The Structural Basis for the Phosphorylation-Induced Activation of Smad Proteins: a Dissertation

Benoy M. Chacko
University of Massachusetts Medical School

Follow this and additional works at: https://escholarship.umassmed.edu/gsbs_diss

Part of the Amino Acids, Peptides, and Proteins Commons, Biochemical Phenomena, Metabolism, and Nutrition Commons, Cells Commons, Enzymes and Coenzymes Commons, Genetic Phenomena Commons, and the Investigative Techniques Commons

Repository Citation

This material is brought to you by eScholarship@UMMS. It has been accepted for inclusion in GSBS Dissertations and Theses by an authorized administrator of eScholarship@UMMS. For more information, please contact Lisa.Palmer@umassmed.edu.
THE STRUCTURAL BASIS FOR THE PHOSPHORYLATION-INDUCED ACTIVATION OF SMAD PROTEINS

A Dissertation Presented

by

Benoy Maramparambil Chacko

Submitted to the Faculty of the
University of Massachusetts Graduate School of Biomedical Sciences, Worcester
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

IN

BIOCHEMISTRY AND MOLECULAR PHARMACOLOGY

February 23, 2004
Parts of this dissertation have appeared as separate publications:


THE STRUCTURAL BASIS FOR THE PHOSPHORYLATION-INDUCED ACTIVATION OF SMAD PROTEINS

A Dissertation Presented
By
BENOY M. CHACKO

Approved as to style and content by:

Celia A. Schiffer, Ph.D., Chair of Committee

David G. Lambright, Ph.D., Member of Committee

Mary Munson, Ph.D., Member of Committee

Alonzo H. Ross, Ph.D., Member of Committee

Rik Derynck, Dr. Sc., Member of Committee

Kai Lin, Ph.D., Thesis Advisor

Anthony Carruthers, Ph.D.,
Dean of the Graduate School of Biomedical Sciences

Department of Biochemistry and Molecular Pharmacology

February 23, 2004
ACKNOWLEDGMENTS

I could not have accomplished the work described in this thesis without the assistance of many others, particularly my fellow members of the Lin Laboratory. First and foremost, I have to thank my advisor, Dr. Kai Lin, for his guidance and support. My abilities as a scientist have grown immeasurably under Kai’s tutelage. Most of the molecular biology and protein purification techniques I used in my studies were taught to me by Suvana Lam, for which I am very grateful. Much of the success of the experiments in this thesis are directly due to Suvana’s efforts. I also thank Dr. Bin Qin, whose considerable expertise of X-ray crystallography was vital to my project, and who kindly took time to instruct me in structure determination. Bin solved the Smad1 crystal structure described in Chapter III, and also helped solve the heteromeric structures described in Chapter IV. Hema Srinath is a superb technician who I have had the privilege of working with, and who made sure I was never lacking for supplies.

I am also grateful to the members of my thesis committee: Drs. Celia Schiffer, David Lambright, Kendall Knight, Alonzo Ross, Mary Munson, and Rik Derynck, who provided invaluable guidance and advice throughout my graduate school experience. Many of their suggestions found their way into experiments and figures in this thesis. I would especially like to thank Celia, for many helpful discussions during the writing phase.

I would also like to thank Dr. Jack Correia, whose expertise in analytical ultracentrifugation was essential to the studies described in Chapter II. The signaling
assays described in Chapters II and III, as well as the immunprecipitation analysis in Chapter IV, were done by Dr. Mark de Caestecker. Dr. Ashutosh Tiwari was helpful in getting the isothermal titration calorimetry studies off the ground, and also assisted with the differential scanning calorimetry analysis. Dr. Lawrence Hayward provided many helpful suggestions with regard the calorimetric experiments.

I also have to thank the friends I have made here at UMass, who made an often stressful experience enjoyable. Last, but certainly not least, I thank my family, to whom I dedicate this work, and without whose constant love and support I could never have completed this degree.
To my family
ABSTRACT

The Smad proteins transduce the signal of transforming growth factor-β (TGF-β) and related factors from the cell surface to the nucleus. Following C-terminal phosphorylation by a corresponding receptor kinase, the R-Smad proteins form heteromeric complexes with Smad4. These complexes translocate into the nucleus, bind specific transcriptional activators and DNA, ultimately modulating gene expression. Though studied through a variety of means, the stoichiometry of the R-Smad/Smad4 complex is unclear. We investigated the stoichiometry of the phosphorylation-induced R-Smad/Smad4 complex by using acidic amino acid substitutions to simulate phosphorylation. Size exclusion chromatography, analytical ultracentrifugation, and isothermal titration calorimetry analysis revealed that the R-Smad/Smad4 complex is a heterotrimer consisting of two R-Smad subunits and one Smad4 subunit. In addition, a specific mechanism for phosphorylation-induced R-Smad/Smad4 complex formation was studied. Although it had been previously established that part of the mechanism through which phosphorylation induces Smad oligomerization is through relieving MH1-domain mediated autoinhibition of the MH2 (oligomerization) domain, it is also evident that phosphorylation serves to energetically drive Smad complex formation. Through mutational and size exclusion chromatography analysis, we established that phosphorylation induces oligomerization of the Smads by creating an electrostatic interaction between the phosphorylated C-terminal tail of one R-Smad subunit in a Smad trimer with a basic surface on an adjacent R-Smad or Smad4 subunit. The basic surface is defined largely by the L3 loop, a region that had previously been implicated in R-Smad
interaction with the receptor kinase. Furthermore, the Smad MH2 domain shares a similar protein fold with the phosphoserine and phosphothreonine-binding FHA domains from proteins like Rad53 and Chk2. Taken together, these results suggest that the Smad MH2 domain may be a distinct phosphoserine-binding domain, which utilizes a common basic surface to bind the receptor kinase and other Smads, and takes advantage of phosphorylation-induced allosteric changes dissociate from the receptor kinase and oligomerize with other Smads. Finally, the structural basis for the preferential formation of the R-Smad/Smad4 heterotrimeric complex over the R-Smad homotrimeric complex was explored through X-ray crystallography and isothermal titration calorimetry. Crystal structures of the Smad2/Smad4 and Smad3/Smad4 complexes revealed that specific residue differences in Smad4 compared to R-Smads resulted in highly favorable electrostatic interactions that explain the preference for the interaction with Smad4.
TABLE OF CONTENTS

Title i
Copyright information ii
Signature page iii
Acknowledgments iv
Dedication vi
Abstract vii
Table of contents ix
List of figures xiii
List of tables xvii

CHAPTER I: INTRODUCTION 1

Overview of TGF-β signaling 1
The Smads are the primary mediators of TGF-β signaling 1
Propagation of the TGF-β signal 5
TGF-β and Smads in human disease 9
Smad complex stoichiometry and mechanism of formation 11
A specific mechanism for phosphorylation-induced activation 15
Preferential formation of the R-Smad/Smad4 complex 16
CHAPTER II: ANALYSIS OF R-SMAD/CO-SMAD Oligomerization Through Pseudophosphorylated Mutants

Introduction

Materials and Methods

Reagents

Construction of expression plasmids and mutagenesis

Protein expression and purification

Size exclusion chromatography

Transcriptional response assays

Analytical ultracentrifugation

Results

R-Smads and Smad4 display different oligomerization tendencies

Phosphorylation promotes R-Smad homotrimerization

Attachment of a pseudophosphorylated tail to Smad4

Stoichiometry of the R-Smad/Smad4 heteromeric complex

Role of the conserved trimer interface

Discussion

CHAPTER III: SMAD Trimerization is Facilitated by the Interaction between the Positively-Charged Loop-Strand Pocket and the C-Terminal Phosphorylated Tail

Introduction

Materials and Methods

Reagents
Construction of expression plasmids and mutagenesis

Protein expression and purification

Size exclusion chromatography

Transcriptional response assays

Crystallization of S4AF(R515S) and Smad1LCS

Results

Mutational analysis of anion-binding pockets in Smad4

Crystallization of S4AF(R515S)

Signaling assays confirm the importance of the L3 loop arginine

Crystal structure of S1LCS

Concerted structural changes drive Smad1 activation

Mutational analysis based on the crystal structure

Discussion

The Smad MH2 is structurally related to other phosphoserine binding domains

A phosphorylation-induced allosteric switch drives Smad dissociation from the receptor complex

CHAPTER IV: CRYSTALLOGRAPHIC AND CALORIMETRIC ANALYSIS OF THE R-SMAD/SMAD4 COMPLEX

Introduction

Materials and Methods

Reagents

Generation of phosphorylated Smad2 and Smad3
Isothermal titration calorimetry

Differential scanning calorimetry

Crystallization of phosphorylated R-Smad/Smad4 complexes and structure determination

Size exclusion chromatography

Immunoprecipitation and Western blots

Results

Generation of phosphorylated forms of Smad2 and Smad3 and isothermal titration calorimetry

Crystal structures of phosphorylated Smad3(LC, 2P)/Smad4(AF) and phosphorylated Smad2(LC, 2P)/Smad4(AF)

Overall Structure of the R-Smad/Smad4 Heterotrimer and Interface Analysis

Asp493 of Smad4 coordinates a buried electrostatic interaction in the BC Interface

Arg378 of Smad4 reinforces phosphoserine binding in the AB Interface

The CA Interface Resembles the Homotrimeric Interface

Structural stabilization of R-Smads by Smad4 in the heterotrimer

Discussion

CHAPTER V: GENERAL DISCUSSION

REFERENCES
# LIST OF FIGURES

## CHAPTER I

**Figure 1-1.** The three classes of Smad proteins and domain organization

**Figure 1-2.** Crystal structure of the Smad4 MH2 domain and location of tumorigenic mutations

**Figure 1-3.** Basic mechanism of TGF-β signaling through the Smad proteins

**Figure 1-4.** Smad mutations found in cancer

## CHAPTER II

**Figure 2-1.** Diagram of the S3LC, S1LC, and S4AF constructs

**Figure 2-2.** Size exclusion chromatography profiles of S3LC and S4AF

**Figure 2-3.** Sedimentation velocity analysis of S4AF, S3LC, and mixtures of the two Smad constructs

**Figure 2-4.** Sedimentation velocity analysis of S1LC and S1LC(2D)

**Figure 2-5.** Weight average $s_{20,w}$ values derived from g(s) analysis of sedimentation velocity runs with S4AF, S4AF(3E), S3LC, and S3LC(3E)

**Figure 2-6.** Sedimentation equilibrium runs of S4AF.

**Figure 2-7.** Transcriptional analysis of pseudophosphorylated Smad3 and Smad1 constructs

**Figure 2-8.** Size exclusion chromatography elution profiles of S3LC and S3LC(3E)

**Figure 2-9.** Sedimentation velocity analysis of S3LC(3E) and mixtures of S4AF and S3LC(3E)

**Figure 2-10.** Sedimentation velocity analysis of S4AF(3E)
Figure 2-11. Size exclusion analysis of S3LC(3E)/S4AF and S1LC(2D)/S4AF complex stoichiometry

Figure 2-12. Sedimentation velocity analysis of S1LCS(2D)/S4AF

Figure 2-13. S4AF crystal structure

Figure 2-14. Structure-based sequence alignment of the MH2 domains of the Smad proteins

Figure 2-15. Mutation of S4AF trimer interface residues disrupts interaction between pseudophosphorylated R-Smads and S4AF

Figure 2-16. Mutation of the equivalent trimer interface residues disrupts R-Smad homotrimerization

CHAPTER III

Figure 3-1. F_o-F_c map displaying the electron density of the sulfate ions bound at individual sulfate-binding sites throughout the previously solved S4AF structure

Figure 3-2. Detailed representation of the four unique sulfate-binding sites

Figure 3-3. Analysis of interaction between S3LC(3E) and S4AF sulfate-binding site mutants

Figure 3-4. Stereo view of the F_o-F_c omit map at the L3 loop of S4AF(R515S)

Figure 3-5. Transcriptional analysis of Smad4 sulfate-binding site mutants

Figure 3-6. Crystal structure of S1LCS

Figure 3-7. Surface electrostatic potential representation of the interaction between the L3 loop/β8 strand pocket and the C-terminal tail; View of L3 loop in tail-bound and unbound forms

Figure 3-8. Mutation of the Smad phosphoserine-binding pocket disrupts both R-Smad homotrimerization and R-Smad/Smad4 heterotrimerization

Figure 3-9. The Smad MH2 domain is structurally homologous to the FHA and IAD domains
Figure 3-10. An allosteric model for Smad activation

CHAPTER IV

Figure 4-1. Intein-mediated peptide ligation strategy used to generate phosphorylated Smad3
Figure 4-2. Purification of phosphorylated S2LC and S3LC
Figure 4-3. Isothermal titration calorimetry analysis of the interaction between S4AF and S2LC(2P) and S3LC(2P)
Figure 4-4. Structure-based sequence alignment of the MH2 domains of the Smad proteins
Figure 4-5. Overall structure of the R-Smad/Smad4 heterotrimeric complexes
Figure 4-6. Superposition of the alpha carbon traces of the Smad2 homotrimer and the Smad3/Smad4 heterotrimer
Figure 4-7. Representative interface from the Smad2 homotrimer structure
Figure 4-8. Smad3/Smad4 BC interface
Figure 4-9. Electrostatic representation of D493 and surrounding residues of the B subunit and the adjacent interacting residues in the C subunit
Figure 4-10. SEC and ITC analysis of S3LC(2P)/S4AF(D493A) interaction
Figure 4-11. Co-immunoprecipitation analysis of Smad2/Smad4 and Smad3/Smad4 interaction
Figure 4-12. Size exclusion chromatography analysis of wild-type Smad3(LC) and S3LC(Y364D)
Figure 4-13. The AB interface of the Smad3-Smad4 crystal structure
Figure 4-14. Representation of surface electrostatic potential of phosphorylated C-terminal tail interaction with L3 loop/β8 strand pocket in the Smad3/Smad4 AB interface and the Smad2 homotrimer interface
Figure 4-15. SEC and ITC analysis of S3LC(2P)/S4AF(R378A) interaction

Figure 4-16. The CA interface of the S3LC(2P)/S4AF heterotrimer

Figure 4-17. DSC analysis of the S3LC(2P)/S4AF interaction
LIST OF TABLES

CHAPTER II

Table 2-1. Trimerization constants for S3LC and S3LC(3E) as determined through either sedimentation equilibrium or sedimentation velocity methods. 33

Table 2-2. Summary of analytical ultracentrifuge studies on Smad1 homomeric interaction 43

CHAPTER III

Table 3-1. Summary of crystal analysis for S4AF(R515S). 73

Table 3-2. Summary of crystal analysis for S1LCS 78

CHAPTER IV

Table 4-1. Summary of isothermal titration calorimetry of phosphorylated Smad3 and Smad2 with Smad4 107

Table 4-2. Summary of crystal analysis for Smad2(LC,2P)/Smad4(AF) and Smad3(LC,2P)/Smad4(AF) 111

Table 4-3. Hydrogen bond comparison between the Smad2 homotrimer interfaces and the three Smad3/Smad4 heterotrimer interfaces 116
CHAPTER I: INTRODUCTION

Overview of TGF-β Signaling

The transforming growth factor-beta (TGF-β) superfamily, includes the TGF-βs, bone morphogenetic proteins (BMPs), and activins, families of secreted polypeptide growth factors that regulate growth and development, in organisms ranging from C. elegans to human (Kingsley 1994; Massague, 1998; Derynck and Zhang 2003). Structurally related by the presence of three conserved disulfide bonds, these growth factors have essential roles in regulating cell growth, differentiation, migration, apoptosis, and homeostasis (Shi, 2001; ten Dijke et al., 2002). The TGF-β ligands are able to elicit numerous biological activities, depending on the cellular context. A well-known effect of TGF-β is its ability to induce cell cycle arrest in epithelial cells, which is linked to its role as a tumor suppressor (Derynck and Zhang, 2003). Other specific activities of these ligands include regulating the immune response, directing embryonic patterning, stimulating extracellular matrix formation, and promoting wound healing (Roberts and Sporn 1990; Moustakas et al., 2001). TGF-β ligands act by binding specific transmembrane receptor kinases, activating the Smad transducer proteins, and ultimately regulating specific transcription (Massague, 1998).

The Smad proteins are the primary mediators of TGF-β signaling

The first downstream mediator of TGF-β signaling to be identified was called Mothers Against Decapentaplegic (Mad), after being identified from a screen for genes that enhance decapentaplegic (dpp) signaling in Drosophila (Raftery et al., 1995; Sekelsky et al., 1995). Homologues to Mad were later discovered in C. elegans and
named Sma-2, Sma-3, and Sma-4, since mutation of these genes resulted in small body size (Savage et al., 1996). The subsequent identification of several vertebrate homologues led to the classification of these proteins as Smads, a combination of *C. elegans Sma* and *Drosophila Mad* (Derynck et al., 1996). The first human homologue was the tumor suppressor DPC4 (Deleted in Pancreatic Carcinoma, locus 4), later named Smad4 (Hahn et al., 1996).

The eight known vertebrate Smads are between 400 and 500 amino acids in length, and are divided into three classes according to function (Figure 1-1A; Derynck and Zhang, 2003). The receptor-regulated Smads, or R-Smads, are direct substrates of corresponding TGF-β receptor kinases. This class of Smads is highly pathway-specific, with Smads 2 and 3 exclusively involved in TGF-β and activin signaling (Zhang et al., 1996), and Smads 1, 5, and 8 active only in BMP signaling (Liu et al., 1996; Nishimura et al., 1998). The common mediator Smad (Co-Smad) subclass includes only one protein in vertebrates, Smad4. Smad4 participates in TGF-β signaling by associating with the R-Smads (Zhang et al., 1996). The third class of Smads is known as the inhibitory Smads, or I-Smads, which includes Smads 6 and 7 (Hata et al., 1998a; Nakao et al., 1997). The I-Smads are structurally divergent from the R-Smads and Smad4, and these proteins negatively regulate TGF-β signaling by preventing R-Smad phosphorylation by the receptor kinase or inhibiting formation of an R-Smad/Smad4 complex (Hata et al., 1998a; Kavsak et al., 2000).
R-Smads and Smad4 have two distinct domains of high sequence similarity (ten Dijke et al., 2000; Figure 1-1). The MH1 (Mad-homology 1) domain of approximately 130 amino acids is located at the N-terminus. This domain has an intrinsic DNA-binding activity in most R-Smads and Smad4, the only exception being Smad2, which has a specific insertion that precludes DNA binding (Shi et al., 1998; Yagi et al., 1999). The MH1 domain is composed of four α helices, six short β strands, and five loops. DNA-binding is accomplished through the β2 and β3 strands, which form a β hairpin structure (Shi et al., 1998). In the basal, unphosphorylated state, the MH1 domain of R-Smads inhibits the oligomerization activity of the MH2 domain through direct binding. Phosphorylation decreases the affinity of the MH1 for the MH2 domain, thereby relieving the auto-inhibition. This domain is also able to bind some transcription factors, although at a lower level than the MH2 domain (Hata et al., 1997; Massague, 1998).

The MH2 domain is approximately 200 amino acids long and located at the C-terminus of Smads. This domain is responsible for mediating the primary activities of the Smad proteins, including interaction with the receptor kinase and DNA-binding proteins, homo- and hetero-oligomerization, and transcriptional activation (Macias-Silva et al., 1996; Chen et al., 1997; Liu et al., 1996; Massague, 1998). The MH2 domain of R-Smads also contains a conserved SSXS motif at the C-terminus, the last two serines of which are the sites of receptor phosphorylation (Abdollah et al., 1997; Souchelnytskyi et al., 1997). Smad4 lacks the consensus phosphorylation sequence, and TGF-β dependent phosphorylation of this protein has accordingly not been reported in mammals. However,
Receptor-regulated Smads

Mad (Drosophila)
- Smad1
- Smad5
- Smad6

Smad2
- Smad3
- Sma-2.3 (C. elegans)

Co-Smads
- Smad4/DPC4
- Medea (Drosophila)
- Sma-4 (C. elegans)

Inhibitory Smads
- Smad6
- Smad7
- Dad (Drosophila)

Signaling specificity
- Dpp via Thickveins
- BMP2/4 via BMPR-I
- TGF-β via TβR-I
- Activin via ActR-Iβ

Receptor phosphorylation
- SSXS motif

Basal state

Regulatory phosphorylations

Activated state

DNA binding

MAP kinase phosphorylation

• SMAD-receptor interaction
• SMAD homomedization
• Receptor phosphorylation
• SSXS

MH1 domain
- Linker
- MH2 domain

• Autoinhibition
• SMAD-Smad4 interaction
• Interaction with DNA-binding proteins
• Activation of transcription

Figure 1-1. (A) The three classes of Smad proteins. Reproduced from Hata et al., 1998b. (B) Smad domain organization and functions. In the basal (unphosphorylated) state, the MH1 and MH2 domains inhibit each other through interaction. R-Smads MH2 domains interact with and are phosphorylated by the activated type I receptor at the C-terminal SSXS motif. Once activated, R-Smads associate with Smad4 and DNA-binding proteins via the MH2 domains, and the MH1 domain is able to mediate DNA binding. Reproduced from Massague, 1998.
TGF-β induced phosphorylation of a *Xenopus* Smad4 isoform, Smad4p, has been shown (Howell et al., 1999). Smad4 is phosphorylated at sites in the linker domain by proteins of the MAP kinase pathway, although the significance of these sites is unknown (Derynck and Zhang, 2003).

The crystal structure of the MH2 domain of Smad4 revealed that it consists of a central β-strand sandwich with twisted antiparallel β-sheets of five and six strands each (Shi et al., 1997). The β-sandwich is flanked on one end by a three-helix bundle and on the other end by three loops and another helix, referred to as the loop-helix domain (Figure 1-2A). The central β-strand sandwich of the MH2 domain shares a similar β strand topology with the phosphoserine/phosphothreonine-binding forkhead-associated (FHA) domain, as well as with the transactivation domain of the interferon regulatory factor IRF-3, indicating that this β sandwich may be a conserved motif for binding phosphorylated serine or threonine residues (Li et al., 2000; Durocher et al., 2000; Qin et al., 2003; Takahasi et al., 2003).

The MH1 and MH2 domains are separated by a linker domain of varying length and sequence (Figure 1-1), the C-terminal 47 residues of which have been shown to be required for mediating signaling responses in Smad4 (de Caestecker et al., 2000). This region also contains phosphorylation sites for Erk-family MAP kinases, mediating cross-talk with other signaling pathways (Kretzschmar et al., 1997).

