ISSN 0006-2979, Biochemistry (Moscow), 2008, Vol. 73, No. 11, pp. 1260-1268. © Pleiades Publishing, Ltd., 2008.
Published in Russian in Biokhimiya, 2008, Vol. 73, No. 11, pp. 1569-1579.
Originally published in Biochemistry (Moscow) On-Line Papers in Press, as Manuscript BM08-074, October 12, 2008.

Identification of Recognition Sites for Myc/Max/Mxd Network Proteins by a Whole Human Chromosome 19 Selection Strategy

S. B. Akopov[1,2], I. P. Chernov[1,2], T. Wahlström[2], M. B. Kostina[1], G. Klein[2], M. Henriksson[2], and L. G. Nikolaev[1]*

[1]*Shemyakin–Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Sciences, ul. Miklukho-Maklaya 16/10, 117997 Moscow, Russia; fax: (495) 330-6538; E-mail: lev@humgen.siobc.ras.ru*
[2]*Department of Microbiology, Tumor and Cell Biology (MTC), Karolinska Institutet, Box 280, SE-17177 Stockholm, Sweden*

Received February 27, 2008
Revision received June 11, 2008

Abstract—In this study, we have identified 20 human sequences containing Myc network binding sites in a library from the whole human chromosome 19. We demonstrated binding of the Max protein to these sequences both *in vitro* and *in vivo*. The majority of the identified sequences contained one or several CACGTG or CATGTG E-boxes. Several of these sites were located within introns or in their vicinity and the corresponding genes were found to be up- or down-regulated in differentiating HL-60 cells. Our data show the proof of principle for using this strategy in identification of Max target genes, and this method can also be applied for other transcription factors.

DOI: 10.1134/S0006297908110138

Key words: Max, Myc, Mxd, target genes, human chromosome 19

The completion of the human genome sequence [1, 2] and sequences of other genomes [3] opened up the possibility for massive analysis of functional genomic elements. At first coding sequences [4] have been analyzed, but to understand the complex interactions between genes and other functional elements the complete annotation of both coding and non-coding parts of the genome are needed. Experimental data on genomic positions of a multitude of regulatory sequences, such as enhancers, silencers, insulators, transcription terminators, and replication origins are very limited, especially at the whole genome level. Since most genomic regulatory elements (e.g. enhancers) are generally gene-, tissue-, or cell-specific, the prediction of these elements by computational methods is difficult and often ambiguous. Therefore, the development of high-throughput experimental approaches for identifying and mapping genomic functional elements is highly desirable [5]. An important example of this kind of elements is DNA sequences recognized and bound by various regulatory proteins, including transcription factors.

The Myc/Max/Mxd network of transcription factors plays an important role in the regulation of cellular behavior [6-8]. *Myc* genes are frequently deregulated in many human tumors and affect development as well as cell cycle progression. In contrast, Mxd (formerly Mad) and Mnt function as Myc antagonists and may even serve as tumor suppressors [9, 10]. The Myc/Max, Mxd/Max, and Mnt/Max complexes share the E-box recognition sequence within the *cis*-regulatory elements where they bind. However, the E-box sequence is frequent in the genome and is recognized also by several other proteins like USF [11], which makes it problematic to understand the *in vivo* relevance of specific E-box elements for the regulation of specific genes by the Myc network members or other E-box binding proteins. Since the Myc and Mxd proteins are expressed at a relatively low level, it has been suggested that Myc/Max or Mxd/Max complexes cannot occupy all available binding sites. Thus, selectivity for specific E-box elements occurs. The identification of *in vivo* target genes for the Myc network proteins will be a basis for understanding the specificity of E-box binding proteins as well as to give insights into the mechanisms of regulation of cell cycle, apoptosis, and differentiation. It still remains to be shown to what extent the different fac-

Abbreviations: BSA) bovine serum albumin; ChIP) chromatin immunoprecipitation assay; DMSO) dimethylsulfoxide; DTT) dithiothreitol; EMSA) electrophoretic mobility shift assay; MPC) magnetic particle concentrator; PBS) phosphate buffered saline.

* To whom correspondence should be addressed.

tors of the Myc network regulate the same or distinct set of genes, since the *in vivo* target genes are not fully identified. It has been shown by cDNA arrays that c-Myc and Mxi1 engage both common and distinct target genes [12]. Probably Myc and Mxd1 behave similarly. In addition, *CCND2* (cyclin D2), *cdk4* (cyclin-dependent kinase 4), *hTERT* (human telomerase reverse transcriptase), and *odc* (ornithine decarboxylase) genes have been shown to be direct targets for both c-Myc and Mnt [13-16]. The known Myc target genes were first identified by empirical, candidate-based approaches and more recently by a number of expression profiling screens, which have shown correlation between the expression of individual genes and altered Myc levels. Another kind of study is genetic screens, which has allowed identification of genes capable of rescuing loss of Myc function.

Recently, the Myc binding sites within the genome has been assessed and thousands of loci bound by Myc were identified. However, the correlation between Myc binding and changes in the transcription of target genes need to be further explored. Among all the genes identified as Myc target genes, only a minority has a well-documented role in specific Myc functions [17]. Therefore, methods for identification of target genes, which also are specific for Myc function, need to be further established. In addition, it has been suggested that other Myc network proteins, in particular Mnt, are of great importance for regulation of Myc target genes. We recently showed that phosphorylation of Mnt at cell cycle entry disrupts the critical Mnt–mSin3–HDAC interaction, which allows relief of Mnt-mediated transcriptional repression of the Myc/Mnt target gene *cyclin D2* [15]. With this in mind, it is of importance to investigate the binding sites shared for the entire network. To address this issue we present an approach for searching and mapping of binding sites of the entire Myc network. The method used is expanded from our previous work, which was based on a genomic sequence selection procedure [18] where libraries of short PCR-amplified fragments from the whole human chromosome were constructed and DNA fragments capable of binding S/MAR elements were selected.

