-Pro-F	TGGGGCGAAAGCTTAGGCAA	Flank M8 motif of RBP-j-kappa gene
RBP-Pro-R	CTCAGACACAACACGGCCGT	
3C-RBP-F	CCCGGGCGCTTTGTCTTTCA	Localize 1.8 kb downstream to the M8 motif of RBP-j-kappa gene
3C-RBP-R	CAAGGCGGCGAAACAAAGGG	
PURA-Pro-F	CCCTGTTACCGGGTCTCGTCTGTCT	Flank M8 motif of PurA gene
PURA-Pro-R	CACAGCCAGTCAGCCA CTCTCG	
PURA-exon-F	TGACAGTTTCTCTTCTTGACTTGC	Localize 1.3 kb downstream to the M8 motif of PurA gene
PurA-Exon-R	TGTTTTATATGGTGGATAAAGTGGAC	

Table 1: Primer pairs used for quantification of chromatin immunoprecipitated M8 motif bearing promoters.

M8 is bound by different TFs *in vitro*

To identify proteins specifically bound to the M8 motif we applied our previously published experimental workflow [3]. For designing the control DNA we aimed to mutate nucleotides in the highly conserved hexameric core of M8 without substantially changing its GC content but eradicating the palindromic nature of the motif. Heavy encoded nuclear extract (^{2}H4 lysine and ^{13}C6 arginine) from 1.5×10^{8} Molt4 was incubated with duplex DNA harboring M8 sequences whereas the mutated control DNA bait was incubated with a non-labeled counterpart (forward experiments). In reverse experiments the labeling was swapped meaning that the control DNA was mixed with the SILAC-labeled nuclear extract. In order to improve the amount of quantifiable peptides we used two labeling channels: ^{2}H4 lysine and ^{13}C6 arginine. By doing so, we were able to identify and quantify 710 proteins from two forward and two reverse pull-downs (Figures 1E and 1F). Most of these proteins possess SILAC ratios around one-to-one, meaning that they bind unspecifically to the DNA oligonucleotide

