Biophysical Characterization of the c-Myb DNA-Binding Domain†

Andreas Ebneth,‡ Olaf Schweers,§ Hubert Thole,|| Ursula Fagin,‡ Claus Urbanke,‡ Günther Maass,‡ and Heiner Wolfes*,‡

Zentrum Biochemie, Medizinische Hochschule Hannover, Konstanty-Gutschow-Strasse 8, D-30623 Hannover, Germany, Max Planck Arbeitsgruppe für strukturelle Molekularbiologie, Notkestrasse 85, D-22607 Hamburg, Germany, and Max Planck Institut für experimentelle Endokrinologie, Feodor-Lynen Strasse 4, D-30625 Hannover, Germany

Received October 28, 1993; Revised Manuscript Received September 8, 1994⊗

ABSTRACT: We have examined proteins containing the DNA-binding domain of c-Myb with biophysical methods. This DNA-binding domain consists of three imperfect repeats (R1, R2, and R3) conserved among many species. Our results indicate that the DNA-binding domain forms unspecific and specific complexes with oligodeoxynucleotides. In the presence of R1, DNA sequences related to a canonical c-Myb-binding site are better discriminated. Furthermore, although R2 and R3 are sufficient for sequence-specific DNA binding, a structural change of the DNA-binding domain upon specific complex formation is induced only when R1 is present. Therefore, R1 might serve as an important element required for secondary structure alteration upon binding and its stabilization as well as for better discrimination between specific and related DNA sequences.

The chicken c-myb protooncogene is the cellular progenitor of the v-myb oncogene, first identified as the transforming component of the two avian leukemia viruses AMV and E26 [for a review, see Lüscher and Eisenman (1990)]. The protooncogene c-myb encodes a DNA-binding protein of approximately 75 kDa. C-Myb is expressed predominantly in undifferentiated cells of the hematopoietic lineage (Westin, 1982). It was demonstrated that v-Myb and c-Myb proteins recognize specifically the DNA sequence YAACKG (Biedenkapp et al., 1988) and activate transcription in mammalian cells (Klempnauer et al., 1989; Nakagoshi et al., 1989). It thus appears that they might function as sequence-specific transcriptional regulatory proteins: mim-1, a gene identified recently (Ness et al., 1989), which has three c-Myb-binding sequences in the 5′ promoter region, was shown to be directly activated by c-Myb proteins. Myb homologs have been identified in many species (Bender & Kuehl, 1986; Gerondakis & Bishop, 1986; Rosson & Reddy, 1986; Katzen et al., 1985; Paz-Ares et al., 1987); the main feature of all these homologs is a high degree of conservation in a region required for specific DNA binding. This region consists of 3 imperfect repeats of ~51 amino acids. In each of these repeats, 3 tryptophans are spaced by 18 or 19 amino acids. Since the spacing of the rare amino acid tryptophan is conserved evolutionary, it was speculated that the DNA-binding domain of c-Myb proteins is organized in a unique stereochemical arrangement (Anton & Frampton, 1988). Mutation analysis of this region revealed that the first repeat of the tripartite region is dispensable for DNA binding, whereas the second and third repeats are essential (Howe et al., 1990; Oehler et al., 1990). Furthermore, it could be demonstrated that the tryptophans play a major role in stabilizing the structure of the DNA-binding motif (Kanei-Ishii et al., 1990). Recently, a model derived from NMR data was proposed for the structure−function relationship of repeat 3 (Ogata et al., 1992; Jamin et al., 1993). According to this model, the third repeat harbors three α-helices, of which two are forming a helix−turn−helix-related motif maintained by the conserved tryptophans which form a hydrophobic core together with other hydrophobic amino acids. In contrast, repeat 2 appears to be more flexible, consisting of two α-helices; the flexible part might alter its secondary structure upon binding to DNA (Jamin et al., 1993; Sarai et al., 1993).

This suggestion is endorsed by a recent publication which used fluorescence emission spectroscopy to study the quench of the tryptophan signals upon binding of DNA to c-Myb (Myrset et al., 1993). Furthermore, it was demonstrated (Tanikawa et al., 1993) that repeat 3 recognizes the core sequence AAC very specifically, whereas the second repeat contacts the DNA only weakly and repeat 1 stabilizes the c-Myb−DNA complex (Ording et al., 1994).

In order to study the interaction of c-Myb proteins with DNA in detail, we have expressed different constructs harboring the three repeats and the last two repeats of the DNA-binding region of c-Myb in *Escherichia coli* and characterized their binding properties with biophysical methods. We report here that a structural change of the DNA-binding domain (i.e., a 10% increase of α-helical content) is detectable only when the entire DNA-binding domain (including repeat 1) is bound to a canonical DNA sequence.

EXPERIMENTAL PROCEDURES

Oligodeoxynucleotides. All synthetic oligodeoxynucleotides and cloning linkers were synthesized on a Milligen Cyclone DNA synthesizer. The reagents were purchased from MWG-Biotech, Ebersberg, and the acetonitrile was from AMBIX, München.

Construction of c-Myb Expression Vectors. For the construction of the vectors pR123, pR123Δ5, and pR23,

† This work has been supported by a grant to A.E. from the Deutsche Forschungsgemeinschaft (No. 120/6-1).

* To whom correspondence should be addressed.
‡ Medizinische Hochschule Hannover.
§ Max Planck Institut für Strukturelle Molekularbiologie.
|| Max Planck Institut für experimentelle Endokrinologie.

⊗ Abstract published in *Advance ACS Abstracts,* November 1, 1994.

0006-2960/94/0433-14586$04.50/0

containing the DNA-binding repeats 1−3 and 2−3, respectively, fused to the glutathione *S*-transferase, the c-myb gene (Rosson & Reddy, 1986) was digested with *Not*I and *Bsp*MII and cloned with *Bam*HI linkers d(GATCCCCGGCGGC), d(GGGCCGCCGCCGG), d(CCGGAATTCAGCTGAATGAATGAATGAG), and d(TTAAGTCGACTTACTTACTCCTG), into the pGex2T vector (Pharmacia) to yield p-Myb. Starting from this full-length clone, pR123, spanning mainly the DNA-binding domain, was constructed by digesting p-Myb with *Pst*I and incubating with T7 DNA polymerase to generate blunt ends. This preparation was digested with *Bam*HI; the fragment was isolated and recloned into a *Bam*HI/*Sma*I-digested pGex2T plasmid. Plasmid pR123Δ5, which expresses a protein lacking the last five amino acids of the third repeat, was derived from p-Myb by digestion with *Eco*RI and ligation. Plasmid pR23 contains repeats 2 and 3 of the DNA-binding domain and was prepared by PCR cloning using the oligodeoxynucleotides d(GGGGGATCCCTTATCAAAGGTCCATGGACTAA) and d(CCCAGATCTCTTCCGGCGCTGGTGGAATTCCA) as primers and p-Myb as a template. The PCR product was digested with *Bam*HI and *Bgl*II and inserted into the *Bam*HI site of pGex2T. All constructs were sequenced using a modified dideoxy sequencing protocol (Geiger et al., 1989).

