Journal of Cellular Biochemistry 56:74–85 (1994)

A Complex Composed of at Least Two HeLa Nuclear Proteins Protects Preferentially One DNA Strand of the Simple $(gt)_n(ga)_m$ Containing Region of Intron 2 in *HLA-DRB* Genes

Winfried Mäueler, Gabi Frank, Marc Muller, and Jörg T. Epplen

Molekulare Humangenetik, Ruhr-Universität, 44780 Bochum, Germany (W.M., G.F., J.T.E.) and Genetisches Institut der Justus-Liebig-Universität, 35392 Gießen, Germany (M.M.)

Abstract Electrophoretic mobility shift assays reveal that HeLa nuclear proteins bind fast and with measurable affinity to target DNAs containing mixed simple repetitive $(gt)_n(ga)_m$ stretches. Preincubation of the proteins at elevated temperature prevents the formation of the major DNA/protein complex in favour of several distinct assemblies. A similar pattern of retarded bands was observed employing higher salt concentrations in the binding reaction. Thus conformational changes of different proteins appear to influence the complex rather than alternating DNA structures. Separation of the total nuclear extract into a water soluble and an insoluble protein fraction leads to a complete loss of target DNA binding capability of the fractions. The binding capacity is restored by combining the two fractions suggesting that at least two protein components are necessary to form a complex with the target sequence. The proteins can be differentiated into heat sensitive, water soluble and temperature stable, water insoluble, respectively. Furthermore, specifically binding polypeptides are not detectable by Southwestern analyses, probably because the essential components are separated during electrophoresis. DNase I footprint analyses yield four different protein binding regions only on the $(gt)_n(ga)_m$ harbouring strand. The footprints cover larger portions of the mixed simple repeat in addition to a portion 5′ of the $(gt)_n$ part. Hence at least two nuclear protein components of unknown biological function have to be present simultaneously to protect preferentially the $(gt)_n(ga)_m$-containing strand of intron 2 in *HLA-DRB* genes.

Key words: DNA/protein interaction, simple repetitive DNA, binding domain, conformation changes, intron 2

Simple repetitive DNA sequences are widely spread across eukaryotic genomes. For most of these elements biological functions are not known [for review see Epplen et al., 1993]. Certain simple trinucleotide repeats appear to cause distinct human genetic diseases [Brook et al., 1992; Fu et al., 1991; Knight et al., 1993; La Spada et al., 1991; Orr et al., 1993; The Huntington's Disease Collaborative Research Group, 1993], although the molecular mechanisms remained largely unclear. To understand the biological meaning of naturally occurring simple tandem dinucleotide repeats we investigated distinct elements in class II genes encoded in the major histocompatibility complex (MHC). The introns of *MHC-DRB* genes contain mixed repeats with the basic motifs $(gt)_n(ga)_m$. This structure is preserved from artiodactyls to man [Schwaiger et al., 1993]. The reasons for this evolutionary stability of the intronic simple repeats are not clear. Recently we have demonstrated that different $(gt)_n(ga)_m$ containing targets isolated from intron 2 of human (*HLA-*) *DRB* alleles bind nuclear proteins from HeLa cells specifically [Mäueler et al., 1992]. In addition, protein binding to the $(gt)_n(ga)_m$ containing DNA was also demonstrated by using extracts from immortalized human B lymphocytes [Epplen et al., 1993]. In order to generate detailed insight into this protein/DNA interaction, we first characterized binding specificities and other biochemical properties of the proteins participating in the protein/DNA complex formed with

Received October 4, 1993; revised February 15, 1994; accepted March 16, 1994.

Address reprint requests to Winfried Mäueler, Molekulare Humangenetik, Ruhr-Universität, 44780 Bochum, Germany.

$(gt)_n(ga)_m$ containing targets and secondly we determined the binding sites in the target DNA responsible for formation of the protein/DNA complex.

METHODS

Target DNAs and 3'-End Labeling

Target DNAs from the following cloned human *HLA-DRB1* genes containing exon 2 and different $(gt)_n(ga)_m$ simple repeats in intron 2 [Rieß et al., 1990] (see Fig. 1) were prepared after restriction enzyme digestion [Zhen and Swank, 1993]: I) Clone A9/1 (*DRB1*0401*) containing $(gt)_{22}(ga)_{15}$; a) A9 138r, a 160 bp EcoO109I × EcoRI intronic fragment containing the simple repeat; b) A9 102r, a 102 bp HinfI fragment containing nearly exclusively the simple repeated DNA. II) Clone CE/1 (allele 1301/01) containing $(gt)_{25}(ga)_{10}ca(ga)_3ca(ga)_6$; a) CE 120r, a 120 bp HinfI fragment corresponding to A9 102r; b) CE 138r, a 180 bp EcoO109I × HindIII fragment corresponding to A9 138r. As a completely unrleated sequence C4 111r, a 130 bp EcoO109I × EcoRI fragment of clone C4 (I. Nanda, C. Epplen, J.T. Epplen; unpublished data) was used harbouring $(gaa)_{24}$. 3' -endlabelling using [α-^{32}P]-dATP or [α-^{32}P]-dCTP (6,000 Ci/mmol or 3,000 Ci/mmol; Amersham, Braunschweig) was carried out as described [Mäueler et al., 1992]. The labelled DNAs were phenolized and purified by Sephadex G-100 chromatography. Incorporation of ^{32}P was measured by scintillation counting. Typical specific activities were 2–6 × 10^6 cpm/pmol DNA.

Binding Assay of Total HeLa Nuclear Proteins to $(gt)_n(ga)_m$ Containing DNA

Preparation of nuclear extracts was performed according to Altschmied et al. [1989] containing all total ammonium sulfate precipitable proteins (total HeLa nuclear extract). Protein concentrations were determined according to the method of Bradford [1976]. Before use each nuclear extract preparation and each labelled target DNA was calibrated for the correct amount of poly(dIdC)-poly(dIdC) (Pharmacia, Freiburg) as unspecific competitor DNA (pdIdC) in standard gel retardation assays. For standard gel retardation assays 250–500 ng pdIdC were mixed with 5 μg total HeLa nuclear extract, 4 μl of 5 × incubation buffer (1 × = 5% glycerol, 20 mM HEPES/KOH, pH 7.6, 20 mM KCl, and 5 mM $MgCl_2$) and 8.5–10.5 μl TE (10 mM Tris/HCl, pH 8.0, and 1 mM EDTA) and preincubated for 20 min on ice; 2–4 μl of labelled target DNA (20 μl total reaction volume) were added and the mixture was again incubated for 20 min on ice. After addition of 1/10 volume of 10 × loading buffer (0.4% bromophenol blue, 0.4% xylene cyanol FF, 25% Ficoll), and 10 × electrophoresis buffer (1 × = 25 mM Tris, 192 mM glycine) probes were directly loaded onto precooled native 5% polyacrylamid gels (20 cm in length) which had been prerun for 1 h at 100 V. Gels were run for 3–4 h at 350 V, fixed (10% acetic acid, 12% methanol), dried, and exposed over night to Fuji RX X-ray films.

