the N domain. Other tumour-suppressor mutations, including missense mutations in the C domains of Smad2 and Smad4, act by disrupting protein stability or effector function¹². Our findings reveal a mechanism of tumour suppressor inactivation which instead involves a gain of autoinhibitory function. Antagonists of SMAD autoinhibition might be useful in reversing the effects of this type of mutation. □

Methods

Construction of expression vectors. To generate human Smad4 and Smad2 mutations, a fragment of the corresponding cDNAs^{1,4} was amplified by PCR. The amplified region was subcloned into full-length Smad4 or Smad2 in pCMV5 for transfection into mammalian cells. The regions amplified by PCR and the presence of missense mutations were confirmed by sequencing.

Yeast two-hybrid system. LexA fusions were created in pBTM 116 (ref. 21) and GAD fusions within pGAD424 (Clontech). Interactions were tested in the strain L40 (ref. 22). Activation of the LexA operator–HIS3 reporter was assayed on medium lacking histidine with increasing concentrations of 3-amino-triazole.

Transfection, immunoprecipitation, immunoblot and metabolic labelling. For Smad2/Smad4 homo- or hetero-complex analysis, COS cells were transiently transfected with the indicated constructs and stimulated with 200 pM TGF- β 1 for 1 h. Cells were lysed in TNE buffer³, immunoprecipitated with anti-Flag M2 monoclonal antibody (IBI; Eastman Kodak), and interacting proteins detected by immunoblotting with the anti-HA monoclonal antibody 12CA5 (Boehringer Mannheim) as described³. Anti-SMAD rabbit polyclonal antibody was raised against full-length Smad1. To study the interaction between the N and C domains of Smad4 or Smad2, transiently transfected COS cells were lysed in LSLD buffer (50 mM HEPES, pH 7.4, 50 mM NaCl, 0.1% Tween 20, 10% glycerol, 1 mM DTT) containing protease and phosphatase inhibitors. Immunoprecipitation and immunoblotting were done as described. COS or R-1B/L17 cells transfected with the indicated constructs were labelled with ³⁵S-methionine or ³²P-orthophosphate and visualized by electrophoresis and autoradiography²³.

In vitro binding. The Smad2 N domain (amino acids 1–184) fused to GST and the Smad2 C domain (amino acids 248–467) protein were expressed in *Escherichia coli* and partially purified by column chromatography. GST fusion proteins were conjugated to glutathione–Sepharose beads (Pharmacia) and incubated with Smad2 C domain protein. Smad2 C domain protein was detected by western blotting using anti-Smad antibody.

Functional assays. For the animal cap assay, RNA (10 nl, 2 ng) was introduced in the animal pole of two-cell *Xenopus* embryos. Animal caps were explanted at the blastula stage and cultured to the tadpole stage. Total RNA from the explants and control sibling embryos was extracted and was amplified with reverse transcription by using muscle actin and *EF-1 α* primers²⁴. In the MDA-MB468 cell experiments, the amount of transfected plasmids was adjusted to render the TGF- β response dependent on both Smad2 and Smad4. Luciferase and growth inhibition were assayed as described³.

Received 24 March; accepted 6 May 1997.

- Hahn, S. A. *et al.* DPC4, a candidate tumor suppressor gene at human chromosome 18q21.1. *Science* **271**, 350–353 (1996).
- Eppert, K. *et al.* MADR2 maps to 18q21 and encodes a TGF- β -regulated MAD-related protein that is functionally mutated in colorectal carcinoma. *Cell* **86**, 543–552 (1996).
- Lagna, G., Hata, A., Hemmati-Brivanlou, A. & Massagué, J. Partnership between DPC4 and SMAD proteins in TGF- β signalling pathways. *Nature* **383**, 832–836 (1996).
- Liu, F. *et al.* A human Mad protein acting as a BMP-regulated transcriptional activator. *Nature* **381**, 620–623 (1996).
- Baker, J. & Harland, R. M. A novel mesoderm inducer, mMad2-2, functions in the activin signal transduction pathway. *Genes Dev.* **10**, 1880–1889 (1996).
- Massagué, J., Hata, A. & Liu, F. TGF- β signalling through the Smad pathway. *Trends Cell Biol.* **7**, 187–192 (1997).
- Zhang, Y., Feng, X.-H., Wu, R.-Y. & Derynck, R. Receptor-associated Mad homologues synergize as effectors of the TGF- β response. *Nature* **383**, 168–172 (1996).
- Hoodless, P. A. *et al.* MADR1, a MAD-related protein that functions in BMP2 signaling pathways. *Cell* **85**, 489–500 (1996).
- Macías-Silva, M. *et al.* MADR2 is a substrate of the TGF β receptor and phosphorylation is required for nuclear accumulation and signaling. *Cell* **87**, 1215–1224 (1996).
- Kretschmar, M., Liu, F., Hata, A., Doody, J. & Massagué, J. The TGF- β family mediator Smad1 is phosphorylated directly and activated functionally by the BMP receptor kinase. *Genes Dev.* **11**, 984–995 (1997).
- Wu, R.-Y., Zhang, Y., Feng, X.-H. & Derynck, R. Heteromeric and homomeric interactions correlated with signaling activity and functional cooperativity of Smad3 and Smad4/DPC4. *Mol. Cell. Biol.* **17**, 2521–2528 (1997).

