

Specific targeting of insect and vertebrate telomeres with pyrrole and imidazole polyamides

Kazuhiro Maeshima, Samuel Janssen and Ulrich K. Laemmli¹

Departments of Biochemistry and Molecular Biology, University of Geneva, 30 Quai Ernest-Ansermet, CH-1211 Geneva 4, Switzerland

¹Corresponding author
e-mail: Ulrich.Laemmli@molbio.unige.ch

K. Maeshima and S. Janssen contributed equally to this work

DNA minor groove-binding compounds (polyamides) that target insect and vertebrate telomeric repeats with high specificity were synthesized. Base pair recognition of these polyamides is based on the presence of the heterocyclic amino acids pyrrole and imidazole. One compound (TH52B) interacts uniquely and with excellent specificity ($K_d = 0.12$ nM) with two consecutive insect-type telomeric repeats (TTAGG). A related compound, TH59, displays high specificity ($K_d = 0.5$ nM) for tandem vertebrate (TTAGGG) and insect telomeric repeats. The high affinity and specificity of these compounds were achieved by bidentate binding of two flexibly linked DNA-binding moieties. Epifluorescence microscopy studies show that fluorescent derivatives of TH52B and TH59 stain insect or vertebrate telomeres of chromosomes and nuclei sharply. Importantly, the telomere-specific polyamide signals of HeLa chromosomes co-localize with the immunofluorescence signals of the telomere-binding protein TRF1. Our results demonstrate that telomere-specific compounds allow rapid estimation of relative telomere length. The insect-specific compound TH52 was shown to be incorporated rapidly into growing Sf9 cells, underlining the potential of these compounds for telomere biology and possibly human medicine.

Keywords: DNA minor groove/DNA sequence-specific polyamides/telomeres/TRF1

Introduction

One of the striking features of eukaryotic chromosomes is their low gene density. Human chromosome 21 is especially gene poor; although it contains 33.55 Mbp, it harbors only an estimated 284 genes (Hattori *et al.*, 2000). One particularly long non-genic DNA stretch of chromosome 21 encompasses 7 Mb and contains only one gene. This kind of gene-poor region embraces almost 10 Mb or one-third of chromosome 21. The functions (if any) of these and other non-genic regions are unknown, and genetic tools to dissect their roles are largely lacking. To overcome this experimental void, we propose that sequence-specific artificial proteins and/or small molecules may serve as tools. Upon binding to the non-genic DNA of interest, it is hoped that such molecules can

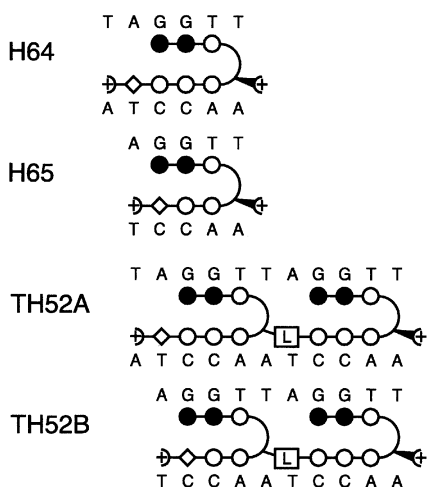
perturb and possibly reveal the functions of the targeted DNA elements.

Attempts at achieving the above goal led to the synthesis of an artificial protein termed MATH (for multi-AT hooks), which bound non-genic, AT-rich regions called SARs (scaffold-associated regions) with high specificity. SARs are candidate *cis*-acting elements of chromosome dynamics (for a review see Hart and Laemmli, 1998). In support of such a role, we observed that MATH20 (an 80 kDa protein containing 20 AT hooks) specifically inhibited chromosome condensation in mitotic *Xenopus* extracts (Strick and Laemmli, 1995). Moreover, expression of MATH20 in *Drosophila melanogaster* was found to suppress the position effect variegation (PEV) phenotype of *white-mottled* (w^{m4}) flies (Girard *et al.*, 1998). We proposed that suppression of PEV is due to the binding of MATH20 to the SAR-like, centric satellite III, ensuring chromatin opening and a reduced spreading of its heterochromatic state toward the *white* gene.

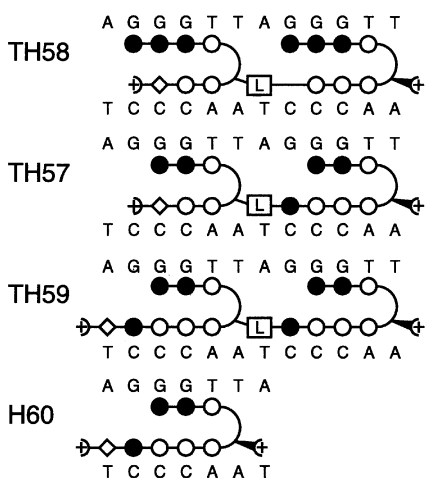
To extend this approach, we synthesized DNA-binding compounds (referred to as polyamides) composed of heterocyclic amino acids (see below), which bind different *D. melanogaster* satellites with high specificity (Janssen *et al.*, 2000b). Remarkably, these drugs, when fed to developing *Drosophila*, caused gain- or loss-of-function phenotypes with *white-mottled* or *brown-dominant* flies, respectively (Janssen *et al.*, 2000a). Both phenomena are explained at the molecular level by chromatin opening (increased accessibility) of the targeted DNA satellites. The biological insights obtained suggested that satellite sequences are not passive evolutionary residues, but essential components of gene regulation circuits. Our observations suggest that sequence-specific artificial proteins and polyamides can serve as powerful and innovative tools for many different applications, thereby yielding important biological information.

The data obtained with the aforementioned polyamides are based on recent important progress, which described the synthesis of such molecules (Geierstanger *et al.*, 1994). Polyamides composed of *N*-methylpyrrole (Py) and *N*-methylimidazole (Im) amino acids can target many predetermined DNA sequences with high specificity (for a recent review see Wemmer, 2000). Specific recognition of the base pairs is based on the principle that these linear compounds can fold to adopt a U-shaped conformation (hairpins) in the minor groove due to the presence of a flexible 'turn monomer' (γ -aminobutyric acid). As a result, the heterocyclic pyrrole and imidazole rings form side-by-side pairs that are accommodated in the minor groove. An Im/Py pair targets a G–C base pair, while a Py/Im pair recognizes C–G. A Py/Py pair is partially degenerate and binds both A–T and T–A base pairs (White *et al.*, 1997). A limitation of these compounds is that when these aromatic rings are coupled contiguously, it seems impossible to

TTAGG Compounds



TTAGGG Compounds



Symbols

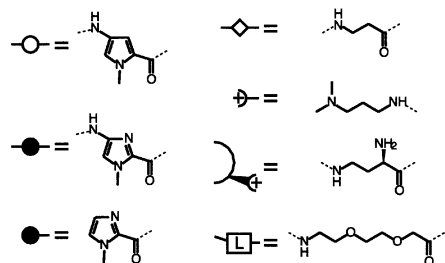


Fig. 1. Schematic structure of telomere compounds used in this study. *N*-methylpyrrole rings are represented by open circles. The flexible, hydrophilic linker (8-amino-3,6-dioxaoctanoic acid) is indicated by a boxed L. The C-terminal dimethylaminopropylamide (Dp) is represented by a plus sign. A diamond represents β -alanine. *N*-methylimidazole is represented by a black circle. The γ -turn monomer (*R*-2,4-diaminobutyric acid) is indicated by a curved line. The solid wedge attached to the plus sign (in the γ -turn monomer) represents the amino substitution at C2. Chemical structures of the symbols used are shown at the bottom.

target stretches of >7 bp without seriously compromising binding specificity and affinity. However, studies demonstrated that longer sequences can be targeted by tethering DNA-binding moieties with flexible linker molecules (Herman *et al.*, 1999; Janssen *et al.*, 2000b).

