60 Honig, L.S. (1981) *Nature* 291, 72–73
61 Parr, B.A. and McMahon, A.P. (1995) *Nature* 374, 350–353
62 Francis, P.H., Richardson, M.K., Brickell, P.M. and Tickle, C. (1994) *Development* 120, 209–218
63 Laufer, E. *et al.* (1994) *Cell* 79, 993–1003
64 Niswander, L., Jeffrey, S., Martin, G.L. and Tickle, C. (1994) *Nature* 371, 609–612
65 Roberts, D.J. *et al.* (1995) *Development* 121, 3163–3174
66 Bitgood, M.J. and McMahon, A.P. (1995) *Dev. Biol.* 172, 126–138
67 Vortkamp, A. *et al.* (1996) *Science* 273, 613–622
68 Bitgood, M., Shen, L. and McMahon, A.P. (1996) *Curr. Biol.* 6, 298–304
69 Perrimon, N. (1996) *Cell* 86, 513–516
70 Nakano, Y. *et al.* (1989) *Nature* 341, 508–513
71 Hooper, J. and Scott, M.P. (1989) *Cell* 59, 751–765
72 Ingham, P.W., Taylor, A.M. and Nakano, Y. (1991) *Nature* 353, 184–187
73 Johnson, J.L., Grenier, J.K. and Scott, M.P. (1995) *Development* 121, 4161–4170
74 Bejsovec, A. and Wieschaus, E. (1993) *Development* 119, 501–517
75 Marigo, V. *et al.* (1996) *Development* 122, 1225–1233
76 Concordet, J.P. *et al.* (1996) *Development* 122, 2835–2846
77 van den Heuvel, M. and Ingham, P.W. (1996) *Nature* 382, 547–551
78 Alcedo, J. *et al.* (1996) *Cell* 86, 221–232
79 Hooper, J.E. (1994) *Nature* 372, 461–464
80 Orenic, T.V., Slusarski, D.C., Kroll, K.L. and Holmgren, R. (1990) *Genes Dev.* 4, 1053–1067
81 Préat, T. *et al.* (1990) *Nature* 347, 87–89
82 Thérond, P. *et al.* (1993) *Mech. Dev.* 44, 65–80
83 Kalderon, D. (1995) *Curr. Biol.* 5, 580–582
84 Epstein, D.J., Martí, E., Scott, M.P. and McMahon, A.P. (1996) *Development* 122, 2885–2894
85 Thérond, P.P., Knight, J.D., Kornberg, T.B. and Bishop, J.M. (1996) *Proc. Natl. Acad. Sci. U. S. A.* 93, 4224–4228
86 Alexandre, C., Jacinto, A. and Ingham, P.W. (1996) *Genes Dev.* 10, 2003–2013
87 Domínguez, M., Brunner, M., Hafen, E. and Basler, K. (1996) *Science* 272, 1621–1625
88 Delmas, V. *et al.* (1994) *Rev. Physiol. Biochem. Pharmacol.* 124, 1–28
89 Vortkamp, A., Gessler, M. and Grzeschik, K-H. (1991) *Nature* 352, 539–540
90 Hahn, H. *et al.* (1996) *Cell* 85, 841–851
91 Johnson, J.L. *et al.* (1996) *Science* 272, 1668–1671
92 Kinzler, K.W. *et al.* (1988) *Nature* 332, 371–374
93 Marigo, V. *et al.* (1995) *Genomics* 28, 44–51

M. Hammerschmidt *is in the Spemann-Laboratories, MPI für Immunobiologie, Stuebeweg 51, 71908 Frieburg, Germany;*
A. Brook *is in the Laboratory of Molecular Oncology, Cancer Center, Massachusetts General Hospital, 13 Street, Building 149, Charlestown, Massachusetts 02129, USA;*
A.P. McMahon *is in the Department of Molecular and Cellular Biology, The Biolabs, Harvard University, 16 Divinity Avenue, Cambridge, Massachusetts 02138, USA.*

TRF1, a mammalian telomeric protein

SUSAN SMITH (smithsu@rockvax.rockefeller.edu)

TITIA DE LANGE (delange@rockvax.rockefeller.edu)

Telomerase adds TTAGGG repeats onto mammalian chromosome ends, replenishing the terminal sequence loss incurred during DNA replication. This maintenance of telomeric DNA preserves binding sites for telomeric proteins, which form a protective nucleoprotein complex at chromosome ends. The recent isolation of TRF1, the mammalian telomeric-repeat binding factor, should now allow the structure and function of the telomeric complex to be examined in detail.

Telomerase, the enzyme that elongates telomeres, has received much attention due to its remarkable enzymology[1] and its involvement in cancer[2,3]. Less thought has been given to its products: the long arrays of telomeric repeats that cap the ends of chromosomes. Telomeres were originally identified functionally, based on cytological and genetic evidence indicating that broken chromosomes are unstable. The current view is that telomeres are protective caps that stabilize chromosome ends and shield them from the cellular surveillance systems that monitor DNA damage[4]. Although the mechanism is unknown, it is clear that cells can discriminate between natural chromosome ends (telomeres) and broken DNA. Telomeres (unlike double-strand breaks) do not activate DNA-damage checkpoint proteins, such as ATM and p53, nor do they serve as substrates for repair enzymes that degrade or ligate DNA ends. Thus, telomeres serve at least two essential functions: they provide the substrate for telomerase-mediated DNA replication and they mask the chromosome end from DNA-damage checkpoints. The highly conserved nature of telomeric DNA predicts that aspects of telomere function will be conserved in widely diverged eukaryotes. Recently, the first vertebrate telomere-repeat binding factor, TRF1, was isolated from human cells and its cDNA cloned. This review summarizes the current information on the structure and function of mammalian TRF1 and draws a comparison with its potential functional analog, the yeast telomeric protein Rap1p.

