Refolding, purification, and characterization of human-Smad8 produced in *E. coli* in form of inclusion bodies

Von der Naturwissenschaftlichen Fakultät der

Gottfried Wilhelm Leibniz Universität Hannover

zur Erlangung des Grades

Doktorin der Naturwissenschaften (Dr. rer. nat.)

genehmigte Dissertation

von

Carla Lizbeth Segovia Trinidad,

Maestria en Ciencias, Universidad Nacional Autónoma de México

2023

Referentin: Prof. Dr. rer. nat. Ursula Rinas

Korreferentin: PD Dr. rer. nat. habil. Antonina Lavrentieva

Tag der Promotion: 11.04.2023

Hope is a waking dream

Aristotle

Table of content

Abstract

Smad proteins are transcriptional regulators that participate in the intracellular signaling of the TGF-β superfamily [1] [2]. The members of this family are involved in cell proliferation, differentiation and death, among others [2]. For example, osteogenic BMPs (BMP-2, 6 and 9) are potent inducers of osteogenic differentiation [3].

Smads are classified into receptor activated Smads (R-Smads: Smad1, 2, 3, 5 and 8/9), Co-Smad (Smad4) and inhibitory Smads (I-Smads: Smad6 and 7) [4]. In general, they all are formed by the high conserved domains MH1 and MH2. Both domains are connected by a linker (L) region [5]. R-Smads transduce the signal from the factors of the TGF-β through the type I and type II receptors, which phosphorylate R-Smads. R-Smads complex with Smad4 and translocate itself to the nucleus. On the other hand, this pathway is negatively regulated by I-Smads [6].

Interestingly, Smad8 displays characteristics of both R-Smad and I-Smads [1, 7]. It is believed that it is part of a fine tuning Smad through a mechanism that is not totally understood [1]. In spite of this, Smad8 has been found to have application in the field of regenerative medicine. A truncated form of Smad8 consisting in the MH2 domain and the linker (Smad8L-MH2, hcSmad8) expressed by murine cells (C3H10T1/2) in combination with BMP-2 is able to induce the formation of tenocytes (tendon cells). C3H10T1/2 cells stimulated by BMP-2 normally self-differentiate into osteocytes (bone cells). Smad8 seems to inhibit the osteogenic pathway induced by BMP-2 while promoting tendon differentiation [8].

In view of the above, Smad8L-MH2 is a promising molecule that can be used for the treatment of tendon lesions, for example. In orthopedic medicine, the use of implants is a common practice. In this regard, the production of Smad8L-MH2 using a recombinant system that circumvent the disadvantages of modified cells is desirable. Additionally, this would offer the possibility of scalability.

Herein, hcSmad8 was modified to display a Transactivation Sequence (TAT) of the HIV virus which allows cell internalization. TAT is a widely-used cell penetrating peptide that permits the translocation of cargos into cells [9]. The resulting protein, called TAT-hcSmad8 is then formed by the MH2 domain, the linker and a TAT sequence.

TAT-hcSmad8 was produced in *E. coli* cells in form of inclusion bodies (IBs). After their production, they were washed and solubilized using chaotropic agents. The solubilized TAT-hcSmad8 was then refolded by dilution into a CHES-based buffer. In this state, the protein was purified by affinity chromatography

(Heparin) with which 96 % purity was achieved. After purification, the biochemical properties of TAT-hcSmad8 were investigated. Thermal shift assays (TSA) indicated that purified TAT-hcSmad8 has a hydrophobic core, which is characteristic of folded proteins. Circular dichroism analysis indicated that it is formed by a mixture of α-helixes and β-sheets which is in accordance with the theoretical model made by homology. The addition of a histidine tag at the N-termini showed not to modify the secondary structure of TAT-hcSmad8. The importance of the purification is highlighted by the fact that the use of 4 M urea during Heparin purification, leads to an unfolded TAT-hcSmad8 that lacks the hydrophobic core as observed by TSA.

In general, the biological activity of TAT-hcSmad8 was studied in murine cells (C3H10T1/2 cells) which were incubated in presence of BMP-2 (0 and 500 ng/mL) and different concentrations of TAT-hcSmad8 (0, 5, 10, 25 and 50 µg/mL). C3H10T1/2 cells stably transfected with the construct called BRE-Luc containing a luciferase gene under the control of a BMP Responsive Element (BRE) [10] indicated that TAT-hcSmad8 inhibited the osteogenic pathway induced by BMP-2 (500 ng/mL). This inhibition was observed for concentrations equal or higher than 10 µg/mL of TAT-hcSmad8. The gene expression of wild type C3H10T1/2 cells was also studied by RT-PCR. With this method a decrease of 50 % in the *Bglap* expression in cells incubated with 25 µg/mL of TAT-hcSmad8 was observed. For the same concentration, a duplication in the expression level of *Scx* and *Sox9* was observed. The global behavior of this group of genes suggest an early differentiation state of cells towards tenocytes.

Justification

Smad8 is an intriguing molecule that found biotechnological application [8, 11] in spite of the still controversial and, to some degree, unknown-role in TGF-β signaling pathway. As briefly mentioned before and later explained more in detail, a truncated variant of Smad8 consisting of the MH2 domain and the linker (Smad8L-MH2: hcSmad8) co-expressed with BMP-2 in C3H10T1/2 cells induced the differentiation of the cells into tenocytes (tendon cells). This effect was observed *in vitro* and *in vivo. In vivo* studies in Achilles tendon partial defect indicated that tendon regeneration was induced by this particular combination of molecules [8]. This made hcSmad8 and BMP-2 promising proteins that can be applied in the clinical medicine.

Tendon and ligament repair is a major clinical challenge due to their slow, painful and ineffective process. The resulting tissue is commonly mechanically, structurally and functionally inferior to the undamaged tissue [12]. Tendons form part of a complex and specialized tissue interface (enthesis) that integrates tendon/ligament in bone. Points of attachments are smooth and perfectly graded transitions that facilitate motion [13]. Since enthesis are vital points that are subjected to constant mechanical stress, the complete restoration of its functionality after an injury is desirable. Until the present, therapeutic options used to treat tendon injuries have not provided a successful long-term solution and usually lead to variable results [14]. There is an evident need for new functional tendon-repairing strategies. In this regard, genetically engineered C3H10T1/2 cells expressing BMP-2 and hcSmad8 is a novel technique. However, the restrictions for its medical application of genetically modified cells in humans and the unpracticality of its handling are major withdraws. The production of these factors in a recombinant system can, though, circumvent this problem. In this context, hcSmad8 would be genetically modified to include a Transactivation Sequence (TAT) of the HIV virus that enables the internalization of cargos into cells [9]. TAT-hcSmad8 could be then produced and loaded in tendon implants.

The work here presented is specifically oriented to the production and characterization of TAT-hcSmad8 as part of a multidisciplinary research group called *Graded Implants* whose main objective is the development of a graded implant for the tendon-bone junction of the rotator cuff [15-18]. The implant is meant to be made of polycaprolacton fibers loaded with a gradient of BMP-2, TGF-β3 and TAT-hcSmad, which would promote the formation of bone, cartilage and tendon, respectively.

Abbreviations

% v / v	Percent volume / volume
% w / v	Percent weight / volume
2D / 3D	Two-dimensional / three dimensional
ALK	Tyrosine kinase receptor
ALK - x	Tyrosine kinase receptor - x
ALP	Alkaline phosphatase
BCIP	5-bromo-4-chloro-3'-indolyphosphate p-toluidine salt
BMP	Bone morphogenetic protein
BMP-2	Bone morphogenetic protein-2
BRE	BMP-specific response elements
BSA	Bovine serum albumin
CD	Circular dichroism
CHES	N-Cyclohexyl-2-aminoethanesulfonic acid
CLSM	Confocal laser scanning microscopy
cSmad8	Constitutively active form of Smad8 consisting of Smad8 linker + MH2 domain
CTB	CellTiter-Blue
CV	Column Volume
Cy3 / Cy5	Cyanine dye 3 / Cyanine dye 5
Da	Dalton
DAPI	4',6-diamidino-2-phenylindole
DMEM	Dulbecco's Modified Eagle Medium
DTT	Dithiothreitol
E. coli	*Escherichia coli*
EDTA	Ethylenediaminetetraacetic acid
FCS	Fetal calf serum
g	Gram
GDFs	Growth differentiation factors
GFP	Green fluorescent protein
GST	Glutathione S-transferase
GuHCl	Guanidine hydrochloride
h	Hour(s)
hcSmad8	His6-Tag + constitutively active form of Smad8 consisting of Smad8 linker + MH2 domain
HEK cells	Human embryonic kidney cells
His6-Tag	Tag consisting of six histidines
HIV	Human immunodeficiency virus
IBs	Inclusion bodies
IMAC	Immobilized metal affinity chromatography
IPTG	Isopropyl β- d-1-thiogalactopyranoside
kDa	Kilodalton
L	Liter
LB	Luria-Bertani broth
LPS	Lipopolysaccharide

Luc	Luciferase
m	meter
M	Molar (mol / liter)
MAPK	Mitogen-activated protein kinase
MEM	Minimum Essential Media
MES	2-(N-morpholino)ethanesulfonic acid
MH1 / MH2	Mad homology 1 / Mad homology 2 (domains of Smads)
min	Minutes
mM	Milimolar
MOPS	3-(N-morpholino)propanesulfonic acid
mS/cm	Millisiemens per centimeter
MSC	Mesenchymal stem cell
NBT	Nitro-blue tetrazolium chloride
NCBI	National Center for Biotechnology Information
nm	nanometer
OD_{600}	Optical density measured with a wavelength of 600 nm
PAH	Pulmonary arterial hypertension
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction
PDB	Protein Data Bank
pI / Mw	Isoelectric point / molecular weight
PMSF	Phenylmethylsulfonyl fluoride
psi	Pounds per square inch
PVDF	Polyvinylidene difluoride
RFU	Relative fluorescence units
RLU	Relative luciferase units
rpm	revolutions per minute
RT-PCR	Real-time PCR
Scx	Scleraxis
SDS	Sodium dodecyl sulfate
SDS-PAGE	Sodium dodecyl-sulfate polyacrylamide gel electrophoresis
Smad8L-MH2	Linker +MH2 domain of Smad8
TAT	Transactivator Sequence
TAT-cSmad8	Transactivator Sequence of human immunodeficiency virus + Smad8 linker + MH2 domain
TAT-hcSmad8	His6 -Tag+ Transactivator Sequence of human immunodeficiency virus + Smad8 linker + MH2 domain
TEV	Tobacco etch virus
TGF-β	Transforming growth factor-beta
TGF-β1	Transforming growth factor-beta 1
Tm	Melting point
TMB	3,3',5,5'-Tetramethylbenzidine
TSA	Thermal shift assay

1. Theoretical background

1.1 Generalities of the transforming growth factor β (TGF-β) superfamily pathway

The TGF-β superfamily are secreted growth factors which regulate many cellular processes, including cell growth, adhesion, migration, cell-fate determination, cell differentiation, and apoptosis [19, 20]. The TGF-β superfamily is large and it continuously expands. In mammals, the family counts so far with 33 members which includes bone morphogenetic proteins (BMP), growth differentiation factors (GDF), activins, nodal, anti-Müllerian hormone (AMH; also known as Müllerian inhibiting substance (MIS)), inhibins, left-right determination factor (Lefty) and glial-derived neurotropic factors (GDNFs) [21, 22]. Since members of the TGF-β superfamily display multiple functions thorough complex signaling pathways where more factors are usually involved, it is not possible to make accurate generalizations. On the other hand, some members have rather functions particularly specific. However, in general it is possible to recognize common features. The BMP/GDF subfamily, for example, regulates osteo-, chrondro-, and tenogenesis [23]. Activins and inhibins are important in wound repair, and testis and skin development. TGF-βs control proliferation and differentiation in many different cells [23].
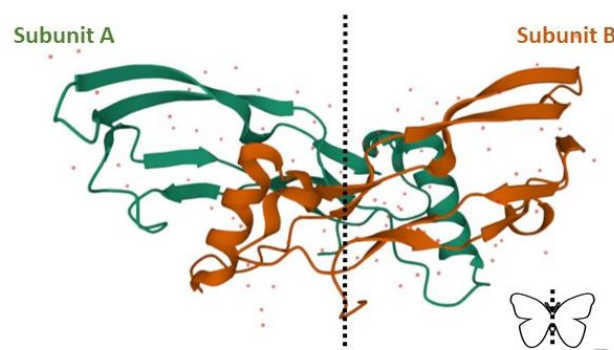

Figure 1. Representation of a butterfly-like shape of TGF-β ligands.
Crystal structure of Bone Morphogenetic Protein-3 (BMP-3) (2QCQ). Figure made according to Yadin et al. (2016) [24].

All TGF-β superfamily ligands are synthesized as precursors [24, 25] that are later matured to form the functional ligand. A key feature is the presence of a cysteine knot that comprises six cysteines. Four cysteines are arranged in the motif C_2-X-G-X-C_3 and C_6-X-C_7-X that form a ring fixed by two disulfide bonds. A third disulfide bond formed between the first and the fifth cysteine crosses and tights the ring [24]. The active protein is then formed by two assembled monomers that display a butterfly-like shape [24] (Figure 1). However, GDF-3, GDF-9, BMP-15, lefty1 and lefty2 likely form non-covalent dimers as they lack the cysteine required for the disulfide bone formation [22].

Proteins of the TGF-β superfamily signal through transmembrane receptors which also act as serine/threonine kinases. They can activate intracellular components that are part of different pathways. The signal transduction of TGF-β superfamily mediated by Smads initiates with the ligand-binding to the extracellular domain of type II receptors (Table 5, Supplemental in Appendix). Ligand engagement induces a complex formation that includes subunits of type II and type I receptors. Withing the complex, the subunits of type II receptors phosphorylate the type I receptor which in turn recruits and phosphorylates receptor-regulated Smads (R-Smads, Smad 1,2,3,5 and 8) [4]. Activins, nodal and TGF-βs signal predominantly through Smad2 and Smad3, whereas BMPs and GDFs signal via Smad1, Smad5 and Smad8 [4]. Phosphorylated R-Smads associate with the common partner Smad4 (Co-Smad) and translocate to the nucleus to control gene transcription [19]. The signaling cascade is negatively regulated by inhibitory Smads (I-Smads, Smad6 and Smad7) [4].

There is a multitude of intracellular pathways in addition to the canonical Smad pathway, through which TGFβ superfamily transduces its signals. These non-Smad pathways includes the p38 and Jun N-terminal kinase (JNK) mitogen-activated protein kinase (MAPK), the phosphoinositide 3-kinase-Akt-mTOR pathway, the small GTPases Rho, Rac, Cdc42, and the Ras-Erk-MAPK pathway [26-29].

Disfunction of the TGF-β superfamily signaling can affect the skeletal, muscular, and/or cardiovascular system. Moreover, its malfunctioning has an important role in various cancers such as colorectal, prostate, breast, ovarian and melanoma [23].

1.2 A closer look at the Smads

1.2.1 General features of Smad proteins

Smad proteins are transcriptional regulators implicated in the intracellular signaling of the proteins of the TGF-β superfamily [4]. They were first described in *Drosophila* and in *Caenorhabditis elegans*. They were named Mad (mothers against decapentaplegic (dpp)) and Sma, respectively. Related proteins in *Xenopus*, humans, mice, rats, and other systems where later identified. Smads were also named as Mad–related (MADR) family or Mad homologe (MADH). However, all family members in all organisms are nowadays known as Smads, which is a contraction of the invertebrate gene names "Smad" and "Mad" [6, 30-32] (Table 1).

Table 1. Smad family members

Name	Previous names	Gene name	Accession number to Unigene
TGF-β/activing-regulated R-Smads			
Smad2	MADR2, JV18-I	MADH2	NM005901
Smad3	JV15-2	MADH3	NM005902
BMP-regulated R-Smads			
Smad1	MADR1, JV4-1 ,Dwarfin A, bsp-1	MADH1	NM005900
Smad5	Dwarfin C, JV5-1	MADH5	NM005903
Smad8	Smad9, MADH6	MADH9	NM005905
Common Smad			
Smad4	DPC4	MADH4	NM005359
Inhibitory Smads			
Smad6	JV15-1	MADH6	AF035528
Smad7		MADH7	NM005904

Table based on Attisano et al. (2001) [6]. Modifications were performed.

The Smad family in humans consist of eight members that are classified into three categories, based on their structures and biochemical functions: R-Smads, Co-Smad and I-Smad. As briefly described in Section 1.1, R-Smads includes Smad1, Smad2, Smad3, Smad5 and Smad8/Smad9[1]. In mammals, only Smad4 has been identified as Co-Smad [33]. However, two Smad4-like molecules (XSmad4α and XSmad4β) have been identified in *Xenopus*, which can cooperate with Smad1 and Smad2 in the intracellular signaling pathway of TGF-β superfamily [33]. Smad6 and Smad7 are I-Smads. Except for sponges, which lack I-Smads and have multiple duplications of other Smads, invertebrates have four Smads (two R-Smads, Smad4 and one I-Smads) [5]. Smad4- and R-Smads-sequence is highly conserved not only within mammals but also across all vertebrates indicating the essential role of Smads in the organisms [4, 5, 34].

1.2.2 R-Smads and Co-Smad

R-Smads and Co-Smad consist of two 'Mad homology' (MH) domains, which are highly conserved globular domains across species [4, 5, 34]. Both MH domains are connected by a linker (Figure 2). The linker has the highest concentration of amino acid differences among vertebrates [5]. Crystallographic structures of full-length Smads have not been obtained so far. However, crystal structures of truncated domains of different Smads are available (Table 4). Structural information for the MH1 domain [5, 35-40] and MH2 domain [5, 41-49] and for the linker regions have revealed important features about Smads.

[1] *Smad8 and Smad9 differ in two amino acids: comparison of mouse Smad8 (GenBank AF175408) and Smad9 (GenBank MN_019483) genes indicates that six nucleotides (5′-cgagtc-3′) are inserted at nucleotide +961 of Smad8 leading to two extra amino acids present in Smad9.*
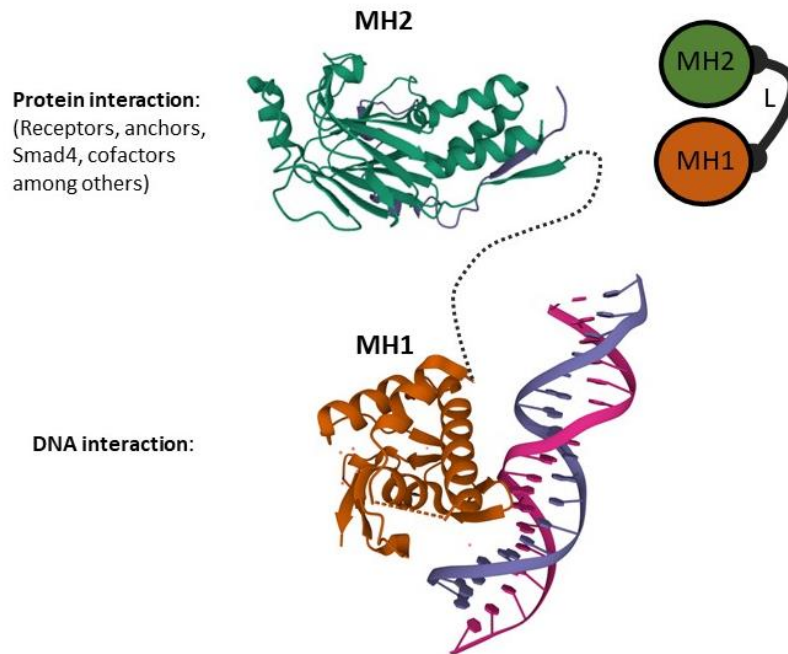
Figure 2. Representation of common domains with typical structure of a Smad.
Since there is no complete crystal structures of any Smad so far, crystal structures 1DEV (green, MH2) and 6H3R (orange, MH1) were taken. The linker has no secondary structure and his high mobility impedes that its structure is determined by common crystallographic methods (dotted line). Figure made according to Macias et al. (2015) [5].

MH1 is the N-terminal domain whose main function is to bind DNA [19]. In R-Smads and Smad4, the MH1 recognizes a double stranded DNA motif 5´-CAGAC-3´ known as Smad binding element (SBE). Rich-G/C motives are also recognized by MH2 domain *in vivo* and *in vitro* [5]. In the obtained crystal structures of the MH1 domains, is observed common features such as four-helices and three sets of antiparallel β-hairpins (Figure 2). Three conserved cysteines and a histidine residue coordinate a Zn^{2+} ion which stabilizes β1- β4 and β5- β6 hairpins. The hairpin formed by β2 and β3 strands protrudes and mediates the interaction with the major groove of DNA. This DNA-binding β-hairpin is highly conserved in Smads. In fact, mutations in and around this loop have been observed in different types of tumors [37]. Interestingly, a short amino-acid insert immediately prior to the DNA-binding β-hairpin is responsible for the functional differences between Smad2 and Smad3 [50]. The β-hairpin in Smad3 allows direct binding to the TGF-β responsive element (CAGA box). Moreover, a 30-aminoacid insert present in Smad2, prevents this to happen due to steric hindrance. In other words, this small difference is responsible for the functional differences between Smad2 and Smad3 [50].

In the absence of activation of the MH2 domain by phosphorylation, the MH1 has been shown to inhibit the MH2 domain by physically binding it. The phosphorylation abolishes the MH1-MH2 interaction and allows R-Smads to form heterooligomers with Smad4 [49]. Genetically modified forms of R-Smads lacking MH1 domain can exert, however, biological activity [8] (Section 1.2.7).

The MH2 domain, for its part, is the C-terminal domain that mediates protein-protein interactions with TGF-β receptors, other R-Smads, Smad4, particular cytoplasmic anchor proteins, lineage-specific DNA-binding cofactors, and chromatin modifiers [5]. In general, the MH2 domain is defined by two sets of antiparallel β-strands (six and five strands respectively) that forms a β-sandwich flanked by a triple-helical bundle and large loops [5]. The L3- and L46-loop mediates interaction with type I receptors [6]. The C-terminus of MH2 domain contains a highly conserved N-S**S**X**S**-C motive where phosphorylation of the last two serines (in position -1 and -3, in bold) is carried out. Interestingly, serine -4 seems not to be a site of phosphorylation although it is important for the activation of the other serins [49, 51]. ALK5 (TGF-β type I receptor (TβRI)) and ALK4 (activin type I receptor (ActRIB)) activate both Smad2 and Smad3. ALK1 and the BMP type I receptors ALK2, ALK3 and ALK6 target Smad1, Smad5 and Smad8 (Table 5 in Appendix). For the case of Smad2 and Smad3, the Smad anchor for receptor activation (SARA) and the related protein endofin, facilitates their anchoring to the type I receptor. The phosphorylation triggers dissociation of the R-Smads from the type I receptor for subsequent complexing with Smad4 [49]. Regarding type II receptors, there are five type II receptors available for the TGF-β/BMPs. The TGF-β receptor TβRII exclusively interacts with TGF-β, whereas the AMH type II receptor AMHRII binds anti-Mullerian hormone. The activin type II receptors ActRIIA and ActRIIA interacts with the majority of the activin, inhibin, BMP and GDF subfamilies. The BMP type receptor BMPRII is shared among all BMP and GDF ligands [24] (Table 5, Supplemental in Appendix).

The linker region is the less conserved region of Smads [5]. However, it contains regulation sites which are essential for the correct functioning of Smads. The linker includes, for example, potential phosphorylation sites for mitogen-activated protein kinases (MAPKs) which negatively regulate Smads. A proline-tyrosine (PY) motif is also found in the linker. The PY motif mediates interaction with WW domains present in various proteins such as Smad-interacting proteins Smurf1 and Smurf2 [6]. Smurf are ubiquitin ligases that catalyze ubiquitin-mediated degradation.

1.2.3 I-Smads

I-Smads are members of the Smad family involved in the regulation of TGF-β superfamily signaling [52]. Smad6 and Smad7 have been identified as I-Smads in vertebrates [34]. In *Drosophila* for example, daughters against DPP (Dad) acts as an I-Smad [34].

Like R-Smads and Co-Smads, I-Smads possess an MH1 and MH2 domain. The MH2 domain of I-Smads is structurally related to the MH2 domain of R-Smads and Co-Smads. However, I-Smads and Co-Smad lacks the SSXS motif which prevents their activation by the type I receptor [4]. On the other hand, the MH1 domain and the linker are highly divergent from the elements of the other Smads. Moreover, amino acid sequence is only partially conserved (36.7 %) between Smad6 and Smad7 [34].

Smad6 preferentially inhibits Smad signaling by inhibiting BMP type I receptors (ALK-3 and ALK-6) whereas Smad7 inhibits both TGF-β- and BMP-induced signaling pathway [34, 52-54].

The negative effect of I-Smads is exerted through multiple mechanisms that include, for example, the prevention of the interactions between R-Smads and type I receptors [53]. For this, I-Smads associate directly with activated type I receptors and impedes activation of R-Smads [4]. Smad6 has been shown to interact with Smad1 and to interfere with the Smad1-Smad4 complex formation [55].

I-Smads also exert their inhibitions effects by promoting degradation of R-Smads and type I receptors [34]. In this process, I-Smads interacts with Smurfs, which induce in turn ubiquitylation and further proteasomal degradation [34]. I-Smads have been shown to cooperate with other proteins that also interferes with the TGF-β signaling. BAMBI, for example, is structurally similar to type I receptors and naturally functions as a negative regulator of the TGF-β pathway [34].

In the nucleus, Smad7 interferes with the formation of R-Smads-DNA complexes [52]. Smad7 interacts with the Smad-binding DNA element through its MH2 domain blocking in this way R-Smads interaction with DNA [52].

I-Smads are regulated post-translationally by ubiquitylation, which induces their degradation. Acetylation confers to Smad7 resistance to ubiquitylation whereas Smad6 methylation promotes its dissociation from type I receptors [34]. Ligand stimulation generates an increase in I-Smad transcription [1, 7]. In response to TGF-β, Smad3/Smad4 complex associates with the promoter of Smad7 which activates Smad7 transcription [34]. BMP for its part, induces Smad7 transcription through Smad1 [34]. Similarly, Smad6 transcription is increased upon stimulation with BMP, which act through Smad1 and Smad5. In other words, I-Smads form an autoinhibitory feedback loop to inhibit TGF-β/Smad signaling [34].

Deregulation of I-Smads, specially Smad7, were found in human diseases such as inflammatory bowel disease and scleroderma [53].

1.2.4 Smads multimerization and the role of SSXS-domain phosphorylation

It is widely accepted that Smad4 serves as common mediator for R-Smads [4, 6, 19, 22]. Activated R-Smads (by phosphorylation) form heterotrimers with Smad4 [49, 56] and translocate to the nucleus to regulate gene expression [4, 6, 19, 22]. Studies of the crystal structures of the MH2 domain of Smads (Smad1, Smad2 and Smad3) suggest that phosphorylation triggers the complex formation between R-Smads and Smad4 [49, 51, 57]. Interestingly, it was determined that the phosphorylation of Ser -1 and -3 in the S**SXS** motif (Ser-1 and -3 are in bold) is required for Smad2 and Smad4 association. Ser -4 (underlined) was not phosphorylated by the receptor [51].

The role of phosphorylation and the stoichiometry of heteromeric complex are still controversial. Some models suggest that phosphorylation uncouples MH1 and MH2 which are thought to interact with each other in absence of phosphorylation serving the MH1 domain as a negative regulator for MH2 [8, 55, 58]. Another model suggests that phosphorylation contributes to subunit assembly by bridging the MH2 subunits in a complex [49, 57]. Qin et al. (2001) [49] reported the crystal structure of the MH2 domain of Smad1. Based on the structure and subsequent observations they propose a general model for R-Smads. In this model, R-Smads are found as monomer in its basal state. Phosphorylation would energetically favor subunit homo-trimerization. However, R-Smads homotrimer would be less stable than the R-Smad/Smad4 heterotrimer. The formation of the complex formed by R-Smad and Smad4 would be formed as it is energetically favored [49]. Biochemical studies indicate that the R-Smad/Smad4 complex consist of two phosphorylated R-Smads and one Smad4 subunit [49, 57].

Some biochemical studies indicate that unphosphorylated-MH2 domains of Smad1 and Smad2 are mainly found as monomers and that trimerization is enhanced by phosphorylation [49, 57]. Crystal structures of trimers (Figure 3) indicate that C-terminal residues of one MH2-domain interact with the L3 loop of the next MH2 subunit [49]. These interacting elements are highly conserved and trimerization seems to have an important role. In fact, mutations that affect subunit interaction are commonly observed in developmental and tumorigenic processes [57].
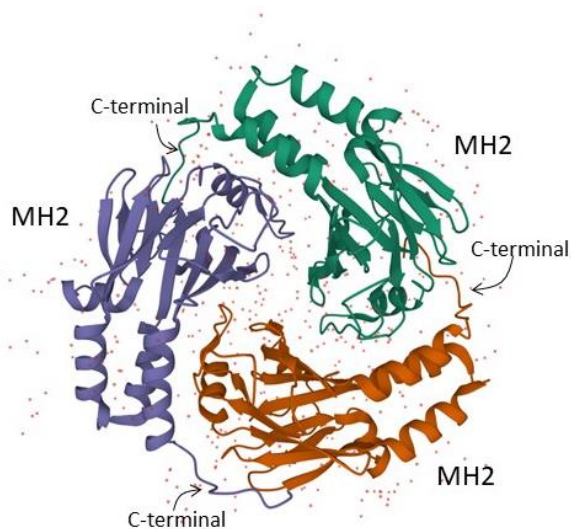
Figure 3. Crystal structure of MH2 trimer (PDB 1KHU).

