Published in final edited form as:
J Mol Evol. 2008 October ; 67(4): 323–333. doi:10.1007/s00239-008-9159-4.

Lysine Conservation and Context in TGF*β* and Wnt Signaling Suggest New Targets and General Themes for Posttranslational Modification

Charlotte E. Konikoff
School of Life Sciences, Arizona State University, Tempe, AZ 85287-4501, USA

Center for Evolutionary Functional Genomics, Arizona State University, Tempe, AZ 85287-4501, USA

Robert G. Wisotzkey
Department of Biological Sciences, California State University—East Bay, Hayward, CA, USA

Stuart J. Newfeld
School of Life Sciences, Arizona State University, Tempe, AZ 85287-4501, USA

Center for Evolutionary Functional Genomics, Arizona State University, Tempe, AZ 85287-4501, USA

Abstract

TGF*β* and Wnt pathways play important roles in the development of animals from sponges to humans. In both pathways posttranslational modification as a means of regulating their function, such as lysine modification by ubiquitination and sumoylation, has been observed. However, a gap exists between the immunological observation of posttranslational modification and the identification of the target lysine. To fill this gap, we conducted a phylogenetic analysis of lysine conservation and context in TGF*β* and Wnt pathway receptors and signal transducers and suggest numerous high-probability candidates for posttranslational modification. Further comparison of results from both pathways suggests two general features for biochemical regulation of intercellular signaling: receptors are less frequent targets for modification than signal transduction agonists, and a lysine adjacent to an upstream hydrophobic residue may be a preferred context for modification. Overall the results suggest numerous applications for an evolutionary approach to the biochemical regulation of developmental pathways, including (1) streamlining of the identification of the target lysine, (2) determination of when members of a multigene family acquire distinct activities, (3) application to any conserved protein family, and (4) application to any modification of a specific amino acid.

Keywords

Phylogenetics; TGF*β*; Smads; Frizzled; Dishevelled; Ubiquitination; Sumoylation

e-mail: E-mail: newfeld@asu.edu.
Charlotte E. Konikoff and Robert G. Wisotzkey contributed equally to this work.

Electronic supplementary material The online version of this article (doi:10.1007/s00239-008-9159-4) contains supplementary material, which is available to authorized users.

Introduction

Posttranslational modifications are a common aspect of eukaryotic biochemistry. Modifications affecting lysine residues such as ubiquitination and sumoylation modulate the activities of numerous proteins including TGFβ and Wnt pathway components. The classic ubiquitin pathway resulting in protein degradation begins with the attachment of the carboxy-terminal glycine of ubiquitin to a cysteine in an E1 activating enzyme. E1 then transfers the ubiquitin to E2, the conjugating enzyme, via a second Gly-Cys bond. The E3 ligase then transfers the ubiquitin from the E2 to a target protein's lysine. A chain of ubiquitin molecules can be created by connecting Lys^{48} of one molecule to the carboxy-terminal glycine of another, yielding a series of Lys-Gly linkages. When a polyubiquitin chain is composed of at least four lysine residues, recognition and degradation of the protein can occur (Meinnel et al. 2006).

Alternatively, if interubiquitin linkages occur at Lys^{63} (rather than Lys^{48}), then polyubiquitination can stimulate a protein's endogenous activity. For example, in the NF-κB pathway, interaction between TAB2 and Lys^{63}-polyubiquitinated TRAF6, RIP, and NEMO appears essential for pathway activation (Kanayama et al. 2004). In addition, monoubiquitination is emerging as a key mechanism for regulating protein function. For example, the transcription factor p53 is normally directed out of the nucleus by monoubiquitination. However, proteins activated by DNA damage such as ARF inhibit p53 ubiquitination, allowing p53 to accumulate in the nucleus, where it inhibits the proliferation of damaged cells (Salmena and Pandolfi 2007).

Like ubiquitination, sumoylation is the attachment of a SUMO molecule to a lysine. E1, E2, and E3 enzymes perform analogous functions in both processes. Sumoylation occasionally confers protein stability by competitively inhibiting Lys^{48}-polyubiquitination. More frequently, like monoubiquitination, it influences a protein's endogenous activity. Unlike ubiquitination, for which no consensus amino acid sequence has been identified, sumoylation by Ubc-9, an E2 conjugating enzyme, occurs at a ΨKxE motif (Yang et al. 2006), though other sumoylation sites exist as well (Long et al. 2004).

The current model for TGFβ signal transduction involves two types of transmembrane receptor serinethreonine kinases. The Type II receptor binds the TGFβ ligand and phosphorylates a Type I receptor. Then the Type I receptor phosphorylates a receptor-associated Smad protein (R-Smad), causing the R-Smad to form a complex with a Co-Smad. The Smad complex translocates to the nucleus and regulates gene expression. The functions of Smad proteins are accomplished via highly conserved regions: the MH1 domain mediates DNA binding and the MH2 domain modulates protein-protein interactions. In addition to R-Smads and Co-Smads, the Smad family contains I-Smads, which antagonize TGFβ signaling and a subfamily that functions only in the Daf pathway of *C. elegans* (Newfeld and Wisotzkey 2006).

Numerous proteins have been proposed as TGFβ pathway ubiquitin ligases including Ectodermin/Tif1γ (Dupont et al. 2005), $SCF^{\beta\text{-TrCP1}}$ (Wan et al. 2004), Smurfs (Morén et al. 2005), and eIF4A (Li and Li 2006). Negative regulation of the pathway by polyubiquitination has been reported for both Type I receptors and Smads. For example, Smurf-mediated polyubiquitination of the Alk-4 receptor is reported to terminate TGFβ signaling (Yamaguchi et al. 2006). Alternatively, monoubiquitination of Lys^{507} in human Smad4 (a Co-Smad) accentuated signaling by enhancing Smad4's ability to form complexes with R-Smads (Morén et al. 2003). Sumoylation also plays positive and negative roles in the TGFβ pathway. When Smad4 is sumoylated at Lys^{113} or Lys^{159} it stimulates its transcriptional activity in HeLa cells (Lin et al. 2003) but represses transcription in COS cells (Long et al. 2004).

The current model for canonical Wnt signal transduction begins with Frizzled seven-pass transmembrane receptors. Upon Wnt binding a Frizzled receptor activates a Dishevelled signal

transducer. Dishevelled then relays the signal to a cytoplasmic protein complex that includes GSK3β, APC, Axin, and βcatenin. Under nonsignaling conditions the other proteins shunt βcatenin to the ubiquitination pathway and destruction. Upon a Dishevelled signal, βcatenin is released from the complex, enters the nucleus, and activates gene expression. Dishevelled proteins also have highly conserved domains: the DIX domain mediates interactions with Axin, the PDZ domain facilitates other protein-protein interactions, and the DEP domain has an unknown function (Wallingford and Habas 2005). In the Wnt pathway, negative regulation of βcatenin by ubiquitination has been extensively studied (e.g., Kitagawa et al. 1999; Hino et al. 2005). However, little is known about ubiquitination or sumoylation of Frizzled receptors and Dishevelled signal transducers. We found a report that human Dishevelled-3 is a target for Lys^{48} polyubiquitination and degradation in HEK293 cells (Angers et al. 2006) but no reports on Frizzled receptors. We could not find any papers on sumoylation of Frizzleds or Dishevelleds.

