Insights into transcription enhancer factor 1 (TEF-1) activity from the solution structure of the TEA domain

Asokan Anbanandam*, Diana C. Albarado*, Catherine T. Nguyen*, Georg Halder†, Xiaolian Gao‡, and Sudha Veeraraghavan*§

*Department of Biochemistry & Molecular Biology, University of Texas Medical School, Houston, TX 77030; †Department of Biochemistry and Molecular Biology, University of Texas M. D. Anderson Cancer Center, Houston, TX 77030; and ‡Department of Biology and Biochemistry, University of Houston, Houston, TX 77204

Edited by Pierre Chambon, Institut de Génétique et de Biologie Moléculaire et Cellulaire, Strasbourg, France, and approved September 13, 2006 (received for review November 28, 2005)

Transcription enhancer factor 1 is essential for cardiac, skeletal, and smooth muscle development and uses its N-terminal TEA domain (TEAD) to bind M-CAT elements. Here, we present the first structure of TEAD and show that it is a three-helix bundle with a homeodomain fold. Structural data reveal how TEAD binds DNA. Using structure-function correlations, we find that the L1 loop is essential for cooperative loading of TEAD molecules on to tandemly duplicated M-CAT sites. Furthermore, using a microarray chip-based assay, we establish that known binding sites of the full-length protein are only a subset of DNA elements recognized by TEAD. Our results provide a model for understanding the regulation of genome-wide gene expression during development by TEA/ATTS family of transcription factors.

DNA-binding protein | gene regulation | NMR structure | development | Scalloped

Transcription enhancer factors (TEFs) and their homologs make a highly conserved family of eukaryotic DNA-binding proteins whose expression patterns have been correlated with the transcription of viral genes (1, 2) and development (3–7) (Fig. 1*A*). Four TEFs have been identified in humans, namely, TEF-1, TEF-3, TEF-4, and TEF-5. TEF-1 binds M-CAT elements and regulates the development of cardiac, skeletal, and smooth muscles (3, 8, 9). Consistent with this, TEF-1 knockout results in defective heart development and embryonic lethal phenotypes in mouse (10). TEF-3 is found to be expressed largely in lungs and liver, but its expression also is induced by mitogenic stimulation in quiescent fibroblasts or during differentiation of myoblasts to myotubes (8). Expressed from the two-cell stage onwards, mTEAD2, hTEF-4 homolog, is sufficient for TEAD-dependent transcriptional activity during early embryonic development (7, 11, 12). TEF-5 is expressed in placenta and regulates human somatomammotropin gene enhancer (13, 14) and influences fetal development. Widespread early expression of mouse TEAD4 becomes restricted at later stages to neural and mesenchymal tissues as well as nephronic regions of kidneys. TEF orthologs of other eukaryotes also regulate developmental processes. Scalloped (SD) is required for wing development (15). The only *Caenorhabditis elegans* TEF homolog, *egl-44*, regulates differentiation of touch-sensitive cells and egg-laying motor neurons (16–18). The less conserved orthologs of *Saccharomyces cerevisiae* and *Aspergillus nidulans*, namely TEC1, regulate hyphal development (filamentous growth) (19, 20). Development of virulence in systemic candidiasis in mouse models also depends on *Candida albicans* TEC1 activity (21). Thus, the requirement for TEA/ATTS family transcription factors is conserved throughout eukaryotic development.

The most conserved region of TEA/ATTS transcription factors is their N-terminal DNA-binding TEA domain (TEAD) (22, 23) (Fig. 1*B*), which is unique to this family of proteins. TEF-1 and SD are 68% identical overall and 99% identical in the TEAD. This sequence conservation is sufficient for functional conservation because TEF-1 substitutes for SD in wing development (24). The gene regulatory activities of TEF-1 are governed by interactions with protein cofactors (25–27). For instance, interactions of TEF-1 with Max and serum response factor (SRF) regulate the α-MHC gene and normal cardiac and smooth muscle development (28, 29). Vestigial (VG), a protein cofactor of SD in *Drosophila*, is required for transactivation of wing-specific genes (30, 31). Notably, the VG–TEAD fusion protein is sufficient for wing development (32), indicating that TEAD is necessary and sufficient for the *in vivo* DNA-binding activity of TEA/ATTS transcription factors.

Two decades since the discovery of TEF-1, the three-dimensional structure of TEAD remains unknown, largely because of difficulties in generating sufficient quantities of protein. Here, we report the first three-dimensional structure of TEAD determined by using solution NMR spectroscopy. This structure contains three α-helices resembling the homeodomain fold and clarifies previously reported mutational effects. The structure provides insights regarding the DNA-binding activity of TEAD, which we have tested by using mutational analyses. We also establish the consensus DNA-binding sequence for TEAD. Based on these findings, we propose a model for TEF activity.

Results and Discussion

Structure Validation. The accuracy of the final set of 25 structures (Fig. 8, which is published as supporting information on the PNAS web site) is presented in terms of violations of the input distance and dihedral restraints (Table 1). Errors in rmsd reflect precision of structure. Procheck NMR (33) and Errat (34) were used for validation (Figs. 9 and 10, which are published as supporting information on the PNAS web site). For the lowest-energy structure shown in Fig. 2, the backbone dihedral angles of the 82 residues exist in most favored (77.5%) and additional allowed (22.5%) regions of the Ramachandran space (Fig. 9).

TEAD Structure. TEAD is a folded globular protein made of three α-helices, H1, H2, and H3 (Fig. 2*A*). H1 and H2 are nearly

Author contributions: S.V. designed research; A.A., D.C.A., C.T.N., G.H., and S.V. performed research; D.C.A., C.T.N., G.H., and X.G. contributed new reagents/analytic tools; A.A., D.C.A., and S.V. analyzed data; and S.V. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS direct submission.

Abbreviations: TEF, transcription enhancer factor; TEAD, TEA domain; SD, Scalloped; SRF, serum response factor; VG, Vestigial; D-LIA, DNA–ligand interaction assay; HSQC, heteronuclear single quantum coherence.

