

**UCSF**

**UC San Francisco Electronic Theses and Dissertations**

**Title**

Structure and plasticity of protein-protein interfaces in factor Xa and the androgen receptor

**Permalink**

<https://escholarship.org/uc/item/4q61s66h>

**Author**

Hur, Eugene

**Publication Date**

2006

Peer reviewed|Thesis/dissertation

Structure and Plasticity of Protein-Protein Interfaces  
in Factor Xa and the Androgen Receptor

by

Eugene Hur

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

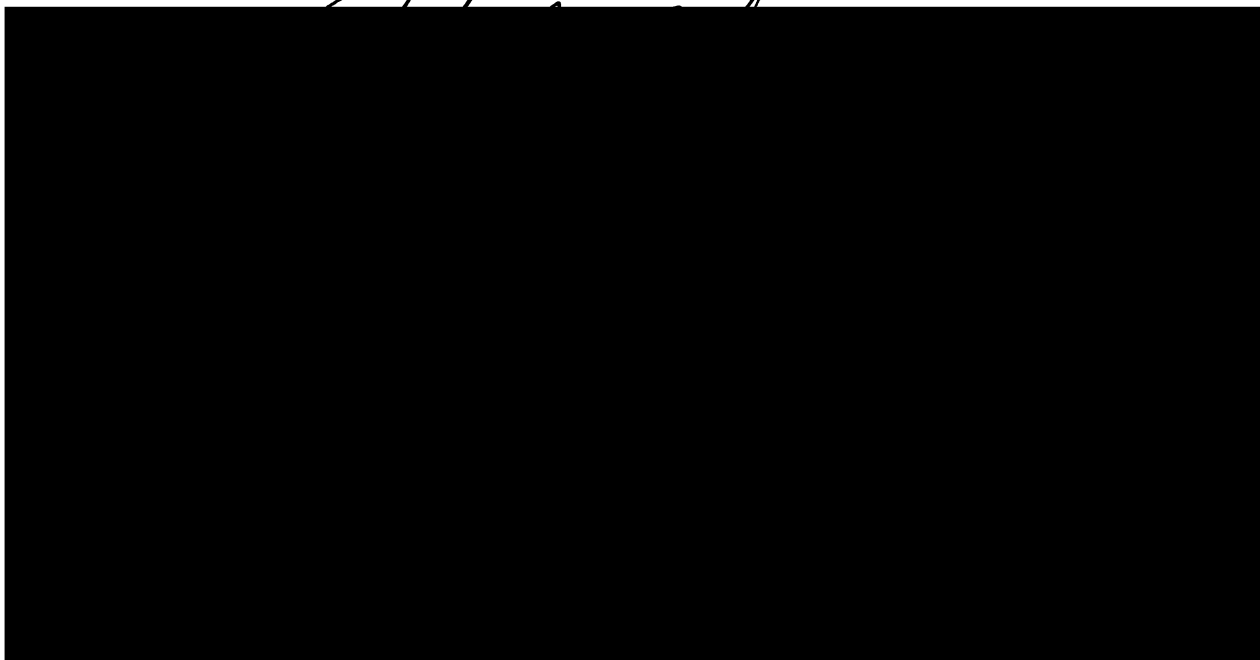
BIOPHYSICS

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, SAN FRANCISCO



Copyright 2006  
by  
Eugene Hur



*To*  
*Mom and Dad*

## PREFACE

Writing this work has given me the opportunity to reflect on my time at UCSF. As I look back, the past six years have been a time of great personal transition. Graduate school has been a long, hard journey, yet exactly because of that, I feel I've grown immeasurably on both a professional and personal level. While I'm still the same person that I was when I started, in many ways that person is forever gone, reincarnated into someone better, more enlightened. I am excited to see what the future holds for this new version of myself.

While often times graduate school can seem like a dark, lonely place, only rarely is the journey a completely solitary one. Over the years I have been helped by numerous people without whose support I would not be writing this now. For those people I have the most sincere feelings of gratitude.

First and foremost among them is my thesis advisor, Robert Fletterick. After a difficult first year, where I often questioned whether graduate school was really for me, Robert welcomed me into the lab with open arms and fostered a warm, supportive environment to pursue my graduate career. Robert always seemed to have the best interests of the people in his lab in mind and that is something that I very much hope to emulate if I ever have my own lab. By working in Robert's lab, I learned not only how to do science, but I also learned how to keep things in perspective, and to keep life balanced. This was reinforced every year on the annual lab ski trip - I'll never forget the time the lab went to Utah, really, just for the heck of it! In fact if there is one thing that I have learned from Robert over the years, it is to always look at the big picture - I've found this to be tremendously useful both inside and outside of the lab.

While Robert was the captain providing leadership and direction, it was the lab who I spent the most time with, and with whom I endured many long days and nights. Over the

years, the people in the Fletterick Lab have been wonderful coworkers, mentors, and friends. Because of them I'll always look back at my graduate school days with fond memories. In particular though, I must thank Stephanie Wang, who took me under her wing when I did a rotation in the lab, and then continued to dispense invaluable advice throughout the years even after she left the lab; Elena Sablin, who was always willing to put up with my questions and be a patient consultant to whatever project I was working on; Peter Hwang, whose kind, patient tutelage and Zen-like approach to science and life has always been an inspiration. Our discussions about everything from music to skiing to squash were always welcome diversions from science; Jennifer Turner, Sabine Borngräber, Eva Estébanez-Perpiñá, and Maia Vinogradava for serving as my in-house life-advisors and therapists. With them I shared both my happiest moments and my darkest lows - without their kind words, compassion, and advice, graduate school would have truly been a dark place; and finally, Chuck Sindelar, Mary-Jane Budny, Sam Pfaff, and Jeremy Wilbur for always brightening my day and making the lab such a fun place to work.

Of course, none of this would have been possible without the fantastic environment at UCSF and the Biophysics program. Julie Ransom was always available to lend an ear to problems and her kind advice through the years (and her god-like ability to wade through all the bureaucratic nonsense at the Graduate Division) always made life easier. I also have to thank the other professors who helped guide me over the past six years: Marc Shuman was a fantastic collaborator on the uPAR project and was pivotal during my early years in the lab. My orals committee, Tom Scanlan, Dave Agard, Holly Ingraham, and Lily Jan put me through a hair-raising hour of the most intense questioning I have ever experienced, but I am grateful for the experience and for their input. Holly, along with Kip Guy served on my

thesis committee and their scrutinizing advice has been invaluable over the past couple of years.

I also must mention my wonderful classmates. I don't think I could have asked for a better group of people to start graduate school with. Without their support, there is no way I would have made it through the first year. In particular I have to thank Mona Sridharan-Kulp, Toral Surti, Alex Schnoes, and Keiko Petrosky for always keeping me balanced and grounded; Adrian Keatinge-Clay for his endless enthusiasm about science and fun discussions in the hallways; and Felix Lam for always keeping me entertained and for being such a great friend through the years.

Last, but not least, I have to thank my parents. Their endless support and patience over the years has gotten me to where I am now - for that I am forever grateful. Thank you.



Structure

## ABSTRACT

Like the  
proteins is this  
be able to recog  
accommodating  
partners in the c  
paradox: how can  
are the mechanis  
more important  
role in the cell? A  
antagonize prote

To better  
cascade, the stru  
macromolecular  
association. Des  
ecotin binds FX  
specific seconda

# Structure and Plasticity of Protein-Protein Interfaces in Factor Xa and the Androgen Receptor

Eugene Hur

## ABSTRACT

Like the cells they operate, proteins are dynamic, plastic entities. Nowhere in proteins is this more evident than in protein-protein interfaces. In many cases, proteins must be able to recognize not just one partner, but many. The ability to be flexible and accommodating, yet at the same time discriminate against the multitude of unwanted partners in the cell is what allows proteins to make life possible. However, this presents a paradox: how can protein recognition be both specific and promiscuous? In particular, what are the mechanisms that allow broad-specificity in a protein-protein interface, and perhaps more importantly, what insights can this give us toward understanding a protein's biological role in the cell? Answering these questions can potentially lead to new compounds that antagonize protein-protein interfaces and cure disease.

To better understand these issues in the context of the proteolytic blood coagulation cascade, the structure of factor Xa (FXa) was solved in complex with a M84R variant of the macromolecular protease inhibitor ecotin. The structure reveals the atomic mechanism of association. Despite not possessing the recognition sequence of canonical FXa substrates, ecotin binds FXa with pico-molar affinity through a combination of induced fit and non-specific secondary site interactions.

Toward

cancer, the struc

that mimic the

structural basis

able to accomm

provide a frame

Finally, t

extreme. Like m

partners - foldin

AR N-terminal c

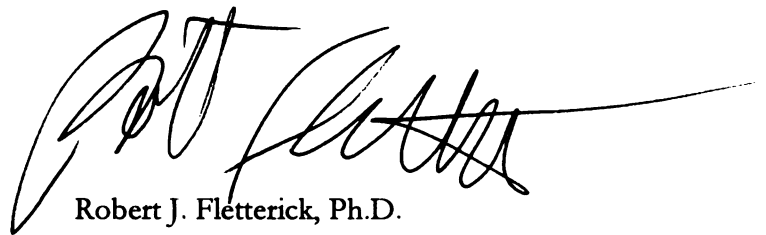
coexpression and

interaction of the

NC6A-62.

Toward the aim of developing alternative anti-androgen compounds for prostate cancer, the structure of the androgen receptor (AR) was solved in complex with peptides that mimic the receptors' interactions with coactivators. The complexes reveal both the structural basis of AR's preference for coactivators containing FxxLF motifs, and how AR is able to accommodate motifs containing other hydrophobic residues as well. The structures provide a framework for the design of AR antagonists.

Finally, the AR-NTD and its coregulators represent protein plasticity at its most extreme. Like many transcription factors, they are disordered in the absence of binding partners - folding is induced by association with partner proteins. To understand how the AR N-terminal domain (NTD) interacts with its coregulator partners, a high-throughput coexpression and domain mapping system was developed. This system was used to map the interaction of the orphan receptor LRH-1 and one of its recently discovered coactivators, NCoA-62.

A handwritten signature in black ink, appearing to read 'R. J. Fletterick', with a long horizontal line extending to the right.

Robert J. Fletterick, Ph.D.

Chair, Thesis Committee

## TABLE OF CONTENTS

<b>Chapter 1</b> .....	<b>1</b>
The Extended Interactions and Gla Domain of Blood Coagulation Factor Xa	
<b>Chapter 2</b> .....	<b>28</b>
Recognition and Accommodation at the Androgen Receptor Coactivator Binding Interface	
<b>Chapter 3</b> .....	<b>66</b>
The AR N-terminal domain and ART-27	
<b>Chapter 4</b> .....	<b>109</b>
LRH-1 and NCoA-62	
<b>Chapter 5</b> .....	<b>122</b>
The Urokinase Plasminogen Activator Receptor	
<b>References</b> .....	<b>132</b>

## LIST OF TABLES

<b>Table 1-1</b> .....	<b>27</b>
Data and Refinement Statistics for the Ecotin M84R-FXa Complex	
<b>Table 2-1</b> .....	<b>34</b>
Rate and Dissociation Constants for the Interaction between the AR LBD and Selected Peptides	
<b>Table 2-2</b> .....	<b>64</b>
Summary of Structures and Crystallographic Statistics for AR-peptide complexes	
<b>Table 3-1</b> .....	<b>72</b>
Gateway Coexpression Vectors	

## LIST OF FIGURES

<b>Figure 1-1</b> .....	9
Ribbon diagram of the tetrameric FXa-ecotin M84R complex	
<b>Figure 1-2A</b> .....	14
Surface representation of the extended interactions between FXa and ecotin M84R	
<b>Figure 1-2B</b> .....	15
Extended interactions between FXa and ecotin M84R shown as a Ligplot	
<b>Figure 1-3</b> .....	16
The S2 and S4 substrate binding pockets of M84R-FXa	
<b>Figure 1-4</b> .....	19
A). Alignment of partial sequences of major serine proteases from the coagulation pathways	
B). Residue 192 and its interactions with the P2 main-chain carbonyl	
<b>Figure 1-5</b> .....	22
The FXa Gla domain	
<b>Figure 1-6</b> .....	23
Superpositions of the Gla domain from various proteases	
<b>Figure 2-1</b> .....	34
AR LBD–Interacting Peptides Selected by Phage Display	
<b>Supplementary Figures</b> .....	36
A). Crystals of the AR LBD-FxxLF peptide complex	
B). Overall structural features of the AR LBD-FxxLF complex	

C). Electron density of the FxxLF peptide bound to the AR LBD

<b>Figure 2-2</b> .....	37
A structural profile of the AR coactivator binding interface	
<b>Figure 2-3</b> .....	42
Interactions of FxxLF and LxxLL with the AR LBD	
<b>Figure 2-4</b> .....	43
Induced fit of the AR AF-2 interface	
<b>Figure 2-5</b> .....	46
Interactions of the tryptophan motifs with the AR LBD	
<b>Figure 2-6</b> .....	47
Interactions of Ser-2 with Glu897	
<b>Figure 2-7</b> .....	49
Interactions of FxxYF and FxxFF with the AR LBD	
<b>Figure 2-8</b> .....	56
Sequence alignment of the AF-2 region of nuclear receptors	
<b>Figure 2-9</b> .....	57
Surface complimentarity of hydrophobic motifs in the AR, ER $\alpha$ , and GR AF-2 clefts	
<b>Figure 3-1</b> .....	78
Gateway primers	
<b>Figure 3-2</b> .....	85
Expression of GST-ART-27	



<b>Figure 3-3</b> .....	85
Coexpression of 6xHis-ART-27 with 6xHis-NTD(142-498)	
<b>Figure 3-4</b> .....	89
Codon usage of ART-27	
<b>Figure 3-5</b> .....	92
Expression of codon optimized GST-ART-27	
<b>Figure 3-6</b> .....	93
Expression of codon optimized MBP-ART-27	
<b>Figure 3-7</b> .....	94
Coexpression of MBP-ART-27 with GST and 6xHis NTD(142-448)	
<b>Figure 3-8</b> .....	97
Cleavage of MBP-ART-27 with TEV protease	
<b>Figure 3-9</b> .....	98
Phenyl HP and HiTrap Q purification of MBP-ART-27	
<b>Figure 3-10</b> .....	100
MBP and 6xHis pulldowns of ART-27 and NTD(142-448)	
<b>Figure 3-11</b> .....	104
Expression of NTD(142-448) in Rosetta 2 cells	
<b>Figure 3-12</b> .....	106
Expression of codon optimized NTD(142-448)	
<b>Figure 3-13</b> .....	107
Expression of codon optimized 6xHis-NTD(142-448) at 37°C	

<b>Figure 3-14</b> .....	<b>108</b>
Talon purification of 6xHis-NTD(142-448) under denaturing conditions	
<b>Figure 4-1</b> .....	<b>113</b>
Expression of MBP-NCoA-62 domains	
<b>Figure 4-2</b> .....	<b>117</b>
Expression of MBP-NCoA-62 fragments	
<b>Figure 4-3</b> .....	<b>118</b>
Coexpression of LRH-1 with NCoA-62 fragments	
<b>Figure 4-4</b> .....	<b>120</b>
Interaction of NCoA-62 fragments with mLRH-1 shown by ELISA	
<b>Figure 5-1</b> .....	<b>129</b>
Overall structure of the uPA-uPAR complex	
<b>Figure 5-2</b> .....	<b>130</b>
Electron density of the uPA-uPAR interaction	
<b>Figure 5-3</b> .....	<b>131</b>
Superposition of 2FD6 with uPAR2	

# Chapter 1

## The Extended Interactions and Gla Domain of Blood Coagulation Factor Xa

Coauthors on this chapter:

Stephanie X. Wang, Carolyn A. Sousa, Linda Brinen, Eric J. Slivka, Robert J. Fletterick

Reproduced with permission from:

Hur E, Wang SX, Sousa CS, Brinen L, Slivka EJ, Fletterick RJ (2003) *The extended interactions and Gla domain of blood coagulation factor Xa*. *Biochemistry* 42(26):7959-66. Copyright 2003

American Chemical Society.

## ABSTRACT

The serine protease factor Xa (FXa) is inhibited by ecotin with pico-molar affinity. The structure of the tetrameric complex of ecotin variant M84R (M84R) with FXa has been determined to 2.8 Å. Substrate directed induced fit of the binding interactions at the S2 and S4 pockets modulates the discrimination of the protease. Specifically, the Tyr at position 99 of FXa changes its conformation with respect to incoming ligand, changing the size of the S2 and S4 pockets. The role of residue 192 in substrate and inhibitor recognition is also examined. Gln 192 from FXa forms a hydrogen bond with the P2 carbonyl group of ecotin. This confirms previous biochemical and structural analyses on thrombin and activated protein C which suggested that residue 192 may play a more general role in mediating the interactions between coagulation proteases and their inhibitors. The structure of ecotin M84R-FXa (M84R-FXa) also reveals the structure of the Gla domain in the presence of  $Mg^{2+}$ . The first eleven residues of the domain assume a novel conformation and likely represent an intermediate folding state of the domain.

## INTRODUCTION

Blood vessel trauma initiates a proteolytic cascade of serine proteases that culminates in the cross-linking of the clotting protein fibrin and the formation of blood clots. The coagulation cascade is composed of two interconnected pathways: the extrinsic and the intrinsic, which converge in the activation of zymogen factor X (FX) to functional factor Xa (FXa). This activation occurs either by the factor VIIa (FVIIa)-tissue factor complex (extrinsic pathway) or by the factor IXa (FIXa)-factor VIII complex (intrinsic pathway). Mature FXa consists of two disulfide linked polypeptide chains: a light chain composed of an N-terminal  $\gamma$ -carboxyglutamate (Gla) containing Gla domain, followed by two epidermal-growth-factor-like (EGF) domains, and a heavy chain harboring a trypsin-like serine protease domain (Davie et al. 1991; Padmanabhan et al. 1993). In the presence of its cofactor, factor Va (FVa), FXa cleaves the zymogen prothrombin to generate active thrombin protease.

Due to its dual role as both the final enzyme of the cascade and a positive feedback regulator of the intrinsic pathway, thrombin has historically been the major protease target of the blood coagulation cascade in the development of anti-coagulant therapies. However, because thrombin is so ubiquitous in the blood coagulation response, many thrombin inhibitors have had low safety to efficacy profiles in clinical trials, with their usage often associated with an increased risk of bleeding complications (Kaiser 2002).

