Science
AAAS

A Shared Docking Motif in TRF1 and TRF2 Used for Differential Recruitment of Telomeric Proteins

Yong Chen, *et al.*

Science **319**, 1092 (2008);
DOI: 10.1126/science.1151804

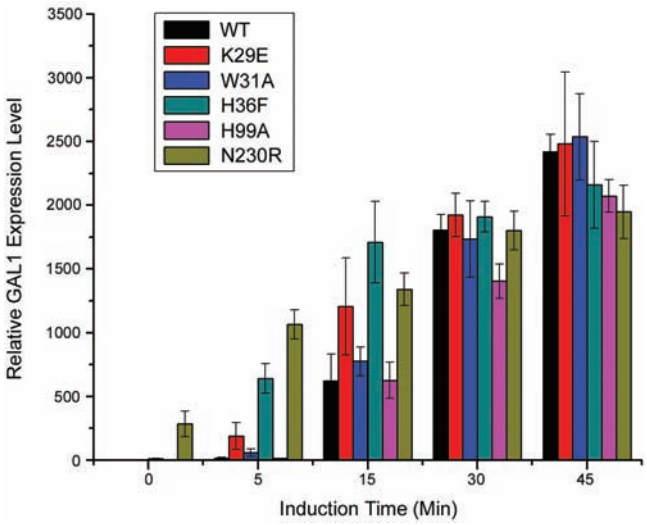


Fig. 3. Alterations in the NADP binding site changes the rate of induction in vivo. *GAL1* mRNA expression as a function of time after galactose induction. Data are shown for wild-type Gal80p and for Gal80p point mutants. All data were normalized to RNA levels measured for a control gene, *PMA1.* A *gal80*Δ mutant has a high expression level even when uninduced—as high as that seen for wild-type Gal80p when fully induced. The dimer mutant, N230R, also shows expression in the uninduced state (SOM text).

both NAD and NADP binding, would therefore disrupt both the stabilizing effect of NAD and destabilizing effect of NADP with a net result of faster induction for the mutants compared to the wild type.

The involvement of dinucleotides and metabolic factors in transcriptional regulation is seen in a few other systems. The coactivator of Oct-1, OCA-S, contains two glycolytic enzymes—glyceraldehyde-3-phosphate dehydrogenase (GAPDH) and lactate dehydrogenase (*10*). The binding of the transcriptional corepressor complex, CtBP, is enhanced by the reduced dinucleotide NADH compared to the oxidized form (*11*) and it possesses a NAD-dependent dehydrogenase activity (*12*). The DNA-binding activity of the transcription factor neuronal PAS domain protein 2 (NPAS2) is sensitive to the oxidation state of NAD, with DNA binding enhanced by the reduced form of the dinucleotide (*13*). Although we do not understand precisely how this trigger for *GAL* regulation functions, nor the involvement of NADP versus NAD, we speculate that switching the cell to a fermentable galactose medium causes a change in NADP/NADPH or NADP/NAD ratios in the cell, and Gal80p effectively senses the metabolic state of the cell. NADP might be acting as a "second messenger" in triggering the system. Alternatively, Gal80p may function as an oxidoreductase, actively converting NADPH to NADP in the presence of a substrate and causing it to disassociate from Gal4p.

References and Notes

1. P. J. Bhat, T. V. S. Murthy, *Mol. Microbiol.* **40**, 1059 (2001).
2. G. Peng, J. E. Hopper, *Mol. Cell. Biol.* **20**, 5140 (2000).
3. J. Ma, M. Ptashne, *Cell* **50**, 137 (1987).
4. L. Keegan, G. Gill, M. Ptashne, *Science* **231**, 699 (1986).
5. S. A. Johnston, J. M. Salmeron Jr., S. S. Dincher, *Cell* **50**, 143 (1987).
6. Y. Han, T. Kodadek, *J. Biol. Chem.* **275**, 14979 (2000).
7. Abbreviations for the amino acid residues are as follows: A, Ala; D, Asp; E, Glu; F, Phe; G, Gly; H, His; K, Lys; L, Leu; N, Asn; Q, Gln; R, Arg; V, Val; and W, Trp.
8. J. B. Thoden, C. A. Sellick, R. J. Reece, H. M. Holden, *J. Biol. Chem.* **282**, 1534 (2007).
9. M. B. Einarson, in *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 2001), pp. 18.55–18.59.
10. L. Zheng, R. G. Roeder, Y. Luo, *Cell* **114**, 255 (2003).
11. C. C. Fjeld, W. T. Birdsong, R. H. Goodman, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 9202 (2003).
12. V. Kumar *et al.*, *Mol. Cell* **10**, 857 (2002).
13. J. Rutter, M. Reick, L. C. Wu, S. L. McKnight, *Science* **293**, 510 (2001).
14. Materials and methods are available as supporting material on *Science* Online.
15. We thank T. Messick, K. Siddiqui, M. Rossman, J. Hicks, A. Gann, S. Harrison, and members of the Joshua-Tor lab for discussions and advice; A. Heroux, M. Becker, and H. Robinson for support at the National Synchrotron Light Source (NSLS) and S. Ginell and N. Duke for support at the Advanced Photon Source (APS). The NSLS and the APS are supported by the U.S. Department of Energy, Office of Basic Energy Sciences. Coordinates and structure factors have been submitted to the Protein Data Bank. Accession numbers: Gal80p^{S0}-ScGal4AD-NAD: 3BTS; Gal80p^{S2}: 3BTU; Gal80p^{S0}: 3BTV. This work was supported by NIH grants GM074075 (to L.J.) and GM55641 (to R.S.).