Data deposition: The atomic coordinates for TEAD have been deposited in the Protein Data Bank, www.pdb.org (PDB ID 2HZD).

§To whom correspondence should be addressed. E-mail: sudha.veeraraghavan@uth.tmc.edu.

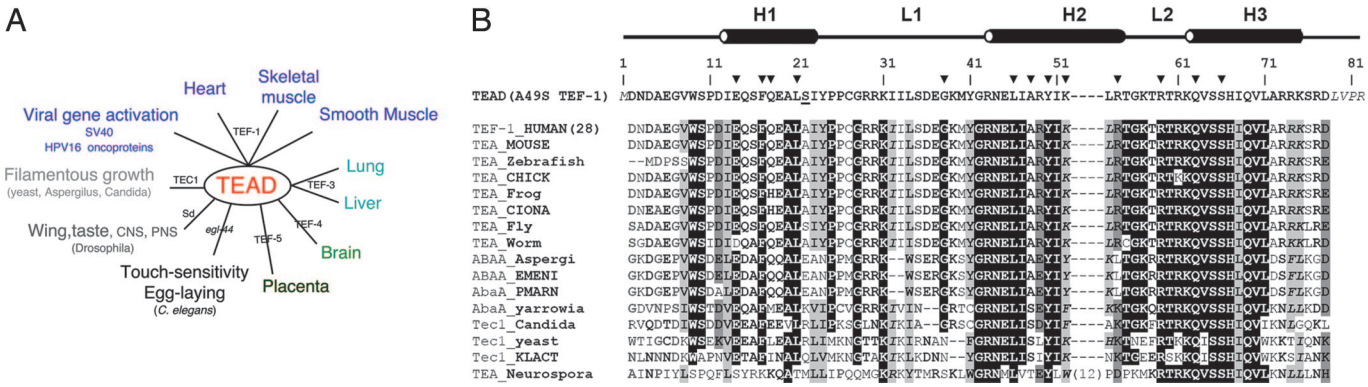

Fig. 1. TEAD transcription factors. (*A*) TEAD proteins regulate tissue-development in eukaryotes. TEF-1 also regulates viral gene activation and TEC1 homolog is essential for *Candida* virulence. (*B*) TEAD sequence alignment shows strict evolutionary conservation of most residues (white letters on black background). TEAD used here contains five extraneous residues (in italics) resulting from cloning procedures. Residue 2 of our TEAD corresponds to residue 28 of the TEF-1. Positions of secondary structural elements, based on our NMR-derived structure, are shown over the sequence. Mutations discussed in the article are identified by ▼.

antiparallel with an interhelical angle of 36° (H1-H2). H1 and H2 pack on either side of the H3 with interhelix angles of 109.8° (H1-H3) and 101.2° (H2-H3). Residues 32–34 and 59–60 exist in β space, which allows loops to reverse direction. Residues 1–11, 24–31, 35–42, 55–58, and 61 are unstructured. Of 28 hydrophobic residues in TEAD, the core consists of only 12 residues (Ile-13, Ile-47, Ile-51, Ile-68; Val-64; Leu-21, Leu-71; Ala-48, Ala-20; Phe-17; and Thr-55, Thr-58; Fig. 2*B*). Of these, L71 makes weak contact to L21 alone. T58 interacts with I13 and T55 to stabilize the L2 loop. The relatively few hydrophobic contacts defining the protein core predicts the thermodynamic stability of TEAD to be low. Consistent with this observation, TEAD unfolds irreversibly with a midpoint for urea denaturation of 2.5 M (Fig. 2*C*). The TEAD surface contains a hydrophobic patch that consists of I23, Y24, L46, Y50, and L53 and is created by the H1- H2 contact (Fig. 2*B*). This surface is likely to be a protein docking site (see below).

Table 1. Structural statistics for 25 final NMR structures of TEAD (1–82)

Statistic	Value
Distance and dihedral constraints	
Distance restraints	
Total NOE	1,121
Intraresidue	585
Sequential ($\|i-j\|=1$)	291
Medium-range ($1<\|i-j\|\leq 4$)	184
Long-range ($\|i-j\|\geq 4$)	60
Dihedral restraints	
ϕ	69 (38 HNHA; 31 TALOS)
ψ	69 (TALOS)
Structure statistics	
Violations, mean ± SD	
Distance constraints, Å	0.031 ± 0.001
Dihedral angle constraints	0.192 ± 0.063
Deviations from idealized geometry	
Bonds, Å	0.004 ± 0.0001
Angles, °	0.363 ± 0.006
Impropers, °	0.190 ± 0.009
Total energy	186.3 ± 6.4
Bond energy	19.8 ± 1.1
Angle energy	49.8 ± 1.8
NOE energy	80.6 ± 4.6
van der Waals energy	31.9 ± 3.7
Impropers	3.9 ± 0.4
rms distance from mean structure	
(H1, H2, and H3: residues 11–23, 43–54, and 62–75)	
Backbone	0.76 ± 0.35
Heavy	1.31 ± 0.28
(H2 and H3: residues 43–54 and 62–75)	
Backbone	0.67 ± 0.35
Heavy	1.31 ± 0.27

TEAD Binds DNA with Nanomolar Affinity. On binding double-stranded 1xGT, the electrophoretic movement of the TEA/DNA complex is retarded relative to the free probe (Fig. 3*A*). The dissociation constant, K_D, is 4–8 nM (Fig. 3*B*). In comparison, K_D for full-length TEAD proteins is 16–45 nM (7). Similarity in these K_D values affirms that TEAD alone confers DNA-binding activity to full-length TEA/ATTS proteins. Like the full-length proteins, isolated TEAD also binds tandem sites (2xGT) cooperatively and with nanomolar affinity (35).