The search for alternative targets has led to FXa. Like thrombin, it is strategically placed to regulate both the intrinsic and extrinsic pathways, but because of its more restricted activity, it is thought that inhibiting the upstream FXa would be more efficient and less likely to elicit the side-effects seen with thrombin inhibitors. However, similarities of structure and substrate selection among the serine proteases of the blood coagulation cascade have challenged the design of specific inhibitors against FXa (Davie et al. 1991;

Strabb

investu

crystal

al 1993.

Magnan

macrom

bond has

substrate s

useful in in

The

broad range

and collagen

that inhibition

domain swap

et al 1997; Wa

two distinct sit

interaction is ch

site cleft of the

involving the c

protease.

E:cotin

$K_d$  of 54 pM

increases inhib:

determined the

Stubbs and Bode 1994). Interactions between FXa and its inhibitors have not been investigated as thoroughly as interactions between thrombin and its inhibitors. X-ray crystallography has been used to define several FXa-inhibitor complexes (Padmanabhan et al. 1993; Brandstetter et al. 1996; Kamata et al. 1998; Wei et al. 1998; Adler et al. 2000; Maignan et al. 2000). However, no structure of FXa with a canonically bound macromolecular inhibitor showing the substrate-like interactions at both sides of the scissile bond has been determined. Because of the importance of extended interactions in FXa substrate specificity (Rezaie and Yang 2001), such a complex may prove to be especially useful in inhibitor design.

The *E. coli* macromolecular protease inhibitor ecotin (Chung et al. 1983) inhibits a broad range of serine proteases of the chymotrypsin fold including trypsin, chymotrypsin, and collagenase (Seymour et al. 1994; Ulmer et al. 1995). Previous structural analyses showed that inhibition by ecotin occurs by the formation of a tetrameric complex consisting of a domain swapped ecotin dimer binding two protease molecules (McGrath et al. 1994; Perona et al. 1997; Waugh et al. 2000; Wang et al. 2001b). Each ecotin molecule binds its target at two distinct sites on the protease: a primary site and a secondary site. The primary site interaction is characterized by substrate-like binding of the 80's loop of ecotin to the active site cleft of the protease. Binding at the secondary site involves less specific interactions involving the distally located 60's loop of ecotin with a flat, hydrophobic patch on the protease.

Ecotin is the most potent reversible inhibitor of FXa, with an inhibition constant ( $K_i$ ) of 54 pM (Seymour et al. 1994). An M84R mutation in the P1 residue of ecotin further increases inhibition by five fold to 11 pM (Seymour et al. 1994). We have crystallized and determined the x-ray structure of human FXa in complex with ecotin M84R (M84R). The

canonical binding site  
extended into  
of thrombin  
binding. War

The s

determined in  
N-terminal d  
factor VII, IN  
Gla domain is  
translationally  
of  $Ca^{2+}$ , the C

the membrane  
proteins, as b  
injury but also  
substrates. Da

Previous

from full length  
1996; Kamata  
2001). The Gla  
structure in the  
Gla domain in  
high affinity, lo  
1976; Prendergast



canonical binding mode of the ecotin 80's loop provides valuable insights into the role of extended interactions in FXa substrate recognition. Comparisons with the existing structure of thrombin-M84R reveal how ecotin adapts to different proteases to achieve high affinity binding (Wang et al. 2001b).

The structure of FXa-ecotin M84R (M84R-FXa) is also the first structure of FXa determined in the presence of an ordered Gla domain. The Gla domain is a highly conserved N-terminal domain found in many of the proteins involved in blood coagulation including factor VII, IX, X, proteins C, S, Z and prothrombin. About 45 amino acids in length, the Gla domain is characterized by 9-12  $\gamma$ -carboxyglutamates which are added post-translationally by a vitamin K-dependent carboxylase (Furie and Furie 1988). In the presence of  $\text{Ca}^{2+}$ , the Gla domain binds negatively charged phospholipids, anchoring the protein to the membrane surface. This is thought to be critical for the proper functioning of these proteins, as binding to the membrane not only assists in localizing these enzymes to sites of injury but also helps properly position them for interaction with activators, cofactors and substrates (Davie et al. 1991).

Previously, the Gla domain of FXa was either disordered or proteolytically cleaved from full length FXa to aid in crystallization (Padmanabhan et al. 1993; Brandstetter et al. 1996; Kamata et al. 1998; Wei et al. 1998; Adler et al. 2000; Maignan et al. 2000; Nar et al. 2001). The Gla domain of M84R-FXa, however, was well ordered. This represents the first structure in the presence of an ordered Gla domain as well as the first crystal structure of the Gla domain in the presence of  $\text{Mg}^{2+}$ . Three bound  $\text{Mg}^{2+}$  ions define the locations of three high affinity, low specificity cation binding sites predicted in earlier studies (Nelsestuen et al. 1976; Prendergast and Mann 1977; Bloom and Mann 1978; Borowski et al. 1986). The

structure thus provides a glimpse of an intermediate folding state of the domain and provides key insights into how the Gla domain attains the membrane bound state.

## **MATERIAL AND METHODS**

### **Purification and Crystallization of the FXa-ecotin M84R Complex**

Ecotin M84R was purified by standard procedures as described (McGrath et al. 1991; Yang et al. 1998), and subsequently sent to Haematologic Technologies Inc. (Essex Junction, VT) to be complexed with purified FXa from human plasma. Equimolar quantities of FXa and ecotin M84R were incubated on ice before being purified by gel filtration in the presence of 150 mM NaCl, 2 mM MgCl<sub>2</sub>, and 20 mM Tris at pH 7.4.

The purified M84R-FXa complex was concentrated to 5.4 mg/ml and crystallized at room temperature by the sitting drop method. Initial crystals were obtained by the PEG/Ion screen (Hampton Research) with a well solution containing 20% PEG3350, and 0.2 M sodium potassium tartrate at pH 7.1. Final crystals were optimized by adding a trace amount of glycerol (to a final concentration of 2%) and 1  $\mu$ l of MPD to 4  $\mu$ l of well solution and 5  $\mu$ l of protein solution.

### **Data Collection, Structure Determination, and Refinement**

Orthorhombic crystals of M84R-FXa diffracted to 2.8 Å. A full data set was collected at the Advanced Light Source (ALS) beam line 5.0.2 with an ADSC Quantum 4 CCD detector. Data sets were integrated and merged using SCALEPACK/DENZO (Otwinowski and Minor 1997a). The symmetry of crystals of M84R-FXa is described by the space group *I*222, with  $a = 66.8\text{Å}$ ,  $b = 108.0\text{Å}$ ,  $c = 186.3\text{Å}$ . Data statistics indicated that one

FXa and one ecotin molecule were present per asymmetric unit. Solvent content was estimated to be 58%.

The structure of M84R-FXa was solved using molecular replacement with rotation and translation functions from CNS 1.0 (Brunger et al. 1998). Data to 3.5 Å were used in all rotation searches. Initially, a hetero-dimer search model of FXa (from PDB: 1FAX) bound to ecotin (from PDB:1ID5) was constructed based on the structure of trypsin-ecotin (McGrath et al. 1994) using Insight II (Accelrys, San Diego, CA). The light chain of FXa was omitted in this model for molecular replacement. This hetero-dimer model, however, did not result in any rotation solutions that could be improved by further translation searches. A second approach was undertaken involving a two step rotation search. Rotation solutions from a first search using the heavy chain of FXa were used to fix the orientation of FXa in a second search using ecotin as the search model. The top 10 solutions from the second rotation search were then subjected to translation searches and rigid body refinement using data to 2.8 Å to produce a starting model of M84R-FXa with an initial R value of 43.5%. The final structure was refined to 2.8Å with a R of 20.7% and a  $R_{free}$  of 23.4% using CNS 1.0/1.1 (Brunger et al. 1998) and Quanta 2000 (Molecular Simulations Inc., San Diego CA). Detailed data and refinement statistics are listed in Table 1-1. The coordinates of M84R-FXa have been deposited in the Protein Data Bank under accession code 1POS.

Buried surface areas between protein domains were calculated by the program NACCESS (Hubbard and Thornton 1993). Superposition data were calculated with LSQMAN (Kleywegt 1996). The relative orientations between superimposed molecules were calculated by the program GEM (Browner et al. 1992). Figure 1-1 was made with Molscript and Raster3D (Merritt and Bacon 1997). Figure 1-2B was generated with Ligplot (Wallace et al. 1995). All other figures were made with PyMOL (DeLano 2002).

RESULT

Overall S

L

molecules.

the active s

binding int

heavy chain

nearly perpe

1B). The GL

neighboring s

to be disorder

density maps.

refinement sta

from the mod

Overall

protease comp

ecotin M84R r

the major ang

## RESULTS AND DISCUSSION

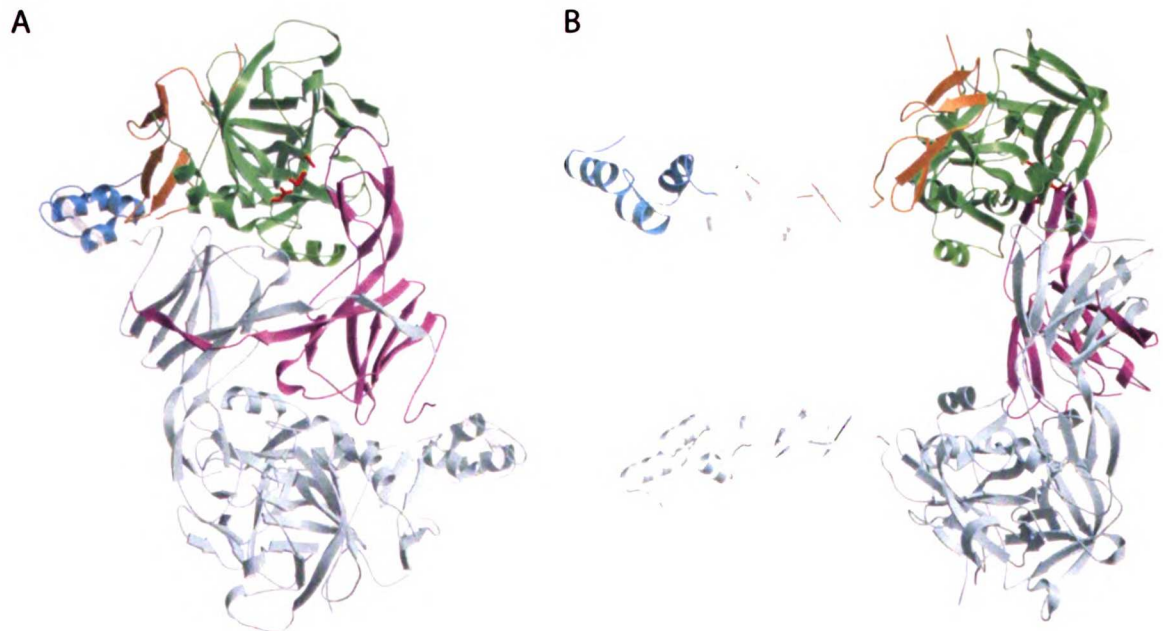
### Overall Structural Features of the FXa-Ecotin M84R Complex

Like other ecotin-protease complexes, ecotin M84R binds as a dimer to two FXa molecules, forming a tetramer via two distinct binding sites on the protease: a primary site at the active site and a secondary site located distally near the C-terminus (Figure 1-1A). Both binding interactions occur on the heavy chain of FXa. The light chain of FXa binds to the heavy chain via the C-terminal EGF domain (EGF2) in an extended conformation that is nearly perpendicular to the complex formed between the heavy chain and ecotin (Figure 1-1B). The Gla domain is involved in crystal contacts primarily with the heavy chain of a neighboring symmetry mate. Surprisingly, the N-terminal EGF domain (EGF1) was found to be disordered. Numerous sparse density peaks in this region were observed in electron density maps, but efforts to trace a backbone through these peaks did not improve refinement statistics or result in improved density. The EGF1 domain was therefore omitted from the model.

Overall, M84R-FXa maintains the general tetrameric features found in other ecotin-protease complexes. However, comparisons of M84R-FXa with the structure of thrombin-ecotin M84R reveal that the relative orientation between ecotin and protease, as measured by the major angle between the major axis of the two proteins, differs by nearly 8 degrees.

**Figure I-1. Ribbon diagram of the tetrameric FXa-ecotin M84R complex**

(A) is rotated 90° with respect to (B). The Gla domain is shown in blue; the EGF2 domain in orange; the catalytic domain in green; ecotin M84R in purple. The catalytic triad is shown in red. The EGF1 domain (transparent white) was disordered and modeled using the EGF1 domain from PDB: 1XKA. The symmetry mate which comprises the other half of the tetrameric complex is shown in grey.



## Extended Interactions Between Ecotin M84R and FXa

It was previously thought that the specificity of FXa relied mostly on the P1<sup>1</sup> Arg residue, making FXa more similar to the digestive protease trypsin than thrombin. However when the reactive site loop of the serpin antithrombin (AT) was replaced by the two corresponding P4'-P4 fragments from the FXa cleavage sites in prothrombin, its capacity in inactivating FXa was significantly altered (Rezaie and Yang 2001). This observation suggested a significant role for extended interactions in FXa substrate recognition whose details small molecule inhibitors may not have revealed. Because ecotin is not hydrolyzed when bound to FXa, it was possible to observe interactions between FXa and inhibitor on both sides of the scissile bond. The entire buried surface area of interactions between the primary site of ecotin and FXa is 1980 Å<sup>2</sup>. Detailed interactions, spanning from ecotin residue 78 to 86, are depicted in Figures 1-2A and 1-2B.

Like most serine proteases from the blood coagulation cascade, FXa prefers substrates with an Arg at the P1 site. In the M84R-FXa structure, the positively charged side chain of Arg84 of ecotin forms two salt bridges with the negatively charged side chain of Asp<sup>11189</sup>, as well as a hydrogen bond with the carbonyl oxygen of Gly<sup>11218</sup> (in order to distinguish FXa from ecotin residues, FXa residue numbers are written as superscript following its chain label: H for heavy chain and L for light chain). The aliphatic part of the Arg side chain is sandwiched between two groups of hydrophobic residues, with the first group formed by Ala<sup>11190</sup>, Cys<sup>11191</sup> and Cys<sup>11220</sup> and the second group formed by Val<sup>11213</sup> and

---

<sup>1</sup> Protease-ligand interaction is described by conventional nomenclature. Residues in the substrate are labeled \*NH<sub>2</sub>-...P4-P3-P2-P1-P1'-P2'-P3'-P4'...-COO'. The P1-P1' peptide bond is cleaved by the protease. The corresponding subsites on the enzyme are labeled ...S4-S3-S2-S1-S1'-S2'-S3'-S4'... [34] so that the P2 residue of the ligand binds in the S2 pocket of the enzyme. Pockets N-terminal and C-terminal to the cleaved bond constitute the non-primed and primed sites, respectively.

Tyr<sup>1228</sup>. The main chain atoms of P1 Arg are coordinated by four hydrogen bonds from three protease residues: Gly<sup>1193</sup>, Ser<sup>1195</sup>, and Ser<sup>1214</sup> (Figures 1-2A & 1-2B).

It is interesting to note that while an M84R mutation in ecotin increases the binding affinity for thrombin by over 10<sup>4</sup> fold (Wang et al. 2001b), the same mutation increases the binding affinity for FXa by only five fold (Seymour et al. 1994). This apparent discrepancy can be explained by the presence of a bulky 60's loop insertion in thrombin. By obstructing access to the active site cleft, this loop hinders the establishment of the full array of interactions necessary for binding. Tighter binding to the S1 pocket caused by an M84R mutation provides energy to displace the 60's loop and ensure the proper register of interactions between the 80's loop of ecotin and the residues of the thrombin active site cleft (Wang et al. 2001b). Thus, in the case of thrombin, the observed increase in inhibition upon an ecotin M84R mutation is not solely attributable to the energetic contribution of a single Arg side chain. In contrast, the absence of a protruding 60's loop in FXa results in a more open active site. This allows the 80's loop of ecotin to establish the full extent of interactions with FXa without having to overcome an energy barrier caused by the displacement of an obstructing surface loop. As a result, the interactions of both wild-type and M84R ecotin with FXa are very similar, resulting in a far less dramatic change in inhibition constants.

The M84R-FXa structure reveals that the hydrophobic S2 pocket is defined by residues His<sup>1157</sup>, Tyr<sup>1160</sup>, Phe<sup>1194</sup>, Tyr<sup>1199</sup>, and Ser<sup>1214</sup> (Figure 1-2A & 1-2B). Previous modeling suggested that larger hydrophobic residues would bind well in the S2 pocket. This was confirmed by *in vitro* substrate scanning analyses that showed Gly, the highly conserved natural P2 residue, was only slightly better or, in some cases, even worse than the much bulkier residues of Trp, Phe or Tyr (Harris et al. 2000; Bianchini et al. 2002). Though existing FXa-inhibitor structures and known natural substrates of FXa argue for a small and



shallow S2 pocket, the M84R-FXa structure reveals an unusual conformation of Tyr<sup>1199</sup> that explains the apparent paradox for the S2 pocket size (Figure 1-3A). In M84R-FXa, Tyr<sup>1199</sup> swings 70° away from the P2 Thr, dramatically increasing the size of the S2 pocket. This unusual conformation of Tyr<sup>1199</sup> is firmly held by two hydrogen bonds between the terminal O<sub>η</sub> from Tyr<sup>1199</sup> and the carbonyl oxygens from ecotin residues Leu52 and Val81. It is clear that in this conformation, Thr83 from ecotin only occupies a small part of the S2 pocket, most of which is still left open to accommodate even larger hydrophobic residues. In its “small molecule inhibitor” bound state, Tyr<sup>1199</sup> will likely deny Thr binding in the S2 pocket, due to a clash between Thr83 and the O<sub>η</sub> from Tyr<sup>1199</sup>, which is less than 2 Å away.

The S3 region, as is usual for serine proteases, is solvated. Aside from Gln at 192, no charged or other discriminatory residues are nearby. It has been hypothesized that the Gln at position 192 may contribute to FXa’s preference for acidic residues at P3, as mutating Glu192 to Gln in both thrombin and activated protein C resulted in both enzymes selectively cleaving substrates with acidic residues at P3 position (Le Bonniec and Esmon 1991; Rezaie and Esmon 1993). Our M84R-FXa structure indicates that a possible polar-charge interaction may occur between Gln<sup>11192</sup> and the P3 residue (Figure 1-2A). The distances between the O<sub>γ</sub> of Ser82 and the N<sub>ε</sub> and O<sub>ε</sub> of Gln<sup>11192</sup> are 3.6 Å and 4.3 Å, respectively, and it is conceivable that in the presence of a negatively charged P3 residue, interactions could occur between the side chain of P3 and Gln<sup>11192</sup>.

