Biochemical and Biophysical Research Communications **278,** 158–166 (2000)
doi:10.1006/bbrc.2000.3783, available online at http://www.idealibrary.com on IDEAL®

A Nuclease Hypersensitive Element in the Human c-*myc* Promoter Adopts Several Distinct *i*-Tetraplex Structures

Tomas Simonsson,*[,1] Marketa Pribylova,† and Michaela Vorlickova†
**Department of Molecular Biotechnology, Lundberg Laboratory, Chalmers University of Technology, P.O. Box 462, SE 405 30 Göteborg, Sweden; and †Institute of Biophysics, Academy of Sciences of the Czech Republic, Kralovopolska 135, CZ 612 65 Brno, Czech Republic*

Received October 10, 2000

Nucleic acid structure-function correlations are pivotal to major biological events like transcription, replication, and recombination. Depending on intracellular conditions *in vivo* and buffer composition *in vitro*, DNA appears capable of inexhaustible structure variation. At moderately acidic, or even neutral pH, DNA strands that are rich in cytosine bases can associate both inter- and intramolecularly to form *i*-tetraplexes. The hemiprotonated cytosine$^+$–cytosine base pair constitutes the building block for the formation of *i*-tetraplexes, and motifs for their formation are frequent in vertebrate genomes. A major control element upstream of the human c-*myc* gene, which has been shown to interact sequence specifically with several transcription factors, becomes hypersensitive to nucleases upon c-*myc* expression. The control element is asymmetric inasmuch as that one strand is uncommonly rich in cytosines and exhibits multiple motifs for the formation of *i*-tetraplexes. To investigate the propensity for their formation we employ circular dichroism (CD) in combination with ultra violet (UV) spectroscopy and native gel electrophoresis. Our results demonstrate the cooperative formation of well-defined *i*-tetraplex structures. We conclude that *i*-tetraplex formation occurs in the promoter region of the human c-*myc* gene *in vitro*, and discuss implications of possible biological roles for *i*-tetraplex structures *in vivo*. Hypothetical formation of intramolecular fold-back *i*-tetraplexes is important to c-*myc* transcription, whereas chromosomal translocation events might involve the formation of bimolecular *i*-tetraplex structures. © 2000 Academic Press

Key Words: **circular dichroism; spectroscopy; DNA; nucleic acid; i-motif; transcription.**

The protein product of the c-*myc* proto-oncogene activates telomerase (1, 2) and alters the activity of transcription factors that force cell growth forward (3, 4). Normally the human c-*myc* gene is tightly regulated and changes in its expression are critical to tumor progression (5). It is overly active in a variety of malignant tumors including lymphomas, leukaemias, and lung, cervical, ovarian, breast, pancreatic, and gastric cancers (6). Conversion of the c-*myc* proto-oncogene into its oncogenic counterpart originates from diverse phenomena like chromosomal translocation (7), gene amplification (8), retroviral transduction (9), proviral insertion (10), and point mutation (11). Not regarding how c-*myc* acquires oncogenic properties, mechanisms behind its deviant transcription are obscured by the fact that the gene utilizes as many as four different promoters (12).

One of the first steps in the activation process of genes is the rearrangement or disruption of the chromatin structure, which otherwise might prevent binding of the transcription machinery to the promoter. It is well established that chromatin unfolding in the promoter regions of genes is preceded by the appearance of nuclease hypersensitive elements (NHEs), whose existence are thought to be a prerequisite for gene expression as opposed to a mere reflection thereof. NHEs are believed to represent regions devoid of canonical nucleosomes in which the DNA is exposed and more sensitive to nucleases, but mechanisms that govern their formation remain elusive. The evolution of five NHEs consistently precedes activation of human c-*myc* (13). One of them, denoted NHE III_1, is a major c-*myc* control element and corresponds to bases 2180–2212 in the sequence of the human c-*myc* locus (14). Irrespective of promoter usage it accounts for up to 95% of total c-*myc* transcription (15, 16). The element has an unusual asymmetry; it is exceptionally guanine-rich in the anti-sense strand of the c-*myc* locus. This strand has been carefully characterized under physiological conditions. It forms a fold-back struc-

[1] To whom correspondence should be addressed at MRC-Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, England. Fax: +44 1223 213 556. E-mail: tomas@mrc-lmb.cam.ac.uk.

0006-291X/00 $35.00

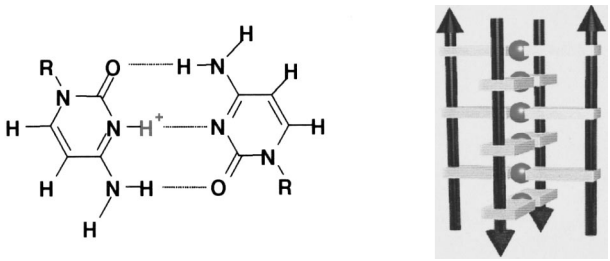

FIG. 1. The left panel shows a hemiprotonated cytosine$^+$–cytosine base pair, which is the fundamental building block for *i*-tetraplexes. H^+ represents the N3 proton and R represents the deoxyribose. The right panel schematically illustrates a fictive four-stranded *i*-tetraplex that can be adopted by d(CCC). Cytosine bases are represented by gray rectangles, N3 protons by spheres, phosphate backbones by black lines, and arrowheads indicate strand polarity.

ture with a core of stacked guanine tetrads, which specifically coordinate potassium ions *in vitro* (17, 18). The formation of such a structure in the guanine rich strand of duplex DNA emancipates the complementary cytosine-rich strand. The cytosine-rich strand of the c-*myc* NHE III_1 has so far escaped characterization, but might undergo formation of another atypical non-B-DNA structure—the cytosine intercalated tetraplex, also known as the *i*-tetraplex.

Cytosine-rich DNA strands can associate both inter- and intramolecularly to form *i*-tetraplexes, whose building block is the hemiprotonated cytosine$^+$–cytosine base-pair (Fig. 1, left). The protonation of N3 is essential for the stability of the structure and enables formation of three hydrogen bonds between the two cytosines. Although early diffraction fiber studies indicated that oligonucleotides containing cytosine tracts could base-pair at acidic pH (19), the precise structure was not revealed until NMR (20) and X-ray crystallography (21, 22) data were available. Surprisingly, it revealed that short cytosine-rich oligonucleotides form four-stranded structures composed of two parallel-stranded duplexes zipped together in antiparallel orientation (Fig. 1, right). The core structure is flat and has a pair of wide grooves and a pair of narrow grooves. Although acidic pH favors the formation of *i*-tetraplexes, and there is yet no conclusive proof that they exist *in vivo,* motifs for their formation are frequent in the human genome and possible biological roles must be given serious consideration. So far, they have been proposed to form in both centromeric (23) and telomeric (20, 24–26) regions of human chromosomes, as well as in the promoter region of the human c-ki-*ras* gene (27). In addition it has been speculated that the formation of *i*-tetraplexes correlates with the development of insulin-dependent diabetes mellitus (28) as well as the expansion of certain triplet repeat sequences associated with neurological disorders (29).

