5050–5058 Nucleic Acids Research, 1993, Vol. 21, No. 22

Single exchanges of amino acids in the basic region change the specificity of N-Myc

Thomas Feldmann, Regina Alex, Jörg Suckow, Renate Dildrop, Brigitte Kisters-Woike and Benno Müller-Hill*
Institut für Genetik der Universität zu Köln, Weyertal 121, 50931 Köln, Germany

Received August 13, 1993; Revised and Accepted September 28, 1993

ABSTRACT

We exchanged specific amino acids in the basic region of the murine N-Myc protein and tested the mutant proteins for their DNA binding specificity. The amino acids we exchanged were chosen in analogy to residues of the homologous basic regions of bHLH and bZIP proteins. Mutant N-Myc peptides were expressed in *Escherichia coli* and specific DNA binding was monitored by gel shift experiments. For this we used palindromic target sequences with systematic base pair exchanges. Several mutants with altered DNA binding specificity were identified. Amino acid exchanges of residues −14 or −10 of the basic region lead to specificity changes (we define leucine 402 of N-Myc as +1; comparable to GCN4 see (1)). The palindromic N-Myc recognition sequence 5'CACGTG is no longer recognized by the mutant proteins, but DNA fragments with symmetrical exchanges of the target sequence are. Exchanges at position −15 broaden the binding specificity. These data were used to build a computer based model of the putative interactions of the N-Myc basic DNA binding region with its target sequence.

INTRODUCTION

The N-Myc protein belongs to the family of Myc-oncoproteins (2, 3) which play a role in differentiation, organogenesis and cell-cycling (4–6). Expression of N-Myc protein is essential for the development of the CNS, lung, kidney and heart. Homozygous disruption of the N-*myc*-gene in the mouse leads to embryonic lethality (5, 6).This indicates that the N-Myc protein cannot be functionally replaced by other members of the Myc-family.

Deregulated expression of the N-*myc*-gene, mostly due to gene amplification or retroviral insertion (7), has been observed in a restricted set of tumors of neuroendocrine- or embryonic origin (8, 9). In transgenic mice, which carry the N-*myc* transgene under the control of immunoglobulin enhancer sequences the high level of N-Myc expression leads to the formation of lymphoid tumors (10, 11).

The Myc-oncoproteins and other eucaryotic transcription factors like the upstream stimulatory factor Usf, the transcription factors TfeB and Tfe3 or the muscle determination factor MyoD, share a common DNA-binding and dimerization domain, the basic-helix-loop-helix (bHLH) motif (12–15). The basic region of this motif mediates DNA binding. The two helices, which are interrupted by a loop-structure are responsible for homo- or heterodimerization of the bHLH proteins (16). A number of bHLH proteins, including the Myc proteins possess two or three leucine heptad repeats, also called leucine-zipper (LZ), immediately C-terminal to the second helix of the bHLH domain (14, 17). The leucine zipper has firstly been identified in a class of transcription factors, the bZip-proteins (18), where it drives homo- or heterodimerization.

All bHLH proteins recognize DNA sequences belonging to the hexameric E-box motif 5'CANNTG. The target sequence specifically recognized by the bHLH domain of the N-Myc (19, 20), c-Myc (21–23) and L-Myc (20) protein has been determined to be 5'CACGTG. The full-length Myc-proteins cannot form active, DNA-binding homodimeric complexes (24, 26), but bacterially expressed truncated Myc-proteins, which possess the complete bHLH-LZ-domain, bind easily to the hexameric target sequence 5'CACGTG (19, 21, 23). The Myc-proteins form *in vitro* and *in vivo* heterodimeric complexes with another bHLH protein which has been named MAX (17). Myc-MAX heterodimers as well as MAX homodimers bind to the same target sequence as the truncated, homodimeric Myc-proteins (17).

Most of the so far known prokaryotic and eukaryotic DNA binding proteins use short α-helices, which cross the major groove to contact the DNA. Prominent examples are the recognition helices of the prokaryotic Helix-turn-helix motifs and the structurally related eukaryotic homeodomains, the basic regions of the bZip proteins and the short α-helices of the zinc-finger proteins (26). To get more information about the complex structure of the contacting α-helices and the recognized basepairs of the target-DNA, different experimental approaches are possible: X-ray or NMR-studies, biochemical studies or genetic analysis. The genetic approach requires the generation of loss-of-function mutations within the DNA contacting α-helix of the protein. Some of these mutations can be complemented by basepair changes in the target sequence. The definition for a change of specificity is as follows: the mutated protein must bind to the variant target sequence better than the wildtype protein, and the wildtype protein should not bind the variant target

* To whom correspondence should be addressed

sequence or should at least bind it less well than the mutant protein. The restored binding activity of the mutated protein to the mutated target sequence identifies functional interactions between amino acids and the target sequence but cannot differentiate between direct contacts (contacts between bases and amino acids) and indirect contacts. Genetic analysis performed with Lac-repressor (27, 28) and GCN4 (1, 29) are examples, how these experiments can provide and add structural information to X-ray or NMR data (30).

The basic regions of various bHLH proteins are highly similar, and there is also, albeit to a lesser degree, homology detectable between the basic regions of bZip- and bHLH-proteins (Figure 1). This observation and mutation experiments performed with the yeast transcription factor CBF1 (31) and the human protein E47 (32) prompted us to introduce amino acid exchanges in 3 positions of the basic region of the N-Myc-bHLH domain. The DNA binding activity of mutated proteins with natural and altered target sequences were tested *in vitro* in electrophoretic mobility shift assays. We used the results to build a computer based model for the sequence specific interactions between the basic region of N-Myc and its target sequence. As we carried out the experiments, the X-ray data of the related MAX-bHLH domain bound to its target sequence were published (33). Our genetic data and our computer model are in agreement with and confirm these data with respect to the DNA-binding of the bHLH-domain.

MATERIAL AND METHODS

Chemicals and enzymes

Restriction enzymes were purchased from New England Biolabs (Bad Schwalbach, Germany) and Boehringer (Mannheim, Germany), vent-polymerase from New England Biolabs (Bad Schwalbach, Germany) and T_7-polymerase kit from Pharmacia (Freiburg, Germany). The enzymes were used according to the suppliers instructions. (^{32}P)-deoxyribonucleotides were obtained from Amersham Buchler (Braunschweig, Germany), the chemicals for automated DNA synthesis from Applied Biosystems (Pfungstadt, Germany), all other chemicals from Sigma (München, Germany) or Merck (Darmstadt, Germany).

Oligonucleotides were synthesised on an Applied Biosystems Synthesizer model 894 and purified on denaturing polyacrylamide gels.