Ubiquitination and sumoylation are reversible modifications and several ubiquitin proteases (deubiquitinases) have recently been identified (e.g., Nijman et al. 2005). In the TGFβ pathway, UCH37 is capable of deubiquitinating the TGFβ Type I receptor via a complex that includes Smad7 in mammalian cells and UCH37 activity serves to promote TGFβ signaling (Wicks et al. 2005). In the Wnt pathway, FAM deubiquitinates βcatenin, preventing its degradation in mammalian cells, but the effect of FAM activity on Wnt signaling was not noted (Taya et al. 1999). Recently, Trabid was shown to deubiquitinate Lys^{63}-polyubiquitin chains on APC in mammalian cells and Trabid activity served to promote the transcription of Wnt pathway target genes in flies and mammalian cells (Tran et al. 2008).

Numerous questions about the role of lysine modification in these two pathways remain. For example, does each signaling pathway have an individual lysine modification regime, or are there common modalities that, when characterized, could provide a basis for more accurate predictions of targeted lysines? Here we address these questions phylogenetically by identifying all absolutely conserved lysine residues in TGFβ and Wnt receptors and signal transducers. This allows us to assess the content and conservation of their amino acid context. For these specific pathways, our analysis identified numerous conserved lysines not previously implicated in ubiquitination or sumoylation, suggesting that additional regulatory events have yet to be discovered. In the bigger picture, our analysis showed that signal transducers are more frequent targets for posttranslational modification than receptors in both pathways, perhaps a general feature of signaling pathway biochemical regulation.

Materials and Methods

Sequences

Protein sequences from *Caenorhabditis elegans* (Ce), *Drosophila melanogaster* (Dm), and *Mus musculus* (Mm) were retrieved from NCBI.

TGFβ Type I receptors were as follows: Ce Sma-6, AAC46790; Ce Daf-1 iso-a, AAP82657; Ce Daf-1 iso-b, AAC19189; Dm Sax iso-a, AAF59189; Dm Sax iso-b, NP_724606; Dm Babo iso-a, AAF59011; Dm Babo iso-b, AAM71094; Dm Tkv iso-a, NP_787989; Dm Tkv iso-b, NP_787991; Dm Tkv iso-c, NP_787990; Dm Tkv iso-d, NP_787992; Mm Alk-1, CAA83484; Mm Alk-2, NP_031420; Mm Alk-3, NP_033888; Mm Alk-4, NP_031421; Mm Alk-5, NP_033396; Mm Alk-6, NP_031586; and Mm Alk-7, NP_001028541.

TGFβ Type II receptors were as follows: Ce Daf-4 iso-a, AAC02726; Ce Daf-4 iso-c, AAN63460; Ce Daf-4 iso-d, AAO61443 (there is no Ce Daf-4 iso-b in www.worm base.org); Dm Wit, NP_524692; Dm Punt, NP_731926; Mm ActR-IIA, NP_031422; Mm ActR-IIB, NP_031423; Mm BmpR-II, NP_031587; Mm TBR-II, Q62312; and Mm MIS-II, Q8K592.

Smads were as follows: Ce Sma-2, NP_498931; Ce Sma-3, NP_498493; Ce Sma-4, NP_001040864; Ce Daf-3 iso-a, AAK68348; Ce Daf-3 iso-b, AAM54188; Ce Daf-3 iso-c, AAM54189; Ce Daf-8, NP_492321; Ce Daf-14, NP_501880; Ce 1L81 (Tag-68), NP_492746; Dm Med iso-a, NP_524610; Dm Med iso-b, NP_733438; Dm dSmad2, NP_511079; Dm Mad, NP_477017; Dm Dad, AAN13728; Mm Smad1, AAG41407; Mm Smad2 iso-a, EDL09484; Mm Smad2 iso-b, EDL09485; Mm Smad3, AAB81755; Mm Smad4 iso-a, EDL09559; Mm Smad4 iso-b, EDL09560; Mm Smad4 iso-c, EDL09561; Mm Smad5, AAC83580; Mm Smad6, AAB81351; Mm Smad7 iso-a, EDL09492; Mm Smad7 iso-b, EDL09493; and Mm Smad8/9, AAN85445.

Frizzled receptors were as follows: Ce Fz-1 (Mig-1), AAF60492; Ce Fz-2, ABA18181; Ce Lin-17, AAF36028; Ce Mom-5, AAC47750; Dm Fz-1, NP_524812; Dm Fz-2, AAF49184; Dm Fz-3, AAF45547; Dm Fz-4, AAN09202; Mm Fz-1, NP_067432; Mm Fz-2, AAH55727; Mm Fz-3, AAC52429; Mm Fz-4, NP_032081; Mm Fz-5, NP_073558; Mm Fz-6, NP_032082; Mm Fz-7, NP_032083; Mm Fz-8, NP_032084; Mm Fz-9, AAD27789; and Mm Fz-10, NP_780493.

Dishevelleds are as follows: Ce Dsh-1, AAM98048; Ce Dsh-2, NP_494937; Ce Mig-5 iso-c, NP_1022318; Dm Dsh, AAA20216; Mm Dvl-1, AAA74049; Mm Dvl-2, CAI35165; and Mm Dvl-3, NP_031915.

Phylogenetics

In the analysis we employed all isoforms of each protein to ensure that we were able to identify absolute conservation for each lysine. Protein sequences were aligned with ClustalW with all parameters set to default values in Mega 4.0 (Tamura et al. 2007). Two exceptions were a gap-opening penalty of 3.0 and a gap-extension penalty of 1.8 for multiple alignments. Gap only sites were deleted and absolutely conserved lysine residues were identified. In the analysis we identify only lysines present in the same position in all sequences of a particular protein. There are 796 homologous positions in the TGFβ Type I receptor alignment of 18 sequences, with a total of 518 lysines. There are 1172 homologous positions in the TGFβ Type II receptor alignment of 10 sequences, with a total of 389 lysines. There are 892 homologous positions in the Smad alignment of 26 sequences, with a total of 584 lysines. There are 997 homologous positions in the Frizzled alignment of 18 sequences, with a total of 413 lysines. There are 1003 homologous positions in the Dishevelled alignment of 7 sequences, with a total of 416 lysines. A total of 79 sequences were analyzed. Alignments are available upon request.

Potential sumoylation sites matching the Ubc-9 consensus ΨKxE motif were identified utilizing SUMOplot (www.abgent.com.cn/doc/sumoplot/login.asp). Candidate sumoylation sites associated with absolutely conserved lysine are reported regardless of the predicted probability. Alternatively, sites associated with nonconserved lysines are reported with a likelihood score of 66% or greater (medium probability in SUMOplot). Sequences were also analyzed for a nonconsensus sequence sumoylated in human Smad4 (VKYC-Lys^{113} [Long et al. 2004]) but none were found.