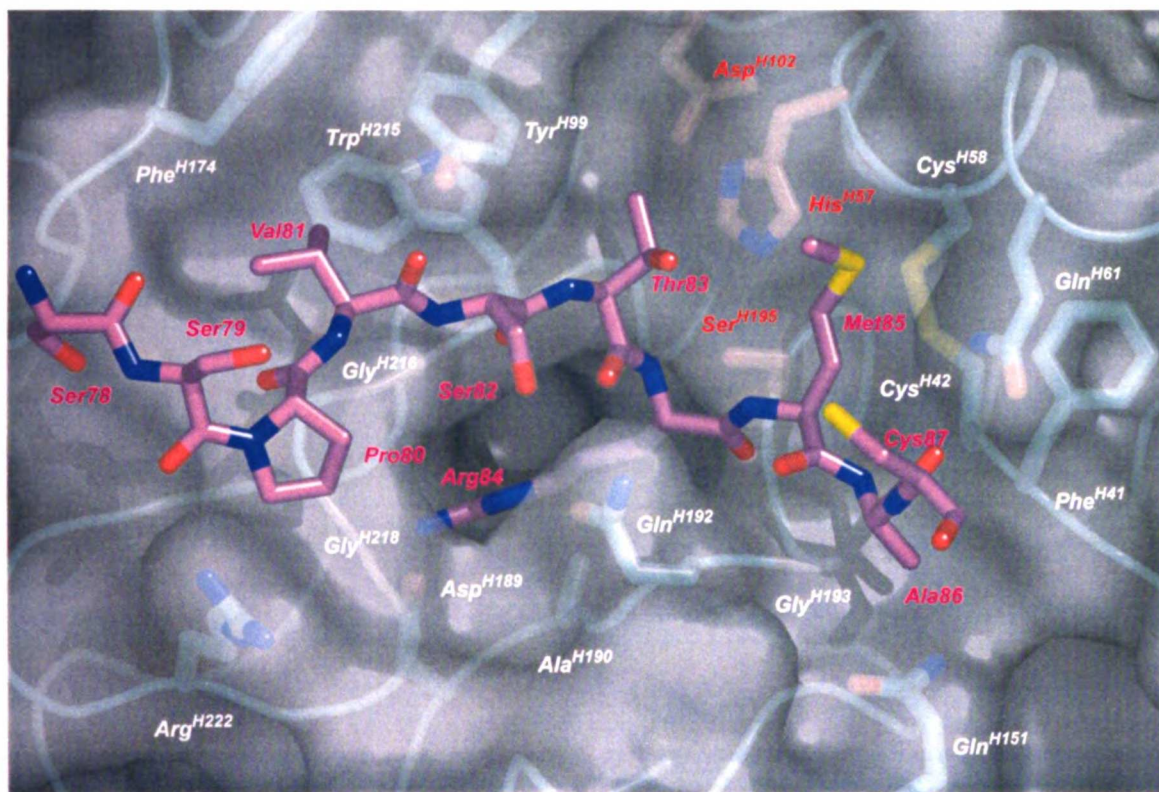
Together with two additional hydrophobic residues, Phe<sup>11174</sup> and Trp<sup>11215</sup>, residue Tyr<sup>1199</sup> also defines the S4 substrate binding pocket (Kamata et al. 1998) (Figures 1-2A and 1-3A). Its macromolecular inhibitor induced conformation inevitably changes the size of the S4 pocket. The long established belief that the S4 pocket of FXa is fit to accommodate large bulky hydrophobic motifs no longer holds, as in M84R-FXa the S4 pocket shrinks to

maximize interactions with a smaller residue such as Val. Though changes in the size of the S4 pocket are due in large part to Tyr<sup>1199</sup>, Phe<sup>1174</sup> also makes a significant contribution by rotating over 120° in response to Val81 from ecotin. The position of Trp<sup>1215</sup> remains unchanged as the large flat floor of the S4 pocket (Figure 1-3A).

Beyond P4, it is difficult to conclude whether there is any significant specificity. Our structure suggests that that ecotin may bind FXa better if its Ser78 is replaced by a large charged residue to establish interactions with Ser<sup>1173</sup>, Glu<sup>1217</sup>, or Lys<sup>1224</sup>. The primed site in FXa is also predominantly hydrophobic, but unlike thrombin the S1' and S2' pockets of FXa are well separated by residue Gln<sup>1161</sup>.

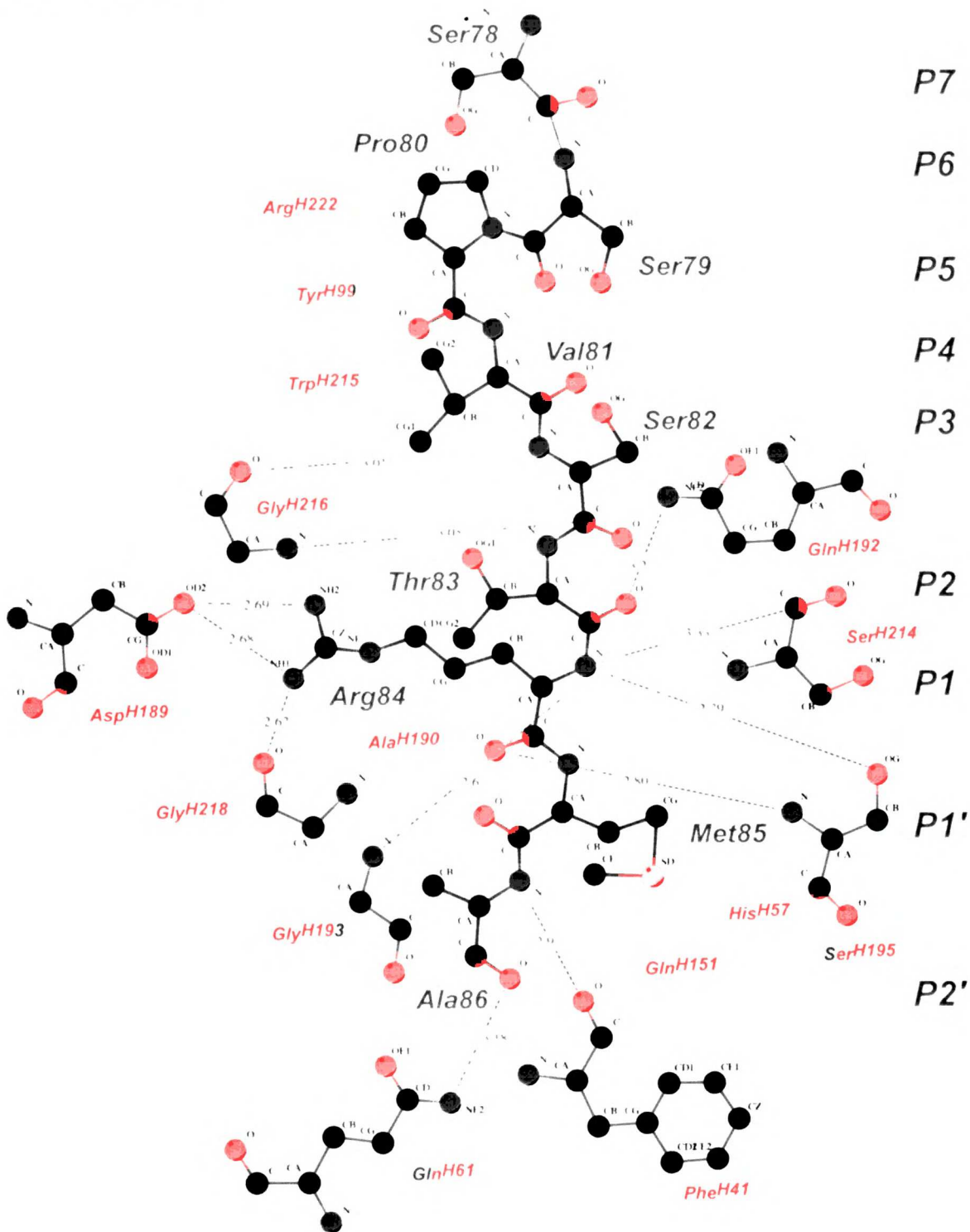
**Figure I-2A**

Extended interactions between FXa and ecotin M84R shown as a three dimensional surface rendering of FXa, with ecotin residues 78-86 in purple, FXa in green, and the catalytic triad highlighted in red.



**Figure I-2B**

Extended interactions between FXa and ecotin M84R shown as a two dimensional Ligplot (Wallace et al. 1995).



### Figure I-3

The S2 and S4 substrate binding pockets of M84R-FXa (FXa-green, ecotin-purple) are shown superimposed in (A) with three FXa-inhibitor complexes (PDB: 1XKA-yellow, 1EZQ-dark grey, and 1F0S-bright green) and in (B) with M84R-thrombin (PDB: 1ID5; thrombin-white, ecotin-orange).



## Residue 192 in Substrate and Inhibitor Discrimination

Besides its possible influence on acidic residues at the P3 position, residue 192 may have a more significant role in the regulation of coagulation proteases (Rezaie and Esmon 1995). Sequence alignments show that for the major serine proteases of the coagulation pathway, the amino acids surrounding the catalytic Ser<sup>1195</sup> are almost completely conserved across species, with the exception of residue 192 and 190 (Figure 1-4A). In contrast to residue 190 which makes a fairly conservative fluctuation between Ala and Ser, residue 192 varies dramatically from the positively charged Lys (FVIIa), to the neutral but polar Gln (FXa) or Asn (mTSP), to the negatively charged Glu or Asp (thrombin and activated protein C (aPC)). In general, there is a trend in the charge of residue 192 from negative at the bottom of the coagulation cascade to positive at the top, which may be related to the order in which these enzymes evolved. (Krem and Cera 2002). Possibly because it served as an intermediate during the evolutionary process, FIXa is the only exception, showing a wide charge distribution among different species. Figures 1-2A and 1-2B reveal that 192 lies within hydrogen bonding distance to the P2 carbonyl oxygen of a substrate or canonically bound inhibitor (in this case ecotin). The presence of a negatively charged residue at this position, as in thrombin and aPC (Figure 1-4A), would cause charge repulsion with any incoming P2 residue (Figure 1-4B). The energy barrier caused by such repulsion would not only render the protease more selective, but also would decrease the likelihood of inhibition from such endogenous protease regulators like tissue factor pathway inhibitor (TFPI) (Broze et al. 1990) and antithrombin III (ATIII) (Kisiel 1979). This would explain why both human thrombin and aPC become highly susceptible to inhibition by bovine pancreatic trypsin inhibitor (BPTI) or TFPI once the negative charge at 192 is eliminated by mutagenesis (Rezaie and Esmon 1993; Guinto et al. 1994; van de Locht et al. 1997). On the other hand, a

positively charged or polar amine containing residue at 192 would be thought to have the opposite effect. In FXa, Gln<sup>1192</sup> forms a hydrogen bond with the carbonyl oxygen of ecotin Ser83 (Figure 1-2B, 1-4B) promoting the proper alignment of interactions between ecotin and FXa. Together with residue Tyr<sup>1199</sup>, it forms a “clamp” to secure tight binding of FXa to ecotin (Figure 1-2A).

While these results and others suggest a critical role for residue 192 in substrate and inhibitor discrimination, the role of other factors such as the presence of multiple extended surface loops and the conformation of subsite residues like residue 99, must not be discounted (Rezaie 1996). In addition, it is necessary to keep in mind that many of these studies were performed *in vitro* in the absence of a complete ensemble of physiological cofactors. While this does not detract from the relevance or utility of such studies toward a better understanding of these enzymes, in particular toward the design of more selective inhibitors, studies demonstrating the importance of exosite binding in the activity and regulation of these enzymes illustrate that a complete understanding of the coagulation proteases can only be achieved by considering the full context of these proteins *in vivo*. (Vijayalakshmi et al. 1994; Dennis et al. 2000; Fuentes-Prior et al. 2000; Rezaie 2000; Chuang et al. 2001).

**Figure I-4**

A). Alignment of partial sequences of major serine proteases from the coagulation pathways.

B). depicts the position of residue 192 and its interactions with the main-chain carbonyl of inhibitor P2 residue. M84R-FXa (green) is superimposed with two other structures, the M84R-thrombin (PDB: 1ID5, pink) and thrombin E192Q-BPTI (PDB: 1BTH, light blue).

**A**

	190 192		190 192
proteinC_rat	DAC <b>D</b> GDSGGP	IXa_human	D <b>S</b> COGDSGGP
proteinC_human	DAC <b>E</b> GDSGGP	IXa_dog	D <b>S</b> COGDSGGP
proteinC_dog	DAC <b>E</b> GDSGGP	IXa_bovine	D <b>S</b> COGDSGGP
proteinC_mouse	DAC <b>D</b> GDSGGP	IXa_sheep	D <b>S</b> COGDSGGP
proteinC_bovine	DAC <b>E</b> GDSGGP	IXa_pig	D <b>S</b> CLGDSGGP
proteinC_pig	DAC <b>E</b> GDSGGP	IXa_rabbit	D <b>S</b> CEGDSGGP
		IXa_rat	D <b>S</b> CEGDSGGP
		IXa_mouse	D <b>S</b> CEGDSGGP
thrombin_bovine	DAC <b>E</b> GDSGGP	Xa_human	DAC <b>O</b> GDSGGP
thrombin_human	DAC <b>E</b> GDSGGP	Xa_rat	DAC <b>O</b> GDSGGP
thrombin_mouse	DAC <b>E</b> GDSGGP	Xa_bovine	DAC <b>O</b> GDSGGP
thrombin_rat	DAC <b>E</b> GDSGGP	Xa_chicken	DAC <b>O</b> GDSGGP
thrombin_rabbit	DAC <b>E</b> GDSGGP	Xa_mouse	DAC <b>O</b> GDSGGP
thrombin_chicken	DAC <b>E</b> GDSGGP		
Kallikrein_human	DAC <b>K</b> GDSGGP	XIa_chimpanzee	D <b>S</b> COGDSGGP
Kallikrein_mouse	DAC <b>K</b> GDSGGP	XIa_human	DAC <b>K</b> GDSGGP
VIIa_human	D <b>S</b> CKGDSGGP	trypsin_rat	D <b>S</b> COGDSGGP
VIIa_bovine	DAC <b>K</b> GDSGGP	trypsin_bovine	D <b>S</b> COGDSGGP
VIIa_rabbit	DAC <b>K</b> GDSGGP		
VIIa_mouse	DAC <b>K</b> GDSGGP	mTSP1_human	D <b>S</b> CNGDSGGP
		granB_human	T <b>S</b> FKGDSGGP

**B**





## Gla Domain - Overall Structure

In previous FXa structures, the Gla domain was either disordered or proteolytically cleaved to aid in crystallization. (Padmanabhan et al. 1993; Brandstetter et al. 1996; Kamata et al. 1998; Adler et al. 2000; Maignan et al. 2000; Nar et al. 2001). In the structure of M84R-FXa all residues except the N-terminal three were ordered and unambiguously traced through electron density maps (Figure 1-5A). The Gla domain forms a bundle of three alpha helices which are stabilized by a core of hydrophobic amino acids composed of Leu<sup>1.5</sup>, Leu<sup>1.13</sup>, Tyr<sup>1.24</sup>, Ala<sup>1.27</sup>, Val<sup>1.30</sup>, Phe<sup>1.31</sup>, Phe<sup>1.40</sup>, Phe<sup>1.41</sup> and Tyr<sup>1.44</sup> (Figure 1-5B) Except for Gla<sup>1.39</sup>, which is not involved in any contacts, all eleven Gla residues are involved in crystal contacts or Mg<sup>2+</sup> binding and presumably would be solvent exposed in solution.