A search for mammalian telomeric proteins

The only *cis*-acting requirement for telomere function in mammalian cells is an array of TTAGGG repeats, orientated such that the G-rich strand runs out to the 3′ end of the chromosome. Long arrays (more than 10 kb) of TTAGGG repeats are found at all human chromosome ends, and this same telomeric repeat is shared by all vertebrates[5] and several unicellular organisms[4].

 0168-9525/97/$17.00
PII: S0168-9525(96)10052-4

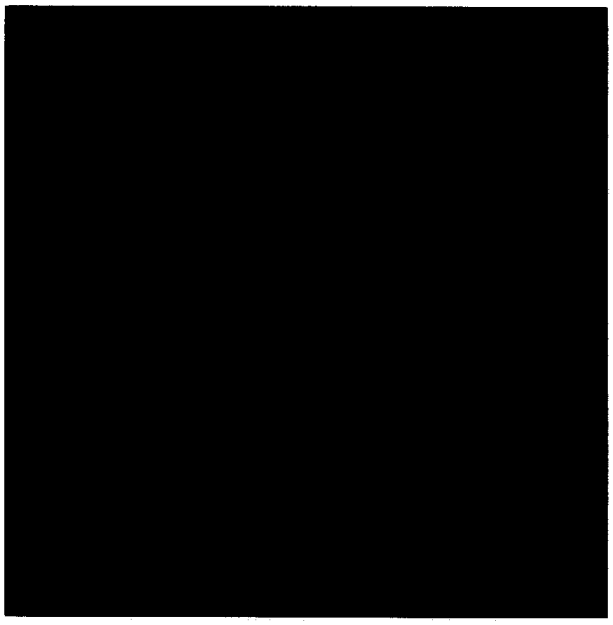

FIGURE 1. Detection of endogenous TRF1 protein at the ends of HeLa metaphase chromosomes. TRF1 was detected with a polyclonal rabbit serum raised against TRF1 and affinity-purified using the acidic N-terminal domain of TRF1.

Several lines of evidence support the view that mammalian chromosome ends are packaged in telomere-specific chromatin (reviewed in Ref. 6). (1) The whole telomeric tract (unlike the bulk of chromosomal DNA) is bound to the nuclear matrix, a proteinaceous subnuclear fraction[7,8]. (2) An altered nucleosomal organization can be observed at chromosome ends[9,10] and telomeric histone H4 is hypoacetylated[11]. (3) Telomere formation displays a stringent requirement for the TTAGGG repeat sequence[12]. These experiments, while indirect, hint at the existence of a TTAGGG-repeat binding protein. Biochemical approaches have, so far, yielded a single protein component of mammalian telomeres, which is discussed here in detail. An overview of all candidate vertebrate telomeric DNA-binding factors can be found in Ref. 6.

The TTAGGG-repeat binding factor, TRF1

Human TRF1 was first identified in electrophoretic-mobility shift assays with HeLa cell nuclear extracts and tandem TTAGGG repeats as a probe[13]. The DNA-binding activity detected, termed TTAGGG-repeat binding factor (TRF1), shows specificity for the TTAGGG arrays found at mammalian telomeres. TRF1 binds double-stranded but not single-stranded telomeric DNA and it does not require the proximity of a DNA terminus for its binding. While TRF1 can recognize probes containing as few as three TTAGGG repeats, it displays increased binding efficiency to probes containing increasing numbers of TTAGGG repeats. Thus, TRF1 appears to bind preferentially to long contiguous repeat arrays, a property consistent with its binding along the length of mammalian telomeres *in vivo*. Finally, TRF1 shows a high degree of sequence specificity, preferring TTAGGG repeats over a number of closely related sequences.

TRF1 was purified from human cells as a 60 kDa protein and its corresponding cDNA cloned using peptide sequence analysis[14]. Comparison of the predicted primary sequence of human TRF1 with the databases indicates that it is not a previously identified protein. Both TRF1 mRNA and protein are detectable in all cells and tissues so far analysed, suggesting that this factor is required in every nucleated cell.

TRF1 is located at telomeres

Indirect immunofluorescence of cells transfected with an epitope-tagged TRF1 cDNA revealed a punctate staining pattern in interphase nuclei that co-localized (in double-labeling experiments) with telomeric DNA detected by TTAGGG-repeat-specific FISH (Refs 8, 14). In human metaphase cells, TRF1 was observed at chromosome ends (Ref. 14; Fig. 1), consistent with co-localization with telomeric DNA. Furthermore, TRF1 co-fractionated with telomeric DNA to the nuclear matrix, and co-localized with telomeres in double-labeling experiments on nuclear matrix preparations[8]. Thus, in interphase as well as in mitosis, TRF1 is a telomeric protein that caps all telomeres *in vivo*.

