A distinct nuclear localization signal in the N terminus of Smad 3 determines its ligand-induced nuclear translocation

Zhan Xiao*, Xuedong Liu*, Yoav I. Henis*†, and Harvey F. Lodish*‡§¶

*Whitehead Institute for Biomedical Research, Nine Cambridge Center, Cambridge, MA 02142; †Department of Neurobiochemistry, The George S. Wise Faculty of Life Sciences, Tel Aviv University, Tel Aviv 69978, Israel; and §Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139

Contributed by Harvey F. Lodish, May 3, 2000

Smad proteins are intracellular mediators of transforming growth factor β (TGF-β) and related cytokines and undergo ligand-induced nuclear translocation. Here we describe the identification of a nuclear localization signal (NLS) in the N-terminal region of Smad 3, the major Smad protein involved in TGF-β signaling. An NLS-like basic motif (Lys^{40}-Lys-Leu-Lys-Lys^{44}), conserved among all pathway-specific Smad proteins, not only is responsible for constitutive nuclear localization of the isolated Smad 3 MH1 domain but also is crucial for Smad 3 nuclear import in response to ligand. Mutations in this motif completely abolished TGF-β-induced nuclear translocation but had no impact on ligand-induced phosphorylation of Smad 3, complex formation with Smad 4, or specific binding to DNA. Hence Smad 3 proteins with NLS mutations are dominant–negative inhibitors of TGF-β-induced transcriptional activation. Smad 4, which cannot translocate into the nucleus in the absence of Smad 3 or another pathway-specific Smad, contains a Glu in place of the last Lys in this motif. Smad 3 harboring the same mutation (K44E) does not undergo ligand-induced nuclear import. Conversely, the isolated Smad 4 MH1 domain does not accumulate in the nucleus but becomes nuclear enriched when Glu^{49} is replaced with Lys. We propose that this highly conserved five-residue NLS motif determines ligand-induced nuclear translocation of all pathway-specific Smads.

Smad proteins are a family of intracellular mediators of the transforming growth factor β (TGF-β) family of cytokines. On activin or TGF-β binding to the respective Type II receptor serine kinases, the Type I receptor becomes phosphorylated. This activated Type I receptor phosphorylates two C-terminal serine residues of Smad 2 or Smad 3, inducing complex formation with Smad 4 and then translocation into the nucleus (1–5). There the complexes interact with other transcription factors (6, 7) to induce or repress transcription of a number of target genes. Like Smads 2 and 3 in the activin/TGF-β pathway, Smads 1, 5, and 8 of the BMP pathway also interact directly with activated receptor kinases and become phosphorylated at their C termini (8). Co-Smads or common-mediator Smads are not substrates of receptor kinases but form complexes with activated pathway-specific Smads to effect transcriptional activation (9). In mammals, the only member of this class is Smad 4. Antagonistic or inhibitory Smads, which counteract the effects of the first two classes, include Smads 6 and 7 (2).

Intensive efforts have focused on two aspects of TGF-β signaling pathways: interactions between Smad proteins and receptor kinases and transcriptional regulation by Smad proteins once inside the nucleus. The central part of the pathway—nuclear import of activated Smads—is poorly understood. The only relevant finding to date is that receptor-regulated Smads, such as Smads 1, 2, and 3, can move independently into the nucleus, but Smad 4 must first complex with one of these Smads to become localized in the nucleus (5).

All transport into the nucleus occurs through nuclear pore complexes, which allow small proteins and solutes of less than 45–55 kDa to diffuse into or out of the nucleus (10, 11). Import of larger proteins usually requires the presence of a nuclear localization signal (NLS), a sequence sufficient and necessary for nuclear import of the host proteins. There are two major types of NLSs: (*i*) a single stretch of five to six basic amino acids, exemplified by the simian virus (SV)40 large T antigen NLS; and (*ii*) a bipartite NLS composed of two basic amino acids, a spacer region of 10–12 amino acids, and a cluster in which three of five amino acids must be basic. This type is typified by nucleoplasmin. For NLS-mediated nuclear import to occur, the NLS first associates with the cytosolic import-receptor proteins importin α and β, which allows docking at the cytoplasmic side of the nuclear pore. Translocation then proceeds through the pore via an energy-driven process (10).

Sequence analysis revealed that Smad proteins contain no classical NLS motif. It is not known whether they have an unconventional NLS that will mediate nuclear translocation, or whether they need to complex with yet-unidentified NLS-containing partners to effect import. Previous studies hint that murine Smad 2 may have an NLS. Intact Smad 2 is normally localized in the cytoplasm, but a LacZ-fusion protein with the isolated Smad 2 C-terminal MH2 domain was localized to the nucleus even in the absence of hormone stimulation (12). These studies implied that Smad 2 contains an NLS within its MH2 domain that is somehow masked or sequestered in the full-length protein.

Because of the intrinsic affinity between the MH1 and MH2 domains (13), we hypothesize that, in the absence of ligand stimulation, any putative NLS in a Smad protein will be buried within the MH1-MH2 intramolecular complex. After receptor phosphorylation, the Smad undergoes a conformational switch that would open up the complex and make the NLS available for nuclear import. Here we show that an isolated Smad 3 MH1 domain contains an NLS-like basic motif, K40KLKK44, which is essential for movement of a fused green fluorescent protein (GFP) into the nucleus regardless of ligand stimulation. More importantly, this basic five-residue motif is required for ligand-induced nuclear import of the intact Smad 3 protein. In contrast, an isolated Smad 4 MH1 domain does not localize to the nucleus; it contains a version of this sequence in which the last lysine residue is replaced by a glutamate. However, a mutant Smad 4 MH1 domain with this glutamate changed to lysine does undergo nuclear accumulation. These results explain why Smad 4 normally requires binding to Smad 3 or another pathway-specific Smad to be transported into the nucleus. Because of its highly conserved nature in all of the

Abbreviations: TGF-β, transforming growth factor β; NLS, nuclear localization signal; GFP, green fluorescent protein.; SV, simian virus.