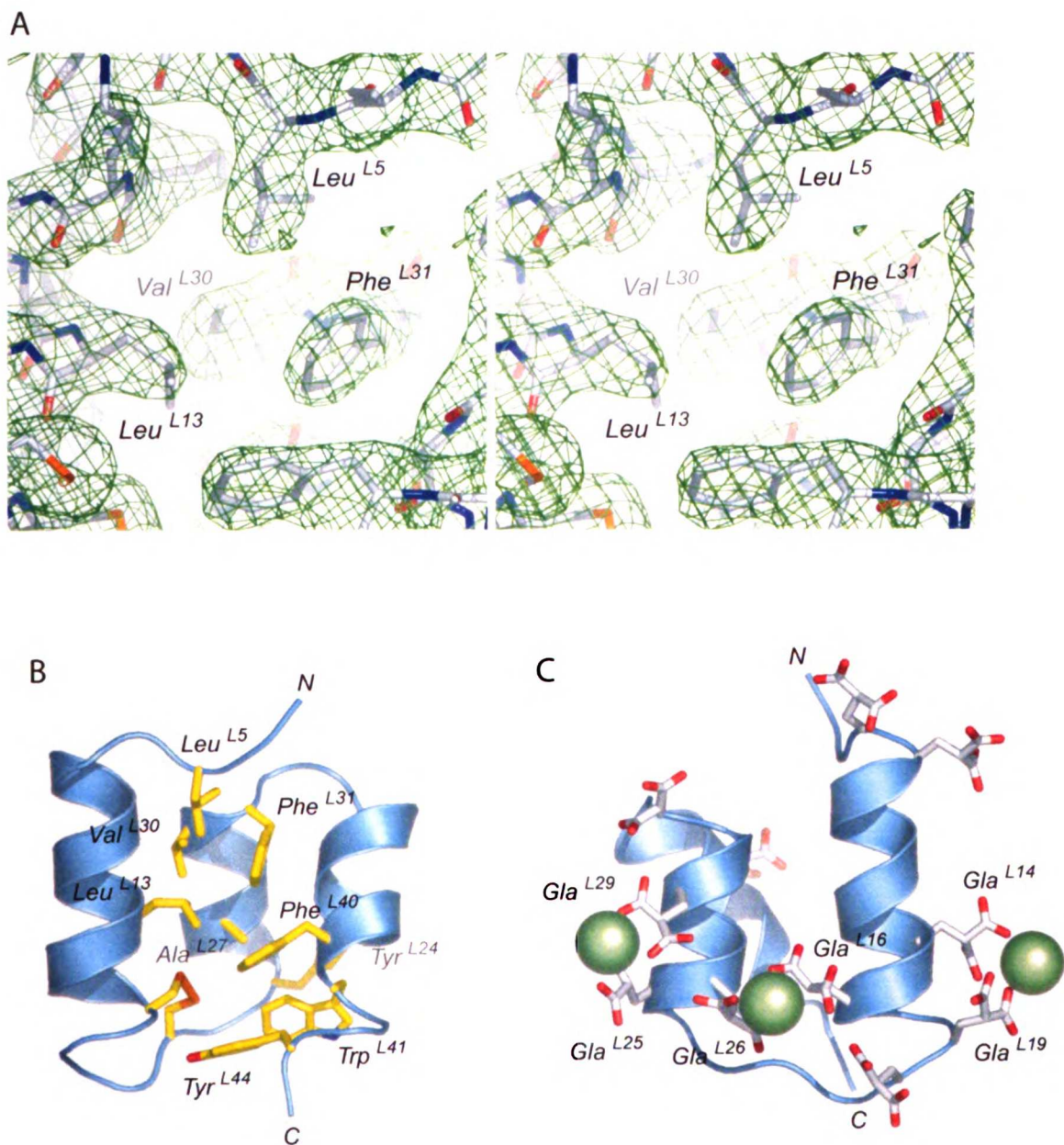
With the exception of the eleven N-terminal residues, the Gla domain of M84R-FXa is quite similar to structures reported for other Gla domains (Figure 1-6) (Soriano-Garcia et al. 1992; Freedman et al. 1995; Banner et al. 1996; Freedman et al. 1996; Mizuno et al. 2001). The rms deviations in the alpha carbon positions for residues 12-44 (FX numbering) between the M84R-FXa Gla domain and the Ca<sup>2+</sup> bound structures of the prothrombin, FVIIa, FIX (lowest energy model), and the FX Gla domain are 0.72 Å, 0.54 Å, 0.87 Å, and 0.56 Å, respectively. The rms deviation for the same residues of the lowest energy NMR structure with Mg<sup>2+</sup> bound is 1.44 Å. Due to the unusual conformation of the N-terminus, the Gla domain of M84R-FXa forms a more globular structure than the flat, ellipsoid structures observed previously. The positions of the N-terminal eleven amino acids deviate quite dramatically from those previously observed. This Gla domain represents a novel conformation. In previous structures with Ca<sup>2+</sup> bound, these residues form an omega loop structure whose Gla residues face toward the protein interior, and together with Gla residues from helix 1 and helix 2, chelate a line of Ca<sup>2+</sup> ions. Three conserved hydrophobic residues

at positions 4, 5, and 8, which are solvent exposed in this loop, are thought to be involved in interactions with the phospholipid membrane (Zhang and Castellino 1994; Christiansen et al. 1995; Sunnerhagen et al. 1995; Freedman et al. 1996; Falls et al. 2001). In the one available structure with  $Mg^{2+}$  bound, the conformation of the N-terminal eleven residues are not defined due to disorder (Freedman et al. 1996). These residues in M84R-FXa assume a largely helical structure, extending helix 1 by almost a full two turns compared to the structures with  $Ca^{2+}$  bound. The remaining N-terminal residues (3-6) form a random coil partially extending over the top of the helical bundle. Leu<sup>1.5</sup> makes hydrophobic contacts with Phe<sup>1.31</sup>, effectively “capping” the top of the Gla domain’s hydrophobic core (Figure 1-5A, 1-5B). The absence of the omega loop structure in the Gla domain of M84R-FXa is probably due to the charge repulsion that would occur between Gla<sup>1.6</sup> and Gla<sup>1.7</sup>, and Gla<sup>1.16</sup>, Gla<sup>1.20</sup>, Gla<sup>1.26</sup>, and Gla<sup>1.29</sup> upon formation of this structure in the presence of unoccupied  $Ca^{2+}$  binding sites.

**Figure I-5**

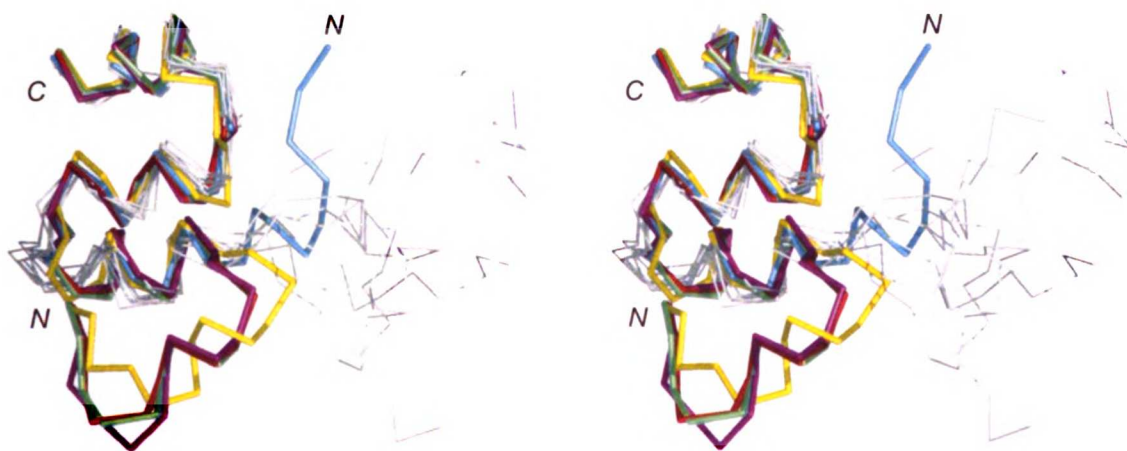
A). Electron density of the Gla domain ( $2F_{\text{obs}} - F_{\text{calc}}$ ) in the region around  $\text{Leu}^{\text{L5}}$  contoured at  $1\sigma$ .

B) and C). Two views of the Gla domain. Residues composing the hydrophobic core are highlighted in (B) while Gla residues and bound magnesium ions are shown in (C).



**Figure I-6**

The Gla domain of M84R-FXa (blue) superimposed with the calcium bound Gla domains of FVIIa (green), prothrombin (purple), FX (red), and FIX (yellow) and the magnesium bound NMR structure of the FIX Gla domain (grey).



## Gla Domain - Mg<sup>2+</sup> Binding

While the Gla domain can bind a number of different cations, only Ca<sup>2+</sup> and Sr<sup>2+</sup> have been shown to induce the state that binds phospholipid membranes (Nelsestuen et al. 1976). Previously, fluorescence, circular dichroism, and conformation-specific antibody studies on the homologous Gla domain of prothrombin have shown that there are two classes of cation binding sites (Nelsestuen et al. 1976; Prendergast and Mann 1977; Bloom and Mann 1978; Borowski et al. 1986). The first is composed of 3-4 cooperative, high affinity, but low specificity sites, while the second is characterized by 3-4 sites of lower affinity, but higher specificity. Based on these observations, two cation-dependent folding transitions have been proposed (Borowski et al. 1986). The first transition could be produced by binding to the non-specific, high affinity sites while the second transition responsible for the final membrane bound conformation may only be triggered by the specific binding of Ca<sup>2+</sup> to the lower affinity sites.

Three large, positive difference electron density peaks between Gla<sup>1.14</sup> and Gla<sup>1.19</sup>, Gla<sup>1.16</sup> and Gla<sup>1.26</sup>, and Gla<sup>1.25</sup> and Gla<sup>1.29</sup> were modeled as Mg<sup>2+</sup> ions due to the presence of 1 mM MgCl<sub>2</sub> in the crystallization buffer (Figure 1-5C). Previously, an NMR structure of the FIX Gla domain was solved in the presence of Mg<sup>2+</sup>. However, due to cross-peak broadening, side chain positions were poorly defined and the locations of the Mg<sup>2+</sup> binding sites could not be deduced (Freedman et al. 1996). Due to the presence of Mg<sup>2+</sup> during crystallization and the finding that Mg<sup>2+</sup> can induce the first transition but not the second (Nelsestuen et al. 1976), we believe that the locations of the three bound Mg<sup>2+</sup> ions represent the three high affinity, low specificity sites predicted in earlier studies and that this structure represents an intermediate folding state of the Gla domain. At first glance, these sites would not be the ones predicted, owing to the fact that they are spread out across the domain and

do not appear to interact with each other. Indeed, the original sites proposed were adjacent to one another and cooperativity was thought to be due to the orientation of Gla residues in a way that facilitated cation binding to an adjacent site in a polymeric array (Soriano-Garcia et al. 1992; Sunnerhagen et al. 1995). Instead, we propose that the cooperativity observed in these high affinity sites is due more to global rearrangements of secondary and tertiary structure than to local rearrangements of side chains. In a scheme similar to that proposed previously (Freedman et al. 1996), binding of a cation to one site would nucleate and stabilize secondary structure elements, organizing these elements into more ordered tertiary structure. This in turn would help properly position adjacent Gla residues for binding to the next site, which would then further order the domain and facilitate the positioning of Gla residues for cation binding at the last site. Binding at this last site would complete the transition to the intermediate state described here. Once in this intermediate state, we speculate that binding of  $\text{Ca}^{2+}$  to the remaining lower affinity sites promotes the ordering of the N-terminal residues into the omega loop structure and the formation of the final membrane bound state through bridging of  $\text{Gla}^{1.6}$  and  $\text{Gla}^{1.7}$  of the omega loop with  $\text{Gla}^{1.16}$ ,  $\text{Gla}^{1.20}$ ,  $\text{Gla}^{1.26}$ , and  $\text{Gla}^{1.29}$  via bound  $\text{Ca}^{2+}$  ions.

This model is supported by NMR and crystal structures of the Gla domain solved in the absence of  $\text{Ca}^{2+}$  (Tulinsky et al. 1988; Freedman et al. 1995; Sunnerhagen et al. 1995). In these structures, the Gla domain has clearly defined elements of secondary structure, but individual helices are truncated and do not form well ordered tertiary structure. Moreover, the NMR structure of the FIX Gla domain in the presence of  $\text{Mg}^{2+}$  shows that while most of the Gla domain is well ordered, forming a globular core essentially equivalent to that seen in the  $\text{Ca}^{2+}$  bound structures, residues 1-11 have no defined structure (Figure 1-6) (Freedman et al. 1996). In addition, equilibrium dialysis studies have shown that in the presence of  $\text{Mg}^{2+}$ ,

$\text{Ca}^{2+}$  affinity is increased (Prendergast and Mann 1977). This observation, combined with studies that have shown that  $\text{Mg}^{2+}$  can promote the first transition but not the second required for phospholipid binding (Nelsestuen et al. 1976), suggests that binding to the high affinity sites primes the Gla domain for  $\text{Ca}^{2+}$  binding at the low affinity sites and subsequent formation of the membrane bound conformation. Although the possibility exists that the unusual conformation of the N-terminal eleven residues observed here does not exist in solution, and may have been induced by crystal contacts, the finding in NMR studies that these residues are disordered in the absence of  $\text{Ca}^{2+}$  implies that the N-terminus is highly flexible and may attain a number of different conformations. The capping of the hydrophobic core of the domain by Leu<sup>15</sup> and the formation of additional secondary structure through the N-terminal extension of helix 1, both energetically favorable interactions, suggest that this conformation is not a mere crystallization artifact and we speculate as to whether this may indeed be biologically relevant. Further experiments will show whether or not this is the case.

## ACKNOWLEDGEMENTS

The authors would like to thank Haematologic Technologies Inc. for their assistance in the purification of the factor Xa-ecotin M84R complex. The authors would also like to thank the UC Biotechnology STAR Project for their financial support.

**Table I-I. Data and Refinement Statistics for the Ecotin M84R-FXa Complex**

<b>Structure</b>	<b>M84R-FXa</b>
Space group	<i>I</i> 222
Unit cell constants (Å)	
A	66.8
B	108.0
C	186.3
Highest resolution (Å)	2.8
Unique reflections	16994
Completeness	92.9%
$\langle I/I_0 \rangle$	11.2
$R_{\text{merge}}^a$	9.3% (three data sets)
Solvent content	57.6%
$R^b$ after molecular replacement	41%
Refinement resolution range	20.0–2.8 (Å)
R	20.7%
free $R^c$	23.4%
water molecules	51
r.m.s.d. in bond length (Å)	0.008
r.m.s.d. in bond angle (°)	1.4

<sup>a</sup>  $R_{\text{merge}} = \sum |I - \langle I \rangle| / \sum I$

<sup>b</sup>  $R = \sum_{h,k,l} (|F_{\text{obs}}(h,k,l)| - k|F_{\text{calc}}(h,k,l)|)^2 / \sum_{h,k,l} |F_{\text{obs}}(h,k,l)|$

<sup>c</sup> free R: Cross-validation R calculated by omitting 10% of the reflections (Kleywegt and Brunger 1996).



## Chapter 2

# Recognition and Accommodation at the Androgen Receptor Coactivator Binding Interface

Coauthors on this chapter:

Samuel J. Pfaff, E. Sturgis Payne, Hanne Grøn, Benjamin M. Buehrer, Robert J. Fletterick

Reprinted with permission from:

Hur E, Pfaff SJ, Payne ES, Grøn H, Buehrer BM, Fletterick RJ (2004) *Recognition and accommodation at the androgen receptor coactivator binding interface*. PLoS Biology 2(9): e274.

## ABSTRACT

Prostate cancer is a leading killer of men in the industrialized world. Underlying this disease is the aberrant action of the androgen receptor (AR). AR is distinguished from other nuclear receptors in that after hormone binding, it preferentially responds to a specialized set of coactivators bearing aromatic-rich motifs, while responding poorly to coactivators bearing the leucine-rich “NR box” motifs favored by other nuclear receptors. Under normal conditions, interactions with these AR-specific coactivators through aromatic-rich motifs underlie targeted gene transcription. However, during prostate cancer, abnormal association with such coactivators, as well as with coactivators containing canonical leucine-rich motifs, promotes disease progression. To understand the paradox of this unusual selectivity, we have derived a complete set of peptide motifs that interact with AR using phage display. Binding affinities were measured for a selected set of these peptides and their interactions with AR determined by X-ray crystallography. Structures of AR in complex with FxxLF, LxxLL, FxxLW, WxxLF, WxxVW, FxxFF, and FxxYF motifs reveal a changing surface of the AR coactivator binding interface that permits accommodation of both AR-specific aromatic-rich motifs and canonical leucine-rich motifs. Induced fit provides perfect mating of the motifs representing the known family of AR coactivators and suggests a framework for the design of AR coactivator antagonists.

## INTRODUCTION

The androgen receptor (AR) is the cellular mediator of the actions of the hormone 5 $\alpha$ -dihydrotestosterone (DHT). Androgen binding to AR leads to activation of genes involved in the development and maintenance of the male reproductive system and other tissues such as bone and muscle. However, it is the pivotal role of AR in the development and progression of prostate cancer that has led to increasing interest in this nuclear receptor. Presently, hormone-dependent prostate cancer is treated with a combination of strategies that reduce circulating levels of androgens, such as the administration of antiandrogens that compete for the androgen-binding pocket in the core of the C-terminal ligand-binding domain (LBD). The benefits of these treatments are typically transient, with later tumor growth associated with increases in expression levels of AR or its cofactors, or mutations that render AR resistant to antiandrogens (Gregory et al. 2001; Culig et al. 2002; Lee and Chang 2003) Alternative approaches to inhibiting AR transcriptional activity may therefore lie in disrupting critical protein associations the receptor needs for full function.

The precise details of how AR binds the dozens of coregulator proteins reported to associate with different regions of AR *in vivo* remain poorly understood (Lee and Chang 2003). Many nuclear receptors activate transcription by binding short leucine-rich sequences conforming to the sequence LxxLL (where “x” is any amino acid), termed nuclear receptor (NR) boxes, which are found within a variety of NR coactivators including the p160 family. Hormone binding to the LBD stabilizes the C-terminal helix of the receptor, helix 12, in a conformation that completes a binding surface for these LxxLL motifs (Darimont et al. 1998; Nolte et al. 1998; Shiao et al. 1998; Bledsoe et al. 2002). The structural elements composing this binding interface, consisting of helices 3, 4, 5, and 12 of the receptor, are

synonymous with a previously defined hormone-dependent activation function that lies within the LBD termed activation function (AF)-2. Association of p160 coactivators allows the recruitment and assembly of a number of other cofactors that together modulate the state of chromatin and interactions with components of the basal transcription machinery to initiate transcription (Glass and Rosenfeld 2000).

AR, however, utilizes multiple mechanisms to activate gene transcription. Generally, AR activity is dependent on contributions from multiple transactivation functions that lie within the N-terminal domain (NTD) collectively called AF-1. Although the AR AF-2 can bind to a restricted set of LxxLL motifs (Ding et al. 1998; He et al. 1999; Needham et al. 2000) and is relatively potent (Wang et al. 2001a), it usually displays weak independent activity at typical androgen-regulated genes, with significant activity observed only in the presence of high levels of p160 coactivators, as detected in some prostate cancers (He et al. 1999; Gregory et al. 2001). Instead, the AR AF-2 exhibits a distinct preference among NRs for phenylalanine-rich motifs conforming to the sequence FxxLF (He et al. 2000; He and Wilson 2003). Such motifs have been identified in the AR NTD and in an AR cognate family of coactivators that includes AR-associated protein (ARA) 54, ARA55, and ARA70 (Swiss-Prot: Q13772) (He et al. 2000; He et al. 2002b; Lee and Chang 2003). The NTD FxxLF motif (residues 23–27) mediates a direct, interdomain, ligand-dependent interaction between the NTD and LBD (N/C interaction) that is thought to facilitate dimerization, stabilize androgen binding, and possibly regulate AF-1 and AF-2 activity (Langley et al. 1998; He et al. 2000). In addition, the NTD also contains a related hydrophobic motif, WxxLF (residues 433–437), that nucleates formation of an alternative N/C interaction that may serve to inhibit AR activity (He et al. 2000; He et al. 2002a; Hsu et al. 2003).

Presently, how the AR AF-2 surface can accommodate residues with bulky aromatic side chains and distinguish FxxLF motifs from LxxLL motifs is not known. To understand the structural basis of this unusual coactivator recognition preference, we characterized the full repertoire of interacting sequences using phage display to define amino acids preferred at the AR coactivator binding interface. Crystal structures of the AR LBD in complex with several phage display–derived peptides reveal the structural basis of FxxLF motif specificity and an induced fit of the receptor that allows accommodation of other related hydrophobic motifs. Comparisons of the structures suggest strategies for the design of AR coactivator antagonists.

