



Received 10 September 2014

Accepted 22 January 2015

 Edited by G. J. Kleywegt, EMBL-EBI, Hinxton,  
England

 ‡ Present address: LRL-CAT Eli Lilly and  
Company, Advanced Photon Source, Argonne  
National Laboratory, Building 438A, 9700 South  
Cass Avenue, Lemont, IL 60439, USA.

**Keywords:** Expansion protein; Smad homology  
domain 2; MH2 domain; Smad/FHA protein  
superfamily; protein–protein interaction;  
*phenix.mr\_rosetta*.

**PDB reference:** Expansion N $\alpha$ -MH2, 4r9p

**Supporting information:** this article has  
supporting information at journals.iucr.org/d

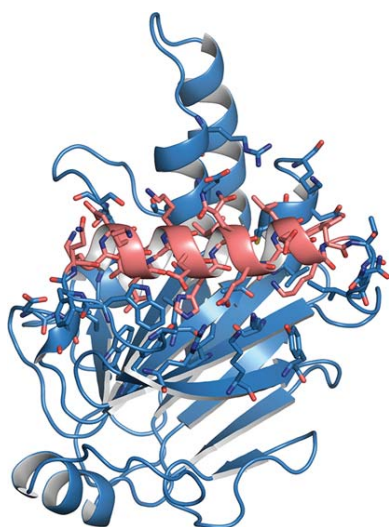
# Structure of the N-terminal domain of the protein Expansion: an ‘Expansion’ to the Smad MH2 fold

 Mads Beich-Frandsen,<sup>a</sup> Eric Aragón,<sup>a</sup> Marta Llimargas,<sup>b</sup> Jordi Benach,<sup>c,‡</sup> Antoni Riera,<sup>a,d</sup> Joan Pous<sup>a,e</sup> and Maria J. Macias<sup>a,f,\*</sup>
<sup>a</sup>Institute for Research in Biomedicine (IRB Barcelona), Baldiri Reixac 10, 08028 Barcelona, Spain, <sup>b</sup>Institut de Biologia Molecular de Barcelona, IBMB–CSIC, Baldiri Reixac 10, 08028 Barcelona, Spain, <sup>c</sup>ALBA Synchrotron, BP 1413, km 3.3, Cerdanyola del Vallès, Spain, <sup>d</sup>Departament de Química Orgànica, Universitat de Barcelona, Martí i Franqués 1-11, 08028 Barcelona, Spain, <sup>e</sup>Platform of Crystallography IBMB–CSIC, Baldiri Reixac 10, 08028 Barcelona, Spain, and <sup>f</sup>Catalan Institution for Research and Advanced Studies (ICREA), Passeig Lluís Companys 23, 08010 Barcelona, Spain.  
\*Correspondence e-mail: maria.macias@irbbarcelona.org

Gene-expression changes observed in *Drosophila* embryos after inducing the transcription factor Tramtrack led to the identification of the protein Expansion. Expansion contains an N-terminal domain similar in sequence to the MH2 domain characteristic of Smad proteins, which are the central mediators of the effects of the TGF- $\beta$  signalling pathway. Apart from Smads and Expansion, no other type of protein belonging to the known kingdoms of life contains MH2 domains. To compare the Expansion and Smad MH2 domains, the crystal structure of the Expansion domain was determined at 1.6 Å resolution, the first structure of a non-Smad MH2 domain to be characterized to date. The structure displays the main features of the canonical MH2 fold with two main differences: the addition of an  $\alpha$ -helical region and the remodelling of a protein-interaction site that is conserved in the MH2 domain of Smads. Owing to these differences, to the new domain was referred to as N $\alpha$ -MH2. Despite the presence of the N $\alpha$ -MH2 domain, Expansion does not participate in TGF- $\beta$  signalling; instead, it is required for other activities specific to the protostome phyla. Based on the structural similarities to the MH2 fold, it is proposed that the N $\alpha$ -MH2 domain should be classified as a new member of the Smad/FHA superfamily.

## 1. Introduction

The *Drosophila* transcription factor Tramtrack (Ttk) is involved in a wide range of processes during development of the tracheal system. The analysis of gene-expression changes in *Drosophila* embryos after inducing Ttk loss of function and gain of function enabled the identification of the *Expansion* gene (CG13188; *Exp*). A search for similar proteins to Expansion led to the identification of CG13183 (recently renamed Rebuf; Reb; Rotstein *et al.*, 2011). These two proteins share 56% amino-acid similarity that is concentrated in the N-terminal part of the sequence. Expansion and Rebuf (Exp/Reb) proteins have been identified in several *Drosophila* species, other insects and arthropods (Iordanou *et al.*, 2014). They are annotated in the NCBI database as modular proteins containing an N-terminal MH2 domain and a variable C-terminal region which does not present sequence similarity to other characterized domains. The sequence similarity to the MH2 domain is quite remarkable since these domains were believed to be exclusively present in Smad proteins, which are the main players in the TGF- $\beta$  signalling pathway in metazoans (Massagué, 2012). Smad proteins comprise two conserved domains separated by a linker that does not adopt a defined tertiary structure. The N-terminal (MH1) domain binds to DNA sites in promoters, while the linker and the


 OPEN  ACCESS

C-terminal (MH2) domain are the protein-interaction sites. As mediators and regulators of cytokine signalling, Smads are involved in many cellular processes from cell homeostasis to differentiation, division and cell death (Massagué *et al.*, 2005).

The presence of an MH2 domain in Exp/Reb proteins has been used as a hallmark to classify them as Smad-like proteins (Iordanou *et al.*, 2014). However, the sequence identity of the Exp/Reb proteins to the Smads is very low and is restricted to the MH2 domain. Furthermore, the differences are not only at the sequence level but also in the localization of the domain: in the N-terminal part in Exp/Reb in contrast to a C-terminal position in Smads. Secondary-structure predictions indicate that the Exp/Reb MH2 domain might contain additional elements of secondary structure preceding the MH2 fold. All of these characteristics suggest that Exp/Reb might constitute a new family of proteins that share the presence of a divergent MH2 domain with the Smads.

To clarify this issue and prompted by the similarities and differences between the Exp/Reb and Smads proteins, we set out to investigate the presence of the Exp/Reb proteins in metazoans and to characterize the structure of this new MH2 domain. Our results reveal that Exp/Reb proteins are restricted to protostomes, whereas Smads are highly conserved in both protostomes and deuterostomes. Regarding the structure, the crystal structure of this domain has been determined at 1.6 Å resolution and represents the first structure of an MH2 domain to be defined outside the Smad family of proteins. Although the structure displays the main features of the MH2 fold, it contains an additional  $\alpha$ -helical region that covers the concave site of the MH2 domain and defines the specific structure of the Exp/Reb MH2 domain. Based on this observation, we refer to the Exp/Reb MH2 domain as an  $N\alpha$ -MH2 domain, a new member of the FHA/Smad superfamily of MH2 domains.

A characteristic of activated Smad proteins is the formation of quaternary structures through interactions of their MH2 domains (Shi & Massagué, 2003). Even if Exp/Reb proteins are different from Smads, the presence of the  $N\alpha$ -MH2 domain led us to hypothesize that Exp/Reb proteins could perhaps also modulate TGF- $\beta$  signalling in protostomes through the formation of heterotrimers using the  $N\alpha$ -MH2 domain as a binding partner for Smads. The functional and structural studies presented here support the nonparticipation of Expansion in the canonical TGF- $\beta$  signalling pathway. Analysis of the  $N\alpha$ -MH2 structure and its comparison with those of Smad proteins provides the basis for understanding the binding differences. Our data are in agreement with the results reported in the literature indicating that Exp/Reb proteins are required for specific activities in protostomes, regulating receptor tyrosine kinase signalling to control terminal branch size and morphology (Iordanou *et al.*, 2014).

