J Mol Med (1999) 77: 403–405


Bioinformatics: Bits and Bytes

Alexey Eroshkin · Arcady Mushegian

# Conserved transactivation domain shared by interferon regulatory factors and Smad morphogens

**Abstract** Interferon regulatory factors (IRFs) regulate the transcription of both interferon-inducible genes and interferons themselves. Along with the N-terminal, DNA-binding, winged-helix domain, most IRFs contain the C-terminal domains that are shown to be related to the C-terminal domains in the proteins of the Smad family that mediate transcription activation in the transforming growth factor response pathway. Comparison of the IRF-Smad alignment to the known three-dimensional structure of human tumor suppressor Smad4 suggests that a conserved loop, equivalent to Loop 3 in Smad 4, is a determinant of protein-protein interaction in IRFs.

A. Eroshkin · A. Mushegian (✉)
Department of Computational Sciences,
Axys Pharmaceuticals,
11099 N. Torrey Pines Road, La Jolla,
CA 92037, USA
e-mail arcady@axyspharm.com
Tel.: +1-619-6468207
Fax: +1-619-4526653

Please send articles to:
Peer Bork
Max-Dehlbrück-Center
for Molecular Medicine (MDC)
Robert-Rössle-Strasse 10
D-13122 Berlin, Germany
and:
EMBL
Meyerhofstrasse 1
D-69117 Heidelberg, Germany
E-mail: bork@embl-heidelberg.de
http://www.embl-heidelberg.de/~bork/

**Key words** Transforming growth factor signaling · Smad · Interferon response factor · Proten-protein interactions · SMIR domain

**Abbreviations** *IRF* Interferon regulatory factor · *TGF*-β Transforming growth factor

## Introduction

Interferon regulatory factors (IRFs) are transcriptional regulators important in controlling the expression of interferons and interferon-inducible genes in response to virus infection and other stimuli in animals (reviewed in [17]). The IRF family members share a conserved N-terminal DNA-binding domain which is characterized by a five-tryptophan signature and also have C-terminal domains, thought to mediate protein-protein interactions with other components of transcriptional machinery. Indeed, one of the IRF proteins, Pip, has been isolated by virtue of the specific interaction of its C-terminal domain with another transcriptional regulator PU.1 [7]. Other interactors of IRFs may include transcriptional activators ATF-2 and NF-κB and the high-mobility group protein HMG I (Y), which are the components of the enhanceosome, a large complex mediating transcription of the interferon β gene (*IFN*-β) [18]. Recently the crystal structure of IRF-1 N-terminal domain in a complex with its cognate DNA segment PRD I from the *IFN*-β promoter, was determined, establishing the winged-helix architecture for this domain [8]. IRF-1 joins the αββαββ structural class to which several other transcription regulators also belong (SCOP: Superfamily: Winged DNA-binding domain. http://scop.mrc-lmb.cam.ac.uk/scop/data/scop.1.001.004.003.html). Apart from this structural assignment, no sequence similarities between IRF proteins and any proteins outside of the family have been reported.

We partitioned sequences of the IRF family members into putative globular and nonglobular domains, using the SEG program with a set of parameters optimized for this purpose [20]. In most of the IRFs the segments homologous to the known IRF-1 DNA-binding domain corresponded to the N-terminal globular portion, typically separated from the C-terminal globular domain with a predicted elongated hinge (data not shown). Multiple alignments of the C-terminal domains of IRF proteins were constructed using the MACAW program [15], and the most conserved regions were converted into profiles [4]. The nonredundant protein sequence database at NCBI was searched, using WiseTools [4] to detect additional related sequences.