It is worth mentioning, that the studies mentioned previously were done mainly with MH2 domains of R-Smads. Besides, the crystal structures of those MH2 domains are purified protein and the observations describe *in vitro* behavior. The stoichiometry of R-Smads *in vivo* is complex and ambiguous, especially because R-Smads interact with a variety of factor and regulators. This in turn depends on the cell type and the overall state of the cell.

Kawabata et al (1998) [59] reports that Smad2, Smad3, Smad4, Smad6 and Smad7 exist as monomers *in vivo* and that R-Smads undergo homo and hetero-oligomerization upon activation by phosphorylation. In contrasts to Kawabata et al. (2000) [59], Jayarama et al. (2000) [60] provided evidence that Smad2 is mainly found as monomer, whereas Smad3 exists in multiple states. Smad4 was found as a clear homo-oligomer [60]. The differences found may be due to the different cells and experimental protocols used in both works.

1.2.5 General features of Smad8/Smad9

Although in the literature Smad8 and Smad9 are sometimes used as synonyms, they are slightly different. Mouse Smad9 is originated by an insertion of six nucleotides at +961 position in the Smad8 gene. This means, that Smad9 has two extra amino-acids when compared to Smad8 [7]. This is relevant because some mouse cells express Smad8, Smad9 or both when stimulated with BMP-4 [7]. For example Smad9 mRNA was detected in mouse H9c2, fibroblast, 3T3-L1, HepG2 and B16, but not in C2C12 cells when treated with BMP-4.

The Smad8 gene was identified for the first time by Watanabe at al. (1997) [61]. The new gene was originally designated MADH6 (13q12-q14) and was derived from human placenta cDNA library. Even

then, they were able to identify two splicing variants of MADH6: a short transcript (MADH6b, 430 amino-acids) and a long transcript (MADH6a, 467 amino-acids). MADH6a displays an insertion of 111 nucleotides which code 37 amino-acids that forms a proline-rich domain. Why the human genome affords two variants is not fully understood.

Weak expression of MADH6 was observed in heart, brain, placenta, lung, skeletal muscle, prostate, testis ovary and small intestine [61]. In human fetal tissues there was a rather stronger expression in brain, lung and kidney [61].

A rat MADH6-homologue was later described by Chen at al. (1997) [62], who formally used the name "Smad8" to describe a protein with a predicted molecular weight of 49 kDa (434 amino-acids) that shared significant homology with R-Smads. The novel Smad8 shared structural similarities with other R-Smads such as the common MH1 and MH2 with surprising highly homology. Both domains were bound by a variable linker domain. The MH2 domain of Smad8 also possessed the SSXS motif present in R-Smads. Interestingly, Nishita et al. (1999) [63] described a human splice variant of Smad8 (Smad8B) lacking 47 amino acids that would code the portion that includes the SSXS phosphorylation site. Smad8B was able to associate with Smad4 but the complex did not translocate into the nucleus. Smad8B exerted a dominant inhibitory effect on BMP-responsible genes. The observed tissue specific mRNA expression of Smad8B might indicate the existence of delicate and tunable inhibition mechanisms in which this molecule participates [63]. The exact mechanism is still not understood.

Chen at al. (1997) [62]determined, that Smad8-phosphorylation by ALK-2 triggered the Smad8-Smad4 association. These results were later confirmed and expanded by Kawai et al. (2000) [64] which determined that Smad8 was activated not only by ALK-2, but also by ALK-3 and ALK-6. Intriguing, Smad8 was reported to inhibit the increase if alkaline phosphatase activity induced by BMP-2 in C3H10T1/2, which was usually observed for I-Smads [64].

At this point, Smad8 pointed out to be an interesting molecule with dual features shared by R-Smads and I-Smads. This observation was deeply studied by Tsukamoto et al. (2014) [1] and Katakawa et al. (2016) [7]. In this studies, Smad8 is also named Smad9, however, they differ in two amino acids.

Whereas I-Smads inhibit R-Smads phosphorylation (Section 1.2.3), Smad9 forms complexes with BMP-regulated R-Smads that successfully binds to BMP-responsive elements (BRE) but that dominantly suppresses the transcriptional activity [1]. Different experiments with chimers formed by the different motives (MH1, MH2 and linker) of Smad1, Smad5 and Smad9, showed that the linker region of Smad9 is responsible of the suppression of the transcriptional activity [1]. Similarly to Smad6 mRNA, Smad9

mRNA increased under BMP-4 stimulation [1, 7]. This indicates that Smad8/9 is transcriptionally regulated which is not observed for other R-Smads.

Although Smad8 biological activity shares features with I-Smads and R-Smads, Smad8 structural-features belong completely to R-Smads (Section 1.2.2). Actually, the amino acid sequence similarity in the MH1 and MH2 domains of Smad8, Smad1 and Smad5 is greater than 91 % [1]. On the other hand, their linker region share less than 40 % similarity [1]. Especially high is the similarity between mouse MH2 of Smad8 and the MH2 of Smad1 or Smad5, whose amino acid similarity is 92 % and 90 %, respectively. Mouse MH1 of Smad8 shares 83 % and 80 % similarity with MH1 domains of Smad1 and Smad5, respectively [64].

In contrast to the Smad8-MH2 domain, whose structure has not been determined, Smad8-MH1 domain has been characterized using X-ray crystallography [40]. Smad8 MH1 holds a Zn^{2+} coordinated by three cysteines and one histidine which is a key feature of R-Smads [5, 40]. In general, Smad8-MH1 is formed by four helices and six strands which forms a four-helical bundle and three anti-parallel pairs, respectively [40]. When MH1 domain of Smad5 and Smad8 are superimposed, the domains showed to be nearly identical [40].

Although Smad8 specific activity in cells is poorly understood, constant new pieces of information have been published. Lee et al. (2012) [65] reported a novel Smad8-binding protein called CREBZF. This protein is a transcription factor that holds basic region-leucine zipper that interacts not only with Smad8, but also with Smad1 and Smad5 in PC-3 (prostate cancer cell line) [65]. Surprisingly, Smads-CREBZF interaction was able to block the inhibition of BMP-6 [65].

Furthermore, Min et al. (2021) [66] reported that the E3 ligase ASB2 specifically targeted Smad8 (and not other Smads) by ubiquitination for proteasomal degradation. This regulation plays a pivotal role in cardiogenesis.

All this information indicates that Smad8 is a molecule that might be part of a fine-tuning process of the BMP signaling pathway.

1.2.6 Smad8 and its role in human diseases

The location of Smad8-gene is in a region were cancer-associated mutations frequently occur [61]. In fact, loss of Smad8 expression is associated with both breast and colon cancer [67]. On the other hand aberrations in the BMP signaling pathway, such as the loss of BMP2, Smad8, and Smad4 expression was associated with the progression of prostate cancer [68].

A missense mutation in Smad9 (c.65T>C, p.Leu22Pro) was found to cause high bone mass (HBM) [69]. HBM is a rare syndrome characterized by the high bone mineral density (BMD) among others [70]. The Smad9 missense mutation generates the disruption of the MH1 DNA-binding domain. Since Smad9 acts as an inhibitory Smad, the Smad9 disruption causes the loss of inhibition which generates high BMD [70]. This place Smad-dependent signaling pathway as a potential target for the treatment of osteoporosis [70].

A nonsense mutation in Smad8 (c.606 C>A, p.C202X) was identified to be the cause of pulmonary arterial hypertension (PAH) [71]. PAH is a progressive disorder that generates an elevation of pulmonary artery pressure due to the abnormal proliferation of endothelial and smooth muscle cells in pulmonary arteries [72]. The nonsense mutation in Smad8 results in a truncated form of Smad8 C202X protein that cannot be phosphorylated by receptors and that was unable to interact with Smad4 [71].

1.2.7 Biotechnological applications of Smad8

Although the exact role of Smad8 in TGFβ-family pathways is not completely understood, Smad8 could find application in the field of regenerative medicine. Hoffmann et al. (2006) [8] reported that murine C3H10T1/2 mesenchymal stem cells (MSCs) co-expressing a truncated variant of Smad8 consisting of MH2 and the linker [cSmad8: MH2+linker] and BMP-2, were able to self-differentiate into tendon-like cells. These genetically engineered cells showed morphological and gene expression profile of tenocytes (Figure 4). Moreover, they induced tendon regeneration *in vivo [8]*.
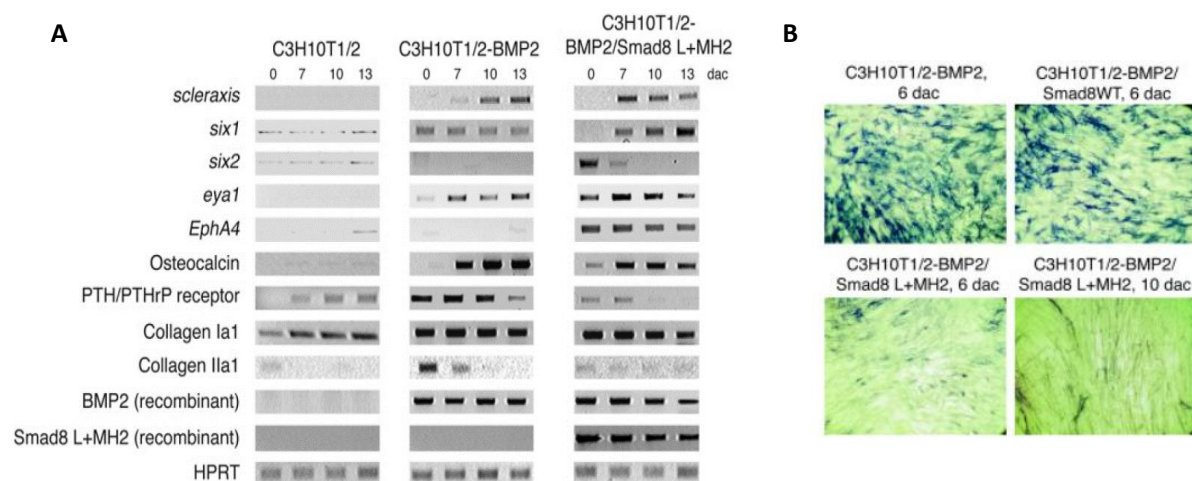

Figure 4. Analysis of the expression levels of genes involved in bone and tendon formation.
In panel A it can be observed an analysis by electrophoresis of reverse-transcription PCR products. The samples that were analyzed (top) includes C3H10T1/2 cells, C3H10T1/2 cells expressing BMP-2 and C3H10T1/2 cells expressing BMP-2 and cSmad8 (in [8], cSmad8 is called Smad8L + MH2). Specific primers were used to amplify gene markers of tendon, bone and cartilage (left). In panel B, it can be

observed C3H10T1/2 cells expressing BMP-2 and C3H10T1/2 cells expressing BMP-2 and cSmad8 stained with NBT/BCIP. With this reagent, ALP-positive cells are stained in blue/purple. It can be seen that cells expressing BMP-2 and cSmad8 produces less ALP (less coloring) and present elongated shape corresponding to tendon cells. *The images were taken from Hoffmann et al. (2006) and with perssion of the editorial [8].*

C3H10T1/2 cells expressing BMP-2 and cSmad8 (Smad8L + MH2) are shown in Figure 4 (Panel B). The cells were stained with NBT (nitro-blue tetrazolium chloride) and BCIP (5-bromo-4-chloro-3'-indolyphosphate p-toluidine salt) with which it is possible to detect alkaline phosphatase (ALP). ALP is widely used as osteogenic marker [73, 74]. C3H10T1/2 cells expressing BMP-2 clearly showed the presence of ALP (Figure 4, panel B) at day 6 [8]. This is comparable to cells expressing Smad8WT and BMP-2. However, C3H10T1/2 cells expressing BMP-2 and cSmad8 (Smad8L+MH2) showed a noticeable decrease in the ALP production (Figure 4, panel B) at day 6 [8]. Moreover, at day 10, the elongated shape of cells was already distinguishable. More interesting is the fact that the effect caused by BMP-2/cSmad8 was not obtained with other members of the TGF-β family such as TGFβ-1, BMP-2 or GDF5 [8]. Moreover, the combination of Smad1 (Smad1L+MH2) or cSmad5 (Smad5L+MH2) with BMP-s did not produce the formation of tendon cells either [8]. The combination of BMP-2 and the truncated form of Smad8 is then a remarkable combination and the only one that generated tendon formation. Hoffman et al. (2006) [8] performed experiments with C3H10T1/2 cells to explain this particular effect of BMP-2 and cSmad8.

The phosphorylation of Smad8 wild type (WT) in C3H10T1/2 cells activated with BMP-2 showed only a marginal activation (phosphorylation) as observed by immunoblotting [8]. The capacity of DNA-binding of Smad8WT was investigated with Gal4-Smad8WT fusions. The level of activation generated by BMP-2 was quantified by using a Gal4 luciferase reporter [8]. Generally explained, the Gal4 DNA-binding domain is genetically attached to Smads. The transactivation potential is measured thanks to the translation of the luciferase under the control of the Gal4 promoter. This assay showed that Gal4-Smad8WT fusions were activated only to a minor extent. In contrast, Gal4-Smad1WT fusion was shown to be clearly phosphorylated in C3H10T1/2 cells [8].

As mentioned in section 1.2.2, the MH1 domain exhibits DNA-binding capacity and interferes with the effector function of MH2 in absence of phosphorylation [6, 8]. The absence of MH1 makes cSmad8 constitutively active [8]. The phosphorylation in this regard is no longer required to activate cSmad8 [8]. Experiments using Gal4 luciferase reporter, indicated that Smad1- and Smad8-constructs consisting of the linker and the MH2 domain possess a higher capacity of transactivation than the constructs formed by the MH2 domain only. These last constructs were even more active than the

constructs with Smads wild type. This indicated, that the constructs with the linker and the MH2 were more efficient than the construct with MH2 due to the presence of the proline-rich linker domain [8]. It was concluded that cSmad8 is essential for tenogenic induction, since BMP-2 marginally activate cSmad8 [8].

1.3 Protein expression in recombinant systems

Production of proteins using recombinant systems has an important impact in science and technology. In early stages, proteins were obtained from the animal or vegetal source using complicated, expensive and time-consuming procedures. Now, it is possible to create, design, modify and finally produce proteins in a simplified, economical and fast way using bacteria, yeast, filamentous fungi, unicellular algae, insect cells or eukaryotic cells [75]. The selection of the expression system depends on many factors such as the final application of the product, the desired yields, the need to post-translational modifications or the origin of the protein (the organism that naturally produces the protein), among others. Extensive reviews covering the characteristics, advantages and disadvantages of different production systems can be found in literature [76-78].

Escherichia coli (*E. coli*) is of particular interest. *E. coli* is a gram-negative bacteria which is a part of the microflora of humans and other mammals [79]. By 1950, it was already considered as the basic working organism in studies of genetics, transcription, translation, and replication [80]. This intensive work generated detailed knowledge about *E. coli* that allowed the development of protocols, tools, media, plasmids, and modified strains that are available nowadays. *E. coli* is a versatile system due to the available technology around it, low production cost, short replication times and high yields [81]. However, this organism also presents some disadvantages. Production of proteins that require post-translational modifications are not carried out in this system. Moreover, the reducing environment of *E. coli* hamper disulfide-bonds formation, which is necessary for the correct folding of some proteins. Its fast and tightly coupled transcription and translation can be disadvantageous for some proteins, which can result in partially folded, unfolded or completely misfolded products [82]. Additionally, its high yields can also lead to insoluble aggregates of proteins called inclusions bodies. Although they were seen as undesired products during a long time, this perception has largely changed. This will be discussed in more detail in the following section.

1.4 Inclusion bodies (IBs)

IBs are proteinaceous amyloid particles ranging from 50 nm to 1000 nm [83]. They were conceived as inactive and useless and with this, the major drawback of bacterial systems [84]. However, this perspective has changed since then. Now it is known that IBs are not randomly formed by inactive

protein. In fact, they possess a complex amyloid structure formed by stereospecific cross-molecular interactions [85] that can hold active proteins [86] with native or native-like conformations [84]. Embedded protein can be released as observed for the secretory granules in the endocrine system [87]. Conveniently, IBs are non-toxic and totally biocompatible materials [87] that are able to cross membranes and internalize into cytoplasm [84]. These features make them promising self-supported vectors of proteins. Moreover, IBs used to decorate 2D and 3D surfaces favored cell adhesion and they promoted cell proliferation and adhesion [88]. Finally, the IBs mechanical properties such as size, morphology, stiffness, zeta potential and density can be modulated depending on the production conditions [89]. The IBs can be recovered by simple mechanical methods such as centrifugation and they can be reused several times.

With the employment of protocols based on the use of soft detergents, it is possible to obtain pure endotoxin-free IBs that conserve their activity or fluorescence [88, 90, 91]. Additionally, endotoxin-free *E. coli* strains are available on the market. ClearColi™ (Lucigen, USA) is an *E. coli* strain that displays a genetically modified lipopolysaccharide (LPS) which does not cause any endotoxic response in human cells [92].

On the other hand, another approach consist of using chaotropic agents such as Guanidine hydrochloride (GuHCl) or urea to solubilize the protein embedded in IBs. Reducing agents such as Dithiothreitol (DTT) or β-mercaptoethanol are commonly used to eliminate disulfide bonds. Prior to the solubilization, a process of IBs cleaning with detergents and low concentrations of chaotropic agents is carried out to get rid of major contaminants. The best conditions to get a soluble, stable and later active product are usually determined by trial-and-error for each protein [84].

The establishment of denaturing conditions allow the protein to go through a homogenous *in vitro* renaturing process. Proteins can be purified by traditional chromatographic methods before or after renaturing.

2. Objectives

Due to the potential of hcSmad8 in the field of the regenerative medicine (see section of Justification), it is desirable to produce it as a recombinant system that can offer the possibility of industrial scaling. *Escherichia coli* is a well-known system that has shown to be fast, economic and with high yields.

The main objective herein was to produce TAT-hcSmad8 in *E. coli,* to purify it and characterize it by physicochemical methods. Additionally, the effect of the purified product on murine cells (C3H10T1/2) forms part of the scope of this work.

Since TAT-hcSmad8 is observed to be produced in form of IBs, some partial objectives were established:

- To explore the possibility of using IBs as a self-supported delivery system.
  Since IBs are amyloid particles that are composed by native-(like) structure variants of the protein they are made of, in principle they could be used as a vector for TAT-hcSmad8.

- To determine if it is possible to produce TAT-hcSmad8 in a soluble form.
  For this, a screening of production conditions was carried out. Different induction temperatures, times of induction, concentration of inducer, types of inducer and strains were explored.

3. Materials and Methods

3.1 Plasmids construction and strains

Construction of pET15b-TAT-hcSmad8. The sequence of human Smad8 transcript variant b (NCBI Reference sequence: NM_005905.6) plus the HIV-TAT sequence (TAT-cSmad8), was codon-optimized for expression in *E. coli*. The TAT-cSmad8 gene was artificially synthetized and introduced into the pMA-T (GeneArt, Germany, commercial service). From this vector, TAT-cSmad8 gene was sub-cloned into the expression vector pET-15b using *Xho*I and *Bam*HI at its 5´-end and 3´-end, respectively. The resulting product (TAT-hcSmad8) includes a His6-tag coded in pET-15 (Table 2) [2].

Construction of pET29c-TAT-cSmad8. For His6-Tag removal, the TAT-cSmad8 sequence was subcloned into pET29c plasmid (Merck Millipore, Germany) using *Nde*I and *Bam*HI restriction sites (Table 2) [2].

pET15b-TAT-hcSmad8 and pET29c-TAT-cSmad8 were designed and constructed by Dr. Andrea Hoffmann and Dr. Bastian Quaas [93], respectively.

pET15b-TAT-hcSmad8 and pET29c-TAT-cSmad8 were used to transform ClearColi® BL21(DE3) electrocompetent cells (Lucigen, Biosearch Technology, USA) and *E. coli* BL21 (DE3) (Thermo Fisher Scientific, USA) [2].

Construction of pBAD-TAT-hcSmad8. The sequence of TAT-cSmad8 optimized for *E. coli* was introduced by artificial gene synthesis into pBAD/His B (GeneArt, Germany, commercial service) using the restrictions sites *Eco*RI (5´-end cutter) and *Hin*dIII (3´-end cutter) (Table 2). This plasmid was used to transform TOP10 *E. coli* (Thermo Fisher Scientific, USA).

Construction of pGEX-TAT-cSmad8. Similarly to what described before, the TAT-cSmad8 sequence was introduced artificially into pGEX-4T-1 (GenScript, USA) using the restrictions sites *Bam*HI (5´-end cutter) and *Not*I (3´-end cutter). The resulting product displays a Glutathione S-transferase (GST) tag (Table 2). This plasmid was used to transform *E. coli* BL21 (DE3) (Thermo Fisher Scientific, USA), Origami2® (DE3)pLysS (Novagen, Germany) and Rosetta® (DE3) (Novagen, Germany).

Construction of pETM30-His6-GST-GFP. This plasmid encoding His6-GST-GFP (54 kDa) was donated by Dr. Zhaopeng Li. The construction of this plasmid is described in Li et al. (2011) [94]. In this work, it will be referred to this sequence as GST-GFP.

Table 2. Sequences and plasmids used in this work.

*The TAT sequence is double underlined, the linker is printed in italics, MH2 is printed in bold, His6-tag colored with red, amino acids included to confer flexibility to the molecule are printed in grey and GST-tag is dashed underlined (red) *Computed with pI/Mw tool – Expasy [95].*

Plasmid
pET29c-TAT-cSmad8 (35.352 kDa)
MLEG<u>YGRKKRRQRRR</u>G*EYNPQLSLLAKFRSASLHSEPLMPHNATYPDSFQQPPCSALPPSPSHAFSQSPCTASYPHSPGSPSEPESPYQHSD FR***PVCYEEPQHWCSVAYYELNNRVGETFQASSRSVLIDGFTDPSNNRNRFCLGLLSNVNRNSTIENTRRHIGKGVHLYYVGGEVYAECV SDSSIFVQSRNCNYQHGFHPATVCKIPSGCSLKVFNNQLFAQLLAQSVHHGFEVVYELTKMCTIRMSFVKGWGAEYHRQDVTSTPCWI EIHLHGPLQWLDKVLTQMGSPHNPISSVS**
pET15b-TAT-hcSmad8 (35.842 kDa)
MGSSHHHHHHSSGLVPRGSHMLEG<u>YGRKKRRQRRR</u>G*EYNPQLSLLAKFRSASLHSEPLMPHNATYPDSFQQPPCSALPPSPSHAFSQSPC TASYPHSPGSPSEPESPYQHSDFR***PVCYEEPQHWCSVAYYELNNRVGETFQASSRSVLIDGFTDPSNNRNRFCLGLLSNVNRNSTIENTRR HIGKGVHLYYVGGEVYAECVSDSSIFVQSRNCNYQHGFHPATVCKIPSGCSLKVFNNQLFAQLLAQSVHHGFEVVYELTKMCTIRMSFV KGWGAEYHRQDVTSTPCWIEIHLHGPLQWLDKVLTQMGSPHNPISSVS**
pBAD-TAT-hcSmad8 (39.222 kDa)
MGGSHHHHHHGMASMTGGQQMGRDLYDDDDKDPSSRSAAGTIWEFLVPRGSMLEG<u>YGRKKRRQRRR</u>G*EYNPQLSLLAKFRSASLHSE PLMPHNATYPDSFQQPPCSALPPSPSHAFSQSPCTASYPHSPGSPSEPESPYQHSDFR***PVCYEEPQHWCSVAYYELNNRVGETFQASSRS VLIDGFTDPSNNRNRFCLGLLSNVNRNSTIENTRRHIGKGVHLYYVGGEVYAECVSDSSIFVQSRNCNYQHGFHPATVCKIPSGCSLKVF NNQLFAQLLAQSVHHGFEVVYELTKMCTIRMSFVKGWGAEYHRQDVTSTPCWIEIHLHGPLQWLDKVLTQMGSPHNPISSVS**
pGEX-TAT-cSmad8 (59.971 kDa)
<u>MSPILGYWKIKGLVQPTRLLLEYLEEKYEEHLYERDEGDKWRNKKFELGLEFPNLPYYIDGDVKLTQSMAIIRYIADKHNMLGGCPKERAEIS MLEGAVLDIRYGVSRIAYSKDFETLKVDFLSKLPEMLKMFEDRLCHKTYLNGDHVTHPDFMLYDALDVVLYMDPMCLDAFPKLVCFKKRIE AIPQIDKYLKSSKYIAWPLQGWQATFGGGDHPPKSDLVPRGSMLEG</u>G<u>YGRKKRRQRRR</u>G*EYNPQLSLLAKFRSASLHSEPLMPHNATYPDSF QQPPCSALPPSPSHAFSQSPCTASYPHSPGSPSEPESPYQHSDFR***PVCYEEPQHWCSVAYYELNNRVGETFQASSRSVLIDGFTDPSNNRN RFCLGLLSNVNRNSTIENTRRHIGKGVHLYYVGGEVYAECVSDSSIFVQSRNCNYQHGFHPATVCKIPSGCSLKVFNNQLFAQLLAQSVH HGFEVVYELTKMCTIRMSFVKGWGAEYHRQDVTSTPCWIEIHLHGPLQWLDKVLTQMGSPHNPISSVS**

3.2 Cultivation

Test of protein production were performed in small volumes according to the protocols suggested by the manufacturer of *E. coli* BL21 (DE3) (Thermo Fisher Scientific, USA), Origami 2(DE3)pLysS (Novagen, Germany), Rosetta (DE3) (Novagen, Germany). Briefly, transfected *E. coli* cells were cultured in LB containing the corresponding antibiotics at 37 °C to an OD_{600} ~ 0.6 at which point Isopropyl β- d-1-thiogalactopyranoside (IPTG) or arabionose (pBAD) was added. A screening of temperatures, inducer concentration and incubation times was carried out. 16 °C, 20 °C, 25 °C, 30 °C, 35 °C and 37 °C were tested. For low temperatures (16 °C-25 °C), short (4 h) and long (16 h) incubation times were tested, while for high temperatures (30 °C – 37 °C) short incubations times were tested (4 h). For IPTG-inducible plasmids, 0.25 mM, 0.5 mM and 1 mM IPTG were explored. For arabinose-inducible pBAD, 0.00002 %, 0.0002 %, 0.002 %, 0.02 %, 0.2 % arabinose and 4 h were explored according to the manufacturer´s manual.

Large scale protein production using *E. coli* BL21 (DE3) was carried out according to [94] in a 10 L stainless steel bioreactor (B. Braun Biotech International, Germany) containing 10 L of S-DAB autoinduction broth. During bioreactor autoinduction, temperature, pH, aeration, agitation speed was set constant at 37 °C, pH 6.8, 1 vvm and 400 rpm, respectively. After consumption of all carbon sources (about 24 hours after inoculation), the culture was harvested by centrifugation and the cell pellet stored at -80 °C until further usage.

ClearColi® BL21(DE3) cells were cultivated in LB-Miller medium to a OD_{600} ~ 0.6 at 37 °C. At this point, 0.25 mM IPTG were added and it was further incubated for 5 hours. The cells were harvested by centrifugation and the cell pellet stored at -80 °C until further usage.

3.3 Cell lysis and IBs washing

The cell pellet was resuspended into precooled cell lysis buffer (50 mM MOPS pH 7.5, 150 mM NaCl, 1 mM EDTA, 5 mM DTT) with a ratio of 1:10 (w/v). Cells were then lysed using a high pressure homogenizer (M-110L Microfluidizer, Microfluidics, Newton, USA). The pressure was set to 13 500 psi and the sample was subjected to 10 cycles. The IBs were recovered by centrifugation at 10 000 g for 30 min. The IBs pellet can be stored at -80 °C for further use [2].

IBs were resuspended twice into precooled washing buffer A (50 mM MOPS pH 7.5, 2 % (v/v) Triton X-100) by stirring or using an ultra turrax at the lowest speed. Afterwards IBs were centrifuged 20 min at 10 000 g and the resuspension process with washing buffer A was repeated. For the third and fourth washing cycle, washing buffer B (50 mM MOPS pH 7.5, 2 M urea) was used. Finally, washed IBs were rinsed once with ddH20, aliquoted and stored at -80 °C [2].