- Shi, Y., Hata, A., Lo, R. S., Massagué, J. & Pavletich, N. P. A structural basis for mutational inactivation of the tumour suppressor Smad4. *Nature* **388**, 87–93 (1997).
- Cárcamo, J. *et al.* Type I receptors specify growth inhibitory and transcriptional responses to TGF- β and activin. *Mol. Cell. Biol.* **14**, 3810–3821 (1994).
- Schutte, M. *et al.* DPC4 gene in various tumor types. *Cancer Res.* **56**, 2527–2530 (1996).
- Kim, S. K. *et al.* DPC4, a candidate tumor suppressor gene, is altered infrequently in head and neck squamous cell carcinoma. *Cancer Res.* **56**, 2519–2521 (1996).
- Nagatake, M. *et al.* Somatic *in vivo* alterations of the DPC4 gene at 18q21 in human lung cancers. *Cancer Res.* **56**, 2718–2720 (1996).
- Barrett, M. T., Schutte, M., Kern, S. E. & Reid, B. J. Allelic loss and mutational analysis of the DPC4 gene in esophageal adenocarcinoma. *Cancer Res.* **56**, 4351–4353 (1996).
- Riggins, G. J. *et al.* Mad-related genes in the human. *Nature Genet.* **13**, 347–349 (1996).
- Uchida, K. *et al.* Somatic *in vivo* alterations of the JV18-1 gene at 18q21 in human lung cancers. *Cancer Res.* **56**, 5583–5585 (1996).
- Graff, J. M., Bansal, A. & Melton, D. A. *Xenopus* Mad proteins transduce distinct subsets of signals for the TGF β superfamily. *Cell* **85**, 479–487 (1996).
- Bartel, P. L., Chien, C.-T., Sternglanz, R. & Fields, S. in *Cellular Interactions in Development: A Practical Approach* (ed. Hartley, D. A.) 153 (Oxford University Press, Oxford, 1993).
- Vojtek, A. B., Hollenberg, S. M. & Cooper, J. A. Mammalian Ras interacts directly with the serine/threonine kinase Raf. *Cell* **74**, 205–214 (1993).
- Wrana, J. L., Attisano, L., Wieser, R., Ventura, F. & Massagué, J. Mechanism of activation of the TGF- β receptor. *Nature* **370**, 341–347 (1994).
- Hemmati-Brivanlou, A. & Melton, D. A. Inhibition of activin receptor signalling promotes neuralization in *Xenopus*. *Cell* **77**, 273–281 (1994).

Acknowledgements. We thank Y. Shi for recombinant Smad2 proteins, and I. Reynisdóttir, J. Doody and S. Lee for advice and technical assistance. G.L. thanks A. Hemmati-Brivanlou for support and advice. This work was supported by NIH Breast Spore and Cancer Center grants. D.W. is the recipient of a postdoctoral fellowship from the Human Frontier Science Program. A.H. is a research associate and J.M. an investigator of the Howard Hughes Medical Institute.

Correspondence and requests for materials should be addressed to J.M. (e-mail: j-massague@ski.mskcc.org).

A structural basis for mutational inactivation of the tumour suppressor Smad4

Yigong Shi, Akiko Hata*, Roger S. Lo*, Joan Massagué* & Nikola P. Pavletich

Cellular Biochemistry and Biophysics Program, *Cell Biology Program and the Howard Hughes Medical Institute, Memorial Sloan-Kettering Cancer Center, New York, New York 10021, USA

The Smad4/DPC4 tumour suppressor¹ is inactivated in nearly half of pancreatic carcinomas² and to a lesser extent in a variety of other cancers^{3–4}. Smad4/DPC4, and the related tumour suppressor Smad2, belong to the SMAD family of proteins that mediate signalling by the TGF- β /activin/BMP-2/4 cytokine superfamily from receptor Ser/Thr protein kinases at the cell surface to the nucleus^{5–7}. SMAD proteins, which are phosphorylated by the activated receptor, propagate the signal, in part, through homo- and hetero-oligomeric interactions^{8–13}. Smad4/DPC4 plays a central role as it is the shared hetero-oligomerization partner of the other SMADs. The conserved carboxy-terminal domains of SMADs are sufficient for inducing most of the ligand-specific effects, and are the primary targets of tumorigenic inactivation. We now describe the crystal structure of the C-terminal domain (CTD) of the Smad4/DPC4 tumour suppressor, determined at 2.5 Å resolution. The structure reveals that the Smad4/DPC4 CTD forms a crystallographic trimer through a conserved protein–protein interface, to which the majority of the tumour-derived missense mutations map. These mutations disrupt homo-oligomerization *in vitro* and *in vivo*, indicating that the trimeric assembly of the Smad4/DPC4 CTD is critical for signalling and is disrupted by tumorigenic mutations.

The conserved C-terminal domain can mediate many of the biological effects of SMAD proteins, and the conserved N-terminal domain can negatively regulate the SMAD activity^{14,15}. When over-expressed in a Smad4/DPC4^{–/–} cell line, the Smad4/DPC4 CTD can activate the transcription of TGF- β responsive genes and result in

Figure 1 The structure of the Smad4/DPC4 CTD consists of a β -sandwich with a three-helix bundle on one end and a collection of three large loops and an α -helix on the other. The view is along the edge of the β -sandwich; the dotted line represents the disordered region between the H3 and H4 helices. Figures were prepared with the programs MOLSCRIPT²⁶ and RASTER3D²⁷.

Figure 2 SMAD C-terminal domains are highly conserved and are targeted by tumorigenic and developmental mutations. **a**, Sequence alignment of C-terminal domains of five human SMAD proteins^{18,19} (Smad 1, 2, 3, 5 and Smad4/DPC4) and homologues from *Drosophila*¹⁸ (Mad) and *C. elegans*¹⁹ (Sma-2, 3, 4), with the Smad4/DPC4 CTD secondary structure elements indicated below the sequences. Residues that are more than 40% exposed to solvent have no significant role in the structure and are conserved in at least 6 out of the 9 aligned sequences are highlighted in cyan. The 14 missense mutations shown above the alignment include tumour-derived Smad4/DPC4 and Smad2 mutations^{1,2,4,12,17,28} (yellow), as well as mutations from *Drosophila* and *C. elegans* genetic screens^{18,19} (developmental mutations, in green). The residues at which these mutations occur are in bold face and underlined. **b**, Mapping of the missense mutations and the highly conserved and solvent-exposed residues identifies the three-helix bundle and the three-loop/helix region as regions likely to be important for the macromolecular recognition that mediates SMAD function. Colour coding is the same as in **a**. The amino-acid substitution and the residue number of the mutation in SMAD family members other than Smad4/DPC4 are shown in parentheses. The three structural mutations (Arg441Pro in Smad4/DPC4, Leu440Arg and Pro445H in Smad2) are not shown.