**Propagation of the TGF-β signal**

TGF-β signaling is initiated by binding of a specific ligand to its corresponding type II transmembrane receptor kinase (e.g., TGF-β1 ligand and TGF-β type II receptor),
Figure 1-2. Crystal structure of the Smad4 MH2 domain, as solved by Shi and co-workers. (A) The Smad4 MH2 domain comprises a central β strand sandwich, capped on one end by a three-helix bundle (H3, H4, H5) and on the other end by a loop-helix region (L1, L2, L3, H1). (B) The Smad4 MH2 domain crystallizes as a trimer. Interface residues known to be mutated in cancer are depicted in yellow. (C) Other mutations map to the L3 loop region, which is located perpendicularly to the trimer interface. All figures reproduced from Shi et al., 1997.
TGF-β Kinase domain RII

Sma sma

Lj TGF-β specific genes

co-factors (e.g., Fast-1)

TGF-β specific genes

Figure 1-3. Basic mechanism of TGF-β signaling through the Smad proteins. Binding of TGF-β ligand to its corresponding type II receptor induces oligomerization with the type I receptor. Constitutively active type II phosphorylates type I receptor at the GS domain, activating the receptor, and creating a binding site for R-Smads. R-Smads bind type I receptor with assistance from the anchoring protein SARA. Upon phosphorylation by the type I receptor, R-Smads dissociate from the receptor-SARA complex and are able to homo-oligomerize, but preferentially hetero-oligomerize with Smad4. (Note: The Smad complexes are depicted as dimers only for simplicity; the stoichiometry of these complexes is a central question of this thesis.) R-Smad/Smad4 complexes translocate into the nucleus, bind other transcription factors, and activate TGF-β specific genes. Adapted from Massague, 1998.
leading to formation of a complex between type II and type I receptor kinases, following which the type I receptor kinase is phosphorylated and activated by the constitutively active type II receptor kinase (Wrana et al., 1994; Figure 1-3). Phosphorylation of the type I receptor occurs at a region just inside the plasma membrane containing a preponderance of glycine and serine residues (GS region) (Huse et al., 2001), creating a binding site for the receptor-regulated Smads, or R-Smads, which are subsequently phosphorylated by the type I receptor kinase (Macias-Silva et al., 1996; Kretzschmar et al., 1997). The interaction is structurally mediated by the L45 loop of the type I receptor kinase and the L3 loop of R-Smads (Feng and Derynck, 1997; Lo et al., 1998; Chen et al., 1998). For certain R-Smads, the interaction with the type I receptor kinase requires the targeting protein SARA (Tsukazaki et al., 1998; Hayes et al., 2002).

Following C-terminal phosphorylation, R-Smads dissociate from the receptor-SARA complex and are able to form homo-oligomeric complexes. However, in the presence of the Co-Smad, Smad4, R-Smads preferentially interact with Smad4 to form hetero-oligomeric complexes (Lagna et al., 1996). Mutations in the L3 loop of Smad4 appeared to inhibit its ability to associate with Smad2 (Shi et al., 1997). Thus, the L3 loop of Smads appears to mediate two key protein-protein interactions that are essential in TGF-β signaling: the receptor-Smad interaction, as well as the Smad-Smad interaction (Massague, 1998).

Activated R-Smad/Smad4 complexes subsequently translocate into the nucleus. For the R-Smads Smad1 and Smad3, this translocation is mediated by a classical nuclear localization sequence (NLS) present in the MH1 domain (Xiao et al., 2001). Although
this sequence is conserved in all Smads, Smad2 utilizes a different method for nuclear translocation, by directly binding components of the nuclear pore complex (Xu et al., 2002). Smad4 is able to constantly shuttle between the nucleus and cytoplasm, due to the presence of both the aforementioned NLS and a nuclear export signal (NES) in its linker domain (Watanabe et al., 2000). The NES is thought to be masked when Smad4 is in complex with R-Smads, allowing R-Smad/Smad4 complexes to accumulate in the nucleus (Derynck and Zhang, 2003). Once in the nucleus, Smads modulate transcription through several mechanisms. Smad3 and Smad4 have direct but low-affinity DNA binding activity to consensus Smad-binding elements (SBEs) (Zawel et al., 1998), therefore Smad proteins typically interact with other DNA-binding factors such as FAST-1 (Chen et al., 1996), TFE3 (Hua et al., 1998) and the p300 and P/CAF families (Janknecht et al., 1998; Itoh et al., 2000). The transactivation functions of Smads are conducted through the MH2 domain (Moustakas et al., 2001), but the interaction of Smad4 with the p300/CBP co-activators requires approximately 40 amino acids of the linker domain known as the Smad-activation domain, or SAD (de Caestecker et al., 2000a)

**TGF-β and Smads in human disease**

As with other signal transduction pathways, aberrant TGF-β signaling can lead to a number of diseases, most notably cancer. Disease-related mutations have been identified at every level of TGF-β signaling, including the TGF-β ligands, type I and II receptor kinases, cytoplasmic and nuclear co-factors, and the Smad proteins (Miyazono, et al., 2001). Deregulation of TGF-β signaling has been implicated in autoimmune
disorders, fibrosis, and vascular diseases such as atherosclerosis, but the majority of pathogenic mutations within the TGF-β signaling pathway have been identified in the context of human cancers (Dennler et al., 2002). The typical response of epithelial cells to TGF-β is growth inhibition, and when this response is lost due to somatic mutations in components of the TGF-β signaling pathway, unregulated cell growth and consequent tumor formation can result (Derynck et al., 2001).

Mutations in the Smad proteins have been found in many cancers (Hata et al., 1998b). As mentioned previously, Smad4 was originally identified as a possible tumor suppressor gene on chromosome 18q21, which is frequently deleted or otherwise mutated in pancreatic carcinomas, and was therefore originally named DPC4 (deleted in pancreatic carcinoma locus 4) (Hahn et al., 1996). Mutations in Smad4 are associated with approximately 50% of pancreatic cancers and 30% of colorectal cancers (Cohen, 2002; Takagi et al., 1996; Thiagalingam et al., 1996). Smad4 has also been found to be the target of inactivating mutations, albeit less frequently, in breast (Xie et al., 2002), ovarian (Schutte et al., 1996), lung (Uchida et al., 1996), prostate (MacGrogan et al., 1997), and skin cancers (Kim et al., 1996; Xie et al., 2003). Smad2 is also coded by a gene located on chromosome 18q21, and has likewise been found to be mutated in colorectal cancer (Eppert et al., 1996). Smad3, despite having an overall 92% sequence homology with Smad2, has not been found to be mutated in human cancer (Massague et al., 2000). This contrasts with the finding that Smad3-null mice become moribund with colorectal tumors (Zhu et al., 1998). Other results have shown that several pathogenic mechanisms utilizing TGF-β signaling require Smad3 (Roberts et al., 2003), leading to
the proposal that while Smad2 and Smad4 may have tumor suppressive functions, Smad3 is likely a mediator of oncogenic TGF-β signaling (de Caestecker et al., 2000).

Inactivation of Smad2 and Smad4 can occur through gene deletion or frameshift, nonsense, or missense mutations. Study of missense mutations in Smads have been particularly fruitful in yielding information about the mechanism of Smad-dependent signaling (Shi et al., 1997; Hata et al., 1998b). The majority of tumor-derived missense mutations in Smads map to the MH2 (oligomerization) domain, with only a small number mapping to the MH1 (DNA-binding) domain (Figure 1-4, Massague et al., 2000). The most significant tumor-derived MH1 domain mutation is the mutation of a conserved arginine, Smad2 (R133C) in colon carcinoma and the equivalent Smad4 (R100T) in pancreatic carcinoma (Eppert et al., 1996; Schutte et al., 1996). These mutations increase the affinity of the MH1 domain for the MH2 domain by 18 to 22-fold, possibly locking the MH2 domain in an inhibited state (Hata et al., 1997). Mutation of this arginine also causes an increase in the ubiquitin-mediated degradation of Smad2 or Smad4 (Xu and Attisano, 2000).

**Smad complex stoichiometry and mechanism of formation**

The solution of the crystal structure of the MH2 domain of Smad4 provided some of the first insights into the mechanism of formation of the active Smad complex (Shi et al., 1997). As mentioned, an overwhelming majority of tumor-derived missense mutations are found in the MH2 domain, suggesting that the loss of the functions of this domain are somehow tumorigenic (Hata et al., 1998b). The Smad4 MH2 domain was found to crystallize as a trimer (Figure 1-1B). Each subunit of the trimer consists of a...
central β strand sandwich, flanked on one end by three helices (three-helix bundle region) and on the other end by three loops and another helix (loop-helix region) (Figure 1-1A). In the overall structure, each interface of the trimer is defined by a three-helix bundle of one subunit contacting the loop-helix region of an adjacent subunit. Tumor-derived missense mutations map to three general areas in the MH2 domain. A small number of mutations map to the central β-strand sandwich (L440R and P445H in Smad2; R441P in Smad4) (Shi et al., 1997), and mutations in this region would be expected to disturb protein folding (Hata et al., 1998b). This was supported by results that showed that some of these proteins containing β sandwich mutations could not be expressed (Eppert et al., 1996).

The majority of tumor-derived mutations in the Smad4 MH2 domain map to the regions mediating interface contact in the crystallographic Smad4 homotrimer. These mutations are found either in the loop-helix region (D351H, R361C, and V370D in Smad4) or in the three-helix bundle region (D450E in Smad2; D493H in Smad4) (Hata et al., 1998b) (Figure 1-3B). It was proposed that trimerization is an essential mechanism in Smad-dependent signaling, and that the presence of mutations at the trimer interface would disrupt signaling. Consistent with this prediction, size exclusion chromatography of purified Smad4 revealed that wild-type Smad4 eluted as a trimer, while several of the tumor-derived Smad4 point mutants eluted as monomers (Shi et al., 1997). A third group of MH2 domain missense mutations map to the L3 loop region (Figure 1-3C). Mutations in the L3 loop have been primarily found as null or developmental defects in Drosophila Mad or C. elegans Sma, although mutation of a conserved L3 loop arginine in Smad2
Figure 1.4. Smad mutations found in cancer. Smad4 is mutated or deleted in approximately 50% of pancreatic carcinomas. Both Smad4 and Smad2 have also been found to be mutated in other cancer types. The majority of tumor-derived mutations map to the MH2 domain. Reproduced from Hata et al., 1998.
(R428S) has been discovered in cervical cancer (Sekelsky et al., 1995; Savage et al., 1996; Maliekal et al., 2003). In the context of the Smad4 MH2 structure, the L3 loop is located perpendicularly to the conserved trimer interface, and thus does not appear to participate in subunit-subunit interactions. It is especially notable that mutation of the L3 loop disrupted the interaction between Smad2 and Smad4, but did not appear to affect homotrimerization of Smad4 (Shi et al., 1997). To account for the presence of missense mutations at both the conserved trimer interface and the L3 loop, that differentially affected Smad homo- and heterotrimerization, it was proposed that both R-Smads and Smad4 were each able to form homotrimers (following activation by the receptor kinase). The R-Smad and Smad4 homotrimers then could interact via their respective L3 loops to form heterohexamers, which as the active signaling complex, could translocate into the nucleus and direct Smad-dependent transcription (Shi et al., 1997).

Another approach to determining the stoichiometry of active Smad signaling complexes was taken by Miyazono and colleagues, who immunoprecipitated overexpressed, epitope-tagged Smads from mammalian cells, and analyzed these complexes through size exclusion chromatography (Kawabata et al., 1998). These studies indicated that the R-Smad/Smad4 complex was a heterotrimer, although the stoichiometric ratio of R-Smad to Smad4 was unclear. Furthermore, while R-Smads were capable of forming homo-oligomeric complexes following phosphorylation, Smad4 appeared to be incapable of homo-oligomerization, which was contrary to the prior findings of Shi and colleagues.
The stoichiometry of endogenous Smad complexes in TGF-β stimulated cells was also assessed through size exclusion chromatography (Jayaraman and Massague, 2000). These studies showed that the size of the Smad2/Smad4 complex was most consistent with a heterodimer. In addition, Smad4 displayed a tendency to oligomerize in TGF-β stimulated cells, contradicting the findings of Miyazono and colleagues.

Our own approach to addressing the controversial issue of Smad complex stoichiometry was to analyze the solution behavior of purified fragments of R-Smads and Smad4. In these studies, described in Chapter II, we evaluated the stoichiometry of R-Smad homo-oligomeric and R-Smad/Smad4 hetero-oligomeric complexes by analyzing pseudophosphorylated proteins, in which physiological phosphorylation was simulated by mutating the sites of serine phosphorylation to acidic amino acid residues.

**A specific mechanism for phosphorylation-induced activation**

Another important question regarding Smad-dependent signaling is the mechanism of phosphorylation-induced activation of the R-Smads. Although part of the mechanism of phosphorylation-induced activation is clearly through relieving MH1 domain inhibition of the MH2 domain (Hata et al., 1997), phosphorylation is known to act as an energetic driver of oligomerization in many proteins. For example, phosphorylation of the STAT proteins leads to dimer formation through reciprocal interaction between phosphotyrosine residues of one subunit with the SH2 domain of the other (Aaronson and Horvath, 2002). Moreover, the central β-sandwich of the MH2 domain of Smads bears a great deal of structural similarity to the phosphoserine and phosphothreonine-binding FHA domain of Rad53 and several other proteins (Durocher et
al, 2000; Li et al., 2000), as well as to the IAD domain of the IRF proteins (Qin et al., 2003; Takahasi et al., 2003). This suggests the possibility that Smad MH2 domain may be a discrete phosphoserine recognition module. Chapter III describes experiments exploring the specific mechanism of phosphorylation-induced R-Smad/Smad4 complex formation, primarily through crystallography and mutational analysis.

**Preferential formation of the R-Smad/Smad4 complex**

Although R-Smads are capable of forming stable homo-oligomeric complexes following phosphorylation, it is well established that R-Smads preferentially associate with Smad4 to form hetero-oligomeric complexes (Lagna et al., 1996). The structural basis for this preferential formation of the R-Smad/Smad4 hetero-oligomer over the R-Smad homo-oligomer will be addressed through crystallographic and calorimetric analysis of two R-Smad/Smad4 structures, described in Chapter IV of this thesis.
CHAPTER II: ANALYSIS OF R-SMAD/CO-SMAD OLIGOMERIZATION THROUGH PSEUDOPHOSPHORYLATED MUTANTS

INTRODUCTION

As discussed in the general introduction to this thesis, a better understanding of the structural basis for TGF-β signaling through the Smad proteins is necessary. Characterization of the stoichiometry of hetero-oligomeric Smad complexes is of primary importance in elucidating the overall mechanism of TGF-β signaling. Mutations in the Smad proteins are associated with a number of diseases, most notably cancer (Blob et al., 2000; Derynck et al., 2001). Therefore it is also important to understand how the mechanisms of Smad-dependent signaling may be subverted in diseases like cancer.

Prior to the studies described in this thesis, it was well known that phosphorylation induces the R-Smad proteins to form homo-oligomers, and form hetero-oligomers in the presence of Smad4 (Lagna et al., 1996). However, the precise stoichiometry of these oligomeric complexes had not been clearly established. The first investigation into the stoichiometry of the Smad complexes was by Shi and coworkers, who solved the crystal structure of the MH2 domain of Smad4 (Shi et al., 1997). The Smad4 MH2 crystallized as a trimer, and coupled with mutational analysis, led the authors to propose that R-Smad phosphorylation promotes the formation of a heterohexamer comprised of two homotrimers: one phosphorylated R-Smad homotrimer and one Smad4 homotrimer.

Subsequent studies of overexpressed, epitope-tagged Smads in mammalian cells led to the proposal that the activated Smad complex was a heterotrimer assembled in a manner
similar to the Smad4 homotrimer, but of which the ratio of R-Smad to Smad4 was unclear (Kawabata et al., 1998). A third model, where the R-Smad and Smad4 form a heterodimer, was based on size exclusion chromatography of endogenous Smads in mammalian cells (Jayaraman and Massague, 2000).

As an initial approach toward determining the stoichiometry of the heteromeric R-Smad/Smad4 complex, we analyzed purified forms of these proteins through size exclusion chromatography and analytical ultracentrifugation. Full-length versions of Smad3 and Smad4 did not express well in *E. coli*, therefore we used truncated forms containing the MH2 oligomerization domain and part or all of the linker region (Figure 2-1). Smad4AF (Active Fragment) had previously been determined to be the minimal Smad4 fragment capable of directing transcription (de Caestecker et al., 1997). The R-Smads Smad3 (TGF-β/activin) and Smad1 (BMP) were studied to identify possible differences between the mechanisms of Smad-dependent signaling in the TGF-β/activin and BMP signaling pathways. Smad3(LC) and Smad1(LC) contained the entire linker and MH2 domains of Smad1 and Smad3. Attempts to obtain phosphorylated forms of S3LC and S1LC through enzymatic means or peptide ligation were not immediately successful; therefore we attempted to mimic phosphorylation by replacing the sites of serine phosphorylation with either glutamic acid or aspartic acid residues.
MATERIALS AND METHODS

Reagents

EDTA and NaCl were purchased from Sigma; DTT from BioRad; HEPES from Fisher; TCEP from Pierce. All restriction enzymes were purchased from New England Biolabs unless otherwise indicated. Oligonucleotides for PCR were purchased from Sigma-Genosys.

Construction of expression plasmids and mutagenesis

The cDNA-derived fragments for Smad3LC (S3LC, residues 145-425), Smad1LC (S1LC, residues 143-465), and Smad4AF (S4AF, residues 273-552) were generated by PCR using primers containing the appropriate restriction sites. S3LC and S1LC were subcloned into the pGEX-6P1 vector and S4AF into the pGEX-4T2 vector (Amersham Biosciences). Vectors and cDNA inserts were treated with the appropriate restriction enzymes to generate complementary ends. Vectors were further treated with alkaline phosphatase to prevent re-ligation of the empty plasmid. cDNA inserts were ligated into vectors with T4 DNA ligase.

Site-directed mutants were produced through a PCR-based approach, primarily with the QuikChange kit (Stratagene). Primers (~30-40 nucleotides long) were designed to include the desired mutation, as well as a silent restriction site mutation that could be used to rapidly determine the success of the PCR mutagenesis. Following a series of PCR cycles, the original template DNA was destroyed through DpnI treatment. The resulting DpnI-treated PCR product was used to transform competent E. coli cells (strains
HB101 or DH5α), and transformed cells were plated on LB-agar plates containing 200 μg/ml of ampicillin. Plasmids from the resulting colonies were isolated with the QIAprep Miniprep kit (QIAGEN) and screened by restriction digest. Confirmation of the desired mutation was obtained by DNA sequencing.

The S4AF(3E) mutant was generated through cassette mutagenesis, in which the nine C-terminal residues of Smad4 (544-552, PIADPQPLD) were replaced by the eleven C-terminal residues of pseudophosphorylated Smad3 (415-425, GSPSIRCEEVE). PCR mutagenesis was used to engineer a new silent NcoI site between codons 542 and 544. Wild-type S4AF in the pGEX-4T2 vector was treated with NcoI and EcoRI, and the C-terminus of S4AF was excised through purification and extraction from agarose gel (QIAquick Gel Purification kit, QIAGEN). The double-stranded DNA cassette containing the pseudophosphorylated Smad3 C-terminal sequence was inserted and ligated into the plasmid as described above.

**Protein expression and purification**

S3LC, S1LC, S4AF, and the mutant derivatives of these proteins were expressed as glutathione-S-transferase (GST) fusion proteins in *E. coli* (HB101 strain). Cells transformed with the appropriate plasmid were grown at room temperature in tryptone phosphate media to an optical density (600 nm) of approximately 0.5, at which point protein production was induced by adding IPTG to a concentration of 0.05-0.2 μM. Cells were allowed to continue growing at room temperature with constant shaking for 30-36 hours. Cells were pelleted by centrifuging at 5000 rpm for 30 minutes, and resuspended...
in harvesting buffer (50 mM Tris-HCl, 100 mM NaCl, 2 mM EDTA, and 1 μM PMSF) before flash freezing in liquid nitrogen and storing at -80°C until needed.

Cells were sonicated, then centrifuged at 15000 rpm for 45 minutes to obtain a clarified extract. DTT was added to the extract at a concentration of 5 mM, and the extract was combined with glutathione sepharose beads (Amersham). The extract/bead mixture was incubated at 4°C for 3 hours, with constant stirring. Following incubation, the mixture was poured over a fritted column and the extract was eluted, and the remaining glutathione sepharose beads were washed extensively with FPLC buffer (20 mM Tris-HCl, 100 mM NaCl, 0.1 mM EDTA). S3LC, S1LC, and the mutants derived from these proteins were cleaved from the GST tag by adding Prescision Protease and incubating at 12°C for approximately 24 hours. S4AF and the mutants derived from it were cleaved with thrombin, which required a 12-hour incubation at room temperature.

The eluted proteins were purified through ion exchange chromatography, using the anion-exchanger DEAE-sepharose. Following release from the glutathione matrix, proteins were dialyzed (overnight, 4°C) in DEAE buffer (10 mM Tris-HCl, 10 mM NaCl, 0.1 mM EDTA). The pH of the buffer was based on the calculated isoelectric point (pI) of the protein and was generally 7.3 for S3LC and mutants, and 8.2 for S1LC, S4AF, and mutants. The DEAE-sepharose column was equilibrated in an identical buffer prior to loading of the dialyzed protein onto the column. After extensive washing of the loaded column with DEAE buffer, a 10-300 mM NaCl gradient was applied to separate the desired protein from impurities. Protein purity was initially analyzed through SDS-PAGE. Proteins were concentrated with an Amicon YM10 filter, and protein
concentrations were determined by measuring the absorbance, $A$, at 280 nm and using the Beer-Lambert law:

$$A = \varepsilon cl$$

where $\varepsilon$ is the molar absorption coefficient (M$^{-1}$ cm$^{-1}$), $c$ is the protein concentration (M), and $l$ is the path length. $\varepsilon$ was calculated from the amino acid sequence using the following equation from the method described by Gill and von Hippel (Gill and von Hippel, 1989):

$$\varepsilon(280) = (#\text{Trp})(5500) + (#\text{Tyr})(1490) + (#\text{Cys})(125)$$

where $#\text{Trp}$, $#\text{Tyr}$, and $#\text{Cys}$ are the respective numbers of tryptophan, tyrosine, and cysteine residues in the protein. Proteins were flash frozen in liquid nitrogen and stored at -80°C.

Size exclusion chromatography

Size exclusion chromatography (SEC) was performed with the Superdex 200 HR analytical gel filtration column on the Akta Explorer 10 FPLC system (Amersham Biosciences). The FPLC buffer used in the size exclusion chromatography experiments consisted of 20 mM HEPES (pH 7.4), 0.1 mM EDTA, 100 mM NaCl, and 1 mM DTT. Prior to loading onto the column, the protein samples were incubated in buffer containing 1 mM TCEP for at least 60 minutes at room temperature. For most of the protein interaction experiments described herein, the final concentration per protein in solution was approximately 50 $\mu$M (~1.5 mg/ml). FPLC operation and data analysis were done with UNICORN software. The column was calibrated with blue dextran (to determine void volume) and molecular weight standards ovalbumin (43 kDa), albumin (67 kDa),
aldolase (158 kDa), catalase (232 kDa), and ferritin (440 kDa). SEC fractions (0.5 ml) were collected in FPLC buffer at room temperature with a flow rate of 0.7 ml/min. Fractions were immediately mixed with Laemmli buffer and run on 12% SDS-PAGE gels. Proteins were visualized by Coomassie blue or silver staining. Individual bands within the Coomassie-stained SDS-PAGE gels were quantified with the Fluor-S MultiImager and MultiAnalyst software (BioRad).

Transcriptional response assays

Transcriptional response assays were performed by Dr. Mark de Caestecker, and compared the difference in signaling between wild-type (full-length) Smad3 or Smad1 and the pseudophosphorylated versions of each that I had created. MDA-MB 468 or NMuMg cells were seeded to 50% confluence and transfected with SBE-Lux, pSV-β, and full-length forms of wild-type or pseudophosphorylated Flag-Smad3 or Flag-Smad1, and Smad4-Myc, using Fugene-6 (Roche) or Lipofectamine (Gibco). MDA-MB 468 cells are a Smad4-null and TGF-β unresponsive cell line derived from breast carcinoma, and NMuMg cells are mouse mammary epithelial cells that undergo epithelial to mesenchymal transition when stimulated by TGF-β. SBE-Lux is a consensus Smad-binding element (Jonk et al., 1998) linked to a luciferase reporter gene. Cells were lysed after 36 hours, and luciferase and β-galactosidase activity determined as described in de Caestecker et al (1997). Luciferase values were corrected for transfection efficiency with β-galactosidase, expressed as the means of three independent transfections. Experiments were repeated at least twice, with similar results. Protein expression levels were determined in parallel experiments using COS-1 cells transfected with identical
proportions of FLAG-Smad3 and Smad4-Myc used in the signaling assays, and cell lysates were immunoblotted with anti-FLAG M2 (Kodak) or anti-Myc (9E10) antibodies.

