Investigation of Kinase Activation in Fibrodysplasia Ossificans Progressiva

Caroline Elizabeth Sanvitale

Green Templeton College

University of Oxford

Trinity Term 2013

Thesis submitted for the degree of Doctor of Philosophy in Clinical Medicine

Declaration

I declare that there are no parts of this thesis or its research herein have been reproduced or accepted for another award or degree or diploma at any other university or learning institution. This thesis contains no other person's work except where stated in text.

Acknowledgements

There are many people who must be acknowledged for their part in making this thesis possible. Firstly my two supervisors for allowing me to undertake this fascinating project; Dr Alex Bullock, whose knowledge, guidance and enduring patience have been pivotal to the success of this thesis; and Professor Jim Triffitt who has provided kind encouragement throughout and insight into the more clinical aspects of the disease.

The GFS group past and present for making every day a pleasure to be in the lab, Eleanor Williams, Georgina Kerr, Peter Canning, Fiona Sorrell, Sarah Dixon-Clarke and Vikram Ayinampudi; and the SGC as a whole for the wealth of knowledge shared by all, in particular James Bray, Ivan Alfano, Grazyna Kochan, Rod Chalk, Apirat Chaikuad, Tracy Keates, and every single member of The Lunch Group.

Green Templeton College, and the many friends who have made my time at Oxford great, especially members of GTSC and GTBC. My family, who continue to support me unconditionally. Amalie Rudolf, whose love and support has kept me going through everything.

Finally I would like thank the University of Oxford FOP Research Fund, which is supported by friends and families of those with FOP to promote research into this rare disease, for funding this project.

Abstract

Fibrodysplasia ossificans progressiva (FOP) is a rare autosomal dominant disease resulting in episodic but progressive extraskeletal bone formation. FOP is caused by missense mutations in the cytoplasmic domain of the type I bone morphogenetic protein (BMP) receptor ACVR1, leading to dysregulated activation. Currently there are no available drug treatments and the structural mechanism of mutant activation is still poorly characterised. To address this, a number of BMP and TGFβ receptors, including FOP mutants of ACVR1 were cloned, expressed and purified for both structural and biophysical experiments. The arginine at the site of most recurrent FOP mutation, R206H, is common across all type I receptors except BMPR1A and BMPR1B which have a lysine at this site. The novel structure of BMPR1B differed to wild-type ACVR1 showing some of the conformational changes expected of the active conformation. However, a variety of disease related ACVR1 mutant structures, including ACVR1 R206H, revealed a surprisingly persistent inactive conformation in the kinase domain. Some conformational changes suggestive of activation were observed in the mutant Q207D affecting the ATP pocket, the β4–β5 hairpin and the activation loop. Additionally, the structure of the Q207E mutant showed a slight release of the regulatory glycine-serine rich domain from its inhibitory position. These subtle changes suggest that the mutant inactive conformation is destabilised and potentially more dynamic. In agreement, all of the ACVR1 mutants showed reduced binding to the inhibitory protein FKBP12. However, mutant phosphorylation of the substrate Smad1 was not constitutive, but dependent on the co-expression of the partner ACVR2, consistent with recent evidence from transgenic knock-out mice. A novel 2-aminopyridine inhibitor scaffold with favourable specificity for ACVR1 was identified using a fluorescence-based thermal shift assay. Further derivatives were characterised with improved potency and selectivity. The crystal structures of ACVR1 bound to these inhibitors showed exquisite shape complementarity, contributing to their favourable specificity. This work has increased the understanding of FOP-associated mutant activation and provided a novel starting scaffold for potential drug development.

Contents

Table of Figures

List of Tables

List of Abbreviations

ACVR1	activin A receptor type I
ade. Cre	Adenovirus expressing Cre
A-loop	Activation loop
AMH	Anti-Müllerian hormone
ATP	Adenosine-5'-triphosphate
BMP	Bone morphogenetic protein
CID	Collision –induced dissociation
DEAE	Diethylaminoethyl
DHB	Dihydroxybenzoic acid
DNA	Deoxyribonucleic acid
DSF	Differential scanning fluorimetry
DTT	Dithiothreitol
ETD	Electron-transfer dissociation
FKBP12	Peptidyl-prolyl cis-trans isomerase FKBP1A
FOP	Fibrodysplasia ossificans progressiva
GDF	Growth differentiation factor
GS domain	Gycine/Serine rich domain
GST	Glutathione S-transferase
HPLC	High-performance liquid chromatography
IC_{50}	Half maximal inhibitory concentration
IDA	Iminodiacetic acid
ITC	Isothermal titration calorimetry

JCSG	Joint Center for Structural Genomics
JPS	Juvenile polyposis syndrome
K_d	Dissociation constant
LC-MS	Liquid chromatography- mass spectrometry
LFS	Ligand friendly screen
LIC	Ligation independent cloning
MH	Mad homology
MOWSE	Molecular weight search
MS/MS	Tandem mass spectrometry
MW	Molecular weight
PCR	Polymerase chain reaction
PAH	Pulmonary arterial hypertension
PEG	Polyethylene glycol
P-loop	Phosphate binding loop
SAR	Structure-activity relationship
SARA	Smad anchor for receptor activation
SBE	Smad binding element
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
Sf9	Spodoptera frugiperda
Smad	Sma and Mad homolog
TAB1	TGF-β-activated kinase binding protein
TAK1	TGF-β-activated kinase
TCEP	tris(2-carboxyethyl)phosphine
TEV	Tobacco etch virus

TFA	Trifluoroacetic acid
TGFβ	Transforming growth factor beta
Tm	Melting temperature
XIAP	X-linked inhibitor of apoptosis protein

1. Introduction

1.1 Fibrodysplasia Ossificans Progressiva

Fibrodysplasia ossificans progressiva (FOP) is a rare autosomal dominant disease resulting in unpredictable, episodic and progressive heterotopic bone formation (ossification) in skeletal muscles, ligaments and tendons [1,2,3] (Figure 1.1A). The disease, affecting 1 in 2 million people [4], is seriously debilitating to patients, ultimately leaving them wheelchair bound as bone growth bridges joints and severely impedes joint mobility [2,5]. To date there is no effective treatment for the disease [6]. The average life expectancy is approximately 45 years and death often results from thoracic insufficiency syndrome as intercostal muscles ossify and restrict effective respiration [3,7]. This thesis aims to address an urgent therapeutic need for this condition by better defining the molecular mechanisms involved in the development of FOP and identifying small molecule FOP inhibitors.

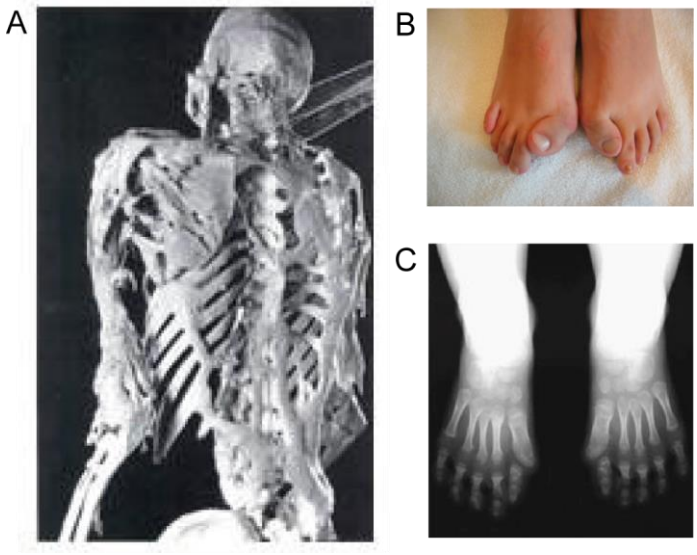


Figure 1.1. Classical symptoms of FOP patients. (A) Skeleton of Harry Eastlack, who lived with FOP to age 39, showing ectopic bone formation (on display at The Mutter Museum of The College of Physicians in Philadelphia). (B) photograph and (C) radiograph recording shortened great toes as an early diagnostic for FOP [8].

The first symptoms of FOP are detectable at birth with malformation of the great toe [2], [3] (Figure 1.1B and 1C). The hallux valgus appearance of the great toe is often caused by deformities present in the first metatarsal and the proximal phalanges [8], which often fuse with the distal phalanges later in life [9]. Periods of ossification usually begin during childhood and can occur as a result of trauma or spontaneously, and are preceded by a period of inflammation [2,3]. Trauma caused by surgical intervention also results in further bone growth [3]. The heterotopic ossification occurs in skeletal muscle, ligaments and tendons, but not in smooth or cardiac muscle [7]. The ectopic bone formed resembles normal bone and develops following the endochondral pathway [7]. The pattern of ossification in FOP appears generally to follow that seen during skeletal development in embryogenesis, beginning in dorsal, axial, cranial and proximal regions, affecting the ventral, appendicular, caudal and distal regions later in life [3,7], although this can be altered by trauma [4].

1.2 Identification of *ACVR1* as the causative gene

Genetic linkage of five small families of FOP patients enabled the disease locus to be mapped to chromosome 2q23-34 [10]. DNA sequencing identified the causative gene as activin A receptor type I (*ACVR1)* [10] which encodes a type I bone morphogenetic protein (BMP) receptor, also known as ALK2. This discovery helped explain earlier suggestions concerning the involvement of BMPs in this condition since BMPs were initially discovered by their ability to induce bone formation in skeletal muscle [11,12]. The ACVR1 protein consists of an extracellular ligand binding domain, a transmembrane helix, a cytoplasmic GS (glycine/serine rich) regulatory domain and a serine/threonine kinase domain [4,10]. The GS domain is an unique regulatory feature of these type I receptors [13], so named for its SGSGSG loop sequence, which is a target for activating

phosphorylations [1,14]. To date all instances of FOP have been linked to missense mutations in the cytoplasmic domains of ACVR1. The most common mutation R206H, observed in approximately 95% of cases, is located in the GS domain. Other point mutations have been identified elsewhere in the GS domain, as well as in the kinase domain [1,15,16,17,18] (Table 1.1 and Figure 1.2).

Table 1.1. Summary of FOP mutations.

ACVR1 mutation	ACVR1 domain	Amino acid change	Age of onset (years)	Clinical course	Clinical Features[a]
c.617G>A	GS	R206H	0.3-14	Moderate/ Severe	Classical FOP
c.587T>C	GS	L196P	21	Slow	FOP-variant (mild)
c.590_592delCTT	GS	P197_F198delinsL	11	Moderate	FOP-variant
c.605G>T	GS	R202I	14	Moderate	FOP-variant*
c.619C>G	GS	Q207E	0.5	Severe	FOP-plus
c.774G>C	Kinase	R258S	4 - 14	Slow	FOP-variant
c.974G>C	Kinase	G325A	47	Moderate	Classical FOP
c.983G>A	Kinase	G328E	2	Slow/ Severe	FOP-variant
c.982G>A	Kinase	G328R	21 - 26	Slow	FOP-variant
c.982G>T	kinase	G328W	2 - 8	Severe	FOP-variant
c.1067G>A	Kinase	G356D	10	Slow	FOP-plus
c.1124G>C	Kinase	R375P	14	Slow	FOP-variant

[a]The classical FOP phenotype has two main identifiers, heterotopic ossification and malformation of the great toes [4]. FOP-plus mutations have additional symptoms. FOP-variants have variations on the two classical features of FOP [3,4,16,17].

*Contains malformation of only one of the great toes [1]. Adapted from Kaplan et al (2009) [1] and Gregson et al (2011) [15].

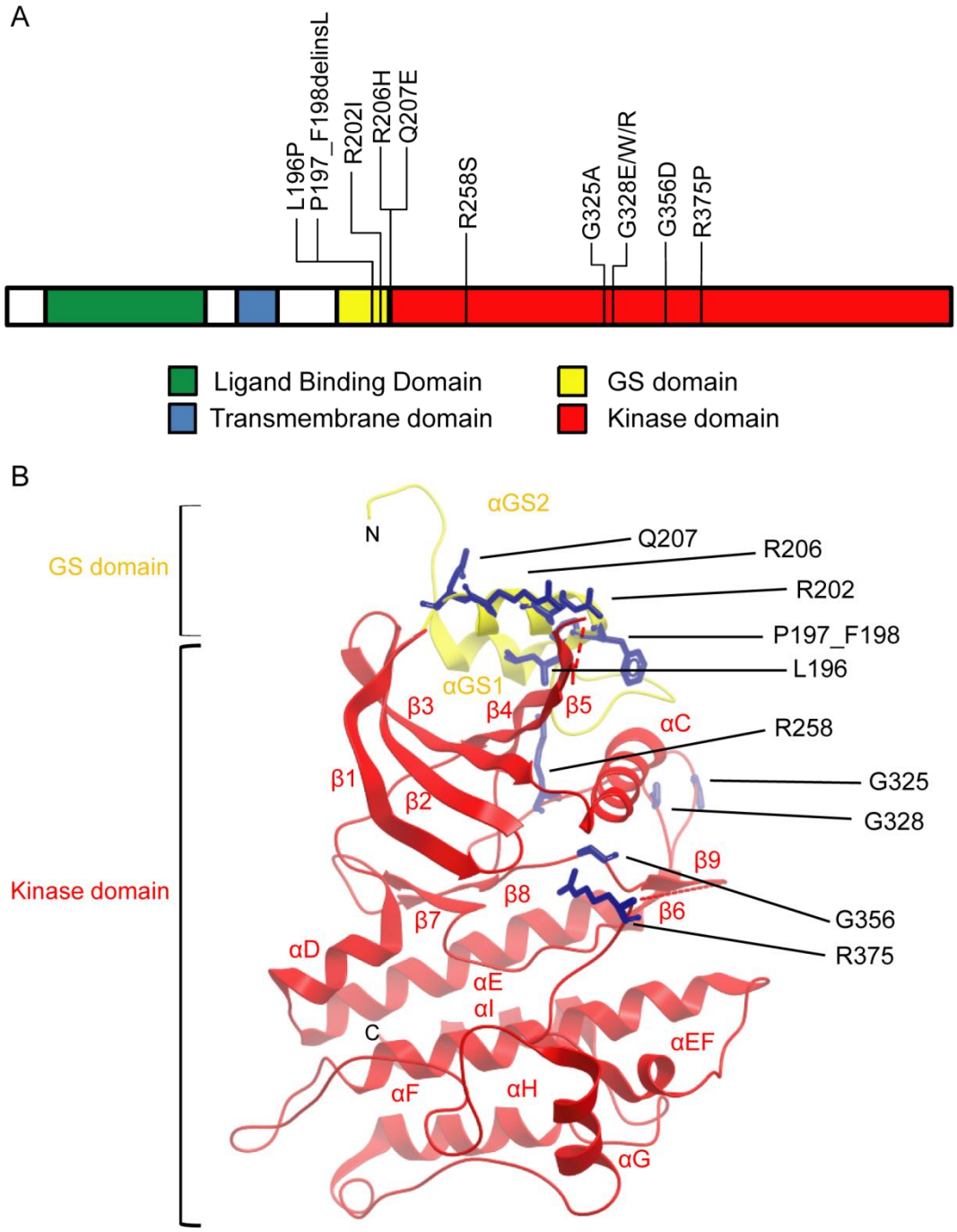


Figure 1.2 The ACVR1 protein. (A) Schematic of the protein domain organization. The various domains of ACVR1 are highlighted and positions of known FOP mutations (listed in Table 1) are indicated. (B) The crystallographic structure of the GS and kinase domains of ACVR1 (pdb 3h9r) showing the known sites of FOP mutations.

1.3 Overview of the ACVR1 signalling pathway

Protein kinases mediate post-translational modifications by phosphorylation and regulate a vast array of signalling processes in the cell [19]. ACVR1 functions in a signalling pathway initiated by BMP ligands of the transforming growth factor beta (TGFβ) superfamily. All BMP/TGFβ .receptors are transmembrane serine/threonine receptor kinases [20,21]. The pathway includes approximately 30 different secreted ligands, 12 different receptors and 5 different substrate Smads. At its simplest, the core canonical pathway can be described by four steps. Ligand binding to the receptor extracellular domains first induces two type I receptors to form a tetrameric complex with two type II receptors. The type II kinase domains then phosphorylate the GS domains of the associated type I receptors [22,23]. This activates the type I receptors promoting their subsequent recruitment and phosphorylation of the receptor-regulated Smads (R-Smads) [24]. Once phosphorylated, R-Smads assemble with Smad4 and translocate to the nucleus where they regulate transcription of target genes such as *Id1* [24] (figure 1.3). Each of these steps is described in more detail in following sections.

Through the receptors BMPs can also stimulate Smad-independent, or non-canonical BMP signalling, most notably through activation of various MAPK pathways [25]. This may occur through receptor binding and autoubiquitination of ubiquitin ligase (E3) TRAF6 for TGFβ receptors [26], or through the binding of X-linked inhibitor of apoptosis protein (XIAP) to the receptor and TGF-β-activated kinase binding protein (TAB1) [27]. This leads to the activation of TGF-β-activated kinase (TAK1) and subsequent phosphorylation of p38 stress-activated protein kinase (figure 1.3) [27].

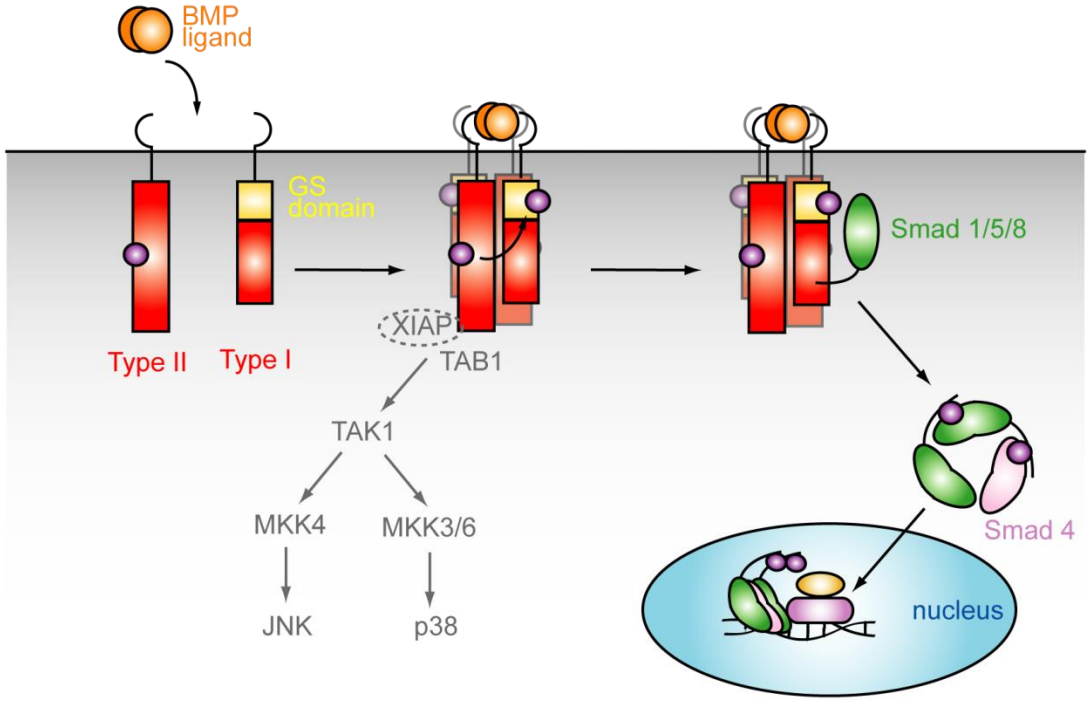


Figure 1.3 BMP signalling pathway. BMPs signal through the Smad1/5/8 signalling pathway. Following ligand binding a tetrameric complex is formed by two type I and two type II receptors. The constitutively active type II receptor phosphorylates the type I receptor (purple circles). R-Smads can then be phosphorylated by the type I receptor. The R-Smad then dissociates from the receptor and forms a trimer with another R-Smad and Smad4 for translocation to the nucleus. Here the Smads bind the DNA at the Smad binding element (SBE) and also transcription factors such as RUNX2 (light purple) and co-activators such as CBP/p300 (light orange). They can also activate various Smad-independent pathways, such as p38 MAPK (grey), however the exact mechanism is unknown. TGFβ ligands induce a similar pathway, but activate Smad2/3.

1.3.1 Bone morphogenetic proteins

The activity of bone morphogenetic proteins (BMPs) was identified by Urist in 1965, reporting on the ability of decalcified human bone matrix to induce bone formation in rabbits [28]. The first four BMPs were then purified and sequenced in 1988 by Wozney et al who identified three (BMP2A, BMP2B, BMP3) of them as part of the TGFβ superfamily by sequence similarity [29]. The fourth, BMP1, was an unrelated molecule that contained an EGF-like domain and was suggested to be a protease involved in

activation of TGFβ1. Since then some 20 BMPs have been identified [30]. BMPs are cytokines synthesized as a dimeric proproteins consisting of an N-terminal signal peptide for secretion, a large prodomain and a C-terminal mature domain containing a cystine-knot motif [29,30]. The proprotein undergoes proteolytic cleavage at an RXXR motif following disulphide bond dimerization. The active mature BMP dimer is then secreted either as a soluble dimer or in complex with the prodomain [30]. BMPs are involved in a wide range of functions, such as body patterning and specification of many organs, as well as the initially identified bone and cartilage formation (figure 1.4). During early embryogenesis, BMPs are involved in axial development [24], cell fate determination and organogenesis [32], and are crucial for normal development and tissue homeostasis. For instance, BMP7 has been shown in mice to be crucial for kidney development, with BMP7 knock-out mice dying shortly after birth due to kidney failure [33,34]. BMP7 is also expressed in the adult kidney. The down regulation of BMP7 under disease conditions results in renal fibrosis, inflammation and apoptosis (figure 1.4) [34]. BMP7 is currently in use in orthopaedic surgery for spinal fusion and fracture healing, however it has the potential to be used therapeutically in renal disease [35].

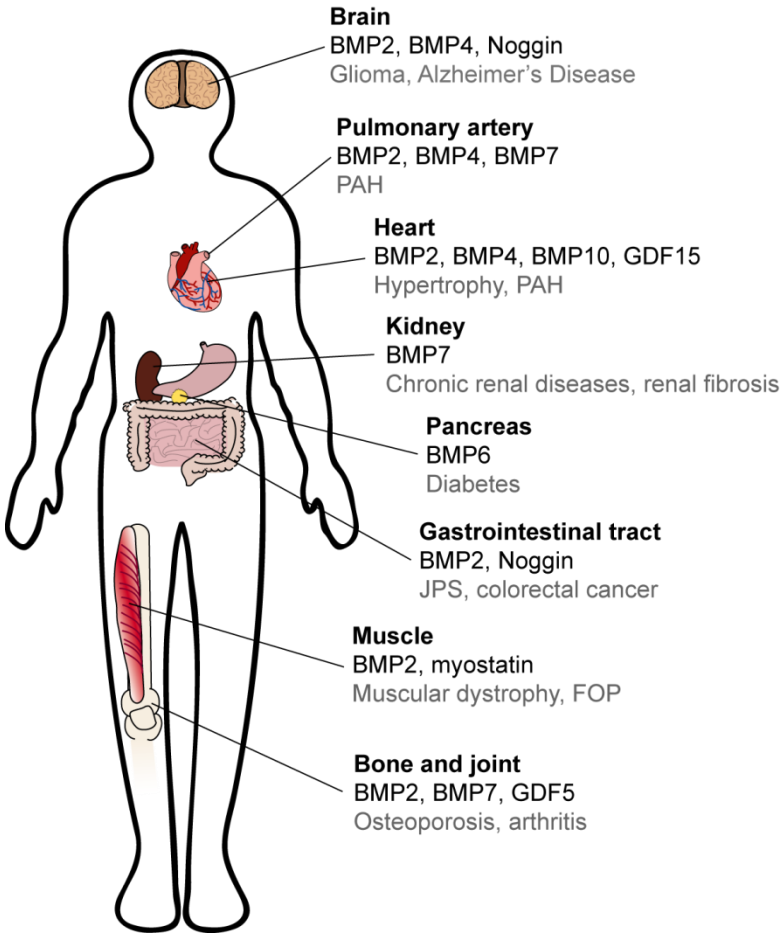


Figure 1.4 Diseases associated with BMPs. The functions of BMPs range from embryonic patterning to regulating many different organs and tissues. Disruption to signalling can result in a wide range of disease (PAH – pulmonary arterial hypertension; JPS – juvenile polyposis syndrome; GDF – growth differention factor). Adapted from Wagner et al (2010) [36]

1.3.2 BMP receptors

Dimeric BMP ligands are recognised by heterotetrameric BMP receptor complexes at the cell membrane which subsequently activate intracellular Smad transcription factors, leading to changes in gene expression [11]. The BMP/TGFβ receptor family consists of seven type I receptors and five type II receptors. The type I receptors, also known as activin receptor-like kinases (ALK1-7), are inactive in the absence of ligand. The type II receptors lack the GS domain and are constitutively active [14]. The various type I and II

receptors have differing binding preferences for the different ligands (Table 1.2), providing some specificity and context-dependent signalling.

Table 1.2 BMP/TGFβ receptors and their ligands.

	Receptor	Ligand
TYPE I	ACVRL1 (ALK1)	TGFβ, BMP9, BMP10
	ACVR1 (ALK2)	BMPs and GDFs
	BMPR1A (ALK3)	BMPs
	ACVR1B (ALK4)	Activins, GDF8/myostatin and GDF11
	TGFBR1 (ALK5)	TGFβ, GDF8/myostatin and GDF11
	BMPR1B (ALK6)	BMPs
	ACVR1C(ALK7)	BMP16/nodal
TYPE II	TGFBR2	TGFβ
	BMPR2	BMPs and GDFs
	ACVR2	Activins, BMPs and GDFs
	ACVR2B	Activins, BMPs, GDFs and BMP16/nodal
	AMHR2	AMH

AMH – anti-Müllerian hormone. Adapted from Drabsch and ten Dijke (2012) [32]

The specificity for BMP/TGFβ ligands is determined by both the type I and type II receptors which assemble as a heteromeric complex, comprising two receptors of each class [25] The ligand-bound extracellular domain complex has been solved for both the BMP and TGFβ receptors (figure 1.4). These structures showed stark differences between the binding orientation of the two types of receptor, however the interaction of the cytoplasmic domain has yet to be defined (figure 1.5). ACVR1 (also known as ALK2) has been shown to mediate the signalling of BMP2/4/6/7/9 through interactions with

various type II receptors, including BMPR2, ACVR2A and ACVR2B [37]. Both receptor classes are required for signalling.

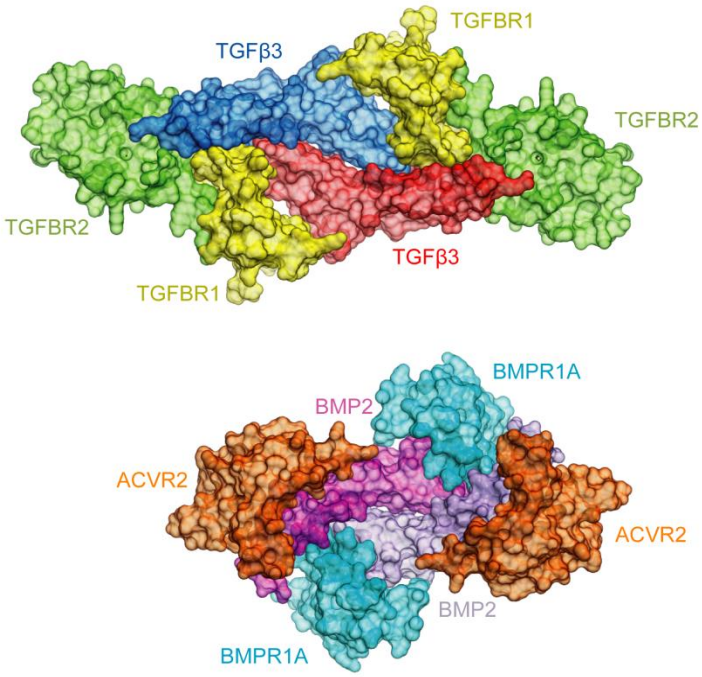


Figure 1.5 Extracellular domain of BMP/TGFβ receptors. Surface representation of example TGFβ (top) and BMP (bottom) complexes with extracellular receptor domains. The ligand dimer binds the heterotetrameric complex of two type I and type II receptors. Different sequence insertions in the BMP/TGFβ receptors provide distinct binding modes for the different ligands [38], [39]. Adapted from Groppe et al (2008).

Due to the variety of cellular functions controlled by BMP/TGFβ signalling it is unsurprising that dysregulation of these signalling pathways results in disease and many cancers [40]. Mutations in the BMP/TGFβ receptors have been identified in many degenerative diseases. As well as the activating mutations found in FOP, loss of function of ACVR1 has been associated with congenital heart defects [41]. Loss of function mutations in other type I BMP receptors have also been identified; for example BMPR1B mutations in A2 brachydactyly [42,43], BMPR1A mutations in juvenile polyposis

syndrome (JPS) [44] and ACVRL1 mutations in hereditary hemorrhagic telangiectasia (HHT) [45,46]. Mutations in the type II receptors have also been identified, such as BMPR2 in pulmonary arterial hypertension (PAH) (figure 1.4) [47]. To date, FOP is the only example of mutant BMP/TGFβ receptor gain of function.

1.3.3 Smads

Smad proteins are sequence-specific transcription factors. Mutational activation of ACVR1 in FOP stimulates the Smad pathway in a ligand-independent manner [48]. In total, there are 3 different classes of Smad, the receptor-regulated Smads (R-Smads, Smads 1/2/3/5/8), the common-partner Smad (co-Smad, Smad4) and the inhibitory Smads (I-Smads, Smad6/7) [49,50] (figure 1.6A).

Structurally they consist of highly conserved N and C-termini containing Mad homology domain 1 and 2 (MH1 and MH2) respectively, linked by a divergent proline-rich linker (figure 1.6A). The MH1 domain facilitates DNA binding in the nucleus [49,50] and mediates further interaction with other transcription factors. The MH2 domain is involved in oligomer formation, accessory protein interactions, interactions with transcription factors, and, in the case of R-Smads, receptor interactions.

The R-Smads are divided by the type I receptor they interact with. TGFβ receptors predominately phosphorylate Smad2/3 but can also phosphorylate Smad1/5 [22,50,53], whereas BMP receptors generally phosphorylate Smad1/5/8 [25,30,52]. The specificity for R-Smad binding is determined by the sequence in the kinase β4-β5 loop (commonly known as the L45 loop) [14,55,56,57], which is recognized by the R-Smad L3 loop [58] in the MH2 domain, although the complex structure is not known. The C-terminus of the R-Smads contains a SSXS motif which is phosphorylated on the last two serines by the type I receptors [52,59]. The site of phosphorylation was determined previously by

Kretzschmar *et al* by mutation of the serines in the SSXS sequence [54]. The R-Smads share a common fold with the well-characterized Forkhead associated (FHA) domain [60]. Sequence alignments show the conservation of a basic patch, corresponding to the known phosphoryl binding site in the FHA domain, that is likely to interact with the phosphorylated GS loop [61].

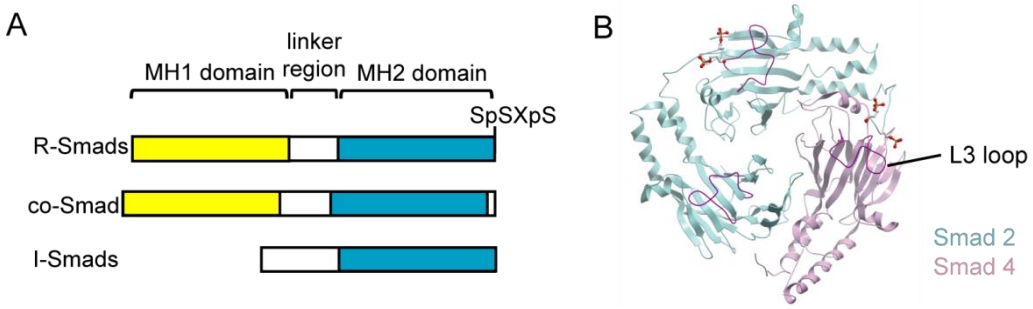


Figure 1.6 Smads (A) Schematic of the three classes of Smad, highlighting the MH1 domain (yellow) and MH2 domain (blue). The R-Smads (Smad1/2/3/5/8) uniquely contain a SSXS motif which is phosphorylated by the BMP/TGFβ type I receptor. I-Smads (Smad6/7) lack the MH1 domain. (B) Crystallographic structure of phosphorylated Smad2/Smad4 heterotrimer (pdb: 1u7v). The L3 loop (purple) interacts with the SpSXpS motif of the adjacent Smad molecule. The same loop is predicted to bind the type I receptors through their phosphorylated GS domains.

Smad4 (co-Smad) is a mediator subunit common to both Smad1/5/8 and Smad2/3 complexes. While Smad4 lacks the SSXS motif, it comprises an MH2 domain that interacts with the phosphorylated R-Smad for signal transduction (figure 1.6B) [54]. Prior to receptor activation, the R-Smads exist as monomers in the cytoplasm [62]. On phosphorylation, the SSXS motif of one Smad monomer interacts with the L3 loop of another R-Smad or Smad4. Conformational changes [21,63] lead to the formation of either an R-Smad homotrimer, or a heterotrimer consisting of 2 R-Smads and Smad4 [49,62] (figure 1.6B). The formation of the trimer was shown in size exclusion chromatography studies using both recombinant protein [64] and COS7 cells [62]. The molecular understanding of the trimer was then provided by various structural studies

[59,61] The heterotrimer translocates to the nucleus, where it activates transcription by binding to Smad binding elements (SBE) in gene promoters and interacts with other DNA-binding transcription factors [47]. Translocation of Smads as heterotrimers to the nucleus in response to TGFβ signalling was shown by the reduced nuclear mobility of the Smads in HeLa cells using fluorescence recovery after photobleaching (FRAP) [66], and *in vitro* import nuclear assays, again in HeLa cells, which demonstrated transport of the heterotrimer dependent on direct contact between the R-Smads and nucleoporins [63,64].

1.3.3.1 Smad accessory proteins

Smad phosphorylation is thought to continue while the receptor-Smad complex is internalised through endocytosis. In TGFβ signalling, Smad2/3 are localized to endosomes by a FYVE domain-containing protein known as Smad Anchor for Receptor Activation (SARA), shown in previous studies by immunoprecipitation of transfected COS cells and immunofluorescence in Mv1Lu cells [69]. This 1425-residue protein contains a Smad-binding domain that blocks premature formation of the Smad trimer as well as a receptor binding domain that may further stabilise the receptor-Smad complex [69]. SARA is specific for Smad2/3 and does not bind the BMP receptor associated Smad1/5/8 [70]. A possible alternative to SARA for BMP-regulated Smads is a closely related FYVE domain protein known as endofin [60,61]. Further studies in COS1 cells and protein-fragment complementation assays using yellow fluorescent protein showed endofin binds preferentially to Smad1 [71]. Endofin also recruits protein phosphatase 1 (PP1), potentially providing a dephosphorylation mechanism for subsequent deactivation of the type I receptor [71].

1.3.4 Regulation of BMP signalling

Receptor activation is regulated by multiple mechanisms at the levels of ligand, receptor and Smad. In BMP signalling, noggin acts as an extracellular antagonist that binds to the BMP ligand to prevent its interaction with receptor [6,11]. Overexpression of noggin has been shown to result in osteoporosis in mice [11]. Other extracellular antagonists include chordin and follistatin [29,62]. These extracellular antagonists cannot be used against FOP, however, as the mutational activation of the intracellular kinase domain is not reliant on binding of BMP ligand.

The immunophilin FKBP12 binds to the cytoplasmic GS domain of the type I receptors[56,74,75] to inhibit its ligand-independent phosphorylation by the type II receptor [10]. FKBP12 is an abundant 12 kDa rapamycin and FK506 binding protein [74, 76,77]. The inhibitory effect of FKBP12 on phosphorylating the type I has been previously shown using TGFBR1 defective cells cotransfected with TGFBR1 and TGFBR2 either with or without FKBP12. Here they showed reduced phosphorylation of TGFBR1 in the presence of FKBP12 which could be restored by the presence of rapamycin [56] On BMP/TGFβ ligand binding, FKBP12 is released [56,74] and the type I receptor is phosphorylated.

Finally, both the inhibitory Smads (Smad6/7) and the E3 ligases Smad ubiquitylation regulatory factors (Smurf1/2), downregulate R-Smad activation [67,68]. Expression of the I-Smads is induced by Smad signalling, forming an autoinhibitory feedback system [69]. Smad6/7 compete with the R-Smads to form interactions with the type I receptors. Smad7 acts as an inhibitor for both BMP and TGFβ receptors whilst Smad6 selectively inhibits the BMP receptors [73]. The receptor bound Smad6/7 associate with Smurfs 1/2 for ubiquitin-dependent degradation of the receptor [80].

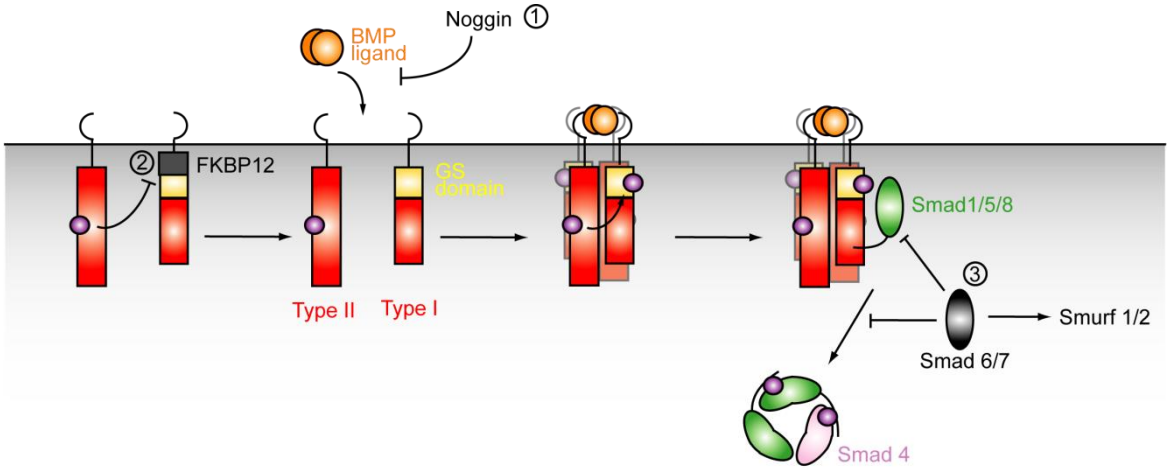


Figure 1.7 Regulation of BMP signalling. BMP signalling can be regulated at various stages in the pathway. The colour coding is the same as in figure 1.3, with the regulatory proteins shown in black. (1) Noggin and other antagonists bind BMP ligand and inhibit interaction between the ligand and the receptor. (2) FKBP12 binds the type I receptor to prevent ligand-independent phosphorylation and activation by the type II. (3) Smad6/7 associate with the receptor to prevent R-Smad binding and phosphorylation, leading to degradation by Smurfs 1/2.

1.4 Receptor kinase activation in FOP

BMP/TGFβ type I receptors are normally inactive in the absence of BMP ligand. The inactive state of these receptors is controlled by both the GS domain and the conformation of the catalytic kinase domain. For activation, the GS loop needs to be exposed to allow transphosphorylation by the type II receptor. This modification may allow both the recruitment of the R-Smad MH2 and the correct positioning of the kinase activation loop and the αC helix to enable ATP binding and substrate (R-Smad SSXS) phosphorylation.

1.4.1 Kinase activation

Structurally, the kinase domain fold comprises two distinct lobes, the N-terminal lobe and the C-terminal lobe: the N-lobe is formed primarily of β-sheet with a single α-helix

(known as αC), whereas the C-lobe is formed primarily of α-helices [19]. The two lobes are joined by a linker known as the hinge region. The interface between the two lobes forms the site of ATP binding. The N-lobe contains a glycine-rich loop (a GXGXΦG motif, where Φ is usually a tyrosine or phenylalanine; in ACVR1 this is a tyrosine (Y219)) that coordinates the phosphates of ATP and is therefore also known as the phosphate-binding loop, or P-loop [13,81]. An essential conserved lysine residue in the β3 strand, known as the catalytic lysine, additionally coordinates the α and β phosphates, and in the active configuration forms a strictly conserved ion pair with a glutamate in the αC helix. The correct positioning of the αC helix is therefore essential for kinase activity. The active state is further stabilised by interactions between the αC helix and the activation loop (A-loop). For many kinases these interactions require an activating phosphorylation in the A-loop, but this is not observed within the BMP/TGFβ receptor family. The A-loop extends from a generally conserved DFG motif (in ACVR1 this is DLG), the aspartate of which forms polar interactions with magnesium ions, to the APE motif. In its active conformation the A-loop contributes part of the docking site for the substrate. In the catalytic loop, a conserved HRD motif coordinates both the substrate and the aspartate of the DFG motif (figure 1.8) [13,19,81,82].

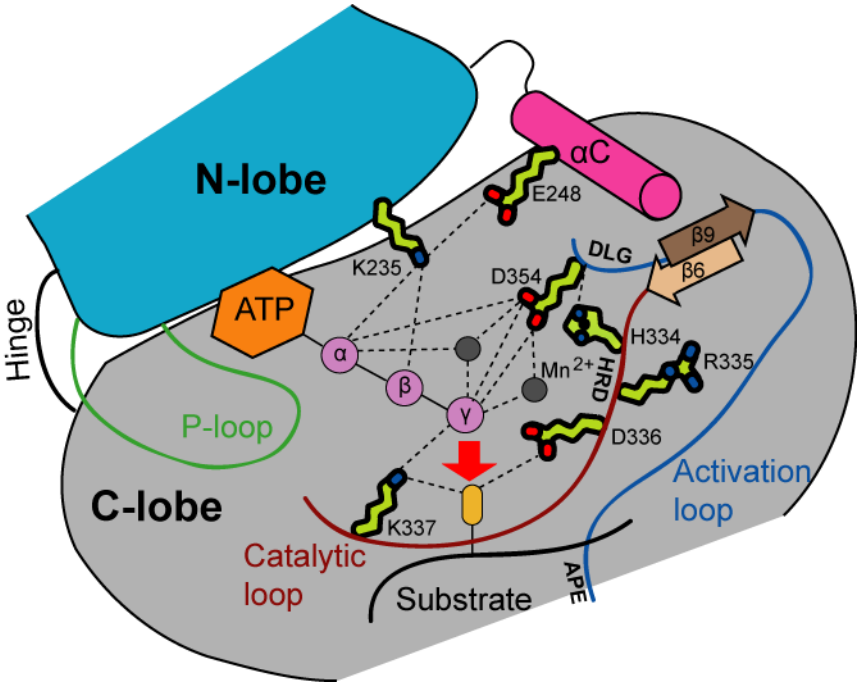


Figure 1.8 Diagram of conserved interactions within ACVR1. Diagram of known interactions between the protein kinase catalytic core, ATP and a substrate (ACVR1 numbering is used). Red arrow indicates the catalysed transfer of the ATP γ-phosphate to the hydroxyl group of a protein substrate. Important polar contacts are shown by dashed lines. Adapted from Kornev et al (2006) [82].

1.4.2 Structural insights into BMP/TGFβ receptor kinase activation

A complete structural description of receptor activation is lacking since there are no structures of the phosphorylated type I receptor, the receptor-Smad complex, or the intracellular assembly of the type I-type II receptor complex [30]. The current preferred model was proposed following the structure determination of the type I TGFβ receptor TGFBR1 (ALK5) (figure 1.9). The first structure of the cytoplasmic domain of TGFBR1 was solved in 1999 in complex with FKBP12 in the absence of inhibitor [83]. This revealed an inactive kinase conformation with FKBP12 bound directly to the Leu-195 and Leu-196 pair in the GS domain and also formed interactions with the αC in the N-lobe of the kinase domain [83]. FKBP12 binding locked the GS loop in a buried position between

the αC and β4 that is inaccessible to the type II receptor thus preventing its transphosphorylation. The insertion of the GS loop into the kinase N-lobe also affected the conformation of the catalytic domain [83]. The correct alignment of the αC helix was blocked as the GS loop pushed the αC C-terminus outwards. The αC N-terminus was conversely pushed in and blocked ATP binding [82,84]. Subsequently, the TGFBR1 structure was solved in the absence of FKBP12, providing insight into the potential activation mechanism of TGFβ receptors [76]. This structure was solved in presence of inhibitor NPC-30345 which, although the compound structure has not been released, was reported to bind in the ATP pocket of the kinase domain [76]. The absence of FKBP12 allowed movement of the GS domain away from the kinase N-lobe, thereby promoting potential GS phosphorylation. There was also a slight shift of both the αC and the activation segment [83] towards the active kinase conformation (figure 1.9). This structure, however, does not represent the fully active kinase as the site of inhibitor binding would be incompatible with the complete conformational change required in activation. Therefore the structure of a fully active BMP/TGFβ type I receptor is still unknown.

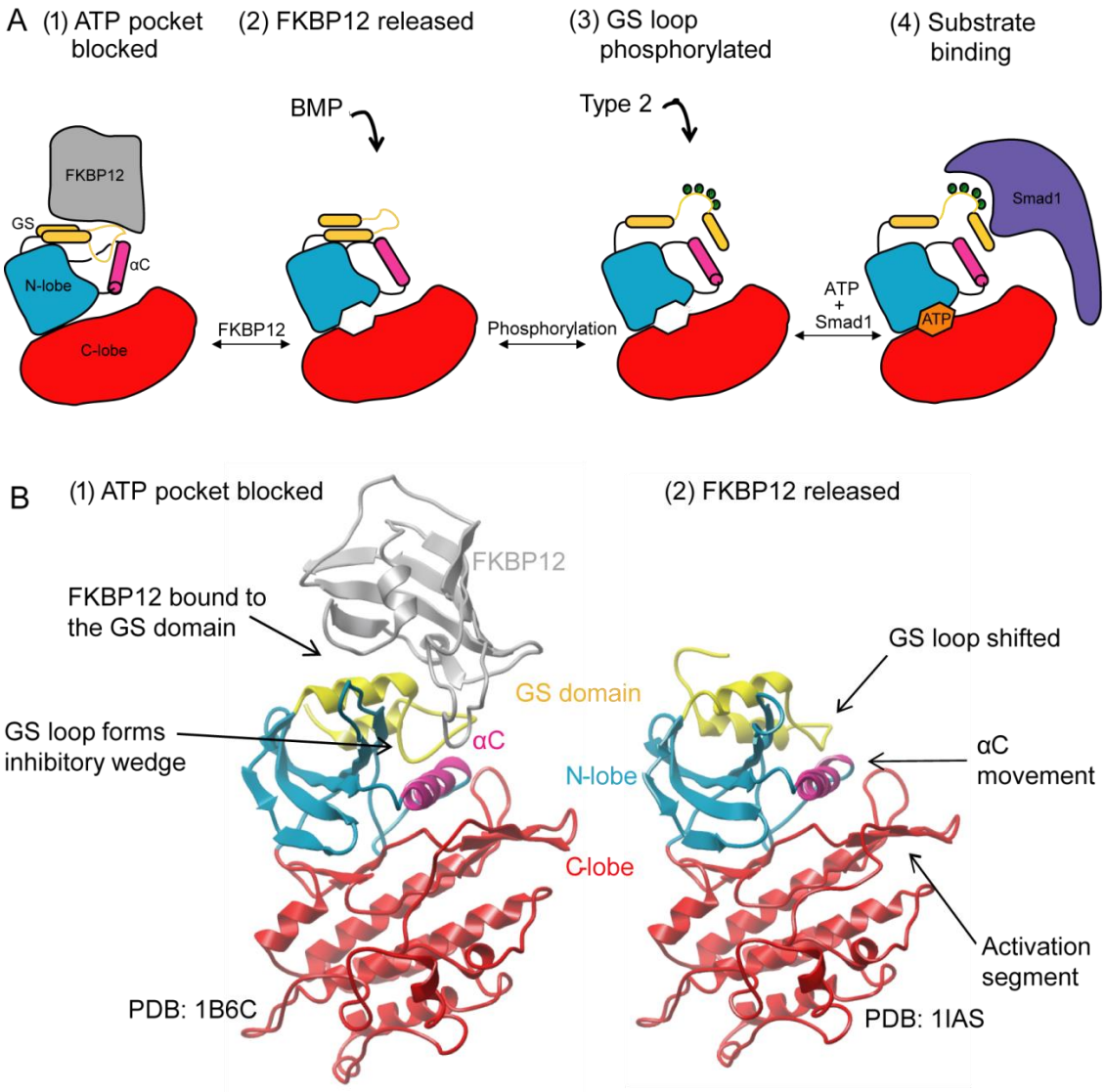


Figure 1.9 Structural insights to receptor activation. (A) Cartoon model of activation (1) FKBP12 binding maintains an inhibited form of the type I receptor. (2) Upon ligand binding to the extracellular domain, FKBP12 is released [22] freeing the GS loop for transphosphorylation (3) and allowing the movement of the αC to free the ATP pocket. (4) ATP and the substrate Smad can now bind to the type I receptor [65,73]. Adapted from Huse et al. (2001) [76] (B) Crystal structures of TGFBR1 in the (1) presence and (2) absence of FKBP12 (equivalent to cartoons (A) 1 and 2, respectively).

1.4.3 FOP-associated mutations promote ligand-independent receptor activation

The mutations in FOP result in mild activation of ACVR1 which induces Smad1/5/8 signalling in the absence of BMP ligand. This has been demonstrated for the R206H mutant in C2C12 myoblast cells [85,86]. Previous studies of cotransfected COS-7 cells have shown that FKBP12 has reduced binding to ACVR1 R206H in comparison to the wild-type receptor [48]. This decrease in regulatory FKBP12 binding may partly explain the mild activation of the receptor.

The crystal structure of the GS and kinase domains of ACVR1 in complex with FKBP12 was solved in 2009, giving structural insights into the possible effects of FOP mutations (figure 1.10). Mutations in the GS domain, such as R206H, are predicted to perturb the inhibitory FKBP12 binding, whereas mutations in the kinase domain, which are mainly clustered around the regulatory sites for activation, are predicted to facilitate the binding of ATP [87]. Previous studies on TGFBR1 have shown that FKBP12 and ATP binding are mutually exclusive demonstrating that structural changes can be relayed throughout the catalytic domain [76].

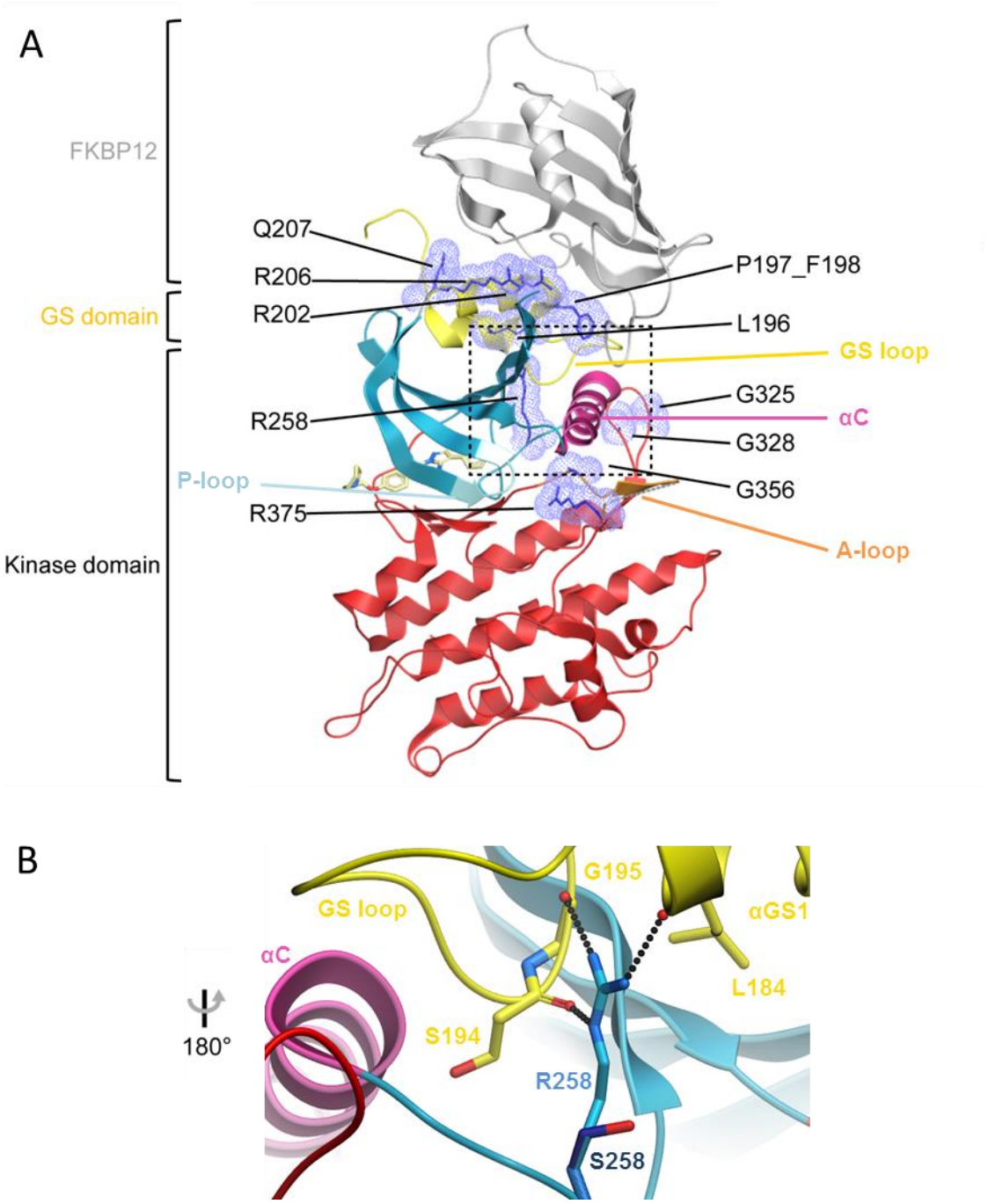


Figure 1.10 The structure of ACVR1 bound to FKBP12 (pdb: 3H9R). (A) FKBP12 (grey) in complex with the GS (yellow) and kinase domains of ACVR1. The known inhibitor dorsomorphin is shown bound to the hinge region of ACVR1. Key structural elements of the kinase domain are highlighted, the N-lobe (blue), C-lobe (red), αC (pink), the activation loop (A-loop: orange) and the glycine rich phosphate binding loop (P-loop: light blue). The side chains of FOP-associated mutations are also displayed in dark blue, clustered around the ATP pocket or the FKBP12 binding site. (B) Example of potential effects of FOP mutation. The hydrogen bonds between the wild-type ACVR1 R258 (yellow) and the GS loop are absent in FOP mutation S258 (dark blue) [87].

1.5 First inhibitors of ACVR1 signalling

Surgical methods to release joints fixed by heterotopic ossification in FOP are ineffective as they simply precipitate further extensive ossification via activation of inflammatory mechanisms. Consequently, the only present treatment used clinically for FOP is the use of corticosteroids and other anti-inflammatory drugs, such as ibuprofen. While these drugs reduce flare ups that occur spontaneously or after trauma, they are unable to alter the general and inevitable progression of the disease [88]. Since ACVR1 mutants are activated independently of extracellular BMP ligand neither the natural antagonist noggin nor antibodies can be used therapeutically. An alternative potential strategy to combat FOP is to use chemical inhibitors that are able to gain access to the intracellular ACVR1 kinase domain [2].