Supporting Online Material
www.sciencemag.org/cgi/content/full/319/5866/1090/DC1
Materials and Methods
SOM Text
Figs. S1 to S7
Tables S1 to S5
References

18 October 2007; accepted 4 January 2008
10.1126/science.1151903

A Shared Docking Motif in TRF1 and TRF2 Used for Differential Recruitment of Telomeric Proteins

Yong Chen,[1] Yuting Yang,[1] Megan van Overbeek,[2]* Jill R. Donigian,[2]* Paul Baciu,[1] Titia de Lange,[2] Ming Lei[1]†

Mammalian telomeres are protected by a six-protein complex: shelterin. Shelterin contains two closely related proteins (TRF1 and TRF2), which recruit various proteins to telomeres. We dissect the interactions of TRF1 and TRF2 with their shared binding partner (TIN2) and other shelterin accessory factors. TRF1 recognizes TIN2 using a conserved molecular surface in its TRF homology (TRFH) domain. However, this same surface does not act as a TIN2 binding site in TRF2, and TIN2 binding to TRF2 is mediated by a region outside the TRFH domain. Instead, the TRFH docking site of TRF2 binds a shelterin accessory factor (Apollo), which does not interact with the TRFH domain of TRF1. Conversely, the TRFH domain of TRF1, but not of TRF2, interacts with another shelterin-associated factor: PinX1.

Shelterin acts in conjunction with many associated factors (*1*–*6*). Most of the shelterin-associated proteins are recruited to telomeres through interactions with TRF1 or TRF2 (*2*–*6*). However, the molecular mechanism of these TRF1- and TRF2-mediated interactions remains unknown. TRF1 and TRF2 share the same molecular architecture, characterized by a C-terminal Myb/SANT DNA binding domain (*7*, *8*) and an N-terminal TRFH domain (*9*). The TRFH domains ($TRF1_{TRFH}$ and $TRF2_{TRFH}$) mediate homodimerization and are required for telomeric DNA binding by TRF1 and TRF2 (*10*, *11*). Several different protein interactions have been mapped to the TRFH domains of TRF1 and TRF2 (*2*, *12*–*14*). The TRFH domains have almost identical three-dimensional structures (*11*); therefore, it is difficult to explain how TRF1 and TRF2 can interact with different proteins.

TRF1 and TRF2 both bind to another shelterin protein: TIN2 (*12*, *15*, *16*). The TRF1-TIN2 interaction was mediated by $TRF1_{TRFH}$ and the C terminus of TIN2 (*12*). Further mapping revealed that a peptide of TIN2—denoted

[1]Department of Biological Chemistry, University of Michigan Medical School, 1150 West Medical Center Drive, Ann Arbor, MI 48109, USA. [2]Laboratory for Cell Biology and Genetics, Rockefeller University, 1230 York Avenue, New York, NY 10065, USA.

*These authors contributed equally to this work.
†To whom correspondence should be addressed. E-mail: leim@umich.edu

as $TIN2_{256–276}$ [$TIN2_{TBM}$: TIN2–TRFH binding motif (TBM)]—retains the $TRF1_{TRFH}$ binding activity with a binding affinity of 314 nM (figs. S1 and S2 and Fig. 1A). To understand how $TIN2_{TBM}$ is recognized by $TRF1_{TRFH}$, we crystallized the $TRF1_{TRFH}$-$TIN2_{TBM}$ complex and solved its structure at 2.0 Å resolution (table S1) (*17*). The electron density map shows that residues 257 to 268 of $TIN2_{TBM}$ assume a well-defined conformation (fig. S3). $TRF1_{TRFH}$ forms homodimers, and each $TRF1_{TRFH}$ interacts with one $TIN2_{TBM}$ peptide (Fig. 1B). $TRF1_{TRFH}$ exhibits essentially the same conformation as unliganded $TRF1_{TRFH}$ except for loop L34 (Fig. 1C) (*11*). Loop L34 is partially disordered