**Analytical ultracentrifugation**

Sedimentation velocity and equilibrium experiments were conducted on a Beckman Optima XLA analytical ultracentrifuge by Dr. John Correia (Correia et al., 2001). Samples were dialyzed against sedimentation buffer comprised of 20 mM HEPES (pH 7.4), 0.1 mM EDTA, 100 mM NaCl, and 0.1 mM DTT. All experiments were repeated in the presence of 2 mM TCEP to remove nonspecific aggregation. Sedimentation velocity experiments were performed at 42000 rpm, and velocity data analyzed with DCDT+ (Philo, 2000; Stafford, 1992). The weight average sedimentation coefficient values were fit to both a monomer-trimer and a monomer-dimer-trimer model with Fitall (MTR Software, Toronto, Canada) as described (Correia, 2000). Based upon similar size and sequence, monomers of S3LC, S3LC(3E), SILC, and SILC(2D) were assumed to have similar s values to that of S4AF, 2.46 S_{20,w} (2.65 S_{app}).

Sedimentation equilibrium experiments were performed at 28K rpm for S4AF, 20K and 22K for S3LC, 20K for S3LC(3E) and S1LC(2D), 16K and 20K for S1LCS, 22K for S1LCS(2D), and 24K for S4AF(3E). Data sets were fit with NONLIN to an appropriate association scheme as described in detail elsewhere (Johnson et al., 1981).
RESULTS

R-Smads and Smad4 display different oligomerization tendencies

Our approach to understanding the molecular basis for Smad oligomerization was to analyze the behavior of the purified proteins in solution. Previous reports have shown through yeast two-hybrid and co-immunoprecipitation analyses that the C-terminal Smad MH2 domain is responsible for mediating both homomeric and heteromeric interactions (Wu et al., 1997). Since the full-length forms of Smad3 and Smad4 formed irreversible aggregates when expressed as GST-fusions in *E. coli*, we used truncated proteins from which the N-terminal MH1 domain was deleted (Figure 2-1). The S3LC construct contained the entire linker and MH2 domains of Smad3 (residues 145-425). S1LC included the identical domain range (residues 143-465), however an insertion in the linker domain made this construct slightly larger than S3LC. The S4AF construct contained the MH2 domain and the 45 C-terminal residues of the linker domain (referred to as the Smad activation domain, or SAD). This construct had previously been defined as the minimal fragment of Smad4 required for transcriptional activity (de Caestecker et al., 1997), and its crystal structure had been previously solved in our laboratory by Dr. Bin Qin (Qin et al., 1999).

Oligomerization of S3LC and S4AF was initially analyzed by size exclusion chromatography (Figure 2-2). Both proteins were applied to the gel filtration column at 4 different concentrations (6, 17, 50, and 150 μM). S3LC elutes as an apparent monomer at the lowest concentration, but displays a concentration dependent tendency toward oligomerization, with a plateau approaching a trimer. S3LC also displayed a tendency to
Figure 2-1. Diagram of the S3LC, S1LC, and S4AF constructs and a domain comparison between Smad3, Smad1, and Smad4. Sites of serine phosphorylation are marked with (*). The constructs used are indicated by black lines underneath the depictions of the full-length forms of the proteins.
form high molecular weight aggregates, a process that was reversed by incubation of the sample with 1 mM TCEP (data not shown). In contrast, S4AF behaves as an apparent monomer, even at the highest concentration tested. The data appear to indicate that R-Smads possess an intrinsic propensity for oligomerization, while Smad4 does not, despite considerable sequence similarity.

The size exclusion chromatography results were verified through sedimentation analysis of the S3LC, S1LC, and S4AF constructs. (Specific experimental details of the sedimentation experiments can be found in Correia et al., 2001.) Sedimentation velocity analysis of S3LC revealed that the sedimentation coefficient patterns, or g(s), shifted to higher values with increasing concentration, indicative of a reversible self-association process (Figure 2-3B). A similar pattern was observed in sedimentation velocity analysis of S1LC (Figure 2-4, top panel). However, as observed in the SEC experiments described above, S3LC displayed a tendency to form high molecular oligomers during sedimentation, which appeared to be reversible by the addition of TCEP. This reversible self-association of S3LC was confirmed by sedimentation equilibrium analysis (Correia et al., 2001). In both the presence and absence of 2 mM TCEP, the sedimentation equilibrium data are best fit by a monomer-trimer model with a global fit producing a trimerization constant (K₃) of (3.09-2.99) x 10⁹ M⁻² (Table 2-1). Fitting to both n and K yielded a value of n=2.96 (+TCEP) or 2.98 (-TCEP), confirming that the trimer is the predominant species in both fits. Including a dimer term, K₂, did not produce an improvement in the fits, indicating that trimer formation is cooperatively favored over dimer formation.
Figure 2-2. Size exclusion chromatography elution profiles of S3LC (solid line) and S4AF (dashed line) at four different protein concentrations (6, 17, 50, and 150 μM). The y-axis (mAU) plots absorption at 280 nm. The calculated subunit molecular weights for S3LC and S4AF are 31565 and 30884 kDa, respectively.
As expected from the SEC data, S4AF displays no shift with increasing concentration and is consistent with a non-interacting, monomeric species in solution (Figure 2-3A). The mean weight average sedimentation coefficient ($S_{20,w}$) for S4AF as calculated from DCDT+ is 2.45 ($\pm$0.04) S (Figure 2-5). The $S_{20,w}$ values obtained from DCDT+ are calculated by a weighted integration covered by the entire range of sedimentation coefficients covered by the $g(s)$ distribution and corrected for the solvent density and viscosity. The average molecular weight from single-species fits was 26905 ($\pm$2197) kDa. An alternate data fitting method using the program SVEDBERG produced similar results. Results for S4AF did not vary significantly $\pm$2 mM TCEP. These results were supported by sedimentation equilibrium runs (Figure 2-6), which produced global best fits of the data consistent with a monomeric species in solution (average MW from equilibrium runs $= 30018 \pm 597$; expected MW $= 30844$).

The ability of unphosphorylated S3LC to interact with S4AF was also analyzed by analytical ultracentrifugation by titrating S4AF into a fixed amount of S3LC. Rather than producing a sum of the S4AF monomer zone and the S3LC interacting zone, the boundary shifts to higher $s$ values, indicative of the formation of a hetero-oligomer between the two proteins (Figure 2-3C). Adding increasing amounts of S4AF causes the associating zone to grow in area (concentration), with the peak consistent with trimer formation. Eventually the addition of increasing S4AF leads to the emergence of a peak near 2.65 $S_{app}$, corresponding to excess monomeric S4AF unable to complex with S3LC. This data is converted to a weight average $S_{20,w}$ value and plotted in Figure 2-5.

Comparing the S3LC (open and closed squares) against the S3LC + S4AF data (-X-X-
Figure 2-3. Sedimentation velocity analysis of S4AF, S3LC, and mixtures of the two Smad constructs. (A) g(s) profiles from sedimentation velocity runs with S4AF; (---) minus TCEP; (+++) plus TCEP. The apparent weight average s value of these data is 2.67 S, corresponding to the left vertical line, and consistent with a noninteracting monomer. (B) g(s) profiles from sedimentation velocity runs with S3LC + 2 mM TCEP. The shift with increasing concentration is consistent with a reversible monomer-trimer self-association with pure trimer corresponding to the right vertical line. (C) g(s) profiles from sedimentation velocity runs with mixtures of a fixed amount of S3LC (7μM) and increasing amounts of S4AF. The bold curve (+++) is S3LC in the absence of S4AF. The addition of S4AF immediately causes an overall shift to higher s values. When there is more S4AF in solution than is able to complex with S3LC, excess S4AF runs as monomer.
Figure 2-4. Sedimentation velocity analysis of S1LC (top panel, 6 to 49 μM) and S1LC(2D) (bottom panel, 2 to 26 μM) and presented as g(s*20,w) vs. s*20,w plots. The data were converted to weight average S20,w values and best fit to a monomer-dimer-trimer model. The fits are shown in the insert, where the weight average S20,w values of S1LC and S1LC(2D) at different concentrations are shown by triangles and squares, respectively. The monomer S value (dashed vertical line in the upper panel) was determined to be 2.72 s20,w by extrapolation of the S1LC data. The trimer S value (dashed vertical line in lower panel) of 5.67 s20,w was determined as described in Correia et al., 2001. These values were used in the fitting and are in agreement with results from sedimentation equilibrium experiments (Table I). Replacing the two C-terminal serines with aspartic acid residues results in a considerable enhancement of trimerization, as the overall trimerization constant K3 increases from 1.11 X 10^7 M^-2 to 2.29 X 10^9 M^-2.
pattern) appears to show that the mixed sample has a larger average size, indicating that the heteromeric S3LC/S4AF complex is tighter (equilibrium is shifted more toward oligomer) than the self-associating S3LC homomeric complex. It is possible that the shape of the heteromeric complex causes the increase, but given the similar size and sequence of the S3LC and S4AF constructs, heteromeric complex should be hydrodynamically similar to the S3LC homomer (Qin et al., 1999). The reason for the decrease in $s_{20,w}$ at higher protein concentrations is due to the contribution of excess S4AF to the average molecular weight (Philo, 2000).

**Phosphorylation promotes R-Smad homotrimerization**

Since phosphorylation of the R-Smads results in the activation of these proteins, we sought to analyze differences in solution behavior between the phosphorylated and unphosphorylated forms of the R-Smads. However, since we were initially unable to produce phosphorylated forms of the R-Smads, we instead replaced the C-terminal serines which are the putative sites of phosphorylation with acidic amino acid residues, either glutamic acid or aspartic acid, in an attempt to approximate phosphorylation. This approach has been previously used by several others to mimic phosphorylation (Thorsness and Koshland, 1987; Maciejewski et al., 1995; Pullen et al., 1995). It had also been previously shown that replacement of the three C-terminal serines of Smad3 with aspartic acid resulted in a constitutively active Smad3 capable of mediating *in vivo* TGF-β responses in the absence of TGF-β stimulation (Liu et al., 1997; Funaba and Mathews, 1997). Although it had been shown that the two C-terminal R-Smad serines are phosphorylated (Abdollah et al., 1997; Souchelnytskyi et al., 1997), the complete
<table>
<thead>
<tr>
<th>Construct</th>
<th>+/- TCEP</th>
<th>$K_3$ (x $10^9$ M$^{-2}$)</th>
<th>$\text{rms}^a$</th>
<th>Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>S3LC</td>
<td>-</td>
<td>3.090 {2.597, 3.691}</td>
<td>0.0056</td>
<td>equilibrium</td>
</tr>
<tr>
<td></td>
<td>+</td>
<td>2.991 {2.588, 3.482}</td>
<td>0.0067</td>
<td>equilibrium</td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>2.097</td>
<td>0.1240</td>
<td>$s_{20,w}$ (1-3)</td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>1.478 ± 0.326</td>
<td>0.0117</td>
<td>SEDFIT (1-3)$^b$</td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>2.109</td>
<td>0.1265</td>
<td>$s_{20,w}$ (1-2-3)</td>
</tr>
<tr>
<td></td>
<td>+</td>
<td>1.150</td>
<td>0.0437</td>
<td>$s_{20,w}$ (1-3)</td>
</tr>
<tr>
<td></td>
<td>+</td>
<td>1.150</td>
<td>0.0465</td>
<td>$s_{20,w}$ (1-2-3)</td>
</tr>
<tr>
<td>S3LC(3E)</td>
<td>-</td>
<td>54.76 {42.99, 70.90}</td>
<td>0.0062</td>
<td>equilibrium</td>
</tr>
<tr>
<td></td>
<td>+</td>
<td>51.94 {40.71, 67.60}</td>
<td>0.0058</td>
<td>equilibrium</td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>74.33</td>
<td>0.0686</td>
<td>$s_{20,w}$ (1-3)</td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>92.04</td>
<td>0.0519</td>
<td>$s_{20,w}$ (1-2-3)</td>
</tr>
<tr>
<td></td>
<td>+</td>
<td>36.33</td>
<td>0.0908</td>
<td>$s_{20,w}$ (1-3)</td>
</tr>
<tr>
<td></td>
<td>+</td>
<td>52.29</td>
<td>0.0356</td>
<td>$s_{20,w}$ (1-2-3)</td>
</tr>
</tbody>
</table>

Table 2-1. Trimerization constants for S3LC and S3LC(3E) as determined through either sedimentation equilibrium or sedimentation velocity ($s_{20,w}$) methods. $^a$rms deviation of the fit; in units of AOD for equilibrium and SEDFIT analysis; in units of AS for weight average fitting. $^b$Average values (10 data sets with S3LC – TCEP) from single-experiment fits to a monomer-trimer model.
Figure 2-5. Weight average $s_{20,w}$ values derived from g(s) analysis of sedimentation velocity runs with S4AF (open and closed circles), S4AF(3E) (closed diamonds), S3LC (open and closed squares), and S3LC(3E) (open and closed triangles). The closed symbols represent data collected in the presence of 2 mM TCEP. S3LC and S3LC(3E) data were best fit to either a monomer-trimer (dotted line) or a cooperative monomer-dimer-trimer model (dashed line) (see Table 2-1). Experiments performed using mixtures of S4AF + S3LC (-x-x-x) or S4AF + S3LC(3E) (---+) contained TCEP. The initial addition of S4AF to S3LC or S3LC(3E) causes a shift to more trimer and a larger $s_{20,w}$ value, indicating that the heterotrimeric interaction is tighter than the homotrimeric interaction.
Sedimentation equilibrium runs of S4AF (28 K, 24.7°C, no TCEP). Loading concentrations were 2, 4, 6, 8, 10, and 12 μM. (A) Equilibrium data and best fits. (B) Distribution of residuals. The best global fit of the data (2 runs) is consistent with a monomeric species in solution giving a measured molecular weight of 30018 ± 597 (expected MW = 30844), with an average rms deviation of 0.00414. Experiments in the presence of TCEP at lower speeds are also consistent with S4AF behaving as a monomer in solution.
conservation of all three serines of the SSXS motif among the R-Smads suggested the possibility that all three serines could be phosphorylated. Also, given that glutamic acid is less charged than phosphoserine, mutating all three serines to glutamic acid might more accurately approximate the charge of the physiologically phosphorylated R-Smad C-terminus, even if only two serines are actually phosphorylated. Therefore, we created two versions of pseudophosphorylated S3LC. S3LC(2E) has the last two serines substituted with glutamic acid, S3LC(3E) has all three serines substituted. (Subsequent studies have shown that phosphorylation of the N-terminal-most of the three serines is energetically destabilizing.) Both the 2E and 3E versions of S3LC behaved similarly in cell signaling assays and in in vitro biochemical analyses. However, most of the S3LC analyses were performed with S3LC(3E), which was more resistant to time-dependent aggregation than S3LC(2E), although we found that the addition of TCEP helped prevent aggregation of these proteins. A pseudophosphorylated form of Smad1 was also created, S1LC(2D), in which the C-terminal serine phosphorylation sites were mutated to aspartic acid residues.

To verify that the pseudophosphorylated R-Smads were suitable mimics of actual phosphorylation, signaling assays were performed to analyze the ability of the pseudophosphorylated mutants to drive Smad-dependent transcription from a consensus Smad binding element. As shown in Figure 2-7A, full-length Smad3(3E) was able to drive transcription with approximately three times the activity of its unphosphorylated counterpart. Smad1(2D) also displayed a similar ability to activate transcription over
wild-type protein (Figure 2-7B), confirming that these pseudophosphorylated mutants are indeed a suitable mimic of phosphorylation.

We further analyzed the solution behavior of the pseudophosphorylated mutants through size exclusion chromatography. S3LC(3E) eluted as a larger molecular weight species than the wild type protein (Figure 2-8), indicative of oligomerization. An identical result was observed with S1LC(2D) and wild-type S1LC (Figure 2-15B, top panels). These results confirm that the pseudophosphorylated R-Smads effectively mimic the homo-oligomerization ability of the phosphorylated proteins.

The SEC results were again verified by sedimentation experiments. Sedimentation velocity data for S3LC(3E) in the presence of TCEP is shown in Figure 2-9A. The g(s) patterns shift to higher values with increasing concentration, indicative of a reversible self-association. The curves are shifted more to trimer formation than the S3LC data (compare Figure 2-3B). The data (±TCEP) were again converted to $s_{20,w}$ values and fit to a monomer-trimer and a cooperative monomer-dimer-trimer model (Table 2-1), and best fits are also shown in Figure 2-5. The best monomer-trimer fits (±TCEP) are consistent with an overall equilibrium constant of $(3.63-7.43) \times 10^{10} \text{ M}^{-2}$, approximately 32 to 35-fold more than the equivalent data for wild-type S3LC (Table 2-1). These results were verified by sedimentation equilibrium experiments (data not shown, see Correia et al., 2001). The best fit is to a monomer-trimer model with $K_3$ varying from $5.19 \times 10^{10}$ to $5.48 \times 10^{10} \text{ M}^{-2}$ (Table 2-1). This agrees well with the weight average data and corresponds to an approximately 17-fold enhancement of trimerization.
Figure 2-7A. The 3E mutant of Smad3 greatly increases Smad3/Smad4 dependent transcription. NMuMg cells were transfected with a consensus Smad-binding element linked to a luciferase reporter gene (SBE-Lux) along with full-length wild-type (WT) Smad3 or the Smad3(3E) mutant with or with full-length WT Smad4. The basal activity observed in the absence of Smad4 was due to the NMuMg cells expressing endogenous Smad2/3 and Smad4, as determined through Western blot (data not shown).
Figure 2-7B. The 2D mutation greatly increases Smad1/Smad4-dependent transcription. NMuMg cells were transfected with SBE-Lux along with full-length wild-type (WT) Smad1 or the Smad1(2D) mutant, in the presence or absence of full-length WT Smad4. Both the Smad1 and Smad4 constructs had similar expression levels. (Bottom) Lysates from NMuMg cells transfected with Flag-tagged Smad1 and Smad4-Myc were separated by SDS-PAGE and immunoblotted using anti-Flag M2 (Sigma) or anti-Myc 9E10 monoclonal antibodies, as indicated.
relative to the wild-type S3LC sedimentation equilibrium data. From the plot of $s_{20,w}$ vs. protein concentration for S3LC(3E) (open and closed triangles in Figure 2-5), it can be seen that trimer formation is 50% complete at 3.3 to 5.2 µM for S3LC(3E), as opposed to 18-29.5 µM for wild-type S3LC (Figure 2-5, open and closed squares). Therefore pseudophosphorylated S3LC(3E) displays a 17 to 35-fold greater tendency to homotrimerize over unphosphorylated S3LC.

The Smad1 constructs S1LC and S1LC(2D) were also examined by sedimentation analysis. These constructs behaved identically to S3LC and S3LC(3E), with the pseudophosphorylated S1LC(2D) displaying a 200-fold greater tendency to homotrimerize over S1LC (Figure 2-4 [bottom]; Table 2-2). A smaller fragment of Smad1, S1LCS, and its pseudophosphorylated mutant S1LCS(2D) were also studied. S1LCS consists of residues 217 to 465 of Smad1, compared to 143 to 465 for S1LC. From sedimentation velocity results, S1LCS(2D) displayed an approximately 4000-fold greater ability to homotrimerize vs. S1LCS (Table 2-2). This result suggests that the N-terminal part of the linker domain may have a role in inhibiting R-Smad trimerization, as has been suggested previously (Hata et al., 1998). These results were also supported by sedimentation equilibrium studies.

**Attachment of a pseudophosphorylated tail to Smad4**

Despite high sequence homology with the R-Smads (Figure 2-14), Smad4 clearly lacks the ability to homo-oligomerize. Since the most apparent structural difference between Smad4 and the R-Smads is the presence of the C-terminal phosphorylation sequence in R-Smads, we examined whether the lack of the phosphorylation site could
Figure 2-8. Size exclusion chromatography elution profiles of S3LC and S3LC(3E), loaded at 50 μM each. The y-axis (mAU) plots absorption at 280 nm. Coomassie-stained SDS-PAGE gels of the eluted fractions are shown.
Figure 2-9. Sedimentation velocity analysis of S3LC(3E) and mixtures of S4AF and S3LC(3E). (A) g(s) patterns from S3LC(3E) in the presence of TCEP. (B) g(s) patterns from a fixed amount of S3LC(3E) (+++, 7 μM) and increasing amounts of S4AF. The $s_{20,w}$ values of the data in these panels are plotted in Figure 2-5. Similar results were obtained in the absence of TCEP.
<table>
<thead>
<tr>
<th>Construct</th>
<th>$1/K_3$ (µM)$^a$</th>
<th>$K_3$ ($x 10^8$ M$^{-2}$)</th>
<th>rms$^b$</th>
<th>Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1LC</td>
<td>322</td>
<td>0.97</td>
<td>0.0093</td>
<td>equilibrium</td>
</tr>
<tr>
<td>300</td>
<td>1.11</td>
<td>0.0049</td>
<td>velocity</td>
<td></td>
</tr>
<tr>
<td>S1LC(2D)</td>
<td>25.6</td>
<td>152</td>
<td>0.0065</td>
<td>equilibrium</td>
</tr>
<tr>
<td>20.9</td>
<td>229</td>
<td>0.0242</td>
<td>velocity</td>
<td></td>
</tr>
<tr>
<td>S1LCS</td>
<td>481</td>
<td>0.43</td>
<td>0.0068</td>
<td>equilibrium</td>
</tr>
<tr>
<td>470</td>
<td>0.45</td>
<td>0.0081</td>
<td>velocity</td>
<td></td>
</tr>
<tr>
<td>S1LCS(2D)</td>
<td>6.4</td>
<td>2410</td>
<td>0.0077</td>
<td>equilibrium</td>
</tr>
<tr>
<td>6.8</td>
<td>2140</td>
<td>0.0321</td>
<td>velocity</td>
<td></td>
</tr>
</tbody>
</table>

Table 2-2. Summary of analytical ultracentrifuge studies on Smad1 homomeric association. $^a$Represents the midpoint of the trimerization reaction in units of protein concentration. $^b$K$_3$ is the overall trimerization constant. $^c$Root mean square deviation of the fit; in units of ΔOD for equilibrium; in units of ΔS for weight average fitting of velocity data. In all other cases, a 1-2-3 model best fits the data; however, for equilibrium experiments with S1LC and S1LCS, a 1-3 model best fits the data. The extent of association in these two cases is only approximately 10%, thus making an accurate determination of K$_3$ from a 1-2-3 model difficult.
solely explain the inability of Smad4 to homotrimerize, in contrast to the R-Smads. To address this question, we engineered a mutant form of S4AF in which the 9 C-terminal residues were replaced by the 11 C-terminal residues of S3LC(3E). This "pseudophosphorylated" Smad4 was designated S4AF(3E). Size exclusion chromatography revealed that this protein was monomeric in the presence of TCEP, but displayed an increased tendency toward nonspecific aggregation in the absence of TCEP (data not shown). This result was verified by sedimentation analysis. In the presence of TCEP, sedimentation velocity and equilibrium data were consistent with a homogeneous monomer, resulting in a weight average s value of 2.7S (the closed diamonds in Figure 2-5 and the left vertical line in Figure 2-10; Correia et al., 2001). Therefore it appears that the phosphorylated tail is not sufficient for mediating Smad homo-oligomerization, and that other structural differences between R-Smads and Smad4 must be present that allow R-Smads to homotrimerize upon phosphorylation. One such difference could be the unique insertion found between helix 3 and helix 4 of Smad4, which is not present in R-Smads. Along with the lack of a C-terminal phosphorylated tail, this Smad4-specific insertion might preclude Smad4 homomer formation.

