Title
A Unique DNA Binding Domain Converts T-Cell Factors into Strong Wnt Effectors

Journal
Molecular and Cellular Biology, 27(23)

ISSN
0270-7306

Authors
Atcha, Fawzia A
Syed, Adeela
Wu, Beibei
et al.

Publication Date
2007-12-01

DOI
10.1128/mcb.02132-06

MOLECULAR AND CELLULAR BIOLOGY, Dec. 2007, p. 8352–8363 Vol. 27, No. 23
0270-7306/07/$08.00+0 doi:10.1128/MCB.02132-06

A Unique DNA Binding Domain Converts T-Cell Factors into Strong Wnt Effectors▽

Fawzia A. Atcha,[1] Adeela Syed,[2] Beibei Wu,[1] Nate P. Hoverter,[1] Noriko N. Yokoyama,[1] Ju-Hui T. Ting,[1] Jesus E. Munguia,[1] Harry J. Mangalam,[3] J. Lawrence Marsh,[2]* and Marian L. Waterman[1]*

Department of Microbiology and Molecular Genetics, University of California, Irvine, Irvine, California 92697[1]; Developmental Biology Center and Department of Developmental and Cell Biology, University of California, Irvine, Irvine, California 92697[2]; and Network and Academic Computing, University of California, Irvine, Irvine, California 92697[3]

Received 14 November 2006/Returned for modification 16 January 2007/Accepted 5 September 2007

Wnt regulation of gene expression requires binding of LEF/T-cell factor (LEF/TCF) transcription factors to Wnt response elements (WREs) and recruitment of the activator β-catenin. There are significant differences in the abilities of LEF/TCF family members to regulate Wnt target genes. For example, alternatively spliced isoforms of TCF-1 and TCF-4 with a C-terminal "E" tail are uniquely potent in their activation of *LEF1* and *CDX1*. Here we report that the mechanism responsible for this unique activity is an auxiliary 30-amino-acid DNA interaction motif referred to here as the "cysteine clamp" (or C-clamp). The C-clamp contains invariant cysteine, aromatic, and basic residues, and surface plasmon resonance (SPR) studies with recombinant C-clamp protein showed that it binds double-stranded DNA but not single-stranded DNA or RNA (equilibrium dissociation constant = 16 nM). CASTing (Cyclic Amplification and Selection of Targets) experiments were used to test whether this motif influences WRE recognition. Full-length LEF-1, TCF-1E, and TCF-1E with a mutated C-clamp all bind nearly identical WREs (TYYCTTTGATSTT), showing that the C-clamp does not alter WRE specificity. However, a GC element downstream of the WRE (RCCG) is enriched in wild-type TCF-1E binding sites but not in mutant TCF-1E binding sites. We conclude that the C-clamp is a sequence-specific DNA binding motif. C-clamp mutations destroy the ability of β-catenin to regulate the *LEF1* promoter, and they severely impair the ability of TCF-1 to regulate growth in colon cancer cells. Thus, E-tail isoforms of TCFs utilize two DNA binding activities to access a subset of Wnt targets important for cell growth.

LEF/T-cell factor (LEF/TCF) proteins are broadly conserved transcription factors necessary for Wnt signaling and control of cell growth. LEF/TCF orthologues in *Caenorhabditis elegans*, *Drosophila melanogaster*, *Hydra magnipapillata*, and *Ciona intestinalis* contain an obligate C terminus called the E-tail, suggesting that this form is the ancestral precursor for mammalian LEF/TCFs (Fig. 1A). Indeed, this tail plays an essential albeit undefined role in Wingless signaling in *Drosophila* flies. A single nucleotide mutation in a highly conserved 30-amino-acid motif within the E-tail behaves genetically as a hypomorph in flies, with lethality at the first larval instar (41). The mutation highlighted the importance of the C-terminal E-tail, and the conserved motif in this region was referred to thereafter as the CRARF domain (after a highly conserved 5-amino-acid motif in the domain) (23, 43). In mammalian TCFs, the E-tail has become one of several alternatively spliced C-terminal tails (40, 43), while the *LEF1* locus has diverged such that an E-tail cannot be produced (22). Interestingly, much of the amino acid sequence in the TCF E-tails is poorly conserved except for the 30-amino-acid motif (67% identical between human TCF-1 and TCF-4 and dTCF/pangolin) (Fig. 1A), suggesting that this motif fulfills an important role in organisms from flies to humans. This role may be the ability to regulate a greater set of Wnt target genes, because we and others have found that activation of the *LEF1* and *CDX1* promoters specifically requires an E-tail containing a TCF isoform for β-catenin activation (3, 20). Interestingly, we have determined that the predominant forms of TCFs in human colon cancer contain E-tails (A. Syed, L. Arce, R. Najdi, F. Atcha, H. Theisen, J.-H. Ting, R. Edwards, M. L. Waterman, and J. L. Marsh, unpublished data). Since *LEF1* gene transcription is aberrantly activated in ~83% of colon tumors and since *CDX1* expression is also often observed in colon cancer, we hypothesize that E-tail isoforms of TCFs are important mediators of Wnt target gene regulation in these tumors.

Recognition of Wnt response elements (WREs) by LEF/TCFs occurs through a highly conserved, 88-amino-acid high-mobility group domain (HMG) (Fig. 1A). This domain consists of a 68-amino-acid HMG box and a 9-amino-acid nuclear localization signal (NLS) separated by a 9-residue linker sequence. The NLS participates in DNA binding through nonspecific contacts between positively charged side chains of the highly basic NLS with the phosphate backbone (14, 26). These additional contacts elevate DNA binding affinity 100-fold such that LEF/TCFs bind to WREs with affinities in the nanomolar range. The HMG DNA binding domain has at least two notable features of DNA recognition. The first is that the HMG box recognizes a specific sequence in the minor groove of DNA

* Corresponding author. Mailing address: Department of Microbiology and Molecular Genetics, University of California, Irvine, Irvine, CA 92697. Phone: (949) 824-2885. Fax: (949) 824-8598. E-mail for Marian L. Waterman: mlwaterm@uci.edu. E-mail for J. Lawrence Marsh: jlmarsh@uci.edu.

▽ Published ahead of print on 24 September 2007.

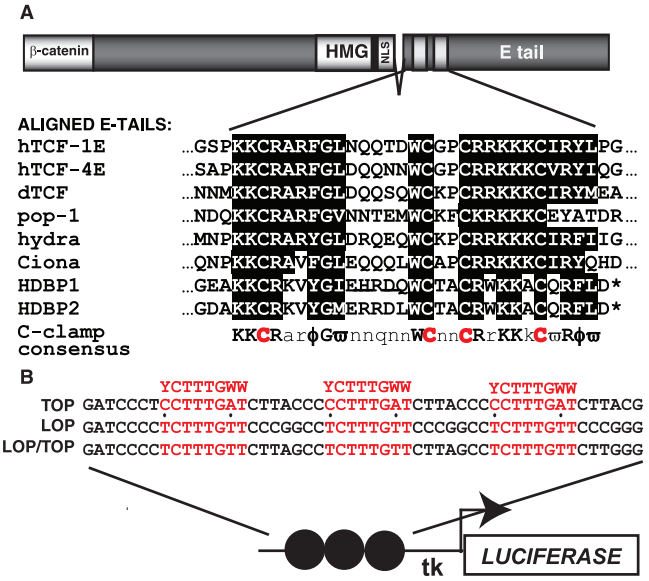

FIG. 1. A 30-amino-acid motif in the TCF E-tail is highly conserved and related to a sequence in two heterologous transcription factors. (A) A general domain structure of LEF/TCFs includes the N-terminal β-catenin binding domain, the high-mobility group DNA binding domain (which also includes the nuclear localization signal [HMG box + NLS]), and the alternatively spliced E-tail. Partial alignment of homologous E-tail sequences in TCF orthologs and the unrelated Huntington's Disease Binding Proteins 1 and 2 shows a high level of sequence conservation of basic, aromatic, and cysteine residues. Regions of sequence conservation are shaded, and a consensus sequence is shown below the alignment, with the four conserved cysteines highlighted in red. hTCF-1E and hTCF-4E are human family members; dTCF, pop-1, hydra, and Ciona represent the single TCF orthologs in *Drosophila*, *C. elegans*, *H. magnipapillata*, and *Ciona intestinalis*, respectively. (B) Three luciferase reporter plasmids used in this study comprise a multimerized Wnt response element (core sequence shown in red) next to the minimal herpesvirus tk promoter.

(core element, YCTTTGWW), and the second is that this domain bends DNA (up to 130°) away from the protein (14, 39). The flexible 9-residue linker sequence between the HMG box and NLS allows the NLS to be placed underneath and inside the crimped major groove for phosphate backbone interaction (26). With the exception of a few studies, most have explored LEF/TCF regulation of gene expression and binding to WREs through interactions with either a consensus WRE or a strong natural binding site. Only a few native sites that differ from the consensus sequence have been studied in detail. Since the HMG DNA binding domain of LEF/TCFs is the most highly conserved (≥94%) feature of these proteins, it has been assumed that all LEF/TCF proteins can bind identically and equivalently to any WRE and activate transcription with β-catenin recruitment. We revisited this issue when it became clear that LEF/TCFs were not equivalent in their abilities to activate a natural Wnt target gene such as *LEF1* or *CDX1* and that the alternative E-tail provided a special activity (3). An early model suggested that the E-tail recruited the transcription coactivator p300 for activation of *CDX1* transcription (20). We therefore tested the hypothesis that the E-tail recruits a cofactor for *LEF1* transcription as well as the hypothesis that the E-tail modifies DNA bending or binding. We found that the 30-amino-acid motif within the E-tail functions as a sequence-specific auxiliary DNA interaction domain and that selected Wnt response elements require it for stable binding and β-catenin activation. We also show, as others have, that overexpression of dominant-negative (dn) TCFs can halt the growth of colon cancer cells, but here we also show that the E-tail is crucial for this activity. We conclude that E-tail isoforms of TCFs are the most potent forms of TCFs in that they have two side-by-side DNA binding domains and are thus able to regulate specific Wnt target genes with lower-affinity WREs. This more potent target activation is important for control of colon cancer cell growth.

MATERIALS AND METHODS

Cell culture. COS-7 cells were cultured in high-glucose Dulbecco modified Eagle medium containing 10% fetal bovine serum (Omega Scientific) and 2 mM L-glutamine. D7p11 cells (a stable DLD-1 cell line; a generous gift from Hans Clevers and Marc van de Wetering) were cultured in RPMI media with 5% fetal bovine serum and 2 mM L-glutamine as described above but also with supplementation of 500 μg/ml Zeocin and 10 μg/ml blasticidin (Invitrogen).