¶To whom reprint requests should be addressed. E-mail: lodish@wi.mit.edu.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

pathway-specific Smad proteins (Smads 1, 2, 3, 5, 8, and 9) and their *Drosophila* and *Caenorhabditis elegans* homologues (Mad, Sma-2, and Sma-3), we propose that this motif acts as a common NLS for Smad proteins.

Materials and Methods

Constructs. GFP-tagged Smad 3 and Smad 4 were made by using the CLONTECH pEGFP-C1, which allows Smad fusion to the C terminus of GFP. Primers corresponding to the N- and C-terminal regions of Smad 3 and Smad 4 MH1 and MH2 domains were tagged with *Bgl*II and *Eco*RI sites, respectively, and used in PCR reactions with human Smad 3 and Smad 4 cDNA as template (Pfu polymerase, Stratagene). The products were digested with *Bgl*II and *Eco*RI and subcloned into the pEGFP vector to allow an in-frame fusion (pEGFP-Smad). The sizes for the various constructs are as follows: Smad 3 MH1 domain (amino acids 1–145), Smad 3 MH2 domain (amino acids 220–425), Smad 4 MH1 domain (amino acids 1–139), and Smad 4 MH2 domain (amino acids 319–552). To generate a retroviral vector encoding a GFP–Smad fusion, the pEGFP-Smad vector was digested with *Age*I and *Sal*I and then ligated into a similarly restricted pMX vector to create the pMX-GFP-Smad (14).

Cell Lines and Transfections. For transient expression, BOSC cells and NMuMG cells (kindly provided by R. Sendak, Creative Biomolecules, Hopkinton, MA) were transfected with Lipofectamine Plus reagent (Life Technologies, Rockville, MD) according to the manufacturer's instructions. To generate stable cell lines expressing various GFP-Smad fusion proteins, BOSC cells were transfected with pMX-GFP-Smad constructs. Two days after transfection, cell supernatant containing the retroviruses was collected and used to infect L20 cells, a Mv1Lu cell line expressing mouse ecotropic viral receptor (14), for 6–9 h in normal medium containing 4 μg/ml polybrene (Sigma).

Site-Directed Mutagenesis. Mutations of specific amino acids in Smad 3 were constructed by using the QuikChange Mutagenesis kit (Stratagene). For substitution of Lys^{43}-Lys^{44} with Asn and Gln (K43N/K44Q), the sense-strand oligo used was: AGCCTGGTCAAGAAACTCAACCAGACGGGGCAGCTGGACGAG. To delete Lys^{43}-Lys^{44} (ΔK43K44), the sense-strand oligo was: AGCCTGGTCAAGAAACTCAC GGGGCAGCTGGACGAG. To change Lys^{44} to Glu to mimic Smad 4 (K44E), the sense-strand oligo used was: GTCAAGAAACTCAAGGAGACGGGGCAGCTG. To delete Lys^{40}-Lys^{41} ($\Delta K^{40}K^{41}$), the sense strand used was: GCGGTCAAGAGCCTGGTCCTCAAGAAGACGGGGCAG.

Fluorescence Microscopy. GFP fluorescence in transfected or retrovirally infected cells was directly visualized with a Nikon TE300 inverted microscope equipped for epifluorescence by using a 488-nm excitation filter and 522- to 535-nm emission filter. Images were recorded with a Hamamatsu (Middlesex, NJ) Orca CCD camera and analyzed with OPENLAB software in the Keck microscope facility at the Whitehead Institute.

Electrophoresis Mobility Shift Assay (EMSA). EMSA was performed essentially as described previously (7).

Luciferase Assays. Twenty-four hours before transfection, Mv1Lu cells were seeded in triplicate at 2×10^5 cells per well in a six-well plate. One microgram each of 3TP-Luciferase and indicated Smad 3 vectors was used to transfect each well by Lipofectamine Plus (Life Technologies); 0.5 μg pSVβ (CLONTECH), encoding β-galactosidase, was included in each sample to control for transfection efficiency. Thirty-six hours after transfection, cells were treated with 200 pM TGF-β or with buffer in serum-free medium for 18–24 h. Luciferase and β-galactosidase activity were measured as described previously (14).