The human chromosome 19 is characterized by having the highest density of genes, and the probability of finding Myc/Max/Mxd regulated genes is therefore higher compared to most of the other human chromosomes. Using this approach, we have identified Max binding to several human sequences that potentially are regulated by the Myc network.

MATERIALS AND METHODS

Cell culture. The human promyelocytic leukemic cell line HL-60 was grown and treated with dimethylsulfoxide (DMSO) to induce differentiation as described previously [19].

General procedures. Growth and transformation of *Escherichia coli*, preparation of plasmid DNA, agarose gel electrophoresis, blot-hybridization, and other standard procedures were performed as described previously [20]. Expression and purification of the GST-Max protein was carried out as described [21].

Preparation of short fragment library and selection of Max binding fragments. A human chromosome 19 library consisting of short DNA fragments with library primer (Table 1) on both ends was generated as described previously [18, 22].

To prepare for selection of the DNA fragments bound by Max protein, 250 µl of Dynabeads M-280 suspension (Dynal, USA) (10^8 particles coated with sheep α-rabbit IgG) were washed twice with 500 µl of phosphate buffered saline/0.1% bovine serum albumin (PBS/BSA) using the Dynal magnetic particle concentrator (MPC) and re-suspended in 800 µl of PBS/BSA. Fifteen micrograms of rabbit anti-Max antibody (sc-197; Santa Cruz Biotechnology, USA) was added to the particles, mixed by vortexing, and incubated for 30 min at room temperature with continuous slow rotation followed by two washes with PBS/BSA. The Dynabeads particles were then washed with 0.2 M triethanolamine-HCl, pH 9.0, and re-suspended in 10 ml of the same buffer. Fifty-two milligrams of solid dimethyl pimelimidate-2HCl was added to the suspension, and the mixture was incubated for 45 min at room temperature with slow rotation to covalently cross-link the primary and secondary antibodies. The supernatant was discarded with the help of MPC, and the Dynabeads particles were re-suspended in 10 ml of the 0.2 M triethanolamine-HCl, pH 9.0, and incubated for additional 2 h. Further, the particles were washed once with 0.1% Triton X-100 and three times with PBS, re-suspended in 1 ml of PBS containing 0.02% sodium azide, and stored at 4°C.

For selection of sites bound by Max, a binding reaction mixture containing 100 ng of the short DNA fragments of chromosome 19 library, 500 ng *Sau*3A-digested phage λ DNA, 2 µg sonicated salmon sperm DNA, 1 µg BSA, 10 ng GST-Max in 40 µl of buffer A (12 mM Hepes-KOH, pH 7.9, 60 mM KCl, 0.12 mM EDTA, 0.3 mM dithiothreitol (DTT), 12% glycerol, and 0.5 mM phenylmethylsulfonyl fluoride) was incubated for 30 min at 25°C. For negative control, a similar reaction mixture without addition of the GST-Max was prepared. Particles were collected with MPC.

Ten microliters (10^6 particles) of the prepared anti-Max Dynabeads were washed and pre-incubated in 0.5 ml of buffer A containing 4 µg sonicated salmon sperm DNA, 5 µg *Sau*3A-digested λ DNA, and 4 µg BSA for 15 min at room temperature followed by collection of the particles by MPC. The binding reaction mixture was added to the Dynabeads pellet and incubated for 15 min at room temperature. The Dynabeads were then washed three times with 150 µl of buffer A containing up to

Table 1. Sequences of oligonucleotides used

	Length of expected PCR product (bp)	Sequence
Library primer	N/A	actgagctcgagtatccatgaaca
RT-PCR primers		
1	104	cacggctaacaagcccac ggagtcagaggttggggag
2	280	gaaatccgggtcaagtgaag cccctccagacagatgagg
3	155	gttcactgggatcagcca ctgctttaatgaggggtgtgt
6	208	cacgggagagagaccctatg caccattgccacagttacca
8	209	atcggcctctgtatggagt ggggatgaagaagagcaccag
11	174	caggcatgatttcaccttca ctcggtacagcttcctctgg
Chromatin immunoprecipitation (ChIP) primers		
3	183	tgcacagacttacaggggtg ggaaagcagcatgaaatacca
4	186	gagcttgtgcagggaaactc agtgcctgatatggtttggc
5	166	acacgaaagaactcacactgg acgaacatactggggacac
6	169	gaacatgcaagcagaacacg aggctgttagtggcaaggtg
8	183	ttgatttgggacacctgtga tctttcctgttctcgcctgt
11	189	tgtaatttggggcttcattca tgccaatcagtgttaaggttg
cyclin D2	209	gagctcgagccacgccatgc ccctgacacgtgctctaacgc

Note: N/A, not applicable.

10 μg/ml BSA and up to 50 μg/ml sonicated salmon sperm DNA and twice with buffer A containing 10 μg/ml BSA, and the final pellet was re-suspended in 100 μl of TE buffer (10 mM Tris-HCl, pH 8.0, 1 mM EDTA) containing 0.5% SDS. The mixture was incubated for 5 min at room temperature with rotation, and the Dynabeads were removed using the MPC. Proteinase K up to 10 mg/ml was added and the reaction mixture was incubated for 1 h at 57°C. After incubation, 3 M sodium acetate, pH 5.0, was added up to 0.3 M, and the DNA was extracted with phenol–chloroform–isoamyl alcohol (25 : 24 : 1) and precipitated with ethanol overnight at –20°C. The DNA bound by Max was collected by centrifugation at 12,000 rpm for 10 min at 4°C, washed with 75% ethanol, dried, and dissolved in 15 μl of TE buffer. Five microliters of DNA was used as a template for PCR amplification with the library primer (Table 1). PCR was performed as follows: 94°C for 20 sec, 60°C for 40 sec, 72°C for 90 sec for 18, 20, and 22 cycles. The minimal number of amplification cycles producing sufficient amount of the product was chosen. For the next round of selection 100 ng of the DNA purified by phenol–chloroform extraction and precipitated with ethanol was used.