Table 1 Statistics from the crystallographic analysis

Data set		Native 1 (8 °C)	Native 2 (−170 °C)	Thimerosal	UO ₂ (OAc) ₂ *	(CH ₃) ₃ PbOAc*		
Resolution (Å)		3.0	2.5	3.0	3.0	3.2		
Observations		30,691	39,125	30,572	23,488	25,150		
Unique reflections		7,189	11,496	7,073	6,765	5,759		
Data coverage (%)		96.8	96.5	96.8	92.8	94.6		
<i>R</i> _{sym} (%)†		6.5	3.7	4.8	8.5	10.0		
MIR analysis (20.0–3.2 Å):								
Mean isomorphous difference‡				0.18	0.14	0.24		
Phasing power§				2.54	1.38	1.03		
Refinement statistics:					R.m.s.d.#			
Resolution (Å)	Reflections (<i>F</i> > 2σ)	Protein atoms	Waters atoms	<i>R</i> -factor (%)	<i>R</i> -free¶ (%)	Bonds (Å)	Angles (°)	<i>B</i> -factor (Å ²)
7.0–2.5	10,359	1,522	129	20.9	28.6	0.010	1.66	3.29

* OAc, acetate.

† $R_{sym} = \sum_i \sum_h |I_{h,i} - \bar{I}_h| / \sum_i \sum_h I_{h,i}$ for the intensity (I) of i observations of reflection h .

‡ Mean isomorphous difference = $\sum |F_{PH} - F_P| / \sum F_{PH}$, where F_{PH} and F_P are the derivative and native structure factors, respectively.

§ Phasing power = $[(F_{calc})^2 / (F_{PH(obs)} - F_{PH(calc)})^2]^{1/2}$.

|| R -factor = $\sum |F_{obs} - F_{calc}| / \sum |F_{obs}|$, where F_{obs} and F_{calc} are the observed and calculated structure factors, respectively.

¶ R -free is the R -factor calculated using 5% of the reflection data chosen randomly and omitted from the start of refinement.

R.m.s.d., root-mean-square deviations from ideal geometry and root-mean-square variation in the B -factor of bonded atoms.

growth arrest in a ligand-independent manner, paralleling the effects of the TGF- β ligand⁹; microinjection of messenger RNAs encoding the CTD of Smad2 into *Xenopus* embryos can induce a mesoderm response that mimics the effects of the full-length protein¹⁶; and the Smad4/DPC4 CTD fused to a heterologous DNA-binding domain can activate gene expression from a reporter construct¹⁴. Consistent with the SMAD C-terminal domain being the main effector domain, the majority (10 out of 13) of tumorigenic missense mutations in Smad4/DPC4 and Smad2, as well as mutations isolated from *Drosophila* and *Caenorhabditis elegans* genetic screens map to the C-terminal domain. To investigate how the SMAD C-terminal domain functions in mediating TGF- β signalling and how its mutation in cancer inactivates the pathway, we determined the crystal structure of the 234-amino-acid Smad4/DPC4 CTD (residues 319–552) at 2.5 Å resolution (Table 1).

The structure consists of a β -sandwich with twisted antiparallel β -sheets of five and six strands each (Fig. 1). One end of the β -sandwich is capped by a three- α -helix bundle (H3, H4 and H5 helices) that extends over the plane of the six-stranded β -sheet, at a roughly perpendicular angle; the other end of the β -sandwich is capped by a group of three large loops and an α -helix (L1, L2, L3 loops and H1 helix; Fig. 1). To simplify the presentation, the three large loops and α -helix, as well as portions of β -strands in their immediate vicinity, will be collectively referred to as the loop/helix region. The three α -helices of the bundle pack in an up-down-up orientation primarily through leucine residues. In between the H3 and H4 helices, a 34-amino-acid sequence that is rich in Ala (39%), Gly and Pro residues and is present only in Smad4/DPC4 and its *C. elegans* homologue Sma-4, is disordered in the crystals (residues 457–491). In the loop/helix region, the L1, L2 and L3 loops of 7, 9 and 18 residues, respectively, and the H1 helix are mostly polar and pack through extended hydrogen-bond networks. These hydrogen bonds are likely to contribute to the rigid structure of this region that is suggested by the well defined electron density.

SMAD proteins are highly conserved within the family and across species, with Smad4/DPC4 and its *C. elegans* homologue, Sma-4, representing a somewhat divergent subtype which still retains about 40% identity with other family members^{5,7} (Fig. 2a). Many of the conserved residues have structural roles. These include the hydrophobic residues that make up the hydrophobic core of the β -sandwich and of the three-helix bundle, as well as many of the polar residues that form the hydrogen-bond networks important for the structure of the loop/helix region. Examples of the latter group are the invariant Arg 372 and Arg 380 residues from the H1 helix that make four and three charge-stabilized hydrogen bonds,

respectively. Many other highly conserved residues are exposed to solvent and do not appear to stabilize the structure. They are thus candidates for functional residues that may mediate macromolecular interactions important for the function of SMAD proteins. The structure reveals that these candidate functional residues (highlighted in Fig. 2b) have a strong tendency to cluster at the loop/helix region and the three-helix bundle.