## 2. Experimental procedures

### 2.1. Sequence alignment and secondary-structure prediction

Sequences corresponding to Expansion proteins were retrieved from the Ensembl Metazoa database ([\[metazoa.ensembl.org\]\(http://metazoa.ensembl.org\)\) with \*PSI-BLAST\* \(Altschul \*et al.\*, 1990\), using the \*D. melanogaster\* Expansion protein \(CG13188\) as the query. Multiple sequence alignments with the query and the target proteins were generated with \*MAFFT\* v.7.164b using the iterative L-INS-i method \(parameters: Blosum62 scoring matrix and gap-opening penalty set to 1.5\).](http://</a></p>
</div>
<div data-bbox=)

Conserved residues were highlighted using the *BoxShade* server v.3.21 (<http://sourceforge.net/projects/boxshade/>) written by K. Hofmann and M. Baron. A graphical representation of secondary structure was added to the alignment using the *ESPrpt* server (Robert & Gouet, 2014).

Prediction of secondary-structure content was performed using the online version of *NetSurfP* (<http://www.cbs.dtu.dk/services/NetSurfP/>; Petersen *et al.*, 2009).

### 2.2. Cloning of the Expansion $N\alpha$ -MH2 domain

Constructs for the Expansion domain were cloned into the pETM-11 expression vector (EMBL) by means of ligation-independent cloning. The initial construct used for NMR screening and preliminary crystallization trials consisted of a fragment comprised of residues 29–240. We also prepared a second construct including an N-terminal extension (residues 3–240), which was used for structural studies.

Inserts were obtained by PCR using the appropriate primers and the *D. melanogaster* Expansion cDNA (CG13188, isoform B) as a template. PCR amplification was carried out using the standard PCR Master Mix with 0.02 unit  $\mu\text{l}^{-1}$  *Taq* DNA polymerase (Thermo Scientific), 1 ng  $\mu\text{l}^{-1}$  template DNA and 0.5  $\mu\text{M}$  of each primer, using an annealing temperature of 321 K for the first ten rounds and 341 K for the last 30 rounds. After subsequent purification of the reaction product using the GeneJET PCR purification kit (Fermentas), ligation-independent cloning was performed in the presence of 6  $\mu\text{g} \mu\text{l}^{-1}$  RecA DNA recombinase in the recommended buffer (New England Biolabs) at 310 K for 15 min. The full recombination reaction was then used for transformation in *E. coli* DH5 $\alpha$  and positive clones were selected on kanamycin agar plates (50  $\mu\text{g} \text{ml}^{-1}$ ). Positives clones were verified by DNA sequencing and subsequently transformed into the expression strain *E. coli* BL21 (DE3) Rosetta (Invitrogen), selecting positive clones on kanamycin (50  $\mu\text{g} \text{ml}^{-1}$ ) + chloramphenicol (34  $\mu\text{g} \text{ml}^{-1}$ ) agar plates.

### 2.3. Protein expression and purification

Cultures were grown at 310 K until an OD of  $\sim 0.6$  was reached; the temperature was then lowered to 293 K and protein expression was induced by the addition of 1 mM isopropyl  $\beta$ -D-1-thiogalactopyranoside (IPTG). Expression took place for 12–15 h and the cells were harvested by centrifugation (4000g, 15 min). Labelled samples for NMR were prepared in a similar manner using minimal medium (M9) enriched with  $^{15}\text{NH}_4\text{Cl}$ .

The cells were washed in TBS buffer and 5 g of cells were resuspended at 277 K in 10 ml buffer consisting of 50 mM Tris-HCl pH 7.2, 500 mM NaCl, 1 mM phenylmethylsulfonyl-fluoride (PMSF), 0.1 mM  $\beta$ -mercaptoethanol (BME), 0.1 mM

EDTA supplemented with 0.1 mg ml<sup>-1</sup> lysozyme, 0.25 mg ml<sup>-1</sup> DNaseI, 25 mM MgCl<sub>2</sub>, 0.01 mg ml<sup>-1</sup> RNaseA for lysis. Lysis was performed in a pressurized cell homogenizer at 277 K and the crude solution was incubated for 20 min on ice. The soluble fraction was then isolated by centrifugation (45 000g, 20 min, 277 K).

Immobilized metal-ion affinity chromatography (IMAC) was performed on an ÄKTA FPLC system at room temperature using a prepacked 1 ml His-tag column (GE Healthcare). The domain of interest was purified using buffer consisting of 50 mM Tris-HCl pH 7.2, 500 mM NaCl, 0.1 mM BME, 0.1 mM EDTA and eluted with a gradient to the same buffer freshly supplemented with 500 mM imidazole (50 mM Tris-HCl pH 7.2, 500 mM NaCl, 0.1 mM BME, 0.1 mM EDTA, 500 mM imidazole). Peak fractions were isolated and diluted with buffer (50 mM Tris-HCl pH 7.2, 200 mM NaCl, 0.1 mM BME) to lower the salt and imidazole concentrations for enzymatic digestion. Subsequently, the N-terminal 6×His tag was removed by overnight digestion at 277 K with *Tobacco etch virus* (TEV) protease and the digested protein was repurified by IMAC as described above. Size-exclusion chromatography (SEC) was performed on a Superdex 200 10/300 column using buffer consisting of 50 mM Tris-HCl pH 7.2, 200 mM NaCl, 0.1 mM BME and the peak fractions were collected and concentrated to ~10 mg ml<sup>-1</sup> using centrifugal filters (Amicon).

#### 2.4. *Drosophila* strains and genetics

The fly strains used are described in FlyBase. Df(2R)ED2247 and Df(2R)BSC879 uncover CG13188 and CG13183 and were used in transheterozygous conditions to analyze the absence of both genes (CG13188+CG13183 mutants). The transgenes used were P(TRiP.HMS01445)attP2 (UAS13188RNAi), P(TRiP.HMS01444)attP2 (UAS13183RNAi), P(TRiP.JF02218)attP2 (UASMedRNAi), UASDad and UASTkvCA.

For overexpression experiments, we used the Gal4/UAS system (Brand & Perrimon, 1993). We used the breathlessGal4 (btlGal4) driver, which drives expression in all tracheal cells, and the nubbinGal4 (nubGal4) driver, which drives expression in the wing disc. Crosses were kept at 29°C to maximize the expression of the transgenes.

To visualize the 'tracheal pattern', the embryos carrying btlGal4 UAS-srcGFP (cell membrane staining) were stained for GFP. The btlGal4 in this combination also drives the other UAS transgenes.

Confocal images were acquired with a Leica TCS-SPE system. Images were post-processed with *ImageJ* and *Adobe Photoshop* and assembled using *Adobe Illustrator*.

#### 2.5. NMR experiments

An HSQC (heteronuclear single-quantum coherence) NMR experiment (eight scans and 128 increments in the indirect dimension) was recorded at 298 K on a Bruker Avance III 600 MHz spectrometer equipped with a z-pulse field gradient unit and a triple (<sup>1</sup>H, <sup>13</sup>C, <sup>15</sup>N) resonance probe