Cloning and analysis of the library of selected fragments. After the fourth round of selection, the mixture of PCR-amplified fragments was cloned using the pGEM-T PCR cloning system (Promega, USA). Transformed *E. coli* cells were plated on X-gal/IPTG agar plates, and white colonies were arrayed on 96-well clusters. PCR-amplified inserts were resolved in agarose gel and blotted to nylon filters (Hybond-N; Amersham Pharmacia Biotech, USA). The filters were hybridized with the ran-

dom primer labeled pool of selected fragments after the fourth round of selection. Clones that gave strong hybridization signal and therefore are abundant in the library were selected for sequencing.

Electrophoretic mobility shift assay (EMSA). The selected fragments were labeled with [α-^{32}P]dCTP in a course of PCR and gel-purified as described previously [18]. The labeled probe was incubated with 2.5 ng of GST-Max protein in 20 μl of EMSA buffer (20 mM Hepes-KOH, pH 7.3, 50 mM KCl, 3 mM $MgCl_2$, 1 mM EDTA, 8% glycerol, 1 mM 2-mercaptoethanol) in the presence of 1 μg sonicated salmon sperm DNA and 1 μg BSA at 30°C for 30 min. For supershifts, 50 ng of anti-Max antibodies (sc-197; Santa Cruz Biotechnology) were pre-incubated with GST-Max in the EMSA buffer for 10 min at 30°C and added to the binding reactions. The DNA–protein complexes were separated from free probe by electrophoresis in 5% polyacrylamide gel in 25 mM Tris-boric acid, pH 8.3, 0.5 mM EDTA at 4°C. Gels were dried and developed with the use of PhosphorImager technology (Molecular Dynamics, USA).

Chromatin immunoprecipitation assay (ChIP). Chromatin fragments were isolated with anti-Max antibodies in HL-60 cells as described previously [19]. The DNA was used as a template for PCR with primer pairs targeted to amplify the 150-200 bp fragments within the putative Max-binding sequence that include the E-box(es). Sequences of corresponding primers are presented in Table 1. Primers targeted at 5′-region of cyclin D2 (*CCND2*) gene were used as a positive control. In parallel, as a negative control, immunoprecipitation using anti-β-galactosidase antibodies was performed. No PCR product was detected at the number of PCR cycles used for amplification of anti-Max immunoprecipitated DNA.

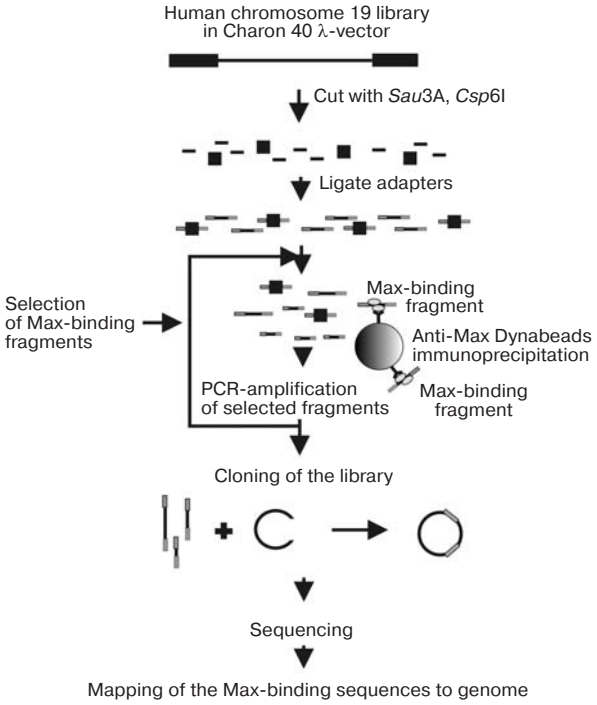
RNA isolation, cDNA synthesis, and real-time PCR. Total cellular RNA was isolated with the RNeasy Mini RNA purification kit (Qiagen, USA). All RNA samples were further treated with DNase I to remove residual DNA. First strand cDNA synthesis was performed using a SuperScript Preamplification System (Life Techologies, GibcoBRL, USA). Control samples without addition of reverse transcriptase were prepared in parallel. Real-time PCR was done with the use of ABI Prism 7000 and SYBR Green PCR master mix (Applied Biosystems, USA) in 50 μl reaction volume for 40 cycles with the following profile: 95°C for 30 sec, 58°C for 30 sec, and 72°C for 60 sec.

Sequencing, GenBank comparison, and identification of genes. Clone inserts were sequenced using an ALFexpress II automated DNA sequencer (Amersham Pharmacia Biotech). The sequences were compared to the GenBank deposits using the BLAST [23] server at NCBI (http://www.ncbi.nlm.nih.gov/blast/). The data were further analyzed with the help of the Draft Human Genome Browser [24] (http://genome.ucsc.edu/cgi-bin/hgGateway?db=hg12, assembly of March, 2006). RepeatMasker (http://www.repeatmasker.org/cgi-bin/WEBRepeatMasker) was used for the analysis of repeated sequences.

Oligonucleotides. Primers for PCR amplification were designed using the Primer 3 WWW server at Whitehead Institute (http://www-genome.wi.mit.edu/cgi-bin/primer/primer3.cgi). Oligonucleotides were synthesized using an ASM-102U DNA synthesizer (Biosset Ltd, Russia).