## RESULTS

### AR Preference for Aromatic Groups in Coregulator Recognition

Phage display has been used to study coactivator recognition specificity and to identify coactivator motif sequence variants preferred by the estrogen receptor (ER), thyroid hormone receptor (TR)  $\beta$ , and most recently AR (Chang et al. 1999; Norris et al. 1999; Paige et al. 1999; Northrop et al. 2000; Hsu et al. 2003). Using phage display, we screened more than  $2 \times 10^{10}$  randomized peptides against DHT-bound AR LBD. Selections identified sequences containing hydrophobic motifs that were primarily aromatic in character, consistent with another recent study (Hsu et al. 2003) (Figure 2-1). Of these aromatic motifs, FxxLF and related motifs with substitutions of phenylalanine or tryptophan for leucine at positions +1, +5, or both, dominated the selections. (Peptide residues are numbered in reference to the first hydrophobic residue of the core motif, which is numbered +1. Residues preceding the first hydrophobic residue are numbered negatively in descending order starting with -1.) Substitutions of tyrosine at the +5 position were also observed, but

to a much lesser extent (unpublished data). At the +4 position, valines, methionines, and even the aromatic residues phenylalanine and tyrosine were observed (Figure 2-1; unpublished data). In general, LxxLL motifs were not selected. The LxxLL motif shown in Figure 2-1 was derived from prior phage selections with ER and subsequently demonstrated to bind AR in FRET based screens *in vitro* (unpublished data).

Preliminary characterization of the subset of AR-interacting peptides shown in Figure 2-1 confirmed that each competed for binding of *in vitro* translated AR cofactors to bacterially expressed AR LBD in pulldown assays, and generally did so with modestly improved efficiency relative to the native FxxLF motif from the AR NTD and significantly greater efficiency than a native LxxLL motif from glucocorticoid receptor-interacting protein 1 (GRIP1) NR box 3 (P. Webb, personal communication). The equilibrium dissociation constants ( $K_d$ ) were directly determined for the interaction between the AR LBD and FxxLF and LxxLL peptides and one variant tryptophan-containing peptide, FxxLW, using surface plasmon resonance (Table 2-1). The  $K_d$  for FxxLF was 1.1  $\mu\text{M}$ , similar to the affinities of physiologically derived FxxLF motifs determined previously by isothermal titration calorimetry (He and Wilson 2003). The affinity of LxxLL was less than 2-fold weaker, with a  $K_d$  of 1.8  $\mu\text{M}$ , but more than three times stronger than the tightest binding p160-derived LxxLL motif, NR box 3 of transcriptional intermediary factor 2 (TIF2) (He and Wilson 2003). Surprisingly, the affinity of FxxLW, with a  $K_d$  of 920 nM, was slightly better than FxxLF, in spite of the presence of the tryptophan residue at the +5 position. Together, our results are consistent with the notion that the phage display peptides interact with the same AR surface that binds FxxLF and LxxLL motifs in native cofactors, and that they do so with similar or improved affinities relative to their natural counterparts.

## Figure 2-1. AR LBD–Interacting Peptides Selected by Phage Display

Hydrophobic residues of the core motif are highlighted in yellow. Residues in bold were ordered in electron density maps.



Table 2-1. Rate and Dissociation Constants for the Interaction between the AR LBD and Selected Peptides

	FxxLF	LxxLL	FxxLW
$k_{d1} \times 10^{-4} \text{ (M}^{-2}\text{s}^{-1}\text{)}$	5.2±1.0	6.2±4.3	5.9±1.5
$k_{d1} \times 10^2 \text{ (M}^{-1}\text{s}^{-1}\text{)}$	2.3±0.03	7.0±0.17	3.4±0.04
$K_{d1} \text{ (nM)}$	440±85	1120±78	580±140
$k_{d2} \times 10^3 \text{ (M}^{-1}\text{s}^{-1}\text{)}$	1.3±0.5	3.2±0.4	3.6±0.3
$k_{d2} \times 10^3 \text{ (M}^{-1}\text{s}^{-1}\text{)}$	3.5±0.06	5.0±0.04	5.6±0.05
$K_{d2}$	2.6±1.0	1.6±0.2	1.6±0.2
$K_{d}^{\text{APP}} \text{ (}\mu\text{M)}$	1.1±0.6	1.8±1.1	0.92±0.3

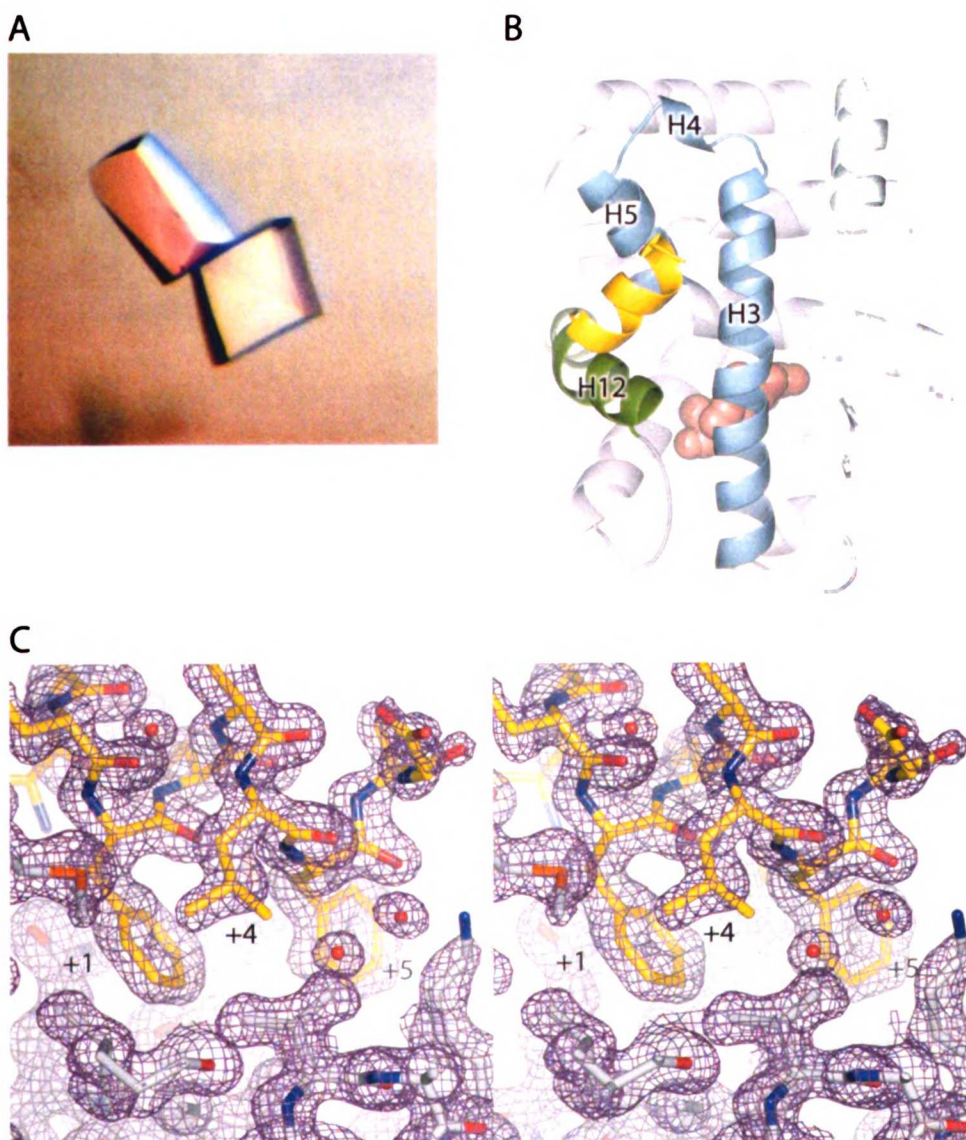
## One Site Fits All

To understand the binding mode of different AR coactivators, we determined the crystal structures of DHT-bound AR LBD without peptide and in complex with each of the seven peptides listed in Figure 2-1. All complexes crystallized in the space group  $P2_12_12_1$  with one molecule per asymmetric unit and unit cell dimensions similar to those observed in previous AR LBD crystal structures (Matias et al. 2000; Sack et al. 2001). Overall structural features of the complexes are shown in Figure 2-2. Peptides assumed short  $\alpha$  helical conformations centered on the core hydrophobic motif and bound in a solvent channel relatively free of crystal contacts on a groove formed by helices 3, 4, 5, and 12 of the receptor (Figure 2-2A). Detailed data collection and refinement statistics, as well as buried surface areas for each complex, are listed in Table 2-2. The structures confirm previous suggestions that AR utilizes a single binding interface for LxxLL and noncanonical aromatic-rich motifs (He et al. 2000; He et al. 2002a). Only side chains move to accommodate the array of peptides, sometimes considerably, with the unbranched side chains of Lys720, Met734, and Met894 making the largest conformational changes upon binding of peptide (Figure 2-2B).



## Supplementary Figures

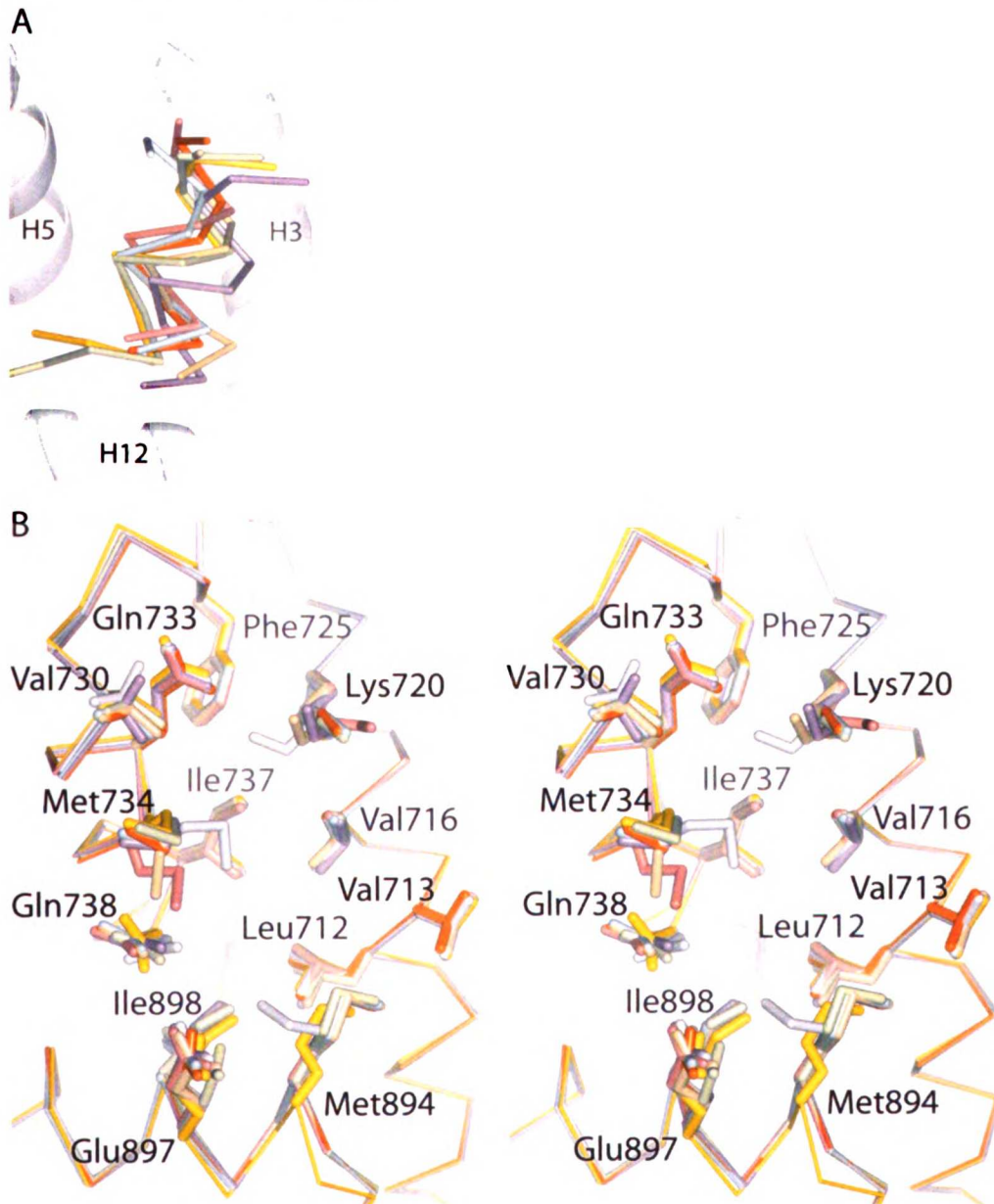
- A). Crystals of the AR LBD-FxxLF peptide complex. Each crystal is about 200 $\mu\text{m}$  across.
- B). Overall structural features of the AR LBD-FxxLF complex. FxxLF peptide is shown in yellow bound to the AR AF2 interface made up of helices 3, 4, 5 (blue), and 12 (green). Bound DHT is shown in salmon.
- C). Electron density ( $2F_o - F_c$ ) of the FxxLF peptide (yellow) bound to the AR LBD (white) contoured at  $1\sigma$ .



## Figure 2-2. A Structural Profile of the AR Coactivator Binding Interface

AR–peptide complexes are colored as follows: FxxLF, yellow; FxxLW, orange; WxxLF, wheat; WxxVW, purple; FxxYF, green; FxxFF, blue; LxxLL, pink; unbound, grey.

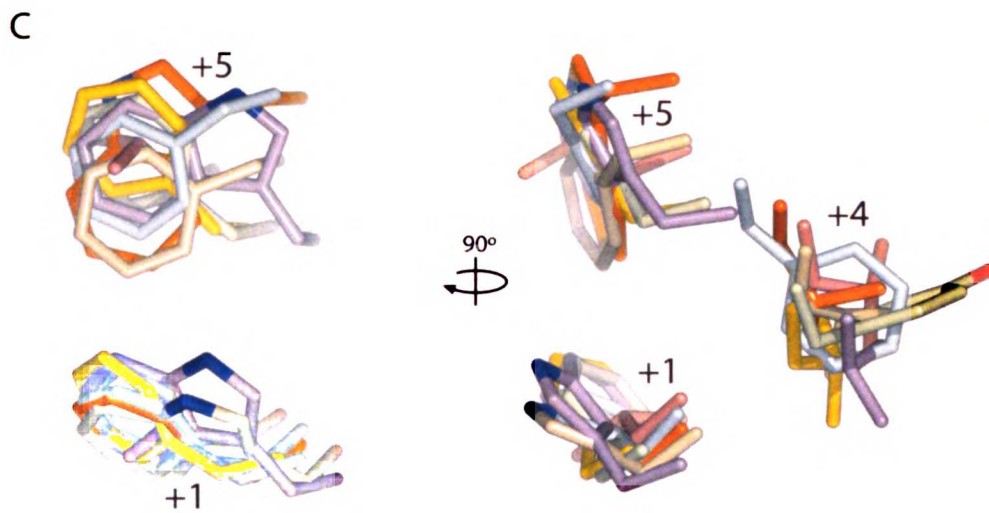
- A). C $\alpha$  trace of the peptides superimposed onto the AF-2. For clarity only the LBD of AR–FxxLF is shown.
- B). Superposition of the LBD of the AR–peptide complexes in the region of the coactivator interface. Backbone atoms are shown as a C $\alpha$  trace. Side chains of residues composing the interface are shown as sticks.



**Figure 2-2. A Structural Profile of the AR Coactivator Binding Interface (cont.)**

AR-peptide complexes are colored as follows: FxxLF, yellow; FxxLW, orange; WxxLF, wheat; WxxVW, purple; FxxYF, green; FxxFF, blue; LxxLL, pink; unbound, grey.

C). Hydrophobic side chains of the core motif superimposed as in (B).



## FxxLF

The mechanisms that permit AR to accommodate motifs with bulky phenylalanine residues were assessed in a crystal structure of the AR LBD in complex with the FxxLF peptide. The FxxLF peptide recapitulates the binding mode of p160-derived LxxLL motifs to other nuclear receptors (Darimont et al. 1998; Nolte et al. 1998; Shiau et al. 1998; Bledsoe et al. 2002). The peptide forms a short  $\alpha$  helix whose hydrophobic face, composed of Phe+1, Leu+4, and Phe+5, binds an L-shaped groove formed by helices 3, 4, 5, and 12 of the LBD that is composed of three subsites that accommodate each hydrophobic residue (Figures 2-2A and 2-3A). The conserved charged residues at either end of the cleft, Lys720 and Glu897, the so-called charge clamp residues, make electrostatic interactions with the main chain atoms at the ends of the peptide helix: Lys720 with the carbonyl group of Phe+5, and Glu897 with the amide nitrogens of Phe+1 and Arg-1 (Figure 2-3C). Glu897 also interacts with the side chain of Arg-1. The two interior residues of the motif, Glu+2 and Ser+3, are solvent exposed and do not interact with the receptor.

Comparison of AR alone and AR in complex with FxxLF (and other aromatic-rich peptides described below) reveals that the AF-2 cleft reorganizes to accommodate the bulky peptide side chains (see Figures 2-2B and 2-4). The unbranched side chains of Lys720 and Met734 move from an extended conformation over the +5 pocket to one almost perpendicular to the surface of the protein. The pockets for Phe+1 and Phe+5 are arranged in a line, forming a deep, extended cleft on the LBD spanning the length of the two side chains on the face of the peptide helix (see Figures 2-3A and 2-4B). Phe+1, almost entirely solvent inaccessible, binds face down at the base of this groove, making hydrophobic contacts with Leu712, Val716, Met734, Gln738, Met894, and Ile898, which define the +1 pocket. The top of the groove, composed of Val716, Lys720, Phe725, Ile737, Val730,

Gln733, and Met734, narrows to form the +5 pocket. Met734 and the aliphatic portion of Lys720 constrict this subsite, forming van der Waals interactions with opposite faces of the Phe+5 benzyl ring. Together, the +1 and +5 residues are almost entirely solvent inaccessible. In contrast, Leu+4 binds in a shallow hydrophobic patch consisting of Leu712 and Val716 lined at the ridges by Val713 and Met894 and is largely solvent exposed.