The *in vivo* drug experiments with *D.melanogaster* discussed above encouraged us to explore further the experimental potential of polyamides as tools and biological interference agents. The ends of chromosomes are capped by structures called telomeres. These subchromosomal structures are for several reasons 'ideal' and interesting polyamide targets: telomeres of most eukaryotes are defined by tandem short DNA repeats and encompass in humans a total length of ~300 kb per haploid genome. Hence, telomeric repeats are considerably less abundant than those of DNA satellites (several megabases) and targeting these subchromosomal regions is therefore more challenging. This problem is compounded by the fact that vertebrate telomeric repeats (TTAGGG) contain three consecutive Gs, which are considered to be a difficult polyamide target. Telomeres are positioned conspicuously at the ends of chromosomes, hence, it should be possible to evaluate unequivocally the specificity of such polyamides using fluorescently tagged compounds and fluorescence microscopy. Telomeres are non-genic sequences whose structure and function are well studied (for a review see McEachern *et al.*, 2000). This system allows a comparison of the biological effect obtained with polyamides and the effects obtained by genetic means. Last, but not least, telomere biology is often altered in cancer cells and is generally manifested by activation of telomerase (for reviews see Prescott and Blackburn, 1999; Oulton and Harrington, 2000). Although the relationship between telomere length, telomerase activity, senescence, and normal and neoplastic growth is a complex issue (Blackburn, 2000), telomere-specific polyamides may serve as new tools to address this issue and may lead to the development of agents that inhibit neoplastic growth.

The most dominant telomeric repeat of vertebrates consists of hexameric TTAGGG repeats (Meyne *et al.*, 1989). A related pentameric repeat (TTAGG) occurs at telomeres of many insect species (Sahara *et al.*, 1999). We describe here the synthesis of polyamides that interact differentially with either insect or vertebrate telomeric repeats with a remarkable specificity.

Results

Targeting insect TTAGG telomeric repeats

Two hairpin polyamides, H64 and H65, were designed to bind an insect TTAGG telomeric repeat. H64 and H65 differ by only one extra pyrrole at the C-terminus (Figure 1). The N-terminus of both hairpins carries two imidazole rings to accommodate the GG dinucleotide. Since a pyrrole ring C-terminal of imidazole (in hairpin polyamides) systematically shows a strong preference for GT over GA, the consensus sequence of H64 and H65 can thus be defined as WWGGTW (W = A or T base) and WGGTW, respectively. The structural basis for this is not fully understood (Kielkopf *et al.*, 2000).

Figure 2A shows a DNase I footprint obtained with H64 on a probe containing four tandem TTAGG repeats and

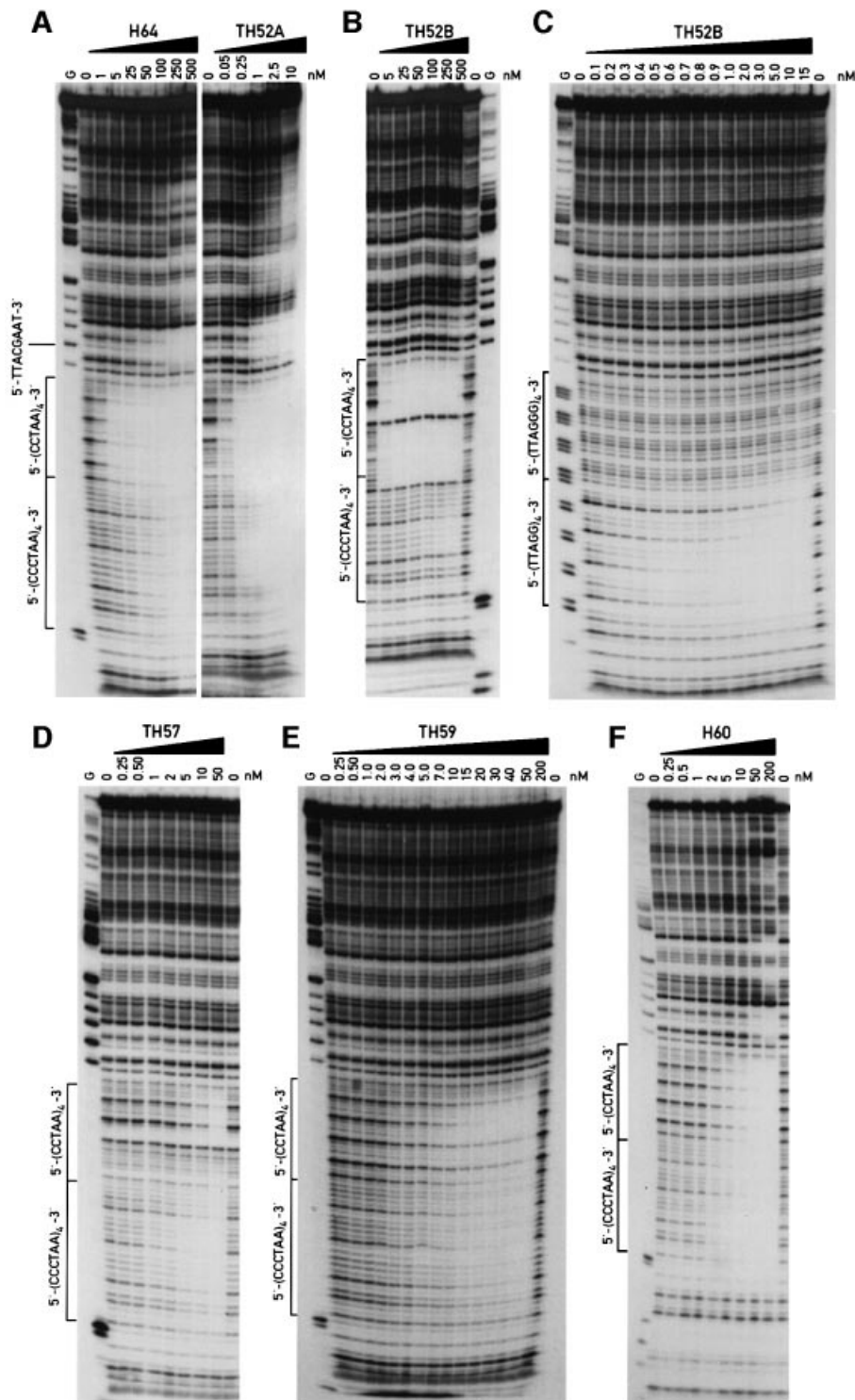


Fig. 2. DNase I footprinting experiments of the telomere-specific compounds. Ligand concentrations (in nM) are indicated at the top of each lane. The letter G refers to a G nucleotide cleavage reaction. (A) The monomeric hairpin H64 or its homodimer TH52A preferably binds TTAGG repeats (the bottom C-rich strand that contains AATCC is labeled) but also protects several mismatch sequences on the vector at higher concentrations. TH52A possesses increased affinity but unimproved specificity. (B) TH52B possesses excellent specificity for TTAGG repeats. Again, the bottom C-rich strand is shown. Unlike H64 and TH52A, high specificity is retained even at a concentration several hundred times that required for protection. The position of the band that is not protected in the CCTAA repeats by TH52B corresponds to the adenine base positioned on the linker. (C) Footprint of TH52B on the top G-rich strand of the same probe as in (A) and (B), showing that TH52B protects TTAGG repeats at subnanomolar concentrations. (D) Insertion of an unpaired imidazole in compound TH57 coupled to the linker generates binding preference for TTAGGG repeats over TTAGG repeats. The bottom C-rich strand is labeled. (E) When an imidazole is also inserted at the same position in the C-terminal hairpin (left), the resulting compound (TH59) has affinity similar to TH57 and binds TTAGG and TTAGGG repeats equally well. (F) The TH59 'monomer' H60 shows lower specificity for telomeric repeats than the tandem hairpin TH59, but similar affinity.