Circular dichroism (CD) is a very sensitive tool to monitor conformational isomerizations of DNA among its different conformers (30), including the interconversion between normal B-DNA and *i*-tetraplex conformations (27). Here we employ CD in combination with ultra violet (UV) absorption spectroscopy and native polyacrylamide gel electrophoresis (PAGE) to investigate the propensity of the cytosine-rich strand of the human c-*myc* NHE III_1 to adopt *i*-tetraplex structures.

MATERIALS AND METHODS

The 33-base-long oligonucleotide 5′-d(TCC CCA CCT TCC CCA CCC TCC CCA CCC TCC CCA)-3′ (PY33), whose sequence corresponds to bases 2180–2212 of the human c-*myc* locus (14) was from Eurogenetec, as were the 27-base-long oligonucleotide 5′-d(CCT TCC CCA CCC TCC CCA CCC TCC CCA)-3′ (PY27) corresponding to bases 2186–2212, the 22-base-long oligonucleotides 5′-d(CC CCA CCC TCC CCA CCC TCC CC)-3′ (PY22) corresponding to bases 2190–2211, and the 16-base-long oligonucleotide 5′-d(CCC TCC CCA CCC TCC C)-3′ (PY16), corresponding to bases 2195–2210 of the human c-*myc* locus. The oligonucleotides were purified on 20% polyacrylamide gels by standard denaturing electrophoresis, recovered by electroelution, and precipitated with ethanol. The two dodecanucleotides 5′-d(CGA CGA CGA CGA)-3′ ($(CGA)_4$) and 5′-d(CAG-CAGCAGCAG)-3′ ($(CAG)_4$), and the hexanucleotide 5′-d(TCC CCC)-3′ (TC5), were synthesized and purified by Genosys and East Port, respectively. The monomer, 2′-deoxycytidine (d(C)) was from Fluka. All oligonucleotides were dissolved in 1 mM sodium phosphate, pH 7.0, to yield stock solutions with approximate nucleoside residue concentrations of 10 mM. The precise sample concentrations were determined at room temperature in Robinson–Britton buffer, pH 8.50, using a UNICAM 5625 UV/Vis spectrophotometer. The molar extinction coefficients of the single stranded fragments were calculated according to Gray (31). At 260 nm the resulting values were 7930 $M^{-1}cm^{-1}$, 8085 $M^{-1}cm^{-1}$, 8130 $M^{-1}cm^{-1}$, 8190 $M^{-1}cm^{-1}$, and 7680 $M^{-1}cm^{-1}$ for PY16, PY22, PY27, PY33, and TC5, respectively.

CD spectra were recorded on Jobin-Yvon Mark IV and VI spectropolarimeters using Hellma cells placed in a thermostated holder. The apparatuses were calibrated by isoandrosterone. A Metrohm electrode connected to a Radelkis pH meter continuously monitored the pH, which was varied by adding HCl or NaOH directly to the samples. All experiments were measured in 1 cm cells at nucleoside concentrations of approximately 0.1 mM.

Native polyacrylamide gel electrophoresis was performed in a thermostated submarine type apparatus (SE 600, Hoefer Scientific). The dimensions of the gel (16%, 29:1 mono:bis ratio) were 14 × 16 × 0.1 cm, and the electrophoresis was run at approximately ±0°C for 17 h at 60 V (about 5 V/cm) in Robinson–Britton buffer. Gels were stained by Stains-all (1-ethyl-2-[3-(1-ethyl-naphthol-[1,2-d]-thiazolin-2-ylidene)-2-methyl-propenyl]-naphtha-[1,2-d]-thiazolium-bromide). Densitometry was performed using Molecular Dynamics Personal Densitometer SI 375 A. Acrylamide, N,N′-methylene-bisacrylamide, ammonium persulfate, TEMED (N,N,N′,N′-tetramethyl-ethylenediamine), TRIS (Tris(hydroxymethyl)aminomethane), boric acid, and EDTA (Ethylenediamine-tetraacetic acid) were from Sigma.

RESULTS AND DISCUSSION

All four oligonucleotides, whose sequences are identical to regions of the pyrimidine-rich strand of the human c-*myc* NHE, exhibit CD spectra with similar features at slightly acidic pH values: One very sharp positive band at 288 nm and one negative, with approx-

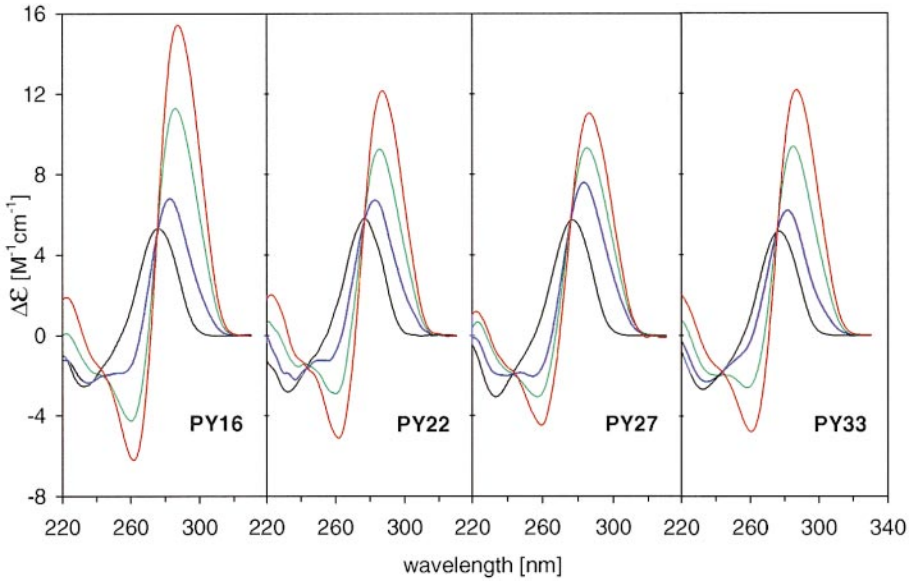

FIG. 2. CD spectra of the pyrimidine-rich oligonucleotides recorded at room temperature in Robinson–Britton buffer. PY16: pH 9.15 (black), pH 6.55 (blue), pH 6.42 (green), and pH 4.94 (red); PY22: pH 8.70 (black), pH 6.68 (blue), pH 6.52 (green), and pH 5.11 (red); PY27: pH 9.06 (black), pH 6.53 (blue), pH 6.29 (green), and pH 5.00 (red); PY33: pH 9.23 (black), pH 7.21 (blue), pH 6.67 (green), and pH 5.00 (red).