The use of small molecule compounds as kinase inhibitors is an established route for drug discovery, particularly in cancers [81,89,90]. Many kinase inhibitors have been designed to bind competitively to the ATP site, either in the kinase active or inactive conformation, or occasionally allosteric sites [90]. The design of highly specific compounds is challenging due to the high conservation of the kinase ATP pocket. As a result, most inhibitors target multiple kinases which can induce unwanted off-target effects [89]. The first reported ACVR1 inhibitor was identified from a phenotypic screen in which BMP inhibitors were identified by their ability to dorsalize zebrafish embryos, detected by the absence of the ventral tail fin [78] (figure 1.11). This led to the discovery of dorsomorphin - a compound previously reported as an inhibitor of AMP-activated kinase (AMPK) and originally referred to as compound C [91] (figure 1.12).

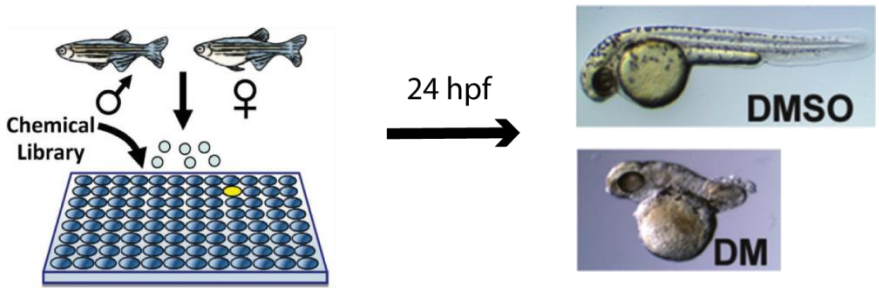


Figure 1.11 Schematic of the high throughput phenotypic screen. The screen was used to identify dorsomorphin (DM), showing the dorsalized phenotype compared to normal (DMSO) embryo development (hpf – hours post fertilization. Adapted from Hong and Yu (2009) [78].

Dorsomorphin is structurally related to pyrazolo[1,5-a]pyrimidine compounds known to inhibit the receptor tyrosine kinase KDR. Indeed, in vitro assays for KDR and AMPK showed IC_{50}'s of 25 nM and 23 nM respectively [78,92]. To confirm the phenotype observed was caused by inhibition of the BMP pathway, embryos were treated with other known (structurally unrelated) inhibitors of KDR/AMPK and did not produce the same phenotype [15]. The concentration of dorsomorphin required to inhibit Smad1/5/8 signalling was lower than that required for inhibition of TGFBR1 Smad2 activation, providing specificity for the Smad1/5/8 pathway [91]. Dorsomorphin also showed apparent selectivity for the BMP type I receptors over BMP type II receptors [91]. Although initially thought to only affect the Smad-dependent pathway [12] subsequent studies have shown additional inhibition of non-Smad receptor signalling, including the p38 MAPK, ERK1/2 and Akt pathways [93].

A structure–activity relationship (SAR) was subsequently performed to optimize dorsomorphin for potency and metabolic stability. This led to the identification of LDN-193189 (figure 1.12), which has a cellular IC_{50} of 5 nM that is 100-fold improved relative to dorsomorphin [78]. This new compound has shown some efficacy in a mouse model for

FOP created by using a transgene of constitutive active ACVR1 Q207D [78,91]. More recently a knock-in R206H mouse, which displays the classic FOP phenotype has been developed [95], that could be used as a further model for testing inhibitors.

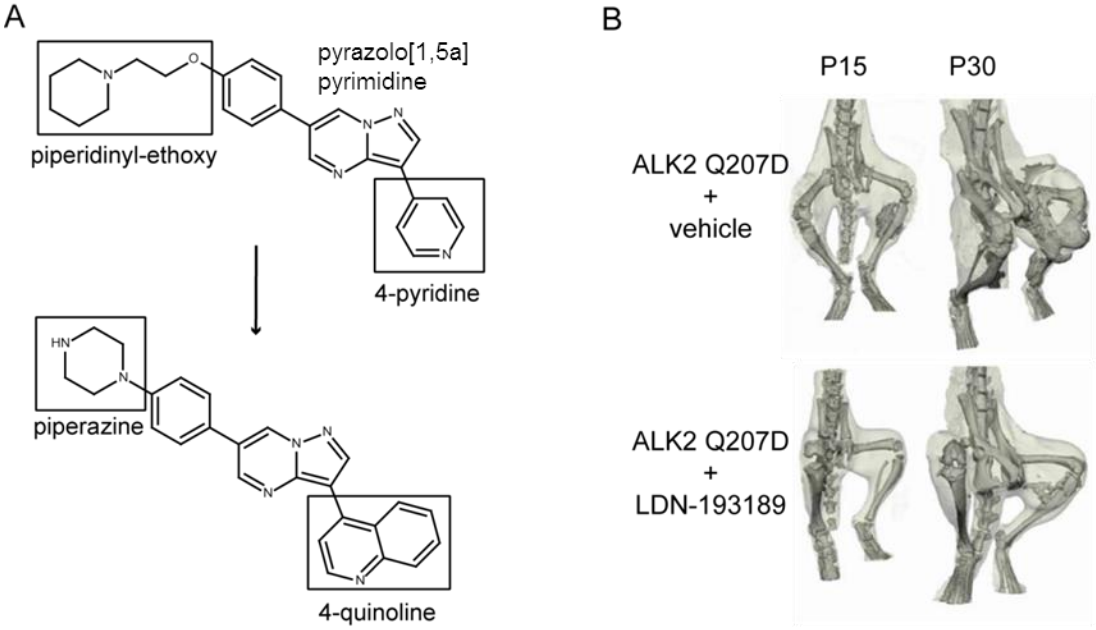


Figure 1.12 A kinase inhibitor reduces heterotopic ossification. (A) The optimization of dorsomorphin to LDN-193189 around the core pyrazolo[1,5a]pyrimidine group. Potency was increased by extending the 4-pyridine group to a 4-quinoline group. Metabolic stability was improved by subsituting the piperidinyl-ethoxy with a piperizine group [96]. (B) LDN-193189 (3 mg/kg) prevented ossification in the hindlimb of an ACVR1 Q207D transgenic mice when analysed 15 and 30 days after induction by ade. Cre [78]. Reproduced with permission from Hong and Yu (2009) [78].

1.6 Aims of the thesis

There are two main aims of the thesis. The first is to gain a better understanding of the disease mechanism in FOP and in particular to investigate how FOP-associated mutations activate the ACVR1 kinase to induce constitutive Smad signalling. While gain of function has been demonstrated in cells, there is little characterisation of the recombinant ACVR1

protein or its mutants. It is of particular interest to address the following questions: whether the mutant ACVR1 protein has lost FKBP12 binding and inhibition? If the mutant ACVR1 protein can phosphorylate Smad and, if so, does this occur in the absence of GS domain phosphorylation? Do the mutants still require transphosphorylation by the type II receptor or is mutant activation a result of autophosphorylation? It is also of particular interest to determine whether there are any structural changes between the mutant and wild-type ACVR1. It would also be of interest to compare structures of related BMP/TGFβ receptors for insights into activation.

Preparation of the recombinant proteins requires a baculoviral expression system due to insoluble expression in *E. coli*. Differences in the activation states of the wild type and mutants will be assessed by their binding to FKBP12, for example by analytical gel filtration, and by phosphorylation assays recorded by mass spectrometry. Conformational changes that occur as a result of mutation, as well as differences between ACVR1 and the BMP receptor BMPR1B, will be addressed by using x-ray crystallography.

The second aim is to advance the development of small molecule inhibitors of ACVR1. It remains of interest to determine the binding mode of the hit compound LDN-193189. Such analysis may provide further avenues for scaffold development. In addition, it is of significant interest to assess whether there are other alternative scaffolds that may confer greater specificity for ACVR1 inhibition. Further structure–activity relationship (SAR) studies are also desirable to identify those chemical groups that confer compound potency as well as specificity. Compound library screening will be performed by using a fluorescence-based thermal shift assay and hits confirmed by isothermal titration calorimetry. The mode of compound binding will be determined by x-ray crystallography.

2. Materials and Methods

2.1 Construct Design

This thesis focuses on the cytoplasmic domain of the BMP-receptor ACVR1, and mutations of this receptor that have been identified in FOP. Two main constructs of this protein were used, containing either the kinase and GS domains or the kinase domain only. All of the BMP/TGFβ receptor constructs, with the exception of BMPR2, were cloned into the pFB-LIC-Bse vector for baculoviral expression (figure 2.1; table 2.1). Constructs of FKBP12, Smad1, Smad2, endofin and SARA were provided by the biotechnology group at SGC for bacterial expression. FKBP12, Smad1 and Smad2 had been cloned into the pGTVL2 vector containing a T7 promoter and both His_6 and GST tag at the N-terminal fusion, followed by TEV protease cleavage site. Endofin and SARA had been cloned in to the pCDF-LIC vector, which will be explained in chapter 4.2.5.1.1. Endofin had been cloned into the pNIC-CTH0 vector containing a T7 promoter and a C-terminal His_6 tag, preceded by a TEV protease cleavage site (table 2.2).

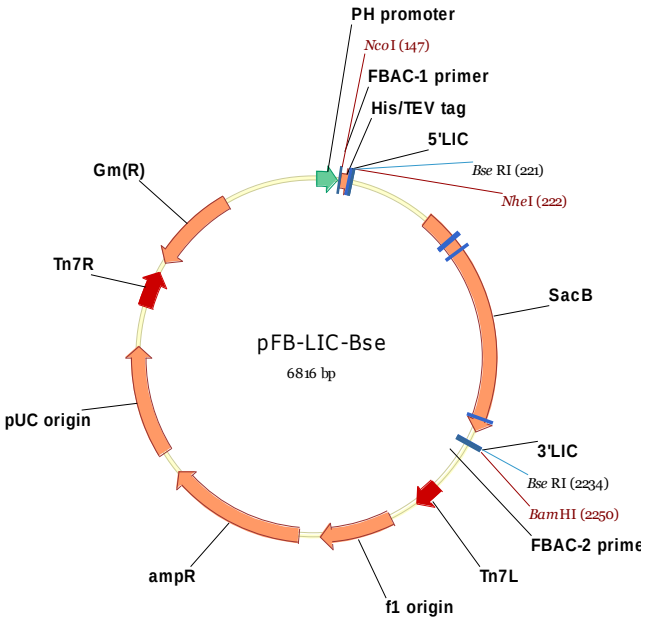


Figure 2.1 The pFB-LIC-Bse vector. Baculovirus transfer vector with His_6 tag in 22-aa N-terminal fusion peptide, with TEV protease cleavage site. Includes sites for ligation independent cloning, and a "stuffer" fragment that includes the SacB gene, allowing negative selection of transformed bacteria on 5% sucrose.

Table 2.1.Table of baculoviral constructs

Vector	Protein	Contruct (a.a.)
pFB-LIC-Bse	ACVR1	172 - 499
	ACVR1 (L196P)	172- 509
	ACVR1 (R206H)	172- 499
		201 - 499
	ACVR1 (Q207D)	172 - 499
		201 - 499
	ACVR1 (R258S)	172 - 509
	ACVR1 (G348E)	172- 509
	ACVR1 (R375P)	172 - 509
	ACVR1B	200 - 505
	ACVRL1	195 - 497
	BMPR1A	198 - 532
	BMPR1B	168 - 502
	TGFBR1	162 - 503
		198 - 503
	ACVR2	191 - 488
	TGFBR2	237 - 547

Table 2.2. Table of bacterial constructs.

Vector	Protein	Contruct (a.a.)
pGTVL2	BMPR2	196 - 517
	FKBP12	1 - 108
	Smad1	247 - 465
	Smad2	239 - 467
pNIC-CTH0	Endofin	725 - 862
pCDF-LIC	Endofin	808 - 855
	SARA	767 - 824

These were generated by members of the SGC Biotechnology group.

2.2 Mutagenesis

Mutants identified in FOP as well as the constitutively active Q207D mutation were generated in ACVR1. This was done using either site directed mutagenesis or using the mega primer method. The same primers were used for both methods to insert the mutation. SGC entry clone ACVR1A-s004, the full length UniProt sequence, was used as a template for mutagenesis. Each mutant ACVR1 was inserted into the pcDNA3 vecor, sequenced and then used as a template for subsequent ligation independent cloning (LIC) into the pFB-LIC-Bse vector.

2.2.1 Primer design

The primers were designed using the following equation found in the QuikChange II XL Site-Directed Mutagenesis protocol:

$$Tm = 81.5 + 0.41(\%\ GC) - 675/N - \%\ mismatch \qquad \text{(equation 2.1)}$$

where N = total number of bases.

Table 2.3 Mutagenesis primers.

Mutation	Base change	Primer sequence		Tm (°C)	GC (%)
L196P	587T>C	Forward:	GGCTCTGGTCCTCCTTTTCTGGTACAAAGAACAG	79.2	50.0
		Reverse:	CTGTTCTTTGTACCAGAAAAGGAGGACCAGAGCC		
P197_F198L	590_592 delCTT	Forward:	CAGGAAGTGGCTCTGGTCTTCTTCTGGTACAAAGAACAGTGGCTCGC CAG	78.1	54.0
		Reverse:	CTGGCGAGCCACTGTTCTTTGTACCAGAAGAAGACCAGAGCCACTTC CTG		
R202I	605G>T	Forward:	GGTCTTCCTTTTCTGGTACAAATAACAGTGGCTCGCC	80.5	48.6
		Reverse:	GGCGAGCCACTGTTATTTGTACCAGAAAAGGAAGACC		
R206H	617G>A	Forward:	CAAAGAACAGTGGCTCACCAGATTACACTGTTGGAGTG	80.5	47.4
		Reverse:	CACTCCAACAGTGTAATCTGGTGAGCCACTGTTCTTTG		
Q207D	CAG>GAC	Forward:	CAAAGAACAGTGGCTCGCGACATTACACTGTTGGAGTG	79.0	50.0
		Reverse:	CACTCCAACAGTGTAATGTCGCGAGCCACTGTTCTTTG		
Q207E	619C>G	Forward:	CAAAGAACAGTGGCTCGCGAGATTACACTGTTGGAGTG	81.6	50.0
		Reverse:	CACTCCAACAGTGTAATCTCGCGAGCCACTGTTCTTTG		
R258S	774G>C	Forward:	GTACAACACTGTGATGCTGAGCCATGAAAATATCTTAGG	78.4	41.0
		Reverse:	CCTAAGATATTTTCATGGCTCAGCATCACAGTGTTGTAC		
G328E	983G>A	Forward:	GATATTTGGGACCCAAGAGAAACCAGCCATTGCCC	80.1	51.4
		Reverse:	GGGCAATGGCTGGTTTCTCTTGGGTCCCAAATATC		
G328R	982G>A	Forward:	GATATTTGGGACCCAAAGGAAACCAGCCATTGCCC	80.4	51.4
		Reverse:	GGGCAATGGCTGGTTTCCTTTGGGTCCCAAATATC		
G328W	982G>T	Forward:	GATATTTGGGACCCAATGGAAACCAGCCATTGCCC	80.4	51.4
		Reverse:	GGGCAATGGCTGGTTTCCATTGGGTCCCAAATATC		
G356D	1067G>A	Forward:	GCATAGCAGATTTGGACCTGGCAGTCATGCATTCCC	81.5	52.8
		Reverse:	GGGAATGCATGACTGCCAGGTCCAAATCTGCTATGC		
R375P	1124G>C	Forward:	GGGAACAATCCCCCTGTGGGCACCAAGCG	81.6	65.5
		Reverse:	CGCTTGGTGCCCACAGGGGGATTGTTCCC		

2.2.2 Mutagenesis by mega-primer PCR

Primers were designed to match both ends of the template ACVR1 sequence. These primers also contained ligation independent cloning (LIC) extensions for insertion into a pc-DNA vector to be used for sequencing and by other group members. LIC will be explained further in section 2.3. The method requires two rounds of polymerase chain reaction (PCR). The first round of PCR was used to generate a mega primer from the forward mutant primer to the 3’ end of the sequence primer. The second round used this mega reverse primer along with the 5’ primer (figure 2.2). The PCR products from each reaction were purified using a Qiagen PCR purification kit. ACVR1 Q207E, G328R, G328W, R375P, R258S, L196P mutants were generated using this method. Wild-type ACVR1 was also inserted into the pc-DNA and used as a template for generating the remaining mutants via site-directed mutagenesis.

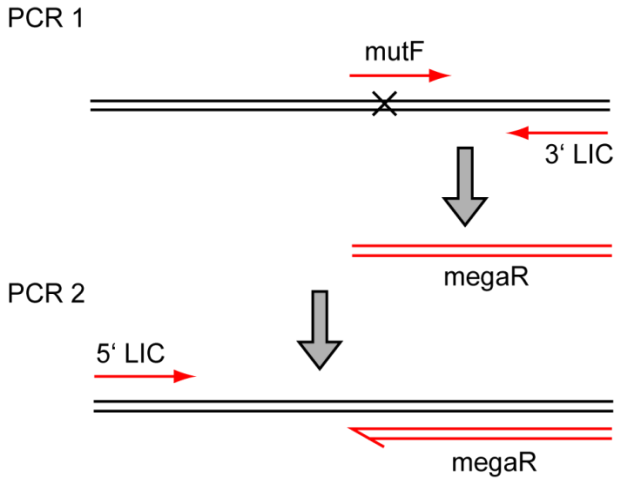


Figure 2.2 Schematic of Mega primer method. PCR1 amplifies a section of the DNA sequence, from the site of mutation (x) to the 3’ end terminus. PCR2 then uses the PCR product from PCR1 as a 3’ primer to amplify the full length of the DNA sequence required.

The PCR protocols were as follows:

1st PCR reaction - to form the MegaR primers

10 x Pfx buffer	5.0 µL
10 x PCR enhancer	5.0 µL
50 mM $MgSO_4$	1.0 µL
10 mM dNTP mix	1.5 µL
Platinum Pfx	0.4 µL
Template DNA (10 ng/mL)	1.25 µL
Forward primer (5 µM)	3.0 µL
Reverse primer (5 µM)	3.0 µL
Water	29.85 µL

Thermocycler conditions:

94 °C, 5 min

(94 °C, 30 sec; 55 °C, 30 sec; 68 °C, 30 sec) x 30 cycles

72 °C, 7 min

15 °C hold

The whole product of the reaction was run on agarose gel. The DNA was recovered and purified using a Qiagen Gel Extraction kit. The product was used as a MegaR primer for the 2nd PCR.

The PCR conditions were the same in the second reaction, with the following exceptions: the volume of the MegaR primer was 10 µL, and the water volume adjusted accordingly to 22.85 µL. In the thermocycler conditions, the extension time was increased to 2 minutes to reflect the increase in DNA product length (figure 2.3).

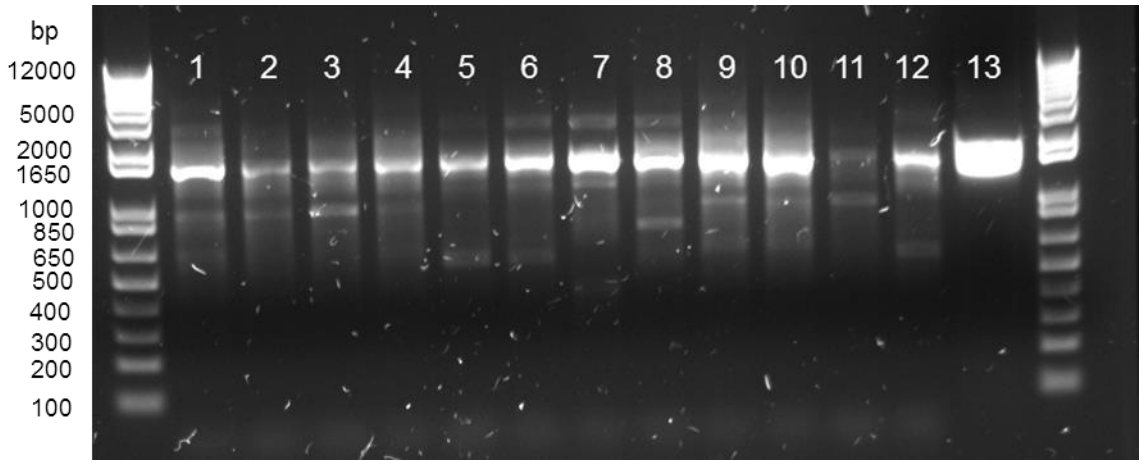


Figure 2.3 Mutagenesis of ACVR1. Example of the products of the second round of PCR for various ACVR1 mutations. (1) Q207D (2) Q207E (3) R206H (4) R202I (5) G328E (6) G328R (7) G328W (8) R375P (9) R258S (10) P197_F198L (11) L196P (12) G356D (13) WT.

2.2.3 Site-directed mutagenesis

This was performed using QuikChange II XL Site-Directed Mutagenesis kit (Agilent) following the protocol but increasing the template concentration 100ng. ACVR1 R202I, R206K, R206H, Q207D, P197_F198L, G356D and G328E mutants were generated using this method.

2.3 Ligation Independent Cloning

2.3.1 Introduction

Ligation independent cloning (LIC) does not require the use of DNA ligase, instead exploiting the activity of T4 DNA polymerase in the presence of a single nucleotide. [97]. T4 DNA polymerase has both polymerase (5' to 3') and exonuclease (3' to 5') activity. In the presence of a single nucleotide the restricted polymerase activity generates a 12 base-pair 5' overhang. The LIC extensions on both the vector and insert ensure that, when treated with complementary nucleotides, large sticky 5' overhangs are produced which allowscomplementary annealing of vector and insert following incubation [98].

2.3.2 Method

A variety of length primers were designed for ACVR1, with and without the GS domain and 2 different C terminal lengths, each containing the appropriate LIC extension. The forward and reverse primers for each length construct were mixed to a concentration of 5 μM each (20 - 50 μl) and aliquoted into a 96-well plate. Each ACVR1 template (wild type

and mutant from previous mutagenesis) was diluted to 2.5 ng/µL and aliquoted into a 96-well plate into wells corresponding to the primer locations. 21 µl of a master mix containing Platinum Pfx polymerase was added to each well of a 96-well PCR plate, followed by primers and templates separately. Standard thermocycler conditions were used (section 2.2.3) with an annealing temperature of 52 °C.

3 µL of the PCR products were analysed on a 1 % agarose gel. The products were treated with *Dpn I* and incubated at 37 ^{o}C for 1 hour. The PCR products were then purified using a Qiagen PCR purification kit and eluted with 50 µl of the kit elution buffer.

To prepare for insertion into the pFB-LIC-Bse vector, the PCR products were treated with T4 DNA polymerase and dCTP, before being incubated for 30 min at 22 °C and then 20 min at 75 °C in a thermocycler.

The pFB-LIC-Bse vector was digested with two cycles of *BseRI* addition, the first at 37 °C in a water bath for 2 hours, the second incubation at 37 °C for 1 hour. The vector was purified using a Qiagen DNA purification kit. The vector was then treated with T4 DNA polymerase as before, except with dGTP in place of the dCTP.

For ligation, 1 µl of treated plasmid was mixed with 2 µl of treated insert. The sample was then incubated for 30 minutes at 22^{o}C and then transferred to ice.

40 µl of MACH-1 competent cells was added to each sample and incubated on ice for a further 30 minutes, then heat-shocked at 42 ^{o}C for 45 seconds. The cells were returned briefly to ice, then 100 µL of SOC medium was added. The plates were then incubated for at least 1 hour at 37 °C. 100 µL of transformation was plated onto Lysogeny broth (LB) agar plates containing 200 µg/mL ampicillin and 5 % sucrose and incubated overnight at 37 °C. A single colony was picked per plate and used to inoculate 1 ml of LB

containing 200 μg/mL. The cultures were grown overnight at 37 °C in a Glas-Col shaker, shaking at 37 °C. 120 μL was removed and mixed with 30 μl 60 %glycerol before being flash frozen at -80 °C to form a glycerol stock. The rest was centrifuged at 2,600 x g for 20 min. The supernatant was removed and the DNA recovered using a Millipore 96-well miniprep kit and stored at -20 °C.

2.4 Buffers

Binding buffer: 50 mM Hepes, pH 7.5; 500 mM NaCl; 5% Glycerol; 5 mM imidazole**,** 0.1mM TCEP

Wash buffer: 50 mM Hepes, pH 7.5; 500 mM NaCl; 5% Glycerol; 25 mM imidazole**,** 0.1mM TCEP

Elution buffer: 50 mM HEPES, pH 7.5; 500 mM NaCl; 5% Glycerol; 50 - 250 mM imidazole, 0.1mM TCEP

Gel Filtration buffer: 50 mM Hepes, pH 7.5; 300 mM NaCl; 0.5 mM TCEP

ITC buffer: 50 mM Hepes, pH 7.5; 150 mM NaCl

10 mM DTT was added after every stage of purification for the BMP receptors to prevent aggregation of the protein.

2.5 Baculoviral Expression

2.5.1 Introduction

All of the BMP/TGFβ receptors used in this thesis were expressed using the baculoviral expression system. A timeline of expression is shown in figure 2.4. Preparation for expression (following initial cloning into the vector) using this system can be broken down into 5 sections; transposition, high-throughput bacmid production, transfection, virus amplification, and expression. The first two steps, transposition and bacmid production, generate the viral DNA. Transposition is a cloning step to transfer the gene of interest into the baculoviral genome using bacterical DH10Bac^{TM} cells which carries viral DNA with a *lacZα* complementation factor. If correctly reconstituted the recombinant bacmid disrupts the introduced *lac*Z cassette, allowing blue/white colour selection [99]. The high-throughput bacmid production isolates the high molecular weight bacmid DNA for transfection. The protocol for each step is described in the following sections.

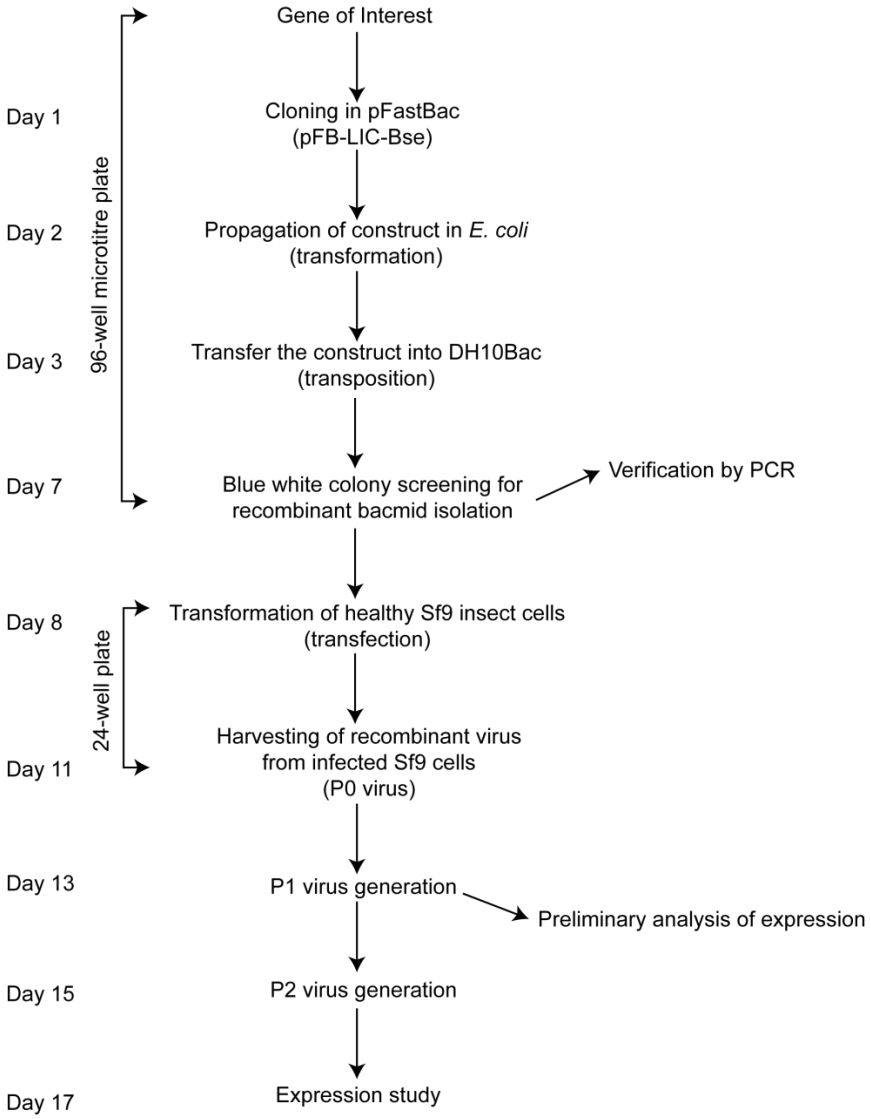


Figure 2.4 Schematic of baculoviral expression. The first three stages have already been described. Adapted from Shrestha *et al* [99].

2.5.2 Transposition

20 µL of DH10Bac™ competent cells was added to 3-5 µL of recombinant DNA on ice and left to rest for 30 min. The cells were then heat shocked for 45 s in a 42 °C water bath and returned to ice before being transferred to 900 µL of pre-warmed 2 x LB medium containing 50 µg/mL kanamycin, 10 µg/mL tetracycline and 1 µg/mL gentamycin. This was then covered with a porous seal and placed in an incubator at 37 °C with shaking at

700 rpm for 5 hours. It was then spread on DH10Bac™ selective agar plates containing 50 μg/mL kanamycin, 7 μg/mL gentamycin, and 10 μg/mL tetracycline, 40 μg/mL IPTG and 100 μg/mL bluo-gal. The plates were incubated at 37 °C for 48 hours covered with foil. White colonies containing the recombinant DNA were selected and streaked to dilution to ensure colour. Plates were left overnight at 37 °C.

2.5.3 High-throughput bacmid production

The white colonies were inoculated with 50 μg/mL kanamycin, 7 μ/mL gentamycin and 10 μg/mL tetracycline in 1 mL of 2 x LB medium. This was incubated overnight at 37°C with shaking at 700 rpm. The following reagents were taken from a Millipore Montage Plasmid Miniprep HTS 96 Kit (LSKP09624): Solution 1 (cell resuspension solution, Millipore (#LSKCRS500)), Solution 2 (cell lysis solution, Millipore (#LSKCLS500)), Solution 3 (neutralization solution (#LSKNS500)) and TE buffer (tris buffer for storage (# LSKCTB500)). The suspension was centrifuged at 2,600 x g for 30 min and the supernatant decanted. 250 μL of Solution 1 was added and the pellet resuspended. 250 μL of Solution 2 was then added and inverted gently 5 times before incubating at room temperature for 10 min. Then 300 μL of Solution 3 was added and mixed by inverting 5 times. The sample was then left on ice for 20 min before being centrifuged at 2,600 x g for 30 min at 4 °C. The clear supernatant was transferred and then centrifuged again. Next 0.8 ml of the supernatant was added to 0.8 mL of isopropanol and mixed by gently pipetting up and down before incubation overnight at 4°C. The sample was then centrifuged at 2,600 x g for 30 min at 4 °C. Under sterile conditions the supernatant was discarded and 500 μL of 70 % ethanol was added to wash the pellet. This was then

centrifuged again, and under sterile conditions, the supernatant removed and left to dry. The bacmid DNA was then resuspended in 50 µL sterile TE buffer (from kit).

2.5.4 Transfection

Sf9 cells were prepared one day in advance by diluting the cell count to 1 x 10^6 cells/mL in Sf900-II medium (Invitrogen) which allowed the cells to reach mid-log phase. The mid-log phase cells were diluted to 2 x 10^5 cells/ml in Sf900-II medium and 1 mL was aliquoted into each well of 24–well tissue culture grade plates. The plates were incubated at 27 °C for 1 hour to allow cell attachment.

50 µL of Grace's insect medium (serum free) was added to a sterile flat-bottomed 96-well microtitre plate. 5 µL of recombinant bacmid DNA was transferred into each well. 5 mL of Grace's medium and 75 µL of linear polyethylenimine (LPEI) (stock 1 mg/mL in water) were mixed and 50 uL of this was added to the bacmid DNA plate. This was diluted by a further 100 µL Grace's insect medium and incubated at room temperature for 30 min.

The cells in the 24-well tissue culture plate were aspirated and 200 µL of the DNA-LPEI as well as 200 µL of Grace's medium were added to the wells. The cells were incubated at 27 °C for 5 hours. 700 µL of Sf900-II insect medium containing 2% (v/v) foetal bovine serum (FBS) and antibiotics Pen/Strep 0.1% (v/v) (Pen/Strep = 50U penicillin and 50 µg streptomycin per ml medium) was then added to each well. The cells were incubated at 27 °C. The viruses were collected when the cells appeared to be infected, appearing larger and elongated, after 96 hours by centrifugation at 2,500 rpm for 20 min at room temperature. This was the P0 virus stock, stored at 4 °C in the dark.

2.5.5 Virus Amplification

Transfer the liquid contents from the 24-well tissue culture plate into a sterile 96-deep well block and centrifuge at 1,500 xg (2,500 rpm) for 20 min at room temperature.

The virus was amplified with 3 ml of Sf9 cells (2 x 10^6 cells/mL) in Sf900-II medium containing 2% (v/v) FBS dispensed into each well of four sterile 24-deep well blocks. The cells were infected with 120 µl of P0 virus stock. The block was covered with a porous seal and incubated at 27 °C, 450 rpm for 72 hours in a Glas-Col incubator. The blocks were centrifuged at 1500 x g for 20 mins and the P1 virus was collected in 2 x sterile 96-deep well blocks. This was repeated for amplification to P2 virus, using the P1 virus in place of the P0 virus.

2.5.6 Expression

Sf9 cells at a density of 2 x 10^6/mL were infected with recombinant baculovirus (virus stock P2; 1 mL of virus stock/500 mL of cell culture). After 48 or 72 hours post-infection the cultures were harvested by centrifugation for 20 min at 900 x g. Cell pellets from each 1L flask were resuspended in 15 mL binding buffer (50 mM Hepes, pH 7.5; 500 mM NaCl; 5 % Glycerol; 5 mM imidazole). Calbiochem protease inhibitor SET V was added to the cell suspension at a 1:2000 dilution before being transferred to 50 mL tubes and stored at -20 °C.

2.6 Escherichia coli expression

All other proteins associated with the BMP/TGFβ receptors were expressed using an *E. coli* expression system.

2.6.1 Transformation

2 µl of plasmid was incubated on ice for 30 minutes with 50 µl of thawed BL21(DE3) competent cells, which contain the bacteriophage T7 RNA polymerase gene. The cells were heat shocked for 45s at 42°C in a water bath. The cells were then returned to ice for a further 10 minutes. 500 µl of LB was added to the cells which were incubated at 37°C while shaking for 1 hour. 90 µl of the suspension was spread onto a 50 µg/mL kanamycin LB-agar plate. The plate was incubated at 37°C overnight. A single colony was picked and grown in LB medium containing 50 µg/mL kanamycin overnight at 37°C to. This cell culture was mixed with an equal volume of sterile glycerol and stored at -80°C to form a glycerol stock.

2.6.2 Bacterial Expression

A glycerol stock was used to inoculate a starter culture containing LB media and 34 µg/mL chloramphenicol and 50 µg/mL kanamycin. The starter culture was grown overnight at 37 °C shaking at 250 rpm. Flasks containing 1 L LB media with 34 µg/mL chloramphenicol and 50 µg/mL kanamycin were inoculated with 5 mL starter culture. Each 1L culture was incubated at 37 °C at 160 rpm until an OD600 nm ≥ 0.5 was reached. The flasks were then cooled down to 18 °C and 0.5 mM Isopropyl β-D-1-thiogalactopyranoside (IPTG) added to induce protein expression overnight. Cells were

harvested by centrifugation at 4500 rpm at 4 °C for 15 min. Cell pellets were resuspended in 30 ml binding buffer, transferred to a 50 mL tube, and stored at -20 °C.

2.7 Protein Purification

Protein purification was the same for all constructs, except where noted. After each purification column the elutions were subjected to sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) to show the presence and purity of the protein. The gels were stained using a Coomassie-based stain (InstantBlue, Expedeon). An example of the data produced is shown in figure 2.5.

2.7.1 Lysis

The frozen cells were thawed and the volume increased to 200 mL with binding buffer. The cells were lysed using an Emulsiflex C5 homogeniser or using a sonicator. The cell lysate was spun down by centrifugation at 21,000 rpm at 4 °C for 1 hour. The supernatant was recovered for purification.

2.7.2 Anion Exchange

Nucleic acid removal with diethylethanolamine **(**DEAE) cellulose (DE52, Whatmann) 10 g of resin was suspended in 50 mL 2.5 M NaCl, and then applied onto a 2.5 x 20 cm column. The resin was then equilibrated with 50 ml binding buffer prior to loading the sample.

The supernatant was first applied onto the column by gravity flow, which was followed by a wash with 50 mL wash buffer. The column flow-through and wash were directly applied onto a Ni-sepharose column.

2.7.3 Ni-Affinity Chromatography

6 mL of 50 % Ni-sepharose slurry was applied onto a 1.5 x 10 cm column. The column was washed with distilled water and then equilibrated with binding buffer (25 mL).

The flow-through from column 1 (DE52) was applied by gravity flow onto the Ni-sepharose column. The bound protein was eluted by applying a step gradient of imidazole – using 10 mL portions of elution buffer with increasing concentration of imidazole (50 mM, 100 mM, 150 mM, 250 mM).

2.7.4 Enzymatic treatment

0.1mg of TEV protease was added to the Ni-eluted protein to remove the tag if required.

2.7.5 Size-exclusion chromatography

S75 or S200 HiLoad 16/60 Superdex (Amersham) columns run on ÄKTA-Express were used. Prior to applying the protein, the Superdex column was washed and equilibrated with gel filtration buffer. The protein was concentrated to 3 mL using an Amicon Ultra-15 filter with an appropriate size cut-off. The concentrated protein was directly applied onto the equilibrated column, and run at a flow-rate of 1 mL/min. Fractions containing the protein were pooled together.

2.7.6 Nickel rebinding

0.5 mL 50% Ni-sepharose slurry was applied onto a Bio-Rad polypropylene disposable drip column. The column was washed with distilled water and then equilibrated with binding buffer. The pooled protein was applied onto the column by gravity flow. This was followed by 8mL wash buffer, then 10 mL 250 mM imidazole to elute anything left on the column.

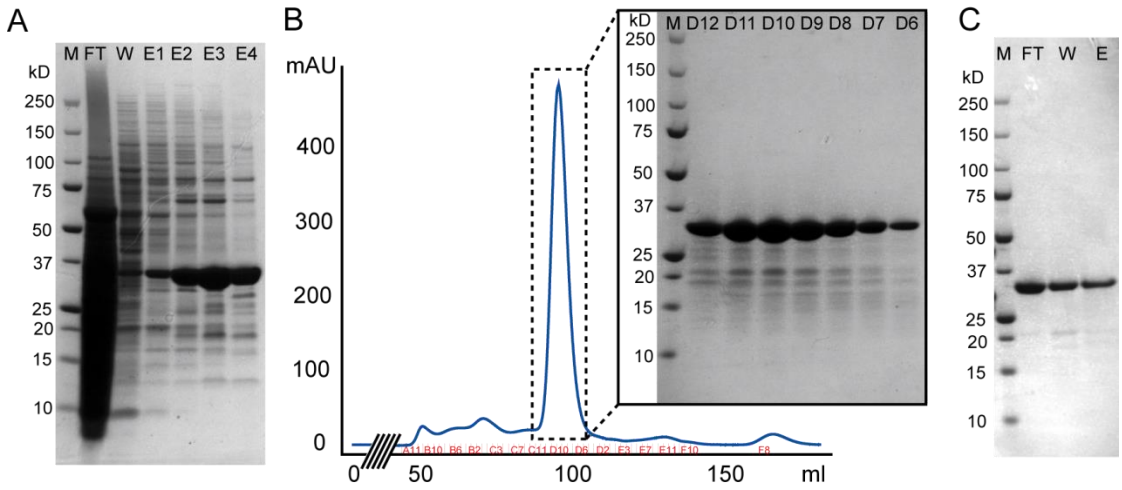


Figure 2.5 Example of the purification protocol. Purification of ACVR1 Q207D (a.a 201 – 499). (A) SDS PAGE showing the elution fractions from a Ni-sepharose column using increasing imidazole concentration. Fractions shown are flow through (FT), wash (W), and numbered elution fractions referring to the imidazole concentration (50 mM, 100 mM, 150 mM, 250 mM labelled as E1-4 respectively). (B) Chromatogram of protein elution following gel filtration. SDS PAGE of the gel filtration elutions showing purity of the protein. (C) SDS PAGE showing the elutions of Ni-rebinding labelled as flow through (FT), wash (W) and elution (E).

2.8 Mass Spectrometry

2.8.1 Intact Mass

Masses were determined by LC-MS, using an Agilent LC/MS TOF system with reversed-phase HPLC coupled to electrospray ionisation and an orthogonal time-of-flight mass analyser. Proteins were desalted prior to mass spectrometry by rapid elution off a C3 column with a gradient of 5-95% isopropanol in water with 0.1% formic acid.

2.8.2 MSMS (Phosphorylation mapping)

Phosphorylation mapping of the proteins was achieved following tryptic digest and identification of phosphopeptides using LC-MS/MS. 100 µL of 100 mM NH_4HCO_3 (pH 8.0) and 1 µL of 1 M DTT was added to the sample and incubated at 56 °C for 40 minutes. 4 µL of 1 M iodoacetamide was then added and the reaction incubated at ambient temperature in the dark for 20 min. A further 1 µL of 1 M DTT, 200 µL of 100 mM NH_4HCO_3 and 1 µL of trypsin solution (sequencing grade, 1 mg/mL in 0.01 M HCl) was then added. Tryptic digestion proceeded at 37 °C for 16 hours and was terminated by addition of 3 µL of formic acid.

Phosphopeptides were enriched offline using TiO/ZrO spin columns. The columns were prepared using Titanosphere TiO 10 µm medium (GL-Sciences) and Alfa Aesar 6 micon sieve ZrO_2 powder. The columns were washed first with an aqueous ammonia solution (25 % ammonium hydroxide stock, 75 % water) and then a solution containing 80 % ACN, 2 % (TFA) and 3 mg/mL dihydroxybenzoic acid (DHB). The sample solution was adjusted to 10 % formic acid before being added to the column. The sample was then

desalted using 40 µL ion trap buffer (2 % ACV, 0.1 % FA). The non-phosphopeptides were eluted from the column using 49 µL 80% ACN, 300 mg/mL DHB, 0.1% TFA. The column was then washed again using 80 % ACN, 0.1 % TFA, and then 80 % CAN, 0.1 % FA. The phosphopeptides were eluted with 40 µL aqueous ammonia solution, 40 % ACN. The eluent was evaporated to dryness and resuspended in 5 µL 2 % ACN, 0.1 % FA.

LC-MS/MS was performed using a *Dionex U3000 nano* HPLC coupled to a *Bruker Esquire HCT* ion trap mass spectrometer. 5 µL of enriched phosphopeptides were loaded on to a 200 µm ID x 5 cm PS-DVB monolith column (*Pepswift, Dionex Corp.*) A linear gradient of 0 % B to 15 % B was developed over 5 min, followed by a second linear gradient from 15 % B to 40 % B over 2 min. The column was washed at 90 % for 2 minutes and then equilibrated at 90 % B for a further 6 min. Solvent A was 2 % (v/v) acetonitrile, 0.1% formic acid in water, solvent B was 80 % acetonitrile 0.1 % formic acid. The flow rate was 2.5 µL/min. The mass spectrometer was operated in positive ion, standard enhanced mode with a scan rate of 8,100 m/z/sec and a scan range of 250 – 1800 m/z. The trap accumulation time was 200 msec and the accumulation target was 200,000 counts. Data dependent peptide fragmentation was performed in Auto MSMS mode. Compound extraction and peptide deconvolution was performed using the DA data analysis program (*Bruker Daltonik*). Database searching was performed using the *Mascot* 2.2.04 search algorithm with the following search parameters: Charge states +1, +2, +3; MS tolerance 1.5 Da; MSMS tolerance 1.3 Da; Carbamidomethyl global modifications; Phospho (ST), Phospho (Y) and Oxidation (M) variable modifications; <4 missed Trypsin cleavages; UniProt_SwissProt database without taxonomic restrictions.

2.9 Western blotting

The protein was transferred onto nitrocellulose membrane (GE Healthcare) and probed with the anti-P-Smad1/5.8 antibody (Cell Signaling (#9511)) at 4°C overnight. After subsequent incubation with HRP-conjugated anti-rabbit antibody, protein bands were detected using ECL (Pierce) and an LAS4000 image reader.

2.10 Differential Scanning Fluorimetry analysis

2.10.1 Introduction

Differential scanning fluorimetry (DSF) is a method of identifying small molecule compounds that bind to a native protein and increase its stability [100]. It is a widely applicable assay that can be used for any type of protein class, and for both active and inactive enzymes. DSF does not require much protein and melting temperature (Tm) values can be measured for many small compounds in a 96 well plate format making this method relatively high throughput. When a protein unfolds it exposes hydrophobic regions that are otherwise buried. Fluorescent dyes, such as Sypro Orange, that are quenched in aqueous solutions can bind to these regions as they are exposed (figure 2.6) [100]. The levels of fluorescence emitted by the dye are measured as a reporter for protein folding; higher levels of fluorescence reflect a greater population of unfolded protein [100]. Protein stability measurements are routinely made from equilibrium unfolding studies, including by thermal denaturation. The temperature at which the amounts of unfolded and folded protein are equal is known as the Tm [101]. When a small molecule binds to the folded state of the protein it causes an increase in the Tm value, referred to as Tm shift [100]. This stability is proportional to the concentration ($[L_f]$) and affinity (K_b) of

the ligand and the stability of the protein in the absence of ligand ($\Delta_U G_{(T)}$) (equation 2.2) [102].

$$\Delta G_{(T)} = \Delta_U G_{(T)} + \Delta_B G_{(T)} = \Delta_U G_{(T)} + RT \ln(1 + K_b [L_f]) \qquad \text{(equation 2.2)}$$

[102]

Where $\Delta G_{(T)}$ is the is the observed stability of the protein, $\Delta_U G_{(T)}$ is the protein stability in the absence of ligand, and $\Delta_B G_{(T)}$ is the additional stabilization free energy due to ligand binding. and $[L_f]$ is the concentration of free ligand. At the Tm value $\Delta G_{(T)} = 0$.

In practice the true thermodynamic equation cannot be reliably fit from DSF data due to the number of unknown parameters [101]; also the denaturation is typically irreversible so equilibrium conditions are not met [101]. However, at high heat rates the Tm approaches the expected equilibrium as the effect of aggregation is minimized. A reasonable estimation of the observed Tm can be obtained from a simple Boltzmann equation (equation 2.3).

$$I = \left(A + \frac{(B-A)}{1+e^{(Tm-T)/C}}\right) \qquad \text{(equation 2.3) [101]}$$

Where *I* is the fluorescence intensity at temperature *T, A* and *B* are pre-transitional and post-transitional fluorescence intensities respectively and *C* is a slope factor.

It is important to note that proteins cannot be compared exactly with each other; a Tm shift with an inhibitor in one protein is not comparable with that of the same inhibitor with another protein. What is comparable is how this inhibitor compares against other inhibitors for that one protein, and this ranking can be compared from protein to protein.

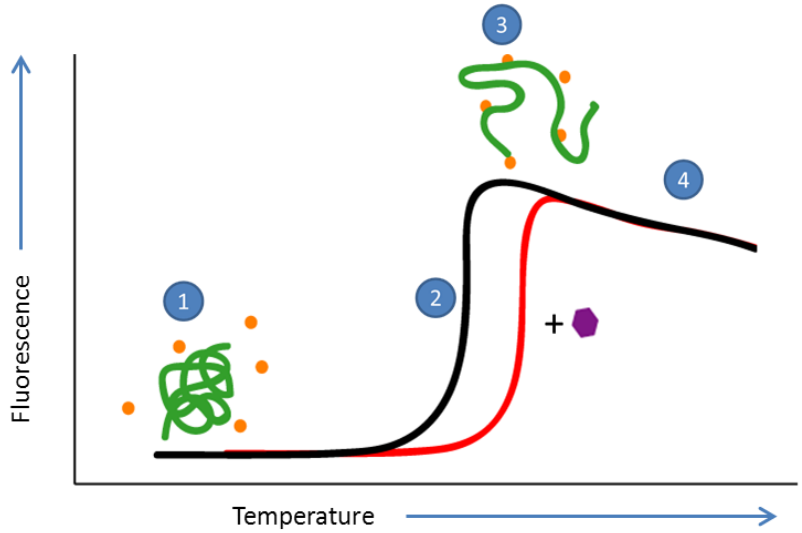


Figure 2.6 Cartoon representation of DSF. (1) The dye binds poorly to the folded protein giving low fluorescence. (2) The Tm or midpoint is the temperature at which the population of unfolded protein is equal to that of folded protein. (3) The dye binds preferentially to unfolded protein with exposed hydrophobic regions giving the highest level of fluorescence. (4) Sloping baselines are expected as fluorescence intensity decreases with increasing temperature. This results from increased molecular collisions at higher temperatures which bleed energy from the excited state that produces fluorescence. In addition, protein may aggregate upon unfolding and fall out of solution again reducing fluorescence at high temperatures. Any ligand binding preferentially to the folded protein (red line) will increase the melting temperature (Tm) as a result of the law of mass action. The size of the Tm shift is proportional to the binding affinity.

Compounds screened were selected from the SGC compound library, specifically directed towards kinases. The library contains about two thousand compounds from commercial libraries of well-defined and published kinase inhibitors. The commercial vendors included Calbiochem. In addition several hundred compounds were purchased from BioFocus,a kinase-directed compound collection which comprises known ATP mimetic chemical scaffolds exploring a wide range of substituents. Finally a large number of compounds are included from academic collaborators which have been shown to have inhibitory properties.

2.10.2 DSF Method

Purified protein was distributed at a concentration of 2µM in 500 mM NaCl, 10 mM HEPES pH 7.5 buffer, containing SYPRO orange, into a 96-well PCR plate. Inhibitors were added at an end concentration of 10 mM. The plate was then briefly spun down and placed into a Real-Time PCR machine (Stratagene Mx3005p). Samples were heated in a series of 71 cycles, each cycle increasing in temperature by 1 °C from 25 °C to 95 °C and the level of fluorescence measured (figure 2.7). The data were then exported and analysed in excel using the Boltzmann equation (equation 2.3).

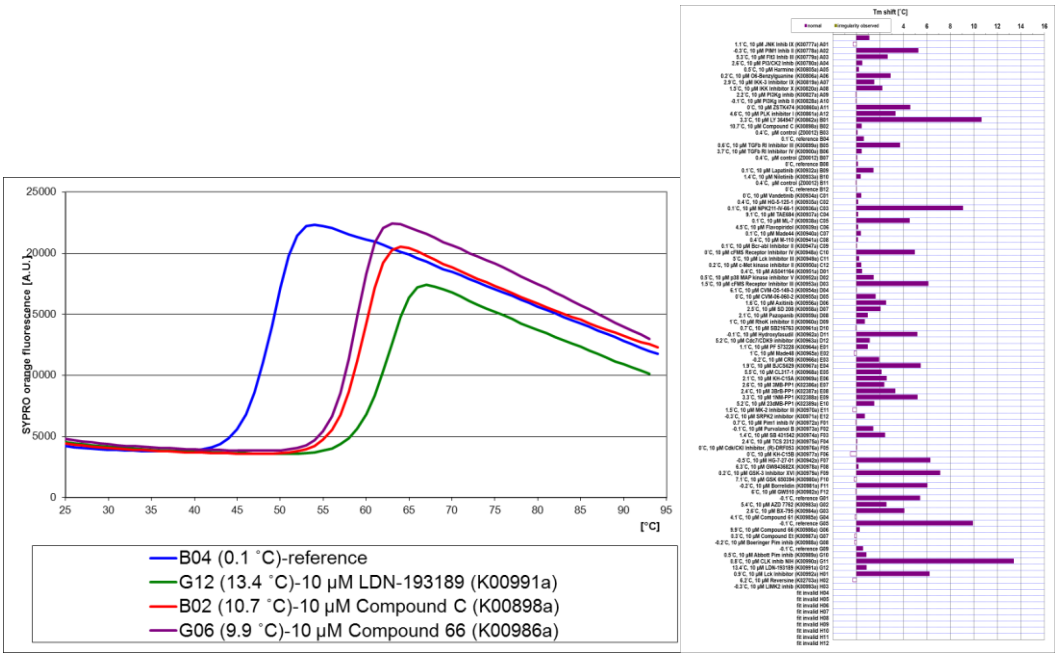


Figure 2.7 Example of DSF analysis. (A) Fluorescence of unfolding and (B) Bar chart of Tm shifts forACVR1 Q207D in the absence and presence of selected kinase inhibitors, including LDN-193189 and dorsomorphin (Compound C).

2.11 Isothermal Titration Calorimetry

All titrations were carried out using a VP-ITC (Microcal) instrument and analysed using a single binding site model implemented in the Origin 7.0 software package provided with the instrument (OriginLab). This allowed the determination of the change in enthalpy (ΔH), binding affinity (K_a) and stoichiometry (n) of the binding reaction. From these parameters the change in both entropy (ΔS) and Gibbs free energy (ΔG) were calculated using the equations:

$$\Delta G = \Delta H - T\Delta S \qquad \text{(equation 2.4) [103]}$$

$$\Delta G = - RT \ln K_a \qquad \text{(equation 2.5) [103]}$$

$$K_d = 1 / K_a \qquad \text{(equation 2.6)}$$

where R is the gas constant taken as 8.314 J.K-1.mol-1, T is the temperature in Kelvin, and K_d is the dissociation constant of the reaction.

The single site binding interaction, titrations are fitted to a curve by Origin software following the equation:

$$Q = \frac{nM_t\Delta HV_O}{2}\left[1 + \frac{X_t}{nM_t} + \frac{1}{nKM_t} - \sqrt{\left(1 + \frac{X_t}{nM_t} + \frac{1}{nKM_t}\right)^2 - \frac{4X_t}{nM_t}}\right] \qquad \text{(equation 2.7)}$$

where Q is the total heat, V_0 is the active cell volume, M_t is the bulk concentration of the protein, X_t is the bulk concentration of the ligand. Over a series of injections this equation can be fitted and the value of Q adjusted for the increased volume due to the injection of ligand into the cell, from this n (the number of binding sites), ΔH (the enthalpy of the reaction) and K (the binding constant) can be determined. [104]

All experiments were carried out at 10 °C, at reference power of 15 μcal/s with a series of 20 injections. The first injection had a volume of 2 μL with a delay of 120 s before the second injection. All subsequent injections had a volume of 15 μL with a delay of 240 s between each and a constant stirring speed of 300 rpm was used. All experiments were conducted with ligand in the cell and protein in the syringe. The concentration ratio of syringe to cell was 10:1. Prior to set up the protein was dialysed in ITC buffer (section 2.4) overnight whilst the compound was diluted in the remaining buffer to ensure an exact buffer match.

2.12 X-ray crystallography

2.12.1 Introduction

X-ray crystallography is a powerful tool for understanding the three-dimensional arrangement of atoms within proteins. This can be used to further the understanding of protein function and also provide information on ligand interaction for use in drug development. Protein crystallography can be divided into three broad stages: protein crystal growth, data collection and data analysis.