TEAD Binds Numerous M-CAT-Like DNA Sequences. TEFs, SD, and TEC1 share the highly conserved TEAD. The high degree of sequence conservation of the DNA-binding TEAD suggests strongly that their tertiary structure and DNA-binding specificity should also be conserved. Differences in their observed binding selectivities could arise from amino acid differences in their C-terminal domains or be attributable to differences in their cofactor preferences. If so, DNA elements recognized by different full-length TEAD-containing proteins will be a subset of the sequences recognized by the isolated TEAD. To test this in a systematic manner using a large array of synthetic elements, we designed a chip-based assay for direct measurement of protein–DNA interaction named "DNA–ligand interaction assay," or D-LIA (Fig. 3*C* and *Supporting Text*, which is published as supporting information on the PNAS web site). D-LIA results (Fig. 3*D*) show that TEAD binds strongly to 1xGT DNA with the core sequence 5′-TGGAATGT-3′. Further, TEAD binds all previously reported TEF-1 and TEC1 binding sites, some with higher affinities than others (Fig. 11, which is published as supporting information on the PNAS web site). D-LIA data also yielded the consensus DNA sequence recognized by the isolated TEAD with high affinity, namely: N, not C, G, not G, A, T, N, T; where, N = A, T, C, or G. Thus, six of eight nucleotides in the M-CAT sequence confer binding selectivity. Our finding is somewhat surprising given that known TEF-1 binding sequences contain the central ATG. Also surprising is the finding that T is favored in the eighth position because many TEF-1 binding sequences end with C. These differences suggest that although TEF-1 binds to sequences with G7, C8, its intrinsic affinity for such a sequence is lower than for strongly selected sequences such as TGGAATTT. It also might reflect alterations in DNA-binding specificity induced in the full-length

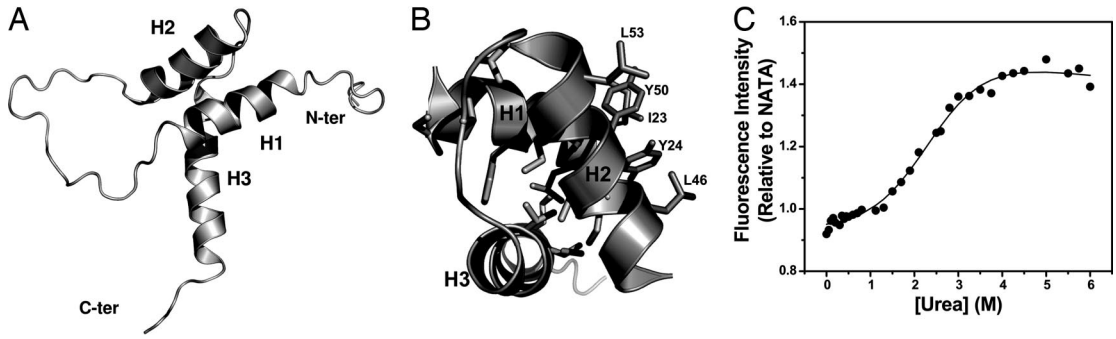

Fig. 2. Solution NMR structure of TEAD. (*A*) Ribbon diagram shows front view of the TEAD fold in which helices H1 and H2 pack against H3. N and C termini are labeled. (*B*) View down the H3 helix illustrates hydrophobic residues (sticks) contributing to core packing and the surface hydrophobic patch. (*C*) TEAD unfolds with a midpoint of 2.5 M urea. Fluorescence intensity is relative to that of an equimolar solution of *N*-acetyl tryptophanamide (NATA).

proteins by conformational changes in TEAD (35), either because of the intact C-terminal domain or an associating protein factor.

We also used D-LIA to determine the length of DNA suitable for NMR experiments to map the protein–DNA interactions. D-LIA results show that TEAD binds 1xGT DNA of different lengths, namely, 8-mer, 12-mer, and 16-mer, equally well (Fig. 3*E*). We chose the 12-mer because it is small enough for NMR studies and expected to be more double-stranded relative to the 8-mer when devoid of 5′ and 3′ flanking nucleotides.

H3 Is the DNA-Recognition Helix. Two-dimensional $^{1}H,^{15}N$-heteronuclear single quantum coherence (HSQC) NMR experiments reveal that ≈30% of TEAD resonances show chemical shift perturbations ($\Delta\delta^{1}H \geq 0.02$ ppm and/or $\Delta\delta^{15}N \geq 0.1$ ppm) upon binding DNA (Fig. 4*A*). Of these, 15 backbone and 3 side-chain resonances are shifted unambiguously (labeled). The majority of these residues are located on the back surface of TEAD and in helix H3 and in the L2 loop immediately preceding H3 (Fig. 4*B*), identifying this as the DNA-binding surface. This result is in agreement with a lack of negative charges and the accessibility of H3 on the DNA-binding surface (Fig. 4*C*). Negatively charged groups on the front surface would preclude a favorable interaction between TEAD and DNA. Thus, H3 is the DNA-recognition helix of TEAD.

We find three serines (Ser-65, Ser-66, and Ser-76) on the DNA-binding surface. Phosphorylation of one or more of these could interfere with DNA-binding activity, by introducing electrostatic repulsion and/or steric hindrance, and help regulate TEA/ATTS transcription factor activity. Biochemical data concur with our structural insights; phosphorylation of Ser-76 (Ser-102 of TEF-1) by protein kinase A (36) or of Ser-65 by protein kinase C (37) diminishes DNA-binding activity.

Structural Homologs. Search for structural homologs of TEAD used DALI (38–40). TEAD most closely resembles the homeodomain protein MatA1 (41) (PDB ID 1AKH; Fig. 5*A*) with an overall backbone rmsd of 1.8 Å (49 residues superimposed, Q score = 0.31, Z score = 3.1). Although large, this root mean square distance (rmsd) is appropriate when comparing the more dynamic DNA-free solution structure of TEAD with the highly stabilized DNA-bound MatA1 in the ordered crystalline form. Structural similarity is even stronger when only residues in helices are compared. The rmsd is 1.4 Å for superimposing residues in the three helices (11–23, 43–54, and 62–75 of TEAD on 79–91, 97–107, and 111–124 of MatA1) or 0.90 Å for residues in H2 and H3 (HTH motif) (Fig. 5*B*). This is in agreement with our finding that H3 of TEAD is the DNA-recognition helix. The structural homology with homeodomains allows modeling of TEAD/DNA interaction. The results presented here indicate that TEAD is a HTH protein and that it belongs to the homeodomain structural family, proving correct the two-decade-old hypothesis of Davidson *et al.* (1).