RESULTS

Selection of sequences bound by the Max protein. In order to identify regulatory sites for the Myc/Max/Mxd network proteins on human chromosome 19, a library of short fragments from this chromosome was pooled with GST-Max protein and Dynabeads coupled with Max antibodies (Scheme). Four successive rounds with purification of sequences bound to Max followed by PCR amplification were performed to select for Max binding fragments (Fig. 1). When separated on agarose gel the initial library DNA produced a ladder of fragments mostly originated from the restriction enzymes digested λ-vector arms. However, after four rounds of selection the pattern



Strategy for selection of Max-binding sequences from a whole human chromosome 19 genomic library. A library of short fragments from chromosome 19 was pooled with full length GST-Max together with anti-Max antibody-coupled Dynabeads. Selection for sequences bound by Max was performed in four successive rounds of Max binding/purification/PCR amplification. After the final round of selection, the PCR-amplified mixture of DNA fragments was cloned into a pGEM-T plasmid

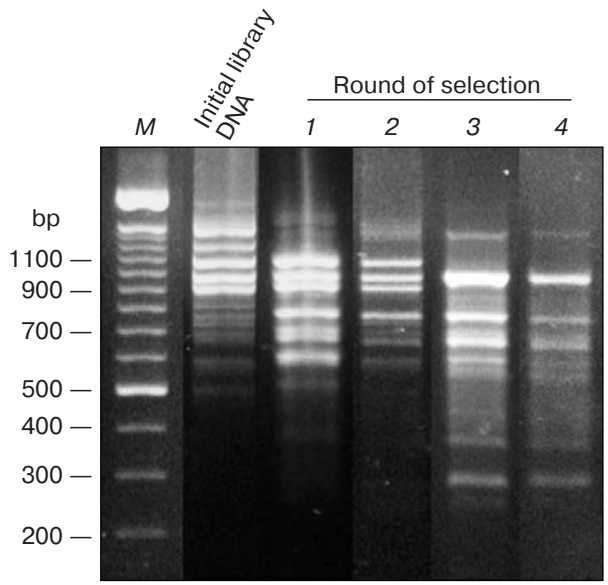


Fig. 1. Selection of Max-binding DNA fragments. The initial library of short fragments of chromosome 19 and PCR-amplified fragments after 1 to 4 rounds of selection resolved in 1.5% agarose gel as indicated. The pattern after the fourth round of selection corresponds to Max-binding fragments.

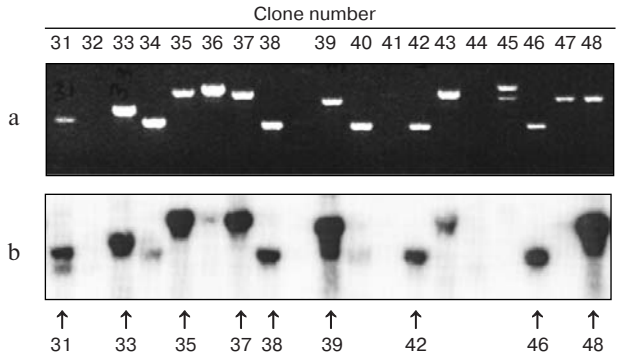


Fig. 2. PCR-amplified inserts of the selected putative Max-binding clones were resolved in 1.5% agarose gel (a), blotted to nylon filter (b), and hybridized with a mix of ^{32}P-labeled fragments obtained after the fourth round of selection (see Fig. 1). Clones that gave rise to strong hybridization signals (indicated by arrows below) were selected for further analysis.

was changed since the human DNA fragments with different electrophoretic mobility were selected for and amplified (Fig. 1). No bands were detected after the same number of PCR amplification cycles in the negative control, which was performed without addition of the GST-Max protein (data not shown). After the final round of selection, the PCR-amplified mixture of DNA fragments was cloned into a pGEM-T plasmid (Scheme). Transformed *E. coli* was plated on X-gal/IPTG agar plates and 192 white colonies were arrayed on two 96-well clusters.

The presence and the size of the inserts in all clones were analyzed by PCR using the library primer (Table 1). One hundred-eighty PCR-amplified inserts were resolved in agarose gel (Fig. 2a), blotted to nylon filters, and hybridized with a pool of labeled fragments obtained after the fourth round of selection (Fig. 2b). Fifty-one clones produced strong hybridization signal (indicated by arrows in Fig. 2b), which showed their abundance in the library. These clones were selected for sequencing. We assumed that clones giving weak hybridization signal, like clone number 34 and 36 (Fig. 2b), represented background and were excluded from further analysis.

Sequencing of the selected clones produced 20 independent human sequences. These were compared to GenBank and were all unambiguously mapped into genomic sites of human chromosome 19. Positions of identified Max binding sites within human chromosome 19 fragments are presented in Table 2.

All selected sequences (except three of them) contained one or more copies of canonical (CACGTG) or non-canonical (CATGTG) sequences of E-box (Table 2). We analyzed the content of CACGTG and CATGTG sequences in ~150 kb of human chromosome 19 (GenBank accessions AC002133 and AC002115). The content of CACGTG was 1/9673 bp (which is not surprising since this hexanucleotide includes the CG dinucleotide), and the content of CATGTG – 1/1909 bp, which is close to its content in random sequence (1/2048 bp). The content of CACGTG and CATGTG in the selected fragments exceeded those in genomic DNA 12.8- and 6.5-fold, respectively. The last difference may reflect different efficiency of Max binding by these two E-boxes. Twelve of the sequences contained more than one E-box, raising the possibility of cooperative Max binding [25].