Preparation of Proteins. Overnight cultures of pR123, pR23, or pR123Δ5 in *E. coli* strain LK 111 (λ) (Zabeau & Stanley, 1982) were diluted 1:20 in 8 L of Luria Bertani medium and grown to a turbidity of $A^{600} = 1$ at 37 °C. At this stage, expression was induced by adding IPTG to a final concentration of 0.1 mM for 4 h. The cells were harvested, resuspended in 400 mL of PBS (10 mM sodium phosphate, pH 7.5, and 130 mM NaCl), and stored at −20 °C. The frozen cells were thawed and sonified twice in a Branson sonifier for 5 min; 1% Triton X-100 was added, and the lysates were centrifuged for 30 min at 13000*g*. Five milliliters of GSH−Sepharose 4B (Pharmacia), gel volume, equilibrated with PBS, was added to the clear supernatant and incubated for 1 h with gentle shaking at room temperature. The solution was centrifuged at 2000*g* for 2 min, the supernatant decanted, and the pellet washed twice with TBS (10 mM Tris-HCl, pH 7.5, 1 mM EDTA, and 150 mM NaCl). After equilibrating with cleavage buffer (TBS/3 mM $CaCl_2$), the fusion protein was cleaved with 100 units of thrombin (Sigma) overnight at 4 °C in a total volume of 15 mL of cleavage buffer. The slurry was filtered and washed with 15 mL of PBS. The cleaved-off c-Myb protein was bound on a DNA−cellulose (Litman, 1968) column (gel bed volume 10 mL, linear gradient of 0.5 × PBS, 75 mM−1.5 M NaCl) to remove DNA associated with it. The protein eluted at approximately 600 mM NaCl; fractions were pooled, concentrated in Centricon microconcentrators, and stored in PEDL [30 mM KP_i, pH 7.2, 1 mM EDTA, 10 mM DTT, and 0.01% (w/v) Lubrol].

This procedure yielded about 1−2 mg of protein which was estimated to be about 98% pure as judged by SDS−polyacrylamide gel electrophoresis. The concentration was determined by UV absorption spectroscopy. The extinction coefficients at 280 nm, ϵ(R123) = 55 500 L mol^{-1} cm^{-1} and ϵ(R23) = 34 400 L mol^{-1} cm^{-1}, empirically determined (Levine & Federici, 1982), were used for the calculation. We took special care that the reducing agent DTT was present in at least 0.1 mM concentration in all steps of the protein purification as well as in all biophysical experiments, to prevent oxidation of Cys130, which was shown to play an important role in DNA binding (Guehmann et al., 1992; Myrset et al., 1993).

Protein Sequencing. Aliquots (30 μL) of peak fractions from DNA−cellulose chromatography were applied to polybrene-coated glass−fiber filters and sequenced by Edman degradation on an Applied Biosystems 477 A sequenator. The phenylthiohydantoin amino acids were analyzed on-line with an Applied Biosystems 120 A narrow-bore HPLC system.

Electrophoretic Mobility Shift Assay (EMSA). The oligodeoxynucleotides listed in Table 1 were labeled with T4 polynucleotide kinase and [γ-^{32}P]ATP (Amersham) according to standard procedures, annealed at 75 °C by mixing equimolar amounts of the complementary strands, and slowly cooled to room temperature. The oligodeoxynucleotides or the 418 bp PCR product was incubated with R123 or R23 in EMSA-binding buffer [10 mM Tris-HCl, pH 7.6, 10 mM DTT, 1 mM EDTA, 5 mM $MgCl_2$, 100 mM NaCl, and 10% glycerol (v/v)] for at least 30 min at room temperature. EMSA was carried out on 20 × 20 cm, 6% polyacrylamide gels in a 0.5 × TBE (50 mM Tris−borate, pH 8.0, and 2 mM EDTA) gel at room temperature at 100 V. The gels were dried and scanned with a Fuji BAS 1000 Mac Bas Phospho-Imager. The binding isotherms were analyzed with a computer program (TITRAT; Dr. C. Urbanke, unpublished results) using a multistep predictor/corrector method (Curth et al., 1993).

PCR Shift Template. A double-stranded oligodeoxynucleotide containing a c-Myb-binding sequence, d(GATCAGAGTTAACGGTTTTTC), was ligated into the *Bam*HI site of pAT153 (Twigg & Sherrat, 1980). Polymerase chain reaction (PCR) was carried out in a TPS 5.4 Thermocycler (Landgraf) with the primers d(TGCGGGATATCGTCCATTCCGACAGCATCGCCAGT) and d(GGTAGAGGAACGTACGTGGTAAGGA). PCR reaction conditions were as follows: 4 nM template DNA, 400 nM of each primer, 0.2 mM dNTP, 10 μCi of [α-^{32}P]dATP, 5 μL of 10 × *Taq* buffer, and 2 units of *Taq* polymerase (Amersham) in a reaction volume of 50 μL; cycle 1 (1×) 390 s at 92 °C; cycle 2 (30×) 90 s at 92 °C, 90 s at 54 °C, 210 s at 72 °C; cycle 3 (1×) 360 s at 72 °C. The labeled product was separated on a 20 × 20 cm, 6% polyacrylamide gel in 0.5 × TBE, radioactive bands were excised, and the 418 bp fragments were eluted with water.