TRF1 functions at telomeres

Insights into TRF1 function emerged from studies on telomere formation *in vivo*[12]. New human telomeres can be formed by transfecting cells with a linear plasmid that carries a telomere seed of TTAGGG repeats at one end[15,16]. In this process, the formation of new telomeres is coupled to terminal deletions and chromosome fragmentation. In this assay telomere formation displays a critical dependence on the TTAGGG repeat sequence (Fig. 2; Ref. 12). This stringent sequence requirement is unlikely to reflect a requirement for telomerase because there is no correlation between the activity of telomere seeds and their interaction with telomerase *in vitro*[12,17]. In contrast, the sequence requirements for *de novo* telomere formation show a perfect match with the binding specificity of TRF1 (Fig. 2; Ref. 12), suggesting that the telomere seeds need to recruit TRF1 in order to form new telomeres (Fig. 2). TRF1 could serve at multiple steps in this process. For instance, binding of TRF1 could protect the telomere seed from enzymes, such as ligases and exonucleases, that normally act on transfected DNA and bring about integration. If the TRF1-bound telomere end of the construct cannot recombine, a cross-over between the non-telomeric end of the transfected DNA and a chromosome internal site will result in the chromosome truncations that are observed in the transfected cells (Fig. 2). In addition, the binding of TRF1 to the telomere seed might be required for the recruitment of telomerase to the new telomere. The proposed activities of TRF1 in telomere formation – protecting the telomere seed and facilitating interaction with telomerase – are in keeping with the proposed functions of TRF1 at endogenous telomeres (see below).

TRF1 contains a MYB-type DNA-binding domain

TRF1 activity is detected in nuclear extracts from human, monkey, rodent and chicken cells, suggesting that all vertebrates have a similar telomeric protein. Isolation and DNA sequence analysis of the full length mouse and human TRF1 (mTRF1 and hTRF1) cDNAs has allowed a prediction of the primary structure and a comparison across species (Fig. 3a; Refs 14, 18). Interestingly, human

and mouse TRF1 are not strongly conserved, suggesting that the mammalian telomeric proteins, as with those in yeast[19] and ciliates[20,21], evolve rapidly. This sequence divergence has allowed the identification of the most conserved domains of mammalian TRF1 proteins. mTRF1 and hTRF1 each contain an acidic amino terminal domain that, while not conserved in actual sequence, exhibits a local pI of about three in both cases. This region is followed by a conserved TRF1-specific domain of 200 amino acids with no obvious signature. The next 100 amino acids (the least conserved domain) contain putative nuclear localization signals. The final 50 amino acids are, again, highly conserved, with strong homology to the DNA-binding domain of the MYB family of proto-oncogenes (Fig. 3b).

MYB oncoproteins are transcriptional activators that bind to CNGTTPu sites using an N-terminal DNA-binding domain[22]. This domain consists of three imperfect tandem repeats of 50 amino acids, each composed of a three-helix bundle. The second and third helices form a helix-turn-helix motif, with the third helix serving as the DNA recognition helix[23]. This is a common motif found in several classes of DNA-binding proteins, including the prokaryotic repressors and homeobox proteins. Recent structural analysis has revealed that, in the MYB proteins, this motif is used in a novel way[23]. DNA binding is achieved through a cooperative interaction between the third helices of the second and third repeats. Although initially identified in MYB proteins, the MYB motif has subsequently been found in many different plant and animal proteins in the form of either one, two or three tandem repeats. Very little is known of how the MYB domain in these (mostly uncharacterized) proteins binds to DNA.