**Table 1**  
Data-collection and refinement statistics.

Values in parentheses are for the highest resolution shell.	
Beamline	XALOC, ALBA
Wavelength (Å)	0.979
Resolution (Å)	31.3–1.59 (1.63–1.59)
Space group	<i>P</i> 2 <sub>1</sub>
Unit-cell parameters (Å, °)	<i>a</i> = 36.8, <i>b</i> = 45.7, <i>c</i> = 54.2, α = 90.00, β = 93.31, γ = 90.00
Molecules per asymmetric unit	1
Unique reflections	23152 (2902)
Completeness (%)	95.5 (74.6)
<i>R</i> <sub>meas</sub> <sup>†</sup>	0.071 (0.983)
<i>R</i> <sub>merge</sub> <sup>‡</sup>	0.069 (0.840)
Multiplicity	5.89 (3.5)
<i>I</i> / <i>σ</i> ( <i>I</i> )	13.6 (1.1)
CC <sub>1/2</sub> <sup>§</sup>	0.998 (0.506)
<i>B</i> <sub>Wilson</sub> (Å <sup>2</sup> )	33.0
<i>R</i> <sub>work</sub> / <i>R</i> <sub>free</sub> <sup>¶</sup>	0.176 (0.323)/0.198 (0.350)
Average <i>B</i> factor (Å <sup>2</sup> )	
Protein	32.9
Water	38.9
R.m.s. deviation from ideal values	
Bond lengths (Å)	0.019
Bond angles (°)	1.462
Ramachandran plot statistics (from <i>MolProbity</i> <sup>††</sup> )	
Outliers (%)	0
Favoured (%)	98.2
<i>MolProbity</i> <sup>‡‡</sup> statistics	
Score	1.35
Clashscore	6.26
Rotamer outliers (%)	1.0
No. of atoms (non-H)	1740
No. of solvent molecules	75
PDB code	4r9p

<sup>†</sup> *R*<sub>meas</sub> =  $\sum_{hkl} \{N(hkl)/[N(hkl) - 1]\}^{1/2} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$ .  
<sup>‡</sup> *R*<sub>merge</sub> is defined according to Kabsch (2010). <sup>§</sup> CC<sub>1/2</sub> is the correlation between intensities from random half data sets (Karplus & Diederichs, 2012). <sup>¶</sup> *R*<sub>free</sub> is the cross-validation *R* factor computed for a test set of reflections (5%) which were omitted from the refinement process. <sup>††</sup> R.m.s. deviation from ideal values in accordance with Engh & Huber (2001). <sup>‡‡</sup> Chen *et al.* (2010).

head. The protein sample (0.7 mM) was equilibrated in 20 mM deuterated Tris-HCl, 150 mM NaCl buffer with 10% D<sub>2</sub>O and the pH adjusted to 7.5. The data were processed using the *XwinNMR* 3.5 software supplied with the Bruker NMR spectrometer.

#### 2.6. Crystallization

Initial crystallization conditions were identified from sparse-matrix screens. A series of gradient screens optimized the final condition and three-dimensional diffraction-quality crystals were finally grown in sitting drops at 293 K from 12%(v/v) 1,2-propanediol, 9%(w/v) PEG 20 000, 0.1 M glycine pH 9. The short protein construct crystallized in similar conditions.

#### 2.7. Data collection

Data were collected at 100 K from a monoclinic crystal using a PILATUS 6M detector on BL13-XALOC at the ALBA Synchrotron Light Source, Barcelona, Spain. 360 images were collected at a wavelength of 0.97949 Å with an exposure time of 1 s and an oscillation of 1°. Data were processed in the *XDS* suite (Kabsch, 2010) and validated using

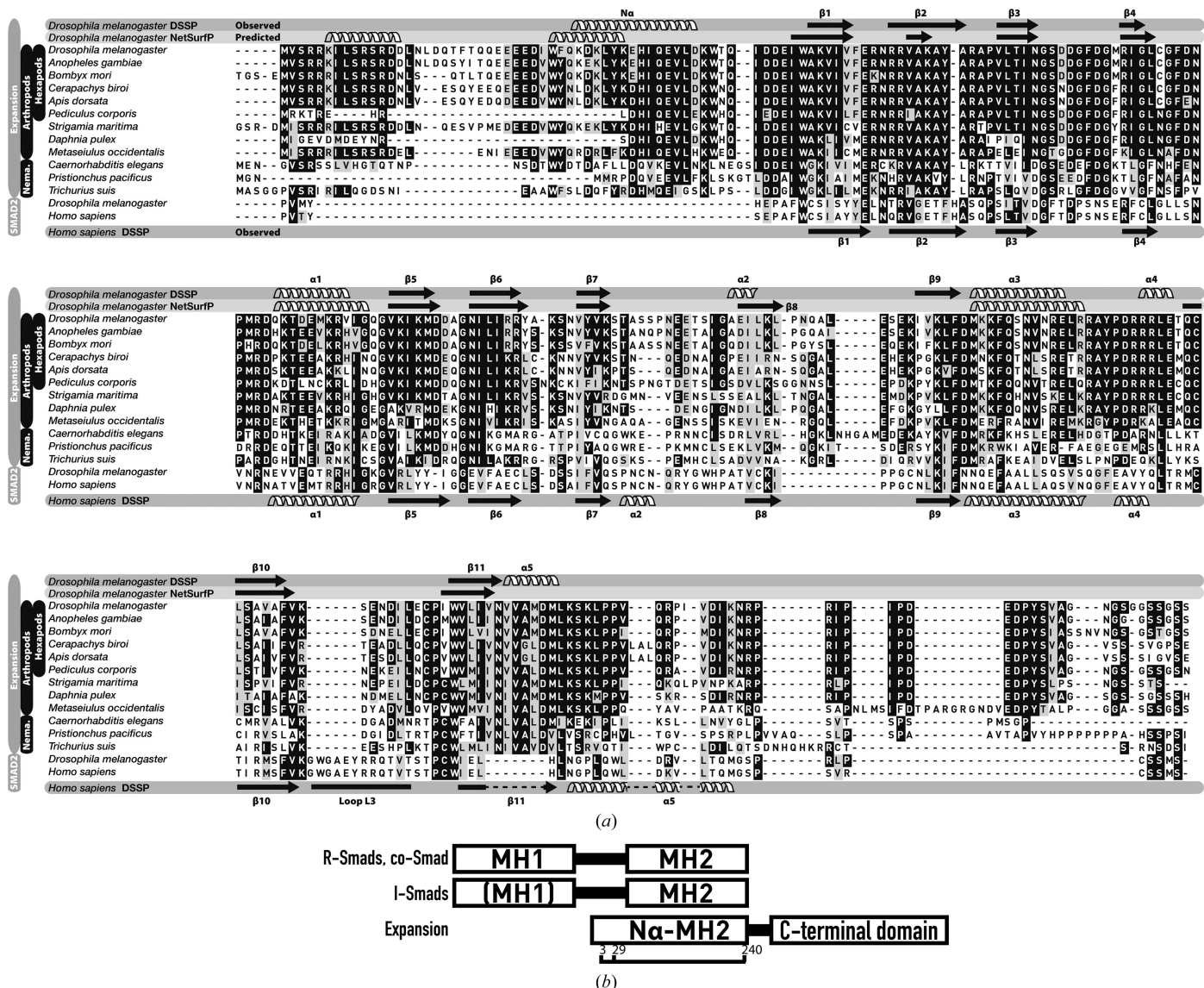
*POINTLESS* (Evans, 2006). Statistics of data-collection and model building are presented in Table 1.

### 2.8. Phasing, model building and refinement

Structure solution and model refinement was performed with programs from the *PHENIX* suite (Adams *et al.*, 2010). A structural solution was obtained by molecular replacement with *Phaser* (McCoy *et al.*, 2007) in the *phenix.mr\_rosetta* pipeline of programs (DiMaio *et al.*, 2009; Terwilliger *et al.*,

2012), using as a search model an ensemble constructed from the known structures of canonical MH2 domains: PDB entries 1kxh (16% sequence identity; Wu *et al.*, 2001), 1mjs (16%; Qin *et al.*, 2002), 1khu (15%; Qin *et al.*, 2001), 1dd1 (15%; Qin *et al.*, 1999) and 3gmj (14%; Wang *et al.*, 2009).