Due to the insolubility of IBs, a solubilization buffer (50 mM MOPS pH 7.5, 6 M GuHCl, 1 mM EDTA, 25 mM DTT) was used with a ratio of 1:16 (w/v) or 1:10 (w/v) for *E. coli* BL21 (DE3) and ClearColi® BL21 (DE3), respectively. The solubilization was carried out by low-speed stirring at 4 °C for 16 h. The mixture was then centrifuged at 15 000 g for 40 min. The supernatant was carefully filtered using a 0.2 μm polycarbonate filter, then aliquoted and finally stored at -80 °C [2].

3.4 Protein refolding and purification

Protein was purified by two different methods, here referred as Method 1 and Method 2. In general, Method 1 maintains soft conditions during the whole purification process. In contrast, Method 2 establishes a denaturing condition during purification by using urea (4 M) [2].

Before purification, solubilized IBs where refolded by dilution. For this, a small volume of solubilized IBs (no more than 0.5 % of the refolding volume mixture) was added to precooled refolding buffer [**Method 1:** 50 mM MES pH 6.0, 500 mM GuHCl, 750 mM CHES. **Method 2:** 44.44 mM CHES, 33.33 mM HEPES, 22.22 mM citric acid pH 6, 300 mM GuHCl, 750 mM CHES] while stirring to achieve a final concentration of about 50 µg/mL. Then, the refolding mixture was incubated for 2 hours at 8°C with no stirring. This cycle was repeated five times to achieve a final protein concentration of 250 µg/mL. The refolding mixture was finally incubated with no stirring for another 20 h at 8°C [2].

Afterwards, the refolding mixture was slowly diluted 1:2 (v/v) with precooled H_2O (**Method 1**) or diluted 1:1 (v/v) with sample buffer (MES 25 mM pH 6.0, urea 8 M) (**Method 2**).

The sample was then loaded to a 5 mL HiTrap Heparin HP column (GE Healthcare, USA) previously equilibrated with 5 CV of equilibration buffer [**Method 1:** 50 mM MES pH 6.0, 250 mM GuHCl and 375 mM CHES. **Method 2:** 50 mM MES pH 6.0, 4 M urea]. After sample application, the column was washed with 5 CV equilibration buffer. Finally, the sample was eluted using a linear gradient from 1-100 % of elution buffer [**Method 1:** 50 mM MES pH 6.0, 250 mM GuHCl, 375 mM CHES, 2 M NaCl. **Method 2:** 50 mM MES pH 6.0, 4 M urea, 2 M NaCl].The samples were analyzed by SDS-PAGE 12 % and the fractions with higher purity were mixed and dialyzed against storage buffer (50 mM MES pH 6.0, 10 % glycerol).

The purification method following the **Method 1** was optimized for TAT-cSmad8 and TAT-hcSmad8. Once that the protein were loaded to a 5 mL HiTrap Heparin HP column (GE Healthcare, USA), the column was washed with 5 CV of washing buffer with a conductivity value of 45 mS/cm (~ 15 % of elution buffer) for TAT-cSmad8. The buffer with a conductivity value of 50 mS/cm (~ 21 % of elution buffer) was used for TAT-hcSmad8. Finally, TAT-cSmad8 can be eluted with 65 mS/cm (~ 37 % of elution buffer) whereas TAT-hcSmad8 can be eluted with 80 mS/cm (~ 47 % of elution buffer).

After purification, the desired fractions were pooled and dialyzed against storage buffer (50 mM MES pH 6.0, 10 % glycerol) for 20 h at 8 °C. Finally, the protein was cleared from precipitates and aggregates by centrifugation, it was filtered with a 0.2 µm PVDF filter under sterile conditions and quantified with Quick Start ® Bradford Protein Assay (BioRad, USA) according to manufacturer´s protocol. Aliquots were stored at -80 °C until further usage [2].

3.5 SDS-PAGE and Western blot

SDS-PAGE. Cell lysis to be analyzed by SDS-PAGE were prepared with BugBuster® protein extraction reagent (Merck Millipore, Germany) according to manufacturer´s protocol. rLysozyme® (Novagen, Merck Millipore, Germany) and Benzonase® (Novagen, Merck Millipore, Germany) were

supplemented to BugBuster® reagent. Cell lysates and samples obtained from purification were analyzed by 12 % SDS-PAGE and visualized by Coomassie Blue Staining [96]. PageRuler® Plus Prestained Protein Ladder 26619 (Thermo Scientific, USA) was used to estimate molecular weights.

Western blot. After the SDS-PAGE run was complete, proteins were transferred to an Immun-Blot® PVDF membrane (Bio-Rad, USA). For this, the PVDF membrane was activated in 96 % (v/v) ethanol for 5 min. Afterwards, the PVDF membrane and the SDS-gel were equilibrated in transfer buffer (48 mM Tris pH 8.3, 40 mM glycine, 0.04 % (w/v) SDS, 5 % (v/v) isopropanol) for 30 min. Protein transference to the PVDF membrane was carried out using a Trans-Blot Semi Dry Transfer Cell (Bio-Rad, USA) set to 15 V for 30 min. The PVDF membrane was later blocked overnight at 8 °C using 5 % (w/v) skim milk (SERVA, Germany) in TBS-T buffer (25 mM Tris-HCl pH 7.4, 150 mM NaCl, 0.15 % (v/v) Tween-20). After blocking, it was washed with TBS-T buffer in 3 cycles of 5 min [2].

Subsequently, the PVDF membrane was incubated with TBS-T buffer containing 2.5 % of skim milk (SERVA, Germany) and 0.1 µg/mL of goat anti-Smad8 antibody AF2309 (R&D, USA) at room temperature for 1 h. The membrane was washed with TBS-T buffer and later incubated with Goat IgG HRP-conjugated Antibody HAF017 (R&D, USA) under the same conditions mentioned before. Finally, the membrane was washed one more time and the protein bands were stained using 3,3',5,5'-Tetramethylbenzidine (TMB) T065 (Sigma Aldrich, Germany) [2].

3.6 Solubility test

To test the solubility of TAT-hcSmad8 under different pH conditions, purified TAT-hcSmad8 in storage buffer (50 mM MES pH 6.0, 10 % glycerol) was diluted (1:10) in different buffers to a final concentration of 50 µg/mL. 127 mM of sodium acetate pH 4.0, MES pH 5.5, MOPS pH 7.0, sodium borate pH 8.0-10.0 and carbonate-bicarbonate pH 9-10.5 in presence or absence of 20 % glycerol or 200 mM NaCl were tested. The tested samples were incubated for 1 h at 8 °C in a transparent pureGrade 96 well plate (Brand, Germany). Finally, the absorbance at 340 nm was measured using a Multiskan GO spectrophotometer (Thermo Fischer Scientific, USA) [2].

3.7 Thermal Shift Assay (TSA)

Purified TAT-hcSmad8 in 50 mM MES pH 6.0 was mixed with SYPRO Orange solution 5000X (Sigma Aldrich, Germany) to a final concentration of 50 µg/mL and 20X SYPRO Orange, respectively. The mixture was incubated for 5 min in ice and the fluorescence was later monitored using a 545-585 nm cut-off filter in an iQ5 Real time PCR (Biorad, USA) while the temperature was raised stepwise from 35 °C to 100 °C by 1 °C per min [2].

3.8 Circular dichroism (CD) and structure prediction

The theoretical structure of TAT-hcSmad8 was developed using the Predictor of Natural Disordered Regions (PONDR) [97] server (accessed on April 2019). The resulting model was visualized with PyMOL Molecular Graphics System, v2.0, Schrödinger, LLC.

On the other hand, secondary structure determination was performed by CD. For this, the CD spectrum of 5 µM TAT-hcSmad8 and TAT-cSmad8 in 10 mM potassium phosphate (K_2HPO_4/KH_2PO_4) pH 5.8 was recorded using a J-815 spectropolarimeter (JASCO, Japan) at 20 °C [2]. The data acquisition were set to an interval of 1 nm, scanning speed 20 nm/min, 1 nm bandwidth, 3 scans averaging and cell path length 0.1 cm [2].

3.9 Cy3 labeling

1 mg/mL of TAT-hcSmad8 in 127 mM MES pH 6.0, 20 % glycerol was added to one stock of Invitrogen Cy3 dye (ThermoFisher Scientific, USA). Some adaptations were performed to the manufacturer´s protocol. Briefly, the reaction mixture was incubated for 3 h at 8 °C in a thermomixer Comfort 1.5 mL (Eppendorf, Germany). Every 30 min, the sample was agitated for 5 min using 300 rpm. Afterwards, the sample was centrifuged at maximum speed for 20 min at 4 °C and then loaded to a PD-10 desalting column (GE Healthcare Life Sciences). Fractions with labeled protein were pooled, filtrated through a 0.22 µM PES filter under sterile conditions, aliquoted and stored at -80 °C [2].

3.10 Cell culture

C3H10T1/2 cells (clone 8), C3H10T1/2-BMP2, C3H10T1/2-TGFβ1, and C3H10T1/2-GDF5 cells, were donated by Prof. Dr. rer. nat. Andrea Hoffmann (Hannover Medical School) [8].

BRE-luc C3H10T1/2 cells were donated by Delphine Logeart (French National Centre for Scientific Research) [10].

C3H10T1/2 cells were cultured in high-glucose Dulbecco's Modified Eagle's Medium (DMEM) (Merck, Germany) supplemented with 10 % heat-inactivated fetal calf serum (FCS) (Biochrom, USA), 2 mM L-glutamine and antibiotics (50 U/mL penicillin and 50 mg/mL streptomycin) (Biochrom, USA) [2].
750 µg/mL G418 (Gibco, Life Technologies Corporation, USA) and 5 µg/mL of puromycin (Gibco, Life Technologies Corporation, USA) were used for C3H10T1/2-BMP2, C3H10T1/2-TGFβ1, and C3H10T1/2-GDF5 cells [8].

C3H10T1/2 cells containing the BRE-Luc reporter gene [10] were cultured in MEM (Merck, Germany) supplemented with 10 % heat-inactivated FCS (Biochrom, USA), 2 mM L-glutamine, 50 U/mL penicillin (Biochrom, USA), 50 mg/mL streptomycin (Biochrom, USA) and 200 µg/mL G418 (Gibco, Life Technologies Corporation, USA) [2].

All cell types were incubated at 37 °C and 5 % CO_2.

3.11 Fluorescence microscopy and flow cytometry

Fluorescence microscopy. C3H10T1/2 cells were seeded at a density of 6 x 10^3 cells/cm^2 on a 96 well plate (Sarstedt, Germany) under the conditions described in Section 4.10. After 48 hours the medium was exchanged with fresh medium containing 10 µg/mL of Cy3-labeled TAT-hcSmad8 or Cy3-labeled TAT-hSmad8 and the cells were incubated for 2, 24, 48 and 36 h at 37 °C and 5 % CO_2. After the incubation period, the cells were rinsed with phosphate-buffered saline (PBS) pH 7.4 (Genaxxon Bioscience GmbH, Germany) and the fluorescence was examined in a Cytation 5 Cell Imaging Multi-Mode Reader (BioTek, USA) [2].
Alternatively, confocal laser scanning microscopy (CLSM) was performed with the confocal laser scanning microscope TCS SP8 (Leica, Germany). Detection channel photomultiplier was set to 555 nm-625 nm to visualize Cy3-labeled protein. Images were analyzed with the software Imaris, version 8.4 (Bitplane).

Flow cytometry. C3H10T1/2 cells were seeded at a density of 6 x 10^3 cells/cm^2 on a 12 well plate (Sarstedt, Germany) under the conditions described in Section 4.10. At confluence, the medium was exchanged with fresh medium containing 1, 10 or 100 µg/mL of Cy3-labeled TAT-hcSmad8. After 0.5 and 2 h incubation, the cells were rinsed with PBS pH 7.4, detached with trypsin (Gibco, Life Technologies Corporation, USA) and the fluorescence of the cells was quantified by flow cytometry (BD Accuri C6, BD Biosciences, USA). By forward scatter, cell debris was excluded. At least 10 000 events per sample were considered for the analysis and the data was processed with BD Accuri C6 software [2].

3.12 Immunocytochemistry

C3H10T1/2 cells were seeded at a density of 6 x 10^3 cells/cm^2 on a 96 well plate (Sarstedt, Germany) under the conditions described in Section 4.10. After 48 hours the medium was exchanged with fresh medium containing 6 µg/mL of TAT-hcSmad8 or TAT-hSmad8 and the cells were incubated for 2 h or 24 h at 37 °C and 5 % CO_2. After the incubation period, the cells were rinsed three times with pre-

warmed PBS pH 7.4 (Genaxxon Bioscience GmbH, Germany). Then, they were fixed with precooled 70 % ethanol for 5 min at -20 °C. The ethanol was discarded and the cells were rinsed three times with PBS pH 7.4. Afterwards, the fixed-cells were blocked with 5 % (w/v) bovine serum albumin (BSA) (Thermo Fisher Scientific, USA) in PBS pH 7.4 for 1 h. After rinsing three times with PBS pH 7.4, the cells were incubated for 24 h at room temperature with a goat anti-Smad8 antibody AF2309 (R&D, USA) diluted 1:100 in PBS pH 7.4 and 3 % BSA. The cells were washed 3 times with PBS pH 7.4 with cycles of 3 min incubation. The cells were then incubated for 1 h at room temperature with PBS pH 7.4 containing 3 % BSA and secondary monkey anti-goat Alexa 488 antibody A11055 (Thermo Fisher Scientific, USA) (dilution 1:100). After washing as described before, cells were incubated 15 min at room temperature with PBS pH 7.4 containing 5 µg/mL of 4',6-diamidino-2-phenylindole (DAPI) and 3 % BSA. After the final washing, the fluorescence was visualized with a Cytation 5 Cell Imaging Multi-Mode Reader (BioTek, USA) with Gen5 8.0 software [2].

3.13 Activity assays

Gene expression analyses by RT-PCR. C3H10T1/2 were seeded at a density of 6×10^3 cells/cm^2 on a 6 well plate (Sarstedt, Germany) as described in 4.10. After reaching confluence, the cells were incubated with BMP-2 (500 ng/mL) and increasing concentrations of TAT-hcSmad8 (0, 5, 10, 25 µg/mL). The DMEM was supplemented with 10 % heat-inactivated fetal calf serum (FCS) (Biochrom, USA), 2 mM L-glutamine and antibiotics (50 U/mL penicillin and 50 mg/mL streptomycin) (Biochrom, USA), 50 µg/mL L-ascorbic acid (Merck, Germany) and 10 mM β-glycerophosphate disodium salt hydrate (Merck, Germany). L-ascorbic acid and β-glycerophosphate disodium salt hydrate were used as osteogenic supplements [98].

The cells were incubated during 12 days and the medium was exchanged every third day. After the incubation period the cellular RNA was prepared using RNeasy ® Kit (QIAGEN, Germany) according to the manufacturer´s protocol.

The RNA was reverse transcribed and the cDNA was subjected to RT-PCR by Prof. Dr. rer. nat. Andrea Hoffmann (Hannover Medical School). RT-PCR was normalized with the house-keeping gene Rsps29.

Histological analysis. Cells were cultivated as described for the gene expression analysis. In this case, after 12 days of incubation, cells were fixed and the alkaline phosphatase (ALP) was visualized by staining the cells with NBT/BCIP Ready-to-Use Tablets (Merck, Germany) according to manufacturer´s manual. Briefly, cells were rinsed with PBS pH 7.4 and later fixed with fresh prepared 4 % (v/v) formaldehyde (Merck, Germany) in PBS pH 7.4 for 20 min at room temperature. The cells were rinsed

with PBS pH 7.4 and incubated with a fresh prepared NBT/BCIP solution for 30 min. The reaction was stopped with repeated washes with PBS pH 7.4.

C3H10T1/2 reporter cells: Mouse cells C3H10T1/2 containing a BMP-responsive element (BRE) fused to the firefly luciferase (Luc) reporter gene [10] were provided by Dr. Delphine Logeart (Université Paris).

BRE-Luc C3H10T1/2 cells were seeded at a density of 6×10^3 cells/cm^2 in a 96 well plate (Sarstedt, Germany) under de conditions and using the medium as described in Section 4.10. After confluence, the medium was exchanged with MEM (Merck, Germany) supplemented with 10 % heat-inactivated FCS (Biochrom, USA), 2 mM L-glutamine, 50 U/mL penicillin (Biochrom, USA), 50 mg/mL streptomycin (Biochrom, USA) and 200 µg/mL G418 (Gibco, Life Technologies Corporation, USA), different concentrations of TAT-hcSmad8 (0, 5, 10, 25 and 50 µg/mL) purified by Method 1 or purified by Method 2. Bovine Serum Albumin (BSA) (Merck Millipore, Germany) was supplemented instead of TAT-hcSmad8 as control. The cells were incubated at 37 °C and 5 % CO_2 for 48 h. After this, the cells were rinsed twice with PBS and the luciferase activity was quantified and a proliferation assay was performed [2].

The proliferation assay was performed by adding CellTiter-Blue (CTB, Promega, USA) according to the manufacturer´s protocol. Shortly, the rinsed cells were incubated for 20 min at at 37 °C and 5 % CO_2 with 100 µL of a solution of 10 % CTB (v/v) diluted with basal MEM. The fluorescence was measured every 20 min at an extinction wavelength of 544 nm and an emission wavelength of 590 nm using a microplate reader (Fluoroskan Ascent, Thermo Fisher Scientific Inc., Waltham, MA, USA). Fluorescence values between 200 nm and 400 nm were obtained after 30 or 40 min and these values were used for data analysis [2].

The luciferase activity was measured by using Firefly Luciferase Assay Kit (Canvax, Spain) according to the manufacturer´s instructions. The relative luciferase units (RLU) were normalized to the relative number of cells obtained in the proliferation assay (RLU/relative number of cells) [2].

3.14 Cytotoxicity assay: Lactate dehydrogenase

The cytotoxicity assay was performed using the cytotoxicity Detection Kit Plus (LDH) (Roche, Germany) according to the manufacturer´s instructions. The cultivation conditions used are the same conditions described in section 4.10. The fluorescence at 492 nm was examined in a Cytation 5 Cell Imaging Multi-Mode Reader (BioTek, USA).

3.15 IBs as a self-supported delivery system: Purification

ClearColi® BL21(DE3) cells were cultivated in LB-Miller medium to a $OD_{600} \sim 0.6$ at 37 °C. At this point, 0.25 mM IPTG were added and it was further incubated for 5 hours. The cells were harvested by centrifugation, then rinsed with a solution of 0.9 % NaCl and the cell pellet stored at -80 °C until further usage.

IBs production were performed according to [91]. For this, the cell pellet was resuspended into precooled cell lysis buffer 2 (50 mM MOPS pH 7.5, 150 mM NaCl, 1 mM EDTA, 5 mM DTT, lysozyme 1 mg/mL, PMSF 100 mM) with a ratio of 1:10 (w/v). The lysis mixture was incubated 2 h at 37°C. Afterwards, Triton X-100 was added to a final concentration of 0.5 % (v/v) and the mixture was incubated for an additional hour at room temperature under agitation. Later, the lysis mixture was sonicated using cycles of 0.5 min and an amplitude of 40 % for 10 min in total. The sample was kept on ice especially during this procedure to avoid overheating. At this point, the viable cell were tested by inoculating an LB-agar plate with the corresponding antibiotics. The plates were incubated 24 h at 37 °C and the lysis mixture was stored at -80 °C. When no growth was detected on the LB-agar plates, Nonident P40 was added to a final concentration of 0.25 % (v/v) [or 5 μL Nonident P40 per 20 mL of lysis mixture] and the mixture was incubated for 1 h at 4 °C under agitation. After the incubation period, DNase and $MgSO_4$ were added to a final concentration of 0.75 μg/ml and 0.75 mM, respectively. In other words, per 20 mL of lysis mixture, 15 μL of DNase (1 mg/mL) and 15 μL of $MgSO_4$ (1 M) were added. The sample was incubated for 45 min at 37 °C under agitation. The IBs were harvested by centrifugation at maximum speed at 4 °C for 15 min. Finally, the IBs were washed with lyses buffer 2 supplemented with 0.5 % Triton X-100. The IBs were harvested and resuspended in PBS. All these procedures were performed using sterile conditions. A final growth-test was performed inoculating LB-medium without antibiotics and in DMEM medium, which were incubated during 24 h at 37 °C. IBs were stored at -80 °C until further usage.

4. Results

4.1 Protein production

E. coli produces TAT-hcSmad8 mainly in form of IBs. Low production rates were observed to improve soluble expression. Low rates expression was achieved by inducing the protein production in *E. coli* Tuner (DE3) with low concentration of IPTG [93]. However, low soluble expression, low purification yields and low purity were obtained with this strategy. Despite of the extensive screening of conditions, it was not possible to find suitable condition for soluble expression [93].
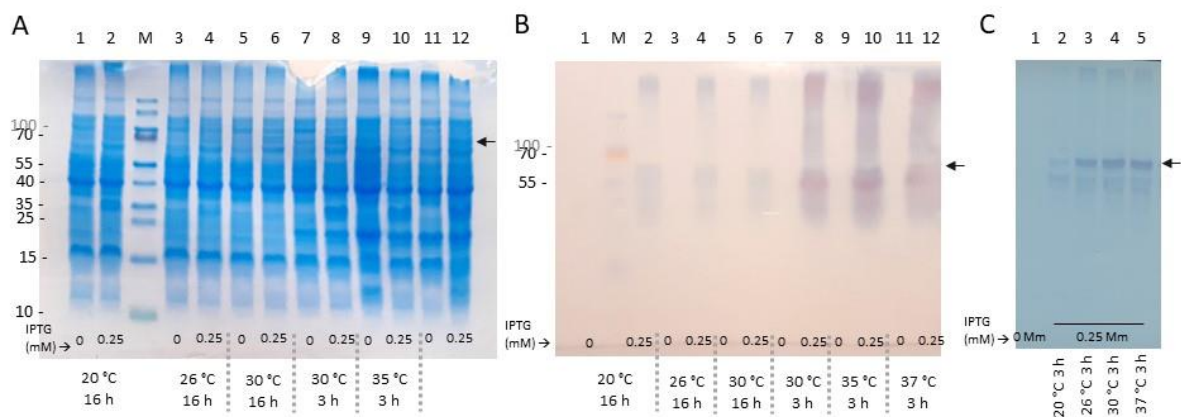
Figure 5. Analysis of protein production by SDS-PAGE and Western blot.
GST-TAT-hSmad8 produced in Rosetta® 2 (DE3) pLysS using 0.25 mM IPTG at different temperatures. Cell lysates were analyzed by SDS-PAGE (A) and Western blot (B). Similarly, GST-TAT-hSmad8 was produced in E. coli BL21 (DE3) and analyzed by Western blot. The protein of interest is indicated with a back arrow at the left side for each panel (in KDa). The expected weights and sequences can be seen in Table 2 in Section 4.1. Additionally, the induction time (after addition of IPTG to a final concentration of 0.25 mM) and the temperature is described under the figures in each panel.

A common strategy to improve the solubility of recombinant protein is the usage of fusion proteins [99]. In this work, glutathione-S-transferase (GST) tag was fused to TAT-cSmad8 and a screening of conditions was performed. Different temperatures (16 °C – 37 °C), incubation times (4 h – 16 h), IPTG concentration (0.25 mM- 1 mM) were tested. These conditions were explored with, *E. coli* BL21 (DE3), Rosetta® 2 (DE3) pLysS and Origami® (DE3). Low temperatures (16 °C – 25 °C) lead to low protein production-levels when compared with higher temperatures and shorter induction times (Figure 5). However, the production of GST-TAT-cSmad8 in form of IBs were observed for the three strains. Moreover, a sub product with a molecular weight of approximately 50 kDa was co-produced in a high proportion in Rosetta® 2 (DE3) pLysS and Origami® (DE3) . In *E. coli* BL21 (DE3) other co-products were obtained (Figure 5, panel B). Similar results were obtained with pBAD-TAT-hcSmad8 in TOP10 *E. coli*

cells (Figure 5, panel C). It was concluded that *E. coli* produced inherently TAT-cSmad8/ TAT-hcSmad8 in form of IBs. To exploit the advantages of IBs, the conditions were optimized to produce high amounts of IBs which later would be purified with chaotropic agents or detergents, refolded and purified by affinity chromatography.
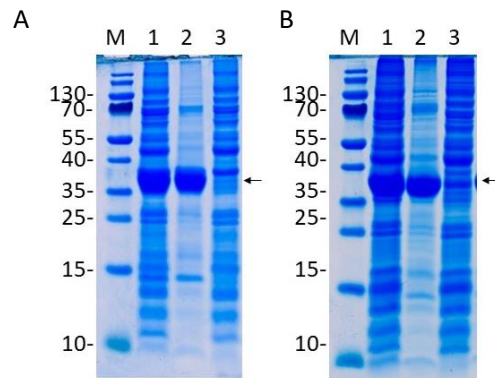

Figure 6. SDS-PAGE of purified protein.
SDS-PAGE analysis of TAT-cSmad8 (35.4 kDa) (A) and TAT-hcSmad8 (35.8 kDa) (B). M: Pre-stained marker; 1: complete cell lysate (soluble + insoluble fraction); 2: insoluble fraction; 3: soluble fraction.

According to best conditions found, protein expression was carried out by culturing *E. coli* cells in LB broth (shake flask, 180 rmp, supplemented with the appropriate antibiotic) at 37 °C to mid-exponential phase (Abs$_{600 nm}$ ~ 0.6), at which point IPTG was added to a final concentration of 0.25 mM. The cell cultures were incubated for three additional hours at same temperature. Afterwards, they were harvested and lysed by French press. By densitometry analysis, it was obtained that around 20 % of the protein present in the cell, corresponds to the desired overexpressed protein. After lysis, it was obtained around 0.8-1.5 g of insoluble material per L of culture, from which around 40 % correspond to the protein of interest (Figure 6). For the protein production using ClearColi® BL21 (DE3), the total induction time was increased to 5 hours due to its lower growth rate.

Large scale protein production using *E. coli* BL21 (DE3) was carried out in a 10 L bioreactor as described in the methods section.

4.2 IB purification

After the production of TAT-cSmad8/TAT-hcSmad8 in *E. coli*, the cells were lysed and the insoluble fraction was cleaned with buffers containing Triton 2 % (washing buffer A) or 4 M urea (washing buffer B) (Figure 7, panel A). The final sample was solubilized in 6 M GuHCl, which is the maximum concentration that the solubility of GuHCl permits. In panel B of Figure 7 the percentage of the protein that can be solubilized under different concentrations of GuHCl (3 M – 6 M) can be observed. The

values were normalized with the highest concentration of solubilized protein (concentration of total protein solubilized with 6 M GuHCl). The use of 5 M GuHCl resulted in an unstable solubilization mixture that led to dramatic precipitation during refolding. Addition of aggregation suppressor such as L-Arginine (0.5 M) did not reduce precipitation. For further steps, a concentration of 6 M GuHCl was used for solubilization.
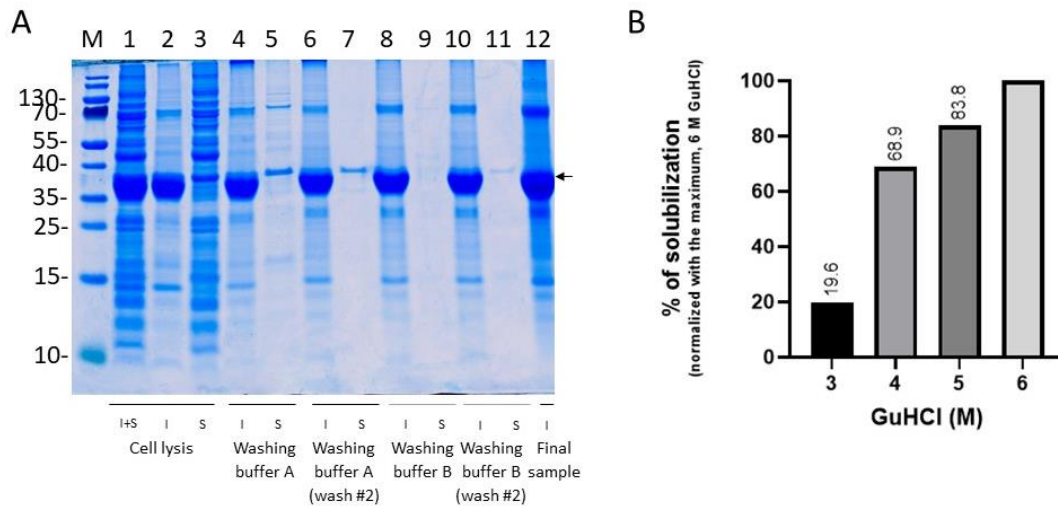

Figure 7. Purification of IBs.
SDS analysis of samples obtained during the cleaning of TAT-cSmad8 IBs (A). M: Pre-stained marker; 1: complete cell lysate [soluble (S) + insoluble fraction (I)]; 2: insoluble fraction; 3: soluble fraction; 4-7: washing with washing buffer A; 8-11: washing with washing buffer B; 12: final sample which was solubilized in solubilization buffer containing 6 M GuHCl. The normalized rates of total solubilized protein are displayed in panel B (6 M GuHCl is considered as 100 % of recovery or maximum recovery)

Under the conditions at which TAT-cSmad8/ TAT-hcSmad8 IBs were solubilized (high concentration of GuHCl and reducing conditions), the protein is thought to be in an unfolded state where disulfide bonds are disrupted. The refolding process was performed by diluting the refolding mixture into refolding buffer. This is a controlled *in vitro* process that ensures that all molecules of TAT-cSmad8/TAT-hcSmad8 are refolded under same condition, leading to a common folding-state. Afterwards, the refolding mixture was purified by Heparin chromatography.