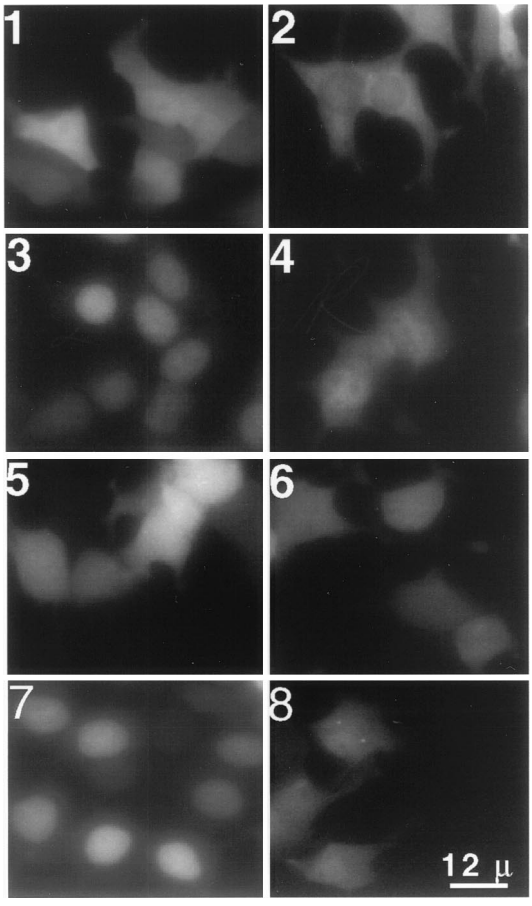

Fig. 1. Smad 3 MH1 domain is constitutively localized to the nucleus after both transient and stable expression. (*1–6*) Transient expression; MH1 and MH2 domains of Smad 3 and Smad 4 proteins were fused to the C terminus of eGFP and transiently expressed in BOSC cells. One day after transfection, the living cells were photographed under the fluorescence microscope to detect the GFP signal. (*1*) GFP; (*2*) GFP-Smad 3 full-length protein fusion; (*3*) GFP-Smad 3 MH1; (*4*) GFP-Smad 3 MH2; (*5*) GFP-Smad 4 MH1; and (*6*) GFP-Smad 4 MH2. Nuclear localizations were confirmed by 4′,6-diamidino-2-phenylindole staining of fixed cells (not shown). (*7–8*) Stable expression. The MH1 domains of Smads 3 and 4 each were fused with GFP and cloned into the retroviral pMX vector. Two days after transfection into BOSC cells, virus-containing media were collected and used to infect L20 cells. Two days later, living cells were photographed under the fluorescence microscope. (*7*) Smad 3 MH1; (*8*) Smad 4 MH1.

Results

MH1 Domain of Smad 3 but Not Smad 4 Is Constitutively Localized in the Nucleus. To locate the putative NLS in Smad proteins, we constructed GFP fusions with the isolated MH1 and MH2 domains of Smad 3 and Smad 4. We fused GFP to the N terminus of each domain in light of our previous observation that blocking the C terminus of Smad 3 disrupts its function (14). We first transiently overexpressed these constructs in BOSC cells and monitored GFP fluorescence signals (Fig. 1 *1–6*). Consistent with previous studies (15), GFP itself exhibited fluorescence evenly distributed between nucleus and cytoplasm (Fig. 1*1*), indicating that it did not contain a NLS. This profile is presumably because of its small size (27 kDa), which allows it to traverse freely through nuclear pores. A GFP fusion with the full-length Smad 3 was predominantly cytoplasmic, although nuclear staining was also visible, especially in high-expressing cells (Fig. 1*2*).

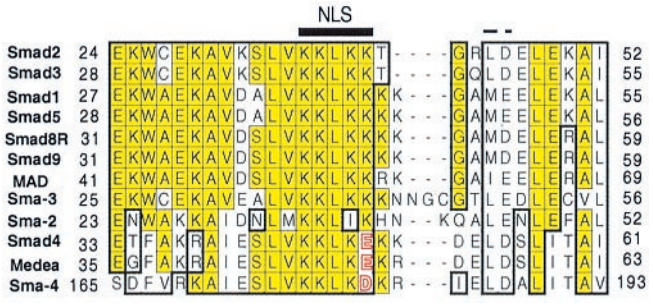

Fig. 2. Multiple sequence alignment of the presumptive NLS regions of Smad proteins. N-terminal regions from pathway-specific Smads 1, 2, 3, 5, 8, and 9 (all human except Smad 8, which is from rat), their *Drosophila* and *C. elegans* homologues (Mad, Sma-2, and Sma-3), and the co-Smad protein Smad 4 and its *Drosophila* and *C. elegans* homologues (Medea and Sma-4) were aligned with CLUSTALX software and displayed by the SEQVU program (using Smad 2 as primary sequence). Identical residues are yellow colored, and homologous residues are boxed. The identified five-residue NLS-like motif is highlighted. Smad 4 and its homologues contain an ''E'' or ''D'' in place of ''K'' in the last position of the NLS; these acidic residues are shaded red. The two conserved hydrophobic/acidic residues at the C-terminal side of the NLS are indicated by a dashed line.

This is in line with our speculation that in the absence of ligand, the putative NLS in Smad 3 does not function in the context of the whole protein.

In contrast, the GFP-Smad 3 MH1 fusion protein displayed a dramatic accumulation in the nucleus (Fig. 1*3*), suggesting that it contains a fully functional NLS. On the other hand, the Smad 3 MH2 domain, although somewhat enriched in the nucleus, was mostly dispersed throughout the cell (Fig. 1*4*). Because Smad 4 cannot independently translocate into the nucleus, our hypothesis predicts that it should not contain a NLS. Correspondingly, neither its MH1 nor its MH2 domains were enriched in the nucleus (Fig. 1 *5* and *6*, respectively). All nuclear localizations were confirmed by 4′,6-diamidino-2-phenylindole staining. Similar results were obtained by transiently expressing these constructs in NMuMG cells (data not shown). For each cell line, GFP immunoblots were performed to confirm that the correct-sized fusion proteins were made at similar expression levels (data not shown).

To prove that the nuclear localization of the GFP-Smad 3 MH1 fusion protein was not caused by transient overexpression, we subcloned the various GFP-Smad fusion constructs into pMX, a retroviral vector allowing stable expression in infected cells. Using the Mv1Lu cell line expressing the ecotropic murine receptor as recipient (14), we generated stable cell lines expressing these GFP fusion proteins. The findings were consistent with the above transient expression results even though the overall signal level was reduced: the Smad 3 MH1 domain was again completely nuclear, whereas the Smad 4 MH1 domain was uniformly distributed between the cytoplasm and nucleus (Fig. 1 *7* and *8*).