four tandem TTAGGG repeats. Examination of this panel demonstrates that H64 (left) protects the TTAGG repeats at subnanomolar concentrations ($K_{app} < 1$ nM), but interacts less well with the TTAGGG units ($K_{app} > 5$ nM). Closer inspection of the H64 footprint data reveals that at higher concentrations a number of additional sequences become protected. The most prominent occurs at the sequence TTACGAAT (indicated). H64 interacts with this sequence similarly well as with the hexameric repeats ($K_{app} > 50$ nM). Unexpectedly, the smaller compound H65 displayed only very weak affinity for either sequence; only at a concentration > 100 nM was some protection of insect repeats observed (not shown).

We previously demonstrated that dimerization of DNA-binding moieties with a flexible linker can result in an improved binding specificity (Janssen *et al.*, 2000b). This enhanced specificity is thought to occur by an enlargement of binding site size and a discrimination against partial match sites. To improve specificity for TTAGG repeats, H64 was therefore converted into a homodimer, termed TH52A, which is expected to bind the 11 bp sequence TAGGTTAGGTT as depicted in Figure 1. However, surprisingly, although TH52A had a 10- to 20-fold increased affinity compared with H64 (Figure 2A), its specificity was unimpressive since it bound many sites along the probe and protected almost the entire probe (coating) at a concentration > 2.5 nM.

In an attempt to increase specificity further, we next synthesized a heterodimer termed TH52B by linking H64 and H65. This dimer is expected to target the 10 bp sequence AGGTTAGGTT (Figure 1). DNase I footprinting revealed a remarkably improved specificity of interaction for TH52B (Figure 2B). Whereas the affinity was not significantly increased as compared with H64, the tandem hairpin TH52B displayed much higher specificity for its target binding site (TTAGG repeats), since it is almost devoid of affinity for hexameric TTAGGG repeats ($K_{app} \sim 500$ nM).

The equilibrium dissociation constant (K_d) for TH52B was determined at 0.12 nM by quantitative DNase I footprinting (Wade *et al.*, 1993; Figure 2C). Most importantly, TH52B, as opposed to its monomer H64 and dimer TH52A, no longer interacted with the TTACGAAT mismatch (or any other sequence on the probe) even at the highest concentration of 500 nM (Figure 2B).

These experiments demonstrate that linking DNA-binding moieties can dramatically improve the binding specificity of polyamides (see also Herman *et al.*, 1999). They also illustrate the subtlety and apparent unpredictability of the DNA recognition rules, as illustrated by the difference in behavior between TH52A and TH52B; although these molecules are very similar, TH52B (lacking only one pyrrole ring) has a much higher specificity than TH52A.

Behavior of the flexible linker

The binding scheme for TH52B (Figure 1) suggests that the ethylene oxide linker of TH52B spans the central A (bold) of its 10 bp AGGTTAGGTT target. We carried out a number of initial footprinting studies to establish the 'linker rules' at this position. This central A was either deleted or replaced with any nucleotide. Footprint studies indicated that T and A are equally well tolerated at this

position, but that replacement of this central A by a G reduced the affinity by ~ 4 -fold and that replacement by a C abolished binding altogether (not shown). Also poorly tolerated was a deletion of this central A or the insertion of any nucleotide 3' of it. Thus, the ethylene oxide linker is best suited to bridge a single W nucleotide.

To determine the effect of linker length, we synthesized two compounds with a shorter, aliphatic, methylene linker (β -alanine and 5-aminovaleric acid) and one tandem hairpin with a longer linker composed of two amphipathic 8-aminodioxo-octanoic acid units. The latter linker adds 18 interatomic bonds and has an amphipathic character due to the presence of the ethylene oxide units. This tandem hairpin showed binding affinity similar to its parent TH52B (containing nine interatomic bonds) but an increased affinity for its 'mismatch' sequence TTAGGG (increased from ~ 500 to 10 nM) (not shown). The shorter β -alanine linker (adding four interatomic bonds) showed both lowered affinity and specificity. Affinity was lowered by 15- to 20-fold and it bound many mismatch sites on the probe, suggesting that this linker is too short to allow proper binding of both hairpins. The valeric acid-spaced tandem hairpin (adding six interatomic bonds) was similar to its parent TH52B in both affinity and specificity, suggesting that the original linker is longer than necessary but that there is no entropic penalty to be paid up to (at least) nine interatomic bonds.

Targeting vertebrate TTAGGG telomeric repeats

Recognition of the additional G of the vertebrate hexameric telomere repeat could, in principle, be obtained easily by the addition of an extra imidazole in each hairpin of the tandem hairpin TH52B. A related sequence containing three consecutive Gs (AGGGA) was targeted previously by a hairpin containing three consecutive imidazoles [ImImImPy- γ -PyPyPyPy- β -Dp; (Swalley *et al.*, 1996), where γ represents the γ -aminobutyric acid turn monomer and Dp represents dimethylaminopropylamide]. We therefore synthesized a new tandem hairpin polyamide termed TH58 containing an additional imidazole in each hairpin (Figure 1). Surprisingly, subsequent binding analysis by DNase I footprinting showed that TH58 did not protect TTAGGG repeats (nor any other sequence) up to the highest concentration tested (50 nM; not shown).

Hence, we considered an alternative drug design for the recognition of TTAGGG repeats by moving one of the three imidazoles to the opposite side of the hairpin and leaving it 'unpaired'. Few data have been published about unpaired imidazoles (without the opposite pyrrole) and their sequence preference. In our experience, unpaired imidazoles recognize both G-C and C-G base pairs (Janssen *et al.*, 2000a), which is in accordance with the central location of the guanine exocyclic amine group in the minor groove (White *et al.*, 1997 and references therein). The structure of the compound in which this principle is applied in the N-terminal hairpin is schematized in Figure 1 (compound TH57). The binding behavior of TH57 was evaluated by DNase I footprinting. Figure 2D shows that binding of TH57 to TTAGGG repeats (the opposite strand, CCCTAA, is labeled) now occurs at low nanomolar concentrations ($K_{app} \sim 3$ nM). At higher

concentrations, binding to TTAGG repeats is also observed ($K_{app} \sim 10$ nM).

