

Determination of the human c-Abl consensus DNA binding site

Marie-Hélène David-Cordonnier^a, Malika Hamdane^a, Christian Bailly^{a,b},
Jean-Claude D'Halluin^{a,*}

^aINSERM U 124 Onco-hématologie moléculaire, Institut de Recherches sur le Cancer de Lille, Place de Verdun, 59045 Lille, France

^bCentre Oscar Lambret, Institut de Recherches sur le Cancer de Lille, Place de Verdun, 59045 Lille, France

Received 3 December 1997

Abstract c-Abl tyrosine kinase, an essential protein of the cell cycle signalling pathways, is implicated in the regulation of RNA polymerase II activity, apoptosis and DNA repair. Its DNA binding activity is important for its biological functions. However, the molecular basis of c-Abl interaction with DNA remains largely unclear. We delimited the human c-Abl DNA binding domain and identified its preferred binding site, 5'-AA^A/C^AAACAA^A/C. The central AAC motif is highly conserved and constitutes the major core element in the binding sites. EMSAs and footprinting experiments were performed to explore how the c-Abl fusion protein recognizes specific sequences in DNA.

© 1998 Federation of European Biochemical Societies.

Key words: CASTing; DNA binding; Tyrosine kinase; Transcription factor

1. Introduction

The p145^{c-Abl} protein is a member of the c-Src non-receptor tyrosine kinase family. In its N-terminal part, c-Abl contains several domains common to c-Src tyrosine kinase but in its C-terminal part, it possesses specific functional regions including the nuclear localization signal [1,2], an actin binding site [3] and a DNA binding domain [4]. Mutations within the last exon encoding the large C-terminal domain are sufficient to activate the transforming potential of c-Abl [5].

Human and mouse c-Abl proteins contain two proline-rich regions within the central portion of their last exon. The DNA binding domain whose activity is regulated during the cell cycle by the cdc2 kinase, contains one of the two proline-rich regions [6]. At early stages of mitosis, cdc2 kinase phosphorylates c-Abl on serine and threonine residues. Hyperphosphorylation of the c-Abl DNA binding domain on Ser-852 and Ser-883 residues results in an inhibition of the DNA binding activity of c-Abl during mitosis [4]. In addition, the c-Abl tyrosine kinase can negatively regulate cell growth via a p53-dependent pathway [7,8]. c-Abl tyrosine kinase activity is also regulated during G2, M and G1 phases of the cell cycle by interaction with the retinoblastoma protein RB in order to abolish the c-Abl tyrosine kinase activity [9,10].

The C-terminal repeat domain (CTD) of the RNA polymerase II large subunit has been identified as a nuclear substrate for c-Abl tyrosine kinase which phosphorylates tyrosine residues present in the 52 amino acid repeats [11]. The tyrosine phosphorylation is believed to participate in the transition from initiation to elongation phases of the basal transcription

complex, as do serine and threonine phosphorylations on the CTD [12].

Although it has been known for some time that c-Abl binds to DNA, the type of sequence targeted by the protein is still a matter of debate. Dikstein et al. [13,14] have shown that c-Abl binds to the palindromic EP sequence of the hepatitis B virus enhancer but others have failed to find this interaction [15,16]. The activation of c-myc transcription by c-Abl [17] apparently depends neither on an EP-like element in the c-myc promoter nor on the DNA binding domain of c-Abl [15]. Therefore, the nature of c-Abl binding sites on DNA remains an area of ongoing controversy and it is the issue we addressed here.

2. Materials and methods

2.1. Binding of proteins mutants to DNA-cellulose

The DNA fragment [Bg/II 1233–Bg/II 3440] of the pSP65BCR-ABL plasmid encoding the C-terminal domain of human Abl protein was cloned into the pcDNA3 plasmid (Invitrogen). Several restriction sites in the c-Abl cDNA sequence were used for generating deletion mutants (Fig. 1). In vitro transcription and translation were performed with 1 µg of the various plasmids in TNT-reticulocyte lysates (Promega) and labelled using [³⁵S]methionine. The labelled products were incubated for 1 h at 4°C with 40 µl of DNA-cellulose (Pharmacia) pre-equilibrated in binding buffer (20 mM Tris-HCl pH7.4, 40 mM NaCl, 0.2% NP40, 1 mM EDTA, 1% aprotinin). After two washes with the same buffer, the protein-DNA complex was eluted with increased saline buffer (binding buffer supplemented with 100 mM, 300 mM, 500 mM and 1 M NaCl). Each fraction was resolved on 15% SDS-PAGE and quantified using a PhosphorImager. The S1A2, S2A2 and S1A1 proteins were obtained from the pAlter-EX1-derived vectors (Promega).