Stoichiometry of the R-Smad/Smad4 heteromeric complex

The stoichiometry of the activated R-Smad/Smad4 complex that translocates into the nucleus and activates transcription of Smad-specific genes is unclear. As stated in the introduction, it has been suggested that this complex is a heterohexamer (Shi et al., 1997), a heterotrimer (Kawabata et al., 1998), or a heterodimer (Jayaraman and Massague, 2000). Our approach to this question was to analyze complexes of pseudophosphorylated
Figure 2-10. (A) Sedimentation velocity g(s) data for S4AF(3E) at 42K, 24.7°C, in the absence (solid lines) and presence (+++ of 2 mM TCEP. The average s_{app} values are 3.04 and 2.70 S, respectively, as indicated by the two vertical lines. These data are consistent with a mixture of monomers and cross-linked dimers in the absence of TCEP, and homogeneous monomers in the presence of TCEP. The s_{20,w} values of the +TCEP data are plotted in Figure 2-5. (B) An example of a two-species Gaussian fit (solid line) performed with DCDT+ demonstrating the quality of the fit with a monomer and a cross-linked dimer. The two species are represented by the dotted lines, while the actual data (every 5th point) are plotted as open circles.
R-Smads with Smad4 through size exclusion chromatography and analytical ultracentrifugation. The interaction of S3LC(3E) and S1LC(2D) with Smad4 were both analyzed in this manner. S4AF directly interacts with S3LC(3E) as evidenced by the co-elution of these proteins on the SEC (Figure 2-11A). When mixed at a 1:1 mole ratio, part of the S4AF protein co-elutes in the same fractions as S3LC(3E). However, a significant amount of S4AF remains uncomplexed and elutes as monomer on the SEC, indicating that the heteromeric Smad complex is not a 1 to 1 dimer. The position of the elution peak of the S3LC(3E)/S4AF complex is coincident with that of the homotrimeric S3LC(3E), suggesting that this heteromeric complex is a trimer as well. Furthermore, the height and position of the heteromeric elution peak remained unchanged when either a four-fold more or three-fold less S4AF was combined with S3LC(3E) (Figure 2-11A, 4th and 5th panels), an indication that Smad3 is the limiting reagent in heteromeric complex formation. In addition, we determined through densitometry of the bands in the Coomassie-stained gels that the amount of S3LC(3E) in the fractions containing complex was almost exactly twice the amount of S4AF, regardless of the amount of S4AF in solution. A similar analysis was applied to the S1LC(2D)/S4AF complex, which produced identical results to the S3LC(3E)/S4AF analysis, indicating a conserved mechanism for all R-Smads (Figure 2-11B). These data strongly suggest that the R-Smad/Smad4 heteromeric complex is a heterotrimer with preferred stoichiometry of two subunits of phosphorylated R-Smad, and one subunit of Smad4. This 2:1 stoichiometry of the heteromeric Smad complex was confirmed through sedimentation velocity analysis.
Figure 2-9B shows the g(s) patterns for mixtures of fixed amounts of S3LC(3E) with increasing amounts of S4AF, and the same experiment repeated with S1LCS(2D) and S4AF is shown in Figure 2-12. As in the equivalent S3LC experiment (Figure 2-3C), the addition of S4AF does not produce a sum of g(s) patterns, but instead causes a shift to a larger extent of association. The boundary, represented by the g(s*), shifts to a faster sedimenting form that with increasing addition of S4AF approaches the sedimentation coefficient of trimer. The trimer zone grows in area (concentration) first, consistent with heterotrimer formation, then addition of increasing amounts of S4AF eventually leads to the emergence of the monomeric S4AF peak near 2.65 S or 2.724 S, as the ratio of S4AF to S1LCS(2D) becomes greater than 1 to 2. From the plot of weight average as a function of protein concentration (Figure 2-5), the S3LC(3E)/S4AF sample initially has a larger average size at the same total protein concentration as S3LC(3E) alone, indicating that the heterotrimer is a tighter complex than the homotrimer, i.e., equilibrium is shifted more toward the oligomeric state. Not surprisingly, as excess monomeric S4AF accumulates, the weight average value decreases. The significant shift in boundary upon addition of S4AF indicates that the formation of heterotrimer is tighter than the formation of pseudophosphorylated homotrimer (Figure 2-9B and Figure 2-12).

Role of the conserved trimer interface

The subunit packing arrangement in the previously solved crystal structures of the Smad4 MH2 domain and S4AF (Shi et al., 1997; Qin et al., 1999) appears to implicate trimer formation as an important mechanism in Smad-dependent signaling. Although Smad4 is monomeric in solution, the subunit packing interfaces of the crystallographic
Figure 2-11. The R-Smad/Smad4 heteromeric complex is a trimer consisting of two R-Smad subunits and one Smad4 subunit. Size exclusion chromatography runs of (A) S3LC(3E) and S4AF and (B) S1LC(2D) and S4AF were performed, with varying S4AF concentrations. Eluted fractions were analyzed through SDS-PAGE. The “total” relative mole ratio refers to the ratio of S4AF to S3LC(3E) or S1LC(2D) in solution, prior to loading on the size exclusion column. The “complex” relative mole ratio refers to the ratio of S4AF to the respective R-Smad in the elution peak of the complex (fractions 15 and 16 for S3LC[3E] or fractions 16 and 17 for S1LC[2D]), as determined from the analysis of the Coomassie-stained bands with the Fluor-S MultiImager and MultiAnalyst software (Bio-Rad). Standard deviations were obtained from multiple measurements of background at different regions of the gels.
Figure 2-12. Sedimentation velocity analysis of S1LCS(2D)/S4AF. S1LCS(2D) (X line, 9.77 μM) was titrated with increasing amounts of S4AF (solid lines, 2.59, 5.79, 10.16, 15.15, and 20.34 μM). Experiment was performed at 24.7°C, 42000 rpm, + 2 mM TCEP, and presented as a plot of g(s*20,w) versus S*20,w to demonstrate the formation of heterotrimer. The vertical dashed line corresponds to trimer (4.682 S20,w). At higher S4AF concentrations, the excess S4AF sediments as monomer, indicated by the vertical dotted line at 2.724 S20,w.
trimer include highly conserved residues to which the majority of tumor-derived missense mutations of S4AF map (Figure 2-13). The MH2 domains of R-Smads and Smad4 share approximately 50% homology, the major difference being the presence of a 35-amino acid insertion between helix 3 and helix 4, found only in Smad4 (Figure 2-14).

Based on the high sequence homology between the R-Smads and Smad4, we would expect these proteins to be structurally similar, and therefore expect both R-Smad/Smad4 heterotrimerization and R-Smad homotrimerization to be mediated by a similar crystallographically conserved trimer interface. To test the idea that heterotrimerization is mediated by a conserved interface, interface residues of S4AF were mutated to their tumorigenic counterparts and the mutants tested for their ability to interact with S3LC(3E) or S1LC(2D). S4AF(D351H) is a mutation of an aspartic acid residue in the B3 strand of the loop-helix region, which abolishes a key intrasubunit salt bridge. S4AF(D537E) is a mutation of an aspartic acid residue in the H5 helix of the three-helix bundle subdomain. This conservative mutation evidently causes the loss of subunit interaction by creating unfavorable packing at the interface. Both residues are completely conserved in all R-Smads and Smad4, even in Drosophila and C. elegans Smads. Using the identical size exclusion chromatography-based interaction assay described earlier, we found that neither mutant was able to interact with S3LC(3E) or S1LC(2D), as the S4AF mutants did not co-elute with the pseudophosphorylated proteins (Figure 2-15). That the heteromeric Smad interaction is dependent on Asp351 and Asp537 of Smad4 supports the idea that the R-Smad/Smad4 heterotrimeric interaction is mediated through an interface identical to that seen in the crystal structure of Smad4.
Figure 2-13. Crystal structure of S4AF showing the homotrimer in the asymmetric unit (as solved by Qin et al., 1999). The three subunits are colored green, blue, and red. Side chains involved in the trimer interface contacts are shown in black. Corresponding residues in Smad3 are given in parentheses. D351 does not directly contact the neighboring subunit, but forms part of a hydrogen bond network with R361 and D537, which directly link the subunits. Features that will be discussed in more detail in Chapter III are the sulfate-binding sites (boxed), and the L3 loops (depicted in yellow).
Figure 2.14. Structure-based sequence alignment of the MH2 domains of the Smad proteins. The secondary structural motifs are shown below the alignment. Residues in Smad4 and Smad2 known to be mutated in cancer are highlighted in yellow. Sites of phosphorylation are shown in green. The residues highlighted in red coordinate phosphorylation and will be discussed in Chapter III. Adapted from Wu et al., 2001b.
To test whether R-Smad homotrimerization occurs through the same Smad4 crystallographic trimer interface, we made mutations in the pseudophosphorylated R-Smads S3LC(3E) and S1LC(2D) that were equivalent to some of the tumorigenic mutations found in Smad4 (Figure 2-16). Of the R-Smads, only Smad2 is known to be mutated in cancers. Although Smad3 and Smad1 are not known to have any tumorigenic mutations, it has been discovered that these proteins can mediate aberrant signaling in cancerous cells, rather than being inactivated (de Caestecker et al., 2000; Roberts et al., 2003). However, if the R-Smads homotrimerize via an interface identical to that seen in the Smad4 crystallographic trimers, creating mutations equivalent to the tumorigenic Smad4 mutations in R-Smads should disrupt homotrimerization.

The mutations equivalent to the tumorigenic β3 strand mutation S4AF(D351H) in pseudophosphorylated Smad3 and Smad1 are S3LC(3E, D258H) and S1LC(2D, D297H). The mutations equivalent to the H5 helix mutation S4AF(D537E) are S3LC(3E, D408E) and S1LC(2D, D448H) (Figures 2-13 and 2-14). Asp 537 in Smad4 has been found to be mutated to either Glu or His in cancers, and we made the S1LC(2D, D448H) mutation anticipating a more pronounced effect than the more conservative Asp to Glu substitution might produce. A third trimer interface mutation was made in both proteins, S3LC(3E, V277D) and S1LC(2D, V317D), a mutation in the L2 loop of the loop-helix subdomain, whose equivalent mutation in Smad4, V370D, is also a cancer-derived interface mutation (Figure 2-13). Assessment of homotrimerization through size exclusion chromatography revealed that all three tumorigenic mutations made in S3LC(3E) and S1LC(2D) prevented homotrimerization, as the mutants eluted from the sizing column as monomers.
(Figure 2-16). Absent a crystal structure, these data support the idea that R-Smad homotrimerization is structurally identical to and occurs via the same interface as Smad4 crystallographic homotrimerization and R-Smad/Smad4 heterotrimerization.

**DISCUSSION**

Determining the stoichiometry of active Smad complexes is of central importance to elucidating the mechanism of TGF-β signaling through the Smad proteins. Prior analyses of Smad complex stoichiometry have yielded disparate results, with evidence suggesting that the active Smad complex could be either a heterodimer, heterotrimer, or heterohexamer (comprising homotrimers of an R-Smad and Smad4) (Jayaraman and Massague, 2000; Kawabata et al., 1998; Shi et al., 1997). Using glutamic acid or aspartic acid substitutions to simulate serine phosphorylation in purified R-Smad proteins, we analyzed the oligomerization state of the R-Smad homo-oligomeric and R-Smad/Smad4 hetero-oligomeric complexes. Based on the size exclusion chromatography and analytical ultracentrifugation analysis described in this chapter, we conclude that the active Smad signaling complex formed between R-Smads and Smad4 is a heterotrimer, consisting of two R-Smad subunits and one Smad4 subunit. Our results are in agreement with those of Kawabata and colleagues, who determined through immunoprecipitation of epitope-tagged Smad 2/3 and Smad4 that the R-Smad/Smad4 complex was a heterotrimer. Although their data suggested that the ratio of R-Smad to Smad4 in the heteromeric complex might be variable, our data strongly indicate that the ratio is immutably 2:1 in favor of R-Smads (Figure 2-11). It is also evident that although
Figure 2-15. Mutation of S4AF trimer interface residues disrupts interaction between pseudophosphorylated R-Smads (LC domains) and S4AF. Either S3LC(3E) (A) or S1LC(2D) (B) were mixed at a 1:1 mole ratio and loaded onto the Superdex 200 size exclusion column (50 μM each). Eluted fractions were analyzed by SDS-PAGE and gels stained with Coomassie blue.
**Figure 2-16.** Mutation of the equivalent trimer interface residues disrupts R-Smad homotrimerization. Proteins were loaded at 50 µM onto the size exclusion column. Eluted fractions were analyzed through SDS-PAGE and gels stained with Coomassie blue. (A) S3LC(3E) and derived mutants (B) S1LC(2D) and derived mutants.
phosphorylated R-Smads are able to form stable homotrimeric complexes, heterotrimeric R-Smad/Smad4 complexes form preferentially in the presence of Smad4 (Figure 2-11). The structural basis for this phenomenon will be addressed in Chapter IV.

Our results directly contradict those of Wu and colleagues, who used similar size exclusion chromatography, analytical ultracentrifugation, and mutational analyses of the Smad2/Smad4 interaction, and determined that this complex was a heterodimer (Wu et al., 2001a). It should be noted that an unphosphorylated form of Smad2 was used in the majority of their analyses, which we believe caused the difference with our results. It appears that even in the absence of phosphorylation, R-Smad/Smad4 interaction can occur, albeit incompletely, and that the heterodimer is the observed product of the incomplete interaction. Our own sedimentation analysis reveals that S3LC and S4AF associate even in the absence of phosphorylation (Figure 2-3C), and analysis of this association by size exclusion chromatography reveals a complex containing only slightly more Smad3 than Smad4, which could be interpreted as a heterodimer (data not shown).

Analysis of (actual) phosphorylated forms of Smad2 and Smad3 in complex with Smad4, described in Chapter IV of this thesis, confirms the core phosphorylated R-Smad/Smad4 complex is a heterotrimer. However, the heterodimer model of interaction was also supported by size exclusion chromatography analysis of endogenous Smad complexes in TGF-β stimulated cells, which revealed that the Smad2/Smad4 complex was an apparent heterodimer (Jayaraman and Massague, 2000). More recent analysis of epitope-tagged R-Smad/Smad4 oligomers, revealed that the Smad2/Smad4 oligomer, when in complex with the transcription factor Fast-1 and bound to DNA, is a heterotrimer, while the
Smad3/Smad4 complex under similar conditions is a heterodimer (Inman et al., 2002). It is also possible that these results reflect different oligomerization tendencies between Smad2 and Smad3. We think that this is improbable, given that Smad2 and Smad3 are approximately 95% identical between their respective MH2 oligomerization domains. It is more likely that this reflects a difference between the core R-Smad/Smad4 complex that translocates into the nucleus, and the DNA- and co-factor-bound form of the complex that activates transcription. This possibility will be addressed more completely in Chapter IV.

We also determined that the structural determinants for R-Smad homotrimerization and R-Smad/Smad4 heterotrimerization are similar to those seen in the crystallographic Smad4 homotrimers (Shi et al., 1997; Qin et al., 1999). In the crystallographic homotrimers, contact is mediated by the interaction of the three-helix bundle of one subunit with the loop helix region of a neighboring subunit, and this is emulated in Smad homotrimers and heterotrimers. Mutations equivalent to tumorigenic mutations found in Smad4 were made in pseudophosphorylated versions of Smad3 and Smad1, and were found to disrupt R-Smad homotrimerization as well as R-Smad/Smad4 heterotrimerization (Figures 2-15 and 2-16). This general pattern of trimeric Smad interaction has subsequently been confirmed by crystal structures of Smad1 (Chapter III; Qin et al., 2001), Smad2 (Wu et al., 2001b), Smad2/Smad4, and Smad3/Smad4 (Chapter IV).

Although it has been previously established that C-terminal phosphorylation of R-Smads promotes oligomerization by relieving MH1-mediated inhibition of the MH2
domain (Hata et al., 1997), it is clear from our studies that phosphorylation also serves to energetically drive Smad MH2 domains toward trimerization. The R-Smad constructs used in these studies lacked the autoinhibitory MH1 domain, containing only the linker and MH2 domains. Although the wild-type versions of the constructs displayed a tendency to oligomerize in a concentration-dependent manner (Figures 2-2, 2-3B, and 2-4[top]), phosphorylation greatly strengthens the ability of Smads to oligomerize (Figures 2-4[bottom], 2-8, and 2-9; Tables 2-1 and 2-2). The structural mechanism through which phosphorylation effects Smad oligomerization is the subject of Chapter III of this thesis. We also confirm the observation of Kawabata and colleagues that Smad4 does not appear to form homo-oligomers under any condition (Kawabata et al., 1997), as opposed to R-Smads. Since a primary structural difference between R-Smads and Smad4 is the presence of the C-terminal phosphorylation sequence in R-Smads (Figures 2-1 and 2-14), we addressed the possibility of whether this was a determinant for Smad4 homo-oligomerization. Replacement of the Smad4 C-terminus with the pseudophosphorylated form of the Smad3 C-terminus did not produce a form of Smad4 that was capable of homo-oligomerization, indicating that other structural differences might be responsible for the inability of Smad4 to homo-oligomerize, such as the Smad4-specific insertion between helices 3 and 4 of the Smad4 MH2 domain (Figure 2-14). It is possible that the observed inability of Smad4 to form homo-oligomers might be an artifact of the construct that was used, which lacked the N-terminal half of the protein (Smad4, 273-552). However, size exclusion chromatography analysis of full-length purified Smad4 revealed that this protein also elutes as a monomer (unpublished observations). However, our
results again contradict analysis of endogenous and epitope-tagged Smad4, which revealed that Smad4 exists as a homo-oligomer, most likely a homo-trimer (Jayaraman and Massague, 2000). This may reflect the presence of other factors bound to Smad4, causing it to run as an apparent oligomer, but additional studies will be required to clarify the homo-oligomerization state of Smad4, if any.
CHAPTER III: SMAD TRIMERIZATION IS FACILITATED BY THE INTERACTION BETWEEN THE POSITIVELY-CHARGED LOOP-STRAND POCKET AND THE C-TERMINAL PHOSPHORYLATED TAIL

INTRODUCTION

The experiments described in Chapter 2 effectively demonstrate that R-Smads homotrimerize upon phosphorylation, and in the presence of Smad4, preferentially form heterotrimers consisting of 2 R-Smad subunits to 1 Smad4 subunit. However, while it is clear that R-Smad phosphorylation induces trimerization, the specific role of phosphorylation in effecting Smad trimerization is not completely understood. Although previous studies have shown that phosphorylation induces the oligomerization activity of the R-Smad MH2 domain by relieving MH1 domain-mediated inhibition (Hata et al., 1997), the results described in Chapter II of this thesis clearly demonstrate that phosphorylation also serves as an energetic driver for R-Smad oligomerization. The studies in Chapter II utilized constructs of R-Smads from which the MH1 domain had been deleted, leaving only the linker and MH2 domains (referred to as S3LC and S1LC). Although the LC domains demonstrated a tendency to oligomerize as a function of concentration (Figures 2-2 and 2-3B), phosphorylation was necessary to induce oligomerization at physiological concentrations (Figures 2-4[bottom] and 2-8). In particular, analytical ultracentrifugation analysis showed that pseudophosphorylation of MH1-deleted R-Smads resulted in a substantial increase in trimerization constant (Tables 2-1 and 2-2). Therefore we sought to determine the specific mechanism for
phosphorylation-induced activation of the Smad MH2 domain. An initial clue toward such a mechanism was obtained from the crystal structure of a Smad4 fragment that had been previously solved in this laboratory (Qin et al., 1999). S4AF had been crystallized in a solution containing lithium sulfate, and the structure revealed that sulfate ions from the crystallization solution were bound at specific sites throughout the protein (Figure 3-1). This led to the idea that perhaps one or more of these positively charged pockets might be able to bind the phosphorylated R-Smad tail in the heterotrimeric complex, helping to facilitate R-Smad/Smad4 interaction. To test this hypothesis, we used similar site-directed mutagenesis and SEC-based interaction assay methods as described in Chapter II. The subsequent solution of the crystal structure of a Smad1 fragment supported this hypothesis. In addition, the Smad MH2 domain was found to share a similar protein fold with two other phosphopeptide-binding domains: the FHA domain of Rad53 and the IAD of IRF-3 (Li et al., 2000; Durocher et al., 2000; Qin et al., 2003; Takahasi et al., 2003), supporting the idea that the Smad MH2 is a conserved phosphoserine binding domain.

MATERIALS AND METHODS

Reagents

Initial SILCS crystal screens were performed using the Crystal Screen I and II and Natrix kits (Hampton Research).
Construction of expression plasmids and mutagenesis

Smad3LC (S3LC, residues 145-425), Smad1LC (S1LC, residues 143-465), Smad1LCS (S1LCS, residues 217-465), and Smad4AF (S4AF, residues 273-552) were generated as described in Chapter II. Site-directed mutagenesis was performed through a PCR-based strategy using the QuikChange kit (Stratagene), also as described in Chapter II.

Protein expression and purification

All versions of S3LC, S1LC, and S4AF described in this chapter were expressed as GST fusions and purified as described in Chapter II.

Size exclusion chromatography

Size exclusion chromatography was used to detect protein interaction, as described in Chapter II.

Transcriptional response assays

Transcriptional response assays were performed by Dr. Mark de Caestecker, and are similar to those described in Chapter II. NMuMg cells were transfected with SBE-Lux, full-length forms of wild-type or pseudophosphorylated Flag-Smad3, and full-length wild-type Smad4-Myc or the point mutants derived thereof.

Crystallization of S4AF(R515S) and Smad1LCS

Crystals of S4AF(R515S) were obtained by the hanging drop vapor diffusion technique. As expected, this protein crystallized under nearly identical conditions to wild-type S4AF (Qin et al., 1999). Purified S4AF(R515S) (40 mg/ml) was combined with an equal volume of reservoir solution containing 100 mM HEPES (pH 7.5), 250 mM
lithium sulfate, and 10% (v/v) PEG 4000. Crystals were transferred to a cryo-solvent consisting of 25% glycerol and 75% reservoir solution, and flash frozen in liquid nitrogen. Diffraction data were collected at -170°C using an R-Axis IV image plate system mounted on a Rigaku rotating anode generator. The data were collected at a detector distance of 140 mm with 1° oscillation per frame and were integrated and reduced using DENZO and Scalepack (Otwinowski and Minor, 1997). The structure was solved by molecular replacement with the CNS software package (Brunger et al., 1998), using the previously solved S4AF structure as a search model (Qin et al., 1999). Rigid body, positional, and simulated annealing refinements were also performed with CNS.