Another indication that the loop/helix region and the three-helix bundle are functionally important comes from an analysis of nine tumour-derived missense mutations, some occurring several times, in the CTDs of the Smad4/DPC4 and Smad2 tumour suppressors. Excluding three mutations that map to structural residues, five of the six tumour-derived missense mutations map to either the loop/helix region or to the three-helix bundle: the Smad4/DPC4 mutations Asp351His², Arg361Cys¹⁷ and Val370Asp¹⁷ map to the loop/helix region, whereas the Smad4/DPC4 mutation Asp493His¹ and the Smad2 mutation Asp450Glu¹² (corresponding to Asp 537 of Smad4/DPC4) map to the three-helix bundle. We presume that these mutations deprive the CTD of critical intermolecular contacts. The one mutation that does not map to either region is Arg420His from Smad4/DPC4, which instead maps to the side of the β -sandwich (H2 helix), a region that is not as well conserved. The remaining three mutations map to structural residues: the Smad2 Leu440Arg mutation (corresponding to Ile 527 of Smad4/DPC4) in the hydrophobic core of the β -sandwich probably disrupts packing in the hydrophobic core; the Smad4/DPC4 Arg441Pro mutation at the three-helix bundle probably disrupts the H3 helix because of the introduction of a proline in the middle of the helix; and the Smad2 Pro445His mutation (corresponding to Ala 532 in Smad4/DPC4), also at the three-helix bundle, probably disrupts the packing between the three-helix bundle and the β -sandwich as there is little space for the larger histidine side chain in this part of the hydrophobic core.

In support of the functional significance of the loop/helix region, mutations in *Drosophila* and *C. elegans* in this region produce null or severe developmental phenotypes^{18,19}. These mutations map to Gly 508 (*Drosophila* Mad, *C. elegans* Sma-2), Gly 510 (Sma-3), and Glu 520 (Mad) of the L3 loop in the loop/helix region (Fig. 2). Thus the location of conserved, solvent-exposed residues and of mutations derived from tumours or from *Drosophila* and *C. elegans* genetic screens together indicate that the loop/helix region and the three-helix bundle are critical for mediating Smad activity.

As SMAD CTDs can mediate most of the effects of the full-length proteins, we tested the Smad4/DPC4 CTD for homo-oligomerization activity. Co-immunoprecipitation using extracts from COS

cells transfected with differentially tagged Smad4/DPC4 CTD constructs showed that the Smad4/DPC4 CTD could still form homo-oligomers when overexpressed¹⁵ (Fig. 3d), suggesting that the CTD may contain a primary homo-oligomerization activity. However, full-length Smad4/DPC4 homo-oligomers are more stable than Smad4/DPC4 CTD homo-oligomers *in vivo*¹⁵, indicating that residues N-terminal to the Smad4/DPC4 CTD may contribute to homo-oligomerization.

To investigate the homo-oligomerization activity of the Smad4/DPC4 CTD further, we examined the crystal packing of the Smad4/DPC4 CTD molecules and identified a trimer that formed through three identical extended protein–protein interfaces, burying 4,800 Å² of surface area (Fig. 3a). Each interface forms through

the interactions of the highly conserved regions of the Smad4/DPC4 CTD that contain most of the candidate functional residues: the loop/helix region of one subunit packs extensively with the three-helix bundle from another subunit while making a few additional contacts to residues from the β-sandwich (Fig. 3a). The only portion of the loop/helix region that does not participate in this interface is the L3 loop.

The trimer interface includes the majority of the conserved residues and the tumour-derived non-structural missense mutations (five out of six). Most noteworthy is an extended intermolecular hydrogen-bond network involving, from one subunit, the Arg 361 and Asp 351 side chains and two backbone amide groups of the loop/helix region, and from another subunit, the Asp 537 side

Figure 3 In the crystals, the Smad4/DPC4 CTD forms a trimer that is disrupted by tumorigenic mutations and is probably important for SMAD function. **a**, The three monomers (red, blue and purple) pack across three identical protein–protein interfaces. Tumour-derived missense mutations map to five amino acids (yellow) that are involved in intermolecular contacts. **b**, Close-up of an intermolecular hydrogen-bond network involving three residues, all of which have been found mutated in cancer. Colouring is as in **a**. **c**, Close-up showing the intermolecular packing of Val 370, mutated to Asp in colon tumours, against Phe 329, Trp 524 and the aliphatic portion of Lys 519. The subunit in which Val 370 is shown is in space-filling representation, whereas the other subunit is shown as a molecular surface (red mesh). Other intermolecular interactions not mentioned in the text include: van der Waals contacts between the L1 loop of the loop/helix region and the H4 and H5 helices of the three-helix bundle (Tyr 353, Val 354 and Pro 356 wedging in between His 530, Leu 533, Leu 536, Leu 540 and His 541); the hydrogen-bond networks between Ser 368 of the L2 loop and Arg 496, Glu 526 and His 528 of the

β-sheet, and between His 371 of the L2 loop and Asp 332 of the β-sheet. The figure was prepared with the program GRASP²⁹. **d**, *In vivo*, tumour-derived trimer interface mutations (D351H, R361C, V730D, D537E) disrupt both homo- and hetero-oligomerization, whereas a developmental mutation in the L3 loop (G508S) disrupts only hetero-oligomerization. To assay for homo-oligomerization, mammalian COS-1 cells were transiently transfected with Flag-tagged wild-type Smad4/DPC4 CTD and HA-tagged WT or mutant constructs. For hetero-oligomerization, cells were transfected with Flag-tagged Smad2 C-terminal domain and HA-tagged Smad4/DPC4 C-terminal domain WT or mutant constructs together with a constitutively active TGF-β type I receptor construct. The cell lysate was immunoprecipitated with anti-Flag antibody and subsequently immunoblotted using anti-HA antibody. Immunoblots indicated that the mutant Smad4/DPC4 CTDs were expressed at levels comparable to those of the wild-type constructs (data not shown). Lower panels, full-length (FL) proteins were treated similarly.