Protein binding affinities were determined using gel retardation assays by including increasing amounts of labelled target DNA in the reaction mix containing 5 μg of total HeLa nuclear extract. Competition experiments were carried out using constant amounts of labelled target DNA (approximately 10 fmol = 35,000 cpm) and total HeLa nuclear extract (5 μg) by including up to 500-fold molar excess of unlabelled competitor DNA. Radioactive bands were cut out from the gels and scintillation counted. The affinity studies were evaluated according to the method of Scatchard [1949]. Competition with unlabelled competitor DNA was calculated relative to the cpm measured for the corresponding

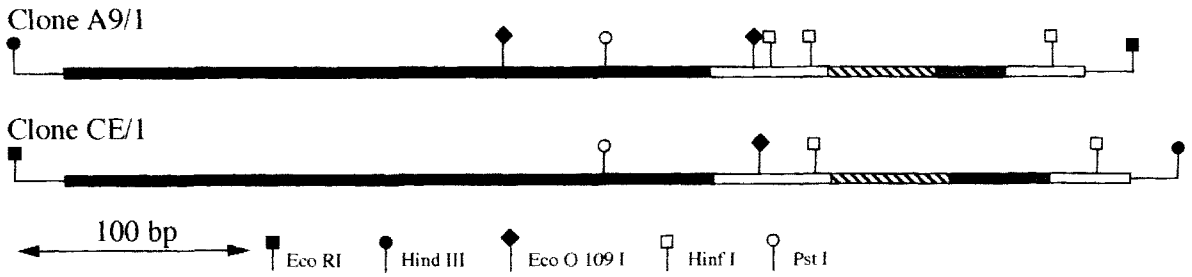

Fig. 1. Partial restriction map of the clones A9/1 and CE/1: Exon ■, intron □ sequences; $(gt)_n$ repeat ▨, $(ga)_m$ repeat ▨.

retarded band of the control (100%). The velocity of the protein/DNA complex formation as well as the stability of the protein binding to the target DNA was determined from 2 min to 4 h at 4°C and at room temperature before native electrophoresis.

Salt and Temperature Stability

Salt stability of the protein/DNA complex was determined by addition of 20–1,000 mM NaCl or KCl directly to the standard/reaction mix prior to the addition of the labelled target DNA. The temperature sensitivity of proteins was studied by preincubating 5 μg aliquots at 42°C and 56°C for 10, 20, 40 min, and at 100°C for 10 min, respectively. Treated proteins were stored at −70°C and then applied for gel retardation assays.

Separation of the Water-Soluble and -Insoluble Protein Fractions of Total Nuclear Extracts From HeLa Cells

Total HeLa nuclear extract was fractionated into a water soluble (FS) and insoluble fraction (FI) by centrifugation (40,000*g*, 30 min, 4°C). The FI proteins were resuspended in dialysis buffer (25 mM HEPES/KOH, pH 7.6, 40 mM KCl, 0.1 mM EDTA, 10% glycerol, and 1 mM DTT) to a concentration of 2.5 μg/μl. Aliquots of the fractions containing equal amounts of protein were remixed and stored together with the fractions at −70°C until further use in the standard gel retardation assay. To study temperature sensitivity of the separated fractions 2.5 μg protein containing aliquots were exposed to 42°C or 56°C for 10, 20, and 40 min each and for 10 min at 100°C. Heat-treated proteins were remixed with the corresponding untreated proteins stored at −70°C and then assayed by gel retardation.

Southwestern Analyses

Thirty to sixty μg of total HeLa nuclear proteins were separated in 11% polyacrylamide/SDS gels [Laemmli, 1970]. Proteins were electrotransferred using a liquid blotting chamber onto PVDF membranes (Millipore, Bedford, MA; 5 h, 400 mA, 4°C; transfer buffer 25 mM Tris, 192 mM glycine, pH 8.6). Transfer efficiency was controlled by staining the membranes with 0.1% amido black and by standard Coomassie staining of the gels. Proteins were denatured 6 M guanidine-hydrochloride in binding buffer [BB: 10% (v/v) glycerol, 20 mM HEPES, pH 7.9, 50 mM KCl, 1 mM DTT, and 0.1% (v/v) NP-40]and renatured in BB. Membranes were blocked with 0.5% acetylated BSA. Incubations with labelled DNA probe were carried out with 10^6 cpm/ml BB containing 12.5 μg pdIdC and 25 μg heat denatured *E. coli* DNA (5 h, 4°C). Membranes were washed under different "stringency" conditions with BB containing 50–250 mM KCl at room temperature and exposed to Fuji RX X-ray films using intensifying screens at −70°C.

DNase I Footprint Assays

For footprint analyses the double stranded probe A9 138r was end labelled either at the EcoO109I-site [$(tc)_{15}(ac)_{22}$-strand] or at the EcoRI-site [$(gt)_{22}(ga)_{15}$-strand]. Assays with total HeLa nuclear exctract were performed essentially as described [Altschmied et al., 1989; Baniahmad et al., 1990] using 20,000–45,000 cpm of labelled target DNA/lane; 0.1–0.25 U of DNase I (RNase free, Boehringer Mannheim) was used for reactions without, as well as 0.75–2 U for reactions with nuclear extract. DNA fragments were separated in 8% sequencing gels containing 8 M urea. Gels were fixed (10% acetic acid and 12% methanol), dried and exposed to Fuji RX X-ray films using intensifying screens at −70°C.

RESULTS

Binding Properties of Proteins From Total HeLa Nuclear Extracts to $(gt)_n(ga)_m$ Stretches

Binding constants for the protein/DNA interaction were determined by using total HeLa nuclear extract and increasing amounts of labelled target DNA in the gel retardation assay system. Retarded bands were quantified and analyzed according to Scatchard [1949]. One typical experiment is presented in Figure 2, using A9 102r target DNA. Binding constants were found to be 1.5×10^{-8} Mol/ltr ($s_x \pm 2.32 \times 10^{-9}$, n = 4) for A9 102r and 1.33×10^{-8} Mol/ltr (n = 1) for CE 120r DNA. To determine the specificity of the complexes, competition experiments (Fig. 3) were carried out by addition of unlabelled specific (CE 138r, Fig. 1) or non-specific (C4 111r) competitor DNA in the gel retardation assay. Complete competition of the complex formed with the probe CE 120r was observed at 500-fold molar excess of competitor CE 138r (Fig. 3A,C, left panel). In contrast, no competition was detectable using the unrelated

sequence C4 111r (Fig. 3B, right panel), suggest-ing that the complex formation is specific for the target DNA. To study the kinetics of the protein/DNA complex formation, time course experiments were performed in gel retardation assays (Fig. 4). Total nuclear extract formed the full pattern of shifted bands within 2–5 min at 4°C. This pattern was stable for up to 2 h during incubation (Fig. 4A and data not shown). At room temperature the full pattern of shifted bands was observed within less than 2 min. After 20 min of incubation the level of the major shifted band decreased (Fig. 4B).

Heat Sensitivity of the Proteins and Salt Stability of the Protein/DNA Complex

To analyze temperature sensitivity of the proteins participating in the protein/DNA complex formation total HeLa nuclear extract was preincubated at higher temperatures for different times and then analyzed for DNA binding by gel retardation. Exposure of the proteins to 42°C for 10 min resulted in significant changes of the normal band shift pattern (Fig. 5). The intensity of one low mobility and of the major retarded band drastically decreased, while that of two minor retarded bands appeared amplified. Preincubation for 20 min at 56°C resulted in a similar although weaker signal pattern. Incubation for longer times or at higher temperatures led to total loss of any retarded band.