imately half the amplitude, at 267 nm (Fig. 2). These CD spectral characteristics match those previously reported for cytosine-rich oligonucleotides at acidic pH, and are distinctive for hemiprotonated cytosine$^+$–cytosine pairs (27). Alkalization of the samples to about pH 8 changes the spectral features similarly for all oligonucleotides. Their positive maxima shift to 277 nm, and the amplitudes decrease to 5 $M^{-1}cm^{-1}$. These spectra are almost congruent with those calculated (32) for the individual oligonucleotides in unstructured single stranded conformations (data not shown). In the region between pH 6 and 7 the spectra intersect in isoelliptic points at about 244 and 276 nm, which is strong evidence for a transition between two discrete conformational states. One state corresponds to unstructured single stranded conformations, whereas the other state corresponds to well-defined ordered structures that are stabilized by hemiprotonated cytosine$^+$–cytosine pairs. The transition between the two states is highly cooperative and reversible for all oligonucleotides. For the three longer oligonucleotides, PY22, PY27, and PY33, the transition kinetics are too rapid to be followed by CD, while it takes more than 5 h for PY16 to attain equilibrium in the vicinity of the transition midpoint. The transition midpoints are within the same range for all four oligonucleotides in Robinson–Britton buffer; it is at pH 6.7 for PY33 and at pH 6.5 for the three shorter oligonucleotides PY27, PY22, and PY16 (Fig. 3). *Nota bene,* lowering ionic strength considerably shifts the midpoint of the transition towards alkaline pH, so that all four oligonucleotides adopt *i*-tetraplex conformations already at pH 7 in 1 mM sodium phosphate. Despite slightly different primary structures, the single stranded forms of the oligonucleotides exhibit similar CD spectra when recorded at pH 8 or above. In contrast, the limiting CD spectra are similar in shape at acidic pH, but reach

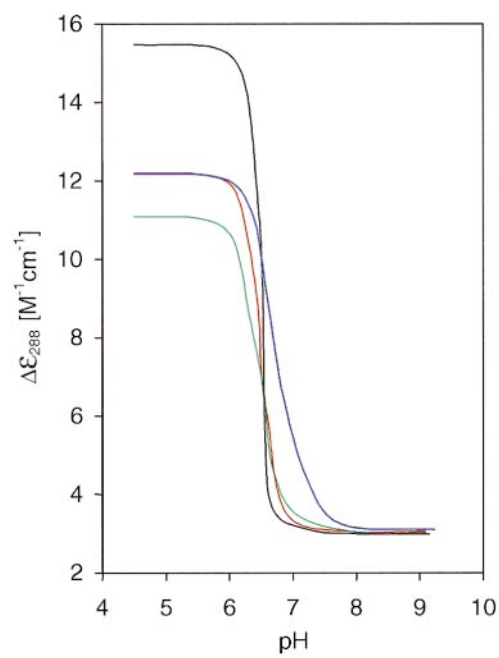

FIG. 3. pH-induced transitions monitored by ellipticity changes at 288 nm of PY16 (black), PY22 (green), PY27 (red), and PY33 (blue). Each curve consists of 10 individual measurements. Experimental conditions are identical to those as described in the legend to Fig. 2.

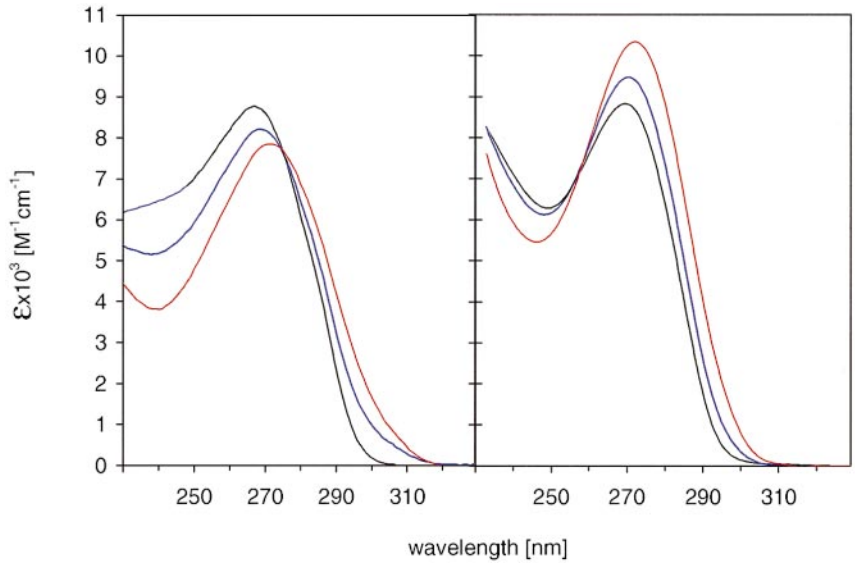

FIG. 4. Changes in UV absorption accompanying the pH-induced transition of the unstructured form of PY33 into its *i*-tetraplex conformation (A) and the protonation of 2′-deoxycytidine (B), which does not involve structural changes. UV spectra were recorded at room temperature in Robinson–Britton buffer; for PY33 at pH 8.5 (black), pH 6.4 (blue), and pH 4.9 (red); and for 2′-cytidine at pH 7.1 (black), 5.3 (blue), and 4.7 (red).

distinct ellipticity values. The shortest, PY16, exhibits the highest ellipticity value per base, PY27 and PY33 give roughly the same ellipticity values per base, whereas PY27 yields the lowest signal per base.

We complemented the CD measurements by monitoring how pH-induced *i*-tetraplex formation alters the UV absorption spectra of the four oligonucleotides. The changes are principally the same for all oligonucleotides: *i*-tetraplex formation is accompanied by decreased absorption maxima and red-shifts (Fig. 4A). As evidenced by the absorption spectrum of deoxycytidine, cytosine protonation without *i*-tetraplex formation also causes a red-shift (Fig. 4B). Albeit this red-shift is accompanied by increased absorption, and the CD

