Structural basis for genome wide recognition of 5-bp GC motifs by SMAD transcription factors

M.J. Macias

IRB Barcelona and ICREA, Barcelona SPAIN, **maria.macias@irbbarcelona.org**

The transforming growth factor (TGF-beta) regulates critical biological processes during embryo development, tissue homeostasis, regeneration, and immune regulation. The main TGF-beta signal transduction mechanism involves Smad transcription factors, and mutations in the components of this pathway are responsible for various inherited and somatic diseases. As transcription factors, Smad proteins interact directly with DNA. However, the motif/s specifically recognized by Smads have been a matter of debate for many years. In the past, an *in vitro* approach identified the 5'–GTCT–3' but its 4-base short length suggested that its presence in many Smad target genes could be random. Recent experiments *in live cells* have revealed that TGF-β target genes bind Smad2/3 and Smad4 complexes in promoters and enhancers lacking GTCT motifs.

Applying a combination of experimental techniques, we identified a consensus GC-motif as the binding site of Smad4, Smad3, Smad5 and Smad8 and we provided several X-ray crystal structures of the protein-DNA complexes determined at 2 and 2.5 Å resolution. All screenings were performed at the HTX facility of the EMBL Grenoble Outstation. Diffraction data were recorded at the ESRF on the beamlines ID23-2 and ID30A-3.

Using CRISPR/Cas9, we demonstrated that binding of Smads to these sites is functional. Remarkably, Smad3 and Smad4 bind to the **GGC(GC)|(CG)** sites reading up to five bases. The flexibility of the DNA binding hairpin characterized by NMR and SAXS, facilitates access to DNA duplexes with slightly distinct topologies and DNA sequences. Using bioinformatic analysis and ChIP-seq data, we also found that the new 5GC-motifs are highly represented as clusters in Smad-bound regions genome-wide. Binding to clusters suggest how three MH1 domains in a trimeric Smad complex may bind to promoters/enhancers at once with high specificity. This capacity endows Smad transcriptional complexes with a level of adaptability that has established TGF-β as one of the most versatile and highly conserved signaling pathways in metazoans. Our findings also highlight regions in promoters affected by tumor mutations, which have not been regarded as Smad targets so far.

The results of our work have been recently published in Nature Communications, Article number: 2070 (2017), doi:10.1038/s41467-017-02054-6.