Arrangement of Max-binding sequences relative to genes. The human genomic loci in which the identified sequences are located can be subdivided into four groups according to mutual positions of the Max binding sequences and cellular genes. Group number 1 includes five loci (sequence ID numbers 4, 12-14, and 17 in Table 2), that did not contain any known genes within 10 kb up- and downstream of the Max binding sites. However, additional genes may still remain to be identified within these loci. Alternatively, it is also possible that sequences capable of binding Max, similar to enhancers, may exert their regulatory potential over significant distances. In this case, identification of their target genes could be a challenge. Group number 2 includes sequence ID numbers 8 and 9 (Table 2), located upstream of known genes, namely the genes encoding the free fatty acid receptor 2 (FFAR2, ~3.5 kb) and the pregnancy specific β-1-glycoprotein 6 (PSG6, ~7 kb), respectively. The third group consists of the only one Max-binding sequence that was found ~150 bp downstream of the gene encoding the zinc finger protein ZNF221 (sequence ID number 11; Table

Table 2. Properties of identified Max-binding sequences

Sequence ID	Length, bp	Position in chromosome 19*	Position relative to known genes	E-box CACGTG	E-box CATGTG	EMSA	ChIP	RT-PCR difference**
1	854	7813810-7814663	1st intron of *EVI5L*	1	12	+	+	no difference
2	546	40480512-40481057	4th intron of *MAG*	0	4	+	–	no expression
3	754	11203956-11204709	22nd intron of *DOCK6*, 4.7 kb upstream of *LOC55908*	1	5	+	+	–2.5
4	663	56630357-56631019	intergenic	2	1	+	+	N/A
5	728	11876765-11877492	4th intron of *ZNF69*	2	0	+	+	N/D
6	409	57088844-57089252	4th intron of *ZNF649*	0	1	+	+	+3.9
7	568	33833216-33833783	5th intron of *BC068609*	1	3	+	N/D	N/D
8	199	40628518-40628716	3.5 kb upstream of *FFAR2* (*GPR43*)	1	0	+	+	–22
9	540	48120136-48120675	~7 kb upstream of *PSG6*	0	1	N/D	N/D	N/D
10	426	57221896-57222321	1st intron of *ZNF614*	1	2	N/D	N/D	N/D
11	608	49163915-49164451	0.5 kb downstream of *ZNF221*	1	2	+	+	+3.7
12	617	34181367-34181982	intergenic	1	3	+	N/D	N/A
13	506	56515805-56516310	»	1	4	N/D	N/D	N/A
14	333	7057671-7058102	»	0	0	+	N/D	N/D
15	443	58275473-58275921	3rd intron of *ZNF160*	0	0	N/D	N/D	N/D
16	446	40481054-40481499	4th intron of *MAG*	0	1	N/D	N/D	N/D
17	442	23424841-23425282	intergenic	1	1	N/D	N/D	N/D
18	1056	38032581-38033636	9th intron of *SLC7A9*	1	0	N/D	N/D	N/D
19	739	9309172-9309910	1-2nd introns, 1st exon of *ZNF559*	2	1	N/D	N/D	N/D
20	1231	33710585-33711815	5th intron of *BC068609*	0	0	N/D	N/D	N/D

Note: ND, no data; NA, not applicable.

* According to Human Genome Browser [52], assembly of March 2006.

** Change of expression of the nearby gene (folds) in proliferating compared to differentiating HL-60 cells.

2). Group number 4 contains the largest number of identified binding sites and includes sequence ID numbers 1-7, 10, 15, 16, and 18-20. These were all located within introns of characterized human genes (Table 2), as previously shown for several Myc-regulated genes [26-29]. Interestingly, four identified genes (three of them with intronic location of the Max binding sites) belong to the superfamily of genes encoding zinc finger containing proteins. Despite the fact that these are over-represented on human chromosome 19 [30], this proportion is much higher than expected from a random gene selection. Sequences belonging to all mentioned groups were selected for further analysis.

Binding of Max to the identified sequences. In order to evaluate binding of the identified sequences to Max, eleven sequences were analyzed by EMSA. All of these were specifically recognized by GST-Max *in vitro*. The data are summarized in Table 2 and a representative

experiment for sequence ID number 1, 2, 4, and 14 is shown in Fig. 3.

To further confirm that the sequences identified were recognition sites for the Max protein *in vivo*, we performed chromatin immunoprecipitation assay (ChIP) [31]. Fragmented interphase chromatin DNA bound by the Max protein *in vivo* were isolated by precipitation with α-Max antibodies and used as template for PCR amplification with primers specific for the selected sequences (Table 1). We employed the human cell line HL-60 for these experiments since different proteins of the Myc/Max/Mxd network are expressed differently depending on the state of these cells. c-Myc is expressed during logarithmic growth whereas Mxd1 is up-regulated during differentiation [19, 32, 33]. As a positive control, we used primers for *CCND2*, which have previously been shown to be targeted by Max, Myc, Mxd, and Mnt in ChIP [15, 34]. Unfortunately, we did not succeed in finding a fragment that is not able to bind Max in HL-60 cells (negative control).

Sequences 1-6, 8, and 11 were analyzed by ChIP, and 3-6, 8, and 11 were found among the fragments bound to the Max proteins *in vivo* (Fig. 4). As seen, the intensity of PCR bands for all the fragments is close or exceeded that of the positive control. At the same time, we were unable to find conditions for efficient amplification of fragments 1 and 2 (Table 2), possibly due to their enrichment with $(TG)_n$ simple repeat.