Protein crystals are formed by precipitating the protein molecules out of solution. These molecules tend to pack in a regular way to achieve the lowest energy state. In order to produce protein crystals, the protein molecules must first separate from solution and then self-assemble into an ordered lattice. This is done either by adding precipitants or by removing water to form a supersaturated protein solution. Nucleation may then take place and under favourable conditions crystals may grow [105] (figure 2.8A).

The solution that crystals are grown from is known as the mother liquor. Protein crystals contain around 50 % solvent on average and must be stored in their mother liquor at all times to prevent dehydration. In this thesis the "sitting drop" vapour diffusion method was employed (figure 2.8B). In this technique, a small drop containing protein mother liquor equilibrates with a much larger volume of reservoir solution by vapour diffusion until the final conditions of the drop are very similar to those of the reservoir solution [105], [106].

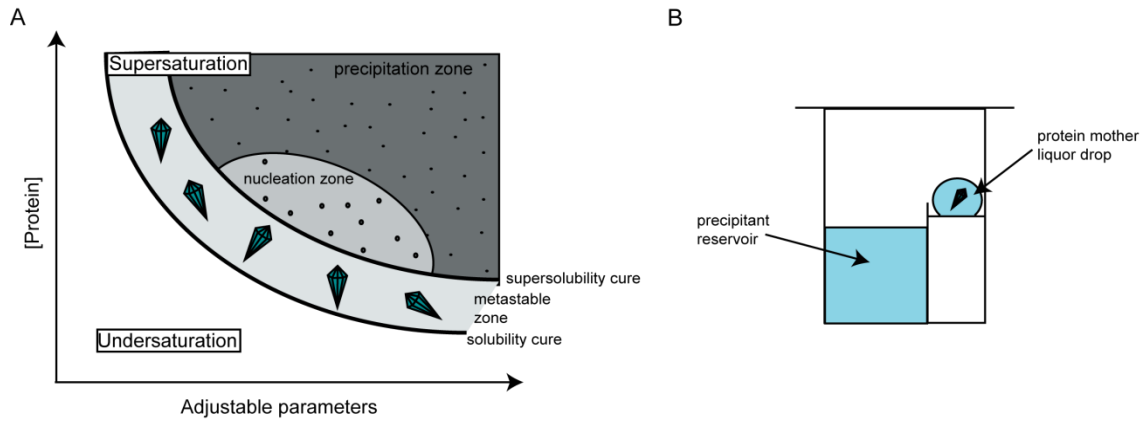


Figure 2.8 Crystal formation. (A) Protein crystallisation phase diagram. There are several zones: a zone of undersaturation, where the protein has fully dissolved and no crystallisation can occur.; a zone of high supersaturation, where protein precipitates; a zone of moderate supersaturation, where spontaneous nucleation will take place; and the metastable zone (just below the nucleation zone) of lower supersaturation, where crystals are stable and may grow but no further nucleation will take place. This final zone is where the best growth of crystals will occur. Adapted from Chayen 2004 [105] (B) Sitting drop method.

Once the crystals have been grown they must first be flash frozen in the presence of cryoprotectant (usually mother liquor and ethylene glycol) to reduce radiation damage during data collection. The crystal is then mounted on the instrument and an x-ray beam is fired through the crystal. The x-rays are diffracted in waves by the atoms, resulting in both destructive and constructive interference. This can be determined by Bragg's Law,

$$2\,d \sin \Theta = n\lambda \qquad \text{(equation 2.8)}$$

where *d* is the spacing between the atomic layers in a crystal, θ is the angle of incidence of the x-ray beam, λ is the wavelength of the incident x-ray beam and *n* is an integer.

During data collection the crystal is rotated many times and diffraction data are collected for each new rotation. Combined, this data is referred to as a dataset. In order to obtain a structure, it is necessary to calculate a Fourier transform using the structure factors amplitude and phase [107]. However, the x-ray crystallography experiment can only provide the reflection amplitudes from measured intensities, but the phases cannot be measured. This is known as the phase problem. There are various techniques for solving this and here molecular replacement was used as the structure of wild-type ACVR1 had previously been solved. Molecular replacement uses a closely related protein structure as a model to provide an initial phasing solution. This technique uses the calculated diffraction data from the known search model and rotates and translates this data within the unit cell of the crystal to obtain the best possible match with the experimental diffraction data [108]. The computational program Phaser, which employs algorithms based on maximum-likelihood probabilities, was used in this thesis to calculate phases [109]. This can then be used to provide an electron density map. Refinement of the model into this map improves the phases, which then improves the map. This continues, taking into account various constraints such as bond angles and lengths, until the model (R_{fact}) is in agreement with the observed data (R_{free}).

2.12.2 Crystallisation of proteins

Protein was buffered in 50 mM HEPES pH 7.5, 300 mM NaCl, 10 mM DTT and 50mM L-arginine, 50 mM L-glutamate. The protein was concentrated to 10-15 mg/ml (calculated using appropriate extinction co-efficients) in the presence of inhibitor (1 mM

final concentration). Sparse matrix screens used were an SGC ligand friendly screen (LFS), which is based on a range of precipitants varying the polyethylene glycol (PEG) size at low salt concentrations, the commercial HCS (Hampton Crystal Screens I + II) and HIN (Hampton Index) screens and the JCSG screen, which is a screen developed by a collection of academics at the Joint Center for Structural Genomics**.** Crystals were grown at both 4°C and 20°C in 150 nl sitting drops at 3 different ratios of protein to precipitant, 2:1, 1:1 and 1:2. Crystals were mounted by SGC staff and cryoprotected with mother liquor plus 20% ethylene glycol before transfer to liquid nitrogen.

2.12.3 Data Collection

This was performed at the Diamond Light Source by SGC staff.

2.12.4 Data Processing

Data were integrated and scaled with MOSFLM and SCALA or AIMLESS from the CCP4 suite [110,111,112]. The phases were solved by molecular replacement using PHASER [109]. The model for molecular replacement of ACVR1 Q207D in complex with LDN-213844 was ACVR1 Q207D in complex with LDN-193189 (pdb 3q4u). Subsequent manual model building was performed using COOT [113] and refined using REFMAC [114]. TLS-restrained refinement was applied in the latter cycles using the input thermal motion parameters determined by the TLSMD server [115]. The final models were verified for geometry correctness with the PHENIX [116] validation tools and MOLPROBITY [117].

3. BMPR1B offers a structural model for mutant ACVR1

3.1 Introduction

Signalling by the type I BMP/TGFβ receptor kinases is controlled by the distinct active and inactive conformations of the kinase domain. To date all kinase structures of the type I receptors are representative of the inactive state. However, some insights into the conformational changes required for activation can be gained from structural comparison of the type II receptor BMPR2 which is constitutively active (figure 3.1). A number of interactions in the crystal structure of the ACVR1-FKBP12 complex are identified as important for the stability of the inactive conformation of the receptor (figure 3.1A). To prevent leaky signalling FKBP12 stabilises this conformation by binding to residues in the GS domain of the type I receptor [76,83]. This traps the GS loop in an inhibitory wedge between the N-lobe β strands and αC preventing GS loop phosphorylation. This packing is held by hydrogen bond interactions between R258 in the kinase domain and the GS loop (figure 3.1). In FOP, the R258S mutation is predicted to be unable to form these bonds and thereby destabilise the inactive conformation. Similarly, release of FKBP12 may allow the GS loop greater freedom to move away from the kinase domain increasing the potential for GS loop phosphorylation. The inhibitory ACVR1-FKBP12 complex also shows a closed conformation of the activation loop (A-loop) which blocks the binding of ATP and substrate. This inactive conformation is held by a salt bridge between R375 (A-loop) and the DFG motif (D354) (figure 3.1). Again the FOP-associated mutations R375P and G356D (DFG motif) are predicted to destabilise the inactive state. As a result, ATP pocket elements including the αC helix, will have freedom to move to complete the

conformational changes required for receptor activation[76] (figure 3.1B). In fact, all sites of mutation identified in FOP are involved in stabilizing the inactive conformation in ACVR1 [87].

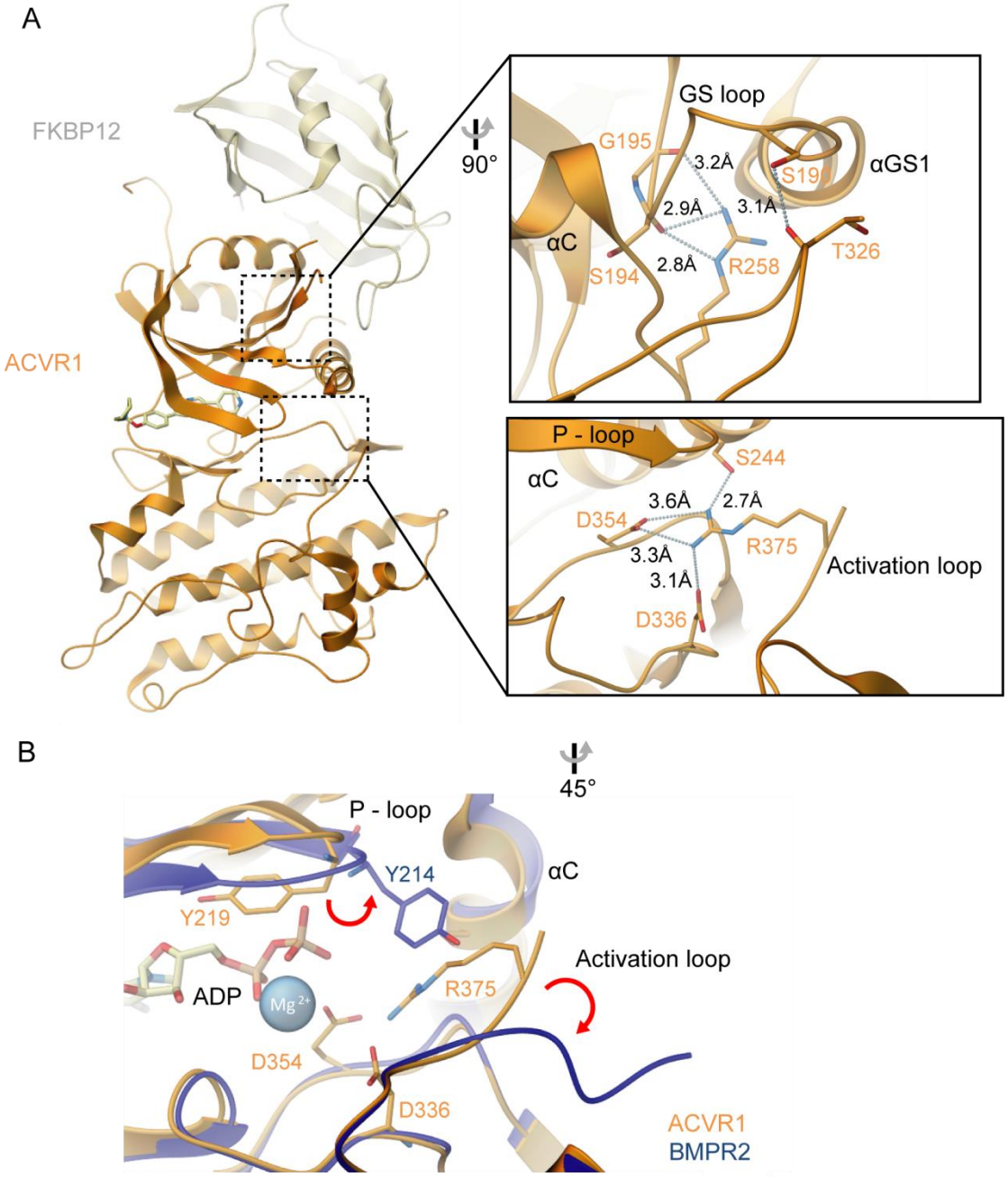


Figure 3.1 Inhibitory interactions within ACVR1. (A) The complex of ACVR1 (orange) bound to FKBP12 (grey) (pdb: 3h9r). The GS domain is bound to FKBP12, burying the GS loop between the αC helix and the kinase domain β4 strand. R258 is a FOP-associated residue that stabilises this inactive conformation (top panel). ATP binding is also blocked by an inhibitory conformation of the activation loop, which is held by another FOP-associated residue R375 (bottom panel). (B) Superposition of ACVR1 and the structure of the BMPR2-ADP complex (dark blue; pdb: 3g2f) shows conformational

changes required for kinase activation and ATP binding. Comparison of the two equivalent tyrosine residues (ACVR1 Y219 and BMPR2 Y214) in the P-loop revealed a 12 Å shift in position of the -OH group. The activation loop C-alpha distance between R375 and BMPR2 equivalent (E376) is 7.7Å to allow ATP binding.

The tight interaction between the GS and kinase domains helps to stabilize the buried GS loop in the FKBP12 complex. Several FOP-associated mutations occur at this interface. The classic R206H mutation in the GS domain is predicted to be unable to form hydrogen bonds with D269 and M270 in the β4 strand of the kinase N-lobe (figure 3.2). The histidine imidazole side chain has a pKa of 6.0 and so at neutral pH it is largely unprotonated. Therefore the R206H mutation could act as a pH-sensitive switch disrupting interactions between the histidine and the kinase domain, with binding only occurring as the side chain is protonated at lower levels of the physiological pH range [118]. As binding to the kinase domain is disrupted the GS domain helices would no longer be held as tightly, altering the site of FKBP12 binding. Thus the change in electrostatic potential is also thought to contribute to the weaker interaction of ACVR1 R206H with FKBP12.

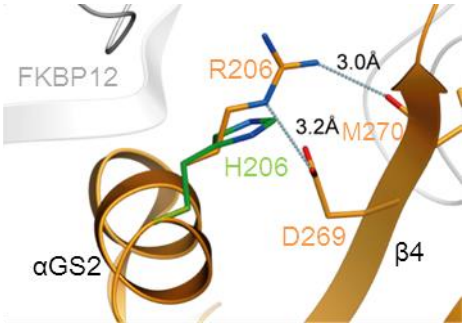


Figure 3.2 Interactions at the R206 site. In WT ACVR1, R206 (orange; pdb:3h9r) forms hydrogen bond interactions with D269 and M270. When modelled using Molsoft ICM-Pro, the mutated H206 (green) side chain is unable to form these same interactions at neutral pH.

The ACVR1 D269 position is completely conserved across all of the type I BMP/TGFβ receptors, while the R206 position is conserved with the exception of the BMP receptors BMPR1A and BMPR1B. Instead, these receptors have a lysine substitution at this position

(for example, BMPR1B K202, figure 3.3). It has been shown that both receptors are still inhibited by FKBP12but the substitution of this arginine should prevent the interactions with the kinase β4 strand observed in ACVR1 (figure 3.2). Therefore, it is of interest to determine the structures of wild-type BMPR1A and BMPR1B to investigate these changes in addition to the efforts to crystallize mutant ACVR1. In particular, there structures may provide insights into how structural changes at this position are propagated throughout the catalytic domain.

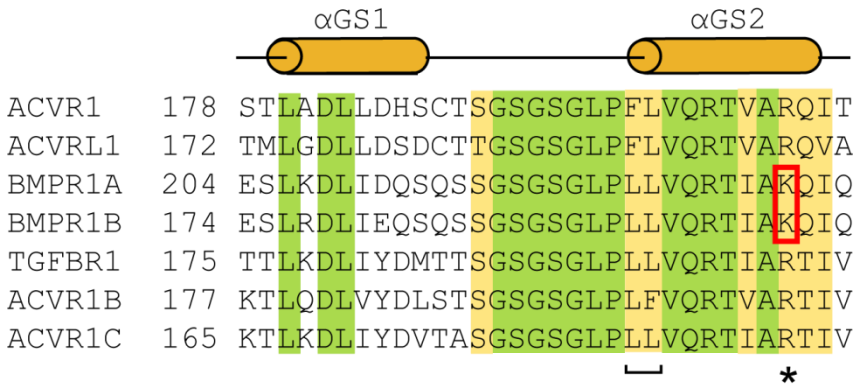


Figure 3.3 Alignment of the GS domain of type I receptors. The GS domain is highly conserved across all members of the family. The ACVR1 R206 position (marked by an asterisk) is conserved across the family except for the BMP receptors BMPR1A and BMPR1B. The FKBP12 binding pair is also highlighted by a bracket.

3.2 Results

3.2.1 Expression testing of BMP/TGFβ receptors

A set of baculoviral expression constructs for the human BMP/TGFβ receptor cytoplasmic domains were cloned by the SGC prior to the start of this project and were available for small scale test expression and purification. These included constructs spanning both the

GS and kinase domains as well as truncations comprising only the kinase domains. The latter also included the artificial constitutive active mutant ACVR1 Q207D and the FOP-associated mutant ACVR1 R206H. Forty constructs were tested in 6 ml *Spodoptera frugiperda* (Sf9) cell cultures, infected with 100 µl P1 virus, and harvested in 3 ml volumes at 48 and 72 hours post infection. Cells were lysed by sonication, clarified by centrifugation and recombinant protein purified by Ni-sepharose pull down as described in the Methods section 2.6 (figure 3.4). Table 3.1 summarises the purified protein yield observed for the various constructs.

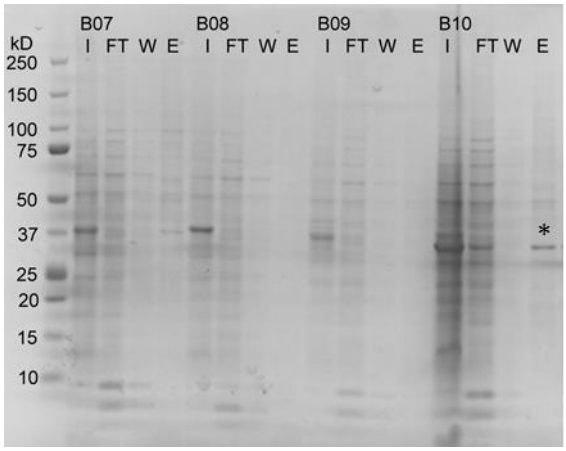


Figure 3.4 Example SDS PAGE analysis of BMP receptor expression testing at 72 hours post infection. Fractions included are I = insoluble, FT = flow through, W = wash, and E = elution. The labels B08-B10 refer to different truncation constructs of BMPR1A and BMP1B as outlined below in Table 3.1. The elution fraction of the construct used in this chapter is highlighted with an asterisk.

Of the constructs tested, the highest soluble expression was observed for the type I BMP receptor BMPR1B, the type I activin receptor ACVR1B and the type II TGFβ receptor TGFBR2. Other receptors, including the ACVR1 kinase domain, were recovered at lower levels (Table 3.1). Constructs of BMPR1B, BMPR1A, TGFBR1 and the two mutated ACVR1 kinase constructs were subsequently expressed at large scale for crystallisation trials (the remaining targets were pursued by other colleagues). Viable crystals were

successfully obtained for BMPR1B (construct B10, discussed below) and the two mutant ACVR1 constructs (A2 and A3). ACVR1 crystallisation will be discussed in chapter 4.

Table 3.1 Table summarising all constructs tested for soluble expression.

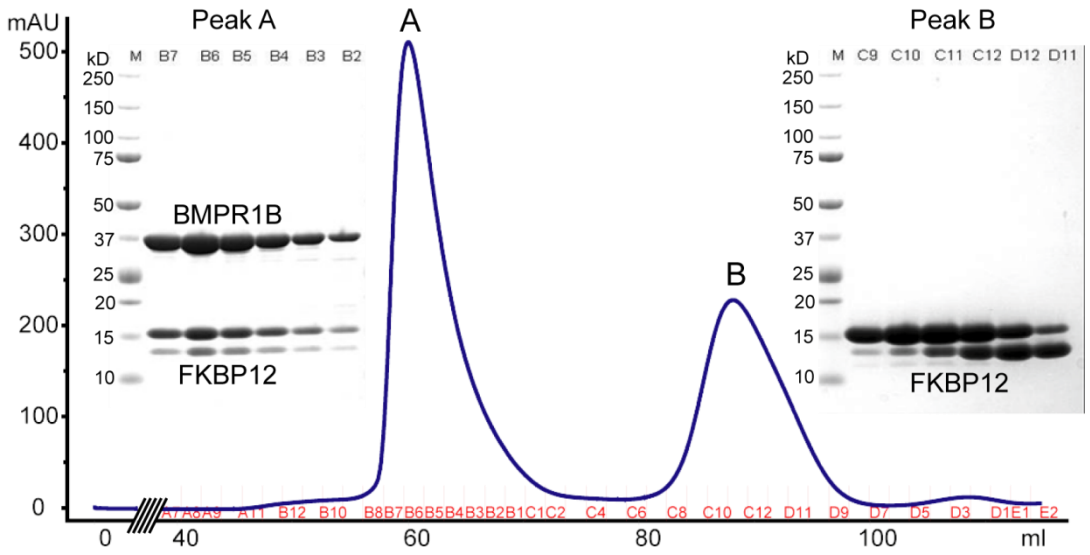
Well	Target	Construct (a.a)	Theoretical Mass (Da)	Expression	
				48 hour	72 hour
A01	ACVR1	201-499	37028.5	Low	Low
A02		201-499 (Q207D)	37015.5	Low	Medium
A03		201-499 (R206H)	37009.4	None	Medium
A04	ACVR1B	172-505	40700.7	None	None
A05		172-501	40245.2	None	None
A06		200-505	37774.4	High	None
A07		200-501	37318.8	High	Low
A08	ACVR1C	159-493	40668	None	None
A09		159-487	39993.2	None	None
A10		188-493	37769.6	None	None
A11		188-487	37094.9	None	Low
A12	ACVRL1	166-502	40708.4	Low	Low
B01		166-497	40142.7	Low	None
B02		166-493	39715.2	Medium	None
B03		195-502	37853.3	None	Low
B04		195-497	37287.6	None	None
B05		195-493	36860.1	None	None
B06	BMPR1A	198-532	40672	Medium	Low
B07		198-526	40001.2	Low	Low
B08		227-532	37774.7	Low	None
B09		227-526	37104	None	None
B10	BMPR1B	168-502	40787.9	Low	High
B11		168-496	40103.1	Medium	Low
B12		197-502	37804.6	Low	None
C01		197-496	37119.8	None	None
C02	AMHR2	189-519	39576.9	None	None
C03		189-514	38969.3	None	None
C04		196-519	38789.1	None	None
C05		196-514	38181.4	None	None
C06		202-519	38079.3	None	None
C07		202-514	37471.7	None	Low
C08	ACVR2	191-494	37695.3	Medium	Low
C09		191-488	37022.6	None	None
C10		186-494	37170.6	None	None
C11		186-488	36497.9	None	None
C12	TGFBR2	232-547	39021.3	Medium	Medium
D01		237-547	38494.8	High	High
D02		241-547	38013.3	Low	Low
D03	TGFBR1	162-503	41606.7	Low	None
D04		198-503	37798.4	None	Low

The purification levels from none to high indicate the amount of soluble protein visible in the SDS PAGE analysis of the elution fraction.

3.2.2 Purification and crystallisation of BMPR1B with the inhibitory protein FKBP12

3.2.2.1 Purification of BMPR1B and FKBP12

A construct of BMPR1B (B10, Table 3.1) containing both the GS and kinase domains (a.a. 168 – 502) was expressed at 3 L scale in Sf9 cells, at 27°C for 72 hours. Separately, a construct of full length human FKBP12 was expressed in *E.coli*. Initially, the two proteins were purified individually by Ni-affinity chromatography. Subsequently, the FKBP12 protein was mixed in excess with a sample of BMPR1B before applying the complex to a Superdex S75 16/60 (Amersham) gel filtration column. FKBP12 co-eluted with BMPR1B with a shift to higher molecular weight (earlier elution time), indicating that the two proteins had bound (figure 3.5). A sample of BMPR1B alone was also purified by gel filtration for a separate crystallisation trial.



igure 3.5 Complex formation of BMPR1B and FKBP12. A complex was formed by mixing BMPR1B (38265.2 Da) with an excess of FKBP12 (12037.8 Da). The 1:1 complex (peak A) was purified and separated from surplus FKBP12 (peak B) by size-exclusion chromatography. The peak fractions were analysed by SDS-PAGE. The double band of FKBP12 is due to incomplete cleavage of the His-tag by TEV protease.

Crystallisation trials were performed to test the effects of precipitant, inhibitors and temperatures to optimise conditions for producing crystals. Both the purified BMPR1B protein and the BMPR1B-FKBP12 complex were concentrated to 15 mg/mL in 50 mM HEPES pH 7.5, 300 mM NaCl, 10 mM DTT, 50 mM L-arginine, 50 mM L-glutamate. The concentrated protein was used in crystallisation trials with two chemical inhibitors: LDN-193189, which was previously identified as an inhibitor of BMP receptors, and K00986a, which was identified as another potent binder by differential scanning fluorimetry (DSF) screening of the SGC compound library (discussed in Chapter 5.2.1). The crystal plates were set up using sitting drops for vapour diffusion. Four different sparse matrix screens were tested at both 4°C and 20°C (see Chapter 2.12.2 for protocol).

Table 3.2 Conditions screened for protein crystallisation of both BMPR1B and the BMPR1B-FKBP12 complex.

Protein	Screen*	Compound	Temperature (°C)
BMPR1B	LFS	LDN-193189	20
	LFS	K00986a	20
	JCSG	LDN-193189	4
	JCSG	K00986a	4
	HCS	LDN-193189	4
	HIN	K00986a	20
BMPR1B-FKBP12	LFS	LDN-193189	4
	LFS	K00986a	4
	JCSG	LDN-193189	20
	JCSG	K00986a	20
	HIN	LDN-193189	4
	HCS	K00986a	20

*LFS, JCSG, HCS and HIN refer to the crystallisation screens used to provide a range of conditions (see methods chapter 2.12.2).

3.2.2.2 Crystallisation and Data Collection

The BMPR1B-FKBP12 complex together with the chemical inhibitor LDN-193189 produced diffraction quality rod-like crystals from a reservoir condition in the coarse screen containing 20% PEG 3500, 0.2 M Na(malonate) pH 7.0 (figure 3.6). There were no crystals of BMPR1B alone. All datasets in this thesis were collected, processed and the resulting structures refined by staff at the Structural Genomics Consortium (SGC), except for the structure of ACVR1 Q207D in complex with LDN-213844, which was refined by myself and will be discussed in chapter 4. The dataset for the BMPR1B-FKBP12 complex

was collected at Diamond Light Source beamline I02 and refined at 2.05 Å. The BMPR1B-FKBP12 complex was crystallised in space group $P2_12_12_1$ with two protein complexes in the asymmetric unit.

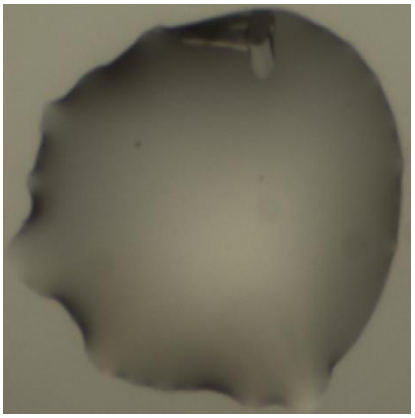


Figure 3.6 Rod-like crystals of BMPR1B-FKBP12. These crystals were observed after 2 days in a 150 nl drop containing 20% PEG 3500, 0.2 M Na(malonate) pH 7.0 and set up at a 2:1 ratio of protein to reservoir solution.

3.2.3 Structural analysis of BMPR1B in complex with FKBP12 and the chemical inhibitor LDN-193189

3.2.3.1 Overview of the BMPR1B-FKBP12 structure

The structure shows FKBP12 bound to the GS domain of BMPRIB as predicted. The GS domain consists of two helices (αGS1 and αGS2) joined by a loop region known as the GS loop. The kinase domain displayed the conserved kinase fold, consisting of two lobes; the smaller N-terminal lobe predominately formed of β strands and the larger C-terminal lobe predominantly α-helical, connected by a hinge region (D280-G285) (figure 3.7). LDN-193189 was bound to the hinge region of the ATP pocket (discussed further in Chapter 5.2.5.1). Throughout this thesis I have employed the commonly used 3.5 Å cutoff for hydrogen bonding and a cutoff of 4 Å for salt bridges, as described by Kumar and Nussinov [119]. In certain instances the side chains are likely to be flexible, such as on the

surface of the protein, so in these cases the bond is described as potential. The overall sequence identity between the solved BMPR1B and ACVR1 structure constructs is 64% (78% sequence similarity), however the overall root mean square deviation (rmsd) between the structure of BMPR1B-FKBP12 complex and ACVR1-FKBP12 complex (pdb:3h9r) is 1.79 Å. The rmsd between the FKBP12 molecules in each compound was only 0.367 Å, indicating that the main variation between the structure of complexes came from the receptors, with an rmsd of 1.71 Å.

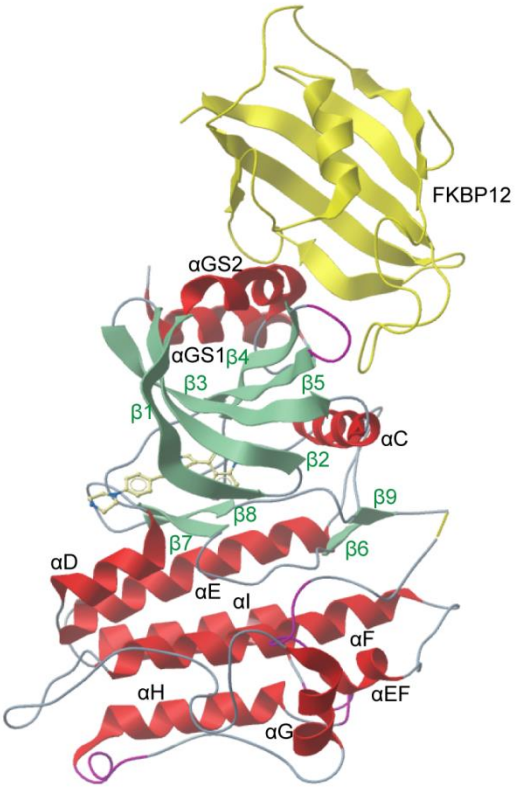


Figure 3.7 Crystal structure of BMPR1B in complex with FKBP12 and the inhibitor LDN-193189 (pdb: 3mdy). FKBP12 is shown in yellow and LDN-193189 is shown in stick representation. BMPR1B is coloured by secondary structure (α helices red, β strands green) and have been labelled accordingly.

FKBP12 is a known binder of the macrolides FK506 and rapamycin (figure 3.8A). The structure of TGFBR1 bound to FKBP12 showed the binding of a Leu-Leu pair in the GS domain to the same macrolide pocket in FKBP12 (figure 3.8B). In ACVR1 this is

replaced by a Phe-Leu pair (as shown by the alignment in figure 3.3) which results in a subtle shift of the binding of FKBP12 in order to accommodate the larger residue [87] (figure 3.8C). BMPR1B contains the conserved Leu-Leu pair, and showed similar binding to TGFBR1 (figure 3.8D).

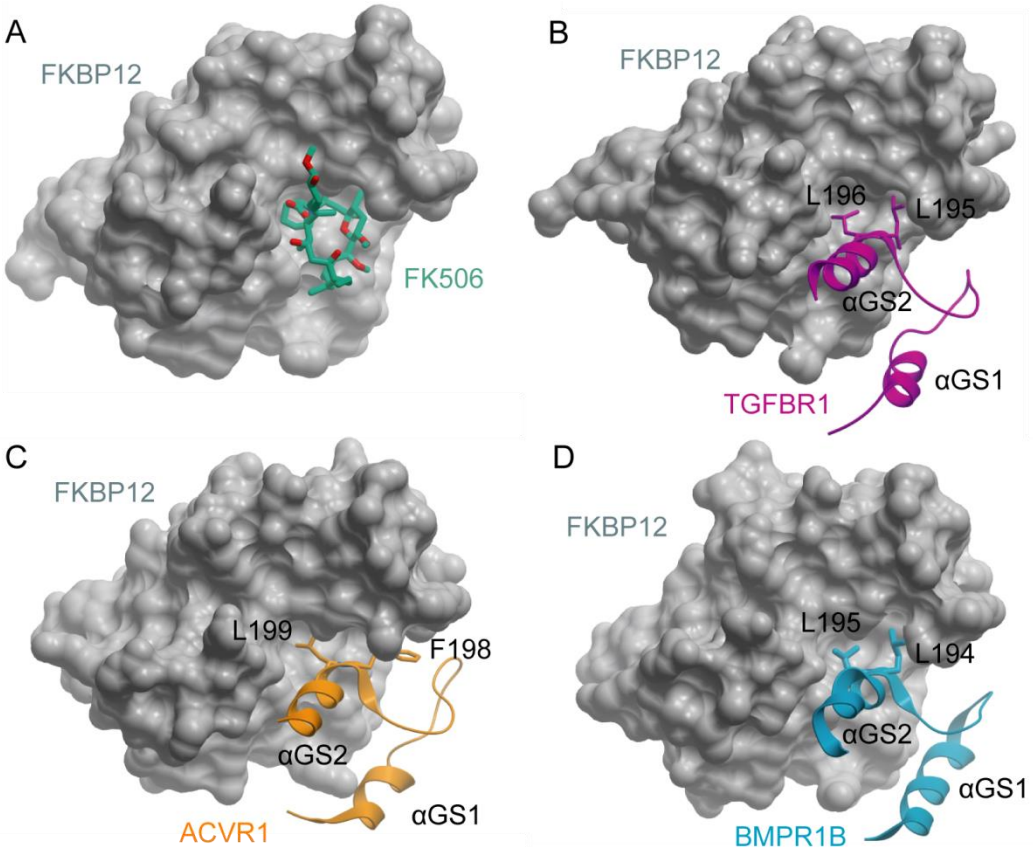


Figure 3.8 Binding of the immunophilin FKBP12. (A) Surface representation of FKBP12 displays a pocket which allows the macrolide FK506 to bind (pdb: 1fkj). (B) The Leu-Leu pair in TGFBR1 binds in the same site (pdb: 1b6c) (C) The equivalent residues in ACVR1, L199 and F198, bind in a slightly shifted position in order to accommodate the larger phenylalanine (pdb: 3h9r). (D) The conserved Leu194 and Leu195 allowed BMPR1B to bind FKBP12 in the same orientation as TGFBR1.

Analysis of the BMPR1B structure showed that K202 (equivalent to ACVR1 R206) did not form interactions with the β4 strand and was instead orientated away from the kinase domain (figure 3.9, and discussed later in figure 3.10). Moreover, key interactions at the equivalent FOP-associated mutation sites were changed across the BMPR1B structure.

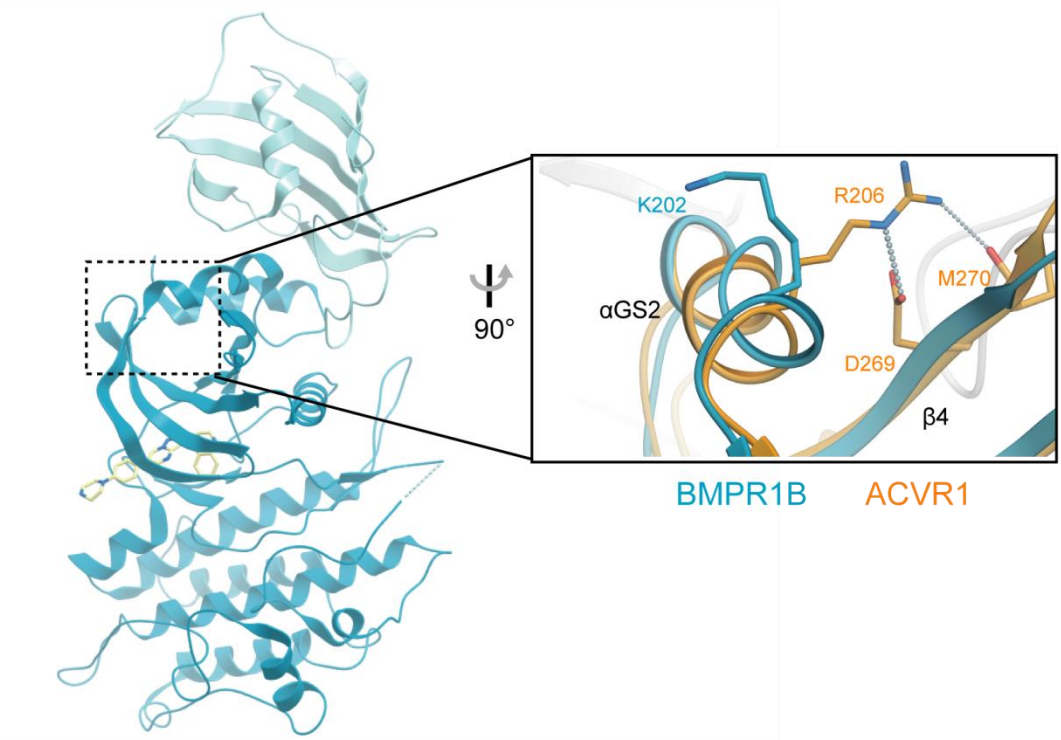


Figure 3.9 Interactions of BMPR1B K202 and the kinase domain. The superposition of BMPR1B (blue) and ACVR1 (orange – pdb: 3h9r) showed that the ACVR1 R206-equivalent residue in BMPR1B, K202, does not interact with the kinase domain.

3.2.3.2 The sites of FOP mutation show conformational shifts in BMPR1B

BMPR1B forms novel alternative interactions with FKBP12 compared to the earlier structures of ACVR1 and TGFBR1. In ACVR1 there are two αGS2 arginines, R202 and R206,that bind D269 in the β4 strand. Both of these are sites of FOP-associated mutations. R206 is the site of classic FOP mutation R206H, in which the interactions with D269 and M270 in the kinase domain are thought to be absent. The equivalent residue in BMPR1B is K202. Although BMPR1B K202 did not contact the kinase domain, it was positioned to form a potential salt bridge with FKBP12 E55 with a intermolecular distance of < 4 Å [119] (figure 3.10A). FKBP12 E55 was also involved in another salt bridge with the equivalent BMPR1B residue to the FOP mutation Q207E (figure 3.10A).

The other arginine, R202, has been found to mutate to isoleucine in FOP which would be unable to form the bonding interaction with D269. In the structure of BMPR1B the equivalent residue, R198, maintained this hydrogen bond but also formed interactions with FKBP12 H88 (figure 3.10B). Due to their proximity in αGS2, these bonds from BMPR1B R198 may have helped stabilize the GS-kinase domain interactions, despite the lack of interaction from K202.

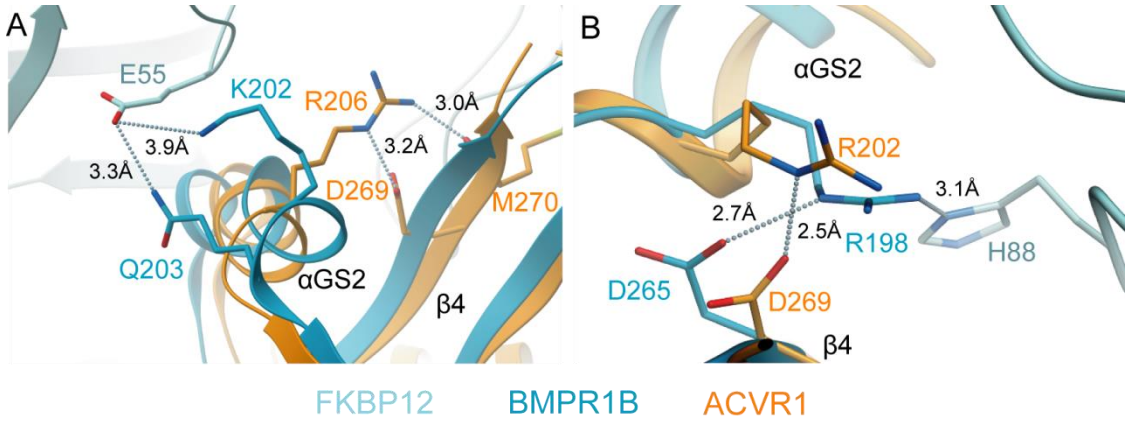


Figure 3.10 Binding of FKBP12 to BMPR1B. FKBP12 (light blue) forms two additional interactions with BMPR1B (blue) compared to ACVR1 (orange, pdb: 3h9r). (A) BMPR1B K202 is positioned to form a potential salt bridge with FKBP12 E55. (B) BMPR1B R198 forms a hydrogen bond with FKBP12 H88 in addition to a stabilizing hydrogen bond with the kinase domain.

Comparison of the ACVR1 and BMPR1B structures revealed a subtle shift in the relative positions of the αGS2 helix and GS loop. Hydrogen bonds from R258 to the GS loop are thought to be absent in the FOP mutation R258S which would destabilise the inhibitory conformation. In BMPR1B, the GS loop was not held by coordination of the equivalent residue R254 (figure 3.11A). Instead, the equivalent arginine in BMPR1B (R254) formed a hydrogen bond within the kinase domain to the main chain carbonyl group of F261 in β4. In addition, there is a subtle movement of the αC, which is a regulatory element often associated with kinase activation (figure 3.11B). In part, this could be due to the shift in conformation of the activation loop. In ACVR1, R375 in the activation loop forms

interactions with a serine (S244) in the αC and the aspartates from the DFG (D354) and HRD (D336) motifs, which would not be present in the FOP mutation R375P. In BMPR1B the equivalent arginine (R271) is flipped out to solvent and the interactions are lost (figure 3.11C).

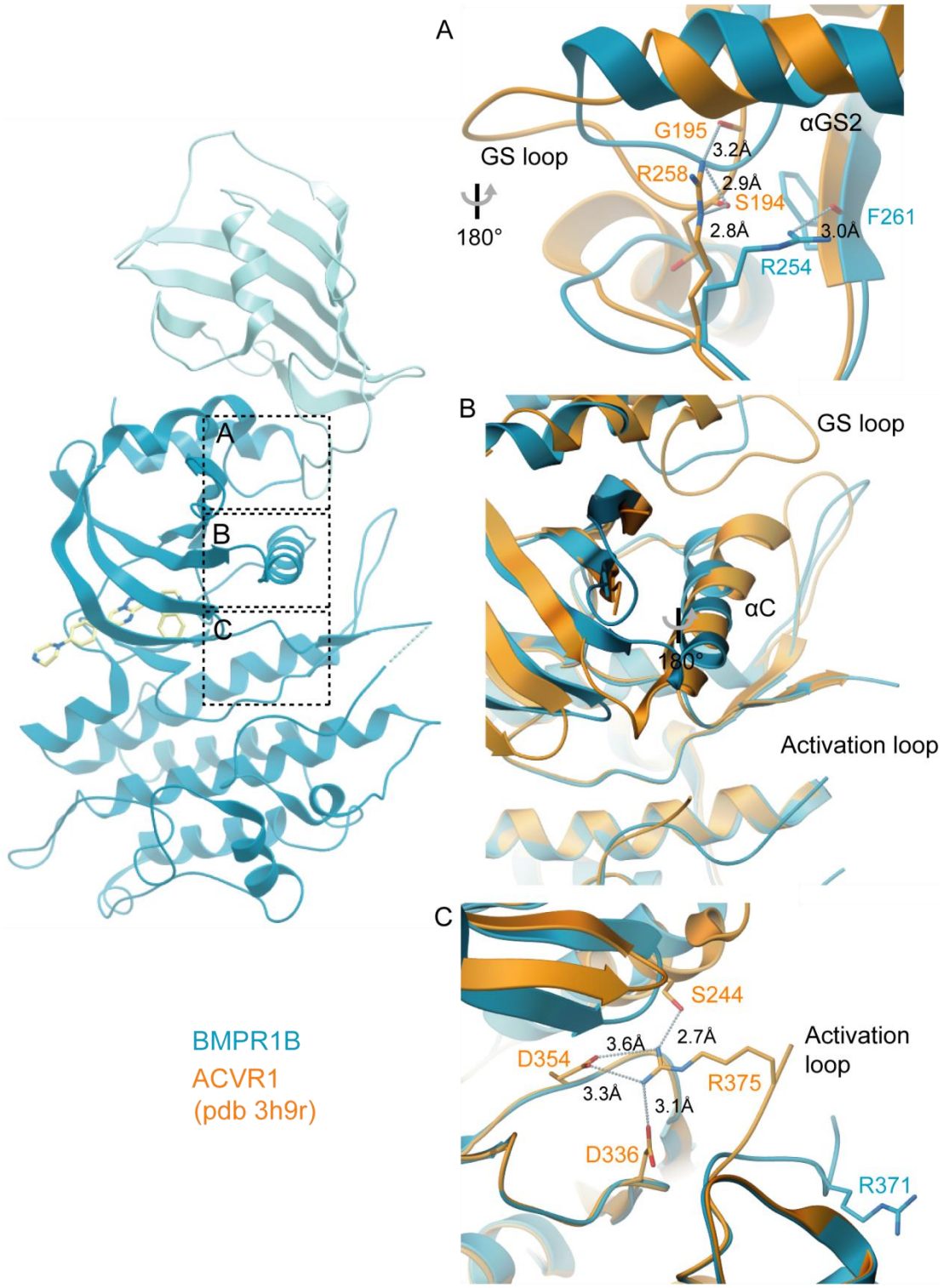


Figure 3.11 Structural comparison of BMPR1B (cyan) and ACVR1 (orange, pdb: 3h9r). (A) The GS loop in BMPR1B is less buried than the GS loop in ACVR1. ACVR1 R258S forms hydrogen bonds with the GS domain S194 and G195. The equivalent interactions are not formed in BMPR1B. (B) The αC helix of BMPR1B also shows a shift towards the active conformation. (C) ACVR1 R375 forms hydrogen bonds with a number of residues to form a "closed" orientation at the ATP pocket. In contrast, the BMPR1B activation loop is flipped out to the solvent in an "open" orientation, with an 8 Å difference between the two conformations.

3.2.2.3 BMPR1B remains inactive despite conformational changes

Despite the open conformation of the activation loop, the ATP pocket remained inaccessible to ATP as the P-loop was collapsed into the pocket. The inactive conformation was stabilized by hydrogen bonds from the conserved R214 (P-loop) to D332 (catalytic loop HRD motif) and T374 (A-loop) (figure 3.12). This is a novel alternative to the inhibitory ACVR1 R375 interaction seen in the previous BMP/TGFβ receptor structures. The kinase is of course also inhibited by its ability to bind to FKBP12, which prevents phosphorylation of the GS loop by the type II receptor, in turn restricting Smad substrate recruitment and activation of the Smad signalling pathway.

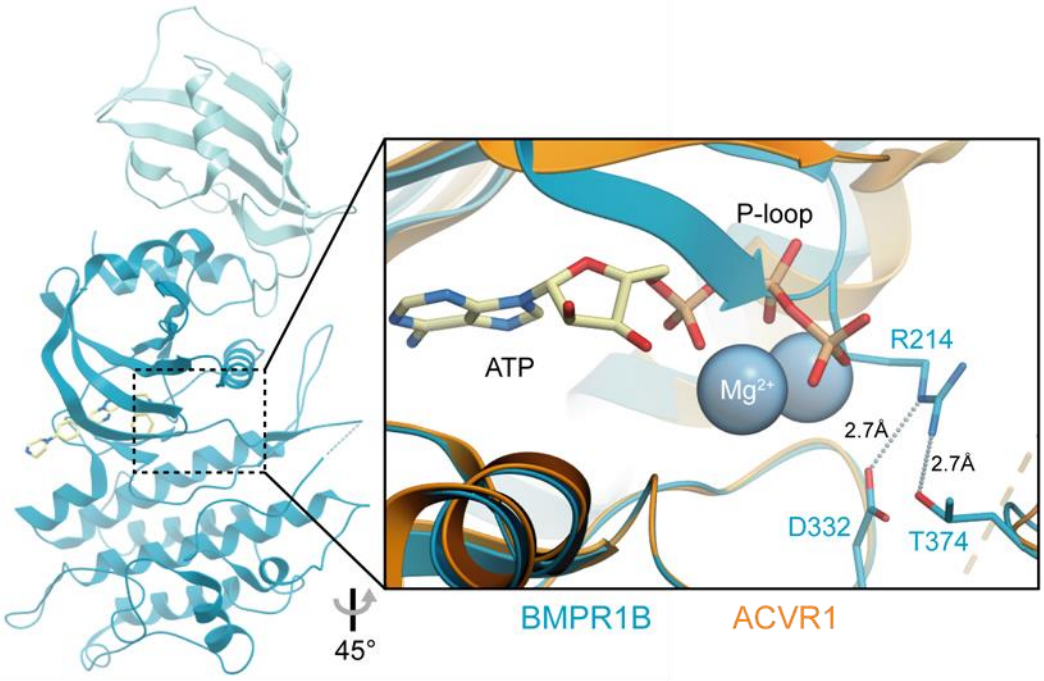


Figure 3.12 The inhibitory conformation of the BMPR1B P-loop. The BMPR1B (blue) ATP pocket is partly occupied by the position of the P-loop which is held in place by hydrogen bonds to D332 (HRD motif) and T374 (A-loop), compared to ACVR1 (orange). The C-alpha atoms of the equivalent arginines (BMPR1B R214 and ACVR1 R218) showed a distance of 6.7 Å. Modelling showed that BMPR1B would clash with ATP. ATP was modelled by superposition of the ATP-bound structure of PKA, pdb: 1ATP.

3.3 Discussion

The structural differences between the ACVR1 and BMPR1B complexes are surprising. Both receptors interact with the same type II receptors, ACVR2, ACVR2B and BMPR2, and are similarly regulated by the antagonist FKBP12. As a BMP receptor, BMPR1B is involved in chondrocyte differentiation and bone formation [42]. Missense mutations in BMPR1B are observed in the disease A2 brachydactyly [42,43], although no activating mutations in BMPR1B have been identified to date. A2 brachydactyly is characterised by the shortening of the fingers [43], which is somewhat akin to the malformation of the great toe found in FOP patients.

Crystals of BMPR1B were only obtained in the presence of FKBP12 and the inhibitor LDN-193189. This made it impossible to explore the proteins in the absence of any bound ligand. It is known that some ligands may stabilise the active or inactive conformation of kinases, for example Src tyrosine kinases form the active conformation in the presence of ligand [13]. Here, the presence of FKBP12 would restrict the active conformation by binding the GS domain. However, the movement of the proposed regulatory GS loop and the αC whilst in the presence of FKBP12 was of interest. Small molecule inhibitors can also affect the conformation of the ATP binding site [90]. However, here the compared structures are shown bound to two similar ATP mimetic and competitive inhibitors.

In comparison to ACVR1, the structural changes to the GS loop, αC and A-loop of BMPR1B are indicative of a shift towards the active state of the receptor. Such changes are suggestive of how single missense mutations in FOP could alter the ACVR1 kinase conformation. Without the inhibitory conformation of the P-loop observed in the BMPR1B structure, this shift towards the active conformation would likely increase the

ability for ATP and substrate to bind, allowing the mild increase in ACVR1 receptor activation and signalling observed in FOP.

The subtle shifts propagated throughout the BMPR1B structure in comparison to ACVR1 are consistent with the hypothesis that the R206H mutation results in destabilization of the inactive conformation. Loss of interactions between the GS and kinase domains from a variant of ACVR1 R206 resulted in the shift of the GS loop from its buried position within the kinase domain. With FKBP12 bound in the BMPR1B structure, the serines in the GS loop are not fully exposed. While the BMPR1B receptor retains this FKBP12 interaction, it has been observed previously that ACVR1 R206H mutants show weaker binding to this inhibitory protein [120]. However, if the FOP mutation R206H results in a similar structural shift as observed in BMPR1B, the reduced FKBP12 binding would expose the GS loop. This could result in type II phosphorylation and substrate binding in this region, in turn promoting downstream Smad signalling.

Intriguingly, all of the sites of conformational change when comparing the ACVR1-FKBP12 structure to this new BMPR1B-FKBP12 structure equate to sites of FOP mutation. This provides clues to the conformational changes that could result as consequence of such mutations. The structure of BMPR1B has provided an intriguing model of the structural effects FOP mutations could have at various locations in the ACVR1 receptor. However, it has also identified an alternative inactive conformation of the P-loop that could also be accessible to ACVR1 and might explain the very mild receptor activation observed in FOP. Further crystallographic studies of the FOP mutants were conducted to explore this hypothesis.

4. Characterisation of ACVR1 FOP mutants

4.1 Introduction

FOP-associated mutations result in the mild activation of ACVR1. To date, all structures of ACVR1 and other BMP receptors are of the inactive conformation. Activation of the type I receptor can result in the stimulation of various signalling pathways. The bone formation observed in FOP is thought to require the Smad1/5/8 pathway [120]. This is achieved by the phosphorylation of R-Smad C-terminal serines by type I receptors, such as ACVR1. Although no structure is resolved for a receptor-Smad complex, various factors involved in Smad binding have been determined. Mutagenesis and domain swap experiments suggest that the interaction involves the L3 loop in the Smad MH2 domain and the L45 loop in the type I receptor kinase domain. R-smads are structurally similar to the FHA domain which binds phosphothreonine sites. Therefore, it is also expected that the R-smad interface will extend to the phosphorylated GS loop of the type I receptor, likely explaining the requirement for this phosphorylation site. The SXS motif at the distal C-terminal tail of the R-Smad can then be phosphorylated at the kinase active site [25,58]. Although the regions critical for binding have been mapped, the actual mode of substrate docking remains unknown, and so it is difficult to predict the structural effects of FOP mutations on Smad binding. It is also unknown whether the FOP mutants are capable of forming a complex with Smad independent of GS phosphorylation and, if so, how efficient this binding is at stimulating Smad signalling. Therefore, the Smad binding efficiency of both wild-type and mutant ACVR1 are of interest, as well as their required phosphorylation states and structural comparisons.

From the structures of both the ACVR1-FKBP12 complex and the BMPR1B-FKBP12 complex it is predicted that FOP mutations would destabilize the inactive state to induce a mild activation of the receptor. The R206H mutation is predicted to break hydrogen bonds with D269 and M270 in the β4 strand of the kinase N-lobe. The tight interaction between the GS and kinase domains at this position is likely to be critical to stabilize the buried GS loop between the β4 and αC when bound to FKBP12. The structural differences between BMPR1B and ACVR1 bound to FKBP12 (chapter 3) agreed with this hypothesis, although to date there are no structures of mutant ACVR1 nor any mutant BMP or TGFβ receptor kinase. The classic mutation R206H has also been shown to weaken the inhibitory binding of FKBP12, which is thought to contribute to leaky signalling [77]. This has been shown both *in vitro,* using analytical gel filtration and optical biosensor studies, and *in vivo* studies, including determining the effects of FKBP12 inhibition on BMP-6-induced BRE-luc activity in Bovine Aortic Endothelial Cells (BAECs). However, the other mutations observed in FOP have not been studied in this respect.

To investigate the effects of FOP mutations on the catalytic state of the protein, the mutant receptors were generated. Crystal structures were obtained to understand the structural effects of mutation. Further experiments were conducted to understand activation caused by mutation *in vitro*. In order to assess the potential loss of FKBP12 binding in FOP, the mutants and wild type were assayed using analytical gel filtration. Mass spectrometry was also employed to understand the phosphorylation requirements of both the receptors and substrate Smad.

4.2 Results

4.2.1 Mutant generation

Primers were designed to generate the FOP mutants of ACVR1 using site directed mutagenesis, either by the QuikChange kit (Stratagene) or a "megaprimer" method (methods section 2.2). The mutations were initially made in the full length ACVR1 coding region and cloned into the pcDNA3-NFlag vector to permit their concurrent use in cellular assays by other colleagues. All mutations were confirmed by DNA sequencing. Ten individual FOP mutations were successfully introduced, including L196P, P197_F198delinsL, R202I, R206H, Q207E, R258S, G328E, G328R, G356D and R375P, as well as the artificial constitutively active mutation Q207D. The ACVR1 Q207D mutation was initially identified by scanning mutation to examine potential phosphorylation sites. The aspartate in this position was shown to be activating in an *in vitro* luciferase response assay [121]. It has since been further characterized by FOP-like bone formation following its transgenic over-expression in mouse models [94].