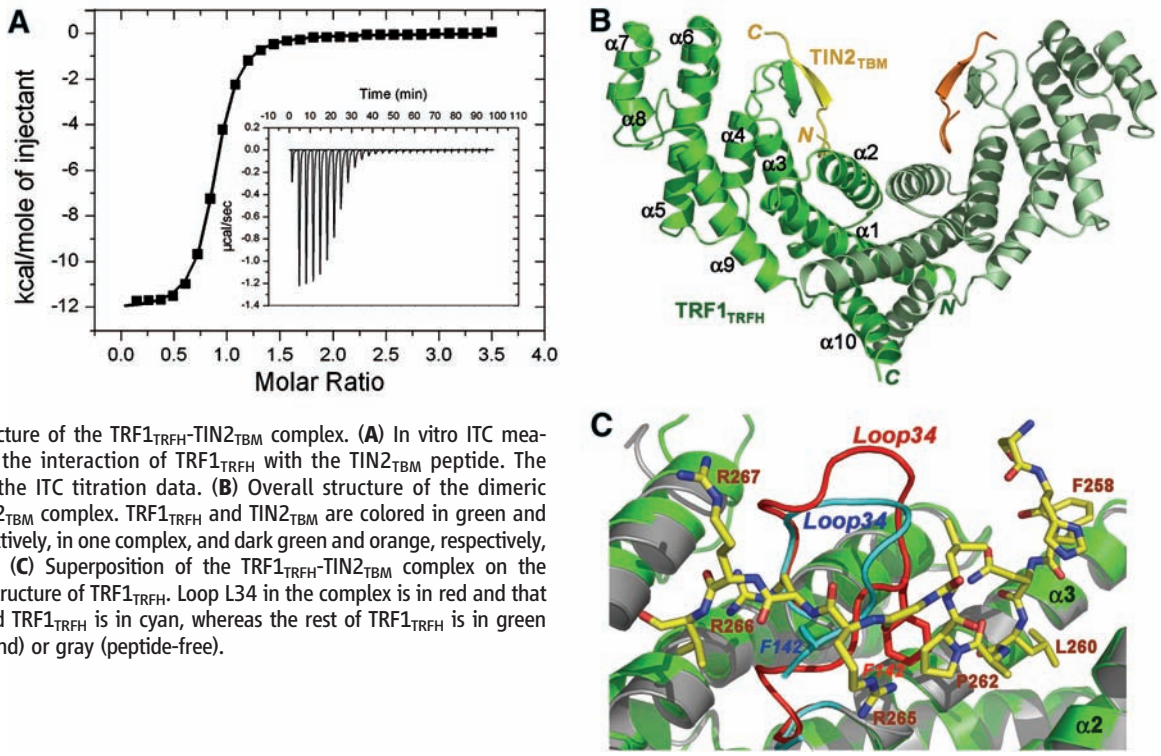
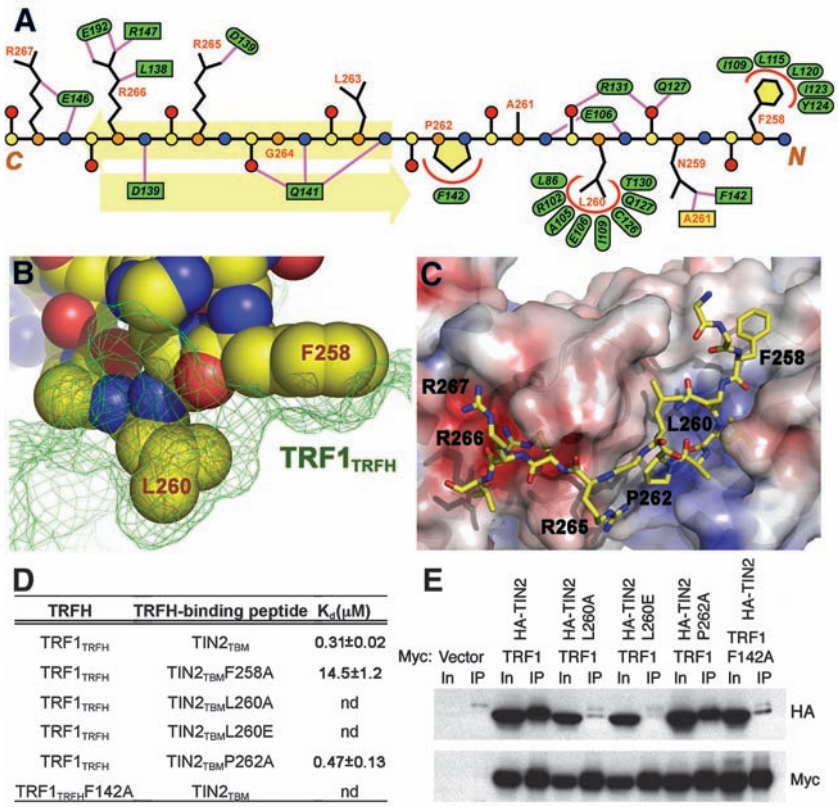


Fig. 1. Structure of the $TRF1_{TRFH}$-$TIN2_{TBM}$ complex. (**A**) In vitro ITC measurement of the interaction of $TRF1_{TRFH}$ with the $TIN2_{TBM}$ peptide. The inset shows the ITC titration data. (**B**) Overall structure of the dimeric $TRF1_{TRFH}$-$TIN2_{TBM}$ complex. $TRF1_{TRFH}$ and $TIN2_{TBM}$ are colored in green and yellow, respectively, in one complex, and dark green and orange, respectively, in the other. (**C**) Superposition of the $TRF1_{TRFH}$-$TIN2_{TBM}$ complex on the unliganded structure of $TRF1_{TRFH}$. Loop L34 in the complex is in red and that of unliganded $TRF1_{TRFH}$ is in cyan, whereas the rest of $TRF1_{TRFH}$ is in green ($TIN2_{TBM}$-bound) or gray (peptide-free).



TRFH	TRFH-binding peptide	K_d(μM)
$TRF1_{TRFH}$	$TIN2_{TBM}$	0.31±0.02
$TRF1_{TRFH}$	$TIN2_{TBM}$F258A	14.5±1.2
$TRF1_{TRFH}$	$TIN2_{TBM}$L260A	nd
$TRF1_{TRFH}$	$TIN2_{TBM}$L260E	nd
$TRF1_{TRFH}$	$TIN2_{TBM}$P262A	0.47±0.13
$TRF1_{TRFH}$F142A	$TIN2_{TBM}$	nd

Fig. 2. The $TRF1_{TRFH}$-$TIN2_{TBM}$ interface. (**A**) Schematic depiction of the $TRF1_{TRFH}$-$TIN2_{TBM}$ interaction. The main-chain atoms of $TIN2_{TBM}$ are shown as circles [carbon in yellow (Cα in orange), oxygen in red, and nitrogen in blue]. Residues of $TRF1_{TRFH}$ are shown as green ovals (side-chain interaction) and square boxes (main-chain interaction). Hydrophilic and hydrophobic interactions are shown as straight magenta lines and curved red lines, respectively. The pale yellow arrows denote the intermolecular β sheet. (**B**) The shape of the hydrophobic pocket of TRF1 (green mesh) complements the side chain of TIN2-L260 well. (**C**) Electrostatic surface potential of the $TIN2_{TBM}$ binding site of $TRF1_{TRFH}$. Positive potential, blue; negative potential, red. (**D**) In vitro ITC binding data of wild-type and mutant $TRF1_{TRFH}$-$TIN2_{TBM}$ interactions. K_d, equilibrium dissociation constant; nd, not detectable by ITC. (**E**) Co-IP of the same sets of mutant TRF1-TIN2 interactions (except the TRF1-TIN2 F258A interaction) as in (D). Lanes marked "In" represent 2.5% of input cell lysate used for the immunoprecipitation.