Expression of putative Myc/Max/Mxd regulated genes. Based on the results for Max binding to the identified sequences, we were able to select several genes that could be potentially regulated by the Myc/Max/Mxd network. The expression levels of two genes with upstream Max-binding site (sequence ID numbers 3 and 8), one with downstream Max-binding site (sequence ID number 11), and three with intronic Max-binding sites (sequence ID numbers 1, 2, and 6) were compared by SYBR Green quantitative real-time PCR in proliferating HL-60 cells and in cells induced to granulocyte differentiation by treatment with DMSO [19]. PCR primers were designed to amplify exon sequences of the potential Myc/Max/Mxd target genes (Table 1). β-Actin, which is not regulated by the Myc network proteins and is expressed at high level in most growing cells [35], was used as loading control. The results are summarized in Table 2. For the *MAG* gene (sequence ID number 2, see Table 2), no expression was detected in either proliferating or DMSO-treated HL-60 cells. However, this finding is not surprising since the *MAG* gene is expressed almost exclusively in nervous tissues [36]. In addition, no difference in expression was detected for the ecotropic viral integration site 5-like (*EVI5L*) gene. The genes *FFAR2* (*GPR43*) [37, 38], *LOC55908/DOCK6* were up-regulated in HL-60 cells treated with DMSO (Table 2). In contrast, genes encoding the zinc finger proteins ZNF649 and ZNF221 were moderately up-regulated in proliferating HL-60 cells. This is in accord with the fact that binding of Max may either enhance or suppress transcription of the target genes [8, 39].

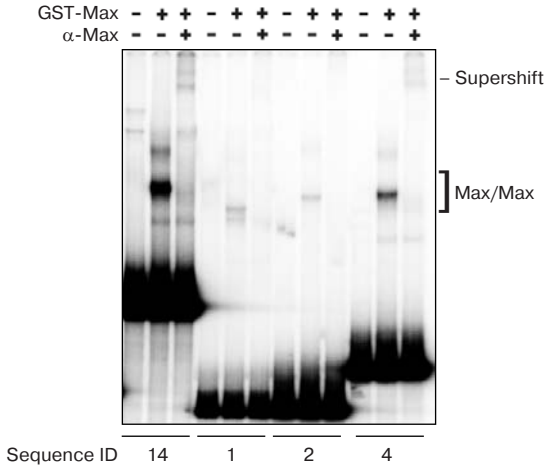


Fig. 3. *In vitro* DNA-binding of Max/Max homodimers to the identified DNA sequences by electrophoretic mobility shift assay (EMSA). Representative data are shown for the sequences with ID number 1, 2, 4, and 14 (see Table 2). For supershift, Max antibodies were added as described in "Materials and Methods".

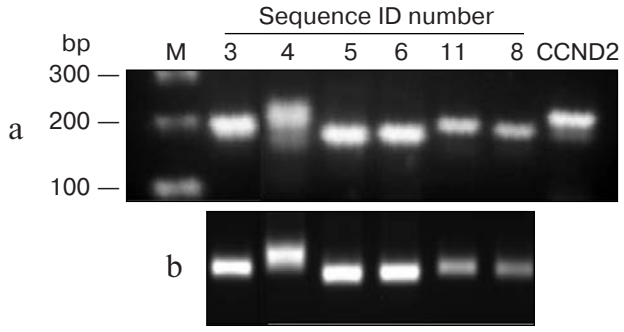


Fig. 4. Binding of Max to identified sequences *in vivo*. Anti-Max antibody was used to immunoprecipitate Max-bound chromatin fragments using ChIP assay. a) Representative data are shown from experiments where primers to the sequences with ID number 3-6, 8, and 11 (see Table 2) were used for PCR amplification on the immunoprecipitated DNA. b) The same but on the input genomic DNA template. Primers for cyclin D2 (*CCND2*) were used as a positive control.

DISCUSSION

Binding of the Myc/Max/Mxd network proteins to E-box sequences in the control region of target genes is a prerequisite for transcriptional regulation. To date, about 4000 Myc target genes have been identified [40] and the question still remains whether all are direct targets. To understand more about the functions of Myc/Max/Mxd network, identification and validation of direct and indi-

rect target genes as well as evaluation of their specificity for Myc function are needed. The majority of the potential Myc/Max/Mxd network targets have been identified through localization of the E-box sequence in their regulatory regions and reporter gene assay (selected references are [41-43]). In addition, selection procedures like chromatin immunoprecipitation and computer-assisted search have been employed [17, 44, 45]. However, the approaches based on localization of the E-box sequence in regulatory regions have limitations since the number of all possible E-box sequences in the genome is more than 1,000,000, which is much higher than the actual number of binding sites for the Myc/Max/Mxd proteins.

With this in mind, Max protein binding should be limited to the specific fraction of the E-boxes that (i) have the appropriate flanking sequences which facilitate formation of a specific DNA structure or binding of other proteins; (ii) are located within chromatin regions with structure favorable for binding; (iii) are not inactivated, e.g. by methylation [46]. In our study, the data were obtained under conditions similar to those used for EMSA where influence of chromatin structure and methylation or other modification of DNA was excluded and the necessity of specific flanking sequences for efficient binding of Max was supported. Most of the identified fragments bound by Max contained more than one E-box sequence, sometimes more than 10 (Table 2), and exact localization of the Max-bound E-boxes within them require further analysis. Moreover, at least one of the identified fragments, bound by Max in EMSA, did not contain any CACGTG or CATGTG sequences, which suggests that Max can also bind other, non-canonical sequences. Indeed, it has been shown that the Mnt/Max heterodimers bind non-canonical E-box sequence CACGCG with higher affinity compared to the canonical CACGTG E-box [47, 48]. It is noteworthy that our strategy, which is based on identifying regulatory sequences bound by Max, allows us to identify putative target genes for the entire Myc/Max/Mxd network.

With the use of RepeatMasker, we analyzed the presence of different repeats in the selected fragments and compared their content with that of the human genome [1]. General content (by the length of the sequence) of all types of repeats in our sequences was 49.9%, which is slightly more than in the human genome (46.4%). However, the content of specific types of repeats was considerably different. The content of both Alu (0.9 against 10% in genome) and LINE (16 against 21%) repeated DNAs was reduced, and the content of human endogenous retroviruses (15 against 8%) and, especially, simple repeats (18 against 2%) was increased. The higher content of clones containing simple repeats can be possibly explained by the enrichment of the selected pool with these sequences during hybridization (see "Results").