Construction of mutant GST-N-Myc genes

We used a previously described *N-myc* gene fragment fused in frame to the 3'-terminus of the glutathione-s-transferase (GST) gene for the construction of mutant *N-myc* genes (19). The c-terminal four amino acids of Lac-repressor were added to the carboxy-terminus of the N-Myc peptide to increase the stability of the expressed fusion proteins in *E.coli*. To obtain a control mutant lacking DNA binding activity, we exchanged the codons of six amino acids in positions −10 to −15 of the basic region for alanine codons by PCR-mutagenesis (34). In the same step we introduced upstream of the altered basic region an *Aat* II restriction site and downstream an *Afl* II restriction site. For this we used the following oligonucleotide primers : primer 1: 5'GCT.GCT.AAC.GAC.GAC.TTA.AGG.TCC.AGC.TTC.-CTG.ACG.CTC.AGG.3' and primer 2: 5'GGC.CGC.CGC.-GGC.GAT.GTT.GTG.GTT.GCG.ACG.TCG.CTC.GC3'. We used the *Aat* II and *Afl* II restriction sites (underlined) for cloning double stranded oligonucleotides into the vector. The oligonucleotides were phosphorylated with polynucleotide kinase (NEB) prior to the PCR reaction using a standard protocol (35). The PCR-reaction was performed on a Perkin Elmer Cetus Thermal DNA Cycler using vent-polymerase (NEB) under the following conditions: 10 ng GST-N-Myc-expression vector was incubated in vent-polymerase reaction buffer (NEB) supplemented with 20 pmoles of primer 1 and 2 each, 15 mM $MgSO_4$ and 20 mM dATP, dCTP, dGTP and dTTP. After 25 cycles, the PCR product was purified with the Quiaex-Kit (DIAGEN) and ligated. After cleaving the mutant expression vector (N-Myc AAAAAA) with *Aat* II and *Afl* II, pairs of synthetic, annealed oligonucleotides (Figure 2) were ligated into the vector. The mutant plasmids were sequenced with a T_7-polymerase sequencing kit (Pharmacia).

Protein synthesis and purification

E.coli K12 strain BL21DE3 (36) was transformed with the various expression constructs and grown in NZCYM medium (35). Purification of the GST-N-Myc wildtype and GST-N-Myc mutant fusion proteins was performed as described (37). Purified proteins were dialysed against MTPBS (150 mM NaCl, 16 mM Na_2HPO_4, 4 mM NaH_2PO_4; pH 7.3) and stored in MTPBS with 10% glycerol. Protein concentrations were determined by the Bradford method (38). The purity of the preparations was estimated from coomassie blue stained SDS-PAGE (39).

Gel retardation assay

For gel shifts we used 19 synthetic self complementary oligonucleotides. They were 23 bases long and were annealed to obtain double stranded DNA (27). The DNA fragments differed symmetrically at positions 1, 2 or 3 of a halfsite (counted from the center of symmetry):

```
                3 2 1    1' 2' 3'
5'-GGCCCATTC    C A C *  G T G   GAATGGGC-3'
3'- CGGGTAAG    G T G *  C A C   CTTACCCGG-5'
                3' 2' 1'  1 2 3
```

In 16 of the 19 fragments position 2 and 3 were symmetrically exchanged with all four nucleotide bases to obtain all possible combinations. In the remaining three variants only the inner base pairs were exchanged. The oligonucleotides were endlabeled using polynucleotide kinase (NEB) (35).

Binding reactions for retardation assays were performed with 2 ng of endlabeled oligonucleotides (10 000 cpm/ng), 40 ng protein and 20 ng poly (dI-dC) as non-specific competitor DNA, in 20 µl binding buffer (10 mM Tris/HCl pH 7.6, 10% glycerol, 1 mM EDTA, 2 µg BSA, 33 mM KCl). The mixtures were incubated for 15 min at room temperature and then loaded onto 5% running polyacrylamide gels. Electrophoresis was performed in 0.5×TBE for 1.5 hours at 5V/cm. Prior to loading, gels were prerun for 4 hours at 14V/cm. The gels were dried on Whatman 3MM paper and autoradiographed with Kodak XAR film at −70°C.

RESULTS

The choice of mutants: sequence comparison of N-Myc with various bZIP andbHLH proteins

In order to decide which positions in the basic region of N-Myc are candidates for specific target recognition we aligned the N-Myc amino acid sequence with the sequences of the bZIP proteins GCN4, C/EBP and TAF1 (Figure 1), for which the amino acids involved in specific DNA recognition are already known (1, 40).

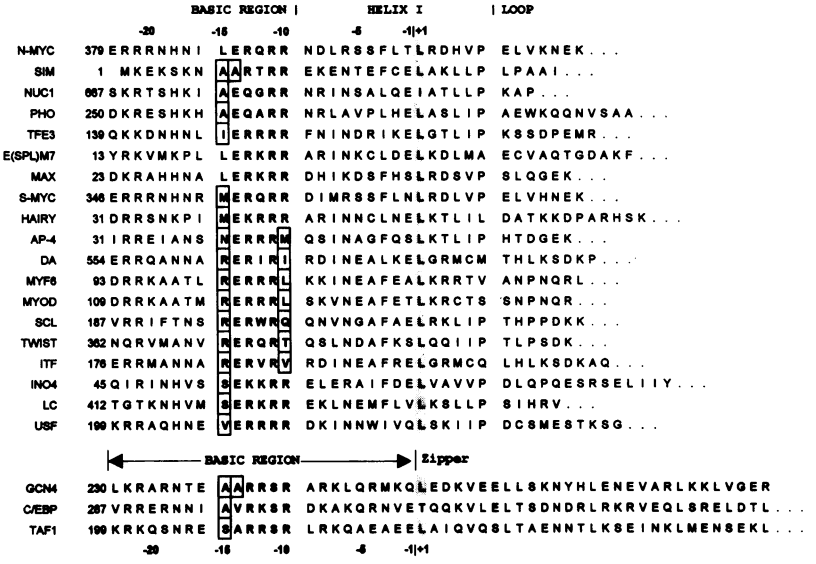


Figure 1. Amino acid sequences of the basic regions of bHLH and bZIP proteins are shown in single letter code. The numbers at the beginning of each protein sequence give the position of the first amino acid in our sequence according to the references. Alignment of bHLH and bZIP was performed to obtain best fit of the DNA binding amino acids. The region in which amino acids were exchanged in N-Myc is represented in bold letters. The numbering above and below the protein sequences corresponds to the numbering used for bZIP proteins (1). This numbering was adapted to bHLH proteins to denominate N-Myc mutants with changed amino acids in order to facilitate comparisons. The hatched leucines at position +1 represent the starting point of numbering. We have chosen the amino acid exchanges according to the boxed amino acids at positions −15, −14 and −10. References are as follows (if available the SwissProt accession number is given): N-MYC (Accession number: P03966), SIM (Accession number: P05709, (41)), NUC1 (Accession number: P20824), PHO (Accession number: P07270), TFE3 (Accession number: P19532), E(SPL)M7 (Accession number: P13097), MAX (Accession number: P25912), S-MYC (Accession number: P23999), HAIRY (Accession number: P14003), AP−4 (47), DA (Accession number: P11420), MYF6 (Accession number: P23409), MYOD (Accession number: P10085), SCL (Accession number: P22091), TWIST (Accession number: P10627), ITF (Accession number: P15806), INO4 (Accession number: P13902), LC (Accession number: P13526), USF (Accession number: P22415), GCN4 (Accession number: P03068, P03069), C/EBP (18), TAF1 (48).