2.2. Determination of predicted secondary structure

The potential secondary structure of the NX-HIS fusion peptide was determined using the GSM [18], Garnier [19], Gibrat [20], Levin [21,22], DPM and SOPMA [23] computer programs.

2.3. Expression and purification of the human c-Abl DNA binding domain in bacteria

The [NcoI 2373–XmaIII 2912] sequence was fused at the 3' end to a (His)₇ coding oligonucleotide, in order to purify the protein on Ni beads, and cloned in the pTRC-99A vector (Pharmacia) or the pAlter-EX1 plasmid. The NcoI site at position 2373 was created by mutagenesis. The pTRC-99A-NX-HIS construct was used to obtain a bacterially expressed NX-HIS protein. The NcoI-MluI DNA fragment was subcloned into pGEX-2T vector (Pharmacia). The GST-NX-HIS fusion protein was expressed in *Escherichia coli* strain JM109 by induction with IPTG, purified using Ni columns under denaturing conditions and then refolded to a native structure by gradual dialysis against the EMSA incubation buffer (50 mM HEPES pH 7.9, 100 mM KCl, 5 mM EDTA, 1 mM PMSF, 20% glycerol). The protein preparation was analyzed by SDS-PAGE. Gels were stained with Coomassie blue or used for immunoblot in standard methods using a polyclonal anti-GST antibody. The capacity of the proteins to bind to DNA was verified via an interaction with DNA-cellulose, prior to be used in the cyclic amplification of sequence targeting (CASTing) experiments.

*Corresponding author. Fax: (33) 3 20 16 92 29.

E-mail: jcd@lille.inserm.fr

2.4. Selection of binding sites by CASTing [24]

5 µg of the R62 oligonucleotide (CAGGTCAGTTCAGCG-GATCCTGTCG(N)₁₂GAGGCGAATTCAGTGCAACTGCAGC) was added to 5 µg of the reverse primer, heated to 95°C and cooled slowly to room temperature prior to adding the dNTPs and the Klenow enzyme (10 units) and then incubated for 30 min at 37°C. The double-stranded 62-mer oligonucleotides were mixed with 50 ng of the purified GST-NX-HIS protein in 100 µl binding buffer (20 mM HEPES, pH 7.9; 100 mM KCl, 20% glycerol, 10 mg/ml BSA, 1 mM PMSF) in the presence of aprotinin (1 µg/ml) and 0.5 µg of poly-(dI-dC)(dI-dC) and then incubated for 30 min at 4°C. 25 µl of pre-equilibrated Ni bonds was added for 15 min at 4°C before washing 5–10 times with the same buffer. The washed pellet was resuspended in 50 µl of PCR buffer (Applied Biosystems), heated for 5 min at 95°C and the supernatant was subjected to 10 cycles of PCR (1 min at 94°C, 1 min at 65°C and 2 min at 72°C) in a final volume of 100 µl PCR buffer containing 15 pM of the reverse and forward primers. An aliquot was analyzed on a 3% agarose gel. When no amplicon was visualized, 10 additional PCR cycles were performed. Nine rounds of binding and amplification were undertaken using reduced protein concentrations and increasing KCl (up to 300 mM) and poly(dI-dC)(dI-dC) (up to 5 µg per reaction). Subsequently, the selected fragments were cloned in the pBS_{II}SK vector and sequenced.

2.5. Electromobility shift assays

The cloned fragments isolated by CASTing were 3' end-labelled with [α -³²P]dCTP (3000 Ci/mM, Amersham, UK) using the Klenow enzyme. The labelled probes (10 000 cpm, 5 ng) were incubated with GST-NX-HIS in EMSA binding buffer (50 mM HEPES pH 7.9; 100 mM KCl; 1 mM DTT; 5 mM MgCl₂; 0.1 mM EDTA; 20% glycerol; 500–1000-fold poly(dI-dC)(dI-dC); 1 mM PMSF). Electrophoresis was carried out on a 6% polyacrylamide gel in a 0.7×TAE buffer (4.7 mM Tris-HCl; 2.3 mM Na-acetate; 0.7 mM EDTA; pH 7.9) at 100 V.