A hydrophobic surface, similar to that of TEAD (Fig. 2*B*), is present in the Matα2/Mat A1 structure and the homeodomain of the Prospero protein. It is used to convert the three-helix homeodomain structure into a four-helix bundle (42, 43). Further, in the Prospero protein, this structure masks the nuclear export signal. Structural homology suggests that the hydrophobic patch of TEAD also might be involved in the formation of a four-helix bundle. The

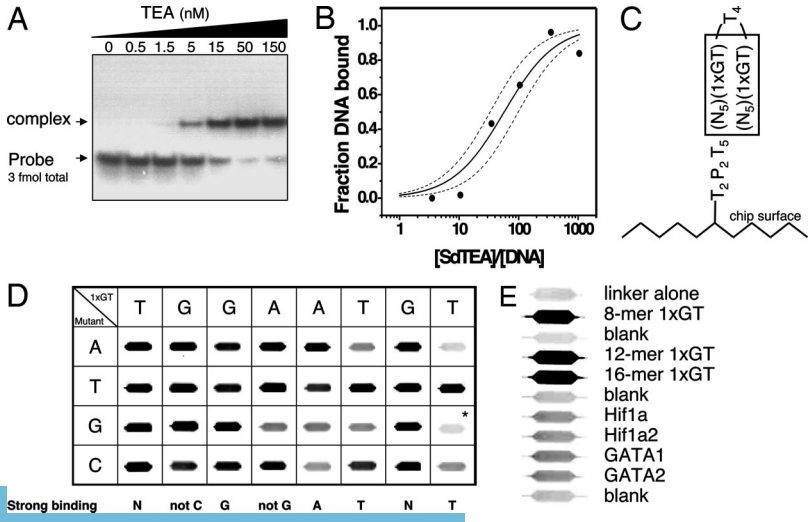

Fig. 3. DNA-binding activity of TEAD. (*A*) EMSA. (*B*) Nonlinear least-squares fit (solid line) of EMSA data (circles) and 95% confidence limits (dotted lines). (*C*) Design of DNA sequences used in D-LIA. DNA is attached to the chip by using a linker containing thymidines (T) and poly(ethylene glycol) (P). N_5 = GCTATG or GGGGG. (*D*) Fluorescent spots extracted from the inverted fluorescence image of a microfluidic chip show TEAD binding to various synthetic double-stranded DNA elements. Dark and light spots indicate strong and weak binding, respectively. *, DNA sequence is given in *Supporting Text*. (*E*) Identification of DNA length suitable for NMR investigations of protein–DNA complex.

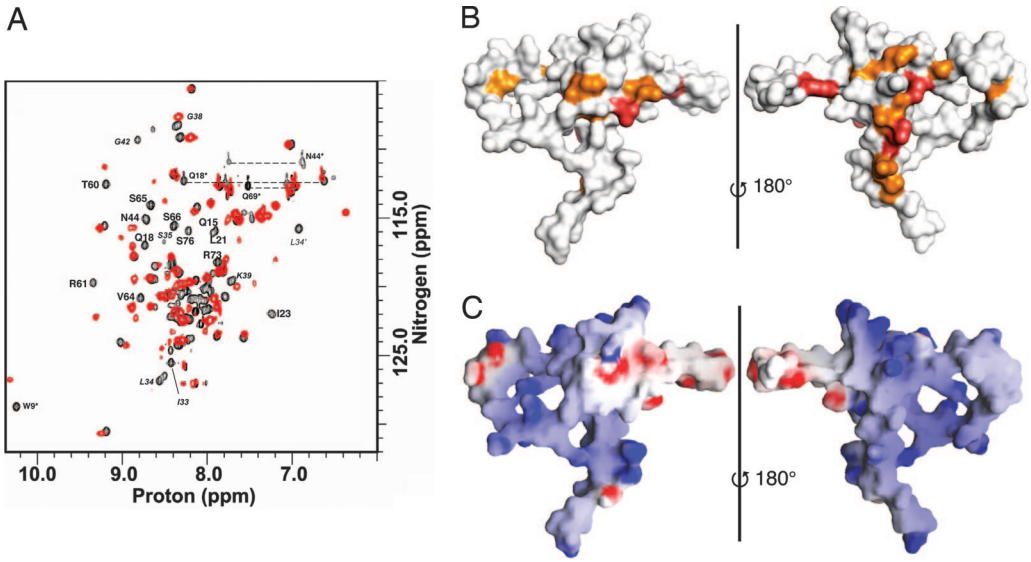

Fig. 4. NMR-mapping of DNA-binding surface. (*A*) $^{1}H,^{15}N$-HSQC spectra of DNA-free (black) and DNA-bound (red) TEAD. Labeled residues show unambiguous and substantial chemical shift changes upon binding DNA. Amino acids in the L1 loop are shown in italics. Asterisks denote side-chain resonances. (*B*) Backbone (orange) and side-chain (red) resonances that respond to DNA binding identify the DNA-binding surface. (*Left*) Front view. (*Right*) Back view. (*C*) Surface electrostatic potential of TEAD. The front surface (*Left*) comprises some negatively charged (red) residues, whereas back surface (*Right*) lacks negative charges.

formation of such a four-helix bundle could increase stability of TEAD, provide additional constraints for DNA-selectivity, or regulate nuclear export of TEAD transcription factors.

Like some other homeodomain proteins, TEF-1 is regulated by and interacts with SRF, a protein that is essential for cardiac development (44–46). Furthermore, the binding of TEAD and SRF to adjacent elements enhances transcription (29). TEAD/SRF interactions, like in the SAP-1/SRF complex (47, 48), may be mediated by contacts between the two DNA-binding domains (29).