Results

To achieve maximum confidence we focused our analysis on species belonging to distinct phyla. We chose two coelomates, animals with three germ layers and a digestive tract with two openings: mice (*M. musculus*; a deuterostome—the blastopore becomes the anus) and flies (*D. melanogaster*; a protostome—the blastopore becomes the mouth). The third species is a nematode (*C. elegans*; a pseudocoelomate—animals with three germ layers and a digestive tract with one opening). The split between deuterostomes and protostomes was roughly 990

million years ago and that between coelomates and pseudocoelomates was about 1.2 billion years ago (Hedges and Kumar 2003).

During the analysis we determined that the Wnt pathway might be slightly older than the TGF*β* pathway. Wnts, Frizzleds, and Dishevelleds are present in a single sponge species, suggesting that this pathway is functional there (Nichols et al. 2006). Sponges are the simplest multicellular animals (a few cell types in a single germ layer) and the split with other animals occurred roughly 1.5 billion years ago (Hedges and Kumar 2003). Alternatively, TGF*β* ligands, receptors, and an R-Smad have been reported in three different sponge species (Nichols et al. 2006; Suga et al. 1999; Adamska et al. 2007), but to date these proteins have not been found in the same sponge species. This implies that TGF*β* signaling is not functional in sponges and pinpoints an organism that needs further study to fully understand the origin of this pathway. The sea anemone *N. vectensis* (a diploblast—an animal with two germ layers [Putnam et al. 2007]) contains a complete TGF*β* pathway (including the three main subfamilies of Smads) and currently is the simplest organism in which both pathways parallel those of flies and mammals.

Conserved Lysines in TGF*β* Receptors and Smads

The TGF*β* Type I receptor Dm Sax iso-a contains 24 lysines, of which 5 (21%) are conserved. All of the conserved lysines occur in the kinase domain (Fig. 1a). Three occur in a cluster (Lys^{385}, Lys^{394}, and Lys^{396}) and one is near the carboxy terminus. Each of these is adjacent to another invariant amino acid and within two residues of a hydrophobic amino acid (Table 1). Lys^{291} in Dm Sax iso-a occurs in a low-probability sumoylation sequence (VKIF; score, 58%). Numerous nonconserved sumoylation sites were predicted (Supplementary Table 1).

The TGF*β* Type II receptor Dm Wit contains 30 lysines, of which 2 (7%) are conserved. Both of the conserved lysines occur in the kinase domain (Fig. 1b). Lys^{251} in Dm Wit occurs in a nearly invariant region with two hydrophobic residues (Table 1). Lys^{347} in Dm Wit occurs in a nonconserved region. Numerous nonconserved sumoylation sites were predicted (Supplementary Table 1).

For Smads, we first confirmed our data (Newfeld and Wisotzkey 2006) that there are only eight Smads in mammals (Smad8 and Smad9 are the same). Then we confirmed the report of Chang et al. (2006) that Smad10 in *Xenopus laevis* is a variant of Smad4 iso-b (these two sequences have > 99% nucleotide identity). The alignment showed that all Smads contain a single conserved lysine in the MH2 (Fig. 2a). Examining the amino acid context for Lys^{726} in Dm Medea iso-a revealed that three of the six surrounding residues are absolutely conserved and an upstream adjacent hydrophobic residue is present in all sequences except the Ce Daf pathway subfamily (Table 1).

The Co-Smad Dm Medea iso-a contains 19 lysines. Two of these are conserved (11%), including the universal Smad lysine, and both occur in the MH2 (Fig. 2b). Examining the context for Co-Smad specific conserved lysine (Lys^{738} in Dm Medea iso-a) revealed that five of the six surrounding residues are absolutely conserved, including an upstream adjacent hydrophobic residue (Table 1). Numerous nonconserved sumoylation sites were predicted (Supplementary Table 1).

The R-Smad Dm Mad contains 20 lysines, of which 7 are conserved (35%), including the universal Smad lysine (Fig. 2c). Six conserved lysines are in the MH1. In Dm Mad, four of these lysines (Lys^{46}, Lys^{53}, Lys^{54}, and Lys^{95}) are also present in Ce Daf-3—an antagonist Smad in the Ce Daf pathway subfamily that functions via a transcriptional mechanism like an R-Smad. In addition, an eighth lysine present in Ce Daf-3 but not in the R-Smad Ce Sma-3 falls in the MH2 (Lys^{362}). Examining the context for each R-Smad specific lysine revealed

that, except for Lys^{35}, each is associated with invariant amino acids and conserved hydrophobic residues either adjacent or one amino acid distant (Table 1). A single sumoylation site was predicted in each fly and mouse TGFβ/Activin subfamily Smad but the sites are not homologous: Dm Smad2 Lys^{126}, Mm Smad2 iso-a Lys^{126}, Mm Smad2 iso-b Lys^{156}, and Mm Smad3 Lys^{116}.

The I-Smad Dm Dad contains 18 lysines but only the universal Smad lysine (6%) is conserved (Fig. 2d). However, an additional lysine in the MH2 lies in a highly conserved context in Ce Daf-3, fly, and mammalian I-Smads but not in Ce 1L81 (Table 1). Lys^{500} in Dm Dad occurs in a low-probability sumoylation site (LKAF; score, 56%).

The Ce Daf pathway Smad Ce Daf-8 contains 22 lysines. Two of these are conserved (9%) including the universal Smad lysine in the MH2 (Fig. 2e). Lys^{462} in Ce Daf-8 occurs in a low-probability sumoylation site (MKVF; score, 45%). Numerous nonconserved sumoylation sites were predicted (Supplementary Table 1).

Conserved Lysines in Frizzled Receptors and Dishevelleds

The Frizzled receptor Dm Frizzled-1 contains 19 lysines, and 1 (5%) is conserved (Fig. 3a). This lysine is found in an intracellular loop between transmembrane domain 3 and transmembrane domain 4 in all sequences. Lys^{369} in Dm Frizzled-1 occurs in a medium-probability sumoylation site (LKWG; score, 73%). A sumoylation site is also found in the homologous position in Ce Mom-5, Dm Frizzled-2, and all Mm Frizzled sequences except Mm Frizzled-6 (scores, 31–73%). Numerous nonconserved sumoylation sites were predicted (Supplementary Table 1).

The Dishevelled protein Dm Dishevelled contains 18 lysines, of which 6 (33%) are conserved (Fig. 3a). In Dm Dishevelled, three (Lys^{41}, Lys^{52}, and Lys^{66}) occur in the DIX domain, Lys^{338} occurs in the PDZ domain, and Lys^{465} occurs in the DEP domain. Lys^{66} in the Dm Dishevelled DIX domain occurs in an absolutely conserved high-probability sumoylation site (VKEE; score, 93%; Table 1). Each of the other lysines is associated with one or more invariant amino acids and all but Lys^{52} have hydrophobic residues nearby. Numerous nonconserved sumoylation sites were predicted (Supplementary Table 1).

Discussion

We propose that phylogenetic analysis of lysine conservation and context can fill the existing gap between easily obtainable immunological evidence that a protein is subject posttranslational modification and laborious identification of the target lysine. To confirm the validity of our approach we discuss several instances of correspondence between our data and experimentally demonstrated sites of lysine posttranslational modification. Alternatively we assess an obvious caveat to our data, sites in which lysine conservation is due to structural rather than posttranslational considerations. We then compare our findings for the two pathways to reveal common features that suggest general principles for the biochemical regulation of intercellular signaling pathways. Finally, we discuss the applicability of computational biochemistry to a number of outstanding issues in protein evolution, regulation, and function.