Crystals of S1LCS were also obtained through the hanging drop method. Purified S1LCS (20 mg/ml) was combined with an equal volume of reservoir solution containing 50 mM MES (pH 6.0) and 6% ethanol. Freezing of crystals, data collection, and indexing of diffraction data was performed as described above. The S1LCS structure was also determined through molecular replacement, using S4AF as a search model. Molecular replacement and subsequent refinements were performed with CNS. Model building was performed with CHAIN (Sack, 1988). Structure figures were prepared with Molscript (Kraulis et al., 1991), Raster3D (Merritt and Bacon, 1997), and GRASP (Nicholls et al., 1991).
RESULTS

Mutational analysis of anion-binding pockets in Smad4

The previously solved crystal structure of S4AF revealed that sulfate ions (from the lithium sulfate used as a crystallization additive) were bound at specific sites throughout the protein (Figure 3-1). Since phosphorylation is required to activate the Smad proteins, and given the chemical similarity between sulfate and phosphate ions, we asked whether one or more of these sulfate-binding sites might be serving as a binding site for the R-Smad phosphorylated tail, thereby facilitating Smad trimerization. Three positions in each subunit (subunits a, b, and c) of the S4AF trimer, named sites 1, 2, and 3, were found to bind sulfate ions, resulting in a total of nine sulfate binding sites (Sa1, Sa2, Sa3, Sb1, Sb2, Sb3, Sc1, Sc2, Sc3). A tenth sulfate-binding site was located at the junction of subunits B and C, and this lone site where the sulfate was coordinated by residues from two subunits was designated site Sbc. All four unique sulfate-binding sites are defined by the presence of arginine residues (Figure 3-2). Sulfate-binding site 1 is located at the base of the Smad4-specific MH2 insertion, and the sulfate ion is coordinated by the side chains of Arg445 and Gln442 of helix H3, and Arg416 from helix H2. None of the three sulfate-coordinating residues are conserved in other Smads (Figure 2-14), so this site seemed unlikely to interact with the phosphorylated R-Smad C-terminal tail. Sulfate-binding site 2 comprises Arg497 from helix H4 and Arg502 from the β10 strand. Arg497 and Arg502 are well conserved among Smad4 isoforms, except that the equivalent to Arg502 in C. elegans is an alanine. Sulfate-binding site 3 is located in the L3 loop, where the side chains of Lys507 and Arg515 coordinate the sulfate ion.
Figure 3-1. $F_o-F_c$ map displaying the electron density of the sulfate ions bound at individual sulfate-binding sites throughout the previously solved S4AF structure. The map was calculated after simulated annealing refinement without including the coordinates of the sulfate ions. The S4AF structure is shown as a Cα trace, with subunits A, B, and C colored blue, green, and red, respectively. Adapted from Qin et al., 1999.
Figure 3-2. The specific amino acid residues that comprise each of the four unique sulfate binding sites, which are colored in purple. The location of the sulfate ions in the context of the entire S4AF structure is shown in Figure 3-1.
These residues are completely conserved in all Smad proteins. As mentioned in the general introduction, the L3 loop had been previously shown to mediate R-Smad interaction with the receptor (Lo et al., 1998). Mutations in the L3 loop found in cancer and developmental disorders have been shown to disrupt R-Smad/Smad4 heterooligomerization, which led Shi and colleagues to suggest that the active Smad signaling complex was a heterohexamer comprising homotrimers of R-Smads and Smad4, interacting via their L3 loops (Hata et al., 1998b; Shi et al., 1997). The final sulfate-binding site, Sbc, is located next to site Sc2, and is created by the side chains of Arg496 and Arg497 of subunit C, and the side chain of Arg372 and the backbone of Asn369 of subunit B. These interactions are conserved among Smad4 isoforms and R-Smads, except that the equivalent to Arg496 in R-Smads is a threonine.

To test the idea that interaction of the C-terminal phosphorylated tail with one or more of the sulfate-binding sites of Smad4 might mediate heterotrimer formation, we used site-directed mutagenesis to change arginine residues within each of the sulfate-binding sites to serine, then analyzed the ability of the mutants to interact with pseudophosphorylated S3LC(3E) through size exclusion chromatography. The results of this analysis are shown in Figure 3-3A. Mutation of Arg416 (site 1), Arg502 (site 2), and Arg496 to serine seemed to have no effect on the interaction between S3LC(3E) and S4AF. However, mutation of Arg515 of the L3 loop appeared to completely disrupt the heterotrimeric interaction (Figure 3-3A). Lys507 of the L3 loop sulfate-binding site was also mutated, resulting in a significant decrease in interaction between S4AF(K507S) and S3LC(3E), but not a complete abolition as with S4AF(R515S) (data not shown). These
results support the hypothesis that interaction between phosphorylated R-Smads and Smad4 is facilitated by a specific electrostatic interaction between the phosphorylated C-terminus of R-Smads with a putative anion-binding site in Smad4, located in the L3 loop.

As the experiments in Chapter II demonstrate, R-Smad homotrimerization and R-Smad/Smad4 heterotrimerization occur via a common interface. The high structural similarity between R-Smads and Smad4 suggest that an identical phosphorylated tail-L3 loop interaction may also mediate R-Smad homotrimerization. This was tested by mutating the L3 loop arginine residue equivalent to Smad4 Arg515 in pseudophosphorylated Smad3 to a serine residue. This mutant, S3LC(3E, R386S) was tested for its ability to homotrimerize through size exclusion chromatography. This single mutation clearly prevents homotrimerization, as the mutant protein elutes as a monomer (Figure 3-3B), further supporting the idea that a direct interaction between the phosphorylated C-terminal tail and the L3 loop facilitates Smad protein interaction. This also supports a key result of the studies described in Chapter II, that R-Smad homotrimerization and R-Smad/Smad4 heterotrimerization share a common structural basis.
Figure 3-3. (A) Analysis of interaction between S3LC(3E) and S4AF sulfate-binding site mutants. S3LC(3E) and sulfate-binding site mutants of S4AF were combined at a 1:1 ratio and loaded on the gel filtration column. The eluted fractions that contained protein were analyzed by SDS-PAGE and the gels stained with Coomassie blue. (B) Mutation of the L3 loop arginine to serine (R386S) abrogates S3LC(3E) homotrimerization.
Crystallization of S4AF(R515S)

We considered the possibility that the loss of interaction between S3LC(3E) and S4AF(R515S) could be due to a gross conformational change caused by the mutation, rather than the loss of a bond between the L3 loop arginine and the phosphorylated tail. To determine whether the loss of interaction caused by the R515S mutation was due to a conformational change, we solved the crystal structure of S4AF(R515S). The protein crystallized in nearly identical conditions as the previously crystallized wild-type S4AF (Qin et al., 1999). The superimposed Fo-Fc omit maps of the L3 loops of both wild-type S4AF and the R515S are shown in Figure 3-4, and the crystallography statistics for the structure are given in Table 3-1. It is apparent from the electron density that no change in conformation exists in the L3 loop of S4AF(R515S), indicating that the loss of interaction between S3LC(3E) and R515S is due solely to the absence of R515 in the L3 loop, supporting the hypothesis that interaction between the L3 loop arginine and the phosphorylated C-terminal tail stabilizes Smad complex formation.

Signaling assays confirm the importance of the L3 loop arginine

Given that the only sulfate-binding site mutation to cause a loss of Smad3/Smad4 interaction was S4AF(R515S), we wanted to establish a functional consequence for this loss of interaction, by determining the effect of this mutation on Smad3/Smad4 dependent transcriptional activity. Signaling assays were performed as described in Chapter II, to compare the ability of each of the sulfate-binding site mutants to activate transcription from a consensus Smad-binding element (SBE) linked to a luciferase gene.
Figure 3-4. Stereo view of the $F_o-F_c$ omit map at the L3 loop of S4AF(R515S). The mutant and wild-type coordinates are shown in black and gray, respectively.
Table 3-1. Summary of crystal analysis for S4AF(R515S).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>S4AF(R515S)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Crystal parameters and crystallographic data</strong></td>
<td></td>
</tr>
<tr>
<td>Space group</td>
<td>I4(1)22</td>
</tr>
<tr>
<td>Unit cell dimensions</td>
<td>a = b = 140.6, c = 193.0</td>
</tr>
<tr>
<td>Diffraction limit (Å)</td>
<td>100-3.0 (3.11-3.0)</td>
</tr>
<tr>
<td>Total reflections</td>
<td>72845</td>
</tr>
<tr>
<td>Unique reflections</td>
<td>19552</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>97.9 (88.0)</td>
</tr>
<tr>
<td>Intensity/Sigma</td>
<td>9.5 (2.3)</td>
</tr>
<tr>
<td>R$_{merge}$ (%)</td>
<td>12.6 (48.5)</td>
</tr>
</tbody>
</table>

**Refinement statistics**

- Protein atoms: 5434
- R factor (%): 20.01
- R$_{free}$ factor (%): 26.27
- Rms deviation from ideal:
  - Bond lengths (Å): 0.007
  - Bond angles (°): 1.30
- B-factor rms deviation:
  - Main chain (Å$^2$): 2.1
  - Side chain (Å$^2$): 2.8

*Values in brackets are for the highest resolution shell.

$^b$ $R_{merge} = \sum |I_{hk0} - <I_{hk0}>|/\sum I_{hk0}$

$^c$ $R$ factor = $\sum_{hk0} |F_{obs}| - |F_{calc}|/\sum_{hk0} |F_{obs}|$ for all data.

$^d$ $R_{free} = \sum_{hk0} |F_{obs}| - |F_{calc}|/\sum_{hk0} |F_{obs}|$ for 10% of the data not used in refinement.
As shown in Chapter II, Smad3(3E) is able to activate the SBE-luciferase gene only when cotransfected with Smad4 (Figure 2-7 and Figure 3-5). Of the four arginine to serine sulfate-binding site mutants, the L3 loop mutant R515S impacted signaling the most, causing a decrease of over 80% in transcriptional activity compared to wild-type Smad4 (Figure 3-5). The R416S (site 1) and R496S (site bc) mutations resulted in an approximately 50% decrease in signaling activity, and the R502S (site 2) mutant decreased signaling by 20% from wild-type Smad4. These data support the in vitro results, confirming that Arg515 of the L3 loop is a critical determinant of Smad3-Smad4 heterotrimerization. The R416S, R496S, and R502S mutations appear to impact signaling as well, although less than the R515S mutation. However, since these other mutants do not appear to affect interaction with pseudophosphorylated Smad3 (Figure 3-3A), the reduction in signaling that is observed may involve components of the signaling pathway other than Smad3 and Smad4. Further studies will be required to elucidate these other components of Smad-dependent signaling.

The results of the previous sections provide strong evidence that the homotrimeric R-Smad and heterotrimeric R-Smad/Smad4 complex formation are facilitated by an electrostatic interaction between the phosphorylated C-terminal tail of an R-Smad subunit with the L3 loop of the adjacent Smad subunit. As can be seen in Figure 2-13, the L3 loop is not involved in intermolecular contacts in the complex but is instead located on a solvent accessible surface of the trimer. However, mutations in the L3 loop have been found in cancer and developmental diseases (Sekelsky et al., 1995; Savage et al., 1996; Maliekal et al., 2003), supporting a key role for this motif in Smad-dependent signaling.
Figure 3-5. The Smad4 (R515S) mutant only weakly activates Smad3/Smad4 dependent transcription. MDA-MB468 cells were transfected with SBE-Lux along with the indicated FLAG-Smad3(3E) and Smad4-Myc mutant constructs.
Although it had previously been shown that the L3 loop is involved in mediating receptor interaction (Lo et al., 1998), these studies clearly demonstrate that it also has a role in mediating subunit interaction in oligomeric Smad complexes. It can be seen from the previously solved Smad4 MH2 domain structures (Shi et al., 1997; Qin et al., 1997) that the L3 loop is located near the C-terminus of the neighboring subunit within the trimer. Comparing the Smad4 structures with the structure of the Smad2 MH2 domain (solved in complex with the R-Smad anchoring protein SARA) (Wu et al., 2000) reveals that the 10 C-terminal residues of R-Smads form flexible extensions from the core β-strand sandwich. This flexibility could allow the 10 C-terminal residues of the R-Smad phosphorylated tail to form an extended conformation to reach the neighboring L3 loop. Direct confirmation of this model was obtained from the crystal structure of S1LCS, described in the following section.

**Crystal Structure of S1LCS**

The structural basis for R-Smad activation was further investigated through the solution of the crystal structure of the Smad1 fragment S1LCS. S1LCS contains the C-terminal 49 residues of the linker domain and the entire MH2 domain (residues 217 to 465). As described in Chapter II, the pseudophosphorylated mutant S1LCS(2D) displays an approximately 4000-fold improvement in ability to trimerize over S1LCS, as determined from sedimentation velocity experiments (Table 2-2). However, attempts to crystallize S1LCS(2D) were unsuccessful, as this pseudophosphorylated protein only produced clusters of small crystals unsuitable for diffraction. Under similar crystallization conditions, unphosphorylated S1LCS produced well-ordered, single
crystals, which diffracted to 2.5 Å resolution (Table 3-2). The structure was determined through molecular replacement using the previously solved S4AF structure as a search model (Qin et al., 1999). The asymmetric unit comprises four Smad1 subunits (Figure 3-6). Three subunits (subunits A, B, and C) pack as a trimer and are related by 3-fold noncrystallographic symmetry (NCS), while the fourth (subunit D) interacts with the 3-fold crystallographic symmetry mates to form the same trimeric arrangement. This trimeric arrangement is similar to the previously solved structures of Smad4 MH2 (Shi et al., 1997) and Smad4 AF (Qin et al., 1999). Residues 217 to 267 of the Smad1 linker region were disordered and therefore not visible. The tail of subunit D (residues 454 to 465) is similarly disordered, but the tails of the three NCS-related subunits are well ordered and assume two distinct structural arrangements (Figure 3-6 and Qin et al., 2001). The tails of subunits A and B extend and interact with the L3 loops of adjacent subunits C and A, respectively. However, the C-terminal tail of subunit C rotates about Gly455 and interacts with the L3 loop of subunit D, leaving the L3 loop of subunit B empty (Figure 3-6). The structure reveals how phosphorylation of the two C-terminal serines facilitates phosphorylation (Figure 3-7A). The serine at the -1 position (Ser465) is located near Lys418 and Arg426 of the adjacent L3 loop, two residues that are conserved in all Smads. As described above, mutation of the equivalent residues in Smad4 (Lys507 and Arg515) resulted in abrogation of the interaction between S3LC(3E) and S4AF. It is evident from the structure that phosphorylation of the -1 serine will introduce favorable electrostatic and hydrogen bonding interactions between the
Table 3-2. Summary of crystal analysis for S1LCS.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>S1LCS</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Crystal parameters and crystallographic data</strong></td>
<td></td>
</tr>
<tr>
<td>Space group</td>
<td>R3</td>
</tr>
<tr>
<td>Unit cell dimensions</td>
<td>a = b = 138.1, c = 199.3</td>
</tr>
<tr>
<td>Diffraction limit (Å)(^a)</td>
<td>2.5 (2.57-2.5)</td>
</tr>
<tr>
<td>Total reflections</td>
<td>77509</td>
</tr>
<tr>
<td>Unique reflections</td>
<td>47621</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>96.8 (98.8)</td>
</tr>
<tr>
<td>Intensity/Sigma</td>
<td>10.4 (2.0)</td>
</tr>
<tr>
<td>R(_{merge}) (%)(^b)</td>
<td>6.5 (35.1)</td>
</tr>
<tr>
<td><strong>Refinement statistics</strong></td>
<td></td>
</tr>
<tr>
<td>Protein atoms</td>
<td>6187</td>
</tr>
<tr>
<td>Water molecules</td>
<td>586</td>
</tr>
<tr>
<td>R factor (%)(^c)</td>
<td>23.9</td>
</tr>
<tr>
<td>R(_{free}) (%)(^d)</td>
<td>27.4</td>
</tr>
<tr>
<td>Rms deviation from ideal</td>
<td></td>
</tr>
<tr>
<td>Bond lengths (Å)</td>
<td>0.01</td>
</tr>
<tr>
<td>Bond angles (°)</td>
<td>1.40</td>
</tr>
</tbody>
</table>

\(^a\)Values in brackets are for the highest resolution shell.

\(^b\)R\(_{merge}\) = \(\sum |I_{hk\ell}| - \langle |I_{hk\ell}| \rangle / \sum I_{hk\ell}\)

\(^c\)R factor = \(\sum_{hk\ell}|F_{obs}^2 - |F_{calc}^2| / \sum_{hk\ell}|F_{obs}^2|\) for all data.

\(^d\)R\(_{free}\) = \(\sum_{hk\ell}|F_{obs}^2 - |F_{calc}^2| / \sum_{hk\ell}|F_{obs}^2|\) for 10% of the data not used in refinement.
Figure 3-6. Crystal structure of S1LCS. The four subunits in the asymmetric unit are shown in ribbon representation. The 3-fold non-crystallographic symmetry axis is perpendicular to the page and indicated by the (x). The 3-fold crystallographic axis is indicated by the horizontal gray dashed line. Areas where a C-terminal tail is in contact with an L3 loop are circled. L3 loops are depicted in black.
Figure 3-7. (A) Surface electrostatic potential representation of the interaction between the L3 loop/β8 strand pocket and the C-terminal tail. The residues lining the tail binding pocket are labeled in black, while the tail residues are labeled in red. (B) (top) Stereo view of the L3 loop/tail interaction. The L3 loop main chain and side chain are colored in cyan and green, respectively. The C-terminal tail is shown in pink. Gly419 is indicated by a sphere. (bottom) Stereo view of the unbound L3 loop.
phosphoserine residue and the lysine and arginine of the L3 loop. The residue at the −2 position of the C-terminus, Val464, is solvent exposed, consistent with the non-conservation of this residue in the R-Smads. Ser463, the residue in the −3 position, is in close proximity to Tyr424 of the L3 loop, and Lys373 of the β8 strand. The structure suggests that phosphorylation of this serine will result in the formation of several hydrogen bonds between the −3 phosphoserine and Tyr424 and Lys373. The serine at the −4 position has been demonstrated not to be phosphorylated from previous phosphopeptide mapping studies (Abdallah et al., 1997; Souchelnytskyi et al., 1997), and that is supported by this structure. The −4 serine points toward the solvent, suggesting that phosphorylation of this serine is unlikely to affect subunit interaction. This was supported by analytical ultracentrifugation experiments performed on the eight combinations of wild-type serine or glutamic acid pseudophosphorylation mutations of the three C-terminal serines at the −1, −3, and −4 positions (SSVS, SSVE, SEVS, ESVS, SEVE, EEVS, ESVE, and EEVE). The studies showed that pseudophosphorylation of only the −1 and −3 serines was strongly stabilizing, with the SEVS and SEVE mutations exhibiting strong trimerization enhancement (ΔΔG = -2.9 to -3.5 kcal/mol), the ESVS mutations exhibiting almost no enhancement (ΔΔG = -0.3 kcal/mol), and the rest (SSVE, ESVE, EEVS, EEVE) exhibiting intermediate enhancement (ΔΔG = -1.5 to -2.3 kcal/mol) (data not shown). Also noteworthy is that the conserved Gly419 in the L3 loop is located at the bottom of phosphotail binding site, within van der Waals contact to the −3 serine (Ser463) (Figure 3-7A). Mutation of the corresponding glycine in Smad2 and Smad4 has been implicated in developmental defects (Hata et al., 1998b). The structure
suggests that mutation of this residue to anything other than glycine would result in a steric clash with the –3 serine, preventing the tail from interacting with the L3 loop. The interaction between the phosphorylated tail and the positively charged pocket defined by the L3 loop and the β8 strand provide additional subunit contact and hydrogen bonds, and may partly explain why R-Smads homotrimerize, but not Smad4. The buried surface area between subunits is 2400 Å², of which 900 Å² is contributed by the final twelve C-terminal residues.

Concerted structural changes drive Smad1 activation

In the crystal structure of S1LCS, the L3 loop is found in two distinct conformations. The L3 loops of subunits A, C, and D are bound to the C-terminal tails of adjacent subunits, whereas the L3 loop of subunit B is unbound (Figure 3-7B). Comparison between the two conformations in which the L3 loop is found reveals that it undergoes significant rerouting upon binding of the C-terminal tail. The root mean square deviation of the L3 loop main chain (residues 420 to 427) between two tail-liganded subunits (A/C, C/D, and D/A pairs) is 0.33 ± 0.14 Å, but is 1.77 ± 0.14 Å between the tail-liganded subunits and the unliganded subunit B (A/B, C/B, and D/B pairs). Most notably, Arg426, the putative –1 phosphoserine binding residue, flips toward the tail to form the phosphoserine binding pocket, with its guanido group moving 13 Å. The adjacent His425 and Gln427 of the L3 loop undergo significant side chain flipping as a result of this interaction. His425 and Asp428 have previously been shown to mediate BMP receptor kinase specificity for Smad1, and switching these residues with those in the equivalent position in Smad2 (Arg427 and Thr430) results in a Smad1 that is
instead phosphorylated by the TGF-β receptor (Lo et al., 1998). Therefore the L3 loop appears able to exist in two distinct conformations, one that enables interaction with the GS domain of activated receptor kinase, and another that allows interaction with other Smads by interacting with the phosphorylated tail.

Structural comparison between the Smad1 structure and a Smad2 monomer solved in complex with the anchor protein SARA reveals that the three-helix bundle of trimeric Smad1 undergoes a substantial shift (Qin et al., 2001). SARA is a membrane-anchoring protein that recruits (monomeric) Smad2 or 3 via its Smad-binding domain (SBD) to the corresponding receptor kinase, and is thought to restrict movement of the Smad three-helix bundle and thus keep Smad 2 or 3 in a monomeric state (Figure 1-3; Tsukazaki et al., 1998). Following phosphorylation by the receptor kinase, R-Smads dissociate from the receptor/SARA complex and subsequently trimerize. Therefore it appears that Smad1 activation is defined by concerted, phosphorylation-induced structural changes, in which flipping of the L3 loop and tilting of the three-helix bundle result in a Smad1 protein with a greatly increased tendency toward forming trimeric Smad complexes.

Mutational analysis based on the crystal structure

The previous mutational analysis described above was based on the sulfate-binding sites that were seen in the S4AF structure. In those studies, only the lysine and arginine of the L3 loop were mutated, as these were the only residues observed to directly bind the sulfate ion. However, the structure of S1LCS reveals that the interaction of the phosphorylated tail with the L3 loop involves more than just those two residues. In
addition to Lys418 and Arg426 of the L3 loop, which bind the phosphorylated serine at the –1 position, Tyr424 of the L3 loop and Arg373 of the β8 strand also appear to bind the phosphorylated serine at the –3 position. Each of these four potential phosphate-coordinating residues was mutated in S1LC(2D) to determine how the loss of the phosphoserine tail/loop-strand pocket interaction would affect homotrimerization. The equivalent residues were also mutated in the Smad4 construct S4AF to determine the effect of mutation on heterotrimer formation with S1LC(2D).