Transient transfection assay. COS-7 cells (2.5×10^5) were transfected using 0.5 μg of reporter plasmids with different, multimerized WREs in the TOPtk backbone (*LEF1* promoter 1 [−672, +314; Lop, Top, Lop/Top [Fig. 1B]), 0.1 μg of cytomegalovirus β-galactosidase plasmid, and FuGENE 6 transfection reagent according to the protocol of the manufacturer (Roche Molecular Biochemicals). Reporter plasmids were cotransfected with expression vectors for β-catenin (pCS2-β-catenin; gift of B. Gumbiner) (0.4 μg), LEF/TCFs (0.2 μg), and other expression constructs indicated in the figure legends. Cells were harvested 18 to 20 h posttransfection and assayed for luciferase activity, and activities were normalized using β-galactosidase activity values. In all cases experiments were performed with duplicate samples, and at least three or more independent experiments were performed for calculation of standard deviation values for error bars.

Western blot analysis. Cell lysates of 1×10^5 DLD1 cells treated with doxycycline (D7p11 cells; 6×10^{-4} μg/ml for dn wild-type TCF-1E [dnTCF-$1E^{WT}$] induction and 1 μg/ml for mutant dnTCF-1E [dnTCF-$1E^{mt}$] induction) were analyzed for induced expression of dnTCF-1 by Western blotting with a polyclonal LEF-1 antibody that detects all LEF/TCF isoforms (21) and secondary anti-rabbit immunoglobulin G conjugated to horseradish peroxidase. Blots were developed with ECL reagent (Amersham).

Sulforhodamine B cell growth assay. A 6-day growth curve experiment was performed with DLD-1 cells, which inducibly express dnTCF-$1E^{WT}$ or dnTCF-$1E^{mt}$. Cells were plated on 96-well plates as eight replicates for each experimental time point. Every 2 days cells were provided with fresh media containing either doxycycline (concentrations indicated above for the Western blot protocol) or water (mock treatment control). Cells were fixed and stained according to published protocols (36). Optical density readings were performed at 490 nm with Spectra Max 340 from the Molecular Devices Corporation. Growth curves were carried out for 6 days, and the standard deviation generated for eight replicates is shown.

EMSA. Electrophoretic mobility shift assay (EMSA) reactions were carried out with 1 to 2 pmol (approximately 200 cps) of oligonucleotide (see sequence listed in Fig. 5) in a final reaction volume of 15 μl containing 50 mM HEPES (pH 8.0), 12.5 mM EDTA, 50% glycerol, 100 mM KCl, 2 mM $MgCl_2$, 1 μg of poly(dI-dC), and 10 mM dithiothreitol. Extracts from COS-1 cells transiently transfected with expression vectors for full-length human TCF-$1E^{WT}$ and TCF-$1E^{mt}$ (CR1 mutant) were used for the assays. Lysates from COS-1 cells were prepared 48 h after transfection by swelling cells on ice, immersing for 15 min in hypotonic lysis buffer (10 mM Tris [pH 7.9], 50 mM KCl, 10 mM Mg_2Cl, 0.01 mM EDTA, 1 mM dithiothreitol, 0.01 mM EGTA, 1 mM phenylmethylsulfonyl fluoride, protease inhibitor cocktail), and douncing. A 10% volume of glycerol was added to the cell lysate for storage at −80°C. Lysates from Colo320 colon cancer cells, which express TCF-1E and TCF-4E, were prepared with the same protocol.

Biacore analysis. E-tail coding sequences of human TCF-1E (amino acids [aa] 436 to 561, excluding 35 residues from the extreme C terminus for stability in bacteria) were subcloned into pGex3X to generate pGEX/E-$tail^{WT}$ or pGEX/E-$tail^{mt}$. Rosetta cells were transformed with the vectors, and the fusion protein was purified by glutathione agarose column chromatography as previously described (32). All SPR studies were performed with a Biacore 3000 instrument and a carboxymethyl (CM5)-coated sensor chip (Biacore AB, Uppsala, Sweden). The sensor chips were conditioned with three consecutive 1 min injections of 1

M NaCl in 50 mM NaOH followed by extensive washing with buffer. Glutathione *S*-transferase (GST) E-tail protein was directly immobilized onto a CM5 chip by use of *N*-ethyl-*N*-(3-diethylaminopropyl)carbodimide and *N*-hydroxysuccinimide chemistry. Synthetic oligonucleotides were annealed. Oligonucleotide sequences were as follows: for TOP, 5′-GATCTAGGGCACCCTTTGAAGCTCT-3′ (sense) and 5′-AGAGCTTCAAAGGGTGCCCTA-3′ (antisense); for FOP (data not shown), 5′-GATCTAGGGCACCATCTGCGGCTCT-3′ (sense) and 5′-AGAGCCGCAGATGGTGCCCTA-3′ (antisense); for LOP, 5′-GAGCCCG GGAACAAAGAGGGGTC-3′ (sense) and 5′-GACCCCTCTTTGTTCCCG G-3′ (antisense); and for MOP, 5′-CAGCCCGGGCGCAGATAGGGGTC-3′ (sense) and 5′-GACCCCTATCTGCGCCCGG-3′ (antisense).

Duplex DNAs (10 nM to 100 nM) were flowed over the immobilized GST E-tail protein. All the Biacore data were collected at 25°C with running buffer at a constant flow of 40 μl/min. Apparent association and dissociation rate constants were calculated using BiaEvaluation software version 3.0 supplied by the vendor (Biacore Inc, Uppsala, Sweden). Values were derived from three independent experiments using two different recombinant protein preparations. For the GST/E-tail^{WT} versus GST/E-tail^{mt} binding and competition assays, biotinylated LOP was immobilized on the streptavidin chip. The LOP mt sense strand (MOP) was used for competition with single-stranded DNA, and the following oligonucleotides were used for double-stranded DNA competition: sense, 5′-A GATCTACGGAGGACTGTCCTCCGT-3′; and antisense, 5′-ACGGAGGAC AGTCCTCCGTAGATCT-3′.

CASTing analysis. CASTing (Cyclic Amplification and Selection of Targets) was performed as described previously with some modifications (details available upon request) (28, 45). A 70-mer GCGTCGACTCTAGACTGCAG-N_{30}-GAA TTCGGATCCCTCGAGCG sequence was synthesized to generate the target oligonucleotide library. A 25-μg volume of the oligonucleotide library was converted to double-stranded DNA, and 5 μg was incubated with 15 μg of COS cell extract (see above [EMSA paragraph]). Binding conditions for the first CASTing experiment were as follows: 15 mM Tris (pH 7.5), 50 mM KCl, 5% glycerol, 5 μM EGTA, 5 μM EDTA, and 5.75 mM $MgCl_2$. Binding conditions for the second CASTing experiment were as follows: 35 mM Tris (pH 7.5), 25 mM KCl, 60 mM NaCl, 0.25% NP-40, 2.5 mM beta-mercaptoethanol, and 0.75 mM $MgCl_2$. The second CASTing experiment used COS-1 cell extracts of overexpressed TCFs in which the cells had been lysed in 50 mM Tris-HCl (pH 7.5)–120 mM NaCl–0.5% NP-40–2 mM Pefabloc SC (AEBSF; Roche) protease inhibitor–5 mM beta-mercaptoethanol–8 mM $MgCl_2$. Overexpressed LEF/TCFs were epitope tagged with either TAP tags (LEF-1) or histidine tags (TCF1-E^{WT} and TCF1-E^{mt}). After a 20-min incubation at room temperature, 2 μl of Talon Dynabeads (Invitrogen) or streptavidin beads (Invitrogen) was added and incubated for 10 min at room temperature. Bead-DNA-protein complexes were washed with 500 μl of phosphate-buffered saline containing 0.1% NP-40 and 0.1% bovine serum albumin and resuspended in 100 μl of PCR mixture. The enriched oligonucleotide pool was amplified by PCR using primers complementary to the fixed flanking sequence (5′-GCGTCGACTCTAGACTGCAG-3′ and 5′-CGCTCGAGGGATCCGAATTC-3′). Sample aliquots of the PCR were removed after 6, 9, 12, and 15 cycles and analyzed on agarose gels. A 10-μl volume from the aliquot which contained the first visible PCR products was used in the next round of CASTing. From the second round on, cytosolic extract was also preincubated with 1 μg poly(dI-dC) as a nonspecific competitor. After six rounds of CASTing, the selected DNA was cloned into a TOPO vector (Invitrogen) and individual colonies were selected for sequencing. Typically, 70 to 96 colonies were grown in LB in 96-well plates, sealed, and sent to Agencourt for Miniprep treatment and sequencing.

RESULTS

The E-tail is necessary for a weak WRE. To understand the mechanism underlying the E-tail requirement for activation of the *LEF1* promoter, we investigated its role in regulation via transient transfection assays. We previously reported that a small, highly conserved 30-amino-acid motif (CRARF) within the E-tail is crucial for *LEF1* promoter activation (Fig. 1A) (3). Amino acid substitution of the CRARF sequence destroys the ability of β-catenin to activate the *LEF1* and *CDX1* promoters (3, 20). In contrast, the synthetic reporter construct TOPtk, which contains three multimerized optimal Wnt response elements (Fig. 1B) (41), can be activated by all LEF/TCF isoforms, including those that do not have E-tails and those with the CRARF amino acid sequence destroyed (3, 20). A major difference between the synthetic TOP reporter plasmids and the *LEF1* reporter is the artificial multimerization of three consensus TOP WREs and their juxtaposition to highly active core promoters (c-*fos* or herpes virus tk) versus the *LEF1* promoter WREs in their natural position downstream of the transcription start site (21). To test whether the multimerization of WREs masks differences among LEF/TCF isoforms and family members, we generated a matching synthetic reporter construct by substitution of the three TOP sites in TOPtk with an identically multimerized triplet of one of the E-tail-requiring WREs from the *LEF1* promoter (LOP; Fig. 1B). The only sequence difference between these 6-kb reporter plasmids is two nucleotides in each of the three multimerized WREs. Cotransfection of β-catenin and LEF/TCFs with the TOPtk reporter construct resulted in similar levels of activation by all forms of LEF/TCFs (TCF-1E, LEF-1, and a LEF-E chimeric fusion protein), confirming that the E-tail is not required for activation of the consensus WRE (Fig. 2A). In contrast, activation of the nearly identical LOPtk required the E-tail, because the reporter was activated only by TCF-1E and the LEF-E chimera (16- and 12-fold, respectively) (Fig. 2B). Another synthetic reporter construct that was generated by converting the GC-rich flanking sequences of LOP to match TOP (LOP/TOP) continued to require an E-tail for activation, albeit the LEF-1E chimera was only weakly active with this reporter (TCF-1E, LEF-1, and LEF-1E exhibited 13-, 2-, and 5-fold activation, respectively) (Fig. 2C). These data suggest that the need for an E-tail is influenced by sequence variation in the core Wnt response element and flanking sequences. Even when fused to the potent activator VP16, a chimeric LEF-VP16 fusion protein that can robustly activate TOP, LEF-1 is unable to activate LOP (286-fold activation of TOP versus 2-fold activation of LOP at maximum levels of expression) (Fig. 2D). These results definitively show that in vivo, LEF-1 is unable to bind well enough or long enough to the LOP sites to allow the potent VP16 domain to activate transcription. The stark contrast in regulation between the two reporter plasmids is not due to reporter plasmid backbone or promoter elements, as these are all identical, but instead correlates with only a few nucleotide differences within the multimerized LOP and TOP sites (Fig. 1B).