After comparing the sequences, we decided to alter the amino acids at position −15, −14 and −10 (see Figure 1). We define leucine 402 of N-Myc as +1. The numbering corresponds to the numbering used for bZIP proteins (1). The importance of amino acid −10 for the recognition of the inner base pair of the palindromic target sequence 5′CANNTG of different bHLH-ZIP proteins had been reported (31, 32). The sequence alignment suggested residues −15 and −14 as most promising candidates for additional specific interactions with the target DNA (Figure 1). The other known bHLH proteins contain either Leu, Ala, Ile, Met, Asn, Arg, Ser or Val at position −15 ; Glu or Ala at position −14, and Arg, Ile, Leu, Met, Asn, Gln, Thr or Val at position −10 (Figure 1). All of these amino acids were subsequently used for single exchanges. As a positive control we exchanged all three positions simultaneously and replaced the residues of N-Myc by the corresponding residues of AP-4. The $Asn_{-15}Arg_{-12}Met_{-10}$ (NERRRM) mutant should specifically recognize the target sequence of AP-4: 5′CAGCTG (41). We also exchanged Glu_{-14} for Lys, an amino acid which is not present at this position in any of the so far known bHLH proteins (Figure 1).

The test system

The GST-fusion proteins were expressed in *E.coli* and purified by affinity chromatography on glutathione-agarose beads. SDS-PAGE allowed an estimation of the molecular weights of the fusion proteins. Although we observed partial degradation of the GST-N-Myc proteins, approximately half of the total protein in the preparation has an apparent molecular weight of 38 kD, which corresponds to the calculated size of the intact fusion protein.

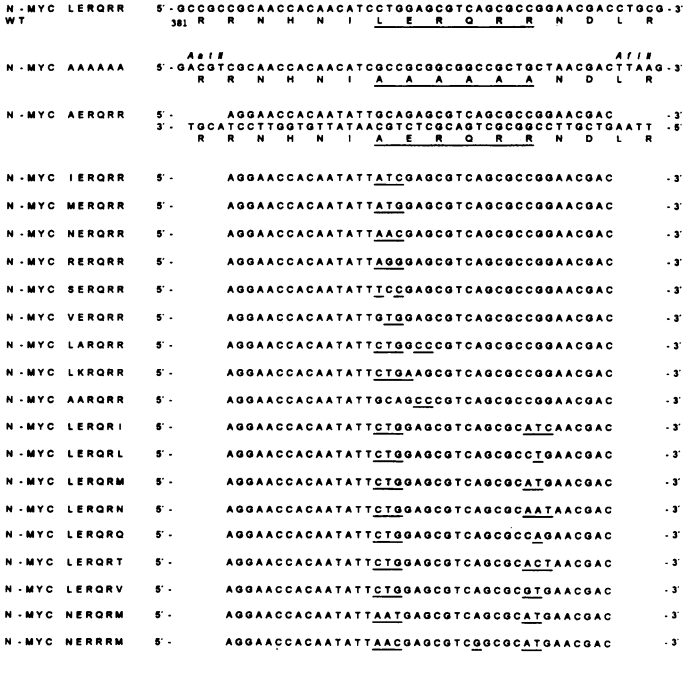
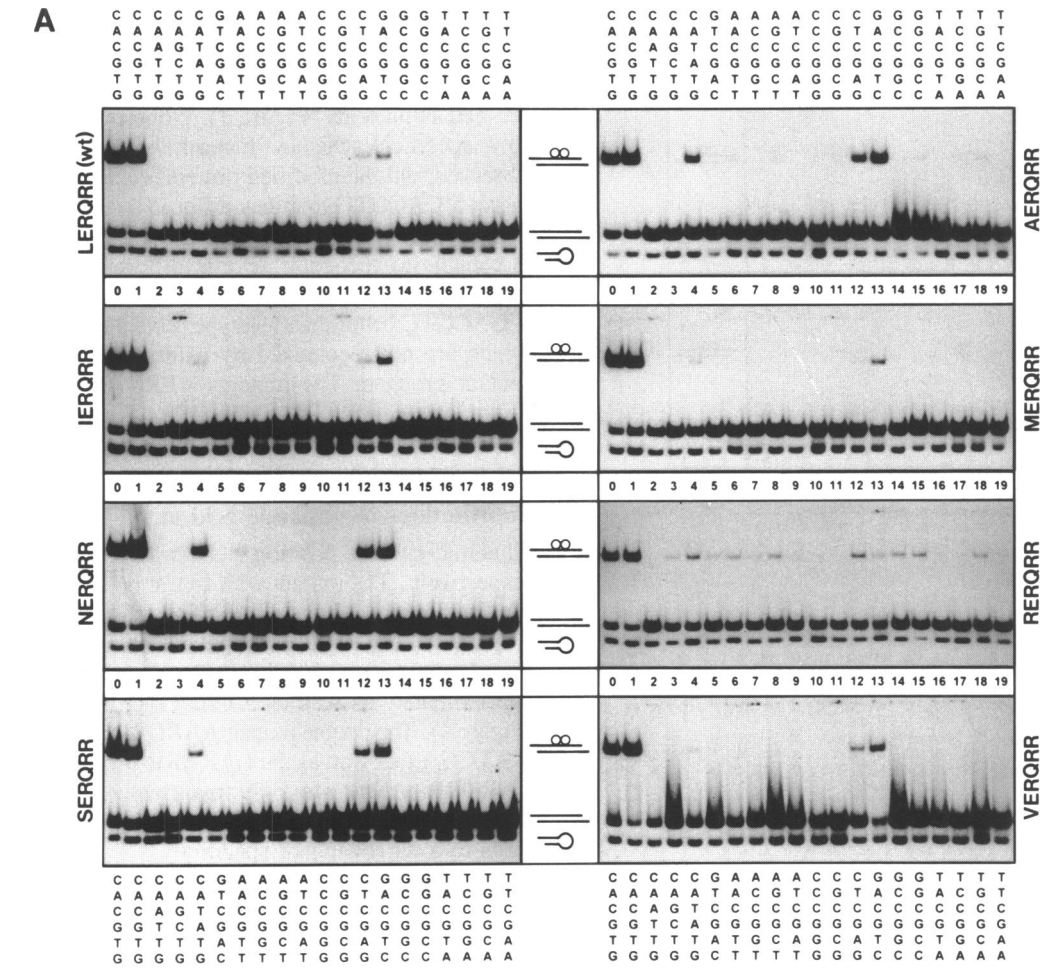


Figure 2. Sequence of the DNA coding for the N-Myc wildtype protein (LERQRR), the non-DNA binding mutant (AAAAAA), and of all oligonucleotides leading to amino acid exchanges of N-Myc mentioned here. The regions of interest are underlined. Amino acids are represented in the single letter code. *Aat* II and *Afl* II restriction sites used for cloning are marked in the N-Myc AAAAAA mutant. As an example for the double stranded oligonucleotides cloned into the vector, the sequence which converts N-Myc wildtype to the mutant AERQRR is shown. All other oligonucleotides are represented by the coding strands in comparison to AERQRR. The exchanged bases are underlined.