TAT-cSmad8 and TAT-hcSmad8 (including a His6-tag) were purified by Heparin chromatography following two different purification protocols. The purification Method 1 kept "native conditions" during the whole process, whereas for the purification Method 2, urea 4 M was used to create "denaturing conditions". The chromatograms of such purifications are shown in Figure 8.
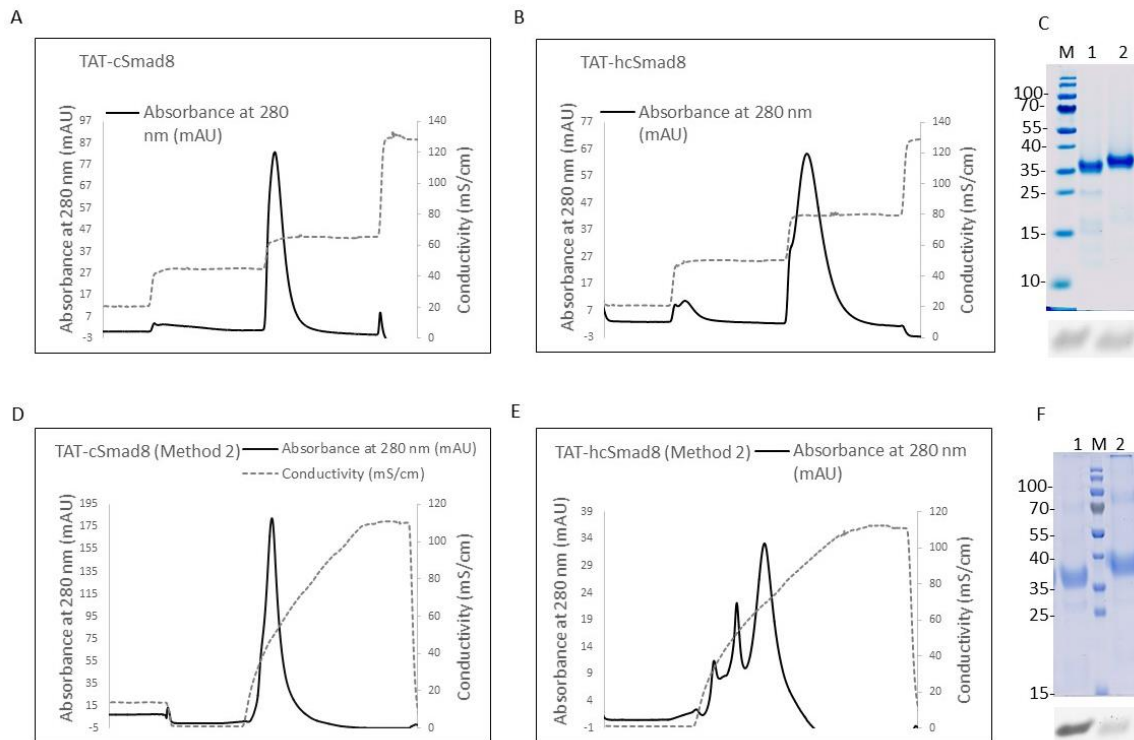
Figure 8. Chromatograms of the purification of TAT-cSmad8 and TAT-hcSmad8.
The purification Method 1 was followed for TAT-cSmad8 (A) and TAT-hcSmad8 (B) leading to highly pure samples analyzed by SDS-PAGE (C, above) and Western blot (C, under); M: marker; 1: purified TAT-cSmad8; 2: TAT-hcSmad8. Similarly, TAT-cSmad8 (D) and TAT-hcSmad8 (E) were purified by Method 2. The samples were analyzed by SDS-PAGE (C, above) and Western blot (C, below); M: marker; 1: purified TAT-cSmad8; 2: TAT-hcSmad8.

4.3 Physicochemical characterization

TAT-cSmad8 and TAT-hcSmad8 purified by Method 1 and Method 2 (urea-based Method) were analyzed by thermal shift assay (Figure 9). With this method, the presence of a hydrophobic core was observed in TAT-cSmad8 and TAT-hcSmad8 purified by Method 1 (Figure 9, panel A). Moreover, the melting point (Tm) corresponded to a value of 46.6 ± 0.1 °C and 47.1±0.1 °C for TAT-cSmad8 and TAT-hcSmad8, respectively (Figure 9, panel B).

The presence of the hydrophobic core was not observed in samples purified by Method 2 (Figure 8, panel C) as no characteristic peak was obtained.
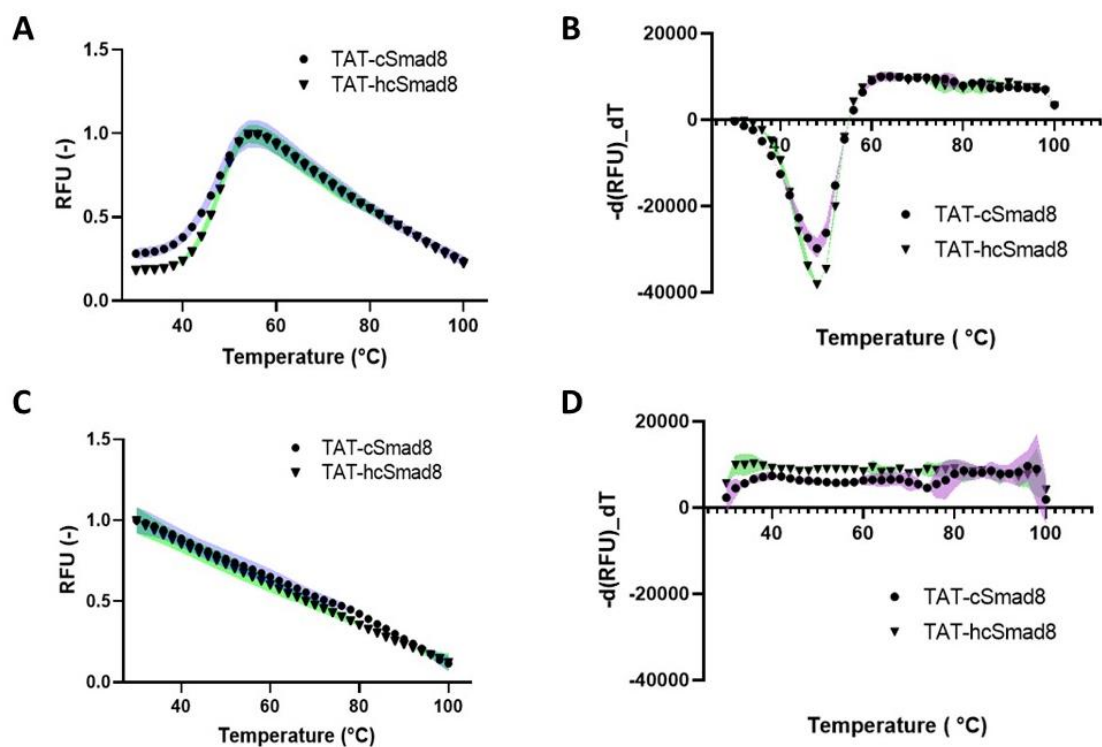
Figure 9. Thermal shift assay analysis of purified samples

Thermal shift assay of TAT-cSmad8 and TAT-hcSmad8 purified by Method 1 (A and B) and Method 2 (C and D). The melting curve is displayed in panel A and C. The second derivative of the function of the melting curve is displayed in panel B and D. The minimum indicates the inflexion point of the melting curve which corresponds to the Tm. Colored areas around data-points indicate standard deviation of three replicates. RFU refers to the relative fluorescence units, whereas –d(RFU)_dT refers to the negative derivative of the function of RFU.with respect to temperature.

More insights of the secondary structure of TAT-cSmad8 and TAT-hcSmad8 purified by Method 1 were obtained by circular dichroism (CD) spectroscopy (Figure 10, panel A).The overlapping CD-spectra of TAT-cSmad8 and TAT-hcSmad8 discarded the possibility that the His6-tag modified the protein pledge.

Essentially, negative circular dichroism values were obtained from 198 nm to 247 nm followed by a minimum at 207 nm. Positive values were obtained at wavelengths lower than 198 nm [100, 101]. This analysis indicated that the structure of TAT-cSmad8 and TAT-hcSmad8 were formed by a mixture of α-helixes and β-sheet. α-Helical structures present both negative values at 208 and 222 nm and positive values at 192 nm. Sheet structures typically display a negative value at 217 nm and a positive value at 195 nm. Unordered structure is discarded as unordered peptides possess a single slump below 200 nm with negative values [100, 101]. Theoretical model obtained with the software PONDR® predicted a

structure consisting of two sets of antiparallel β-strands arranged as a β-sandwich flanked by a triple-helical bundle on one side and a large loop (Figure 10, panel B).

A

B
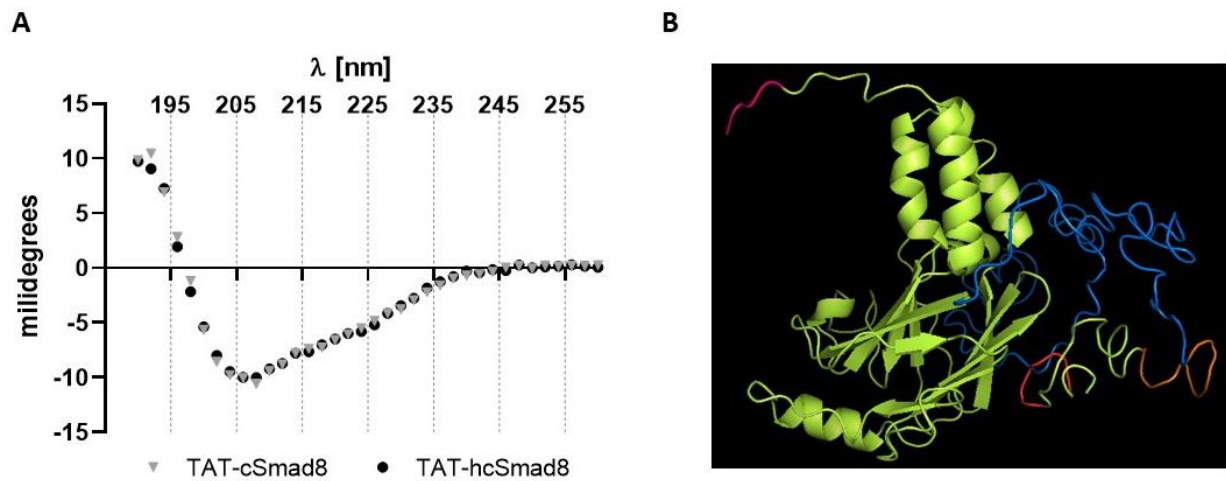
Figure 10. Structural analysis and prediction
CD spectra of TAT-cSmad8 and TAT-hcSmad8 purified by Method 1 (A). Theoretic structure of TAT-hcSmad8 made with PONDR ® software (B). SSVS motif in pink; main structure in green; loop in blue; TAT sequence in orange; histidine tag in red.

The physicochemical analyses indicated that TAT-cSmad8 and TAT-hcSmad8 are structurally equivalent. However, in the praxis the histidine tag seemed to have a stabilizing effect. In general, the yields of purification, labeling and stability were higher for TAT-hcSmad8. Therefore, further analysis was performed with TAT-hcSmad8.

Solubility assays were performed by incubating purified TAT-hcSmad8 into different buffers. The formation of aggregates or precipitates was followed by measuring the absorbance of the samples at 340 nm (Figure 11). Basic pH was set by sodium borate buffer (Figure 11, panel A) and by carbonate-bicarbonate buffer (Figure 11, panel B). The aggregation observed at basic pH was suppressed by addition of glycerol (20 % final concentration). On the other hand, addition of NaCl increased aggregation.
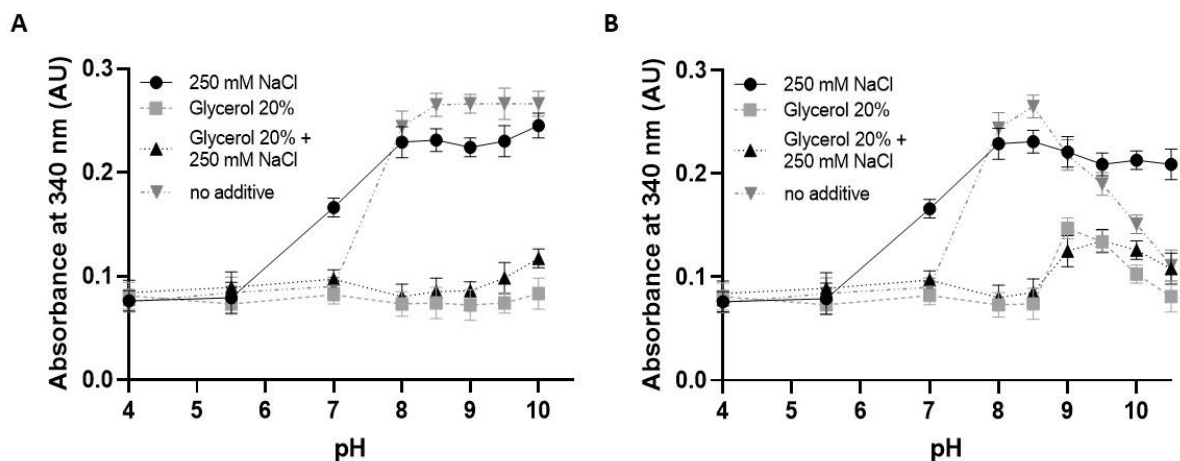
Figure 11. Solubility test of TAT-hcSmad8.

Purified protein was diluted into sodium acetate (pH 4.0), MES (pH 5.5), MOPS (pH 7.0), sodium borate (pH 8.0 – pH 10.0) in A and in carbonate-bicarbonate (pH 9 – pH 10.5, instead of sodium borate) in B. Values are presented as mean ± SD. Figure taken from Segovia et al., 2021 [2].

4.4 Internalization test

Since Smad8 is an intracellular factor, great effort was conducted to the verification of TAT-hcSmad8 internalization.

C3H10T1/2 cells were incubated with 1, 10 and 100 µg/mL of TAT-hcSmad8 previously labeled with Cy3. The incubated cells were detached and analyzed by flow cytometry (Figure 12, panel A). Significant fluorescence was even detected after 30 min incubation with 1 µg/mL. The intensity of fluorescence was proportional to the concentration of Cy3-labeled TAT-hcSmad8 and to the incubation time. This results suggested interaction between Cy3-labeled TAT-hcSmad8 and C3H10T1/2 cells.

To verify the complete internalization of TAT-hcSmad8 into C3H10T1/2 cells and to discard that the detected fluorescence originated by the superficial interaction of TAT-hcSmad8 with the membrane of the cells, confocal scanning microscopy was performed. With this technique, the presence of Cy3-labeled TAT-hcSmad8 was clearly detected in the inner layers of cells. TAT-hcSmad8 was found mainly in the cytoplasm and in less proportion in the nucleus of cells (Figure 13).
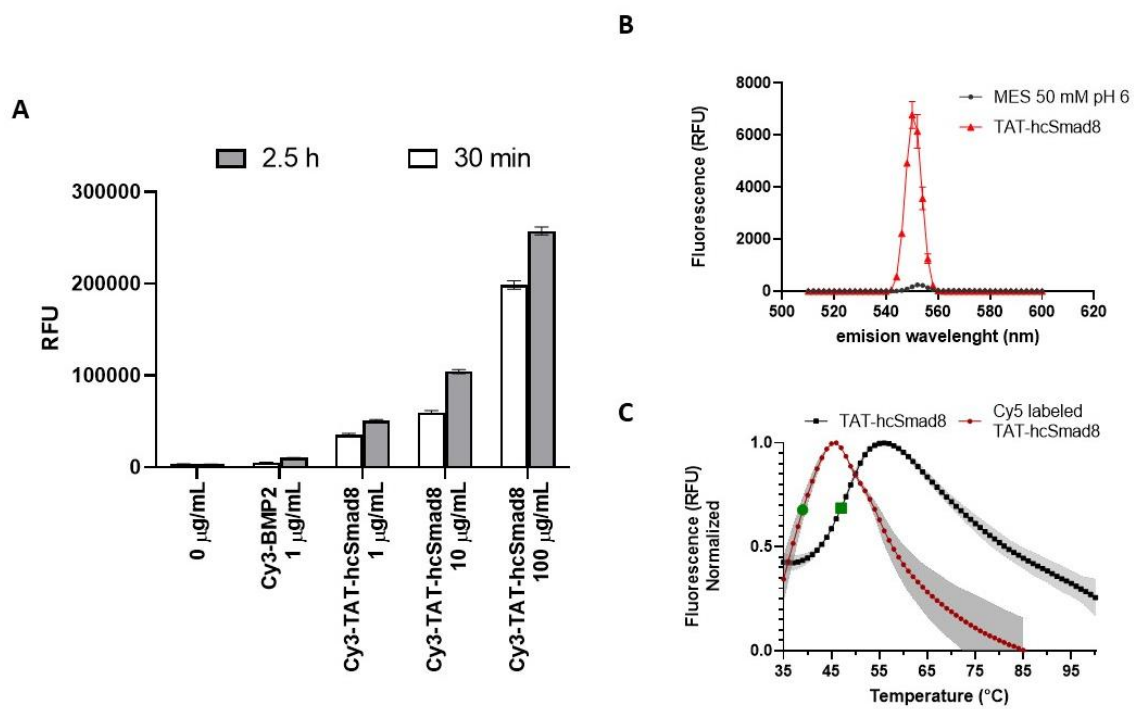
Figure 12. Flow cytometry analysis.
C3H10T1/2 cells incubated with Cy3-labeled TAT-hcSmad8 and Cy3-labeled BMP-2 (negative control) for 30 min and 2.5 h (A). Fluorescence spectra of Cy3-labeled TAT-hcSmad8 (B) and its corresponding thermal shift analysis (C). Colored areas around data points indicate standard deviation of three replicates.
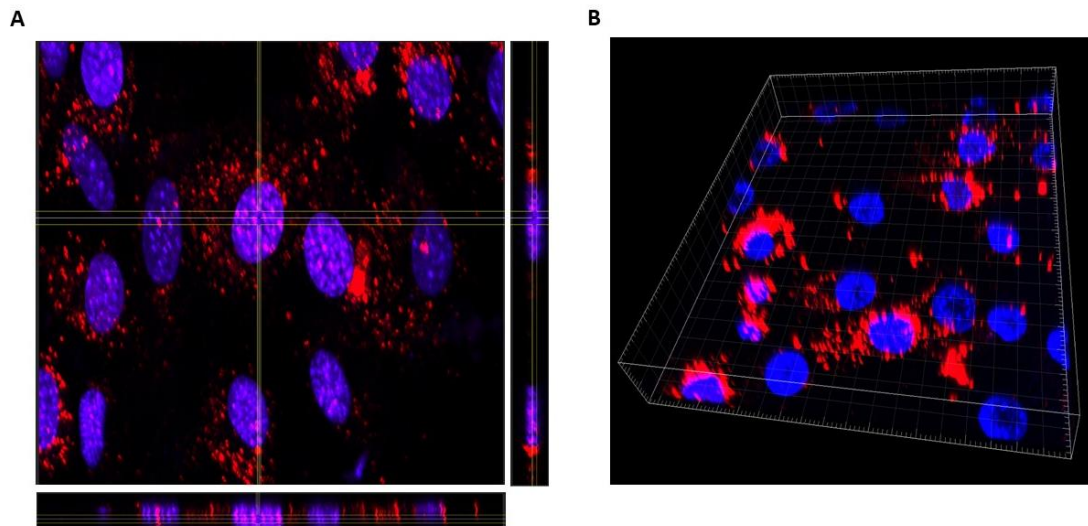

Figure 13. Images of cells analyzed by confocal scanning microscopy.
The presence of Cy3-labeled TAT-hcSmad8 (in red) was found in the middle stack of C3H10T1/2 (A), mainly in the cytoplasm and in less proportion, in the nucleus (in blue). 3D-model (B) obtained with the software IMARIS ®.

TSA performed to Cy5-labeled TAT-hcSmad8 showed the decrease of the Tm of labeled protein (Figure 12, panel C) in around 10 degrees. The procedure of labeling itself, or the formation of the covalent bound between the Cy3 (and Cy5) and the amino groups of the protein could negatively affect the folding of TAT-hcSmad8. The decrease of the Tm might indicate that there was a slight unfolding process which in turn could generate the overexposure of the TAT sequence. Antibodies used in dot blot analysis [93] directed against His-tag were able to detect the his6-Tag. This indicated that at least the external part which comprises the His-tag is exposed. Since the His-tag is followed by the TAT sequence, it is probable that the TAT sequence is also exposed.

To verify that the unlabeled protein is also able to be internalized by cells, C3H10T1/2 cells were incubated with 3 µg/mL of TAT-hcSmad8 purified by Method 1 and Method 2 (urea-based Method).
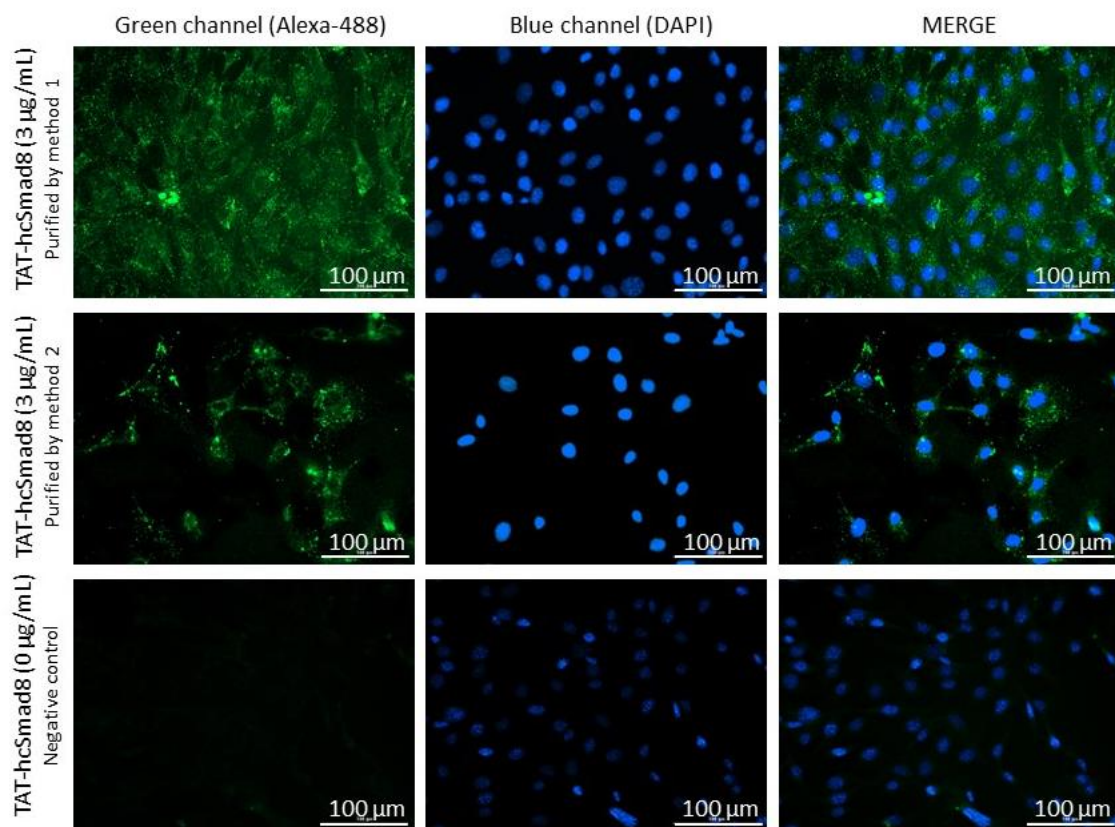
Figure 14. Immunocytochemistry of C3H10T1/2 cells incubated with TAT-hcSmad8.
TAT-hcSmad8 was purified by Method 1 and Method 2 (urea-based Method). Green fluorescence is derived from donkey anti-goat Alexa 488 antibody coupled to goat anti-Smad8 antibody. Counterstaining of cell nuclei was performed with DAPI.

The presence of TAT-hcSmad8 purified by both methods was clearly observed in the intracellular part of cells (Figure 14). Moreover, analysis by confocal microscopy confirmed the presence of TAT-hcSmad8 in the internal space of the cells (Figure 15).
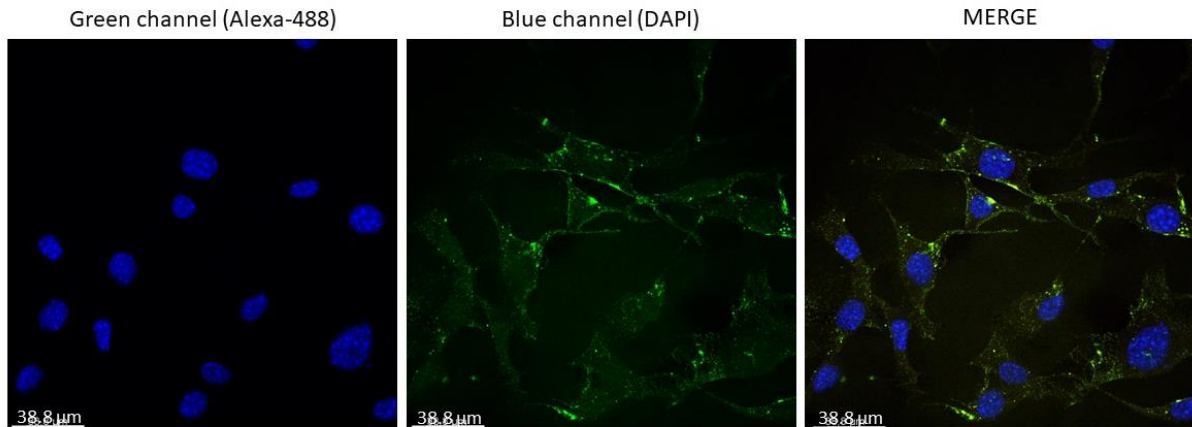

Figure 15. Confocal analysis of a middle stack of C3H10T1/2 cells.
The cells were incubated with TAT-hcSmad8 purified by Method 1. Green fluorescence is derived from donkey anti-goat Alexa 488 antibody coupled to goat anti-Smad8 antibody. Counterstaining of cell nuclei was performed with DAPI.

4.5 Life time of TAT-hcSmad8 in C3H10T1/2 cells expressing TGFβ1, GDF5 and BMP2.

Once that the successful internalization of TAT-hcSmad8 by C3H10T1/2 was determined, the life time of this molecule inside the cell was explored. For this, C3H10T1/2 cells and C3H10T1/2 cells expressing TGFβ1, GDF5 and BMP2 were incubated with 15 µg/mL of Cy3-labeled TAT-hcSmad8 for 24 h. After this initial incubation time, the medium of cells was replaced with fresh medium without Cy3-labeled TAT-hcSmad8. The internalization was analyzed after one, two and three days. This experiment was performed three times (Figure 16). The life time of Cy3-labeled TAT-hcSmad8 was determined to be two days for a concentration of 15 µg/mL. Concentrations of 5 µg/mL are no longer detectable after 24 hours. This information was important to determine how often the medium containing TAT-hcSmad8 should be exchanged in further activity test. Besides, the different cells producing different factors can be used as control for further activity test of TAT-hcSmad8 (not included here). In a previous work [8], for example, it was demonstrated that cells co-expressing cSmad8 and BMP-2 were able to self-differentiate into tendon cells. Interestingly, no other member of the TGF-β such as TGFβ1 or GDF5 was able to generate tenogenesis in combination to BMP-2.
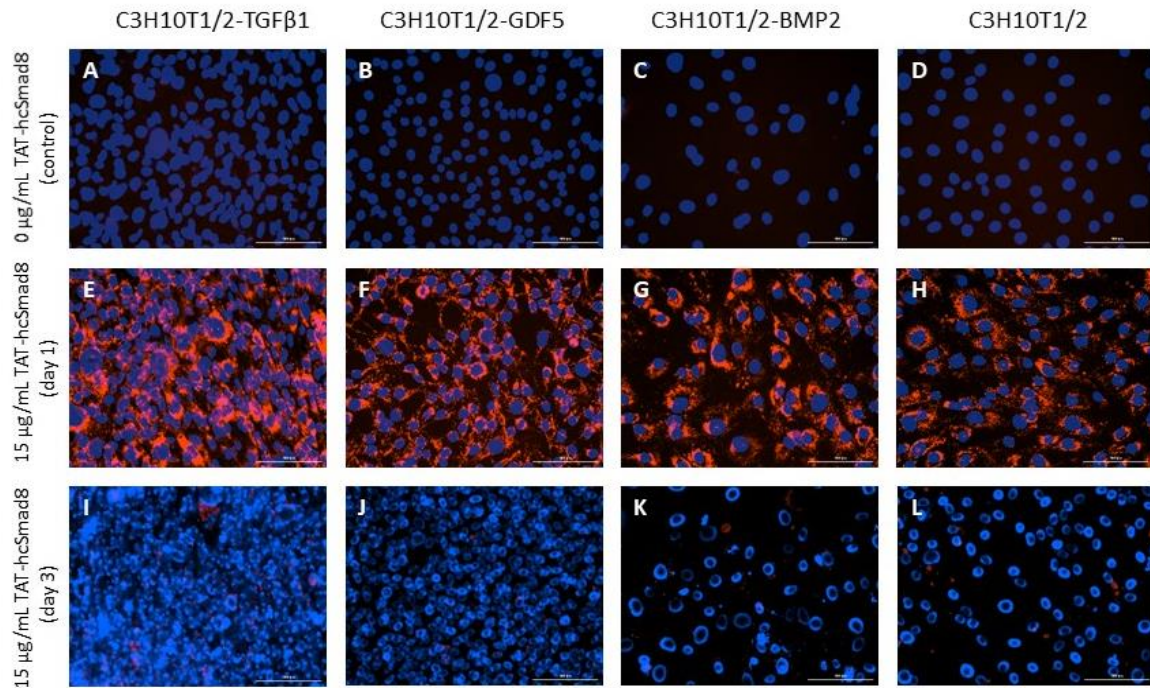
Figure 16. Life time assay of TAT-hcSmad8 in different cell types.
C3H10T1/2 cells and C3H10T1/2 cells constitutively expressing TGF β1, GDF5 and BMP2 incubated with 15 µg/mL of Cy3-labeled TAT-hcSmad8 (red). Nuclei were stained with DAPI (blue).