in the peptide-free structure (Fig. 1C). However, once $TIN2_{TBM}$ is bound, loop L34 folds back upon helices α3 and α4, sandwiched between the helices and $TIN2_{TBM}$ (Fig. 1C).

The structure of the complex reveals two adjacent but structurally distinct interaction modes. The N terminus of $TIN2_{TBM}$ [His^{257}-Phe-Asn-Leu-Ala-Pro^{262} (H257-F-N-L-A-P262)] (*18*) adopts an extended conformation stabilized by an extensive intermolecular hydrogen-bonding network (Fig. 2A and fig. S4). The side chain of L260 is therefore positioned into a deep hydrophobic pocket of $TRF1_{TRFH}$ (Fig. 2, B and C). In addition, F258 and P262 also make hydrophobic contacts with $TRF1_{TRFH}$: F258 sits on a concave hydrophobic surface, whereas P262 stacks with TRF1-*F142* (italics are used here for individual residues of TRF1 and TRF2) (Fig. 2, A to C, and fig. S4). In contrast, the C terminus of $TIN2_{TBM}$ (L263-G-R-R-R-V268) is positioned on the surface of loop L34 through formation of an antiparallel β sheet with *D139-A-Q141* of $TRF1_{TRFH}$ (Fig. 2A and fig. S4) so that R265-R-R267 of $TIN2_{TBM}$ contacts $TRF1_{TRFH}$ through electrostatic interactions (Fig. 2C). In particular, R266 is nested within an acidic depression on the surface of loop L34 through a network of salt bridges and hydrogen bonds (Fig. 2, A and C, and fig. S4).

To investigate the importance of the TRF1-TIN2 interaction, we first measured the binding of different mutant $TIN2_{TBM}$ peptides to $TRF1_{TRFH}$ by isothermal titration calorimetry (ITC). Substitution of L260 with either an alanine or a glutamate abolished the binding (Fig. 2D). Similarly, mutant $TIN2_{TBM}$–F258→A258 ($TIN2_{TBM}$-F258A) substantially impaired the interaction (Fig. 2D). By contrast, mutant $TIN2_{TBM}$-P262A, designed to eliminate a stacking interaction with TRF1-*F142*, had a wild-type binding affinity, indicating that loss of this interaction is not essential for binding (Fig. 2D). However, substitution of TRF1-*F142* with an alanine completely abrogated the binding to $TIN2_{TBM}$ (Fig. 2D). We then tested the interactions of mutant proteins transiently expressed in human embryonic kidney 293T cells, and the coimmunoprecipitation (Co-IP) results are consistent with the in vitro ITC measurements (Fig. 2E). We therefore conclude that the TRFH interaction motif in TRF1 is necessary for the TRF1-TIN2 interaction both in vitro and in vivo.

Given the sequence and structural similarities of the TRFH domains of TRF1 and TRF2, we expected that TRF2 would also bind to TIN2 through the TRFH domain (figs. S5 and S6). However, Co-IP studies of a specific mutant in TRF2 (TRF2-*F120A*, where TRF2-*F120* is structurally equivalent to TRF1-*F142*), which was predicted to abolish TIN2 binding to $TRF2_{TRFH}$, did not have the expected effect (Fig. 3A). Therefore, $TRF2_{TRFH}$ is not required for the stable association with TIN2 in vivo. In order to define the actual TIN2 binding site, we tested an array of glutathione *S*-transferase–TRF2 fusion fragments in a Far-Western assay for their ability to interact with TIN2. The result showed that a short peptide of TRF2 ($TRF2_{352–365}$) can mediate an efficient interaction with TIN2 (Fig. 3B). In addition, purified $TRF2_{350–366}$ comigrated with $TIN2_{1–220}$ in gel-filtration chromatographic analysis, indicating that $TIN2_{1–220}$ is sufficient for binding (fig. S7). Furthermore, Co-IP data showed that a deletion mutant of TRF2 (TRF2-Δ352–367) that retains the entire TRFH domain but lacks the TIN2 binding site failed to associate with TIN2 (Fig. 3A). Therefore, $TRF2_{TRFH}$ does not mediate a stable interaction with TIN2 in vivo. Collectively, we conclude that, although TRF1 binds TIN2 through its TRFH domain, TRF2 interacts with TIN2 through a short motif in its C terminus.