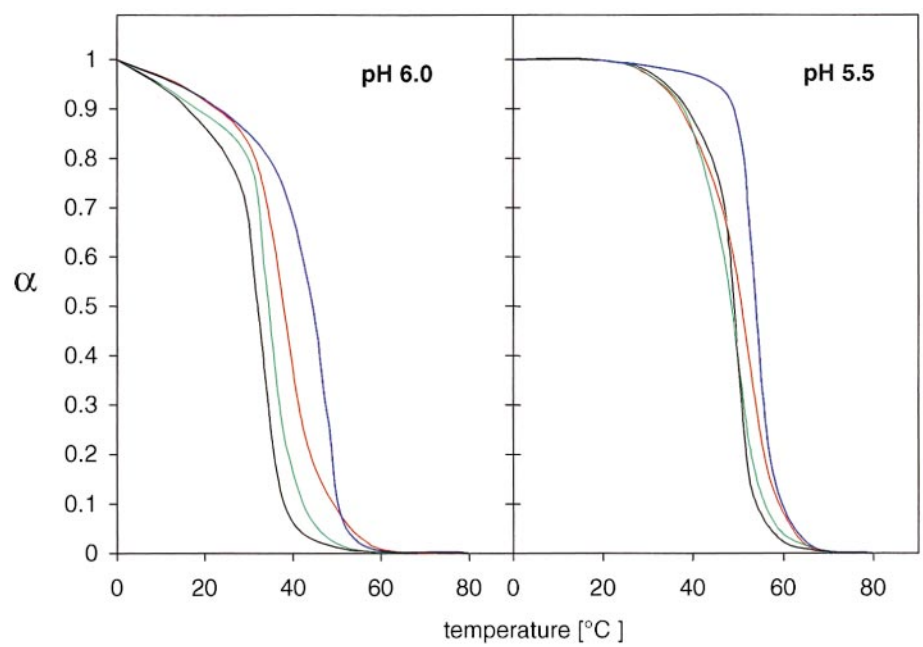

FIG. 6. Temperature-induced melting curves of PY16 (black), PY22 (red), PY27 (green), and PY33 (blue) in Robinson–Britton buffer, pH 6.0 (left) and pH 5.5 (right).

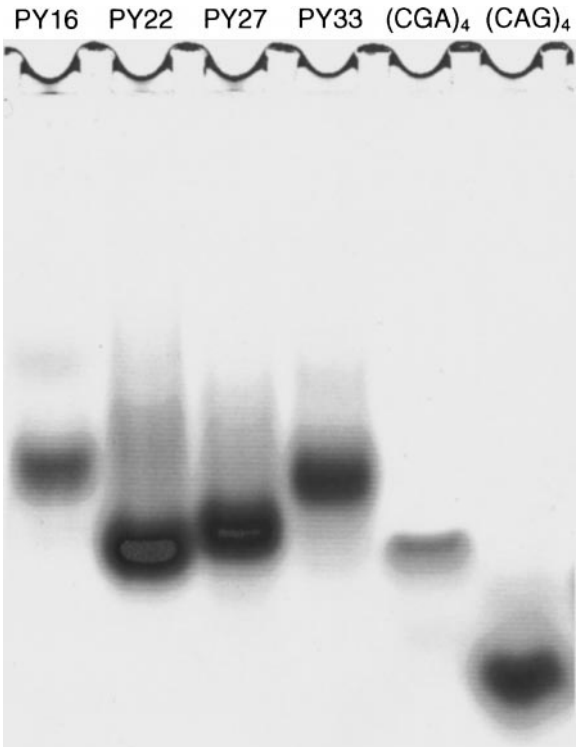

FIG. 5. Polyacrylamide gel electrophoresis performed in Robinson–Britton buffer pH 4.5 at 0°C. Included as markers are $(CGA)_4$ that forms a bimolecular duplex and $(CAG)_4$, which forms a stable intramolecular fold-back structure under the experimental conditions.

spectrum of deoxycytidine in the same pH region only exhibits an insignificant amplitude decrease.

Next, we analyzed the electrophoretic mobility of the oligonucleotides in native polyacrylamide gels at acidic pH (Fig. 5). Variations in migration distance reveal that the shortest oligonucleotide PY16 primarily is a bimolecular entity, whereas the three longer ones, PY22, PY27, and PY33, principally are unimolecular species. This was further confirmed by melting temperature measurements performed at low oligonucleotides concentrations. The PY16 melting temperature depends on concentration, whereas those of PY22, PY27, and PY33 does not (data not shown). We conclude that the ordered structure, which is stabilized by hemiprotonated cytosine$^+$–cytosine base pairs and formed by a single oligonucleotide is an intramolecularly fold-back *i*-tetraplex. Thus, the longer molecules form unimolecular *i*-tetraplexes, which is consistent with the observed fast kinetics of their formation and dissociation. The shortest oligonucleotide is a two-strand complex, which agrees well with the long lasting kinetics observed for the *i*-tetraplex formation and dissociation of PY16. The electrophoresis also reveals that a smaller population of the PY16 forms four-stranded complexes, whereas a minor fraction of the PY22 forms two-stranded complexes (Fig. 5).

Analogously to changes in pH, temperature changes also cause a reversible, cooperative (Fig. 6), and two-state (data not shown) transition of the *i*-tetraplexes into their unstructured single-stranded forms. Thus, both the intramolecular and intermolecular *i*-tetraplexes evolve directly from their unstructured single strands. Again, PY16 melts with exceedingly slow kinetics (e.g., attaining equilibrium after changing the temperature from +20°C to +35°C, at pH 6, takes about two hours), it exhibits the lowest T_m value (Table I) and its melting is the most cooperative. The other three oligonucleotides melt rapidly, without measurable kinetics. At pH 6, the T_m values of the oligonucleotides increase in the order PY16, PY27, PY22, and PY33 (Fig. 6A), whereas their melting becomes similar at pH 5.5 (Fig. 6B). Nevertheless, PY33 again forms the most stable and PY27 the least stable intramolecular *i*-tetraplex.

While the limiting CD spectra of the *i*-tetraplex conformations of the four studied oligonucleotides have

TABLE I

Some Characteristic Features of the Different *i*-Tetraplex Structures

Oligonucleotide	Strand stoichiometry	T_m [°C] (pH 6)	T_m [°C] (pH 5.5)	$\Delta\epsilon_{288}$ [$M^{-1}cm^{-1}$] (pH 5)	$\Delta\epsilon_{288}$ [$M^{-1}cm^{-1}$] (pH 8.5)	$\Delta\Delta\epsilon_{288}$ [$M^{-1}cm^{-1}$] (per tetraplex)		$\Delta\epsilon_{310}$ [$M^{-1}cm^{-1}$]	$\Delta\epsilon_{310}$ [$M^{-1}cm^{-1}$] (per tetraplex)	*n*
TC5	4			15.8	3.2			3.50	84.0	10.0
PY16	2	32.5	49.2	15.4	3.1	12.3	393.6	2.98	95.4	11.3
PY22	1	38.0	50.7	12.1	3.3	8.8	193.6	1.99	43.8	5.6
PY27	1	35.2	48.0	11.0	3.1	7.9	213.3	1.90	51.3	6.1
PY33	1	44.6	54.3	12.1	3.2	8.9	293.7	2.21	72.9	8.4

Note. Type of oligonucleotide; strand stoichiometry; melting temperature (T_m) in Robinson–Britton buffer at pH 5.5 and 6, respectively; ellipticity ($\Delta\epsilon$) at 288 nm of the unstructured single-stranded oligonucleotide at pH 8.5, and of the corresponding *i*-tetraplex structure at pH 5.0; ellipticity change ($\Delta\Delta\epsilon$) at 288 nm reflecting tetraplex formation related to molar nucleoside concentration and to molar concentration of the whole tetraplex associates, respectively. For each oligonucleotide the $\Delta\epsilon_{310}$ values represent averages of fifteen ellipticity measurements around 310 nm of the *i*-tetraplex structures measured at pH 5. At pH 8.5, $\Delta\epsilon_{310}$ is zero for all oligonucleotides. The value *n* corresponds to the number of hemiprotonated cytosine$^+$–cytosine base pairs.