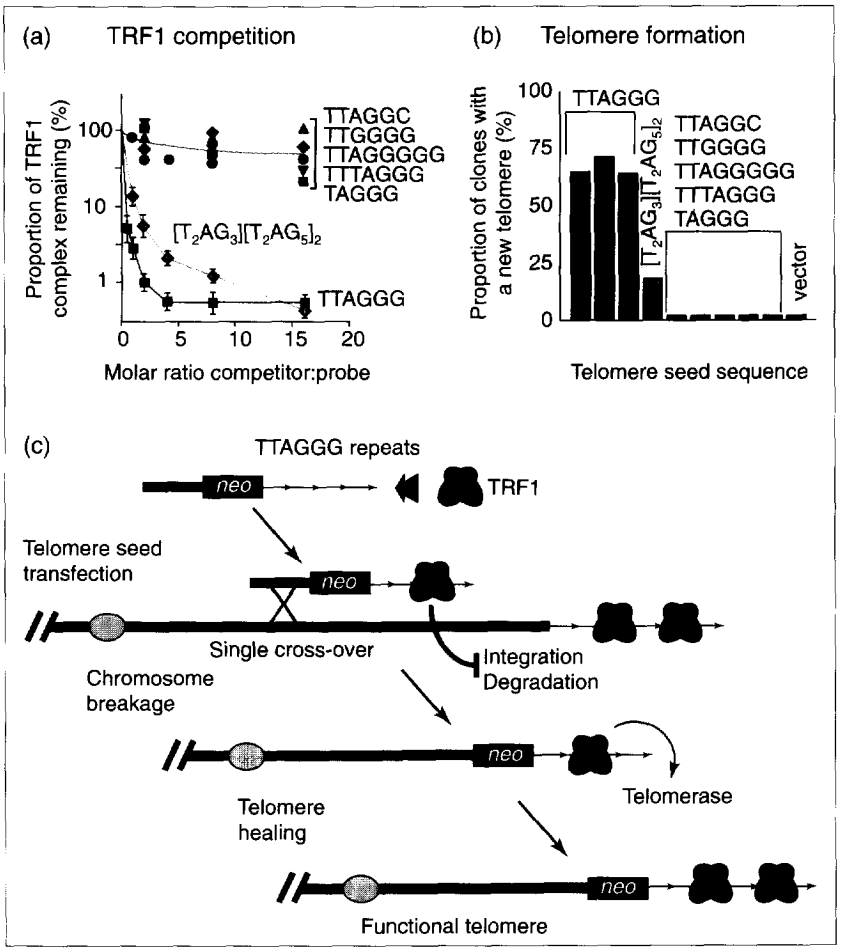

FIGURE 2. Involvement of TRF1 in the formation of new human telomeres by transfection of telomeric DNA (adapted from Ref. 12). (a) TRF1 binding *in vitro* is competed efficiently by TTAGGG repeats but not by heterologous telomeric DNAs composed of related G-rich repetitive sequences. (b) *De novo* telomere formation (as detected by genomic blotting of individual clones) occurs in approximately 70% of human cells that are stably transfected with a linear DNA carrying a *neo* marker and a stretch of 0.5–3.0 kb of TTAGGG repeats. By contrast, telomere seeds carrying heterologous telomeric DNAs do not lead to telomere formation. (c) Model for telomere formation in human cells. The initial step is speculated to require TRF1 binding to protect the end from recombination. Subsequent steps might require TRF1 to protect the end from degradation and to recruit telomerase.

TRF1 binds TTAGGG repeats as a dimer

The DNA-binding domain of TRF1 differs from MYB oncoproteins in at least two ways: TRF1 contains one (rather than three) MYB repeats and it is located at the extreme carboxy (rather than amino) terminus of the protein. In the light of the way MYB proteins use two repeats for DNA binding, it is interesting that TRF1, which contains only one MYB motif, binds DNA as a dimer (A. Bianchi, S. Smith and T. de Lange, unpublished). TRF1 is a dimer in solution as well as when bound to DNA, and it is likely that both MYB domains of the dimer are involved in site recognition. Yeast two-hybrid analysis indicates that the dimerization domain is located in the N-terminal part of TRF1, suggesting a function for the highly conserved domain of TRF1 (Fig. 3a). It is tempting to speculate that the use of a dimeric DNA-binding factor with two identical MYB domains is designed to recognize the tandem repeats typical of telomeric DNA. However, Rap1p, which also binds to a direct repeat, contains two MYB repeats that are not identical (see below).

Other TRF1-like MYB-type DNA-binding motifs

Although the TRF1 MYB domain is clearly related to the repeated motif in the vertebrate MYB oncoproteins, the MYB proteins are not TRF1's closest relatives[24]. There is one anonymous human cDNA in the database that encodes a TRF1-like MYB region, and similar MYB domains are found in a group of plant proteins and in the yeast factor Tbf1p (Refs 25, 26). In each case, the similarity to TRF1 is limited to the MYB homology region, particularly, the third (DNA recognition) helix (Fig. 3c). Bilaud *et al.* have suggested that their common