Circular Dichroism Experiments. Circular dichroism (CD) spectra were recorded in PEDL at room temperature in 0.01 cm cuvettes in a Jobin Yvon Dichrograph R. J. Mark III at a concentration of 10 μM. Complexes of c-Myb proteins and oligodeoxynucleotides were preincubated for 5 min at room temperature. Ten independent measurements were performed, and the data were accumulated using the program AKKUPROG (kindly provided by B. Kindler, MHH). The spectra were analyzed using the program CIRCULAR (kindly provided by Dr. F. Peters and Dr. J. Greipel, MHH).

Analytical Ultracentrifugation. Sedimentation velocity runs were carried out in PEDL buffer at 60 000 rpm at 25 °C in a Spinco Model E analytical ultracentrifuge (Beckman), equipped with electronic speed control, a high-intensity ultraviolet illumination system, a photoelectric scanner, and an electronic multiplexer. Sedimentation profiles were recorded at 280 nm. Band positions were determined by second-moment analysis (Schachman, 1959; Fujita, 1975)

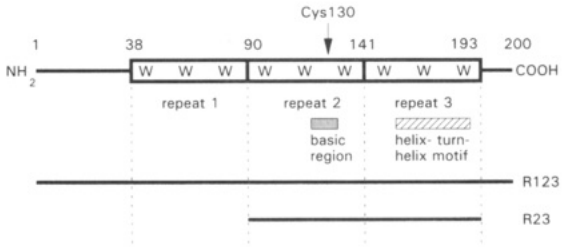

FIGURE 1: Schematic view of the first 200 amino acids of c-Myb. The positions of the tryptophans are marked by the single-letter symbol "W", while the repeats are represented by three white rectangular boxes and the basic region and helix−turn−helix motif by shaded boxes. The constructs used in this study are depicted as solid bars.

using the program UZAUS (Dr. C. Urbanke, unpublished results). The different samples were run in an An F-Ti-Rotor in parallel at a concentration of 5 μM.

Nitrocellulose Filter Binding Assay. Nitrocellulose filter binding experiments were performed as described (Ausubel et al., 1993) using a Bio-dot apparatus (Bio-Rad). Protein−DNA complexes were preformed in a total volume of 160 μL in binding buffer (10 mM Tris-HCl, pH 7.6, 1 mM DTT, 1 mM EDTA, 5 mM $MgCl_2$, and 100 mM NaCl) for 30 min at room temperature. Filters (Schleicher & Schüll BA 85 nitrocellulose) were prewetted and rinsed with binding buffer. The concentrations of the proteins were varied from 0 to 13 μM, the concentration of the oligodeoxynucleotides from 20 to 4500 nM. The complexes were applied to the filtration apparatus and washed once with 200 μL of binding buffer. The radioactivity retained on the NC filter was determined by scintillation counting. The binding isotherms were evaluated with the TITRAT computer program (Dr. C. Urbanke, unpublished results) using a multistep predictor/corrector method (Curth et al., 1993).

Trypsin Digestion Assay. R23 protein (buffer: 25 mM Tris-HCl, pH 8.0, 50 mM KCl, 10 mM $CaCl_2$, 10 mM $MgCl_2$, and 0.5 mM DTT) was incubated on ice for 30 min at a concentration of 5 μM with a 2-fold molar excess of oligodeoxynucleotide in a reaction volume of 20 μL. Two micrograms of trypsin (Sigma) was added, and the mixture was incubated at 25 °C for 30 min. The reaction was stopped with 10 μL of SDS−PAGE sample buffer [50 mM Tris-HCl, pH 6.8, 2% sodium dodecyl sulfate (w/v), 10% glycerol (v/v), 0.1% bromphenol blue, and 100 mM β-mercaptoethanol]; 20 μL of each sample was subjected to SDS−PAGE analysis.

RESULTS

Analysis of DNA−Protein Complexes. The proteins utilized in our study were expressed as GST fusion proteins; the GST segment was cleaved off with thrombin in the purification procedure. The constructs are represented schematically in Figure 1. Our preparations were at least 98% pure as judged by SDS−polyacrylamide gel electrophoresis. No DNA contaminations can be detected by UV spectroscopy (Figure 2). Due to the high tryptophan content (nine tryptophans in the binding region), the UV spectrum of the protein solution exhibits a pronounced shoulder at 290 nm. Although the DNA sequence of pR123Δ5 had been verified to code for the three DNA-binding repeats of the c-Myb protein, the purification scheme outlined under Experimental Procedures always yielded a protein of apparent

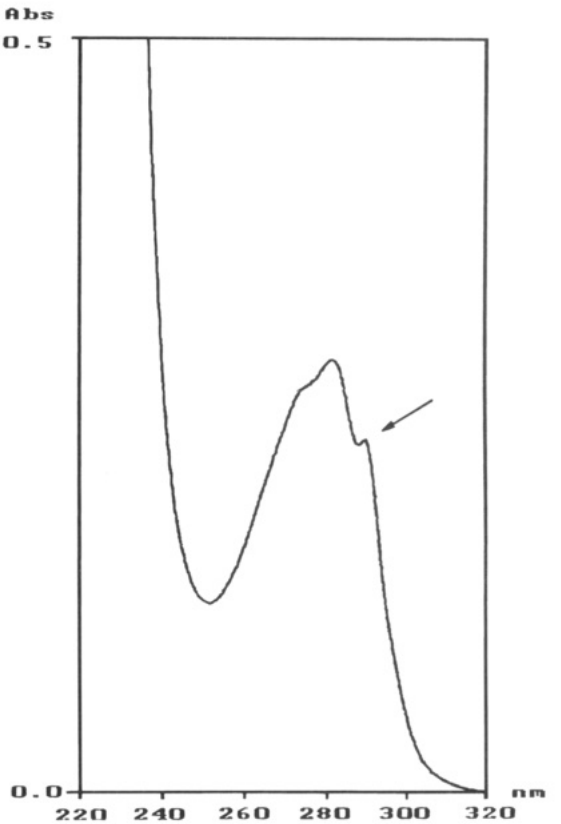

FIGURE 2: UV spectrum of a purified R213 protein. The arrow points to the shoulder of the tryptophan signal at 290 nm. With UV spectroscopy, no DNA is detectable at 260 nm.