The E-tail is not an independent regulatory domain. One possibility for E-tail action is that it enables recruitment or stable association of TCF-1E to a variant WRE through interactions with a DNA-bound factor or component of the basal transcription machinery. We asked whether the E-tail could mediate this recruitment in the absence of a functioning HMG DNA binding domain. Two mutant constructs were generated for these experiments. The first mutant contains a two-amino-acid substitution in the HMG DNA binding domain that abrogates the ability of the HMG box to bind to DNA (Fig. 3A; KK/EE) (13). The second mutant encodes a chimeric protein composed of the β-catenin binding domain of TCF-1, the Gal4 DNA binding domain from yeast (*Saccharomyces cerevisiae*), and the E-tail (Fig. 3A) (β-cat/Gal4/E-tail). Cotransfection of either of these mutant proteins with a β-catenin expression vector was unable to activate any of the WRE reporters, including the *LEF1* promoter. Thus, activation by TCF-1E re-

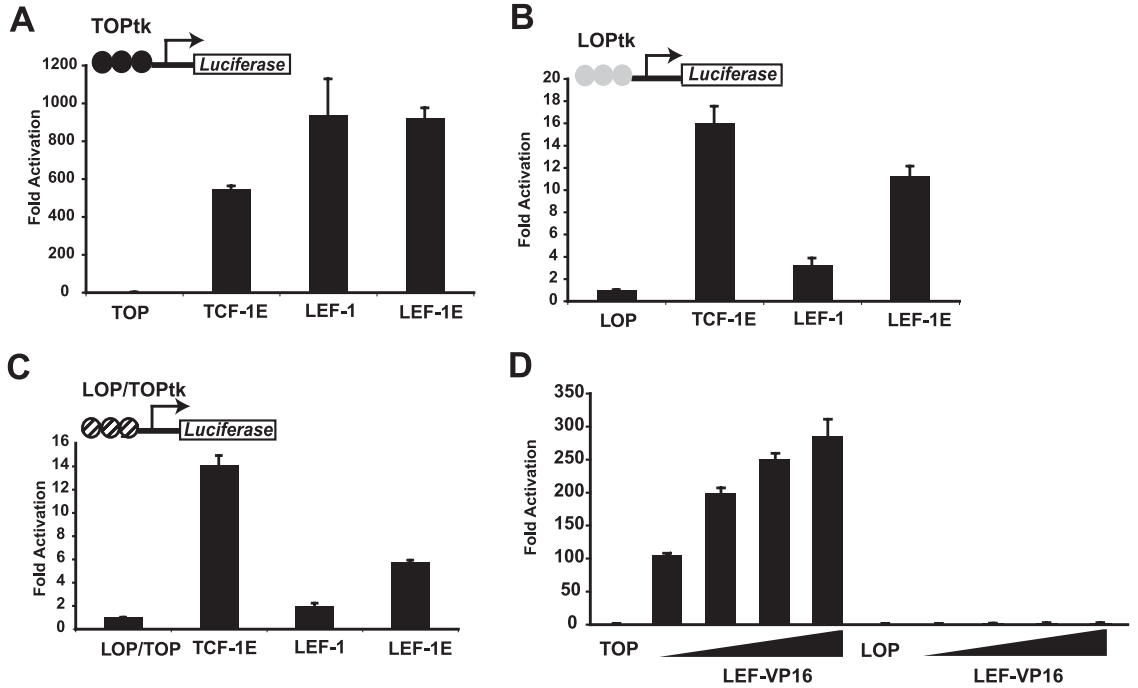

FIG. 2. The E-tail is necessary to activate transcription from weak Wnt response elements. Cos-1 cells were transiently transfected with TOP (A), LOP (B), or LOP/TOP (C) luciferase reporters with expression vectors for the indicated LEF/TCFs and β-catenin (cotransfected in all conditions). (D) TOP and LOP were cotransfected with increasing concentrations (200, 400, 800, and 1,000 ng) of LEF-VP16 expression vector. At the highest concentration of LEF-VP16, activation of LOP was increased twofold. Error bars represent standard deviations derived from the results of three or more experiments.

tains a strict requirement for both the HMG DNA binding domain and the E-tail.

An independent interaction between the E-tail of human TCF-4E and the transcription coactivator p300 has been previously reported (20). The CRARF domain of the TCF-4 E-tail was shown to be involved in this interaction and subsequent activation of the *CDX1* promoter. We explored the possibility that the E-tail recruits a coactivator independently and/or is itself recruited by another protein to the *LEF1* promoter. To address the former possibility, we coexpressed a Gal4/E-tail fusion protein with full-length TCF-1E and β-catenin in transient transfections (Fig. 3B). Gal4/E-tail cannot bind to the *LEF1* reporter, because there are no Gal4 sites, but it could potentially inhibit reporter activation if it competes for an E-tail interacting protein (squelching). We observed that excess Gal4/E-tail was not able to reduce *LEF1* reporter activation. While lack of squelching is not definitive evidence, it is suggestive that the E-tail is not independently engaged with proteins bound to the *LEF1* promoter.

The E-tail contains a novel cysteine motif. The five-amino-acid CRARF substitution that destroys transcription activation removes basic residues and the first of four conserved cysteine residues (Fig. 3C). The four cysteines are 100% conserved in their sequences and spacing between all TCF orthologs (except for TCF-3), suggesting that these residues are important for functioning. The spacing between cysteines does not conform to known metal chelating finger motifs but could be important for a novel structure or finger. To determine whether individual cysteines are essential for activity and whether histidine residues could perform substitutions, the second and third conserved cysteines were individually changed to either alanine or histidine within the context of full-length TCF-1E and tested for activity in transient transfection assays. Single amino acid substitution of the cysteines completely abrogates the ability of these mutants to activate the *LEF1* promoter but not TOPtk (Fig. 3C). Histidine substitution also completely inactivated the proteins, suggesting that replacing cysteines with alternative residues that can chelate metals is either not sufficient or not compatible with a functional structure. These results suggest that the cysteines play a fundamental role in TCF-1E action by establishing a tertiary structure optimal for reporter gene activation.

The E-tail is a novel DNA interaction domain. An alternative possibility for a mechanism for E-tail action is that it augments binding of the HMG DNA binding domain to nucleic acids. Support for this possibility was provided by a recent study which identified Huntington's Disease Binding Protein 1 (HDBP1) and HDBP2, two highly related transcription factors that bind to a unique GC motif (GCCGGCG) in the *HD* promoter (37). The DNA binding domain was delimited to a region that has remarkable amino acid sequence similarity to the 30-amino-acid cysteine motif in the TCF-1 and TCF-4 E-tails (see alignment in Fig. 1A). Interestingly, GC-rich sequences are near each of the two WREs in the *LEF1* promoter (CCCGGG shown in Fig. 1B and GCCGGCG seven nucleotides downstream of the second WRE element in the *LEF1* promoter; see Fig. 5). Thus, a novel type of DNA binding domain may lie very close to the well-described HMG DNA binding domain and may have sequence-specific recognition properties.