Various baculoviral expression constructs were then prepared for them by ligation-independent cloning focussing on the cytoplasmic GS and kinase domains, as previously solved for wild-type ACVR1. Additional constructs were prepared for the kinase domain only to maximise the chances of soluble protein expression and subsequent crystallisation. Constructs in the pFB-LIC-Bse vector were then transformed into DH10bac^{TM} cells for the production of bacmid DNA, which is needed to transfect Sf9 insect cells for the generation of P1 baculoviruses and a 3ml test expression (chapter 2.4). A summary of the constructs is shown in table 4.1. All mutants showed some expression, although this was low for G356D and R202I (figure 4.1). The soluble proteins were tested by mass spectrometry and the viruses confirmed by DNA sequencing.

Table 4.1 Table of mutant ACVR1 generation.

Mutant	Domains	Construct	Cloned into pFB-LIC	Bacmid DNA	Soluble protein
WT[a]	GS + Kinase	172-509	✓	✓	✓
		172-499	✓	✓	✓
	Kinase	201-509	✓	✓	✓
		201-499	✓	✓	
L196P	GS + Kinase	172-509	✓	✓	✓
		172-499	✓	✓	
P197_F198delinsL	GS + Kinase	172-509	✓	✓	
		172-499			
R202I	GS + Kinase	172-509	✓	✓	✓ (low)
		172-499			
R206H	GS + Kinase	172-509	✓	✓	
		172-499	✓	✓	✓
	Kinase[b]	201-499	✓	✓	✓
Q207D	GS + Kinase	172-509	✓	✓	✓
		172-499	✓	✓	✓
	Kinase[b]	201-499	✓	✓	✓
Q207E	GS + Kinase	172-509	✓	✓	✓
		172-499	✓	✓	
R258S	GS + Kinase	172-509	✓	✓	✓
		172-499	✓	✓	
	Kinase	201-509	✓	✓	
		201-499	✓	✓	
G328E	GS + Kinase	172-509	✓	✓	✓
		172-499	✓	✓	
	Kinase	201-509			
		201-499			
G328R	GS + Kinase	172-509	✓	✓	
		172-499			
	Kinase	201-509	✓	✓	
		201-499	✓	✓	
G356D	GS + Kinase	172-509	✓	✓	
		172-499			
	Kinase	201-509	✓	✓	
		201-499	✓	✓	
R375P	GS + Kinase	172-509	✓	✓	✓
		172-499			
	Kinase	201-509			
		201-499	✓	✓	

Ticks denote that this stage was reached with the construct.

[a]Although soluble constructs were reproduced, the previously crystallised construct of WT ACVR1 (a.a. 172 - 499) was used throughout the thesis. [b]These constructs were designed by Alex Bullock. Their test expression was described in Chapter 3.1.

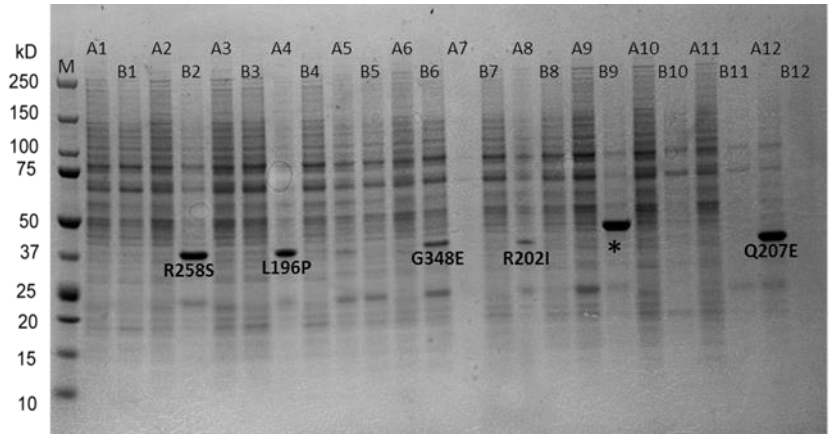


Figure 4.1 Test expression of a number of the mutant ACVR1 constructs. The Ni-sepharose eluted fractions of constructs number A1 – B12 were run on SDS PAGE to show soluble proteins. *indicates a contaminant.

Following the test expression, all the P1 viruses, except for R202I and G356D which had low expression, were amplified to produce P2 viruses for larger scale studies. 50 ml expressions were collected at both 48 and 72 hours, and tested for optimal expression time, which was determined as 72 hours (figure 4.2). The ACVR1 mutant proteins were then expressed and purified at large scale for use in biophysical characterisation and crystallisation studies. Proteins containing the kinase domain mutations R258S, G348E or R375P were observed to precipitate during purification presenting a challenge for the preparation of these crystallisation trials.

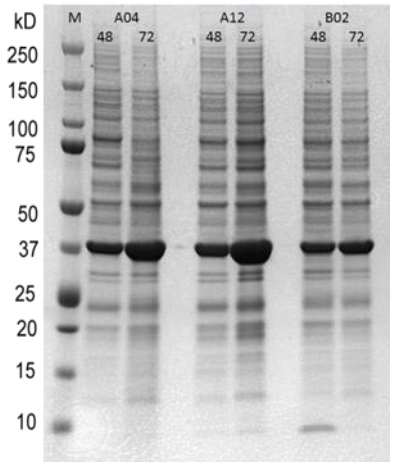


Figure 4.2 Expression time-point testing. Elution fractions from a test expression of L196P (A04), Q207E (A12), and R258S (B02) at 48 and 72 hours post infection. The highest protein yield was observed after 72 hours, and so this was used for all future expressions.

4.2.2 Crystallographic structures of ACVR1 mutants

4.2.2.1 Crystal structure of FOP mutant ACVR1 Q207E containing both GS and kinase domains displays an inactive kinase conformation

A number of the purified GS and kinase domain FOP mutants were put into crystallisation trials (table 4.2). The ACVR1 Q207E mutant in complex with the chemical inhibitor K02288 (but not FKBP12) produced diffraction quality pyramid-like crystals from a reservoir condition in the coarse HCS screen (methods chapter 2.12.2) condition containing 0.1M MES pH 6.5 and 12% PEG 20000 (figure 4.3).

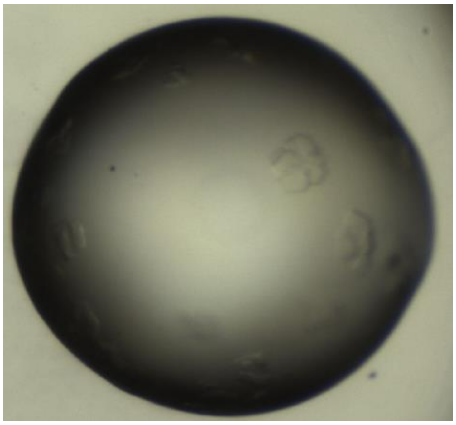


Figure 4.3 Crystals of ACVR1 Q207E (a.a. 172-509) with K02288. The crystals were grown at 20°C, in a sitting drop with a mix of 50 nl protein solution and 100 nl of a reservoir solution containing 0.1 M MES pH 6.5 and 12% PEG 20000.

The structure of both the GS and kinase domain of FOP mutant ACVR1 Q207E (a.a. 172-509) was solved in complex with the inhibitor K02288. The diffraction data were collected by SGC staff at Diamond Light Source beamline I03. The structure has been refined at 2.1 Å by Eleanor Williams at SGC (PDB deposition pending). The complex crystallised in space group C121 with a single protein chain in the asymmetric unit.

Table 4.2 Table of crystallisation trials for the various ACVR1 mutants.

ACVR1 mutation	Construct (a.a)	Concentration (mg/ml)	Screen	Compound	Temperature (°C)
L196P	172 - 509	9.0	JCSG	K02288	4
			JCSG	LDN-193189	20
			LFS	K02288	20
			LFS	LDN-193189	4
			HIN	LDN-193189	4
			HCS	K02288	20
R206H	172 - 499	6.0	JCSG	K02288	20
			HIN	K02288	4
			HCS	K02288	4
Q207D	172 - 499	7.5	JCSG	K02288	20
			JCSG	AMPPNP	4
			LFS	K02288	4
			LFS	AMPPNP	20
			HIN	K02288	20
			HIN	AMPPNP	4
			HCS	K02288	4
			HCS	AMPPNP	20
Q207E	172 - 509	10.0	JCSG	K02288	4
			JCSG	K02288	20
			JCSG	LDN-193189	4
			JCSG	LDN-193189	20
			LFS	K02288	4
			LFS	K02288	20
			LFS	LDN-193189	4
			LFS	LDN-193189	20
			HIN	K02288	4
			HIN	LDN-193189	20
			HCS	K02288	20
			HCS	LDN-193189	4
R258S	172 - 509	20.0	JCSG	K02288	20
			LFS	K02288	4
			HIN	K02288	20
			HIN	K02288	4
			HCS	K02288	20
			HCS	K02288	4
R375P + FKBP12*	172 - 499	6.6	JCSG	LDN-193189	20
			LFS	LDN-193189	4
		7.5	JCSG	LDN-193189	4
			LFS	LDN-193189	20
			HIN	LDN-193189	4
			HCS	LDN-193189	20

*R375P precipitated on concentration in the absence of FKBP12. The preparation of the complex was the same as described for BMPR1B-FKBP12 in chapter 3.

The overall kinase domain fold was essentially unchanged compared to the earlier structure of ACVR1 in complex with FKBP12 (figure 4.4A). The electron density allowed the complete chain to be traceable from S178, except residues 186 – 190 in the GS loop. The site of mutation, Q207E, was ordered and maintained a hydrogen bond interaction with W227 in the kinase domain (figure 4.4B). However, the GS loop was largely disordered and higher B-factors were observed for the surrounding ordered regions compared to the wild-type structure. This suggests that there is significant flexibility in this region which is likely to expose the serines in the GS loop to phosphorylation (figure 4.4C). The other sites of activation, the αC helix and the activation loop were in the inactive conformation observed in the wild type structure, although the activation loop was ordered in this structure, forming two β-strands. Other sites of FOP mutation still formed the inhibitory interactions observed in the wild-type structure, such as R206 binding D269 and M270, in contrast to the BMPR1B structure (chapter 3).

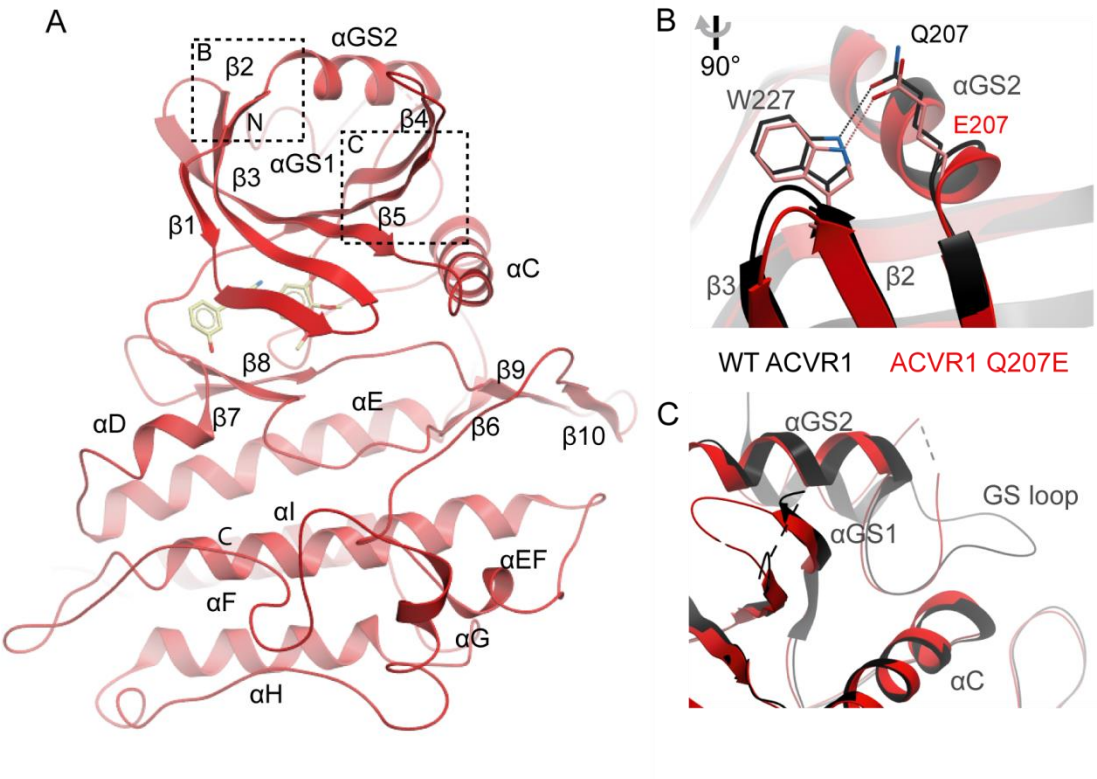


Figure 4.4 Overview of ACVR1 Q207E crystal structure (red). (A) The structure shows the same overall conformation as the wild-type ACVR1 receptor, but reveals also an ordered β9-β10 hairpin in the activation segment. (B) Superposition of the wild-type ACVR1 structure (black; pdb 3h9r) showed that the interaction between the mutated Q207E residue and W227 in the kinase domain was conserved (C) The GS loop in the mutated structure was largely disordered.

4.2.2.2 Kinase domain structures of ACVR1 Q207D and ACVR1 R206H mutants also adopt an inactive conformation

A number of structures of the ACVR1 kinase domain were solved to determine the binding mode of various compounds which will be discussed further in chapter 5. Constructs of ACVR1 comprising residues 201 to 499 were found to crystallise readily and contained the R206H or Q207D mutation for additional investigation of the mutant state.

ACVR1 Q207D is commonly mutated by recombinant methods to generate an artificial highly active mutant receptor [122]. The Q207D mutation is predicted to break hydrogen bond interactions with E55 in FKBP12 as well as with W227 in the ACVR1 kinase domain, but the structural effects on the catalytic domain remains unknown. A kinase domain construct of ACVR1 (a.a. 201 – 499) containing the Q207D mutation but lacking most of the GS domain was purified and screened for crystallization following the protocol described in chapter 2.12.1.

Diffraction quality crystals were obtained with four different inhibitors, K00507, K02288, LDN-193189 and LDN-213844 (figure 4.5, and chapter 6.6 for LDN-213844 (a derivative of K02288 with a piperazine group and lacking the primary amine). Datasets were collected at Diamond Light Source by members of SGC staff. The first three of these complexes were refined by Apirat Chaikaud (SGC) whilst I refined the structure of ACVR1 Q207D in complex with LDN-213844. These compounds and their binding will be discussed in more detail in chapters 5 and 6. Crystallisation and structure determination of ACVR1 Q207D with LDN-213844 is detailed in chapter 6.

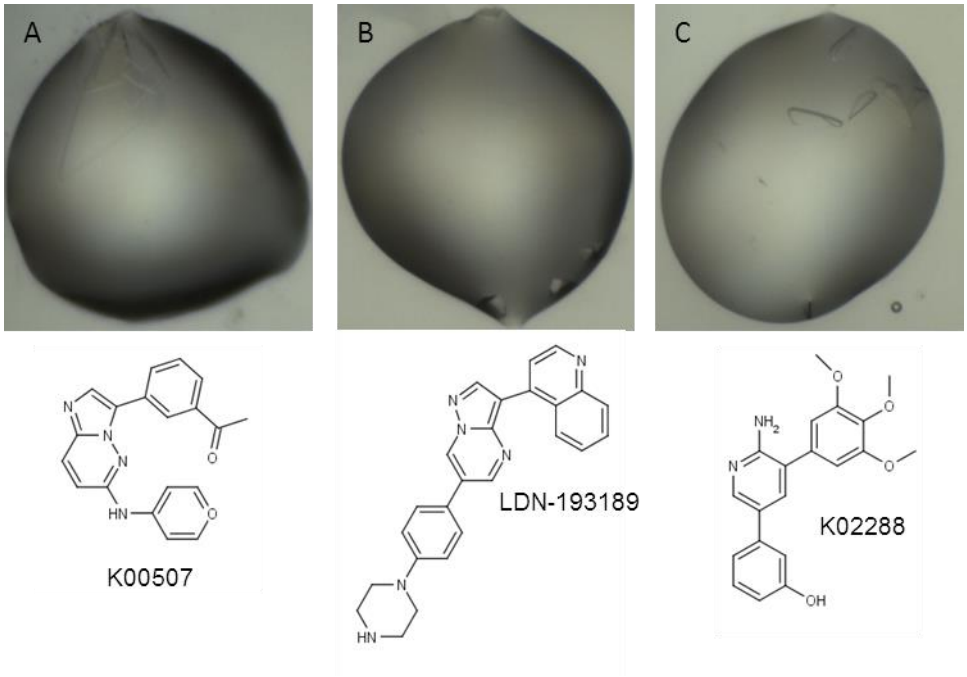


Figure 4.5 Crystals of ACVR1 Q207D with a range of inhibitors. (A) ACVR1 Q207D in complex with K00507. The crystals were grown at 20°C, in a sitting drop with a mix of 50 nL protein solution and 100 nL of a reservoir solution containing 0.20M Na/KPO_4, 20.0% PEG 3350 and 10.0% ethylene glycol. (B) ACVR1 Q207D in complex with LDN-193189. The crystals were grown at 20°C, in a sitting drop with a mix of 75 nL protein solution and 75 nL of a reservoir solution containing 20% PEG 3350 and 0.20M $(NH_4)_2H$(citrate). (C) ACVR1 Q207D in complex with K02288. The crystals were grown at 20°C, in a sitting drop with a mix of 50 nL protein solution and 100 nL of a reservoir solution containing 1.6 M Na/KPO_4, 0.1M HEPES pH 7.5

Overall, the four structures are almost identical to the Q207E ACVR1 structure and the wild-type ACVR1 structure (figure 4.A). The conformation of the β4 and β5 strands (the L45 loop), however, is strikingly different in both the LDN-193189 complex and the LDN-213844 complex compared to the other structures. This change results from a novel crystallographic dimer interface formed across the L45 loops of two adjacent ACVR1 molecules in the crystal packing (figure 4.6B). Interestingly, the observed dimer interface is sterically incompatible with FKBP12 binding. Therefore, it is interesting to speculate whether such a conformation could be adopted by the L45 loop as it switches from the FKBP12 complex to Smad recognition. Alternatively, it should be taken into consideration that the two structures containing the alternative L45 loop conformation

were crystallised bound to inhibitors containing a piperazine group (LDN-193189 and LDN-213844) which extend towards the solvent. These inhibitors could be influencing crystal contacts within the asymmetric unit that affect this conformation.

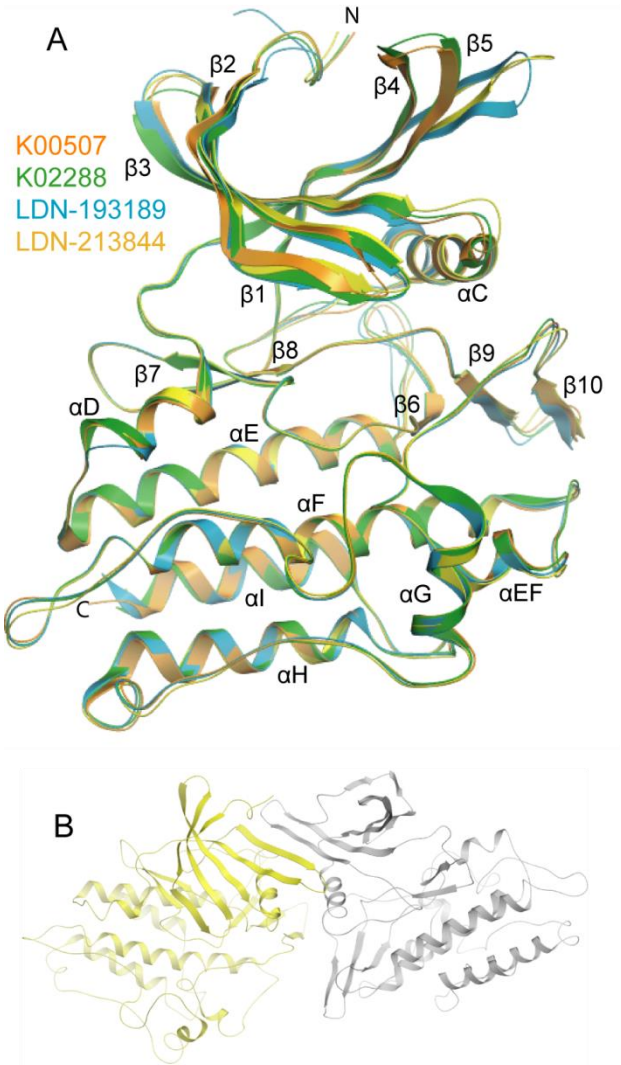


Figure 4.6 Crystallographic structure of ACVR1 Q207D. (A) Overview of the structures of ACVR1 Q207D (a.a. 201-499). The four structures are overlaid in complex with K00507a (orange; pdb: 3OOM), K02288a (green; pdb: 3MTF), LDN-193189 (blue; pdb: 3Q4U) and LDN-213844 (yellow; pdb: 4BGG). The inhibitors have been omitted for clarity. The only significant structural change is at the L45 loop. (B) Two molecules of the ACVR1-LDN-213844 complex packed as a crystallographic dimer in the asymmetric unit, the other two as monomers.

As predicted the mutated residue Q207D is unable to form the same interaction with W227 as observed in the previous wild type or Q207E structures (figure 4.7A) due to the shorter aspartate side chain. Interestingly, Q207D is known to be a more potent activating mutation than those identified in FOP [87]. Despite this, all four mutant complexes presented here show a comparable inactive kinase conformation to wild type as demonstrated by the interactions formed by the A-loop residue R375 (figure 4.7B). However, one of the four protein molecules within the asymmetric unit of the ACVR1-LDN-193189 complex shows a shift in conformation in both the P-loop and A-loop, suggesting a relaxation of the inactive state (figure 4.7C). So, although the mutation does not result in a fully active receptor, the subtle changes suggest that the inactive conformation is destabilised and therefore more dynamic following mutation than in the wild-type receptor. Combined with the reported disruption to FKBP12 binding and the alterations at the GS and L45 loops, these conformational changes could explain the mild activation of ACVR1 in FOP.

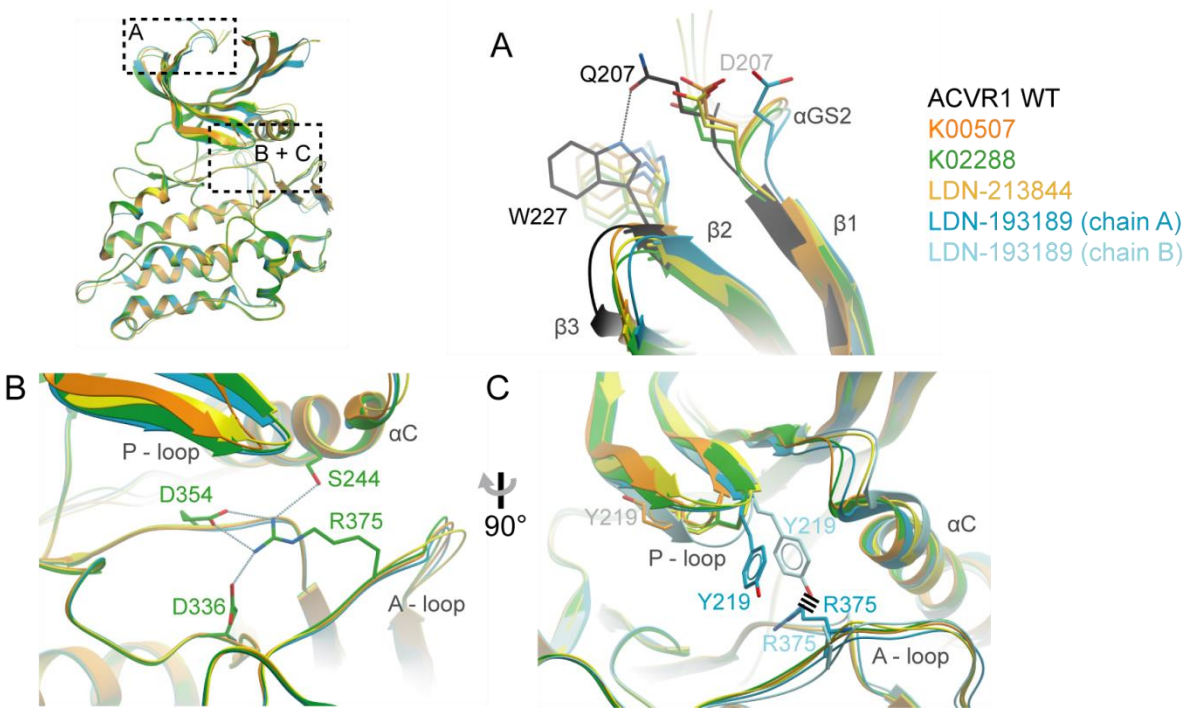


Figure 4.7 Interactions within the ACVR1 Q207D structure. (A) Interactions between Q207 and W227 are broken by mutation to D207. (B) Inhibitory interactions with R375 in the activation loop (A-loop) are generally conserved in the structures. All the mutants show the same conformation but here only the side chains from the K02288 complex structure are shown for clarity. (C) In the complex with LDN-193189, the P-loop residue Y219 is flipped out towards an active conformation (light blue, chain B) in contrast to the inactive conformation observed in other ACVR1 structures. Moreover, this P-loop conformation remains compatible with ATP binding, unlike the conformation observed in BMPR1B. The LDN-193189 co-structure also lacked electron density for the A-loop residue R375. Modelling suggested that the typical inhibitory conformation of R375 would be sterically incompatible due to a clash (shown by the black dashed line) with Y219. Again this reflects a shift away from the inactive conformation, similar to the BMPR1B A-loop in chapter 3 (figure 3.9C).

R206H is the most recurrent mutation observed in FOP [7]. Mutation of this site is predicted to break R206 interactions with D269 and M270 in the kinase domain, as described in chapter 3. Diffraction quality crystals were formed at a 2:1 precipitant to protein ratio in 1.6M $MgSO_4$, 0.1M MES pH 6.5 at 20°C with the inhibitor K00135 (figure 4.8). The structure of the ACVR1 R206H (a.a. 201 – 499) complex was solved subsequently by Apirat Chaikuad at SGC at a resolution of 2.42 Å. The description of the inhibitor binding is discussed further in chapter 5.

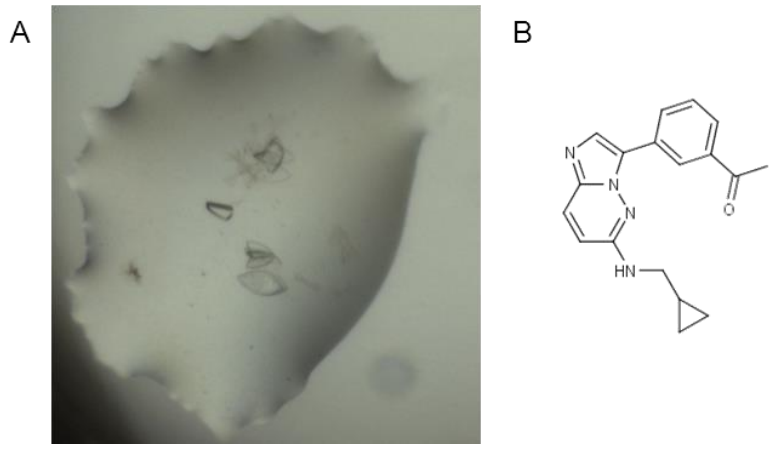


Figure 4.8 Crystals of ACVR1 R206H in complex with K00135. Crystals were grown at 20°C, in a sitting drop with a mix of 75 nl protein solution and 75 nl of a reservoir solution containing 1.60M $MgSO_4$ and 0.1M MES. (B) Scaffold of K00135.

The overall structure was very similar to that of the Q207D structure bound to K00507 (calculated Cα RMSD was 0.49 Å)(figure 4.9A). Unfortunately, the R206H residue was disordered. However, this in itself confirmed that the stabilising interaction of this position with the kinase domain observed in the wild-type receptor was broken by the mutation. Interestingly, the kinase domain retained the inhibitory conformation observed in wild-type ACVR1 rather than the alternative conformation of BMPR1B (figure 4.9B). The A-loop of ACVR1 R206H adopted a closed position with the αC helix positioned into the ATP pocket. The P-loop in ACVR1 R206H also adopted a similar orientation to the other ACVR1 structures with the residue Y219 blocking access to the ATP pocket, compared to the entire P-loop of BMPR1B which folded into the ATP pocket forming an unusual inhibitory conformation (figure 4.9B).

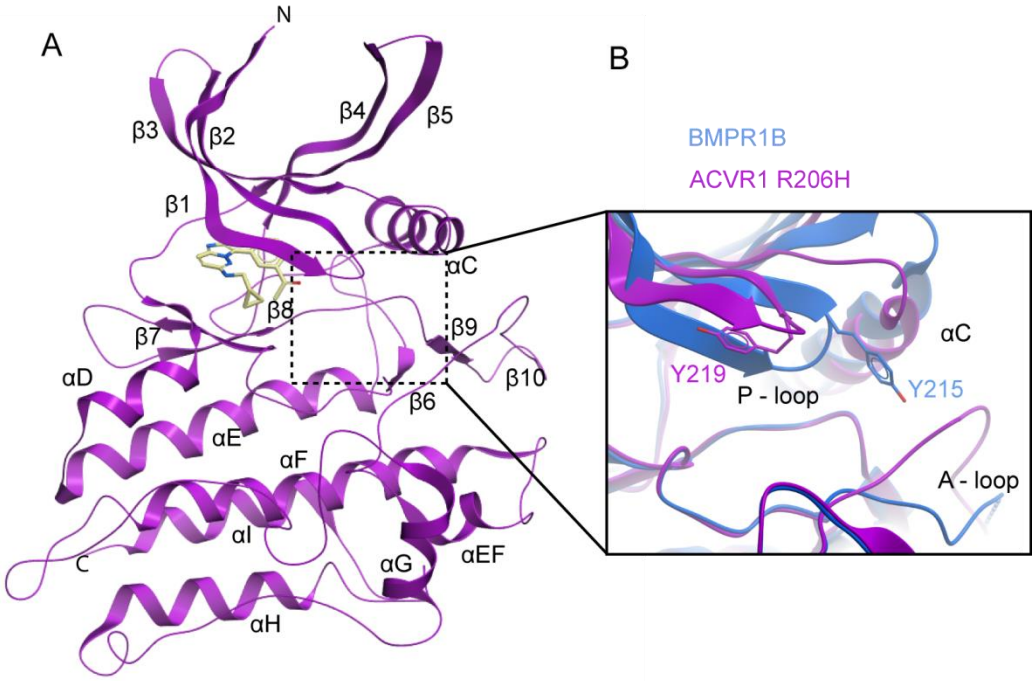


Figure 4.9 Crystal structure of the kinase domain of ACVR1 R206H (a.a. 201- 499). (A) Overview of the structure (pdb: 4DYM). (B) Comparison of ACVR1 R206H (purple) and BMPR1B (blue). The observed conformation matched that of the wild-type ACVR1 rather than BMPR1B. For example, the A-loop was in the "closed" conformation.

4.2.3 ACVR1 mutant interaction with FKBP12

The transient binding of FKBP12 to the ACVR1 mutant R206H has often been cited as an explanation for the leaky activation of receptor activation in FOP. Although the binding of R206H to FKBP12 has been characterized, the interaction of other mutants has not been studied. FKBP12 binds to the GS domain, so mutations in this domain are predicted to reduce FKBP12 binding, as has been shown for R206H. However, the effect of kinase domain mutants is unknown.

To assess FKBP12 interactions with the ACVR1 mutants, analytical gel filtration was used to observe the potential shift in retention time indicative of protein binding. If the receptor was able to bind then the increase in size of the complex, compared to the individual proteins, would reduce the elution volume and shift the peak to the left. FOP mutants L196P, R206H, Q207E, R258S, G348E and R375P were tested for their binding

to FKBP12. As a control, wild-type ACVR1 was also measured. Each ACVR1 variant was injected at a concentration of 50 μM, whilst FKBP12 was in slight excess at 75 μM to ensure any uncomplexed ACVR1 was not due to loading artefacts. The receptors and FKBP12 were passed through a Shodex KW-803 column on a Dionex micro-HPLC system at a flow rate of 0.5 mL/min, both as complexes and as single proteins. All experiments were performed in duplicate. The samples were run in standard gel filtration buffer (methods chapter 2.7) at 4 °C. Despite each receptor being prepared to the same concentration, the absorbance of R375P was far lower than observed for the other proteins. This is probably due to precipitation of the mutant over the course of the experiment.

Overlay of the chromatograms for the complexes and each protein separately allowed comparison of the elution points (and therefore the apparent size of the eluted protein) for each variant of the receptor (figure 4.10). When mixed with FKBP12, the wild-type ACVR1 showed the greatest shift with 0.25 mL reduction in retention volume compared to the receptor alone retention volume of 11.2 mL. Surprisingly, L196P had no change in retention volume whether ran alone or mixed with FKBP12, indicating that no binding between the proteins had occurred. This is consistent with L196P's position adjacent to the FKBP12 binding pocket (chapter 3, figure 3.8). G328E showed nearly the same shift observed for wild type with a 0.2 mL reduction in retention time. All the other mutants displayed changes in retention volume that lay somewhere between G328E and L196P. The variation in retention volume likely reflects the equilibrium of the population bound to FKBP12 for each mutant. Attempts were made to quantify the affinity between mutant ACVR1 and FKBP12. One method was isothermal titration calorimetry (ITC), however this was problematic as FKBP12 aggregated in both cell and syringe. Additionally an Octet optical biosensor (ForteBio) was used however non-specific binding was observed

creating a high background. Both methods could potentially be optimised for further investigation. Alternatively differences in affinity between FKBP12 binding ACVR1 compared to the various ACVR1 mutants could be measured using electrophoretic mobility shift assays or by fluorescence resonance energy transfer.

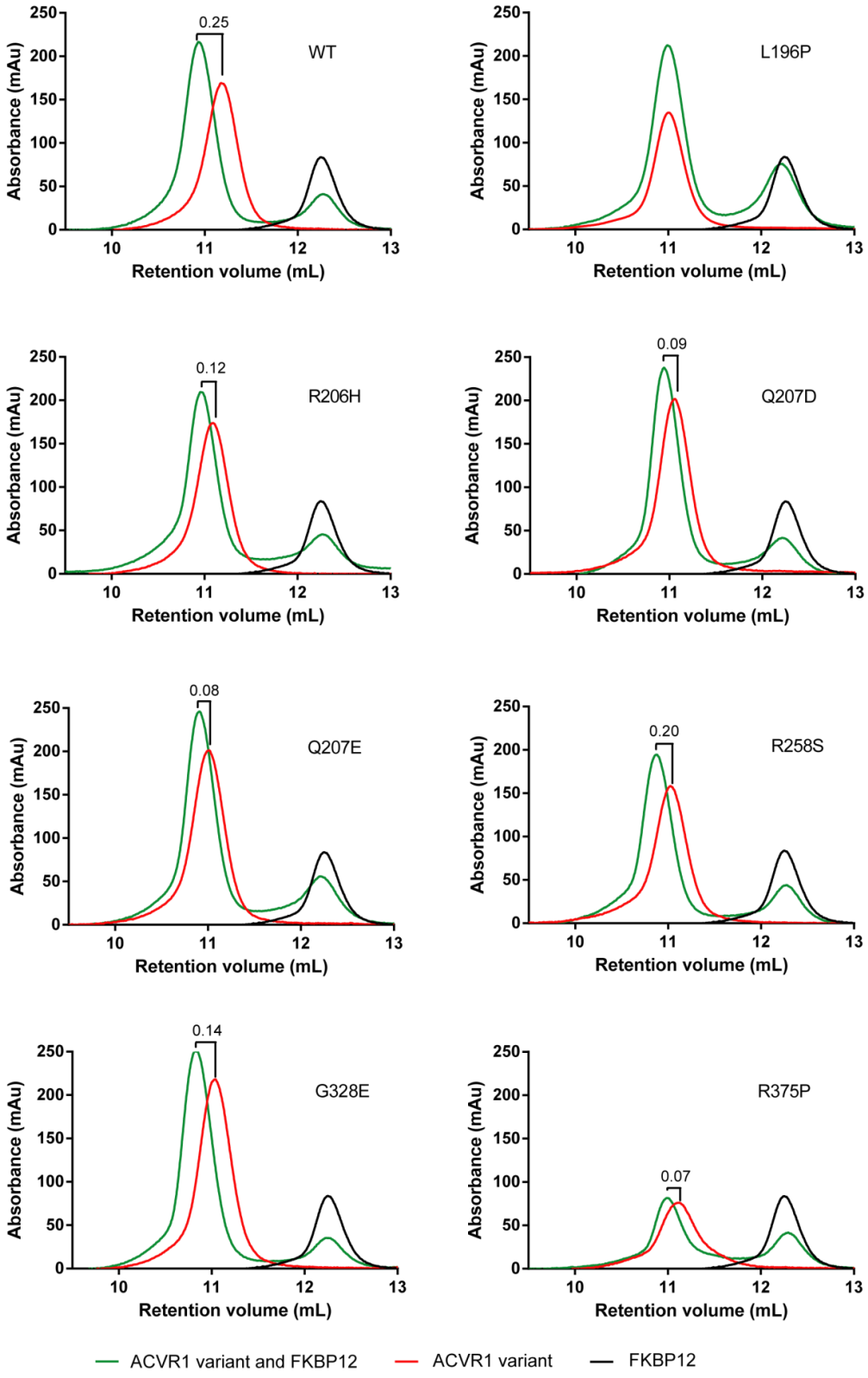


Figure 4.10 Analytical gel filtration chromatograms of the various FOP mutants binding FKBP12. FKBP12 alone is represented by a black line, receptor alone by a red line, and the mixed sample of both proteins is green. The difference in elution volume is shown on each graph. All samples were run twice and showed excellent replication. The average variation was only 0.017ml giving a signal to noise ratio <7%. An exception was R375P which is consistent with protein precipitation.

4.2.4 Requirements for activation of ACVR1

ACVR1 is a receptor kinase, a protein which upon ligand activation signals through the transfer of a phosphate group onto a substrate. The BMP/TGFβ type I receptors require the transphosphorylation of the GS domain to phosphorylate their substrate Smad. In wild-type receptor signalling, this GS phosphorylation comes from the associated type II receptor following BMP/TGFβ ligand binding. In FOP patient-derived cell lines, activation of the Smad pathway does not require this extracellular ligand binding, but the requirements for Smad phosphorylation in this instance are unknown.

4.2.4.1 Autophosphorylation of ACVR1

On expression neither wild-type nor mutant ACVR1 showed phosphorylation by mass spectrometry. An *in vitro* kinase assay was performed to assess the potential autophosphorylation activity of the wild-type ACVR1 and a number of mutant receptors (L196P, Q207E and Q207D). The proteins were incubated with 1 mM ATP, 1 mM $MnCl_2$ and 5 mM $MgCl_2$ at room temperature for 2 hours. Samples were analysed by intact mass spectrometry using an Agilent LC/MS TOF system with reversed-phase HPLC linked to an ESI and TOF analyser (chapter 2.8.1). This showed a heterogeneous solution of up to 4 autophosphorylations, demonstrated by successive mass shifts of +80 Da, for all proteins (figure 4.11).

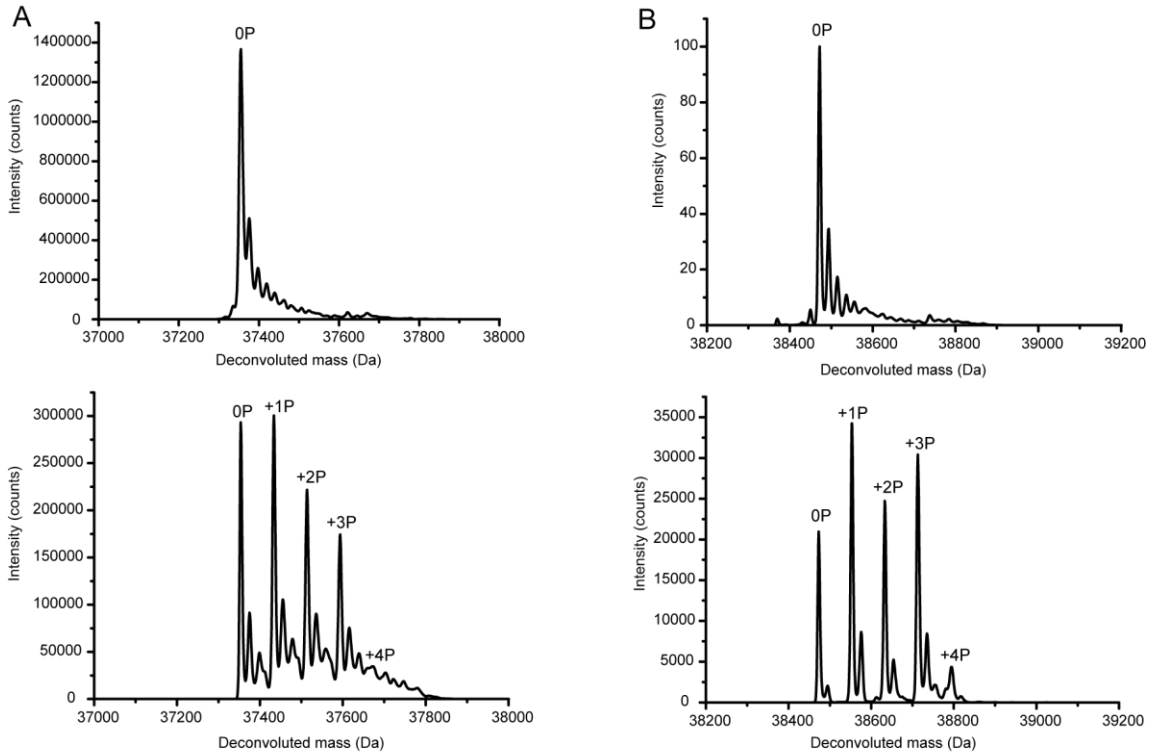


Figure 4.11 Autophosphorylation of ACVR1. (A) Wild-type ACVR1 and (B) ACVR1 Q207E. On expression neither protein was phosphorylated (top panel). Following incubation *in vitro* with ATP both receptors showed mass shifts indicative of up to 4 phosphorylations (lower panel).

The sites of autophosphorylation in wild-type ACVR1 and the constitutively active mutant Q207D were mapped by ESI-MS/MS (see methods 2.8.2). Analyses of both proteins identified the same phosphorylated peptides (figure 4.12). The peptides were identified using the Mascot server, based on the MOWSE [123] significance score (methods section 2.8.2). Although there was some ambiguity for the precise sites of phosphorylation, none of the proteins were autophosphorylated in the GS domain, suggesting that both the mutants and the wild type may require further components for full activation and Smad phosphorylation.

A

-EpTELYNTVMLR-

-HENILGFIASDMpTSR-

-MTpTNVGDpSpTLADLLDHSCTSGSGSGLPFLVQR- with T2,T3 ambiguity
-MpTpTNVGDpSpTLADLLDHSCTSGSGSGLPFLVQR-

B

GS loop

SM**TT**NVGD**ST**LADLLDHSCTSGSGSGLPFLVQRTVARQITLLECVGKGRY
GEVWRGSWQGENVAVKIFSSRDEKSWFRE**T**ELYNTVMLRHENILGFIASD
M**T**SRHSSTQLWLITHYHEMGSLYDYLQLTTLDTVSCLRIVLSIASGLAHL
HIEIFGTQGKPAIAHRDLKSKNILVKKNGQCCIADLGLAVMHSQSTNQLD
VGNNPRVGTKRYMAPEVLDETIQVDCFDSYKRVDIWAFGLVLWEVARRMV
SNGIVEDYKPPFYDVVPNDPSFEDMRKVVCVDQQRPNIPNRWFSDPTLTS
LAKLMKECWYQNPSARLTALRIKKTLTKID

C

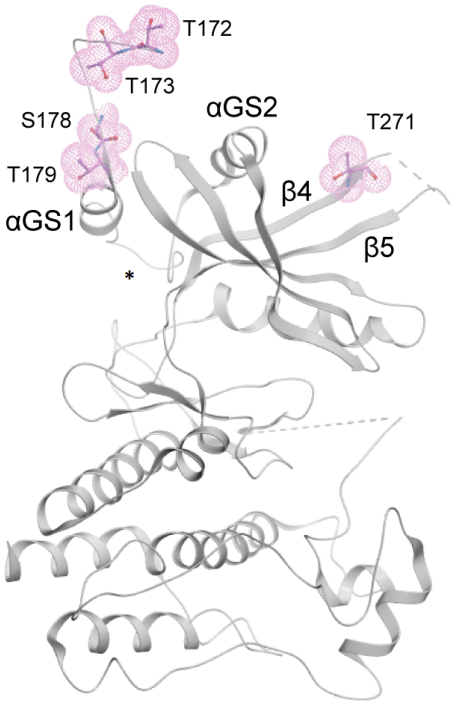
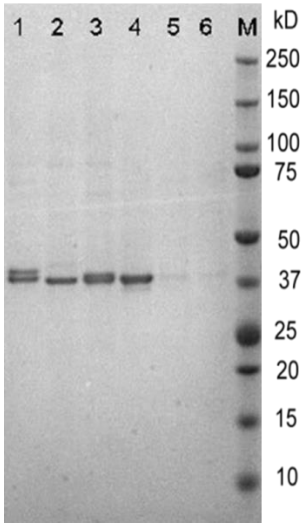


Figure 4.12 Mapped sites of autophosphorylation in both WT ACVR1 and ACVR1 Q207D. (A) MSMS analyses detected the same phosphorylated peptides in both constructs. The tri- and tetra-phosphorylated peptides showed strong evidence that the phosphorylated residues were not part of the GS loop. (B) The construct sequence of wild-type ACVR1 with the phosphorylated residues highlighted in purple. (C) The sites of phosphorylation were mapped onto the structure of wild-type ACVR1 (pdb 3h9r). This revealed that the phosphorylation sites were exposed regions of the kinase. The GS loop is highlighted with an asterisk.

4.2.4.2 Transphosphorylation of the type I receptors by the BMP/TGFβ type II receptors

4.2.4.2.1 Co-expression of various type I and type II receptors

On ligand binding the type I receptors are phosphorylated in the GS domain by the type II receptors. Initial experiments of co-expression of the soluble cytoplasmic domains of the BMP/TGFβ type I and type II receptors showed a difference in levels of phosphorylation of the type I receptor. 50 ml insect cell cultures (Sf9) were obtained from the SGC biotechnology group following baculoviral co-expression of various type I and type II receptors. The cultures were harvested and the histagged proteins purified by Nickel-affinity chromatography (figure 4.13).



Sample	Type I	Mass (Da)	Type II	Mass (Da)
1	TGFBR1	41606.7	TGFBR2	38494.8
2	ACVRL1	40708.4	ACVR2	36497.9
3	BMPR1B	40836.0	ACVR2	36497.9
4	ACVR1	39875.6	ACVR2	36497.9
5	BMPR1A	40672.0	AMHR2	38969.3
6	BMPR1A	40672.0	AMHR2	38079.3

Figure 4.13 Baculoviral co-expression of type I and type II receptors and table of the receptor constructs used. SDS PAGE showed a good level of purification, but the similar MW of the two receptors made a quantitative assessment of the relative expression levels difficult, except perhaps in sample 1, although both receptor types were detected by mass spectrometry. Previous studies at the SGC had revealed poor expression levels from a vector encoding for a larger GST tag.

The purified proteins were then analysed by electrospray mass spectrometry (chapter 2.8.1 for methods). High levels of type I receptor phosphorylation were found in sample 1 containing TGFBR1 and TGFBR2 (figure 4.14), whereas limited phosphorylation was observed in other BMP receptor samples. Mass spectrophotometry of the intact proteins revealed a heterogeneous TGFBR1 sample containing up to four phosphorylations, as well as the monophosphorylation of the TGFBR2 receptor. At least two TGFBR1 phosphorylations were subsequently mapped to the GS loop by MS/MS (figures 4.15 – 4.18). The sites of phosphorylation were mapped following tryptic digestion and LC-MS/MS (methods chapter 2.8.2). Peptide deconvolution was performed using the DA data analysis program (Bruker Daltonik). Database searching was performed using the *Mascot* 2.2.04 MOWSE (Molecular Weight Search) probability algorithm with the search parameters as described in chapter 2.8.2. This online search tool produces a report of the identified fragment peptides, giving each peptide a MOWSE score - the higher the score, the greater the significance (figures 4.16 and 4.17). The two highest scored monophosphorylated peptides in this example were for the same peptide but showing two possible phosphorylation sites (figure 4.16). The peptide view of each identified peptide displayed a table of theoretical masses of the charged ions. These were b and z ions, depending on end of the fragment the charge is retained, with numerous charge states. If the mass has been detected it is highlighted in red (figure 4.16) and a series of detected ions increases confidence in the peptide. This example showed high confidence the peptide identified is significant but there was some ambiguity for the phosphorylation site. As one of the peptides identified the unphosphorylated mass of one of the potential sites it seemed likely that the other site had been phosphorylated. There remains, however, the possibility that there were two forms of the monophosphorylated peptide. Similar reasoning was applied to the diphosphopeptide (figure 4.17).

The deconvoluted CID data was searched to prove the existence of the various charge states of the peptide DLIYDMTTSGSGSGLPLLVQR (figure 4.18). The neutral mass of each phosphorylation state of the peptide was calculated, and then the additional masses of the charge states for each. The data was then searched for these masses to confirm the presence of the phosphopeptide. This confirmed that the mono-, di- and triphosphorylated peptide (figure 4.18)

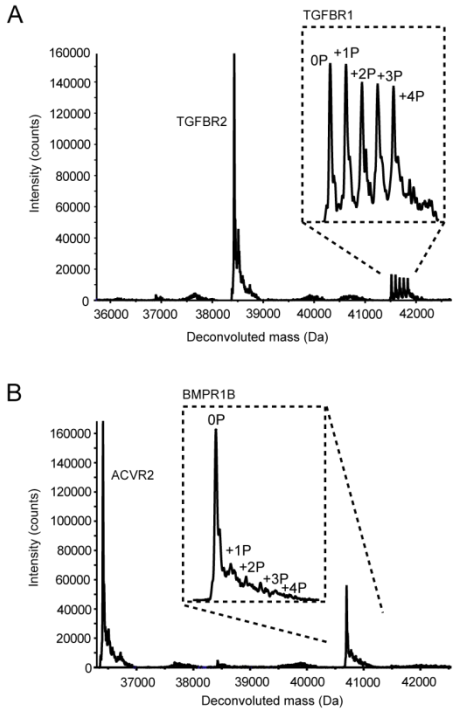


Figure 4.14 Mass spectrometry following co-expression in a baculoviral expression system. (A) TGFβ receptors TGFBR1 and TGFBR2. The deconvoluted mass spectrum showed the presence of multiple states of TGFBR1 phosphorylation from unphosphorylated (0P) to four phosphorylations (B) BMP receptors BMPR1B and ACVR2. Although this showed the presence of various phosphorylation states the occupancy of phosphorylated BMPR1B was low. Other BMP-receptor spectra were similar (See Appendix figure A1).

Figure 4.15 Mapped sites of GS phosphophorylation in TGFBR1. MSMS analyses of sample 1 detected three distinct phosphorylation sites, but only a monophosphopeptide and diphosphopeptide could be mapped with confidence. Further analyses are given in figures 4.16 – 4.18.