Nuclear Localization of the Smad 3 MH1 Domain Depends on an N-Terminal Basic Motif. To pinpoint the presumptive NLS motif within the Smad 3 MH1 domain, we scanned its sequence and noticed an SV40 T-antigen NLS-like basic motif in the N-terminal part of the protein (residues 40–44: KKLKK). A multiple sequence alignment of mammalian pathway-specific Smads (Smads 1, 2, 3, 5, 8, and 9), together with their *Drosophila* and *C. elegans* homologues (Mad, Sma-2, and Sma-3), demonstrated that, with a minor exception in the case of Sma-2, this five-residue motif is highly conserved (Fig. 2). Interestingly, when Smad 4 and its invertebrate homologues Medea and Sma-4 were also aligned, we noticed that the C-terminal Lys in this

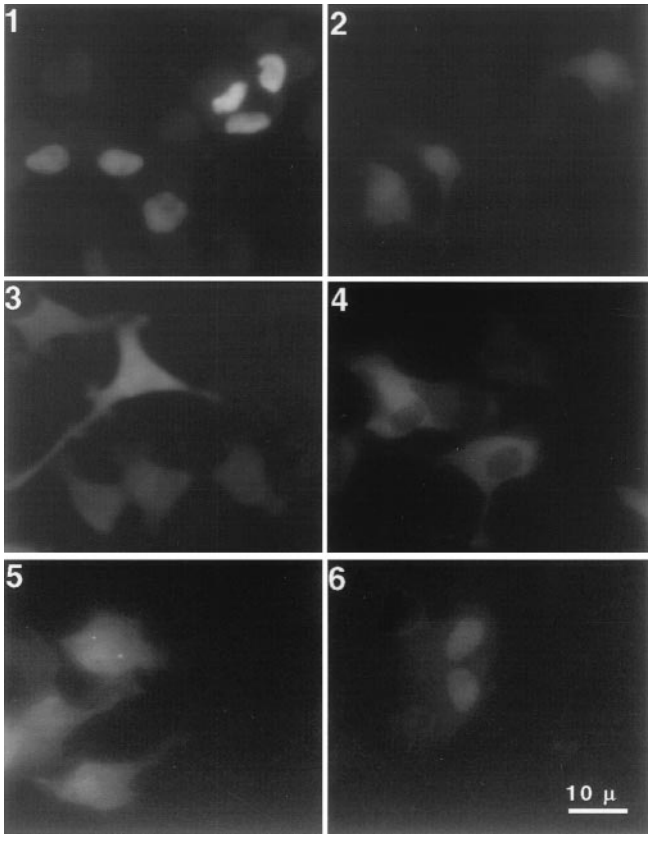

Fig. 3. Mutations in the NLS motif interfere with nuclear localization of Smad 3 MH1. The five-residue NLS motif in the Smad 3 MH1-GFP fusion protein was mutated in different ways, and the mutant proteins were stably expressed in L20 cells by retrovirus infection. Two days after infection, cells were viewed under the fluorescence microscope and the GFP signal recorded. (*1*) Wild- type Smad 3 MH1; (*2*) Smad 3 MH1 K43N/K44Q; (*3*) Smad 3 MH1 ΔK40K41; (*4*) Smad 3 MH1 ΔK43K44; (*5*) wild- type Smad 4 MH1; (*6*) Smad 4 MH1 E49K.

motif was changed to either Glu or Asp, a dramatic charge reversal (red-colored residues). This might explain why Smad 4, containing a defective NLS, is unable to move into the nucleus by itself.

To prove that this motif is essential for nuclear localization of the Smad 3 MH1 domain, it was disrupted by either point mutations or small deletions. After expression in stable cell lines (L20), these Smad 3 MH1 NLS mutants demonstrated dramatically altered subcellular distributions (Fig. 3). The wild-type Smad 3 MH1 domain was exclusively nuclear (Fig. 3*1*), whereas the three NLS mutants tested were either distributed throughout the cell or excluded from the nucleus (Fig. 3 *2–4*). K43N/K44Q (Fig. 3*2*) was significantly less enriched in the nucleus compared with wild type, and the deletion mutant ΔK40K41 (Fig. 3*3*) became evenly distributed throughout the cell. Deletion mutant ΔK43K44 (Fig. 3*4*) was almost excluded from the nucleus. Transient expression of these NLS mutants in BOSC or NMuMG cell lines provided essentially the same disrupted profiles (data not shown). These data establish that the identified NLS motif plays a critical role in the nuclear enrichment of the isolated Smad 3 MH1 domain.

We additionally confirmed that Smad 4 MH1 domain was, as before, uniformly distributed throughout the cell under stable expression (Fig. 3*5*). More importantly, a mutant Smad 4 MH1 domain (E49K), in which the deviant Glu in the last position of the NLS-homologous sequence was changed to Lys, exhibited a dramatic concentration in the nucleus (Fig. 3*6*). Thus, the