The distinctive specificity of the TRFH domains of TRF1 and TRF2 suggested that subtle structural differences are responsible for the ability of TIN2 to distinguish between these two paralogous proteins. ITC measurement showed that $TRF2_{TRFH}$ interacts with $TIN2_{TBM}$ in vitro, but with a much lower affinity (6.49 μM) (fig. S8A). To understand this binding specificity, we solved the crystal structure of the $TRF2_{TRFH}$-$TIN2_{TBM}$ complex at 2.15 Å resolution (fig. S8B and table S1). Although the overall conformations of $TIN2_{TBM}$

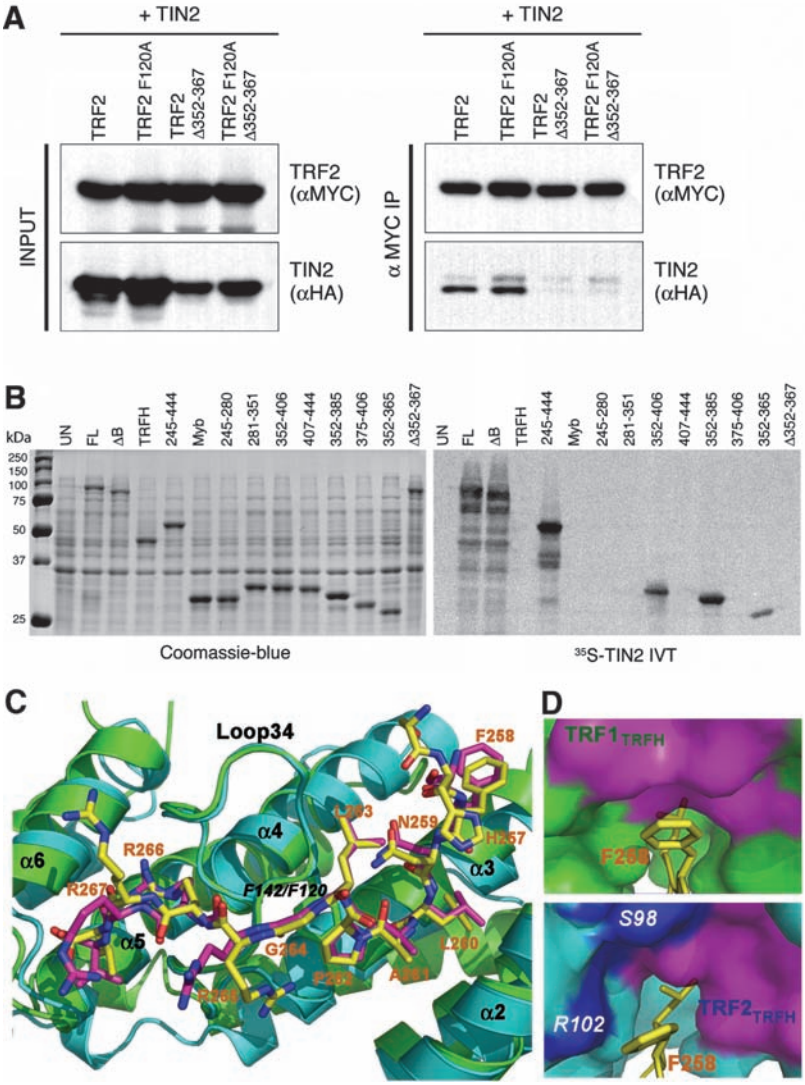


Fig. 3. The TRF2-TIN2 interaction. (**A**) Co-IP of TIN2 with cotransfected wild-type and mutant TRF2. (**B**) Far-Western analysis of the TIN2 binding region of TRF2 (FL, full-length; TRF2-ΔB, TRF2-Δ1–42). (**C**) Superposition of the $TIN2_{TBM}$ binding sites in the $TRF1_{TRFH}$-$TIN2_{TBM}$ and $TRF2_{TRFH}$-$TIN2_{TBM}$ complexes. $TRF1_{TRFH}$ and $TRF2_{TRFH}$ are in green and cyan, respectively. The $TIN2_{TBM}$ peptides bound to $TRF1_{TRFH}$ and $TRF2_{TRFH}$ are shown in stick model format and in yellow and magenta, respectively. (**D**) TIN2-F258 interacts less efficiently with TRF2 than with TRF1. The F258 binding surfaces of $TRF1_{TRFH}$ (top panel) and $TRF2_{TRFH}$ (bottom panel) are shown in magenta (hydrophobic patch) and blue (hydrophilic patch). The rest of $TRF1_{TRFH}$ and $TRF2_{TRFH}$ is in green and cyan, respectively.

bound to $TRF1_{TRFH}$ and $TRF2_{TRFH}$ are very similar (Fig. 3C), subtle differences can explain the difference in affinities of the two complexes (Fig. 3D and fig. S8, C and D). In the $TRF1_{TRFH}$-$TIN2_{TBM}$ complex, TIN2-F258 sits snugly on a hydrophobic surface of $TRF1_{TRFH}$ (Fig. 3D). In contrast, F258 rotates away from the interface and packs less efficiently with $TRF2_{TRFH}$, because the edge of the interaction surface is partially occupied by polar residues *S98* and *R102* (Fig. 3D). In addition, TRF1-*E192*, which is key for $TIN2_{TBM}$ binding, is replaced by a lysine residue in TRF2 (*K173*), resulting in loss of two ion-pairing interactions and an electrostatically unfavorable contact between TIN2-R266 and TRF2-*K173* (figs. S6 and S9).