As demonstrated by size exclusion chromatography analysis in Chapter II, the pseudophosphorylated S1LC(2D) construct displays a tendency toward trimerization over its unphosphorylated counterpart, S1LC. Figure 3-8A shows the effect of mutation of the putative phosphoserine-binding sites on homotrimerization of S1LC(2D). Mutation of all four sites either reduces or prevents trimerization. The K373S, K418S, and R426S mutations display a marked decrease in homotrimerization. The conservative Y424F mutation displayed only a slightly reduced homotrimerization from S1LC(2D). This may be because the Y424F mutation only leads to the loss of a sole hydrogen bond mediated by the tyrosine hydroxyl group, whereas the other mutations result in the loss of positively charged residues that likely maintain strong electrostatic interactions with phosphoserine residues. The three equivalent mutations of the charged residues (K426S in the β8 strand, and K507S and R515S in the L3 loop) in S4AF significantly reduced the heterotrimeric interaction with S1LC(2D) (Figure 3-8B). Taken together, these data strongly support the model that both R-Smad homotrimerization and R-Smad/Smad4 heterotrimerization require the interaction between the phosphoserine residues of the R-
Smad C-terminus and a positively charged pocket in adjacent R-Smad or Smad4 subunits, defined by the L3 loop and β8 strand.
Figure 3-8. Mutation of the Smad phosphoserine-binding pocket disrupts both R-Smad homotrimerization and R-Smad/Smad4 heterotrimerization. (A) Mutation of the four residues comprising the phosphoserine-binding pocket of S1LC(2D). (B) Mutation of three of the residues comprising the phosphoserine-binding pocket of S4AF.
DISCUSSION

The studies described in this chapter reveal the structural mechanism through which C-terminal phosphorylation of the MH2 domain activates R-Smads. Although the SILCS crystal structure was unphosphorylated, biochemical evidence and other observations indicate that this structure represents an active R-Smad conformation. The size exclusion chromatography and sedimentation analysis described in Chapter II showed that Smad3 and Smad1 both displayed a concentration-dependent tendency toward trimerization. It is therefore reasonable to suppose that under the high protein concentration of crystallization, unphosphorylated Smad1 can form an active trimer. Also, the interactions between the C-terminal tail and the L3 loop involve highly conserved residues, consistent with the conserved phosphorylation mechanism of R-Smads. This was verified by the mutagenesis experiments shown in Figure 3-8, the locations of which were derived from the crystal structure. The structure also explains the occurrence of tumorigenic and developmental mutations in the L3 loop, and the critical role of the L3 loop in Smad homomeric and heteromeric interaction. This work reveals that the activation mechanism employs specific docking of the phosphorylated C-terminal SVS sequence to the L3 loop phosphoserine binding pocket of the neighboring Smad molecule to tighten a trimeric scaffold (Figure 3-10). The trimerization event induces concerted structural changes in the MH2 domain, which function as an allosteric switch in signaling. The phosphoserine binding pocket and the trimer interface are conserved among all R-Smad proteins, suggesting a conserved structural mechanism.
The Smad MH2 is structurally related to other phosphoserine binding domains

The Smad MH2 domain consists of a central β-sandwich with twisted antiparallel β-sheets of five and six strands each (Figure 1-2A; Shi et al., 1997; Qin et al., 1999). A strikingly similar β-sandwich protein fold is seen in two other protein domains that have putative phosphoserine or phosphothreonine binding activity, the forkhead-associated domain (FHA), found in Rad53, Chk2, and numerous other proteins (Li et al., 2000; Durocher et al., 2000; Li et al., 2002), and the IRF-association domain (IAD) of the interferon regulatory factors (IRFs) (Erushkin and Mushegian, 1999; Qin et al., 2003; Takahasi et al., 2003) (Figure 3-9). The positively charged phosphoserine or phosphothreonine-binding surface (defined by the L3 loop and β8 strand in the Smad MH2 domain) is conserved between the MH2, FHA and IAD domains, despite low sequence homology (Qin et al., 2003). Other studies have shown that the same positively-charged surface of the MH2 domain of R-Smads binds the L45 loop and phosphorylated GS domain of the activated TGF-β type II receptor, with receptor-Smad specificity determined by the L45-L3 loop interaction (Huse et al., 2001; Lo et al., 1998). Therefore the Smad MH2 domain, through a basic surface defined largely by the L3 loop, appears to be a protein-protein interaction module capable of performing two distinct phosphoserine binding functions in TGF-β signaling, 1) binding to phosphorylated TGF-β type I receptor and 2) binding of phosphorylated C-terminal tails of adjacent subunits in Smad trimers (Huse et al., 2001; Qin et al., 2002).
Figure 3-9. The Smad MH2 domain is structurally homologous to the FHA and IAD domains. (A) (left) Superimposed β-sandwiches of the Smad2 MH2 (yellow) and Rad53 FHA (red). (right) Secondary structure topology diagrams of the MH2 and FHA domains. Reproduced from Durocher et al., 2000. (B) Secondary structure topology of Smad MH2 (left) and IAD from IRF-3 (right). Reproduced from Qin et al., 2003.
A phosphorylation-induced allosteric switch drives Smad dissociation from the receptor complex

Prior studies have shown that two distinct structural mechanisms direct the specific phosphorylation of R-Smads by receptor kinase complexes. First, the L45 loop of the TGF-β receptor kinase domain specifically interacts with the L3 loop of an R-Smad protein (Feng and Derynck, 1997; Chen et al., 1998; Lo et al., 1998; Persson et al., 1998). Second, the receptor-associated recruiting molecule, such as SARA in the TGF-β/activin pathway, specifically recruits an R-Smad protein to the receptor kinase for phosphorylation (Tsukazaki et al., 1998). Although these mechanisms ultimately lead to specific phosphorylation of the R-Smads, how the recruited R-Smads disengage the receptor kinase/recruiting molecule complex after phosphorylation is unknown. The structure of the Smad1 homotrimer, in comparison with other R-Smad structures, provides insight into how this might occur (Figure 3-10). The structure of the trimeric Smad1 reveals that the L3 loop and the three-helix bundle structure undergo concerted conformational changes upon activation, which may induce Smad1 dissociation from the receptor kinase complex. In the basal state, in which Smad1 is monomeric, the L3 loop of Smad1 interacts with the L45 loop of the receptor kinase. In addition, the basic surface defined by the L3 loop and the β8 strand concurrently interacts with the phosphorylated GS domain of the receptor kinase (Huse et al., 2001). Phosphorylation-induced activation of R-Smads effects two distinct structural changes: a shift in the three-helix bundle and a flipping of the L3 loop (Figure 3-7B). This results in formation of a trimer, in which the L3 loop of an R-Smad interacts with the phosphorylated C-terminal
Figure 3-10. An allosteric model for Smad activation. In the basal, unphosphorylated state, the R-Smad is a monomer. Following activation of the receptor kinase through ligand binding, the R-Smad is recruited to the receptor kinase. The R-Smad interacts with the phosphoserine binding site and L45 loop of the receptor through the basic pocket defined by the L3 loop and β8 strand. After phosphorylation of the R-Smad by the receptor kinase, the R-Smad dissociates and is capable of forming homotrimers, but preferentially form heterotrimers in the presence of Smad4.
tail of another R-Smad molecule. Thus, by employing distinct conformations to interact in a mutually exclusive manner with multiple signaling partners, the L3 loop can serve as a switch for R-Smad dissociation from the receptor. Consistent with this model, the interaction between the receptor kinase complex and R-Smad is stronger when either the C-terminal phosphorylation sites of the R-Smad are mutated or when kinase activity is rendered inactive by a catalytic site mutation (Lo et al., 1998; Macias-Silva et al., 1996). Furthermore, the trimerization-induced tilting of the three-helix bundle structure toward the subunit interface may also serve as a conformational switch to direct R-Smad dissociation from the receptor complex (Figure 3-10). In the analogous TGF-β/activin pathway, SARA recruits and stabilizes the monomeric form of Smad2 or Smad3. Structural comparison between the trimeric Smad1 structure and the Smad2/SARA 1:1 complex structure reveals that SARA inhibits Smad2 trimerization by restricting movement of the three-helix bundle, without which Smad trimerization cannot occur (Qin et al., 2002; Tsukazaki et al., 1998). That the three-helix bundle exists in different conformations between monomeric and trimeric Smads was confirmed by the structure of an unliganded form of Smad3 (Qin et al., 2002). As was expected, the three-helix bundle of that monomeric protein was found in an identical conformation as the SARA-bound forms of Smad2 and Smad3 (Tsukazaki et al., 1998), whereas the three-helix bundles of trimeric Smad1 and Smad2 display a substantial shifting (Qin et al., 2001; Wu et al., 2001b). We suggest that phosphorylation energetically favors trimerization, and that formation of trimers is sterically incompatible with receptor/SARA association. Similar
mechanisms may exist for Smad1 through other as yet undiscovered receptor-associated molecules.
CHAPTER IV: CRYSTALLOGRAPHIC AND CALORIMETRIC ANALYSIS OF THE R-SMAD/SMAD4 COMPLEX

INTRODUCTION

The studies described to this point indicate that formation of the R-Smad/Smad4 heterotrimer is energetically preferred over R-Smad homotrimerization, and that heterotrimer formation occurs through an interface identical to that seen in the crystallographic homotrmers of Smad4 (Shi et al., 1997; Qin et al., 1999), Smad1 (Qin et al, 2001), and Smad2 (Wu et al., 2001b). Contrary to our results, however, data from other studies appears to show that the heteromeric R-Smad/Smad4 complex is a heterodimer, rather than a heterotrimer. Size exclusion chromatography analysis of endogenous Smad2/Smad4 complexes from TGF-β-stimulated cells revealed that these complexes were heterodimers (Jayaraman et al., 2000). Smad2 and Smad3 were also discovered to have very different endogenous homo-oligomerization states, but the oligomerization state of the Smad3/Smad4 complex was not determined in that study. Subsequent biochemical analysis of the MH2 domains of Smad2 and Smad4 also appeared to support a heterodimeric model (Wu et al., 2001a). Additional mutational analysis showed that some mutations in the putative heterotrimeric interfaces of Smad2 and Smad4 did not abolish interaction between those two proteins, which would support the notion that the Smad2/Smad4 hetero-oligomer is not mediated by the crystallographic trimer interface (Wu et al., 2001b). While these studies may suggest a fundamental difference between the Smad3/Smad4 (heterotrimer) and Smad2/Smad4 (heterodimer)
complexes, analysis of nuclear extracts of cells transiently co-transfected with
differentially-tagged Smads revealed the opposite. Smad2 and Smad4 formed a
heterotrimer when bound to the cofactors Fast-1 or Fast-3 and DNA, while the DNA-bound Smad3/Smad4 complex was a heterodimer (Inman et al., 2002).

In addition to the issue of subunit stoichiometry, the mechanism of R-
Smad/Smad4 complex formation is similarly unclear. As the experiments detailed in
Chapter II have shown, R-Smad phosphorylation results in homotrimerization, in the
absence of Smad4 (Kawabata et al., 1998; Chacko et al., 2001; Qin et al., 2001; Wu et al.,
2001b). The R-Smad homotrimer is strongly stabilized by the interaction between
phosphoserine residues of the C-terminal tail and positively charged residues in the L3
loop/β8 strand of an adjacent subunit, and through extensive hydrogen bonds and van der
Waals contacts at the trimer interface, as described in Chapter III and elsewhere (Qin et
al., 2001; Wu et al., 2001b). However, when Smad4 is present, the R-Smad/Smad4
heteromer clearly forms preferentially over the R-Smad homotrimer. This occurs despite
Smad4 not having a phosphorylated C-terminus, and thus not being able to donate
phosphoserine residues to the L3 loop/β8 strand of its neighboring R-Smad. Therefore
Smad4 is presumably able to contribute other favorable interactions that can compensate
for the loss of the interaction between the phosphorylated C-terminal tail and the loop-
strand pocket.

Therefore, to resolve the stoichiometric identity of the heteromeric R-
Smad/Smad4 complex, and understand the basis for its preferential formation over the R-
Smad homotrimer, we determined the crystal structure of both Smad3/Smad4 and
Smad2/Smad4 complexes. The structures, in corroboration with functional studies, reveal the presence of unique favorable electrostatic interactions within the heteromeric interfaces, and support a unifying mechanism of heterotrimeric Smad assembly, consisting of two R-Smads and one Smad4, in TGF-β signaling.

MATERIALS AND METHODS

Reagents

The pTXB intein expression vector, E. coli strain ER2566, and chitin sepharose beads were obtained from New England Biolabs. The phosphorylated peptides corresponding to Smad2 residues 463 to 467 (Cys-Ser-pSer-Met-pSer) or Smad3 residues 421 to 425 (Cys-Ser-pSer-Val-pSer) were obtained from the UCSF peptide sequencing facility.

Generation of phosphorylated Smad2 and Smad3

Phosphorylated forms of Smad2 and Smad3 were produced through an intein-mediated peptide ligation (IPL) method (Figure 4-1). PCR was used to generate fragments of Smad2 and Smad3 containing the linker and MH2 domains with the 5 C-terminal residues deleted (S3LCΔtail, residues 145-420 and S2LCΔtail, residues 186-462). PCR-generated flanking Ndel and Sapl restriction sites were used to subclone the fragments into the pTXB vector. E. coli ER2566 cells were transformed with S3LC(Δtail)- or S2LC(Δtail)-pTXB and grown in tryptone-phosphate media. Cells were grown at room temperature to an optical density (600 nm) of 0.5-0.7, and induced with 0.1-0.25 μM IPTG, then allowed to grow for an additional 25-36 hours. Cells were
Figure 4-1. Intein-mediated peptide ligation strategy used to generate phosphorylated Smad3. Adapted from Muir, 2003.
harvested, lysed through sonication, and centrifuged to obtain the clarified cell extract as described in Chapter II. The cell extract was applied a chitin sepharose column. After extensive washing of the column with chitin buffer (50 mM Na HEPES, 500 mM NaCl, 1 mM EDTA), the column was flushed with approximately 100 mL of chitin buffer containing 50 mM 2-mercaptoethanesulfonic acid (MESNA), and allowed to incubate for 16-40 hours at 4°C. Following the MESNA-induced cleavage, the cleaved protein was eluted from the chitin sepharose column with chitin buffer (+50 mM MESNA). After concentrating the eluted protein to ~20 mg/ml (~667 μM), the peptide corresponding to the phosphorylated C-terminus (C-S-pS-V-pS for Smad3 and C-S-pS-M-pS for Smad2) was added in 4-fold molar excess. Following overnight incubation at 4°C, phosphorylated (trimeric) protein was separated from unphosphorylated (monomeric) protein through size exclusion chromatography on the Superdex 200 column. Smad4AF (S4AF, residues 273-552) was expressed and purified as described in Chapter II.

Isothermal titration calorimetry

Binding of Smad4AF (273-552) to phosphorylated Smad2LC (186-467) or phosphorylated Smad3LC (145-425) was measured by isothermal titration calorimetry (Wiseman et al., 1989) using a VP-ITC calorimeter (MicroCal, Northampton, MA). All samples were dialyzed against ITC buffer (20 mM HEPES, pH 7.0, 100 mM NaCl, 0.1 mM EDTA, 1 mM TCEP), and degassed prior to titration. Experiments were initially performed at 25°C, but no binding enthalpy could be observed at this temperature. Subsequent experiments at 37°C produced a measurable binding enthalpy. 100 μM
Smad2 or Smad3 was titrated by addition of 300 μl of 400 μM Smad4 over 30 injections. Data was analyzed with Origin 7.0 software (MicroCal), assuming a one-site binding model. ΔH, ΔS, and Kᵣ values were experimentally determined, and ΔG was calculated from these values (ΔG = -RT ln Kᵣ = ΔH - TΔS). Heat capacity (ΔCₚ) values were determined for S3(LC, 2P) by performing additional ITC experiments at 34 and 40°C, and calculating ΔCₚ from (ΔΔH/ΔT).

**Differential scanning calorimetry**

Differential scanning calorimetry measurements were conducted on a VP-DSC calorimeter (MicroCal), with cell volumes of 0.5 mL. The protein concentration for all samples was 1 mg/ml. Reversibility of sample denaturation was examined through comparison of thermograms from consecutive DSC scans of the same sample. Scan-rate dependence, or susceptibility to time-dependent denaturation, was assessed by repeating DSC runs at varying scan rates ranging from 5-90°C/hr. All Smad protein constructs tested displayed irreversible thermal transitions and scan-rate dependence. Standard DSC buffer used in most experiments was 20 mM HEPES, pH 7.0, 100 mM NaCl, 0.1 mM EDTA, and 1 mM TCEP, although experiments were performed in a variety of buffer and denaturant conditions in an attempt to create thermodynamically reversible conditions. Samples were dialyzed extensively against the DSC buffer and degassed prior to experiment. Instrumental baselines were recorded by filling both cells with buffer, and were subtracted from the experimental traces to obtain heat capacity curves. Deconvolution analysis of DSC curves was performed using Origin software (MicroCal) with the assistance of Dr. Ashutosh Tiwari.
Crystallization of phosphorylated R-Smad/Smad4 complexes and structure determination

To form the heteromeric Smad complex, S4AF was added in 2-fold molar excess to phosphorylated S2LC or S3LC. Each heteromeric Smad complex was first purified by ion-exchange chromatography on a DEAE anion exchange column, during which the uncomplexed Smad4 eluted in the flow-through fractions, while the heteromeric Smad complex eluted in a 100-300 mM NaCl gradient. Since the resulting Smad complex could not be crystallized, it was subjected to limited chymotrypsin digest (1:100 w/w) for 24 hours on ice, and further purified by Superdex 200 size exclusion column (Amersham). N-terminal sequencing and mass spectrometry analysis revealed that the final heteromeric Smad complex lost most of the linker sequence as a result of the chymotrypsin digest. Smad3 extends from residue Glu228 to the C-terminus. Smad2 extends from the equivalent residue Glu270 to the C-terminus. Smad4 contains residues 309 to the C-terminus in two fragments, as a result of a single internal chymotrypsin cut after Leu484. Leu484 is located within a flexible, solvent accessible region connecting helix 3 and helix 4 in the Smad4 MH2 domain structure.

Crystals of the chymotrypsin-treated S3(LC,2P)/S4AF and S2(LC,2P)/S4AF complexes were obtained using the hanging drop vapor diffusion technique. The protein solution (15 mg/ml) was mixed with an equal volume of well solution containing 50 mM Tris-HCl (pH 7.5), 0-15 mM MgCl₂, and 5-15% ethanol. Crystals were transferred to a cryosolvent comprised of 24% glycerol and 76% well solution, and flash frozen in liquid nitrogen. Diffraction data for the S3(LC,2P)/S4AF complex were collected at the
BioCARS beamline at the Advanced Photon Source at Argonne National Laboratory. Diffraction data for the S2(LC,2P)/S4AF complex were collected at -170°C using an R-Axis IV image plate system mounted on a Rigaku rotating anode generator, at a detector distance of 200 mm with 1 oscillation per frame. Data were integrated and reduced using DENZO and Scalepack (Otwinowski and Minor, 1997). The structures were determined by molecular replacement using the CNS software package (Brunger et al., 1998), using the phosphorylated Smad2 MH2 domain (Wu et al., 2001b) as the search model. The search located three MH2 domains in the asymmetric unit, designated subunits A, B and C, which arrange in the same manner as the structures of the previously solved crystallographic homotrimers of Smad4 (Shi et al., 1997; Qin et al., 1999), Smad1 (Qin et al., 2001), and Smad2 (Wu et al., 2001b. Examination of the initial 2f_o-f_e electron density map revealed features indicating that subunits A and C correspond to Smad3 while subunit B corresponds to Smad4. For example, the electron density of the phosphorylated C-terminal sequence of Smad3 is readily visible in subunits A and C but is completely absent in subunit B. Also, extra electron density corresponding to the two-residue insertion in Smad4 (Ser344 and Cys345) between the β2 and β3 strands was present in subunit B but not in subunits A and C. The same observations were obtained when the Smad4 MH2 domain monomer was used as a search model to locate the three subunits. Rigid body refinement of the three MH2 domain subunits using all possible combinations of Smad3 and Smad4 also resulted in the same conclusion. The correct combination of pSmad3-Smad4-pSmad3, corresponding to subunit A-B-C, gave an initial
$R_{\text{free}}$ of 37.5%, while the next best combination, pSmad3-pSmad3-pSmad3, gave an initial $R_{\text{free}}$ of 40.7%.

The initial solutions were further refined through the CNS rigid body, simulated annealing, and B-factor refinement protocols (Brunger et al., 1998). Structures were rebuilt with CHAIN (Sack, 1998) and O (Jones et al., 1991).

**Size exclusion chromatography**

Size exclusion chromatography was used to analyze the interaction between S3LC(2P) or S2LC(2P) and S4AF, or various mutant forms derived thereof, as described in Chapter II.

**Immunoprecipitation and Western blots**

Immunoprecipitations were performed by Dr. Gen Shi. COS-1 cells were transfected with the indicated constructs with or without the activated type I TGF-β receptor point mutant (T204D) Alk5 CA (Wieser et al., 1995). After 24 hours, cells were switched to 0.1% serum overnight, and lysed in triton X lysis buffer (1% triton X, 150mM NaCl, 10% glycerol, 5mM EDTA and 25mM HEPES, pH 7.5) in the presence of protease and phosphatase inhibitors. Lysates were either directly separated by SDS-PAGE and transferred onto Immobilon-P membranes (Millipore), or first immunoprecipitated for 2 hours using epitope specific rabbit anti-Flag antibodies (Sigma). Immunoblots were performed using mouse 9E10 anti-Myc or anti-Flag M2 antibodies (Sigma), as indicated, and detected using the appropriate horseradish peroxidase-conjugated secondary antibody, and visualized by chemiluminescence (Pierce).
RESULTS

Generation of phosphorylated forms of Smad2 and Smad3 and isothermal titration calorimetry

Intein-mediated peptide ligation was used to generate phosphorylated forms of Smad2 and Smad3 (Materials and Methods; Figures 4-1 and 4-2A). Native-PAGE and size exclusion chromatography (Figure 4-2B and 4-2C) were used to confirm that ligation of the phosphorylated tail was successfully accomplished. The SEC elution profile of phosphorylated S2LC and S3LC was consistent with trimerization, with the peak occurring in fractions 14 and 15, compared to fraction 20 for the unligated proteins. The phosphorylated proteins were also tested for their ability to bind Smad4. SEC analysis of 1:1 mixtures of phosphorylated S2LC or S3LC with S4AF revealed that these proteins formed complexes with S4AF of apparent 2:1 stoichiometry, identical to the pseudophosphorylated forms of Smad3 and Smad1 described in Chapter II (Figures 4-2D, 2-11, and 2-12).

Although the experiments described above and in Chapter II strongly support that the R-Smad/Smad4 complex is a heterotrimer, this idea was contradicted by other studies. Size exclusion chromatography analysis of endogenous Smad2/Smad4 complexes from TGF-β-stimulated cells suggested that these complexes were heterodimers (Jayaraman et al., 2000). Smad2 and Smad3 were also discovered to have very different endogenous homo-oligomerization states in these studies. Using size exclusion chromatography and analytical ultracentrifugation methods similar to those
A  

[Image of SDS-PAGE and Native PAGE gels with Smad3LC (Δtail) indicated]

B  

<table>
<thead>
<tr>
<th>SDS-PAGE</th>
<th>Native PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>S3LC</td>
<td>S3LC</td>
</tr>
<tr>
<td>S3LC(2P)</td>
<td>S3LC(2P)</td>
</tr>
</tbody>
</table>

C  

[Graph showing assays with mAU on the y-axis and ml on the x-axis]
Figure 4-2. (A) Expression and purification of Smad3LC(Δ421-425). Lane 1: Molecular weight standards; Lane 2: Cell extract containing Smad3LC(Δ421-425)-intein-CBD fusion protein; Lanes 3 and 4: Washes following adsorption of the fusion protein to chitin column; Lane 5: Fusion protein on chitin beads after washing; Lane 6: Fusion and cleaved protein on chitin beads following 2-mercaptoethanesulfonic acid-mediated release of protein; Lane 7: protein as eluted from the chitin column.

(B) SDS-PAGE (left) and native-PAGE (right) of Smad3LC(Δ421-425) prior to and following the addition of the phosphorylated C-terminal tail. The shift in protein migration on native-PAGE following addition of the phosphorylated C-terminus indicates that ligation has been achieved.

(C) Size exclusion chromatography elution profile of Smad2LC(Δ463-467) prior to (dashed green line) and following (solid blue line) the addition of the phosphorylated C-terminal tail. The shift in protein oligomerization state is indicative of successful peptide ligation.