Of the genes identified, the *FFAR2* (*GPR43*) gene [38], a member of G-protein coupled receptors superfamily, was most pronouncedly induced in differentiated HL-60 cells (Table 2). The *FFAR2* gene is located within a region that is looped out into the nuclear halo in proliferating cells. However, upon DMSO induced differentiation the looped fragment becomes associated with the nuclear matrix, and this change in the spatial organization correlated with transcriptional activation of the *FFAR2* gene [49]. In addition, it has been shown that the mouse *FFAR2* homolog, *LSSIG* is induced in M1 leukemia cells when stimulated by the leukemia inhibitory factor (LIF) to differentiate into macrophages [50]. The reasons for the detected up-regulation of *FFAR2* in DMSO-treated cells is not clear, but one explanation could be that these genes are also regulated by other, not related to Myc/Max/Mxd network, transcription factors. Thus, further investigations are needed to clarify the regulation of this gene.

The zinc finger proteins 649 (*ZNF649*) and 221 (*ZNF221*) genes were shown to be up-regulated in proliferating cells HL-60 cells compared to cells induced to differentiate. ZNF649 has been suggested to act as a transcriptional repressor in mitogen-activated protein kinase signaling pathways to mediate cellular functions. In addition, over-expression of *ZNF649* in cells was shown to inhibit the transcription factors SRE and AP-1 [51]. One possibility is therefore that some of the large number of genes that have been shown to be affected by Myc could be regulated indirectly via zinc finger proteins and their regulatory activity.

In conclusion, the approach used in this study could with minor modifications be applied for identification of target genes regulated by other transcription factors, both in selected chromosomes and from the whole genome.

We are grateful to A. Carrano and colleagues from Lawrence Livermore National Laboratory (USA) for the human chromosome 19 library, to Bernhard Lüscher for the GST-Max expression plasmid, and to B. O. Glotov for critical reading of the manuscript.

This work was supported by grants from the Swedish Royal Academy of Science (M. H.), the Swedish Research Council (M. H.), and the Russian Foundation for Basic Research (01-04-48264) (L. N). S. Akopov, I. Chernov, and T. Wahlström were all recipients of fellowships from Cancer Research Institute and Concern Foundation, USA.

REFERENCES

1. Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., et al. (2001) *Nature*, **409**, 860-921.
2. Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., Smith, H. O., Yandell, M., Evans, C. A., Holt, R. A., et al. (2001) *Science*, **291**, 1304-1351.