Table 1: Oligodeoxynucleotide Sequences

canonical sequence	AACATTATAACGGTTTTTTAAT
GTT motif sequence	AACATTATTTCGGTTTTTTAAT
unspecific sequence	AACATTATTTGGCTTTTTTAAT

molecular mass of only 18 kDa. Protein sequencing revealed that in this construct a cryptic thrombin cleavage site (PK) at amino acid position 31 was used. Apparently, the deletion of five C-terminal amino acids in the third repeat changes dramatically the secondary structure of the GST−c-Myb fusion protein and renders a cryptic thrombin site accessible to proteolytic cleavage. Although this protein covers repeats 1−3 except for the last five amino acids, it is completely devoid of DNA-binding activity. This is consistent with the results of Howe (Howe et al., 1990), who used an analogous construct in their assays.

In order to study the complexes of the isolated c-Myb-binding domain with oligodeoxynucleotides containing a canonical c-Myb-binding site, a minimal GTT motif, and an unspecific sequence (Table 1), we used the nitrocellulose filter binding assay (NC) and electrophoretic mobility shift assay (EMSA). All assays were performed at least in triplicate. The individual constants of the various complexes were used to calculate the ratios of the canonical vs GTT motif binding constants. The results are summarized in Table 2.

Figure 3 shows the binding isotherm of one of these assays. We calculated a K_{ass} of 1.5×10^9 M^{-1} (±28%) for the R23 protein bound to the specific oligodeoxynucleotide, which is in the same order of magnitude reported by Tanikawa et al. (1993).

Table 2: Nitrocellulose Filter and Electrophoretic Mobility Shift Assay Analysis of the Complexes

method	protein	ratio of canonical vs GTT motif binding constants
NC filter	R23	1
NC filter	R123	8
EMSA	R23	1.2
EMSA	R123	3.6

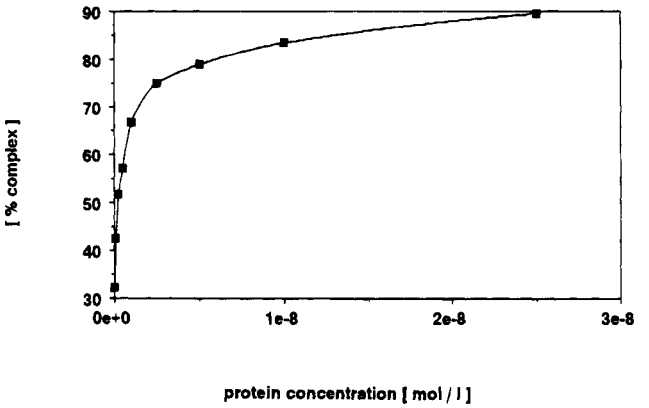

FIGURE 3: Titration of the canonical 22 bp oligodeoxynucleotide (0.2 nM) with R23 protein. The complexes were run on a 6% polyacrylamide gel in 0.5 × TBE. The gels were analyzed with a phospho-imager; the binding constant of 1.5×10^9 M^{-1} (±28%) was determined with the computer program TITRAT as described under Experimental Procedures.

We then determined the relative DNA-binding activity of protein R23 to a canonical sequence and an oligodeoxynucleotide containing a GTT core. Within the error margin of approximately ±30%, both sequences are bound equally. In contrast, there is a slight, but significant, difference for the binding behavior of the whole binding domain R123 to the two oligodeoxynucleotides: with both methods, the canonical sequence is bound better than the minimal GTT sequence. During the preparation of this paper, Ording et al. (1994) reported that R123 yields a 5-fold higher binding constant than R23, which is in accord with our result.

It was not possible to detect a complex between any of the proteins and an oligodeoxynucleotide lacking a GTT motif with the nitrocellulose filter binding assay or the electrophoretic mobility shift assay. However, a preformed complex can be competed with an excess of unspecific oligodeoxynucleotide lacking a GTT motif: the addition of a 200-fold molar excess of competitor to the preformed specific complex caused an approximately 25% decrease of the complex detected in an EMSA assay (data not shown). From this result, we estimated that the affinity of the R123 protein to unspecific DNA is about 1000-fold lower than to specific DNA. It can be expected that this difference is even more pronounced if one takes into account that oligodeoxynucleotides instead of synthetic poly(dI-dC) were used as competitor in our experiment. Thus, our result is consistent with the data of Myrset et al. (1993), who determined in the presence of poly(dI-dC) a difference of 3−4 orders of magnitude between the binding of R23 to specific and unspecific oligodeoxynucleotides.

In binding experiments with protein R123 and the 418 bp DNA fragment, which harbors six GTT motifs in addition to one YAACKG site (Figure 4), we detected one band for the complex at about equimolar concentrations of protein and DNA. With increasing amounts of protein, multiple unspecific complexes are formed, indicating an only moderate capability of the protein to distinguish between specific, minimal, and unspecific sites. At 130-fold molar excess of protein over DNA, one band with low mobility is detected, presumably representing DNA fully occupied with protein.

Analytical Ultracentrifugation Experiments. We studied the binding of protein R123 to specific and unspecific oligodeoxynucleotides. All experiments were carried out at a 5 μM concentration of R123 and DNA. The sedimentation velocity profiles are represented in Figure 5; the sedimentation coefficients are summarized in Table 3.

Free R123 migrates at a sedimentation coefficient of 1.8 S, unbound oligodeoxynucleotide at 2.1 S. The ultracentrifugation sedimentation runs detect a R123 protein interacting with an unspecific oligodeoxynucleotide at a sedimentation coefficient of 3.8 S. An s_{20w}, value of 3.8 S is compatible with a monomeric complex. The sedimentation coefficient of the specific complex is 3.7 S. Apparently, there is no difference in the Stoke's radii between a c-Myb protein bound to an unspecific or specific oligodeoxynucleotide. This result is at variance with EMSA and NC filter binding experiments, in which no binding to the unspecific sequence could be detected. This discrepancy can be explained by the fact that EMSA and NC filter assays are heterophasic techniques where the equilibrium of the reaction partners may be disturbed when free DNA and protein are separated from the complex during the analysis. Different from the assays described above, sedimentation velocity runs in the analytical ultracentrifuge allow experiments to be carried out in homogeneous solution where in the fast moving zone both complex and free components are present.