Sequence alignment and selection of homologues was performed with *HHpred* (Söding *et al.*, 2005), and *Sculptor* (Bunkóczi & Read, 2011) was used to edit non-identical side chains (Schwarzenbacher *et al.*, 2004), subsequently superimposing the structures in *Ensembler*, which also



**Figure 1** Sequence alignment of the N-terminal region of Expansion and Smad MH2 domains. (a) Sequence alignment of the N-terminal region of the *D. melanogaster* Expansion protein with orthologous proteins found in the Hexapoda, Arthropoda and Nematoda phyla. These sequences were obtained using *BLAST*. Using a *PSI-BLAST* search, we also detected sequence similarity to the MH2 domain of Smads. The *Homo sapiens* and *Drosophila* Smad2 MH2 domains are included in the alignment for comparison. Additional entries corresponding to uncharacterized proteins were not included in the alignment. A sequence alignment of the full-length proteins belonging to the Expansion family and a list of the entries included in the alignment are given as Supporting Information. The elements of secondary structure corresponding to the Expansion N $\alpha$ -MH2 domain are highlighted at the top of the alignment, with the elements of secondary structure labelled. The results of the secondary-structure prediction used to guide us in the definition of the protein boundaries are shown below. For comparison, the secondary-structure elements corresponding to the human Smad2 domain are depicted below the alignment. (b) Schematic representation of the domain organization in Smad proteins and Expansion, indicating the boundaries of the prepared N $\alpha$ -MH2 constructs.

removes nonconserved loops (<http://www.phenix-online.org/documentation/reference/enssembler.html>). The ensemble and the final model are superimposed and displayed in Supplementary Fig. S1.

Initial molecular-replacement solutions were rebuilt using *Rosetta* model completion and relaxation in *phenix.mr\_rosetta*, and automated rebuilding yielded the best results starting from Smad2 MH2 (PDB entry 1kxh), which resulted in a model with a free *R* factor of 30%.

Refinement was performed with *phenix.refine* (Afonine *et al.*, 2012) employing simulated-annealing and energy-minimization cycles. Initially tight geometry restraints were applied, which were released gradually as the  $R_{\text{work}}$  and  $R_{\text{free}}$  factors diverged. After several rounds of iterative refinement and model building in OMIT maps (Terwilliger *et al.*, 2008) and feature-enhanced maps (FEMs; Afonine *et al.*, 2015) in *Coot* (Emsley & Cowtan, 2004), density representing the N-terminal helix of the molecule could easily be identified in a difference electron-density map ( $mF_o - DF_c$ ). Iterative refinement and model building was continued until no significant improvement in *R* factors could be obtained. The refined model was validated using *MolProbity* (Chen *et al.*, 2010) and the *wwPDB Validation Server* (<http://wwpdb-validation.wwpdb.org>; Berman *et al.*, 2003). The Ramachandran plot is shown in Supplementary Fig. S2.

## 2.9. Structural analysis

Structural superimposition was performed with the superpose algorithm in *PyMOL* (Schrödinger) based on the  $\beta$ -strand content of each structure as calculated by *DSSP* (Kabsch & Sander, 1983). Sequence alignments with the *EMBOSS Needle* algorithm returned values for sequence identity and similarity (Rice *et al.*, 2000).

## 3. Results and discussion

### 3.1. Identification of a non-Smad MH2-like domain in protostomes

The sequence identity of the Exp/Reb *Drosophila* proteins (also known as CG13188 and CG13183, respectively) is mostly restricted to the MH2 domain (~30%). A comparison with Smads reveals a similarity about 16%, which is limited to the MH2 domain. This similarity suggests that the Exp/Reb and Smads proteins share the presence of the MH2 domain and that the Exp/Reb proteins may constitute a new protein family with a specialized function in a subset of metazoan phyla.

To clarify this issue, we searched the Ensembl Metazoa database (<http://metazoa.ensembl.org>) using *PSI-BLAST* with the Expansion full-length sequence as the query. Our search retrieved two sets of matches. The first set reflects similarity to the entire query in nematodes, arthropods and other hexapods, suggesting the presence of orthologous proteins in these organisms. The second set reflects similarity of the N-terminal part of the sequence query to only the C-terminal part of Smads, specifically to the MH2 domain, as reported in the annotation of these proteins in the NCBI database. Remark-

ably, the similarity in the Exp/Reb subfamily covers an additional region of similarity preceding the canonical MH2 domain which is absent in the Smad sequences. The alignment comparing the sequences of the Exp/Reb domains with those of the MH2 domain of Smads is shown in Fig. 1(a).

Secondary-structure predictions using *NetSurfP* (Petersen *et al.*, 2009) corroborated the sequence similarity of the new domain to that of the Smad MH2 structure and indicated that the additional conserved region preceding the canonical MH2 fold might adopt a helical structure. Of the two predicted helices, the sequence that lies adjacent to the canonical MH2 domain is more conserved than the fragment predicted at the most N-terminal part of the protein. The prediction is depicted at a 0.4 level of probability in the sequence alignment (Fig. 1a). A schematic representation of the domain organization of the Exp/Reb proteins and the similarity to Smads is shown in Fig. 1(b). An alignment based on the full sequence of the Expansion proteins is shown in Supplementary Fig. S3.

### 3.2. *Drosophila* Exp/Reb do not participate in the TGF- $\beta$ pathway

In addition to the structural work, we investigated the functional implication of this new family of proteins using cellular and genetic approaches in *Drosophila*. The TGF- $\beta$  signalling pathway has been shown to play a key role in tracheal formation, specifying the most dorsal and ventral branches (Llimargas & Casanova, 1997; Ribeiro *et al.*, 2002; Vincent *et al.*, 1997). When the pathway is downregulated (by overexpressing the inhibitory Smad), the formation of dorsal and ventral tracheal branches is compromised. However, when the pathway is constitutively activated (by overexpressing a constitutively active Thick Veins receptor;  $\text{TkV}^{\text{CA}}$ ) all branches migrate along the dorso-ventral axis (Supplementary Figs. S4a, S4b and S4c). We have observed that the pattern of dorsal and ventral branches was correct not only when Exp/Reb were downregulated using RNAi (Supplementary Figs. S4d and S4e) but also when both genes were removed using a combination of chromosome deficiencies that uncovers both of them (Supplementary Fig. S4f). In fact, we detected a completely different phenotype in loss-of-function conditions for these genes (Supplementary Figs. S4d, S4e and S4f), related to the accumulation of an apical chitinous extracellular matrix (aECM; Araújo *et al.*, 2005; Devine *et al.*, 2005; Tønning *et al.*, 2005; Moussian *et al.*, 2015). To further test any possible involvement of Exp/Reb in the TGF- $\beta$  pathway, we also analyzed the phenotypes of their downregulation in the wing (also obtaining a negative effect) and compared them with the defects in the downregulation of the control Smad4/Medea. As depicted in Supplementary Figs. S4(g), S4(h) and S4(i), the detected phenotypes of Exp/Reb downregulation are very different from those of the control, confirming that Exp/Reb do not appreciably transduce TGF- $\beta$  signals. Our results are consistent with recently reported experiments (Iordanou *et al.*, 2014) and suggest that this new family of proteins are involved in different functional pathways to the canonical TGF- $\beta$  pathway.

**Table 2**  
Human Smad MH2 domains used as search models for MR.

PDB code	Description	Identity (%)
1kxh	Human SMAD2 MH2 domain	16
1mjs	Human SMAD3 MH2 domain	16
1khu	Human SMAD1 MH2 domain	15
1dd1	Human SMAD4 MH2 domain	15
3gmj	<i>D. melanogaster</i> MAD MH2 domain	14

Overall, these results support our hypothesis that the Exp/Reb proteins define a new family of proteins specific for protostomes that have the MH2 domain in common with Smads.