Salt stability of the protein/DNA complex was studied in the gel retardation system by raising the concentration to up to 1 M NaCl or KCl in the standard reaction mix (Fig. 6). Increasing concentrations of both salts resulted in a more complex pattern of retarded bands mainly due to a splitting of the major retarded band into a lower and two higher mobility bands. This phenomenon was more obvious for KCl than for NaCl. Except for the highest mobility complex, the pattern strongly resembles to the one detected after preincubation of the proteins at higher temperatures indicating similar mechanisms leading to the appearance of the novel retarded bands. Comparable results were obtained using the A9 102r probe as target DNA (not shown).

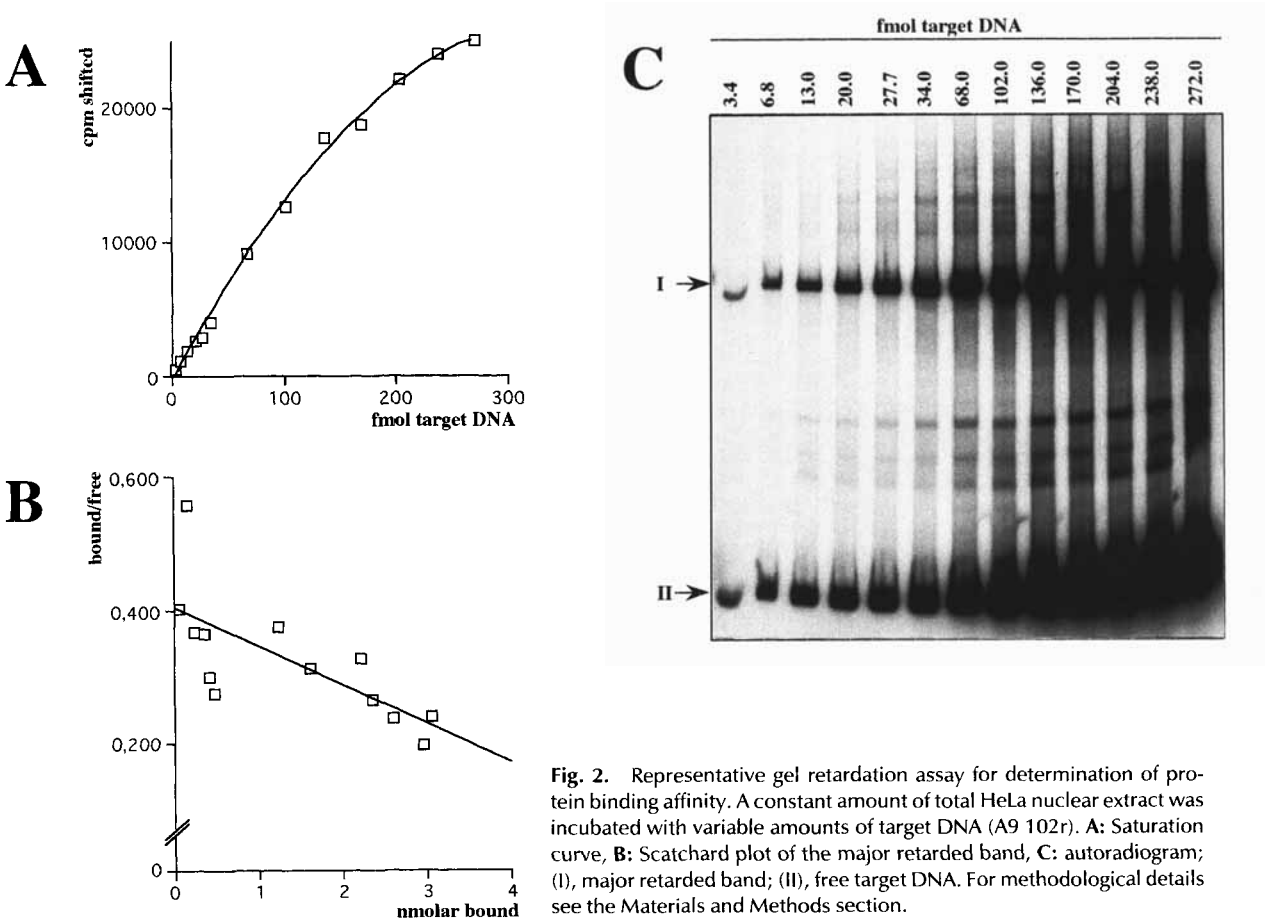

Fig. 2. Representative gel retardation assay for determination of protein binding affinity. A constant amount of total HeLa nuclear extract was incubated with variable amounts of target DNA (A9 102r). **A:** Saturation curve, **B:** Scatchard plot of the major retarded band, **C:** autoradiogram; (I), major retarded band; (II), free target DNA. For methodological details see the Materials and Methods section.