## LxxLL

The preference of AR for motifs with aromatic groups over leucine-rich motifs was assessed with a crystal structure of the AR LBD in complex with the LxxLL peptide. The structure reveals similarities between the binding modes of the LxxLL and FxxLF motifs to AR, and other LxxLL motifs to other nuclear receptors. The LxxLL motif adopts a helical conformation, and interactions of the motif with the AF-2 cleft are predominantly hydrophobic, with the three leucine residues of the motif contributing most of the interactions. However, significant differences can be seen between the binding mode of the LxxLL motif to AR and that of p160-derived LxxLL motifs to other nuclear receptors. First, flanking residues were largely disordered, with only two N-terminal flanking residues and one C-terminal residue visible in electron density maps (see Figures 2-1 and 2-3B). This contrasts with extended structures seen in the p160-derived LxxLL motifs in complex with their cognate receptors (Darimont et al. 1998; Nolte et al. 1998; Shiao et al. 1998; Bledsoe et al. 2002). Second, the LxxLL peptide backbone forms hydrogen bonds with only one of the two conserved charge clamp residues, Lys720. A shift in the position of the LxxLL peptide helix precludes direct interactions with Glu897 (see Figures 2-2A and 2-3D). This shift results from changes in the geometry of the +1 and +5 subsites mediated by Met734, which moves 2.5Å toward the +1 pocket (see Figures 2-2B and 2-4C) and enables binding of a

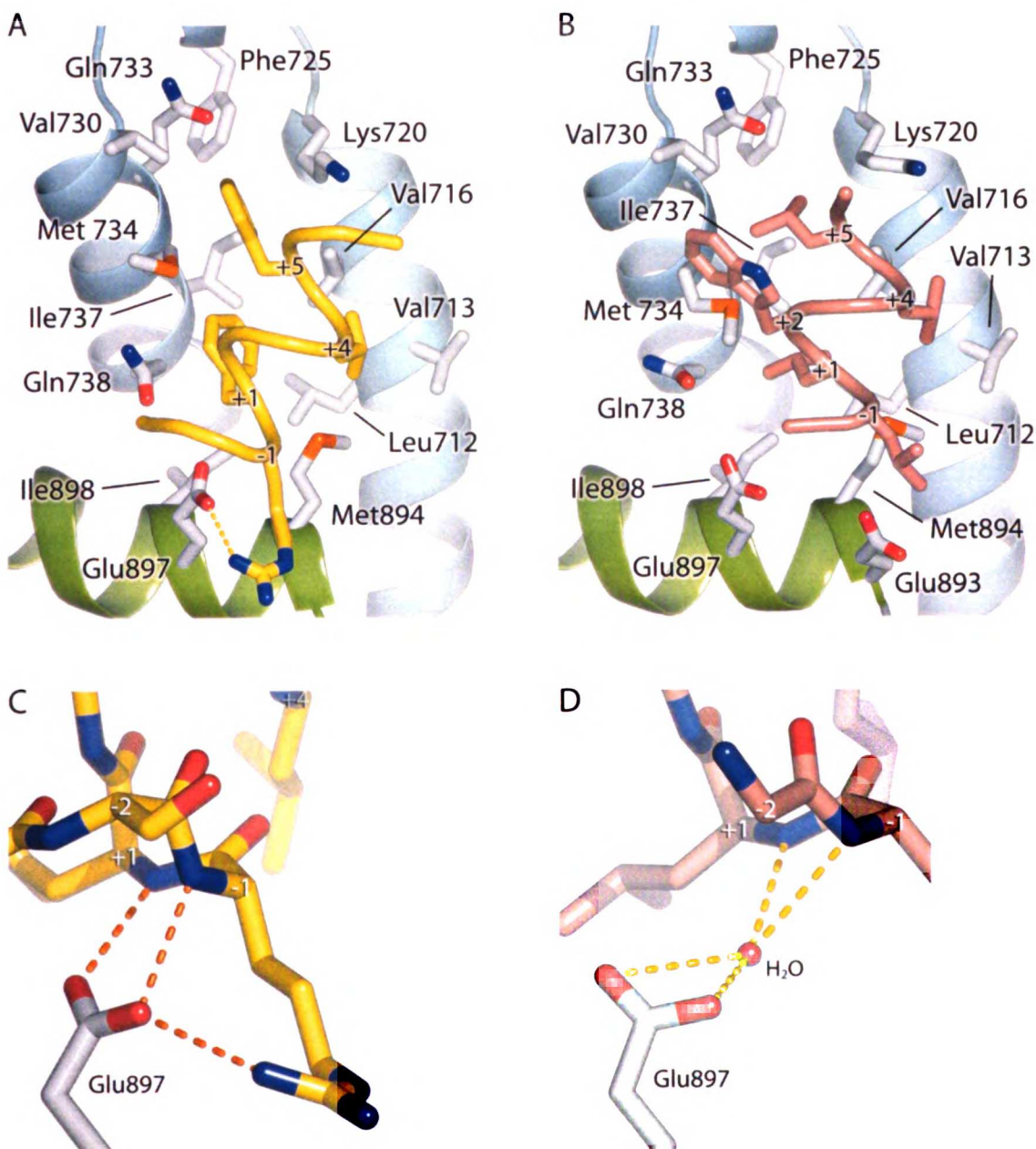
leucine at the +5 subsite by a simultaneous widening and shallowing of the pocket. This movement of Met734 causes displacement of the +1 residue, resulting in a rotation of the peptide helix away from helix 12, toward helix 3. A slight translation of the peptide helix also occurs away from helix 12 because of the shorter side chain length of leucine (see Figure 2-2A).

Side chains of residues flanking the first leucine of the motif make additional hydrophobic interactions with the AR surface (see Figure 2-3B). Trp+2 reaches over Met734, clamping the methionine in between itself and Leu+1. Leu-1 extends over Met894, abutted against Glu893. These interactions likely explain the moderate affinity of AR for this particular LxxLL motif despite suboptimal complementarity with the residues of the core motif (as discussed below) and the loss of main chain interactions with Glu897.

### Figure 2-3. Interactions of FxxLF and LxxLL with the AR LBD

A and B). FxxLF (A) and LxxLL (B) bound to the AR AF-2 interface. FxxLF and LxxLL are shown as yellow and pink  $\alpha$  coils, respectively. Helices 3, 4, and 5 of the LBD are shown as blue ribbons; Helix 12 is shown in green. LBD residues interacting with peptides are depicted as white sticks. For clarity only peptide side chains making significant interactions with the LBD are shown.

C and D). Hydrogen-bonding interactions between backbone atoms of FxxLF (C) and LxxLL (D) with Glu897 of the LBD. Peptide alpha carbons are labeled.



### Figure 2-4. Induced Fit of the AR AF-2 Interface

Surface representations of the AR AF-2 interface. The unbound structure is shown in (A), the FxxLF bound in (B), and the LxxLL bound in (C). Side chains of the hydrophobic residues of the core motifs of FxxLF and LxxLL are shown as spheres.

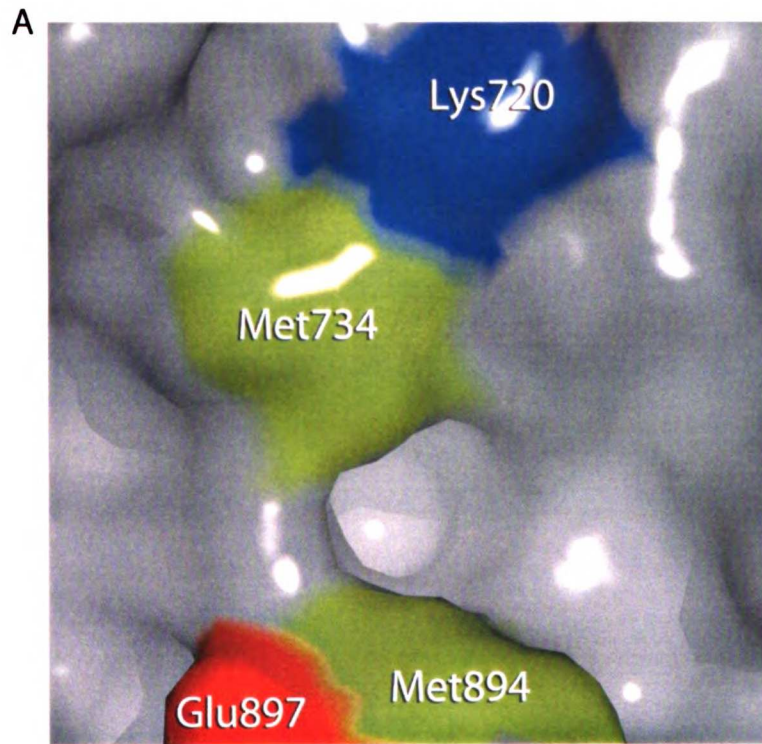
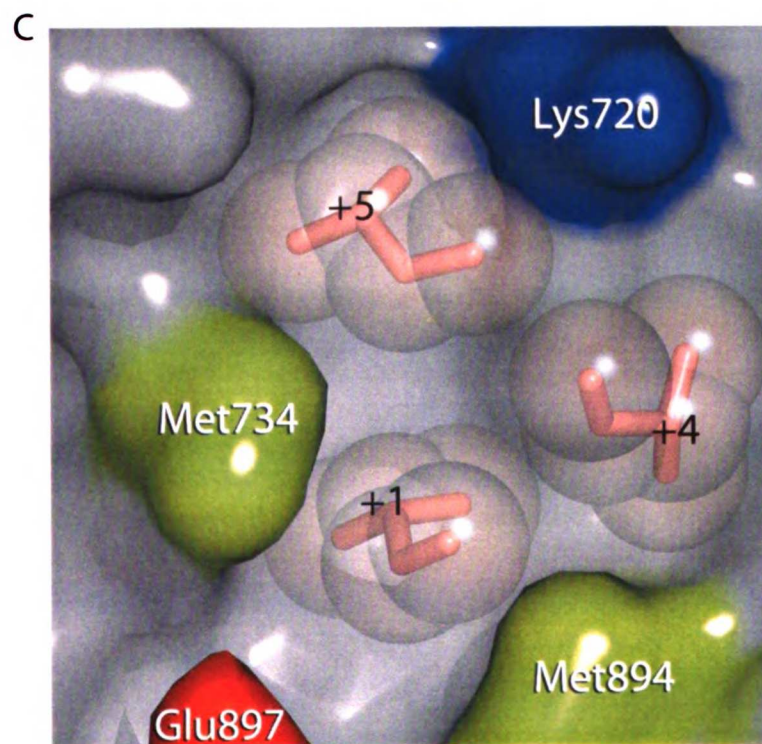
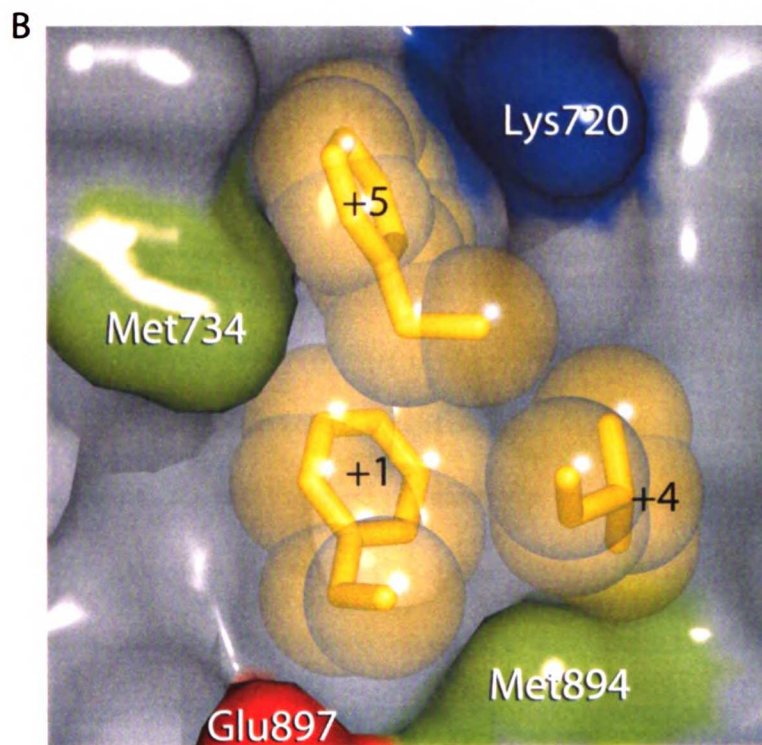




Figure 2-4. Induced Fit of the AR AF-2 Interface (cont.)

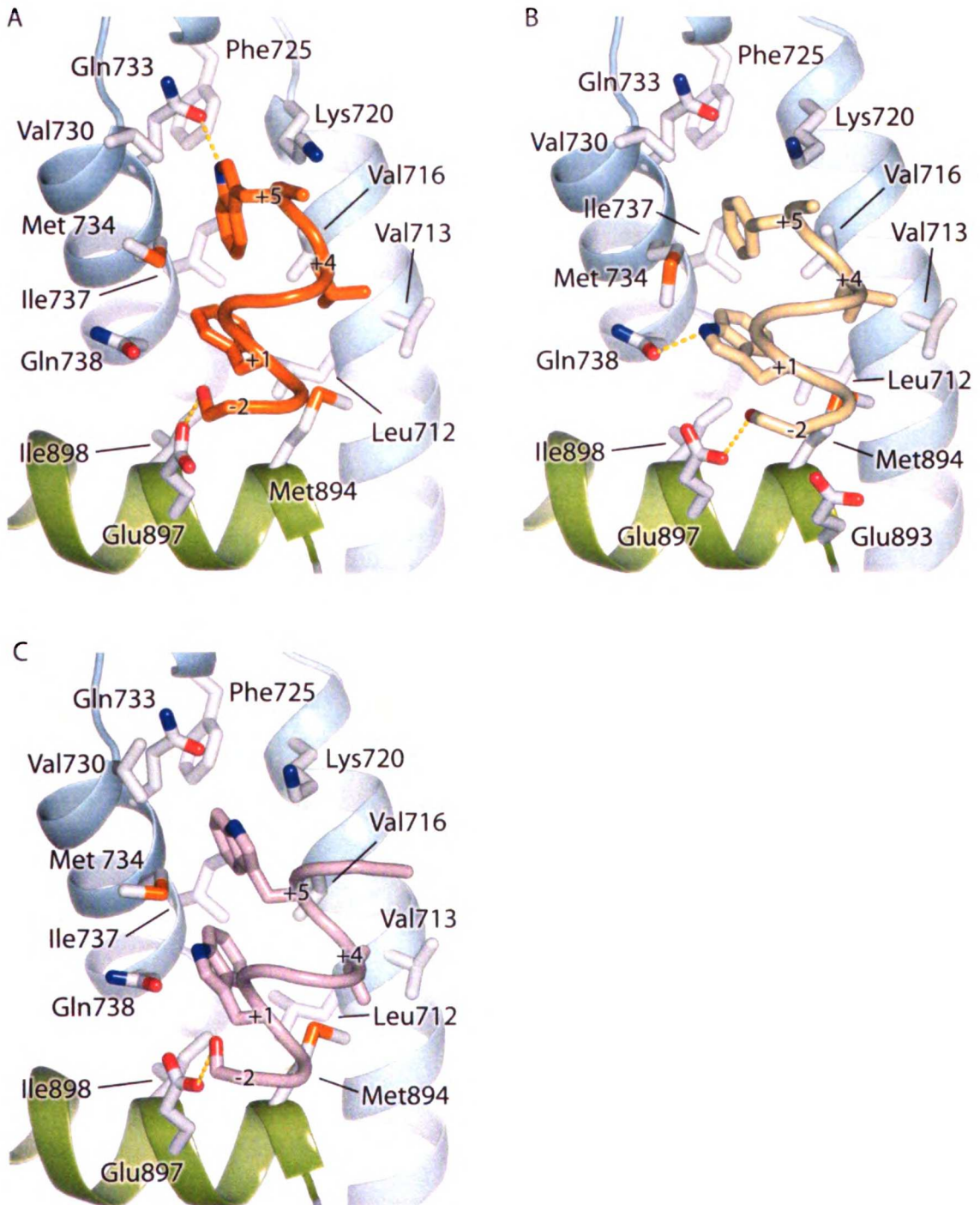


### WxxLF, FxxLW, and WxxVW

To understand how the AR AF-2 accommodates tryptophan residues, structures of AR in complex with peptides containing tryptophan substitutions at the +1 or +5 position, or both, were determined (Figure 2-5). Surprisingly, WxxLF, analogous to the only tryptophan-containing motif known *in vivo*, WHTLF in the AR NTD, was relatively disordered, with the peptide displaying the highest B-factor and least well defined density, suggesting that it binds with the lowest affinity (Table 2-2). Nonetheless, each of the tryptophan peptides adopted similar helical conformations. As described above for the LxxLL motif, substitutions at the +1 and +5 positions for non-phenylalanine residues result in shifts of the peptide helix (see Figure 2-2A). Consequently, backbone interactions with Lys720 are maintained, but interactions with the other charge clamp residue, Glu897, are lost. Once again, however, flanking residues within the peptide make additional contacts with the AR surface, and, unlike the LxxLL peptide, these contacts include Glu897. In FxxLW and WxxVW, the -2 serine (Figure 2-6) forms a bidentate hydrogen-bonding interaction, making hydrogen bonds to both Glu897 and the backbone amide group of the +2 residue. Ser-2 of WxxLF similarly interacts with Glu897, but is too distant for helical-capping interactions with the +2 amide group. Instead, Glu893, in a more typical interaction with the +1 amide nitrogen, caps the WxxLF helix (Figure 2-6B). Thus, tryptophan substitutions are tolerated, but they induce a shift in the peptide backbone that precludes interactions with one of the charge clamp residues. This suboptimal interaction is compensated partially by interactions of flanking residues with the AR surface.