	Delta	Miss	Score	Expect	Rank	Unique	Peptide
	1.1812	0	86	3e-06	1	U	K.NGTCCIADLGLAVR.H
	-0.9957	0	(48)	0.063	1	U	R.HDSATDTIDIAPNHR.V
	0.2191	0	102	1.1e-07	1	U	R.HDSATDTIDIAPNHR.V
	0.2594	0	(47)	0.034	1	U	R.HDSATDTIDIAPNHR.V
(1)	0.1763	0	77	4.9e-05	1	U	K.DLIYDMTTSGSGSGLPLLVQR.T + Phospho (ST)
(2)	1.1573	0	(67)	0.00071	1	U	K.DLIYDMTTSGSGSGLPLLVQR.T + Phospho (ST)
	0.0843	0	(41)	0.18	1	U	K.DLIYDMTTSGSGSGLPLLVQR.T + Phospho (ST); Phospho (Y)
	0.2067	0	(30)	2.1	1	U	K.DLIYDMTTSGSGSGLPLLVQR.T + 2 Phospho (ST)
	2.4969	0	(26)	4.7	1	U	K.DLIYDMTTSGSGSGLPLLVQR.T + Phospho (ST); Phospho (Y)
	0.0608	0	(33)	1.3	1	U	K.DLIYDMTTSGSGSGLPLLVQR.T + 3 Phospho (ST)

(1) Top scoring peptide matches to query 208
BTSpecNo=188
Score greater than 46 indicates identity

Score	Expect	Delta	Hit	Protein	Peptide
76.6	4.9e-05	0.1763	2	TGFBR1A-c003	K.DLIYDMTTSGSGSGLPLLVQR.T
76.6	4.9e-05	0.1763	2	TGFBR1A-c003	K.DLIYDMTTSGSGSGLPLLVQR.T
68.8	0.0003	0.1763	2	TGFBR1A-c003	K.DLIYDMTTSGSGSGLPLLVQR.T

#	b	b^{++}	b^{*}	b^{*++}	b^{0}	b^{0++}	Seq.	y	y^{++}	y^{*}	y^{*++}	y^{0}	y^{0++}	#
1	116.0342	58.5207			98.0237	49.5155	D							21
2	229.1183	115.0628			211.1077	106.0575	L	2090.0951	1045.5512	2073.0685	1037.0379	2072.0845	1036.5459	20
3	*342.2023*	171.6048			324.1918	162.5995	I	1977.0110	989.0091	1959.9844	980.4959	1959.0004	980.0039	19
4	*505.2657*	253.1365			487.2551	244.1312	Y	1863.9269	932.4671	1846.9004	923.9538	1845.9164	923.4618	18
5	620.2926	310.6499			602.2821	301.6447	D	1700.8636	850.9354	1683.8370	842.4222	1682.8530	841.9302	17
6	*751.3331*	376.1702			733.3225	367.1649	M	1585.8367	793.4220	1568.8101	784.9087	1567.8261	784.4167	16
7	*852.3808*	426.6940			834.3702	417.6887	T	1454.7962	727.9017	1437.7696	719.3884	1436.7856	718.8964	15
8	*953.4285*	477.2179			935.4179	468.2126	T	1353.7485	677.3779	1336.7219	668.8646	1335.7379	668.3726	14
9	1040.4605	520.7339			1022.4499	511.7286	S	1252.7008	626.8540	1235.6743	618.3408	1234.6902	617.8488	13
10	1097.4820	549.2446			1079.4714	540.2393	G	1165.6688	583.3380	1148.6422	574.8248	1147.6582	574.3327	12
11	1166.5034	583.7553			1148.4929	574.7501	S	1108.6473	554.8273	1091.6208	546.3140	1090.6368	545.8220	11
12	1223.5249	612.2661			1205.5143	603.2608	G	1039.6259	520.3166	1022.5993	511.8033	1021.6153	511.3113	10
13	1310.5569	655.7821			1292.5463	646.7768	S	*982.6044*	491.8058	965.5778	483.2926	964.5938	482.8006	9
14	1367.5784	*684.2928*			1349.5678	675.2875	G	*895.5724*	448.2898	878.5458	439.7765			8
15	*1480.6624*	740.8349			1462.6519	731.8296	L	*838.5509*	419.7791	821.5244	411.2658			7
16	1577.7152	*789.3612*			1559.7046	780.3560	P	*725.4668*	363.2371	708.4403	354.7238			6
17	1690.7993	845.9033			1672.7887	836.8980	L	*628.4141*	314.7107	611.3875	306.1974			5
18	1803.8833	*902.4453*			1785.8728	893.4400	L	*515.3300*	258.1686	498.3035	249.6554			4
19	1902.9517	*951.9795*			1884.9412	942.9742	V	*402.2459*	201.6266	385.2194	193.1133			3
20	2031.0103	1016.0088	2013.9838	1007.4955	2012.9998	1007.0035	Q	*303.1775*	152.0924	286.1510	143.5791			2
21							R	175.1190	88.0631	158.0924	79.5498			1

(2) Top scoring peptide matches to query 149
BTSpecNo=200
Score greater than 45 indicates homology
Score greater than 48 indicates identity

Score	Expect	Delta	Hit	Protein	Peptide
66.7	0.00071	1.1573	2	TGFBR1A-c003	K.DLIYDMTTSGSGSGLPLLVQR.T
59.5	0.0037	1.1573	2	TGFBR1A-c003	K.DLIYDMTTSGSGSGLPLLVQR.T

#	b	b^{++}	b^{*}	b^{*++}	b^{0}	b^{0++}	Seq.	y	y^{++}	y^{*}	y^{*++}	y^{0}	y^{0++}	#
1	116.0342	58.5207			98.0237	49.5155	D							21
2	229.1183	115.0628			211.1077	106.0575	L	2090.0951	1045.5512	2073.0685	1037.0379	2072.0845	1036.5459	20
3	*342.2023*	171.6048			324.1918	162.5995	I	1977.0110	989.0091	1959.9844	980.4959	1959.0004	980.0039	19
4	505.2657	253.1365			487.2551	244.1312	Y	1863.9269	932.4671	1846.9004	923.9538	1845.9164	923.4618	18
5	620.2926	310.6499			602.2821	301.6447	D	1700.8636	850.9354	1683.8370	842.4222	1682.8530	841.9302	17
6	*751.3331*	376.1702			733.3225	367.1649	M	1585.8367	793.4220	1568.8101	784.9087	1567.8261	784.4167	16
7	*852.3808*	426.6940			834.3702	417.6887	T	1454.7962	727.9017	1437.7696	719.3884	1436.7856	718.8964	15
8	*953.4285*	477.2179			935.4179	468.2126	T	1353.7485	677.3779	1336.7219	668.8646	1335.7379	668.3726	14
9	*1022.4499*	511.7286			1004.4394	502.7233	S	1252.7008	626.8540	1235.6743	618.3408	1234.6902	617.8488	13
10	1079.4714	540.2393			1061.4608	531.2341	G	1183.6793	592.3433	1166.6528	583.8300	1165.6688	583.3380	12
11	1166.5034	583.7553			1148.4929	574.7501	S	1126.6579	563.8326	1109.6313	555.3193	1108.6473	554.8273	11
12	1223.5249	612.2661			1205.5143	603.2608	G	1039.6259	520.3166	1022.5993	511.8033	1021.6153	511.3113	10
13	1310.5569	655.7821			1292.5463	646.7768	S	*982.6044*	491.8058	965.5778	483.2926	964.5938	482.8006	9
14	1367.5784	*684.2928*			1349.5678	675.2875	G	*895.5724*	448.2898	878.5458	439.7765			8
15	1480.6624	*740.8349*			1462.6519	731.8296	L	*838.5509*	419.7791	821.5244	411.2658			7
16	*1577.7152*	*789.3612*			1559.7046	780.3560	P	*725.4668*	363.2371	708.4403	354.7238			6
17	1690.7993	845.9033			1672.7887	836.8980	L	*628.4141*	314.7107	611.3875	306.1974			5
18	1803.8833	*902.4453*			1785.8728	893.4400	L	*515.3300*	258.1686	498.3035	249.6554			4
19	1902.9517	*951.9795*			1884.9412	942.9742	V	*402.2459*	201.6266	385.2194	193.1133			3
20	2031.0103	1016.0088	2013.9838	1007.4955	2012.9998	1007.0035	Q	*303.1775*	152.0924	286.1510	143.5791			2
21							R	175.1190	88.0631	158.0924	79.5498			1

Figure 4.16 Evidence for the monophosphopeptide DLIYDMTT(p)SGSGSGLPLLVQR from CID data. In (1) the MOWSE scores were identical for both pS9 and pS11 (residue numbers taken from peptide position). In combination with peptide (2) which, despite a lower MOWSE score, has captured the pS9

b ion it is clear that this is the site of phosphorylation. The table gives predicted masses for all possible peptide ions. Masses in bold italic red contributed to the series score. Bold red indicates that the number of matches in the ion series is greater than would be expected by chance, suggesting that the ion series is present. Non-bold red indicates that the number of matches in the ion series is no greater than would be expected by chance. The blue brackets (right) indicate a significant series of identified ions. The bold red arrow (left) identifies the confident assignment of phosphorylation to S9 in peptide 2. The two blue arrows indicate two potential sites for the single phosphorylation in peptide 1. The dashed blue arrow indicates an unlikely site.

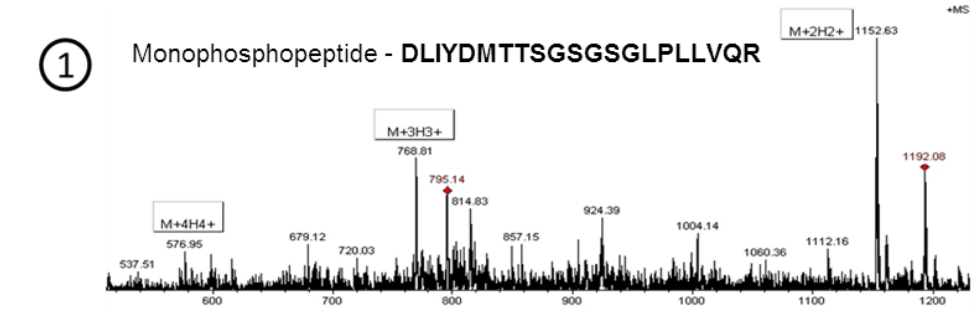
Delta	Miss	Score	Expect	Rank	Unique	Peptide
1.1812	0	86	3e-06	1	U	K.NGTCCIADLGLAVR.H
-0.9957	0	(48)	0.063	1	U	R.HDSATDTIDIAPNHR.V
0.2191	0	102	1.1e-07	1	U	R.HDSATDTIDIAPNHR.V
0.2594	0	(47)	0.034	1	U	R.HDSATDTIDIAPNHR.V
0.1763	0	77	4.9e-05	1	U	K.DLIYDMTTSGSGSGLPLLVQR.T + Phospho (ST)
1.1573	0	(67)	0.00071	1	U	K.DLIYDMTTSGSGSGLPLLVQR.T + Phospho (ST)
0.0843	0	(41)	0.18	1	U	K.DLIYDMTTSGSGSGLPLLVQR.T + Phospho (ST); Phospho (Y)
0.2067	0	(30)	2.1	1	U	K.DLIYDMTTSGSGSGLPLLVQR.T + 2 Phospho (ST)
2.4969	0	(26)	4.7	1	U	K.DLIYDMTTSGSGSGLPLLVQR.T + Phospho (ST); Phospho (Y)
0.0608	0	(33)	1.3	1	U	K.DLIYDMTTSGSGSGLPLLVQR.T + 3 Phospho (ST)

Top scoring peptide matches to query 214
BTSpecNo=184
Score greater than 35 indicates homology
Score greater than 46 indicates identity

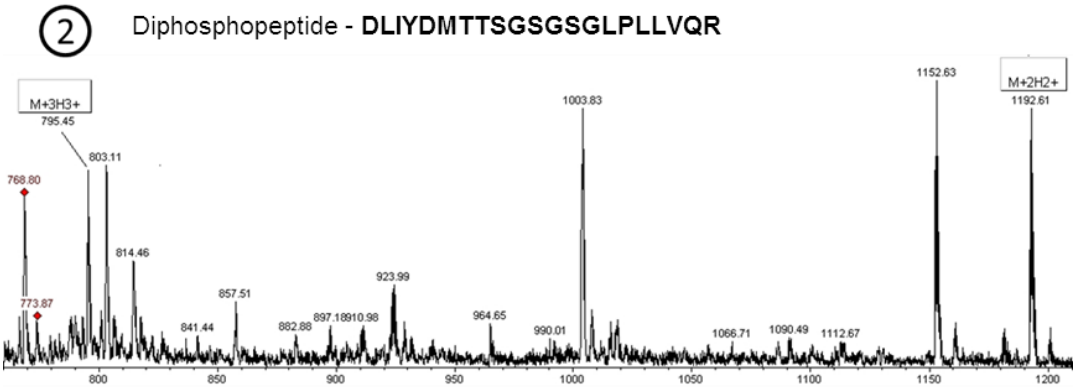
Score	Expect	Delta	Hit	Protein	Peptide
30.4	2.1	0.2067	2	TGFBR1A-c003	K.DLIYDMTTSGSGSGLPLLVQR.T
30.4	2.1	0.2067	2	TGFBR1A-c003	K.DLIYDMTTSGSGSGLPLLVQR.T
29.7	2.4	0.2067	2	TGFBR1A-c003	K.DLIYDMTTSGSGSGLPLLVQR.T

#	b	b^{++}	b^{*}	b^{*++}	b^{0}	b^{0++}	Seq.	y	y^{++}	y^{*}	y^{*++}	y^{0}	y^{0++}	#
1	116.0342	58.5207			98.0237	49.5155	D							21
2	229.1183	115.0628			211.1077	106.0575	L	2072.0845	1036.5459	2055.0579	1028.0326	2054.0739	1027.5406	20
3	342.2023	171.6048			324.1918	162.5995	I	1959.0004	980.0039	1941.9739	971.4906	1940.9899	970.9986	19
4	505.2657	253.1365			487.2551	244.1312	Y	1845.9164	923.4618	1828.8898	914.9485	1827.9058	914.4565	18
5	620.2926	310.6499			602.2821	301.6447	D	1682.8530	841.9302	1665.8265	833.4169	1664.8425	832.9249	17
6	751.3331	376.1702			733.3225	367.1649	M	1567.8261	784.4167	1550.7995	775.9034	1549.8155	775.4114	16
7	852.3808	426.6940			834.3702	417.6887	T	1436.7856	718.8964	1419.7591	710.3832	1418.7750	709.8912	15
8	953.4285	477.2179			935.4179	468.2126	T	1335.7379	668.3726	1318.7114	659.8593	1317.7274	659.3673	14
9	1022.4499	511.7286			1004.4394	502.7233	S	1234.6902	617.8488	1217.6637	609.3355	1216.6797	608.8435	13
10	1079.4714	540.2393			1061.4608	531.2341	G	***1165.6688***	583.3380	1148.6422	574.8248	1147.6582	574.3327	12
11	1148.4929	574.7501			1130.4823	565.7448	S	1108.6473	554.8273	1091.6208	546.3140	1090.6368	545.8220	11
12	1205.5143	603.2608			1187.5038	594.2555	G	***1039.6259***	520.3166	1022.5993	511.8033	1021.6153	511.3113	10
13	1292.5463	646.7768			1274.5358	637.7715	S	982.6044	491.8058	965.5778	483.2926	964.5938	482.8006	9
14	1349.5678	675.2875			1331.5572	666.2823	G	***895.5724***	448.2898	878.5458	439.7765			8
15	1462.6519	731.8296			1444.6413	722.8243	L	***838.5509***	419.7791	821.5244	411.2658			7
16	1559.7046	780.3560			1541.6941	771.3507	P	***725.4668***	363.2371	708.4403	354.7238			6
17	1672.7887	836.8980			1654.7781	827.8927	L	628.4141	314.7107	611.3875	306.1974			5
18	1785.8728	893.4400			1767.8622	884.4347	L	***515.3300***	258.1686	498.3035	249.6554			4
19	1884.9412	942.9742			1866.9306	933.9689	V	***402.2459***	201.6266	385.2194	193.1133			3
20	2012.9998	1007.0035	1995.9732	998.4902	1994.9892	997.9982	Q	303.1775	152.0924	286.1510	143.5791			2
21							R	175.1190	88.0631	158.0924	79.5498			1

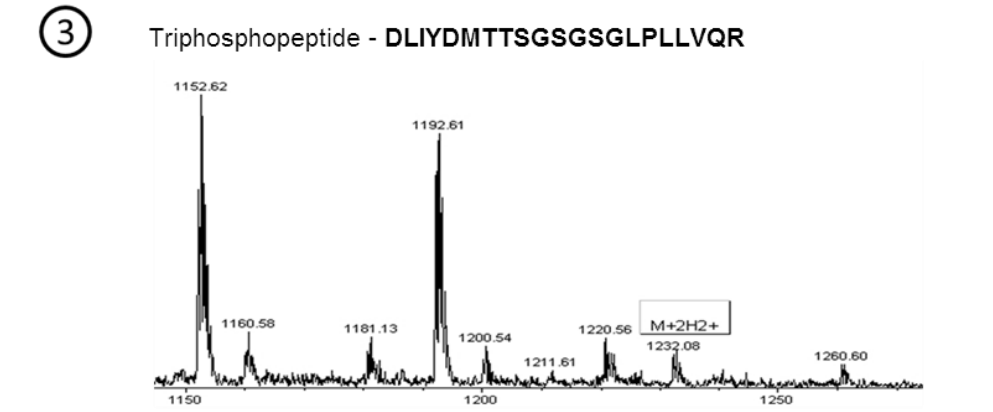
Figure 4.17 Evidence for the diphosphopeptide DLIYDMTT(p)SG(p)SGSGLPLLVQR. The presence of the pS11 ion can be determined with reasonable confidence (red arrow) from the presence of the neighbouring ions. The MOWSE score indicates that pS9 or pT8 are equally likely (blue arrows). Since pS9 was also identified in a monophosphorylated peptide (figure 4.17) it might suggest that this is the most likely second phosphorylation.



Charge state exact mass calculator							
Neutral mass	proton mass	M+H+	M+2H2+	M+3H3+	M+4H4+	M+5H5+	M+6H6+
2302.0916	1.0073	2303.0989	1152.0531	768.3711	576.5302	461.4256	384.6892



Charge state exact mass calculator							
Neutral mass	proton mass	M+H+	M+2H2+	M+3H3+	M+4H4+	M+5H5+	M+6H6+
2382.0579	1.0073	2383.0652	1192.0362	795.0266	596.5218	477.4189	398.0169



Charge state exact mass calculator							
Neutral mass	proton mass	M+H+	M+2H2+	M+3H3+	M+4H4+	M+5H5+	M+6H6+
2462.0243	1.0073	2463.0316	1232.0194	821.6820	616.5134	493.4121	411.3447

Figure 4.18 Evidence for the existence of the DLIYDMTTSGSGSGLPLLVQR phosphopeptide as: (1) monophosphopeptide, the (2) diphosphopeptide and the (3) triphosphopeptide through the observed charged fragment masses. For each phosphopeptide, the "neutral mass" corresponding to the mass of the uncharged peptide plus the respective additional phosphorylations is shown as a table together with individual charge states. Peaks matching these masses are identified by boxed labels, with the charge state mass of each phosphopeptide labelled as M (matching the mass calculator).

4.2.5.2.2 FOP mutants show similar levels of phosphorylation as the wild type on co-expression

The mutant ACVR1 constructs were then co-expressed with the type II receptor ACVR2 to see whether these might be more susceptible to transphosphorylation than the wild type. However, all of the mutants showed very little phosphorylation similar to wild type (figure 4.19A). The purified proteins from co-expression were then incubated with 1 mM ATP, 1 mM $MnCl_2$ and 5 mM $MgCl_2$ for 2 hours at room temperature to see if *in vitro* phosphorylation activity was altered between wild type and mutant. Again mass spectrometry revealed that the level of phosphorylation was the same in all ACVR1 proteins (figure 4.19B).

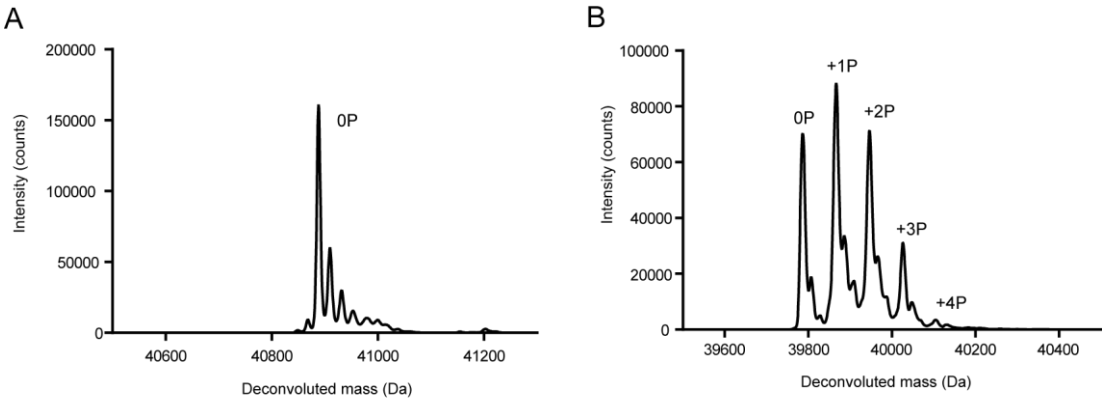


Figure 4.19 Co-expression of ACVR1 and ACVR2. (A) ACVR1 L196P showed no phosphorylation. (B) Following incubation *in vitro* with ATP, ACVR1 Q207E revealed mass shifts indicative of up to 4 phosphorylations

4.2.5 Interaction between mutant ACVR1 and its substrate Smad1

4.2.5.1 Purified Smad1 does not form a stable complex with the kinase domain of ACVR1 Q207D.

To identify the structural basis for receptor-Smad interaction both the MH2 domain of Smad1 and the constitutively active ACVR1 Q207D kinase domain construct (a.a.201 - 499) were expressed in bacterial and baculoviral expression systems, respectively, and purified to homogeneity. Various methods were subsequently employed to test the binding of the substrate. Initially, binding was tested in a pull-down assay utilizing the His6-GST-tagged Smad1 to capture the His-tagged ACVR1. However, the Smad1 protein bound poorly to the glutathione sepharose (GSH) beads hindering analysis. Smad1 and ACVR1 Q207D were then mixed at a 1:1 ratio and concentrated for preparative gel filtration. SDS-PAGE analysis of the elution fractions indicated co-elution of the two proteins, but no evidence of a robust shift in retention volume to suggest the formation of a higher MW complex (figure 4.20). Since crystallisation screens are performed at far higher concentrations than used in gel filtration, the co-eluted protein mix was concentrated and assayed in the JCSG4 and LFS4 crystal screens at 20°C in the presence of the ATP-competitive inhibitor LDN-193189. Diffraction quality crystals were obtained in 0.20 M NaI, 20.0% PEG 3350 and 10.0% ethylene glycol. However, subsequent dataset collection and structure refinement by the SGC crystallography group determined that the crystals were of the Smad1 trimer only, which was solved previously.

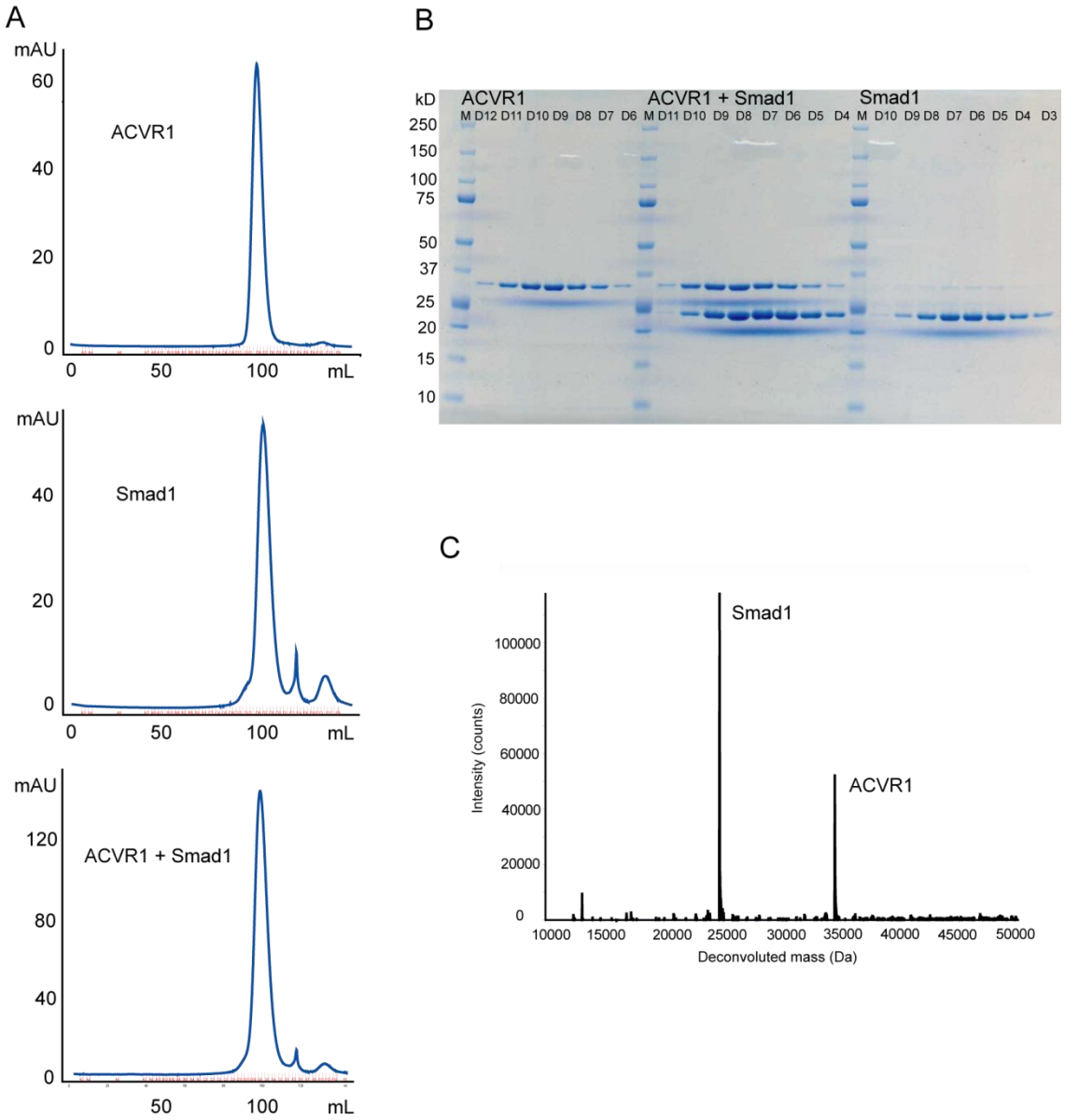


Figure 4.20 Size exclusion chromatography analysis of Smad-receptor interaction. (A) Gel filtration elution profiles of the individual proteins and their 1:1 mixture. The peak of the mix appears to be only slightly shifted (eluted at 98ml) compared to Smad1 (eluted at 100ml), casting doubt on the formation of the larger complex. (B) SDS PAGE analysis of the eluted fractions did confirm co-elution of the two proteins. The photograph of the gel was taken using a scanner which resulted in a blue smeared shadow below the protein bands and a colour picture. (C) An aliquot from the co-elution was analysed by intact mass spectrometry. Deconvolution showed the presence of both Smad1 and ACVR1 Q207D.

4.2.5.1.1 Endofin as an accessory protein for Smad1?

For Smad to bind the active type I receptor it most likely requires a non-trimeric state. The Smads2/3 proteins, which bind TGFBR1, are localized to the membrane by a FYVE domain-containing protein known as Smad Anchor for Receptor Activation (SARA) [69]. This protein binds directly to the R-Smad, preventing the formation of the Smad trimer, and also interacts directly with the receptor to facilitate Smad phosphorylation [69]. The homologue endofin has been suggested as a similar accessory protein for Smad1/5/8 signalling [71], but this finding has been reproduced. To test this interaction, Smad1 and endofin were expressed separately in *E. coli* and purified by nickel-affinity chromatography. The two proteins were then mixed and ran on a Superdex 16/60 S75 gel filtration column using standard gel filtration buffer (chapter 2.7). The two proteins appeared to elute separately, although the two elution peaks were overlapping (figure 4.21A).

For further analysis, endofin and SARA expression constructs in the pCDF-LIC vector were provided by the SGC Biotechnology group. These contained the CloDF13-derived CDF replicon allowing the co-expression of His-tagged endofin or SARA with Smad1 or Smad2, which contained N-terminal His6-GST tags. Following lysis, the clarified supernatant was passed through a GSH drip column, followed by a Nickel-sepharose column to collect any flow through protein that was not bound to the first column. Each column was then separately washed and eluted with glutathione or imidazole, respectively. Endofin was not observed in the eluate that contained Smad1 from the GSH column, indicating that the two proteins had not bound (figure 4.21B). In contrast, Smad2 and SARA were co-eluted from the GSH column confirming their interaction.

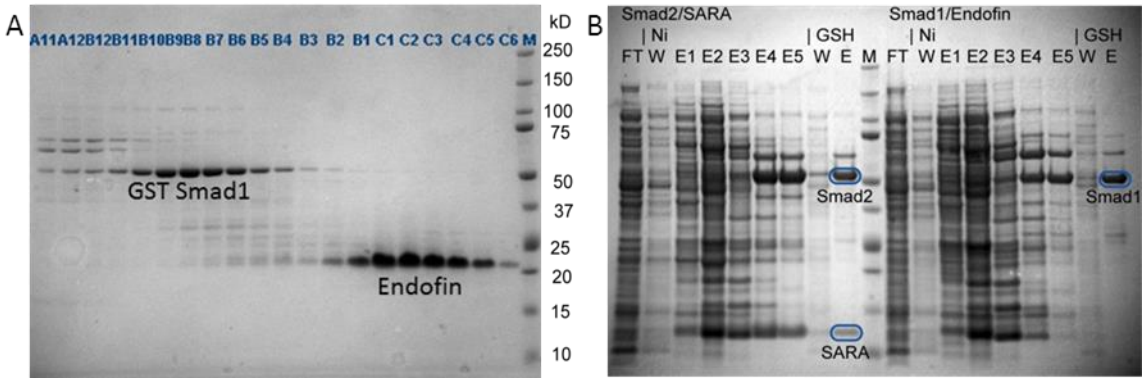


Figure 4.21 SDS PAGE analysis of endofin-Smad1 binding. (A) Size-exclusion chromatography of GST-tagged Smad 1 and endofin. The proteins were mainly eluted in two peaks, however there was some minor overlap. A11 – C6 relates to the elution fraction, and M is the marker lane. The same size marker was used for both gels. (B) Pull down analyses following co-expression of His-GST-tagged Smad1 with His-tagged endofin, and His-GST–tagged Smad 2 with His-tagged SARA as a control. The lane This showed robust binding of Smad 2 and SARA (shown by the GSH elution (lane labelled GSH E). However, no binding was observed between the test proteins Smad1 and endofin. The endofin expressed was captured on the nickel column and eluted.

4.2.5.2 Phosphorylation of Smad1 by mutant ACVR1

4.2.5.2.1 Mutant ACVR1 alone appears unable to phosphorylate Smad1

It was shown that the activating mutation Q207D could not induce a stable complex with Smad 1. However, this substrate interaction may be weak or only transient, so the mutated receptor could still be capable of phosphorylating Smad1. The various ACVR1 mutants were mixed with Smad1 at a highly favourable 1:1 ratio together with 1 mM ATP, 1 mM $MnCl_2$ and 5 mM $MgCl_2$, and incubated at room temperature for 2 hours. This resulted in an increase in the proportion of the receptor in higher states of phosphorylation (figure 4.22), compared to previous autophosphorylation experiments (figure 4.11). However, there was no observable phosphorylation of Smad1 (figure 4.22). This same was observed in the wild type (See Appendix figure A2).

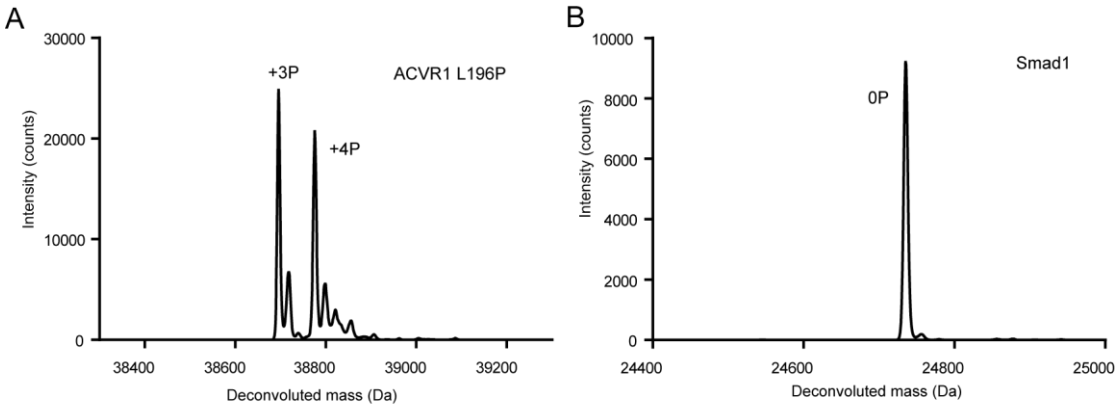


Figure 4.22 Mass spectra of the Smad1 and ACVR1 L196P mix following an in vitro kinase assay. (A) ACVR1 L196P showed + 3 and + 4 phosphorylations (B) Smad1 was not phosphorylated.

4.2.5.2.2 In vitro phosphorylation of Smad1 by mutant ACVR1 was not induced by adding purified recombinant ACVR2

The Smad signalling pathway requires the presence of both the type I and type II receptor. The purified type I ACVR1 mutant and wild-type receptors were therefore assayed together with the purified type II receptor ACVR2A and the substrate Smad1, at a 1:1:1 ratio and incubated with 1 mM ATP, 1 mM $MnCl_2$ and 5 mM $MgCl_2$ for 2 hours at room temperature. Again, mass spectrometry of the samples revealed multiple phosphorylation states of the type I receptor (similar to the autophosphorylation samples), but no phosphorylation of Smad1 (figure 4.23).

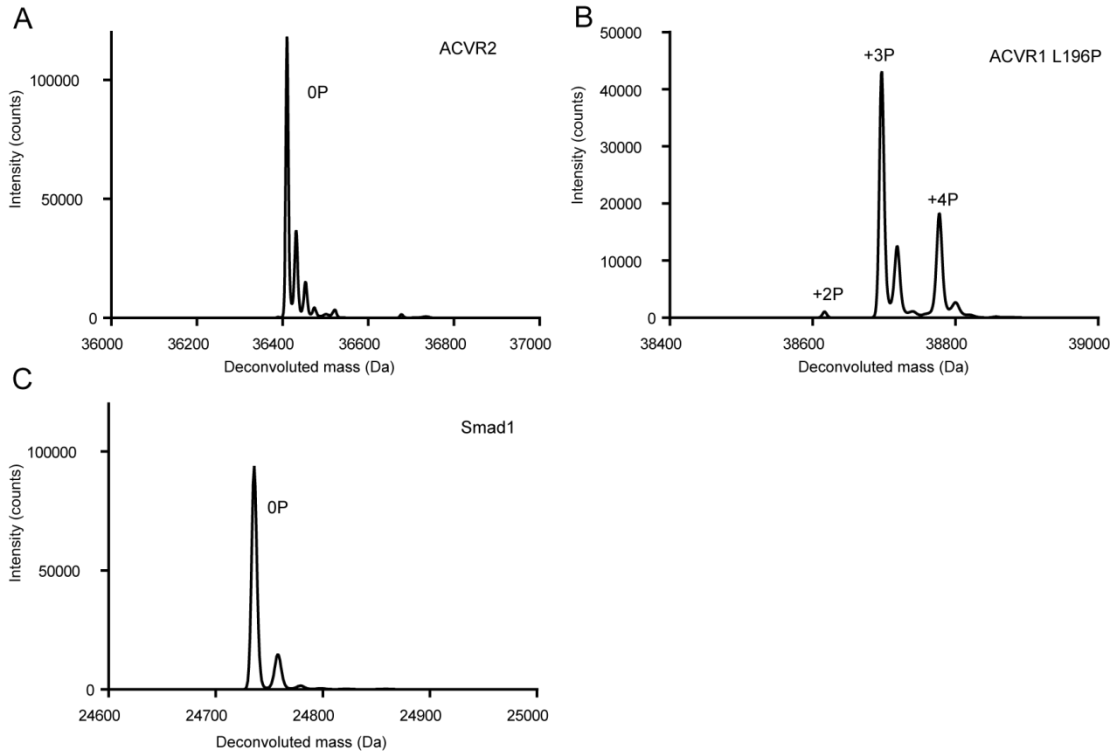


Figure 4.23 Mass spectra of mixed ACVR1 L196P, ACVR2 and Smad1 following an in vitro kinase assay. The receptor proteins were mixed after purification (A) ACVR2 showed no phosphorylation (B) ACVR1 L196P showed a heterogenous mix of + 2, + 3 and + 4 phosphorylations (C) Smad1 was not phosphorylated.

4.2.5.3 Co-expression of mutant ACVR1 with ACVR2 promotes Smad1 phosphorylation

Following previous negative results, co-expression of the ACVR2 and ACVR1 receptors was trialled to assess whether this could promote the subsequent phosphorylation of Smad1 *in vitro*. The co-expressed and purified mutant ACVR1 L196P and ACVR2 were incubated with Smad1, 1 mM ATP, 1 mM $MnCl_2$ and 5 mM $MgCl_2$ at room temperature for 2 hours. Mass spectrometry revealed the expected 2 phosphorylations of Smad1 (figure 4.24A). Western blotting using a phospho-specific Smad1 antibody confirmed that the phosphorylations were at the C-terminal SSXS motif (figure 4.24B). However, there were difficulties reproducing this result both with the mutant and wild-type ACVR1. This prevented the further comparison of the activities of different mutants as well as that of

wild-type ACVR1. Such difficulty may result from the variability of infection when co-expressing two viruses in Sf9 insect cell cultures or that the activity is only weak resulting in marginal detection.

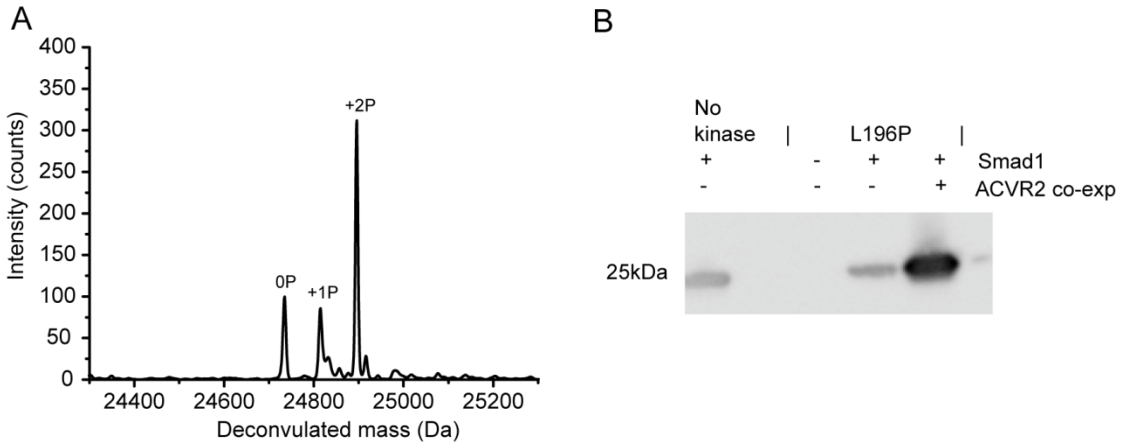


Figure 4.24 Co-expressed ACVR1 L196P and ACVR2 phosphorylate Smad1. (A) Deconvoluted mass spectrum of Smad1 showing multiple phosphorylation states, from unphosphorylated up to 2 phosphorylations. (B) Western blot using a phospho-specific antibody against the Smad1 C-terminal SSXS motif (Cell Signaling anti-P-Smad1/5/8 (#9511)). The recombinant Smad1 was added at 1 mg/ml equally in each 100 μL kinase reaction. 10 μL of each reaction was loaded onto SDS PAGE for the blot. Smad1 showed phosphorylation only in the sample containing ACVR1 L196P that was co-expressed with ACVR2. The other faint bands are likely a product of overloading and background antibody binding, as seen in the no kinase lane.

4.3 Discussion

The mutations of ACVR1 found in FOP patients are known to activate the Smad1 pathway. However, the presented structures of mutant ACVR1 mostly displayed an inactive conformation similar to the wild-type receptor. This is somewhat surprising since all of the structures were solved in the absence of the inhibitor protein FKBP12 and two of the mutant structures, Q207D and R206H, were lacking the regulatory GS domain. In the first crystal structure of TGFBR1 without FKBP12 there were several structural changes which hinted at a more active conformation when compared to the structure of TGFBR1 in complex with FKBP12 [76]. In chapter 3, BMPR1B also showed similar structural changes to the GS loop and to the αC and activation loop in the kinase domain. All of the

structures could be additionally influenced by the presence of an inhibitor in the ATP site. However, the structures were solved with a number of different inhibitors with different properties showed the same conformation of the activation loop and αC helix. Whilst some of the inhibitors form a water-mediated bond to the middle of the αC, this would not exclude the rotation of the αC expected in activation of the receptor (see chapter 5.2.5 for a full description on inhibitor binding mode). Additionally, one of the small molecule inhibitors, LDN-193189 bound to mutant ACVR1, was also crystallised in the BMPR1B-FKBP12 complex structure (chapter 3.2.3). In that structure the active conformation of a number of sites in the kinase were observed, including the activation loop which formed the "open" orientation expected of an active receptor kinase. These were not seen in the ACVR1 mutant structure with the same inhibitor bound.

Despite the overall persistence of the inhibitory interactions, there were examples of conformational shifts that hinted at a more dynamic mutant structure. For example, some structural shifts were observed in the kinase domain of the Q207D mutant in the L45 loop (a proposed site of Smad interaction) and around the ATP pocket in the P-loop and activation loop. In addition, the GS loop in the ACVR1 Q207E structure showed some disorder and high B-factors that suggest greater freedom of the serines for potential phosphorylation by the type II receptor. However, a structure of wild-type ACVR1 in the absence of FKBP12 is desirable for strict comparison. Overall, the apparent stability of the inactive state suggests that other factors may influence the conformation of the catalytic domain. For example, the binding of the type II receptor or the globular Smad MH2 domain may be necessary for a conformational switch. This scaffolding role would mirror that observed upon cyclin binding to the cyclin-dependent kinases (CDKs).

The inactive structures of the mutants were also consistent with their general inability to phosphorylate Smad1. Only upon ACVR2 co-expression was Smad1 phosphorylation

observed. This could be due to the local concentration of the two proteins being greater during expression which would encourage interactions. Additionally the presence of chaperone proteins and general cellular milieu could also increase protein contacts and accessibility of the GS loop for phosphorylation [124]. This could also explain why Smad1 phosphorylation was only proven in one experiment. A number of autophosphorylation sites were identified in both the wild-type and mutant ACVR1, which were mapped to the GS and kinase domains, but not to the important GS loop. It is widely documented that kinases may contain sites of inhibitory phosphorylation as well as activating sites [125]. The identified phosphorylation of T271 at the start of the L45 loop is interesting in this respect as it forms part of the predicted Smad1 binding site. Phosphorylation of this residue could potentially interfere with Smad1 interaction, perhaps explaining the difficulties encountered in the Smad1 phosphorylation assays even upon receptor co-expression.

Recent studies in Drosophila have also indicated that hyperactive signalling by ACVR1 R206H is dependent on both the type II receptor and GS phosphorylation [126]. Studies of knock-out mice have alternatively suggested that the FOP mutants only require a kinase dead type II receptor [127]. This implied that the scaffold of the type II was simply required for a conformational change in ACVR1 or for correct Smad docking or endosomal trafficking [127]. Similar results were obtained with specific knock-out cell lines. However, it cannot be ruled out that an alternative type II receptor (for example TGFBR2) or another kinase may be acting on mutant ACVR1 in these models. In light of these recent studies, it would be interesting to observe whether the *in vitro* phosphorylation assays examined here would show similar results with the kinase dead type II receptor.

The poor binding of FKBP12 to R206H, resulting in the leaky signalling of the receptor, has been proposed as an explanation of the mild activation observed in FOP [77], [128]. However, in C2C12 mouse myoblast cells FK506, which competitively binds FKBP12, greatly enhanced signalling from ACVR1 R206H compared to wild type [87]. This suggests that there is also an inherent activation of the mutant ACVR1 kinase beyond any effects of FKBP12, although dependent still on the type II receptor. These findings are generally supported by the results presented here.

5. Small molecule inhibitors of ACVR1

5.1 Introduction

There are currently no effective therapies for FOP. Current treatment is restricted to easing other features of FOP, such as inflammation, rather than preventing further bone growth. Small molecule chemical inhibitors are commonly used in drug development to target the protein kinase family [81], which includes ACVR1. The TGFβ signalling pathway, in particular TGFBR1, has been targeted over recent years due to its involvement in cancer and fibrosis [129]. However, the BMP-receptors have not been studied as intensely. The first inhibitor of ACVR1, dorsomorphin, was identified from zebrafish embryo phenotypic screening [91]. This was then optimised to LDN-193189 [96]. During this optimisation, Cuny *et al* noted the importance of the nitrogen in position 4 of the pyrazolo[1,5-a] pyrimidine ring for inhibitor potency [96]. This was explained by the structure of ACVR1 in complex with dorsomorphin, which showed that this nitrogen forms a hydrogen bond to H286 in the kinase hinge region [87] (figure 5.1). It was also found that inhibition was increased by replacing the 4-pyridine ring with 4-quinoline. Substitution of a piperizine group in place of the solvent exposed piperidinyl-ethoxy group also provided an increase in metabolic stability [96].

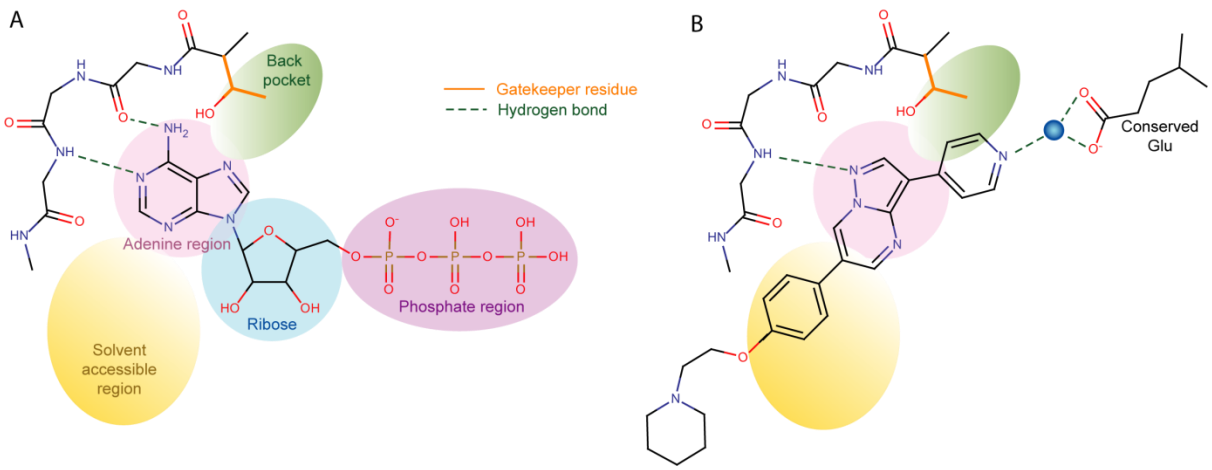


Figure 5.1 Schematic representation of the ACVR1 ATP pocket. Adapted from Zuccotto et al (2010)[90] (A) ATP binding to the hinge. (B) Dorsomorphin binding showing the core pyrazolo[1,5-a]pyrimidine binding to the hinge. The 4-pyridine group explores the back pocket due to the small gatekeeper residue (T283) while the piperazine group extends out into the solvent accessible region.

Although dorsomorphin and LDN-193189 are directed to ACVR1 they still target some other kinase families as well as other members of the BMP/TGFβ receptor family [93], [96]. Kinase inhibitor specificity is strongly influenced by the hinge gatekeeper residue [96]. Uniquely, type I TGFβ and activin B receptors have a small gatekeeper residue (S280 in TGFBR1) that allows inhibitor access to a deep hydrophobic back pocket and therefore the development of highly selective inhibitors [130]. In type I BMP receptors the gatekeeper residue is a threonine (ACVR1 T283) which establishes a slightly smaller back pocket (figure 5.2A). The importance of this for inhibitor selectivity has been demonstrated by the substitution of the gatekeeper residue of ACVR1 with that of TGFBR1 (i.e. ACVR1 T283S). This results in the inhibition of ACVR1 by the TGFβ inhibitor SB431542, whereas the TGFBR1 S280T mutant is no longer susceptible to inhibition [131] (figure 5.2B).

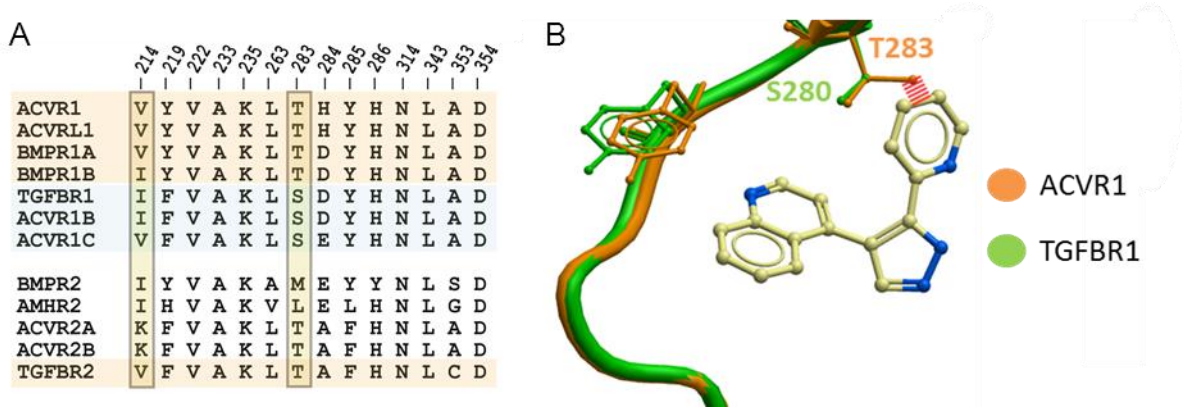


Figure 5.2 Comparison of the binding pocket of ACVR1 and TGFBR1. (A) Pseudo alignment of the ATP pocket region of all the BMP/TGFβ receptors. (B) Superimposition of the hinge region of ACVR1 (pdb: 3H9R) and TGFBR1 (pdb: 1PY5). The TGFBR1 inhibitor LY364947 would clash with the ACVR1 gatekeeper residue T283, shown by the dashed red line.

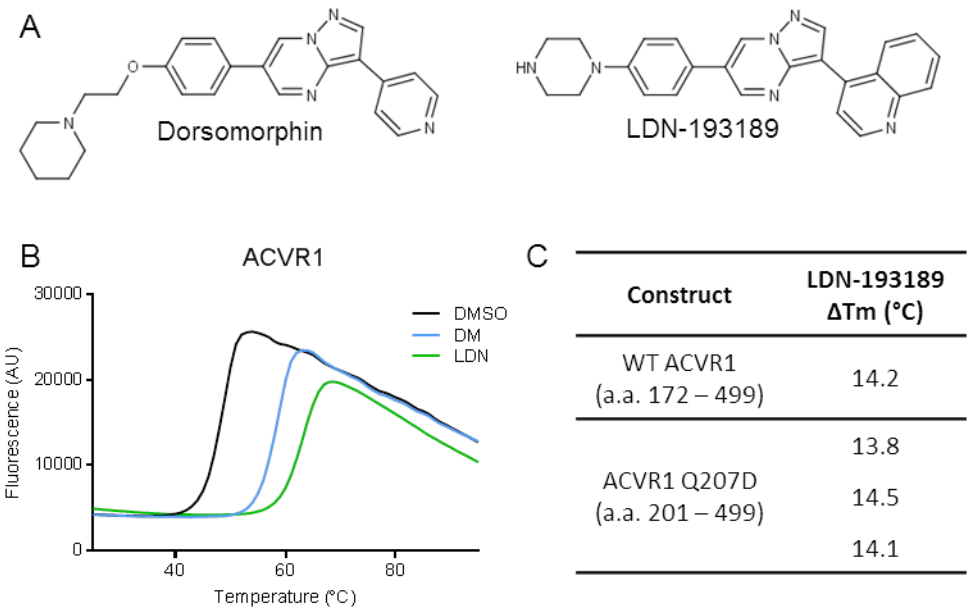
Despite the previous identification of LDN-193189, there is still a need for more highly selective, potent molecules in order to generate an investigational drug for FOP clinical trials. Here, differential scanning fluorimetry (DSF) was employed as a high throughput screening assay to identify further chemical scaffolds that bind to ACVR1. The assay uses a fluorescent dye (sypro orange) which binds to exposed hydrophobic regions of the protein. In aqueous solution the dye is quenched but as the protein unfolds upon heating more hydrophobic regions are exposed and the fluorescence increases. The midpoint of the unfolding defines the apparent the melting temperature (Tm). When a compound binds specifically to the folded protein, the thermodynamic stability increases proportionally to the binding affinity, resulting in a Tm shift (for full details see methods section, 2.10) [100]. The assay is convenient as it applicable to a high throughput setting, screening 96 compounds per plate and does not require substrate knowledge or optimisation between proteins. However, DSF can give rise to artefacts due to interactions between the compound and fluorescent dye, interference in optical detection due to coloured compounds, and high background fluorescence from hydrophobic surface residues [132].

The technique is limited by its reliance on two-state unfolding, so multi-domain proteins and protein complexes can give ambiguous results, For most kinases Tm shifts > 4 °C typically equate to an IC50 of < 1 μM whereas > 8 °C equates to an IC50 ≤ 100 nM [133]. However, the precise Tm shift range can vary slightly across different proteins, as they are influenced by their individual thermodynamic properties, such as heat capacity. Nonetheless, these values provide an indication of compound potency as well as selectivity for ACVR1 over other TGFβ receptors and the wider kinome. Following the identification of a novel hit compound, kinase assays and isothermal titration calorimetry (ITC) were employed for validation. Finally, compounds of interest were co-crystallised with ACVR1 to understand their binding mode to aid further inhibitor optimisation.

5.2 Results

5.2.1 Differential Scanning Fluorimetry (DSF)

In order to identify potent binding compounds, the wild-type ACVR1 construct (a.a. 172 – 499) and ACVR1 Q207D construct (a.a. 201 – 499) were screened using DSF against the SGC kinase targeted compound library as described in the Methods section (chapter 2.10). This screening included the known inhibitors dorsomorphin and LDN-193189 (figure 5.3A and B). The two ACVR1 constructs gave similar Tm shift values for all compounds tested (figure 5.3C). The average apparent Tm shifts of dorsomorphin and LDN-193189 were 10.3 °C and 14.2 °C respectively (figure 5.3B).



Construct	LDN-193189 ΔTm (°C)
WT ACVR1 (a.a. 172 – 499)	14.2
ACVR1 Q207D (a.a. 201 – 499)	13.8
	14.5
	14.1

Figure 5.3 DSF screening of dorsomorphin and LDN-193189. (A) Scaffolds of dorsomorphin and LDN-193189 (B) Raw DSF data for dorsomorphin (DM; blue) and LDN-193189 (LDN; green) against ACVR1 Q207D (a.a. 201 – 499). The compound stocks were dissolved in 100 % DMSO, but upon dilution were used at a final concentration of 2.5 % DMSO in each experiment. The final inhibitor concentration was 10 µM and the protein concentration was 2 µM. The Tm shifts induced by compound binding were determined against a reference well also containing 2.5 % DMSO (black). (C) Comparison of ACVR1 WT and ACVR1 Q207D Tm shift data.

The highest Tm shifts from the initial DSF screen are listed in table 5.1. This list included a number of compounds with the ASC prefix, which were synthesized in the lab of K. Shokat and designed to be potent pan-protein kinase inhibitors for cross-linking kinase and substrate [134], and cdk 1/2 inhibitor III which is known to be toxic, and so these compounds were not taken further. The next highest Tm shift observed, after LDN-193189 (ranked 2nd), was a novel compound K02288 (ranked 5th). The screening also identified a series of compounds with the imidazo[1,2-b]pyridazine scaffold (table 5.1).