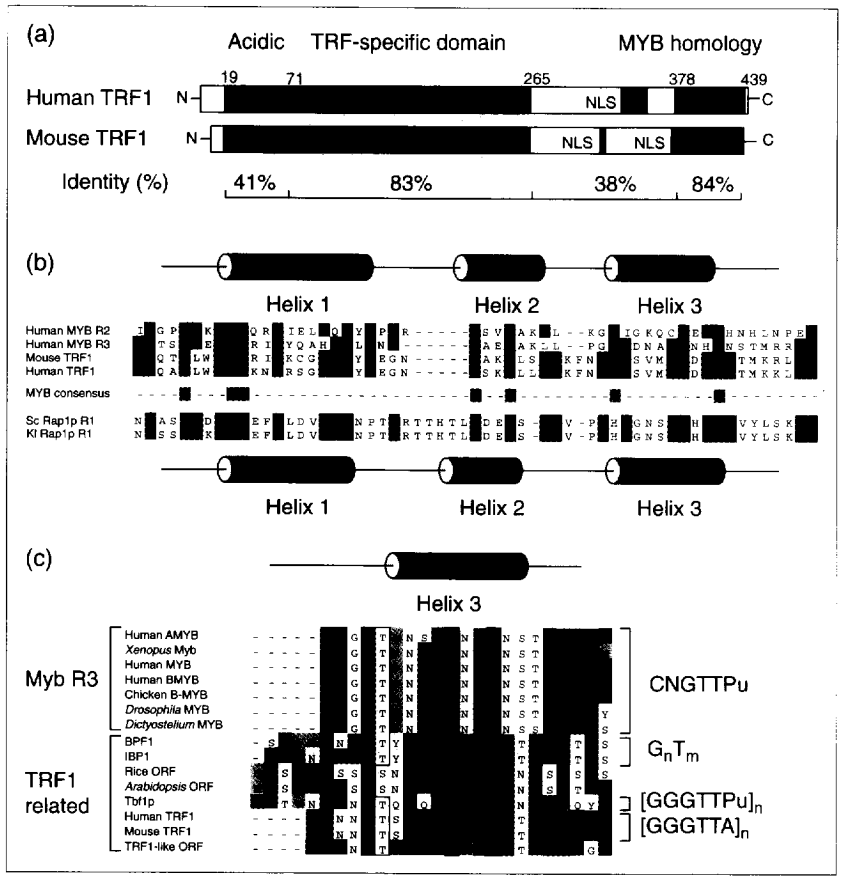

FIGURE 3. Primary structure of TRF1 and comparison with MYB-related proteins. (a) Domain structure of human and mouse TRF1 (Ref. 18). (b) Alignment of the two three-helix bundles (R2 and R3) in the DNA-binding domain of human MYB, the single MYB repeat in human and mouse TRF1, the first MYB repeat (R1) of *Saccharomyces cerevisiae* (Sc) and *Kluyveromyces lactis* (Kl) Rap1p, and the MYB consensus. Conserved residues are highlighted. (c) Alignment of the recognition (third) helices of the third repeat (R3) of MYB proteins and a group of TRF1-related proteins. Highly conserved positions are boxed. Amino acids of similar nature (nonpolar, uncharged polar, acidic and basic) are given the same color to underscore similarities in the sequences. Human MYB binds to CNGTTPu, BPF1 and IBP1 bind to G/T rich sites, Tbf1p binds to TTAGGG and TTGGGG repeats, and TRF1 proteins are specific for TTAGGG repeats. All protein sequences are available in GenBank. The GenBank accession numbers of the sequences derived from cDNA fragments are: Rice ORF, D23805; *Arabidopsis* ORF, Z26064; TRF-like ORF, T58911. Abbreviation: ORF, open reading frame.

structure might reflect a shared function at telomeres[24]. Therefore, it will be interesting to see which of these proteins is actually located at telomeres. Because some of these factors are known to bind DNAs with clusters of G residues in one strand, their resemblance might also simply reflect certain sequence constraints in MYB-type recognition helices that recognize G/C-rich sequences. Furthermore, it should be noted that the MYB repeats of Rap1p (below) do not belong to the TRF1 family[24], indicating that the MYB regions of telomeric proteins are not particularly related in terms of their amino acid sequence.

Comparison of TRF1 with yeast Rap1p, a potential functional analog

To date, only one other duplex telomeric DNA-binding protein has been recognized, the yeast repressor/activator protein, Rap1p (Ref. 27). Initially identified as a transcriptional regulator in *Saccharomyces cerevisiae*, Rap1p binds to the promoters of many yeast genes to activate transcription, and associates with specific elements of the silent mating type genes to repress their expression. However, the highest affinity Rap1p sites are found within the TG_{1-3} tracts at the ends of yeast chromosomes. Each telomere can bind to in the order of ten Rap1p molecules, which, in turn, recruit a number of other telomere-associated factors to the telomeric complex. The binding of Rap1p to yeast telomeres controls telomere length, telomeric silencing (telomere position effect), and telomere stability (reviewed in Ref. 28). Rap1p contains two MYB repeats and is not the yeast homolog of TRF1. In fact, the yeast genome does not contain an open reading frame with significant sequence similarity to TRF1. However, as TRF1 and Rap1p are the only two *bona fide* duplex telomeric DNA-binding proteins identified so far, they are worthy of comparison.

Related DNA-binding domains

Although TRF1 and Rap1p are not homologous proteins, they have related DNA-binding domains. The recently solved crystal structure of the Rap1p DNA-binding domain (DBD)[29] revealed a tandem repeat of a three-helix bundle, most closely related in structure to the MYB motif (Fig. 3b). Although their DBDs can be superimposed, the mode of DNA recognition by Rap1p is quite different from MYB. In MYB, two repeats interact with each other and bind as one unit within the major groove. In contrast, the two repeats in Rap1p are separated by a linker that spans the minor groove, positioning the binding units to make major groove contacts with the two direct repeats in the Rap1p recognition site. Thus, while TRF1 has structural similarity to the DNA-binding fold of Rap1p and MYB, the dissimilar docking of Rap1p and MYB on DNA indicates that insight into TRF1's mode of DNA-binding awaits determination of its structure in complex with DNA. Recently, a third duplex telomeric DNA-binding protein has been identified in *Schizosaccharomyces pombe*, which, like Rap1p and TRF1, carries a MYB-like DNA-binding motif (J.C. Cooper, E.R. Nimmo, R.C. Allshire and T.R. Cech, unpublished), corroborating the view that while telomeric proteins evolve rapidly, their DNA-binding regions are structurally conserved.

In keeping with their relationship to the MYB oncoproteins, TRF1 and Rap1p both bend DNA. The bending angle for TRF1-bound DNA is in the order of 120° (A. Bianchi, S. Smith, L. Chong, P. Elias and T. de Lange,

unpublished), a distortion slightly less pronounced than that seen with MYB (Ref. 30) and Rap1p (Refs 31, 32). Bending might suggest that this is a common feature of MYB-type DNA-binding proteins. It is of interest to note, however, that for MYB and Rap1p the DBD, alone, is insufficient for full bending. Whether the bend induced by TRF1 is also dependent on domains outside the three-helix bundle is not known at this stage. The ability of TRF1 to bend DNA could play a role in generating a higher-order structure at the telomere (see below).

TRF1 and Rap1p carry transactivation domains

TRF1, like Rap1p, contains a putative transcriptional activation domain. When targeted to a promoter via a heterologous DNA-binding domain, the acidic N-terminus of TRF1 and the TRF-specific conserved domain will weakly activate transcription of a reporter gene (S. Smith and T. de Lange, unpublished). While the role of Rap1p in transcriptional regulation has been established *in vivo*, the significance of the transactivation activity of TRF1 is unclear. Specifically, there is as yet no evidence that implicates TRF1-binding sites in transcriptional control. Further work will be required to establish whether transactivation is a common theme in telomeric proteins, and what its relevance is to telomere function.