We next synthesized a tandem hairpin polyamide where both hairpins have the same design as the N-terminal hairpin of TH57 (with the third Im moved to the 'bottom'). The schematic structure of this new compound (termed TH59) is shown in Figure 1. The chemical structure of TH59 and a speculative binding model is shown in Figure 3. DNase I footprint analysis of TH59 (Figure 2E) showed that this compound bound with slightly increased affinity to its target sequence TTAGGG, but cannot discriminate between the two different telomeric sequences. The equilibrium dissociation constant (K_d) for TH59 was determined by quantitative DNase I footprinting (Wade *et al.*, 1993) at 0.51 nM. We also evaluated the binding behavior of the TH59 'monomer' and observed that this compound (termed H60, see Figure 1) bound with similar affinity but lower specificity (see Figure 2F).

Staining telomeres with fluorescently tagged polyamides

Using fluorescently tagged polyamides and epifluorescence microscopy, we previously highlighted the position of different DNA satellites in nuclei and chromosomes. This staining method provides specificity information for polyamides on chromatin rather than on naked DNA. Molecules TH52B and TH59 bind two telomeric repeats. Since the length of eukaryotic telomeres varies between a few and tens of kilobase pairs, one might ultimately expect a hundred to a few thousand molecules bound per telomere. Is it possible to observe this signal above the background?

Insect telomeres. To obtain fluorescently tagged polyamides, a primary amine was introduced at the C-terminus, which was subsequently acylated using a commercially available *N*-hydroxysuccinimide active ester of Texas Red (Janssen *et al.*, 2000b). Insect chromosomal material was then isolated from Sf9 cells, a cell line derived from the armyworm, *Spodoptera frugiperda* (Vaughn *et al.*, 1977), which is expected to contain insect-type telomeric repeats (Okazaki *et al.*, 1993). Classical metaphase chromosomal spreads were prepared from Sf9 cells and then double stained with 4',6-diamidino-2-phenylindole (DAPI) and TH52B-T (T for Texas Red). Figure 4A shows in blue (DAPI) the metaphase chromosomes and in red striking foci, which represent the subchromosomal signals of TH52B-T. Karyotypes of Sf9 cells are very complex and poorly characterized, consisting of innumerable mini-chromosomes. Generally, two foci are observed at each chromosomal end, suggesting that TH52B-T highlights telomeres as expected. Included in Figure 4A is a black and white inset showing the TH52B-T telomeric signal separately. Note that although generally low 'background' signal is observed along the chromosomal body, one can also observe some subtelomeric signals

Figure 4B shows a representative image of an Sf9 nucleus stained with DAPI and TH52B-T, which again yields sharp red foci. Interestingly, the DAPI signal of stained Sf9 nuclei shows an unusual 'grape-like' structure rather than displaying the generally homogeneous appearance of eukaryotic nuclei. Closer examination of these

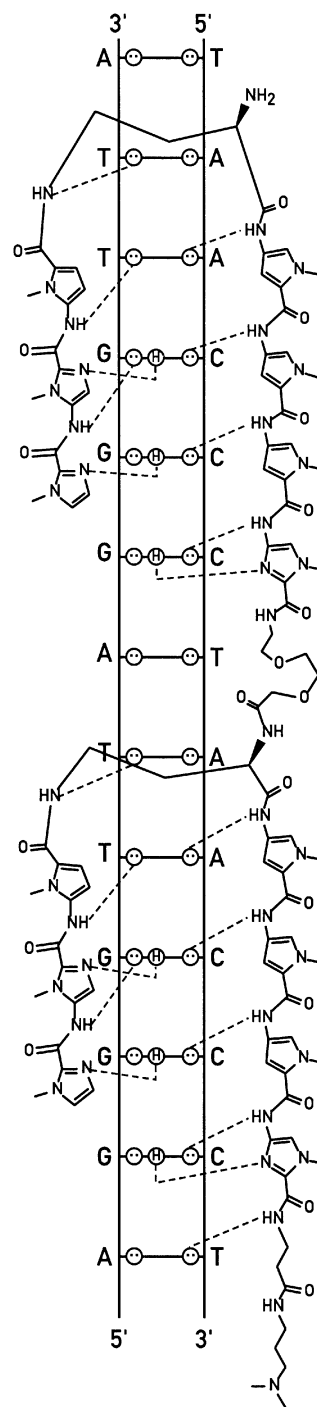


Fig. 3. Putative binding model of TH59. Proposed binding model for the complex of TH59 with 5'-AGGGTTAGGGTT-3'. Circles with two dots represent lone pairs of electrons on N3 of purines and O2 of pyrimidines at the edges of the bases. Circles containing an H represent the N2 hydrogen of guanine. Putative bifurcated hydrogen bonds to the amide NHs are illustrated by dashed lines.

images reveals that the red TH52B-T foci are often abutting blue grape-like domains, perhaps representing interphase chromosomal territories. Importantly, in line with our footprint data, no red foci were observed when HeLa nuclei were stained with TH52B-T (Figure 4C).

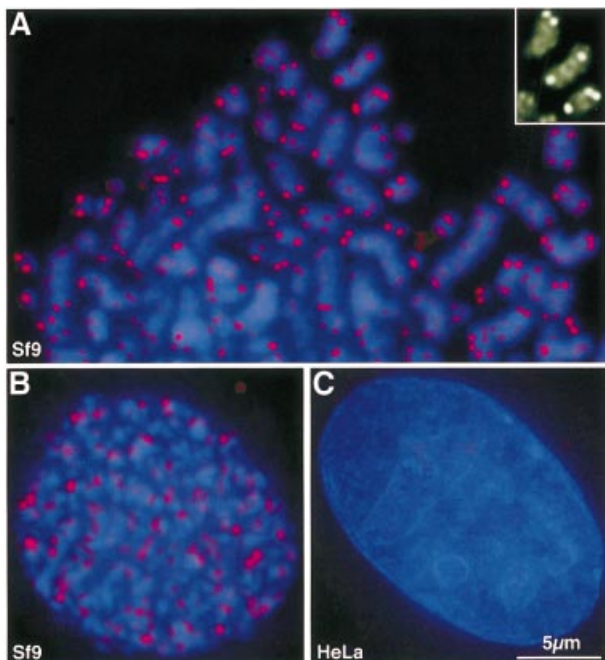


Fig. 4. Staining of insect-type telomere repeats (TTAGG) with TH52B-T. Chromosomes or nuclei prepared from Sf9 and HeLa cells were co-stained with TH52B-T (red) and DAPI (blue). Note that TH52B-T sharply highlights red foci in Sf9 (B) but not HeLa nuclei (C). The two images (B and C) were obtained under identical conditions and are shown on an identical intensity scale. TH52B-T also stains the ends of chromosomes, as observed in metaphase spreads derived from Sf9 cells (A). A number of non-telomeric signals can also be noted (black and white inset). Scale bars represent 5 μ m.

As mentioned above, the footprinting data indicated that TH59, in contrast to TH52B, bound both the penta- and hexameric telomeric repeats. Chromosomal staining studies confirmed this notion. We observed that TH59-T stains telomeres of Sf9 chromosomes, although its signal was less sharp than that obtained with TH52B-T owing to a higher general background (not shown).

Vertebrate telomeres. We assessed the specificity of the telomere polyamide TH59-T by staining chromosomal material derived from a variety of vertebrate cells. These studies collectively demonstrated that TH59-T (not TH52B-T) specifically stained telomeres of all these cell lines. A series of micrographs is shown in Figure 5. All images were obtained by double staining chromosomes or nuclei with DAPI and TH59-T. Figure 5A impressively shows that TH59-T sharply highlights telomere foci in HeLa cell nuclei. Telomeres are also marked robustly by TH59-T in spreads of HeLa metaphase chromosomes (Figure 5D), on isolated Indian Muntjac (Figure 5B) and on *Xenopus laevis* (Figure 5C) chromosomes.