Conserved Lysines in the TGFβ Pathway

TGFβ Type I receptors possess five conserved lysines in the kinase domain. It is known that Type I receptors are polyubiquitinated and targeted for degradation (e.g., Yamaguchi et al. 2006) but the affected lysines have not been identified. An examination of the crystal structure for human Alk5 suggests that three of the conserved lysines (Mm Alk5 Lys^{232}, Lys^{335}, and

Lys^{337}) have internal locations near the ATP binding pocket and catalytic domain, respectively (Huse et al. 1999). The other two lysines (Mm Alk5 Lys^{326} and Lys^{490}) are exposed and remain candidates for ubiquitination targets. Type II receptors contain two conserved lysines in the kinase domain. A recent report of the crystal structure of human Act R-IIB indicates they are both deeply buried within the protein and inaccessible to modification (Han et al. 2007).

All Smads contain a single conserved lysine in the MH2 domain. In addition, three of the six surrounding residues are absolutely conserved and an upstream adjacent hydrophobic residue is present in all sequences except members of the Ce Daf pathway subfamily. This lysine in human Smad4 (Lys^{507}) is known be a major site of ubiquitination in HEK293 cells (Morén et al. 2003). The concordance of conservation and experimental data suggests that this lysine in all Smads is a strong candidate for regulation by ubiquitination.

Co-Smads contain two conserved lysines including the universal Smad lysine in the MH2 domain. Five of the six surrounding residues for Lys^{738} in Dm Medea are absolutely conserved, including an upstream adjacent hydrophobic residue. The conservation of the context for Lys^{738} in Dm Medea (homologous to human Smad4 Lys^{519}) is even more impressive than that of the universal Smad Lys. This suggests, in the strongest possible terms, that the Co-Smad specific lysine is also regulated by ubiquitination.

From this perspective we should note that recent experiments in our lab also implicate ubiquitination in TGFβ signaling. In a standard *Drosophila* assay (e.g., Nicholls and Gelbart 1998) we found that an allele of *fat facets*, a deubiquitinase, acts as a maternal enhancer of the embryonic lethality associated with recessive mutation in *dpp*, a TGFβ family member (S.J.N., unpublished data). The aspect of *dpp* signaling surveyed in this assay involves Mad and Medea and a potential explanation is that Fat Facets deubiquitinates the universal Smad lysine in Mad or Medea or perhaps the Co-Smad specific lysine (Lys^{738} in Medea).

Neither of the conserved lysines in Co-Smads lies within a sumoylation consensus. However, Lys^{185} in Dm Medea iso-a (score, 93%) is not conserved in Ce Smad4 but is conserved as Lys^{159} in human Smad4, a residue shown to be sumoylated in COS cells (VKDE [Long et al. 2004]). Lys^{113} in human Smad4 is also sumoylated in COS cells (VKYC [Long et al. 2004]) and conserved in Dm Medea iso-a (Lys^{141}) but not in Ce Sma-4. When human Smad4 is sumoylated at Lys^{113} or Lys^{159} it represses this function in COS cells (Long et al. 2004) but it stimulates transcriptional activity in HeLa cells (Lin et al. 2003). Given that the contexts for both lysines are absolutely conserved between human Smad4 and Dm Medea, this suggests, again in the strongest possible terms, that both lysines are sumoylated in Dm Medea.

R-Smads contain eight conserved lysines including the universal Smad lysine. Six of these are in the MH1 domain and two in the MH2 domain. Notwithstanding an extensive literature on ubiquitination and sumoylation of R-Smads, we could not find a single report that identifies the targeted lysine. Thus, the universal Smad lysine shown to be ubiquitinated in human Smad4 is a strong candidate for ubiquitination in all R-Smads. The crystal structure of phosphorylated human Smad2 showed that the second MH2 lysine (Lys^{362} in Dm Mad) is also surface exposed (Wu et al. 2001). Several of the lysines in the MH1 are viable candidates for posttranslational modification. Lys^{53} and Lys^{57} in Dm Mad are exposed, as shown in the crystal structure of the Smad3 MH1 bound to DNA. Alternatively, Lys^{95} in Dm Mad is a DNA binding residue and likely conserved for structural reasons. The three remaining MH1 lysines (Lys^{35}, Lys^{46}, and Lys^{54} in Dm Mad) are partially exposed and their accessibility to modification is difficult to predict (Shi et al. 1998).

I-Smads contain only the universal Smad lysine. Several studies demonstrate ubiquitination of Smad7 but none report the affected lysine. Again, we propose the universal Smad lysine as the primary candidate. Ce Daf pathway Smads contain two lysines including the universal Smad

lysine, and the Ce Daf pathway specific lysine matches the sumoylation consensus. We could find no reports of posttranslational modification for this Smad subfamily but our data are suggestive.

Conserved Lysines in the Wnt Pathway

Frizzled receptors contain a single conserved lysine that resides in an intracellular loop between transmembrane domain 3 and transmembrane domain 4. Further, this lysine resides in a sumoylation consensus site in virtually all Frizzled sequences. We could find neither reports of posttranslational modification nor a crystal structure for any Frizzled cytoplasmic domain. Thus, it is difficult to make predictions but a conserved lysine in an intracellular loop that may be enzymatically accessible suggests that post-translational modification is a possibility.

Dishevelled proteins contain six conserved lysines. We found one report that human Dishevelled-3 is a target for polyubiquitination and degradation in HEK293 cells (Angers et al. 2006) but the specific lysine was not identified. An examination of the crystal structure for Mm Dishevelled-2 suggests that the three DIX domain lysines (Lys^{44}, Lys^{54}, and Lys^{68} in Mm Dishevelled-2) are exposed in αhelix1, *β*sheet3, and *β*sheet4, respectively (Schwarz-Romond et al. 2007). Examining the context for each lysine revealed that the Lys^{68} in the Mm Dishevelled-2 DIX domain occurs in an absolutely conserved and high-probability sumoylation site (VKEE; score, 93%). We could find no reports of sumoylation for Dishevelleds but the presence of a conserved lysine in a high-probability site is strongly suggestive. This suggestion is supported by the fact that mutation of this lysine eliminates signaling (Schwarz-Romond et al. 2007).

Common Features of Lysine Conservation in the TGF*β* and Wnt Pathways

With regard to lysine conservation, in both pathways signal transduction agonists (R-Smads and Dishevelleds) have more conserved lysines in absolute terms and a greater proportion of their total lysine content is conserved than in either receptors or other signal transducers (I-Smads, Co-Smads, and Ce Daf pathway Smads). After correcting for structural considerations, R-Smads have seven and Dishevelled proteins have six conserved lysine candidates for posttranslational modification, versus two or fewer candidates for TGF*β* Type I and Type II receptors and Frizzleds. While a sample of two pathways is not conclusive, it suggests that biochemical regulation of agonist signal transducers is more frequently employed as a mechanism for influencing intercellular signaling than posttranslational regulation of receptors or other classes of signal transducers.