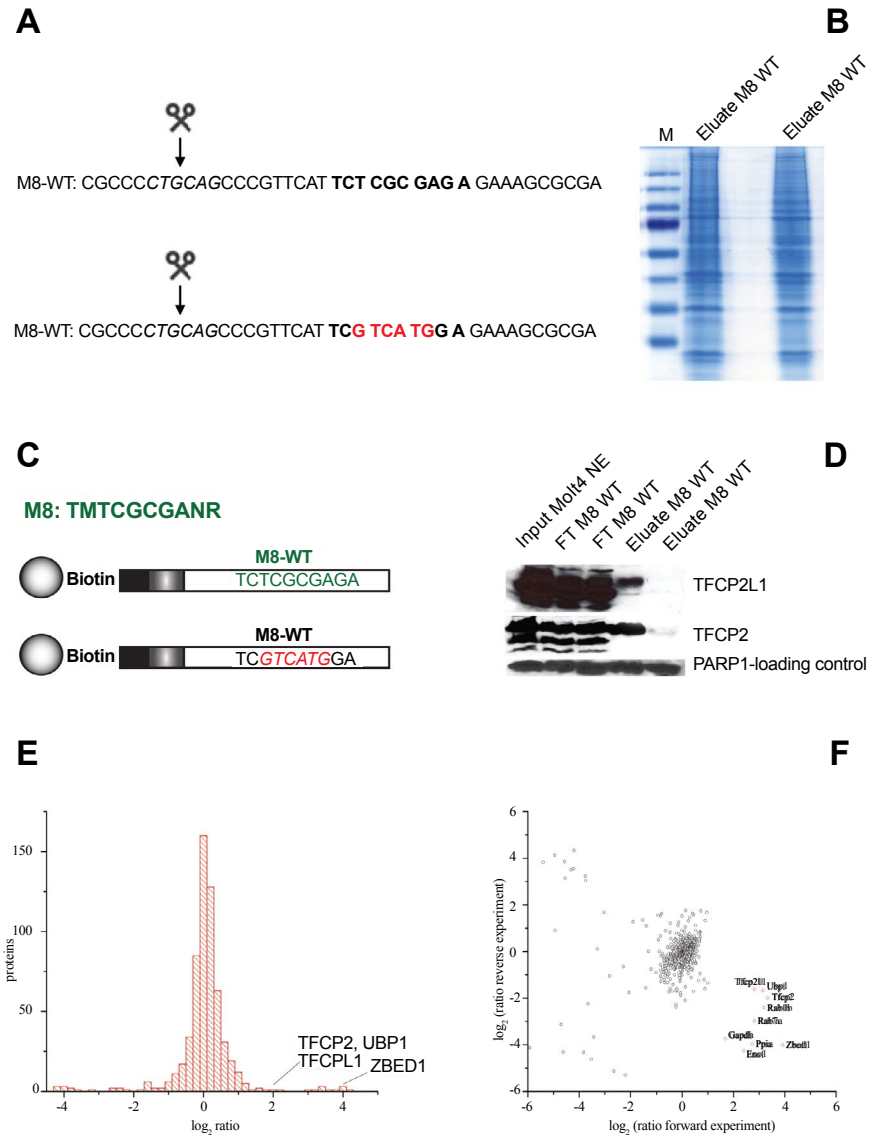

Figure 1: Properties of M8 motif interactome. (A) The M8-cis regulatory element is shown in bold letters, whereas the point mutations designed to disrupt DNA binding are highlighted in red. (B) Specific M8 interaction partners cannot be identified by one-dimensional SDS-PAGE and colloidal coommassie staining comparing the eluted material from the affinity columns carrying either a M8 wild-type (WT) or mutant (MT) sequence. (C) Scheme of biotinylated wild-type and mutant M8 sequences used for immobilization. The PstI restriction site (CTGCAG) is indicated (italics). (D) Immunoblot analysis of selected M8-specific interaction partners confirms results from the SILAC analysis. Equal amounts (20% of total) of input (IN) and flow-through (FT) as well as eluted (Elution) material (20% of total) from the M8 wild-type (WT) and mutant (MT) columns, respectively, were analyzed by Western-blot with antibodies directed against TFCP2 and TFCP2L1. The Poly [ADP-ribose] polymerase 1 protein (PARP1) serves as a control for equalized total protein loading amounts and unspecific DNA background binding. (E) SILAC ratio distribution of all proteins identified at least once in a forward and reverse experiment. Proteins containing at least two peptides with log2 SILAC ratios (lysine-$^{2}H_{4}$ or arginine $^{13}C_{6}$ vs lysine-$^{1}H_{4}$ or arginine $^{12}C_{6}$) greater than 1.0 are initially considered specific. Likewise, in reverse experiments (see text) the log2 of inverted ratios was used. Scatter plot of SILAC ratios of all proteins identified at least once in a forward and reverse experiment. The mean ratios of proteins which were found in at least one forward and one reverse experiment were plotted. Proteins enriched on the wild-type bait were initially considered as positive hits and further analyzed by statistics.

baits. However, nine out of 710 proteins exhibited log2 SILAC ratios substantially different from one in at least one of the four DNA affinity chromatography experiments. These are zinc finger BED domain-containing protein 1 (ZBED1), alpha globin transcription factor CP2 (TFCP2), transcription factor CP2 like-1 (TFCP2L1), upstream binding protein 1 (UBP1), alpha-enolase, glyceraldehyde-3-phosphate dehydrogenase (GAPDH), peptidyl-prolyl *cis*-trans isomerase A, Ras-related protein Rab-7a, and Ras-related protein Rab-1B. To account for this, we applied more specific selection conditions, in which only those that appeared in all four pull-down experiments with a Max Quant significance B value < 0.05 (p(forw) × p(rev) < 0.05) were considered as *bona fide* M8 motif-binding candidates. By doing so, the last five proteins mentioned above were filtered out from the list. The remaining TFs ZBED1, TFCP2, TFCP2L1 and UBP1 exhibiting SILAC ratios of 15.465, 4.4744, 3.5161, and 3.8396, respectively (Figures 1E and 1F and Table 2), were considered for further investigation.

ZBED1 was first identified as DNA replication – Related Element binding Factor (DREF). In *Drosophila*, ZBED1 binds to replication – related element [5'-TATCGATA-3'] [19]. In humans, ZBED1 binds to the A box consensus palindromic sequence [5'-TGTCG(C/T) GA(C/T)A-3'] acting as a positive regulator of cell proliferation and ribosomal gene expression. The A box sequence is quite similar to the M8 motif. Furthermore, previous ChIP data from different groups have confirmed the functional binding of ZBED1 to many ribosomal M8 harboring genes such as RPS6, RPL10A, and RPL12 [20]. Therefore, we believe DREF/ZBED1 is one of the important factors that regulate gene expression of M8 bearing genes.