TRF1 does not complement Rap1p defects

In the light of the striking similarities between TRF1 and Rap1p, we designed experiments to determine if there is functional similarity between these two proteins (S. Smith, T. de Lange, A. Lustig, J. Stavenhagen and V. Zakian, unpublished). We asked if hTRF1 could mimic Rap1p-dependent silencing, a process that involves recruitment of the silencing complex to sites where Rap1p is bound. To this end, we explored the effects of hTRF1 expression in a yeast strain containing TTAGGG repeats upstream of the *URA3* gene. In this context TRF1 was unable to silence transcription of *URA3*. Similarly, we asked if TRF1 could function (like Rap1p) in silencing near telomeres by targeting a LexA–TRF1 fusion protein to a silencer containing LexA sites upstream of a subtelomeric *URA3* gene. Again, hTRF1 did not have detectable activity. By two-hybrid analysis, we found no evidence for interactions between hTRF1 and Sir3p, Sir4p and Rif1p, which are components of the silencing complex that interact with Rap1p in this setting[27]. Finally, overexpression of human TRF1 in yeast for approximately 100 generations had no effect on telomere length. Thus, while there are parallels between TRF1 and Rap1p, these two proteins are not sufficiently similar to allow functional complementation.

Insight into telomere function

Proteins that bind to telomeres have been implicated in two aspects of telomere function: the formation of a protective cap and the regulation of telomere length. TRF1 might participate in either of these activities. In addition, it is prudent to ask whether TRF1 contributes to the role telomeres play in tumorigenesis.

Telomere protection

In ciliates and yeast, the G-rich telomeric strand protrudes as a single-stranded 3′ overhang at chromosome ends[33]. Proteins that bind to this protrusion have been isolated from hypotrichous ciliates and are thought to function as the protective cap *in vivo* (reviewed in Ref. 34). Although a similar telomere terminus-binding activity has been identified in *Xenopus* egg extracts[35], there is no evidence yet for an analogous structure (a 3′ overhang and cognate binding protein) at mammalian telomeres. An alternative, but not necessarily exclusive, model for a protective cap at the mammalian telomere involves the double-stranded telomeric repeat and its cognate DNA-binding proteins. A distinguishing feature of the mammalian telomere is its extreme length (2–150 kb of tandem repeats) compared with the very short (less than 100 bp) telomeric tracts found in organisms that carry a terminal telomere protein. Such long arrays of tandem repeats might be required to assemble a higher-order structure at the telomere, burying the chromosome end and effectively hiding it from the cell's surveillance system. This could be achieved through the binding of TRF1 at discrete sites along the telomeric repeat. The intrinsic properties of TRF1, its ability to interact with itself and to bend DNA, could promote crosslinking of the telomeric DNA. Indeed, ultrastructural analysis indicates that human telomeres are highly condensed, TRF1-bound elements[8]. An important test of the protective cap model is what happens if TRF1 function is compromised. If the loss of TRF1 unmasks the chromosome end, it might appear to the cell as damaged DNA, and activate checkpoint controls. A critical question to be addressed in the future is whether the checkpoint that senses an unmasked telomere is the same as other DNA-damage checkpoints, or whether there is a telomere-specific checkpoint.

Telomere length regulation

Mammalian cells have a mechanism to measure and regulate the length of individual telomeres. For example, in telomere seed experiments (described above) the final length of a newly formed telomere is dictated by the length of the telomeres in the host cell[16]. Telomere-length regulation is also apparent in certain stable human cell lines, which maintain all their telomeres within the same size range, despite the presence of high levels of telomerase[36]. Furthermore, the length of telomeres is species specific and ranges from 5–20 kb in man to 20–150 kb in mice[37].

Hints at the molecular mechanisms underlying telomere length regulation come from studies in budding yeast. In *S. cerevisiae*, mutations in Rap1p or overexpression of its C-terminus result in changes in telomere length[28,38,39]. Studies in *Kluyveromyces lactis* show that mutations in the telomeric repeat array (probably altering the *K. lactis* Rap1p-binding sites) result in inappropriate telomere elongation[40,41]. A model for yeast telomere-length regulation is proposed whereby binding of a duplex telomeric DNA-binding protein directly, or indirectly, inhibits telomerase. By analogy with the yeast models, TRF1 would be the probable candidate for the regulation of telomere homeostasis in mammalian cells.

Tumor suppression

Human telomeres undergo programmed shortening in the soma[42]. This process appears to constitute a tumor suppressor mechanism, limiting the replicative life-span of normal human cells, and restraining the

growth of transformed clones[36]. In agreement with this, the immortalization of human cells is accompanied by the activation of a mechanism that restores telomere length, and telomerase is activated in the majority of human cancers (Refs 2, 36, 43; reviewed in Ref. 3). While there is a good correlation between the length of the TTAGGG repeat array and the number of divisions before arrest of a primary culture[44], it is unclear how cells sense their shortened telomeres. One possibility is that critically shortened telomeres fail to recruit sufficient telomeric protein (including TRF1) and become exposed. Presumably, such exposed ends could activate DNA-damage checkpoints and lead to cell-cycle arrest. An interesting alternative possibility is that as the telomeres shorten, unbound TRF1 accumulates and signals a telomere-specific checkpoint. These models are now testable.