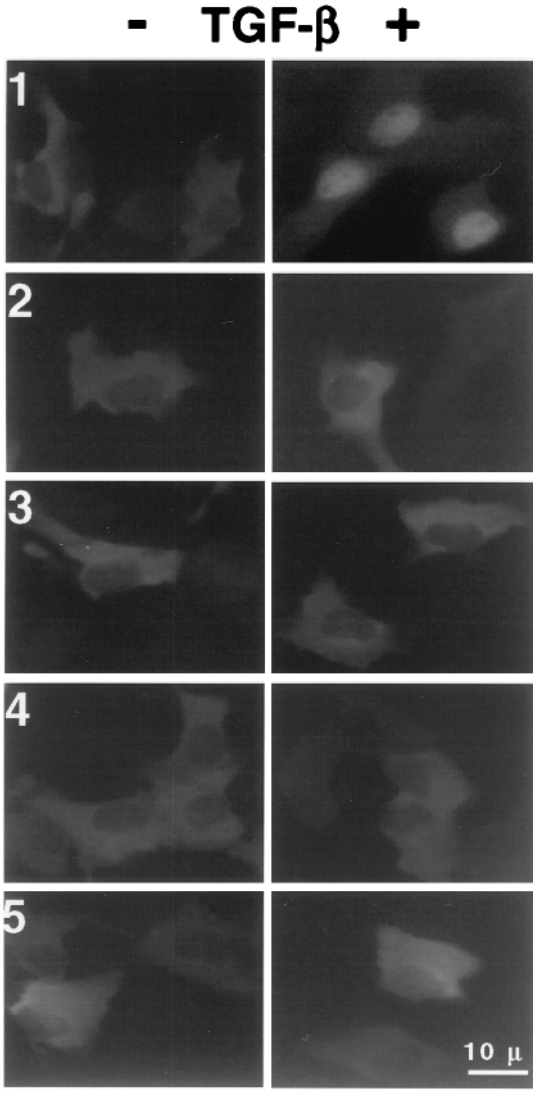

Fig. 4. NLS-mutated Smad 3 proteins failed to translocate to nucleus in response to TGF-β. L20 (Mv1Lu) cells stably expressing wild-type or mutant forms of GFP-Smad 3 were first starved in low-serum medium (0.5% FCS in DME) for 2–3 h and then treated with 200–500 pM TGF-β. After 1 h, the cells were then photographed under the fluorescence microscope. *Left* for each sample shows the image before stimulation; *Right* shows the image after TGF-β addition. (*1*) Wild-type Smad 3; (*2*) Smad 3 K43N/K44Q; (*3*) Smad 3 ΔK43K44; (*4*) Smad 3 K44E; (*5*) Smad 3 ΔK40K41.

Lys-to-Glu change in this position in Smad 4 accounts for the nonnuclear distribution of its MH1 domain and presumably also explains the inability of full-length Smad 4 to translocate into the nucleus in the absence of Smad 2 or Smad 3.

Smad 3 Proteins with Mutant NLS Sequences Fail to Undergo TGF-β-Induced Nuclear Translocation. To ascertain the relevance of this presumptive NLS in the nuclear import of intact Smad 3, we made similar point mutations and small deletions in wild-type full-length Smad 3 fused with GFP. To achieve Smad 3 nuclear translocation induced by physiological levels of TGF-β, cell lines stably expressing GFP-Smad 3 had to be constructed. Transient-expressing cells were poorly responsive to TGF-β stimulation, probably because Smad 3 level was too high to show any ligand-dependent change in its localization profile (data not shown). In the resting state (Fig. 4 *Left*), wild-type GFP-Smad 3 and all of the Smad 3 NLS mutants were excluded from the nucleus. One hour after stimulation with 200 pM TGF-β, wild-type Smad 3 (Fig. 4*1*) displayed almost complete nuclear accumulation, consistent with our previous studies on native Smad 3 protein (14). This translocation was detectable only 5–10 min after ligand addition and approached completion within 30–40 min (data not shown). In contrast, the NLS mutants were totally defective in nuclear import (Fig. 4 *2–5*), and no accumulation was observed even after prolonged incubations up to 6 h (data not shown). The mutations were similar to those shown in Fig. 3. Note that the K44E mutation, which mimics the Smad 4 NLS sequence, abrogates Smad 3 nuclear import (Fig. 4*4*). These results demonstrate that the basic motif not only is essential for constitutive nuclear localization of isolated MH1 domains but also is necessary for ligand-induced translocation of intact Smad 3. It is noteworthy that, whereas substitution mutations were less disruptive in MH1 domain nuclear targeting than were deletion mutations (Fig. 3, compare *2* with *3* and *4*), they caused the same severe damage to nuclear import of the intact Smad 3 protein (Fig. 4, compare *2* and *3*).

Smad 3 NLS Mutants Are Dominant-Negative Inhibitors of TGF-β Transcriptional Activation. Because of their inability to translocate to the nucleus in response to TGF-β stimulation, the NLS-mutated Smad 3 proteins are likely to be defective in activating transcription of target genes. Such mutants may even be dominant negative if they are still able to couple with the Type I receptor or bind to Smad 4. To investigate this possibility, we cotransfected Mv1Lu cells with a 3TP-luciferase reporter construct and wild-type or NLS-mutated Smad 3 constructs, all as GFP-fusions. One day after transfection, cells were treated with or without 200 pM TGF-β to study ligand-induced reporter gene activation (Fig. 5*A*). Vector-transfected controls (Mv1Lu) displayed the expected 15- to 20-fold induction of luciferase activity after TGF-β treatment. Cotransfection with both wild-type Smad 3 and the GFP-Smad 3 fusion construct produced moderate increases in induction compared with the vector control, consistent with previous results (14), and indicated that in this assay, GFP-Smad 3 fusion behaved similarly to untagged Smad 3. In contrast, the two Smad 3 NLS mutants (K43N/K44Q and ΔK43K44) showed over 65% reduction in ligand-induced reporter gene activity, demonstrating that not only are they defective in transcriptional activation, but they also dominantly inhibit functions of the endogenous wild-type Smad 3. As expected, the Smad 3 R74A mutant, which is defective in DNA binding (16), also significantly decreased TGF-β-induced gene activation in a dominant-negative fashion.