With regard to lysine context, in both pathways conserved lysines are often associated with conserved adjacent upstream hydrophobic residues. Results for all proteins identified 25 distinct conserved lysines in the analysis (e.g., the universal Smad lysine is counted only once). Of these, 6 are most likely inaccessible to modification. Of the remaining 19 lysines, 9 have a conserved hydrophobic amino acid immediately upstream—including the universal Smad lysine shown to be ubiquitinated (Morén et al. 2003). Numerous proteins shown to be ubiquitinated, but whose target lysine is unidentified (e.g., TGF*β* Type I receptors, Smads, and Dishevelleds), have protein specific lysines with a conserved upstream adjacent hydrophobic residue.

While a sample of one (the VK pair of the ubiquitinated universal Smad lysine) is too small to generate confidence, in the absence of any other clues, we would consider each of these conserved hydrophobic-lysine pairs as top candidates for posttranslational modification (or in the case of Smads, considered immediately after the universal Smad lysine). The pairs are associated with Lys^{549} in Dm Sax iso-a, Lys^{738} in Dm Med iso-a, Lys^{53} in Dm Mad, Lys^{500} in Dm Dad, Lys^{462} in Ce Daf-8, Lys^{41} and Lys^{66} in Dm Dishevelled, and Lys^{369} in Dm

Frizzled1. If ubiquitination is confirmed for any of these hydrophobic-lysine pairs, a conserved adjacent upstream hydrophobic residue may be the first predictive context for ubiquitination.

From an evolutionary perspective, a common feature of the two pathways is that receptors and signal transducers in mice and flies are more similar to each other than either is to *C. elegans* proteins. For example, our alignments show that in the Frizzled family Ce Mom-5 is distinct from all other family members and in the Dishevelled family the three *C. elegans* proteins cluster together. This finding for the Wnt pathway is consistent with our analyses of the TGF*β* pathway (Newfeld and Wisotzkey, 2006; Newfeld et al. 1999) and defies the logic of an ecdysozoan phylum. To further support this interpretation, a complete phylogenetic analysis of the Wnt pathway is under way (C.K., R.G.W., and S.J.N).

Application of Phylogenetics to Biochemical Regulation of Developmental Pathways

The wealth of potential targets we predict and the confirmation of several predicted targets for mammalian Smad4 suggest that this approach can be applied to a number of open questions on the evolution, regulation, and function of multigene families. First, ubiquitination and sumoylation of proteins are typically identified in cell culture assays with ubiquitin and sumo antibodies leaving the target lysine in the protein unknown. Our approach provides a rational method for prioritizing lysines for analysis, streamlining the current tedious method of random (especially in the case of ubiquitination, where no consensus exists) single-lysine mutagenesis to identify targets.

Second, for those lysines that have been identified in cell culture, a phylogenetic analysis can suggest how wide-spread that modification is. For example, a strongly conserved lysine with a strongly conserved context is likely an ancient target and important in other organisms. An example is the universal lysine in Smads, and conservation information could serve as a point of departure for a mutational analysis of the homologous lysine in a model organism. Alternatively, modestly conserved or even unconserved lysines that are modified may be species specific, pathway specific, or cell type specific targets. An example is a report of the sumoylation of the human TGF*β* Type I receptor Alk-5 (also known as TBRI [Kang et al. 2008]).

In our alignments this sumoylated lysine (Lys^{392} in Mm Alk-5) is present in Mm Alk-5, Mm Alk-4, all isoforms of Dm Tkv, and Ce Sma-6. This pattern of conservation (presence in worms, flies, and mammals) suggests that sumoylation of this lysine is an ancient mechanism of regulation for the TGF*β* Type I receptor family. This information could serve as a point of departure for an analysis of Type I receptor sumoylation in flies and worms. Interestingly, human TBRI is more closely related to Dm Babo and Dm Sax, neither of which have the conserved lysine, than to Dm Tkv (Newfeld et al. 1999). This suggests that the sumoylated lysine was lost independently three times: from Dm Babo, from the Mm Alk-3/Mm Alk-6 pair, and from the cluster containing Dm Sax, Mm Alk-1, and Mm Alk-2. Thus, it appears that lysine conservation, when coupled with experimental data, can be employed to identify when closely related proteins gain (or lose) a particular mechanism of biochemical regulation.

At the most general level, phylogenetic biochemistry can be employed to predict posttranslational modification of lysines in any conserved protein family. More importantly, it can be applied to any posttranslational modification that targets a specific amino acid. In summary, our data suggest that potential applications of phylogenetic methods in biochemistry are limited only by the investigator's imagination.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Jay Groppe and David Wotton for valuable discussions. C.E.K. thanks Sudhir Kumar for support (R01HG002516). This work was also supported by grants to S.J.N. from the NIH (R01CA095875) and SFAz (CAA0285-08).

References

Adamska M, Degnan S, Green K, Adamski M, Craigie A, Larroux C, Degnan B. Wnt and TGF*β* Expression in the sponge Amphimedon queenslandica and the origin of metazoan embryonic patterning. PLoS ONE 2007;2:e1031. [PubMed: 17925879]

Angers S, Thorpe C, Biechele T, Goldenberg S, Zheng N, MacCoss M, Moon R. The KLH12-Cullin-3 ubiquitin ligase negatively regulates the Wnt-ßcatenin pathway by targeting Dishevelled for degradation. Nat Cell Biol 2006;8:348–357. [PubMed: 16547521]

Brummel T, Twombly V, Marqués G, Wrana J, Newfeld S, Attisano L, Massagué J, O'Connor M, Gelbart W. Characterization and relationship of Dpp receptors encoded by the *saxophone* and *thick veins* genes in *Drosophila*. Cell 1994;78:251–261. [PubMed: 8044839]

Chang C, Brivanlou A, Harland R. Function of the two *Xenopus* Smad4 s in early frog development. J Biol Chem 2006;281:30794–30803. [PubMed: 16908518]

Dupont S, Zacchigna L, Cordenonsi M, Soligo S, Adorno M, Rugge M, Piccolo S. Germ-layer specification and control of cell growth by Ectodermin, a Smad4 ubiquitin ligase. Cell 2005;121:87–99. [PubMed: 15820681]

Han S, Loulakis P, Griffor M, Xie Z. Crystal structure of activin receptor Type IIB kinase domain from human at 2A resolution. Protein Sci 2007;16:2272–2277. [PubMed: 17893364]

Hedges S, Kumar S. Genomic clocks and timescales. Trends Genet 2003;19:200–206. [PubMed: 12683973]

Hino S, Tanji C, Nakayama K, Kikuchi A. Phosphorylation of *β*catenin by cAMP-PK stabilizes *β*catenin through inhibition of ubiquitination. Mol Cell Biol 2005;25:9063–9072. [PubMed: 16199882]

Huse M, Chen Y, Massagué J, Kuriyan J. Crystal structure of the cytoplasmic domain of the Type I TGF*β* receptor in complex with FKBP12. Cell 1999;96:425–436. [PubMed: 10025408]

Kanayama A, Seth R, Sun L, Ea C, Hong M, Shaito A, Chiu Y, Deng L, Chen Z. TAB2 and TAB3 activate the NF-*κ*B pathway through binding to polyubiquitin chains. Mol Cell 2004;15:535–548. [PubMed: 15327770]