4.6 Activity test

A common bioassay for BMPs is the measurement of the production of alkaline phosphatase (ALP). For this, C3H10T1/2 cells were incubated during 12 days with different concentrations of TAT-hcSmad8 (0, 5, 10, 25, 50 µg/mL) and BMP-2 (0, 50, 100, 250, 500, 1000 µg/mL). After the incubation period, the cells were fixed and the ALP was visualized by staining the cells with NBT/BCIP Ready-to-Use Tablets (Figure 17). In panel A-F/ε-κ of figure 17, it can be observed that the ALP production (purple) increased proportionally to the concentration of BMP-2. ALP is a marker of osteoblast differentiation as it is produced by cells upon BMP-2 stimulation. On the other hand, Smad8 has shown to inhibit BMP signaling by a mechanism that is not completely understood [1]. In this context, a decrease in the ALP production was expected. However, colorimetric analysis did not show significant decrease of the coloration of samples treated with TAT-hcSmad8 and BMP2 (Figure 17). Additional to the colorimetric shift, the morphology of cells is also a parameter that can be used to qualitatively evaluate differentiation. The initial star-like phenotype of C3H10T1/2 cells change to elongated tenocyte-like phenotype during differentiation (Hoffmann et al., 2006). A change in the morphology of C3H10T1/2 cells treated with BMP-2/TAT-hcSmad8 was not observed (Figure 18). The lack of differentiation was also observed in C3H10T1/2 cells expressing BMP-2 (Figure 19). Undesired formation of protein

precipitates visible at concentrations of 25 and 50 µg/mL of TAT-hcSmad8 could be the reason that the cells do not go into differentiation. Since the presence of TAT-hcSmad8 inside the cells was clearly determined, it was hypothesized that loss in form of precipitation of TAT-hcSmad8 generated a stimulus that was not strong enough to be observed or quantified by histological analysis. It was also hypothesized that more sensitive and quantitative assays would be necessary to study the biological activity of TAT-hcSmad8.
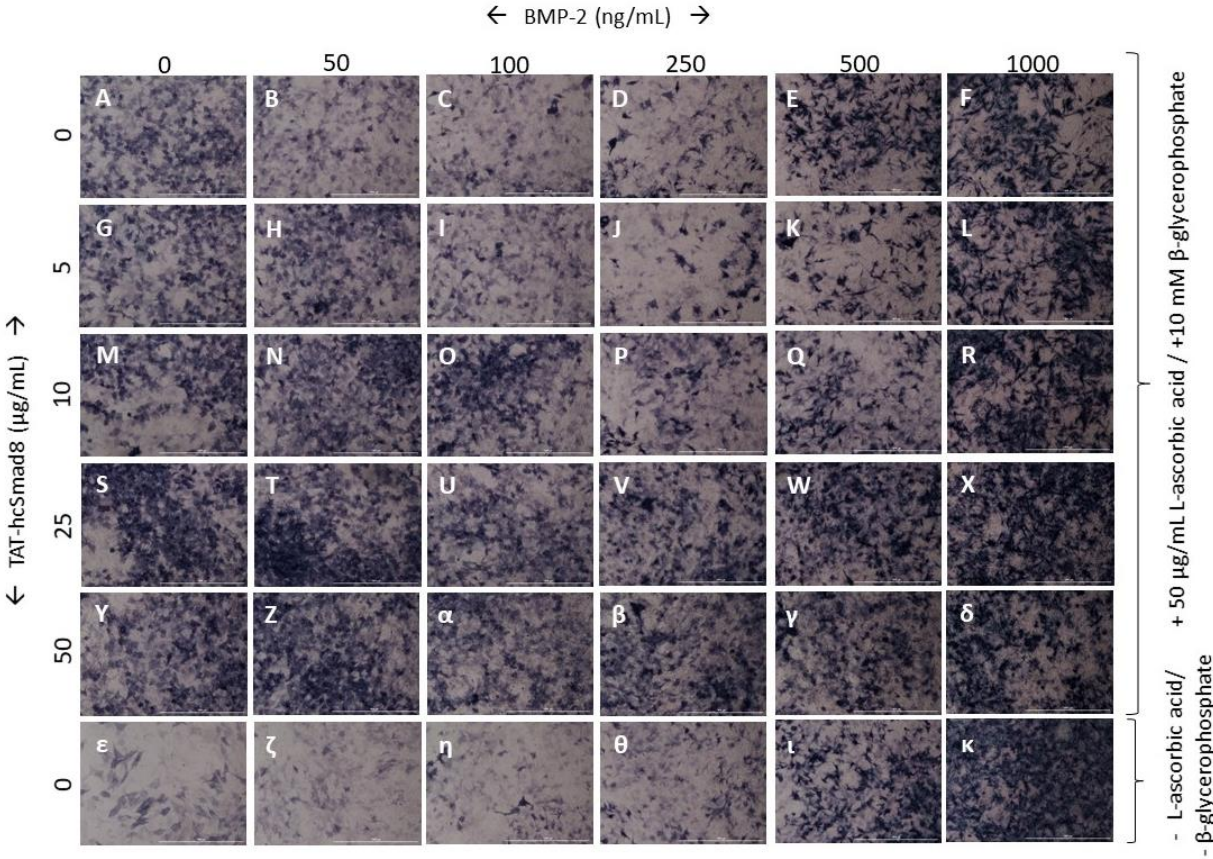
Figure 17. ALP-staining of C3H10T1/2 cells (4x magnification).
The cells were incubated with TAT-hcSmad8 for 11 days.
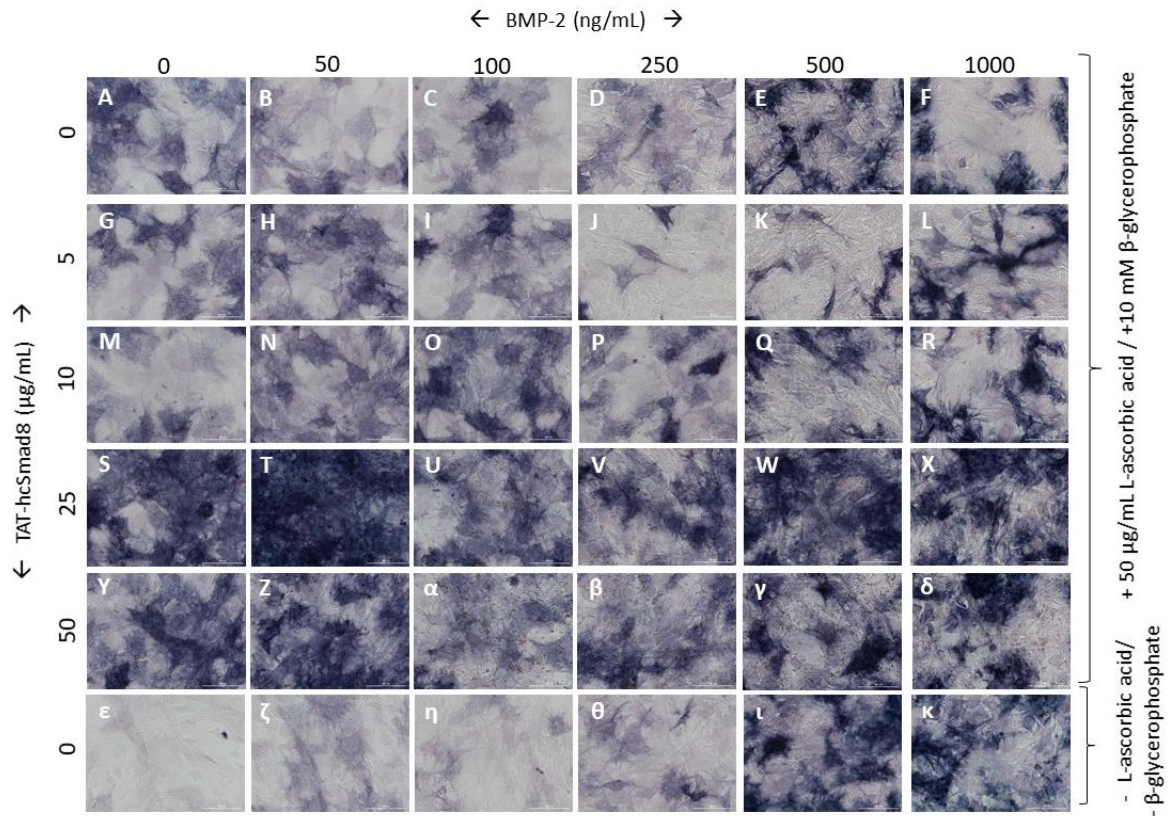
Figure 18. ALP-staining of C3H10T1/2 cells (20x magnification).
The cells were incubated with TAT-hcSmad8 for 11 days.
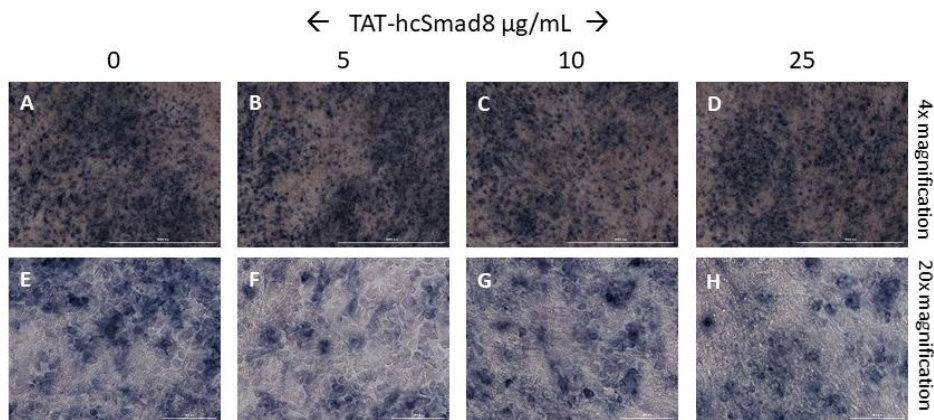
Figure 19. ALP-staining of C3H10T1/2-BMP2 cells (4x and 20 magnification).
The cells were incubated with TAT-hcSmad8 for 10 days.

To further test the activity of TAT-hcSmad8, a BRE-Luc C3H10T1/2 reporter cell-line was used. These cells were stably transfected with the firefly luciferase reporter gene under the control of a BMP-responsive element (BRE). BRE is formed by different sequence-elements from an Id1 promoter [10]. This system is widely used to test BMP-2 activity. Moreover, similar luciferase-based reporter assays

were used before to test Smad8 activity (Tsukamoto et al., 2014) due to the down-stream role of Smads in the BMPs signaling pathway. Smad8 is described in the literature to have an inhibitory effect on the BMP signaling pathway [1, 8]. It was hypothesized that the signal generated by BMP-2 in BRE-Luc C3H10T1/2 cells would be decreased by the presence of Smad8.

A standard curve using different concentrations of BMP-2 (Figure 20, panel A) revealed that a concentration of 500 µg/mL is enough to reach the maximum of RLU under the incubation-conditions used for the cells. The presence of BSA, the concentration of FCS used during the incubation of the cells or the incubation time of the cells with BMP-2 influence the intensity of the signal [10].
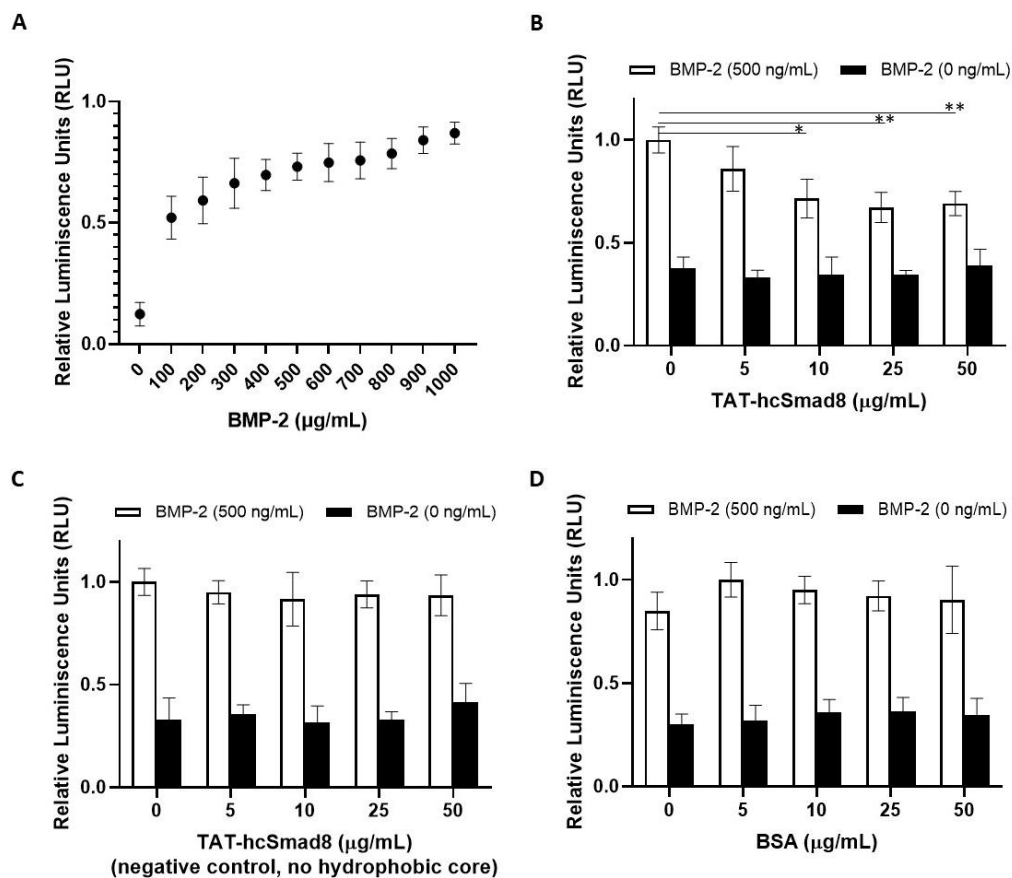
Figure 20. Reporter-based activity assays using BRE-luc C3H10T1/2 cells.
*Standard curve of BMP-2 (A). Activity test of TAT-hcSmad8 purified by Method 1 (B) and Method 2 (C). TAT-hcSmad8 purified by Method 2 and BSA (D) were taken as negative controls. The values were normalized with the highest value of each group. The data is represented as the mean ± SD. *, p<0.05; ** p<0.01.*

The inhibitory effect of TAT-hcSmad8 can already be observed in Figure 20 (panel B) at a concentration of 10 µg/mL. TAT-hcSmad8 purified by Method 2 (urea-based) was used as negative control (Figure 20, panel C) in which no hydrophobic core was observed. Additionally, a BSA was taken as a further

negative control (Figure 20, panel D). Moreover, a cytotoxicity test showed that TAT-hcSmad8 exerts no cytotoxicity effect on C3H10T1/2 cells (Figure 21).
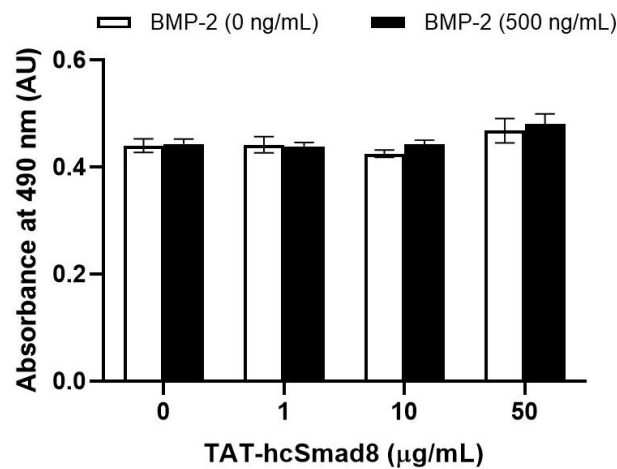

Figure 21. Cytotoxicity assay.
C3H10T1/2 cells incubated with different concentrations of TAT-hcSmad8 in presence or absence of BMP-2.

The expression level of genes used as chondrogenic, osteogenic, and tenogenic markers in murine C3H10T1/2 cells were explored. *Scleraxis* (*Scx*) is the only marker gene for tendon. Moreover, the expression of this gene seemed almost nonexistent in undifferentiated proliferating mesenchymal cells [8]. In a fist wide gene-screening, the expression level of *Bglap, Scx, Six1, Sox9, Eya-1* and *Epha4* was analyzed by RT-PCR for C3H10T1/2 cells incubated with TAT-hcSmad8 (0, 5, 10 and 25 μg/mL) in the presence of BMP-2 (0 and 500 ng/mL) for 4, 7, 10 and 12 day (Figure 33, Supplemental in Appendix). *Bglap, Sox9, Six1* and *Scx* showed to be the most promising genes for samples corresponding of day 12 (Figure 22). *Bglap, Sox9* and *Scx* showed statistical differences at concentrations of 25 μg/mL of TAT-hcSmad8. End-point analysis of *Bglap, Sox9 and Scx* showed the expected tendency (Figure 22). *Bglap (osteocalcin)* is used as osteogenic marker. Consistently with the results of luciferase-assay, an inhibition of this pathway is expected. In figure 22 (panel A) and 23 (panel A), a decrease in the expression level of *Bglap* is reduced almost by half in cells incubated with 25 μg/mL of TAT-hcSmad8 could be observed. For this concentration a notable increase of *Sox9* was observed (Figure 22, panel B; Figure 23 panel B). Similar behavior was observed for *Scx (Figure 22, panel D; Figure 23, panel D)*. It is concluded that the behavior of the genes is in accordance with the expression tendency observed during tenogenesis.
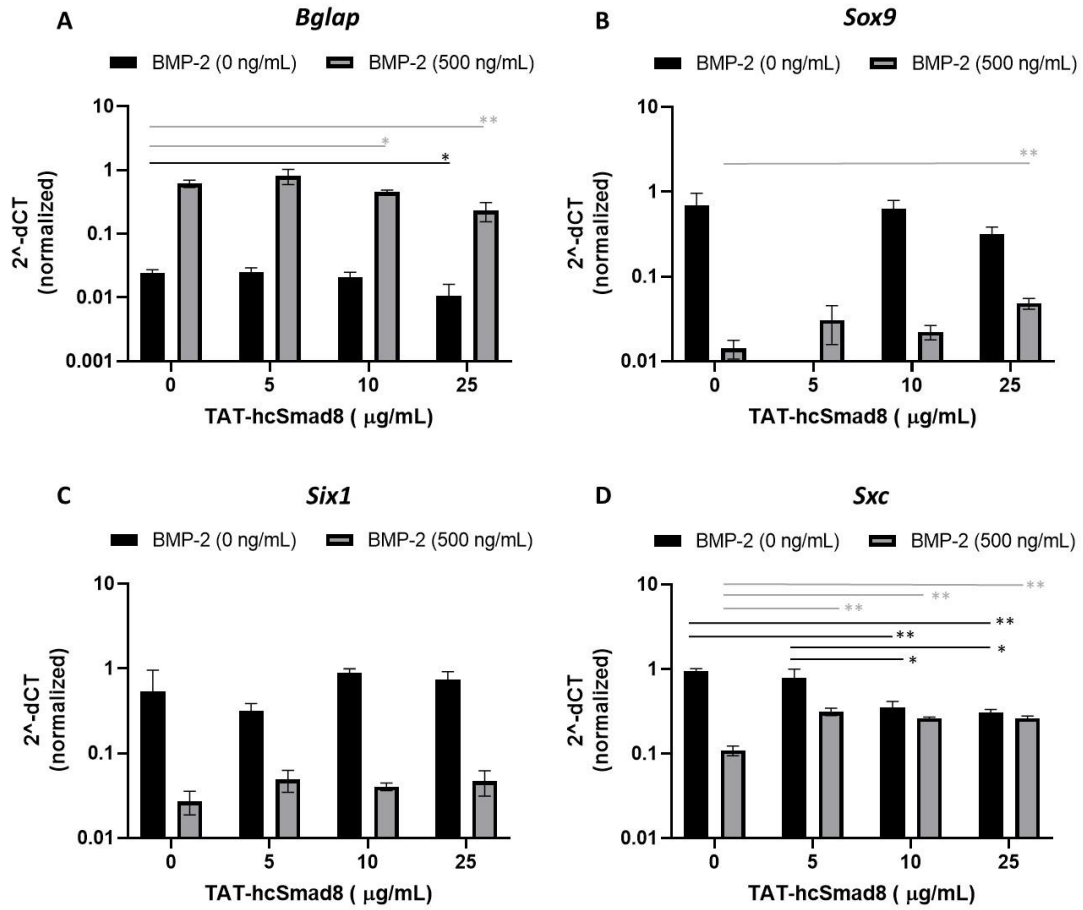
Figure 22. RT-PCR analysis for Bglap, Sox9, Six1 and Scx.
*The data is represented as the mean ± SD. *, p<0.05; ** p<0.01. Data was plotted (y axis) as $\log_{10}$ scale for a better representation due to the difference (range) of values obtained for samples treated with 500 ng/mL BMP-2 (grey bar) and 0 ng/mL BMP-2 (black bar).*
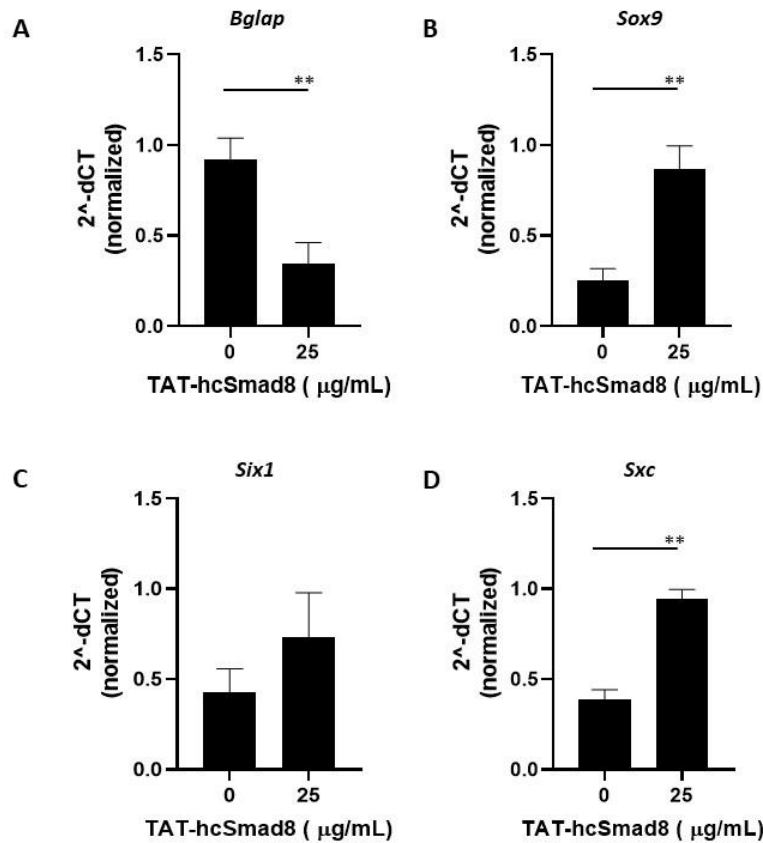
Figure 23. RT-PCR analysis for Bglap, Sox9, Six1 and Scx.
*This figure is based on figure 24. For aims of comparison, the concentration 0 and 25 µg/mL of TAT-hcSmad8 were taken (only). The data is represented as the mean ± SD. *, p<0.05; ** p<0.01.*

4.7 Inclusion bodies as self-supported vector for proteins

The possibility of using IBs as a self-supported delivery was explored herein. The methods followed for the study of solubilized TAT-hcSmad8 were extended to the study of IBs.

The production of IBs of TAT-hcSmad8 and of GST-GFP (control) was carried out in *E. coli* BL21 (DE3) and ClearColi ®. The optimization was carried out using *E. coli* BL21 (DE3) (Figure 24, panel A, B and C). The best condition found for the production of TAT-hcSmad8 was 3 h induced with 0.25 mM IPTG. The yields decreased considerably under longer incubation times (even at low temperatures). The degradation of the antibiotic (Amp) might be the reason for this as it is observed a clear production of protein after induction with IPGT that is later not observed. The production conditions were used for the production of TAT-hcSmad8 and GST-GFP in ClearColi®. In figure 24 (panel D) can be observed that the genetic modified ClearColi® presents slower growth rates.
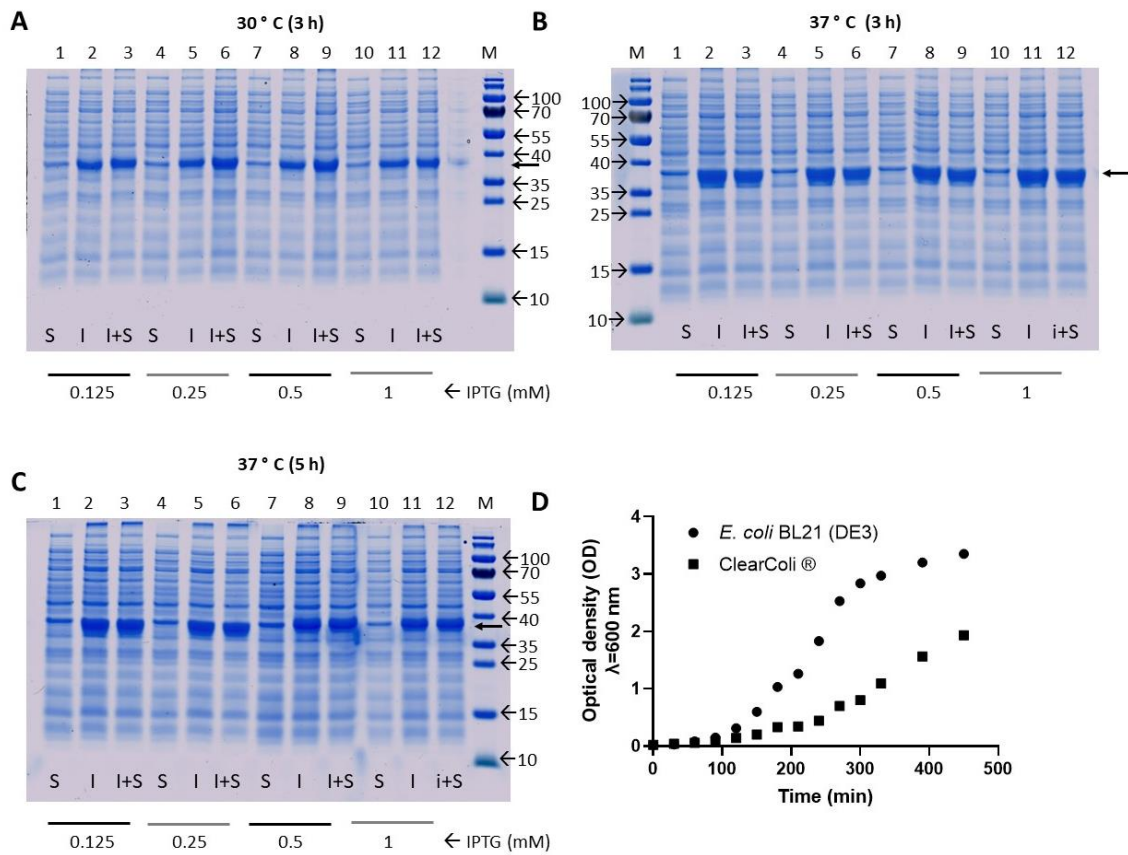
Figure 24. SDS-PAGE of lysates of E. coli BL21 (DE3) expressing TAT-hcSmad8 (indicated with a black arrow, 35.8 kDa).