chain of the three-helix bundle (Fig. 3b). The Asp351, Arg 361, and Asp 537 residues are essentially invariant, apart from a conservative Arg-to-Lys substitution in Smad2 (Fig. 2a), and all three are mutated in cancer (Fig. 2): Asp351His and Arg361Cys mutations have been found in Smad4/DPC4 in ovarian² and colon cancers¹⁷, respectively, and the Asp450Glu mutation, corresponding to Asp 537 of Smad4/DPC4, has been isolated from Smad2 in colon cancer¹². These mutations all disrupt the intricate hydrogen-bond network at the interface. Also noteworthy are the intermolecular van der Waals contacts between Val 370 on the L2 loop of the loop/helix and Trp 524, Phe 329, and the aliphatic portion of the Lys 519 side chain on the β -sheet at the base of the three-helix bundle (Fig. 3c). The two aromatic residues are invariant, except for a conservative Tyr-to-Phe substitution in Smad4/DPC4 (Fig. 2a). Also, Val 370 is mutated to Asp in colon cancer¹⁷, causing the hydrophobic portion of the trimer interface to be destabilized by the charged amino acid. The Smad4/DPC4 Asp493His mutation found in pancreatic tumours¹ also maps to the trimer interface (Fig. 3a) and may interfere with the electrostatic packing of Asp 493 of one subunit with Arg 496 and Arg 497 of another subunit at the trimer interface. Asp 493 is near the disordered region of the H4 helix and its interaction with the arginine residues is not well defined.

Many of the other trimer–interface contacts are also conserved in the SMAD family (Fig. 3c legend), suggesting that other SMAD CTDs may form a similar trimeric structure. Not all residues in the Smad4/DPC4 CTD trimer interface are conserved in all SMAD, and those that differ may contribute to subtype specificity. An example is an intermolecular hydrogen-bond contact between His 371 and Asp 332: this pair is conserved in the *C. elegans* Smad4/DPC4 homologue Sma4, but is an invariant Asn–Asn pair in the pathway-restricted SMADs (Fig. 2).

If the trimeric Smad4/DPC4 CTD assembly in the crystals is part of the *in vivo* homo-oligomer, then mutations of residues that make intermolecular contacts at the interface, particularly tumour-derived mutations, should disrupt or reduce homo-oligomerization *in vivo*. Figure 3d shows the results of co-immunoprecipitation experiments using extracts from COS cells transfected with differentially tagged mutant Smad4/DPC4 molecules. All four of the tumorigenic mutations that are important at the trimer interface, Asp 351, Arg 361, Val 370 and Asp 537, disrupt homo-oligomerization of the Smad4/DPC4 CTD and of full-length Smad4/DPC4 (Fig. 3d). Conversely, the *Drosophila/C. elegans* developmental mutation Gly508Ser (Fig. 2a) had no effect on homo-oligomerization (Fig.

3d). This mutation maps to the L3 loop, which is the only portion of the loop/helix region not involved in the trimer interface.

If Smad4/DPC4 CTD forms a trimer, then so should full-length Smad4/DPC4. Recombinant full-length Smad4/DPC4, purified to near homogeneity, elutes from a gel-filtration column with an apparent relative molecular mass (M_r) of ~ 180 K, consistent with the calculated size (181K) of the Smad4/DPC4 trimer (Fig. 4a). This large M_r is probably due to trimerization, because the M_r is reduced by a factor of ~ 3 in the tumour-derived trimer-interface mutants (Fig. 4b). Conversely, the *Drosophila/C. elegans* developmental mutation Gly508Ser in the L3 loop has no effect on the large M_r of Smad4/DPC4 (Fig. 4b). Note that the Smad4/DPC4 CTD elutes as a monomer from a gel-filtration column, which is consistent with residues N-terminal to the CTD contributing to homo-oligomerization¹⁵.

In principle, the full-length Smad4/DPC4 protein could assume oligomeric states other than a trimer but have a gel-filtration mobility approximating that of a trimer. However, our *in vivo* and *in vitro* results with the trimer-interface mutants, both with the CTD and the full-length proteins, indicates that the trimeric protein–protein interface in the crystals also participates in homo-oligomerization *in vivo*.

The Smad4/DPC4 CTD also supports hetero-oligomerization, as shown by the co-immunoprecipitation of overexpressed Smad4/DPC4 CTD and Smad2 CTD from COS cells¹⁵ (Fig. 3d), and by their association on electrophoresis in non-denaturing gels (data not shown). Furthermore, both the tumour-derived trimer-interface mutants and the L3-loop mutants abolished hetero-oligomerization between the CTDs of Smad4/DPC4 and Smad2 (Fig. 3d); results were similar with full-length Smad4/DPC4. As the L3-loop developmental mutation, which does not significantly affect homo-oligomerization, disrupts hetero-oligomer formation, the L3 loop may participate in hetero-oligomerization. Also, mutations preventing homo-oligomerization that also disrupt hetero-oligomerization indicate that the former is a prerequisite for the latter.