Smad 3 NLS Mutants Are Activated Normally by the Type I TGF-β Receptor and Form Normal DNA-Binding Complexes with Smad 4. Despite their inabilities to activate gene expression, the NLS-mutated Smad 3 proteins may still act as normal substrates of the Type I receptor kinase because their C-terminal phosphorylation sites remain intact. Competitive binding to receptors could be the underlying reason for their dominant-negative actions. To examine the functionality of the NLS-mutant Smad 3 proteins, we carried out DNA-binding gel-shift assays by using extracts from cells expressing wild-type Smad 3 or a Smad 3 NLS mutant, K43N/K44Q, which is deficient in ligand-induced nuclear localization. The DNA probe used was the Smad-binding element that contains two tandem copies of the consensus Smad 3-binding palindromic sequence (GTCTAGAC) (17). As expected, extracts from cells expressing wild-type Smad 3, GFP- Smad 3 wild type, or GFP-Smad 3 K43N/K44Q alone (without active Type I receptor) showed no specific DNA binding (Fig. 5*B*, lanes 1, 4, and 7), consistent with the notion that in the resting state, intramolecular complex formation between MH1 and MH2 not only sequesters NLS but also masks the DNA-binding site on the MH1 domain. The constitutively active Type I receptor Tβ-RI/

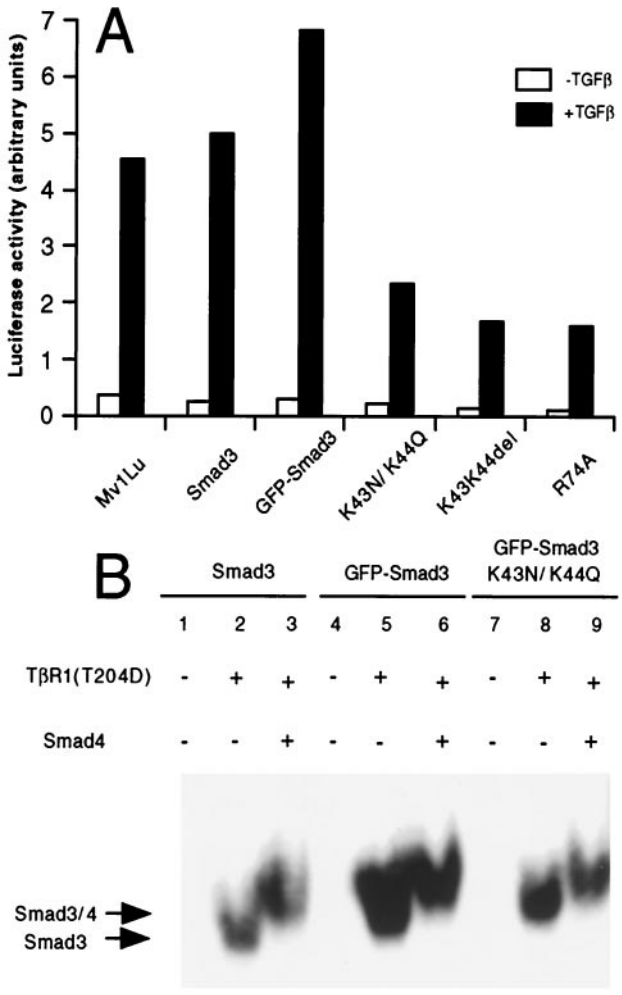

Fig. 5. (*A*) Smad 3 NLS mutants are dominant-negative inhibitors of TGF-β transcriptional activation. 3TP-Luciferase assay was performed as described previously (14). Cells were treated with 200 pM TGF-β (+) or buffer (−) for 24 h, then lysed in reporter assay buffer (Promega) and subjected to luciferase assay. All results are averages of triplicate tests (sample-to-sample variations within the triplicate are all less than 8%). Mv1Lu: cells transfected with vectors only; K43N/K44Q: Smad3 K43N/K44Q; K43K44del: Smad 3 ΔK43K44; R74A: Smad 3 R74A (DNA-binding mutant). (*B*) Smad 3 NLS mutants are activated normally by the Type I TGF-β receptor and form normal DNA-binding complexes with Smad 4. BOSC cells transiently expressing Smad 4 (+) or not (−); constitutively activated TβRI T204D (+) or not (−); and either wild-type Smad 3, GFP-Smad 3, or GFP-Smad 3 NLS mutant K43N/K44Q were lysed in gel-shift buffer. Clarified lysates were then used in gel-shift assays to bind a ^{32}P-labeled Smad-binding element. Samples were resolved on a 5% 1× TEB buffer (90 mM Tris/64.6 mM boric acid/2.5 mM EDTA, pH 8.3)/PAGE. Positions of Smad 3-DNA complex (Smad 3) and Smad 3/Smad 4/DNA complex (Smad 3/4) are indicated by arrows.

T204D is able to phosphorylate Smad 3 at its C terminus and functionally mimics ligand-stimulated conditions (18). Hence in its presence all three constructs exhibited intense protein–DNA complexes (lanes 2, 5, and 8), and all three complexes displayed slower mobility when the cells also coexpressed Smad 4 (lanes 3, 6, and 9). From this gel, we can draw several conclusions. First, a comparison of lanes 1–3 with 4–6 shows that GFP fusion with Smad 3 does not interfere with its receptor-dependent phosphorylation, DNA-binding, and Smad 4-complexing functions. The migration of the DNA–protein complexes is somewhat slower for the fusion protein (lanes 5 and 6) than the corresponding species containing the native Smad 3 (lanes 2 and 3), consistent with the extra GFP moiety in the fusion protein. Second, GFP-Smad 3 K43N/K44Q displayed the same profile of phosphorylation-dependent DNA binding and enhancement by Smad 4 (lanes 7–9), as did the GFP fusion with wild-type Smad 3, indicating that, except for nuclear translocation, it is fully functional in all other aspects of signaling.