2.6. DNase I experiments

DNase I footprinting was performed as described [25]. The *Xba*I-*Xho*I-labelled fragment of pBS_{II}SK-05 was incubated with the purified protein and subjected to the nicking activity of DNase I. DNA cleavage products were resolved by electrophoresis on a 8% polyacrylamide gel containing 8 M urea. Electrophoresis was performed at 85 W in TBE buffer (100 mM Tris-HCl; 100 mM boric acid; 2 mM EDTA; pH 8.3).

3. Results

In previous studies using the mouse c-Abl protein [4], the DNA binding domain was localized in the C-terminal region, between the kinase domain and the actin binding site (Fig. 1). To delimit the position of the DNA binding domain in the human protein, we performed multiple deletion mutants (see Fig. 1). The so-called BB and S1A2 fragments correspond to the large C-terminal domain of c-Abl. The deletion mutants were transcribed and translated in reticulocyte lysates using T7 RNA-polymerase and the labelled proteins were tested for their DNA-cellulose affinity (Fig. 1). The *Bgl*III-*Sma*I and *Apa*I-*Bgl*III segments located on the 5' and 3' sides of the largest BB fragment, respectively, are not absolutely required for the interaction with DNA. Indeed, deletion of either the 3' segment (mutant BBΔ*Apa*) and the 5' segment (mutant S1A2) has little or no effect on the interaction of the protein with DNA-cellulose. The simultaneous deletion of those two segments (mutant S1A1) confirms that the DNA binding domain is located between the *Sma*I 1469 and the *Apa*I 2752 sites. The protein mutant BBΔ*Sma* with a large deletion between the two *Sma*I sites shows a slightly reduced affinity for DNA whereas deletion between the two *Pst*I sites (mutant BBΔ*Sma*-Δ*Pst*) almost completely abolishes the DNA binding capacity of the protein. The analysis of the various deletion

mutants suggests that the DNA binding domain of human c-Abl is mainly located between the *Sma*I 2564 and the *Apa*I 2752 sites. However, the sequences on the 5' side of the *Sma*I 2564 site may also be important for the interaction with DNA (compare mutants S1A2 and S2A2). For this reason and to make sure that the fusion protein will interact with DNA, we cloned a fragment *Nco*I (2373)–*Xma*III (2912) which is slightly longer than the required *Sma*I (2564)–*Apa*I (2752) segment. The NX-HIS protein, produced in reticulocyte lysate, and the protein GST-NX-HIS produced in bacteria show the same DNA binding capacity as the full length c-Abl C-terminal domain (BB) (Fig. 1), whereas the GST protein alone failed to bind to DNA-cellulose (unpublished control). The protein GST-NX-HIS which can be produced in large quantity in bacteria was used for the different experiments described here below.

Five computer programs were employed to predict the secondary structure of the DNA binding domain of c-Abl. The agreement between the different methods is satisfactory. They all predicted roughly the same structure. A consensus structure was deduced from the complementary analysis. As indicated in Fig. 2, the DNA binding domain of c-Abl would be composed of three α -helices and three potential β -sheets separated by coiled regions of variable lengths.

CASTing experiments were undertaken to identify the sequence preferentially recognized by the minimal DNA binding domain of the human c-Abl protein. The GST-NX-HIS fusion protein was incubated with 62-mers containing 12 random base pairs, the bound sequences were subjected to nine rounds of selection-amplification and the last amplicons were cloned. Among the 47 selected sequences (Fig. 3), 41 showed a high homology and were used to display out the consensus binding

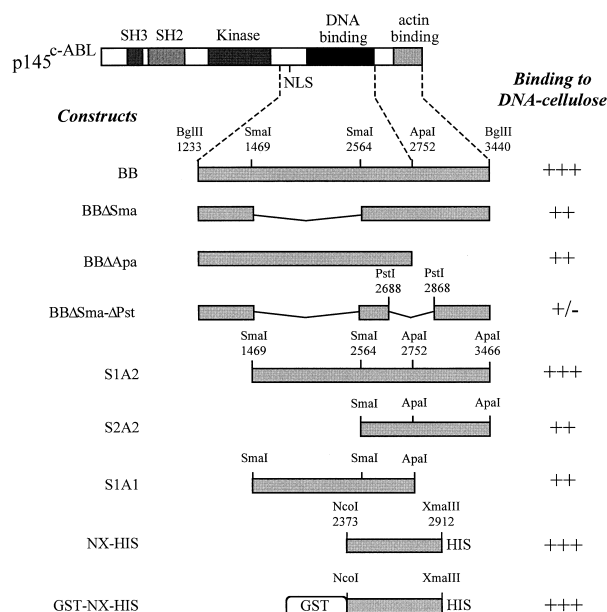
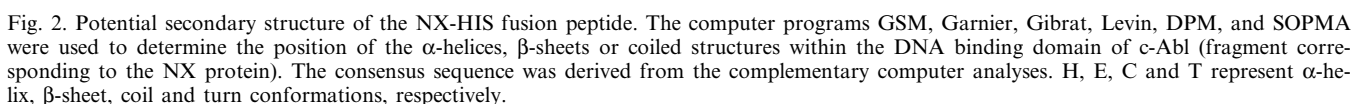