TFCP2, TFCP2L1 and UBP1 are three closely related but distinct proteins, which belong to the *Drosophila* grainyhead transcription factor family [21]. These TFs do not possess any well-characterized DNA recognition motif like zinc finger, helix-turn-helix, homeo-box or leucine zipper. They do, however, bind to the consensus DNA sequence CNRG-N_6-CNRG in context with a promoter element, overlapping with the CCAAT box [22] as homo- or hetero-dimers [23]. It seems that the short sequence CNRG is part of the M8 motif (*CGAG*), and that the M8 motif itself is a palindromic DNA sequence. As a result, it would absolutely make sense if its associated transcription factors are binding to M8 as homo- or heterodimers. This also suggests a potential pre-complex formation between these TF family members prior to loading on DNA. We have detected unique peptides for all three individual proteins (Supplementary Table 1), which demonstrates that all three CP2 family member proteins interact, in a direct or indirect manner, with the M8 motif *in vitro*.

Results obtained by the SILAC-proteomics-based-approach are consistent with conventional immunoblotting detection

To confirm the results obtained by our quantitative proteomics pipeline, we re-performed the pull-down with M8 and mutant M8 oligonucleotide baits similarly as used in the proteomic screening experiments. The bound material from the DNA ligands was released separately and subjected to western blot analysis. As shown in Figures 1A-1D both TFCP2 and TFCP2L1 are interacting specifically with the wild-type M8 motif-bearing DNA sequence. Interestingly, only the slowest migrating form of TFCP2 present in the Molt4 nuclear extract is able to bind to DNA. This was also observed in the case of TFCP2L1. Both proteins appear as multiple isoforms in Molt4 cells, but only the largest isoforms (presumably harboring an intact DNA binding domain) are able to interact specifically with the M8 motif.

TFCP2 interacts with M8 *in vivo*

In order to demonstrate the biological relevance of our M8 binding candidates, we performed chromatin immunoprecipitation (ChIP) experiments. Due to the lack of suitable ChIP grade antibodies ChIP analysis was only carried out for TFCP2. As shown in Figure 2, the two selected M8 motif bearing promoters of the RBP-j-kappa and PurA genes, respectively, were robustly and specifically enriched by the anti-TFCP2 antibody in comparison to the input DNA, the pre-immuneserum control (IgG) or the exon DNA control regions. Hence we conclude that our M8 motif interaction partner list is of very high quality and does therefore qualify for further biological studies that are aiming to address the functional role of the M8 motif in mammalian gene expression regulation.

Advantages compared the classical electrophoretic mobility shift assay (EMSA)

In our particular situation, the three highly-related but distinct TF proteins TFCP2, UBP1, and TFCP2L1 have quite similar molecular weights ranging from 54 to 57 kDa (Supplementary Table 1). They all make contact with the same DNA consensus sequence either in a direct or indirect manner. Importantly, their protein-DNA complexes display similar electrophoretic mobility, making the differentiation between individual EMSA complexes virtually impossible (data not shown). Strikingly, by using our SILAC quantitative proteomics approach combined with high resolution mass spectrometry, it is possible to resolve this issue, demonstrating that our methodology indeed represents a robust reverse CHIP [24] approach.

Protein Names	Uniprot	Gene Names	Unique Peptides	Sequence Coverage [%]	SILAC ratio, normalized	Significance A (corr)	Significance B (corr)
Zinc finger BED domain-containing protein 1	O96006	ZBED1; ALTE; DREF; KIAA0785; TRAMP	35	63.7	15.465	4.1457E-07	1.5767E-11
Alpha-globin transcription factor CP2, isoform 1	Q12800-1	TFCP2; LSF; SEF	24	62.7	4.4744	0.01275003	0.00091001
Upstream-binding protein 1	Q9NZI7-1	UBP1; LBP1	23	61.9	3.5161	0.05167777	0.0084517
Transcription factor CP2-like protein 1	Q9NZI6	TFCP2L1; CRTR1; LBP9	22	56.2	3.8396	0.03198858	0.00406242
Alpha-enolase	P06733-1	ENO1; ENO1L1; MBPB1; MPB1	18	47.5	17.756	9.4386E-08	4.1214E-14
Glyceraldehyde-3-phosphate dehydrogenase	P04406	GAPDH; GAPD; CDABP0047; OK/SW-cl.12	12	50.1	11.871	5.3191E-06	1.4037E-09
Peptidyl-prolyl cis-trans isomerase A	P62937	PPIA; CYPA	13	75.2	11.23	8.8744E-06	3.8145E-10
Ras-related protein Rab-7a	P51149	RAB7A; RAB7	5	30.9	9.3246	4.6223E-05	0.05613949
Ras-related protein Rab-1B	Q9H0U4	RAB1B;RAB1C;RAB1A;RAB1	5	32.8	8.0692	0.00015555	9.6487E-08