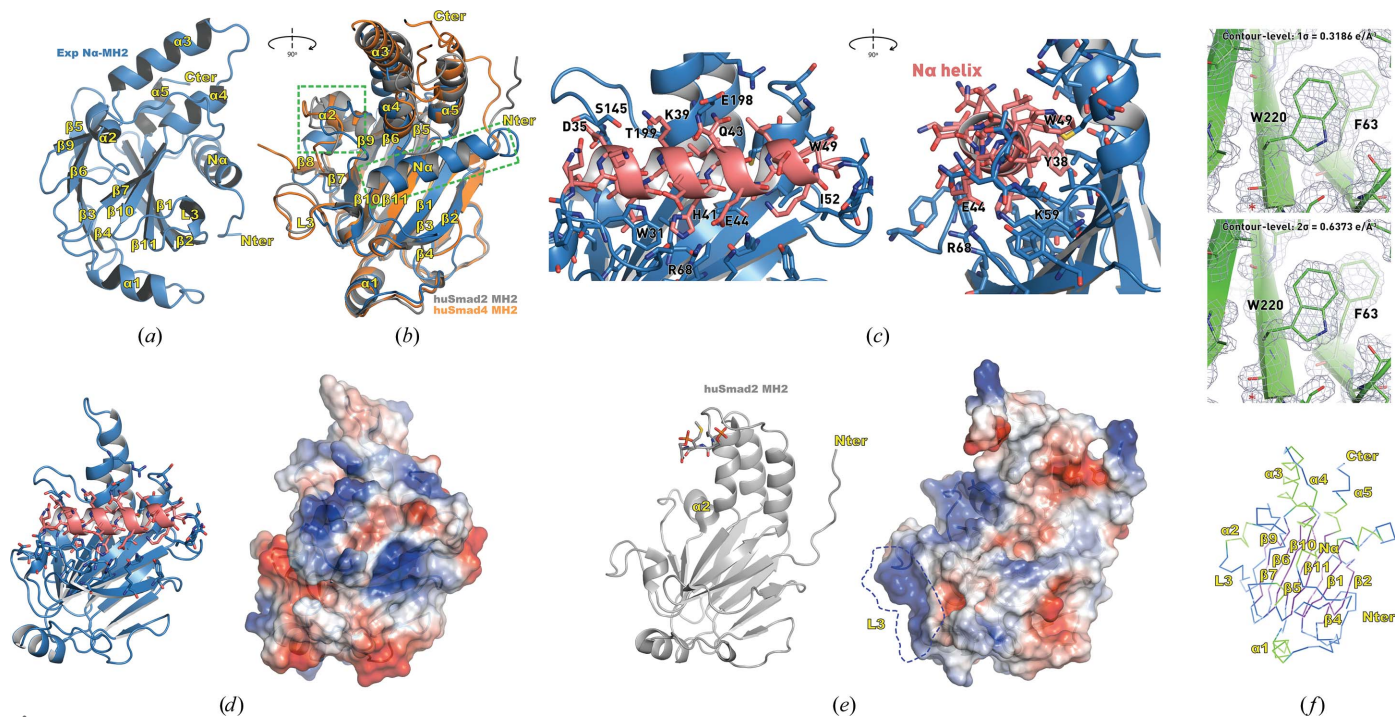
### 3.3. Recombinant expression and structural determination of the $N\alpha$ -MH2 domain of an Exp/Reb protein

**3.3.1. Protein expression, NMR and crystallographic screening.** Based on the sequence conservation and on the secondary-structure predictions, we selected two different domain boundaries for structural studies: a construct including the most conserved predicted helix (amino acids 29–240) and a second construct comprising nearly the full N-terminal region of the protein (amino acids 3–240) (Fig. 1*b*). These two recombinant proteins were soluble and eluted as monomers

in size-exclusion chromatography (Supplementary Fig. S5). According to NMR experiments the construct consisting of amino acids 29–240 was folded (Supplementary Fig. S6), and the initial crystallographic results were obtained using this construct. Since diffraction-quality crystals were also obtained from the larger construct, we focused the structural work on this construct consisting of amino acids 3–240. This would allow us to elucidate the role of the entire N-terminal region.

Diffraction data were collected from a monoclinic wedge-shaped crystal with approximate dimensions of  $10 \times 30 \times 150 \mu\text{m}$ . The data were processed in space group  $P2_1$  to a maximum resolution of 1.6 Å. Matthews coefficient analysis indicated a solvent content of  $\sim 27\%$  for one protomer ( $\sim 28$  kDa) in the asymmetric unit, which was consistent with the tight packing observed in the structure.

**3.3.2. Structural determination of the  $N\alpha$ -MH2 domain.** Owing to the anticipated structural resemblance of the new domain to the canonical Smad MH2 domain, we first attempted to solve the structure by conventional molecular replacement (MR) using *Phaser* (McCoy *et al.*, 2007) with several human Smad MH2 domains as search models (Table 2). Using this program we could identify a few potential MR solutions that could not be further refined, probably owing to the low sequence identity between the Expansion MH2 domain and human Smad MH2 ( $\sim 16\%$ ; Schwarzenba-



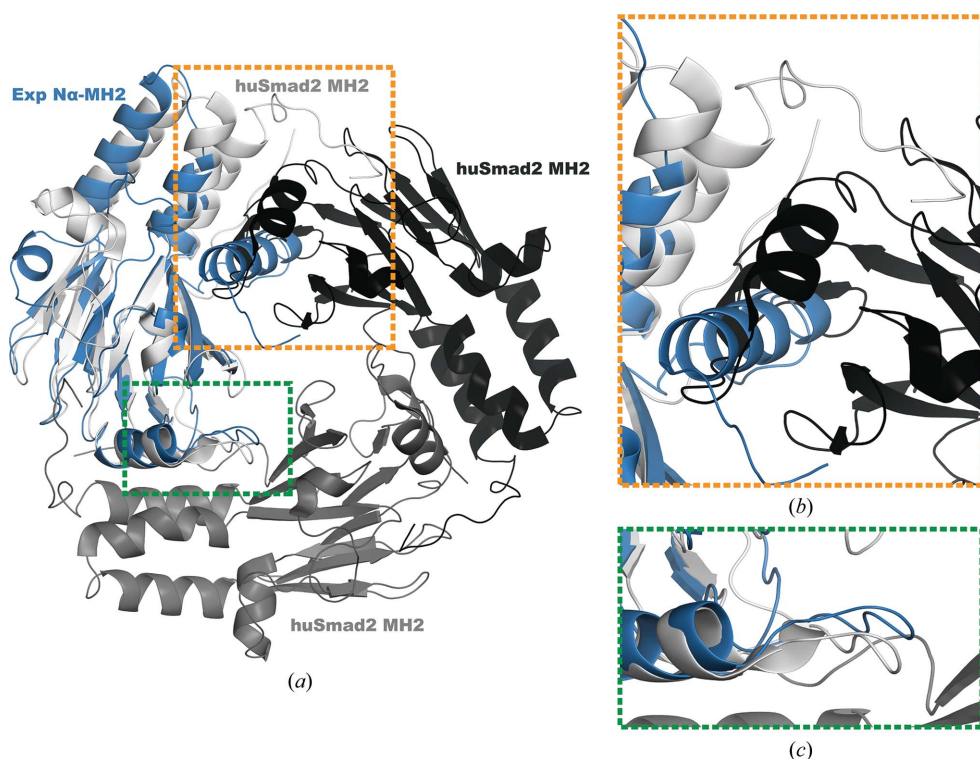
**Figure 2**  
Structure of the Expansion  $N\alpha$ -MH2 domain and its comparison with the human Smad2 (huSmad2) and Smad4 (huSmad4) MH2 domains. (a) Refined structure of the Expansion  $N\alpha$ -MH2 domain including residues 27–236. The structure is displayed as a cartoon representation. Elements of secondary structure are labelled according to the Smad canonical MH2 fold, despite differences in the secondary-structure content in the N-terminus (the  $N\alpha$  helix present in Expansion) and in the region of  $\alpha 2$  and  $\beta 8+9$ . (b) Superposition of the huSmad2 MH2 domain (PDB entry 1mjs; grey), the human Smad4 MH2 domain (PDB entry 1dd1; orange) and the refined structure of the  $N\alpha$ -MH2 domain of Expansion (PDB entry 4r9p; blue). This view is rotated by  $90^\circ$  with respect to (a). (c) Close-up view of the  $N\alpha$  helix of the Expansion  $N\alpha$ -MH2 domain. The  $N\alpha$  helix (depicted in salmon) and the side chains that contribute to its packing are represented as sticks and labelled. (d) Cartoon and electrostatic potential surface distribution of the Expansion  $N\alpha$ -MH2 domain. (e) Same as in (d) for the huSmad2 MH2 domain (PDB entry 1kxh). The L3 region required for phosphorylation recognition is highlighted. (f) The  $C^\alpha$  trace of the  $N\alpha$ -MH2 domain and electron-density maps at contour levels of  $1\sigma$  ( $0.3186 \text{ e} \text{ \AA}^{-3}$ ) and  $2\sigma$  ( $0.6372 \text{ e} \text{ \AA}^{-3}$ ). All figures were prepared using *PyMOL* (Schrodinger).