Despite its requirement for normal development, nanomolar affinity for DNA, homeodomain structure, and cooperation with SRF, TEAD is distinct from known homeodomains. First, the amino acid sequence of TEAD is distinct enough from the homeodomain consensus that primary structure analyses do not identify TEAD as a homeodomain family member. Second, the activities of TEAD transcription factors cannot be substituted by other homeodomain proteins, or else, TEF-1 and SD knockout would not produce the reported cardiac or wing phenotypes. Third, the DNA-selectivity of TEAD is different from that of homeodomains. Fourth, unlike homeodomain proteins, both the isolated TEAD and the full-length proteins bind cooperatively to tandem, nonpalindromic, and enhancer elements (1, 35, 49). Thus, TEAD is a divergent member of the homeodomain structural family. The fold conservation suggests an evolutionary advantage in using the same architecture to produce divergent functions.

Structural Correlations to TEAD Activity. Some residues in the L1 loop respond to DNA binding (Fig. 4*A*), suggesting that the L1 loop becomes structured and/or contacts DNA. To test whether the L1 loop is a determinant of the DNA-binding activity, we deleted a portion of the L1 loop (residues 26–37) to create the ΔL1 TEAD mutant. The resultant protein is structured (Fig. 12, which is published as supporting information on the PNAS web site) and is competent for DNA binding (Fig. 6*A*, 1xGT data). However, unlike wild-type TEAD, ΔL1 TEAD does not bind cooperatively to 2xGT DNA. Thus, the L1 loop is essential for cooperative loading of TEAD to tandem sites. Earlier work by DePamphilis and coworkers (7, 50) found a TEF-3 mutation (Gly-38-Asp) that renders the protein DNA-binding-deficient. In light of our findings about the L1 loop, we posit that the Asp-38 TEF-3 mutant does not bind DNA either because of (*i*) modified L1 structure and/or (*ii*) unfavorable charge–charge interactions with DNA.

In an earlier report, proline substitutions were used to establish identity of the DNA-recognition helix (51). Our TEAD structure indicates that helix I mutant (F17P, L21P) is DNA-binding-deficient because both F17 and L21 are part of the hydrophobic core (Fig. 6*B*). The mutations likely disrupt secondary and tertiary structures of TEAD. Helix III mutant (Q63P, S66P) weakened but did not abolish DNA binding presumably because these side chains point away from the DNA-binding surface, and their replacement does not disrupt tertiary structure or other key determinants of DNA binding. Helix II mutant L46 and Y50 did not affect DNA-binding activity of TEAD *in vitro* (51), because these residues are part of the surface hydrophobic patch (Fig. 2*B*) and are not involved in core formation or DNA interactions. Disruption of SRF-binding in the A48, K52 double-mutant (29) might have resulted from destabilization of tertiary structure because A48 also is part of the hydrophobic core and packs against V64 of the recognition helix.

Two other mutations within TEAD have physiological consequences in *Drosophila* and *C. elegans*. First, Sd31H mutant (R59K in TEAD) of *Drosophila* is wing-formation-deficient (52). R59 is located in the invariant and structurally ordered L2 loop immediately preceding H3 helix. The guanidino group of R59 is likely to hydrogen-bond to DNA backbone. This is supported by interactions

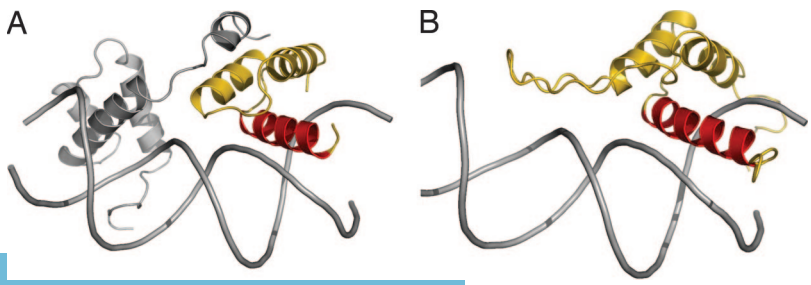

Fig. 5. Models of TEAD structure when bound to DNA. (*A*) The MatA1 homeodomain (gold) in the MatA1/Matα2/DNA complex (PDB ID 1AKH; ref. 62) is the closest structural homolog of TEAD. (*B*) Model of TEAD/DNA complex generated by superimposition of H2 and H3 helices of TEAD on MatA1.

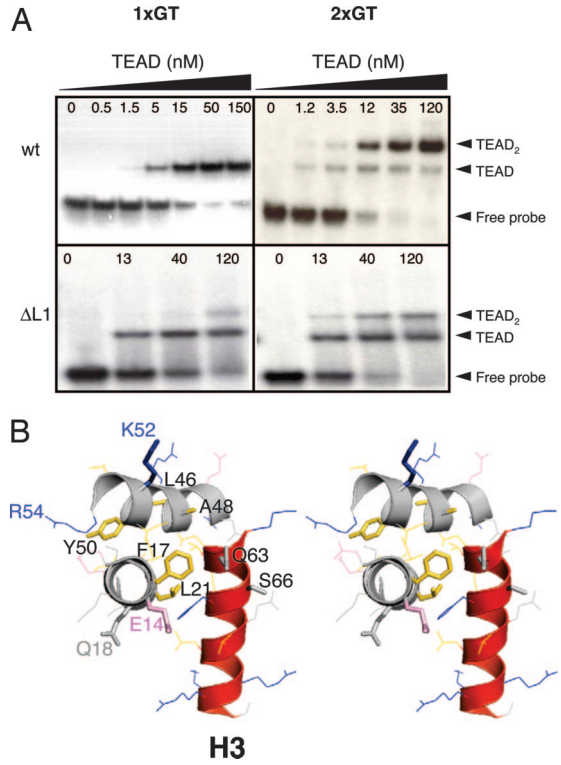

Fig. 6. Correlating TEAD structure and function. (*A*) EMSA of wild-type and mutant TEAD. Wild-type TEAD molecules bind cooperatively to tandem M-CAT-like elements (2xGT). The L1 deletion mutant of TEAD binds M-CAT DNA stochastically, and cooperativity is lost. (*B*) Stereoview of TEAD structure is shown. Amino acid side chains in the three helices are shown as lines or sticks (mutants are also labeled). Loops are not shown for clarity of illustration. Gly-28 resides in loop L1. DNA-recognition helix, H3, is shown in red. Side-chain colors: pink, acidic; blue, basic; gold, hydrophobic; gray, other.