Table 2: Proteins binding with a log2 SILAC ratio greater than one to the DNA column bearing the wild type M8 motif: The Uniprot number, gene name, observed unique peptide number, sequence coverage, normalized SILAC ratio and significance A and B values of potential specific binders were shown.

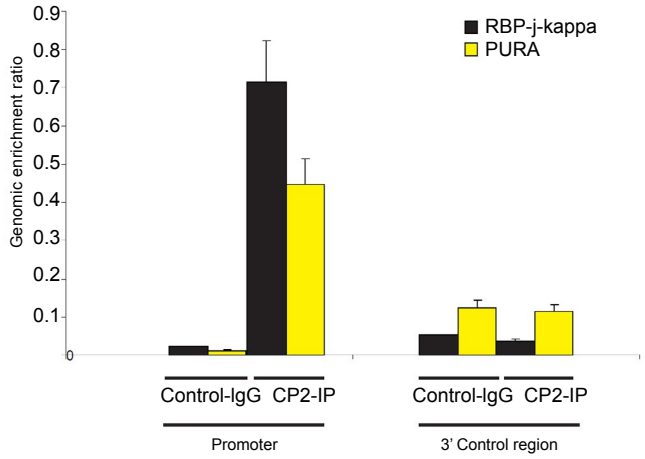

Figure 2: Chromatin immunoprecipitation (ChIP) analysis of TFCP2 protein (CP2-IP) demonstrating specific binding of TFCP2 to promoter regions of two selected M8 motif-bearing genes (RBP-j-kappa and PurA) *in vivo*. One selected region, localizing 1.8 downstream to the M8 motif of the RBP-j-kappa gene and a region, localizing 1.3kb downstream to the M8 motif of the PurA gene served as controls for background DNA binding. Rabbit pre-immune serum (Control-IgG) was used as the specificity control for the anti TFCP2 antibody. The experiments were conducted in triplicate (error bars show SD).

Potential role of M8 motif in erythrocyte development

Until now there has been no systematic study to identify how ribosomal proteins change during erythrocyte differentiation. It is known, however, that a defect in ribosomal protein S19 blocks red blood cell development [25]. The S19 gene is one of eight ribosomal M8 motif harboring genes, including RPL10A, RPL12, RPL17, RPL26, RPS15A, RPS19, RPS6 and RPS7. In addition, the three CP2 family M8 binders (TFCP2, UBP1, and TFCP2L1) are well known to be involved in regulating one of the most important erythrocyte-specific genes, the alpha globin locus [26]. In context of the latter TFCP2 seems to bind an element ($CNRGN_{5\text{-}6}CNRG$) that is also found in other loci like the serum amyloid A3 and Aα-fibrinogen gene promoter [27]. Therefore, we predict that the M8 motif and M8-associated factors may have ubiquitous roles but might also represent a hitherto uncharacterized regulatory circuit of the erythrocyte differentiation program (Supplementary Figure 1).

Authors' Contributions

Gerhard Mittler supervised the study and designed the experiments together with Trung Tat Ngo and performed the liquid chromatography mass spectrometric analysis. Rudolf Engelke carried out the bioinformatic and statistics analysis.