The imidazo[1,2-b]pyridazine scaffold, closely related to the pyrazolo[1,5-a]pyrimidine of dorsomorphin, had been previously identified at the SGC and bound ACVR1 with similar potency (figure 5.4). A further set of compounds in this series had been acquired since

initial identification allowing the identification of several more potent hits, such as K02271. This imidazo[1,2-b]pyridazine scaffold contained a trimethoxyphenyl R1 group (Tm shift 11.5°C) (figure 5.4C). A small structure-activity relationship (SAR) study was determined from the Tm shift data. This showed the importance of all three methoxy groups for binding potency (figure 5.4C)

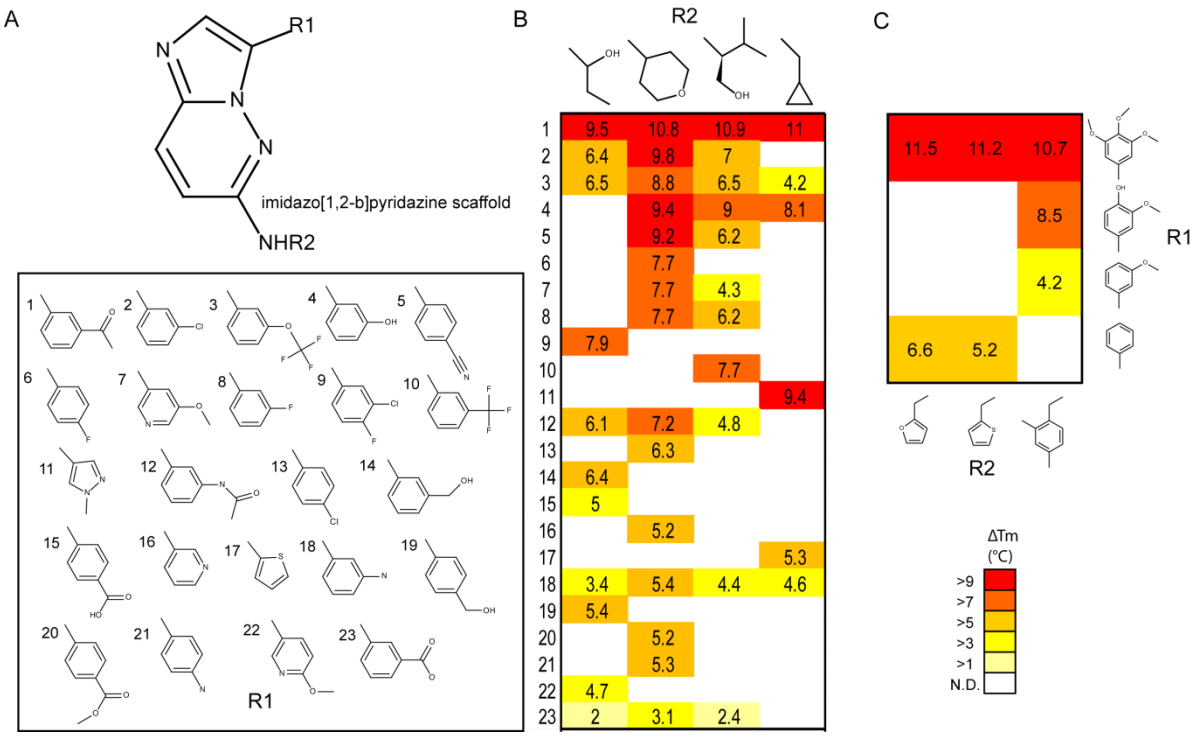
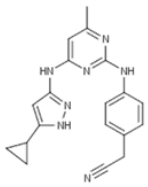
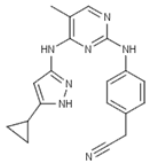
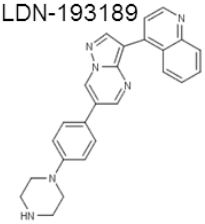
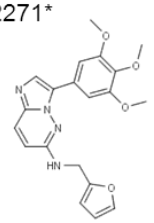
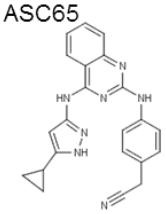
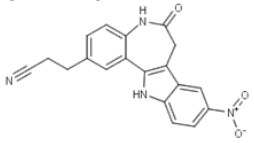
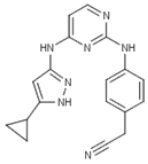
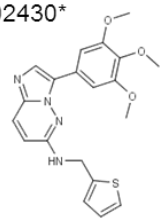
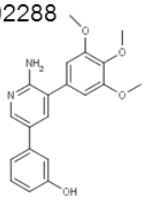
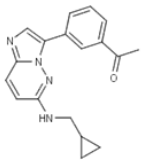
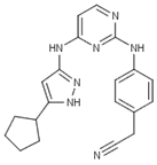
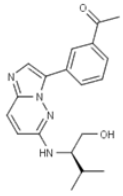
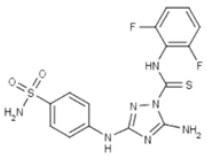
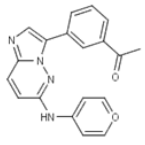


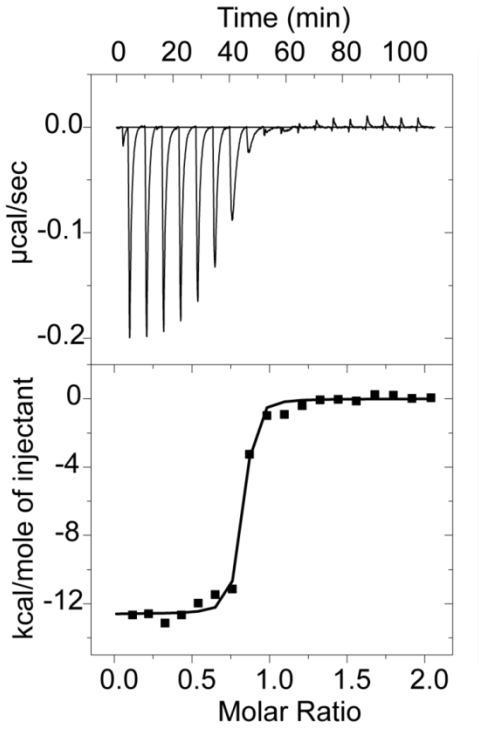
Figure 5.4 Tm shift heatmap for the imidazo[1,2-b]pyridizine compounds assayed against ACVR1. (A) The imidazo[1,2-b]pyridazine scaffold. (B) Tm shift heatmap of the early imidazo[1,2-b]pyridazine compounds. The highest Tm was 11 °C from compound K00135. (C) Heatmap of the methoxyphenol imidazo[1,2-b]pyridazine compounds.

Table 5.1 Top 20 compounds with the highest Tm shifts against ACVR1

Compound	ΔTm (°C)	Compound	ΔTm (°C)
(1) ASC69 	14.5	(8) ASC68 	12.5
(2) LDN-193189 	14.3	(9) K02271* 	11.5
(3) ASC65 	13.3	(10) Alsterpaullone, 2-cyanoethyl 	11.3
(4) ASC24 	13.2	(11) K02430* 	11.3
(5) K02288 	13.1	(12) K00135* 	11
(6) ASC67 	12.6	(13) K00518* 	10.9
(7) Cdk1/2 Inhibitor III 	12.5	(14) K00507* 	10.8

The ASC compounds are highly promiscuous. Compounds LDN-193189 and K02288 are the next most potent and most interesting. The imidazo[1,2-b]pyridazine series of compounds (highlighted with an asterisk) also showed high potency.

The trimethoxyphenyl group was also identified in the novel compound K02288. This compound gave a high Tm shift of 13.1 °C for ACVR1. K02288 has a 2-aminopyridine scaffold (table 5.1), which is also found in known inhibitors of kinases such as CHK2 [135] and NEK2 [136] as well as in crizotinib, a clinically-approved drug targeting the tyrosine kinase ALK [137]. To further characterise this novel compound, the binding affinity of K02288 to wild-type ACVR1 (a.a. 172 – 499 containing both GS and kinase domain) was measured by ITC (for methods see chapter 2.11). This gave a K_d of 8 nM showing tight binding as expected from the Tm shift data (figure 5.5).



Construct	Compound	Temperature (°C)	Stoichiometry (N)	Kd (nM)	ΔH (kJ mol^{-1})	TΔS (kJ mol^{-1})	ΔG (kJ mol^{-1})
ACVR1 (a.a.172-499)	K02288	10	1 ± 0.0	8 ± 0.0	-52.8 ± 0.8	-8.8	-43.9

Figure 5.5 ITC of ACVR1 and K02288. ITC trace showing the raw (top panel) and integrated (lower panel) data. 70 µM ACVR1 was titrated into the cell containing 7 µM K02288 at 10°C. Samples were buffered in 50 mM Hepes pH 7.5, 150 mM NaCl. 1 mM DTT. Data were fit using the single binding site model within the Origin software provided with the instrument, which also provides the errors to the fit.

5.2.2 Selectivity within the BMP/TGFβ receptor family

Other members of the BMP/TGFβ receptor family were also screened against the compound library. In particular, it was of interest to screen the TGFβ type I receptors, namely ACVR1B and TGFBR1. Plots of the Tm shift data for ACVR1 against those of ACVR1B and TGFBR1 gave an indication of compound selectivity between the receptor types, where the dashed diagonal line denotes non-selective compounds (figure 5.6). The library contained several TGFβ-specific compounds such as A-83-01, SD-208 and LY364947 [129]. These compounds showed clear selectivity for the TGFβ receptors over ACVR1, likely due to the larger gatekeeper residue in ACVR1 (figures 5.2 and 5.6). The comparisons also highlighted compounds showing an overall preference for ACVR1 (figure 5.6). In particular K02288 (green) showed clear selectivity for ACVR1 over ACVR1B (figure 5.6A).

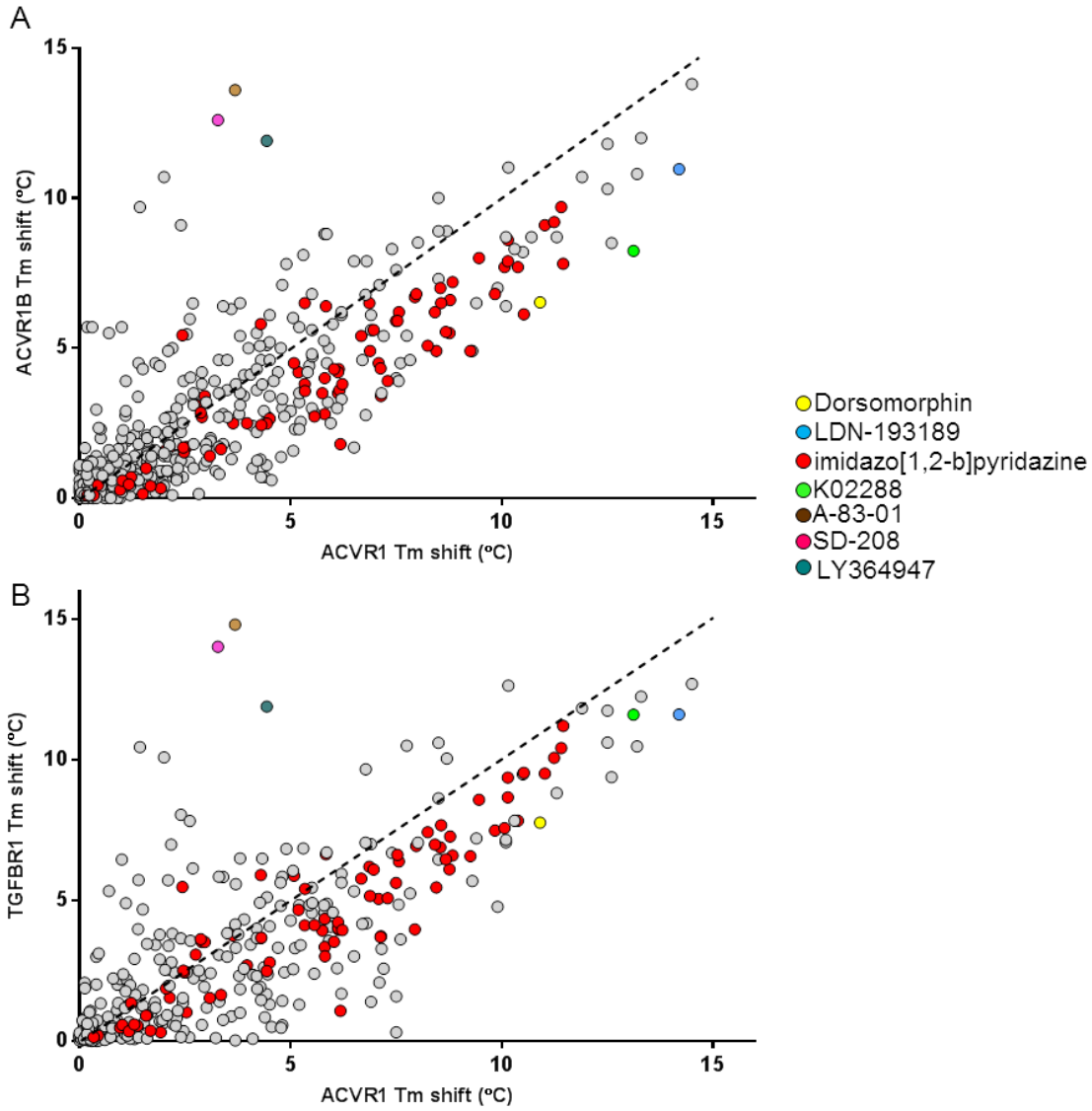


Figure 5.6 Scatter plot of Tm shift assay results for ACVR1 compared to TGFβ receptors ACVR1B and TGFBR1. Dorsomorphin (yellow), LDN-193189 (blue), K02288 (green) and the imidazo[1,2-b]pyridazine compound series (red) are highlighted, whilst all other tested compounds are grey. TGFBR1-specific compounds A-83-01 (brown), SD-208 (pink) and LY364947 (teal) were also highlighted for comparison. A dashed line is provided as a rough guide for BMP verses TGFβ receptor selectivity; compounds below the line appear more selective for ACVR1. (A) Plot of ACVR1 data against ACVR1B data. (B) Plot of ACVR1 data against TGFBR1 data.

The most promising novel compounds from the screen, including K02288 and the imidazo[1,2-b]pyridazine compounds K00135 and K00507, are compared with known binders in table 5.2. All of these compounds were highly selective against the BMP type II

receptor BMPR2. From comparison of the structures it is evident that binding in BMPR2 is unfavourable due to a steric clash with its large gatekeeper residue M279 (figure 5.2A). The apparent Tm shifts for TGFBR2 appear higher than those for ACVR2 although both residues have the same threonine gatekeeper residue as ACVR1; however, ACVR2 has a larger lysine group packing on the top of the binding pocket compared to the valine in TGFBR2 and ACVR1, which could affect compound binding. Interestingly, TGFBR2 showed a significantly reduced Tm shift with K02288 compared to the other inhibitors.

Table 5.2 Tm shift assay results for selected compounds screened against members of the BMP/TGFβ receptor family.

	Target Tm shift (°C)								
	Type I						Type II		
Compound	ACVR1	ACVRL1	BMPR1A	BMPR1B	ACVR1B	TGFBR1	ACVR2	BMPR2	TGFBR2
LDN-193189	14.2	16.2	14	16.2	11	12.2	8.6	0	16.3
K02288	13.1	12.2	12.8	13.3	8.2	11.9	8.3	3.5	8
K00135	11	11.1	10.9	11.5	9.1	9.4	2.8	7.7	13.8
K00507	10.8	11.6	9.5	10.4	9.2	10.4	2.4	4.2	12.5
Dorsomorphin	10.3	10.6		10.6	6.1	7.4	6.1	1.5	16.6
A-83-01	3.7	4.4	2.7	3.5	13.9	15	4.2	0	10.9

Red indicates strong binding, green indicates weaker binding. A-83-01 is included as an example of a TGFβ-specific inhibitor.

5.2.3 Specificity of LDN-193189 and K02288 against the kinome

The DSF data accumulated at the SGC provided reasonable coverage across the kinome (80-131 kinases tested, depending on the inhibitor) (table 5.3). From a review of these data (figure 5.7A) it was clear that dorsomorphin was less selective than the other compounds, as it gave a high Tm shift (>8 °C) for multiple kinases such as CLK1, CLK4, STK17A and STK17B. LDN-193189 showed slightly greater specificity (Figure 5.7B),

potentially due to the increase in the size of the 4-quinoline group (figure 5.3A). However, for some off-target kinases that can accommodate this group, the Tm shift was also increased relative to dorsomorphin. AMPKα2 was one example and was perhaps not surprising having been the original target of dorsomorphin [91]. K02288, however, appeared to be far more selective for the BMP/TGFβ receptor family than other kinases (figure 5.7C). From 80 kinases screened outside the BMP/TGFβ receptor family only 3 kinases showed a Tm shift above 4°C, TNIK, GAK and CK1ε (table 5.3).

Table 5.3 Table of the available SGC DSF data for dorsomorphin, LDN-193189 and K02288.

Kinase	Tm shift (°C)		
	Dorsomorphin	LDN-193189	K02288
AAK1	3.8		0.5
ActRIIA	6.1	8.6	8.3
AKT3	0.5		0.5
ALK1	10.6	15.7	14.2
ALK2	10.3	14.2	13.1
ALK3		14.0	12.8
ALK4	6.1	11.0	8.2
ALK5	7.4	12.0	11.6
ALK6	12.4	17.1	13.9
AMPKα1	3.9	4.3	0.2
AMPKα2	7.1	8.3	1.2
BMPRII	1.5	-0.4	3.5
BMX	1.4		1.0
BRK	2.8		0.0
CAMK1δ	-0.3		-0.3
CDK2	2.5	0.7	0.1
CDK8	0.5		0.1
CDKL1	2.0		0.0
CHEK2	1.7	1.4	0.0
CK1γ1	0.4	0.3	0.7
CK1γ3	1.2		2.0
CK1ε	5.9		4.5
CK2α1	3.8		0.3
CLK1	8.3		-0.1
CLK2	5.0	3.3	0.0
CLK3	3.7	0.7	0.0
CLK4	8.7	5.3	0.3
DAPK3	1.2		0.2
DCAMKL1	0.3		0.1
DRAK1	9.0		1.0
DRAK2	10.0	2.1	0.6
DYRK1A	4.4		-0.9
DYRK2	5.1		0.3
ERK1	0.0	0.2	0.2
FES	0.5	0.4	0.4
GAK	4.3		6.9
GRK1	0.4	-0.7	0.0
GRK3	-0.1		0.1
GUCY2D	-0.1	-0.2	
Haspin	2.3	0.4	0.3

Kinase	Tm shift (°C)		
	Dorsomorphin	LDN-193189	K02288
JNK1	0.0	0.5	0.2
JNK2	0.4	-0.6	-0.1
LOK	2.5	3.1	2.1
MEK2	-0.1	2.1	2.4
MEK6	0.1		1.9
MPSK1	5.3	5.3	0.3
MRCKβ	0.0	0.6	0.4
MST3	4.6	7.9	0.3
MST4	0.7	0.8	0.2
MYLK4	4.3		1
NDR1	2.8		0.1
NEK1	0.1	0.2	0.2
NEK11	2	-0.3	0.3
OSR1	1.6	0.9	0.0
p38β	1.1		0.5
p38δ	0.0	-0.4	-0.2
PCTAIRE1	4.7		-1.8
PIM1	0.6	0.5	0.1
PIM2	0.6	-1.3	0.1
PIM3	0.6	-0.2	-0.2
PKCη	0.9		1.2
PKD2	0.1		0.2
PKD3	0.9		1.2
PKMYT1	0.1		-0.4
PKN1	2.3	2.3	0.5
PLK4	4.2	3.7	0.0
PRKX	2.9		0.2
PTK7	-1.2	-0.3	-1.3
RSK1	2.8	3.3	1.1
RSK4	2.7		1.4
SLK	1.6	1.2	1.0
SRPK2	0.2		0.2
STK33	4.2	6.5	1.1
TNIK	4.2	4.6	6.1
TOPK	0.3		-0.2
TTK	1.0		1.1
VRK1	0.4		0.5
VRK2	0.2	-0.2	0.6
YANK1	0.5	1.2	0.2
ZAK	0.5	2.7	-0.1

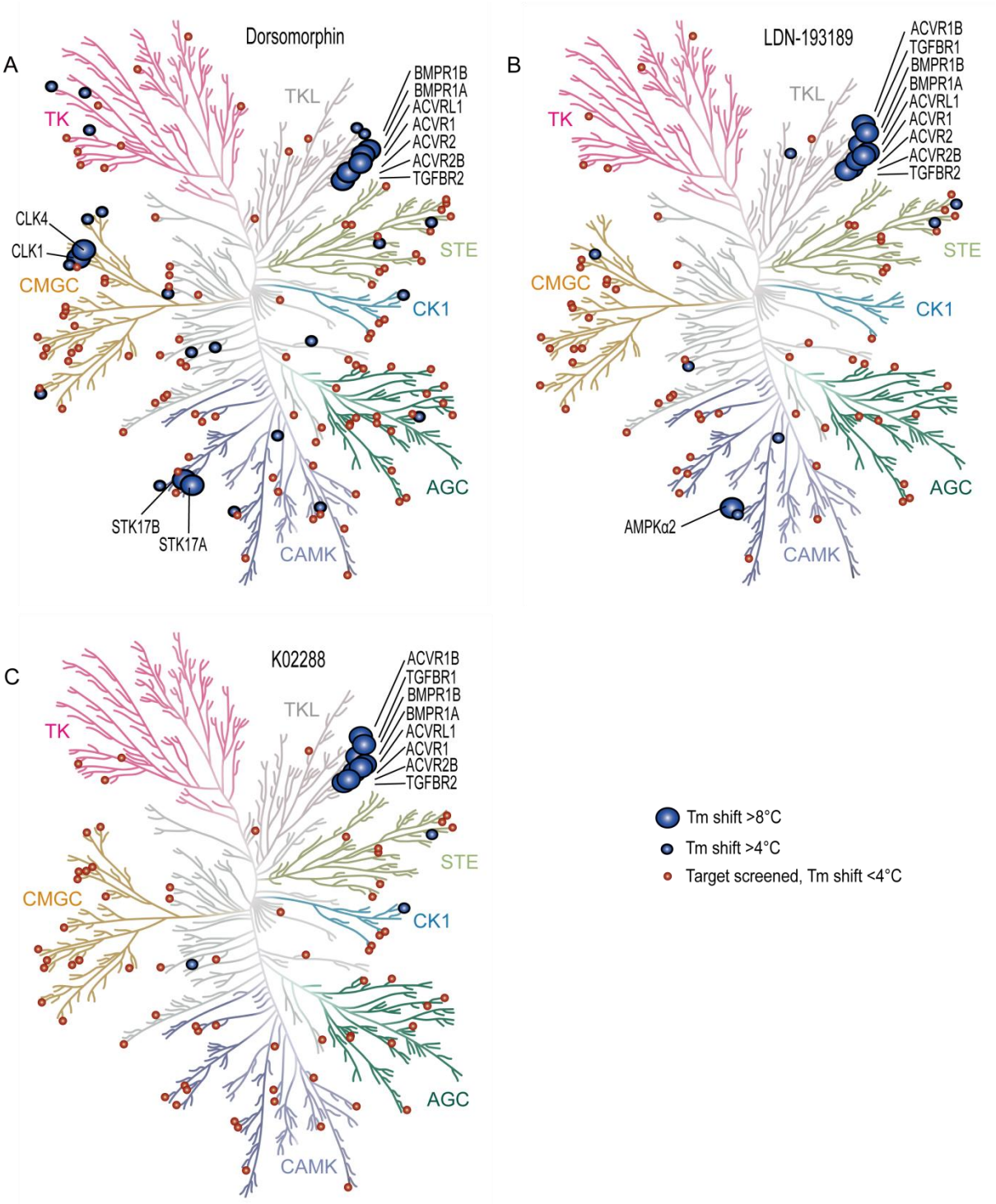


Figure 5.7 Evolutionary tree representation of the kinome highlighting inhibitor selectivity. (A) Dorsomorphin, 131 kinases screened; B) LDN-193189, 80 kinases screened; C) K02288, 80 kinases screened. For most kinases, Tm shifts >4°C (small blue circle) typically equate to an IC50 of <1μM whereas > 8°C equates to an IC50 ≤100nM (large blue circle)[133]. The kinases are grouped as follows: AGC- containing PKA, PKG, PKC families; CAMK - calcium/calmodulin-dependent protein kinases; CK1 - casein kinases; CMGC - containing CDK, MAPK, GSK3, CLK families; STE - homologs of yeast Sterile 7, Sterile 11, Sterile 20 kinases; TK - tyrosine kinases; TKL - tyrosine kinase–like.

K02288 and LDN-193189 were sent to Nanosyn (www.nanosyn.com) for enzymatic screening using a microfluidic capillary electrophoresis screening platform. This method directly measures the proportion of phosphorylated peptide by utilising the difference in electrophoretic mobility of the substrate and phosphorylated product. This was used to assay a panel of 200 kinases measuring percentage inhibition of the kinase determined at both 0.1 μM and 1 μM inhibitor concentration (Table 5.4). Analysis of the percentage inhibition revealed a general trend of increased inhibition by LDN-193189 at both concentrations when compared to K02288 (figure 5.8A). At 0.1 μM only 2 kinases for each compound were inhibited at >50 %. K02288 inhibited ABL and ARG (ABL2), whereas LDN-193189 inhibited ABL and SIK2. However at 1 μM this was increased to 21 kinases for LDN-193189 compared to only 8 kinases inhibited at >50 % using K02288. A plot to show correlation of the inhibition at 1 μM between the two inhibitors revealed that perhaps only half of the kinases were similarly inhibited by both compounds; these included ABL, ARG, MAP4K2 and MAP4K4 (figure 5.8B). It is interesting to note the low percentage inhibition of K02288 against KDR compared to LDN-193189, a protein kinase previously identified as a target for LDN-193189 and its precursor dorsomorphin. Some 14 % of the kinases were screened in both the Nanosyn assay and the SGC DSF assay which showed significant correlation between the two. The one exception was the tyrosine kinase BRK (breast tumor kinase), which showed a 0 °C Tm shift for K02288, yet showed inhibition > 50 % at 1 μM inhibitor concentration. The Tm shift data for kinases outside the BMP/TGFβ family is accumulated over years and this result suggests that in this instance the data should be repeated.

Table 5.4 Nanosyn data of enzymatic inhibition.

	Compound			
	K02288 (μM)		LDN-193189 (μM)	
Kinase	0.1	1	0.1	1
ABL1	79	89	62	84
AKT1	-2	-1	0.8	1.6
AKT2	-2	-1	1	1.8
AKT3	-2	1.2	0.2	1.5
ALK	-1	2.5	3.7	12
AMPKα1	-1	-0	5.2	23
ARG	66	85	33	73
ARK5	1.1	4.1	16	56
AURKA	-2	-1	4.2	23
AURKB	0.3	1.5	2.3	15
AURKC	-1	3.2	4.2	25
AXL	0.9	1.5	-4	1.1
BLK	3.2	27	3.9	24
BMX	0.5	19	0.2	4
BRAF	-0	12	3.5	5.2
BRK	13	66	13	57
BRSK1	2.3	-0	3	17
BRSK2	0.9	2.7	4.4	22
BTK	0.5	6.6	0.5	1.8
CAMK1δ	-2	-1	0.8	0.9
CAMK2α	-2	-2	-0	-8
CAMK2δ	-1	1.4	3.1	8.7
CAMK4	-1	-1	0.1	-0
CDK1/cycB	1	1.5	0.5	5
CDK2/cycA	-0	-0	1.2	4.1
CDK2/cycE	0.2	1.5	2.8	5.3
CDK3/cycE	1.6	-91	1.6	7
CDK4/cycD	0.1	1.3	0.8	8.3
CDK5/p35	0.3	1.9	1.4	8.5
CDK6/cycD3	1.7	2.8	7.2	22
CHEK1	-1	-2	2.9	0.3
CHEK2	-1	2.5	1.5	2
CK1α	-0	7.2	14	56
CK1γ1	-0	6.1	3.1	14
CK1γ2	-1	5.5	1.6	3.9
CK1γ3	2.7	19	2.3	5.5
CK2	6.2	2.8	0.7	4.2
CLK2	-0	0	2.3	35
CLK3	1.4	1.5	0.5	2.6
CRAF	-5	-8	0.6	0.1

	Compound			
	K02288 (μM)		LDN-193189 (μM)	
Kinase	0.1	1	0.1	1
CSK	3.2	13	3.3	6
DAPK1	-1	-1	1.5	0.4
DCAMKL2	-2	2.3	0.8	3.4
DDR2	-0	11	5.5	27
DYRK1A	1.9	4.1	10	53
DYRK1B	2.6	6	7.1	41
DYRK2	-1	-3	2.3	0.2
EGFR	2.4	12	1.5	4.8
EPH-A2	-2	3	-2	-3
EPHA3	-6	4.3	2.2	6.3
EPH-A4	-2	17	7.1	10
EPHB2	-0	21	8.4	24
EPHB3	-2	-1	5.6	4.3
EPHB4	-0	5.9	-1	6.6
ERB-B2	-5	-3	2.1	2.1
ERBB4	-5	0.9	6.1	11
FER	0.8	7.5	2.8	3.7
FES	-1	-1	2.2	2.4
FGFR1	-1	-3	3.4	19
FGFR2	0.3	1.1	5	18
FGFR3	-2	0.2	5	16
FGFR4	4	-0	0.3	2.4
FGR	6.7	37	13	53
FLT1	0	-2	0	17
FLT3	-2	0.4	3.4	12
FLT4	0.6	1.4	3.4	13
FMS	6.1	33	3.7	26
FYN	10	43	18	60
GRK6	0.2	1.4	-0	0.2
GRK7	0.2	-3	1.8	0.8
GSK3α	0.6	2.1	3.2	1.1
GSK3β	0.5	1.6	2.7	2.1
HCK	8.9	51	9.7	33
HIPK1	1.7	3.1	0.6	4.4
HIPK4	0.8	4.2	-0	4.9
IGF1R	-3	-1	4.1	5.4
IKK-α	7.4	2.7	3	-2
IKK-β	-2	1.3	0.3	1.4
IKK-ε	-6	-2	7.5	7.4
INSR	-4	0.7	3.8	4.5

	Compound			
	K02288 (μM)		LDN-193189 (μM)	
Kinase	0.1	1	0.1	1
IRAK4	-1	0.3	0.7	1.9
IRR	-1	-1	1.9	2.2
ITK	-0	-0	0.9	3.6
JAK1	-0	0.4	2.7	4.2
JAK2	-0	-0	2	0.5
JAK3	5.6	2.3	-0	0.1
JNK2	0.9	-0	3.9	1.3
KDR	0.5	3.9	19	69
KIT	4	9.2	7	41
LCK	13	58	12	46
LOK	3.5	25	8.3	38
LRRK2	7.1	27	8.7	27
LTK	3.6	31	3.6	4
LYNA	9.7	49	9.1	39
LYNB	6	43	7.4	34
MAP4K2	18	53	21	65
MAP4K4	46	89	31	81
MAPK1	-3	9.9	2.6	1.4
MAPK3	-0	-1	1.9	-2
MAPKAPK2	-2	-1	0.7	-0
MAPKAPK3	-1	-2	1.6	0.5
MARK1	1.3	4.5	13	62
MARK3	2.6	5.6	11	63
MARK4	2.4	7.2	13	61
MEK1	-1	10	1.4	1
MER	5.1	40	4.4	10
MET	0.6	-2	6.2	8.4
MINK	36	86	26	71
MKNK1	0.7	1.3	7.1	37
MNK2	1.9	3.2	3.7	31
MRCKα	2	6.9	0.8	6.8
MRCKβ	0.6	2	-0	2.1
MSK1	-2	-1	1.1	1.3
MSK2	-1	1	0.9	0.8
MSSK1	1.8	-0	-3	1.5
MST1	-1	1.1	4	7.9
MST2	-2	-1	5.1	6
MST4	0.4	0	-1	-5
MUSK	1.2	-1	5.8	26
NEK1	-1	0.8	2.1	3.4

	Compound			
	K02288 (μM)		LDN-193189 (μM)	
Kinase	0.1	1	0.1	1
NEK2	-2	-2	2.8	1.2
NEK6	-2	-1	2.3	-1
NEK7	0.2	-3	7.9	1.7
NEK9	5.3	-2	1.9	6.6
P38α	-3	1.5	4.4	3.3
P38β	-0	9.7	0.4	1.5
P38γ	-1	-1	1.2	1.3
P38δ	-3	-3	1.8	0.5
P70S6KB1	-2	-1	1	0.6
PAK1	-0	0.6	4.3	4.9
PAK2	-1	-1	-0	-0
PAK3	-3	-2	2.9	1.8
PAK5	-1	-1	6.2	8.5
PAK6	0.8	1.9	8.8	5.1
PAR1Bα	2.3	7	6.9	43
PASK	1.1	0.6	1.9	1.4
PDGFRα	5.3	37	18	56
PDGFRβ	7.7	43	39	75
PDK1	-0	0.5	1.1	1.9
PHKγ1	1.8	4.4	5.2	27
PHKγ2	-3	-3	3.5	4.9
PI3-K-α	2.6	2	2.1	1.5
PI3-K-δ	0.8	2.5	-1	10
PI4-K-β	0.8	-4	3.5	5.8
PIM-1	0.2	1.5	-1	-1
PIM-2	0.9	2.3	-0	-0
PIM3	2.3	5.9	2	-1
PKA	0.3	6.1	1.4	2.1
PKCα	0.2	2.2	1	1.1
PKCβ1	-1	0.7	-2	0.2
PKCβ2	-0	-1	1	1.8
PKCγ	-4	-1	2.1	4.1
PKCη	-2	-4	2.2	-2
PKCθ	-0	0.1	1.4	1.3
PKCι	0.3	1.6	1.4	1.5
PLK1	-2	-2	0.8	1.3
PRAK	-0	-0	0.8	0.1
PRKD1	-0	6.1	4.6	15
PRKD2	0.6	6.6	4.5	13
PRKD3	-1	3.2	3.2	13

	Compound			
	K02288 (μM)		LDN-193189 (μM)	
Kinase	0.1	1	0.1	1
PRKG1	0.3	3.3	1	1.2
PRKG2	1	5	1.7	9.3
PRKX	-1	-1	0.2	1.2
PTK5	2.2	22	4.1	14
PYK2	-1	0	4.6	9
RET	2.1	3.9	4	16
ROCK1	0.5	3.1	0	0.7
ROCK2	-1	0.8	1.4	1.5
RON	3	-1	-2	12
ROS	1.8	6.8	3.8	3.7
RSK1	0.8	1.8	0.6	8
RSK2	0.8	2.2	1.2	5.4
RSK3	0.6	1.1	1.5	5.6
RSK4	0.1	1	0.8	3.8
SGK1	1.1	-2	1.1	0.8
SGK2	2.7	1.3	2.3	0.2
SGK3	-2	-1	1.6	2.3
SIK	9.4	57	63	96
SPHK1	-4		-1	-11
SPHK2	-2	4.4	-1	-3
SRC	5.1	40	8	33
SRMS	0.8	17	-1	8.9
SRPK1	3.2	2.7	-1	-2
SYK	-0	-2	-1	-12
TBK1	1.5	3.7	4.5	16
TEC	0.3	14	1.8	5.5
TIE2	1	3	2	16
TNK1	0.4	10	18	63
TNK2	3.6	31	2	10
TRKA	1.5	2.8	1.9	7.2
TRKB	1.8	3.9	0.8	3
TRKC	2.7	2.4	1.6	3.9
TSSK1	-1	-2	3.5	5.2
TSSK2	-1	-3	0	0.5
TTK	-4	-1	4.5	7.6
TXK	2.8	36	4.1	16
TYK2	-0	1.8	-1	-2
TYRO3	1.8	19	13	58
YES	4.9	34	12	50
ZAP70	-2	1	4.7	1

Screening of 200 kinases against LDN-193189 and K02288 at both 0.1 μM and 1 μM compound concentrations. The red highlights show significant percentage inhibition (over 50 %).

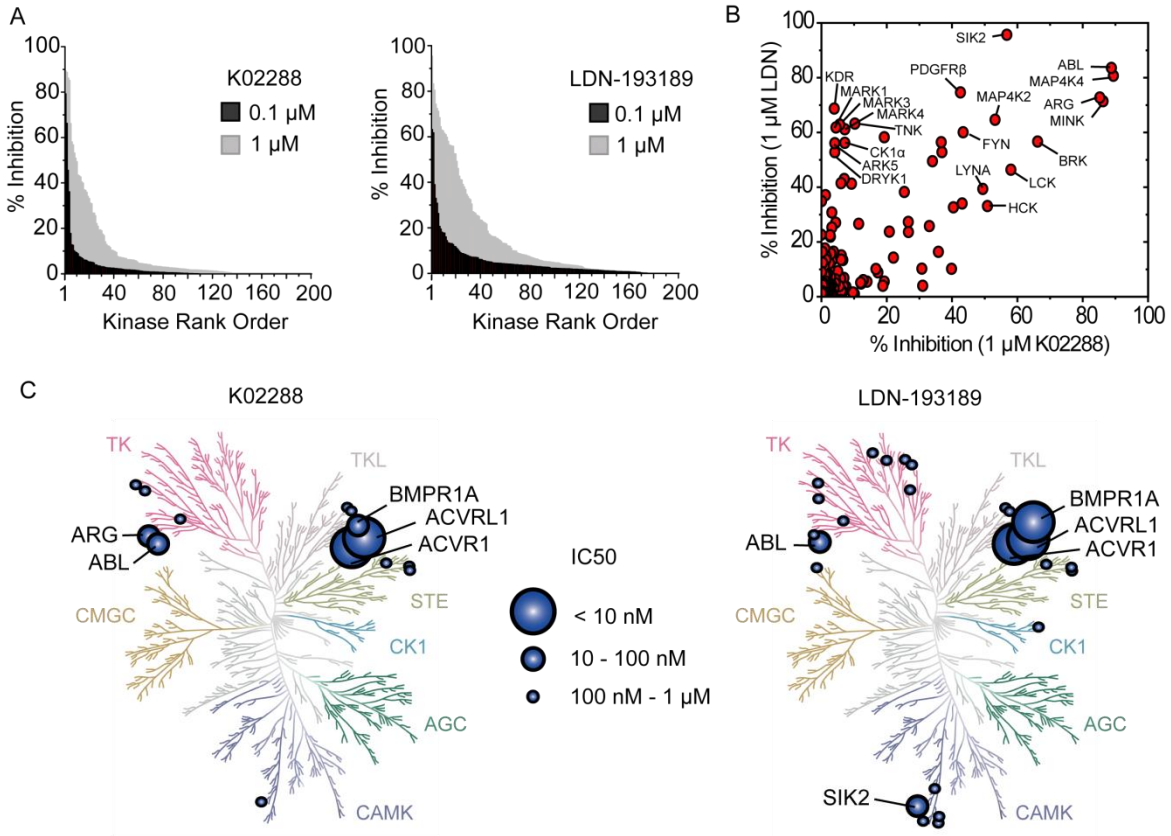


Figure 5.8 Analysis of K02288 and LDN-193189 inhibition from a screen of 200 kinases. (A) Percentage kinase inhibition by both compounds at 1 µM and 0.1 µM, in rank order. (B) A scatter plot of the percentage kinase inhibition at 1 µM inhibitor concentration. There is some correlation between the kinases targeted by the inhibitors, however, LDN-193189 targets a large number of kinases more strongly than K02288. (C) Evolutionary tree representation of the kinome mapping the Nanosyn screening data for K02288 and LDN-193189. The data for the BMP/TGFβ receptors were provided by the Yu lab, Harvard.

5.2.4 Effects of construct variance on inhibitor binding

5.2.4.1 Mutant ACVR1 compound binding

As the small molecule compounds would be intended for use within the FOP disease model, it is important to ensure the binding observed for the wild-type ACVR1 is still observed against the mutant proteins. The Tm shift for each mutant generated in chapter 4 was measured by DSF against a number of BMP inhibitors, with those of interest displayed in figure 5.9.

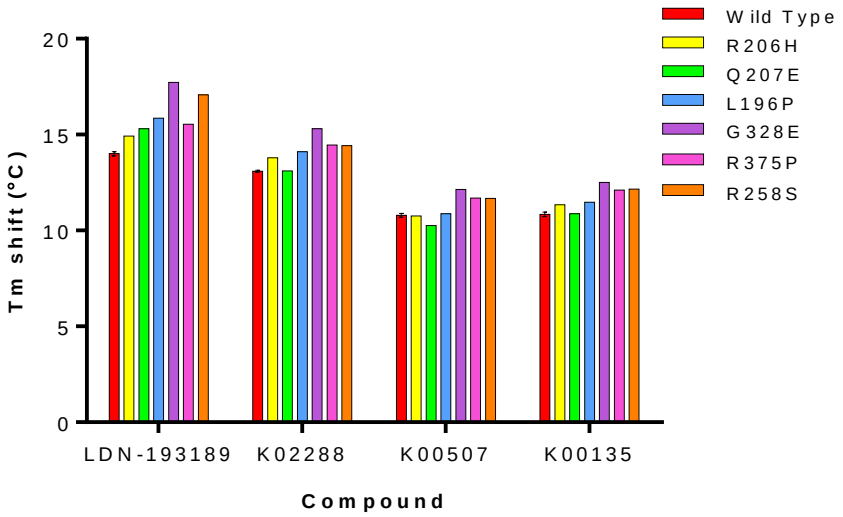


Figure 5.9 Tm shifts of ACVR1 FOP mutants compared to wild type ACVR1. The graph showed data for the wild-type protein from three independent experiments ± S.E.M.

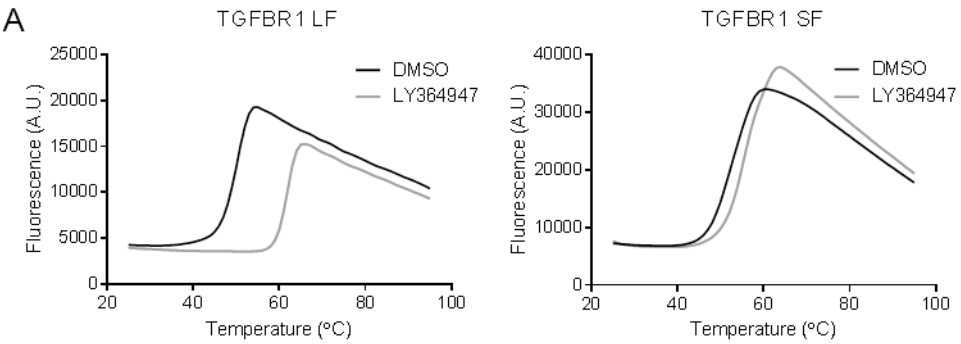
All of the inhibitors bound to the mutants and yielded similar Tm shifts to those of the wild-type ACVR1 receptor. As a general trend, the kinase domain mutants (R375P, G328E and R258S) displayed marginally higher Tm shifts than the other variants. The apparent melting termperatures for the mutant apoproteins were similar to wild-type apoprotein at 47.6 °C, ranging between 45.5 °C – 48.7 °C. The exception to this was ACVR1 R375P which was destabilized by 5.3 °C (Tm = 42.3 °C) compared to the wild type. This perhaps explains the observed aggregation of ACVR1 R375P during purification, as described in chapter 4.

5.2.4.2 TGFBR1 domain boundary affects inhibitor binding affinity

It was previously noted that the GS and kinase domain construct of ACVR1 (a.a. 172 – 499) gave similar values in DSF assays as the kinase domain ACVR1 Q207D construct (a.a. 201 – 499). In screening TGFBR1, however, it was noted that the kinase domain construct (a.a. 198-503) gave strikingly lower Tm shift values than the GS and kinase

containing construct (a.a 162 – 503) (figure 5.10). The raw data curves for both TGFBR1 constructs showed a smooth transition between native and denatured protein, indicating that both proteins are properly folded (figure 5.10A). The comparable Tm values also implied that the altered Tm shifts were reflective of a true change in binding affinity rather than an artefact of the assay.

A



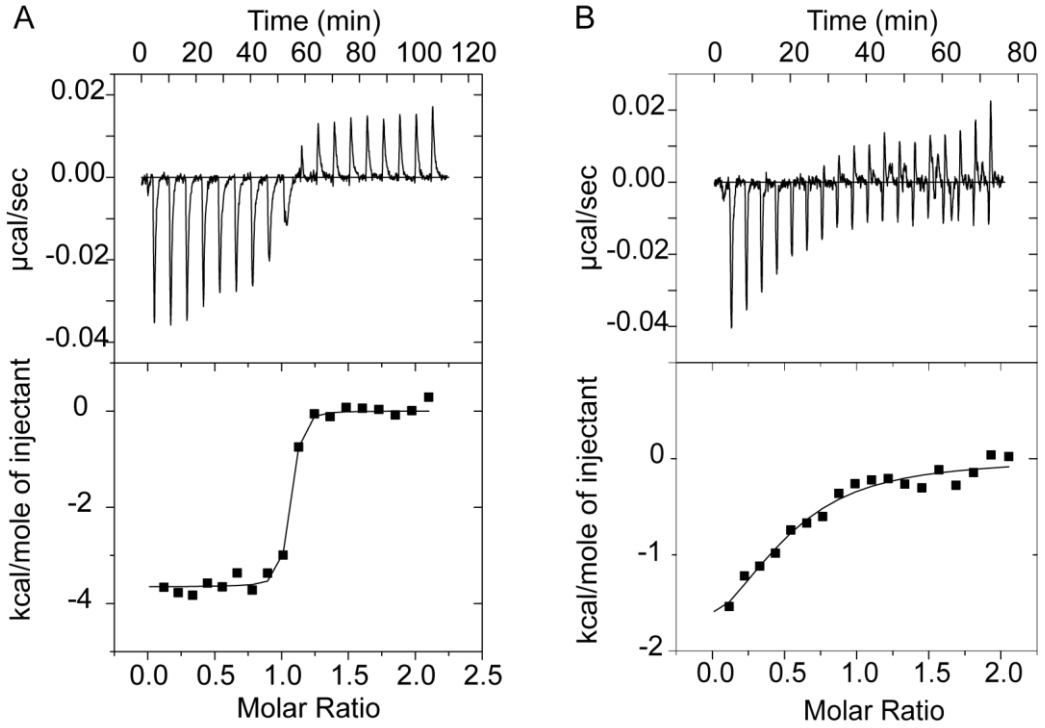
B

TGFBR1 Construct	Compound	Tm (°C)
LF (a.a. 162 – 503)	-	49.4
	LY364947	61.5
SF (a.a. 198 – 503)	-	52.4
	LY364947	55.6

Figure 5.10 An example of the change in Tm shift observed between two constructs of TGFBR1; long form (LF; 162 – 503) and short form (SF; 198 – 503). (A) The raw DSF data for LF and SF TGFBR1 with known TGFBR1 inhibitor LY364947. (B) Tm values for the reference (DMSO) and the compound LY364947. The Tm shifts observed for this compound were LF = 12.1 °C, SF = 3.2 °C.

To further discern whether this was an artefact of the assay, isothermal titration calorimetry (ITC) measurements were conducted to determine the K02288 dissociation constants for each construct of TGFBR1 (figure 5.11). Again, the longer construct (a.a. 162 – 503) containing the GS domain displayed far tighter binding (K_d = 4 nM) compared to the shorter construct (a.a. 198 – 503; K_d = 2.36 μM)(figure 5.11). Although the shorter construct did appear to reach saturation, the results showed poor stoichiometry, which

could be a result of slight precipitation formed during the experiment. The striking difference in the dissociation constants agrees with the earlier observations with the DSF assay.



Construct	Compound	Temperature (°C)	Stoichiometry (N)	Kd (nM)	ΔH (kJ mol^{-1})	TΔS (kJ mol^{-1})	ΔG (kJ mol^{-1})
TGFBR1 (a.a.162-503)	K02288	10	1 ± 0.1	4 ± 0.0	-15.3 ± 0.2	30.2	-45.5
TGFBR1 (a.a.198-503)	K02288	10	0.5 ± 0.1	2.4x10^{3} ± 0.8	-10.3 ± 2.3	20.1	-30.5

Figure 5.11 ITC measurements of K02288 binding to two constructs of TGFBR1. ITC traces showing the raw (top panel) and integrated data (lower panel). (A) 50 µM TGFBR1 (a.a. 162 – 503) injected into 5 µM K02288.(B) 88 µM TGFBR1 (a.a. 198 – 503) injected into 9 µM K02288. Both experiments were performed at 10°C. Samples were buffered in 50 mM Hepes pH 7.5, 150 mM NaCl, 1 mM DTT.

5.2.5 Co-crystal structures of ACVR1 and inhibitor compounds

The co-crystal structures of the compounds were solved in complex with the kinase domain ACVR1 Q207D construct (a.a. 201 – 499), with the exception of the imidazo[1,2-b]pyridazine compound K00135 which was co-crystallised with the equivalent ACVR1 R206H construct (a.a. 201 – 499). All the diffraction data were collected at Diamond Light source by staff at the SGC and the structures in this chapter were refined by Apirat Chaikuad (SGC). The overall structures of the protein were discussed in chapter 4.

5.2.5.1 Binding mode of LDN-193189

The complex of LDN-193189 and ACVR1 Q207D formed diffraction quality crystals at 20 °C, in a sitting drop with a mix of 75 nL protein solution and 75 nL of a reservoir solution containing 20 % PEG 3350 and 0.20 M $(NH_4)_2H$(citrate). The diffraction dataset was collected on the Diamond I03 beamline and the structure refined to 1.82 Å.

The structure showed the binding of the pyrazolopyrimidine N4 position to the hinge residue H286, consistent with dorsomorphin structure, and also showed conservation of the water mediated hydrogen bond to the αC helix residue E248 (figure 5.12A). The increased interaction surface created by the larger quinoline was also evident with increased hydrophobic interactions with ACVR1 residues V222, L263, L343 and A353 (figure 5.12B).

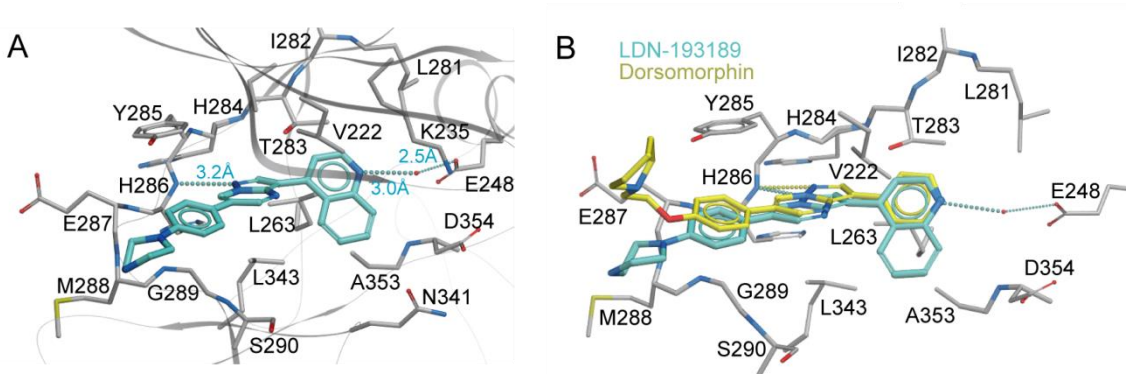


Figure 5.12 Binding of LDN-193189 to ACVR1. (A) LDN-193189 forms a hydrogen bond to H286 in the hinge, and a water mediated hydrogen bond to E248. (B) Overlay of LDN-193189 and dorsormorphin (pdb 3H9R) showing the increased interaction surface provided by the quinoline group.

5.2.5.2 Binding of imidazo[1,2-b]pyridazine compounds

Two compounds of the imidazo[1,2-b]pyridazine series that gave high Tm shift values against ACVR1 were K00135 and K00507 (table 5.1, compounds 12 and 14 respectively). K00135 was co-crystallised with ACVR1 R206H (a.a. 201 – 499). Crystals were grown at 20 °C, in a sitting drop with a mix of 75 nL protein solution and 75 nL of a reservoir solution containing 1.60 M $MgSO_4$ and 0.1 M MES. The diffraction dataset was collected on the Diamond I02 beamline and the structure refined to 2.42 Å.

In a previous co-structure with PIM1, K00135 showed an unusual non-hinge binding mode likely due to a hinge proline residue (P123) that prevented the typical hydrogen bonding pattern (figure 5.13A). Unlike the structure of PIM1, K00135 was able to bind as expected to the hinge region of ACVR1 with a hydrogen bond to H286. In addition, a water mediated hydrogen bond was made to the E248 in the αC helix (figure 5.13B).

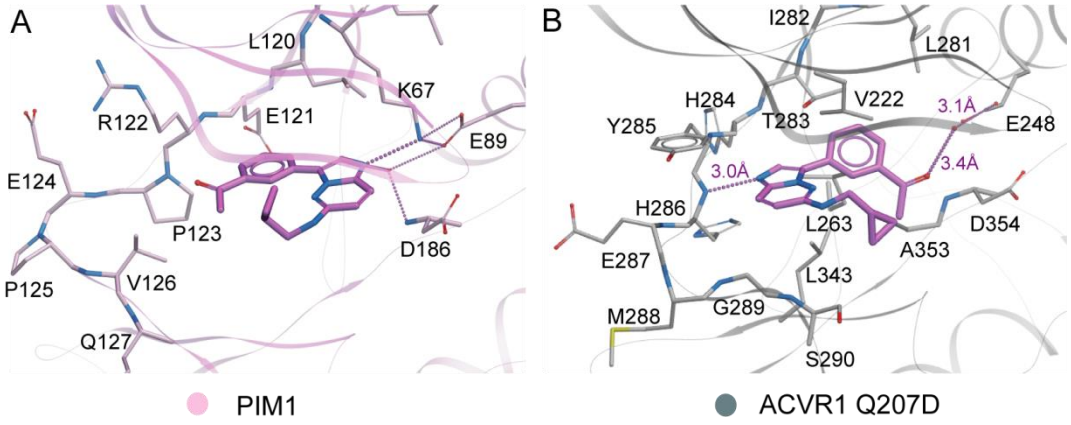


Figure 5.13 Binding mode of K00135 (A) The binding mode of K00135 in PIM1 (pdb 2c3i) (B) Binding of K00135 to ACVR1 showed the conserved hydrogen bond to H286 and a water-mediated bond to E248 in the αC.

Co-crystals of ACVR1 Q207D and K00507 were grown at 20°C, in a sitting drop with a mix of 50 nL protein solution and 100 nL of a reservoir solution containing 0.20 M Na/KPO_4, 20.0 % PEG 3350 and 10.0 % ethylene glycol. A diffraction dataset was collected on the Diamond I03 beamline and the structure refined to 2.00 Å.

The structure showed that K00507 also bound to H286 in the hinge, similar to all previously observed ACVR1 inhibitor interactions. The density of K00507, however, indicated that the phenyl-ethanone group was able to bind in two different conformations, showing the availability of potential hydrophobic interactions and space within the back pocket that could be explored further (figure 5.14). Compared to the binding of dorsomorphin and LDN-193189, the solvent facing groups of both K00135 and K00507 orient themselves into the pocket, rather than along the hinge (figure 5.14B).

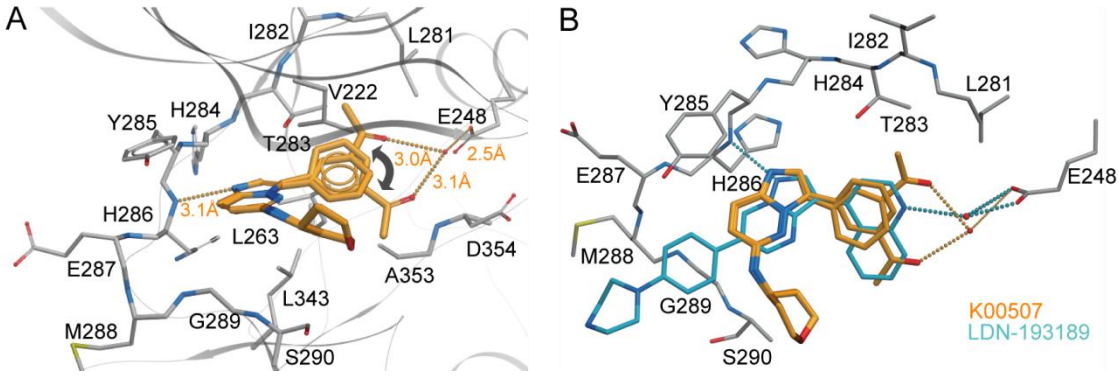


Figure 5.14 Binding mode of K00507 (A) The compound bound ACVR1 H286 in the hinge and showed two conformations of the phenyl-ethanone, both of which were capable of forming water mediated hydrogen bonds to E248. (B) Superposition of ACVR1 co-structures with LDN-193189 and K00507.

5.2.5.3 Binding mode of novel compound K02288

The DSF screening identified a novel compound, K02288, with potent binding to ACVR1. Furthermore, this compound showed some selectivity for ACVR1 over the TGFβ receptors, as well as excellent specificity against the kinome. To investigate the binding mode of the 2-aminopyridine scaffold, K02288 was co-crystallised with ACVR1 Q207D. Diffraction quality crystals were grown at 20 °C, in a sitting drop with a mix of 50 nL protein solution and 100 nL of a reservoir solution containing 1.6 M Na/KPO_4, 0.1 M HEPES pH 7.5. A diffraction dataset was collected on the Diamond I04 beamline and the structure refined to 2.15 Å.