Although several models of hetero-oligomerization could explain our results, one that is suitable from a structural perspective is the formation of a heterohexamer between Smad4/DPC4 and Smad2 trimers. The trimer structure resembles a disk, with the L3 loops forming undulations on the face of the disk (Fig. 5a), so two disks could fit together face to face through their L3 loops (Fig. 5b), explaining why L3-loop mutations disrupt hetero-oligomerization. In this model, heterohexamer formation would

Figure 4 Size-exclusion chromatography showing that wild-type full-length Smad4/DPC4, but not tumour-derived mutants, has an apparent M_r consistent with that of a trimer. **a**, Recombinant Smad4/DPC4 protein, purified to near homogeneity, was applied to a Superdex-200 gel-filtration column, and it eluted at M_r 180K. Fractions were visualized with Coomassie blue staining. **b**, *In vitro*,

tumour-derived trimer-interface mutations disrupt homo-oligomerization, but a developmental mutation in the L3 loop does not. Gel-filtration fractions of partially purified wild-type and mutant Smad4/DPC4 proteins were analysed by immunoblotting with anti-Smad4/DPC4 antibody.

Figure 5 One face of the disk-like trimer structure may mediate hetero-oligomerization. **a**, Mutations outside the trimer interface map primarily to L3-loop residues, with the exception of Arg 420, which is outside the L3 loop. The face of the trimer shown is opposite to that shown in Fig. 3a. **b**, Model of hetero-oligomer formation depicting the Smad4/DPC4 and Smad2 CTD trimers as disks. The approximate positions of the Smad4/DPC4 L3 loops and of the Smad2 sites that get phosphorylated by the receptor kinase³⁰ are shown in yellow and green, respectively.

depend on homotrimer formation, explaining how tumorigenic mutations that disrupt homo-oligomerization can prevent the formation of a functional hetero-oligomeric complex and interfere with signal transduction. □

Methods

Protein expression and purification. Recombinant Smad4/DPC4 CTD, corresponding to residues 319–552, was overexpressed at room temperature in *Escherichia coli* using a pET vector (Novagen). The Smad4/DPC4 CTD in the soluble fraction of the *E. coli* lysate was partially purified on a Q-Sepharose column, concentrated by ultrafiltration, and further purified by gel-filtration chromatography (on a Superdex-75 column) and anion-exchange chromatography (Source-15Q column).

Crystallization. Crystals were initially grown at 4°C by the hanging-drop vapour-diffusion method by mixing the protein solution (10–15 mg ml⁻¹) with an equal volume of reservoir solution containing 100 mM MES, 25% monomethylether PEG5000 (mPEG5000) and 200 mM (NH₄)₂SO₄, pH 6.5.

Crystals suitable for diffraction were grown using streak seeding and macroseeding²⁰. The crystals form in the cubic space group *F*₄32, with *a* = *b* = *c* = 199.6 Å, and contain one molecule in the asymmetric unit.

Data collection and processing. Diffraction data were collected using an R-AXISIIIC imaging plate detector mounted on a Rigaku 200HB generator. Native 1 and derivative data were collected at 8°C, and native 2 data were collected at -170°C with a crystal flash-frozen in buffer containing 20% glycerol and 25% mPEG5000. Heavy-atom soaks were done in 50 mM HEPES, 25% mPEG, 160 mM (NH₄)₂SO₄, 100 mM NaCl, pH 6.1, containing one of the following heavy-atom solutions: 1.2 mM thimerosal (for 12 h), 3.0 mM (CH₃)₃PbOOCCH₃ (for 3 d), or 2.0 mM uranyl acetate (for 19 h).

MIR analysis, model building and refinement. The heavy-atom sites of the thimerosal derivative were determined by direct methods with the program SHELXS-90 (ref. 21), and the heavy-atom sites of the other derivatives were identified by difference Fourier methods. Initial MIR phases calculated with the program MLPHARE²² had a mean figure of merit of 0.62 to 3.2 Å, and were improved with solvent flattening and histogram matching with the program SQUASH²³. The MIR electron density maps had continuous electron density for most of the Smad4/DPC4 CTD polypeptide, apart from a 34-amino-acid region between helices H3 and H4. A model was built into MIR electron density maps with program O (ref. 24), refined by simulated annealing with the program X-PLOR²⁵, and checked by calculating X-PLOR omit maps in which 5–7% of the structure was deleted in each calculation; simulated annealing was used to reduce model bias. The refined model contains residues 319–543 of human Smad4/DPC4 and 129 water molecules. Residues 544–552 at the C terminus and residues 457–491 between helices H3 and H4 have no electron density in the maps—we presume that these regions are disordered in the crystals.

In vivo oligomerization assays. Full-length Smad4/DPC4 and Smad2, Smad4/DPC4 CTD encoding amino acids 294–552, and the Smad2 CTD encoding amino acids 248–467 were subcloned into the mammalian expression vector pCMV5. All Smad4/DPC4 point mutations were generated by a polymerase chain reaction (PCR)-based method and were confirmed by DNA sequencing. Mammalian COS-1 cells were transiently transfected with the indicated Flag- and HA-tagged constructs by the DEAE-dextran method. Two days after transfection, cells were incubated with 200 pM TGF-β1 for one hour. Cells were lysed, the lysate immunoprecipitated and immunoblotted as described⁹. Wash buffers contained 150 mM NaCl for all immunoprecipitations, except for the homo-oligomerization assay of full-length wild-type (WT) Smad4/DPC4 and point mutants, in which 250 mM NaCl gave better differentiation of WT and mutant activities.

In vitro oligomerization assays. Full-length Smad4/DPC4 proteins, both wild-type and point mutants, were overexpressed at room temperature in *E. coli* using a pET vector (Novagen). Smad4/DPC4 protein in the soluble fraction of the *E. coli* lysate was partially purified by ion-exchange chromatography on Q-Sepharose and by gel filtration on Superdex-200 in 50 mM Tris, 200 mM NaCl, 5 mM DTT, pH 8.0. Aliquots from fractions corresponding to *M*_r standards between 440K and 25K were taken for immunoblotting with a rabbit polyclonal antibody raised against the Smad4/DPC4 CTD. Results were visualized by western blot analysis and ECL detection (Amersham). WT full-length Smad4/DPC4 was also cloned as a glutathione *S*-transferase fusion protein and purified to near homogeneity over a glutathione column; the fusion protein was then cleaved with thrombin and the Smad4/DPC4 protein further purified by anion-exchange chromatography (Source-15Q column).