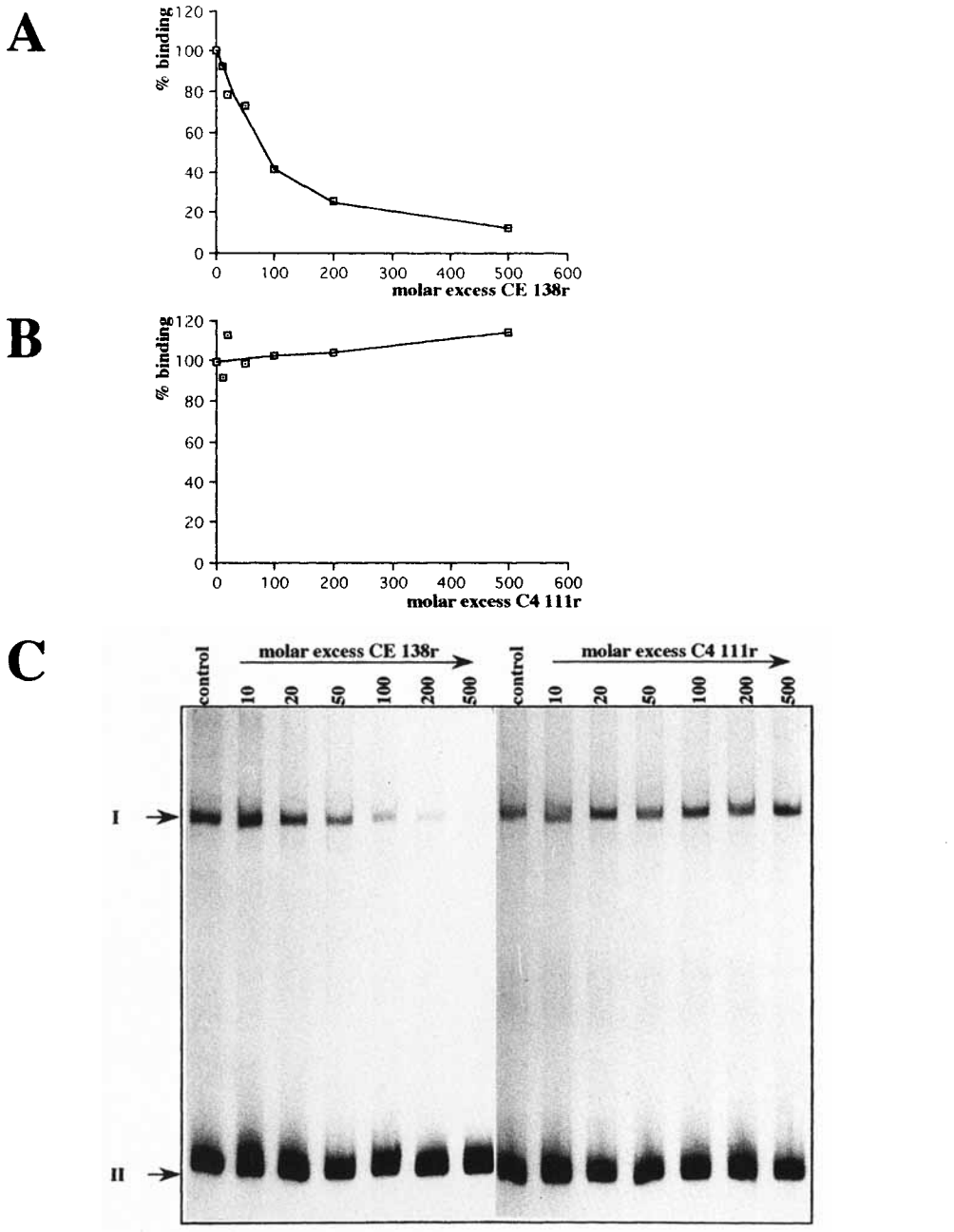

Fig. 3. Competition for protein binding of total HeLa nuclear extract to the CE120r target DNA using CE 138r (**A**) and C4 111r (**B**) as unlabeled competitor DNA. (**C**) Autoradiogram; (I) major retarded band, (II) free target DNA.

At Least Two Protein Components Bind to $(gt)_n(ga)_m$ Containing Target DNAs

In order to determine the molecular weight of nuclear proteins binding to $(gt)_n(ga)_m$ targets, Southwestern experiments using CE 120r and A9 138r target DNAs were performed. Besides very small amounts of higher and lower molecular weight polypeptides, two major distinct polypeptides were labeled by both targets exhibiting molecular weight of 60 and 57 kD (Fig. 7). A control experiment using the unrelated $(gaa)_{24}$ containing C4 111r as target DNA revealed, among others, the same strong signals corresponding to 60 and 57 kD polypeptides. Similar results were obtained using two small HinfI fragments of pUC18, two probes consisting of 5′-flanking regions of the A9 $(gt)_n(ga)_m$, and a $(gt)_{24}$ containing T-cell receptor sequence (data not shown). Since the probe C4 111r did not

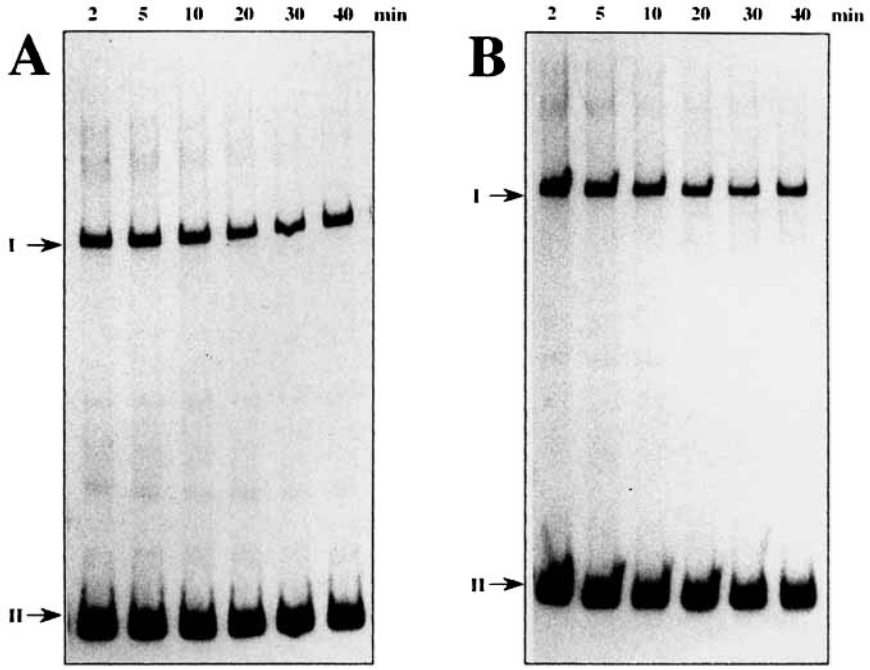

Fig. 4. Kinetics of DNA/protein complex formation using total HeLa nuclear extract at 4°C (**A**) and at room temperature (**B**) as analysed in gel retardation assays using A9 102r as target DNA. (I) major retarded band; (II) free target DNA.

compete for protein binding to the $(gt)_n(ga)_m$ containing target DNAs, binding to the 60 and 57 kD polypeptides has to be classified as non-specific. No signal specific for a $(gt)_n(ga)_m$ harbouring target DNA has been identified. These results suggest that either two or more components of different size are required for specific binding or binding capacity is lost during the denaturing/renaturing procedure.

In a first attempt to separate the putative components of the binding complex, total HeLa nuclear extract was fractionated by centrifugation into a water soluble (FS) and an insoluble protein fraction (FI). Both fractions were analyzed for complex formation by gel retardation (Fig. 8). Neither fraction produced the major retarded band obtained with the total nuclear extract. Yet the major retarded band is restored upon remixing of equal protein amounts of both fractions. These data underscore the aforementioned interpretation that the major protein/DNA complex is composed of at least two components, one of which is water soluble and the other one is water insoluble.

In order to determine whether one or both protein components are responsible for the observed temperature sensitivity (Fig. 5), fractions FS and FI were independently heat treated and remixed with the respective untreated proteins (Fig. 9). Gel retardation experiments using the mixtures containing the heated FS proteins revealed that this component is subject to virtually complete inactivation within 10 min at 56°C. In contrast, FI proteins incubated for 10 min at 100°C still gave rise to the protein/DNA complex as observed with normal extract.

Four Binding Sites Are Limited to One Strand in the Double Stranded $(gt)_m(ga)_m$ Target

In order to localize more precisely the DNA target sequence(s) bound by the nuclear proteins, DNase I footprinting analyses using total HeLa nuclear extract and A9 138r as double stranded target DNA were carried out. No protection was observed in the $(tc)_{15}(ac)_{22}$-containing strand (Fig. 10A). In contrast, when the $(gt)_{22}(ga)_{15}$ was labelled, four clearly protected regions in the target DNA were detected (Fig. 10B). Region I spans some 24 nucleotides 5′ of the $(gt)_{22}$ repeat and includes the first $(gt)_2$ unit. Region II includes $(gt)_6$ and region III $(gt)_{8-9}$. The fourth binding site is located in the $(ga)_{15}$ part of the mixed repeat and comprises at least $(ga)_8$. This binding site is bordered by a hypersensitive site for DNase I (Fig. 10B,C). Similar results were obtained for the $(gt)_n(ga)_m$-containing strand of the probe CE 138r (data not shown).