with neighboring residues as well as chemical shift changes exhibited by T60 and R61 on binding DNA (Fig. 4*A*). Therefore, although conserved, the R→K mutation likely reduces hydrogen-bonding interactions and weakens DNA binding. Second, the *egl-44* mutation in *C. elegans* (R54Q in TEAD) converts FSL neuron cells to touch-cell fate (16). R54Q, located at the beginning of the L2 loop, also might affect the DNA-binding function. Deficiency in DNA binding by the R54Q mutant (*egl-44*) would be consistent with lower expression of *egl-46* corepressor and observed derepression of touch-cell fate.

Regulation of TEAD Activity. TEF-1 and SRF, a MADS box protein like MEF-2, induce transcription in a cooperative manner, and SRF can bind directly to TEAD (29). To determine whether functional cooperativity requires cooperativity in DNA-binding activity of these proteins, we investigated the ability of TEAD and SRF MADS box to bind to the CArG–M-CAT synthetic DNA. We find that although each of the proteins is DNA-binding-competent, the two proteins do not load cooperatively (Fig. 7). It suggests that cooperativity in enhancing transcription may be independent of cooperativity in DNA binding. It is, however, possible that functional cooperativity requires direct or indirect (cofactor-mediated) interactions between the two full-length proteins (TEF-1 and SRF).

Using the microarray-based D-LIA, we have shown that the isolated TEAD can bind a large number of M-CAT-like sequences. Our finding supports the idea that TEFs may act as repressors, either because of increased occupancy on promoters or because of a shortage of its protein cofactors (4). For instance, in *Drosophila*, the ectopic expression of VG in the head is sufficient to induce

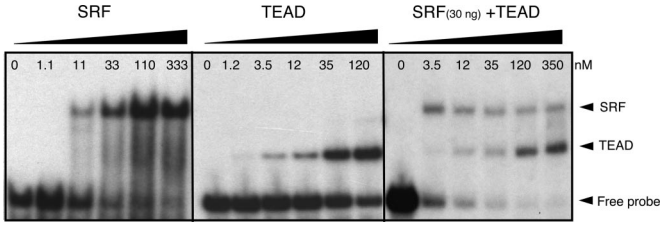

Fig. 7. SRF and TEF-1 DNA-binding domains do not load on to adjacent elements cooperatively.

sprouting of wing-like structures (31). Similarly, in mammalian cells, the VG-like-2 protein promotes skeletal muscle differentiation and activates muscle gene expression by switching DNA-binding properties of TEF-1 factors (27, 53). However, in *C. elegans*, the only TEAD protein, *egl-44*, requires *egl-46* to repress touch-cell fate (16). Thus, the ability of TEAD proteins to bind many M-CAT-like DNA elements makes it a robust transcription factor, whereas its interactions with protein cofactors provide the fine-tuning necessary for tissue-specific action. It is tantalizing to speculate that TEFs, with their promiscuous DNA-binding TEAD, early zygotic expression (12), and ability to interact with p160/SRC (26), might represent a master switch in eukaryotic development.

Materials and Methods

Protein Construct and Sample Preparation. TEF-1 TEAD (A49S mutant, S49 is found in Sd TEAD) was subcloned into pET21d plasmid (Invitrogen, Carlsbad, CA). A thrombin cut site was introduced between the C terminus of the TEAD and the His_6 tag. The L1 mutant was generated by using the QuikChange method (Stratagene, La Jolla, CA) in which residues Pro-26-Arg-37 were deleted. TEAD proteins were overexpressed in CodonPlus *Escherichia coli* cells (Novagen, San Diego, CA). ^{15}N- and ^{13}C-labeled proteins were grown in minimal media with $^{15}NH_4Cl$ as the only nitrogen source as described previously (54). Up to 1.5 mg of purified TEAD was obtained per liter of cell culture (see *Supporting Text* for protein purification details). Molecular weight of the 82-aa thrombin cut A49S TEAD (Fig. 1*B*) is 9,513. TEAD was labeled with Cy-dyes according to the manufacturer's protocol (Amersham Biosciences, Piscataway, NJ). We used an SRF construct provided by Robert Schwartz (Institute of Biosciences and Technology, Texas A&M University, Houston, TX) to similarly clone the MADS box (residues 132–223) into pET21a plasmid and purified the *E. coli*-expressed protein by using nickel-affinity chromatography.

Urea Unfolding of the TEAD. Equilibrium denaturation of TEAD was carried out by using ultra-pure urea in 1× binding buffer [15 mM Hepes (pH 7.9) and 150 mM KCl, 1 mM EDTA, 8% glycerol, and 10 mM β-mercaptoethanol]. Individual solutions of 3 μM TEAD were prepared at different urea concentrations and equilibrated overnight at 4°C, and intrinsic fluorescence intensities were recorded by using an SLM Aminco SPF-500 spectrofluorometer (final temperature was ≈12°C). All other procedures were as described elsewhere (55).

Electrophoretic Mobility-Shift Assay (EMSA). DNA-binding activity of TEAD was established with EMSA. Synthetic 1xGT, 2xGT, and **SRE**-M-CAT DNA probes used for EMSA contained double-stranded DNA made of two complementary strands. One strand for each is as follows: 1xGT (35): 5′-TTCGATACACTTGTGGAATGTGTTTGATTTGTTAGCCCCG-3′; 2xGT (35): 5′-TTCGATACACTTGTGGAATGTGTGGAATGTGTTAGCCCCG-3′; and **SRE**–M-CAT (29): 5′-TGCCTGCTG**CCTAAATTTGG**AATGTTCTGCTGGGACAA-3′.