Circular Dichroism Experiments. The analysis of the CD spectrum (Figure 6) of the c-Myb DNA-binding domain R123 according to Chen (Chen et al., 1974) yielded an α-helical content of 49%. To detect a structural change in the protein by binding the DNA, oligodeoxynucleotides and protein were allowed to form a complex at room temperature at a concentration of 10 μM each. The CD spectrum of the specific complex was recorded, and the CD spectrum of the specific oligodeoxynucleotide was subtracted. The composite spectrum (Figure 6) is represented side-by-side with a CD spectrum of free R123 protein. The comparative analysis indicates that specific complex formation is accompanied by an increase of about 10% α-helical content of the protein. In contrast, no increase is observed when R123 is complexed with an unspecific oligodeoxynucleotide. The results are shown in Table 4. Interestingly, we did not record any difference in the CD signals when we compared specific and unspecific complexes of R23 and oligodeoxynucleotides at the same concentration.

It is known [for a review, see van der Vliet and Verrijzer (1993)] that protein binding to DNA may cause structural changes in the DNA (unwinding, kinking, bending, etc.). In order to rule out the possibility that the observed effect is due to conformational changes in the oligodeoxynucleotide structure induced by the interaction with the protein, we recorded CD spectra of the DNA and the complex up to 350 nm. Changes in DNA helicity or base stacking should give rise to a substantial change of the CD signal in the range of 250−350 nm. We do not detect structural changes of the DNA in this range; however, this does not rule out the

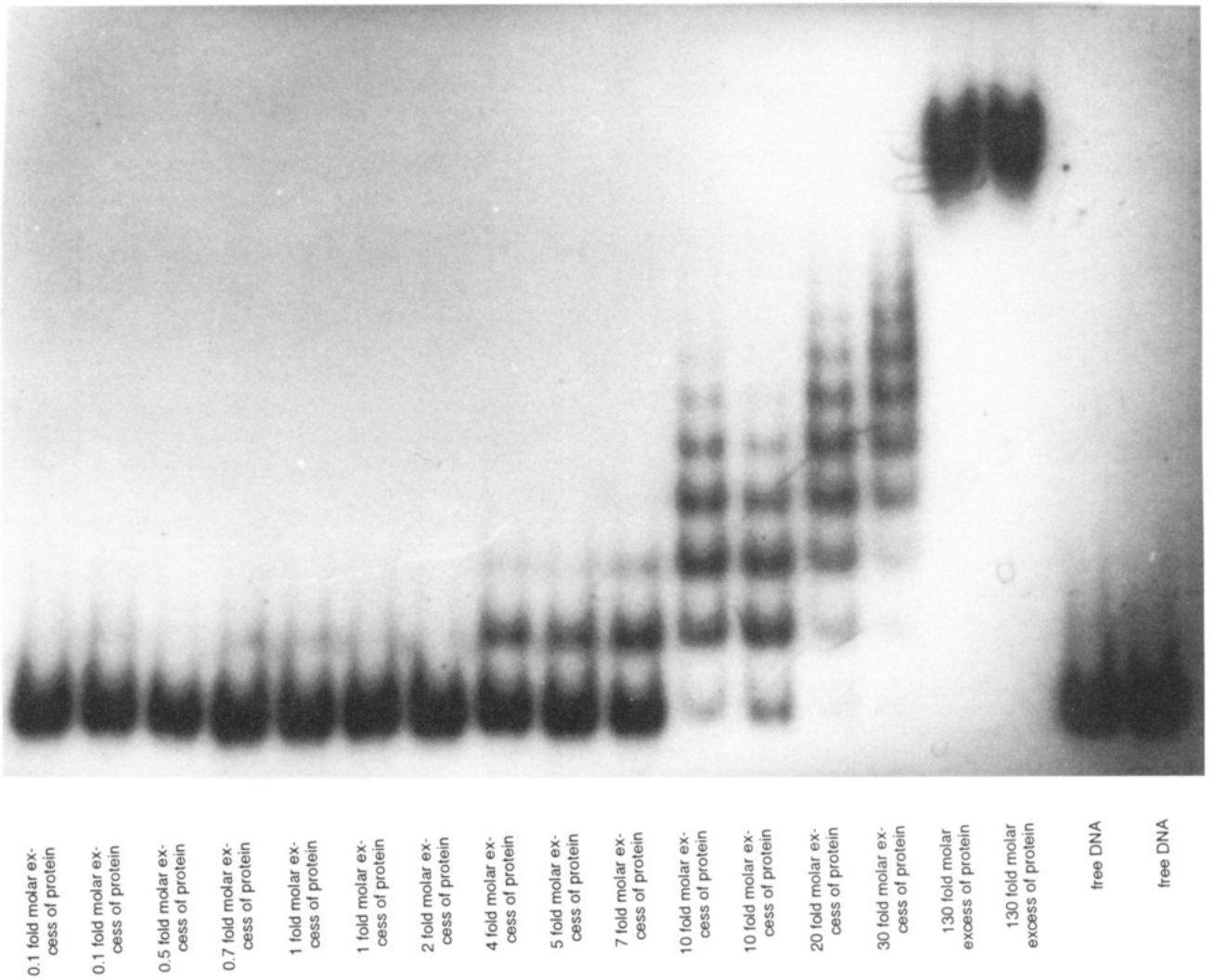

FIGURE 4: Nonspecific binding of protein R23 to a 418 bp DNA fragment containing one YAACKG site. Protein R23 and DNA were prepared as described under Experimental Procedures and run on a 6% polyacrylamide gel in 0.5 × TBE.

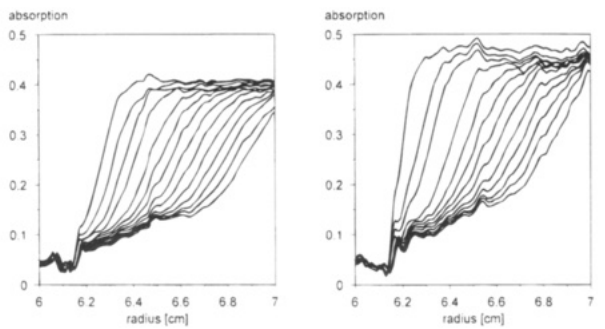

FIGURE 5: Sedimentation velocity profiles of the ultracentrifugation analysis. (Left) specific R123−oligodeoxynucleotide complex; (right) unspecific R123−oligodeoxynucleotide complex. Total sedimentation time was 7000 s. $s_{20,w}$ values (cf. Table 3) were determined from a ln radius (cm) vs time plot.

possibility that other minor conformational changes, which are not being recorded by CD spectroscopy, are induced upon binding of the protein.