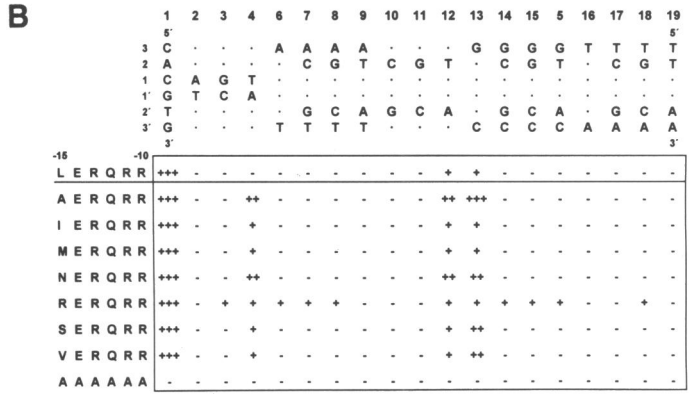


Figure 3. A. Autoradiographs of gel retardation assays with purified N-Myc-GST fusion protein and mutants with exchanges at position −15. As a reference gel, a retardation with wildtype N-Myc (LERQRR) is shown. The wildtype target 5′CACGTG was shifted with wildtype N-Myc as a positive control on each retardation gel with mutated proteins (lane 0). The DNA structures (hairpin or double stranded free DNA) and the protein−DNA complexes that cause the respective bands are symbolized between the autoradiographs. The sequences of the targets are indicated at the top and bottom of the figure. The numbers between the autoradiographs identify the lanes and the names of the used N-Myc mutants are written to the left or right of each corresponding autoradiography. The faint bands which are sometimes visible at the top of the autoradiographs correspond to radioactive material retained in the slots. **B.** Summary of the results shown in Figure 3A. Left column: Relevant protein sequences of the N-Myc protein and its variants. The first line gives the sequence of wildtype N-Myc (LERQRR). The bottom line shows the non-DNA binding mutant (AAAAAA). The mutants in the intervening lines carry substitutions at position −15. Upper panel: The numbers indicate the lanes of Figure 3A, the sequences represent the N-Myc target (E-box motif) and its variants. For the complete DNA sequence of the target fragments see Material and Methods. Only the bases which differ from wildtype N-Myc target are indicated. Lower right panel: Quantification of the individual complexes between target and protein. The relative amounts of shifted DNA fragments by mutant proteins are always compared to the positive control on the same gel. The amounts of shifted DNA are classified as follows: '+++': retardation is as good or almost as good as the retardation of wildtype target 5′CACGTG by wildtype N-Myc protein; '++': the retardation reaches 70−30% of the wildtype binding; '+': the retardation reaches 30−5% of wildtype binding; '−': less than 5% binding i.e. no retardation detectable.

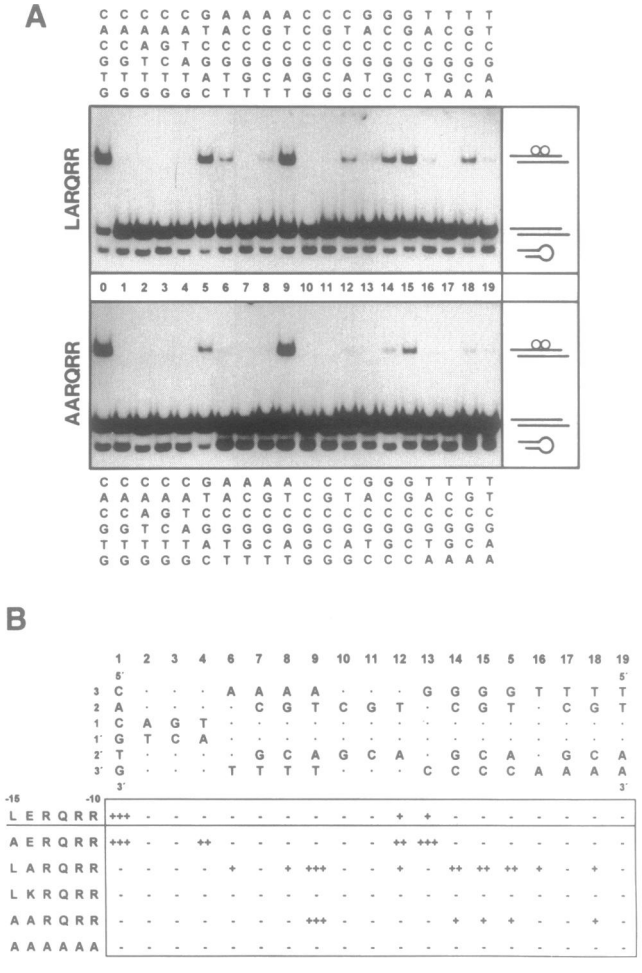


Figure 4. A. Autoradiographs of gel retardation assays with purified N-Myc mutants with exchanges at position −14. On each retardation gel the wildtype target 5′CACGTG was shifted with wildtype N-Myc as a positive control (lane 0). The DNA structures (hairpin or double stranded free DNA) and the protein−DNA complexes that cause the respective bands are symbolized to the right of the autoradiographs. Numbering is the same as in figure 3A. **B.** Summary of the results of figure 4A. Wildtype protein (LERQRR) and non-DNA binding protein (AAAAAA) results are shown in the first and last line as references. For comparison the results obtained with the mutant AERQRR (see Figure 3A) are included. For other details see figure 3B.

We first tested the assay system with the unaltered purified fusion protein (LERQRR) as positive control and the non-DNA binding protein (AAAAAA) as negative control. The unaltered wildtype protein showed the expected binding to the known N-Myc target 5′CACGTG (19) and minor interactions with the targets 5′CTCGAG and 5′GACGTC (Figure 3). The negative control did not bind to any of the DNA targets (Figure 3B).

The binding activities of the different mutants were estimated from two independent experiments. On each retardation gel the wildtype target 5′CACGTG was shifted with wildtype N-Myc as a positive control. The relative amounts of shifted DNA fragments complexed with mutant proteins were always compared with the positive control on the same gel. The amounts of shifted DNA are classified as follows: '+++': retardation is as good or almost as good as the retardation of wildtype target 5′C-ACGTG by wildtype N-Myc protein; '++': the retardation reaches 70−30% of the wildtype binding; '+': the retardation amounts to 30−5% of wildtype binding; '−': less then 5% binding i.e. no retardation detectable.