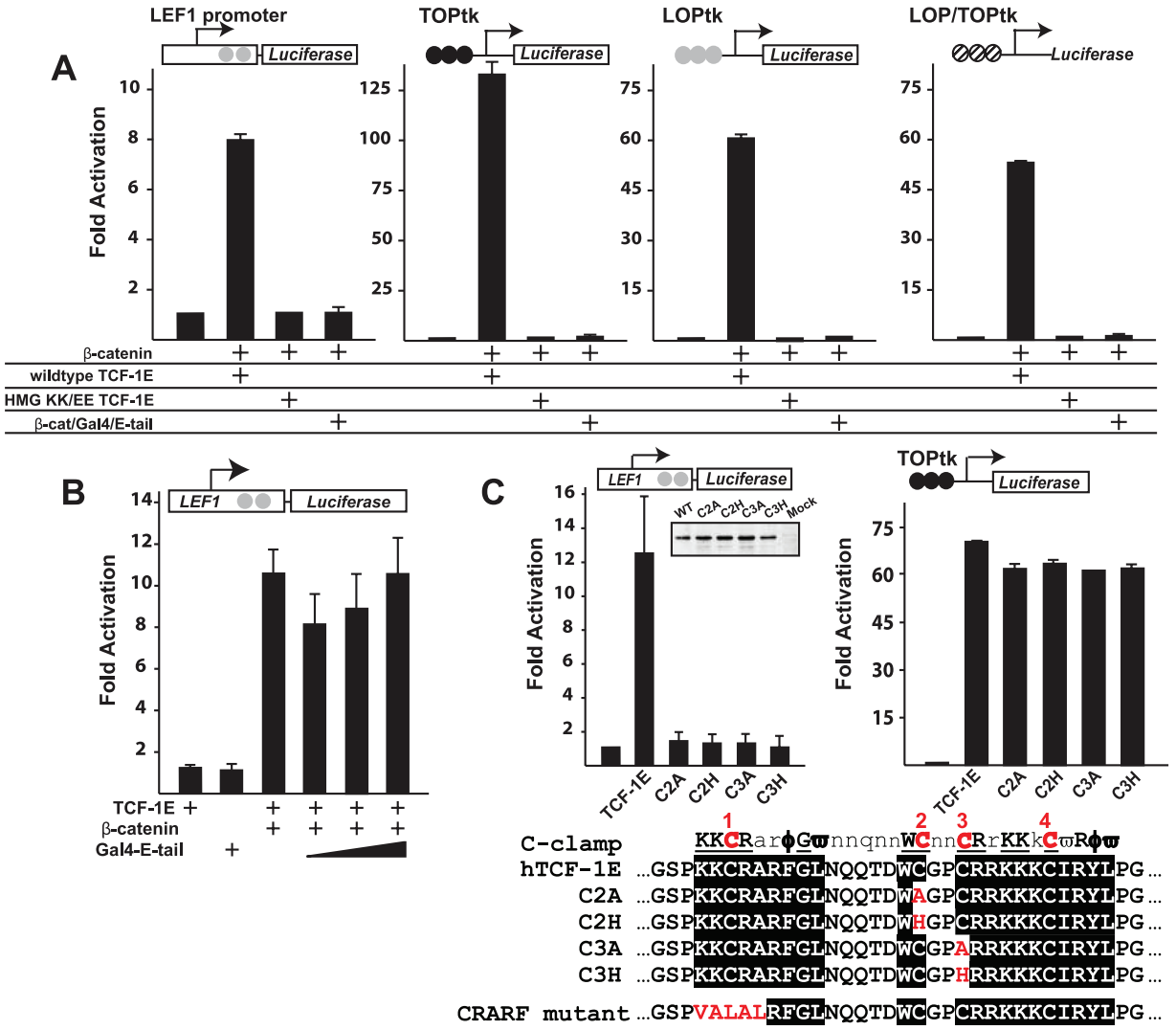

FIG. 3. A cysteine motif in the E-tail cooperates with the HMG DNA binding domain for *LEF1* regulation. (A) Activation of Wnt responsive reporter plasmids *LEF1*, TOP, LOP, and LOP/TOP requires the HMG DNA binding domain. A two-amino-acid substitution in the HMG DNA binding domain (KK changed to EE) prevents reporter gene activation even though the E-tail remains the wild type. The E-tail alone is not sufficient to recruit β-catenin to any of the reporter plasmids (β-cat/Gal4/E-tail denotes a fusion protein in which the Gal4 DNA binding domain is fused at the N terminus to the β-catenin binding domain and at the C terminus to the E-tail). (B) Transient transfection assays showed that coexpression of increasing amounts of a Gal4/E-tail fusion protein (Gal4 DNA binding domain fused to the 138-amino-acid E-tail of TCF-1E) does not inhibit the ability of TCF-1E and β-catenin to activate transcription from the Wnt response elements in the *LEF1* promoter 1. (C) A novel cysteine motif is required for *LEF1* regulation but not TOP reporter regulation. Cos cells were cotransfected with β-catenin and mutant TCF-1E expression vectors and the *LEF1* reporter plasmid (left panel) or the TOPtk reporter plasmid (right panel). Mutations are single amino acid substitutions in the positions indicated in red type in an E-tail alignment (shown below). Error bars indicate standard deviations for the results of three experiments. Mock, mock treatment.

SPR measurements can detect real-time interactions between proteins and nucleic acids, and we used this method to assess E-tail interactions with nucleic acids. We fused the coding sequences of GST to the coding sequences for 125 residues of the E-tail and produced this recombinant fusion protein in bacteria. We modified the surface of SPR chips by covalent conjugation of purified GST or GST/E-tail protein and introduced increasing concentrations of double-stranded oligonucleotides with different sequences, including LOP (which includes the CCCGGG element downstream), the consensus WRE (TOP), and the mutant version of the LOP site (MOP). Rapid binding to all three oligonucleotides was detected with GST/E-tail but not GST (Fig. 4A and data not shown). A quantitative analysis of the binding shows nearly identical kinetics and affinity for binding to all three oligonucleotides, yielding an overall equilibrium dissociation constant (K_D) = 1.55 (± 0.33) × 10^{-8} M (Fig. 4A). The overall K_D places binding affinities in the nanomolar range, but the association and dissociation rates are both rapid. The fast on-off rates may explain why our DNase I footprinting reactions with TCF/E-tail isoforms and GST/E-tail protein do not detect protection of the flanking GC sequences in the *LEF1* promoter (reference 3 and data not shown). Similar binding kinetics results for any double-stranded DNA sequence demonstrate that the E-tail,

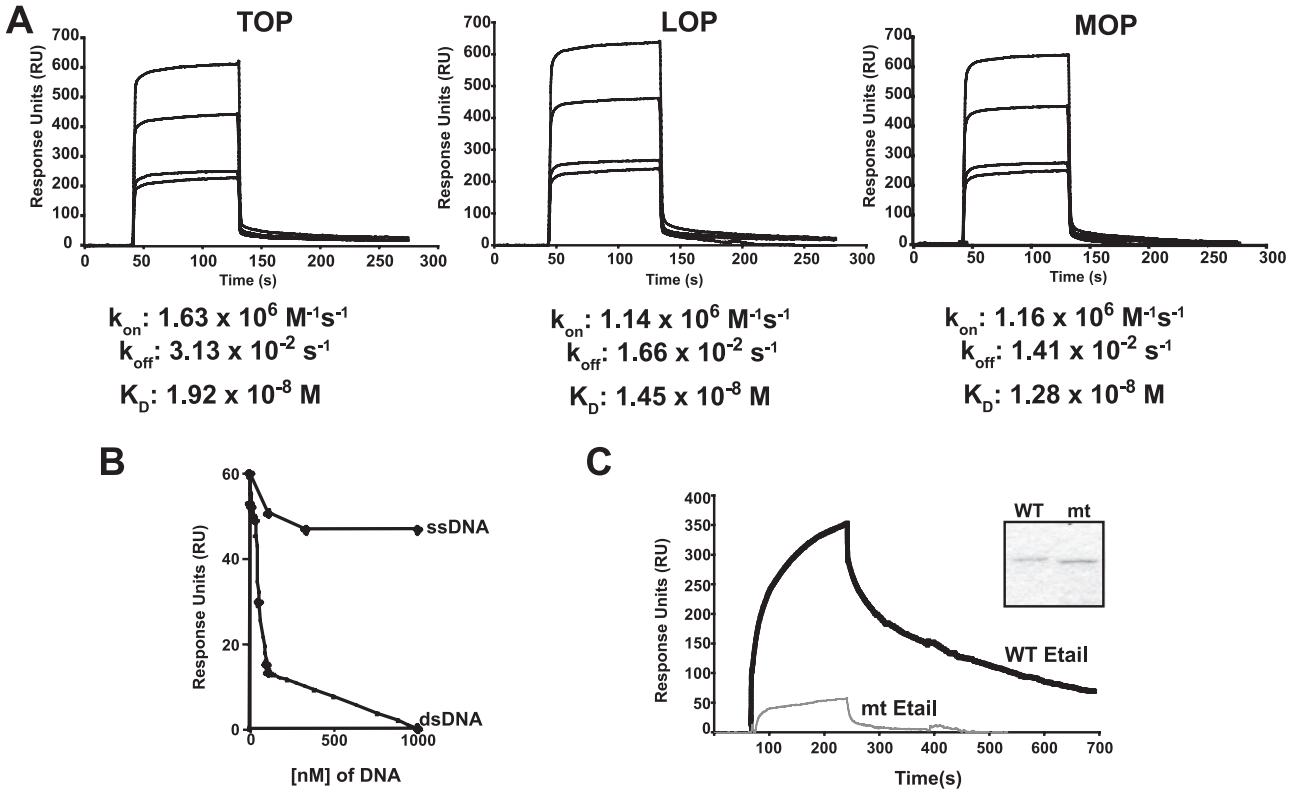

FIG. 4. The E-tail is a new type of DNA binding domain. (A) Overlay sensorgrams for SPR analysis of TOP, LOP, or mutant LOP sequence (MOP) binding to immobilized GST E-tail. DNAs were injected at concentrations ranging from 10 nM to 100 nM over immobilized GST/E-tail at a flow rate of 40 μl/min at 25°C. Results are expressed in resonance units (RU) as a function of time in seconds. Increases in RU indicate binding of the double-stranded oligonucleotides to the GST/E-tail-modified surface. (B) Binding competition with a concentration of up to 1,000 nM of either single-stranded DNA or double-stranded DNA was performed with GST/E-tail bound to 100 RUs of immobilized DNA. (C) Equivalent concentrations of purified, recombinant GST/E-tail wild-type protein and GST/E-tail^{mt} mutant protein were compared with respect to their binding activities with 1,800 RUs of immobilized LOP DNA. The E-tail mutation carries the CRARF substitution shown in Fig. 3C. Sensorgrams are representative of specific interactions.

at least as a separate recombinant protein fragment, has no sequence specificity in its interaction. Interestingly, we observed that for all three profiles, binding is completely eliminated with excess heterologous double-stranded DNA and is only minimally affected by excess single-stranded DNA or tRNA (Fig. 4B and data not shown). These data suggest that the E-tail preferentially binds to specific structural features of double-stranded DNA and not to other nucleic acids. To more rigorously test the idea that residues in the 30-amino-acid motif are important for DNA interactions, the five-amino-acid CRARF substitution that destroys E-tail-dependent activation of transcription (3) (CRARF mutant; Fig. 3C) was introduced into the GST/E-tail protein and tested using SPR. The binding profile presented in Fig. 4C shows that this five-amino-acid substitution destroys DNA binding. We conclude from these experiments that the cysteine-rich motif in the E-tail is a DNA binding motif.

Sequence-specific binding of TCF-1E. We examined the actions of the E-tail in EMSAs, analyzing it both as an independent protein fragment and within the context of full-length wild-type TCF-1E^{WT}. Full-length TCF-1E^{mt} (CRARF mutation) was also used in the assays so that the contribution of this novel DNA binding domain could be assessed in its natural location near the HMG DNA binding domain. We assessed binding to the second Wnt response element in the *LEF1* promoter because, like the first element (LOP), it requires E-tail isoforms for activation and, more importantly, because a perfect match to the HDBP1/HDBP2 GC element occurs seven nucleotides downstream of the core WRE (Fig. 5 [GC CGGCG]). We refer to this extended nucleotide sequence as LOP2+GC. We tested whether the CRARF mutation in the cysteine motif would affect binding of TCF-1E to the LOP2+GC sequence. The results in Fig. 5A show that compared to wild-type TCF-1E, TCF-1E^{mt} is significantly compromised for binding, and therefore, in the context of full-length TCF-1E, the E-tail augments DNA binding. The binding activities of recombinant GST/E-tail with this same probe yielded results similar to those obtained in the SPR studies; the E-tail bound weakly but independently to LOP2+GC, and the CRARF mutation destroyed binding (Fig. 5B [GST/E-tail^{mt}]).