Fig. 1. Binding of c-Abl deletion mutants to DNA-cellulose. The deletion and fusion proteins were prepared with the indicated restriction enzymes. The relative affinity of the deletion mutants is given with respect to that of the BB protein used as a control (eluted with 300 mM NaCl). Symbols +++ and ++ indicate that the DNA binding capacity is equivalent and slightly reduced compared to that of the BB protein, respectively. The symbol ± denotes weak binding capacity. The double deletion mutant BBΔ*Sma*-Δ*Pst* binds to the DNA-cellulose only at low stringency (100 mM NaCl).



In order to test the ability of GST-NX-HIS fusion protein to preferentially bind to one or another selected sequences, a few clones (01, 02, 03, 05, 10, 17 and the control SK that does not contain any potential binding site) were employed in EMSAs using radiolabelled *XbaI-XhoI* fragments (73 or 78 bp) spanning the cloned selected sequences. Protein-DNA complexes can be identified with clones 01, 02, 03 and 05 and to a much lower extent with clone 10 but not with clone 17 and the control probe SK that lacks the target site but does contain the same flanking sequences (Fig. 4). The *XbaI-XhoI* frag-

The exact position of the protein binding site within probe 05, which presents the highest binding ability for GST-NX-

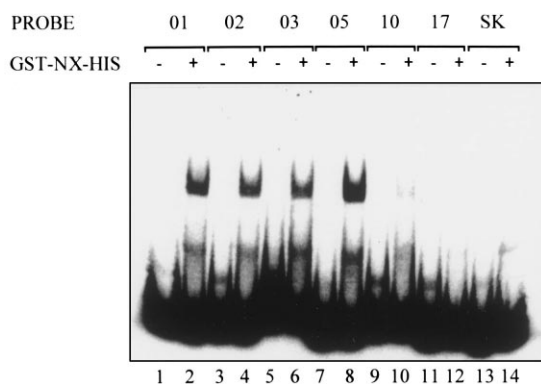


Fig. 4. EMSA showing the binding of the DNA binding domain of c-Abl to the *XbaI-XhoI* fragments containing the selected 01, 02, 03, 05, 10, 17 or the control SK sequences. The duplex oligonucleotide was incubated with (+) or without (–) 1 μ M of GST-NX-HIS in the presence of a 500-fold molar excess of poly(dI-dC)·(dI-dC). The SK probe containing the pBS_{II}SK plasmid sequence but lacking the protein binding site was used as a control.

^A/_TAACAA^A/_T, respectively [25–27]. The DNA binding domain of human p145^{c-Abl}, which we have located from a series of deletion mutants, appears to be homologous to that previously identified for the mouse p150^{c-Abl} protein [4]. This domain does not share any homology with other known DNA binding domains such as basic leucine zipper or zinc finger motifs, but presents a high proline content (the NX peptide contains 15% prolines). Using five computer programs, the $\beta 1$ - $\alpha 1$ - $\beta 2$ - $\alpha 2$ - $\alpha 3$ - $\beta 3$ predictive secondary structure was determined (Fig. 2). However, the potential α -helices are much shorter than those usually found in the DNA binding domain of proteins with a helix-loop-helix motif. Therefore, c-Abl apparently belongs to a distinct class of DNA binding proteins. In a recent study, Miao and Wang [16] reported that the mouse c-Abl DNA binding domain contains three HMG-related domains. Our own analysis with the human protein does not agree with this view. Not only do the three domains described by Miao and Wang differ significantly from one another but in addition they are devoid of tryptophan, tyrosine and threonine residues that are almost always found in the DNA binding domain of HMG and the related HMG-like proteins [29].