Received 22 January; accepted 8 May 1997.

- Hahn, S. A. et al. DPC4, a candidate tumor suppressor gene at human chromosome 18q21.1. *Science* **271**, 350–353 (1996).
- Schutte, M. et al. DPC4 gene in various tumor types. *Cancer Res.* **56**, 2527–2530 (1996).
- Kim, S. K. et al. DPC4, a candidate tumor suppressor gene, is altered infrequently in head and neck squamous cell carcinoma. *Cancer Res.* **56**, 2519–2521 (1996).
- Nagatake, M. et al. Somatic *in vivo* alterations of the DPC4 gene at 18q21 in human lung cancers. *Cancer Res.* **56**, 2718–2720 (1996).
- Massague, J. TGF-β signalling: receptors, transducers, and mad proteins. *Cell* **85**, 947–950 (1996).
- Derynck, R. et al. Nomenclature: vertebrate mediators of TGFβ family signals. *Cell* **87**, 173 (1996).
- Derynck, R. & Zhang, Y. Intracellular signalling: the Mad way to do it. *Curr. Biol.* **6**, 1226–1229 (1996).
- Zhang, Y., Feng, X.-H., Wu, R.-Y. & Derynck, R. Receptor-associated Mad homologues synergize as effectors of the TGF-β response. *Nature* **383**, 168–172 (1996).
- Lagna, G., Hata, A., Hemmati-Brivanlou, A. & Massague, J. Partnership between SPC4 and SMAD proteins in TGF-β signalling pathways. *Nature* **383**, 832–836 (1996).
- Yingling, J. M. et al. Mammalian dwarfins are phosphorylated in response to transforming growth factor β and are implicated in control of cell growth. *Proc. Natl Acad. Sci. USA* **93**, 8940–8944 (1996).

11. Hoodless, P. A. *et al.* MADR1, a MAD-related protein that functions in BMP2 signaling pathways. *Cell* **85**, 489–500 (1996).
12. Eppert, K. *et al.* MADR2 maps to 18q21 and encodes a TGF β -regulated MAD-related protein that is functionally mutated in colorectal carcinoma. *Cell* **86**, 543–552 (1996).
13. Lechleider, R. J., de Caestecker, M. P., Dehejia, A., Polymeropoulos, M. H. & Roberts, A. B. Serine phosphorylation, chromosomal localization, and transforming growth factor- β signal transduction by human bsp-1. *J. Biol. Chem.* **271**, 17617–17620 (1996).
14. Liu, F. *et al.* A human Mad protein acting as a BMP-regulated transcriptional activator. *Nature* **381**, 620–623 (1996).
15. Hata, A., Lo, R. S., Wotton, D. & Massagué, J. Mutations increasing autoinhibition inactivate the tumour suppressors Smad2 and Smad4. *Nature* **388**, 82–87 (1997).
16. Baker, J. C. & Harland, R. M. A novel mesoderm inducer, Madr2, functions in the activin signal transduction pathway. *Genes Dev.* **10**, 1880–1889 (1996).
17. Thiagalingam, S. *et al.* Evaluation of candidate tumour suppressor genes on the chromosome 18 in colorectal cancers. *Nature Genet.* **13**, 343–346 (1996).
18. Sekelsky, J. J., Newfield, S. J., Raftery, L. A., Chartoff, E. H. & Gelbart, W. M. Genetic characterization and cloning of *Mothers against dpp*, a gene required for decapentaplegic function in *Drosophila melanogaster*. *Genetics* **139**, 1347–1358 (1995).
19. Savage, C. *et al.* *Caenorhabditis elegans* genes *smad-2*, *smad-3* and *smad-4* define a conserved family of transforming growth factor β pathway components. *Proc. Natl Acad. Sci. USA* **93**, 790–794 (1996).
20. Stura, E. A. & Wilson, I. A. In *Crystallization of Nucleic Acids and Proteins* (eds Ducruix, A. & Giegé, R.) 99–126 (Oxford Univ. Press, 1992).
21. Sheldrick, G. in *Patterson Interpretation and the Use of Macromolecular Delta-F Data* (Daresbury, UK, 1991).
22. Collaborative Computational Project, B. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
23. Zhang, K. Y. J. SQUASH-combining constraints for macromolecular phase refinement and extension. *Acta Crystallogr. D* **49**, 213–222 (1993).
24. Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–110 (1991).
25. Brunger, A. T. *X-PLOR, a system for crystallography and NMR* (Yale Univ. Press, New Haven, CT, 1991).
26. Kluwe, P. J. Molscript: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24**, 946–950 (1991).
27. Merritt, E. A. & Murphy, M. E. Raster3D Version 2.0: a program for photorealistic molecular graphics. *Acta Crystallogr. D* **50**, 869–873 (1994).
28. Uchida, K. *et al.* Somatic *in vivo* alterations of the JV18-1 gene at 18q21 in human lung cancers. *Cancer Res.* **56**, 5583–5585 (1996).
29. Nicholls, A., Sharp, K. A. & Honig, B. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins Struct. Funct. Genet.* **11**, 281–296 (1991).
30. Macias-Silva, M. *et al.* MADR2 is a substrate of the TGF β receptor and its phosphorylation is required for nuclear accumulation and signalling. *Cell* **87**, 1215–1224 (1996).