The staining results with TH52 and TH59 are impressive since the micrographs revealed high signal-to-noise ratios. This is particularly evident by inspection of the black and white inset in Figure 5D; note that TH59-T yields little general background signal along the chromosomal body. These staining data semi-quantitatively extend the footprint studies from naked DNA to chromatin.

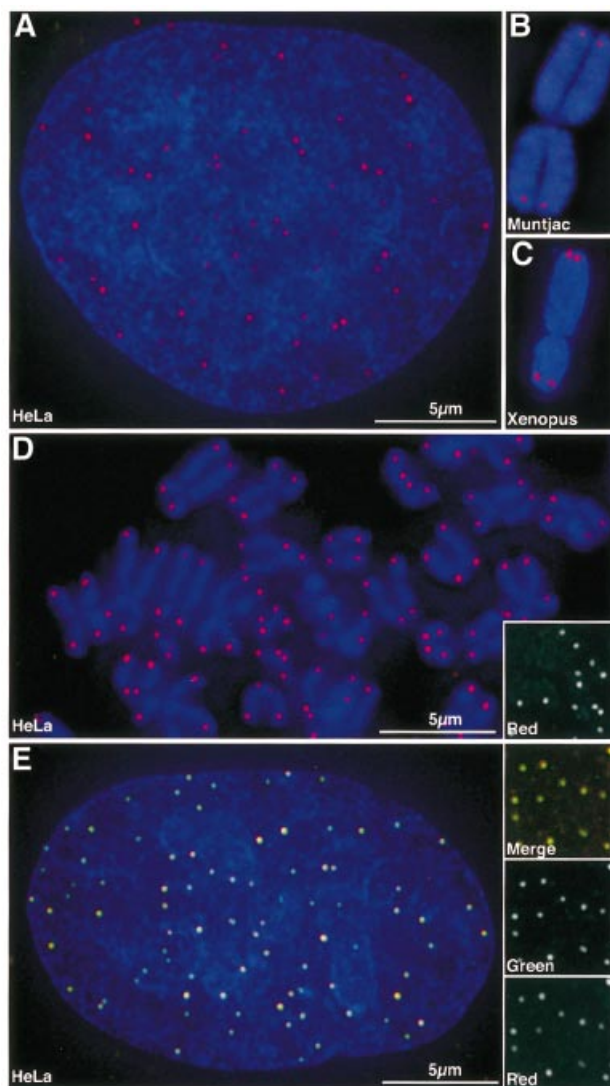


Fig. 5. Staining of vertebrate telomeric repeats (TTAGGG) with TH59-T and co-localization with TRF1. Chromosomal material prepared from vertebrate cell lines was stained with polyamide TH59-T (red) and DAPI (blue). Note that TH59-T sharply highlights red foci in HeLa cell nuclei (A) and metaphase chromosomes derived from Indian Muntjac (B), *X.laevis* (C) or HeLa cells (D). (E) TH59-T foci co-localize with green telomere spots revealed by indirect immunofluorescence with TRF1-specific antibodies. The black and white inset shows the green and red signals separately. Scale bars represent 5 μ m.

The signals of telomere-specific polyamides and those of a telomere-binding protein, TRF1, co-localize

The footprint data and the conspicuous staining results obtained with polyamides TH52 and TH59 (or derivatives thereof) strongly suggest that these compounds are localized at telomeres. We proved this notion by immunofluorescence. Human telomeres contain two related TTAGGG repeat-binding factors, TRF1 and TRF2 (Broccoli *et al.*, 1997). [Both proteins have a C-terminal Myb-like helix–turn–helix domain and a central domain involved in the formation of homodimers (O'Reilly *et al.*, 1999).] TRF1 and TRF2 are located predominantly at chromosomal ends where they contribute to the maintenance of telomere structure. TRF1 and TRF2

should therefore co-localize with TH59-T. Figure 5E shows a HeLa cell nucleus stained with TH59-T (red) together with the immunofluorescence signal of TRF1 (green). Examination of this micrograph shows that TH59-T and anti-TRF1 signals co-localize perfectly since the resulting overlapping spots appear yellow. This conclusion is also apparent from inspection of the insets of Figure 5E displaying the TH59-T and anti-TRF1 signals separately, in black and white. Similar observations were also obtained by staining chromosomal spreads. In addition, the immunofluorescence signals of TRF2 also co-localized with TH59-T (data not shown). These observations establish that TH59-T marks the position of telomeres in nuclei and chromosomes.

Rapid estimation of relative telomere length

In telomere research, it is often of interest to estimate and follow telomere length quantitatively. Three protocols are currently used for this purpose. One of these depends on digestion of the genomic DNA using enzymes with restriction sites in subtelomeric DNA, and Southern blotting (van Steensel and de Lange, 1997). This rather time-consuming procedure provides an estimate of telomere lengths of all cells of a sample. Two related alternative procedures are based on *in situ* hybridization using either chromosomal spreads or permeabilized cells followed by epifluorescence microscopy (Zijlmans *et al.*, 1997) or flow cytometry (Hultdin *et al.*, 1998). The latter two protocols generally use fluorescently labeled peptide nucleic acid (PNA) complementary to the telomere repeats as probes. We propose that this laborious, cumbersome hybridization step can be replaced by a simple staining protocol using fluorescent telomere-specific polyamides.

To examine this issue, metaphase chromosome spreads were prepared from human lymphocytes and two closely related HeLa subclones that differ in telomere lengths. HeLa1.2.11 (termed HeLa-L) has long (L) telomeres of ~15–40 kb (Smogorzewska *et al.*, 2000), whereas HeLaII (HeLa-S) has shorter (S) telomeres of 3–6.5 kb (Smogorzewska *et al.*, 2000). Slides prepared with these cell lines were double stained with DAPI and TH59-T under identical conditions, and several optical sections of chromosomal spreads were recorded with an image restoration, wide field fluorescence microscope (DeltaVision). In order to analyze the maximum intensity of each telomere, the various optical sections were combined by a maximum brightness projection. Visual inspection of these projected images indicated that HeLa-L telomeres were labeled more intensively by TH59-T than the telomeres of lymphocytes or HeLa-S cells (Figure 6A and B; data not shown).

We carried out particle analysis to determine quantitatively the integrated signal intensity of telomeres encompassing entire chromosomal spreads. For that purpose, the telomere spots were contoured appropriately. Examples of chromosomes with contoured telomeres are shown in Figure 6. The total integrated telomere signals within the contoured telomeres were determined subsequently for all chromosomes of entire spreads. The signal intensity per telomere varies over a wide range, yielding an intensity distribution for metaphase spreads derived from lymphocytes and HeLa-L and HeLa-S cells as shown in Figure 7.