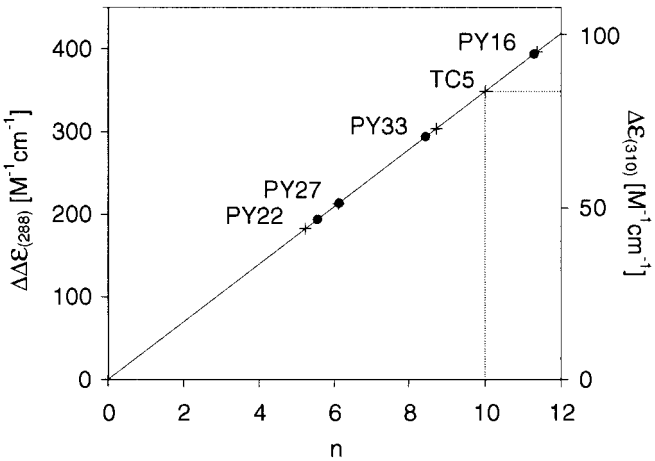

FIG. 7. Determination of the number of hemiprotonated cytosine$^+$–cytosine base pairs (n) in the different *i*-tetraplex structures. Changes in ellipticity upon pH-induced *i*-tetraplex formation were monitored at 288 nm (circles) and 310 nm (crosses). The slope of the line is given by the origin and the ellipticity change at 310 nm that corresponds to formation of the four-stranded TC5 *i*-tetraplex, which has ten hemi-protonated cytosine$^+$-cytosine base pairs (see text for details).

almost identical shapes, their amplitudes are significantly dissimilar. If from the individual CD spectra of the oligonucleotides in their *i*-tetraplex conformations (pH 5.0) we subtract the CD spectra of the corresponding single stranded forms (pH 8.5), the differential spectra depend linearly on the number of hemiprotonated cytosine$^+$–cytosine base pairs. Thereby it is possible to estimate the relative number of bases involved in the different *i*-tetraplexes (Table I). To determine the absolute number of hemi-protonated cytosine$^+$–cytosine base pairs in each of the *i*-tetraplexes we need a reference molecule of known structure. For this purpose, we have chosen the TC5 hexanucleotide, which forms a well-defined four-stranded *i*-tetraplex that is stable at acidic pH and oligonucleotide concentrations higher than 10 μM. Apart from the endmost thymines, the $(TC5)_4$ *i*-tetraplex consist exclusively of hemi-protonated cytosine$^+$–cytosine base pairs. Hence, the ellipticity change generated by this molecule upon *i*-tetraplex formation corresponds to the precise co-ordination of ten protons, and discloses the number of hemi-protonated cytosine$^+$–cytosine base-pairs involved in the other *i*-tetraplexes (Fig. 7). To minimize the influence of irrelevant oligonucleotide structures, we compared the ellipticities at longer wavelengths. Except for protonated cytosines, none of the bases absorb light at 310 nm. However, the ellipticity values at this wavelength are small and not as precise as at the CD maximum (Fig. 7). Models of the most likely *i*-tetraplex structures that conform to calculated data are presented in Fig. 8.

Our experiments support the idea that *i*-tetraplex formation occurs in the promoter region of the human c-*myc* gene *in vitro*. Let us now consider possible bio-

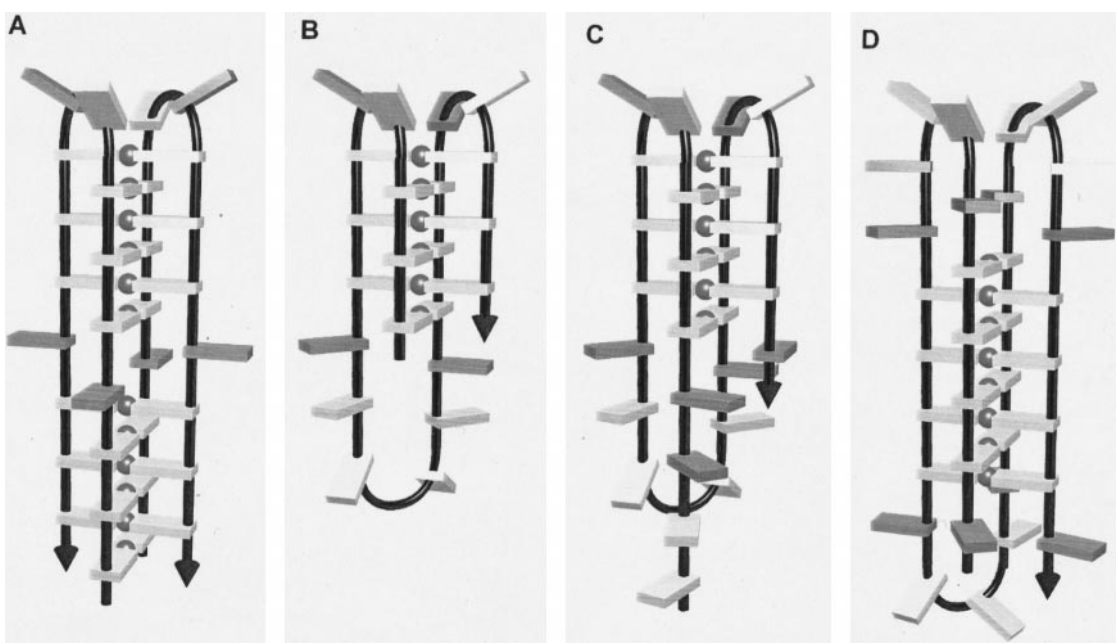

FIG. 8. Schematic illustration of putative *i*-tetraplex structures adopted by the oligonucleotides at slightly acidic conditions. Cytosine bases are represented by light gray rectangles, noncytosine bases by darker gray rectangles, N3 protons by spheres, phosphate backbones by black lines, and arrowheads indicate strand polarity. (A) PY16 adopts a bimolecular *i*-tetraplex with a core of 12 hemiprotonated base pairs. (B) PY22 adopts an intramolecular fold-back *i*-tetraplex with a core of 6 hemiprotonated base pairs. (C) The *i*-tetraplex core of PY27 is identical to that of PY22. (D) PY33 adopts an intramolecular fold-back *i*-tetraplex with a core of 8 hemiprotonated base-pairs.

logical roles thereof *in vivo*. We first turn the focus on how activation of c-*myc* transcription possibly involves formation of intrastrand fold-back *i*-tetraplexes.