Substitutions of leucine in position −15

We decided to substitute leucine −15 for Ala, Ile, Met, Asn, Arg, Ser and Val after comparing the DNA binding domains of the bHLH-proteins (Figure 1). None of these exchanges leads to a specificity change. Instead, broadening of specificity is observed. All the modified proteins recognize well the wildtype target 5′CACGTG, but they differ in their ability to bind to other DNA sequences. Especially the mutant with Arg at position −15 is capable to recognize several different targets. Some of the mutants are able to bind to the sequences 5′CTCGAG and 5′GACGTC better than wildtype protein (Figure 3). Some targets which are not recognized by wildtype protein are retarded by mutant proteins: The mutants AERQRR, NERQRR and to a minor extent IERQRR, MERQRR, RERQRR, SERQRR and VERQRR bind to 5′CATATG (Figure 3) a sequence which is not recognized by wildtype N-Myc.

Substitutions of glutamic acid at position −14

Glutamic acid at position −14 was replaced by Ala or Lys respectively. The mutant with Lys at position −14 has no DNA binding activity, whereas LARQRR leads to a specificity change (Figure 4). The wildtype target is not retarded by this mutant, but the sequence 5′ATCGAT is recognized by LARQRR approximately as well as 5′CACGTG by the wildtype protein (Figure 4). The double mutant AARQRR recognizes the sequence 5′ATCGAT as well as the LARQRR-protein but all other active LARQRR-targets are only marginally bound by AARQRR (Figure 4). Thus the introduction of a second amino acid exchange leads to a narrowing of specificity.

Substitution of arginine at position −10

The substitution of Arg at position −10 by Leu, Met, Asn or Thr also leads to specificity changes (Figure 5). The specific recognition is shifted from 5′CACGTG to 5′CATATG. No other targets are bound by these mutants. N-Myc proteins with Ile, Gln or Val in position −10 are unable to recognize any DNA target at all (Figure 5).

Substitution of the basic region of N-Myc by the basic region of AP−4

For further analysis of the specificity determinants of N-Myc and AP−4 we exchanged the LERQRR-sequence of N-Myc by the NERRRM-sequence of AP−4. This hybrid protein retards the target sequences 5′CAGCTG and CATATG. The target 5′CATATG is recognized better than 5′CAGCTG (Figure 6). The double mutant NERQRM is only able to bind the target 5′CATATG (Figure 6). In our test system we cannot achieve a change of specificity from the N-Myc target 5′CACGTG to the AP4-target 5′CAGCTG by a single amino acid exchange (Arg at position −10 to Met). Even the double mutant NERQRM is not able to recognize the AP−4 target. Only the triple mutant NERRRM can bind this target. But even this protein does not exhibit a higher affinity for 5′CAGCTG than for the sequence 5′CATATG. The single amino acid substitution of Leu_{-15} by Asn_{-15} does not change the specificity at all as mentioned before (Figure 6).

DISCUSSION

The test system

The use of purified GST fusion proteins for gel shift assays is a well established method to determine the binding properties of proteins (19, 42). The wildtype N-Myc-fusion protein showed

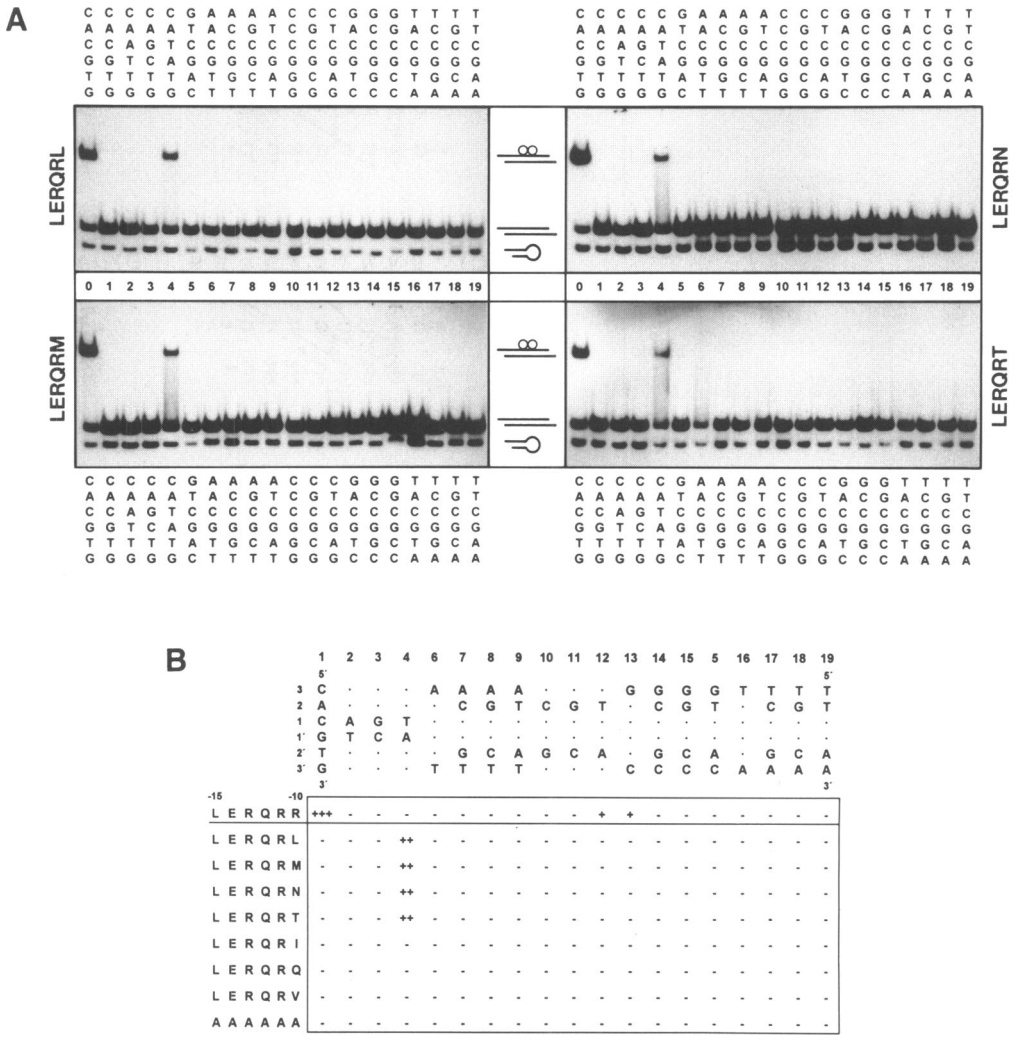


Figure 5. A. Autoradiographs of gel retardation assays with purified N-Myc mutants exchanges at position −10. On each retardation gel the wildtype target 5′C-ACGTG is shifted with wildtype N-Myc as a positive control (lane 0). The DNA structures (hairpin or double stranded free DNA) and the protein−DNA complexes that cause the respective bands are symbolized between the autoradiographs. The numbering is the same as in figure 3A. Autoradiographs of gels with non-DNA binding mutants (LERQRI, LERQRQ and LERQRV) are not shown. **B.** Summary of the results of figure 5A. As reference the results of wildtype (LERQRR) and the negative control are given at top and bottom lines, respectively. For numbering and further details see legend of figure 3B.

the expected binding to the target 5′CACGTG. Minor interactions with the targets 5′CTCGAG and 5′GACGTC are additionally detectable. As expected, the mutant N-Myc-AAAAAA does not bind any of the offered targets.