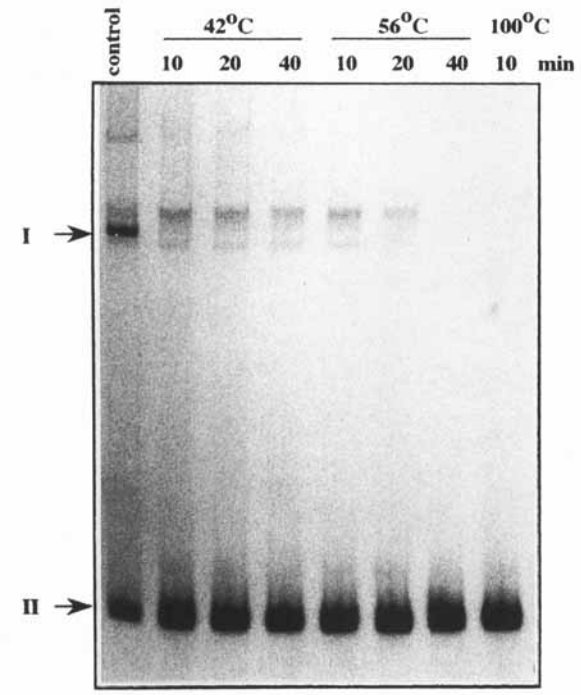

Fig. 5. Temperature sensitivity of total HeLa nuclear extract analysed in a gel retardation assay using A9 102r as target DNA. Control, untreated nuclear extract; (I), major retarded band of the control; (II), free target DNA.

These data demonstrate that nuclear proteins bind to a substantial part of the mixed repeat including an adjacent 5′ non-repetitive sequence preferentially protecting one strand.

DISCUSSION

In this study we demonstrate that proteins of total HeLa nuclear extracts bind rapidly and with a measurable, not random affinity [for comparison see Meisterernst et al., 1988] to simple $(gt)_n(ga)_m$ repeats in their natural DNA environment. In competition experiments the formation of the protein/DNA complex is prevented by the respective unlabelled DNA whereas the unrelated $(gaa)_{24}$ containing competitor DNA had no effect. Thus, specific interaction with the proteins can be assumed.

Application of increasing amounts of salt induced several retarded bands in addition to the major protein/DNA complex. This effect was already visible at physiological salt concentrations (100–150 mM). A large portion of the target DNA consists of $(gt)_n$ repeats, i.e., alternating purin/pyrimidine stretches which could give rise to the formation of Z-DNA under these circumstances [Haniford and Pulleyblank, 1983; Herbert and Rich, 1993; Nordheim and Rich, 1983]. Perfect $(ga)_m$ repeats (polypurines) are

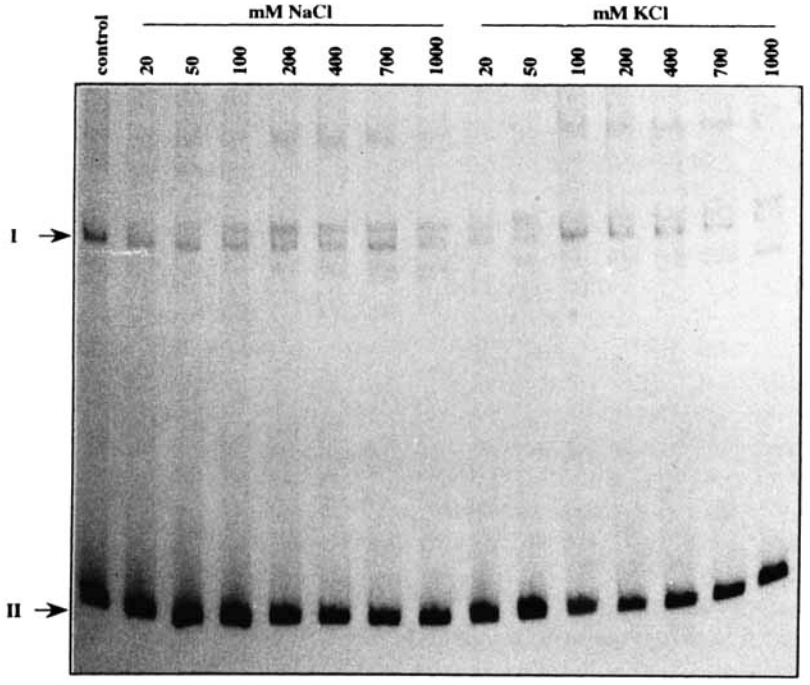

Fig. 6. Influence of increasing concentrations of NaCl and KCl on the pattern of retarded bands with total HeLa nuclear extract using CE 120r as target DNA. Control, pattern without additional salt; (I) major retarded band; (II) free target DNA.

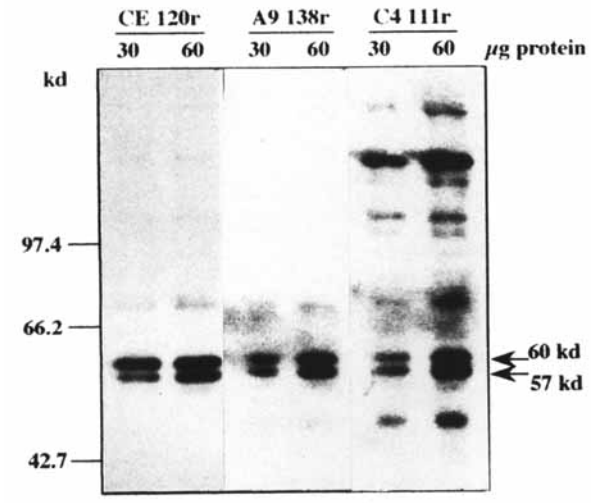

Fig. 7. Southwestern analyses of total HeLa nuclear extract using $(gt)_n(ga)_m$ containing (CE 120r, A9 138r) and $(gaa)_{24}$ containing (C4 111r) target DNAs. Membranes were incubated with 1.5×10^6 cpm/ml and washed four times for 5 min in BB at room temperature. Exposure times were 24 h with screen for C4 111r and A9 138r, respectively, 336 h with screen for CE 120r. Molecular weight markers are indicated at the left. Arrows point to the major polypeptides detected.