DNaseI mapping early indicated that the NHE III_1 in the c-*myc* promoter was pivotal to activation of c-*myc* transcription (13, 33, 34). Subsequent characterization of the NHE III_1 by *in vivo* foot-printing has shown that the occurrence of DNaseI hypersensitivity at the NHE III_1 correlates temporally with expression of the c-*myc* gene (35, 36). Additionally, it has been excellently demonstrated by MNase digestion that displacement of the nucleosome comprising the NHE III_1 precedes c-*myc* activation (35, 37). Further S1 nuclease mapping of the c-*myc* gene, both as naked DNA (38) and in its native chromatin configuration (39), revealed that S1 cleavage co-localize with DNaseI and MNase hypersensitivity at the NHE III_1. Conclusively, the c-*myc* NHE III_1 alters between a traditional helix conformation and a stable atypical DNA structure, which evolves upon initiation of c-*myc* transcription *in vivo*. We believe that *i*-tetraplex formation in the cytosine-rich strand of the c-myc NHE III_1, together with G-quartet formation in the complementary guanine-rich strand, account for the above observations.

Can we thus conclude that activation of c-*myc* transcription involves formation of intrastrand fold-back *i*-tetraplexes? Of course not, but we may arrive closer to the real *in vivo* situation if we take into account the human proteins that have been found to recognize the c-*myc* NHE III_1.

The human c-*myc* promoter has been the subject of innumerous studies aiming to reveal the molecular mechanism(s) by which expression is transcriptionally regulated. The hnRNP A1, A2, and B1 (40), hnRNP K (41–44), NDPK-B (15, 45, 46), CNBP (47), NSEP-1 (48), Sp1 (49, 50), Sp3 (50, 51), CTCF (52), MAZi (53), THZif-1 (54, 55), and c-MYB (56, 57) all exemplify human proteins that recognize the NHE III_1 either *in vitro* or *in vivo*. A striking feature of the majority of these proteins is that they do not recognize the double stranded B-DNA conformation of the NHE III_1, but bind sequence specifically to either the pyrimidine-rich or the purine-rich strand in separate non-B-DNA conformations.

Two proteins from the above list, heterogeneous nuclear ribonucleoprotein K (hnRNP K) and nucleoside diphosphate kinase isoform B (NDPK-B), have been more extensively characterized than others with respect to their roles in activation of c-*myc* transcription. The hnRNP K interacts with the RNA polymerase II transcription machinery *in vivo* (58) and activates c-*myc* transcription from the NHE (41–44). The hnRNP K does not recognize the double-stranded form of the NHE (58) but binds sequence specifically to the pyrimidine-rich strand (40). Its nucleic acid binding domain is evolutionary conserved among mammalian species (59) and is common in proteins that recognize pyrimidine-rich nucleic acids (60). The precise hnRNP K binding motif comprises to two CCCTCCCCA-repeats separated by a spacer (43), which both PY27 and PY33 exhibit. We conjecture that the non-B-DNA conformation that is recognized by hnRNP K *in vivo* is the *i*-tetraplex structure adopted by PY33.

Despite its pivotal role in c-*myc* expression the hnRNP K cannot initiate transcription on its own, since it is oblivious to the B-DNA conformation of the NHE III_1. For proper c-*myc* expression it is imperative that the pyrimidine-rich strand of the NHE III_1 be able to adopt the non-B-DNA conformation that is recognized by the hnRNP K. This can be accomplished in two ways that are not necessarily mutually exclusive: the NHE III_1 might be capable of strand separation on its own, perhaps with the aid of torsional stress but without the help of direct protein catalysis. Such a mechanism would imply that the NHE III_1 and similar sequence motifs possess what might be described as inherent helicase properties. The ability to melt DNA and produce nuclease hypersensitive elements then strictly depends on sequence, and possibly explains why hypersensitivity is more confined to cytosine-rich sequences. The other conceivable mechanism is that the NHE III_1 is unwound by a helicase that provide the hnRNP K with the pyrimidine-rich strand in a non-B-DNA conformation. Even though the former mechanism is plausible, the latter is supported by the discovery of a multi-function protein with helicase activity that recognizes the NHE III_1. This is the NDPK-B, also known as the human nonmetastatic 23 isoform 2 protein (NM23-H2) or purine-binding factor (PUF) (46), which was the first protein that was reported to augment c-*myc* transcription from the NHE III_1 (15, 45). The NDPK-B binds with moderate affinity to the double-stranded B-DNA form of the NHE (61), unwinds it and binds both the pyrimidine-rich (62) and the purine-rich (63) strand in separate non-B-DNA conformations. It is clear that the role of the NDPK-B in c-*myc* expression is not that of a conventional transcription factor (64). However, it was recently demonstrated that the NDPK-B predominantly resides in the nuclei of dividing cells, which further supports its role in c-*myc* transcription (65). Additionally, *in vivo* foot-printing has shown that the NDPK-B activates chromosomally translocated c-*myc* alleles in Burkitt's lymphomas (66). We believe that the helicase activity of the NDPK-B facilitates *i*-tetraplex formation *in vivo*.

Is there also a possible biological role for bimolecular *i*-tetraplexes? Well, what perhaps is a motif for strand separation in the promoter regions of proto-oncogenes may turn out to be a two-edged sword. It is true that for initiation of c-*myc* expression only intramolecular fold-back *i*-tetraplexes are likely to be pertinent. Yet, the c-*myc* allele is normally present in two copies, and at four copies during mitosis, which in principle permits formation of both bimolecular and quadrimolecular

i-tetraplex structures. Furthermore, the formation of bimolecular *i*-tetraplexes does not exclusively depend on cytosine$^+$–cytosine base pairing between two DNA strands of identical sequence. Two stretches of consecutive cytosines, interspersed with the bases that will protrude as loops, are sufficient to form a bimolecular *i*-tetraplex core of intercalating hemiprotonated cytosine$^+$–cytosine base pairs. Notably the *i*-tetraplex motif is not unique to the c-*myc* promoter, but recurs as nuclease hypersensitive elements upstream of many proto-oncogenes; c-*abl* (67), c-*ets* (68, 69), c-*fes/fps* (70), c-*fgr* (71), c-*fos* (72), c-*jun* (73), c-*kit* (74), c-*mos* (75), c-*myb* (76), c-*rel* (77), c-*sis* (78), c-*src* (79), c-*yes* (80), and the *vav* proto-oncogene (81) are all cellular proto-oncogenes from man that exhibit the *i*-tetraplex motif in their upstream regulatory regions. It is tempting to speculate that the preponderance of motifs for tetraplex formation may serve as a molecular basis for chromosomal translocations of proto-oncogenes, thereby converting them into their oncogenic counterparts. Although other candidates exist, the aforementioned NDPK-B could be instrumental to the generation of chromosomal breakpoints. In addition to its role in c-*myc* transcription the NDPK-B not only associates with other sequences bearing motifs for *i*-tetraplex formation (82), but also generates double-strand breaks (83) and appears to drive expression of chromosomally translocated c-*myc* alleles *in vivo* (66).