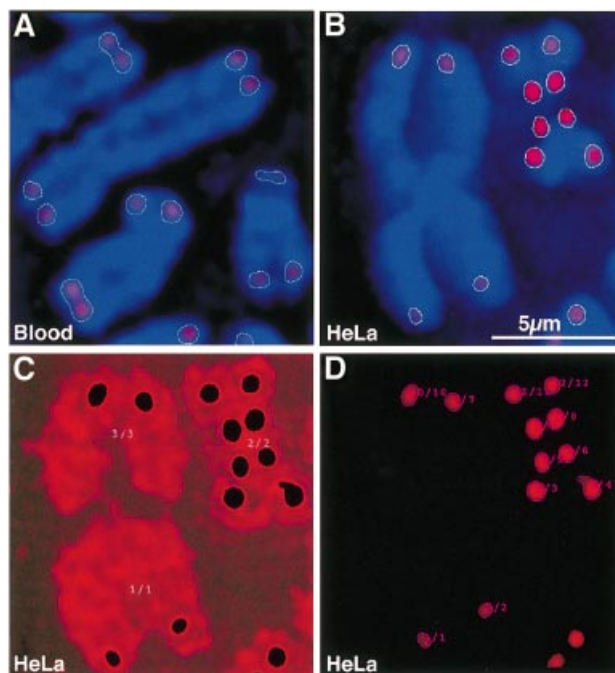


Fig. 6. Quantitative aspect of vertebrate telomere signals. Metaphase chromosome spreads were prepared from human lymphocytes and HeLa-L cells, a HeLa subclone with long telomeres. Subregions of metaphase spreads are shown derived from human lymphocytes (A) and HeLa-L cells (B). The two images (A and B) were obtained under identical conditions and are shown on the same intensity scale. For quantification, the telomere foci were contoured appropriately (indicated) and the total integrated signal intensities were then determined for entire spreads. The total telomere signal intensity distribution obtained for different chromosomal spreads is shown in Figure 7. The fractional amount of polyamide bound at telomeres relative to that encompassing the chromosomal body was determined. An example of this is shown for HeLa-L chromosomes (B, C and D). The telomere spots contoured in (B) were extracted to yield (D). The image lacking telomere spots was then contoured again using a lower threshold level so as to encompass the chromosomal bodies (C). (C) and (D) are shown on a strongly enhanced intensity scale to visualize the staining of chromosomal bodies. The total integrated signal derived from TH59-T encompassing either telomeres (D) or chromosomal bodies (C) was then obtained and corrected for general background level. Scale bar represents 5 μm .

Examination of these data demonstrates that the distribution of the total telomere signal is strongly skewed toward higher intensity for telomeres of HeLa-L chromosomes as compared with those obtained from HeLa-S cells and lymphocytes. Note that the telomere intensity distributions are similar for the latter two chromosomes. The weight averages of these distributions were also determined (inset Figure 7). The values suggest that HeLa-L telomeres are ~3-fold longer than those derived from lymphocytes or HeLa-S cells. These observations demonstrate that specific fluorescent polyamides can provide the basis for a rapid procedure to estimate relative telomere lengths.

In vivo incorporation of TH52-T into Sf9 cells

Can the above *in vitro* studies be extended to live cells? The relatively high molecular weight of TH52B and TH59 (1795 and 2204 Da, respectively) could represent an obstacle in terms of permeability. Hence, it was of

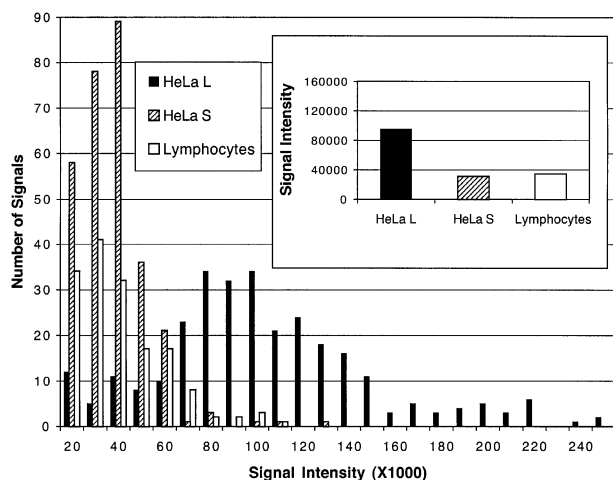


Fig. 7. Distribution of telomere signals of different chromosomal spreads. Metaphase chromosome spreads were prepared from human lymphocytes and two closely related HeLa subclones (HeLa-L and HeLa-S) that differ with regard to telomere length. The bar graph shows the intensity distribution of the telomere signals obtained for the different chromosomal spreads. Note that the distribution of the telomere signals is strongly skewed toward higher intensity for telomeres of HeLa-L chromosomes as compared with those obtained from HeLa-S and lymphocytes. The weight averages of these distributions were also determined and are shown in the inset.

importance to determine whether cells are permeable to these molecules and whether these molecules are capable of binding telomeres *in vivo*. These experiments were carried out with the insect telomere-specific polyamide TH52B-T (mol. wt 2520 Da).

For this purpose, Sf9 cells were grown in the presence of different concentrations of TH52B-T for different lengths of time. Subsequently, cells were fixed with cold methanol. This demonstrated that TH52B-T is incorporated into cells rapidly and that specific telomere staining is obtained.

We first determined the polyamide concentration in the media required to obtain telomere foci. In Figure 8A, the telomere signal strength is plotted as a function of polyamide concentration added to the media for a period of 16 h. Telomere foci can be observed at the lowest TH52B-T concentration used (100 nM), but the signal intensity increases with concentration and approaches a plateau around a concentration of 2 μ M.

We also carried out a time course experiment by incubating cells with a fixed concentration (2 μ M) of TH52B-T and Hoechst 33258. Hoechst 33258 is known to be cell permeable and to provide a general staining of chromosomal material. Slides were prepared at timed intervals and the intensity of the Hoechst 33258 and telomere-specific (TH52B-T) signals was determined. Figure 8B shows that telomeres are stained rapidly with TH52B-T, approaching a maximal signal intensity after 3–4 h of incubation. Hoechst 33258 also enters cells rapidly, but the signal intensity appears to reach a plateau more slowly than for TH52B-T. Quite remarkably, this experiment demonstrated that telomere-specific staining is achieved with an incubation period as short as 15 min. In conclusion, TH52B-T enters cells rapidly, resulting in the telomere-specific staining. The addition of the above

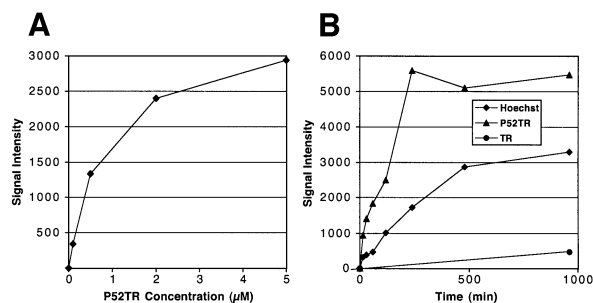


Fig. 8. TH52B-T rapidly highlights telomere foci of Sf9 cells *in vivo*. TH52B-T and Hoechst 33258 were added to the media of growing Sf9 cells at different concentrations and periods of incubation. Cells were then fixed by cold methanol and examined by fluorescence microscopy to determine the strength of the TH52B-T telomere signals and that of Hoechst 33258 (general DNA stain). (A) The average telomere signal intensity obtained after 16 h incubation in media at the indicated concentrations of TH52B-T. (B) A time course of incorporation. The averages of the telomere signal intensity of TH52B-T (telomere foci, triangles), Hoechst 33258 (general stain, diamonds) and Texas Red alone (background, circles) are shown. Note that specific foci were already observed after a 15 min exposure of Sf9 cells to TH52B-T.

compounds appeared not to affect the growth of Sf9 cells during the course of the experiment (data not shown).