Kang J, Saunier E, Akhurst R, Derynck R. The Type I TGF*β* receptor is covalently modified and regulated by sumoylation. Nat Cell Biol 2008;10:654–664. [PubMed: 18469808]

Kitagawa M, Hatakeyama S, Shirane M, Matsumoto M, Ishida N, Hattori K, Nakamichi I, Kikuchi A, Nakayama K, Nakayama K. An F-Box protein, FWD1, mediates ubiquitin-dependent proteolysis of *β*catenin. EMBO J 1999;18:2401–2410. [PubMed: 10228155]

Li J, Li W. A novel function of *Drosophila* eIF4A as a negative regulator of Dpp/BMP signaling that mediates SMAD degradation. Nat Cell Biol 2006;8:1407–1414. [PubMed: 17115029]

Lin X, Lang M, Liang Y, Brunicardi F, Feng X. Sumo-1/Ubc-9 promotes nuclear accumulation and metabolic stability of Smad4. J Biol Chem 2003;278:31043–31048. [PubMed: 12813045]

Long J, Wang G, He D, Liu F. Repression of Smad4 transcriptional activity by Sumo modification. Biochem J 2004;379:23–29. [PubMed: 14750902]

Meinnel T, Serero A, Giglione C. Impact of the N-terminal amino acid on targeted protein degradation. Biol Chem 2006;387:839–851. [PubMed: 16913833]

Morén A, Hellman U, Inada Y, Imamura T, Heldin CH. Differential ubiquitination defines the functional status of Smad4. J Biol Chem 2003;278:33571–33582. [PubMed: 12794086]

Morén A, Imamura T, Miyazono K, Heldin CH, Moustakas A. Degradation of the tumor suppressor Smad4 by WW and HECT ubiquitin ligases. J Biol Chem 2005;280:22115–22123. [PubMed: 15817471]

Newfeld, S.; Wisotzkey, R. Molecular evolution of Smad proteins. In: Heldin, CH.; ten Dijke, P., editors. Smad signal transduction. Springer; Dordrecht: 2006. p. 15-35.

Newfeld S, Wisotzkey R, Kumar S. Molecular evolution of a developmental pathway: phylogenetic analyses of TGFβ family ligands, receptors and Smad signal transducers. Genetics 1999;152:783–795. [PubMed: 10353918]

Nicholls R, Gelbart W. Identification of chromosomal regions involved in *dpp* function in *Drosophila*. Genetics 1998;149:203–215. [PubMed: 9584097]

Nichols S, Dirks W, Pearse J, King N. Early evolution of animal cell signaling and adhesion genes. Proc Natl Acad Sci USA 2006;103:12451–12456. [PubMed: 16891419]

Nijman S, Luna-Vargas M, Velds A, Brummelkamp T, Dirac A, Sixma T, Bernards R. A genomic and functional inventory of deubiquitinating enzymes. Cell 2005;123:773–786. [PubMed: 16325574]

Povelones M, Howes R, Fish M, Nusse R. Genetic evidence that Frizzled controls planar cell polarity and Armadillo by a common mechanism. Genetics 2006;171:1643–1554. [PubMed: 16085697]

Putnam N, Srivastava M, Hellsten U, Dirks B, Chapman J, Salamov A, Terry A, Shapiro H, Lindquist E, Kapitonov V, Jurka J, Genikhovich G, Grigoriev I, Lucas S, Steele R, Finnerty J, Technau U, Martindale M, Rokhsar D. Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization. Science 2007;317:86–94. [PubMed: 17615350]

Salmena L, Pandolfi P. Changing venues for tumor suppression: balancing destruction and localization by monoubiquitylation. Nat Rev Cancer 2007;7:409–413. [PubMed: 17508027]

Schwarz-Romond T, Fiedler M, Shibata N, Butler P, Kikuchi A, Higuchi Y, Bienz M. The DIX domain of Dishevelled confers Wnt signaling by dynamic polymerization. Nat Struct Mol Biol 2007;14:484–492. [PubMed: 17529994]

Shi Y, Wang Y, Jayaraman L, Yang H, Massagué J, Pavletich N. Crystal structure of a Smad MH1domain bound to DNA: insights on DNA binding in TGFβ signaling. Cell 1998;94:585–594. [PubMed: 9741623]

Suga H, Ono K, Miyata T. Multiple TGFβ receptor related genes in sponge and ancient gene duplications before the parazoan-eumetazoan split. FEBS Lett 1999;453:346–350. [PubMed: 10405173]

Tamura K, Dudley J, Nei M, Kumar S. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol 2007;24:1596–1599. [PubMed: 17488738]

Taya S, Yamamoto T, Kanai-Azuma M, Wood S, Kaibuchi K. The deubiquitinating enzyme Fam interacts with and stabilizes βcatenin. Genes Cells 1999;4:757–767. [PubMed: 10620020]

Tran H, Hamada F, Schwarz-Romond T, Bienz M. Trabid, a new positive regulator of Wnt-induced transcription with preference for binding and cleaving K63-linked ubiquitin chains. Genes Dev 2008;22:528–542. [PubMed: 18281465]

Yamaguchi T, Kurisaki A, Yamakawa N, Minakuchi K, Sugino H. FKBP12 functions as an adapter of Smad7-Smurf1 on Activin Type I receptor. Mol Endocrin 2006;36:569–579.

Yang S, Galanis A, Witty J, Sharrocks A. An extended consensus motif enhances the specificity of substrate modification by Sumo. EMBO J 2006;25:5083–5093. [PubMed: 17036045]

Wallingford J, Habas R. The developmental biology of Dishevelled: an enigmatic protein governing cell fate and cell polarity. Development 2005;132:4421–4436. [PubMed: 16192308]

Wan M, Tang Y, Tytler E, Lu C, Jin B, Vickers S, Yang L, Shi X, Cao X. Smad4 protein stability is regulated by ubiquitin ligase SCFβ-TrCP1. J Biol Chem 2004;279:14484–14487. [PubMed: 14988407]

Wicks S, Haros K, Maillard M, Song L, Cohen R, ten Dijke P, Chantry A. The deubiquitinating enzyme UCH37 interacts with Smads and regulates TGFβ signaling. Oncogene 2005;24:8080–8084. [PubMed: 16027725]

Wieser R, Attisano L, Wrana J, Massagué J. Signaling TGFβ Type II receptors lacking specific domains in the cytoplasmic region. Mol Cell Biol 1993;13:7239–7247. [PubMed: 8246946]

Wu J, Hu M, Chai J, Seoane J, Huse M, Li C, Rigotti D, Kyin S, Muir T, Fairman R, Massagué J, Shi Y. Crystal structure of phosphorylated Smad2: recognition of phosphoserine by the MH2 domain and insights on function in TGFβ signaling. Mol Cell 2001;8:1277–1289. [PubMed: 11779503]