Comparison of various IRFs shows high similarity in the C-terminal domains of IRFs 3–7. PSI-BLAST searches with the C-terminal domains of the mum-1/IRF4/ICSAT subgroup also detected similarity to the C-terminal domains of Smad proteins, animal proteins involved in development, and specifically mutated in pancreatic carcinomas and other tumors in humans [3, 13]. In particular, search with the C-terminal domain of mum-1 (gi1 698625; amino acids 111–451) re-

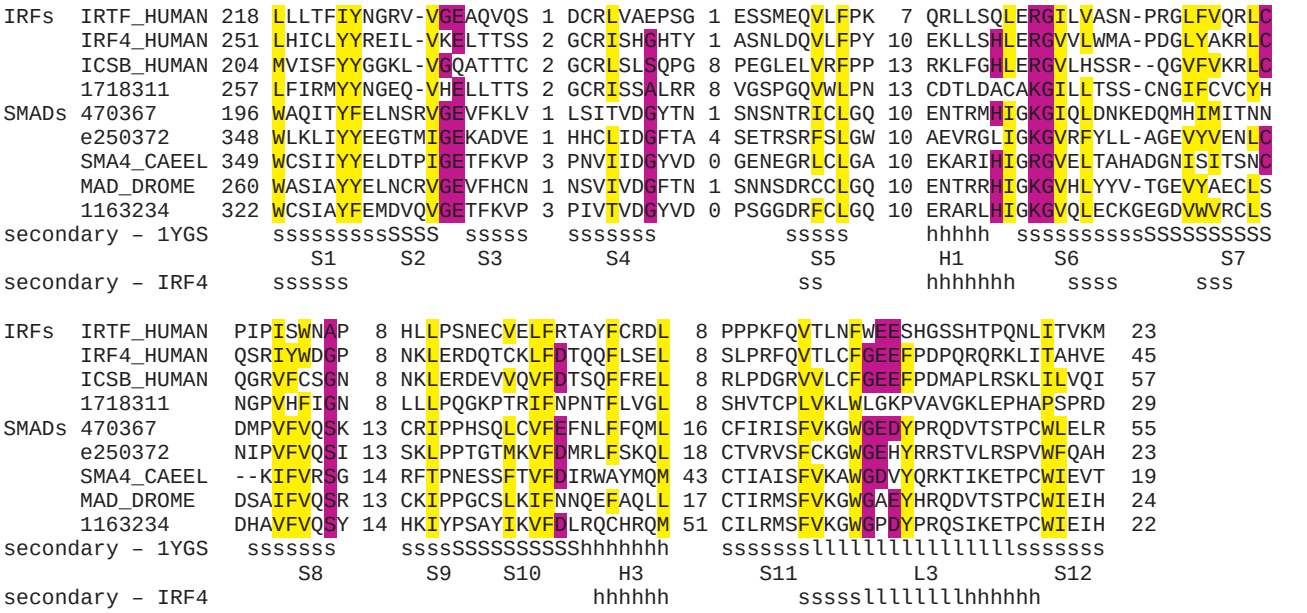


**Fig. 1** SMIR domain in IRF and Smad proteins. Unique identifier for each sequence in SWISSPROT or GenBank is shown. Distances between the most conserved blocks and to both termini of each protein are shown. *Yellow shading* conserved bulky hydrophobic residues (I, F, L, M, V, Y, and W). *Pink shading* Other highly conserved residues (including small side chain residues A, G, and S; positively charged R and K; and negatively charged E and D). Secondary structure elements in the known X-ray structure of Smad-4 are shown (*1YGS*) as well as secondary structure of IRF4 predicted by the PHD program [14] (only prediction with reliability index 6 or above were included). In the secondary structure line: *s* strand; *h* helix; *l* loop. 1718311, ORF K-9 product, Kaposi's sarcoma-associated herpesvirus; 1163234, deleted in pancreatic carcinoma (DPC-4) tumor suppressor (human), 470367, R13F6.9 (*C. elegans*), e250372, R05D11.1 (*C. elegans*)