The trimethoxy group showed excellent shape complementarity to ACVR1, forming many hydrophobic contacts in the ATP pocket, including with A353, L343, L263, V222 (figure 5.15A). The nitrogen of the pyridine ring formed a hydrogen bond to the hinge residue H286, similar to the interaction observed in other inhibitor bound structures. K02288 also formed a water mediated hydrogen bond to the catalytic lysine (K235). This water molecule also bound a second water molecule, which further bound both the DFG motif

aspartate (D354) and the catalytic loop asparagine (N341) (figure 5.15 A and B). The binding potency may also be explained by the formation of a hydrogen bond from the amine group to H284 in the hinge region as well as a hydrogen bond from the phenol group located in the solvent channel of the ATP pocket.

The superposition of K02288 and LDN-193189 showed that the orientation of K02288 to the hinge is slightly altered to that of LDN-193189 (figure 5.15C). This allowed the trimethoxyphenol group to form similar interactions within the pocket as the large quinoline group. The trimethoxyphenol group also extended further back into the pocket towards the gatekeeper residue (T283) which may increase specificity of the compound (figure 5.15C).

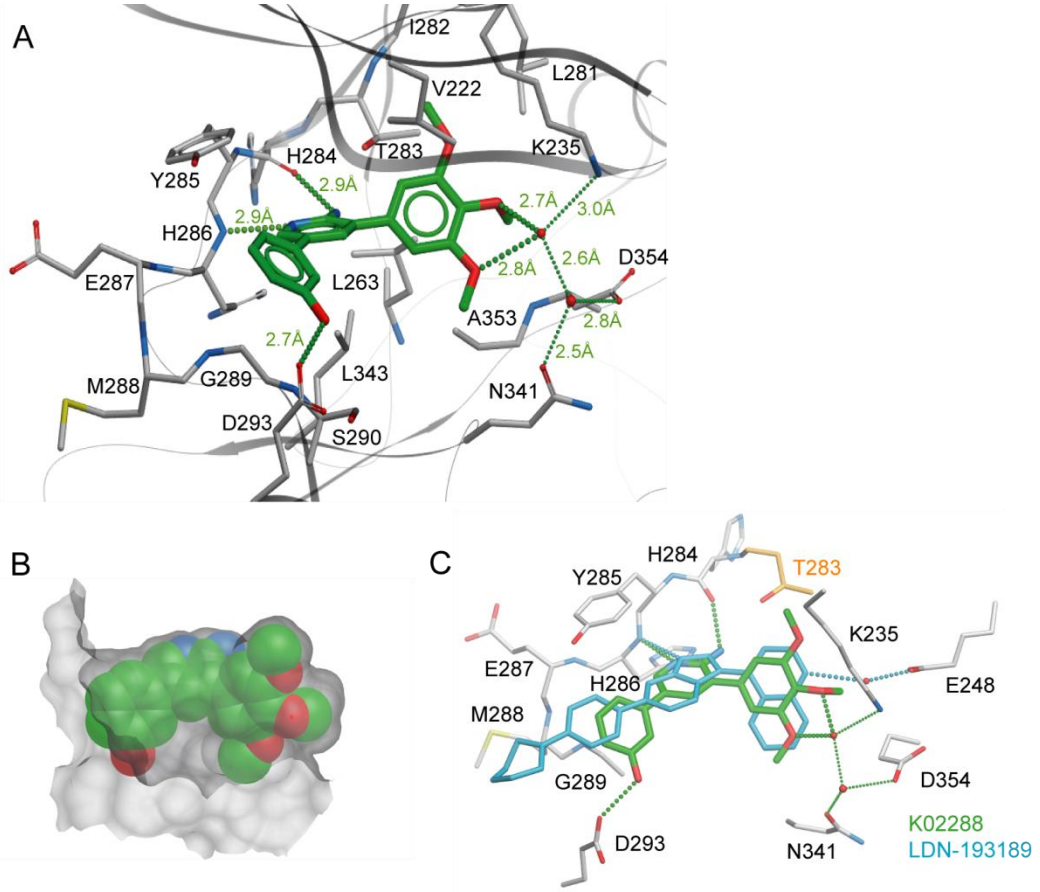


Figure 5.15 Binding mode of K02288 (A) Binding of the 2-aminopyridine to H286 and H284 in the hinge region as well as the phenol group to D293. The trimethoxyphenyl group forms a water-mediated hydrogen bond to the catalytic lysine K235. (B) A spacefill representation of K02288 shown against a surface mesh view of ACVR1 demonstrating the shape complementarity. (C) Superposition of K02288 (green) and LDN-193189 (blue) in the ATP pocket of ACVR1. The methoxy group extends towards the gatekeeper residue T283 (orange) with a 3.5 Å distance between the closest atom contacts.

5.3 Discussion

The high conservation of the ATP pocket in human kinases presents a major challenge for the design of selective kinase inhibitors. This is particularly the case for selectivity between the BMP and TGFβ receptors due to their conserved structures and sequence similarity. Differential scanning fluorimetry provided a useful tool to identify potent binders from a large compound library. This method identified the imidazo[1,2-

b]pyridazine series of compounds, and a novel compound, K02288, as potential inhibitors of ACVR1, as well as confirming the potency of LDN-193189.

The imidazo[1,2-b]pyridazine scaffold has been identified previously as a potent kinase inhibitor. One such compound, SGI-1776, entered clinical trials as an inhibitor of PIM1 but it has been discontinued due to cardiac toxicity [138]. Whilst this compound only gave a Tm shift of 6.6 °C for ACVR1 (compound kindly provided by Angela Russell, Oxford), further optimisation of other imidazo[1,2-b]pyridazine compounds (such as K00135) would have to be approached with caution.

K02288 is a relatively low molecular weight compound, but gave a high Tm shift with ACVR1. K02288 was also shown to be highly selective for ACVR1 by both DSF data, and kinase inhibition assays. This was potentially explained by the trimethoxyphenyl group which, as well as forming water-mediated hydrogen bonds to the β3 strand catalytic lysine (K235), formed hydrophobic interactions that perfectly complemented the shape of the ATP pocket. The 2-aminopyridine scaffold of K02288 has previously been used in inhibitors of the NEK and CHK2 kinases, as well as in the clinically-approved inhibitor crizotinib, which targets the ALK tyrosine kinase [136,137,139, 140]. Comparison of the binding of these compounds shows that they all bind in a similar fashion to the kinase hinge region. Recently a new 2-aminopyridine compound was identified as an anti-malarial drug (figure 5.17) [141]. The initial hit compound was from the same Biofocus library as K02288 and contained a similar scaffold. An optimised derivative of this compound is currently being considered for clinical trials (figure 5.17). This finding highlights the potential for the similar optimisation of K02288. As a low molecular weight compound (MW = 352.1 Da), compared to LDN-193189 (MW = 406.2 Da), the addition

of larger functional groups for improved potency and metabolic stability would still keep K02288 within the recommended limits for successful drugs [142].

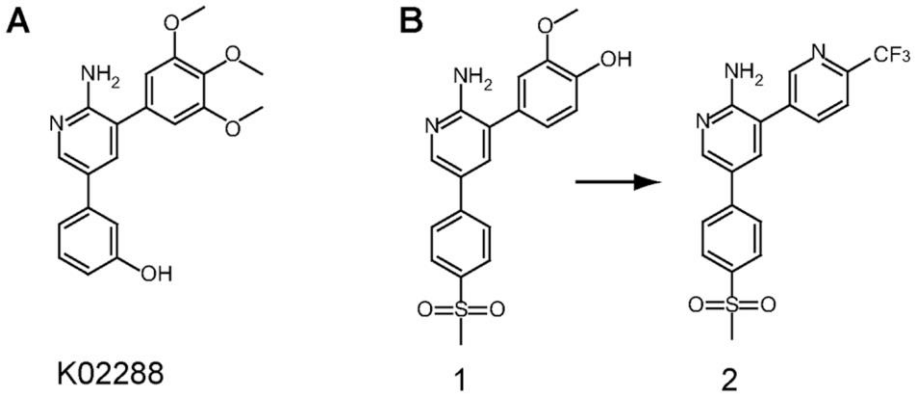


Figure 5.16 The K02288 scaffold is similar to an anti-malarial compound. (A) The 2-aminopyridine scaffold of K02288. (B) A similar initial hit 1 was identified in a recent anti-malarial screen and optimized to the more divergent lead 2 in development of the pre-clinical candidate MMV390048.

The potency and selectivity of LDN-193189 and K02288 was also shown in cellular assays by Georgina Kerr at the SGC (figure 5.18). The ability of K02288 and LDN-193189 to inhibit Smad phosphorylation was confirmed using C2C12 cells and the ligand BMP6, a preferred ligand for ACVR1 [143]. Stimulation with BMP6 led to phosphorylation of Smad1/5/8 that was reduced by both inhibitors in a dose dependent manner, with an apparent IC_{50} of 50 nM for K02288 and 10 nM for the control compound LDN-193189 (figure 5.18). The 5-fold weaker activity of K02288 again suggests that the K02288 scaffold could be optimized further for its cellular and metabolic stability. Both compounds were also tested against other ligands in the TGFβ superfamily to confirm their specificity (figure 5.18). Below 1 μM concentration, neither compound inhibited TGF-β–induced phosphorylation of Smad2, in agreement with their observed specificity for ACVR1 over TGFBR1 (figure 5.18). Interestingly, both compounds showed some weaker inhibition of the Activin A-induced pathway, with near complete inhibition of

Smad2 phosphorylation by 0.5 µM K02288 or LDN-193189 (figure 5.19). This activity likely reflects some weak inhibition of ACVR1B or perhaps the type II activin receptors.

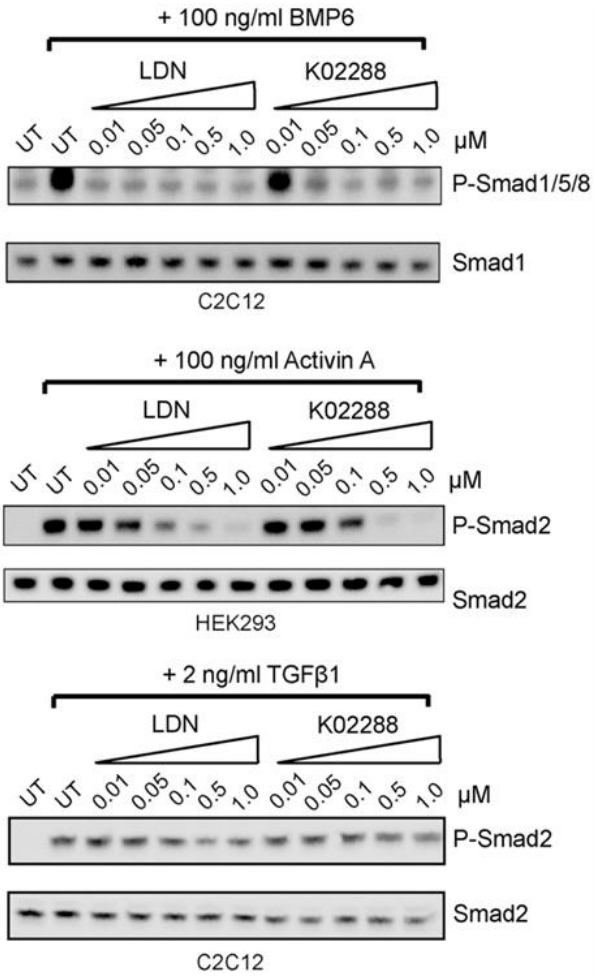


Figure 5.17 K02288 selectively inhibits BMP signalling. This experiment was performed by Dr Georgina Kerr. C2C12 and HEK293 cells were grown in DMEM supplemented with 10% FCS (PAA) and seeded at a density of 1 x 10^5 cells per well in 6-well plates. The next day the cells were starved in DMEM containing 1% FCS for 5 hours. Inhibitors were added at the indicated concentrations for 30 minutes followed by 10 ng/mL BMP4 (Peprotech), 100 ng/mL BMP6 (Peprotech), 100 ng/mL Activin A (Peprotech) or 2 ng/mL TGF-β1 (Cell Signaling) for 1 hour. Cells were washed in PBS and lysed on ice for 30 minutes in lysis buffer (20 mM Tris-HCl pH 7.5, 150 mM NaCl, 1% Triton X-100, 25 mM NaF, 25 mM Na b-glycerophosphate, 2 mM Na3VO4) containing protease inhibitors. Protein concentration in the clarified lysate was determined by Bradford Assay. K02288 and LDN-193189 potently inhibited BMP6-induced Smad1/5/8 phosphorylation in C2C12 cells showing inhibition at 0.05 µM and 0.01 µM inhibitor respectively, but had no effect on TGF-β-induced phosphorylation of Smad2. Activin A-induced P-Smad2 in HEK293 cells was weakly inhibited by K02288 and LDN-193189, showing inhibition at 0.5 µM and 0.1 µM respectively.

The use of DSF identified K02288, a novel scaffold for ACVR1 inhibition. The structures showed similarities in binding for all the compounds, including a water-mediated bond to the αC. As a highly specific small molecule compound, K02288 has potential for further optimisation to increase the potency of compound.

6. Characterisation of K02288 derivatives

6.1 Introduction

The identification of K02288 as a hit compound provides a novel starting point for drug design. In the first instance, it was desirable to increase the selectivity of the compound for ACVR1 whilst maintaining or even increasing the potency. High selectivity is useful to have confidence in the mechanism of action of an investigational drug. Moreover, it may help to limit any potential toxicity arising from off-target effects. However, toxicity and many other important aims for drug design, including metabolic stability, were not addressed at this early stage. Structure Activity Relationship (SAR) studies are used to determine which parts of the hit compound are important for activity. This is achieved through the synthesis of a number of compounds in which functional groups are removed or altered to determine which groups are essential for biological effect.

A number of K02288 derivatives, including hybrids of K02288 and LDN-193189, were synthesized by the group of Greg Cuny at Harvard University. The compounds had been assessed by HPLC and were provided at a purity of ≥ 95%.

In the crystallographic structure of K02288 bound to ACVR1, the pyridine and amide nitrogens in the 2-aminopyridine scaffold were bound to the hinge region of ACVR1. Therefore, the 2-aminopyridine scaffold was largely conserved, while the majority of hybrids explored other ring substituents (R-groups) systematically in order to identify those groups providing potency for ACVR1, as opposed to TGFBR1, and to replace the metabolically vulnerable phenol group of K02288. Binding of these compounds to

ACVR1 and a number of the BMP/TGFβ receptors was measured using DSF as a thermal shift assay. The results of the assay were analysed by looking at the SAR between the compounds.

6.2 Results

6.2.1 Structure Activity Relationship (SAR) Study of K02288

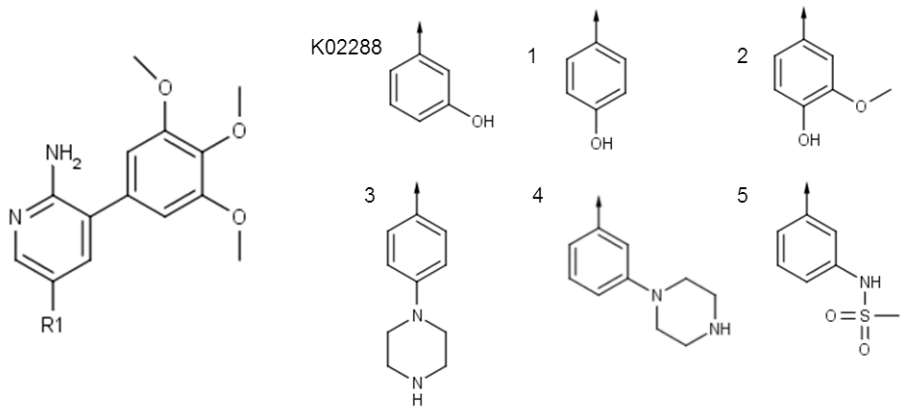
To establish data for a SAR study, 23 derivatives of K02288 were screened using DSF against the primary target WT ACVR1 and the related type I and type II BMP/TGFβ receptors ACVRL1, BMPR1B, TGFBR1, ACVR2B, BMPR2 and TGFBR2. In addition, all derivatives were screened against FOP-associated mutants of ACVR1 to confirm their usefulness for preclinical development.

6.2.1.1 SAR of the solvent-facing R1 group of K02288

The solvent-facing phenol group of K02288 contains a hydroxyl in the meta position that is potentially susceptible to chemical change which could hinder its efficacy *in vivo*. Derivatives at the R1 site (figure 6.1) were compared for their potency towards ACVR1 (as determined by an increased Tm shift) as well as their selectivity against other BMP/TGFβ receptors. In particular, the difference between the Tm shifts observed for ACVR1 and TGFBR1 was taken as a rough guide of selectivity, with a large difference indicating good selectivity (figure 6.1). Shifting the hydroxyl group to the para position (compound 1) produced a slight increase in potency against ACVR1, but resulted in a complete loss in selectivity. Interestingly, selectivity was restored by adding a further methoxy group in the meta position (compound 2). Piperazine substitutions of the phenyl

ring were also tested following the success of this moiety in LDN-193189. A piperazine substituent at the para position (compound 3) was found to more favourable than at the meta position (compound 4). In fact, compound 3 showed the highest potency of any of the R1 compound series, whilst also maintaining reasonable selectivity. Finally, a sulphonamide at the meta position (compound 5) was also well tolerated, but marginally less potent. From these comparisons compound 3 was used as a fixed R1 group for further chemical derivatives.

A



B

	Tm shift (°C)							
	Type I		Difference	Type I		Type II		
Compound	ACVR1	TGFBR1		ACVRL1	BMPR1B	ACVR2B	BMPR2	TGFBR2
K02288	13.1	11.8	1.3	14.2	13.8	12.2	3.5	8.1
1	13.5	13.9	- 0.4	14.8	14.9	12.8	4.0	9.1
2	13.9	12.2	1.7	15.3	15.7	12.3	4.7	9.7
3	15.1	13.9	1.2	16.2	16.3	12.0	1.5	9.7
4	14.5	13.7	0.8	15.0	15.9	12.2	4.6	8.2
5	14.4	13.4	1.0	16.0	16.7	10.9	4.2	8.7

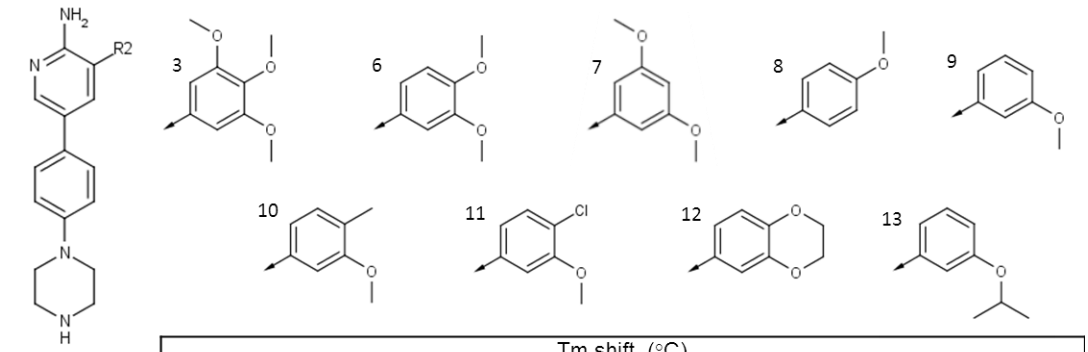
Figure 6.1 SAR of the solvent-facing R1 group of K02288. (A) The K02288 scaffold R1 position and groups (B) The Tm shifts of the corresponding compounds against ACVR1 and TGFBR1, and the difference between the two shifts to highlight variations in selectivity. Comparable Tm shifts are presented for other Type I and Type II receptors.

6.2.1.2 Potency and selectivity of R2 modifications

Using compound 3 as the base scaffold, derivatives exploring changes to the trimethoxyphenyl group (R2) were examined (figure 6.2). The removal of at least one methoxy group did increase the specificity. However, all three methoxy groups were required for high potency, in agreement with the earlier SAR of the imidazo[1,2-b]pyridazine series (chapter 5.2.1). For example, compounds 6 to 9 which lose either one or two of the methoxy groups showed Tm shifts reduced by between 1.2 and 6.3°C. The location of the methoxy group was also particularly important for potency. In the compounds containing a single methoxy group, the meta substitution (compound 9) produced a significantly higher Tm shift than the para substitution (compound 8). Similarly, in the dimethoxy compounds, compound 7 containing two meta substituents produced higher Tm shifts than compound 6, which contains both a meta and a para substitution. Potency of the compounds could be explained by the structure of the ACVR1-K02288 complex as the increasing number of methoxy groups fills the available space in the pocket. However, the para position of the methoxy group appears to provide some improved selectivity for ACVR1 over TGFBR1. The addition of alternative methyl or chlorine groups in the para position with the meta methoxy (compounds 10 and 11) did not increase specificity or potency when compared to compound 9. Other extensions to the methoxy group led to a severe reduction in potency (figure 6.2). The reduced potency of compounds 12 and 13 can be explained by inability of the group to form the water-mediated hydrogen bond to the αC that has been observed in all ACVR1-inhibitor complex structures to date.

There appears to be a trade-off between potency and selectivity in the R2 position. Here compound 7 provided a good compromise of the criteria, and was greater in potency and selectivity than K02288.

A



B

	Tm shift (°C)							
	Type I		Difference	Type I		Type II		
Compound	ACVR1	TGFBR1		ACVRL1	BMPR1B	ACVR2B	BMPR2	TGFBR2
3	15.1	13.9	1.2	16.2	16.3	12.0	1.5	9.7
6	11.5	7.2	4.3	12.6	11.2	9.9	0.1	5.8
7	13.9	10.4	3.5	14.9	13.7	13.4	0.2	7.0
8	8.8	4.5	4.3	9.5	7.2	5.9	0.3	3.8
9	12.1	8.3	3.8	12.5	11.2	11.4	0.3	10.3
10	11.2	7.3	3.9	10.7	10.8	9.9	0.0	5.7
11	11.9	9.2	2.7	11.2	12.5	11.9	0.4	3.7
12	9.5	5.5	4.0	10.3	8.2	5.0	0.2	7.0
13	8.5	7.6	0.9	8.4	6.0	6.3	0.2	5.7

Figure 6.2 SAR of the R2 group. (A) The inhibitor scaffold and various R2 groups (B) The Tm shifts of the corresponding compounds against ACVR1 and TGFBR1, and the difference between the two shifts to highlight variations in specificity. Also shown are the Tm shifts of other members of the BMP/TGFβ receptor family.

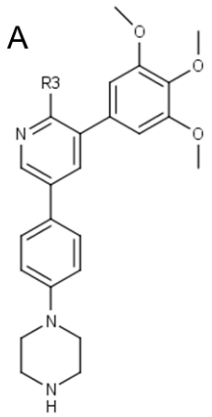
6.2.1.3 Potency and selectivity of R3 modifications

Substitutions of the amine at the R3 position (figure 6.3) were also explored. Their analysis indicated that bulkier R3 groups (compounds 17-19) were not tolerated as judged by Tm shifts of just 0.3-1.1°C (figure 6.3). This could be explained by predicted steric hindrance from the crystal structure of the ACVR1-K02288 complex. The larger methyl and ethyl substituents would preferentially orient away from the adjacent

trimethoxyphenol group. This would disfavour any hydrogen bonding to the scaffold pyridine nitrogen from the hinge region as the compound orientates away to prevent a clash with the gatekeeper residue T283.

Although slightly reduced potency was observed due to the absence of the hydrogen bond from the primary amine nitrogen to the main chain amide of H284 in the hinge region of ACVR1, interestingly, deletion of the amine (compound 14) significantly increased selectivity for ACVR1 over TGFBR1 (which has a smaller serine gatekeeper residue). Also, both methyl (compound 15) and chlorine (compound 16) substitutions were well tolerated with a similar level of selectivity as compound 14 but with reduced potency.

A



B

		Tm shift (°C)							
		Type I		Difference	Type I		Type II		
Compound	R3 group	ACVR1	TGFBR1		ACVRL1	BMPR1B	ACVR2B	BMPR2	TGFBR2
3	NH_2	15.1	13.9	1.2	16.2	16.3	12.0	1.5	9.7
14	H	14.1	10.4	3.7	15.2	13.6	10.8	0.3	7.4
15	CH_3	13.7	9.7	4.0	14.9	12.6	11.3	0.5	7.5
16	Cl	12.8	9.0	3.8	14.0	11.0	9.8	0.0	6.0
17	OCH_3	1.1	1.3	- 0.2	1.1	0.2	-0.2	-0.1	0.2
18	$NHCH_2CH_3$	0.6	0.5	0.1	0.6	0.4	0.0	0.1	-0.2
19	$NHCH_3$	0.3	0.9	- 0.6	0.7	0.1	0.5	03	-0.1

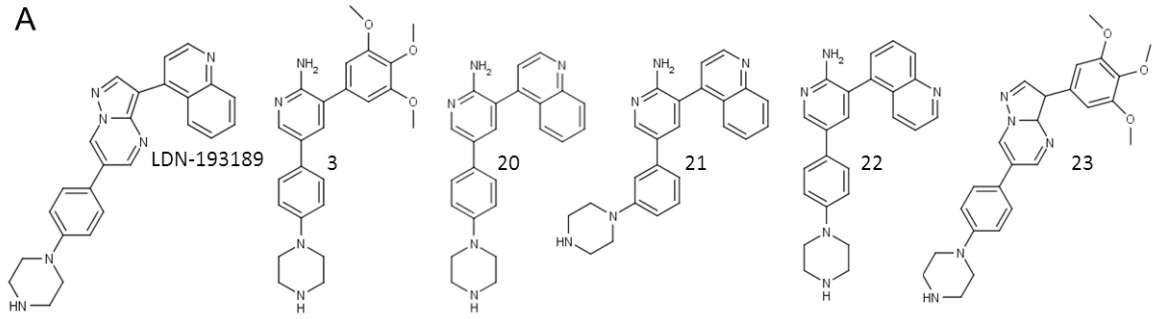
Figure 6.3 SAR at the R3 position. (A) The inhibitor scaffold and R3 position (B) The various R3 groups and the Tm shifts of the corresponding compounds against different BMP/TGFβ receptors. The difference between the Tm shifts of ACVR1 and TGFBR1 is shown to highlight variations in selectivity.

6.2.1.4 Potency and selectivity of other hybrid derivatives

A number of compounds were synthesized to explore multiple substitutions including hybrids between K02288 and LDN-193189. Two compounds were made using a quinoline group at the R2 position. Surprisingly, a 4-quinoline (compound 20), as present in LDN-193189, resulted in a 5°C reduction in Tm shift (figure 6.4). However, movement of the nitrogen atom from the 4-quinoline to the 5-quinoline position (compound 22) did increase the specificity against TGFBR1 (figure 6.4B).

Compared to compounds 20-22, the trimethoxyphenyl group was well tolerated in the LDN-193189 scaffold as an alternative to the 4-quinoline (compound 23). This compound showed high potency similar to that of LDN-193189 with a slightly increased level of specificity. However, specificity was still greater in other compounds including compound 14 (figure 6.3).

A



B

	Tm shift (°C)							
	Type I		Difference	Type I		Type II		
Compound	ACVR1	TGFBR1		ACVRL1	BMPR1B	ACVR2B	BMPR2	TGFBR2
LDN-193189	14.3	12.0	2.3	15.7	17.1	14.3	0.0	15.2
3	15.1	13.9	1.2	16.2	16.3	12.0	1.5	9.7
20	10.4	6.9	3.5	11.1	9.0	11.3	0.3	5.4
21	9.9	7.5	2.4	11.3	9.4	11.9	0.4	3.7
22	9.5	5.3	4.2	10.6	7.8	8.1	0.1	3.1
23	14.2	11.6	2.6	14.6	15.9	12.0	0.7	11.5

Figure 6.4 Comparison of hybrids of K02288 and LDN-193189. (A) Scaffolds of the hybrid compounds. (B) The Tm shifts of the compounds against ACVR1 and TGFBR1, and the difference between the two shifts to highlight variations in specificity. Also shown are the Tm shifts of other members of the BMP/TGFβ receptor family.

6.2.1.5 Binding of FOP mutants

The ultimate aim is to identify a suitable compound for use as a new drug entity to treat patients with FOP. Therefore, the compound must bind to mutant ACVR1 as well as the wild type. To test this, the K02288 derivatives were screened by DSF against the FOP mutants generated in chapter 4. Importantly, this showed that all of the compounds bound to the ACVR1 mutants similarly to the wild type (figure 6.5).

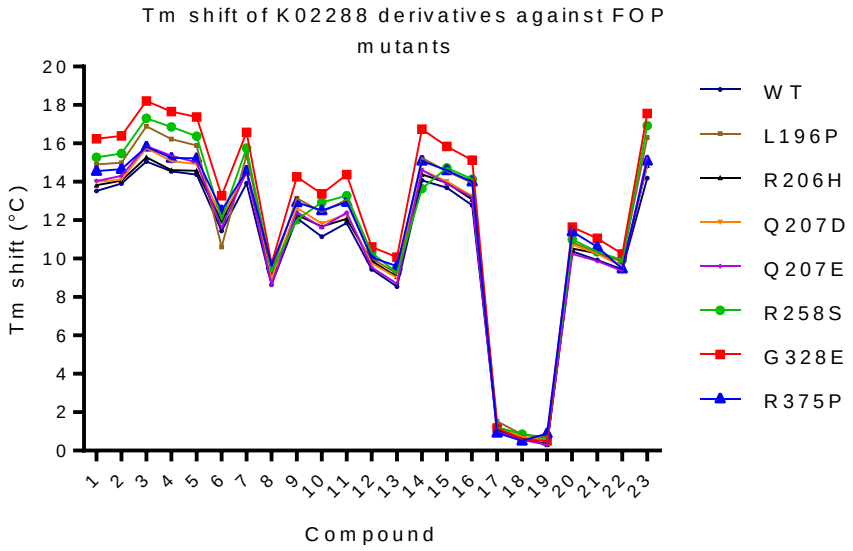


Figure 6.5 Plot of Tm shifts for the different K02288 derivatives tested against wild-type and mutant ACVR1. All the mutants displayed an equal or greater Tm shift against the derivative compounds.

6.2.2 Structural analysis of the binding mode of LDN-213844

It was previously noted that K02288 and LDN-193189 bind in a slightly different orientation to the hinge of ACVR1 (chapter 5.2.5.3). In order to determine the mode of binding of the hybrid compounds, a number of the compounds were set up for co-crystallisation (Table 6.2) with the ACVR1 Q207D construct (a.a. 201-499), as previously used for co-crystallisation of K02288 and LDN-193189. Microcrystals were obtained

from the initial coarse screens for each of the compounds tested in co-crystallisation. These precipitant conditions were used to select appropriate fine screens, based on the chemistry surrounding the condition.

Table 6.1 Conditions and compounds used in co-crystallisation trials with ACVR1.

Compound	Screen	Protein conc. (mg/ml)	Temperature (°C)
3	LFS	10	4
	JCSG	10	20
	JCSG-H11	7.5	20
	LFS-G08	7.5	4
	LFS-G08	8.5	4
	JCSG-F09	8.5	20
7	LFS	10	20
	JCSG	10	4
	JCSG-A03	7.5	4
	JCSG-F09	7.5	4
	JCSG-A03	8.5	4
	JCSG-F09	8.5	4
14	LFS	10	20
	JCSG	10	4
	JCSG-A05	7.5	4
	JCSG-F09	7.5	4

The suffix refers to the screening plate well containing the original hit.

Compound 14 was successfully co-crystallised with ACVR1 Q207D. Diffraction quality crystals were obtained at 4°C in the JCSG coarse screen using a reservoir condition containing 0.2 M ammonium citrate and 20% PEG 3350 (figure 6.6). A dataset was collected at the Diamond Light Source beamline I04-1 by staff at the SGC. The other compounds were not taken any further.

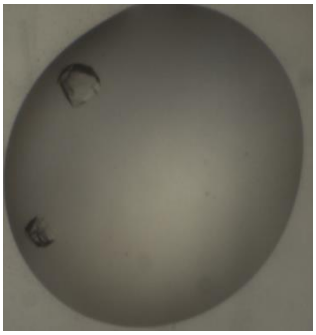


Figure 6.6 Crystals of ACVR1 Q207D and Compound 14. Crystals were grown at 4°C, in a sitting drop with a mix of 100 nl protein and 50 nl of a reservoir solution containing 0.2 M ammonium citrate and 20% PEG 3350.

I solved the structure by molecular replacement using PHASER [109] and ACVR1 Q207D in complex with LDN-193189 (pdb 3q4u) as a search model. The protein crystallised in space group I121 with four protein complexes in the asymmetric unit. The structure was refined at 2.64 Å using Coot [113] and REFMAC [114] (full details in chapter 2.12.4). The structure was then checked for geometric constraints using MolProbity [117]. Diffraction data and refinement statistics are provided in table 6.2.

Table 6.2 Diffraction data and refinement statistics

PDB accession code	**4BGG**
Data Collection	
Beamline	Diamond I04-1
Wavelength (Å)	0.92
Resolution[a] (Å)	52.91-2.64 (10.23-2.56)
Spacegroup	*1* 1 2 1
Cell dimensions	*a* = 85.4, *b* = 99.9, *c* = 187.4 Å
	α = γ = 90.0°, β = 92.9°
No. unique reflections[a]	49,238 (4,579)
Completeness[a] (%)	97.0 (98.5)
I/σI[a]	10.9 (3.2)
R_{merge}[a] (%)	6 (31)
Redundancy[a]	3.6 (3.7)
Refinement	
Ligands	Compound 14 (LDN-213844)
No. atoms in refinement (P/L/O)[b]	8916/120/242
R_{fact} (%)	21.7
R_{free} (%)	24.7
B_f (P/L/O)[b] ($Å^2$)	63/44/55
rms deviation bond[c] (Å)	0.015
rms deviation angle[c] (°)	1.6
Molprobity	
Ramachandran favour	97.6%
Ramachandran allowed	99.7%

a Values in brackets show the statistics for the highest resolution shells.

b P/L/O indicate protein, ligand molecules presented in the active sites, and other (water and solvent molecules), respectively.

c rms indicates root-mean-square.

The structure showed compound 14 was bound in an ATP-mimetic fashion to the hinge (figure 6.7), as observed in the previous co-structures of K02288 and LDN-193189. Here, the pyridine nitrogen formed a conserved hydrogen bond with H286 in the hinge (figure 6.7B). The water-mediated hydrogen bond to the catalytic lysine, K235, observed in the co-structure of K02288 was conserved. Interestingly, the second water-mediated bond to N341 and D354 was also conserved (figure 6.7B), but only in two of the four molecules of ACVR1 in the asymmetric unit, which indicated that this is not a requirement of binding.

The observed electron density for compound 14 provided an accurate assignment of its binding mode (figure 6.7C). Structural superposition showed that compound 14 adopted a similar hinge binding orientation as K02288, rather than LDN-193189 (figure 6.7D). This suggests that the 2-aminopyridine scaffold, rather than the piperazine, determined the precise hinge binding orientation (figure 6.7D).

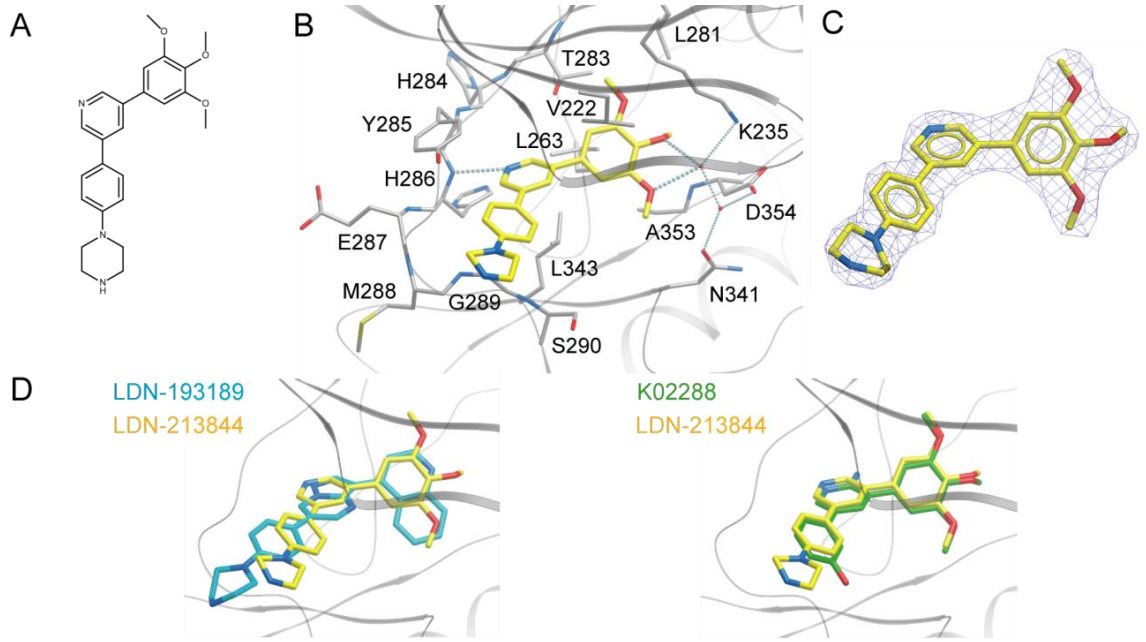
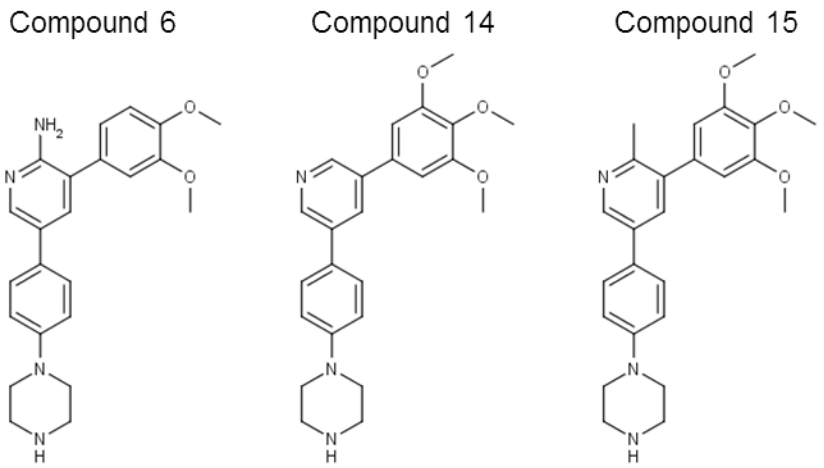


Figure 6.7 Binding mode of compound 14 (A) Chemical scaffold of compound 14 (B) Binding pocket of ACVR1 showing the hydrogen bond to the hinge and the water-mediated bonds to the catalytic lysine (K235), and N341 and D354. (pdb: 4bgg) (C) Electron density (2Fc-2Fo map) for compound 14 is shown as a blue mesh contoured at 2σ (D) Overlay of the binding of compound 14 with LDN-193189 (pdb: 3q4u) and K02288 (pdb: 3mtf) shows that compound 14 follows K02288 binding.

6.4 Discussion

K02288 had been identified as a promising new inhibitor scaffold for optimisation against ACVR1. The SAR of K02288 derivatives provided an insight into the important features for binding to ACVR1. The piperazine group, used also in LDN-193189, was identified as a suitable replacement for the potentially vulnerable phenol group of K02288, which was

solvent exposed in the ACVR1 co-structure. The importance of the trimethoxy group in the R2 position was highlighted for high potency, although this did reduce apparent selectivity. Large groups were unsuitable at the amine position and replacement with hydrogen actually increased selectivity. Several new compounds of interest were identified, for example compounds 6, 14 and 15 (figure 6.8). The mixed dimethoxy group of compound 6 significantly increased selectivity from the trimethoxy (compound 3), shown by a Tm shift difference between ACVR1 and TGFBR1 of 4.5°C as compared to 1.2°C. At the R3 position, the substitution of a methyl group (compound 15) for the amine increased selectivity similarly to hydrogen (compound 14), whilst also retaining good potency (figure 6.8). From these analyses, compound 14 (later designated LDN-213844) was selected as a preferred compound due to its higher potency for ACVR1.



	ACVR1	TGFBR1	ACVR1	TGFBR1	ACVR1	TGFBR1
Tm shift (°C)	11.5	7.2	13.7	9.7	14.1	10.4

	BMP6	TGFβ	BMP6	TGFβ	BMP6	TGFβ
IC_{50} (nM)*	143	13.1 µM	26	1.3 µM	98	16 µM

Figure 6.8 Summary of data for compounds 6, 14 and 15. * IC_{50} data provided by the Yu group, Harvard, testing cellular inhibition of BMP6 or TGFβ signalling by luciferase reporter assay in C2C12 cells.

The same compounds were tested by the Harvard group of Dr Paul Yu using luciferase-based reporter assays to gain IC_{50} values in C2C12 cells for BMP6 or TGF-β signalling. The selected compounds 6, 14 and 15 gave IC_{50} values for BMP6 signalling of 143 nM, 26 nM and 98 nM, respectively (figure 6.8). This is improved on the IC_{50} of 421 nM they determined for K02288 in the same assay for BMP6.

Comparison of the two methods across a wider selection of compounds showed a good correlation, further confirming the SAR and the efficacy of the compounds (figure 6.9). Scatter plots for both ACVR1 and TGFBR1 Tm shifts against the equivalent pIC_{50} [144] values were produced (figure 6.9). The equation derived from the plot can provide

estimates for IC_{50} values for future Tm shift data. However, compounds with poor cell permeability, stability or solubility might deviate from the prediction.

A

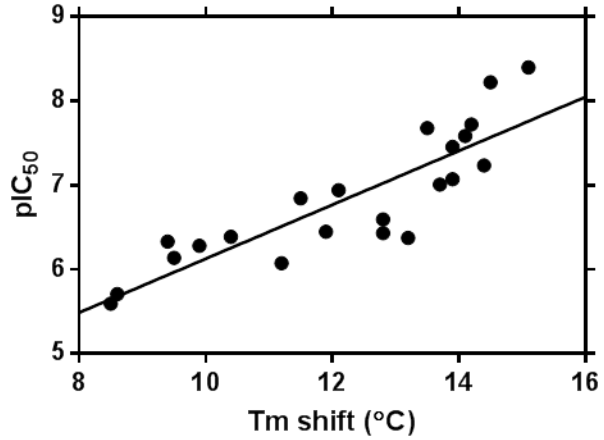


Table of predicted values	
Tm shift (°C)	IC_{50} (nM)
16	9
14	39
12	171
10	744
8	3252

B

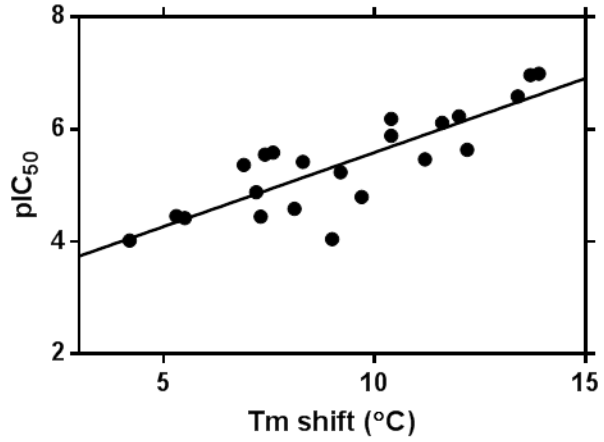


Table of predicted values	
Tm shift (°C)	IC_{50} (nM)
16	67
14	226
12	761
10	2564
8	8646

Figure 6.9 Correlation plots of the cellular pIC50 values against Tm shift values. (A) Graph of pIC_{50} for BMP 6 against Tm shifts for ACVR1. A clear correlation between the two was observed ($R^2 = 0.75$). The linear fit gave the equation $pIC_{50} = 0.3201 \times \Delta Tm + 2.927$ from which a series of estimates for IC_{50} were calculated. (B) Graph of pIC_{50} for TGFβ against Tm shifts for TGFBR1. The correlation was not as strong compared to ACVR1, with $R^2 = 0.70$. The linear fit gave the equation $pIC_{50} = 0.2639 \times \Delta Tm + 2.952$ which was used to provide equivalent IC_{50} predictions.

The structure of compound 14 bound to ACVR1 gives some insight explaining the results seen in the SAR. For example, it is reasonable to expect the single methoxy group in the R1 position (compound 9) to be able to rotate in the pocket, effectively exploring both

meta positions. This was observed also with the binding of K00507 in chapter 5.2.5.1 where the phenyl-ethanone group bound in these two orientations. Exchange of the quinoline for the trimethoxyphenyl group in the 2-aminopyridine scaffold was not particularly well tolerated (compound 20). In comparison the reverse exchange, a trimethoxyphenyl group on the LDN-193189 scaffold, was tolerated (compound 23). This could be explained by the slightly different binding positions of the two core scaffolds (figure 6.10). Compound 20 would have less hydrophobic interaction with the back pocket than either K02288 or LDN-193189 (figure 6.10), whereas the trimethoxyphenyl provides greater flexibility in terms of both hydrogen bonding and hydrophobic interaction.

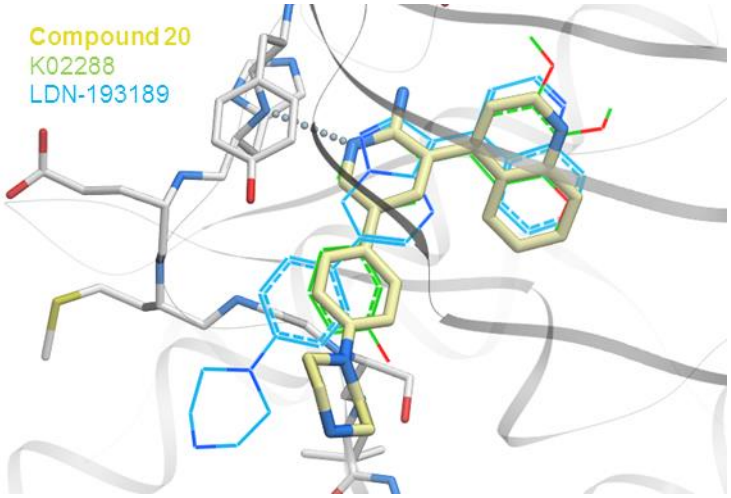
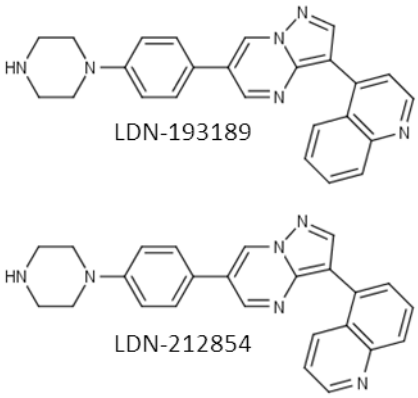


Figure 6.10 Model of compound 20 in complex with ACVR1. The model was built by Dr Brian Marsden at the SGC using the ICM software package for molecular docking. The compound is predicted to bind similarly to K02288 (green).

The movement of the nitrogen atom from a 4-quinoline group to a 5-quinoline on the 2-aminopyridine scaffold (compounds 20 and 22, respectively) provided a notable increase in ACVR1 selectivity. A similar increase in specificity was observed for the same 5-quinoline substitution in a derivative of LDN-193189 developed by the Yu lab (figure 6.11). This derivative, LDN-212854, also inhibited ectopic bone formation in FOP mouse models [145].



	Tm shift (°C)		
Compound	ACVR1	TGFBR1	Difference
LDN-193189	14.2	12.2	2.0
LDN-212854	13.6	9.5	4.1

Figure 6.11 Structure and Tm shift data for LDN-212854. Substitution of the 4-quinoline to the 5-quinoline group in the Yu lab compounds LDN-193189 and LDN-213854 showed a marked increase in specificity towards ACVR1 over TGFBR1, similar to that observed between compounds 20 and 22.

Overall, both the specificity and potency were improved upon from the original hit compound K02288. In particular, the solvent exposed phenol group was replaced with a piperazine group resulting in increased potency. However, this SAR only investigates a limited number of substitutions, so each substituent position could be explored further, an example being the simple deletion of the K02288 phenol hydroxyl group.

7. Discussion

The aims of this thesis were to better understand the activation of ACVR1 in the disease fibrodysplasia ossificians progressiva (FOP), and to identify new chemical inhibitors with potential for therapeutic development. The first was addressed by the generation of FOP-associated mutants of ACVR1 which were used in structural and biophysical experiments. The mutants of ACVR1 were predicted to destabilize the inactive conformation of the kinase by breaking bonds which would either disrupt binding of the inhibitory protein FKBP12, or disrupt the inactive conformation of the ATP pocket, depending on the site of mutation.

Comparison of a panel of mutant structures with the wild-type ACVR1 showed that the inactive kinase conformation was surprisingly persistent. However, some structural changes were identified to indicate a subtle shift towards an active conformation. The activation of the GS and kinase domains could be considered separately, with the activation of both domains necessary for full activation of ACVR1. At the GS domain, the Q207E structure contained a movement of the GS loop that released it from its buried position within the kinase domain. Looking at the kinase domain, some of the Q207D structures displayed conformational changes in the Smad recognition L45 loop and the P-loop. All of the structures obtained were in the absence of the inhibitory protein FKBP12. Whilst lack of FKBP12 binding did not result in the relaxation of the overall inactive structure, it is apparent that its absence impacts on the exposure of the GS loop. The gel filtration data showed that FKBP12 does not bind to the mutant receptors as efficiently as the wild type, so in FOP patients the proportion of uninhibited ACVR1 would be greater,

increasing the likelihood of leaky signalling. This mild activation could be exacerbated by inflammatory signs resulting in the episodic nature of the disease.

Despite being BMP-independent, the FOP-associated mutants appeared to be dependent on the presence of a type II receptor for activation and Smad phosphorylation. However it remains unclear whether the type II receptor is still necessary for GS loop phosphorylation or is simply required to provide a scaffolding function [126,127]. It was recently demonstrated in HEK293T cells that a type II receptor with a kinase-dead mutation (BMPR2 K230R) was sufficient for Smad1 signalling by ACVR1 FOP mutants, but not by wild-type ACVR1 [126]. This suggests that perhaps the type II receptor provides an essential scaffolding function to induce an active kinase conformation or for Smad to bind, but is potentially not necessary for the phosphorylation of mutant ACVR1. It would be interesting to explore this hypothesis through *in vitro* phosphorylation assays.

The activation of ACVR1 has interesting parallels with oncogenic kinases, such as the receptor tyrosine kinase c-Kit. Firstly, the autoinhibition of c-Kit is regulated by its juxtamembrane region. This binds between the N and C-lobes of the kinase domain, preventing the kinase A-loop from forming its active conformation, which thereby inhibits the binding of both ATP and substrate [146]. This mechanism is structurally distinct, but analogous to the burial of the GS loop in ACVR1. Secondly, c-Kit requires ligand-induced homodimerisation for activating transautophosphorylation of Y568 and Y570 in the juxtamembrane region [124,125]. Likewise, heterodimerisation of ACVR1 facilitates activation via transphosphorylation of the juxtamembrane GS loop. Finally, in both kinases, disease mutations in either the juxtamembrane or kinase regions break bonds required for the inhibitory conformation to promote unwanted activation (figure 7.1).

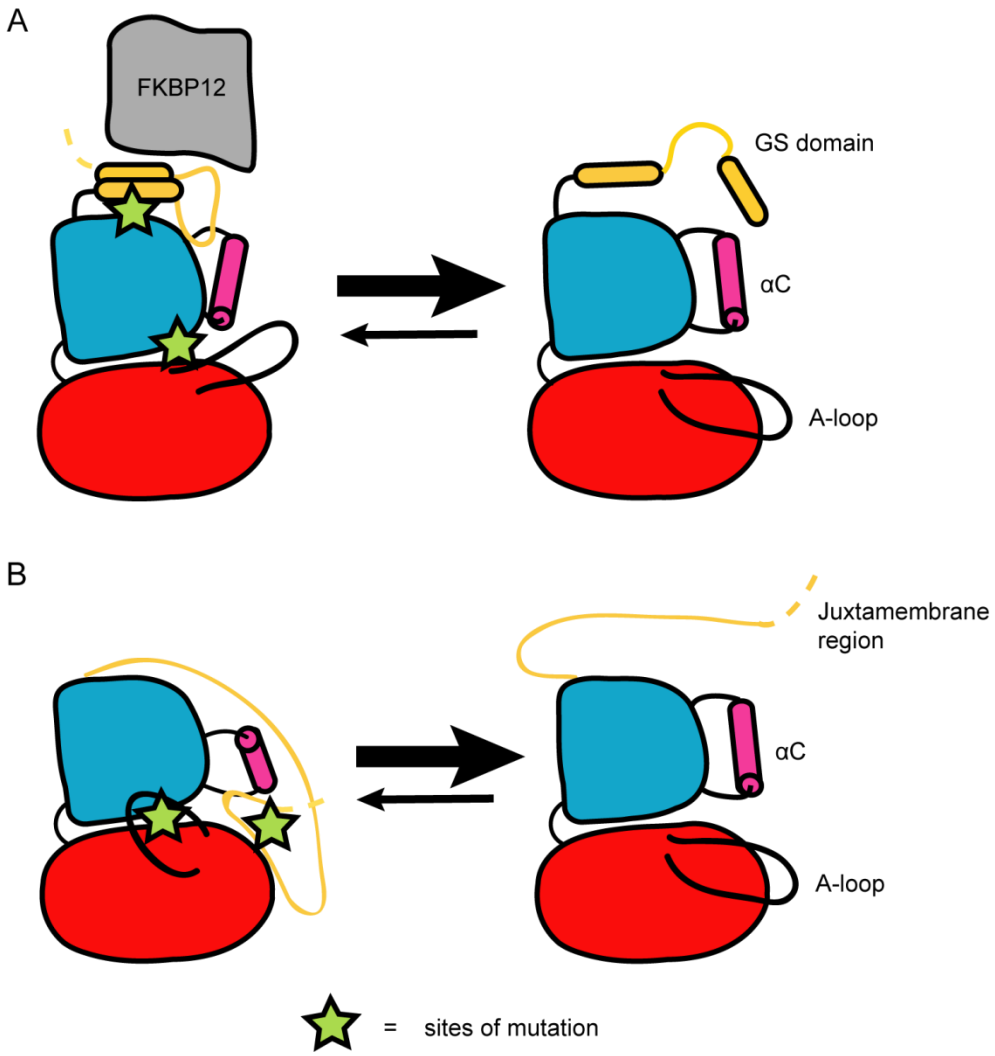


Figure 7.1 Cartoon of mutant activation (A) FOP mutations occur in sites of receptor regulation, the GS domain and around the ATP pocket of the kinase domain. Disruption to either of these regions can cause destabilisation of the inactivate structure. This results in an increase in dynamics of the receptor conformation. (B) Mutations in cKIT can also occur in the regulatory juxtamembrane region and the kinase domain. Whilst the site of mutation results in different disease phenotypes, both mutation sites would disrupt the inhibitory binding conformation, resulting in an increased proportion of the receptor in the active conformation.

Another interesting observation was the difference in conformational dynamics between BMP and TGFβ receptors. Overall, the mode of activation of both receptor classes is thought to be the same. However, there were suggestions throughout this thesis of TGFBR1 being more susceptible to activation than ACVR1. The first structure of TGFBR1 in the absence of FKBP12 showed more significant structural changes towards

the active conformation than the six structures of mutant ACVR1 presented here. On co-expression with the TGFβ type II receptor, TGFBR1 showed multiple phosphorylations within the GS domain, an essential site for TGFBR1 activation and the subsequent phosphorylation of Smad2/3. However, after similar co-expression of the BMP type II receptor, there was comparably little phosphorylation of the type I receptor (chapter 4.2.4.2). Finally, the inhibitor screening of the receptors provided further evidence of variations in the dynamics of the two receptor classes. Deletion of the GS domain in ACVR1 had no effect on the affinity of the inhibitor binding, consistent with the similar conformations of the longer and shorter forms of the receptor. On the contrary, the absence of this regulatory domain in TGFBR1 resulted in a 1000-fold reduction in the dissociation constant (chapter 5.2.4.2). Binding could be diminished by an increase in the dynamics of the kinase domain, or simply by a shift towards the active conformation of TGFBR1.