3. Waterston, R. H., Lindblad-Toh, K., Birney, E., Rogers, J., Abril, J. F., Agarwal, P., Agarwala, R., Ainscough, R., Alexandersson, M., An, P., et al. (2002) *Nature*, **420**, 520-562.
4. Carninci, P. (2006) *Trends Genet.*, **22**, 501-510.
5. Nikolaev, L. G., Akopov, S. B., Chernov, I. P., and Sverdlov, E. D. (2007) *Curr. Genom.*, **8**, 137-149.
6. Henriksson, M., and Luscher, B. (1996) *Adv. Cancer Res.*, **68**, 109-182.
7. Luscher, B. (2001) *Gene*, **277**, 1-14.
8. Grandori, C., Cowley, S. M., James, L. P., and Eisenman, R. N. (2000) *Annu. Rev. Cell. Dev. Biol.*, **16**, 653-699.
9. Hurlin, P. J., and Huang, J. (2006) *Semin. Cancer Biol.*, **16**, 265-274.
10. Wahlstrom, T., and Henriksson, M. (2007) *Adv. Cancer Res.*, **97**, 61-80.
11. Luscher, B., and Larsson, L. G. (1999) *Oncogene*, **18**, 2955-2966.
12. O'Hagan, R. C., Schreiber-Agus, N., Chen, K., David, G., Engelman, J. A., Schwab, R., Alland, L., Thomson, C., Ronning, D. R., Sacchettini, J. C., et al. (2000) *Nat. Genet.*, **24**, 113-119.
13. Hurlin, P. J., Zhou, Z. Q., Toyo-oka, K., Ota, S., Walker, W. L., Hirotsune, S., and Wynshaw-Boris, A. (2003) *Embo J.*, **22**, 4584-4596.
14. Nilsson, J. A., and Cleveland, J. L. (2004) *Cell Cycle*, **3**, 588-590.
15. Popov, N., Wahlstrom, T., Hurlin, P. J., and Henriksson, M. (2005) *Oncogene*, **24**, 8326-8337.
16. Walker, W., Zhou, Z. Q., Ota, S., Wynshaw-Boris, A., and Hurlin, P. J. (2005) *J. Cell Biol.*, **169**, 405-413.
17. Zeller, K. I., Zhao, X., Lee, C. W., Chiu, K. P., Yao, F., Yustein, J. T., Ooi, H. S., Orlov, Y. L., Shahab, A., Yong, H. C., et al. (2006) *Proc. Natl. Acad. Sci. USA*, **103**, 17834-17839.
18. Nikolaev, L. G., Tsevegiyn, T., Akopov, S. B., Ashworth, L. K., and Sverdlov, E. D. (1996) *Nucleic Acids Res.*, **24**, 1330-1336.
19. Xu, D., Popov, N., Hou, M., Wang, Q., Bjorkholm, M., Gruber, A., Menkel, A. R., and Henriksson, M. (2001) *Proc. Natl. Acad. Sci. USA*, **98**, 3826-3831.
20. Sambrook, J., and Russell, D. W. (eds.) (2001) *Molecular Cloning. A Laboratory Manual*, 3rd Edn., CSHL Press, Cold Spring Harbor.
21. Bousset, K., Henriksson, M., Luscher-Firzlaff, J. M., Litchfield, D. W., and Luscher, B. (1993) *Oncogene*, **8**, 3211-3220.
22. Chernov, I. P., Akopov, S. B., Nikolaev, L. G., and Sverdlov, E. D. (2002) *J. Cell. Biochem.*, **84**, 590-600.
23. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) *Nucleic Acids Res.*, **25**, 3389-3402.
24. Kent, W. J., Sugnet, C. W., Furey, T. S., Roskin, K. M., Pringle, T. H., Zahler, A. M., and Haussler, D. (2002) *Genome Res.*, **12**, 996-1006.
25. Walhout, A. J., Gubbels, J. M., Bernards, R., van der Vliet, P. C., and Timmers, H. T. (1997) *Nucleic Acids Res.*, **25**, 1493-501.
26. Tsuneoka, M., Nakano, F., Ohgusu, H., and Mekada, E. (1997) *Oncogene*, **14**, 2301-2311.
27. Gaubatz, S., Meichle, A., and Eilers, M. (1994) *Mol. Cell. Biol.*, **14**, 3853-3862.
28. Greasley, P. J., Bonnard, C., and Amati, B. (2000) *Nucleic Acids Res.*, **28**, 446-453.
29. Horikawa, I., Cable, P. L., Afshari, C., and Barrett, J. C. (1999) *Cancer Res.*, **59**, 826-830.
30. Shannon, M., Ashworth, L. K., Mucenski, M. L., Lamerdin, J. E., Branscomb, E., and Stubbs, L. (1996) *Genomics*, **33**, 112-120.
31. Orlando, V. (2000) *Trends Biochem. Sci.*, **25**, 99-104.
32. Larsson, L. G., Pettersson, M., Oberg, F., Nilsson, K., and Luscher, B. (1994) *Oncogene*, **9**, 1247-1252.
33. Sommer, A., Bousset, K., Kremmer, E., Austen, M., and Luscher, B. (1998) *J. Biol. Chem.*, **273**, 6632-6642.
34. Bouchard, C., Dittrich, O., Kiermaier, A., Dohmann, K., Menkel, A., Eilers, M., and Luscher, B. (2001) *Genes Dev.*, **15**, 2042-2047.
35. Femino, A. M., Fay, F. S., Fogarty, K., and Singer, R. H. (1998) *Science*, **280**, 585-590.
36. Konat, G. W. (1996) *Acta. Neurobiol. Exp.*, **56**, 281-285.
37. Sawzdargo, M., George, S. R., Nguyen, T., Xu, S., Kolakowski, L. F., and O'Dowd, B. F. (1997) *Biochem. Biophys. Res. Commun.*, **239**, 543-547.
38. Brown, A. J., Goldsworthy, S. M., Barnes, A. A., Eilert, M. M., Tcheang, L., Daniels, D., Muir, A. I., Wigglesworth, M. J., Kinghorn, I., Fraser, N. J., et al. (2003) *J. Biol. Chem.*, **278**, 11312-11319.
39. Schreiber-Agus, N., and DePinho, R. A. (1998) *Bioessays*, **20**, 808-818.
40. Dang, C. V., O'Donnell, K. A., Zeller, K. I., Nguyen, T., Osthus, R. C., and Li, F. (2006) *Semin. Cancer. Biol.*, **16**, 253-264.
41. Mac, S. M., D'Cunha, C. A., and Farnham, P. J. (2000) *Mol. Carcinog.*, **29**, 76-86.
42. Ben-Porath, I., Yanuka, O., and Benvenisty, N. (1999) *Mol. Cell. Biol.*, **19**, 3529-3539.
43. Kyo, S., Takakura, M., Taira, T., Kanaya, T., Itoh, H., Yutsudo, M., Ariga, H., and Inoue, M. (2000) *Nucleic Acids Res.*, **28**, 669-677.
44. Grandori, C., Mac, J., Siebelt, F., Ayer, D. E., and Eisenman, R. N. (1996) *Embo J.*, **15**, 4344-4357.
45. Lee, L. A., and Dang, C. V. (2006) *Curr. Top. Microbiol. Immunol.*, **302**, 145-167.
46. Prendergast, G. C., and Ziff, E. B. (1991) *Science*, **251**, 186-189.
47. Hurlin, P. J., Queva, C., and Eisenman, R. N. (1997) *Genes Dev.*, **11**, 44-58.
48. Meroni, G., Reymond, A., Alcalay, M., Borsani, G., Tanigami, A., Tonlorenzi, R., Nigro, C. L., Messali, S., Zollo, M., Ledbetter, D. H., et al. (1997) *Embo J.*, **16**, 2892-2906.
49. Iarovaia, O. V., Akopov, S. B., Nikolaev, L. G., Sverdlov, E. D., and Razin, S. V. (2005) *Nucleic Acids Res.*, **33**, 4157-4163.
50. Senga, T., Iwamoto, S., Yoshida, T., Yokota, T., Adachi, K., Azuma, E., Hamaguchi, M., and Iwamoto, T. (2003) *Blood*, **101**, 1185-1187.
51. Yang, H., Yuan, W., Wang, Y., Zhu, C., Liu, B., Yang, D., Li, Y., Wang, C., Wu, X., and Liu, M. (2005) *Biochem. Biophys. Res. Commun.*, **333**, 206-215.
52. Kuhn, R. M., Karolchik, D., Zweig, A. S., Trumbower, H., Thomas, D. J., Thakkapallayil, A., Sugnet, C. W., Stanke, M., Smith, K. E., Siepel, A., et al. (2007) *Nucleic Acids Res.*, **35**, D668-D673.