Amino acid exchanges at position −15 do not lead to an altered specificity

The results of all substitutions at position −15 (Figure 3) indicate that no change of specificity for the wildtype target 5′CACGTG occurs. The minor interactions of the wildtype N-Myc protein to the targets 5′CTCGAG and 5′GACGTC are conserved by the mutant proteins and even enhanced by Ala_{-15}, Asn_{-15}, Ser_{-15} and Val_{-15}. The mutant proteins Ala_{-15}, Arg_{-15}, Asn_{-15}, Ile_{-15}, Met_{-15}, Ser_{-15} and Val_{-15} additionally bind to the target 5′CATATG. According to the crystal structure of MAX, the recognition of an A:T pair at position 1 of the target cannot be due to a direct contact between the amino acid at position −15 of the protein and the nucleotide bases. We suggest that the protein−DNA interface may be flexible and this may result in the recognition of adenine 1′ of the first and thymine 1 of the other halfside by Arg_{-10}. Thus it seems that leucin at position −15 constrains the wildtype protein to exclude indirectly N-Myc binding to the mutant target 5′CATATG. The protein RERQRR exhibits a weak but broad binding to several targets (Figure 3). This might be due to direct amino acid-basepair contacts of Arg to the basepairs at position 1 and 2. Moreover, Arg in position −15 might stabilize complexes by DNA backbone contacts.

Alanine at position −14 leads to an altered specificity

One of the known bHLH proteins, called sim (43) bears alanine residues at positions −15 and −14 (Figure 2). In analogy to sim we exchanged Glu at position −14 for Ala and observed a specificity change. The wildtype N-Myc target 5′CACGTG is no longer retarded, the preferred target is now 5′ATCGAT. Some other targets are weakly bound (Figure 4A). With the double mutant AARQRR we can spot a narrowing of specificity in comparison to LARQRR, but 5′ATCGAT is still bound by AARQRR as strong as by LARQRR. This change of specificity may be reconciled with the X-ray structure of the MAX−DNA complex where Glu_{-14} is shown to make specific contacts to

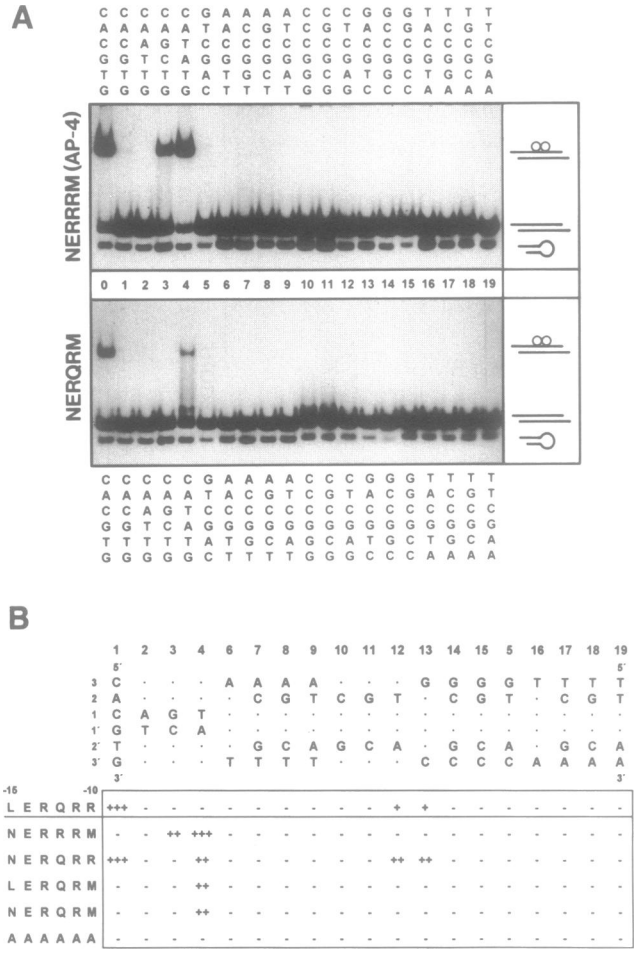


Figure 6. A. Autoradiographs of the double mutant NERQRM and the triple mutant NERRRM. Numbering and symbols are the same as in figure 4A. **B.** Summary of the results of figure 6A. As references the results obtained with wildtype (LERQRR) and non-DNA binding protein (AAAAAA) are shown in the top and bottom lines. The mutants NERQRR (Figure 3) and LERQRM (Figure 5) are included for comparison. For further explanations see legend to figure 3B.

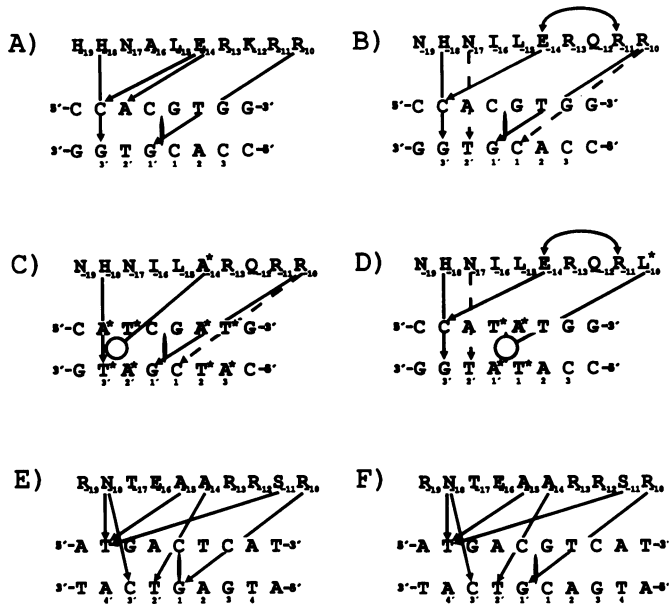


Figure 7. Schematic summary of DNA contacts from MAX, N-Myc and mutants with amino acid exchanges at position −14 and −10. The DNA contacts of GCN4 to its symmetric and asymmetric target are summarised for comparison (as adapted from (1, 40) Contacts of amino acid side chains to nucleotide bases are represented by straight, continuous arrows, circles indicate putative hydrophobic interactions. Neighbourhoods between amino acids and nucleotide bases are represented by broken arrows. Interactions between Glu_{-14} and Arg_{-11} are shown as bend, continuous arrows between these amino acids. **A.** The specific contacts between amino acids and nucleotide bases of the MAX−DNA cocrystal are as previously described (33). **B.** The model of protein−DNA interactions of N-Myc is built on the basis of the crystal data of MAX and the genetic data deduced for N-Myc. **C.** Putative interactions of the mutant LARQRR carrying Ala in position −14. **D.** Putative interactions of the position −10 mutant LERQRL to the target 5′CATATG. **E.** The model of protein−DNA interactions on the basis of the X-ray data of GCN4 (40). **F.** Model of DNA−protein interactions of GCN4 to a symmetrisized target (1, 40) The position of the axis of dyad symmetry is indicated by a black ellipsoid. Substituted amino acids and nucleotide bases are marked by asterisks.For E and F please notice the convenient DNA-target numbering which differs from (1, 40).