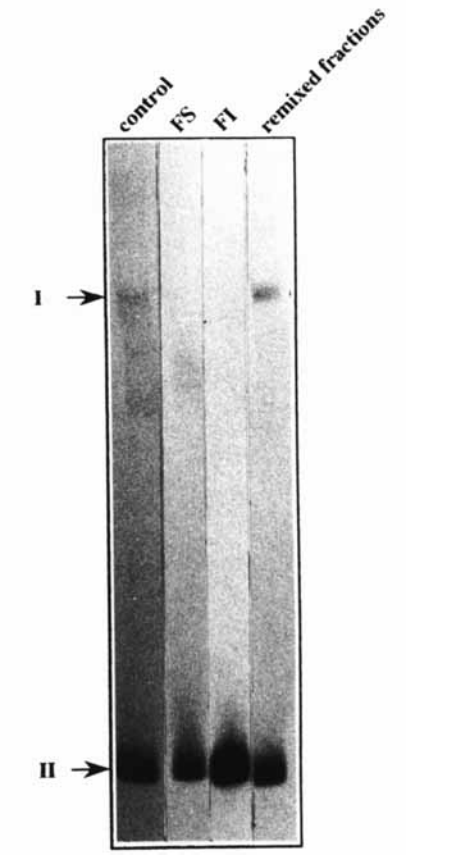

Fig. 8. Gel retardation assay of FS, FI, and recombined proteins using CE 120r as target DNA. Control was performed with total nuclear extract. 5 μg of protein were analyzed per lane. (I), major retarded band; (II) free target DNA.

able to form H- or H*-DNA under certain conditions [Lyamichev et al., 1985; Martinez-Balbas and Azorin 1993]. Thus it is possible that the different complexes observed at higher salt concentrations are due to changes in DNA structure. Extracts which had been previously heat treated gave rise to similar banding patterns. While higher salt concentrations may induce changes in DNA conformation, heat-treated protein fractions can exclusively induce changes in the protein structure by stepwise denaturation. Therefore it is likely that the observed salt induced changes are due to variations in protein rather than in DNA conformation. The diversified pattern of retarded bands suggests that several polypeptides participate in the protein/DNA complex formation.

The attempt to separate the protein components of the binding complex by centrifugation resulted in two inactive protein fractions (FS and FI). Binding activity of FS and FI was restored when both fractions were recombined. Thus, at least two different polypeptides are necessary for DNA binding. The reconstitution experiments showed that one component is temperature sensitive and water soluble, the other one is heat stable and water insoluble. Separation of the components by SDS PAGE could also explain the lack of specific signal in Southwestern experiments. These data underscore the notion that at least two polypeptides are necessary for binding to $(gt)_n(ga)_m$ containing DNA.

In order to determine the protein binding sites in our double stranded target DNA, DNase I footprint analyses were performed. Four different protected regions were protected only on the $(gt)_{22}(ga)_{15}$ strand: two regions were found to include nearly the complete $(gt)_{22}$, and another region comprises more than half of the $(ga)_{15}$ part, suggesting that this composition may be essential for binding of the particular protein complex. This hypothesis is supported by the fact that neither single stranded $(gt)_8$ and/or $(ga)_8$ nor the corresponding double stranded oligonucleotides compete for protein binding (Mäueler et al. unpublished observations). In addition, a $(gt)_{24}$ containing intronic T-cell receptor sequence did not compete, although this sequence has been found to bind other HeLa nuclear protein(s) (Mäueler et al., unpublished data). Thus, the mixed $(gt)_n(ga)_m$ repeat seems to be required for the formation of the complex

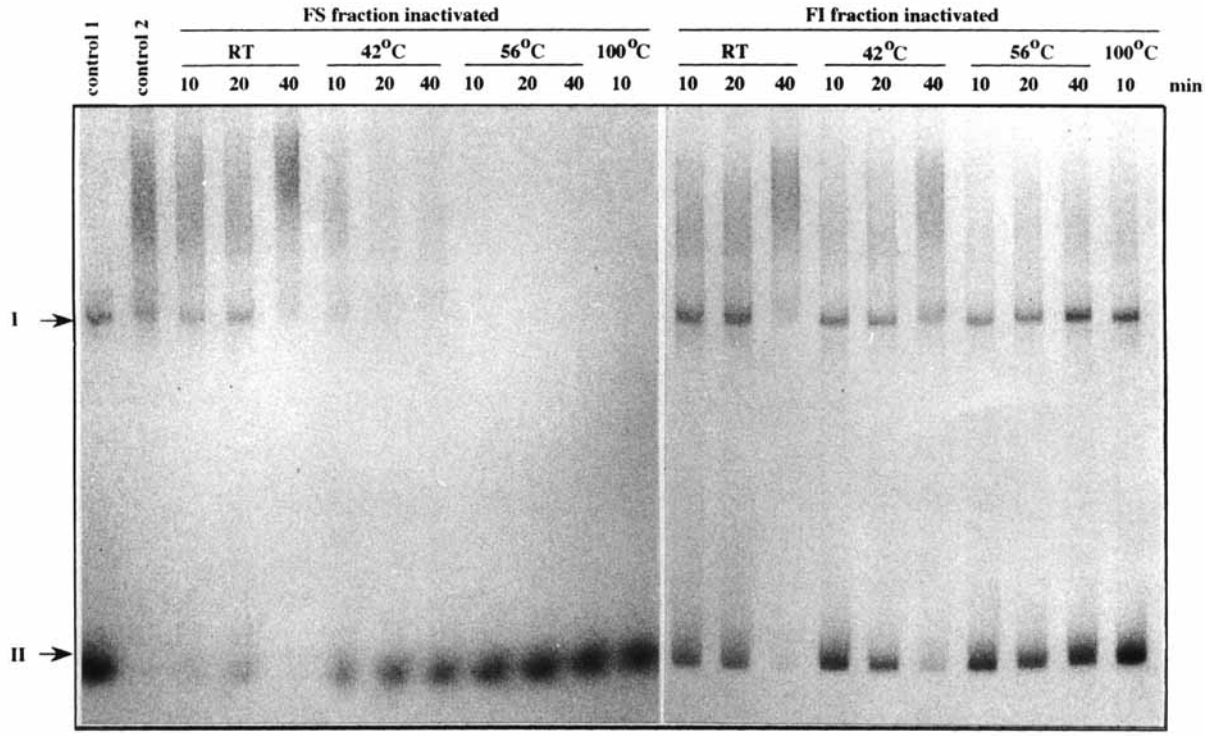

Fig. 9. Temperature sensitivity of FS and FI components of total HeLa nuclear extract in gel retardation assays using CE 120r as target DNA. Control 1 was carried out with 5 μg of total cellular HeLa extract, control 2 using 5 μg of untreated recombined nuclear protein. (I) major retarded band; (II) free target DNA. Note: Reconstitution of the combined FS and the FI proteins leads to increased unspecific protein/DNA interactions.

described here. Besides the protected $(ga)_{8-9}$, the protected region located 5′ of the $(gt)_n$-repeat may be involved in specifying the protein(s) or protein complexes bound by the simple repeated DNAs. The $(gt)_{22}(ga)_{15}$ containing *HLA-DRB* intron 2 and the $(gt)_{24}$-harbouring T-cell receptor intron consist of completely different 5′ ends [Gomolka et al., 1993; Rieß et al., 1990].

So far little is known about the function of simple repetitive sequences. Few reports describe interactions between nuclear proteins and single or double stranded $(gt)_n$ and/or $(ga)_m$ sequences (either isolated or embedded in flanking sequences). Initially, moderate enhancer activity of $(gt)_{15}$ to $(gt)_{25}$ stretches has been observed in transfection experiments [Hamada et al., 1984]. On the other hand, simple $(gt)_n$ repeats have been proposed as recombination hotspots [Pardue et al., 1987], but biochemical evidence is not available [Stallings et al., 1991]. Recently binding activity to double stranded $(ca)_n/(gt)_n$ oligonucleotides was detected in HeLa nuclear extract [Richards et al., 1993]. Since their nuclear extracts did not contain the water insoluble protein fraction of the total ammonium sulfate precipitable HeLa nuclear proteins, we assume that the protein(s) bound to the oligonucleotides are different from those bound to our mixed repeat containing sequence.