FOP is a seriously debilitating disease with no cure. This thesis has attempted to address this need with the identification of the novel compound, K02288 which showed potent binding to ACVR1 whilst being highly selective against other kinases. This provided an initial small scaffold allowing for future drug design. K02288 contained a 2-aminopyridine core scaffold that is also present in the ALK tyrosine kinase inhibitor, crizotinib, which is approved by the US Food and Drug Administration (FDA) for the treatment of non-small cell lung cancer [140] as well as in preclinical inhibitors of CHK2 [136], NEK2 [139] (figure 7.2). 2-aminopyridine is an ATP-mimetic group that has generally been considered a promiscuous scaffold. However, it appears that some substituent groups can be identified that confer selectivity for certain kinases.

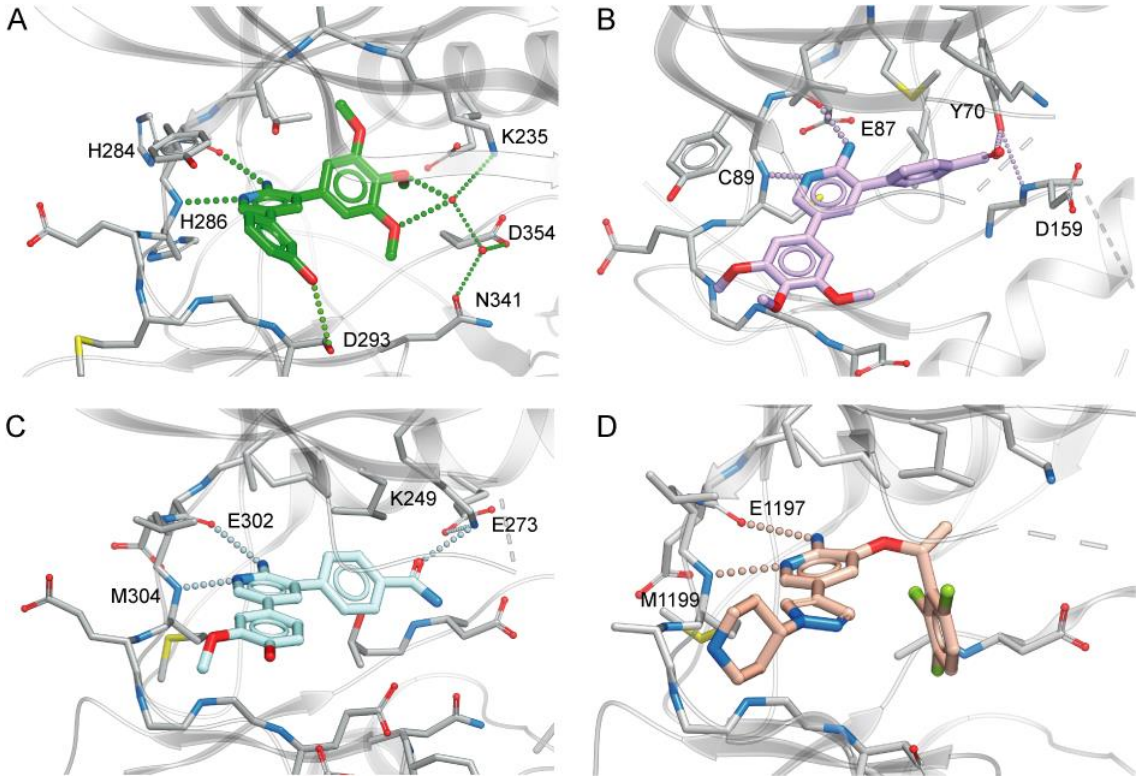


Figure 7.2 Comparison of 2-aminopyridine co-structures. All compounds are bound to the hinge region of the protein. (A) K02288 in complex with ACVR1 Q207D (pdb 3mtf). (B) NEK2 co-structure (pdb 2wqo). (C) CHK2 co-structure (pdb 2wtc). (D) Crizotinib in complex with the ALK tyrosine kinase (pdb 2xp2).

K02288 also showed binding to other BMP receptors, such as ACVRL1. Interestingly, biological inhibitors of ACVRL1 have entered clinical trials for anti-angiogenesis in human cancers [126,127]. It will be of interest to see whether small molecule BMP inhibitors could also be relevant for this indication. Moreover, angiogenesis is important for new bone growth and therefore ACVRL1 inhibition could be additionally relevant for FOP treatment.

In summary, the effects observed relating to FKBP12 binding and mutant structure were consistent with exaggerated leaky signalling by ACVR1. Ultimately, a receptor-Smad co-structure would be required to fully understand the role of the GS domain in receptor signalling. The example of crizotinib provides hope for K02288-like molecules inhibiting ACVR1 in FOP. Importantly, the FOP-associated mutants of ACVR1 also showed high

affinity for the compounds. Finally, the initial structural activity relationship study made early progress towards the development of molecules suited for *in vivo* proof of principle testing.

References

[1] F. S. Kaplan, M. Xu, P. Seemann, J. M. Connor, D. L. Glaser, L. Carroll, P. Delai, E. Fastnacht-Urban, S. J. Forman, G. Gillessen-Kaesbach, J. Hoover-Fong, B. Köster, R. M. Pauli, W. Reardon, S.-A. Zaidi, M. Zasloff, R. Morhart, S. Mundlos, J. Groppe, and E. M. Shore, "Classic and atypical fibrodysplasia ossificans progressiva (FOP) phenotypes are caused by mutations in the bone morphogenetic protein (BMP) type I receptor ACVR1.," *Hum. Mutat.*, vol. 30, no. 3, pp. 379–90, Mar. 2009.

[2] F. S. Kaplan, D. L. Glaser, R. J. Pignolo, and E. M. Shore, "A new era for fibrodysplasia ossificans progressiva: a druggable target for the second skeleton.," *Expert Opin. Biol. Ther.*, vol. 7, no. 5, pp. 705–12, May 2007.

[3] F. S. Kaplan, D. L. Glaser, E. M. Shore, J. M. Connor, and J. A. Kitterman, "The Phenotype of Fibrodysplasia Ossificans Progressiva," *Clin. Rev. Bone Miner. Metab.*, vol. 3, no. 3, pp. 183–188, 2005.

[4] E. M. Shore and F. S. Kaplan, "Insights from a rare genetic disorder of extra-skeletal bone formation, fibrodysplasia ossificans progressiva (FOP).," *Bone*, vol. 43, no. 3, pp. 427–33, Sep. 2008.

[5] F. S. Kaplan, R. J. Pignolo, and E. M. Shore, "The FOP metamorphogene encodes a novel type I receptor that dysregulates BMP signaling.," *Cytokine Growth Factor Rev.*, vol. 20, no. 5–6, pp. 399–407, 2009.

[6] P. Billings and J. Fiori, "Dysregulated BMP signaling and enhanced osteogenic differentiation of connective tissue progenitor cells from patients with fibrodysplasia ossificans progressiva ," *J. Bone Miner. Res.*, vol. 23, no. 3, pp. 305–313, 2007.

[7] F. S. Kaplan, S. A. Chakkalakal, and E. M. Shore, "Fibrodysplasia ossificans progressiva: mechanisms and models of skeletal metamorphosis.," *Dis. Model. Mech.*, vol. 5, no. 6, pp. 756–62, Nov. 2012.

[8] R. J. Pignolo, E. M. Shore, and F. S. Kaplan, "Fibrodysplasia ossificans progressiva: clinical and genetic aspects.," *Orphanet J. Rare Dis.*, vol. 6, no. 1, p. 80, Jan. 2011.

[9] Y. Nakashima, N. Haga, H. Kitoh, J. Kamizono, K. Tozawa, T. Katagiri, T. Susami, J. Fukushi, and Y. Iwamoto, "Deformity of the great toe in fibrodysplasia ossificans progressiva.," *J. Orthop. Sci.*, vol. 15, no. 6, pp. 804–9, Nov. 2010.

[10] E. M. Shore, M. Xu, G. J. Feldman, D. a Fenstermacher, T.-J. Cho, I. H. Choi, J. M. Connor, P. Delai, D. L. Glaser, M. LeMerrer, R. Morhart, J. G. Rogers, R. Smith, J. T. Triffitt, J. A. Urtizberea, M. Zasloff, M. a Brown, and F. S. Kaplan, "A recurrent mutation in the BMP type I receptor ACVR1 causes inherited and sporadic fibrodysplasia ossificans progressiva.," *Nat. Genet.*, vol. 38, no. 5, pp. 525–7, May 2006.

[11] D. Chen, M. Zhao, and G. R. Mundy, “Bone morphogenetic proteins.,” *Growth Factors*, vol. 22, no. 4, pp. 233–41, Dec. 2004.

[12] X. Cao and D. Chen, “The BMP signaling and in vivo bone formation.,” *Gene*, vol. 357, no. 1, pp. 1–8, Aug. 2005.

[13] M. Huse and J. Kuriyan, “The conformational plasticity of protein kinases.,” *Cell*, vol. 109, no. 3, pp. 275–82, May 2002.

[14] J. Massagué, “TGF-beta signal transduction.,” *Annu. Rev. Biochem.*, vol. 67, pp. 753–91, Jan. 1998.

[15] C. L. Gregson, P. Hollingworth, M. Williams, K. A. Petrie, A. N. Bullock, M. A. Brown, J. H. Tobias, and J. T. Triffitt, “A novel ACVR1 mutation in the glycine/serine-rich domain found in the most benign case of a fibrodysplasia ossificans progressiva variant reported to date.,” *Bone*, vol. 48, no. 3, pp. 654–8, Mar. 2011.

[16] R. Bocciardi, D. Bordo, M. Di Duca, M. Di Rocco, and R. Ravazzolo, “Mutational analysis of the ACVR1 gene in Italian patients affected with fibrodysplasia ossificans progressiva: confirmations and advancements.,” *Eur. J. Hum. Genet.*, vol. 17, no. 3, pp. 311–8, Mar. 2009.

[17] K. A. Petrie, W. H. Lee, A. N. Bullock, J. J. Pointon, R. Smith, R. G. G. Russell, M. a Brown, B. P. Wordsworth, and J. T. Triffitt, “Novel mutations in ACVR1 result in atypical features in two fibrodysplasia ossificans progressiva patients.,” *PLoS One*, vol. 4, no. 3, p. e5005, Jan. 2009.

[18] M. P. Whyte, D. Wenkert, J. L. Demertzis, E. F. DiCarlo, E. Westenberg, and S. Mumm, “Fibrodysplasia ossificans progressiva: middle-age onset of heterotopic ossification from a unique missense mutation (c.974G>C, p.G325A) in ACVR1.,” *J. Bone Miner. Res.*, vol. 27, no. 3, pp. 729–37, Mar. 2012.

[19] L. N. Johnson and R. J. Lewis, “Structural basis for control by phosphorylation.,” *Chem. Rev.*, vol. 101, no. 8, pp. 2209–42, Aug. 2001.

[20] G. Manning, D. B. Whyte, R. Martinez, T. Hunter, and S. Sudarsanam, “The protein kinase complement of the human genome.,” *Science*, vol. 298, no. 5600, pp. 1912–34, Dec. 2002.

[21] Y. Shi and J. Massagué, “Mechanisms of TGF-β Signaling from Cell Membrane to the Nucleus,” *Cell*, vol. 113, no. 6, pp. 685–700, Jun. 2003.

[22] P. ten Dijke and C. S. Hill, “New insights into TGF-beta-Smad signalling.,” *Trends Biochem. Sci.*, vol. 29, no. 5, pp. 265–73, May 2004.

[23] J. Wrana, L. Attisano, R. Wieser, F. Ventura, and J. Massague, “Mechanism of activation of the TGF-b receptor,” *Nature*, vol. 370, pp. 341-347, 1994.

[24] M. Y. Wu and C. S. Hill, “Tgf-beta superfamily signaling in embryonic development and homeostasis.,” *Dev. Cell*, vol. 16, no. 3, pp. 329–43, Mar. 2009.

[25] R. Derynck and Y. E. Zhang, “Smad-dependent and Smad-independent pathways in TGF-beta family signalling.,” *Nature*, vol. 425, no. 6958, pp. 577–84, Oct. 2003.

[26] A. Sorrentino, N. Thakur, S. Grimsby, A. Marcusson, V. von Bulow, N. Schuster, S. Zhang, C.-H. Heldin, and M. Landström, “The type I TGF-beta receptor engages TRAF6 to activate TAK1 in a receptor kinase-independent manner.,” *Nat. Cell Biol.*, vol. 10, no. 10, pp. 1199–207, Oct. 2008.

[27] J.-H. Shim, M. B. Greenblatt, M. Xie, M. D. Schneider, W. Zou, B. Zhai, S. Gygi, and L. H. Glimcher, “TAK1 is an essential regulator of BMP signalling in cartilage.,” *EMBO J.*, vol. 28, no. 14, pp. 2028–41, Jul. 2009.

[28] M. Urist, “Bone: formation by autoinduction.,” *Sci.*,vol. 150, no. 3698, pp. 893-899, 1965.

[29] J. Wozney, V. Rosen, and A. Celeste, “Novel regulators of bone formation: molecular clones and activities.,” *Sci.*, vol. 242, no. 4885, pp. 1528–1534, 1988.

[30] B. Bragdon, O. Moseychuk, S. Saldanha, D. King, J. Julian, and A. Nohe, “Bone morphogenetic proteins: a critical review.,” *Cell. Signal.*, vol. 23, no. 4, pp. 609–20, Apr. 2011.

[31] T. D. Mueller and J. Nickel, “Promiscuity and specificity in BMP receptor activation.,” *FEBS Lett.*, vol. 586, no. 14, pp. 1846–59, Jul. 2012.

[32] Y. Drabsch and P. ten Dijke, “TGF-β signalling and its role in cancer progression and metastasis.,” *Cancer Metastasis Rev.*, vol. 31, no. 3–4, pp. 553–68, Dec. 2012.

[33] A. T. Dudley, K. M. Lyons, and E. J. Robertson, “A requirement for bone morphogenetic protein-7 during development of the mammalian kidney and eye.,” *Genes Dev.*, vol. 9, no. 22, pp. 2795–2807, Nov. 1995.

[34] X.-M. Meng, A. C. K. Chung, and H. Y. Lan, “Role of the TGF-β/BMP-7/Smad pathways in renal diseases.,” *Clin. Sci. (Lond).*, vol. 124, no. 4, pp. 243–54, Feb. 2013.

[35] T. Q. Nguyen and R. Goldschmeding, “Bone morphogenetic protein-7 and connective tissue growth factor: novel targets for treatment of renal fibrosis?,” *Pharm. Res.*, vol. 25, no. 10, pp. 2416–26, Oct. 2008.

[36] D. O. Wagner, C. Sieber, R. Bhushan, J. H. Börgermann, D. Graf, and P. Knaus, “BMPs: from bone to body morphogenetic proteins.,” *Sci. Signal.*, vol. 3, no. 107, p. mr1, Jan. 2010.

[37] J. Massagué, “TGFβ signalling in context.,” *Nat. Rev. Mol. Cell Biol.*, vol. 13, no. 10, pp. 616–30, Oct. 2012.

[38] J. Groppe, C. S. Hinck, P. Samavarchi-Tehrani, C. Zubieta, J. P. Schuermann, A. B. Taylor, P. M. Schwarz, J. L. Wrana, and A. P. Hinck, “Cooperative assembly of TGF-beta superfamily signaling complexes is mediated by two disparate mechanisms and distinct modes of receptor binding.,” *Mol. Cell*, vol. 29, no. 2, pp. 157–68, Feb. 2008.

[39] J. E. Zuniga, U. Ilangovan, P. Mahlawat, C. S. Hinck, T. Huang, J. C. Groppe, D. G. McEwen, and A. P. Hinck, “The TβR-I pre-helix extension is structurally ordered in the unbound form and its flanking prolines are essential for binding.,” *J. Mol. Biol.*, vol. 412, no. 4, pp. 601–18, Sep. 2011.

[40] D. Grcević, R. Kusec, N. Kovacić, A. Lukić, I. K. Lukić, S. Ivcević, D. Nemet, R. S. Seiwerth, S. K. Ostojić, P. I. Croucher, and A. Marusić, “Bone morphogenetic proteins and receptors are over-expressed in bone-marrow cells of multiple myeloma patients and support myeloma cells by inducing ID genes.,” *Leuk. Res.*, vol. 34, no. 6, pp. 742–51, Jun. 2010.

[41] I. C. Joziasse, K. A. Smith, S. Chocron, M. van Dinther, V. Guryev, J. J. van de Smagt, E. Cuppen, P. Ten Dijke, B. J. Mulder, C. L. Maslen, B. Reshey, P. A. Doevendans, and J. Bakkers, “ALK2 mutation in a patient with Down’s syndrome and a congenital heart defect.,” *Eur. J. Hum. Genet.*, vol. 19, no. 4, pp. 389–93, Apr. 2011.

[42] K. Lehmann, P. Seemann, J. Boergermann, G. Morin, S. Reif, P. Knaus, and S. Mundlos, “A novel R486Q mutation in BMPR1B resulting in either a brachydactyly type C/symphalangism-like phenotype or brachydactyly type A2.,” *Eur. J. Hum. Genet.*, vol. 14, no. 12, pp. 1248–54, Dec. 2006.

[43] K. Lehmann, P. Seemann, S. Stricker, M. Sammar, B. Meyer, K. Süring, F. Majewski, S. Tinschert, K.-H. Grzeschik, D. Müller, P. Knaus, P. Nürnberg, and S. Mundlos, “Mutations in bone morphogenetic protein receptor 1B cause brachydactyly type A2.,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 100, no. 21, pp. 12277–82, Oct. 2003.

[44] K. Miyazono, Y. Kamiya, and M. Morikawa, “Bone morphogenetic protein receptors and signal transduction.,” *J. Biochem.*, vol. 147, no. 1, pp. 35–51, Jan. 2010.

[45] E. Pardali, M.-J. Goumans, and P. ten Dijke, “Signaling by members of the TGF-beta family in vascular morphogenesis and disease.,” *Trends Cell Biol.*, vol. 20, no. 9, pp. 556–67, Sep. 2010.

[46] S. A. Townson, E. Martinez-Hackert, C. Greppi, P. Lowden, D. Sako, J. Liu, J. A. Ucran, K. Liharska, K. W. Underwood, J. Seehra, R. Kumar, and A. V Grinberg, “Specificity and structure of a high affinity activin receptor-like kinase 1 (ALK1) signaling complex.,” *J. Biol. Chem.*, vol. 287, no. 33, pp. 27313–25, Aug. 2012.

[47] J. W. Lowery and M. P. de Caestecker, “BMP signaling in vascular development and disease.,” *Cytokine Growth Factor Rev.*, vol. 21, no. 4, pp. 287–98, Aug. 2010.

[48] Q. Shen, S. C. Little, M. Xu, J. Haupt, C. Ast, T. Katagiri, S. Mundlos, P. Seemann, F. S. Kaplan, M. C. Mullins, and E. M. Shore, "The fibrodysplasia ossificans progressiva R206H ACVR1 mutation activates BMP-independent chondrogenesis and zebrafish embryo ventralization.," *J. Clin. Invest.*, vol. 119, no. 11, pp. 3462–3472, 2009.

[49] B. Schmierer and C. S. Hill, "TGFbeta-SMAD signal transduction: molecular specificity and functional flexibility.," *Nat. Rev. Mol. Cell Biol.*, vol. 8, no. 12, pp. 970–82, Dec. 2007.

[50] A. Moustakas, S. Souchelnytskyi, and C. H. Heldin, "Smad regulation in TGF-beta signal transduction.," *J. Cell Sci.*, vol. 114, no. Pt 24, pp. 4359–69, Dec. 2001.

[51] Y. Shi, Y.-F. Wang, L. Jayaraman, H. Yang, J. Massagué, and N. P. Pavletich, "Crystal Structure of a Smad MH1 Domain Bound to DNA," *Cell*, vol. 94, no. 5, pp. 585–594, Sep. 1998.

[52] M. Kretzschmar and J. Massague, "SMADs: Mediators and regulators of TGF-b signaling," *Oncog. cell Prolif.*, pp. 103–111, 1998.

[53] M. Macias-Silva, P. Hoodless, S. Tang, M. Buchwald, and J. Wrana, "Specific activation of Smad1 signaling pathways by the BMP7 type I receptor, ALK2," *J. Biol. Chem.*, vol. 273, no. 40, pp. 25628–25636, 1998.

[54] M. Kretzschmar, F. Liu, A. Hata, J. Doody, and J. Massague, "The TGF-beta family mediator Smad1 is phosphorylated directly and activated functionally by the BMP receptor kinase.," *Genes Dev.*, vol. 11, no. 8, pp. 984–995, Apr. 1997.

[55] N. A. Armes, K. A. Neal, and J. C. Smith, "A short loop on the ALK-2 and ALK-4 activin receptors regulates signaling specificity but cannot account for all their effects on early Xenopus development.," *J. Biol. Chem.*, vol. 274, no. 12, pp. 7929–35, Mar. 1999.

[56] Y. G. Chen, F. Liu, and J. Massague, "Mechanism of TGFbeta receptor inhibition by FKBP12.," *EMBO J.*, vol. 16, no. 13, pp. 3866–76, Jul. 1997.

[57] X. H. Feng and R. Derynck, "A kinase subdomain of transforming growth factor-beta (TGF-beta) type I receptor determines the TGF-beta intracellular signaling specificity.," *EMBO J.*, vol. 16, no. 13, pp. 3912–23, Jul. 1997.

[58] Y. Chen, A. Hata, R. Lo, and D. Wotton, "Determinants of specificity in TGF-beta signal transduction," *Genes Dev.*, vol. 12, pp. 2144 – 2152, 1998.

[59] S. Abdollah, "Tbeta RI Phosphorylation of Smad2 on Ser465 and Ser467 Is Required for Smad2-Smad4 Complex Formation and Signaling," *J. Biol. Chem.*, vol. 272, no. 44, pp. 27678–27685, Oct. 1997.

[60] J. Li, G. I. Lee, S. R. Van Doren, and J. C. Walker, "The FHA domain mediates phosphoprotein interactions.," *J. Cell Sci.*, vol. 113 Pt 23, pp. 4143–9, Dec. 2000.

[61] B. Y. Qin, B. M. Chacko, S. S. Lam, M. P. de Caestecker, J. J. Correia, and K. Lin, “Structural basis of Smad1 activation by receptor kinase phosphorylation.,” *Mol. Cell*, vol. 8, no. 6, pp. 1303–12, Dec. 2001.

[62] M. Kawabata, H. Inoue, A. Hanyu, T. Imamura, and K. Miyazono, “Smad proteins exist as monomers in vivo and undergo homo- and hetero-oligomerization upon activation by serine/threonine kinase receptors.,” *EMBO J.*, vol. 17, no. 14, pp. 4056–65, Jul. 1998.

[63] J. W. Wu, M. Hu, J. Chai, J. Seoane, M. Huse, C. Li, D. J. Rigotti, S. Kyin, T. W. Muir, R. Fairman, J. Massagué, and Y. Shi, “Crystal structure of a phosphorylated Smad2. Recognition of phosphoserine by the MH2 domain and insights on Smad function in TGF-beta signaling.,” *Mol. Cell*, vol. 8, no. 6, pp. 1277–89, Dec. 2001.

[64] B. Chacko, B. Qin, and J. Correia, “The L3 loop and C-terminal phosphorylation jointly define Smad protein trimerization,” *Nat. Struct. Bio.*, vol. 8, no. 3, pp. 248–253, 2001.

[65] B. Y. Qin, B. M. Chacko, S. S. Lam, M. P. de Caestecker, J. J. Correia, and K. Lin, “Structural Basis of Smad1 Activation by Receptor Kinase Phosphorylation,” *Mol. Cell*, vol. 8, no. 6, pp. 1303–1312, Dec. 2001.

[66] B. Schmierer and C. S. Hill, “Kinetic Analysis of Smad Nucleocytoplasmic Shuttling Reveals a Mechanism for Transforming Growth Factor ᴺᴸ -Dependent Nuclear Accumulation of Smads †,” vol. 25, no. 22, pp. 9845–9858, 2005.

[67] H. B. Chen, J. G. Rud, K. Lin, and L. Xu, “Nuclear targeting of transforming growth factor-beta-activated Smad complexes.,” *J. Biol. Chem.*, vol. 280, no. 22, pp. 21329–36, Jul. 2005.

[68] C. S. Hill, “Nucleocytoplasmic shuttling of Smad proteins.,” *Cell Res.*, vol. 19, no. 1, pp. 36–46, Jan. 2009.

[69] X.-H. Feng and R. Derynck, “Specificity and versatility in TGF-beta signaling through Smads.,” *Annu. Rev. Cell Dev. Biol.*, vol. 21, pp. 659–93, Jan. 2005.

[70] G. Wu, Y-G. Chen, B. Ozdamar, C. A. Gyuricza, P. A. Chong, J. L. Wrana, J. Massagué, Y. Shi, “Structural Basis of Smad2 Recognition by the Smad Anchor for Receptor Activation,” *Science*, vol. 287, no. 5450, pp. 92–97, Jan. 2000.

[71] W. Shi, C. Chang, S. Nie, S. Xie, M. Wan, and X. Cao, “Endofin acts as a Smad anchor for receptor activation in BMP signaling.,” *J. Cell Sci.*, vol. 120, no. Pt 7, pp. 1216–24, Apr. 2007.

[72] C. Murphy, “Endo-fin-ally a SARA for BMP receptors.,” *J. Cell Sci.*, vol. 120, no. Pt 7, pp. 1153–5, Apr. 2007.

[73] A. Moustakas and C.-H. Heldin, “The regulation of TGFbeta signal transduction.,” *Development*, vol. 136, no. 22, pp. 3699–714, Nov. 2009.

[74] T. Wang, B.-Y. Li, P. D. Danielson, P. C. Shah, S. Rockwell, R. J. Lechleider, J. Martin, T. Manganaro, and P. K. Donahoe, "The Immunophilin FKBP12 Functions as a Common Inhibitor of the TGFβ Family Type I Receptors," *Cell*, vol. 86, no. 3, pp. 435–444, 1996.

[75] T. Okadome, E. Oeda, M. Saitoh, H. Ichijo, H. L. Moses, K. Miyazono, and M. Kawabata, "Characterization of the Interaction of FKBP12 with the Transforming Growth Factor-beta Type I Receptor in Vivo," *J. Biol. Chem.*, vol. 271, no. 36, pp. 21687–21690, Sep. 1996.

[76] M. Huse, T. W. Muir, L. Xu, Y. G. Chen, J. Kuriyan, and J. Massagué, "The TGF beta receptor activation process: an inhibitor- to substrate-binding switch.," *Mol. Cell*, vol. 8, no. 3, pp. 671–82, Sep. 2001.

[77] G.-A. Song, H.-J. Kim, K.-M. Woo, J.-H. Baek, G.-S. Kim, J.-Y. Choi, and H.-M. Ryoo, "Molecular consequences of the ACVR1(R206H) mutation of fibrodysplasia ossificans progressiva.," *J. Biol. Chem.*, vol. 285, no. 29, pp. 22542–53, Jul. 2010.

[78] C. C. Hong and P. B. Yu, "Applications of small molecule BMP inhibitors in physiology and disease.," *Cytokine Growth Factor Rev.*, vol. 20, no. 5–6, pp. 409–18, 2009.

[79] L. Attisano and J. L. Wrana, "Signal transduction by the TGF-beta superfamily.," *Science*, vol. 296, no. 5573, pp. 1646–7, May 2002.

[80] G. Murakami, T. Watabe, K. Takaoka, K. Miyazono, and T. Imamura, "Cooperative inhibition of bone morphogenetic protein signaling by Smurf1 and inhibitory Smads.," *Mol. Biol. Cell*, vol. 14, no. 7, pp. 2809–17, Jul. 2003.

[81] L. N. Johnson, "Protein kinase inhibitors: contributions from structure to clinical compounds.," *Q. Rev. Biophys.*, vol. 42, no. 1, pp. 1–40, Feb. 2009.

[82] A. P. Kornev, N. M. Haste, S. S. Taylor, and L. F. Ten Eyck, "Surface comparison of active and inactive protein kinases identifies a conserved activation mechanism.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 103, no. 47, pp. 17783–8, Nov. 2006.

[83] M. Huse, Y. Chen, J. Massagué, and J. Kuriyan, "Crystal structure of the cytoplasmic domain of the type I TGF Î^2 receptor in complex with FKBP12," *Cell*, vol. 96, pp. 425–436, 1999.

[84] J. Massagué, "How cells read TGF-beta signals," *Nat. Rev. Mol. Cell Biol.*, vol. 1, no. December, 2000.

[85] T. Fukuda, M. Kohda, K. Kanomata, J. Nojima, A. Nakamura, J. Kamizono, Y. Noguchi, K. Iwakiri, T. Kondo, J. Kurose, K. Endo, T. Awakura, J. Fukushi, Y. Nakashima, T. Chiyonobu, A. Kawara, Y. Nishida, I. Wada, M. Akita, T. Komori, K. Nakayama, A. Nanba, Y. Maruki, T. Yoda, H. Tomoda, P. B. Yu, E. M. Shore, F. S. Kaplan, K. Miyazono, M. Matsuoka, K. Ikebuchi, A. Ohtake, H. Oda, E. Jimi, I. Owan, Y. Okazaki, and T. Katagiri, "Constitutively activated ALK2 and

increased SMAD1/5 cooperatively induce bone morphogenetic protein signaling in fibrodysplasia ossificans progressiva.," *J. Biol. Chem.*, vol. 284, no. 11, pp. 7149–56, Mar. 2009.

[86] M. van Dinther, N. Visser, D. J. J. de Gorter, J. Doorn, M.-J. Goumans, J. de Boer, and P. ten Dijke, "ALK2 R206H mutation linked to fibrodysplasia ossificans progressiva confers constitutive activity to the BMP type I receptor and sensitizes mesenchymal cells to BMP-induced osteoblast differentiation and bone formation.," *J. Bone Miner. Res.*, vol. 25, no. 6, pp. 1208–15, Jun. 2010.

[87] A. Chaikuad, I. Alfano, G. Kerr, C. E. Sanvitale, J. H. Boergermann, J. T. Triffitt, F. von Delft, S. Knapp, P. Knaus, and A. N. Bullock, "Structure of the bone morphogenetic protein receptor ALK2 and implications for fibrodysplasia ossificans progressiva.," *J. Biol. Chem.*, vol. 287, no. 44, pp. 36990–8, Oct. 2012.

[88] A. Hughes, F. Monsell, and M. Gargan, "Fibrodysplasia ossificans progressiva," *Curr. Orthop.*, vol. 22, no. 1, pp. 48–51, Feb. 2008.

[89] J. Eswaran and S. Knapp, "Insights into protein kinase regulation and inhibition by large scale structural comparison.," *Biochim. Biophys. Acta*, vol. 1804, no. 3, pp. 429–32, Mar. 2010.

[90] F. Zuccotto, E. Ardini, E. Casale, and M. Angiolini, "Through the 'gatekeeper door': exploiting the active kinase conformation.," *J. Med. Chem.*, vol. 53, no. 7, pp. 2681–94, Apr. 2010.

[91] P. B. Yu, C. C. Hong, C. Sachidanandan, J. L. Babitt, D. Y. Deng, S. A. Hoyng, H. Y. Lin, K. D. Bloch, and R. T. Peterson, "Dorsomorphin inhibits BMP signals required for embryogenesis and iron metabolism.," *Nat. Chem. Biol.*, vol. 4, no. 1, pp. 33–41, Jan. 2008.

[92] J. Hao, J. Ho, and J. Lewis, "In Vivo Structure Activity Relationship Study of Dorsomorphin Analogues Identifies Selective VEGF and BMP Inhibitors," *ACS Chem. Biol.*, vol. 5, no. 2, pp. 245–253, 2010.

[93] J. H. Boergermann, J. Kopf, P. B. Yu, and P. Knaus, "Dorsomorphin and LDN-193189 inhibit BMP-mediated Smad, p38 and Akt signalling in C2C12 cells.," *Int. J. Biochem. Cell Biol.*, vol. 42, no. 11, pp. 1802–7, Nov. 2010.

[94] P. B. Yu, D. Y. Deng, C. S. Lai, C. C. Hong, G. D. Cuny, M. L. Bouxsein, D. W. Hong, P. M. Mcmanus, T. Katagiri, C. Sachidanandan, N. Kamiya, T. Fukuda, Y. Mishina, R. T. Peterson, and K. D. Bloch, "BMP type I receptor inhibition reduces heterotopic [corrected] ossification.," *Nat. Med.*, vol. 14, no. 12, pp. 1363–9, Dec. 2008.

[95] S. a Chakkalakal, D. Zhang, A. L. Culbert, M. R. Convente, R. J. Caron, A. C. Wright, A. D. a Maidment, F. S. Kaplan, and E. M. Shore, "An Acvr1 R206H knock-in mouse has fibrodysplasia ossificans progressiva.," *J. Bone Miner. Res.*, vol. 27, no. 8, pp. 1746–56, Aug. 2012.

[96] G. D. Cuny, P. B. Yu, J. K. Laha, X. Xing, J.-F. Liu, C. S. Lai, D. Y. Deng, C. Sachidanandan, K. D. Bloch, and R. T. Peterson, "Structure-activity relationship study of bone morphogenetic protein (BMP) signaling inhibitors.," *Bioorg. Med. Chem. Lett.*, vol. 18, no. 15, pp. 4388–92, Aug. 2008.

[97] C. Aslanidis and P. J. de Jong, "Ligation-independent cloning of PCR products (LIC-PCR).," *Nucleic Acids Res.*, vol. 18, no. 20, pp. 6069–74, Oct. 1990.

[98] C. Aslanidis and P. J. de Jong, "Ligation-independent cloning of PCR products (LIC-PCR).," *Nucleic Acids Res.*, vol. 18, no. 20, pp. 6069–74, Oct. 1990.

[99] B. Shrestha, C. Smee, and O. Gileadi, "Baculovirus expression vector system: an emerging host for high-throughput eukaryotic protein expression.," *Methods Mol. Biol.*, vol. 439, pp. 269–89, Jan. 2008.

[100] F. H. Niesen, H. Berglund, and M. Vedadi, "The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability.," *Nat. Protoc.*, vol. 2, no. 9, pp. 2212–21, Jan. 2007.

[101] U. B. Ericsson, B. M. Hallberg, G. T. Detitta, N. Dekker, and P. Nordlund, "Thermofluor-based high-throughput stability optimization of proteins for structural studies.," *Anal. Biochem.*, vol. 357, no. 2, pp. 289–98, Oct. 2006.

[102] D. Matulis, J. K. Kranz, F. R. Salemme, and M. J. Todd, "Thermodynamic stability of carbonic anhydrase: measurements of binding affinity and stoichiometry using ThermoFluor.," *Biochemistry*, vol. 44, no. 13, pp. 5258–66, Apr. 2005.

[103] R. Perozzo, G. Folkers, and L. Scapozza, "Thermodynamics of Protein–Ligand Interactions: History, Presence, and Future Aspects," *J. Recept. Signal Transduct.*, vol. 24, no. 1–2, pp. 1–52, Jan. 2004.

[104] C. Yang, "ITC Data Analysis in Origin v. 2.9," *MicroCal, Northamp-ton, MA*, no. January, 1993.

[105] N. E. Chayen, "Turning protein crystallisation from an art into a science.," *Curr. Opin. Struct. Biol.*, vol. 14, no. 5, pp. 577–83, Oct. 2004.

[106] I. Russo Krauss, A. Merlino, A. Vergara, and F. Sica, "An overview of biological macromolecule crystallization.," *Int. J. Mol. Sci.*, vol. 14, no. 6, pp. 11643–91, Jan. 2013.

[107] A. Wlodawer, W. Minor, Z. Dauter, and M. Jaskolski, "Protein crystallography for non-crystallographers, or how to get the best (but not more) from published macromolecular structures.," *FEBS J.*, vol. 275, no. 1, pp. 1–21, Jan. 2008.

[108] A. Wlodawer, W. Minor, Z. Dauter, and M. Jaskolski, "Protein crystallography for aspiring crystallographers or how to avoid pitfalls and traps in macromolecular structure determination.," *FEBS J.*, vol. 280, no. 22, pp. 5705–36, Nov. 2013.

[109] A. J. McCoy, R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni, and R. J. Read, “Phaser crystallographic software.,” *J. Appl. Crystallogr.*, vol. 40, no. Pt 4, pp. 658–674, Aug. 2007.

[110] A. G. W. Leslie, “The integration of macromolecular diffraction data.,” *Acta Crystallogr. D. Biol. Crystallogr.*, vol. 62, no. Pt 1, pp. 48–57, Jan. 2006.

[111] P. Evans, “Scaling and assessment of data quality.,” *Acta Crystallogr. D. Biol. Crystallogr.*, vol. 62, no. Pt 1, pp. 72–82, Jan. 2006.

[112] M. D. Winn, C. C. Ballard, K. D. Cowtan, E. J. Dodson, P. Emsley, P. R. Evans, R. M. Keegan, E. B. Krissinel, A. G. W. Leslie, A. McCoy, S. J. McNicholas, G. N. Murshudov, N. S. Pannu, E. a Potterton, H. R. Powell, R. J. Read, A. Vagin, and K. S. Wilson, “Overview of the CCP4 suite and current developments.,” *Acta Crystallogr. D. Biol. Crystallogr.*, vol. 67, no. Pt 4, pp. 235–42, Apr. 2011.

[113] P. Emsley, B. Lohkamp, W. G. Scott, and K. Cowtan, “Features and development of Coot.,” *Acta Crystallogr. D. Biol. Crystallogr.*, vol. 66, no. Pt 4, pp. 486–501, Apr. 2010.

[114] G. N. Murshudov, A. A. Vagin, and E. J. Dodson, “Refinement of macromolecular structures by the maximum-likelihood method.,” *Acta Crystallogr. D. Biol. Crystallogr.*, vol. 53, no. Pt 3, pp. 240–55, May 1997.

[115] J. Painter and E. a Merritt, “Optimal description of a protein structure in terms of multiple groups undergoing TLS motion.,” *Acta Crystallogr. D. Biol. Crystallogr.*, vol. 62, no. Pt 4, pp. 439–50, Apr. 2006.

[116] P. D. Adams, P. V Afonine, G. Bunkóczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L.-W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty, R. Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger, and P. H. Zwart, “PHENIX: a comprehensive Python-based system for macromolecular structure solution.,” *Acta Crystallogr. D. Biol. Crystallogr.*, vol. 66, no. Pt 2, pp. 213–21, Feb. 2010.

[117] I. W. Davis, A. Leaver-Fay, V. B. Chen, J. N. Block, G. J. Kapral, X. Wang, L. W. Murray, W. B. Arendall, J. Snoeyink, J. S. Richardson, and D. C. Richardson, “MolProbity: all-atom contacts and structure validation for proteins and nucleic acids.,” *Nucleic Acids Res.*, vol. 35, no. Web Server issue, pp. W375–83, Jul. 2007.

[118] J. C. Groppe, J. Wu, E. M. Shore, and F. S. Kaplan, “In vitro analyses of the dysregulated R206H ALK2 kinase-FKBP12 interaction associated with heterotopic ossification in FOP.,” *Cells. Tissues. Organs*, vol. 194, no. 2–4, pp. 291–5, Jan. 2011.

[119] S. Kumar and R. Nussinov, “Close-range electrostatic interactions in proteins,” *ChemBioChem*, pp. 604–617, 2002.

[120] Q. Shen, S. C. Little, M. Xu, J. Haupt, C. Ast, T. Katagiri, S. Mundlos, P. Seemann, F. S. Kaplan, M. C. Mullins, and E. M. Shore, “The fibrodysplasia ossificans

progressiva R206H ACVR1 mutation activates BMP-independent chondrogenesis and zebrafish embryo ventralization.," *J. Clin. Invest.*, vol. 119, no. 11, pp. 3462–72, Nov. 2009.

[121] R. Wieser, J. L. Wrana, and J. Massagué, "GS domain mutations that constitutively activate T beta R-I, the downstream signaling component in the TGF-beta receptor complex.," *EMBO J.*, vol. 14, no. 10, pp. 2199–208, May 1995.

[122] A. Nakao, T. Imamura, S. Souchelnytskyi, M. Kawabata, A. Ishisaki, E. Oeda, K. Tamaki, J. Hanai, C. H. Heldin, K. Miyazono, and P. ten Dijke, "TGF-beta receptor-mediated signalling through Smad2, Smad3 and Smad4.," *EMBO J.*, vol. 16, no. 17, pp. 5353–62, Sep. 1997.

[123] D. Pappin, P. Hojrup, and A. Bleasby, "Rapid identification of proteins by peptide-mass fingerprinting," *Curr. Biol.*, vol. 3, no. 6, pp. 327 – 331, 1993.

[124] A. Shrestha, G. Hamilton, E. O'Neill, S. Knapp, and J. M. Elkins, "Analysis of conditions affecting auto-phosphorylation of human kinases during expression in bacteria.," *Protein Expr. Purif.*, vol. 81, no. 1, pp. 136–43, Jan. 2012.

[125] K. Luo and H. F. Lodish, "Positive and negative regulation of type II TGF-beta receptor signal transduction by autophosphorylation on multiple serine residues.," *EMBO J.*, vol. 16, no. 8, pp. 1970–81, Apr. 1997.

[126] V. Q. Le and K. A. Wharton, "Hyperactive BMP signaling induced by ALK2(R206H) requires type II receptor function in a Drosophila model for classic fibrodysplasia ossificans progressiva.," *Dev. Dyn.*, vol. 241, no. 1, pp. 200–14, Jan. 2012.

[127] J. Bagarova, A. J. Vonner, K. A. Armstrong, J. Börgermann, C. S. C. Lai, D. Y. Deng, H. Beppu, I. Alfano, P. Filippakopoulos, N. W. Morrell, A. N. Bullock, P. Knaus, Y. Mishina, and P. B. Yu, "Constitutively Active ALK2 Receptor Mutants Require Type II Receptor Cooperation.," *Mol. Cell. Biol.*, vol. 33, no. 12, pp. 2413–24, Jun. 2013.

[128] J. C. Groppe, J. Wu, E. M. Shore, and F. S. Kaplan, "In vitro analyses of the dysregulated R206H ALK2 kinase-FKBP12 interaction associated with heterotopic ossification in FOP.," *Cells. Tissues. Organs*, vol. 194, no. 2–4, pp. 291–5, Jan. 2011.

[129] R. J. Akhurst and A. Hata, "Targeting the TGFβ signalling pathway in disease.," *Nat. Rev. Drug Discov.*, vol. 11, no. 10, pp. 790–811, Oct. 2012.

[130] J. M. Yingling, K. L. Blanchard, and J. S. Sawyer, "Development of TGF-beta signalling inhibitors for cancer therapy.," *Nat. Rev. Drug Discov.*, vol. 3, no. 12, pp. 1011–22, Dec. 2004.

[131] A. A. Ogunjimi, E. Zeqiraj, D. F. Ceccarelli, F. Sicheri, J. L. Wrana, and L. David, "Structural basis for specificity of TGFβ family receptor small molecule inhibitors.," *Cell. Signal.*, vol. 24, no. 2, pp. 476–83, Feb. 2012.

[132] N. Garbett and J. Chaires, “Thermodynamic studies for drug design and screening,” *Expert Opin. Drug Discov.*, vol. 7, no. 4, pp. 299–314, 2012.

[133] O. Fedorov, B. Marsden, V. Pogacic, P. Rellos, S. Müller, A. N. Bullock, J. Schwaller, M. Sundström, and S. Knapp, “A systematic interaction map of validated kinase inhibitors with Ser/Thr kinases.,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 104, no. 51, pp. 20523–8, Dec. 2007.

[134] A. V Statsuk, D. J. Maly, M. A. Seeliger, M. A. Fabian, W. H. Biggs, D. J. Lockhart, P. P. Zarrinkar, J. Kuriyan, and K. M. Shokat, “Tuning a three-component reaction for trapping kinase substrate complexes,” *J. Am. Chem. Soc.*, vol. 130, no. 51, pp. 17568–17574, 2008.

[135] M. W. Richards, L. O’Regan, C. Mas-Droux, J. M. Y. Blot, J. Cheung, S. Hoelder, A. M. Fry, and R. Bayliss, “An autoinhibitory tyrosine motif in the cell-cycle-regulated Nek7 kinase is released through binding of Nek9.,” *Mol. Cell*, vol. 36, no. 4, pp. 560–70, Nov. 2009.

[136] S. Hilton, S. Naud, J. J. Caldwell, K. Boxall, S. Burns, V. E. Anderson, L. Antoni, C. E. Allen, L. H. Pearl, A. W. Oliver, G. Wynne Aherne, M. D. Garrett, and I. Collins, “Identification and characterisation of 2-aminopyridine inhibitors of checkpoint kinase 2.,” *Bioorg. Med. Chem.*, vol. 18, no. 2, pp. 707–18, Jan. 2010.

[137] J. J. Cui, M. Tran-Dubé, H. Shen, M. Nambu, P.-P. Kung, M. Pairish, L. Jia, J. Meng, L. Funk, I. Botrous, M. McTigue, N. Grodsky, K. Ryan, E. Padrique, G. Alton, S. Timofeevski, S. Yamazaki, Q. Li, H. Zou, J. Christensen, B. Mroczkowski, S. Bender, R. S. Kania, and M. P. Edwards, “Structure based drug design of crizotinib (PF-02341066), a potent and selective dual inhibitor of mesenchymal-epithelial transition factor (c-MET) kinase and anaplastic lymphoma kinase (ALK).,” *J. Med. Chem.*, vol. 54, no. 18, pp. 6342–63, Sep. 2011.

[138] N. Ogawa, H. Yuki, and A. Tanaka, “Insights from Pim1 structure for anti-cancer drug design.,” *Expert Opin. Drug Discov.*, vol. 7, no. 12, pp. 1177–92, Dec. 2012.

[139] D. K. Whelligan, S. Solanki, D. Taylor, D. W. Thomson, K.-M. J. Cheung, K. Boxall, C. Mas-Droux, C. Barillari, S. Burns, C. G. Grummitt, I. Collins, R. L. M. van Montfort, G. W. Aherne, R. Bayliss, and S. Hoelder, “Aminopyrazine inhibitors binding to an unusual inactive conformation of the mitotic kinase Nek2: SAR and structural characterization.,” *J. Med. Chem.*, vol. 53, no. 21, pp. 7682–98, Nov. 2010.

[140] L. Gandhi and P. A. Jänne, “Crizotinib for ALK-rearranged non-small cell lung cancer: a new targeted therapy for a new target.,” *Clin. Cancer Res.*, vol. 18, no. 14, pp. 3737–42, Jul. 2012.

[141] Y. Younis, F. Douelle, T.-S. Feng, D. González Cabrera, C. Le Manach, A. T. Nchinda, S. Duffy, K. L. White, D. M. Shackleford, J. Morizzi, J. Mannila, K. Katneni, R. Bhamidipati, K. M. Zabiulla, J. T. Joseph, S. Bashyam, D. Waterson, M. J. Witty, D. Hardick, S. Wittlin, V. Avery, S. a Charman, and K. Chibale, “3,5-Diaryl-2-aminopyridines as a novel class of orally active antimalarials

demonstrating single dose cure in mice and clinical candidate potential.," *J. Med. Chem.*, vol. 55, no. 7, pp. 3479–87, Apr. 2012.

[142] C. A. Lipinski, F. Lombardo, B. W. Dominy, and P. J. Feeney, "Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings.," *Adv. Drug Deliv. Rev.*, vol. 46, no. 1–3, pp. 3–26, Mar. 2001.

[143] K. Lavery, P. Swain, D. Falb, and M. H. Alaoui-Ismaili, "BMP-2/4 and BMP-6/7 differentially utilize cell surface receptors to induce osteoblastic differentiation of human bone marrow-derived mesenchymal stem cells.," *J. Biol. Chem.*, vol. 283, no. 30, pp. 20948–58, Jul. 2008.

[144] C. Selvaraj, S. Tripathi, K. Reddy, and S. Singh, "Tool development for Prediction of pIC50 values from the IC50 values-A pIC50 value calculator," *Curr. Trends Biotechnol. Pharm.*, vol. 5, no. 2, pp. 1104–1109, 2011.

[145] A. H. Mohedas, X. Xing, K. A. Armstrong, A. N. Bullock, G. D. Cuny, and P. B. Yu, "Development of an ALK2-Biased BMP Type I Receptor Kinase Inhibitor.," *ACS Chem. Biol.*, vol. 8, no. 6, pp. 1291–1302, Apr. 2013.

[146] C. D. Mol, D. R. Dougan, T. R. Schneider, R. J. Skene, M. L. Kraus, D. N. Scheibe, G. P. Snell, H. Zou, B.-C. Sang, and K. P. Wilson, "Structural basis for the autoinhibition and STI-571 inhibition of c-Kit tyrosine kinase.," *J. Biol. Chem.*, vol. 279, no. 30, pp. 31655–63, Jul. 2004.

[147] C. D. Mol, K. B. Lim, V. Sridhar, H. Zou, E. Y. T. Chien, B.-C. Sang, J. Nowakowski, D. B. Kassel, C. N. Cronin, and D. E. McRee, "Structure of a c-kit product complex reveals the basis for kinase transactivation.," *J. Biol. Chem.*, vol. 278, no. 34, pp. 31461–4, Aug. 2003.

[148] D. Mitchell, E. G. Pobre, A. W. Mulivor, A. V Grinberg, R. Castonguay, T. E. Monnell, N. Solban, J. a Ucran, R. S. Pearsall, K. W. Underwood, J. Seehra, and R. Kumar, "ALK1-Fc inhibits multiple mediators of angiogenesis and suppresses tumor growth.," *Mol. Cancer Ther.*, vol. 9, no. 2, pp. 379–88, Feb. 2010.

[149] L. A. van Meeteren, M. Thorikay, S. Bergqvist, E. Pardali, C. G. Stampino, D. Hu-Lowe, M.-J. Goumans, and P. ten Dijke, "Anti-human activin receptor-like kinase 1 (ALK1) antibody attenuates bone morphogenetic protein 9 (BMP9)-induced ALK1 signaling and interferes with endothelial cell sprouting.," *J. Biol. Chem.*, vol. 287, no. 22, pp. 18551–61, May 2012.

APPENDIX

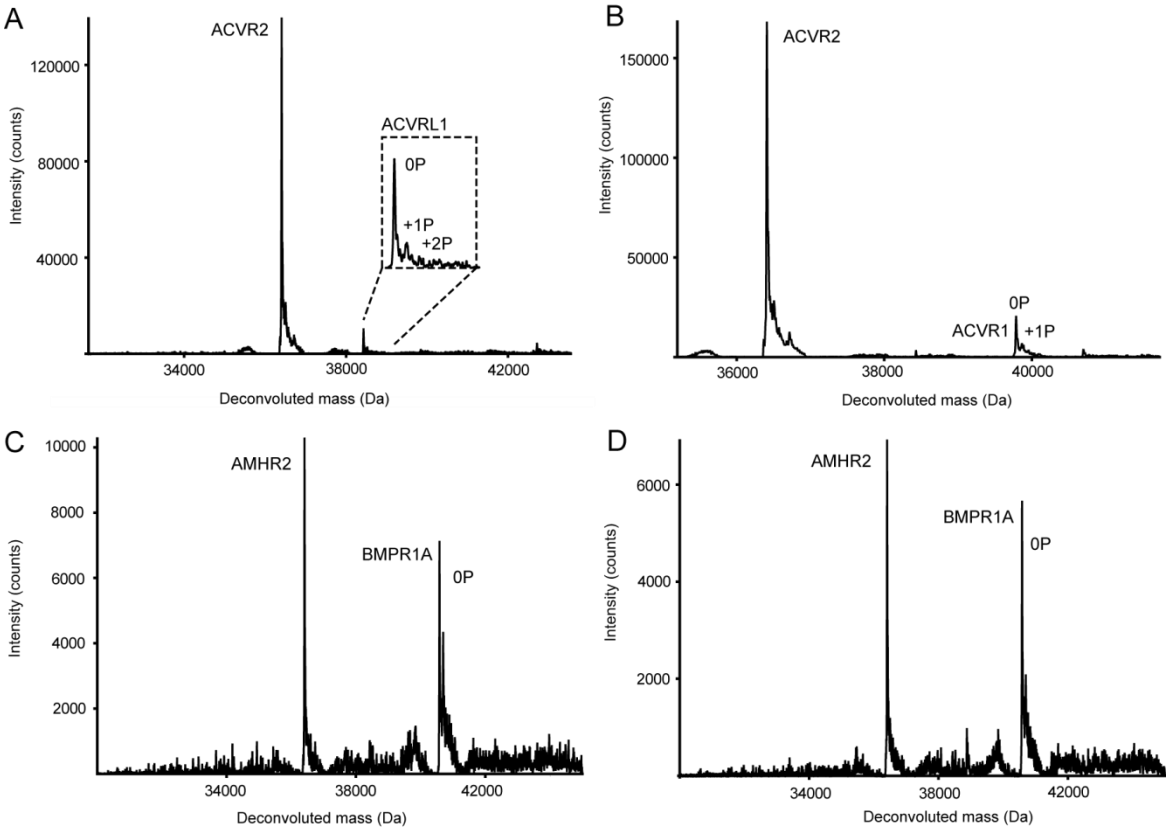


Figure A1 Mass spectrometry following co-expression. (A) ACVRL1 and ACVR2 showed up to 2 phosphorylations of the type I receptor (B) ACVR1 and ACVR2.showed up to 1 phosphorylation of the type I receptor. (C) and (D) Two different constructs of AMHR2 showed no phosphorylation of BMPR1A after expression.

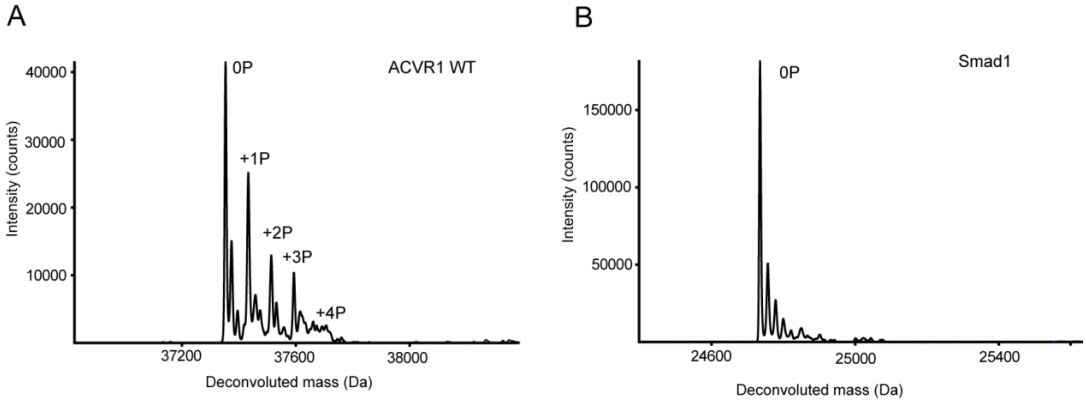


Figure A2 Mass spectra of the Smad1 and wild-type ACVR1 mix following an in vitro kinase assay. (A) ACVR1 showed up to 4 phosphorylations (B) Smad1 was not phosphorylated.