DNA was labeled with [^{32}P]ATP (30). Each reaction contained

0, 0.1, 0.3, 1, 3, 10, 30, 100, or 256 ng of protein and 2–4 fmol of labeled DNA. Additional details can be found in the supporting information. SRE–M-CAT element (29) uses the M-CAT sequence TGGAATGT instead of TGGAGTCA. Cooperativity between SRF MADS box and TEAD were carried out by first preparing a complex of SRF (30 ng) with labeled SRE–M-CAT DNA (4 fmol). Because SRF is reported to bind directly to TEAD, we added an equimolar amount of unlabeled SRE to remove excess SRF that was not bound to the SRE–M-CAT probe. This established an equilibrium in which SRF was distributed between the two probes, which yields free radioactively labeled SRF–M-CAT probe. TEAD was then added to this mixture.

D-LIA. D-LIA chips were designed as shown in Fig. 3*C*. Each chip contained a total of 3,968 (4 K) wells and tested binding of TEAD to ≈1,800 unique DNA sequences including 1xGT sequence. All steps, except optical scanning, were performed at 4°C. Images were scanned by using the GenePix 4000B DNA chip reader and GenePix Pro 4.1 software (Molecular Devices, Sunnyvale, CA). Fluorescence intensities were determined by using ArrayPro Analyzer (Media Cybernetics, Silver Spring, MD). The intensities were sorted and background subtracted with Microsoft Excel. Mean fluorescence intensity in the blank wells was designated as background. Wells containing linker alone showed background (blank) level signal. The fluorescence intensity in each of the 3,968 wells of a given chip was normalized relative to the average intensity at the 1xGT DNA (5′-TGGAATGT-3′) containing wells in that chip. The normalized fluorescence intensities were used to compile the binding preferences of the TEAD to each DNA sequence on the chips. The results were derived by using the data from three different chips, each containing at least two copies of a given sequence ($n \geq 6$).

NMR Spectroscopy and Structure Calculation. Bruker (Billerica, MA) Avance 600- and 800-MHz spectrometers, equipped with cryoprobes, were used. Center of the proton dimension was referenced relative to water; those of carbon and nitrogen dimensions were indirectly referenced (56). Nitrogen and carbon dimensions were centered at 118 ppm and 75 ppm in the $^{1}H,^{15}N$-HSQC and $^{1}H,^{13}C$-HSQC spectra. Data were processed and analyzed using FELIX (Accelrys, San Diego, CA). Resonance assignments (^{1}H, ^{13}C, and ^{15}N) used standard triple-resonance methods (57) (see *Supporting Text*). Valine and leucine methyls were stereospecifically assigned by using 10% ^{13}C-enriched sample (58). Distance restraints for structure calculations were obtained from proton–proton homonuclear NOESY and three-dimensional ^{15}N- and ^{13}C-separated NOESY spectra. Φ angle restraints were derived from HNHA data (38) or TALOS (59) (Table 1). Structure calculations used crystallography and NMR software (60). Of the final set of 50 structures, 25 lowest-energy structures were selected for further analyses (Fig. 8) and analyzed by using AQUA and PROCHECK-NMR (61).

Chemical-shift mapping used TEAD-DNA complex at 1 mM TEAD and 1.1 mM DNA. Samples were prepared as described elsewhere (60). The double-stranded DNA contained 5′-AG***TGGAATGT***GC-3′ and its complementary strand. Sample conditions: 100 mM sodium phosphate and acetate-d_4 (pH 5.8), 100 mM NaCl, 10 mM βME-d_6, and 0.01% NaN_3 at 25°C.

S.V. thanks C. S. Raman for insightful discussions during manuscript preparation. S.V. thanks the late Fred Rudolf for providing laboratory space immediately after Hurricane Allison (2001) and colleagues at Rice University (Houston, TX) for generously providing space for our NMR laboratory during the recovery period. Youlin Xia is acknowledged for valuable assistance in using the 800-MHz NMR spectrometer located at the University of Houston. Xiaochuan Zhou generously shared his microarray equipment. We thank Pierre Nioche for valuable suggestions regarding cloning and Dong-Sun Lee and Minjun Li for generating the MADS box construct and purifying the corresponding protein, respectively. We thank Eugene Krissinel and Alexey Murzin for discussions on DALI. S.V. gratefully acknowledges the Cold Spring Harbor Laboratories for providing a fellowship to attend their first "Making and Using of DNA Microarrays" course, which spurred on parts of this work. This work was supported in part by American Heart Association Beginning-Grant-in-Aid AHA 0365134Y (to S.V.).