The remarkable amino acid sequence similarities between the cysteine motif in the E-tail and the DNA binding domain of HDBP1/HDBP2 suggested that the E-tail could specifically recognize the GC element present in LOP2+GC. We designed a competition EMSA experiment with GST/E-tail to test this possibility. We observed that the results with respect to competition for GST/E-tail binding were similar at 10-fold and 50-fold molar excess no matter what cold competitor was used,

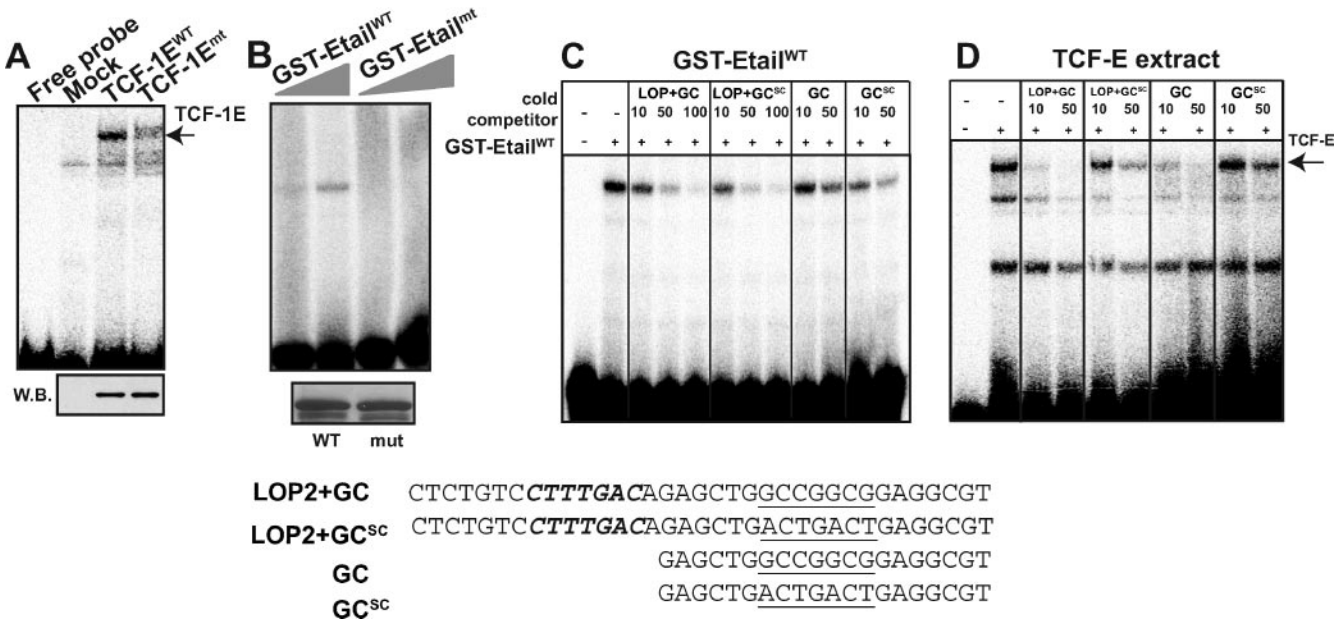

FIG. 5. The E-tail is a sequence-specific DNA binding motif in the context of full-length TCF-1. (A) EMSA with extracts from COS-1 cells expressing full-length histidine-tagged TCF-1E TCF-1E^{WT} or TCF1E^{mt} (10 μg protein) showed that the CR1 mutation in the E-tail (see Fig. 3C for sequence of mutation) decreases TCF-1E binding to an extended probe that encodes the second Wnt response element (bold italics in sequences at bottom of figure) in the *LEF1* promoter and a GC-rich motif (underlined) seven nucleotides downstream of the core (LOP2+GC). A Western blot probed with antibody for the histidine tag shows equal levels of wild-type and mutant proteins in the extracts. (B) Purified recombinant GST/E-tail (GST/E-tail^{WT}) can bind directly to the LOP2 + GC probe, but a mutant GST/E-tail with the CR1 mutation is inactive for DNA binding (GST/E-tail^{mt}). A Coomassie-stained gel shows amounts of purified recombinant proteins equal to those used in the EMSA. (C) In a competition assay for binding to LOP2+GC, the indicated molar excess of cold competitor oligonucleotides shows that the downstream GC-rich sequence is not specifically recognized by recombinant purified GST/E-tail (GC^{SC} refers to a "scrambled" mutation of this GC motif). These data are consistent with the SPR experiments (Fig. 4), which show the E-tail binds double-stranded DNA in a non-sequence-specific manner. (D) Extracts of Colo320 colon cancer cells (15 μg protein), which express high levels of TCF-1E and TCF-4E, were used in a competition EMSA as described for panel C. In this competition with endogenous full-length TCF-E isoforms, mutation of the GC element (as in LOP+GC^{SC} and GC^{SC}) reduces the ability of the oligonucleotide to compete for binding to the LOP2+GC probe. Both the WRE and the GC element can compete for some but not all of the binding activity.

even when oligonucleotides where the GC element was scrambled or the WRE was missing were used (Fig. 5C [compare LOP2+GC^{SC} with GC and GC^{SC}]). In contrast, EMSA studies with full-length TCF-1E protein showed that effective competition for TCF-1E binding was best with wild-type LOP2+GC, which included both the WRE and an intact GC element (Fig. 5D [LOP2 + GC]). An oligonucleotide with a scrambled GC element in the context of the WRE (LOP2+GC^{sc}) was a less effective competitor at 10- and 50-fold molar excess. Surprisingly, even the GC element alone, but not a scrambled mutant version, competed for TCF-1E binding (Fig. 5D [compare competition at 10-fold molar excess between GC and GC^{sc}]). Competition by GC appeared to be more effective than competition by the WRE alone, a result that matches the large drop in binding activity seen when the "cysteine clamp" (C-clamp) is destroyed (see Fig. 5A) and our assessment that LOP is a weak WRE on its own. However, neither the WRE alone (LOP+GC^{SC}) nor the GC element alone was as effective for competition as the wild-type sequence which carried both elements. These results suggest that stable binding of TCF-1E to a variant or weaker WRE such as the one encoded in LOP2+GC requires recognition of both the WRE and the GC element, with a heavy reliance on the presence of the GC sequence. This dual requirement is entirely consistent with our transient transfection data, showing that both the HMG DNA binding domain and the C-clamp motif are required for binding to the reporter plasmid in vivo (Fig. 2).

Taking the SPR and EMSA data together, we conclude that the cysteine motif in the E-tail functions as a DNA interaction motif. Its unique arrangement of cysteines and basic residues suggests that it is a novel nucleic acid binding structure. We propose calling this motif a "C-clamp" to refer to its conserved cysteines and role as an auxiliary DNA binding domain. As an isolated, independent protein fragment, the C-clamp does not exhibit sequence specificity. However, in its native context in TCF-1E, where it is juxtaposed to the HMG DNA binding domain, this motif can augment overall DNA binding by providing additional nucleic acid contacts with specificity for GC-rich sequences. For some WREs, C-clamp function may be as important as the HMG DNA binding domain.

CASTing analysis of LEF/TCFs shows the C-clamp has sequence selectivity. If the C-clamp exhibits sequence selectivity within the context of full-length TCF-1 protein, then recognition of Wnt target sequences may be significantly influenced. We used CASTing to assess whether the DNA binding specificity of TCF-1 was altered by the neighboring C-clamp motif (see Materials and Methods). Previous systematic evolution of ligands by exponential enrichment experiments have been used to define LEF/TCF consensus binding sequences, but in every case, purified HMG DNA binding domain fragments were

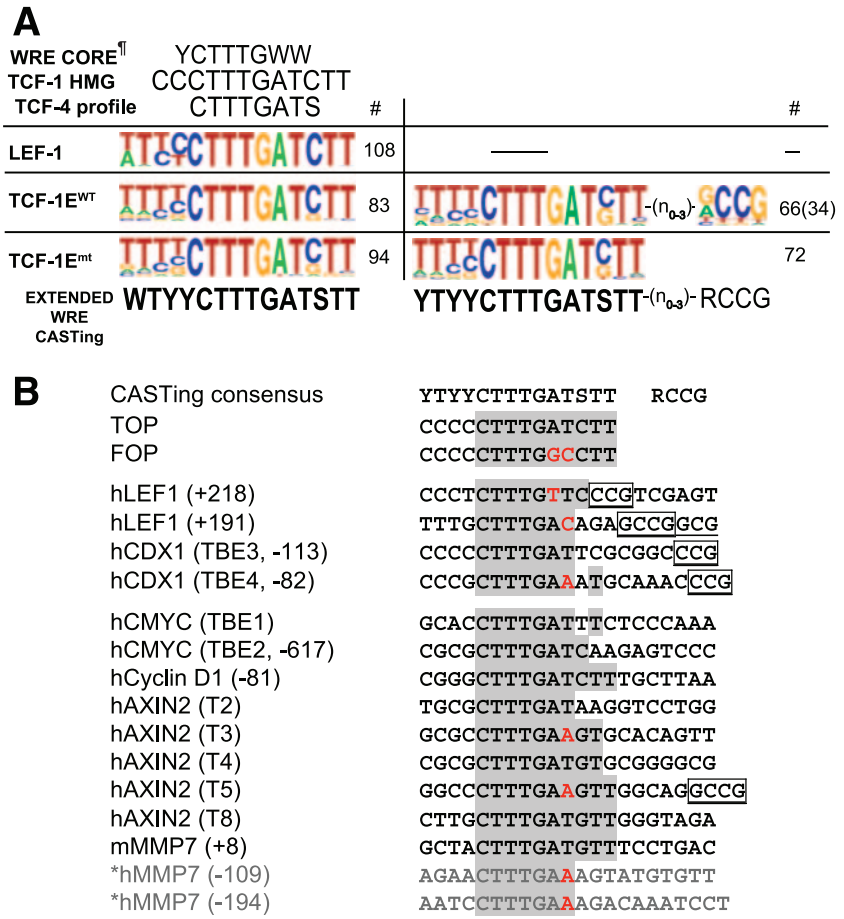

FIG. 6. CASTing analysis of full-length LEF-1 and TCF-1E. (A) Crude cell extracts containing overexpressed, epitope-tagged LEF-1 and TCF-1E were used in a CASTing analysis with a randomly synthesized library (see Materials and Methods). Two independent CASTing analyses were performed (left and right panels). The number of independent sequences used in the sequence alignment is shown for each replicate. Alignments are summarized as sequence logos in which the height of the nucleotide designation represents the frequency of occurrence of that nucleotide. A small GC element was enriched with wild-type TCF-1E for the second replicate CASTing (34 out of 66 independent sequences contained the RCCG motif downstream with variable spacing between 0 to 3 nucleotides). The consensus sequence from a previous systematic evolution of ligands by exponential enrichment experiment is shown (TCF-1 HMG) (39) as well as an affinity profile obtained with a TCF-4/Renilla luciferase fusion protein (TCF-4 profile) (18). W and S are International Union of Pure and Applied Chemistry nomenclature for A or T nucleotides (W) and C or G nucleotides (S). (B) Nucleotide alignment of experimentally validated Wnt response elements. Information about these response elements was derived from studies of *LEF1* (3), *CDX1* (20), *CMYC* (19), cyclin D1 (35), AXIN 2 (24), and *MMP7* (17). Matches to the TOP sequences are shaded gray; mismatched nucleotides are shown in red. RCCG motifs are boxed. Asterisks denote two weak WRE motifs in the human *MMP7* promoter that do not respond to LEF-1/β-catenin when multimerized in a reporter similar to TOPtk and LOPtk (17). ¶ refers to a WRE core sequence that has been experimentally validated in a recent genome-wide survey of sequences occupied by β-catenin (46).