Trypsin Digestion Assay. Another approach to detect conformational changes of the protein by binding to DNA is to compare the time course of tryptic digestion of the protein with or without DNA. During 30 min of incubation under the conditions chosen, the free protein is totally digested by the protease, as well as a protein bound to unspecific oligodeoxynucleotide. However, proteins complexed with a canonical or GTT motif-containing oligodeoxynucleotide are protected from digestion (Figure 7).

Table 3: Sedimentation Coefficients of R123, Oligodeoxynucleotide, and the Complexes

protein	oligodeoxynucleotide	$s_{20,w}$ (S)
R123		1.8
	specific	2.1
R123	specific	3.7
R123	unspecific	3.8

DISCUSSION

Many transcription factors are known to exhibit a modular organization with separate domains for DNA-binding and protein−protein interaction. In "swapping experiments", these modular domains can be exchanged between homologous proteins without impairing the function of the proteins (Brent & Ptashne, 1985). The c-Myb protein is considered to act as a transcription factor with a DNA-binding domain and a transactivation domain. In order to better understand and characterize its DNA-binding activity, we performed a detailed biophysical study of proteins carrying the c-Myb DNA-binding domain. Our first goal was the characterization of complexes between different isolated c-Myb DNA-binding domains and 22 bp oligodeoxynucleotides. The

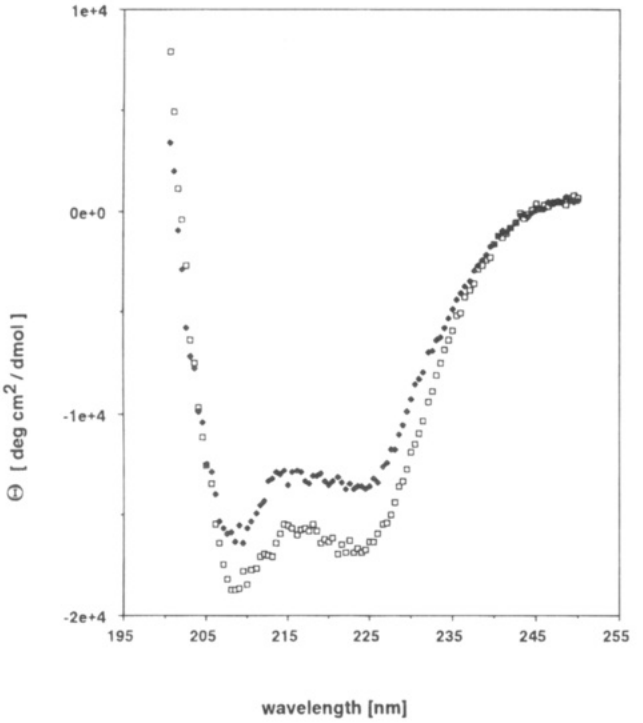

FIGURE 6: R123 changes its conformation in the presence of specific DNA: CD spectrum of R123 (solid boxes) and R123 in a specific complex (open boxes). The protein concentration is 10 μM in PEDL buffer. The α-helical content of the different complexes is compiled in Table 4.

Table 4: α-Helix Content (%) of the Complexes

	no DNA	unspecific	specific
R23	46	43	46
R123	49	48	60

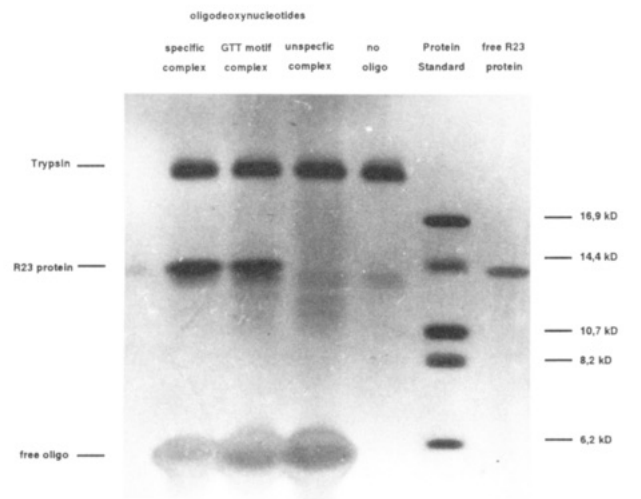

FIGURE 7: Trypsin digestion assay of R23. The presence of a specific oligodeoxynucleotide or a GTT motif containing oligodeoxynucleotide protects R23 from proteolytic cleavage.

c-Myb-binding site YAACKG was first published by Biedenkapp (Biedenkapp et al., 1988); later Zobel (Zobel et al., 1991) suggested the palindromic-binding site AACNGTT while Weston (1992) specified YAACKGHH as the consensus sequence in a PCR-site selection assay. All three proposals share the mutual sequence AACNG. As represented in Table 1, we used three oligodeoxynucleotides: the first oligodeoxynucleotide contained a TAACGGTT sequence, including the underlined palindromic sequence; in oligodeoxynucleotide 2, this sequence is shortened to a GTT motif with the same flanking sequences, while oligodeoxynucleotide 3 was an unspecific control.

As indicated in Table 2, the R23 protein does not distinguish between oligodeoxynucleotides 1 and 2, although oligodeoxynucleotide 2 diverges in three positions from the YAACKG sequence and contains only one GTT motif of the palindromic sequence proposed by Zobel et al. (1991). Albeit the binding affinity of protein R123 was determined to be of similar magnitude as compared to R23, the protein R123 has the capacity to discriminate between an oligodeoxynucleotide containing only one GTT motif from the canonical binding sequence. Therefore, the ability to distinguish must be contributed to repeat 1.

It must be emphasized that neither EMSA nor the NC filter binding method could detect binding of the c-Myb proteins to unspecific oligodeoxynucleotide. The fact that specific complexes can be competed with an excess of unspecific oligodeoxynucleotides (data not shown) illustrates that unspecific binding must occur. This property is even more apparent in regard to band shift assays using a 418 bp DNA fragment: the results reveal a distinct pattern of unspecific binding of the proteins. This finding is supported by the sedimentation analysis where specific and unspecific complexes are detected in solution. Furthermore, it was demonstrated in a Southwestern blot experiment (Oehler et al., 1990) that repeats 2 and 3 have the ability to retain unspecific DNA.