Discussion

Telomere-specific polyamides containing two hairpin DNA-binding moieties, flexibly linked, were synthesized. These dimeric compounds were found to be cell permeable and to bind insect or vertebrate telomeric repeats with impressive specificity. Particularly convincing are the results obtained by staining of chromosomal material with fluorescently tagged derivatives. Visual inspection of such micrographs suggests that telomere signals obtained with polyamides are of a specificity comparable to those observed by immunodetection of the telomere-binding proteins TRF1 and TRF2.

TH52B exclusively binds insect telomere repeats

TH52B is a heterodimer that binds two insect telomere repeats (TTAGG) at subnanomolar concentrations ($K_d = 0.12$ nM). This compound has essentially no affinity for vertebrate (TTAGGG) telomeric repeats up to 500 nM (Figure 2B). Most importantly, even at this high concentration, no other mismatch interaction sites were observed for TH52B. This result contrasts with that of its monomer H64. This compound prefers insect repeats ($K_{app} < 1$ nM), but it also binds to vertebrate repeats ($K_{app} < 50$ nM) and several mismatch sites (Figure 2A). The other monomer of TH52B (H65) binds insect repeats very poorly ($K_{app} > 200$ nM). Nevertheless, tethering H64 (a strong binder but with rather low specificity) to H65 resulted in a dimer (TH52B) that is considerably more specific than its best half (H64). The impressive specificity of TH52B is proposed to arise from the cooperative interaction of the flexibly linked hairpin moieties with two insect repeats (bidentate binding). In contrast, the low affinity of TH52B for mismatch sites must arise from energetic penalties that apply if only one DNA-binding moiety can be accommodated (monodentate binding) or if both moieties are positioned unfavorably.

The putative interaction scheme of TH52B is depicted in Figure 1. It follows closely the most favorable 'interaction rules' that can be deduced from polyamide studies. Published reports implicitly revealed that, in hairpins, imidazoles yield the highest G nucleotide selectivity if positioned at the N-terminal position. The γ -turn (2,4-diaminobutyric acid) of TH52B is, according to binding studies, optimally positioned on a W (Swalley *et al.*, 1999). TH52 and TH59 adhere to this rule. Previous studies further established that hairpin polyamides show a strong preference for GT (as in telomere repeats) over GA (Kielkopf *et al.*, 2000). The C-terminal Dp moiety of TH52B is also positioned on a W. This is known to be preferred over G or C nucleotides (Swalley *et al.*, 1999). Moreover, the linker rules established here demonstrated that the ethylene oxide linker is best suited to bridge a W. Hence, the high affinity of TH52B can be explained by a binding scheme adhering to 'established' interaction rules and its high specificity by a bidentate binding mode.

The binding scheme for TH52A (Figure 1) suggests that this compound also obeys the above interaction rules. However, despite this conformity, TH52A (not TH52B) binds many mismatches and tends to coat the entire DNA probe (Figure 2A). The structural rationale for this 'misbehavior' is unclear to us. Clearly, the interaction rules need further refinements.

TH59 binds vertebrate and insect telomere repeats

Targeting of the extra G nucleotide of vertebrate telomere repeats (TTAGGG) was achieved by positioning in each hairpin a single unpaired imidazole in the bottom strand (Figures 1 and 3). The resulting compound, TH59, binds vertebrate and insect telomeres with high affinity ($K_d = 0.51$ nM). The proposed binding scheme depicted for TH59 follows the general rules discussed above and the additional notion that unpaired imidazoles are accommodated preferably on G or C nucleotides. To explain the interaction of TH59 with insect repeats, we propose a somewhat unusual but noteworthy binding mode. Binding of TH59 to insect TTAGG repeats implies that the unpaired imidazoles are accommodated by As and that the ethylene oxide linker is somehow looped out, opposing no nucleotide. This binding mode is in line with our previous observations that unpaired, internal (not N-terminal) imidazoles are quite well tolerated by Ws, particularly if preceded by an N-terminal pyrrole but not if preceded (N-terminal) by a β -alanine (Janssen *et al.*, 2000b).

Telomere staining

Staining of insect and vertebrate metaphase chromosomes or nuclei with fluorescent derivatives of TH52B and TH59 revealed sharp foci. These foci were found at telomeric locations in stained metaphase chromosome (at both ends of each sister chromatid). In contrast, these foci appeared unorganized throughout the volume of nuclei, with no apparent peripheral tethering. The TH59-T foci were identified unequivocally as telomeres since their signals co-localized with those obtained by immunofluorescence detection of the TRF1 or TRF2 telomere-binding proteins. Visually, the telomeric signals are impressively 'sharp' owing to a low general background of the non-telomeric chromosomal body. Indeed, determination of the average

pixel intensity over the chromosomal body and at telomeres yields a signal-to-noise ratio of 15 for TH59-T and a ratio of 30 for the immunodetection signal of TRF1. Hence, these measurements confirm the visual impression that the specificity of detection by immunofluorescence with a good antibody and by polyamide staining can be of similar quality.

Our results demonstrated that telomere-specific polyamides could be used conveniently to determine relative telomere length. For this purpose, chromosomal spreads were stained with DAPI and TH59-T, recorded by a fluorescence microscope and the total telomere signal determined with the help of an appropriate software package. Our analysis demonstrates that the mean telomere length of chromosomes derived from HeLa-L cells is ~ 3 times longer than those obtained from human lymphocytes or HeLa-S cells (Figure 7). This is in line with previous telomere length studies that established a mean length of 23 kb for the telomeres of HeLa-L and 6 kb for those of HeLa-S chromosomes (Griffith *et al.*, 1999; Smogorzewska *et al.*, 2000). In addition, it should be noted that our staining procedure can be accomplished within 3 h and is much more rapid when compared with the hybridization methods (Hultdin *et al.*, 1998). The staining procedure described here to determine telomere length could be used conveniently for a variety of experiments in research and possibly diagnostics, such as the screening of tissues for cells with increased telomere length. The specific stains should be equally suitable for telomere length determination by flow cytometry.

Specificity considerations of telomere targeting

Polyamides are potentially interesting medicinal agents. Toward this long-term end, it was of importance to estimate the relative enrichment of the targeted polyamides on telomere versus bulk DNA. The data collected above to determine telomere length serve this purpose (Figure 6). The TH59-T telomere signals of stained HeLa-L chromosomes were contoured using an appropriate threshold level as discussed above (Figure 6B). The telomere spots were then extracted from the source image of Figure 6B to yield Figure 6D. The source image lacking telomere spots was then contoured again using a lower threshold level so as to encompass the chromosomal bodies, whose signals were strongly enhanced for printing (Figure 6C). The total integrated signal derived from TH59-T encompassing either telomeres (Figure 6D) or chromosomal bodies (Figure 6C) was then obtained and corrected for general background level.

This quantitative analysis was carried out over entire chromosomal spreads and established that the fractional amount of TH59-T associated with telomeres is remarkably high. The values obtained for chromosomes derived from HeLa-L cells, HeLa-S cells or lymphocytes were ~ 33 , 14 and 18%, respectively. Thus, in line with the increased telomere length of HeLa-L chromosomes, we obtained a higher fractional value of association (33%). These values represent averages over chromosomal spreads that depend on chromosome size. Hence, the fractional value of telomere-bound TH59-T amounts to $\sim 15\%$ for the large chromosome shown on the left in Figure 6B and $\sim 78\%$ for the two small chromosomes on the right of this panel.