Acknowledgements

Matthias Schäfer is thanked for contributing the experiment in Fig. 1. We thank Art Lustig (Memorial Sloan-Kettering Cancer Center), Daniela Rhodes (MRC, Cambridge), members of the de Lange laboratory, and Tom Meier (Albert Einstein College of Medicine) for helpful criticism of this review. Richard Lang (New York University Medical Center) is thanked for help with the figures. We would like to thank Tom Cech (University of Colorado, Boulder) for communicating unpublished information. Work in this laboratory is supported by grants from the NIH, the Rita Allen Foundation, the Irma T. Hirschl Foundation, and the Sandoz Foundation. Susan Smith is a Leukemia Society of America Special Fellow.

References

1 Blackburn, E.H. (1993) in *The RNA World* (Gesteland, R.F. and Atkins, J.F., eds), pp. 557–576, Cold Spring Harbor Laboratory Press
2 Autexier, C. and Greider, C.W. (1996) *Trends Biochem. Sci.* 21, 387–391
3 de Lange, T. (1995) in *Telomeres* (Blackburn, E.H. and Greider, C.W., eds), pp. 265–295, Cold Spring Harbor Laboratory Press
4 Zakian, V.A. (1995) *Science* 270, 1601–1607
5 Meyne, J., Ratlifff, R.L. and Moyzis, R.K. (1989) *Proc. Natl. Acad. Sci. U. S. A.* 86, 7049–7053
6 de Lange, T. (1996) *Semin. Cell Biol.* 7, 23–29
7 de Lange, T. (1992) *EMBO J.* 11, 717–724
8 Ludérus, M.E.E. *et al. J. Cell Biol.* (in press)
9 Makarov, V.L., Lejnine, S., Bedoyan, J. and Langmore, J.P. (1993) *Cell* 73, 775–787
10 Tommerup, H., Dousmanis, A. and de Lange, T. (1994) *Mol. Cell. Biol.* 14, 5777–5785
11 O'Neill, L.P. and Turner, B.M. (1995) *EMBO J.* 14, 3946–3957
12 Hanish, J.P., Yanowitz, J. and de Lange, T. (1994) *Proc. Natl. Acad. Sci. U. S. A.* 91, 8861–8865
13 Zhong, Z., Shiue, L., Kaplan, S. and de Lange, T. (1992) *Mol. Cell. Biol.* 13, 4834–4843
14 Chong, L. *et al.* (1995) *Science* 270, 1663–1667
15 Farr, C., Fantes, J., Goodfellow, P. and Cooke, H. (1991) *Proc. Natl. Acad. Sci. U. S. A.* 88, 7006–7010
16 Barnett, M. *et al.* (1993) *Nucleic Acids Res.* 21, 27–36
17 Morin, G.B. (1989) *Cell* 59, 521–529
18 Broccoli, D. *et al. Hum. Mol. Genetics* (in press)
19 Larson, G.P., Castanotto, D., Rossi, J.J. and Malafa, M.P. (1994) *Gene* 150, 35–41
20 Fang, G. and Cech, T.R. (1991) *Nucleic Acids Res.* 19, 5515–5518
21 Wang, W., Skopp, R., Scofield, M. and Price, C. (1992) *Nucleic Acids Res.* 20, 6621–6629
22 Weston, K. (1990) *Semin. Cancer Biol.* 1, 371–382
23 Ogata, K. *et al.* (1994) *Cell* 79, 639–648
24 Bilaud, T. *et al.* (1996) *Nucleic Acids Res.* 24, 1294–1303
25 Brigati, C. *et al.* (1993) *Mol. Cell. Biol.* 13, 1306–1314
26 Liu, Z. and Tye, B-K. (1991) *Genes Dev.* 5, 49–59
27 Shore, D. (1994) *Trends Genet.* 10, 408–412
28 Zakian, V.A. (1995) in *Telomeres* (Blackburn, E.H. and Greider, C.W., eds), pp. 107–138, Cold Spring Harbor Laboratory Press
29 Konig, P., Giraldo, R., Chapman, L. and Rhodes, D. (1996) *Cell* 85, 125–136
30 Saikumar, P., Gabriel, J.L. and Reddy, E.P. (1994) *Oncogene* 9, 1279–1287
31 Vignais, M-L. and Sentenac, A. (1989) *J. Biol. Chem.* 264, 8463–8466
32 Muller, T. *et al.* (1994) *J. Struct. Biol.* 113, 1–12
33 Henderson, E. (1995) in *Telomeres* (Blackburn, E.H. and Greider, C.W., eds), pp. 11–35, Cold Spring Harbor Laboratory Press
34 Fang, G. and Cech, T.R. (1995) in *Telomeres* (Blackburn, E.H. and Greider, C.W., eds), pp. 69–106, Cold Spring Harbor Laboratory Press
35 Cardenas, M.E., Bianchi, A. and de Lange, T. (1993) *Genes Dev.* 7, 883–894
36 Counter, C.M. *et al.* (1992) *EMBO J.* 11, 1921–1929
37 Kipling, D. and Cooke, H.J. (1990) *Nature* 347, 347–402
38 Conrad, M.N., Wright, J.H., Wolf, A.J. and Zakian, V.A. (1990) *Cell* 63, 739–750
39 Lustig, A.J., Kurtz, S. and Shore, D. (1990) *Science* 250, 549–553
40 McEachern, M.J. and Blackburn, E.H. (1995) *Nature* 376, 403–409
41 Krauskopf, A. and Blackburn, E.H. (1996) *Nature* 383, 354–357
42 Cooke, H.J. and Smith, B.A. (1986) *Cold Spring Harbor Symp. Quant. Biol.* 51, 213–219
43 Kim, N.W. *et al.* (1994) *Science* 266, 2011–2015
44 Allsopp, R. *et al.* (1992) *Proc. Natl. Acad. Sci. U. S. A.* 89, 10114–10118