Myrset et al. determined in their paper (Myrset el al., 1993) a ratio of 3−4 orders of magnitude between unspecific and specific binding of their c-Myb constructs to DNA. As pointed out under Results, these data are not conflicting with our results.

For these reasons and with the caveat that the EMSA and nitrocellulose filter binding methods examine equilibrium binding constants in heterophasic systems under nonequilibrium conditions, where the reaction partners are separated during the analysis, we interpret the EMSA and nitrocellulose filter binding assays only as an approach to determine relative differences in the DNA−protein interaction of various complexes in parallel. Although we have measured in an EMSA experiment a K_{ass} of 1.5×10^9 M^{-1} (±28%) for the specific complex, the absolute value of the affinity constants can thermodynamically correctly only be determined under homophasic equilibrium conditions. Fluorescence emission spectroscopy titration experiments (manuscript in preparation) indicate that a K_{ass} of approximately 10^9 M^{-1} can be measured under homophasic conditions.

Since it is known that the conformation of DNA-binding proteins can be affected by binding their target sequence (Thompson et al., 1993; Weiss et al., 1993), we were interested whether the interaction with DNA influences the secondary structure of the c-Myb proteins.

The CD spectra of the DNA-binding region indicate a high level of α-helical structure. Our finding of 46% (R23) and 49% (R123) α-helical content is in good agreement with the results of Sarai (Sarai et al., 1993). Upon complex formation with a specific oligodeoxynucleotide, we detect an increase of approximately 10% of the α-helical signal of R123, which is not observed with an unspecific oligodeoxynucleotide. Interestingly, no effect is detectable when R23 binds to specific or unspecific DNA. Apparently, the

increase of the α-helical content is detectable only when repeat 1, present in R123, is interacting with specific DNA. In this respect, our results and those of Myrset et al. (1993), who also detect an α-helical increase upon binding, diverge from those of Sarai et al. (1993), who could not find a significant alteration in the presence of DNA. However, the experimental conditions are not strictly comparable: The construct pR123 codes for the three DNA-binding repeats plus 39 N-terminal amino acids, and, furthermore, we employed a buffer system containing a reducing agent.

From these results, it can be concluded that a structural change of the DNA-binding domain takes place upon binding to specific DNA sequences, which in our hands is stabilized only in the presence of repeat 1.

In recent publications (Ogata et al., 1992; Jamin et al., 1993; Myrset et al. 1993; Tanikawa et al., 1993), models for the structure−function relationship of the c-Myb DNA-binding domain were proposed. From the NMR data and computer modeling of different constructs of the c-Myb DNA-binding domain (Frampton et al., 1991), it was suggested that three α-helices in repeat R3 are clustered around a hydrophobic core formed by the conserved tryptophans and hydrophobic amino acids. In this structure, helix 2 and helix 3 are arranged in a helix−turn−helix-related motif. In contrast, repeat R2 covers two α-helices and a flexible region homologous to the basic region of jun, fos, and GCN4 (Carr & Mott, 1991). In this region, the conserved cysteine residue 130 is located, which is believed to play an essential role in the DNA-binding process (Guehmann et al., 1992; Grässer et al., 1992; Myrset et al., 1993). Although repeat 1 appears to possess the most stable structure of the DNA-binding region (Sarai et al., 1993), it is widely accepted that the first repeat of the DNA-binding region is not required for specific DNA binding, since it can be deleted without substantial loss of DNA-binding activity. Ording et al. (1994) reported that the deletion of repeat 1 results in a 5-fold decrease of the affinity constant of the DNA-binding domain toward a specific oligodeoxynucleotide.

However, we detect a conformational change of the c-Myb DNA-binding region upon binding the specific oligodeoxynucleotide only when repeat 1 is present. Repeat 3, on the other hand, cannot be deleted without loss of DNA-binding activity, because no DNA binding could be detected with pR123Δ5. In addition, it is not possible to obtain retarded bands of specific oligodeoxynucleotides in EMSA experiments with a protein consisting solely of repeats 1 and 2 (Howe et al., 1990). Furthermore, repeat 1 seems to be responsible for discriminating a canonical sequence from a binding site reduced to a GTT motif. This lends support to a model, proposed by several authors (Sarai et al., 1993; Tanikawa et al., 1993; Jamin et al., 1993; Myrset et al., 1993; Ording et al., 1994), where repeat 3 might serve as an anchor in contacting the DNA while repeat 2, which exhibits a pronounced flexibility, alters its conformation and gains stability only in the presence of repeat 1 once the target sequence is recognized. In addition to the function of the DNA-binding domain to recognize its specific target sequence, the domain is responsible for interaction with other transcription factors such as C/EBPβ (Burk et al., 1993; S. Mink, personal communication). Although the contact surface between C/EBPβ and c-Myb has not been mapped precisely (K. H. Klempnauer, personal communication), it is plausible that the R1-assisted structural change we have observed in our CD experiments could play a crucial role in regard to modulation of binding strength for specificity by protein−protein interactions of the two transcription factors and their cooperative transactivation of the mim-1 promoter (Burk et al., 1993). The regulatory potential of R1 and the additional N-terminal amino acids of the c-Myb DNA-binding domain becomes even more apparent, if one takes into account that serines 11 and 12, potential phosphorylation sites of casein kinase II (Lüscher et al., 1990), are deleted during oncogenic activation of c-Myb. It is conceivable that the deletion of the N-terminal amino acids and the majority of R1 in the activated v-Myb results in the loss of the secondary-structure alteration by sequence-specific binding and cross-talk to other proteins.

In conclusion, although the presence of repeat 1 and the N-terminal region of c-Myb does not substantially contribute to an increase in DNA affinity, its deletion renders the protein less specific and prevents a conformational change of the protein.

ACKNOWLEDGMENT

We are indebted to Dr. Dan Rosson, who supplied the c-myb gene, and Dr. Peter Jungblut for utilization of his protein-sequencing and phospho-imager facilities. We thank Dr. Alfred Pingoud for stimulating discussions and encouragement, as well as Dr. Jürgen Alves for critical reading of the manuscript. We acknowledge Mrs. Claudia Buntenkötter and Mrs. Hella Scharnhorst for expert technical assistance.

REFERENCES

Anton, I., & Frampton, J. (1988) *Nature 336*, 719−719.