cher *et al.*, 2004). Among these solutions, the best was that obtained using the human Smad3 MH2 domain (PDB entry 1mjs). The selected solution (using *Phaser*) reported values of LLG = 56.1, TFZ = 4.3 and  $R_{\text{val}} = 58.5$  at 3 Å resolution, and allowed us to partially trace the map. However, we were unable to perform rigid-body refinement as we could not improve the  $R$  factors beyond  $R_{\text{work}} = 0.5207$  and  $R_{\text{free}} = 0.5215$ . Similar values were obtained for other resolution ranges.

Since this approach was unsuccessful, we decided to apply the *MR-Rosetta* algorithm, which has recently been demonstrated to facilitate molecular replacement in cases where only search models of low sequence identity are available. Using the methods compiled in the *MR-Rosetta* pipeline (DiMaio *et al.*, 2011; Terwilliger *et al.*, 2012) from the *PHENIX* package (Adams *et al.*, 2010) and an ensemble constructed from the known structures of canonical MH2 domains as a search model, a structural solution could be obtained. The ensemble and the final model are displayed in Supplementary Fig. S1. The data were cut following the recent recommendations by Karplus & Diederichs (2012) at a conservative resolution of  $CC_{1/2} > 0.5$ , which resulted in the data-collection and model-refinement statistics reported in Table 1. The high-quality diffraction data allowed the modelling of residues 27–236, with  $R_{\text{work}} = 0.1765$  and  $R_{\text{free}} = 0.1977$ .

As expected, the refined model revealed a tertiary structure with a striking resemblance to the canonical Smad MH2 fold (Figs. 2*a* and 2*b*). The structure comprises a  $\beta$ -sandwich core of twisted antiparallel  $\beta$ -sheets capped at one end by a three-helix bundle and at the other by a region containing an  $\alpha$ -helix and several loops. In Smad MH2 domains this region is commonly referred to as the ‘loop–helix region’ (Shi *et al.*, 1997).

At a first glance, the most obvious structural difference when superimposed on Smad2 MH2 domains is the presence of the N $\alpha$  helix, which is formed by the N-terminal residues 34–47 and covers one side of the core  $\beta$ -sandwich (Figs. 2*c* and 2*d*). This new element of secondary structure is named N $\alpha$  to indicate its position at the N-terminus of the domain and, most importantly, to maintain the canonical nomenclature of MH2 domains (Fig. 2*c*). A  $C^\alpha$  trace of the N $\alpha$ -MH2 domain and the electron-density map at contour levels of  $1\sigma$  and  $2\sigma$  are shown in Fig. 2(*f*). A few additional structural differences are observed in the length of the helices that form the helical bundle and in the area adjacent to the ‘loop–helix region’ comprised of helix  $\alpha 2$ , a loop and strand  $\beta 8$ , referred to here as the ‘H2 region’ (Fig. 2*b*, green rectangles). Also significant is the reduced length of the L3 loop that connects  $\beta 10$  and  $\beta 11$ . This loop in Smad MH2 domains comprises an interaction motif for phosphorylated residues, and is fundamental to receptor binding and to the quaternary structure and function of the Smad proteins (Lo *et al.*, 1998). In Smad domains loop L3 comprises 17 residues, whereas in Expansion this loop is shorter (11 residues) and is very different in sequence (Fig. 1*a*). To observe the effect of the differences in and around the L3 area, we have represented the surface charge distribution of the Expansion N $\alpha$ -MH2 domain and that of Smad2 for comparison (Figs. 2*d* and 2*e*) and highlighted the presence of positively charged patches in the Smad2 MH2 domain that are absent in Expansion.



**Figure 3** Comparison of the huSmad2 MH2 homotrimer with the Expansion N $\alpha$ -MH2 domain. (*a*) Structural superposition of Expansion N $\alpha$ -MH2 (blue) and the huSmad2 MH2 homotrimer. Smad2 MH2 domains in the trimer are represented as cartoons coloured three shades of grey to distinguish each monomer in the homotrimeric structure. The structures are superimposed using the  $\beta$ -sandwich (r.m.s.d. of 0.636 Å on 43  $C^\alpha$  atoms). All elements of secondary structure of Expansion and of one monomer of the human Smad2 trimer overlap, but the N $\alpha$  helix of Expansion is superimposed on a loop of an adjacent MH2 domain of huSmad2. The superimposed N $\alpha$  helix and the H2 region are marked with rectangles (gold and green, respectively). Detailed views of these regions are shown in (*b*) and (*c*).

3.3.3. The structure of the N $\alpha$ -MH2 domain does not support the formation of homotrimers or heterotrimers. The N-terminal helix observed in Expansion N $\alpha$ -MH2 (Figs. 2*a* and 2*c*) is a new addition to the canonical MH2 fold characteristic of Smad proteins. This novel structural element packs against and interacts with the triple-helical bundle and covers the outer surface of the  $\beta$ -sandwich core, which represents the interface between

adjacent monomers in the functional trimer of the Smad proteins. Previous results have suggested how cancer-derived mutations, which map to the same area in the Smad proteins, inhibit the formation of the functional heterotrimer (Shi *et al.*, 1997). When superposing the  $N\alpha$ -MH2 domain (shown in blue in Figs. 3*a*, 3*b* and 3*c*) onto one monomer of the Smad2 MH2 homotrimer (shown in light grey), the  $N\alpha$  helix overlaps with the area occupied by the 'loop-helix region' of the adjacent monomer in the trimer (in the figure this second MH2 is shown in dark grey). This structural 'clash' most certainly compromises the formation of homotrimers by the  $N\alpha$ -MH2 domain in a manner similar to that observed for the Smad MH2 domains. Furthermore, it will also prevent the formation of heterotrimers with the MH2 domain of Smad proteins.

In Smad proteins, where the MH2 domain is located at the protein C-terminus (Fig. 1*b*), the linker preceding the MH2 domain does not adopt a defined tertiary structure. However, in several structures of Smad complexes segments of the linker have been observed to adopt different conformations when folding upon interaction with other proteins (Aragón *et al.*, 2011, 2012; Shi & Massagué, 2003).

The secondary-structure predictions of the Expansion family of proteins suggested the presence of a second helix at the very N-terminus of the protein (Fig. 1*a*). However, no electron density could be observed for this predicted helix, likely reflecting a degree of flexibility in this area. Patches of positive electron density observed in a difference-density map ( $mF_o - DF_c$ ) suggested that the region comprising the first 24 residues extends into a solvent channel of the crystal structure; however, any effort to improve the model in this region did not improve the refinement statistics and the model was therefore truncated at the N-terminus.