Panel A corresponds to induction temperature of 30 °C for 3 h with different concentrations of IPTG (described at the bottom). Similarly it is shown in panel B the analysis of protein induction at 37 °C for 3 h and in panel C, at 37 °C during 5 h. In panel D the growing curves of E. coli BL21 (DE3) and ClearColi® are compared. S, refers to the soluble part of the cell lysate after centrifugation. I, refers to insoluble part and I+S is the complete cell lysate.

The purification of IBs was carried out according to Carmona et al. 2010 [91]. With this method, purities of 60 % were reached (Figure 25). After purification, CH310T1/2 were incubated with IBs and the cytotoxicity and internalization were studied.
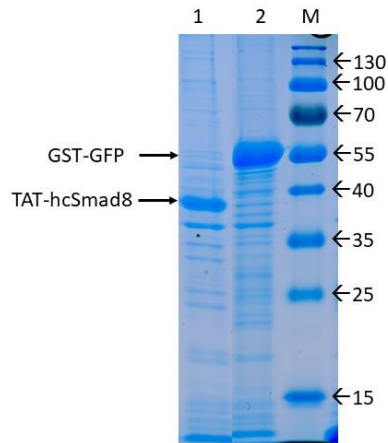
Figure 25. SDS-PAGE of purified IBs of TAT-hcSmad8 and GST-GFP.

Similar to the experimental procedure followed for refolded TAT-hcSmad8, the internalization was investigated though immunocytochemistry. In figure 26, the nuclei of cell are perfectly delimited (remarked) by the green coloration that corresponds to the antibodies used to detect TAT-hcSmad8. In the case that TAT-hcSmad8 inclusion bodies were superficially embedded or interacting with the membrane, a rather homogeneous layer on the cell would be observed. In this case, however, the space of the nucleus is clearly marked (Figure 26, green channel).
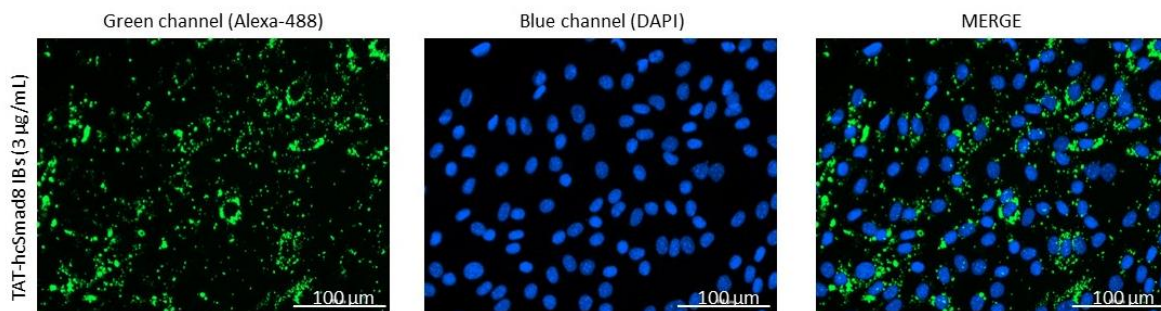

Figure 26. Immunocytochemistry of C3H10T1/2 cells incubated with TAT-hcSmad8 IBs purified.
Green fluorescence is derived from donkey anti-goat Alexa 488 antibody coupled to goat anti-Smad8 antibody. *Counterstaining of cell nuclei was performed with DAPI.*

After confirming the internalization of the IBs, histological analysis was performed. For this, C3H10T1/2 expressing BMP-2 were incubated with 0, 5, 10, 25 and 50 µg/mL of TAT-hcSmad8 IBs. As negative control, GST-GFP IBs were used. ALP staining was carried out for days 4, 8, 11 and 15. In figure 27, the ALP-staining is shown for day 15. No elongated shape related to tenocytes were found with any condition tested. A decrease in ALP production was observed for samples treated with TAT-hcSmad8 and GST-GFP IBs indicating a disturbance of the BMP-2 dependent pathway caused intrinsically by IBs. Though cytotoxicity assays were carried out (Figure 28), it was determined that IBs exert a cytotoxic effect on C3H10T1/2. Moreover, since it was observed that GST-GFP (used as a negative control)

produced a decrease in ALP production it was concluded that the negative effect in the ALP production is rather a nonspecific effect of IBs. Together with the cytotoxicity assays, it was concluded that IBs possess a cytotoxic effect on cells. Lysis and cleaning protocols described by Stamm et al. (2018) [90] or by BugBuster® manufacturer did not show to eliminate the cytotoxicity effect of IBs on C3H10T1/2.

Although it was not investigated here the reason of the cytotoxic effect of IBs, it can be hypothesized that cells can be sensitive to membrane disruption by big particles. Another reason, could be the sensitivity of the cells to the presence of detergents (even at negligible amount), especially during differentiation. Extensive rinsing of IBs with PBS did not showed any improvement.
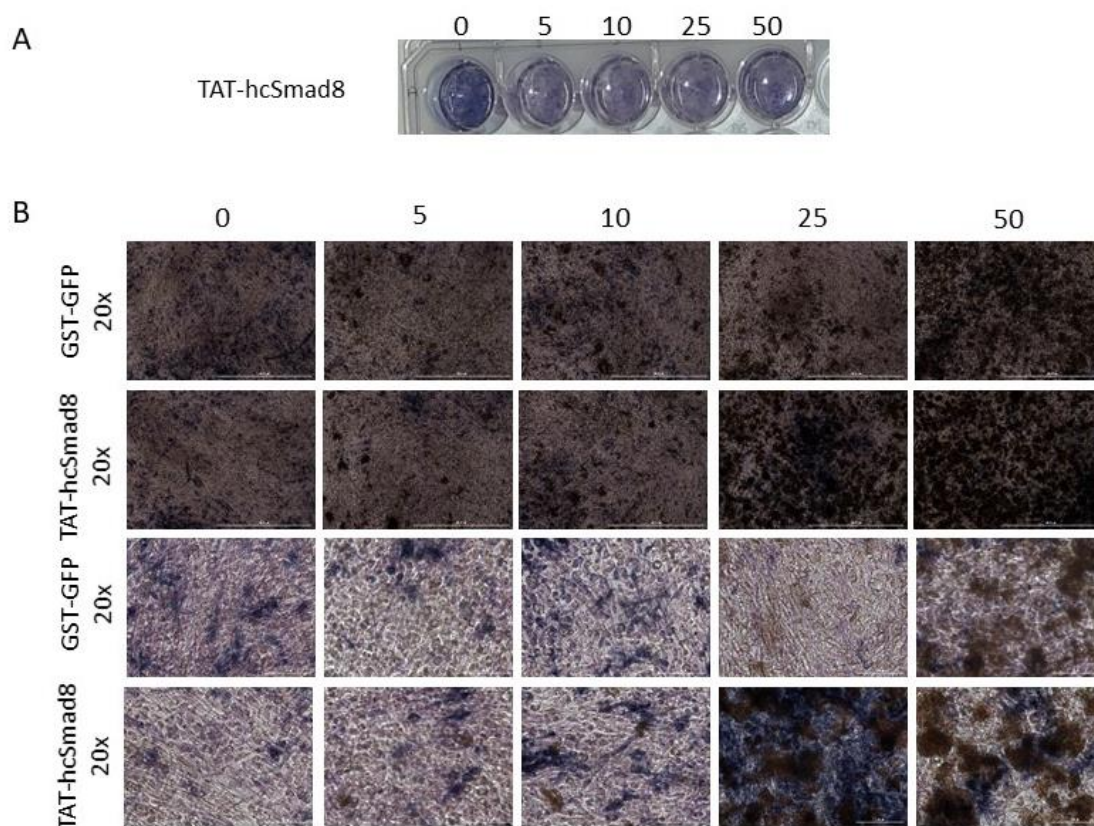

Figure 27. ALP-staining of C3H10T1/2-BMP2 cells incubated with TAT-hcSmad8 IBs and GST-GFP for 15 days.
The cytotoxicity effect due to the present of toxins or impurities proper of *E. coli* was discarded by using endotoxin-free ClearColi®.
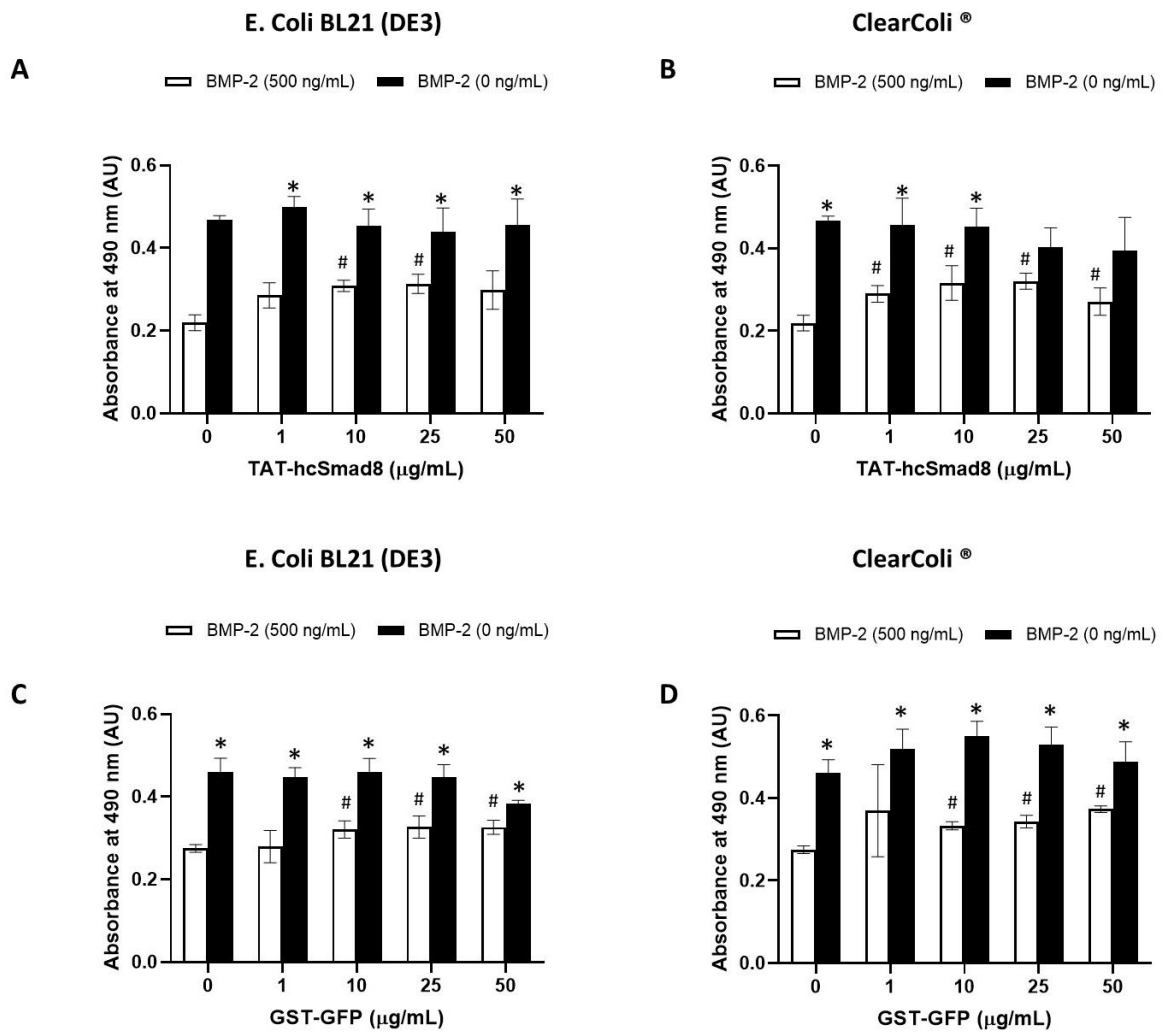
Figure 28. Cytotoxicity assay of C3H10T1/2 cells incubated with different concentrations of TAT-hcSmad8 IBs and GST-GFP in presence or absence of BMP-2.

Significant differences of $p < 0.05$ between pairs are marked: * indicate the difference between the incubation conditions 500 ng/mL BMP-2 and 0 ng/mL BMP-2 of the same IB concentration; # indicate statistical difference between 0 µg/mL of IBs and 1, 10, 25 or 50 µg/mL IBs in presence or absence of BMP-2.

5. Discussion

Production and purification of TAT-hcSmad8

Herein, the possibility of using TAT-hcSmad8 refolded from IBs for biotechnological applications was explored. Two strategies in particular were described (Figure 29). One strategy consisted in the use of purified IBs as a self-supported delivery system. For this strategy, the IBs can be seen as self-organized porous materials that holds mainly the overexpressed protein. The desired protein can display different protein-folding including the native-like folding. The other strategy here described was based in the traditional method of solubilizing IBs using chaotropic agents in order to later refold the protein *in vitro* and to finally purify it. It is also common to carry out the purification of protein using the same conditions used for the solubilization (denaturing/reducing conditions). The pure protein can be then refolded *in vitro* [102, 103].
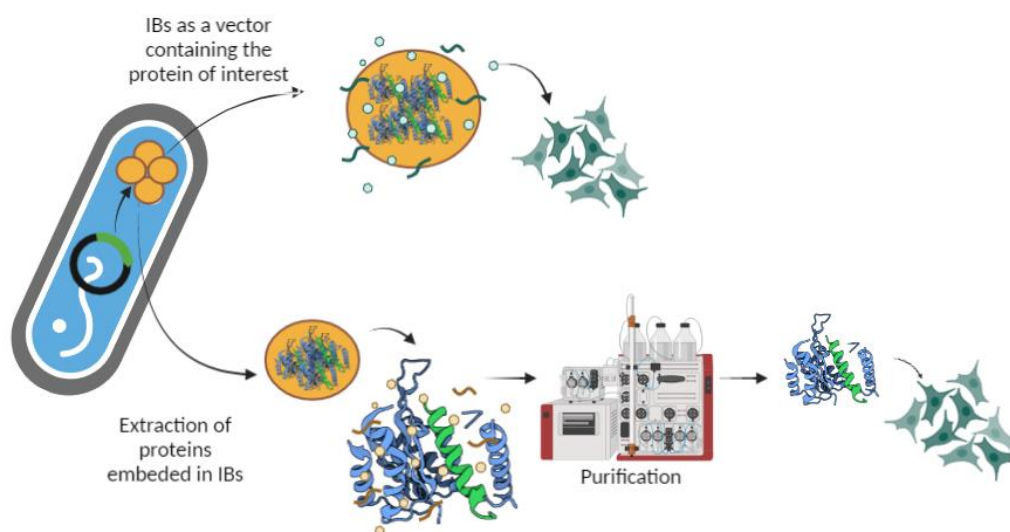

Figure 29. Representation of IBs application. IBs are represented as orange circles.
This Figure was created with BioRender®.

Extensive screenings of conditions were performed to increase the solubility of TAT-hcSmad8. The exploration of different temperatures, concentration of inducer agent an induction time was included in the screening. Additionally, different strains such as *E. coli* Origami2® (DE3)pLysS (Novagen, Germany), Rosetta®(DE3), BL21 (DE3), ClearColi®were also tested. By using pBAD/His B and TOP10 *E. coli* tighter and slower protein production was explored as well.

A common strategy to promote solubility and folding of proteins is the genetic addition of fusion proteins to the working protein. In this work, the GST tag was explored. However, the tendency of TAT-

hcSmad8 to form IBs when expressed in *E. coli* is remarkably high in all this conditions. The propensity of a protein with a eukaryotic origin to form IBs is not surprising. The rates of expression of eukaryotic proteins in *E. coli* are tend to be low, most post-translational modifications are absent and the product is often in form of IBs [104]. In fact, up to 70 % of eukaryotic proteins expressed in *E. coli* are insoluble [105]. However, *E. coli* is still the expression system of choice due to its versatility. In this context, production of TAT-hcSmad8 IBs in *E. coli* showed to be fast, economic and with high yields. After 3 hours of induction of cells with IPTG, the desired TAT-hcSmad8 reached a 16 % of the total amount of protein inside the cells. After cell lysis and IBs washing, TAT-hcSmad8 represented around 30 % and after purification, 96 % purity was reached (Table 3).

Table 3. Brief purification summary.
Taken from [2].

Step	Total protein [a] (mg)	TAT-hcSmad8 (mg)	Yield [c]	Purity (%)
Crude lysate [b]	76.0	12.2	100	16.0
IBs washing and solubilization.	25.0	7.2	59	28.8
Heparin purification	2.2	2.1	17	96.0

Approximately 3.5 mg of TAT-hcSmad8 (total, purified) per gram of cells can be obtained.
[a] Protein concentration was determined by Bradford assay using BSA as standard.
[b] Obtained from ~0.6 g of wet weight of *E. coli* cell pellet. After cell lysis and cell pellet washing, ~0.15 g of IBs were obtained.
[c] Referred to the total amount of TAT-hcSmad8 in g/mL.

The purification of TAT-hcSmad8 by affinity chromatography (Heparin) showed to be the best purification method. Moreover, the presence of a His6-tag permits the purification of TAT-hcSmad8 by IMAC (HisTrap® HP) under denaturing conditions using the solubilization buffer which includes 6 M GuHCl as equilibration buffer (Figure 34, Supplemental in Appendix). The purification of TAT-hcSmad in a refolded state using IMAC showed poor yields. Since Heparin chromatography led to a high degree of purity and because it is a simple and economic method, this method was preferred over IMAC.

Structure prediction and CD analysis revealed that the addition of the His6-tag did not modify TAT-hcSmad8 pledge. The disruption of the general structure or pledge of TAT-cSmad8 was improbable as His6-tag and TAT sequence were attached to the N-termini where the linker region is displayed. The linker region lacks secondary structure. The attachment of the His6-tag/TAT-sequence at the C-termini of TAT-hcSmad8 was discarded as the phosphorylation site SSXV is found at the C-termini. This motif is important for the formation of the trimers between R-Smads and Smad4. This explains why this

motive is highly conserved in all R-Smads. In fact, routine sequencing have revealed that mutations affecting the trimer interface are present in cancer samples [5]. Besides, the central part of Smads, formed by sets of α-helixes and β-sheets, follows the SSXS motif. In conclusion, the modification of the C-termini could led to the disruption of these secondary structures.

The presence of the His6-tag slightly modified the physicochemical properties of TAT-hcSmad8 when compared to TAT-cSmad8 (without His6-tag). When comparing the conductivity values at which the proteins elute, it is noticeable that TAT-hcSmad8 (~80 mS/cm) binds more firmly to the Heparin column than TAT-cSmad8 (~60 mS/cm) probably because of the additional interaction of positively charged groups of the histidine residues ($pK_{aside\ chain}$ = 6.0 and $pK_{amino\ group}$=9.17) with the negatively charged Heparin of the column [106]. TAT-hcSmad8 was also more resistant to precipitation during labeling with Cy3. Moreover, it presented higher yields during purification and during Cy3-labeling. Since TAT-hcSmad8 was easier to handle than TAT-cSmad8, most of the experiments were performed with TAT-hcSmad8. An additional advantage is that the His-tag offers an extra detection method (immunochemistry). The presence of a TEV restriction site between the His-Tag and TAT-cSmad8 permits the elimination of the His-tag for future medical applications.

Physicochemical characterization of TAT-hcSmad8

The purified samples of TAT-cSmad8/TAT-hcSmad8 were analyzed by TSA (Figure 9). Although with this assay is not possible to verify the correct folding of the protein, it can give insights about the presence of hydrophobic core, protein-protein interactions or ligand interactions [107]. TSA showed the presence of a hydrophobic core in samples of TAT-hcSmad8/TAT-cSmad8 purified by Method 1. In samples purified by Method 2 (urea-based Method) the distinctive peak indicating the presence of a hydrophobic core is not obtained. Since most of natively folded proteins possess a hydrophobic core, TSA is a versatile tool to perform preliminary folding studies which give insights about the folding state of proteins [105, 108].

To get more insights about the secondary structure, purified TAT-cSmad8/TAT-hcSmad8 was analyzed by CD (Figure 10, panel A). CD spectra showed the presence of a mixture of α-helixes and β-sheets. In this regard, further optimization of testing conditions could help to get a stronger signal. With a higher concentration of protein, a more stylized profile can be obtained. The analysis of the stronger and clearer signal with an appropriate software could give more information about the exact percentage of α-helixes and β-sheets [109]. However, it can also lead to distorted signal when impurities are present. For the case of TAT-hcSmad8, higher concentrations of protein led to aggregation.

In this work, a theoretical model was obtained according to the crystallographic structures solved for Smads using the software PONDR®. Other predictor softwares were unable to find a suitable structure due to the large loop present in TAT-hcSmad8. To solve this, the region comprising the core structure can be modeled separately from the loop with standard software. With PONR®, however, it is possible to model the protein introducing the whole sequence (including the loop). The model observed in the Figure 10 (Panel B) obtained by homology is thought to be an accurately predicted model due to the extremely high degree of conservation of R-Smads. Across subfamilies, the MH2 domain present amino acid sequences identities ranging from 32 % to 97 %. The similarity of mouse Smad8 with Smad1 and Smad5 is, for example, 92 % and 90 %, respectively [64]. The level of conservation is especially high in critical amino acids which are vital for protein-protein interaction. Particular mutation that cause premature truncations that the C-termini of Smad2 were observed in breast, kidney, lung, and pancreatic carcinomas. The truncated C-termini of Smad2 prevents receptor-mediated activation [5]. As it can be seen, the high level of Smad8 conservation particular across species might indicate that Smads and the TGF-β signaling emerged with metazoans and remained intact due to the critical role of this proteins and this signal [5].

Currently there is no crystallographic structure solved for the complete length of Smads (including the linker region). The high mobility of the disordered region (linker) represents a challenge for crystallization by traditional methods. However, crystallographic structures of individual domains of human Smads have been already reported (Table 4). MH1 and MH2 domains of different Smads that were produced in *E. coli* have been widely used for structural analysis (Table 4).

Table 4. Summary of solved structures of Smads.

Smad	Domain	PDB	Crystalized co-molecule	Expression system	Year	Reference
Smad1	MH2	1KHU	-	E. coli	2001	[49]
	MH2	5ZOK	MAN1	E. coli	2018	[46]
Smad2	MH2	1DEV	Smad binding domain of SARA	E. coli	2000	[45]
	MH2	1U7V	Smad4	E. coli	2004	[56]
	MH2	1KHX	-	E. coli	2018	[46]
	MH2	5XOD	Ski	E. coli	2018	[46]
Smad3	MH1	1MHD	DNA	E. coli	1998	[37]
	MH2	1MK2	SDB	E. coli	2002	[48]
	MH2	1MJS	-	E. coli	2002	[48]
	MH1	1OZJ	DNA	E. coli	2003	[36]
Smad4	MH1	5NM9	GGCGC DNA site	E. coli	2017	[38]
Smad5	MH1	6TBZ	GGCGC DNA site	E. coli	2021	[40]
Smad8	MH1	6TBZ	GGCGC DNA site	E. coli	2021	[40]

Ruiz et al. (2020) [40] solved the structure of the MH1 domain of Smad5 and Smad8 bound to DNA by X-ray crystallography. It was concluded that Smad5 and Smad8 MH1 domains are nearly identical and very similar to the MH1 domain of Smad1 [40]. This feature can be observed in Figure 30 (Panel A). A structural comparison of MH1 and MH2 was performed (panel B). Even though different Smads are compared, the secondary structure is highly conserved. This indicates that the theoretical model generated for the MH2 of Smad8 by homology with PONDR ® software is highly precise.
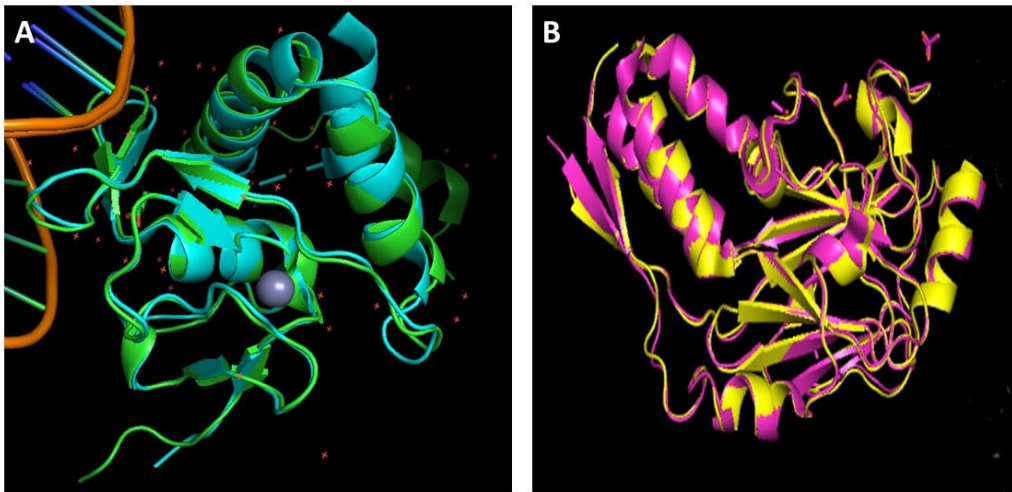

Figure 30. Structural comparison of different Smads.
In panel A, MH1 domain of Smad5 (cyan color, PDB 5x6h) and Smad8 (green color, PDB 6fzt) are overlapped. Both domains are bound to GGCGC sites. The grey circle represents the coordinated Zn^{2+}. In panel B, the MH2 domain of Smad1 (yellow color, PDB 1khu) and Smad2 (pink color, PDB 1mk2). Superposition was done with the software PyMOL®.

Additionally to the structural comparisons of Smad8, its structure was analyzed with the free software "Disulfide by Design®" [110]. A structural model is provided to the software which rapidly pairs all residues. According to the proximity and geometry of the cysteine amino acids, disulfide bonds can be identified. With this analysis, no disulfide bonds were predicted for TAT-hcSmad8. This is consisting with the literature as no disulfide bond has been reported for any of the MH2 domains of Smad8 reported to date.

Once that it was confirmed that the purified TAT-hcSmad8 displayed secondary structure congruent with the theoretical model, the stability of the protein was explored. A simple method for measuring protein aggregation is UV/ VIS spectroscopy. A wavelength between 320 nm and 400 nm can be used to measure turbidity. When protein forms aggregates, turbidity increases. Due to its simplicity, this method is particularly suitable for high throughput processes in which, for example, the protein concentration remains constant and the influence of other parameters such as pH, buffer or

temperature on protein aggregation is to be investigated. Herein, a wavelength of 340 nm was used to measure turbidity.

A pH ranging from pH 4.0 to pH 10.0 (Figure 11) was explored. Sodium acetate was used for stablishing pH 4.0, MOPS for pH 7.0 and sodium borate for pH 8.0 to 10.0 (Panel A). Alternatively, carbonate-bicarbonate buffer was used for pH 9.0 to pH 10.0. All these conditions were also tested in presence of 250 mM NaCl and 20 % glycerol.

In general it can be observed that TAT-hcSmad8 is stable from pH 4.0-7.0. The solubility decreases dramatically at neutral (pH 7.0) and basic (8.0-10.0) pH. The increase of turbidity decreases at pH 9.0 to 10.0 when carbonate-bicarbonate buffer is used. This effect is not seen for sodium borate buffer. The presence of 250 mM of NaCl increases aggregation, which is visible even at pH 6.0. The presence of 20 % glycerol counteract the effects of the basic pH and NaCl as the samples showed low turbidity. Glycerol is a cosolvent that is known to inhibit aggregation while enhancing the stability of proteins [111]. Vagenende et al. (2009) [111] propose that glycerol interacts with hydrophobic patches of proteins acting as an amphiphilic interface between the polar solvent and the hydrophobic surface [111]. In this scenario, glycerol would inhibit aggregation by inhibiting protein-unfolding by stabilizing aggregation-prone intermediates [111].