trieved a mouse Mad-related protein (gi1658159) with the probability of matching by chance $10^{-4}$ at the first iteration. In a reciprocal search with the C-terminus of gi1658159 (amino acids 231–410), mum-1 was retrieved at the first iteration with the probability $5{\cdot}10^{-3}$. Additional members of IRF and Smad families, without false positives, continued to be retrieved in the latter search until convergence at the seventh iteration. Altogether, more than 45 Smad-related proteins and more than 35 IRF-related proteins were detected in this experiment. No new homologs could be detected in database scan using a combined profile built from the complete set of aligned IRF and Smad proteins. Thus, C-terminal domains of IRFs and of Smad proteins comprise a distinct family, thus far recognized only in animals (and in an animal herpesvirus associated with Kaposi's sarcoma).

Smad proteins participate in transducing signals from extracellular ligands, such as transforming growth factor (TGF-β) and bone morphogenetic proteins. Specific members of Smad family are phosphorylated by ligand-activated receptor kinases, resulting in formation of hetero-oligomers, which are translocated into the nucleus where they are thought to directly affect transcription of the target genes [2, 11, 21]. The N-terminal domains of Smad, thus far found only in this protein family, are able to interact with specific DNA sequences [21], although very few specific targets of Smad-dependent transcriptional activation have been characterized thus far [19]. The three-dimensional structures of a C-terminal domain of Smad-4 in a monomer and in a homotrimer have been resolved [16]. The core structure consists of an 11-strand β-sandwich, with flanking α-β subdomains that lock subunits within a trimer.

Multiple sequence alignment of IRF and Smad proteins was compared with the known structure of Smad-4. The regions of the highest conservation include most of the β-strands and loop 3. Interestingly, the sequence is less conserved in the regions corresponding to the strands 3 and 4, and the helices 1, 2, and 3 in the Smad-4 structure (Fig. 1, and data not shown). Since this set of elements is involved in the intersubunit contacts, one may speculate that the monomers of IRF-C and Smad adopt similar folds, but intermolecular interactions of IRF-C differ from those in Smad proteins. Secondary structure prediction for IRF proteins is largely compatible with these data, predicting strands 1, 5, 6, 7, and 11 as well as helices 1 and 3, in the appropriate positions. The C-terminal regions in IRFs 1 and 2 are distantly related to each other but appear to be unrelated to other IRFs or to Smad proteins. Secondary structure prediction for IRF1 suggests at least three beta-strands, indicating a probability of an all-beta fold in this area, but it remains to be seen whether this region adopts the same fold as the Smad C-terminal domain.

Of particular interest is the region that forms the L3 loop in Smad4. The homologous regions in closely related Smad-1 and Smad-2 directly interact with the TGF receptor and the bone morphogenetic protein receptor, respectively, and have been implicated in additional interactions at the downstream stages of signal transduction [12]. The L3 loop is partially conserved in IRF C-terminal domains (Fig. 1), indicating that it may be a major

determinant of IRF interaction with other proteins.

Transcriptional regulators of IRF family consist of two distinct modules, the N-terminal winged-helix DNA-binding domain and the C-terminal protein-binding domain, present in most IRFs but not defined in IRF-1 and IRF-2. We refer to this domain as SMIR (*Sm*ad and *IR*F). The SMIR domain appears to be a portable protein-protein interaction module that can be associated with at least two different classes of DNA-binding modules. Interestingly, all known interaction partners of IRF proteins have nuclear functions [18], whereas the best studied interacting partner of the L3 loop in Smad proteins is the cytoplasmic portion of TGF receptor [12], although a nuclear interactor has also been recently described [11]. It is tempting to speculate that all SMIR domains or their L3 loops are able to interact with both nuclear and cytoplasmic proteins, either as monomers, or as hetero-oligomers, with distinct subunits involved in compartment-specific interactions.