**3.3.4. Protein interactions in MH2 domains.** The divergent sequence of loop L3 in the  $N\alpha$ -MH2 domain indicates a significant difference in the function of the Expansion  $N\alpha$ -MH2 domain with respect to canonical Smad MH2 domains. In the TGF- $\beta$  pathway signals are propagated through receptor activation of R-Smads by phosphorylation of the S-X-S motif at the C-terminus of the MH2 domain, leading to heterotrimer formation through the MH2 domain of two R-Smads with the common Smad4. Formation of the heterotrimer triggers the subsequent translocation of Smads to the nucleus. Mutations in the L3 loop of the Smad proteins abolish the formation of heterotrimers and hence signalling in the TGF- $\beta$  pathway (Wu *et al.*, 2001). Whereas unphosphorylated Smad proteins are able to form dimers and trimers in a concentration-dependent manner (Shi & Massagué, 2003), the presence of the additional helix and the lack of the L3 loop in the  $N\alpha$ -MH2 domain could effectively prevent this domain from supporting the formation of similar dimers and trimers.

The Smad MH2 domain is commonly known to support protein interactions, and interestingly the overall composition of the Expansion  $N\alpha$ -MH2 domain remains largely the same despite the low degree of sequence identity. However, the H2 region that has been established as a protein-protein interaction site in Smad proteins (Qin *et al.*, 1999; Wu *et al.*, 2002) is completely different in the Expansion  $N\alpha$ -MH2 domain. In

the refined model of the  $N\alpha$ -MH2 domain the  $\beta$ -sandwich core is comprised by two antiparallel  $\beta$ -sheets each having five strands, whereas in Smad MH2 the upper sheet comprises six strands (Figs. 2*a* and 2*b*). In the Expansion  $N\alpha$ -MH2 domain the region of amino acids 162–165 that would correspond to the  $\beta$ 8 strand in Smads lacks a defined secondary structure; it instead bears characteristics of random coil similar to most of the adjacent H2 region, apart from helix  $\alpha$ 2. In Smad proteins the H2 region extends from strand  $\beta$ 7 in the upper  $\beta$ -sheet, with helix  $\alpha$ 2 followed by a short loop that connects to strand  $\beta$ 8 continuing to strand  $\beta$ 9. In the  $N\alpha$ -MH2 domain strand  $\beta$ 7 is followed by a large loop connected to a reoriented helix  $\alpha$ 2, followed by an extended region (corresponding to strand  $\beta$ 8) continuing to strand  $\beta$ 9 (Figs. 2*a* and 2*b*).

In the Smad proteins the H2 region and the  $\beta$ 8 strand are implicated in protein-protein interactions by  $\beta$ -sheet augmentation, similar to common PDZ domains (Cowburn, 1997; Doyle *et al.*, 1996; Morais Cabral *et al.*, 1996; Schultz *et al.*, 1998), by annealing an additional  $\beta$ -strand to an existing  $\beta$ -sheet. This type of coordination is common in protein-protein interactions (Remaut & Waksman, 2006) and has been characterized structurally in three cases for Smad proteins. The structure of the isolated human Smad4 MH2 domain was solved from a construct with a protracted N-terminal boundary containing the first part of the linker region (for the 'common' Smad4 this is also known as the Smad-activation domain). As mentioned above, the structure revealed that part of the linker adopts an extended conformation and interacts with the H2 region by  $\beta$ -sheet augmentation (PDB entry 1dd1; Qin *et al.*, 1999). Similarly, the structure of Smad4 in complex with the repressor protein c-Ski (PDB entry 1mr1) also shows that the interaction occurs through  $\beta$ -sheet augmentation of strand  $\beta$ 8 (Wu *et al.*, 2002). Moreover, the interaction between Smad2 and SARA (Smad anchor for receptor activation) has been characterized structurally and was also revealed to involve  $\beta$ -sheet augmentation, but not in the H2 region (PDB entry 1dev; Wu *et al.*, 2000). It is possible that secondary-structural changes induced by ligand binding could stabilize the  $\beta$ 8 structure. Indeed, *NetSurfP* predicts the presence of strand  $\beta$ 8 in Expansion, indicating (to some extent) an intrinsic folding property of this area. Of interest in the remodelling of the H2 region of the Expansion  $N\alpha$ -MH2 domain as a potential protein-interaction site is the consideration of the degree of specificity that this remodelling might provide.

**3.3.5. Structural classification in the Smad/FHA family.** Structural homology to the Smad MH2 domain has also been found in FHA domains and the C-terminal regulatory domain of IRF-3. Despite a complete absence of sequence conservation between these proteins, an evolutionary link has previously been suggested owing to the structural and distant functional similarities between the  $\beta$ -sandwiches in these proteins (Durocher *et al.*, 2000; Huse *et al.*, 2001; Takahashi *et al.*, 2003). Indeed, these proteins have all been classified into the same SCOP superfamily: the Smad/FHA domain. The superimposed structures of the human MDC1 FHA domain, the C-terminal regulatory domain of IRF-3 and the Expansion



N $\alpha$ -MH2 domain are shown in Supplementary Fig. S7. Since the MH2 domain of Smads is similar to the IRF-3 regulatory domain and to the FHA domain, the N $\alpha$ -MH2 domain of Expansion is also a member of the same family of structures.

IRF proteins are only found in vertebrates, whereas the Smad MH2 domain and the FHA domain co-exist in metazoans and the FHA domain is also found in prokaryotes. Durocher *et al.* (2000) found the minimal  $\beta$ -sandwich of the FHA domain to comprise alternative protein-interaction sites, similar to those identified in the Smad MH2 domain, and Takahashi *et al.* (2003) speculated that the different flanking regions of the  $\beta$ -sandwich in the C-terminal regulatory domain of IRF-3 and Smad-MH2 developed on the  $\beta$ -sandwich scaffold of the FHA protein to facilitate phosphorylation signalling in higher organisms.

#### 4. Conclusions

We have determined the structure of the N $\alpha$ -MH2 domain present in Expansion at 1.6 Å resolution. The addition of the N-terminal helix, the differences in the L3 loop and its lack of a role in the canonical TGF- $\beta$  signalling pathway support the classification of Expansion as a new family of proteins that share the presence of the MH2 domain with Smads. Furthermore, the structural differences between the Smad MH2 and Expansion N $\alpha$ -MH2 domains could have evolved to host a different range of protein-interaction partners, with implications for different cellular functions which apparently have been conserved in protostomes (Mollusca, Annelida and Arthropoda phyla).

For these reasons, we suggest that Expansion should not be termed a ‘Smad-like’ protein. Furthermore, we propose that the Smad/FHA family of structures should be ‘expanded’ to also include the Expansion N $\alpha$ -MH2 domain.

#### Acknowledgements

We acknowledge the support provided by the IRB Barcelona and also the Mass Spectrometry core facility and the PAC IRB-IBMB-CSIC Platform of Crystallography. X-ray diffraction data were collected on BL13-XALOC at the ALBA Synchrotron Light Source with the support of the ALBA staff. This work was supported by the Spanish National Research Program (MINECO, SAF2011-25119). MBF was supported by a Marie Curie Action (COFUND) within the European Union Seventh Framework Programme. MJM is an ICREA Programme Investigator.

#### References

Adams, P. D. *et al.* (2010). *Acta Cryst.* **D66**, 213–221.  
 Afonine, P. V., Grosse-Kunstleve, R. W., Echols, N., Headd, J. J., Moriarty, N. W., Mustyakimov, M., Terwilliger, T. C., Urzhumtsev, A., Zwart, P. H. & Adams, P. D. (2012). *Acta Cryst.* **D68**, 352–367.  
 Afonine, P. V., Moriarty, N. W., Mustyakimov, M., Sobolev, O. V., Terwilliger, T. C., Turk, D., Urzhumtsev, A. & Adams, P. D. (2015). *Acta Cryst.* **D71**, doi:10.1107/S1399004714028132.  
 Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). *J. Mol. Biol.* **215**, 403–410.  
 Aragón, E., Goerner, N., Xi, Q., Gomes, T., Gao, S., Massagué, J. & Macias, M. J. (2012). *Structure*, **20**, 1726–1736.