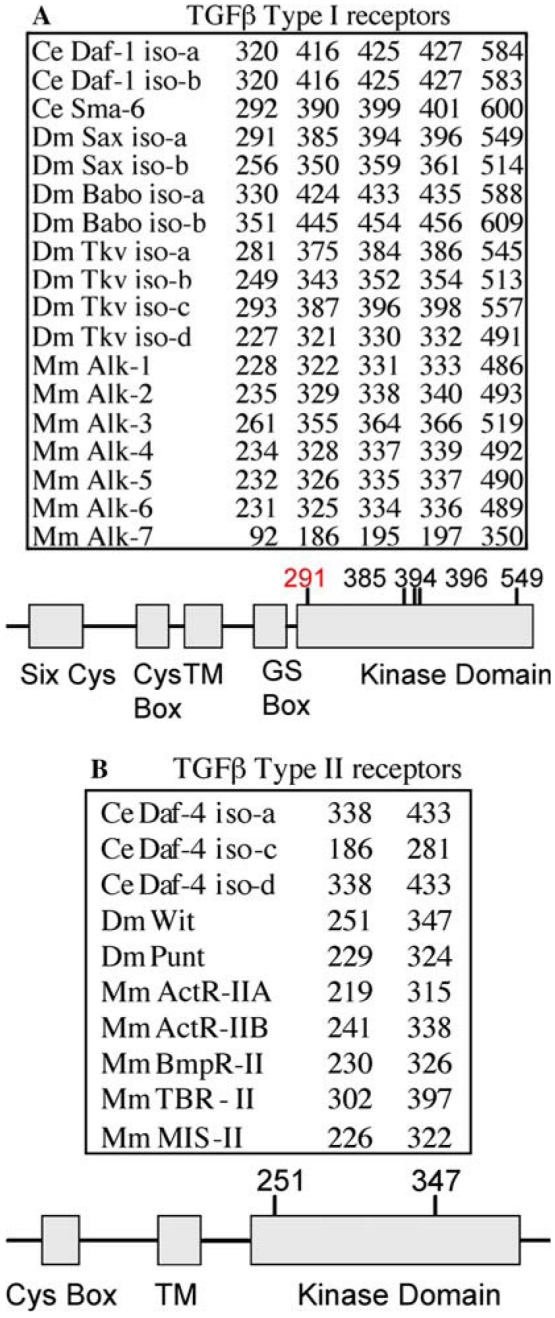

Fig. 1.

Conserved lysines in TGF*β* receptors. (a) Type I receptors contain five conserved lysines, with one in a weak Ubc-9 consensus sumoylation site (red). A table with amino acid numbers for each conserved lysine in all Type I receptors and a schematic of their location in Dm Sax iso-a are shown. Conserved domains for Type I receptors are as described (TM represents the transmembrane domain, with the cytoplasmic kinase domain to the right; [Brummel et al. 1994]). (b) Type II receptors contain two conserved lysines. A table with amino acid numbers for each conserved lysine in all Type II receptors and a schematic of Dm Wit are shown with conserved domains (TM represents the transmembrane domain, with the cytoplasmic kinase domain to the right [Wieser et al. 1993])

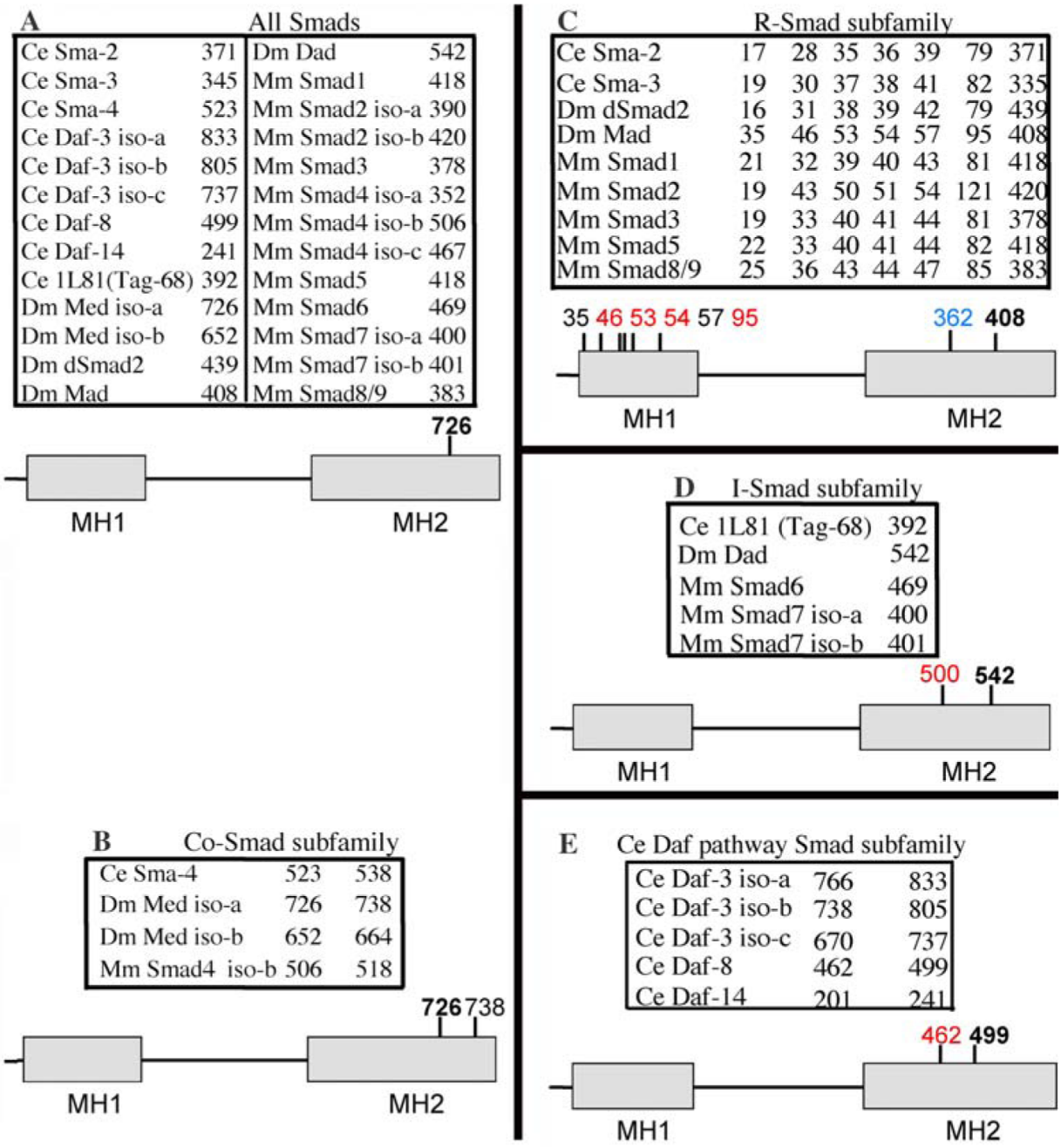

Fig. 2.