(D) SDS-PAGE of size exclusion chromatography of S3LC(2P) and S4AF, combined at a 1:1 protein ratio (approximately 50 μM each). Analysis of the fractions that eluted as a complex (fractions 15-17) through densitometry revealed that the ratio of S3LC(2P) to S4AF was approximately 2:1.
described in Chapter II, Wu and colleagues determined that the MH2 domains of Smad2 and Smad4 appeared to form heterodimers. However, an unphosphorylated version of Smad2 was used in those studies (Wu et al., 2001a). These studies may suggest a fundamental difference between the Smad3/Smad4 (heterotrimer) and Smad2/Smad4 (heterodimer) complexes, but analysis of nuclear extracts of cells transiently co-transfected with differentially-tagged Smads revealed the opposite. Smad2 and Smad4 formed a heterotrimer when bound to the cofactors Fast-1 or Fast-3 and DNA, while the DNA-bound Smad3/Smad4 complex was a heterodimer (Inman et al., 2002).

Isothermal titration calorimetry (ITC) was used to analyze the stoichiometry and the thermodynamics of binding of the phosphorylation-induced Smad3/Smad4 and Smad2/Smad4 interactions (Table 4-1; Figure 4-3). S4AF (400 μM) was titrated into the ITC sample cell containing either S2LC(2P) or S3LC(2P) (100 μM). Smad4 titrates both S2LC(2P) and S3LC(2P) at a relative molar ratio of approximately 0.5, confirming the stoichiometry observed from previous studies utilizing the pseudophosphorylated R-Smad mutants (Chapter II). The apparent dissociation constant (Kd) values are 60 nM for Smad3/Smad4, and 300 nM for Smad2/Smad4. The change in Gibbs’ free energy (ΔG) resulting from complex formation is ~10 kcal/mol for Smad3/Smad4, and ~9 kcal/mol for Smad2/Smad4 (Table 4-1). However, it is important to note that S2LC(2P) and S3LC(2P) exist as homotrimers prior to titration with Smad4, whereupon a monomeric Smad4 subunit displaces a subunit in the R-Smad homotrimer to form the R-Smad/Smad4 heterotrimer. Therefore the reaction heat that is detected by the calorimeter
Table 4-1. Summary of isothermal titration calorimetry of phosphorylated Smad3 and Smad2 with Smad4.

<table>
<thead>
<tr>
<th></th>
<th>Smad3/Smad4</th>
<th>Smad2/Smad4</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>0.50 ± 0.001</td>
<td>0.49 ± 0.011</td>
</tr>
<tr>
<td>ΔH (kcal/mol)</td>
<td>-7.5 ± 0.15</td>
<td>-8.2 ± 0.19</td>
</tr>
<tr>
<td>ΔS (cal/mol/K)</td>
<td>8.8 ± 0.57</td>
<td>3.7 ± 0.86</td>
</tr>
<tr>
<td>ΔG (kcal/mol)</td>
<td>-10.0 ± 0.06</td>
<td>-9.1 ± 0.35</td>
</tr>
<tr>
<td>Kd (nM)</td>
<td>58 ± 2.9</td>
<td>296 ± 165.8</td>
</tr>
<tr>
<td>T (K)</td>
<td>310</td>
<td>310</td>
</tr>
<tr>
<td>TAS (kcal/mol)</td>
<td>2.728</td>
<td>1.147</td>
</tr>
</tbody>
</table>

Figure 4-3. Isothermal titration calorimetry analysis of the interaction between S4AF and S2LC(2P) (left) and S3LC(2P) (right).
is a difference between association and dissociation events, and the actual $K_d$ and $\Delta G$ of the R-Smad/Smad4 interaction are likely to be more favorable than what is determined from the ITC experiment. The change in free energy resulting from complex formation is driven primarily by a highly favorable enthalpy ($\Delta H$), although the Smad3/Smad4 interaction has a significantly more favorable entropy ($T\Delta S$) term than Smad2/Smad4 (Table 4-1). The slight difference in the thermodynamics of binding between these two proteins, which are approximately 94% identical between their respective MH2 domains, is somewhat surprising, but may be partly explained by the approximately 16% difference in sequence between the linker regions of the two constructs. That heteromeric complex formation is enthalpically driven indicates that preferential interaction between R-Smads and Smad4 is primarily mediated by polar interactions (Fersht et al., 1985; Bhat et al., 1994; Ye and Wu, 2000).

**Crystal structures of phosphorylated Smad3(LC, 2P)/Smad4(AF) and phosphorylated Smad2(LC, 2P)/Smad4(AF)**

The constructs used to analyze the R-Smad/Smad4 interaction in these studies include the Smad MH2 domains, which mediate Smad oligomerization. The phosphorylated Smad3 and Smad2 constructs include the linker and MH2 domains. The phosphorylated Smad3 construct extends from residues 145-425 and is referred to as S3LC(2P). The phosphorylated Smad2 construct extends from residues 186-467 and is referred to as S2LC(2P). The Smad4 construct used comprises the MH2 domain and part of the linker domain, extending from residues 273-552, and is referred to as S4AF, as previously described in Chapters II and III of this thesis (Chacko et al., 2001). The
heteromeric Smad complex consisting of S3LC(2P) and S4AF was obtained by combining S4AF in 2-fold molar excess over S3LC(2P) in solution, then separating the complex from free S4AF through size exclusion chromatography. Attempts to crystallize this complex were unsuccessful. The complex was then subjected to a limited chymotrypsin digest (see Materials and Methods) and further purified through size exclusion chromatography. This chymotrypsin treatment produced a form of the complex that yielded diffraction-quality crystals. SDS-PAGE, N-terminal sequencing, and mass spectrometry analysis revealed that the final heteromeric Smad complex lost most of the linker sequence as a result of proteolysis. In the crystal structure, Smad3 extends from residue 228 to the C-terminus. Smad4 contains residues 309 to the C-terminus in two fragments, as a result of a single internal chymotrypsin cut after Leu484. Leu484 is located within a flexible, solvent accessible region connecting helix 3 and 4 in the Smad4 MH2 domain structure (Figure 4-4). This region is presumably also flexible in the heteromeric Smad complex, as indicated by its susceptibility to chymotrypsin digestion. An identical method was used to produce a crystallizable form of the Smad2/Smad4 complex. Both structures were solved through molecular replacement, using the previously solved structure of phosphorylated Smad2 (Wu et al., 2001b) as a search model (see Materials and Methods and Table 4-2). The Smad3/Smad4 structure was refined to 2.5 Å, and the Smad2/Smad4 structure was refined to 2.8 Å. The crystal structures reveal that both protein complexes are heterotrimeric, consisting of two R-Smad subunits and one Smad4 subunit (Figure 4-5). This is in agreement with our prior
Figure 4-4. Structure-based sequence alignment of the MH2 domains of the Smad proteins. The secondary structural motifs are shown below the alignment. Residues in Smad4 and Smad2 known to be mutated in cancer are highlighted in yellow. Sites of phosphorylation are shown in green. Residues coordinating the phosphorylated tail are highlighted in red. The Smad4-specific insertion is located between helix 3 and helix 4 (residues 450-488). Smad4 residues 456-489 are missing in the S2/S4 and S3/S4 structures described in this chapter. Adapted from Wu et al., 2001b.
Table 4-2. Summary of crystal analysis for Smad2(LC,2P)/Smad4(AF) and Smad3(LC,2P)/Smad4(AF)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>S2(LC,2P)/S4AF</th>
<th>S3(LC,2P)/S4AF</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Crystal parameters and crystallographic data</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Space group</td>
<td>P2(1)2(1)2(1)</td>
<td>P2(1)2(1)2(1)</td>
</tr>
<tr>
<td>Unit cell dimensions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>a = 49.24</td>
<td>a = 48.47</td>
<td></td>
</tr>
<tr>
<td>b = 59.95</td>
<td>b = 60.55</td>
<td></td>
</tr>
<tr>
<td>c = 207.44</td>
<td>c = 205.78</td>
<td></td>
</tr>
<tr>
<td>Diffraction limit (Å)(^a)</td>
<td>100-2.8 (2.9-2.8)</td>
<td>100-2.5 (2.6-2.5)</td>
</tr>
<tr>
<td>Total reflections</td>
<td>78952</td>
<td>89892</td>
</tr>
<tr>
<td>Unique reflections</td>
<td>15933</td>
<td>16539</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>85.3</td>
<td>96.5</td>
</tr>
<tr>
<td>Intensity/Sigma</td>
<td>14.4</td>
<td>17.7</td>
</tr>
<tr>
<td>R(_{\text{merge}}) (%)(^b)</td>
<td>7.7</td>
<td>5.2</td>
</tr>
</tbody>
</table>

**Refinement statistics**

| Protein atoms | 4777 | 4985 |
| R factor (%)\(^c\) | 24.3 | 23.2 |
| R\(_{\text{free}}\) factor (%)\(^d\) | 27.9 | 24.6 |
| Rms deviation from ideal | | |
| Bond lengths (Å) | 0.0106 | 0.0079 |
| Bond angles (°) | 1.590 | 1.658 |
| B-factor rms deviation | | |
| Main chain (Å\(^2\)) | 2.960 | 1.445 |
| Side chain (Å\(^2\)) | 4.410 | 1.742 |

\(^a\)Values in brackets are for the highest resolution shell.
\(^b\)R\(_{\text{merge}}\) = \(\sum I_{hkl} - <I_{hkl}>/\sum I_{hkl}\)
\(^c\)R factor = \(\sum_{hkl} |F_{\text{obs}}| - |F_{\text{calc}}| / \sum_{hkl} |F_{\text{obs}}|\) for all data.
\(^d\)R\(_{\text{free}}\) = \(\sum_{hkl} |F_{\text{obs}}| - |F_{\text{calc}}| / \sum_{hkl} |F_{\text{obs}}|\) for 10% of the data not used in refinement.
biochemical analysis that showed, using acidic amino acid substitutions to mimic phosphorylation, that the Smad3/Smad4 and Smad1/Smad4 complexes were heterotrimers (Figure 2-11; Chacko et al., 2001; Qin et al., 2001). The R-Smad subunits in both crystal structures were designated subunits A and C, and the Smad4 subunit was designated subunit B. These three subunits arrange in a manner similar to the structures of the previously solved homotrimers of Smad4 (Shi et al., 1997; Qin et al., 1999), Smad1 (Figure 3-6; Qin et al., 2001), and Smad2 (Wu et al., 2001b) (Figure 4-5)

**Overall Structure of the R-Smad/Smad4 Heterotrimer and Interface Analysis**

Each of the subunits of the R-Smad/Smad4 heterotrimeric complex consists of a central β sandwich flanked on one side by three helices (three-helix bundle) and on the other side by three loops and another helix (loop-helix region). The previously solved homotrimeric structures of Smad1 (Qin et al., 2001) and Smad2 (Wu et al., 2001b) revealed that subunit-subunit contact is mediated specifically by the interaction of helix 4, helix 5, and the phosphorylated C-terminal tail of one subunit with helix 1, the β4 strand, and the L3 loop/β8 strand pocket, respectively, of a neighboring subunit. This general feature of interface formation is conserved in the R-Smad/Smad4 heterotrimer, with the exception that there are only two phosphorylated C-terminal tail interactions in the heterotrimers, one between subunit A and B and the other between subunit C and A, due to Smad4 lacking a phosphorylated tail (Figure 4-7 and 4-8). In addition, Smad4 possesses a unique conformation that results in closer contacts between subunits B and C.
Figure 4-5. Overall structure of the R-Smad/Smad4 heterotrimeric complexes. (A) Smad2(LC, 2P)/Smad4(AF) complex. (B) Smad3(LC, 2P)/Smad4(AF) complex. The R-Smad subunits are designated A and C. The L3 loops are in yellow, with the C-terminal sites of serine phosphorylation depicted in stick representation.
in the heterotrimer. Superposition of the heterotrimer subunits to those of the previously solved Smad2 homotrimers (Wu et al., 2001b) reveals that the three-helix bundle of Smad4 subunit in the heterotrimer undergoes a significant shift toward the neighboring R-Smad subunit when compared to the corresponding subunit in the Smad2 homotrimer (Figure 4-6). The unique conformation of Smad4 likely results in improved contacts between subunits B and C interactions that contribute to the preferential formation of the R-Smad/Smad4 heterotrimer over the R-Smad homotrimer.

To gain insights into the structural mechanism of preferential heterotrimer formation, the interface contacts of the heterotrimeric crystal structures were compared with those of the previously solved crystal structure of the Smad2 homotrimer. The presence of Smad4 in the heterotrimer results in three interfaces that are distinct from each other, designated AB (between subunit A and B), BC (between subunit B and C), and CA (between subunit C and A). This is in contrast to the Smad2 homotrimer, which contains three identical interfaces. Since the structures of the Smad3/Smad4 and Smad2/Smad4 heterotrimeric complexes are virtually identical, with an overall rms deviation of only 1.5 Å, most of the structural analysis described below was done on the higher resolution Smad3/Smad4 structure. Overall, there appear to be fewer hydrogen bond interactions in the BC and AB interfaces compared to the homotrimeric interfaces, although all three heterotrimeric interfaces possess a common subset of hydrogen bond interactions that are also present in the homotrimeric structure (Table 4-3). Most notably, Smad4 possesses unique amino acid residues not present in the R-Smads that result in new electrostatic interactions in the BC and AB interfaces. These favorable electrostatic
Figure 4-6. Superposition of the alpha carbon traces of the Smad2 homotrimer (gray) and the Smad3/Smad4 heterotrimer. The Smad3 subunits of the heterotrimer are depicted in green, the Smad4 subunit in blue. (Inset) Close-up view of helix 3 and helix 4 of the Smad4 (B) subunit of the Smad3/Smad4 heterotrimer and the superimposed Smad2 homotrimer. Helix 3 is located above helix 4. Helix 5 of both structures is removed for clarity.
Table 4-3. Hydrogen bond comparison between the Smad2 homotrimer interfaces and the three Smad3/Smad4 heterotrimer interfaces. Hydrogen bonds that are conserved between the R-Smad homotrimer and R-Smad/Smad4 heterotrimer interfaces are boxed. Residues that are conserved in all four interfaces are shown in red. Interactions that are unique to one interface are marked with an asterisk. A soft cut-off distance of 3.5 Å was used.
interactions appear to compensate for the fewer hydrogen bonds, accounting for the favorable enthalpy change upon hetero-oligomerization.

**Asp493 of Smad4 coordinates a buried electrostatic interaction in the BC Interface**

The smaller number of hydrogen bonds in the BC interface is due largely to Smad4 lacking a phosphorylated C-terminal tail; therefore the interaction between the phosphorylated tail and L3 loop/β8 strand pocket that is characteristic of the R-Smad homotrimeric interfaces is not present at this interface (Figures 4-7 and 4-8). Despite the lack of a phosphorylated tail, and the corresponding reduction in number of hydrogen bonds, the most striking feature of this interface is a network of salt bridges and hydrogen bonds centered on Asp493 of helix 4 of the Smad4 (B) subunit. Asp493 is situated in the center of four arginine residues, Arg321/279 and Arg329/287 (in this convention, the residue number for Smad2 is on the left, Smad3 is on the right) of the C subunit, and Arg496 and Arg497 of the B subunit (Figures 4-8 and 4-9). Asp493 is able to form intermolecular salt bridges with the two arginine residues of the C subunit, as well as an intramolecular salt bridge with Arg496. In addition, the backbone oxygen of Asp493 forms an intramolecular hydrogen bond with Arg497 (Figure 4-8). Asp493 appears to coordinate the positive charge contributed by the four surrounding arginine residues, maintaining a charge balance that allows the interface to form. Asp493 of Smad4 is frequently found to be mutated in pancreatic and other cancers (Hahn et al., 1996), suggesting that this residue has a crucial role in mediating Smad-dependent signaling.
Figure 4-7. Representative interface from the Smad2 homotrimer structure solved by Wu et al., 2001b. The three helix bundle of the C subunit is shown in green, and the interacting loop-helix region of the A subunit is shown in orange. This interface is identical between all three subunits in R-Smad homotrimers.
Figure 4-8. (A) Smad3/Smad4 BC interface. The Smad4 (B) subunit is depicted in blue, and the Smad3 (C) subunit is depicted in purple. (B) Close-up view of the helix 4-helix 1 interaction in the Smad2 homotrimer (left) and the BC interface of the Smad2/Smad4 heterotrimer (right). The presence of the aspartic acid at position 493 of Smad4 (as opposed to Tyr406 of Smad2), coupled with the shift in helix 4, results in highly favorable electrostatic interactions between Asp493 and four surrounding arginine residues. This results in significantly improved interface contacts over those in the Smad2 homotrimeric interface.

Figure 4-9. Electrostatic representation of D493 and surrounding residues of the B subunit (left) and the adjacent interacting residues in the C subunit (right). The lone aspartic acid residue at position 493 of Smad4 forms intermolecular salt bridges with two arginines of the C subunit (R279 and R287), an intramolecular salt bridge with R496, and an intramolecular hydrogen bond with R497.
To examine the role of Asp493 in the R-Smad/Smad4 interaction, the ability of a D493A mutant to form heterotrimeric complexes was analyzed through size exclusion chromatography and ITC. As opposed to wild-type S4AF, S4AF(D493A) did not coelute with S3LC(2P) or S2LC(2P) on the size exclusion column, indicating that the Smad4(D493A) mutant is unable to interact with phosphorylated R-Smads to form heterotrimers (Figure 4-10A). The potency of this mutation is underscored by ITC experiments, which showed that S4AF(D493A) has very little interaction ($K_d = 2.9 \, \mu M$, $\Delta H = -3 \, \text{kcal/mole}$) with S3LC(2P) (Figure 4-10B). These results were further confirmed by immunoprecipitation experiments. COS cells were transfected with full-length versions of Smad2 or Smad3, with full-length WT Smad4 or the Smad4(D493A) mutant, in the presence or absence of a constitutively active TGF-β type I receptor (Wieser et al., 1995). Immunoprecipitation of Smad2 or Smad3 revealed that the Smad4(D493A) mutant did not associate with either R-Smads, unlike wild-type Smad4 (Figure 4-11). As discussed in Chapter II, previous biochemical analysis has suggested a dimeric model for phosphorylation-dependent Smad2/Smad4 interaction, consisting of only subunit A and B. This model would predict that Asp493 of Smad4 would not play a role in heteromeric Smad assembly, as it is located on the surface that is not in contact with the R-Smad. The critical role of Asp493 of Smad4 revealed here strongly opposes the heterodimeric model. The heterodimer between unphosphorylated Smad2 and Smad4 observed in other studies (Wu et al., 2001a) could be the product of the highly favorable interaction between the B and C subunits in the crystal structure. It is likely that the potent electrostatic interaction in the BC interface allows formation of an intermediate
heterodimeric complex even in the absence of R-Smad phosphorylation, while phosphorylation drives the complex toward trimerization.

The residue in the position equivalent to Asp493 in the R-Smads is Tyr406/364. This residue forms a hydrogen bond with Arg329/287 in the R-Smad homotrimer (Figure 4-7; Table 4-3). Of the four arginine residues that interact with Asp493 of Smad4, three are conserved in the R-Smads (Figures 4-7 and 4-8B). Therefore it appears possible that replacement of the R-Smad helix 4 tyrosine residue (Y364) with aspartic acid might approximate the heterotrimeric BC interface, thus resulting in a stronger homotrimeric interaction. To test this possibility, we mutated Tyr364 to aspartic acid in unphosphorylated S3LC, and analyzed the behavior of this protein on the size exclusion column (Figure 4-12). Wild-type S3LC elutes as an apparent monomer, with the peak occurring at fraction 19. The S3LC(Y364D) mutant formed high molecular weight aggregates, with the elution peak at fraction 9. The S3LC(Y364D) mutant was also incapable of forming heteromeric complexes with Smad4 (data not shown). The results suggest that this sole tyrosine to aspartic acid mutation is sufficient to produce a significantly stronger homomeric interaction. Although the Y364D mutation did not result in trimer formation, it is likely that the presence of the buried aspartic acid residue in the homomeric interface results in a strong charge interaction with surrounding arginines, leading to aggregate formation. Taken as a whole, the data show that the electrostatic interactions mediated by Asp493 of Smad4 contribute to the preferential R-Smad/Smad4 heterotrimer formation over R-Smad homotrimer.
Figure 4-10. (A) SEC analysis of interaction between S4AF or D493A mutant with S3LC(2P) or S2LC(2P). (B) ITC analysis of S3LC(2P)/S4AF(D493A) interaction.
**Figure 4-11.** Co-immunoprecipitation analysis of Smad2/Smad4 and Smad3/Smad4 interaction. COS-1 cells were transfected with full-length wild-type versions of Flag-tagged Smad2 or Smad3 with or without Myc-tagged Smad4 or the Smad4 (D493A) point mutant, and with or without the constitutively active TGF-β type I receptor point mutant (T204D) Alk5 CA (Wieser et al., 1995). Complexes were immunoprecipitated from cell extracts with an anti-Flag antibody and stained as indicated.

**Figure 4-12.** Size exclusion chromatography analysis of wild-type Smad3(LC) and S3LC(Y364D). Fraction numbers are indicated on the bottom. The Y364D mutation induces the formation of high molecular aggregates in Smad3.
In addition to the favorable electrostatic interactions, the BC interface also displays better interface packing compared to the R-Smad homotrimeric interfaces. Interface shape complementarity was measured using the gap volume index statistic from the Protein-Protein Interaction Server (Jones and Thornton, 1996). This statistic is a ratio of the interface gap volume to interface buried surface area. More complementary interfaces are indicated by a smaller gap volume index. The gap volume index value for the heterotrimeric BC interface is 2.55, compared to 2.48 for the Smad2 homotrimeric interface. However, the slightly smaller value for the homotrimer is due largely to the favorable interactions between the phosphorylated C-terminal tail and the L3 loop/β8 strand pocket. Deletion of the 5 C-terminal residues of R-Smads results in a homotrimer interface gap volume index of 3.35. This indicates that the core interaction of helices 4 and 5 of Smad4 and helix 1 and the β4 strand of Smad3 in the heterotrimer is significantly tighter than the equivalent interaction in the R-Smad homotrimer. The improvement in surface complementarity of the BC interface is due in part to a shift in the helices 4 and 5 of the three-helix bundle of Smad4 in the heterotrimer as compared to the homotrimer, as described above (Figure 4-6).

Arg378 of Smad4 reinforces phosphoserine binding in the AB Interface

The AB interface of the R-Smad/Smad4 heterotrimer also contains fewer hydrogen bonds than the R-Smad homotrimeric interfaces, primarily due to the lack of hydrogen bond interactions between helix 4 (A subunit) and helix 1 (B subunit), which are present in all Smad interfaces except this one (compare Figures 4-7, 4-8, and 4-16 with 4-13). The unique feature of this interface is the presence of an extensive hydrogen
**Figure 4-13.** The AB interface of the Smad3-Smad4 crystal structure. The three-helix bundle of the Smad3 A subunit is depicted in green. The loop-helix region of the Smad4 B subunit is depicted in blue. The lack of hydrogen bonds between helix 4 and helix 1 is seen only in this interface, but is partly compensated for by the hydrogen bonds mediated by R378 of helix 1 with the phosphorylated tail of the A subunit.
bond interaction network between Arg378 (helix 1) of the Smad4 subunit and the tail phosphoserine residue at position 467/425 of the A subunit. This interaction occurs in the R-Smad/Smad4 heterotrimer, but not in the R-Smad homotrimer because the corresponding arginine residue is not conserved in the R-Smads (Figure 4-4). The presence of Arg378 appears to make Smad4 a better receptor for the phosphorylated C-terminal tail of the adjacent R-Smad subunit (A subunit) by increasing the basic electrostatic potential surface of the phosphoserine binding pocket (Figure 4-14). This additional helix 1/phosphorylated tail interaction likely compensates for the lack of hydrogen bonds between helix 4 and helix 1.