A particular transcription factor (GAGA factor) recognizes $(ga)_m$ promoter elements in *Drosophila* [Biggin and Tjian 1988; Lu et al., 1993]. The longest $(ga)_m$ element binding the GAGA factor was $(ga)_{10}$, located in the promoter region of *Drosophila* histone 3 and 4 genes [Gilmour et al., 1989]. These $(ga)_m$ elements do not harbor any $(gt)_n$ units and they are embedded in sequences completely different from our target DNAs. Vertebrate homologues of the GAGA-transcription factor are not known. Specific protein binding to single stranded $(tc)_{16}$ oligonucleotides was described, but not to the double stranded equivalents [Yee et al., 1991]. The reason for the discrepancy to the data of Richards et al. [1993; see above] with HeLa cell extracts is unclear. Both, the GAGA-transcription factor and the single strand $(tc)_{16}$ binding factor seem to be distinct from the proteins reported here. A $(ga)_{27}$ repeat has been identified in the origin of replication of the hamster rhodopsin gene locus [Gale et al., 1992] as well as a mixed $(ca)_{19} \times (ga)_{20} \times (ga)_{28}$ repeat in an amplified hamster

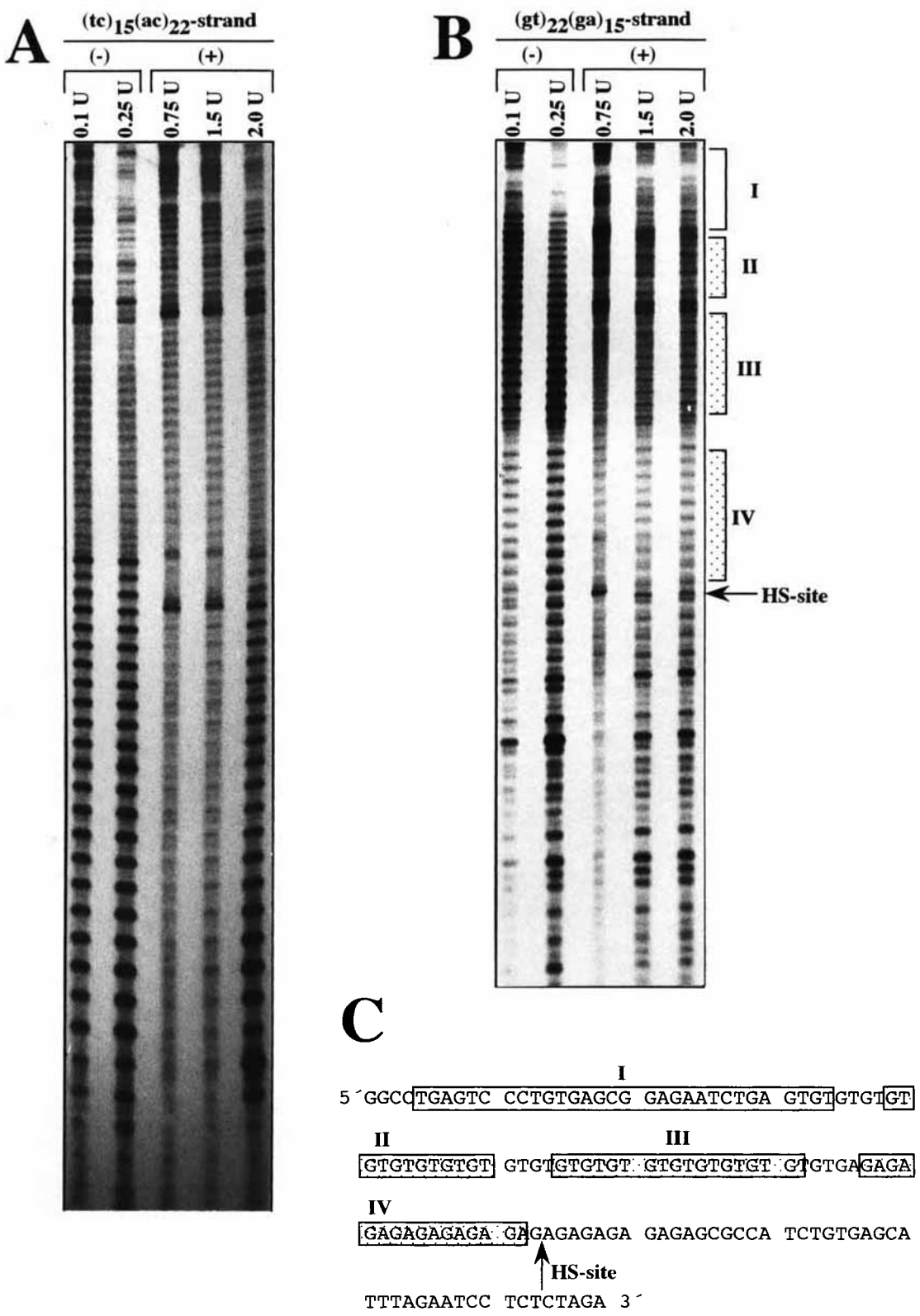

Fig. 10. DNase I footprinting of both the $(tc)_{15}(ac)_{22}$ (**A**) and the $(gt)_{22}(ga)_{15}$ (**B**) strands of A9 138r using 80 μg of total HeLa nuclear extract and variable amounts of DNase I. 15.3 fmol (A) and 30.6 fmol (B) of endlabeled target-DNA was included in each sample. (**C**) DNA-sequence of the protein binding strand; empty boxes indicate a protein binding site outside the repeats; shadowed boxes show protein binding sites within the mixed simple repeat region (+) with, (−) without protein; U, units DNase I; I–IV, footprints I–IV; HS-site, DNase I hypersensitive site.

dihydrofolate reductase replicon [Caddle et al., 1990]. The authors suggest an implication of these simple repetitive elements in replication. Yet no specific interactions of nuclear protein(s) with these sequences are evident. So far the other simple repeat binding proteins have neither been characterized with respect to salt or temperature sensitivity nor have they been fractionated. In addition, the specific recognition site(s) in the target DNAs are not thoroughly analyzed. Yet it is unlikely that the binding factors described here are identical to one of the described proteins. It should be stressed that most of the binding studies reported elsewhere were performed using synthetic $(gt)_n$ or $(ga)_m$ oligonucleotides, while we used cloned genomic DNA fragment harbouring mixed $(gt)_n(ga)_m$ repeats plus flanking sequences. As these mixed repeats as well as additional 5′ and 3′ sequences characterize *MHC-DRB* genes, further studies may finally lead to an understanding of the biological role of these enigmatic elements.

ACKNOWLEDGMENTS

We thank C. Plehn for documentation of the data, Dr. J. Altschmied (Würzburg) for help with the Southwestern technique, Dr. O. Rieß for critical reading of the manuscript, and an anonymous reviewer for thoughtful comments. This work was supported by grants of the VW- and Sander-Stiftung.

REFERENCES

Altschmied J, Muller M, Baniahmad A, Steiner C, Renkawitz R (1989): Cooperative interaction of chicken lysozyme enhancer sub-domains partially overlapping with a steroid receptor binding site. Nucleic Acids Res 17:4975–4991.