These results suggested that TRF2 might use its TRFH domain peptide docking site to recruit one or more of the shelterin accessory factors (*2*, *13*, *19*). $TRF2_{TRFH}$ is known to interact with Apollo, which functions together with TRF2 in protecting telomeres during S phase (*2*, *13*). $TRF2_{TRFH}$ directly binds to the C terminus of Apollo ($Apollo_{496–532}$) (fig. S10A) (*13*). We confirmed this interaction using the ITC binding assay (Fig. 4A). Under the same conditions, no binding enthalpy was measurable between $Apollo_{496–532}$ and $TRF1_{TRFH}$, indicating that $Apollo_{496–532}$ binding is specific for TRF2 (Fig. 4A). To understand how TRF2 recognizes Apollo, we determined the crystal structure of the $TRF2_{TRFH}$-$Apollo_{496–532}$ complex at 2.5 Å resolution (Fig. 4B and table S1). The structure clearly shows electron density corresponding to the 12 N-terminal residues of $Apollo_{496–532}$ (amino acids 498 to 509), referred to as $Apollo_{TBM}$ (fig. S10, A and B). The structure reveals that $Apollo_{TBM}$ interacts with $TRF2_{TRFH}$ through the same molecular surface as in the $TRF1_{TRFH}$-$TIN2_{TBM}$ complex (Fig. 4C). Overlay of the two complexes reveals many similarities between the C terminus of $Apollo_{TBM}$ (Y504-L-L-T-P-V509) and the N terminus of $TIN2_{TBM}$ (F258-N-L-A-P-G265). First, two peptides are almost identical in overall conformation (Fig. 4C and fig. S10, C and D). Second, most of the hydrogen bonds in the $TRF2_{TRFH}$-$Apollo_{TBM}$ complex are conserved in $TRF1_{TRFH}$-$TIN2_{TBM}$ (fig. S10, C and E). Third, L506 and P508 of Apollo interact with $TRF2_{TRFH}$ in the same fashion as do their counterparts of $TIN2_{TBM}$ (Fig. 4C and fig. S10D). It is noteworthy that the TBMs of TIN2 and Apollo share the sequence Y/F-X-L-X-P (where X is any amino acid).

Despite the high degree of similarity between the $TRF1_{TRFH}$-$TIN2_{TBM}$ and $TRF2_{TRFH}$-$Apollo_{TBM}$ interactions, substantial structural variations are evident outside the Y/F-X-L-X-P motif. Unlike $TIN2_{TBM}$, the Y-X-L-X-P motif resides at the C terminus of $Apollo_{TBM}$, and $Apollo_{TBM}$ lacks a C-terminal polyarginine tail (Fig. 4C). Instead, it has a six-residue extension preceding the Y/F-X-L-X-P motif, which adopts a short helical conformation (Fig. 4, C and D) and packs on loop L23 and helices α2 and α3 of $TRF2_{TRFH}$ through hydrophobic contacts (Fig. 4D and fig. S10, C and E). Apollo-Y504 rotates ~ 90° relative to TIN2-F258 in the $TRF1_{TRFH}$-$TIN2_{TBM}$ complex to fit into a hydrophobic cleft formed by *L101* and *R102* of TRF2 (Fig. 4D). This reorientation of Y504 is coupled with a partial refolding of loop L23 of TRF2: TRF2-*E94* rotates ~ 180° relative to its position in the peptide-free conformation and makes two electrostatic interactions with K503 and Y504 of Apollo (Fig. 4D and fig. S10E). These marked conformational differences suggest that a tyrosine residue is preferred at the N-terminal position of the F/Y-X-L-X-P motif for efficient binding to $TRF2_{TRFH}$, whereas a phenylalanine is preferred for $TRF1_{TRFH}$. Furthermore, superposition of the $TRF1_{TRFH}$-$TIN2_{TBM}$ and the $TRF2_{TRFH}$-$Apollo_{TBM}$ complexes shows that the space occupied by L500 and Y504 of $Apollo_{TBM}$ is occluded in $TRF1_{TRFH}$, which explains why $Apollo_{TBM}$ binding is $TRF2_{TRFH}$-specific (Fig. 4E and fig. S11). Given the close similarity of the TRFH do-