The importance of Arg378 was confirmed by mutating Arg378 to alanine and analyzing the effect of the mutant on heteromeric interaction with the phosphorylated R-Smads. S4AF(R378A) is able to only weakly interact with S3LC(2P), as evidenced by its decreased co-elution with S3LC(2P) and S2LC(2P) on the size exclusion column, compared to wild-type S4AF (Figure 4-15A). This result was supported by ITC analysis, which showed that the ability of the S4AF(R378A) mutant to interact with phosphorylated Smad2 or Smad3 was substantially reduced from the wild-type protein (K_a = 790 nM) (Figure 4-15B).

The overall packing of the AB interface is similar to that of the homotrimeric interface as indicated by the gap volume index statistics. The gap volume index for the AB interface is 2.42, which is comparable to the value for the homotrimeric interface. Despite the similarity in the packing statistics, the AB interface is slightly more non-polar than the other interfaces. While the hydrophobic content of the BC, AC or the R-Smad
Figure 4-14. Representation of surface electrostatic potential of phosphorylated C-terminal tail interaction with L3 loop/β strand pocket in (A) Smad3/Smad4 AB interface and (B) Smad2 homotrimer interface. The presence of R378 in helix 1 of Smad4 results in an improved interaction with the phosphorylated tail.
Figure 4-15. (A) SEC analysis of interaction between S4AF or R378A mutant with S3LC(2P) or S2LC(2P). (B) ITC analysis of S3LC(2P)/S4AF(R378A) interaction.
homomeric interfaces are between 50 and 53%, the AB interface has a 57% hydrophobic content, due to the higher percentage of hydrophobic residues contributed by Smad4. Thus it is possible that the hydrophobic effect in the AB interface also contributes to favorable heteromeric Smad assembly. Two residues that contribute to the slight increase in hydrophobicity of the AB interface are Val354 (L1 loop, Thr 303/261 in Smad 2 and Smad3) and Leu381 (Helix 1, Arg 330/288 in Smad2 and Smad3). Both of these residues mediate hydrophobic contacts in the AB interface of the heterotrimer that are not present in the equivalent positions of the homotrimer interfaces. Leu381 of helix 1 of Smad4 sits in a pocket surrounded by R-Smad residues Phe402/360 and Tyr 406/364 of helix 4 (data not shown). Therefore, although there are no hydrogen bonds between helix 4 and helix 1 of the heterotrimeric AB interface, this is at least partially compensated by the presence of unique hydrophobic interactions between the two helices. In addition, Val 354 in loop 1 of Smad4 forms a van der Waals network with three leucine residues (Leu 446/404, 449/407, and 453/411) of the neighboring R-Smad helix 5.

The CA Interface Resembles the Homotrimeric Interface

The CA interface formed by two R-Smad subunits is the only interface that is conserved in both the homotrimer and heterotrimer, and is essentially identical in both structures. The hydrogen bond networks that mediate the interaction of helix 5 and the phosphorylated tail of the C subunit with the β4 strand and L3 loop/β8 strand pocket of the A subunit are nearly the same between homotrimer and heterotrimer (Figures 4-7 and 4-15A). A subtle difference in the two interfaces is found in the hydrogen bonds present between helix 4 (C subunit) and helix 1 (A subunit). The hydrogen bond that exists
between Gln407 (helix 4) of the C subunit and Arg330 (helix 1) of the A subunit of the Smad2 homotrimer is not present in the heterotrimer, but is replaced by an apparent salt bridge between Glu403/361 (Smad2/Smad3) of the C subunit with Arg330/288 of the A subunit (Figure 4-15B). Gln407/365 remains involved in the heterotrimeric interface by forming an intramolecular hydrogen bond with Glu403/361. In the homotrimer, Glu403/361 is positioned toward the solvent, and therefore has no role in the homotrimeric AC interface. In the heterotrimer, however, Glu403/361 undergoes a conformational change that causes its side chain to flip approximately 6 Å to come in contact with Arg330/288 of the adjacent A subunit. However, mutation of Glu361 to an alanine residue does not result in a significant defect in the phosphorylation-induced R-Smad/Smad4 interaction as determined through gel filtration. ITC analysis revealed only a 2 to 3-fold increase in $K_d$ compared to wild-type (data not shown). The result indicates that the conformational change of Glu403/361 in the AC interface, and its associated change in hydrogen bonding pattern, does not have a salient effect on the preferential formation of the heterotrimer over the homotrimer. This suggests that the primary drivers of preferential R-Smad/Smad4 heterotrimer formation are the electrostatic interactions mediated by the Smad4 residues Asp493 and Arg378 of the BC and AB interfaces, respectively.
Figure 4-16. (A) The CA interface of the S3LC(2P)/S4AF heterotrimer. (B) Comparison between the R-Smad/R-Smad helix 4-helix 1 interaction in the Smad2 homotrimer (left) and the equivalent interaction in the Smad2/Smad4 heterotrimer (right). The flipping of E403/361 results in a novel salt bridge interaction at this interface.
Structural stabilization of R-Smads by Smad4 in the heterotrimer

The thermal stability of Smad3 and Smad4, both individually and in complex, was assessed through differential scanning calorimetry (DSC) (Figure 4-17). The scan range used was 5 to 95°C. Initial DSC scans of S3LC and S3LC(2P), as well as S4AF, revealed that both proteins displayed irreversible denaturation after heating to 95°C. Furthermore, both proteins displayed scan rate dependence, i.e., the DSC thermogram varied widely depending on temperature increment used. Because of the thermal irreversibility and scan rate dependence of the proteins, equilibrium thermodynamic analysis could not be performed. Therefore the data obtained could not be used to calculate the ratio of calorimetric enthalpy ($\Delta H_{\text{m}}$), and the unfolding transition curves could only be compared in terms of peak transition temperature ($T_m$). We determined that heating the proteins at a rate of 5-10°C/hr resulted in optimal reproducibility and differentiation between the various samples.

DSC analysis of unphosphorylated S3LC (residues 145 to 425) revealed that this protein had a $T_m$ of 48.2°C (Figure 4-17B). The phosphorylated form, S3LC(2P), displayed an identical peak $T_m$ of 48.2°C, however, the emergence of a trailing left edge (compare “Smad3 monomer” and “Smad3 trimer” peaks in Figure 4-16A) indicated that decomposition of the Smad3 trimer occurred prior to unfolding of the protein (Figure 4-17C). Deconvolution analysis revealed that the trimer to monomer transition occurred at approximately 45.9°C. DSC analysis of the Smad4 construct S4AF (residues 273 to 552), revealed that it unfolded at a significantly higher temperature ($T_m= 59.4$).
Figure 4-17. (A) Overlay of heat capacity curves of unphosphorylated S3LC (magenta), phosphorylated S3LC(2P) (black), S3LC(2P)/S4AF heterotrimeric complex (blue), and S4AF (red), (B)-(E) Heat capacity curves of unphosphorylated S3LC, phosphorylated S3LC(2P), S4AF, and the heterotrimeric S3LC(2P)/S4AF complex. Deconvolution of each curve is shown in red.
than the Smad3 constructs (Figure 4-17D).

The increased thermal stability of Smad4 relative to Smad3 appears to be due to stronger intramolecular interactions in Smad4. It is possible that increased hydrophobic interactions play a role, as the S4AF construct is composed of approximately 13% more nonpolar residues than S3LC. As the numerous Smad crystal structures have shown, the MH2 domains of R-Smads and Smad4 are very similar structurally, comprising a core β-strand sandwich that is capped by a three-helix bundle on one end and a loop-helix region on the other end (Shi et al., 1997; Qin et al., 1999; Wu et al., 2001b; Qin et al., 2001). Aside from the phosphorylated tail, the major structural difference between the two Smad classes is the presence in Smad4 of a specific insertion between helix 3 and helix 4, rich in glutamine and alanine residues (Qin et al., 1999), and the increased thermal stability of S4AF relative to S3LC may be a product of this insertion as well. Additional analysis will be required to determine the precise basis for this phenomenon.

Analysis of the thermal unfolding of the (phosphorylated) S3LC(2P)/S4AF heterotrimeric complex reveals that the presence of Smad4 in the complex contributes an increased stability to the Smad3 subunits within the complex (Figure 4-17E). The heat capacity curve reveals two distinct unfolding events, Smad3 unfolds first, followed by Smad4. Deconvolution of the first peak on the curve indicates that it is the sum of two distinct endothermic processes, the breakdown of the heterotrimeric complex and the thermal unfolding of phosphorylated Smad3. The $T_m$ of the deconvoluted curve that corresponds to the thermally-induced dissociation of the Smad3/Smad4 complex appears
to occur at 49.1°C. Furthermore, the T_m of unfolding for Smad3 in the heterotrimeric complex is 50.6°C, up from 48.2°C in the Smad3 homotrimer. These results appear to show that Smad4 imparts significant thermal stability to the Smad3 subunits in the heterotrimeric complex, increasing both the T_m of trimer dissociation as well the unfolding temperature of Smad3. Given that Smad4 is intrinsically more thermostable than Smad3, it appears that Smad4 may be serving as a sort of heat sink, drawing external heat away from Smad3, allowing Smad3 to exist in a folded state slightly longer than it can as a monomer or homotrimer. More study will be needed to assess whether the increased thermal stability of the R-Smad/Smad4 heterotrimer over the R-Smad homotrimer is a function of increased energetic preference for its formation.

DISCUSSION

We have shown in the studies presented in this chapter that both the Smad2/Smad4 and Smad3/Smad4 complexes are heterotrimers comprising two R-Smad subunits and one Smad4 subunit. This is the first structural demonstration of the heterotrimeric Smad assembly by X-ray crystallography. Structural and functional analysis of the Smad heterotrimer suggests that favorable electrostatic interactions within the BC and AB interfaces play a significant role in the preferential formation of heterotrimer over the R-Smad homotrimer. These electrostatic interactions can exist only in the heterotrimer due to specific charged residues in the Smad4 subunit, Asp493 and Arg378, which mediate complementary electrostatic interactions with the neighboring R-Smad subunits. Furthermore, the unique conformation of Smad4, where the three-helix
bundle structure is tilted more towards the neighboring R-Smad subunit, likely facilitates better subunit-subunit contacts in the heterotrimer.

The stoichiometry of the heteromeric Smad complex and the structural basis for its preferential formation revealed here can likely be generalized to the Smad family of proteins as a whole. Although only the TGF-beta and activin-specific Smads, Smad2 and Smad3, are analyzed here, previous biochemical analysis of a BMP-specific Smad, Smad1, also revealed a 2:1 ratio between pseudophosphorylated Smad1 and Smad4 (Chapter II; Qin et al., 2001). This hypothesis is supported by the structural similarity of the R-Smads and the universal role of Smad4 as a common mediator in Smad signaling. Furthermore, the residues in Smad2 and Smad 3 that form complementary electrostatic interactions with Asp493 and Arg378 of Smad4 are identical in all R-Smads.

Although this study suggests a general model of heterotrimeric Smad assembly in TGF-β signaling, the possibility that other forms of heteromeric Smad assembly can exist 

*in vivo* in the presence of other signaling partners cannot be ruled out. For example, Inman and colleagues (Inman et al., 2002) have reported, using gel supershifting of differentially tagged R-Smads by antibodies, that while the DNA-bound Smad2/Smad4 complex is a heterotrimer when bound to the co-factors Fast-1 or Fast-3, the Smad3/Smad4 complex under similar conditions is a heterodimer. Also, Jayaraman and Massague reported, based on size-exclusion chromatography, that the endogenous Smad2/Smad4 complex from TGF-β stimulated cells is heterodimeric. The difference in these results from our own may represent a difference between the core R-Smad/Smad4 complex, and one that has translocated into the nucleus and is bound to various cofactors.
These results suggest the possibility that Smad cofactors, upon binding the R-Smad/Smad4 complex in the nucleus, can disrupt the core heterotrimer in such a way that one of the Smad subunits is knocked out, resulting in a Smad heterodimer. It has been proposed that the transcription factor menin, which is known to interact with the MH2 domain of Smad3, might be an example of such a co-factor (Inman et al., 2002; Kaji et al., 2001). The structural homology with the interferon regulatory factor IAD domain suggests that Smad MH2 domains may be able to interact with these domains, or with other proteins containing FHA domains (Moustakas and Heldin, 2003). A recent report has shown a direct physical interaction between Smad3 and IRF-7 (Qing et al., 2004), raising the possibility that IRF-7 might replace one of the Smad3 subunits to form a Smad3/Smad4/IRF-7 complex, which would produce the Smad3/Smad4 heterodimer observed by Inman and Hill (Inman and Hill, 2002). Finally, the FHA-containing protein SNIP, which has been shown to interact with both Smad1 and Smad2, may also be an example of this type of protein (Kim et al., 2000). The structural homology between the Smad MH2 domain and the IRF IAD domain suggests the possibility that these domains may also be able to interact.

It should also be noted that both immunoprecipitation and size exclusion analyses are subject to potential artifact. The antibody approach assumes that the epitope tags are all equally exposed. However, it is possible that lack of a supershift in the case of the DNA bound Smad3/Smad4 complex may be the consequence of a tag being buried in the DNA-bound transcriptional complex. The size-exclusion chromatography approach alone, in general, cannot be used to unambiguously determine the oligomerization state,
as the result is affected by both the shape and the mass of the macromolecular complex. Further work is necessary to investigate the nature of novel heteromeric Smad complexes, and the structural basis for their assembly.

Given the apparent simplicity of the Smad system, it is intriguing that the TGF-βs can mediate such a diversity of biological activity. The formation of a heterotrimeric Smad complex consisting of two R-Smads and one Smad4 may provide part of the answer, as this assembly mechanism greatly increases the potential diversity of the Smad signaling system. From a structural standpoint, it is feasible for the heterotrimeric Smad complex to include two different R-Smad molecules, as the trimer interface among the R-Smads is highly conserved. Such a combinatorial mechanism can greatly increase the number of distinct signaling species with different Smad constitutions. Each of these distinct Smad complexes may be able to perform different tasks, capable of recognizing and activating of a specific set of gene promoters. This combinatorial mechanism of Smad heterotrimerization is consistent with several biological observations, which have shown that active Smad complexes can consist of more than one R-Smad (Nakao et al., 1997; Labbe et al., 1998; Feng et al., 2000; Goumans et al., 2003).

Finally, Smad trimerization can also increase its specificity for promoter recognition. The promoters of TGF-β responsive genes generally contain Smad binding elements (SBE), each recognized by the N-terminal MH1 domain of an R-Smad or Smad 4 (Zawel et al., 1998; Shi et al., 1998). Smad2 is an exception since its MH1 domain contains an insertion that blocks DNA recognition. Thus, a heterotrimeric Smad complex consisting of two R-Smads and one Smad4 can contain anywhere from one to three
functional DNA binding domains, depending on the number of Smad2 subunits present. This combinatorial mechanism of Smad assembly may explain in part why the TGF-β responsive promoters contain different number of copies of SBE. The MH1 domains within the heterotrimeric Smad complex can also be at different relative distances to their MH2 domains due to each Smad having a unique linker domain length. This asymmetric property of the heterotrimeric Smad complex may explain in part the different spacing of SBE observed in the TGF-β responsive promoters. Future studies on how different heterotrimeric Smad complexes contact DNAs and transcriptional co-regulators will further clarify the role of Smad trimerization in TGF-β signaling.
CHAPTER V: GENERAL DISCUSSION

This thesis examined the phosphorylation-induced R-Smad/Smad4 complex, which propagates the signals of the TGF-β family of ligands. The experiments in Chapter II were designed to assess the stoichiometry of R-Smad/Smad4 complexes through the use of pseudophosphorylated versions of two different R-Smads, Smad3 and Smad1, in biochemical analyses. Smad3 is involved exclusively in TGF-β and activin signaling, while Smad1 participates only in BMP signaling. These two Smads were chosen to determine whether mechanistic differences might exist between the TGF-β/activin and BMP pathways. Constructs of both proteins lacked the MH1 domain, consisting only of the linker and MH2 domains. The Smad4 construct used in these experiments, S4AF, consisted of the MH2 domain and the C-terminal portion of the linker domain, as this had been previously defined as the minimal Smad4 fragment required to drive Smad-dependent transcription (de Caestecker et al., 2000a). As detailed in Chapter II, size exclusion chromatography and analytical ultracentrifugation revealed that both the Smad3/Smad4 and Smad1/Smad4 complexes were heterotrimers, consisting of 2 R-Smad subunits and 1 Smad4 subunit. This result was later confirmed by ITC and by crystallography of the Smad2/Smad4 and Smad3/Smad4 complexes. This is a confirmation of the prior results of Kawabata and colleagues (Kawabata et al., 1998), who determined that R-Smad/Smad4 complexes were heterotrimeric through immunoprecipitation of epitope-tagged Smads. Our results contradict findings of Jayaraman and Massague (Jayaraman and Massague, 2000), who analyzed Smad complexes from TGF-β stimulated cells through size exclusion chromatography, and
determined these complexes were heterodimers. More directly, our results contradict studies by Wu and colleagues (Wu et al., 2001a; Wu et al., 2001b), who used similar size exclusion chromatography and analytical ultracentrifugation analysis of purified Smad constructs and found that the Smad2/Smad4 complex was a heterodimer. The primary difference between these studies and our own was that Wu and colleagues used unphosphorylated Smad2 constructs, while our studies were done using pseudophosphorylated versions of Smad3 and Smad1. It is our opinion that the heterodimer seen in the studies of Wu and colleagues was the formation of the heteromeric BC interface, as described in Chapter IV. The electrostatic interactions at this interface are so strong that partial complex formation can occur even in the absence of phosphorylation. It appears that phosphorylation then allows the complex to fully trimerize.

More recent analysis of the heteromeric R-Smad/Smad4 complex by Inman and Hill produced contradictory results. Through immunoprecipitation of DNA-bound R-Smad/Smad4 heteromers in complex with other transcriptional co-activators, they determined that the Smad2/Smad4 complex was a heterotrimer (2 Smad2 to 1 Smad4), while the Smad3/Smad4 complex was a heterodimer (Inman and Hill, 2002). We think it is likely that this result reflects a difference between the core R-Smad/Smad4 complex that forms immediately upon phosphorylation and the complex that activates transcription in the nucleus. Much like Smad4 is able to knock out an R-Smad subunit out of the R-Smad homotrimer to form the R-Smad/Smad4 heterotrimer, a nuclear co-factor might be able to knock out another R-Smad subunit out of the heterotrimer to form a new R-
Smad/Smad4/co-factor heterotrimer. Such a complex would appear to be a heterodimer from the supershift analysis employed by Inman and Hill. Some potential candidates for this type of co-factor include menin, which has been demonstrated to interact with the Smad3 MH2 domain (Inman and Hill, 2002; Kaji et al., 2001), and the interferon regulatory factors (IRFs). IRF-7 has been demonstrated to directly interact with Smad3 (Qing et al., 2004), and possesses a similar core β-strand sandwich to the Smad MH2 domain, which further supports the possibility that this protein can replace an R-Smad subunit in the R-Smad/Smad4 heterotrimer.

The studies described in Chapter III were designed to determine a specific mechanism for the phosphorylation-induced Smad trimer formation. It had been established that part of the mechanism through which C-terminal phosphorylation effects the activation of R-Smads is by decreasing the affinity of the MH1 domain for the MH2 domain, thereby relieving autoinhibition (Hata et al., 1997). However, it is also clear from the experiments in Chapter II that phosphorylation also serves to energetically drive the Smads toward trimerization. The constructs used in the studies in Chapter II lacked the autoinhibitory MH1 domain, yet phosphorylation was still required to fully effect trimerization. Based on prior crystallographic analysis of Smad4(AF) (Qin et al., 1999), we proposed that phosphorylation might induce Smad trimerization by creating an interaction between the phosphorylated C-terminal tails and positively-charged pockets on adjacent Smads. This was confirmed by mutational analysis which revealed that altering an arginine residue in the L3 loop disrupted the heteromeric R-Smad/Smad4 interaction. Subsequent solution of a Smad1 crystal structure revealed that the C-
terminal sites of serine phosphorylation were proximal to a basic surface defined by the L3 loop and β8 strand, and supported the idea that C-terminal phosphorylation would create favorable electrostatic interactions between the phosphorylated tail and basic surface. Mutation of residues in the basic surface disrupted the heteromeric R-Smad/Smad4 interaction and supported this notion. It is also noteworthy that the L3 loop is a primary structural mediator of R-Smad interaction with the receptor, and that a basic surface defined partly by the L3 loop is responsible for interacting with the phosphoserine residues on the GS domain of the receptor kinase, suggesting that the Smad MH2 might be a general phosphoserine-binding domain. This was supported by structural comparison of the Smad MH2 to the FHA domain of Rad53 (Durocher et al., 2000). The FHA domain is a known phosphoserine and phosphothreonine-binding domain, and possesses a similar central β-strand sandwich as the Smad MH2.

Comparison of the structure of homotrimeric Smad1 with a SARA-bound (monomeric) Smad2 revealed that the three-helix bundle and L3 loop assumed different conformations between monomer and trimer, suggesting that phosphorylation may induce allosteric structural changes in R-Smads, allowing it to use a common basic surface to mediate phosphoserine-binding events with the receptor kinase and other Smads.

The structural basis for the preferential formation of the R-Smad/Smad4 heterotrimer over the R-Smad homotrimer was addressed in Chapter IV. The crystal structures of the Smad2/Smad4 and Smad3/Smad4 complexes confirmed the stoichiometry that had been observed through other means, and also revealed that specific residue differences between R-Smads and Smad4 led to the formation of highly favorable
electrostatic interactions that result in the R-Smad/Smad4 heterotrimer being energetically favored over the R-Smad homotrimer. Specifically, the presence of aspartic acid (D493) in helix 4 of Smad4, as opposed to a corresponding tyrosine residue in R-Smads, results in highly favorable electrostatic interactions between the aspartic acid residue and four surrounding arginine residues at the heterotrimeric BC interface. These interactions at the BC interface are facilitated by a shift in the three-helix bundle of Smad4 toward the neighboring helices in R-Smads. In addition, the presence of an arginine residue (R378) in helix 1 of Smad4, replacing a methionine or leucine residue in R-Smads, results in a novel interaction between helix 1 of Smad4 and the R-Smad phosphorylated tail. Differential scanning calorimetry analysis revealed that Smad4 is a considerably more thermostable protein than Smad3, and that the presence of Smad4 in the complex raising the melting temperature of Smad3. This suggests another mechanism for explaining how the R-Smad/Smad4 complex is energetically favored over the R-Smad homotrimer; that the presence of Smad4 imparts stability onto the R-Smad subunits in the heterotrimeric complex.

There are many directions in which the research described herein can be logically taken. The idea that a transcriptional co-activator can replace one of the R-Smad subunits in the R-Smad/Smad4 heterotrimer is an exciting possibility. It would be of interest to determine the identity of these co-factors, and ultimately obtain structural knowledge of these R-Smad/Smad4/co-factor complexes. That others have seen heterodimeric Smad complexes (Jayaraman and Massague, 2000; Inman and Hill, 2002) presumes that this is the mechanism accounting for the existence of both heterotrimeric
and heterodimeric Smad complexes, but it is possible that another mechanism may be responsible, and it would be of interest to define whether another such mechanism exists.

Also, given that the L3 loop has been implicated in both Smad binding to receptor (Lo et al., 1998; Huse et al., 2001) and, from our work, in Smad/Smad interactions, it would be intriguing to determine whether this same positively charged surface is responsible for mediating other phosphoserine-binding events. A structure of an R-Smad/receptor interaction would also be of interest, providing information about the mechanism through which the L3 loop is able to allosterically direct these distinct binding events.
REFERENCES