Acknowledgements. This work was supported by the NIH, the Pew Charitable Trusts, the Arnold and Mabel Beckman Foundation, the Dewitt Wallace Foundation, the Samuel and May Rudin Foundation, and the Howard Hughes Medical Institute.

Correspondence and requests for materials should be addressed to N.P.P. (e-mail: nikola@xray2.mskcc.org). Coordinates are being deposited with the Brookhaven Protein Data Bank.

DNA-replication checkpoint control at the *Drosophila* midblastula transition

Ody C. M. Sibon, Victoria A. Stevenson & William E. Theurkauf

Department of Biochemistry and Cell Biology and the Institute for Cell and Developmental Biology, State University of New York at Stony Brook, Stony Brook, New York 11794-5215, USA

Embryogenesis is typically initiated by a series of rapid mitotic divisions that are under maternal genetic control¹. The switch to zygotic control of embryogenesis at the midblastula transition is accompanied by significant increases in cell-cycle length and gene transcription, and changes in embryo morphology^{2,3}. Here we show that mutations in the *grapes* (*grp*) checkpoint 1 kinase homologue⁴ in *Drosophila* block the morphological and biochemical changes that accompany the midblastula transition, lead to a continuation of the maternal cell-cycle programme, and disrupt DNA-replication checkpoint control of cell-cycle progression. The timing of the midblastula transition is controlled by the ratio of nuclei to cytoplasm (the nucleocytoplasmic ratio), suggesting that this developmental transition is triggered by titration of a maternal factor by the increasing mass of nuclear material that accumulates during the rapid embryonic mitoses^{5–9}. Our observations support a model for cell-cycle control at the midblastula transition in

which titration of a maternal component of the DNA-replication machinery slows DNA synthesis and induces a checkpoint-dependent delay in cell-cycle progression¹⁰. This delay may allow both completion of S phase and transcription of genes that initiate the switch to zygotic control of embryogenesis.

The rapid maternally controlled divisions that initiate embryogenesis in *Drosophila melanogaster* proceed without cytokinesis, and produce a syncytial blastoderm embryo¹¹. We used a screen for mutations that disrupt mitotic chromosome segregation¹² to identify a maternal-effect mutation, initially designated *fs(A)4*, that produced severe mitotic defects during the later syncytial divisions.

To analyse directly the effects of this mutation on spindle morphogenesis and cell-cycle progression, we micro-injected embryos derived from homozygous *fs(A)4* females with rhodamine-tubulin conjugates, and visualized microtubule reorganization and cell-cycle progression by time-lapse confocal microscopy^{13,14} (Fig. 1). During the final syncytial divisions before the midblastula transition (MBT) (divisions 11–13), the duration of interphase increases from 6 to 14 min (Table 1). Interphase 14 is prolonged, lasting over 60 min, during which time invaginating membranes surround the majority of nuclei to form a cellular blastoderm¹¹. Blastoderm cellularization is the first important morphogenetic event that requires zygotic transcription, and thus cytologically marks the *Drosophila* MBT⁵. Our *in vivo* analyses of the syncytial mitoses demonstrate that embryos derived from homozygous *fs(A)4* females do not show a significant increase in interphase length during syncytial divisions 11–13 (Fig. 1, Table 1). Furthermore, mutant embryos proceed through at least two extra syncytial cycles of spindle assembly and disassembly and nuclear envelope breakdown and formation after mitosis 13, and cellularization is never observed (Fig. 1; data not shown). The *fs(A)4* mutation thus blocks nearly all of the changes in cell-cycle dynamics and embryo morphology associated with the *Drosophila* MBT.

The changes in cell-cycle length at the *Drosophila* MBT are associated with changes in phosphorylation of the cyclin-dependent kinase Cdc2 (ref. 15). Termination of the rapid syncytial divisions normally occurs between 2 and 3 h after egg deposition, and is accompanied by accumulation of a low electrophoretic-mobility inhibitory tyrosine-phosphorylated form of Cdc2 (ref. 15) (form 4, Fig. 2a). In *fs(A)4* mutant embryos, this tyrosine-phosphorylated form of Cdc2 does not accumulate (Fig. 2a). Studies in yeast and vertebrates indicate that the Cdc2–cyclin complex is activated by the Cdc25 phosphatase, which removes the inhibitory tyrosine phosphates and thus drives the cell into mitosis¹⁶. In *Drosophila*, a dramatic decrease in the level of String, the *Drosophila* Cdc25 homologue, accompanies the increase in cell-cycle time at the MBT^{15,17}. In *fs(A)4* mutants, String protein levels do not decrease significantly during the later syncytial mitoses (Fig. 2b). The biochemical changes in Cdc2 that normally accompany the MBT are therefore blocked by the *fs(A)4* mutation.

These observations raise the possibility that the *fs(A)4* gene product slows the cell cycle by inducing a downregulation of the String phosphatase, thereby allowing accumulation of inhibitory tyrosine-phosphorylated forms of Cdc2. It is also possible that coordinate activation of the *Drosophila* wee 1 kinase, which is in part responsible for inhibitory tyrosine phosphorylation of Cdc2 (ref. 18), may also be involved at the MBT.

The MBT is characterized by a sudden increase in zygotic transcription. In *Drosophila*, the MBT is accompanied by transcription of segmentation genes that are expressed in spatially restricted domains. We therefore used whole-mount *in situ* hybridization to assay the onset of zygotic segmentation gene expression in *fs(A)4* mutant embryos. Before the MBT, the segmentation genes *runt*, *fushi tarazu* (Fig. 3) and *giant* (not shown) are expressed at relatively low levels over broad regions of the embryo¹⁹, and the early patterns of *runt*, *fushi tarazu* and *giant* expression are observed in *fs(A)4* mutants. After mitosis 13, these genes are normally expressed at