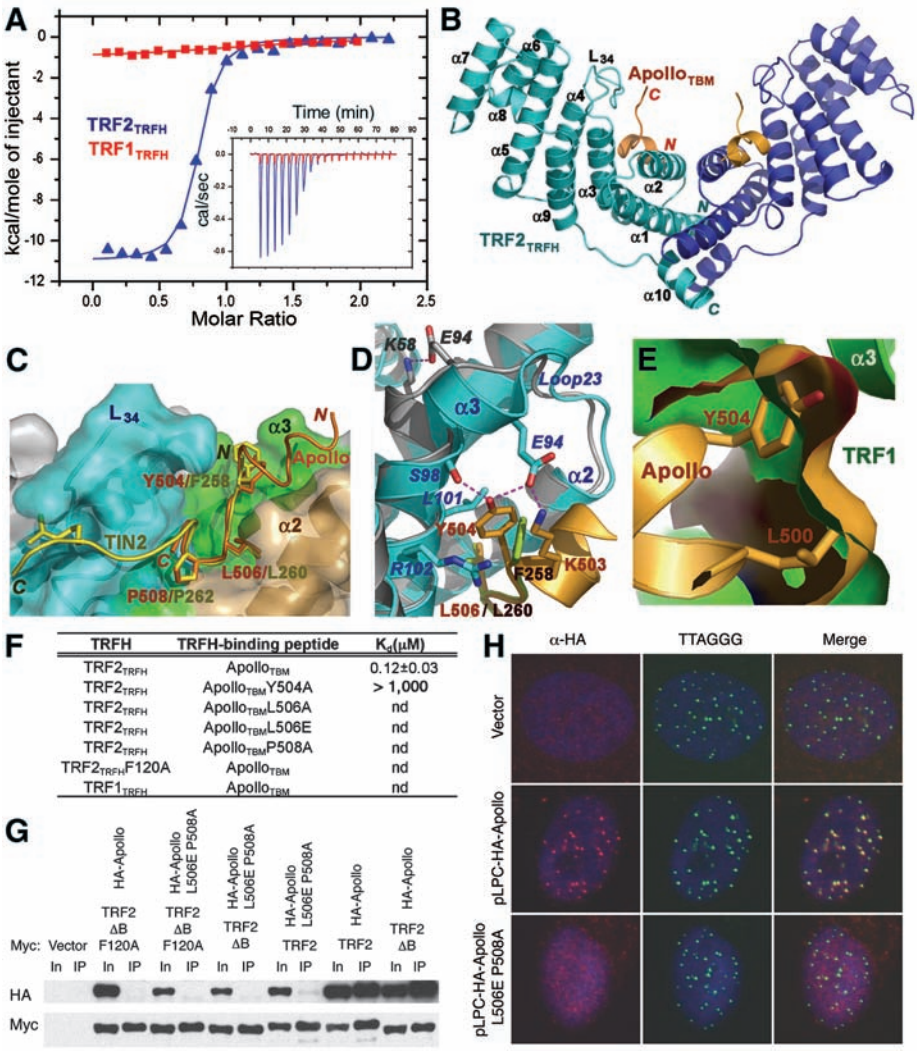


Fig. 4. The TRF2-Apollo interaction. (**A**) ITC measurement of the interactions of $TRF1_{TRFH}$ (red) and $TRF2_{TRFH}$ (blue) with the $Apollo_{TBM}$ peptide. (**B**) Overall structure of the dimeric $TRF2_{TRFH}$-$Apollo_{TBM}$ complex. (**C**) Superposition of $Apollo_{TBM}$ (orange) and $TIN2_{TBM}$ (yellow) reveals a shared F/Y-X-L-X-P motif. (**D**) Superposition of the $TRF2_{TRFH}$-$Apollo_{TBM}$ and the $TRF2_{TRFH}$-$TIN2_{TBM}$ complexes in the vicinity of the Apollo helix. The $TRF2_{TRFH}$ molecules are colored in cyan ($Apollo_{TBM}$-bound) and gray ($TIN2_{TBM}$-bound), respectively. (**E**) $Apollo_{TBM}$ binding is $TRF2_{TRFH}$-specific. The surface representations show that there is no room for Apollo L500 and Y504 to fit into the peptide binding site of $TRF1_{TRFH}$. (**F**) In vitro ITC binding data of wild-type and mutant $TRF2_{TRFH}$-$Apollo_{TBM}$ interactions. (**G**) Co-IP data show that Apollo double-mutant L504E/P506 and TRF2 single-mutant *F120A* disrupt the in vivo TRF2-Apollo interaction. (**H**) Localization of retrovirally expressed HA-tagged wild type and L506E/P508A double mutant of Apollo in BJ-hTERT cells.

mains of TRF1 and TRF2, these structural variations emphasize that the TRFH domain is a versatile framework for interactions with different proteins.

The crystal structure of the $TRF2_{TRFH}$-$Apollo_{TBM}$ complex is corroborated by mutagenesis. Mutations of the conserved hydrophobic residues of Apollo (F504, L506, and P508) or TRF2 (*F120*) completely abolished the interaction both in vitro and in vivo (Fig. 4, F and G). We further assayed the cellular localization of wild-type and mutant Apollo by expressing hemagglutinin (HA)–tagged proteins in human telomerase reverse transcriptase (hTERT)–immortalized human BJ fibroblasts. Although wild-type Apollo showed the expected telomere localization, the L506E/P508A double mutant was distributed throughout the nucleoplasm with no obvious accumulation at telomeres (Fig. 4H). This result confirms the structural information and indicates that the binding of Apollo to the TRFH domain of TRF2 is required for the telomeric localization of Apollo.

We next asked whether other shelterin-associated proteins might contain the F/Y-X-L-X-P motif suggestive of an interaction with the TRFH domain of TRF1 or TRF2. We identified this motif in PinX1, originally identified as a TRF1-interacting protein in a yeast two-hybrid screen (*6*). An 11-residue fragment of PinX1 (R287-D-F-T-L-K-P-K-K-R-R297), referred to as $PinX1_{TBM}$, closely resembles $TIN2_{TBM}$ (fig. S12A), suggesting that it may bind to $TRF1_{TRFH}$ in the same fashion as does $TIN2_{TBM}$. ITC data confirmed the $TRF1_{TRFH}$-$PinX1_{TBM}$ interaction, whereas no measurable interaction was observed between $TRF2_{TRFH}$ and $PinX1_{TBM}$ (fig. S12B). Mutagenesis studies showed that PinX1-L291 and TRF1-*F142* are critical for the interaction, whereas PinX1-P293 is not (fig. S12C). These results are consistent with those of the $TRF1_{TRFH}$-$TIN2_{TBM}$ interaction (Fig. 2D) and indicate that PinX1, like TIN2, binds the TRFH domain of TRF1 but not TRF2. Protein sequence database searches showed many instances of telomere-associated proteins containing the F/Y-X-L-X-P motif (fig. S13). Future studies are needed to address whether this motif mediates the TRF1/TRF2 binding of these telomere-associated proteins in vivo.

Our results indicate that binding to the TRFH docking site involves the sequence F/Y-X-L-X-P in shelterin-associated proteins, which contacts the same molecular recognition surface of the TRFH domains of TRF1 and TRF2 with distinct specificities. Because TRF1 and TRF2 play different roles in telomere length homeostasis and telomere protection (*1*), we propose that the TRFH domains of TRF1 and TRF2 function as telomeric protein docking sites that recruit different shelterin-associated factors with distinct functions to the chromosome ends.