Discussion

Smad NLS Sequences. We have identified a short basic motif (40KKLKK44) in the N-terminal domain of Smad 3 that is absolutely required for its nuclear import. If this motif is mutated by either substitutions or small deletions, Smad 3 fails to migrate to the nucleus in response to TGF-β stimulation. Consistent with this finding, whereas the wild-type Smad 3 MH1 domain is targeted to the nucleus, the MH1 domain of NLS-mutated Smad 3 does not localize exclusively in the nucleus and becomes mistargeted to the cytoplasm. Our conclusions are supported by both transient-transfection and stable-expression studies, so they are not cell-line specific or complicated by different levels of expression.

All four lysine residues in the NLS motif seem to be important for Smad 3 nuclear import (Fig. 4). Even though the substitution mutation K43N/K44Q resulted in relatively modest inhibition of nuclear localization of the isolated MH1 domain compared with the deletion mutant, it caused the same severe defect in TGF-β-induced nuclear translocation of the full-length Smad 3 protein. This indicates that ligand-dependent nuclear import is a finely tuned process that requires a higher degree of NLS integrity than does the import of isolated MH1 domains.

We used the criterion of complete GFP targeting to the nucleus to detect the presumptive NLS. Because GFP is a relatively small protein (molecular mass ≈27 kDa), it can freely diffuse into the nucleus, and this accounts for its even distribution between cytoplasm and nucleus. In contrast, the GFP fusion with the full-length Smad 3 has a size of ≈80 kDa, which greatly diminishes simple diffusion through nuclear pores. This leads to its relative exclusion from the nucleus in the basal state, even though high levels of overexpression can still force some of the protein to "spill" into the nucleus. The molecular masses of the MH1 and MH2 domain fusions to GFP are 42 and 48 kDa, respectively, so it is possible for these proteins to traverse nuclear pores by diffusion, although not as easily as GFP. Correspondingly, the unregulated state for both proteins should be an even dispersion throughout the cells, indicating the absence of either a NLS or a nuclear export signal. Because the Smad 3 MH1 domain displayed exclusive nuclear localization but the Smad 4 MH1 domain was everywhere in the cell, we conclude that Smad 3 contains a NLS motif in its MH1 domain that is not conserved in Smad 4.

Because our newly identified NLS motif is highly conserved in all pathway-specific Smads, it may also be responsible for ligand-stimulated nuclear import of Smads 1, 2, 5, 8, and 9. In unpublished studies, we found that the GFP fusion with an isolated Smad 1 MH1 domain is also predominantly localized to the nucleus, and mutations in its NLS also totally disrupted its nuclear enrichment (data not shown). It will be interesting to see whether these mutations also block nuclear translocation of intact Smad 1 in response to stimulation by BMP2 or 4. It is noteworthy that the NLSs from Smads 1, 5, 8, and 9 contain two more lysines than does the Smad 3 motif (Fig. 2) and thus are more similar to the classical T-antigen NLS.

In contrast, the Smad 4 NLS-like region harbors a drastic charge reversal—a change of the C-terminal Lys to Glu (KKLKK>KKLKE). Consequently, the Smad 4 MH1 domain does not accumulate in the nucleus (Fig. 1). The corresponding mutation in Smad 3 (K44E) eliminates ligand-induced nuclear import of intact Smad 3 protein (Fig. 4). Strikingly, conversion of the Glu back to Lys in the isolated Smad 4 MH1 motif, to

match the "consensus" NLS sequence, allows it to accumulate in the nucleus (Fig. 3, 6). In addition, the NLS-like motif in the Smad 4 MH1 domain is followed by two more acidic residues (45KKLKEKKDE53), probably further attenuating any NLS function. It would appear that Smad 4 and the invertebrate co-Smads are "designed" with a defective NLS such that they can translocate into the nucleus only after binding to a Smad protein with an active exposed NLS.

The SV40 large T-antigen NLS (P**KKKRK**VE) typifies the monopartite NLS consisting of one cluster of basic amino acids (bold), which is often preceded by a proline and followed by a hydrophobic or aliphatic amino acid and an acidic residue (underlined). A bipartite NLS, exemplified by nucleoplasmin, consists of two clusters of basic residues separated by a variable spacer region. Our identified Smad 3 NLS bears more resemblance to the SV40-type NLS, especially in light of the fact that it is also bounded on the C-terminal side by a hydrophobic and an acidic residue [40**KKLKK**TGRLD49 underlined (shown in Fig. 2 under dashed line)].

Conditional Nuclear Import of Smad 3. The SV40 T-antigen and nucleoplasmin are representatives of nuclear proteins that are constitutively imported after synthesis in the cytoplasm. Thus, their NLSs might represent "strong" signals leading to persistent nuclear retention. However, no paradigm has emerged to explain conditional nuclear translocation of proteins that localize to the cytosol in the resting state but are efficiently transported into the nucleus after stimulation by an extracellular signal. An NLS responsible for this more dynamic type of import may or may not be similar to the classical NLSs.

The nuclear entry of STAT1 represents a well-characterized example of conditional nuclear import. Like the Smad proteins, STAT1 resides in the cytoplasm in the resting state but is actively imported into nucleus after receptor activation and STAT phosphorylation (19). It was originally expected that a particular basic segment in STAT1 acts as the NLS to mediate the protein's recognition by the importin complex. However, mutagenesis studies showed that this region was dispensable for nuclear import, thus indicating that STAT1 does not possess a "classical" basic NLS (19). Because we have shown that a NLS-like basic motif is absolutely required for Smad 3 nuclear import, this suggests that different cytoplasmic transcription factors use divergent signals to translocate into the nucleus.