used and neither the E-tail domain nor any other part of LEF/TCF protein was included. In the CASTing experiment reported here, epitope-tagged, full-length LEF-1, TCF-1E^{WT}, and TCF-1E^{mt} proteins were overexpressed in COS-1 cells and extracts of these cells used for in vitro binding assays with a randomized library of oligonucleotides. The proteins were left unpurified to maintain maximum specific activity for DNA binding. The oligonucleotide library was designed with fixed 5′ and 3′ sequences for amplification, and these regions flanked a randomly synthesized core of 30 nucleotides. After six rounds of epitope-tag purification of the LEF/TCFs and PCR amplification of copurified oligonucleotides, the enriched oligonucleotide pool was subcloned and individual clones were sequenced. An extensive set of independent sequences (108 for LEF-1, 83 for TCF-1E^{WT}, and 94 for TCF-1E^{mt}) were aligned with ClustalW. Alignment files were submitted to WebLogo (http://weblogo.berkeley.edu/) to generate sequence logos that depict base frequencies at each position (Fig. 6A, left panel). The results show that an optimal recognition sequence for LEF-1 TCF-1E^{WT} and TCF-1E^{mt} is an extended 14-nucleotide T-rich sequence with only minor differences in nucleotide frequencies between the three motifs. This extended sequence is highly similar to the consensus sequence previously determined with a recombinant fragment of the TCF-1 HMG DNA binding domain (Fig. 6A [TCF-1 HMG]) but different in flanking positions from the positions determined in a serial analysis experiment (examining chromatin occupancy) which identified β-catenin occupancy sites in the genome (Fig. 6A [WRE core]) (46). The extended sequence is much more constrained than is the case with native WREs and therefore represents an opti-

mal binding site (perhaps it also represents T/A-rich oligonucleotides with phased A tracts for DNA bending). Importantly, an independent repeat of the CASTing experiment with different extracts of TCF-1E^{WT} and TCF-1E^{mt} produced an interesting modification to the consensus sequence. In contrast to the procedure employing hypotonic lysis of cells in the first experiment, cells expressing TCF-1E^{WT} and TCF-1E^{mt} were lysed with nonionic detergent and reducing agent. This second CASTing protocol yielded 66 and 72 independent sequences and identical 14-nucleotide T-rich sequences for TCF-1E^{WT} and TCF-1E^{mt} (Fig. 6A, right panel). However, in more than half of the sequences enriched with wild-type TCF-1E^{WT}, a short element (5′-RCCG-3′) was present downstream of the core WRE with variable spacing of up to three nucleotides. None of the sequences isolated with TCF-1E^{mt} had this small motif. It is notable that this RCCG sequence matches the first half of the HDBP response element recognized by HDBP1/HDBP2. Taking these results together with the EMSA data highlighting the importance of the downstream GC element, we conclude that within the context of full-length TCF-1E, the C-clamp can engage in sequence-selective interactions with DNA. A sequence alignment of experimentally determined Wnt response elements from human target genes showed that the occurrence of a RCCG motif is infrequent but appears downstream of each Wnt response element in the *LEF1* and *CDX1* promoters (Fig. 6B).

The C-clamp is involved in regulation of cell proliferation. The C-clamp is highly conserved throughout the animal kingdom, but compared to its ancestral orthologs where the E-tail or C-clamp is obligate, mammalian E-tails in the LEF/TCF family exhibit greater variance. The E-tail is an alternatively spliced exon in TCF-1 and TCF-4, and in TCF-3, half the C-clamp motif is missing. For LEF-1, the E-tail or C-clamp is completely absent. The biological consequence of retaining or losing the C-clamp is not clear, but we hypothesize that a subset of Wnt target genes are selectively regulated by E-tail isoforms and that loss of regulation of this gene set has a consequence in Wnt-regulated cell phenotypes. To test this hypothesis we used DLD-1 colon cancer cells, which have high levels of beta-catenin/TCF complexes in the nucleus. We have determined that the E-tail isoforms of TCF-1 and TCF-4 are predominant isoforms in human colon cancer cells (Syed et al., unpublished), and Clevers, van de Wetering, and associates have shown that inducible overexpression of either dnTCF-4E or dnTCF-1E causes a strong and immediate stall in the G_1 phase of the cell cycle which lasts for several days without much cell death (5, 42). We established a matching DLD-1 cell line that inducibly overexpresses mutant dnTCF-1E^{mt} with the five-amino-acid CRARF substitution in the C-clamp. As described above, this mutation inhibited both transcription activation of the *LEF1* promoter and DNA binding by the C-clamp. Doxycycline induction of dnTCF-1E^{WT}, dnTCF-1E^{mt}, and a mock-treated DLD-1 cell line was carried out for 6 days, and cell growth was monitored on days 2, 4, and 6. Western blot analysis confirmed that equivalent amounts of dnTCF-1E^{WT} and dnTCF-1E^{mt} were induced (Fig. 7 [Western blot inset]). Cell growth was inhibited by wild-type dnTCF-1E^{WT}, resulting in a nearly complete cessation of proliferation. In contrast, rates of cell proliferation in dnTCF-1E$^{\mathbf{mt}}$-expressing cells were no different than those seen with mock-treated cells until day 4 of the induction, when the rate of growth slowed down slightly (Fig. 7). We conclude that the C-clamp motif is important for the growth-arresting function of dnTCF-1E$^{\mathbf{WT}}$ and is important for proper regulation of genes that affect cell growth.

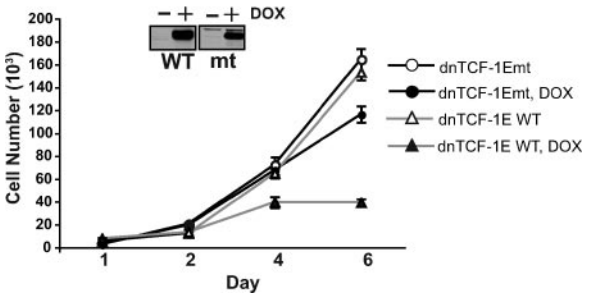

FIG. 7. The E-tail is involved in regulation of colon cancer cell growth. The results of doxycycline (DOX)-induced expression of dominant-negative TCF-1E^{WT} and TCF-1E$^{\mathbf{mt}}$ in DLD-1 colon cancer cells are shown. Quantitation of cell number with or without doxycycline was performed using a sulforhodamine B cell proliferation assay (see Materials and Methods). Error bars depict standard deviations of the results obtained with eight replicates. The inset shows the results of a Western blot analysis of the induced levels of expression of dnTCF-1E^{WT} and dnTCF-1E$^{\mathbf{mt}}$ proteins.

DISCUSSION

Transcription factors with auxiliary DNA binding domains. We conclude from the results of the work presented here that E-tail isoforms of TCF-1 and TCF-4 bind to WREs through bipartite recognition of DNA. Primary binding occurs through sequence-specific recognition of Wnt response elements by the HMG domain, and auxiliary binding occurs via C-clamp interaction with DNA. The C-clamp appears to possess modest sequence specificity for a short GC-rich motif, but we propose that it provides stability of binding through additional DNA contacts whether or not this motif is present. We further propose that E-tail isoforms of TCFs are more potent in their DNA binding activities and may regulate a distinct set of Wnt target genes, including those with weaker Wnt response elements. Some of the targets containing weak WREs are important for cell growth and cycling, as mutating the C-clamp damages the ability of dnTCF-1E to inhibit cell cycling.

A transcription factor with two DNA binding domains is unusual but not unique. For example, alternative splicing at the 3′ end of the p53 tumor suppressor gene generates a longer protein that carries a second DNA binding domain called the CTD (for C-terminal domain) (reviewed in reference 1). The CTD region is enriched in basic residues and binds to DNA without any sequence specificity. Originally thought to be a negative regulator of p53 DNA binding, the CTD has since been shown to stabilize DNA binding when presented with chromatinized templates, long naked DNA fragments, minicircular plasmids, or stem-loop structures (12, 15, 30). The CTD also promotes linear diffusion along DNA and efficient recognition and occupancy of p53 target sites in vivo (12, 29). Results of a study using a recent model for the structure of full-length p53 binding to DNA suggested that the CTD dis-

courages binding to nonspecific sites but stabilizes binding to specific p53 elements (25). It is tempting to speculate that the E-tail plays a similar role for TCF-1 and TCF-4.

A new class of DNA binding domain. The C-clamp motif is highly similar to a 30-amino-acid sequence in the extreme C terminus of HDBP-1 and HDBP-2, two transcription factors that bind to the Huntingtin gene promoter (37). This motif may be uncommon, because a comprehensive BLAST search does not detect similarities to proteins in the database other than the HDBP transcription factors. HDBPs were first cloned as DNA binding proteins that bind to a promoter in the papillomavirus promoter (PRF-1) (7) and an enhancer in the GLUT4 gene (GEF) (31). However, through analysis of HDBP1/HDBP2 binding to specific sequences in the Huntingtin promoter, a DNA binding domain was delimited to the region that includes the cysteine-rich 30-amino-acid motif (37). EMSA as well as DNase I footprinting experiments with a GST fusion of the C-terminal 75 aa or 37 aa from HDBP-1 and HDBP-2 showed that both of these small fusion proteins specifically recognize a 7-bp recognition element (GCCGGCG) in the HD promoter. We assessed whether the TCF C-clamp could bind to specific sequences by use of a fusion of GST to the E-tail of TCF-1 in surface plasmon resonance and EMSA studies. Both approaches showed that the E-tail binds equally well to double-stranded oligonucleotides that encode mutant or wild-type WREs as well as a double-stranded oligonucleotide encoding a Gal4 binding element (data not shown). Therefore, at least as an independent, recombinant protein, the C-clamp does not bind to any specific sequence motif and carries moderate affinity for any double-stranded DNA sequence. There may be other activities carried out by the C-clamp, such as interaction with the transcription coactivator p300. It would be interesting to determine whether HDBPs also interact with p300 through this domain (20).