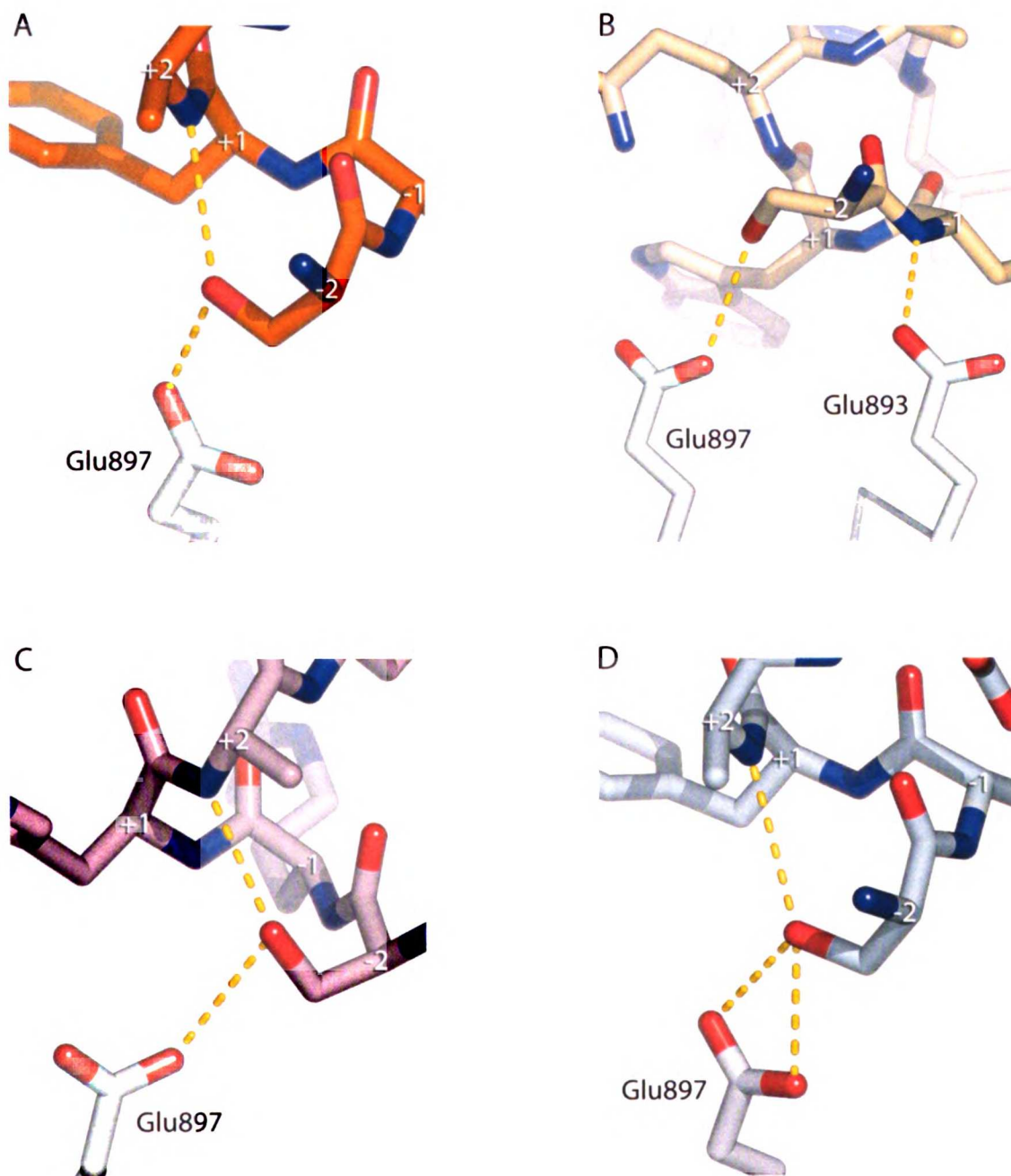
### Figure 2-5. Interactions of the Tryptophan Motifs with the AR LBD

FxxLW (A), WxxLF (B), and WxxVW (C) bound to the AR AF-2 interface. FxxLW, WxxLF, and WxxVW are shown as orange, beige, and purple  $\alpha$  coils, respectively. The LBD is depicted as in Figure 2-3.



**Figure 2-6. Interactions of Ser-2 with Glu897**

Interactions between Ser-2 of the peptides (A) FxxLW, (B) WxxLF, (C) WxxVW, and (D) FxxFF and Glu897 of the LBD. Peptide alpha carbons are labeled.



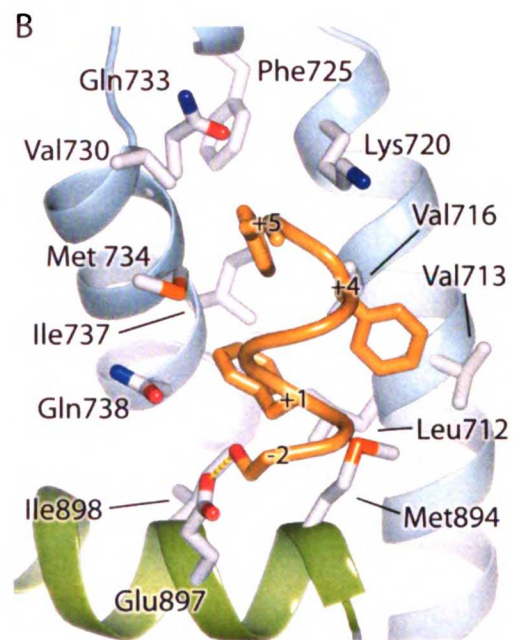
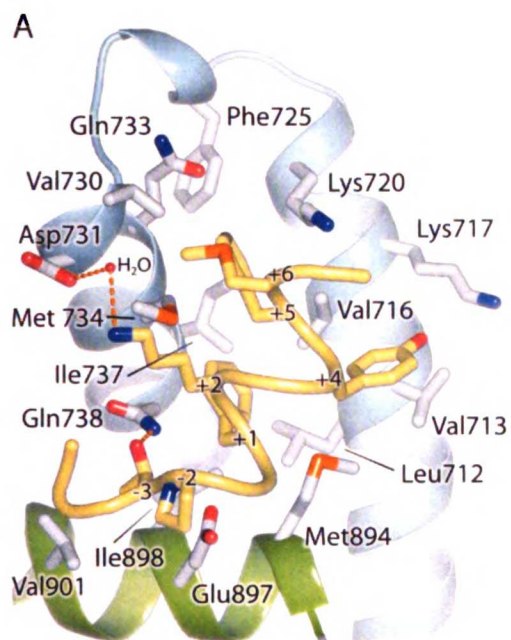
## **FxxFF and FxxYF**

Finally, effects of substitutions at the +4 position were assessed in structures of AR in complex with peptides containing FxxFF and FxxYF motifs (Figure 2-7). Surprisingly, the binding mode of FxxFF to AR resembled that of the tryptophan peptides more closely than the binding mode of FxxLF (see Figures 2-2A and 2-7B). Like the tryptophan peptides, interactions with Glu897 are mediated by Ser-2 instead of the peptide backbone (see Figure 2-6D). Deviations from ideal helical geometry allow Phe+4 to bind facedown in the +4 pocket with the benzyl ring stacked against Val713.

By contrast, the conformation of FxxYF was the closest to FxxLF (see Figure 2-2A). Other than FxxLF, only FxxYF makes direct backbone interactions with Glu897. Unlike the facedown orientation of Phe+4 observed in the FxxFF peptide, Tyr+4 is bound edgewise into the shallow +4 pocket, making interactions with Val713, Val716, and the aliphatic portion of Lys717. FxxYF was the most ordered of all the peptides, with 12 out of 15 residues observed in the electron density (see Figures 2-1 and 2-7A). Significant interactions were observed involving residues other than hydrophobic residues of the motif. Lys+2 and Met+6 are predominantly solvent exposed, extending out over the protein surface. Met+6 is bound on top of Phe+5, while Lys+2 makes a water-mediated hydrogen bond with Asp731. Thr-3 of the peptide defines a new subsite, with the hydroxyl group forming a hydrogen bond to Gln738 and the methyl group making hydrophobic contacts in a pocket formed by Glu897, Ile898, and Val901. Similar interactions were observed in the glucocorticoid receptor (GR)-TIF2 complex involving the -3 glutamine of the TIF2 NR box 3 motif (Bledsoe et al. 2002). However a valine to asparagine substitution at the residue corresponding to 901 in AR creates a pocket with a more polar character in GR (Figure 2-8).

**Figure 2-7. Interactions of FxxYF and FxxFF with the AR LBD**

FxxYF (A) and FxxFF (B) bound to the AR AF-2 interface. FxxYF and FxxFF are shown as yellow and orange  $\alpha$  coils, respectively. The LBD is depicted as in Figure 2-3.



## Restrictions of the Three Subsites

Together, the structures described above permit an assessment of the way that individual subsites of the AR AF-2 cleft accommodate hydrophobic groups. The indole rings of tryptophan and the phenyl rings of phenylalanine fit into their pockets analogously with the +1 and +5 residues bound facedown and edgewise, respectively, into the AF-2 cleft. On the other hand, the position of the +4 residue is variable, with binding in this shallow pocket largely dictated by the position of the peptide backbone caused by the bound conformations of the +1 and +5 residues (see Figure 2-2C). Small shifts in the position of the N-terminal of helix 12 can be seen, which reposition Met894 for more optimal contacts with +4 residues bound at that subsite (see Figure 2-2B).

The binding mode detected in the +1 pocket is the most conserved of the three hydrophobic subsites (see Figure 2-2C). The benzyl moiety of the indole side chains superimpose with the corresponding benzyl side chains of the phenylalanine-rich motifs, effectively mimicking interactions of a phenylalanine residue. However, the presence of a hydrogen-bonding partner on the indole side chain enables an additional polar interaction not seen in the phenylalanine-rich motifs between the indole nitrogen and Gln738 (see Figure 2-5B). Unexpectedly, this additional interaction in the +1 pocket does not occur with Trp+1 of WxxVW (see Figure 2-5C). While similarly distanced to make the same interaction, the plane of the indole ring is rotated about 20° relative to that of WxxLF, causing it to be at a poor angle for strong hydrogen bonding to Gln738.

Binding of tryptophans in the +5 pocket is slightly more variable (see Figure 2-2C). Trp+5 of WxxVW is bound similarly to phenylalanine residues at the same position. Only the six-membered ring of the indole group is fully buried in the pocket. The five-membered ring of the indole side chain sticks out, solvent exposed. In contrast, the +5 indole group of

FxxLW is rotated almost 90°, resulting in burial of both rings of the indole group, as well as the formation of a strong hydrogen bond between the indole nitrogen and Gln730 (see Figure 2-5A). Binding in this orientation appears to be highly favorable, as the FxxLW peptide deviates from helical geometry at the +5 position to do so.

## DISCUSSION

The crystal structures reported here reveal how AR binds coactivator motifs with bulky aromatic hydrophobic groups and permit construction of a profile of the AR coregulator interface (see Figure 2-2). In some ways, this interface resembles those of other nuclear receptors: it is an L-shaped hydrophobic cleft comprised of three distinct subsites that bind hydrophobic groups at the +1, +4, and +5 positions in cognate peptides. Moreover, the so-called charge clamp residues (Lys720 and Glu897) bracket the cleft. Nonetheless, the AR coregulator recognition site is unique in that it rearranges upon motif binding to form a long, deep, and narrow groove that accommodates aromatic residues at the +1 and +5 positions (Figure 2-9). Sequence alignments of AR with other NRs suggest that a unique combination of substitutions at Val730, Met734, and Ile737 combine to permit the formation of a smoother, flatter interaction surface that displays a higher complementarity to aromatic substituents than to branched aliphatic (see Figure 2-8). Of these, methionine, the only unbranched hydrophobic amino acid and the most accommodating, at a key position between the +1 and +5 sites, allows the AR AF-2 interface to vary the size and shape of its pockets to associate with a more diverse set of coregulators. GR also contains a methionine residue at this position, raising the possibility that it may also employ induced fit to broaden motif recognition. While naturally occurring mutations in AR have yet to be observed at Met734, it is interesting to note that mutations at



Val730 and Ile737 have been reported in patients with prostate cancer and androgen insensitivity, respectively (Newmark et al. 1992; Quigley et al. 1995; Gottlieb et al. 1998).

The same characteristics that make the AR AF-2 ideal for binding of longer, aromatic side chains also make it less well suited for binding of shorter, branched side chains. Although changes in the position of Met734 widen the groove towards the +5 subsite to permit binding of leucine residues, the gross features of the groove remain largely the same (see Figure 2-9B). As a result, the +1 and +5 leucines bind in a smooth, elongated groove and interactions between the +1 and +5 residues on the face of the peptide helix, or with a hydrophobic “bump” present in other receptors caused by a isoleucine to leucine substitution between the +1 and +5 subsites, are absent. Thus, a smaller proportion of the available surface area is available for van der Waals interactions.

Unlike the conserved interaction modes of aromatic residues with the +1 and +5 sites, binding interactions at the +4 site are variable and characterized by nonspecific interactions. This finding agrees with the relatively high conservation of residues at the +1 and +5 positions of AR-interacting motifs and suggests that these residues drive peptide interaction with the LBD, whereas the +4 site is less critical. Indeed, the +4 pocket is shallow, surface exposed, and relatively featureless, explaining the assortment of residues selected at the +4 position. It is likely that any hydrophobic residue that does not clash with surrounding residues would be suitable at this subsite.

While peptide motif recognition is governed by hydrophobic interactions, polar interactions from backbone atoms and residues outside the core motif also contribute. With the exception of FxxFF, motifs containing phenylalanines at the +1 and +5 positions present canonical main chain interactions with both charge clamp residues, Lys720 and Glu897. This finding stands in contrast to predictions of previous studies (Alen et al. 1999;

He et al. 1999; Slagsvold et al. 2000; He and Wilson 2003), which concluded that Lys720 was dispensable for FxxLF binding and that Glu897 was required for binding to FxxLF and LxxLL motifs. Lys720 comprises a significant portion of the +5 subsite, making important van der Waals interactions with the Phe+5 benzyl group in addition to hydrogen bonds to the motif backbone. These results suggest that Lys720 is required for binding of FxxLF motifs. However, it may be that enough binding energy is provided by the other residues of the +5 subsite (i.e., Met734), as well as by the other subsites themselves, such that removal of Lys720 would have little effect on binding. Observations that Lys720 plays a greater role in LxxLL motif binding are likely due to the fact that there is less surface area contributing to van der Waals contacts in LxxLL motifs. Disrupting binding contributions from Lys720 would thus have a more detrimental effect on binding.

On the other hand, Glu897 interacts with the FxxLF peptide backbone, but is disengaged from the LxxLL peptide backbone. One possible explanation for the apparent requirement for Glu897 in LxxLL binding is that it might interact with residues outside of the core motif. The corresponding glutamate of GR, Glu 755, forms hydrogen bonds with the -3 asparagine of TIF2 NR box 3 (Bledsoe et al. 2002) and Glu897 of AR participates in noncanonical interactions with the hydroxyl group of a Ser-2 residue that was selected in all of our tryptophan-containing peptides. This is especially intriguing given that the only WxxLF motif known *in vivo*, located in the AR NTD, also possesses a Ser-2 residue. WxxLF also makes backbone interactions with an alternate charge clamp residue, Glu893, pointing towards adaptability in AR AF-2 charge clamp formation.

Sequence alignment of NR coactivator sequences shows that positively charged residues are favored N-terminal to the core hydrophobic motif while negatively charged residues are favored C-terminal to the motif (He and Wilson 2003). Our phage-selected

peptides are consistent with this trend. Arginines and lysines were observed at the N-terminal -1 position in all peptides, except for LxxLL, in which Arg was present at the -3 position. Moreover, four out of seven peptides contained negatively charged aspartate or glutamate residues C-terminal to the core motif. While previous studies have shown that complementary interactions between charged residues flanking coactivator signature motifs of coactivators and charged residues surrounding the AF-2 cleft modulated binding to the receptor (He and Wilson 2003), we find that the flanking charged residues are typically disordered in the electron density, with only Arg<sup>-1</sup> of FxxLF interacting with Glu897, and Lys<sup>+2</sup> of FxxYF forming a water-mediated hydrogen bond to Asp731. Thus, if charge-charge interactions between flanking peptide residues and the AR surface occur, they are too weak to be detected crystallographically.

Finally, the AR AF-2 surface is an attractive target for pharmaceutical design. Selective peptide inhibitors that bind the AF-2 surface of liganded ER $\alpha$ , ER $\beta$ , and TR $\beta$  have been developed (Geistlinger and Guy 2003) and similar  $\alpha$ -helix-mediated protein-protein interfaces have successfully been targeted with tight binding small molecule inhibitors (Asada et al. 2003; Vassilev et al. 2004). Drugs that directly interfere with coactivator binding or formation of the AR N/C interaction would likely inhibit AR activity, perhaps even in androgen-resistant prostate cancers in which conventional therapies have failed. Strategies for designing AR coactivator antagonists are revealed in spite of the changes to the structure at the interface. Together the +1, +4, and +5 subsites contribute the majority of buried surface area of the peptide-LBD interaction (Table 2-2). Inhibitors may be designed by varying hydrophobic constituents at these hotspots. The +1 and +5 subsites of AR have a unique preference for aromatic side chains and provide the most viable starting points for designing AR-specific inhibitors. Aromatic groups, possibly with

polar constituents to exploit hydrogen bonding interactions with Gln733 and Gln738 in the +1 and +5 subsites, respectively, may provide promising leads. Indeed, initial screens have yielded compounds that bind to the +1 subsite in such a manner (E. Estébanez-Perpiñá, personal communication). Poorly conserved binding and a lack of strong structural features at the +4 subsite suggest that this site may be incorporated for achieving other characteristics important for inhibitors besides fit. Synthetic strategies that link together groups that bind with moderate affinity to the +1, +5, and possibly +4 subsites may yield tight binding inhibitors of AR coactivator association.

## Figure 2-8. Sequence Alignment of the AF-2 Region of NRs

Residues composing the coactivator interface of AR are highlighted in yellow. The absolutely conserved glutamate and lysine composing the charge clamp are highlighted in pink and blue, respectively. Residue numbering is that of AR.