The analysis of the sequence specificity of c-Abl shows unambiguously that the human protein exhibits a high preference for sequences containing an AAC motif. Here again, the results are not totally consistent with those reported with the mouse protein. Miao and Wang indicated that p150^{c-Abl} interacts preferentially with AT-rich tracts as do HMG proteins [28], but that the mouse p150^{c-Abl} protein does not recognize any particular sequences. In our hands, the human protein binds to certain AT-rich sequences but largely prefers AAC-containing sequences. All the binding data concur that AAC-based sequences are much more preferred than purely AT-rich sequences. In this respect, it is worth noting that the human c-Abl protein was previously shown to bind to the EP palindromic element of the HBV enhancer which corresponds to the sequence CGTTGCTcgGCAACG [13]. Interestingly, we note that this EP element is not AT-rich but contains two AAC motifs. The highly conserved AAC motif is likely important for the protein-DNA recognition process to occur but it is surely not sufficient. The nature of the flanking nucleotides of the consensus sequence is important as well.

The observations that (i) the fusion protein binds more tightly to 73-mer sequences than to 32-mer containing the same binding core and (ii) the sequences selected by CASTing are frequently longer than the 12 bp randomized sequence suggest that the protein recognizes a central core as well as a particular DNA structure. The flanking sequences likely contribute to a reinforcement of the affinity of the protein for its target site. We also performed EMSA using GST-NX-HIS and the 05-selected sequence included in DNA fragments of several lengths and found that the binding affinity increased up to a size of 66 bp (data not shown). The same binding ability was observed with the c-Myb minimal DNA binding domain which needs a 3' extension of more than 10 bp to allow binding to the MRE (Myb responsive element). This extension of one turn of the helix is important for stabilizing the protein-DNA complex and increasing its half-life [30].

c-Abl has been implicated in DNA recombination and repair [31] and very recently in the regulation of DNA damage-induced apoptosis [32]. So far, it has not been possible to establish a direct connection between these biological effects and the particular DNA binding capacity of the protein. However, the interaction of c-Abl with its DNA consensus binding site may permit the correct positioning of c-Abl tyrosine kinase within the transcription complex so as to facil-

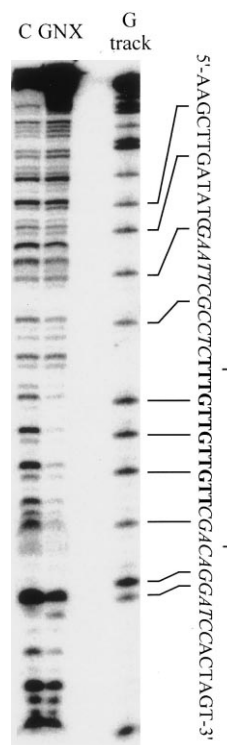


Fig. 5. DNase I footprinting of the GST-NX-HIS protein on the *XbaI-XhoI* fragment of clone 05. The 73 bp DNA fragment was 3' end-labelled at the *XbaI* site. The GST-NX-HIS fusion protein (lane GNX) or BSA in the control (lane C) (10 μ M each) was incubated with the DNA for 15 min at 30°C before digestion with 0.05 units of DNase I. The products of the DNase I digestion were identified by reference to the guanine markers (lane G). A portion of the DNA sequence is indicated on the side of the gel. The sequence protected from DNase I cleavage is bracketed and the sequence of oligonucleotide 05 that binds to the protein is shown in bold letters.

itate the phosphorylation of the C-terminal repeat domain of RNA polymerase II.

Acknowledgements: We thank Dr. E. Canaani for the pSP65BCR-ABL plasmid, H. Cordonnier for his assistance with the computer analysis of the selected sequences and G. Vergoten for his help in the prediction of secondary structure of the DNA binding domain. M.-H.D.-C. is supported by the Institut de Recherches sur le Cancer de Lille, the Conseil Régional du Nord-Pas de Calais and the ARC. M.H. is supported by the CHRU de Lille, the Conseil Régional du Nord-Pas de Calais, the Institut de Recherches sur le Cancer de Lille and the Ligue Nationale contre le Cancer. This work was supported by the INSERM, the Université de Lille II, the ARC (Grant 415-89) and the Comité du Pas-de-Calais of the Ligue Nationale contre le Cancer.