Conserved lysines in Smads. (a) All Smads contain a single conserved lysine (boldface). A table with amino acid numbers for the conserved lysine in all Smads and a schematic of Dm Medea iso-a are shown with conserved domains (Newfeld and Wisotzkey 2006). (b) Co-Smads contain two conserved lysines including the universal Smad lysine (boldface). A table with amino acid numbers for each conserved lysine in all Co-Smads and a schematic of Dm Medea iso-a are shown. (c) R-Smads contain seven conserved lysines including the universal Smad lysine (boldface). A table with amino acid numbers for each conserved lysine in all R-Smads and a schematic of Dm Mad are shown. Four of the conserved lysines are also present in Ce Daf-3, an antagonist Smad that functions via a transcriptional mechanism like an R-Smad, and

are shown in red. An additional lysine is present in Ce Daf-3 but not in the R-Smad Ce Sma-3 is shown in blue. (d) I-Smads contain only the universal Smad lysine (boldface). A table with amino acid numbers for each conserved lysine in all I-Smads and a schematic of Dm Dad are shown. An additional conserved lysine present in Ce Daf-3 and fly and mammalian I-Smads but not in Ce 1L81 (Tag-68) that resides in a weak sumoylation site is shown in red. (e) Ce Daf pathway Smads contain two lysines including the universal Smad lysine (boldface). A table with amino acid numbers for each conserved lysine in all Ce Daf pathway Smads and a schematic of Ce Daf-8 are shown. A Ce Daf pathway specific lysine that resides in a weak sumoylation site is shown in red

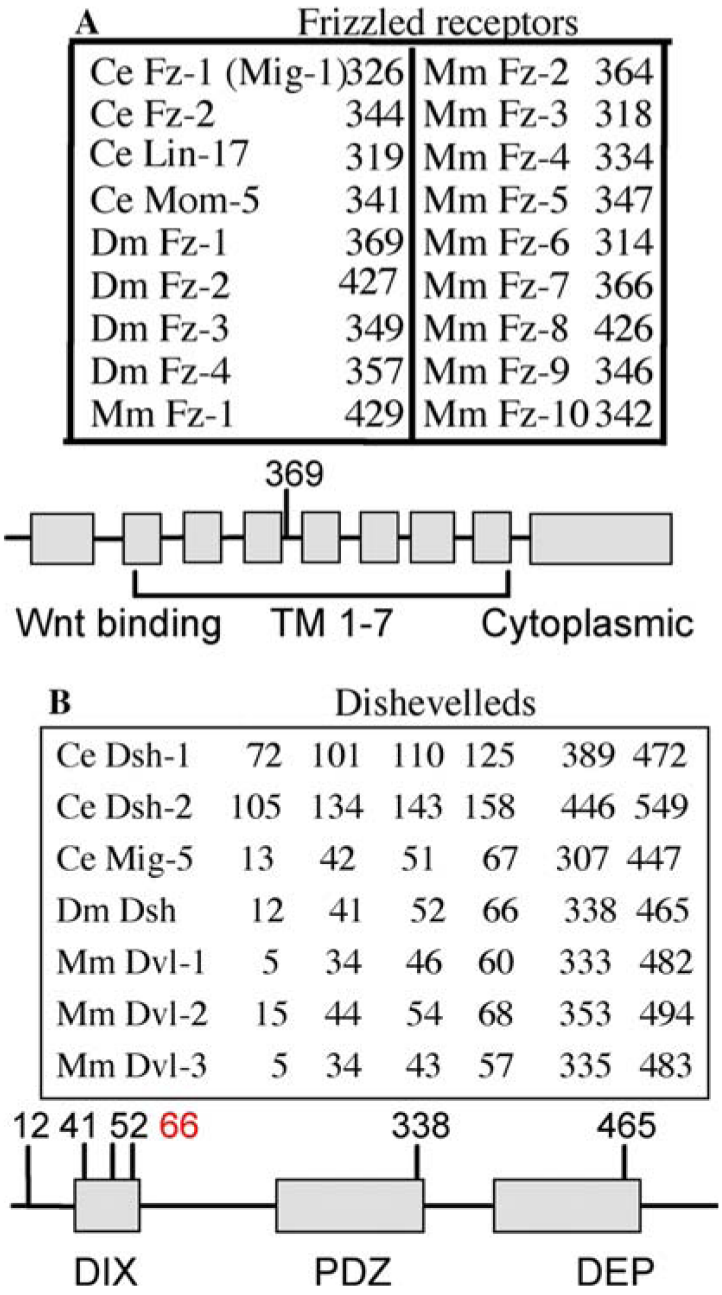

Fig. 3.

Conserved lysines in Frizzled receptors and Dishevelleds. (a) Frizzled receptors contain one conserved lysine in an intracellular loop between transmembrane domain 3 and transmembrane domain 4. A table with amino acid numbers for the conserved lysine in all Frizzled receptors and a schematic of Dm Frizzled-1 are shown with conserved domains (TM 1–7 represent the seven transmembrane domains [Povelones et al. 2006]). (b) Dishevelleds contain six conserved lysines, with three in the DIX domain, one in the PDZ domain, and one in the DEP domain. A table with amino acid numbers for each conserved lysine in all Dishevelleds and a schematic of Dm Dishevelled are shown with conserved domains (Wallingford and Habas 2005)

Table 1

Summary of amino acid numbers and contexts for absolutely conserved lysine residues in TGFβ and Wnt pathway receptors and signal transducers

Protein	Lysine[a]	Lysine Context[b]	Comments[c]
Type I receptors	291	V-A-V-K-V/I/T-F	
	385	G-K-P-A-I/M-A	
	394, 396	I/L-K-S-K-N-I-L	
	549	R-I/V-K-K-T-I/L	only K in Ce Daf-1
Type II receptors	251	V-A-V-K	
	347	K	
All Smads	**726**	S-F-V-K-G-W-G	
Co-Smads	**726**	S-F-V-K-G-W-G	
	738	S-I-K-E-T-P-C	
R-Smads	35	W-K-Q/K	
	46	E-K-A-V	Q-K-A in Ce Daf-3
	53, 54, 57	V-K-K-L-K-K	V-K-K-L-K-D in Ce Daf-3
	95	H-R-K-G-L-P	G-R-K-G-F-P in Ce Daf-3
	362	V-C-K-I/V	in Ce Daf-3 not Ce-Sma-3
	408	S-F-V-K-G-W-G	
I-Smads	500	S/C-I/L-K-A/T-F	in Ce Daf-3 not Ce 1L81
	542	S-F-V-K-G-W-G	
Ce Daf pathway	462	I/M-K-T/I/V-F	
Smads	**499**	S-F-C-K-G-W/F-G	
Frizzled receptors	369	L/M-K-W-G	
Dishevelleds	12	T-K-I/V-I/Y	
	41	D-F/V-K-x-V	
	52	Y/F-K-Y/F-Y/F	
	66	V-K-A/E-E	

Protein	Lysine[a]	Lysine Context[b]	Comments[c]
	338	T-V-A-K-x-x-D	
	465	V-N-K-x-T-F	

[a] Boldface indicates the universal conserved lysine in all Smads. See the figure associated with each protein for the name of the representative sequence from which the lysine number is drawn

[b] Red type indicates the lysine present in all sequences. Underlined amino acids are present in all sequences. Amino acids not underlined are present in all but one or two sequences. Amino acids separated by a slash are present at roughly equal frequencies

[c] Note that Ce Daf-3 is an atypical Smad with features of both I-Smads and R-Smads