cytosine 3 of the target. We suggest that an alanine in position −14 can be involved in hydrophobic interactions with thymines 2 and 3′ which replace adenine 2 and guanine 3′ in the mutated target 5′ATCGAT. Although at first sight alanine seems to be too small to reach any base pair, it seems plausible that either the protein or the target DNA possess sufficient flexibility for the necessary spatial adjustment. Phasing studies have demonstrated DNA bending by bHLH proteins (44) and suggest that bending angles of up to 80° may be achieved. In a bended DNA−protein complex hydrophobic contacts between Ala_{-14} and the methyl groups of thymines 2 and 3′ may occur.

The LARQRR protein seems to be not only less restricted in its specificity than the AARQRR protein, but in some cases the respective complexes are more stable. This might reflect an increase of the hydrophobic interface of LARQRR in comparison to AARQRR. Depending upon the target sequence which is recognized, minor adaptations of the protein relative to the DNA surface might happen. This flexibility is only possible with alanine, not with glutamic acid in position −14. The wildtype glutamic acid is probably not only involved in the specific target recognition. Modelling suggests that it can also capture the side chain of Arg_{-11} in a fixed orientation. Comparable Asp-Arg interactions are known from other DNA−protein structures (45). According to the X-ray data of the MAX−DNA complex (33) this interaction may be expected for the N-Myc complex too, since Arg_{-11} together with Glu_{-14} is extremely well conserved in bHLH proteins (Figure 1). With alanine in position −14 the Glu-Arg interaction is lost and the protein surface should concomitantly gain flexibility for a different DNA-recognition.

Amino acid exchanges at position −10 lead to a specificity change

We investigated the binding properties of mutants with Ile, Leu, Met, Gln, Thr, Asn and Val at position −10 (Figure 5). The amino acids Ile, Gln and Val abolish binding of the N-Myc mutants to any tested DNA. Upon introduction of Leu, Met, Asn or Thr in place of Arg_{-10} we only observed binding to the DNA target 5′CATATG, but the complex is less stable than the wildtype complex. These results imply that amino acid residue −10 interacts with the central base pairs 1 and possibly 1′ and thus are in agreement with the data from the X-ray structure concerning the role of amino acid −10 (33).

From previous attempts to determine residues that were involved in specific contacts, it was already known that a substitution of Arg_{-10} by Val abolishes the binding, but a double mutant in which Leu_{-15} was exchanged for Arg and Arg_{-10} for Val had been reported to possess altered specificity. This C-Myc mutant recognizes 5'CAGCTG instead of 5'CACGTG (32).

Similar experiments performed with the basic region of the AP−4 protein also resulted in a specificity change (31). The AP−4 protein recognizes the target 5'CAGCTG. When Met_{-10} is replaced by Arg the Myc target 5'CACGTG is recognized by the AP−4 protein. We performed the reciprocal experiment and introduced a Met instead of Arg in position −10 but the observed specificity was changed from 5'CACGTG to 5'CATATG. We did not observe any binding to the target 5'CAGCTG (Figure 6A). Modelling indicates that the arginine in position −10 not only forms hydrogen bonds with guanine 1', but seems to interact beyond the axis of target symmetry with cytosine 1. We would like to discuss the change of specific recognition as a result of hydrophobic interactions between the amino acid in position −10 and the thymine 1. Only in the case of methionine in position −10 thymine 1 on the other strand might contribute to the stabilization of the complex. Our results confirm that the amino acid in position −10 of the N-Myc protein is the major determinant for the specific recognition of the inner base pair of the target sequence.

A single amino acid exchange from Arg to Met at position −10 is not sufficient to change specificity from 5'CACGTG to 5'CAGCTG

Because of the outcome of the LERQRM exchange we decided to construct and test the double mutant NERQRM and the triple mutant NERRRM, where residues −15, −12 and −10 of N-Myc are replaced by the corresponding residues of AP−4. Residues −14, −13 and −11 are identical in both proteins (Figure 2). Only the triple mutant is able to recognize the targets 5'CATATG and 5'CAGCTG (Figure 6A). The double mutant still binds to 5'CATATG but does not recognize the AP−4 target 5'CAGCTG (Figure 6A). These results suggest that either Gln_{-12} of N-Myc or Arg_{-12} of AP−4 is involved in specific recognition, at least in an indirect manner. In agreement with the X-ray structure of MAX and our own results we propose changes in the topology of the protein−DNA complex caused by the different amino acid sequences of AP−4 and N-Myc. Due to the different amino acid composition of the basic regions both α-helices might recognize their targets in a slightly different manner. Such effects have been previously described for bZip proteins (46).

Comparison of the bZIP and bHLH basic regions reveals a strong similarity

By comparing the amino acid sequence of the basic region of N-Myc with the bZIP proteins GCN4, C/EBP and TAF1, a certain sequence analogy becomes obvious (Figure 1), especially if one looks at the positions responsible for specific DNA contacts (Figure 7). The specific interactions of GCN4 have been described previously (1, 29, 40). They are established by protein residues −18, −15, −14, −11 and −10. In the MAX−DNA cocrystal, specific interactions are described for residues −18, −14 and −10 (33). If one compares the interactions of the amino acid side chains of GCN4 and N-Myc similar contacts can be seen. In both cases Arg_{-10} contacts guanine of the inner basepair. Furthermore we propose for the N-Myc mutant LARQRR hydrophobic contacts to the base pairs 2 and 3 of the target ATCGAT. This contact is similar to the interaction of the amino acid Ala_{-14} of GCN4 to thymine 2 of the symmetrisized GCN4 target. Finally the amino acid at position −18 of both GCN4 and N-Myc contacts base pair 3 of their specific target. The other interactions of GCN4 to its target sequence have no analogy in N-Myc, which may be due to the fact that GCN4 has to interact to a 8 base pair recognition sequence in contrast to N-Myc which binds to a 6 base pair target sequence.

ACKNOWLEDGEMENTS

We thank S.K.Burley for access to the crystal data of the Max−DNA complex. We thank B.von Wilcken-Bergmann and S.Oehler for discussion. This work was supported by the Deutsche Forschungsgemeinschaft through SFB 243 and SFB 274 and by the BMFT 'Molekulare Gen- und Zelltechnologie'.