Once that protein was physically and chemically characterized, the biological properties of the protein were tested. For this aim, TAT-hcSmad8 produced in ClearColi® was used. ClearColi® was genetically modified to display a precursor of lipid IV_A that does not trigger an endotoxin response [92] (Figure 31).
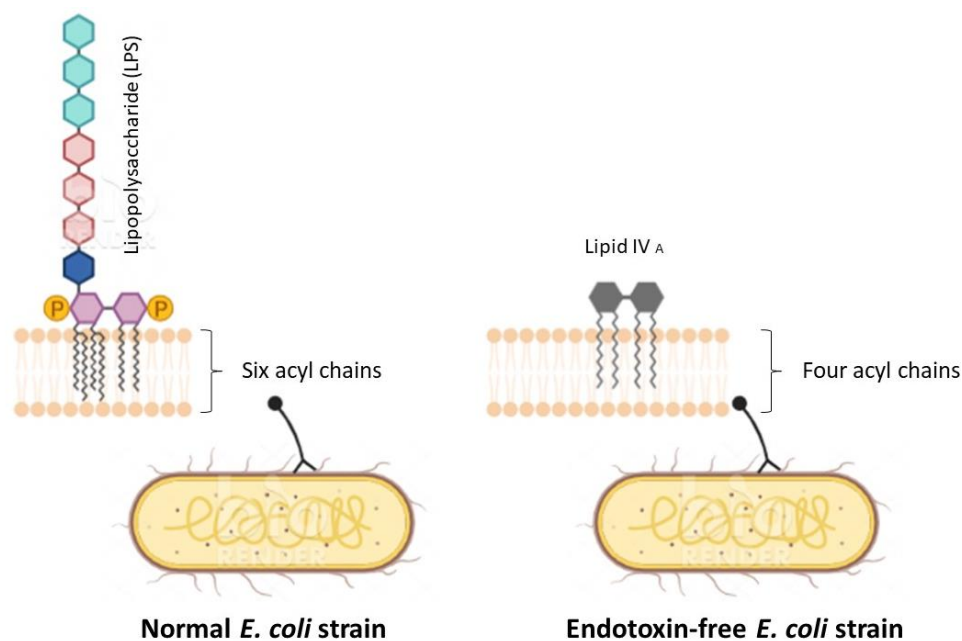

Figure 31. Comparison of the normal E. coli strain and the endotoxin-free E. coli strain.

Normal LPS activates immune system. Specifically, the six acyl chains of LPS are recognized by the human toll-like receptor 4 (hTLR4) and myeloid differentiation factor-2 (MD-2) complex, leading to the activation of NF-κB and cytokines. The endotoxin precursor lipid IV_A (four acyl chains) does not trigger endotoxin response. Figure made with BioRender®

Internalization and biological test

The first approximation to determine the internalization of TAT-hcSmad8 was performed by flow cytometry. Since TAT-hcSmad8 was labeled with Cy3, it was possible to quantitatively compare the fluorescence of C3H10T1/2 cells incubated with Cy3-labeled TAT-hcSmad8 and without it. Increasing concentration of TAT-hcSmad8 (1 µg/mL-100 µg/mL) leads to increasing relative fluorescence units (RFU). Increasing incubation times (0.5 h- 2.5 h) lead to higher RFU in a non-proportional ratio. As negative control, C3H10T1/2 cells incubated in absence of Cy3-labeled TAT-hcSmad8 and C3H10T1/2 cells incubated with Cy3-labeled BMP-2 were taken. BMP-2 is known to interact with type II receptors. In this regard, the signal produced by cells incubated with Cy3-labeled TAT-hcSmad8 can be compared to the signal generated by the cells incubated with Cy3-labeled BMP-2. As expected, the signal produced by the first cells are notably higher than the second ones.

The fact that important RFU is detected after short incubation times such as 0.5 h, indicates high affinity of the TAT-hcSmad8 for the membrane. In this case, it was used the Trans-Activator of Transcription (TAT) of the HIV. TAT is a cationic cell penetrating peptide (CPP) (GRKKRRQRRRPPQ) that has been widely used to achieve internalization of cargos [112]. The arginine-rich content of TAT allows electrostatic interaction with negatively charged cellular membranes, which would explain the high avidity showed by TAT-hcSmad8. In a study performed by Patel et al. 2019 [9], it was observed that C3H10T1/2 cells showed the highest level of uptake of cationic peptides (Penetratin, R8 and TAT) when compared to HeLa, HEK and HepG2 [9]. Interestingly, this study also showed that cyclisation of CPP significantly improves internalization (compared to the lineal CPPs). In further works, the cycled-TAT could be tested for TAT-hcSmad8 to optimize the internalization process.

Flow cytometry measurements provide information about the degree of interaction between TAT-hcSmad8 and the C3H10T1/2 cells. This includes, protein interacting transiently with the membrane, partially embedded protein and completely internalized protein. To distinguish all this possible states of TAT-hcSmad8, the cells were analyzed by confocal analysis. The confocal analysis of C3H10T1/2 cells incubated with Cy3-labeled TAT-hcSmad8 confirmed its presence in the intracellular space. Most of the protein was localized in the cytoplasm and to a lower extent, in the nucleus.

Thermal shift assays revealed a decrease of the Tm of Cy3-labeled TAT-hcSmad8. The labeling process might disturb the secondary and tertiary structure of labeled proteins as a result of the formation of the covalent bond between the Cy3 and the amino-groups of the protein. This can cause an exposition/over exposition of the TAT sequence. To discard that TAT-hcSmad8 is able to internalize itself into cells due to an exposure/overexposure of TAT generated by the labeling process, further internalization through immunocytochemistry with non-labeled TAT-hcSmad8 was performed. For this, C3H10T1/2 cells were incubated with TAT-hcSmad8. A primary antibody would be used to detect TAT-hcSmad8. Then, a green fluorescently-labeled antibody (secondary antibody) directed against the primary antibody would allow the visualization of TAT-hcSmad8 (Figure 14 and Figure 15). Cells were analyzed by fluorescence microscopy and by confocal microscopy. With this methods a punctate signal was observed, which is an indicative of endocytosis [9]. This dots can specially be observed in the intracellular space near the membrane of the cells. For some cells, TAT-hcSmad8 border the inner part of the membrane. Direct penetration and endocytosis pathway are the common internalization ways of cell penetrating peptides such as TAT [113]. Direct penetration might fall under the detection limits of microscopy, which would explain that endosomes containing several molecules of TAT-hcSmad8 are more easily detected. The results obtained by immunocytochemistry using the antibodies directed against TAT-hcSmad8 were congruent with the results obtained with the confocal analysis using Cy3-TAT-hcSmad8.

Once that the internalization of TAT-hcSmad8 was confirmed by different methods, its life-time was determined. For this, Cy3-labeled TAT-hcSmad8 was used due its easy detection. However, this detection strongly depends on the initial concentration of Cy3-TAT-hcSmad8 used to incubate cells. C3H10T1/2 cells expressing TGF-β1, GDF5 and BMP-2 were incubated with 5 and 15 µg/mL of Cy3-TAT-hcSmad8. Concentrations of 5 µg/mL are no longer detectable after 24 hours in cells expressing TGF-β1, GDF5 and BMP-2. An explanation for this, would be the higher metabolism generated by the expressed factors or by the different cell metabolism when compared with the standard C3H10T1/2 cells. On the other hand, the life time of Cy3-TAT-hcSmad8 was determined to be two days for a concentration of 15 µg/mL (Figure 16). This information was useful to set the best conditions to test the activity of TAT-hcSmad8 in cells. Due to the high amount of protein needed to perform the test on C3H10T1/2 cells, the optimization of the time at which the protein had to be added to medium was desirable.

A vital point after protein purification is to test the biological function. For fluorescent proteins, the fluorescence is normally measured and compared to a control. For enzymes, it is common to follow

the progression of the reaction measuring the generation of products or the conversion of the substrates. *p*-nitrophenyl phosphate, for example, is commonly used to measure phosphatase activity [114]. The *p*-nitrophenyl that is produced due to the enzymatic activity can be easily estimated by measuring the absorbance at 480 nm. In all cases, it is desirable to determine the retention of the fluorescence/activity by precise, simple and fast methods. For factors, the determination of the biological activity involves more sophisticated detection methods such as the use of reporter genes. For factors involved in cell differentiation, marker genes can also be used. Marker genes are characteristic genes that are indicative of certain states in the cell. For example, alkaline phosphatase (ALP) has been associated to matrix maturation. Osteocalcin has been associated for the matrix mineralization during osteogenesis. In this process, three distinct phases can be distinguished: a proliferation phase is followed by a matrix maturation phase and finally the extracellular matrix becomes mineralized [73]. Calcium deposits can be visualized by staining with alizarin red. Calcium cations are chelated by the oxygen and hydroxyl groups of the alizarin red. This process generates a distinctive orange-red precipitates. However, other metals such as magnesium, manganese, iron, barium and strontium can be also chelated by alizarin red [115]. Alternatively, a variety of kits are now available to measure calcium by colorimetry. Since calcium is rather secreted in late states of bone formation, this method is not suitable for the study of TAT-hcSmad8. Smad8 is rather meant to prevent this late state of differentiation (bone formation) and to direct it to cartilage/tendon differentiation.

On the other hand, ALP is an early osteogenic marker [74] that can be easily detected using BCIP/NBT solution. The combination of NBT (nitro-blue tetrazolium) and BCIP 5-bromo-4-chloro-3'-indolyphosphate p-toluidine salt) produces an intense, insoluble black-purple precipitate when it reacts with ALP. This method was used by Hoffman et al. (2006) [8] to characterize Smad8L-MH2 (Smad8 composed by the linker and the MH2 domain). In this publication, it is reported an inhibition of the BMP-2 signaling pathway by Smad8L-MH2. This inhibition was experimentally detected as a decrease in the ALP production (decrease of the blue color).

Herein, C3H10T1/2 cells were incubated in presence of TAT-hcSmad8 and BMP2. Tenogenesis was additionally promoted by adding 10 mM β-glycerophosphate and 50 µg/mL ascorbic acid as specified by Owen et al. (1990) [98] and as used previously by Hoffmann et al. (2006) [8]. A screening of conditions including different concentrations of TAT-hcSmad8 and BMP-2 (Figure 17) was carried out. Additionally, C3H10T1/2 cells expressing BMP2 were explored (Figure 19). Smad8L-MH2 was observed to inhibit the ALP production which is generated by BMP-2 [8] by a mechanism not determined so far for this variant. In this publication, the C3H10T1/2 cells expressing Smad8L-MH2 and BMP-2 suffered a decrease in the ALP production. The color of cells stained with BCIP/NBT solution was visibly reduced.

However, the cells analyzed herein did not show such decrease. Colorimetric analysis did not show statistically significant difference in the staining of cells treated with and without TAT-hcSmad8. The reason for this could be that TAT-hcSmad8 precipitates at these conditions (Figure 32). In Figure 32 it can be observed brown dots that are also observed in cells incubated with IBs (Figure 27). The fact that aggregates proportionally increase with the concentration of TAT-hcSmad8 indicates that the precipitates are formed from TAT-hcSmad8. With all this information, it was concluded that this kind of test was not sensible enough to detect any activity of TAT-hcSmad8. For this reason, further tests were performed.
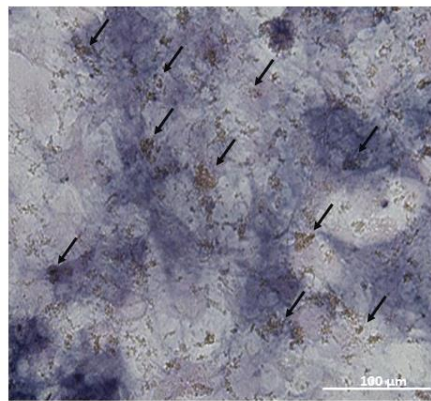
Figure 32. ALP staining of C3H10T1/2 cell incubated with TAT-hcSmad8 (50 µg/mL) and BMP-2 (100 ng/mL).
The black arrows indicate protein precipitates.

To further investigate TAT-hcSmad8 activity, C3H10T1/2 BRE-Luc cells were used. This cell line was developed and reported by [10]. C3H10T1/2 cell line was stably transfected with an expression construct (BRE-Luc) containing the firefly luciferase reporter gene under the control of a BMP-responsive element (BRE). The BRE was built by the multimerization of different sequences from the Id1 promoter. The *Id1, Id2 and Id3* genes are early response genes in BMP signaling comprised by conserved GGCGCC sequences. When the cells are stimulated with BMPs (BMP-2, BMP-4 or BMP-7) the luciferase is produced proportionally to the stimulus. The luciferase can be later quantified by measuring the luminescence generated when the luciferin (substrate of luciferase) is added. This assay is a fast, highly sensitive assay for quantification of bioactive BMP. It was hypothesized that the inhibition of BMP-2 pathway by TAT-hcSmad8 would generate a decrease in the luminescence. As it can be seen in Figure 20, the expected reduction of the signal was observed in cells treated with TAT-hcSmad8. The cytotoxicity assays indicated that TAT-hcSmad8 does not possess a cytotoxic effect on cells (Figure 21).

TAT-hcSmad8 (purified by Method 2) and BSA were taken as a control. The BSA would act as a control for protein crowding outside cells. Inactive TAT-hcSmad8 would exert the control for protein internalization and protein crowding inside the cells. Protein crowding and penetration of protein are conditions that can damage the cell. With these controls, it was concluded that the decreasing signal observed for active TAT-hcSmad8 (purified by Method 1) is not due to cytotoxicity, crowding of proteins or disturbance of the membrane.

The decrease of the signal in this kind of reporter cell lines are in accordance to literature [1].

To know more about the processes occurring in cells, the gene regulation was studied by RT-PCR. An extensive screening of genes (*Bglap, Scx, Six1, Sox9 and Eya-1*), incubation times (4, 7, 10 and 12 days) and concentrations of TAT-hcSmad8 (0, 5, 10 and 25 µg/mL) was performed. Unfortunately, for some conditions the signal was too low. This made the study of such conditions difficult. The reason for this could be a combination of factors such as degradations of the protein in the lysosomes and the precipitation of the protein that was observed (Figure 32). It was hypothesized that due to the loss of protein, a slight stimulation generated by TAT-hcSmad8 was in fact exerted on the cells. The most promising genes (*Bglap, Scx, Six1* and *Sox9)* for day 12 were further investigated.

Cell fates are complex states of the cells where multiple processes are simultaneously happening. In particular, there is not specific markers for tendon formation. Therefore, this process should be rather observed as a combination of genes. In other words, a group of genes are necessary to determine the fate of cells.

Scx is a highly specific marker for connective tissues between muscle and bones, such as tendon and ligaments [116]. In other words, *Scx* is an early marker gene expressed in both tendon/ligament progenitors and tenocytes/ligamentocytes [117]. Moreover, *Scx* expression is practically nonexistent in undifferentiated mesenchymal progenitors [8].

ALP and osteocalcin are routinely used as markers for the mature osteoblast phenotype [73]. The human osteocalcin gene, *bglap*, is considered an indicator of osteoblast differentiation as it is synthetized and secreted exclusively by osteoblastic cells in the late stage of maturation [118]. In fact, osteocalcin is the major component of bone extracellular matrix [118]. On the other hand, *osteocalcin* was downregulated moderately in C3H10T1/2 cells in presence of BMP-2 and Smad8L-MH2 [8].

Even though *Sox9* is a key regulator of chondrogenesis [119], it is also expressed in tendon, bone- and muscle progenitor cells [120]. Tendons express *Sox9* during early stage in development, but not in completely developed tendon cells [120]. *Sox9* was found to be expressed in a subset of tendon and cartilage progenitor cells of mouse [120]. Interestingly, *Sox9* is able to mediate the direct conversion

of tenocytes to chondrocytes through an intermediate state in which both differentiation programs are active [119].

In Figure 23, it can be observed that the expression of the genes observed for C3HT0T1/2 cells treated with BMP-2 and TAT-hcSmad8 has the expected behavior when compared with C3HT0T1/2 cells treated with BMP-2 only. It is observed a reduction of the *bglap* expression in about 50 %, a duplication of the expression level of *Sox9* and most important, and a duplication in the expression level of *Scx*. In general, the global behavior of this group of genes suggest an early differentiation state of cells towards tenocytes. Further studies focused in cell differentiation should be carried out to optimize conditions to obtain the mature and full differentiated tenocytes. However, herein, it is given insights about TAT-hcSmad8 production and purification that led to a promising product that display bioactivity in C3H10T1/2 cells.

6. Conclusions

TAT-hcSmad8 could be successfully produced in *E. coli.* In this expression system, TAT-hcSmad8 behaves as an aggregation-prone protein as it forms IBs. This clear tendency remains under a range of temperatures ranging from 16 °C – 37 °C, induction concentration, type of inducer, expression time or strain. It was found no condition at which production of soluble protein can be obtained.

In vitro-refolded protein extracted from TAT-hcSmad8 under the condition here described, followed by purification by Heparin chromatography led to a product that displays a hydrophobic core (melting point (47.1 °C ± 0.1 °C). Moreover, a folding state comprised by secondary structured elements, was determined by circular dichroism. The mixture of α-helixes and β-sheets present in the purified TAT-hcSmad8 is consistent with the theoretical model elaborated by homology.

Produced TAT-hcSmad8 showed elevated solubility at acid pH. At neutral pH the solubility decreases although the propensity to form aggregates can be decreased with 20 % glycerol.

Form the biological point of view, TAT-hcSmad8 is able to successfully self-internalize into C3H10T1/2-cells. Confocal analysis of C3H10T1/2 incubated with Cy3-labeled TAT-hcSmad8 clearly showed the presence of the protein in the cytoplasmic space of cells and in less amount, in the nucleus. This observation was confirmed by immunocytochemistry using non-labeled TAT-hcSmad8.

Cytotoxicity assays demonstrated that refolded TAT-hcSmad8 is not cytotoxic for the cell, contrary to the IBs (not refolded) which presented intrinsic cytotoxicity in C3H10T1/2 cells.

Using luciferase-based reporter assay, it was determined that purified TAT-hcSmad8 inhibited the osteogenic effect generated by BMP-2 (500 ng/mL) on C3H10T1/2 incubated with 10 µg/mL of TAT-hcSmad8. This effect is in accordance to the activity described in other works [1, 9]. The expression of genes related to osteogenic pathways, such as *Bglap* (osteocalcin) decrease under the conditions mentioned before when compared to the C3H10T1/2 not treated with TAT-hcSmad8. It is observed a reduction of the *bglap* expression in about 50 %, a duplication of the expression level of *Sox9* and most important, and a duplication in the expression level of *Scx.* In general, the global behavior of this group of genes suggest an early differentiation state of cells towards tenocytes.

7. Comments and Outlook

The production and purification of proteins are challenging processes that depends on a lot of factors, especially when proteins are produced as IBs. Due to the insolubility of the IBs, detergents and chaotropic are routinely applied. The concentration used for the purification and solubilization are specific for each protein. The best conditions for a protein are found by trial and error. The refolding and purification of the solubilized protein is especially critical due to the prone of IBs to aggregation. And although the protein can remain soluble, the folding and activity can still be compromised. In this scenario, a protein that can be used as a reference is very helpful. As a reference, proteins produced in another system can be used. The comparison of the own purified proteins with commercial proteins is also common. However, this comparison is not possible so far for the specific Smad8L-MH2. Unfortunately, there are no reports of the purification of such specific variant.

Another important aspect is that the specific mechanism of Smad8L-MH2 is also not completely understood. In this sense, a cell can be seen as a black box that is stimulated by Smad8L-MH2 and BMP2 and turns into tenocytes. Knowing the mechanism through which this happens could increase the chances to characterize TAT-hcSmad8 herein produced. Determining for example if Smad8L-MH2 is phosphorylated at its SSXV motif would offer the possibility of exploring if TAT-hcSmad8 is phosphorylated after being internalized. Knowing for example if Smad8L-MH2 in fact complexes with Smad4 could offer a simple test the binding capacity of purified TAT-hcSmad8 to Smad4 by TSA or immunoprecipitation. In this complex, if it is known if other Smads are part of the complex as reported before, immunoprecipitation studies directed against other R-Smads.

In this regard, it would be necessary to produce soluble TAT-hcSmad8 in other systems. Eukaryotic systems such as HEK cells would be a possibility. Extensive characterization would be necessary to determine basically its structure, stability conditions and most importantly, the biological activity. In further studies, the conditions for the obtaining full differentiated tenocytes can be determined. Once that the desired tenocytes are obtained, it could be established that certain concentration of BMP-2 and TAT-hcSmad8 with well establish insights of its structure (such as circular dichroism) produces a certain effect on cells. This could be, gene expression patterns or the BRE-luc assay. Then, the IBs can be tested and compared. Finally, the virtues of IBs can be really exploited.

One should keep in mind of course, that the structure and features of TAT-hcSmad8 can diverge due to the change of system where it is produced. The chances of getting an active protein when it is produced in soluble form are greater than when it is produced in insoluble form. The purification

process could be simplified in the sense that the protein does not have to go through IBs washing and solubilization. Contact with strength reducing and denaturing conditions should be avoided.

More research needs to be done around Smad8 in order to understand better its role in cells and with this fully exploit Smad8 potential.

8. References

1. Tsukamoto, S., et al., *Smad9 is a new type of transcriptional regulator in bone morphogenetic protein signaling.* Sci Rep, 2014. **4**: p. 7596.

2. Segovia-Trinidad, C.L., et al., *Refolding, purification, and characterization of constitutive-active human-Smad8 produced as inclusion bodies in ClearColi® BL21 (DE3).* Protein Expr Purif, 2021. **184**: p. 105878.

3. Yang, J., et al., *Bone morphogenetic proteins: Relationship between molecular structure and their osteogenic activity.* Food Science and Human Wellness, 2014. **3**(3): p. 127.

4. Miyazono, K., Y. Kamiya, and M. Morikawa, *Bone morphogenetic protein receptors and signal transduction.* J Biochem, 2010. **147**(1): p. 35.

5. Macias, M.J., P. Martin-Malpartida, and J. Massagué, *Structural determinants of Smad function in TGF-β signaling.* Trends Biochem Sci, 2015. **40**(6): p. 296.

6. Attisano, L. and S.T. Lee-Hoeflich, *The Smads.* Genome Biol, 2001. **2**(8): p. REVIEWS3010.

7. Katakawa, Y., M. Funaba, and M. Murakami, *Smad8/9 is regulated through the BMP pathway.* J Cell Biochem, 2016. **117**(8): p. 1788.

8. Hoffmann, A., et al., *Neotendon formation induced by manipulation of the Smad8 signalling pathway in mesenchymal stem cells.* J Clin Invest, 2006. **116**(4): p. 940.

9. Patel, S.G., et al., *Cell-penetrating peptide sequence and modification dependent uptake and subcellular distribution of green florescent protein in different cell lines.* Sci Rep, 2019. **9**(1): p. 6298.

10. Logeart-Avramoglou, D., et al., *An assay for the determination of biologically active bone morphogenetic proteins using cells transfected with an inhibitor of differentiation promoter-luciferase construct.* Anal Biochem, 2006. **349**(1): p. 78.

11. Huang, Z., et al., *Defective pulmonary vascular remodeling in Smad8 mutant mice.* Human Molecular Genetics, 2009. **18**(15): p. 2791.

12. Walden, G., et al., *A clinical, biological, and biomaterials perspective into tendon injuries and regeneration.* Tissue Eng Part B Rev, 2017. **23**(1): p. 44.

13. Font Tellado, S., E.R. Balmayor, and M. Van Griensven, *Strategies to engineer tendon/ligament-to-bone interface: Biomaterials, cells and growth factors.* Adv Drug Deliv Rev, 2015. **94**: p. 126.

14. Gaspar, D., et al., *Progress in cell-based therapies for tendon repair.* Adv Drug Deliv Rev, 2015. **84**: p. 240.

15. Sundermann, J., et al., *ELISA- and activity assay-based quantification of BMP-2 released in Vitro can be biased by solubility in "physiological" buffers and an interfering effect of chitosan.* Pharmaceutics, 2021. **13**(4): p. 582.

16. Sundermann, J., et al., *Varying the sustained release of BMP-2 from chitosan nanogel-functionalized polycaprolactone fiber mats by different polycaprolactone surface modifications.* J Biomed Mater Res A, 2021. **109**(5): p. 600.

17. Reifenrath, J., et al., *TGF-β3 loaded electrospun polycaprolacton fibre scaffolds for rotator cuff tear repair: An in vivo study in rats.* Int J Mol Sci, 2020. **21**(3): p. 1046.

18. Gniesmer, S., et al., *Vascularization and biocompatibility of poly(ε-caprolactone) fiber mats for rotator cuff tear repair.* PLoS One, 2020. **15**(1): p. e0227563.

19. Shi, Y., *Structural insights on Smad function in TGFbeta signaling.* Bioessays, 2001. **23**(3): p. 223.

20. Schmierer, B. and C.S. Hill, *TGFbeta-SMAD signal transduction: molecular specificity and functional flexibility.* Nat Rev Mol Cell Biol, 2007. **8**(12): p. 970.

21. Poniatowski Ł, A., et al., *Transforming growth factor Beta family: insight into the role of growth factors in regulation of fracture healing biology and potential clinical applications.* Mediators Inflamm, 2015. **2015**: p. 137823.

22. Martinez-Hackert, E., A. Sundan, and T. Holien, *Receptor binding competition: A paradigm for regulating TGF-β family action.* Cytokine Growth Factor Rev, 2021. **57**: p. 39.

23. Kwiatkowski, W., P.C. Gray, and S. Choe, *Engineering TGF-β superfamily ligands for clinical applications.* Trends Pharmacol Sci, 2014. **35**(12): p. 648.

24. Yadin, D., P. Knaus, and T.D. Mueller, *Structural insights into BMP receptors: Specificity, activation and inhibition.* Cytokine Growth Factor Rev, 2016. **27**: p. 13.

25. Harrison, C.A., S.L. Al-Musawi, and K.L. Walton, *Prodomains regulate the synthesis, extracellular localisation and activity of TGF-β superfamily ligands.* Growth Factors, 2011. **29**(5): p. 174.

26. Mu, Y., S.K. Gudey, and M. Landström, *Non-Smad signaling pathways.* Cell Tissue Res, 2012. **347**(1): p. 11.

27. Zhang, Y.E., *Non-Smad pathways in TGF-beta signaling.* Cell Res, 2009. **19**(1): p. 128.

28. Zhang, Y.E., *Non-Smad signaling pathways of the TGF-β family.* Cold Spring Harb Perspect Biol, 2017. **9**(2): p. a022129.

29. Moustakas, A. and C.H. Heldin, *Non-Smad TGF-beta signals.* J Cell Sci, 2005. **118**(Pt 16): p. 3573.

30. Gordon, K.J. and G.C. Blobe, *Role of transforming growth factor-beta superfamily signaling pathways in human disease.* Biochim Biophys Acta, 2008. **1782**(4): p. 197.

31. Padgett, R.W., *TGFbeta signaling pathways and human diseases.* Cancer Metastasis Rev, 1999. **18**(2): p. 247.

32. Raftery, L.A. and D.J. Sutherland, *TGF-beta family signal transduction in Drosophila development: from Mad to Smads.* Dev Biol, 1999. **210**(2): p. 251.

33.   Masuyama, N., et al., *Identification of two Smad4 proteins in Xenopus. Their common and distinct properties.* J Biol Chem, 1999. **274**(17): p. 12163.

34.   Miyazawa, K. and K. Miyazono, *Regulation of TGF-β family signaling by inhibitory Smads.* Cold Spring Harb Perspect Biol, 2017. **9**(3): p. a022095.

35.   Chai, N., et al., *Structural basis for the Smad5 MH1 domain to recognize different DNA sequences.* Nucleic Acids Research, 2015. **43**(18): p. 9051.

36.   Chai, J., et al., *Features of a Smad3 MH1-DNA complex. Roles of water and zinc in DNA binding.* J Biol Chem, 2003. **278**(22): p. 20327.

37.   Shi, Y., et al., *Crystal structure of a Smad MH1 domain bound to DNA: insights on DNA binding in TGF-beta signaling.* Cell, 1998. **94**(5): p. 585.

38.   Martin-Malpartida, P., et al., *Structural basis for genome wide recognition of 5-bp GC motifs by SMAD transcription factors.* Nat Commun, 2017. **8**(1): p. 2070.

39.   BabuRajendran, N., et al., *Structure of Smad1 MH1/DNA complex reveals distinctive rearrangements of BMP and TGF-beta effectors.* Nucleic Acids Res, 2010. **38**(10): p. 3477.

40.   Ruiz, L., et al., *Unveiling the dimer/monomer propensities of Smad MH1-DNA complexes.* Comput Struct Biotechnol J, 2021. **19**: p. 632.

41.   Wang, C., et al., *Crystal structure of the MH2 domain of Drosophila Mad.* Sci China C Life Sci, 2009. **52**(6): p. 539.

42.   Beich-Frandsen, M., et al., *Structure of the N-terminal domain of the protein Expansion: an 'expansion' to the Smad MH2 fold.* Acta Crystallogr D Biol Crystallogr, 2015. **71**(Pt 4): p. 844.

43.   Hao, R., et al., *Structure of Drosophila Mad MH2 domain.* Acta Crystallogr Sect F Struct Biol Cryst Commun, 2008. **64**(Pt 11): p. 986.

44.   Wu, J.W., et al., *Crystal structure of a phosphorylated Smad2. Recognition of phosphoserine by the MH2 domain and insights on Smad function in TGF-beta signaling.* Mol Cell, 2001. **8**(6): p. 1277.

45.   Wu, G., et al., *Structural basis of Smad2 recognition by the Smad anchor for receptor activation.* Science, 2000. **287**(5450): p. 92.

46.   Miyazono, K.I., et al., *Structural basis for receptor-regulated SMAD recognition by MAN1.* Nucleic Acids Res, 2018. **46**(22): p. 12139.

47.   Murayama, K., et al., *Structural basis for inhibitory effects of Smad7 on TGF-β family signaling.* J Struct Biol, 2020. **212**(3): p. 107661.