Sequence variation in Wnt response elements. Despite a lack of sequence specificity in SPR assays and EMSAs, the C-clamp exhibits preferential binding to a GC element near the WRE core sequence when in the context of full-length TCF-1E protein. Matches to the GCCGGCG Huntingtin sequence element recognized by HDBP-1 and HDBP-2 are present in the *LEF1* promoter downstream of each of the two WRE elements by one and seven nucleotides (LOP, CCCG GCT; LOP2, GCCGGCG) (22). Also, the *CDX1* promoter, another E-tail requiring Wnt target, has significant matches to the GCCGGCG element one and seven nucleotides downstream of two of its WREs (CGCGGCC and CCCGCCT) (20). At least for LOP2, this element is important for full-length TCF-1E binding, because its mutation in LOP2 reduced binding in the EMSA (Fig. 5D [LOP2+GC^{SC}]). We did not test this element in a reporter plasmid, because the GC element is too far downstream to permit multimerization of the element and still maintain identical spacing with the TOPtk reporter plasmid. Instead, we multimerized the neighboring WRE from *LEF1*, which has a partial GC element (CCCGGCT). Mutation of this GC-rich sequence to match the flanking sequences in TOPtk does not prevent TCF-1E from activating this reporter but does reduce the level of activation by the chimeric LEF-1E (see Fig. 1B and 2C). Thus, the GC element may enhance target recognition and regulation when the core WRE is itself a variant. An alignment of the *LEF1* and *CDX1* WREs with other known WREs in target genes shows that most other targets do not have a flanking GC element (Fig. 6B). The sequences of these elements coincide more closely with the extended consensus derived from HMG fragments of TCF-1 and TCF-4. There is no natural sequence variation in the CTTTG core but some variation in the two nucleotides that follow this core. In fact, sequence deviation from "AT" of these next two positions correlates with WREs that have been experimentally determined to be weaker WREs (*LEF1*, *CDX1*, and *hMMP7*) (this study and reference 17, 20). Thus, variation in the two positions that follow the CTTTG core comes at a cost but does not preclude Wnt regulation. A recent survey of genomic sites occupied by β-catenin in colon cancer cells identified putative WREs with the core sequence CTTTGWW. Equal frequencies of sequence variations at the "WW" positions were observed followed by preferred nucleotides at 3′ positions (46). One might predict that variation at these positions confers LEF/TCF/β-catenin regulation only through cooperative interactions with specific LEF/TCFs and neighboring factors or via TCF/E-tail isoforms that are more potent and independent. Indeed, LEF-1 regulates *MMP7* via cooperative interactions with c-Jun and PEA3 (9, 11), it cooperatively activates *CDX1* with CDX-1 itself, and it activates its own *LEF1* expression by cooperating with PITX2 (38). Determining how common the downstream GC element is in natural targets, and whether it predicts E-tail-specific regulation, will require identification of a larger set of TCF/E-tail targets.

We used EMSA analysis to test whether the TOP sequence is a stronger WRE than that of LOP or LOP2+GC. We observed that TOP was indeed a stronger binding site (approximately 10-fold and 3-fold more probe shifted than with LOP and LOP2+GC, respectively [F. A. Atcha and B. Wu, data not shown]), supporting the model based on the idea that strong WRE sequences do not require the additional C-clamp. It should be pointed out, however, that the observed decrease in DNA binding with a mutant C-clamp is less dramatic than the effects seen in the transient transfection assays where there is a complete loss of transcription regulation. Likewise, LEF-1 protein can bind to these variant LOP WREs in EMSA and DNase I footprinting assays (3; F. Atcha, unpublished observations), but it is clearly unable to bind productively on its own in vivo. Obviously, recognition of WREs presents more of a challenge in vivo, and therefore the DNA binding activities of the E-tail may be more pronounced with supercoiled, chromatinized templates. This is the case for the auxiliary DNA binding domain in p53, where the CTD actually inhibits DNA binding in vitro but facilitates strong binding to chromatin in vivo (12, 15, 30).

DNA binding affinities and bending of full-length LEF/TCFs. Analysis of recombinant GST/E-tail in surface plasmon resonance experiments showed that the E-tail binds double-stranded DNA directly with an affinity 10- to 20-fold weaker (approximately 16 nM) than the affinity of the HMG DNA binding domain for WREs (approximately 1 nM). Two side-by-side DNA binding domains might be predicted to give E-tail isoforms of TCFs very high affinities for WREs, and we did observe differences in the binding of wild-type and mutant TCF-1E to LOP2+GC (Fig. 5A and D) but not differences that suggest that binding is cooperative. It will be important to measure the DNA binding affinity of full-length TCF-1E^{WT}

and TCF-1E^{mt}, but so far this has proven technically challenging. Only one study has been published in which DNA binding affinities of full-length, purified LEF/TCFs, including TCF-3E (which does not have an intact C-clamp), were measured (33). K_D values for all of the proteins fell in the micromolar range, whereas, interestingly, removal of the E-tail from TCF-3 slightly improved binding (1.5 μM versus 4.5 μM). Micromolar K_D values are much weaker than the nanomolar affinities measured for purified HMG DNA binding domains. This large disparity in DNA affinity measurements could be due to the fact that full-length LEF/TCF proteins suffer a large loss (down to approximately ≤0.1%) in specific activity when they are purified, a property that stems from the fact that purification yields unfolded, unstructured polypeptides (10, 27). The goal of future studies will be to overcome these technical issues so that the contribution of the C-clamp domain to DNA binding can be measured.

A hallmark activity of HMG boxes is that they introduce sharp bends in DNA. Bending measurements have been carried out with protein fragments of the HMG DNA binding domain but not with full-length proteins and never with the juxtaposed E-tail (14, 16, 26). We tested whether the C-clamp influenced DNA bending by use of circular permutation EMSAs with full-length LEF-1, LEF-1E, TCF-1E^{WT}, and TCF-1E^{mt} (F. Atcha, data not shown). Full-length LEF-1 protein induces a bend angle of 132° ± 6.8°, and fusing an E-tail onto the end of LEF-1 does not significantly alter DNA bending (121° ± 6.6°). Interestingly, full-length TCF-1E^{WT} bends DNA at roughly half the angle seen with LEF-1 protein (i.e., 67° ± 2.0° and 70.3° ± 1.0° in independent determinations), but the CR1 mutation does not alter this bending (66.5° ± 1.9° and 70.5° ± 1.3° in independent determinations). Differences in bending could certainly influence the architectural roles of LEF-1 and TCF-1 at sites where this activity is important. Indeed, TCF-1 bending is more similar to the 70° and 80° bend angles induced by SRY, SOX-5, and SOX-2 than it is to that seen with LEF-1 (8, 34, 44). However, we conclude that bending is not a critical part of the differences between LEF-1 and TCF-1E action with respect to *LEF1* and *CDX1* promoter activation because the E-tail, which is the only observed requirement, does not influence DNA bending. Once again, this is similar to the p53 CTD results in that its presence does not appear to alter the conformation of the sequence-specific DNA binding domain (4).

Biological relevance of the C-clamp to LEF/TCF actions. Our findings do not rule out the possibility that LEF-1, TCF-3, and other isoforms of TCF-1 and TCF-4 missing the C-clamp are able to regulate Wnt targets such as *LEF1* or *CDX1*. Rather, our study showed that LEF/TCF isoforms without the E-tail must achieve stable interactions through other mechanisms, such as cooperative interactions with other factors. This is highly likely, since LEF/TCFs are known to be context-dependent transcription regulators that interact with other proteins at various targets (for examples, see reference 2). In fact, a recent study by Béland et al. showed that even though the E-tail-requiring *CDX1* promoter cannot be activated by LEF-1 in transient transfection experiments, LEF-1 can activate *CDX1* transcription in vivo through cooperative interactions with the CDX-1 protein itself (6). Others have shown that LEF-1 can activate transcription of its own promoter when it is coexpressed with PITX2 isoforms (38). Whether these examples of cooperativity are due to stabilization of LEF-1 DNA binding or whether they involve a different mode of complex formation is not known. However, for *CDX1*, a LEF-1/β-catenin fusion protein was not active on its own unless coexpressed with CDX-1, implying that CDX-1 might indeed be functioning to stabilize LEF-1 interactions with this promoter.

It is becoming increasingly clear that vertebrate LEF/TCFs exhibit unique patterns of activities and that the C-clamp is a feature of these specializations. Since the C-clamp motif is as conserved among ancient orthologues as the β-catenin binding domain and the HMG DNA binding domain, one can hypothesize that there are cell phenotypes dependent on the presence of this motif. We asked here whether the E-tail carries any biological significance for TCF activities in DLD-1 colon cancer cells. DLD-1 cells contain very high levels of β-catenin in the nucleus and high concentrations of β-catenin/TCF complexes. Overexpressed dominant-negative TCFs compete with the high concentrations of β-catenin/TCF complexes and downregulate Wnt target gene expression, including that of genes involved in the cell cycle and cell growth (5, 42). Overexpression of dominant-negative TCF-1E (dnTCF-1E^{WT}) in DLD-1 cells causes a stall in the cell cycle but overexpression of dominant-negative TCF with a mutation in the C-clamp (dnTCF-1E^{mt}) does not (Fig. 6). This means that some of the growth control genes downregulated by dnTCF-1E^{WT} contain Wnt response elements that require an intact E-tail for stable association. Since overactivation of growth control genes by TCF/β-catenin complexes can lead to cancer, interference with the function of the C-clamp might undercut these oncogenic actions. LEF-1 and TCF-3, vertebrate family members that do not have an intact C-clamp, either have lost the ability to carry out these essential functions or have compensated by acquiring the ability to cooperate with other factors and pathways. Further studies of the structure of the C-clamp and its influence on target gene selection will be important for fully understanding its role in Wnt signaling.

ACKNOWLEDGMENTS

Don Senear, Klemens Hertel, G. Wesley Hatfield, Jay Gralla, and Rob Steele provided suggestions and critique. Members of the Waterman and Marsh laboratories also provided useful input, critique, and reagents.

This work was funded by National Institutes of Health grants CA108697 and CA096878 to M.L.W. and HD36081 to J.L.M. We gratefully acknowledge the support of the Optical Biology Shared Resource and the Cancer Center Support grant (CA-62203) at the University of California, Irvine, for the SPR studies.

REFERENCES

1. **Ahn, J., and C. Prives.** 2001. The C-terminus of p53: the more you learn the less you know. Nat. Struct. Biol. **8:**730–732.
2. **Arce, L., N. N. Yokoyama, and M. L. Waterman.** 2006. Diversity of LEF/TCF action in development and disease. Oncogene **25:**7492–7504.
3. **Atcha, F. A., J. E. Munguia, T. W. Li, K. Hovanes, and M. L. Waterman.** 2003. A new beta-catenin dependent activation domain in T cell factor. J. Biol. Chem. **278:**16169–16175.
4. **Ayed, A., F. A. Mulder, G. S. Yi, Y. Lu, L. E. Kay, and C. H. Arrowsmith.** 2001. Latent and active p53 are identical in conformation. Nat. Struct. Biol. **8:**756–760.
5. **Batlle, E., J. T. Henderson, H. Beghtel, M. M. van den Born, E. Sancho, G. Huls, J. Meeldijk, J. Robertson, M. van de Wetering, T. Pawson, and H. Clevers.** 2002. β-Catenin and TCF mediate cell positioning in the intestinal epithelium by controlling the expression of EphB/EphrinB. Cell **111:**251–263.