Ausubel, F., Brent, R., Kingston, R., Moore, D., Seidmann, J., Smith, J., & Struhl, K., Eds. (1993) *Current Protocols in Molecular Biology*, Wiley, New York.

Bender, T., & Kuehl, W. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 3304−3308.

Biedenkapp, H., Borgmeyer, U., Sippel, A., & Klempnauer, K. H. (1988) *Nature 335*, 835−837.

Brent, R., & Ptashne, M. (1985) *Cell 43*, 729−736.

Burk, O., Mink, S., Ringwald, M., & Klempnauer, K. (1993) *EMBO J. 12*, 2027−2038.

Carr, M., & Mott, R. (1991) *FEBS Lett. 282*, 293−294.

Chen, H., Yang, J., & Chau, K. (1974) *Biochemistry 13*, 3350−3359.

Curth, U., Greipel, J., Urbanke, C., & Maass, G. (1993) *Biochemistry 32*, 2585−2591.

Frampton, J., Gibson, T., Ness, S., Döderlein, G., & Graf, T. (1991) *Protein Eng. 4*, 891−901.

Fujita, H. (1975) *Foundations of Ultracentrifugal Analysis*, Academic Press, New York.

Garcia, A., LaMontagne, K., Reavis, D., Stober-Grässer, U., & Lipsick, J. (1991) *Oncogene 6*, 265−273.

Geiger, R., Rüter, T., Alves, J., Wolfes, H., Pingoud, V., Urbanke, C., Maass, G., Pingoud, A., Düsterhöft, C., & Kröger, M. (1989) *Biochemistry 28*, 2667−2677.

Gerondakis, S., & Bishop, M. (1986) *Mol. Cell. Biol. 6*, 3677−3684.

Grässer, F., LaMontagne, K., Whittaker, L., Stohr, S., & Lipsick, J. (1992) *Oncogene 7*, 1005−1009.

Guehmann, S., Vorbrueggen, G., Kalkbrenner, F., & Moelling, K. (1992) *Nucleic Acids Res. 20*, 2279−2286.

Härd, T., Dahlman, K., Duke, J., Gustafsson, J., & Rigler, R. (1990) *Biochemistry 29*, 5358−5364.

Hassanain, H., Dai, W., & Gupta, S. (1993) *Anal. Biochem. 213*, 162–167.

Howe, K., Reakes, C., & Watson, R. (1990) *EMBO J. 9*, 161–169.

Jamin, N., Gabrielsen, O., Gilles, N., Lirsac, P., & Toma, F. (1993) *Eur. J. Biochem. 216*, 147–154.

Kanei-Ishii, C., Sarai, A., Sawazaki, T., Nakagoshi, H., He, D., Ogata, K., Nishimura, Y., & Ishii, S. (1990) *J. Biol. Chem. 32*, 19990–19995.

Katzen, A., Kornberg, T., & Bishop, M. (1985) *Cell 41*, 449–456.

Klempnauer, K. H., Arnold, H., & Biedenkapp, H. (1989) *Genes Dev. 3*, 1582–1589.

Levine, R., & Federici, M. (1982) *Biochemistry 21*, 2600–2606.

Litman, R. (1968) *J. Biol. Chem. 243*, 6222–6226.

Lüscher, B., & Eisenman, R. (1990) *Genes Dev. 4*, 2235–2241.

Lüscher, B., Christenson, E., Litchfield, D., Krebs, E., & Eisenman, R. (1990) *Nature 344*, 517–522.

Myrset, A., Bostad, A., Jamin, N., Lirsac, P., Toma, F., & Gabrielsen, O. (1993) *EMBO J. 12*, 4625–4633.

Nakagoshi, H., Nakase, T., Ueno, Y., & Ishii, S. (1989) *Nucleic Acids Res. 17*, 7315–7324.

Nakagoshi, H., Takemoto, Y., & Ishii, S. (1993) *J. Biol. Chem. 268*, 14161–14167.

Ness, S., Marknell, A., & Graf, T. (1989) *Cell 59*, 1115–1125.

Oehler, T., Arnold, H., Biedenkapp, H., & Klempnauer, K. H. (1990) *Nucleic Acids Res. 18*, 1703–1710.

Ogata, K., Hojo, H., Aimoto, S., Nakai, T., Nakamura, H., Sarai, A., Ishii, S., & Nishimura, Y. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 6428–6432.

Ording, E., Kvavik, W., Bostad, A., & Gabrielsen, O. (1994) *Eur. J. Biochem. 222*, 113–120.

Paz-Ares, J., Ghosal, D., Wienand, U., Peterson, P., & Saedler, H. (1987) *EMBO J. 6*, 3553–3558.

Rosson, D., & Reddy, P. (1986) *Nature 319*, 604–606.

Sarai, A., Uedaira, H., Morii, H., Yasukawa, T., Ogata, K., Nishimura, Y., & Ishii, S. (1993) *Biochemistry 32*, 7759–7764.

Sauer, T., Jordan, S., & Pabo, C. (1990) *Adv. Protein Chem. 40*, 1–61.

Schachman, H. (1959) *Ultracentrifugation in Biochemistry*, Academic Press, New York.

Tanikawa, J., Yasukawa, T., Enari, M., Ogata, K., Nishimura, Y., Ishii, S., & Sarai, A. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 9320–9324.

Thompson, K., Vinson, C., & Freire, E. (1993) *Biochemistry 32*, 5491–5496.

Twigg, A., & Sherrat, D. (1980) *Nature 283*, 216.

van der Vliet, P., & Verrijzer, P. (1993) *BioEssays 15*, 25–32.

Weiss, M., Ellenberger, T., Wobbe, R., Lee, J., Harrison, S., & Struhl, K. (1990) *Nature 347*, 575–578.

Westin, E., Gallo, R., Arya, S., Eva, A., Souza, L., Baluda, M., Aaronson, S., & Wong-Staal, F. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 2194–2198.

Weston, K. (1992) *Nucleic Acids Res. 20*, 3043–3049.

Zabeau, M., & Stanley, K. (1982) *EMBO J. 1*, 1217–1224.

Zobel, A., Kalkbrenner, F., Guehmann, S., Nawrath, M., Vorbrueggen, G., & Moelling, K. (1991) *Oncogene 6*, 1397–1407.

Zhang, X., Duggan, L., & Gottlieb, P. (1993) *Anal. Biochem. 214*, 580–582.