Our finding, that a nuclease hypersensitive element in the promoter region of the human c-*myc* gene forms *i*-tetraplex structures, further elaborates the current working model for activation of c-*myc* transcription and provides an explanation for how nuclease hypersensitive elements evolve at the molecular level.

ACKNOWLEDGMENTS

Tomas Simonsson is a Swedish Cancer Society fellow receiving support from the Royal Swedish Academy of Sciences, FRF, Stiftelsen Jubileumsklinikens Forskningsfond mot Cancer, and Lars Hiertas Minnesfond. Michaela Vorlickova is supported by Grant 204/98/1027, awarded by the Grant Agency of the Czech Republic.

REFERENCES

1. Wu, K. J., Grandori, C., Amacker, M., Simon-Vermot, N., Polack, A., Lingner, J., and Dalla-Favera, R. (1999) *Nat. Genet.* **21,** 220–224.
2. Wang, J., Xie, L. Y., Allan, S., Beach, D., and Hannon, G. J. (1998) *Genes Dev.* **12,** 1769–1774.
3. Grandori, C., and Eisenman, R. N. (1997) *Trends Biochem. Sci.* **22,** 177–181.
4. Dang, C. V. (1999) *Mol. Cell. Biol.* **19,** 1–11.
5. Henriksson, M., and Luscher, B. (1996) *Adv. Cancer Res.* **68,** 109–182.
6. Slamon, D. J., deKernion, J. B., Verma, I. M., and Cline, M. J. (1984) *Science* **224,** 256–262.
7. Klein, G., and Klein, E. (1985) *Immunol. Today* **6,** 208–215.
8. Alitalo, K., Schwab, M., Lin, C. C., Varmus, H. E., and Bishop, J. M. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 1707–1711.
9. Collins, S. J., Robertson, K. A., and Mueller, L. (1990) *Mol. Cell. Biol.* **10,** 2154–2163.
10. Graham, M., Adams, J. M., and Cory, S. (1985) *Nature* **314,** 740–743.
11. Rabbitts, T. H., Hamlyn, P. H., and Baer, R. (1983) *Nature* **306,** 760–765.
12. Marcu, K. B., Bossone, S. A., and Patel, A. J. (1992) *Annu. Rev. Biochem.* **61,** 809–860.
13. Siebenlist, U., Hennighausen, L., Battey, J., and Leder, P. (1984) *Cell* **37,** 381–391.
14. Gazin, C., Dupont de Dinechin, S., Hampe, A., Masson, J. M., Martin, P., Stehelin, D., and Galibert, F. (1984) *Embo J.* **3,** 383–387.
15. Berberich, S. J., and Postel, E. H. (1995) *Oncogene* **10,** 2343–2347.
16. Davis, T. L., Firulli, A. B., and Kinniburgh, A. J. (1989) *Proc. Natl. Acad. Sci. USA* **86,** 9682–9686.
17. Simonsson, T., and Sjoback, R. (1999) *J. Biol. Chem.* **274,** 17379–17383.
18. Simonsson, T., Pecinka, P., and Kubista, M. (1998) *Nucleic Acids Res.* **26,** 1167–1172.
19. Langridge, R., and Rich, A. (1963) *Nature* **198,** 725–728.
20. Gehring, K., Leroy, J. L., and Gueron, M. (1993) *Nature* **363,** 561–565.
21. Kang, C. H., Berger, I., Lockshin, C., Ratliff, R., Moyzis, R., and Rich, A. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 11636–11640.
22. Chen, L., Cai, L., Zhang, X., and Rich, A. (1994) *Biochemistry* **33,** 13540–13546.
23. Gallego, J., Chou, S. H., and Reid, B. R. (1997) *J. Mol. Biol.* **273,** 840–856.
24. Ahmed, S., Kintanar, A., and Henderson, E. (1994) *Nat. Struct. Biol.* **1,** 83–88.
25. Leroy, J. L., Gehring, K., Kettani, A., and Gueron, M. (1993) *Biochemistry* **32,** 6019–6031.
26. Leroy, J. L., Gueron, M., Mergny, J. L., and Helene, C. (1994) *Nucleic Acids Res.* **22,** 1600–1606.
27. Manzini, G., Yathindra, N., and Xodo, L. E. (1994) *Nucleic Acids Res.* **22,** 4634–4640.
28. Catasti, P., Chen, X., Deaven, L. L., Moyzis, R. K., Bradbury, E. M., and Gupta, G. (1997) *J. Mol. Biol.* **272,** 369–382.
29. Vorlickova, M., Zimulova, M., Kovanda, J., Fojtik, P., and Kypr, J. (1998) *Nucleic Acids Res.* **26,** 2679–2685.
30. Vorlickova, M. (1995) *Biophys. J.* **69,** 2033–2043.
31. Gray, D. M., Hung, S. H., and Johnson, K. H. (1995) *Methods Enzymol.* **246,** 19–34.
32. Cantor, C. R., Warshaw, M. M., and Shapiro, H. (1970) *Biopolymers* **9,** 1059–1077.
33. Dyson, P. J., and Rabbitts, T. H. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 1984–1988.
34. Siebenlist, U., Bressler, P., and Kelly, K. (1988) *Mol. Cell. Biol.* **8,** 867–874.
35. Albert, T., Mautner, J., Funk, J. O., Hortnagel, K., Pullner, A., and Eick, D. (1997) *Mol. Cell. Biol.* **17,** 4363–4371.
36. Arcinas, M., and Boxer, L. M. (1994) *Oncogene* **9,** 2699–2706.
37. Pullner, A., Mautner, J., Albert, T., and Eick, D. (1996) *J. Biol. Chem.* **271,** 31452–31457.
38. Boles, T. C., and Hogan, M. E. (1987) *Biochemistry* **26,** 367–376.
39. Michelotti, G. A., Michelotti, E. F., Pullner, A., Duncan, R. C., Eick, D., and Levens, D. (1996) *Mol. Cell. Biol.* **16,** 2656–2669.
40. Takimoto, M., Tomonaga, T., Matunis, M., Avigan, M., Krutzsch,

H., Dreyfuss, G., and Levens, D. (1993) *J. Biol. Chem.* **268,** 18249–18258.