***S. Smith** and **T. de Lange** are in the Rockefeller University, 1230 York Avenue, New York, NY 10021, USA.*

The molecular epidemiology of *P53* gene mutations in human breast cancer

ARNDT HARTMANN
(arndt.hartmann@klinik.uni-regensburg.de)

HAGEN BLASZYK (blaszyk.hagen@mayo.edu)

JOHN S. KOVACH (jkovach@smtplink.coh.org)

STEVE S. SOMMER (ssommer@smtplink.coh.org)

The P53 tumor-suppressor gene is an advantageous tool for analyzing the molecular epidemiology of cancer. We describe the utility of the P53 gene as a 'mutagen test' and a prognostic indicator in breast cancer. Aspects of study design and methodology are discussed. Two major conclusions emerge: (1) there is an extraordinary diversity of mutational patterns among cohorts, hinting that the unique biology of mammary cells results in exposure to high doses of a diversity of ingested lipophilic mutagens; and (2) mutations in the P53 gene predict poor outcome in breast cancer.

Cancer arises from somatic mutations that derive from mutagen exposure and/or endogenous processes[1]. Classical epidemiological studies have sought associations between cancer(s) in high-risk populations and exposures to putative mutagens. Mutagens have been implicated consistently for a few cancers, but reproducible associations have been elusive for many, including breast cancer.

The observed patterns of mutation in the *P53* gene and possibly other cancer-related genes might implicate specific mutagens in cancers for which classical epidemiological studies have been equivocal. This approach is being utilized in breast cancer, a disease in which there is at least a fourfold variation in incidence between racially and/or geographically diverse populations[2].

The *P53* gene as a 'mutagen test'

The precise biological function(s) of *P53* is unclear, but substantial data suggest that *P53* is a transcription factor that regulates cell proliferation and apoptosis (reviewed in Refs 3–5; Fig. 1). Most *P53* gene mutations found in human cancers produce missense proteins with altered or absent transcriptional regulation activities and are associated with high concentrations of P53 protein detectable by immunohistochemistry[6]. Missense p53 proteins might act in a dominant-negative fashion by binding to and inactivating wild-type P53, while null mutations are presumably recessive at the cellular level[7]. The loss of *P53* function eliminates growth arrest in response to certain DNA-damaging agents and enhances the frequency of gene amplification, suggesting a role for *P53* in the control of the cell-cycle checkpoint and in the maintenance of the integrity of the genome[8]. However, analysis of the Big Blue™ transgenic mouse mutation detection system suggests that nullizygosity for the *P53* gene does not increase the frequency of point mutations or small deletions in normal tissues, or in lymphomas that arise in these mice[9,10].

The *P53* gene is well suited as a 'mutagen test'[11–13] for several reasons: (1) mutations often occur early in the neoplastic process[14–17] and should reflect chronic mutagen exposures; (2) mutagens leave specific molecular fingerprints[18], which facilitate their detection; (3) *P53* gene mutations occur frequently in nearly all tumor types[19], allowing comparisons of the patterns of mutation in different populations within and between tumor types; (4) a large variety of missense as well as null *P53* mutations are found in cancers; and (5) the regions of probable functional significance are of a convenient size for the purposes of mutation screening.

The patterns of *P53* gene mutation in high risk cohorts have provided strong confirmatory evidence for carcinogenic mutagens, previously implicated by classical epidemiological studies, for example, G:C→T:A transversions and lung cancer in smokers[20]; G:C→T:A transversions at codon 249 in hepatocellular carcinoma associated with aflatoxin B_1 exposure[21,22]; A:T→T:A transversions in angiosarcomas associated with vinyl chloride exposures[23]; and CC:GG→TT:AA tandem dipyrimidine transitions as well as C:G→T:A transitions in carcinomas from sun-exposed skin[24,25]. In the above instances, only one mutagen predominated. However, there might be multiple mutagens for some cancers, and/or genetic differences might cause specific patterns of mutations in different populations exposed to the same mutagen[26]. Analysis of *P53* gene mutations might be helpful in untangling these complexities.

Methodological issues

The mutations observed in tumors depend on the underlying pattern of mutation within the gene, that is, the pattern that would be observed if mutations could be defined in segments of genomic DNA, under little if any selective pressure. However, the biology of the gene as well as the biology of the tumor generally skew the pattern of observed mutations. If the biology of the gene is constant when a given tumor type is examined, differences in the pattern should reflect an altered balance of endogenous mutagenic processes and exogenous mutagen exposures.

The optimal analysis of the *P53* gene as a mutagen test requires careful attention to patient ascertainment biases, methods of sample processing and mutation detection (Table 1). With the exception of population-based ascertainment, a number of breast cancer studies have utilized optimal or nearly optimal methodology for this purpose. For example, DNA is amplified by utilizing either touch preparations[27] or tissue sections from consecutive tumors with good clinical data, and genomic DNA is analyzed in all coding regions (or at least exons 5–8), by direct sequencing or a screening method in which virtually 100% sensitivity had been demonstrated by blind analyses[28–35]. When comparing mutational patterns, it is critical to compare mutations from identical exons because the patterns vary by exon,

 0168-9525/97/$17.00
PII: S0168-9525(96)10043-3