6. **Béland, M., N. Pilon, M. Houle, K. Oh, J. R. Sylvestre, P. Prinos, and D. Lohnes.** 2004. Cdx1 autoregulation is governed by a novel Cdx1-LEF1 transcription complex. Mol. Cell. Biol. **24:**5028–5038.
7. **Boeckle, S., H. Pfister, and G. Steger.** 2002. A new cellular factor recognizes E2 binding sites of papillomaviruses which mediate transcriptional repression by E2. Virology **293:**103–117.
8. **Connor, F., P. D. Cary, C. M. Read, N. S. Preston, P. C. Driscoll, P. Denny, C. Crane-Robinson, and A. Ashworth.** 1994. DNA binding and bending properties of the post-meiotically expressed Sry-related protein Sox-5. Nucleic Acids Res. **22:**3339–3346.
9. **Crawford, H. C., B. Fingleton, M. D. Gustavson, N. Kurpios, R. A. Wagenaar, J. A. Hassell, and L. M. Matrisian.** 2001. The PEA3 subfamily of Ets transcription factors synergizes with β-catenin-LEF-1 to activate matrilysin transcription in intestinal tumors. Mol. Cell. Biol. **21:**1370–1383.
10. **Daniels, D. L., and W. I. Weis.** 2005. Beta-catenin directly displaces Groucho/TLE repressors from Tcf/Lef in Wnt-mediated transcription activation. Nat. Struct. Mol. Biol. **12:**364–371.
11. **El-Tanani, M., A. Platt-Higgins, P. S. Rudland, and F. C. Campbell.** 2004. Ets gene PEA3 cooperates with beta-catenin-Lef-1 and c-Jun in regulation of osteopontin transcription. J. Biol. Chem. **279:**20794–20806.
12. **Espinosa, J. M., and B. M. Emerson.** 2001. Transcriptional regulation by p53 through intrinsic DNA/chromatin binding and site-directed cofactor recruitment. Mol. Cell **8:**57–69.
13. **Giese, K., A. Amsterdam, and R. Grosschedl.** 1991. DNA-binding properties of the HMG domain of the lymphoid-specific transcriptional regulator LEF-1. Genes Dev. **5:**2567–2578.
14. **Giese, K., J. Cox, and R. Grosscheldl.** 1992. The HMG domain of lymphoid enhancer factor 1 bends DNA and facilitates assembly of functional nucleoprotein structures. Cell **69:**185–196.
15. **Göhler, T., M. Reimann, D. Cherny, K. Walter, G. Warnecke, E. Kim, and W. Deppert.** 2002. Specific interaction of p53 with target binding sites is determined by DNA conformation and is regulated by the C-terminal domain. J. Biol. Chem. **277:**41192–41203.
16. **Grosschedl, R., K. Giese, and J. Pagel.** 1994. HMG domain proteins: architectural elements in the assembly of nucleoprotein structures. Trends Genet. **10:**94–100.
17. **Gustavson, M. D., H. C. Crawford, B. Fingleton, and L. M. Matrisian.** 2004. Tcf binding sequence and position determines beta-catenin and Lef-1 responsiveness of MMP-7 promoters. Mol. Carcinog. **41:**125–139.
18. **Hallikas, O., K. Palin, N. Sinjushina, R. Rautiainen, J. Partanen, E. Ukkonen, and J. Taipale.** 2006. Genome-wide prediction of mammalian enhancers based on analysis of transcription-factor binding affinity. Cell **124:**47–59.
19. **He, T. C., A. B. Sparks, C. Rago, H. Hermeking, L. Zawel, L. T. da Costa, P. J. Morin, B. Vogelstein, and K. W. Kinzler.** 1998. Identification of c-MYC as a target of the APC pathway. Science **281:**1509–1512.
20. **Hecht, A., and M. P. Stemmler.** 2003. Identification of a promoter-specific transcriptional activation domain at the C terminus of the Wnt effector protein T-cell factor 4. J. Biol. Chem. **278:**3776–3785.
21. **Hovanes, K., T. W. Li, J. E. Munguia, T. Truong, T. Milovanovic, J. Lawrence Marsh, R. F. Holcombe, and M. L. Waterman.** 2001. Beta-catenin-sensitive isoforms of lymphoid enhancer factor-1 are selectively expressed in colon cancer. Nat. Genet. **28:**53–57.
22. **Hovanes, K., T. W. H. Li, and M. L. Waterman.** 2000. The human *LEF-1* gene contains a promoter preferentially active in lymphocytes and encodes multiple isoforms derived from alternative splicing. Nucleic Acids Res. **28:**1994–2003.
23. **Hurlstone, A., and H. Clevers.** 2002. T-cell factors: turn-ons and turn-offs. EMBO J. **21:**2303–2311.
24. **Jho, E.-H., T. Zhang, C. Domon, C.-K. Joo, J. N. Freund, and F. Costantini.** 2002. Wnt/β-catenin/Tcf signaling induces the transcription of Axin2, a negative regulator of the signaling pathway. Mol. Cell. Biol. **22:**1172–1183.
25. **Kitayner, M., H. Rozenberg, N. Kessler, D. Rabinovich, L. Shaulov, T. E. Haran, and Z. Shakked.** 2006. Structural basis of DNA recognition by p53 tetramers. Mol. Cell **22:**741–753.
26. **Love, J. J., X. Li, D. A. Case, K. Giese, R. Grosschedl, and P. E. Wright.** 1995. Structural basis for DNA bending by the architectural transcription factor LEF-1. Nature **376:**791–795.
27. **Love, J. J., X. Li, J. Chung, H. J. Dyson, and P. E. Wright.** 2004. The LEF-1 high-mobility group domain undergoes a disorder-to-order transition upon formation of a complex with cognate DNA. Biochemistry **43:**8725–8734.
28. **Maeda, T., R. M. Hobbs, T. Merghoub, I. Guernah, A. Zelent, C. Cordon-Cardo, J. Teruya-Feldstein, and P. P. Pandolfi.** 2005. Role of the proto-oncogene Pokemon in cellular transformation and ARF repression. Nature **433:**278–285.
29. **McKinney, K., M. Mattia, V. Gottifredi, and C. Prives.** 2004. p53 linear diffusion along DNA requires its C terminus. Mol. Cell **16:**413–424.
30. **McKinney, K., and C. Prives.** 2002. Efficient specific DNA binding by p53 requires both its central and C-terminal domains as revealed by studies with high-mobility group 1 protein. Mol. Cell. Biol. **22:**6797–6808.
31. **Oshel, K. M., J. B. Knight, K. T. Cao, M. V. Thai, and A. L. Olson.** 2000. Identification of a 30-base pair regulatory element and novel DNA binding protein that regulates the human GLUT4 promoter in transgenic mice. J. Biol. Chem. **275:**23666–23673.
32. **Prieve, M. G., K. L. Guttridge, J. E. Munguia, and M. L. Waterman.** 1998. Differential importin—a recognition and nuclear transport by nuclear localization signals within the high-mobility-group DNA binding domains of lymphoid enhancer factor 1 and T-cell factor 1. Mol. Cell. Biol. **18:**4819–4832.
33. **Pukrop, T., D. Gradl, K. A. Henningfeld, W. Knochel, D. Wedlich, and M. Kuhl.** 2001. Identification of two regulatory elements within the high mobility group box transcription factor XTCF-4. J. Biol. Chem. **276:**8968–8978.
34. **Scaffidi, P., and M. E. Bianchi.** 2001. Spatially precise DNA bending is an essential activity of the sox2 transcription factor. J. Biol. Chem. **276:**47296–47302.
35. **Shtutman, M., J. Zhurinsky, I. Simcha, C. Albanese, M. D'Amico, R. Pestell, and A. Ben-Ze'ev.** 1999. The cyclin D1 gene is a target of the beta-catenin/LEF-1 pathway. Proc. Natl. Acad. Sci. USA **96:**5522–5527.
36. **Skehan, P., R. Storeng, D. Scudiero, A. Monks, J. McMahon, D. Vistica, J. T. Warren, H. Bokesch, S. Kenney, and M. R. Boyd.** 1990. New colorimetric cytotoxicity assay for anticancer-drug screening. J. Natl. Cancer Inst. **82:**1107–1112.
37. **Tanaka, K., J. Shouguchi-Miyata, N. Miyamoto, and J. E. Ikeda.** 2004. Novel nuclear shuttle proteins, HDBP1 and HDBP2, bind to neuronal cell-specific cis-regulatory element in the promoter for the human Huntington's disease gene. J. Biol. Chem. **279:**7275–7286.
38. **Vadlamudi, U., H. M. Espinoza, M. Ganga, D. M. Martin, X. Liu, J. F. Engelhardt, and B. A. Amendt.** 2005. PITX2, beta-catenin and LEF-1 interact to synergistically regulate the LEF-1 promoter. J. Cell Sci. **118:**1129–1137.
39. **van Beest, M., D. Dooijes, M. van De Wetering, S. Kjaerulff, A. Bonvin, O. Nielsen, and H. Clevers.** 2000. Sequence-specific high mobility group box factors recognize 10-12-base pair minor groove motifs. J. Biol. Chem. **275:**27266–27273.
40. **Van de Wetering, M., J. Castrop, V. Korinek, and H. Clevers.** 1996. Extensive alternative splicing and dual promoter usage generate Tcf-1 protein isoforms with differential transcription control properties. Mol. Cell. Biol. **16:**745–752.
41. **van de Wetering, M., R. Cavallo, D. Dooijes, M. van Beest, J. van Es, J. Loureiro, A. Ypma, D. Hursh, T. Jones, A. Bejsovec, M. Peifer, M. Mortin, and H. Clevers.** 1997. Armadillo coactivates transcription driven by the product of the *Drosophila* segment polarity gene *dTCF*. Cell **88:**789–799.
42. **van de Wetering, M., E. Sancho, C. Verweij, W. de Lau, I. Oving, A. Hurlstone, K. van der Horn, E. Batlle, D. Coudreuse, A. P. Haramis, M. Tjon-Pon-Fong, P. Moerer, M. van den Born, G. Soete, S. Pals, M. Eilers, R. Medema, and H. Clevers.** 2002. The beta-catenin/TCF-4 complex imposes a crypt progenitor phenotype on colorectal cancer cells. Cell **111:**241–250.
43. **van Noort, M., and H. Clevers.** 2002. TCF transcription factors, mediators of Wnt-signaling in development and cancer. Dev. Biol. **244:**1–8.
44. **Werner, M., J. Huth, A. Gronenborn, and G. Clore.** 1995. Molecular basis of human 46X, Y sex reversal revealed from the three-dimensional solution structure of the human SRY-DNA complex. Cell **81:**705–714.
45. **Wright, W. E., M. Binder, and W. Funk.** 1991. Cyclic amplification and selection of targets (CASTing) for the myogenin consensus binding site. Mol. Cell. Biol. **11:**4104–4110.
46. **Yochum, G. S., S. McWeeney, V. Rajaraman, R. Cleland, S. Peters, and R. H. Goodman.** 2007. Serial analysis of chromatin occupancy identifies beta-catenin target genes in colorectal carcinoma cells. Proc. Natl. Acad. Sci. USA **104:**3324–3329.