The MH1 and MH2 domains from the same Smad protein form intramolecular complexes under resting conditions (13). We speculate that ligand-induced C-terminal phosphorylation opens up the complex and makes the previously masked NLS motif accessible for nuclear import. A similar case has been made for NF-κB, a transcription factor involved in immediate early gene activation during immune responses (20).

Separation of DNA Binding and NLS Motifs. Our results also show that the Smad 3 DNA-binding domain and NLS motif are physically distinct and nonoverlapping, consistent with a previous study demonstrating that the NLS of the human vitamin D receptor is outside of its DNA-binding pocket (21). This suggests that the efficient nuclear accumulation of Smad3 is not simply because of its binding to DNA, which would shift the nucleus/cytosol equilibrium in favor of the nucleus. Indeed, the crystal structure of the Smad 3 MH1 domain bound to DNA revealed that an 11-residue β hairpin, from Leu^{71} to Gly^{82}, is responsible for DNA binding. In contrast, the NLS motif is at the highly conserved C-terminal part of helix-2 (Gln^{21} to Gly^{46}, named "basic helix"), one of the most solvent-exposed regions of the protein (16). This presumably would confer on the NLS motif a ready accessibility to cellular import apparatus.

Thus we have provided a functional definition of the nuclear translocation signal for Smad proteins. We show that an NLS-like basic motif within the MH1 domain (Lys^{40}-Lys-Leu-Lys-Lys^{44}) of Smad 3, highly conserved among all of the pathway-specific Smads, is crucial for nuclear import in response to ligand stimulation. Mutations in this motif disrupt its nuclear translocation but have no impact on its hetero-oligomerization with Smad 4 or DNA binding. Such mutants are dominant-negative inhibitors of TGF-β-induced transcription. We have also demonstrated that the inability of Smad 4 to translocate into nucleus without pathway-specific Smads is probably because of a single amino acid substitution within the identified motif.

We thank members of the Lodish group, especially Xianxin Hua and Bill Schiemann, for stimulating discussions. We also thank Drs. Paul Matsudaira and Yin Sun for critically reviewing the manuscript. This work was supported by National Institutes of Health (NIH) grant CA63260 to H.F.L., by a postdoctoral fellowship from the Damon Runyon–Walter Winchell Cancer Research foundation to Z.X., by a postdoctoral fellowship from the NIH, by the U. S. Army Breast Cancer Research Program to X.L., and by a grant from the Israel Science Foundation administered by the Israel Academy of Sciences and Humanities to Y.I.H.

1. Whitman, M. (1998) *Genes Dev.* **12,** 2445–2462.
2. Massague, J. (1998) *Annu. Rev. Biochem.* **67,** 753–791.
3. Lagna, G., Hata, A., Hemmati-Brivanlou, A. & Massague, J. (1996) *Nature (London)* **383,** 832–836.
4. Macias-Silva, M., Abdollah, S., Hoodless, P.-A., Pirone, R., Attisano, L. & Wrana, J.-L. (1996) *Cell* **87,** 1215–1224.
5. Liu, F., Pouponnot, C. & Massague, J. (1997) *Genes Dev.* **11,** 3157–3167.
6. Chen, X., Rubock, M.-J. & Whitman, M. (1996) *Nature (London)* **383,** 691–696.
7. Hua, X., Liu, X., Ansari, D. & Lodish, H.-F. (1998) *Genes Dev.* **12,** 3084–3095.
8. Hoodless, P. A., Haerry, T., Abdollah, S., Stapleton, M., O'Connor, M.-B., Attisano, L. & Wrana, J.-L. (1996) *Cell* **85,** 489–500.
9. Heldin, C. H., Miyazono, K. & ten Dijke, P. (1997) *Nature (London)* **390,** 465–471.
10. Mattaj, I. & Englmeier, L. (1998) *Annu. Rev. Biochem.* **67,** 265–306.
11. Görlich, D. & Kutay, U. (1999) *Annu. Rev. Cell Dev. Biol.* **15,** 607–660.
12. Baker, J.-C. & Harland, R.-M. (1996) *Genes Dev.* **10,** 1880–1889.
13. Hata, A., Lo, R., Wotton, D., Lagna, G. & Massague, J. (1997) *Nature (London)* **3;388,** 82–87.
14. Liu, X., Sun, Y., Constantinescu, S., Karam, E., Weinberg, R.-A. & Lodish, H.-F. (1997) *Proc. Natl. Acad. Sci. USA* **94,** 10669–10674.
15. Kain, S.-R., Adams, M., Kondepudi, A., Yang, T., Ward, W. & Kitts, P. (1995) *BioTechniques* **19,** 650–655.
16. Shi, Y., Wang, Y., Jayaraman, L., Yang, H., Massague, J. & Pavletich, N.-P. (1998) *Cell* **94,** 585–594.
17. Zawel, L., Dai, J., Buckhaults, P., Zhou, S., Kinzler, K.-W., Vogelstein, B. & Kern, S.-B. (1998) *Mol. Cell* **4,** 611–617.
18. Wieser, R., Wrana, J.-L. & Massague, J. (1995) *EMBO J.* **14,** 2199–2208.
19. Sekimoto, T., Imamoto, N., Nakajima, K., Hirano, T. & Yoneda, Y. (1997) *EMBO J.* **16,** 7067–7077.
20. Henkel, T., Zabel, U., van Zee, K., Muller, J.-M., Fanning, E. & Baeuerle, P.-A. (1992) *Cell* **68,** 1121–1133.
21. Michigami, T., Suga, A., Yamazaki, M., Shimizu, C., Cai, G., Okada, S. & Ozono, K. (1999) *J. Biol. Chem.* **274,** 33531–33538.