References and Notes

1. T. de Lange, *Genes Dev.* **19**, 2100 (2005).
2. M. van Overbeek, T. de Lange, *Curr. Biol.* **16**, 1295 (2006).
3. X. D. Zhu, B. Kuster, M. Mann, J. H. Petrini, T. de Lange, *Nat. Genet.* **25**, 347 (2000).
4. X. D. Zhu *et al.*, *Mol. Cell* **12**, 1489 (2003).
5. S. Smith, I. Giriat, A. Schmitt, T. de Lange, *Science* **282**, 1484 (1998).
6. X. Z. Zhou, K. P. Lu, *Cell* **107**, 347 (2001).
7. T. Nishikawa *et al.*, *Structure* **9**, 1237 (2001).
8. R. Court, L. Chapman, L. Fairall, D. Rhodes, *EMBO Rep.* **6**, 39 (2005).
9. B. Li, S. Oestreich, T. de Lange, *Cell* **101**, 471 (2000).
10. A. Bianchi, S. Smith, L. Chong, P. Elias, T. de Lange, *EMBO J.* **16**, 1785 (1997).
11. L. Fairall, L. Chapman, H. Moss, T. de Lange, D. Rhodes, *Mol. Cell* **8**, 351 (2001).
12. S. H. Kim, P. Kaminker, J. Campisi, *Nat. Genet.* **23**, 405 (1999).
13. C. Lenain *et al.*, *Curr. Biol.* **16**, 1303 (2006).
14. T. H. Lee, K. Perrem, J. W. Harper, K. P. Lu, X. Z. Zhou, *J. Biol. Chem.* **281**, 759 (2006).
15. S. H. Kim *et al.*, *J. Biol. Chem.* **279**, 43799 (2004).
16. J. Z. Ye *et al.*, *J. Biol. Chem.* **279**, 47264 (2004).
17. Materials and methods are available as supporting material on *Science* Online.
18. Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.
19. P. Fotiadou, O. Henegariu, J. B. Sweasy, *Cancer Res.* **64**, 3830 (2004).

20. Coordinates and structure factor amplitudes have been deposited in the Protein Data Bank with access numbers 3BQO ($TRF1_{TRFH}$-$TIN2_{TBM}$), 3BU8 ($TRF2_{TRFH}$-$TIN2_{TBM}$), and 3BUA ($TRF2_{TRFH}$-$Apollo_{TBM}$). We thank F. Wang and K. Wan for assistance. Work was supported by an NIH grant (to T.de L.) and an American Cancer Society Research Scholar grant and a Sidney Kimmel Scholar award (to M.L.). Use of Life Sciences Collaborative Access Team Sector 21 was supported by the Michigan Economic Development Corporation and the Michigan Technology Tri-Corridor (grant 085P1000817). Use of the Advanced Photon Source was supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, under contract no. DE-AC02-06CH11357.

Supporting Online Material
www.sciencemag.org/cgi/content/full/1151804/DC1
Materials and Methods
SOM Text
Figs. S1 to S14
Table S1
References

16 October 2007; accepted 7 January 2008
Published online 17 January 2008;
10.1126/science.1151804
Include this information when citing this paper.

Clonal Integration of a Polyomavirus in Human Merkel Cell Carcinoma

Huichen Feng, Masahiro Shuda, Yuan Chang,* Patrick S. Moore*

Merkel cell carcinoma (MCC) is a rare but aggressive human skin cancer that typically affects elderly and immunosuppressed individuals, a feature suggestive of an infectious origin. We studied MCC samples by digital transcriptome subtraction and detected a fusion transcript between a previously undescribed virus T antigen and a human receptor tyrosine phosphatase. Further investigation led to identification and sequence analysis of the 5387–base-pair genome of a previously unknown polyomavirus that we call Merkel cell polyomavirus (MCV or MCPyV). MCV sequences were detected in 8 of 10 (80%) MCC tumors but only 5 of 59 (8%) control tissues from various body sites and 4 of 25 (16%) control skin tissues. In six of eight MCV-positive MCCs, viral DNA was integrated within the tumor genome in a clonal pattern, suggesting that MCV infection and integration preceded clonal expansion of the tumor cells. Thus, MCV may be a contributing factor in the pathogenesis of MCC.

Polyomaviruses have been suspected as potential etiologic agents in human cancer since the discovery of murine polyoma virus (MuPyV) by Gross in 1953 (*1*). However, although polyomavirus infections can produce tumors in animal models, there is no conclusive evidence that they play a role in human cancers (*2*). These small double-stranded DNA viruses [~5200 base pairs (bp)] encode a variably spliced oncoprotein, the tumor (T) antigen (*3*, *4*), and are divided into three genetically distinct groups: (i) avian polyomaviruses, (ii) mammalian viruses related to MuPyV, and (iii) mammalian polyomaviruses related to simian virus 40 (SV40) (*5*). All four known human polyomaviruses [BK virus (BKV), JCV, KIV, and WUV (*6*, *7*)] belong to the SV40 subgroup. In animals, integration of polyomavirus DNA into the host genome often precedes tumor formation (*8*).

Merkel cell carcinoma (MCC) is a neuroectodermal tumor arising from mechanoreceptor Merkel cells (Fig. 1A). MCC is rare, but its incidence has tripled over the past 2 decades in the United States to 1500 cases per year (*9*). It is one of the most aggressive forms of skin cancer; about 50% of advanced MCC patients

Molecular Virology Program, University of Pittsburgh Cancer Institute, University of Pittsburgh, 5117 Centre Avenue, Suite 1.8, Pittsburgh, PA 15213, USA.

*These authors contributed equally to this work. To whom correspondence should be addressed. E-mail: yc70@pitt.edu (Y.C.); psm9@pitt.edu (P.S.M.)