Baniahmad A, Steiner C, Köhne AC, Renkawitz R (1990): Modular structure of a chicken lysozyme silencer: Involvement of an unusual thyroid hormone receptor binding site. Cell 61:505–514.

Biggin M, Tjian R (1988): Transcription factors that activate the *ultrabithorax* promoter in developmentally staged extracts. Cell 53:699–711.

Bradford MM (1976): A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal Biochem 72:248–254.

Brook JD, McCurrach ME, Harley HG, Buckler AJ, Church D, Aburatani H, Hunter K, Stanton VP, Thirion JB, Hudson T, Sohn R, Zemelman B, Snell RG, Rundle SA, Crow S, Davies J, Shelbourne P, Buxton J, Jones C, Juvonen V, Johnson K, Harper PS, Shaw DJ, Housman DE (1992): Molecular basis of myotonic dystrophy: expansion of a trinucleotide (CTG) repeat at the 3′end end of a transcript encoding a proteinkinase family member. Cell 68:799–808.

Caddle MS, Lussier RH, Heintz NH (1990): Intramolecular DNA triplexes, bent DNA and DNA unwinding elements in the initiation region of an amplified dihydrofolate reductase replicon. J Mol Biol 211:19–33.

Epplen C, Melmer G, Siedlaczck I, Schwaiger FW, Mäueler W, Epplen JT (1993): On the essence of "meaningless" simple repetitive DNA in eukaryote genomes. In Pena SDJ, Chakraborty R, Epplen JT, Jeffreys AJ (eds): "DNA Fingerprinting: State of the Science." Basel: Birkhäuser Verlag, pp 29–45.

Fu YH, Kuhl DPA, Pizzuti A, Pieretti M, Sutcliffe JS, Richards S, Verkerk AJM, Holden JJA, Fenwick RG, Warren ST, Oostra BA, Nelson DL, Caskey CT (1991): Variation of the CGG repeat at the fragile X site results in genetic instability: Resolution of the Sherman paradox. Cell 67:1047–1058.

Gale JM, Tobey RA, D'Anna JA (1992): Localization and DNA sequence of a replication origin in the rhodopsin gene locus of chinese hamster cells. J Mol Biol 224:343–358.

Gilmour DS, Thomas GH, Elgin SCR (1989): *Drosophila* nuclear proteins bind to regions of alternating C and T residues in gene promoters. Science 245:1487–1490.

Gomolka M, Epplen C, Buitkamp J, Epplen JT (1993): Novel members and germline polymorphisms in the human T-cell receptor *Vb6* family. Immunogenetics 37:257–265.

Hamada H, Seidman M, Howard BH, Gorman CM (1984): Enhanced gene expression by the poly(dT-dG)·poly(dC-dA) sequence. Mol Cell Biol 4:2622–2630.

Haniford DB, Pulleyblank DE (1983): Facile transition of poly[d(TG)·d(CA)] into a left-handed helix in physiological conditions. Nature 302:632–634.

Herbert AG, Rich A (1993): A method to identify and characterize Z-DNA binding proteins using a linear oligodeoxynucleotide. Nucleic Acids Res 21:2669–2672.

Knight SJL, Flannery AV, Hirst MC, Campbell L, Christodoulou Z, Phelps SR, Pointon J, Middleton-Price HR, Barnicoat A, Pembrey ME, Holland J, Oostra BA, Bobrow M, Davies KE (1993): Trinucleotide repeat amplification and hypermethylation of a CpG island in FRAXE mental retardation. Cell 74:127–134.

Laemmli UK (1970): Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227:680–685.

La Spada AR, Wilson EM, Lubahn DB, Harding AE, Fischbeck KH (1991): Androgen receptor gene mutations in X-linked spinal and bulbar muscular atrophy. Nature 352:77–79.

Lu Q, Wallrath LL, Granok H, Elgin SCR (1993): $(CT)_n$·$(GA)_n$ repeats and heat shock elements have distinct roles in chromatin structure and transcriptional activation of the *Drosophila hsp26* gene. Mol Cell Biol 13:2802–2814.

Lyamichev VI, Mirkin SM, Frank-Kamenetskii MD (1985): A pH-dependent structural transition in the homopurine-homopyrimidine tract in superhelical DNA. J Biol Struct 3:327–338.

Martinez-Balbas A, Azorin F (1993): The effect of zinc on the secondary structure of $(dGA·TC)_n$ DNA sequences of different length: A model for the formation *H-DNA. Nucleic Acids Res 21:2557–2562.

Mäueler W, Muller M, Köhne AC, Epplen JT (1992): A gel retardation assay system for studying protein binding to simple repetitive DNA sequences. Electrophoresis 13:7–10.

Meisterernst M, Gander I, Rogge L, Winnacker EL (1988): A quantitative analysis of nuclear factor I/DNA interactions. Nucleic Acids Res 16:4419–4435.

Nordheim A, Rich A (1983): The sequence $(dC\text{-}dA)_n\cdot(dT\text{-}dG)_n$ forms left-handed Z-DNA in negatively supercoiled plasmids. Proc Natl Acad Sci USA 80:1821–1825.

Orr HT, Chung M, Banfi S, Kwiatkowski TJ, Servadio A, Beaudet AL, McCall AE, Duvick LA, Ranum LPW, Zoghbi HY (1993): Expansion of an unstable trinucleotide CAG repeat in spinocerebellar ataxia type 1. Nature Genet 4:221–226.

Pardue ML, Lowenhaupt K, Rich A, Nordheim A (1987): $(dC\text{-}dA)_n$ $(dT\text{-}dA)_n$ sequences have evolutionarily conserved chromosomal locations in *Drosopholia* with implications for roles in chromosome structure and function. EMBO J 6:1781–1789.

Richards RI, Holman K, Yu S, Sutherland GR (1993): Fragile X syndrome unstable element, $p(CCG)_n$, and other simple tandem repeat sequences are binding sites for specific nuclear proteins. Hum Mol Genet 2:1429–1435.

Rieß O, Kammerbauer C, Roewer L, Steimle V, Andreas A, Albert E, Nagai T, Epplen JT (1990): Hypervariability of intronic simple $(gt)_n(ga)_m$ repeats in *HLA-DRB* genes. Immunogenetics 32:110–116.

Scatchard G (1949): The attractions of proteins for small molecules and ions. Ann NY Acad Sci 51:660–672.

Schwaiger FW, Weyers E, Epplen C, Brün J, Ruff G, Crawford A, Epplen JT (1993): The paradox of *MHC-DRB* exon/intron evolution: α-helix and β-sheet encoding regions diverge while hypervariable intronic simple repeats coevolve with β-sheet codons. J Mol Evol 37:260–272.

Stallings RL, Ford AF, Nelson D, Torney DC, Hildebrand CE, Moyzis RK (1991): Evolution and distribution of $(gt)_n$ repetitive sequences in mammalian genomes. Genomics 10:807–815.

The Huntington's Disease Collaborative Research Group (1993): A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's Disease chromosomes. Cell 72:971–983.

Yee HA, Wong AKC, van DeSande JH, Rattner JB (1991): Identification of novel single-stranded $d(TC)_n$ binding proteins in several mammalian species. Nucleic Acids Res 19:949–953.

Zhen L, Swank RT (1993): A simple and high yield method for recovering DNA from agarose gels. BioFeedback 14: 894–898.