## References


1. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402
2. Attisano L, Wrana JL (1998) Mads and Smads in TGF-β signalling. Curr Opin Cell Biol 10:188–194
3. Baker JC, Harland RM (1997) From receptor to nucleus: the Smad pathway. Curr Opin Genet Dev 7:467–473
4. Birney E, Thompson JD, Gibson TJ (1996) PairWise and SearchWise: finding the optimal alignment in a simultaneous comparison of a protein profile against all DNA translation frames. Nucleic Acids Res 24:2730–2739
5. Chothia C, Hubbard T, Brenner S, Barns H, Murzin A (1997) Protein folds in all-beta and all-alpha classes. Annu Rev Biophys Biomol Struct 26:597–627
6. Davies C, Gerstner RB, Draper DE, Ramakrishnan V, White SW (1998) The crystal structure of ribosomal protein S4 reveals a two-domain molecule with an extensive RNA-binding surface: one domain shows structural homology to the ETS DNA-binding motif. EMBO J 17:4545–4558
7. Eisenbeis CF, Singh H, Storb U (1995) Pip, a novel IRF family member, is a lymphoid-specific, PU.1-dependent transcription activator. Genes Dev 9:1377–1387
8. Escalante CR, Yie J, Thanos D, Aggarwal AK (1998) Structure of IRF-1 with bound DNA reveals determinants of interferon regulation. Nature 391:103–106
9. Harrison CJ, Bohm AA, Nelson HC (1994) Crystal structure of the DNA binding domain of the heat shock transcription factor. Science 263:224–227
10. Hubbard TJP, Murzin AG, Brenner SE, Chothia C (1997) SCOP: a structural classification of proteins database. Nucleic Acids Res 25:236–239
11. Labbe E, Silvesti C, Hoodless PA, Wrana JL, Attisano L (1998) Smad2 and Smad3 positively and negatively regulate TGFβ-dependent transcription through the forkhead DNA-binding protein FAST2. Mol Cell 2:109–120
12. Lo RS, Chen Y-G, Shi Y, Pavletich NP, Massague J (1998) The L3 loop: a structural motif determining specific interactions between SMAD proteins and TGF-β receptors. EMBO J 17:998–1005
13. Reiss M (1997) Transforming growth factor-β and cancer: a love-hate relationship? Oncol Res 9:447–457
14. Rost B (1996) PHD: predicting one-dimensional protein structure by profile-based neural networks. Methods Enzymol 266:526–539
15. Schuler G, Altschul SF, Lipman DJ (1991) The workbench for multiple alignment construction and analysis. Protein Struct Funct Genet 9:180–190
16. Shi Y, Hata A, Lo RS, Massague J, Pavletich NP (1997) A structural basis for mutational inactivation of the tumor suppressor Smad4. Nature 388:87–93
17. Taniguchi T, Harada H, Lamphier M (1995) Regulation of the interferon system and cell growth by the IRF transcription factors. J Cancer Res Clin Oncol 121:516–520
18. Tsanos D, Maniatis T (1995) Virus induction of human IFNβ gene expression requires the assembly of the enhanceosome. Cell 83:1091–1100
19. Vindevoghel L, Kon A, Lechleider R J, Uitto J, Roberts AB, Mauviel A (1998) SMAD-dependent transcriptional activation of human type VII collagen gene (COL7A1) promoter by transforming growth factor-β. J Biol Chem 273:13053–13057
20. Wootton JC, Federhen S (1996) Analysis of compositionally biased regions in sequence databases. Methods Enzymol 266:554–573
21. Zawel L, Dai JL, Buckhaults P, Zhou S, Kinzler KW, Vogelstein B, Kern SE (1998) Human Smad3 and Smad4 are sequence-specific transcription activators. Molecular Cell 1:611–617


**Note added in proof** Recently, it has been shown that interferon gamma inhibits TGF-beta/Smad signalling (Ulloa *et al*., 1999 Nature, 397:710–713), suggesting that the components of the two pathways may interact, possibly by heterooligomerization of SMIR domains from Smad and IRF proteins, or competition for the L3 loop interactors between the two classes of activators.