REFERENCES

1. Suckow, M., von Wilken-Bergmann, B. and Müller-Hill, B. (1993) *EMBO J.* , 12, 1193−1200
2. Kohl, N.E., Kanda, N. Schreck, R.R., Bruns, G., Latt, S.A., Gilbert F. and Alt, F.W. (1983) *Cell* , 35, 359−367
3. Schwab, M., Alitalo, K., Klempnauer, K., Varmus, H.E., Bishop, J.M., Gilbert F., Brodeur, G.M., Goldstein, M. and Trent, J. (1983) *Nature*, 305, 245−248
4. Lüscher, B. and Eisenman, R.N. (1990) *Genes Dev.*, 4, 2025−2035
5. Stanton, B.R., Perkins, A.S., Tessarollo, L., Sassoon, D.A. and Parada, L.F. (1992) *Genes Dev.*, 6, 2235−2247
6. Moens, C.B., Auerbach, A.B., Conlon, R.A., Joyner, A.L. and Roussant, J. (1992) *Genes Dev.*, 6, 691−704
7. van Lohuizen, M., Breuer, M. and Berns, A. (1989) *EMBO J.*, 8, 133−136
8. Kohl, N.E., Gee, C.E. and Alt, F.W. (1984) *Science*, 226, 1335−1337
9. Lee, W.H., Murphree, A.L. and Benedict, W.F. (1984) *Nature*, 309, 458−460
10. Rosenbaum, H., Webb, E., Adams, J.M., Cory, S. and Harris, A.W. (1989) *EMBO J.*, 8, 749−755
11. Dildrop, R., Ma, A., Zimmerman, K., Hsu, E., Tesfaye, A., Depinho, R. and Alt, F.W. (1989) *EMBO J.*, 8, 1121−1128
12. Beckmann, H., Su, L.-K. and Kadesch, T. (1990) *Genes Dev.* , 4, 167−179
13. Carr, C.S. and Sharp, P.A. (1990) *Mol. Cel. Biol.*, 10, 4384−4388
14. Gregor, P.D., Sawadogo, M. and Roeder, R.G. (1990) *Genes Dev.*, 4, 1730−1740
15. Murre, C., Schonleber McCaw, P. and Baltimore, D. (1989) *Cell*, 56, 777−783
16. Davis, R.L., Cheng, P.-F., Lassar, A.B. and Weintraub, H. (1990) *Cell*, 60, 733−746
17. Blackwood, E.M., and Eisenman, R.N. (1991) *Science*, 251, 1211−1217
18. Landschulz, W.H., Johnson, P.F., Adashi, E.Y., Graves. B.J. and McKnight, S.L. (1988) *Genes Dev.*, 2, 786−800
19. Alex, R., Sözeri, O., Meyer, S. and Dildrop, R. (1992) *Nucleic Acids Res.*, 20, 2257−2263
20. Ma, A., Moroy, T., Collum, R., Weintraub, H., Alt, F.W. and Blackwell, T.K. (1993) *Oncogene*, 8, 1093−1098
21. Blackwell T.K. and Weintraub, H. (1990) *Science*, 250, 1104−1110
22. Kerkhoff, E., Bister, K and Klempnauer, K.H. (1991) *Proc. Natl. Acad. Sci. USA*, 88, 4323−4327
23. Prendergast, G.C. and Ziff, E.B. (1991) *Science*, 251, 186−189
24. Amati, B., Dalton, S., Brooks, M.W., Littlewood, T.D., Evan, G.I. and Land, H. (1992) *Nature*, 359, 423−429
25. Kato, G.J., Lee, W.M.F., Chen, L. and Dang, C.V. *Genes. Dev.* (1992), 6, 81−92
26. Pabo, C.O. and Sauer, R.T. (1992) *Annu. Rev. Biochem.*, 61, 1053−1095
27. Lehming, N., Sartorius, J., Kisters-Woike, B., von Wilcken-Bergmann, B. and Müller-Hill, B. (1990) *EMBO J.*, 9, 615−621
28. Lehming, N., Sartorius, J., Kisters-Woike, B., von Wilcken-Bergmann, B.. and Müller-Hill, B. (1991) In: Eckstein, F. and Lilley, D.M.J. (eds), *Nucleic Acids and Molecular Biology*. Springer Verlag Berlin, 5, 114−125

29. Suckow, M., von Wilcken-Bergmann, B. and Müller-Hill, B. (1993) *Nucleic Acids Res.*, 21, 2081–2086
30. Kisters-Woike, B., Lehming, N., Sartorius, J., von Wilcken-Bergmann, B. and Müller-Hill, B. (1991) *Eur. J. Biochem.*, 198, 411–419
31. Dang, C.V., Barret, J., Villa-Garcia, M., Resar, L.M., Kato, G.J. and Fearon, E.R. (1991) *Mol. Cell. Biol.*, 11, 954–962
32. Halazonetis, T.D. and Kandil, A.N. (1992) *Science*, 255, 464–466
33. Ferré-d'Amaré, A.R., Prendergast, G.C., Ziff, E.B. and Burley, S.K. (1993) *Nature,* 363, 38–45
34. Hemsley, A., Arnheim, N., Toney, M.D., Cortopassi, G. and Galas, D.J. (1989) *Nucleic Acids Res.*, 17, 6545–6551
35. Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular Cloning 2. Ed.*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York
36. Studier, F.W. and Moffatt, B.A. (1986) *J. Mol. Biol.*, 189, 113–130
37. Smith, D.B. and Johnson, K.S. (1988) *Gene*, 67, 31–40
38. Bradford, M.M. (1976) *Anal. Biochem.* , 72, 248–254
39. Laemmli, U.K.(1970) *Nature.* , 227, 680–685
40. Ellenberger, T.E., Brandl, C.J., Struhl, K. and Harrison, S.C. (1992) *Cell*, 71, 1223–1237
41. Mermod, D., Williams, T.J. and Tjian, R. (1988) *Nature*, 332, 557–561
42. Blackwell, T.K., Kretzner, L., Blackwood, E.M., Eisenman, R.N. and Weintraub, H. (1990) *Science*, 250, 1149–1151
43. Nambu, J.R., Lewis, J.O., Wharton, K.A.jr, Crews, S.T. (1991) *Cell*, 67, 1157–1167
44. Fisher, D.E., Parent, L.A. and Sharp, P.A. (1992) *Proc. Natl. Acad. Sci.*, 89, 11779–11783
45. Pavletich, N.P. and Pabo, C.O. (1991) *Science*, 252, 809–817
46. Kim, J., Tzamarias, D., Ellenberger, T., Harrison, S.C. and Struhl, K. (1993) *Proc. Natl. Acad. Sci.*, 90, 4513–4517
47. Hu, Y.-F.,Lüscher, B.,Admon, A.,Mermod, N. and Tjian, R. (1990) *Genes Dev.*, 4, 1741–1752
48. Oeda, K., Salinas, J. and Chua, N.-H. (1991) *EMBO J.*, 10, 1793–1802