References

- Baird,E.E. and Dervan,P.B. (1996) Solid phase synthesis of polyamides containing imidazole and pyrrole amino acids. *J. Am. Chem. Soc.*, **118**, 6141–6146.
- Blackburn,E.H. (2000) Telomere states and cell fates. *Nature*, **408**, 53–56.
- Broccoli,D., Smogorzewska,A., Chong,L. and de Lange,T. (1997) Human telomeres contain two distinct Myb-related proteins, TRF1 and TRF2. *Nature Genet.*, **17**, 231–235.
- Geierstanger,B.H., Mrksich,M., Dervan,P.B. and Wemmer,D.E. (1994) Design of a G-C-specific DNA minor groove-binding peptide. *Science*, **266**, 646–650.
- Girard,F., Bello,B., Laemmli,U.K. and Gehring,W.J. (1998) *In vivo* analysis of scaffold-associated regions in *Drosophila*: a synthetic high-affinity SAR binding protein suppresses position effect variegation. *EMBO J.*, **17**, 2079–2085.
- Griffith,J.D., Comeau,L., Rosenfield,S., Stansel,R.M., Bianchi,A., Moss, H. and de Lange,T. (1999) Mammalian telomeres end in a large duplex loop. *Cell*, **97**, 503–514.
- Hart,C.M. and Laemmli,U.K. (1998) Facilitation of chromatin dynamics by SARs. *Curr. Opin. Genet. Dev.*, **8**, 519–525.
- Hattori,M. *et al.* (2000) The DNA sequence of human chromosome 21. The chromosome 21 mapping and sequencing consortium. *Nature*, **405**, 311–319.
- Herman,D.M., Baird,E.E. and Dervan,P.B. (1999) Tandem hairpin motif for recognition in the minor groove of DNA by pyrrole–imidazole polyamides. *Chem. Eur. J.*, **5**, 975–983.
- Hultdin,M., Gronlund,E., Norrback,K., Eriksson-Lindstrom,E., Just,T. and Roos,G. (1998) Telomere analysis by fluorescence *in situ* hybridization and flow cytometry. *Nucleic Acids Res.*, **26**, 3651–3656.
- Janssen,S., Cuvier,O., Muller,M. and Laemmli,U.K. (2000a) Specific gain- and loss-of-function phenotypes induced by satellite-specific DNA-binding drugs fed to *Drosophila melanogaster*. *Mol. Cell*, **6**, 1013–1024.
- Janssen,S., Durussel,T. and Laemmli,U.K. (2000b) Chromatin opening of DNA satellites by targeted sequence-specific drugs. *Mol. Cell*, **6**, 999–1011.
- Kielkopf,C.L., Bremer,R.E., White,S., Szewczyk,J.W., Turner,J.M., Baird,E.E., Dervan,P.B. and Rees,D.C. (2000) Structural effects of DNA sequence on T-A recognition by hydroxypyrrole/pyrrole pairs in the minor groove. *J. Mol. Biol.*, **295**, 557–567.
- Lee,J.J., Warburton,D. and Robertson,E.J. (1990) Cytogenetic methods for the mouse: preparation of chromosomes, karyotyping and *in situ* hybridization. *Anal. Biochem.*, **189**, 1–17.
- McEachern,M.J., Krauskopf,A. and Blackburn,E.H. (2000) Telomeres and their control. *Annu. Rev. Genet.*, **34**, 331–358.
- Meyne,J., Ratliff,R.L. and Moyzis,R.K. (1989) Conservation of the human telomere sequence (TTAGG)_n among vertebrates. *Proc. Natl Acad. Sci. USA*, **86**, 7049–7053.
- Okazaki,S., Tsuchida,K., Maekawa,H., Ishikawa,H. and Fujiwara,H. (1993) Identification of a pentanucleotide telomeric sequence, (TTAGG)_n, in the silkworm *Bombyx mori* and in other insects. *Mol. Cell Biol.*, **13**, 1424–1432.
- O'Reilly,M., Teichmann,S.A. and Rhodes,D. (1999) Telomerases. *Curr. Opin. Struct. Biol.*, **9**, 56–65.
- Oulton,R. and Harrington,L. (2000) Telomeres, telomerase and cancer: life on the edge of genomic stability. *Curr. Opin. Oncol.*, **12**, 74–81.
- Prescott,J.C. and Blackburn,E.H. (1999) Telomerase: Dr Jekyll or Mr Hyde? *Curr. Opin. Genet. Dev.*, **9**, 368–373.
- Sahara,K., Marec,F. and Traut,W. (1999) TTAGG telomeric repeats in chromosomes of some insects and other arthropods. *Chromosome Res.*, **7**, 449–460.
- Saitoh,Y. and Laemmli,U.K. (1994) Metaphase chromosome structure: bands arise from a differential folding path of the highly AT-rich scaffold. *Cell*, **76**, 609–622.
- Smogorzewska,A., van Steensel,B., Bianchi,A., Oelmann,S., Schaefer, M.R., Schnapp,G. and de Lange,T. (2000) Control of human telomere length by TRF1 and TRF2. *Mol. Cell Biol.*, **20**, 1659–1668.
- Strick,R. and Laemmli,U.K. (1995) SARs are *cis* DNA elements of chromosome dynamics: synthesis of a SAR repressor protein. *Cell*, **83**, 1137–1148.
- Swalley,S.E., Baird,E.E. and Dervan,P.B. (1996) Recognition of a 5'-(A,T)GGG(A,T)2-3' sequence in the minor groove of DNA by an eight-ring hairpin polyamide. *J. Am. Chem. Soc.*, **118**, 8198–8206.
- Swalley,S.E., Baird,E.E. and Dervan,P.B. (1999) Effects of γ -turn and β -tail amino acids on sequence specific recognition of DNA by hairpin polyamides. *J. Am. Chem. Soc.*, **121**, 1113–1120.
- van Steensel,B. and de Lange,T. (1997) Control of telomere length by the human telomeric protein TRF1. *Nature*, **385**, 740–743.
- Vaughn,J.L., Goodwin,R.H., Tompkins,G.J. and McCawley,P. (1977) The establishment of two cell lines from the insect *Spodoptera frugiperda* (Lepidoptera; Noctuidae). *In Vitro*, **13**, 213–217.
- Wade,W.S., Mrksich,M. and Dervan,P.B. (1993) Binding affinities of synthetic peptides, pyridine-2-carboxamidotropin and 1-methylimidazole-2-carboxamidotropin, that form 2:1 complexes in the minor groove of double-helical DNA. *Biochemistry*, **32**, 11385–11389.
- Wemmer,D.E. (2000) Designed sequence-specific minor groove ligands. *Annu. Rev. Biophys. Biomol. Struct.*, **29**, 439–461.
- White,S., Baird,E.E. and Dervan,P.B. (1997) On the pairing rules for recognition in the minor groove of DNA by pyrrole–imidazole polyamides. *Chem. Biol.*, **4**, 569–578.
- Zijlmans,J.M., Martens,U.M., Poon,S.S., Raap,A.K., Tanke,H.J., Ward, R.K. and Lansdorp,P.M. (1997) Telomeres in the mouse have large inter-chromosomal variations in the number of T2AG3 repeats. *Proc. Natl Acad. Sci. USA*, **94**, 7423–7428.

Received March 2, 2001; revised April 11, 2001;
accepted April 23, 2001