41. Matunis, M. J., Michael, W. M., and Dreyfuss, G. (1992) *Mol. Cell. Biol.* **12,** 164–171.
42. Tomonaga, T., and Levens, D. (1995) *J. Biol. Chem.* **270,** 4875–4881.
43. Tomonaga, T., and Levens, D. (1996) *Proc. Natl. Acad. Sci. USA* **93,** 5830–5835.
44. Michael, W. M., Eder, P. S., and Dreyfuss, G. (1997) *Embo J.* **16,** 3587–3598.
45. Postel, E. H., Mango, S. E., and Flint, S. J. (1989) *Mol. Cell. Biol.* **9,** 5123–5133.
46. Postel, E. H., Berberich, S. J., Flint, S. J., and Ferrone, C. A. (1993) *Science* **261,** 478–480.
47. Michelotti, E. F., Tomonaga, T., Krutzsch, H., and Levens, D. (1995) *J. Biol. Chem.* **270,** 9494–9499.
48. Kolluri, R., Torrey, T. A., and Kinniburgh, A. J. (1992) *Nucleic Acids Res.* **20,** 111–116.
49. DesJardins, E., and Hay, N. (1993) *Mol. Cell. Biol.* **13,** 5710–5724.
50. Majello, B., De Luca, P., Suske, G., and Lania, L. (1995) *Oncogene* **10,** 1841–1848.
51. Hagen, G., Muller, S., Beato, M., and Suske, G. (1994) *Embo J.* **13,** 3843–3851.
52. Filippova, G. N., Fagerlie, S., Klenova, E. M., Myers, C., Dehner, Y., Goodwin, G., Neiman, P. E., Collins, S. J., and Lobanenkov, V. V. (1996) *Mol. Cell. Biol.* **16,** 2802–2813.
53. Tsutsui, H., Sakatsume, O., Itakura, K., and Yokoyama, K. K. (1996) *Biochem. Biophys. Res. Commun.* **226,** 801–809.
54. Sakatsume, O., Tsutsui, H., Wang, Y., Gao, H., Tang, X., Yamauchi, T., Murata, T., Itakura, K., and Yokoyama, K. K. (1996) *J. Biol. Chem.* **271,** 31322–31333.
55. Yokoyama, K., Tsutsui, H., and Fujita, A. (1995) *Nippon Rinsho* **53,** 2827–2836.
56. Nakagoshi, H., Kanei-Ishii, C., Sawazaki, T., Mizuguchi, G., and Ishii, S. (1992) *Oncogene* **7,** 1233–1240.
57. Zobel, A., Kalkbrenner, F., Vorbrueggen, G., and Moelling, K. (1992) *Biochem. Biophys. Res. Commun.* **186,** 715–722.
58. Michelotti, E. F., Michelotti, G. A., Aronsohn, A. I., and Levens, D. (1996) *Mol. Cell. Biol.* **16,** 2350–2360.
59. Dejgaard, K., and Leffers, H. (1996) *Eur. J. Biochem.* **241,** 425–431.
60. Siomi, H., Matunis, M. J., Michael, W. M., and Dreyfuss, G. (1993) *Nucleic Acids Res.* **21,** 1193–1198.
61. Postel, E. H. (1996) *Curr. Top. Microbiol. Immunol.* **213,** 233–252.
62. Hildebrandt, M., Lacombe, M. L., Mesnildrey, S., and Veron, M. (1995) *Nucleic Acids Res.* **23,** 3858–3864.
63. Agou, F., Raveh, S., Mesnildrey, S., and Veron, M. (1999) *J. Biol. Chem.* **274,** 19630–19638.
64. Michelotti, E. F., Sanford, S., Freije, J. M. P., MacDonald, N. J., Steeg, P. S., and Levens, D. (1997) *J. Biol. Chem.* **272,** 22526–22530.
65. Kraeft, S. K., Traincart, F., Mesnildrey, S., Bourdais, J., Veron, M., and Chen, L. B. (1996) *Exp. Cell. Res.* **227,** 63–69.
66. Ji, L., Arcinas, M., and Boxer, L. M. (1995) *J. Biol. Chem.* **270,** 13392–13398.
67. Shtivelman, E., Lifshitz, B., Gale, R. P., Roe, B. A., and Canaani, E. (1986) *Cell* **47,** 277–284.
68. Mavrothalassitis, G. J., Watson, D. K., and Papas, T. S. (1990) *Proc. Natl. Acad. Sci. USA* **87,** 1047–1051.
69. Reddy, E. S., and Rao, V. N. (1988) *Oncogene Res.* **3,** 239–246.
70. Roebroek, A. J., Schalken, J. A., Verbeek, J. S., Van den Ouweland, A. M., Onnekink, C., Bloemers, H. P., and Van de Ven, W. J. (1985) *Embo J.* **4,** 2897–2903.
71. Patel, M., Leevers, S. J., and Brickell, P. M. (1990) *Oncogene* **5,** 201–206.
72. van Straaten, F., Muller, R., Curran, T., Van Beveren, C., and Verma, I. M. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 3183–3187.
73. Angel, P., Hattori, K., Smeal, T., and Karin, M. (1988) *Cell* **55,** 875–885.
74. Vandenbark, G. R., deCastro, C. M., Taylor, H., Dew-Knight, S., and Kaufman, R. E. (1992) *Oncogene* **7,** 1259–1266.
75. Watson, R., Oskarsson, M., and Vande Woude, G. F. (1982) *Proc. Natl. Acad. Sci. USA* **79,** 4078–4082.
76. Majello, B., Kenyon, L. C., and Dalla-Favera, R. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 9636–9640.
77. Brownell, E., Mittereder, N., and Rice, N. R. (1989) *Oncogene* **4,** 935–942.
78. Ratner, L., Thielan, B., and Collins, T. (1987) *Nucleic Acids Res.* **15,** 6017–6036.
79. Bonham, K., and Fujita, D. J. (1993) *Oncogene* **8,** 1973–1981.
80. Matsuzawa, Y., Semba, K., Kawamura-Tsuzuku, J., Sudo, T., Ishii, S., Toyoshima, K., and Yamamoto, T. (1991) *Oncogene* **6,** 1561–1567.
81. Katzav, S., Cleveland, J. L., Heslop, H. E., and Pulido, D. (1991) *Mol. Cell. Biol.* **11,** 1912–1920.
82. Nosaka, K., Kawahara, M., Masuda, M., Satomi, Y., and Nishino, H. (1998) *Biochem. Biophys. Res. Commun.* **243,** 342–348.
83. Postel, E. H. (1999) *J. Biol. Chem.* **274,** 22821–22829.