48.   Qin, B.Y., et al., *Smad3 allostery links TGF-beta receptor kinase activation to transcriptional control.* Genes Dev, 2002. **16**(15): p. 1950.

49.   Qin, B.Y., et al., *Structural basis of Smad1 activation by receptor kinase phosphorylation.* Mol Cell, 2001. **8**(6): p. 1303.

50. Dennler, S., S. Huet, and J.M. Gauthier, *A short amino-acid sequence in MH1 domain is responsible for functional differences between Smad2 and Smad3.* Oncogene, 1999. **18**(8): p. 1643.

51. Abdollah, S., et al., *TbetaRI phosphorylation of Smad2 on Ser465 and Ser467 is required for Smad2-Smad4 complex formation and signaling.* J Biol Chem, 1997. **272**(44): p. 27678.

52. Yan, X., Z. Liu, and Y. Chen, *Regulation of TGF-beta signaling by Smad7.* Acta Biochim Biophys Sin (Shanghai), 2009. **41**(4): p. 263.

53. Park, S.H., *Fine tuning and cross-talking of TGF-beta signal by inhibitory Smads.* J Biochem Mol Biol, 2005. **38**(1): p. 9.

54. Hanyu, A., et al., *The N domain of Smad7 is essential for specific inhibition of transforming growth factor-beta signaling.* J Cell Biol, 2001. **155**(6): p. 1017.

55. Hata, A., Y. Shi, and J. Massagué, *TGF-beta signaling and cancer: structural and functional consequences of mutations in Smads.* Mol Med Today, 1998. **4**(6): p. 257.

56. Chacko, B.M., et al., *Structural basis of heteromeric smad protein assembly in TGF-beta signaling.* Mol Cell, 2004. **15**(5): p. 813.

57. Chacko, B.M., et al., *The L3 loop and C-terminal phosphorylation jointly define Smad protein trimerization.* Nat Struct Biol, 2001. **8**(3): p. 248.

58. Hata, A. and Y.G. Chen, *TGF-β signaling from receptors to Smads.* Cold Spring Harb Perspect Biol, 2016. **8**(9): p. a022061.

59. Kawabata, M., et al., *Smad proteins exist as monomers in vivo and undergo homo- and hetero-oligomerization upon activation by serine/threonine kinase receptors.* EMBO J, 1998. **17**(14): p. 4056.

60. Jayaraman, L. and J. Massague, *Distinct oligomeric states of SMAD proteins in the transforming growth factor-beta pathway.* J Biol Chem, 2000. **275**(52): p. 40710.

61. Watanabe, T.K., et al., *Cloning and characterization of a novel member of the human Mad gene family (MADH6).* Genomics, 1997. **42**(3): p. 446.

62. Chen, Y., A. Bhushan, and W. Vale, *Smad8 mediates the signaling of the ALK-2 [corrected] receptor serine kinase.* Proc Natl Acad Sci U S A, 1997. **94**(24): p. 12938.

63. Nishita, M., N. Ueno, and H. Shibuya, *Smad8B, a Smad8 splice variant lacking the SSXS site that inhibits Smad8-mediated signalling.* Genes Cells, 1999. **4**(10): p. 583.

64. Kawai, S., et al., *Mouse smad8 phosphorylation downstream of BMP receptors ALK-2, ALK-3, and ALK-6 induces its association with Smad4 and transcriptional activity.* Biochem Biophys Res Commun, 2000. **271**(3): p. 682.

65. Lee, J.H., et al., *CREBZF, a novel Smad8-binding protein.* Mol Cell Biochem, 2012. **368**(1-2): p. 147.

66. Min, K.D., et al., *ASB2 is a novel E3 ligase of SMAD9 required for cardiogenesis.* Sci Rep, 2021. **11**(1): p. 23056.

67.   Cheng, K.H., J.F. Ponte, and S. Thiagalingam, *Elucidation of epigenetic inactivation of SMAD8 in cancer using targeted expressed gene display.* Cancer Res, 2004. **64**(5): p. 1639.

68.   Horvath, L.G., et al., *Loss of BMP2, Smad8, and Smad4 expression in prostate cancer progression.* Prostate, 2004. **59**(3): p. 234.

69.   Gregson, C.L., et al., *A rare mutation in SMAD9 associated with high bone mass identifies the SMAD-dependent BMP signaling pathway as a potential anabolic target for osteoporosis.* J Bone Miner Res, 2020. **35**(1): p. 92.

70.   Gregson, C.L. and E.L. Duncan, *The genetic architecture of high bone mass.* Front Endocrinol (Lausanne), 2020. **11**: p. 595653.

71.   Shintani, M., et al., *A new nonsense mutation of SMAD8 associated with pulmonary arterial hypertension.* J Med Genet, 2009. **46**(5): p. 331.

72.   Drake, K.M., et al., *Altered MicroRNA processing in heritable pulmonary arterial hypertension: an important role for Smad-8.* Am J Respir Crit Care Med, 2011. **184**(12): p. 1400.

73.   Vaes, B.L., et al., *Comprehensive microarray analysis of bone morphogenetic protein 2-induced osteoblast differentiation resulting in the identification of novel markers for bone development.* J Bone Miner Res, 2002. **17**(12): p. 2106.

74.   Kirsch, M., et al., *Xeno-free in vitro cultivation and osteogenic differentiation of hAD-MSCs on resorbable 3D printed RESOMER(®).* Materials (Basel), 2020. **13**(15): p. 3399.

75.   Rosano, G.L. and E.A. Ceccarelli, *Recombinant protein expression in Escherichia coli: advances and challenges.* Front Microbiol, 2014. **5**: p. 172.

76.   Brondyk, W.H., *Selecting an appropriate method for expressing a recombinant protein.* Methods Enzymol, 2009. **463**: p. 131.

77.   Andersen, D.C. and L. Krummen, *Recombinant protein expression for therapeutic applications.* Curr Opin Biotechnol, 2002. **13**(2): p. 117.

78.   Assenberg, R., et al., *Advances in recombinant protein expression for use in pharmaceutical research.* Curr Opin Struct Biol, 2013. **23**(3): p. 393.

79.   Kaper, J.B., J.P. Nataro, and H.L. Mobley, *Pathogenic Escherichia coli.* Nat Rev Microbiol, 2004. **2**(2): p. 123.

80.   Blount, Z.D., *The unexhausted potential of E. coli.* Elife, 2015. **4**: p. e05826.

81.   Murthy, A.M., et al., *Production and evaluation of virus-like particles displaying immunogenic epitopes of porcine reproductive and respiratory syndrome virus (PRRSV).* Int J Mol Sci, 2015. **16**(4): p. 8382.

82.   Francis, D.M. and R. Page, *Strategies to optimize protein expression in E. coli.* Curr Protoc Protein Sci, 2010. **Chapter 5**(1): p. Unit 5 24 1.

83.   Seras-Franzoso, J., et al., *Cellular uptake and intracellular fate of protein releasing bacterial amyloids in mammalian cells.* Soft Matter, 2016. **12**(14): p. 3451.

84. Rinas, U., et al., *Bacterial inclusion bodies: Discovering their better half.* Trends Biochem Sci, 2017. **42**(9): p. 726.

85. King, J., et al., *Thermolabile folding intermediates: inclusion body precursors and chaperonin substrates.* FASEB J, 1996. **10**(1): p. 57.

86. Quaas, B., et al., *Stability and biological activity of E. coli derived soluble and precipitated bone morphogenetic protein-2.* Pharm Res, 2019. **36**(12): p. 184.

87. Cano-Garrido, O., et al., *Supramolecular organization of protein-releasing functional amyloids solved in bacterial inclusion bodies.* Acta Biomater, 2013. **9**(4): p. 6134.

88. Seras-Franzoso, J., et al., *Functionalization of 3D scaffolds with protein-releasing biomaterials for intracellular delivery.* J Control Release, 2013. **171**(1): p. 63.

89. Díez-Gil, C., et al., *The nanoscale properties of bacterial inclusion bodies and their effect on mammalian cell proliferation.* Biomaterials, 2010. **31**(22): p. 5805.

90. Stamm, A., et al., *Positive in vitro wound healing effects of functional inclusion bodies of a lipoxygenase from the Mexican axolotl.* Microb Cell Fact, 2018. **17**(1): p. 57.

91. Rodríguez-Carmona, E., et al., *Isolation of cell-free bacterial inclusion bodies.* Microb Cell Fact, 2010. **9**: p. 71.

92. Mamat, U., et al., *Detoxifying Escherichia coli for endotoxin-free production of recombinant proteins.* Microb Cell Fact, 2015. **14**: p. 57.

93. Bastian, Q., *Rekombinante Produktion, Aufreinigung und Charakterisierung von Signalproteinen für die Regeneration von Sehnen-Knochen-Defekten (original title in German)*, in *Institute of Technical Chemistry* 2019, Gottfried Wilhelm Leibniz Universität Hannover: German National Library of Science and Technology (TIB).

94. Li, Z., et al., *Simple defined autoinduction medium for high-level recombinant protein production using T7-based Escherichia coli expression systems.* Appl Microbiol Biotechnol, 2011. **91**(4): p. 1203.

95. Gasteiger, E., et al., *ExPASy: The proteomics server for in-depth protein knowledge and analysis.* Nucleic Acids Res, 2003. **31**(13): p. 3784.

96. Candiano, G., et al., *Blue silver: a very sensitive colloidal Coomassie G-250 staining for proteome analysis.* Electrophoresis, 2004. **25**(9): p. 1327.

97. Xue, B., et al., *PONDR-FIT: a meta-predictor of intrinsically disordered amino acids.* Biochim Biophys Acta, 2010. **1804**(4): p. 996.

98. Owen, T.A., et al., *Progressive development of the rat osteoblast phenotype in vitro: reciprocal relationships in expression of genes associated with osteoblast proliferation and differentiation during formation of the bone extracellular matrix.* J Cell Physiol, 1990. **143**(3): p. 420.

99. Ghosh, S., et al., *Method for enhancing solubility of the expressed recombinant proteins in Escherichia coli.* Biotechniques, 2004. **37**(3): p. 418.

100. Greenfield, N.J., *Using circular dichroism spectra to estimate protein secondary structure.* Nat Protoc, 2006. **1**(6): p. 2876.

101. Gopal, R., et al., *Applications of circular dichroism for structural analysis of gelatin and antimicrobial peptides.* Int J Mol Sci, 2012. **13**(3): p. 3229.

102. Sánchez-Rodríguez, S.P., et al., *Human parvovirus B19 virus-like particles: In vitro assembly and stability.* Biochimie, 2012. **94**(3): p. 870.

103. Sánchez-Rodríguez, S.P., et al., *In vitro encapsulation of heterologous dsDNA into human parvovirus B19 virus-like particles.* Mol Biotechnol, 2015. **57**(4): p. 309.

104. Hammarström, M., et al., *Rapid screening for improved solubility of small human proteins produced as fusion proteins in Escherichia coli.* Protein Sci, 2002. **11**(2): p. 313.

105. Wang, Y., et al., *A systematic protein refolding screen method using the DGR approach reveals that time and secondary TSA are essential variables.* Sci Rep, 2017. **7**(1): p. 9355.

106. Ziegler, A. and J. Seelig, *Interaction of the protein transduction domain of HIV-1 TAT with heparan sulfate: binding mechanism and thermodynamic parameters.* Biophys J, 2004. **86**(1 Pt 1): p. 254.

107. Huynh, K. and C.L. Partch, *Analysis of protein stability and ligand interactions by thermal shift assay.* Curr Protoc Protein Sci, 2015. **79**: p. 28 9 1.

108. Boivin, S., S. Kozak, and R. Meijers, *Optimization of protein purification and characterization using Thermofluor screens.* Protein Expr Purif, 2013. **91**(2): p. 192.

109. Micsonai, A., et al., *Accurate secondary structure prediction and fold recognition for circular dichroism spectroscopy.* Proc Natl Acad Sci U S A, 2015. **112**(24): p. E3095.

110. Craig, D.B. and A.A. Dombkowski, *Disulfide by Design 2.0: a web-based tool for disulfide engineering in proteins.* BMC Bioinformatics, 2013. **14**: p. 346.

111. Vagenende, V., M.G. Yap, and B.L. Trout, *Mechanisms of protein stabilization and prevention of protein aggregation by glycerol.* Biochemistry, 2009. **48**(46): p. 11084.

112. Rizzuti, M., et al., *Therapeutic applications of the cell-penetrating HIV-1 Tat peptide.* Drug Discov Today, 2015. **20**(1): p. 76.

113. Derakhshankhah, H. and S. Jafari, *Cell penetrating peptides: A concise review with emphasis on biomedical applications.* Biomed Pharmacother, 2018. **108**: p. 1090.

114. Tabatabai, M.A. and J.M. Bremner, *Use of p-nitrophenyl phosphate for assay of soil phosphatase activity.* Soil Biology and Biochemistry, 1969. **1**(4): p. 301.

115. Paul, H., A.J. Reginato, and H.R. Schumacher, *Alizarin red S staining as a screening test to detect calcium compounds in synovial fluid.* Arthritis Rheum, 1983. **26**(2): p. 191.

116. Schweitzer, R., et al., *Analysis of the tendon cell fate using Scleraxis, a specific marker for tendons and ligaments.* Development, 2001. **128**(19): p. 3855.

117.   Shukunami, C., et al., *Scleraxis is a transcriptional activator that regulates the expression of Tenomodulin, a marker of mature tenocytes and ligamentocytes.* Sci Rep, 2018. **8**(1): p. 3155.

118.   Sila-Asna, M., et al., *Osteoblast differentiation and bone formation gene expression in strontium-inducing bone marrow mesenchymal stem cell.* Kobe J Med Sci, 2007. **53**(1-2): p. 25.

119.   Takimoto, A., et al., *Direct conversion of tenocytes into chondrocytes by Sox9.* Experimental Cell Research, 2012. **318**(13): p. 1492.

120.   Nagakura, R., et al., *Switching of Sox9 expression during musculoskeletal system development.* Sci Rep, 2020. **10**(1): p. 8425.

9. Appendix

Table 5. Ligands and their respective activation elements.

Taken from [24]

	Ligand	Type I receptor	Type II receptor	R-Smad activation
BMPs/GDFs	BMP2/4	BMPRIa BMPRIb	BMPRII ActRIIA ActRIIB	Smad1/5/8
	BMP5/6/7 /8	Alk2 BMPRIa BMPRIb	ActRIIA ActRIIB BMPRII	Smad1/5/8
	GDF5/6/7	BMPRIa BMPRIb	BMPRII ActRIIA ActRIIB	Smad1/5/8
	BMP9/10	Alk1	BMPRII ActRIIB ActRIIA	Smad1/5/8
	GDF1/3	ActRIb	ActRIIA ActRIIB	Smad2/3
	BMP3/GDF10	ActRIb	ActRIIA ActRIIB	Smad2/3
	GDF8 (myostatin)/ GDF11	ActRIb TβRI	ActRIIA ActRIIB	Smad2/3
	BMP15/ GDF9	BMPRIb ActRIb TβRI	BMPRII BMPRII	Smad1/5/8 Smad2/3
TGFβs	TGFβ1/2/3	TβRI Alk1	TβRII	Smad2/3 Smad1/5/8
activins	activing A	ActRIb	ActRIIA ActRIIB	Smad2/3
	activing B	ActRIb Alk7	ActRIIA ActRIIB	Smad2/3
	nodal	ActRIb Alk7	ActRIIB	Smad2/3
outsider	AMH	BMPRIa BMPRIb Alk2	AMHRII	Smad1/5/8
	GDF15	-	-	-
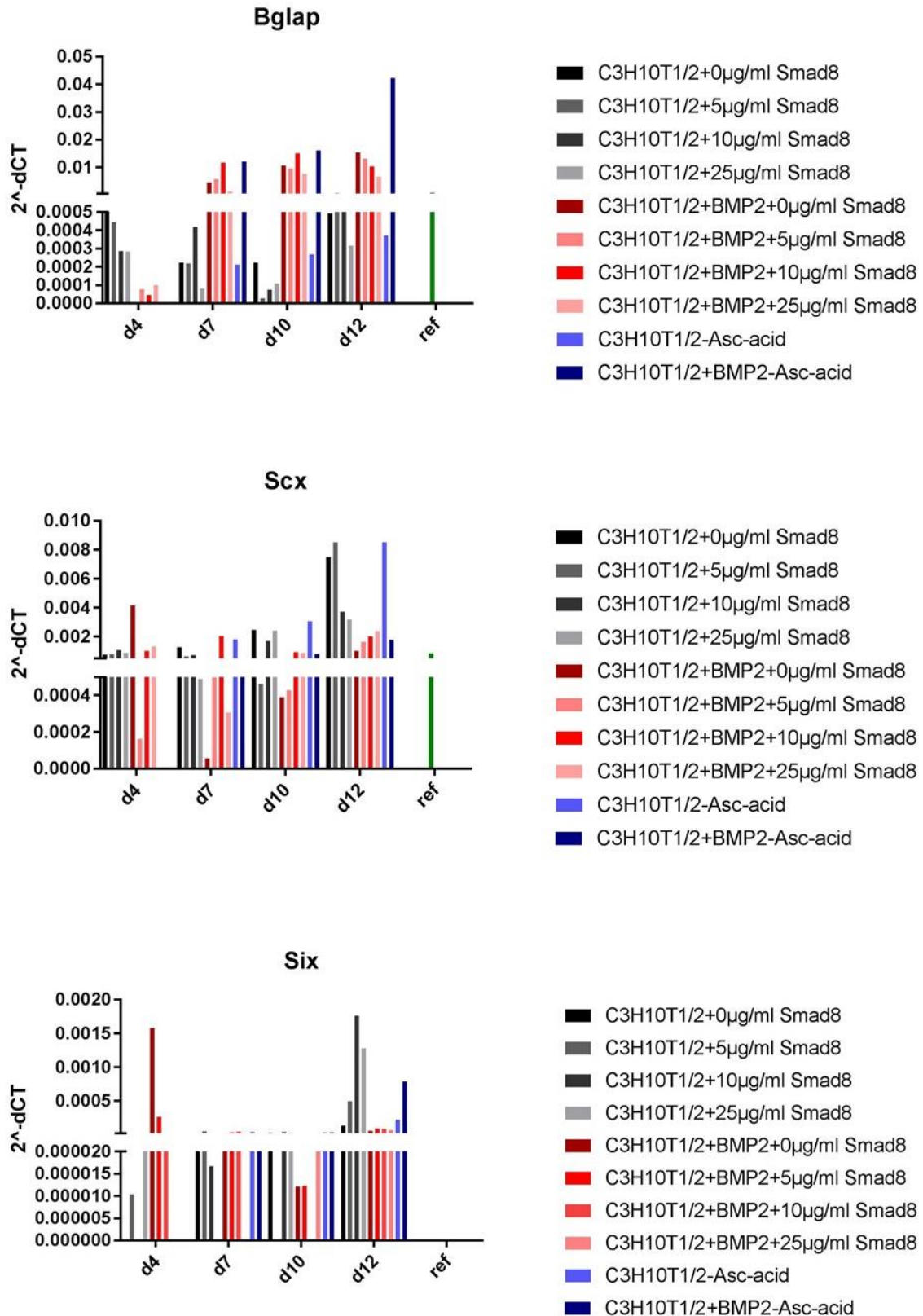
Figure 33. Gene expression screening of C3H10T1/2 cells incubated with different concentrations of TAT-hcSmad8 and BMP-2 during 4, 7, 10 and 12 days.

Sox9

Eya-1

Epha4
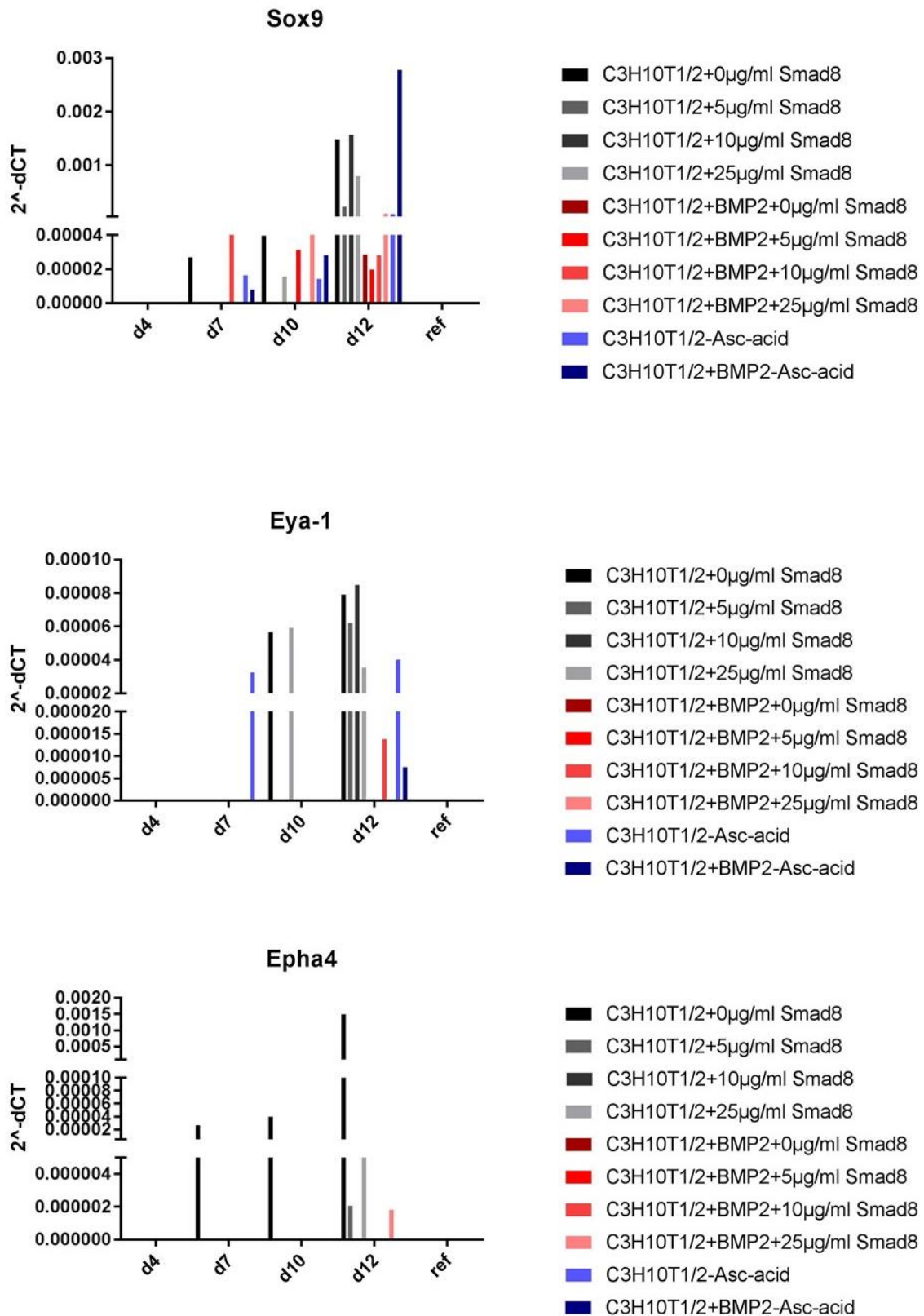
Figure 34 (continuation). Gene expression screening of C3H10T1/2 cells incubated with different concentrations of TAT-hcSmad8 and BMP-2 during 4, 7, 10 and 12 days.
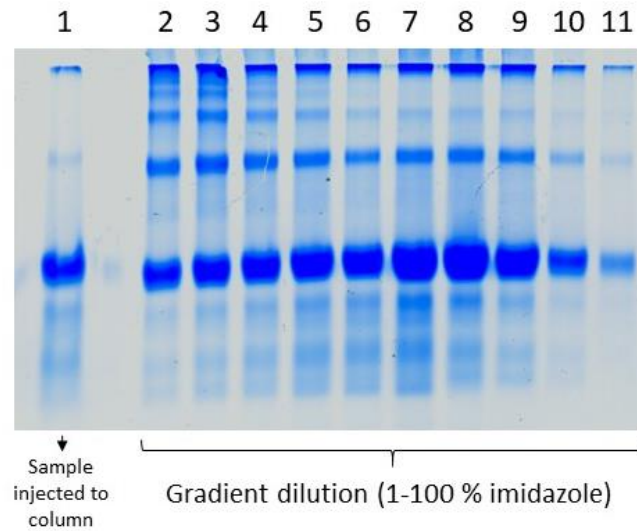
Figure 35. SDS-PAGE of fractions obtained during the purification by immobilized metal affinity chromatography (IMAC) of TAT-hcSmad8.
For this, a HisTrap HP 1 mL was equilibrated with solubilization buffer without EDTA [50 mM MOPS pH 7.5, 6 M GuHCl , 25 mM DTT]. The column was washed with the same buffer and finally eluted with a gradient 1-100 % of imidazole in *solubilization buffer without EDTA.*

10. Index of Figures and Tables

Figures

Tables

11. Acknowledgements

I would like to infinitely thank my parents Carlos and Elizabeth for all their love, support, and guidance. I would not be here if it was not for their unconditional love and support. To my sister Carina because I love her very much.

To my dear professors who accompanied me throughout my academic life. Especially to my professor Luis. Thanks to a short research stay under your supervision I became interested in biochemistry. To my professor Ismael, for allowing me to work with virus-like particles. To my teacher Gloria, for her hard but wise criticism that boosted me to finish my bachelor thesis. To my high school teacher Aquino who prepared me for a mysterious exam that turned out to be the selection exam for Mexico's most prestigious university.

I would like to thank my university, the National Autonomous University of Mexico, whose excellent educational quality opened the doors to knowledge, science, and student exchanges. Through this, I discover my desire to travel and to learn. I became open and curious.

In this path of life, I would like to thank those people who, although they were not the final destination, rightly or wrongly, guided me to the one that would be. To those friends who left, but from whom I learned and discovered a lot at their side.

Dear Uschi, thank you for giving me the opportunity to have a better life. For allowing me to compete for the DAAD scholarship. Thank you for your patience and support.

Thank you DAAD for giving me the financial means to learn German and to do my doctorate. This changed my life by giving me the opportunity to immigrate to another beautiful country.

To my lab partners Bastian, Li, Maria, and Francisco. Thank you for your friendship and help.

Thank you Martina and Martin for the technical support. To Dr. Tonya, thank you very much for welcoming me to your cell culture lab and helping me with my project.

To my dear husband Eike. Thank you for your love and support. I have never met someone as beautiful as you. I hope to spend the rest of my life with you. To my parents-in-law, Katrin and Günther, thank you for your care and support.

Versión en español

Me gustaría agradecer infinitamente a mis padres Carlos y Elizabeth por todo su amor, apoyo y guía. Yo no estaría aqui de no haber sido por su amor y apoyo incondicional. A mi hermana Carina porque la quiero mucho.

A mis queridos profesores que me acompanaron a lo largo de mi vida academica. En especial a mi profesor Luis, que gracias a una estancia corta de investigación me interecé por la bioquímica. A mi profesor Ismael, por permitirme trabajar con las partículas tipo virus. A la maestra Gloria, por su dura pero atinada crítica que fomentó mi crecemiento. A mi maestro de preparatoria Aquino, que me preparó para un misterioso examen que resultó ser el examen de admisión a la más prestigiosa universidad de México.

Quisiera agradecer a mi universidad, a la Universidad Nacional Autónoma de México cuya excelente calidad educativa me abrió las puertas al conociemito, a la ciencia y a los intercambios estudiantiles que me permitieron descubrir mi deseo de viajar y conocer. De ser abierta y curiosa.

En ese camino de la vida, me gustaría agradecer a esas personas que si bien no fueron el destino final, bien que mal me guiaron al que sí lo sería. A esos amigos que se fueron pero que a su lado aprendí y descubrí mucho.

Estimada Uschi, gracias por haberme brindado la oportunidad de mi vida al permitirme concursar por la beca del DAAD. Gracias por tu paciencia y apoyo.

Gracias DAAD por haberme ofrecido los recursos económicos para aprender alemán y para realizar mi doctorado. Con ello, cambió mi vida al tener la oportunidad de emigrar a otro país tan hermoso como lo es Alemania.

A mis companeros de laboratorio Bastian, Li, Maria y Francisco. Gracias por su amistad y ayuda. Gracias Martina y Martin por el apoyo técnico. A la doctora Tonya, muchas gracias por recibirme en tu laboratorio de cultivo celular y por ayudarme en mi proyecto.

A mi querido esposo Eike. Gracias por tu amor y apoyo. Nunca había conocido alguien tan lindo como tu. Espero la vida me regale muchos años a tu lado. A mis suegros, Katrin y Gunther, por su cariño y apoyo.

12. Curriculum vitae

2007 - 2010	**High School – Exact Sciences** John J. Sparks	Veracruz, Mexico
2010-2015	**Bachelor – Chemical engineering** National Autonomous University of Mexico (UNAM)	Mexico City, Mexico
2016-2018	**Master of Science – Biochemistry** National Autonomous University of Mexico (UNAM)	Mexico City, Mexico
2018-2022	**PhD – Natural Sciences** Leibniz University Hannover (LUH)	Hannover, Germany