References

- [1] van Etten, R., Jackson, P. and Baltimore, D. (1989) *Cell* 58, 669–678.
- [2] Wen, S.T., Jackson, P.K. and van Etten, R.A. (1996) *EMBO J.* 15, 1583–1595.
- [3] McWhirter, J.R. and Wang, J.Y.J. (1993) *EMBO J.* 12, 1533–1546.
- [4] Kipreos, E.J. and Wang, J.Y.J. (1992) *Science* 256, 382–385.
- [5] Goga, A., McLaughlin, J., Pendergast, A.M., Parmar, K., Muller, A., Rosenberg, N. and Witte, O.N. (1993) *Mol. Cell. Biol.* 13, 4967–4975.
- [6] Kipreos, E.J. and Wang, J.Y.J. (1990) *Science* 248, 217–220.
- [7] Goga, A., Liu, X., Hambuch, T.M., Senechal, K., Major, E., Berk, A.J., Witte, O.N. and Sawyers, C.L. (1995) *Oncogene* 11, 791–799.
- [8] Sawyers, C.L., McLaughlin, J., Goga, A., Havlik, M. and Witte, O.N. (1994) *Cell* 77, 121–131.
- [9] Wang, J.Y.J. (1993) *Curr. Opin. Genet. Dev.* 3, 35–43.
- [10] Welch, P.J. and Wang, J.Y.J. (1995) *Mol. Cell. Biol.* 15, 5542–5551.
- [11] Baskaran, R., Dahmus, M.E. and Wang, J.Y.J. (1993) *Proc. Natl. Acad. Sci. USA* 90, 11167–11171.
- [12] Young, R.A. (1991) *Annu. Rev. Biochem.* 60, 689–715.
- [13] Dikstein, R., Heffetz, D., Ben-Neriah, Y. and Shaul, Y. (1992) *Cell* 69, 751–757.
- [14] Dikstein, R., Amagi, R., Heffetz, D. and Shaul, Y. (1996) *Proc. Natl. Acad. Sci. USA* 93, 2387–2391.
- [15] Arcinas, M., Sizer, K.C. and Boxer, L.M. (1994) *J. Biol. Chem.* 269, 21919–21924.
- [16] Miao, Y.-J. and Wang, J.Y.J. (1996) *J. Biol. Chem.* 271, 22823–22830.
- [17] Wong, K.K., Hardin, J.D., Boast, S., Cooper, C.L., Merrel, K.T., Doyle, T.G., Goff, S.P. and Calame, K.L. (1995) *Oncogene* 70, 705–711.
- [18] Gascuel, O. and Golmard, J.L. (1988) *Comput. Applic. Biosci.* 4, 357–365.
- [19] Garnier, J., Osguthorpe, D.J. and Robson, B. (1978) *J. Mol. Biol.* 120, 97–120.
- [20] Gibrat, J.F., Garnier, J. and Robson, B. (1987) *J. Mol. Biol.* 198, 425–443.
- [21] Levin, J.M., Robson, B. and Garnier, J. (1986) *FEBS Lett.* 205, 303–308.
- [22] Levin, J.M. and Garnier, J. (1988) *Biochim. Biophys. Acta* 955, 283–295.
- [23] Geourjon, C. and Deleage, G. (1995) *Comput. Applic. Biosci.* 11, 681–684.
- [24] Harley, V.R., Lovell-Badge, R. and Goodfellow, P.N. (1994) *Nucleic Acids Res.* 22, 1500–1501.
- [25] Graves, B.J., Jonhson, P.J. and McKnight, S.L. (1986) *Cell* 44, 565–576.
- [26] Ferrari, S., Harley, V.R., Pontiggia, A., Goodfellow, P.N., Lovell-Badge, R. and Bianchi, M.E. (1992) *EMBO J.* 11, 4497–4502.
- [27] Giese, K., Cox, J. and Grosschedl, R. (1992) *Cell* 69, 185–195.
- [28] Bianchi, M.E., Beltrame, M. and Paonessa, P. (1989) *Science* 243, 1056–1059.
- [29] Landsman, D. and Bustin, M. (1993) *BioEssays* 15, 539–546.
- [30] Gabrielsen, O.S., Sentenac, A. and Fromageot, P. (1991) *Science* 253, 1140–1143.
- [31] Chen, Y.-Y., Wang, L.C., Huang, M.S. and Rosenberg, N. (1994) *Genes Dev.* 8, 688–697.
- [32] Yuan, Z.-M., Huang, Y., Ishiko, T., Kharbanda, S., Weichselbaum, R. and Kufe, D. (1997) *Proc. Natl. Acad. Sci. USA* 94, 1437–1440.