References

1. Kadonaga JT (2004) Regulation of RNA polymerase II transcription by sequence-specific DNA binding factors. Cell 116: 247-257.
2. Blanchette M, Bataille AR, Chen X, Poitras C, Laganière J, et al. (2006) Genome-wide computational prediction of transcriptional regulatory modules reveals new insights into human gene expression. Genome Res 16: 656-668.
3. Mittler G, Butter F, Mann M (2009) A SILAC-based DNA protein interaction screen that identifies candidate binding proteins to functional DNA elements. Genome Res 19: 284-293.
4. Gordân R, Shen N, Dror I, Zhou T, Horton J, et al. (2013) Genomic regions flanking E-box binding sites influence DNA binding specificity of bHLH transcription factors through DNA shape. Cell Rep 3: 1093-1104.
5. Bulyk ML (2003) Computational prediction of transcription-factor binding site locations. Genome Biol 5: 201.
6. Elnitski L, Jin VX, Farnham PJ, Jones SJ (2006) Locating mammalian transcription factor binding sites: a survey of computational and experimental techniques. Genome Res 16: 1455-1464.
7. Hannenhalli S (2008) Eukaryotic transcription factor binding sites--modeling and integrative search methods. Bioinformatics 24: 1325-1331.
8. Badis G, Berger MF, Philippakis AA, Talukder S, Gehrke AR, et al. (2009) Diversity and complexity in DNA recognition by transcription factors. Science 324: 1720-1723.
9. Nakagawa S, Gisselbrecht SS, Rogers JM, Hartl DL, Bulyk ML (2013) DNA-binding specificity changes in the evolution of forkhead transcription factors. Proc Natl Acad Sci U S A 110: 12349-12354.
10. Xie X, Lu J, Kulbokas EJ, Golub TR, Mootha V, et al. (2005) Systematic discovery of regulatory motifs in human promoters and 3' UTRs by comparison of several mammals. Nature 434: 338-345.
11. Déjardin J, Kingston RE (2009) Purification of proteins associated with specific genomic Loci. Cell 136: 175-186.
12. Ong SE, Blagoev B, Kratchmarova I, Kristensen DB, Steen H, et al. (2002) Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. Mol Cell Proteomics 1: 376-386.
13. Dignam JD, Lebovitz RM, Roeder RG (1983) Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. Nucleic Acids Res 11: 1475-1489.
14. Shevchenko A, Tomas H, Havlis J, Olsen JV, Mann M (2006) In-gel digestion for mass spectrometric characterization of proteins and proteomes. Nat Protoc 1: 2856-2860.
15. Rappsilber J, Mann M, Ishihama Y (2007) Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. Nat Protoc 2: 1896-1906.
16. Cox J, Mann M (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat Biotechnol 26: 1367-1372.
17. Cox J, Matic I, Hilger M, Nagaraj N, Selbach M, et al. (2009) A practical guide to the MaxQuant computational platform for SILAC-based quantitative proteomics. Nat Protoc 4: 698-705.
18. Hubner NC, Mann M (2011) Extracting gene function from protein-protein interactions using Quantitative BAC InteraCtomics (QUBIC). Methods 53: 453-459.
19. Hirose F, Yamaguchi M, Handa H, Inomata Y, Matsukage A (1993) Novel 8-base pair sequence (Drosophila DNA replication-related element) and specific binding factor involved in the expression of Drosophila genes for DNA polymerase alpha and proliferating cell nuclear antigen. J Biol Chem 268: 2092-2099.
20. Yamashita D, Sano Y, Adachi Y, Okamoto Y, Osada H, et al. (2007) hDREF regulates cell proliferation and expression of ribosomal protein genes. Mol Cell Biol 27: 2003-2013.
21. Wilanowski T, Tuckfield A, Cerruti L, O'Connell S, Saint R, et al. (2002) A highly conserved novel family of mammalian developmental transcription factors related to Drosophila grainyhead. Mech Dev 114: 37-50.
22. Lim LC, Fang L, Swendeman SL, Sheffery M (1993) Characterization of the molecularly cloned murine alpha-globin transcription factor CP2. J Biol Chem 268: 18008-18017.
23. Yoon JB, Li G, Roeder RG (1994) Characterization of a family of related cellular transcription factors which can modulate human immunodeficiency virus type 1 transcription in vitro. Mol Cell Biol 14: 1776-1785.
24. Rusk N (2009) Reverse CHIP. Nature Methods 6.
25. Flygare J, Kiefer T, Miyake K, Utsugisawa T, Hamaguchi I, et al. (2005) Deficiency of ribosomal protein S19 in CD34+ cells generated by siRNA blocks erythroid development and mimics defects seen in Diamond-Blackfan anemia. Blood 105: 4627-4634.
26. Bosè F, Fugazza C, Casalgrandi M, Capelli A, Cunningham JM, et al. (2006) Functional interaction of CP2 with GATA-1 in the regulation of erythroid promoters. Mol Cell Biol 26: 3942-3954.
27. Bing Z, Reddy SA, Ren Y, Qin J, Liao WS (1999) Purification and characterization of the serum amyloid A3 enhancer factor. J Biol Chem 274: 24649-24656.