1. Davidson I, Xiao JH, Rosales R, Staub A, Chambon P (1988) *Cell* 54:931–942.
2. Ishiji T, Lace MJ, Parkkinen S, Anderson RD, Haugen TH, Cripe TP, Xiao JH, Davidson I, Chambon P, Turek LP (1992) *EMBO J* 11:2271–2281.
3. Mar JH, Ordahl CP (1990) *Mol Cell Biol* 10:4271–4283.
4. Xiao JH, Davidson I, Matthes H, Garnier JM, Chambon P (1991) *Cell* 65:551–568.
5. Campbell S, Inamdar M, Rodrigues V, Raghavan V, Palazzolo M, Chovnick A (1992) *Genes Dev* 6:367–379.
6. Jacquemin P, Hwang JJ, Martial JA, Dolle P, Davidson I (1996) *J Biol Chem* 271:21775–21785.
7. Kaneko KJ, DePamphilis ML (1998) *Dev Genet* 22:43–55.
8. Hsu DK, Guo Y, Alberts GF, Copeland NG, Gilbert DJ, Jenkins NA, Peifley KA, Winkles JA (1996) *J Biol Chem* 271:13786–13795.
9. Farrance I, Ordahl CP (1996) *J Biol Chem* 271:8226–8274.
10. Chen Z, Friedrich GA, Soriano P (1994) *Genes Dev* 8:2293–2301.
11. Yockey CE, Smith G, Izumo S, Shimizu N (1996) *J Biol Chem* 271:3727–3736.
12. Kaneko KJ, Cullinan EB, Latham KE, DePamphilis ML (1997) *Development (Cambridge, UK)* 124:1963–1973.
13. Jacquemin P, Martial JA, Davidson I (1997) *J Biol Chem* 272:12928–12937.
14. Jiang SW, Wu K, Eberhardt NL (1999) *Mol Endocrinol* 13:879–889.
15. Bray S (1999) *Curr Biol* 9:R245–R247.
16. Wu J, Duggan A, Chalfie M (2001) *Genes Dev* 15:789–802.
17. Desai C, Horvitz HR (1989) *Genetics* 121:703–721.
18. Desai C, Garriga G, McIntire SL, Horvitz HR (1988) *Nature* 336:638–643.
19. Gavrias V, Andrianopoulos A, Gimeno CJ, Timberlake WE (1996) *Mol Microbiol* 19:1255–1263.
20. Andrianopoulos A, Timberlake WE (1994) *Mol Cell Biol* 14:2503–2515.
21. Schweizer A, Rupp S, Taylor BN, Rollinghoff M, Schroppel K (2000) *Mol Microbiol* 38:435–445.
22. Burglin TR (1991) *Cell* 66:11–12.
23. Andrianopoulos A, Timberlake WE (1991) *Plant Cell* 3:747–748.
24. Deshpande N, Chopra A, Rangarajan A, Sashidhara LS, Rogdrigues V, Krishna S (1997) *J Biol Chem* 272:10664–10668.
25. Mahoney WM, Jr, Hong JH, Yaffe MB, Farrance IK (2005) *Biochem J* 388:217–225.
26. Belandia B, Parker MG (2000) *J Biol Chem* 275:30801–30805.
27. Maeda T, Chapman DL, Stewart AF (2002) *J Biol Chem* 277:48889–48898.
28. Gupta MP, Amin CS, Gupta M, Hay N, Zak R (1997) *Mol Cell Biol* 17:3924–3936.
29. Gupta M, Kogut P, Davis FJ, Belaguli NS, Schwartz RJ, Gupta MP (2001) *J Biol Chem* 276:10413–10422.
30. Halder G, Polaczyk P, Kraus ME, Hudson A, Kim J, Laughon A, Carroll S (1998) *Genes Dev* 12:3900–3909.
31. Simmonds AJ, Liu X, Soanes KH, Krause HM, Irvine KD, Bell JB (1998) *Genes Dev* 12:3815–3820.
32. Srivastava A, MacKay JO, Bell JB (2002) *Genesis* 33:40–47.
33. Laskowski RA, Rullmann JA, MacArthur MW, Kaptein R, Thornton JM (1996) *J Biomol NMR* 8:477–486.
34. Colovos C, Yeates TO (1993) *Protein Sci* 2:1511–1519.
35. Halder G, Carroll SB (2001) *Development (Cambridge, UK)* 128:3295–3305.
36. Gupta MP, Kogut P, Gupta M (2000) *Nucleic Acids Res* 28:3168–3177.
37. Jiang SW, Dong M, Trujillo MA, Miller LJ, Eberhardt NL (2001) *J Biol Chem* 276:23464–23470.
38. Holm L, Sander C (1995) *Trends Biochem Sci* 20:478–480.
39. Holm L, Sander C (1999) *Nucleic Acids Res* 27:244–247.
40. Holm L, Sander C (1993) *J Mol Biol* 233:123–138.
41. Li T, Stark MR, Johnson AD, Wolberger C (1995) *Science* 270:262–269.
42. Ryter JM, Doe CQ, Matthews BW (2002) *Structure (London)* 10:1541–1549.
43. Yousef MS, Matthews BW (2005) *Structure (London)* 13:601–607.
44. MacLellan WR, Lee TC, Schwartz RJ, Schneider MD (1994) *J Biol Chem* 269:16754–16760.
45. Carson JA, Schwartz RJ, Booth FW (1996) *Am J Physiol* 270:C1624–C1633.
46. Maeda T, Gupta MP, Stewart AF (2002) *Biochem Biophys Res Commun* 294:791–797.
47. Mo Y, Ho W, Johnston K, Marmorstein R (2001) *J Mol Biol* 314:495–506.
48. Hassler M, Richmond TJ (2001) *EMBO J* 20:3018–3028.
49. Jiang SW, Desai D, Khan S, Eberhardt NL (2000) *DNA Cell Biol* 19:507–514.
50. Vassilev A, Kaneko KJ, Shu H, Zhao Y, DePamphilis ML (2001) *Genes Dev* 15:1229–1241.
51. Hwang JJ, Chambon P, Davidson I (1993) *EMBO J* 12:2337–2348.
52. Srivastava A, Simmonds AJ, Garg A, Fossheim L, Campbell SD, Bell JB (2004) *Genetics* 166:1833–1843.
53. Chen HH, Maeda T, Mullett SJ, Stewart AF (2004) *Genesis* 39:273–279.
54. Veeraraghavan S, Mello CC, Lee KM, Androphy EJ, Baleja JD (1998) *J Biomol NMR* 11:457–458.
55. Veeraraghavan S, Holzman TF, Nall BT (1996) *Biochemistry* 35:10601–10607.
56. Wishart DS, Bigam CG, Yao J, Abildgaard F, Dyson HJ, Oldfield E, Markley JL, Sykes BD (1995) *J Biomol NMR* 6:135–140.
57. Clore GM, Gronenborn AM (1991) *Prog NMR Spect* 23:43–92.
58. Neri D, Szyperski T, Otting G, Senn H, Wuthrich K (1989) *Biochemistry* 28:7510–7516.
59. Cornilescu G, Delaglio F, Bax A (1999) *J Biomol NMR* 13:289–302.
60. Veeraraghavan S, Mello CC, Androphy EJ, Baleja JD (1999) *Biochemistry* 38:16115–16124.
61. Laskowski RA, Rullmannn JA, MacArthur MW, Kaptein R, Thornton JM (1996) *J Biomol NMR* 8:477–486.
62. Li T, Jin Y, Vershon AK, Wolberger C (1998) *Nucleic Acids Res* 26:5707–5718.