Aragón, E., Goerner, N., Zaromytidou, A. I., Xi, Q., Escobedo, A., Massagué, J. & Macias, M. J. (2011). *Genes Dev.* **25**, 1275–1288.  
 Araújo, S. J., Aslam, H., Tear, G. & Casanova, J. (2005). *Dev. Biol.* **288**, 179–193.  
 Berman, H., Henrick, K. & Nakamura, H. (2003). *Nature Struct. Biol.* **10**, 980.  
 Brand, A. H. & Perrimon, N. (1993). *Development*, **118**, 401–415.  
 Bunkóczi, G. & Read, R. J. (2011). *Acta Cryst.* **D67**, 303–312.  
 Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2010). *Acta Cryst.* **D66**, 12–21.  
 Cowburn, D. (1997). *Curr. Opin. Struct. Biol.* **7**, 835–838.  
 Devine, W. P., Lubarsky, B., Shaw, K., Luschnig, S., Messina, L. & Krasnow, M. A. (2005). *Proc. Natl Acad. Sci. USA*, **102**, 17014–17019.  
 DiMaio, F., Terwilliger, T. C., Read, R. J., Wlodawer, A., Oberdorfer, G., Wagner, U., Valkov, E., Alon, A., Fass, D., Axelrod, H. L., Das, D., Vorobiev, S. M., Iwai, H., Pokkuluri, P. R. & Baker, D. (2011). *Nature (London)*, **473**, 540–543.  
 DiMaio, F., Tyka, M. D., Baker, M. L., Chiu, W. & Baker, D. (2009). *J. Mol. Biol.* **392**, 181–190.  
 Doyle, D. A., Lee, A., Lewis, J., Kim, E., Sheng, M. & MacKinnon, R. (1996). *Cell*, **85**, 1067–1076.  
 Durocher, D., Taylor, I. A., Sarbassova, D., Haire, L. F., Westcott, S. L., Jackson, S. P., Smerdon, S. J. & Yaffe, M. B. (2000). *Mol. Cell*, **6**, 1169–1182.  
 Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* **D60**, 2126–2132.  
 Engh, R. A. & Huber, R. (2001). *International Tables for Crystallography*, Vol. F, edited by M. G. Rossmann & E. Arnold, pp. 382–392. Dordrecht: Kluwer Academic Publishers.  
 Evans, P. (2006). *Acta Cryst.* **D62**, 72–82.  
 Huse, M., Muir, T. W., Xu, L., Chen, Y.-G., Kuriyan, J. & Massagué, J. (2001). *Mol. Cell*, **8**, 671–682.  
 Iordanou, E., Chandran, R. R., Yang, Y., Essak, M., Blackstone, N. & Jiang, L. (2014). *Dev. Biol.* **393**, 93–108.  
 Kabsch, W. (2010). *Acta Cryst.* **D66**, 125–132.  
 Kabsch, W. & Sander, C. (1983). *Biopolymers*, **22**, 2577–2637.  
 Karplus, P. A. & Diederichs, K. (2012). *Science*, **336**, 1030–1033.  
 Llimargas, M. & Casanova, J. (1997). *Development*, **124**, 3273–3281.  
 Lo, R. S., Chen, Y.-G., Shi, Y., Pavletich, N. P. & Massagué, J. (1998). *EMBO J.* **17**, 996–1005.  
 Massagué, J. (2012). *Nature Rev. Mol. Cell Biol.* **13**, 616–630.  
 Massagué, J., Seoane, J. & Wotton, D. (2005). *Genes Dev.* **19**, 2783–2810.  
 McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.  
 Morais Cabral, J. H., Petosa, C., Sutcliffe, M. J., Raza, S., Byron, O., Poy, F., Marfatia, S. M., Chishti, A. H. & Liddington, R. C. (1996). *Nature (London)*, **382**, 649–652.  
 Moussian, B., Letizia, A., Martínez-Corrales, G., Rotstein, B., Casali, A. & Llimargas, M. (2015). *PLoS Genet.* doi:10.1371/journal.pgen.1004939.  
 Petersen, B., Petersen, T. N., Andersen, P., Nielsen, M. & Lundegaard, C. (2009). *BMC Struct. Biol.* **9**, 51.  
 Qin, B. Y., Chacko, B. M., Lam, S. S., de Caestecker, M. P., Correia, J. J. & Lin, K. (2001). *Mol. Cell*, **8**, 1303–1312.  
 Qin, B. Y., Lam, S. S., Correia, J. J. & Lin, K. (2002). *Genes Dev.* **16**, 1950–1963.  
 Qin, B., Lam, S. S. W. & Lin, K. (1999). *Structure*, **7**, 1493–1503.  
 Remaut, H. & Waksman, G. (2006). *Trends Biochem. Sci.* **31**, 436–444.  
 Ribeiro, C., Ebner, A. & Affolter, M. (2002). *Dev. Cell*, **2**, 677–683.  
 Rice, P., Longden, I. & Bleasby, A. (2000). *Trends Genet.* **16**, 276–277.  
 Robert, X. & Gouet, P. (2014). *Nucleic Acids Res.* **42**, W320–W324.  
 Rotstein, B., Molnar, D., Adryan, B. & Llimargas, M. (2011). *PLoS One*, **6**, e28985.

- Schultz, J., Hoffmüller, U., Krause, G., Ashurst, J., Macias, M. J., Schmieder, P., Schneider-Mergener, J. & Oschkinat, H. (1998). *Nature Struct. Mol. Biol.* **5**, 19–24.
- Schwarzenbacher, R., Godzik, A., Grzechnik, S. K. & Jaroszewski, L. (2004). *Acta Cryst. D* **60**, 1229–1236.
- Shi, Y., Hata, A., Lo, R. S., Massagué, J. & Pavletich, N. P. (1997). *Nature (London)*, **388**, 87–93.
- Shi, Y. & Massagué, J. (2003). *Cell*, **113**, 685–700.
- Söding, J., Biegert, A. & Lupas, A. N. (2005). *Nucleic Acids Res.* **33**, W244–W248.
- Takahashi, K., Suzuki, N. N., Horiuchi, M., Mori, M., Suhara, W., Okabe, Y., Fukuhara, Y., Terasawa, H., Akira, S., Fujita, T. & Inagaki, F. (2003). *Nature Struct. Biol.* **10**, 922–927.
- Terwilliger, T. C., DiMaio, F., Read, R. J., Baker, D., Bunkóczi, G., Adams, P. D., Grosse-Kunstleve, R. W., Afonine, P. V. & Echols, N. (2012). *J. Struct. Funct. Genomics*, **13**, 81–90.
- Terwilliger, T. C., Grosse-Kunstleve, R. W., Afonine, P. V., Moriarty, N. W., Adams, P. D., Read, R. J., Zwart, P. H. & Hung, L.-W. (2008). *Acta Cryst. D* **64**, 515–524.
- Tonning, A., Hemphälä, J., Tång, E., Nannmark, U., Samakovlis, C. & Uv, A. (2005). *Dev. Cell*, **9**, 423–430.
- Vincent, S., Ruberte, E., Grieder, N. C., Chen, C.-K., Haerry, T., Schuh, R. & Affolter, M. (1997). *Development*, **124**, 2741–2750.
- Wang, C., Chen, L., Wang, L. & Wu, J. W. (2009). *Sci. China C Life Sci.* **52**, 539–544.
- Wu, G., Chen, Y.-G., Ozdamar, B., Gyuricza, C. A., Chong, P. A., Wrana, J. L., Massagué, J. & Shi, Y. (2000). *Science*, **287**, 92–97.
- Wu, J. W., Hu, M., Chai, J., Seoane, J., Huse, M., Li, C., Rigotti, D. J., Kyin, S., Muir, T. W., Fairman, R., Massagué, J. & Shi, Y. (2001). *Mol. Cell*, **8**, 1277–1289.
- Wu, J. W., Krawitz, A. R., Chai, J., Li, W., Zhang, F., Luo, K. & Shi, Y. (2002). *Cell*, **111**, 357–367.