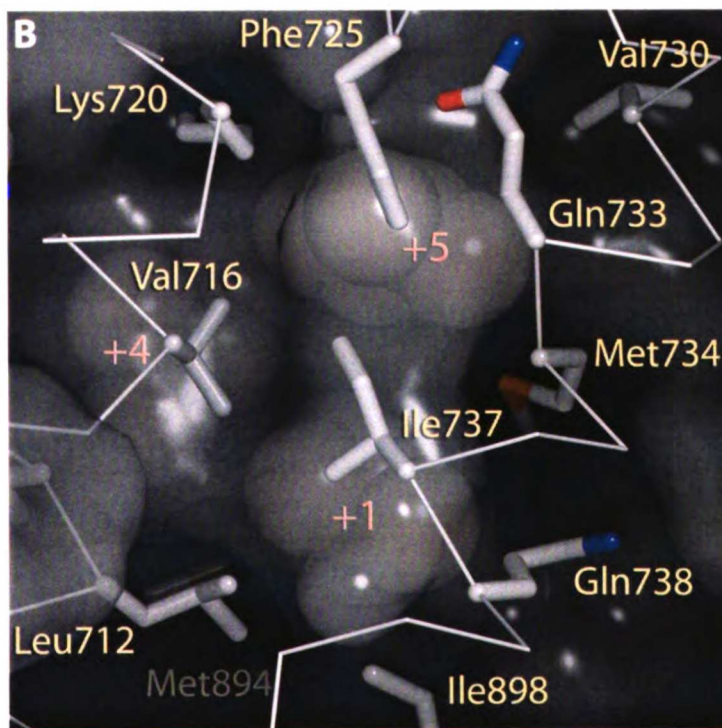
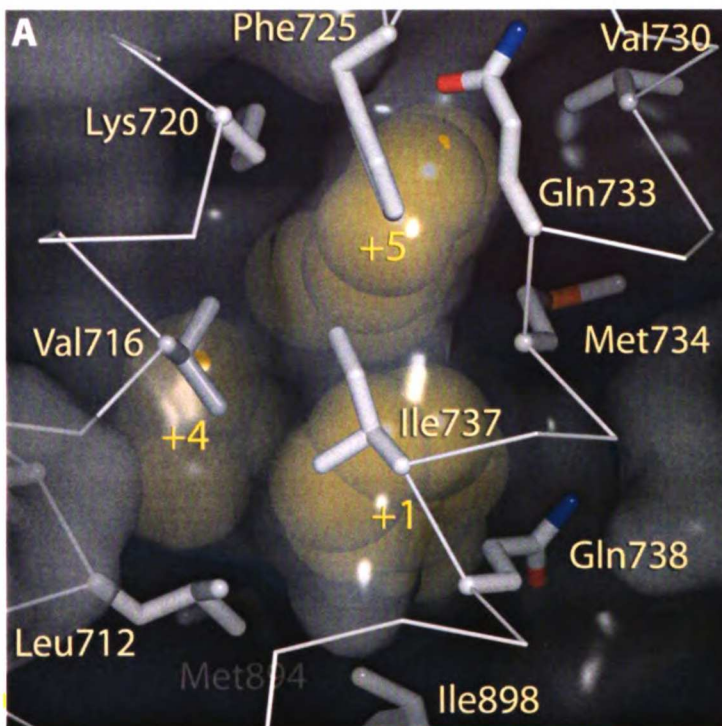
	710	720	730	891
AR	RQLVHVVKWAKALPGFRNLHVDDQMAVIQYS			FPEMMAEIIISV
GR	RQVIAAVKWKAKAIPGFRNLHLDDQMTLLQYS			FPEMLAEIITN
PR	RQLLSVVKWSKSLPGFRNLHIDDQITLIQYS			FPEMMSEVIAA
MR	KQMIQVVKWAKVLPGFKNLPLEDQITLIQYS			FPAMLVEIISD
ER $\alpha$	RELVHMINWAKRVPGFVDLTLHDQVHLLCA			LYDLLLEMLDA
ER $\beta$	KELVHMISWAKKIPGFVELSLFDQVRLLESC			VYDLLLEMLNA
RAR $\alpha$	KCIIKTVEFAKQLPGFTTLTIADQITLLKAA			MPPLIQEMLN
RAR $\beta$	KCIIKIVEFAKRLPGFTGLTIADQITLLKAA			MPPLIQEMLN
RAR $\gamma$ 1	KCIIKIVEFAKRLPGFTGLSIADQITLLKAA			MPPLIREMLN
RAR $\gamma$ 2	KCIIKIVEFAKRLPGFTGLSIADQITLLKAA			MPPLIREMLN
TR $\alpha$ 1	PAITRVVDFAKKLPMFSELPCEDQIILLKGC			FPPLFLEVFEF
TR $\beta$ 1	PAITRVVDFAKKLPMFCELPCEPQIILLKGC			FPPLFLEVFEF
TR $\beta$ 2	PAITRVVDFAKKLPMFCELPCEPQIILLKGC			LPPLFLEVFEF
PPAR $\alpha$	ETVTELTTEFAKAIIPGFANLNDLNDQVTLLKYG			LHPLLQEIYRD
PPAR $\delta$	ETVRELTEFAKSIPSFSSFLNDQVTLLKYG			LHPLLQEIYKD
PPAR $\gamma$	EAVQEITEYAKSIPGFVNLDLNDQVTLLKYG			LHPLLQEIYKD
RXR $\alpha$	KQLFTLVEWAKRIPHFSSELPDDQVILLRAG			IDTFLMEMLEA
RXR $\beta$	KQLFTLVEWAKRIPHFSSELPDDQVILLRAG			IDTFLMEMLEA
RXR $\gamma$	KQLFTLVEWAKRIPHFSDLTLEDQVILLRAG			IDTFLMEMLET

H3	—	H4	—	H5	—	H12
----	---	----	---	----	---	-----

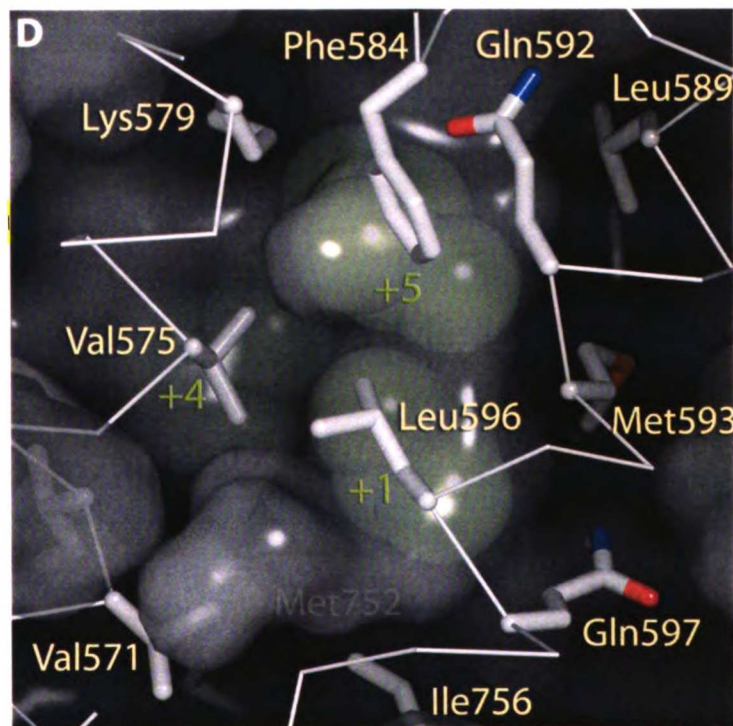
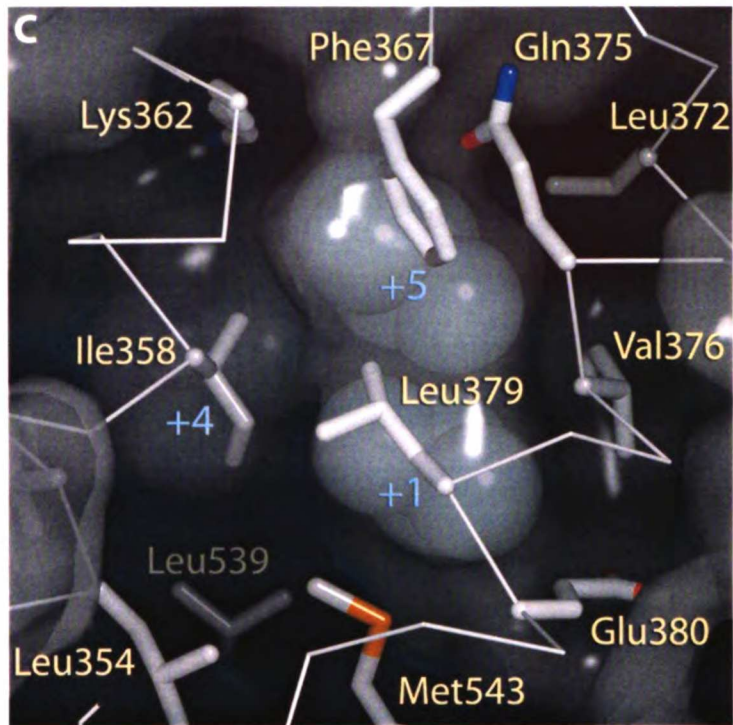
**Figure 2-9. Surface Complimentarity of Hydrophobic Motifs in the AR, ER $\alpha$ , and GR AF-2 Clefts**

A). AR-FxxLF, B). AR-LxxLL, C). ER $\alpha$ -GRIP1 (LxxLL) (Shiau et al. 1998), and D). GR-TIF2 (LxxLL) (Bledsoe et al. 2002). The inside surfaces of the AF-2 cleft in AR, ER $\alpha$ , and GR are depicted. The LBD is additionally shown as a C $\alpha$  trace with key side chains shown as white sticks. Phenylalanines and leucines of the FxxLF and LxxLL motifs are shown as spheres.



**Figure 2-9. Surface Complimentarity of Hydrophobic Motifs in the AR, ER $\alpha$ , and GR AF-2 Clefts (cont.)**

A). AR-FxxLF, B). AR-LxxLL, C). ER $\alpha$ -GRIP1 (LxxLL) (Shiau et al. 1998), and D). GR-TIF2 (LxxLL) (Bledsoe et al. 2002). The inside surfaces of the AF-2 cleft in AR, ER $\alpha$ , and GR are depicted. The LBD is additionally shown as a C $\alpha$  trace with key side chains shown as white sticks. Phenylalanines and leucines of the FxxLF and LxxLL motifs are shown as spheres.



## MATERIALS AND METHODS

### Protein Purification

Expression and purification of the AR LBD for crystallization were performed essentially as described (Matias et al. 2000). The cDNA encoding the chimp AR LBD (residues 663–919—human numbering), which displays 100% identity to the human form in protein sequence, was cloned into a modified pGEX-2T vector (Amersham Biosciences, Piscataway, New Jersey, United States) and expressed as a glutathione S-transferase (GST) fusion protein in the *E. coli* strain BL21 (DE3) STAR in the presence of 10  $\mu$ M DHT. Induction was carried out with 30  $\mu$ M IPTG at 17 °C for 16–18 h. *E. coli* cells were lysed in buffer (10 mM Tris, [pH 8.0], 150 mM NaCl, 10% glycerol, 1 mM TCEP, 0.2 mM PMSF) supplemented with 0.5  $\mu$ g/ml lysozyme, 5U/mL benzonase, 0.5% CHAPS, and 10  $\mu$ M DHT. All buffers for further purification steps contained 1  $\mu$ M DHT. Soluble cell lysate was adsorbed to Glutathione Sepharose 4 Fast Flow resin (Amersham Biosciences), washed with buffer containing 0.1% n-octyl  $\beta$ -glucoside, and eluted with 15 mM glutathione. After cleavage of the GST moiety with thrombin, final purification of the AR LBD was carried out using a HiTrap SP cation exchange column (Amersham Biosciences). Eluted AR LBD was dialyzed overnight at 4 °C against buffer containing 50 mM HEPES (pH 7.2), 10% glycerol, 0.2 mM TCEP, 20  $\mu$ M DHT, 150 mM Li<sub>2</sub>SO<sub>4</sub>, and 0.1% n-octyl  $\beta$ -glucoside, then concentrated to greater than 4 mg/ml for crystallization.

Purification of AR LBD for use in phage affinity selection was carried out as above without the final dialysis and concentration steps. The expression construct contained the AR LBD as an inframe fusion with GST in a modified pGEX-2T vector containing both a flexible region and an AviTag sequence (Avidity, Denver, Colorado, United States) allowing *in vivo* biotinylation. The GST–AR LBD fusion expression plasmid was cotransformed with a



plasmid-encoding *E. coli* biotin ligase (Avidity) into BL21 (DE3) STAR cells. Protein expression was carried out as above but with induction supplemented with 50  $\mu$ M biotin to ensure quantitative biotinylation of AR LBD.

### Phage Affinity Selections and Peptide Identification

Phage affinity selections were performed essentially as described (Paige et al. 1999). Biotinylated AR LBD (10 pmol/well) was incubated in streptavidin-coated Immulon 4 96-well plates (Dynatech International, Edgewood, New Jersey, United States) in TBST (10 mM Tris-HCl [pH 8.0], 150 mM NaCl, 0.05% Tween 20) with 1  $\mu$ M DHT for 1 h at 4 °C. Affinity selections were performed in TBST containing 1  $\mu$ M DHT. M13 phage distributed among 24 libraries displaying a total of greater than  $2 \times 10^{10}$  different random or biased amino acid sequences were added to the wells containing immobilized AR LBD and incubated for 3 h at 4 °C. After washing, bound phage were eluted using pH 2 glycine. Enrichment of phage displaying target-specific peptides was monitored after each round of affinity selection using an anti-M13 antibody conjugated to horseradish peroxidase in an ELISA-type assay.

Synthetic peptides corresponding to the deduced amino acid sequences from receptor-specific phage were tested for their ability to interact with purified AR LBD using a FRET-based assay format. Peptides were synthesized according to the deduced amino acid sequence displayed on phage with an additional C-terminal amino acid sequence consisting of SGSGK to allow the attachment of a biotin tag (Anaspec, San Jose, California, United States). Fluorophore conjugates were prepared by incubating either biotinylated peptides with streptavidin-criptate (Cis Bio International, Bagnols Sur Ceze Cedex, France), or biotinylated AR LBD with streptavidin-XL665 (Cis Bio). Interaction between peptide and

AR LBD was monitored by the ratio of energy transfer by excitation at 320 nm and emission at 625 nm and 665 nm.

### Surface Plasmon Resonance

Affinities of peptides to the AR LBD were determined with a Biacore (Piscataway, New Jersey, United States) 2000 instrument. A peptide derived from silencing mediator for RXR and TR 2 (SMRT2) served as a negative control. 1 mM peptide stock solutions in DMSO were diluted into HBS-P buffer (10 mM HEPES [pH 7.4], 150 mM NaCl, 0.005% Surfactant P20) to generate 10  $\mu$ M working solutions. HBS-P buffer was flowed through the cells to achieve a stable baseline prior to immobilization of the biotinylated peptides. To achieve the binding of approximately 250 RU of peptides to individual cells, working solutions of peptides were diluted to 100 nM in HBS-P buffer. Unbound streptavidin sites were blocked by injection of a 1 mM biotin solution at a rate of 10  $\mu$ l/min.

Purified AR LBD was diluted into HBS-P buffer to a concentration of 10  $\mu$ M and injected into all four Flowcells using the Kinject protocol at a flow rate of 10  $\mu$ l/min (contact time 360 s, dissociation time 360 s). Following the dissociation phase, the surface of the chip was regenerated to remove residual AR LBD by QuickInject of buffer containing 10 mM HEPES and 50% ethylene glycol (pH 11). Following the establishment of a stable baseline, the same procedure was repeated using a series of AR LBD dilutions (5  $\mu$ M, 1  $\mu$ M, and 300 nM) in an iterative manner. Analysis of the data was performed using BIAevaluation 3.0 software (Biacore). The SMRT2 signals were subtracted as background from the three remaining peptide signals. Data were best fit using the two-state conformational change model (Warnmark et al. 2001; Warnmark et al. 2002).

## Crystallization, Data Collection, and Refinement

Purified, concentrated AR LBD was combined with 3x to 6x molar excess of peptide and incubated 1 h at room temperature before crystallization trials. Complexes were crystallized using the hanging drop vapor diffusion method. Protein-peptide solution was combined in a 1:1 ratio with a well solution consisting of 0.6–0.8 M sodium citrate and 100 mM Tris or HEPES buffer (pH 7–8). Crystals typically appeared after 1–2 d, with maximal size attained within 2 wk. For data collection, crystals were swiped into a cryo-protectant solution consisting of well solution plus 10% glycerol before flash freezing in liquid nitrogen. The addition of ethylene glycol to a well concentration of 10%–20% was later found to both improve crystal quality and enable the freezing of crystals directly out of the drop.

Datasets were collected at 100K at the Advanced Light Source (Lawrence Berkeley Laboratory, Berkeley, California, United States), beamline 8.3.1, with either a ADSC Quantum 315 or Quantum 210 CCD detector. Data were processed using Denzo and Scalepack (Otwinowski and Minor 1997b). Molecular replacement searches were performed with rotation and translation functions from CNS (Brunger et al. 1998). Initial searches for AR-FxxLF were performed using the structure of AR-R1881 (PDB: 1E3G) with R1881 omitted from the search model. Subsequent searches for all other complexes were performed using the refined LBD structure from the AR-FxxLF complex. To minimize the possibility of model bias, FxxLF peptide and DHT were omitted from all molecular replacement searches. Protein models were built by iterative rounds of simulated annealing, conjugate gradient minimization, and individual B-factor refinement in CNS followed by manual rebuilding in Quanta 2000 (Accelrys, San Diego, California, United States) using  $\sigma_A$ -weighted  $2F_o - F_c$ ,  $F_o - F_c$ , and simulated annealing composite omit maps. Superposition of structures was performed with LSQMAN (Kleywegt 1996). Buried surface area calculations

were performed with CNS. All figures were generated with PyMOL (DeLano 2002).

Coordinates and structure factors for all complexes have been deposited in the Protein Data Bank. Accession numbers are listed in Table 2-2.

## **SUPPORTING INFORMATION**

### **Accession Numbers**

The Swiss-Prot (<http://www.ebi.ac.uk/swissprot>) accession numbers for the gene products discussed in this paper are AR (P10275), ARA54 (Q9UBS8), ARA55 (Q9Y2V5), ARA70 (Q13772), ER (P03372, Q92731), glucocorticoid receptor-interacting protein 1 NR box 3 (Q61026), GR (P04150), NR box 3 of TIF2 (Q15596), and TR  $\beta$  (P10828).

The Protein Data Bank (<http://www.rcsb.org/pdb>) accession numbers for the structures used in this paper are FxxFF (1T73), FxxLF (1T7R), FxxLW (1T79), FxxYF (1T7M), LxxLL (1T7F), unbound (1T7T), WxxLF (1T74), and WxxVW (1T76).

### **Acknowledgments**

We would like to thank Erin Anderson-Chisenhall for assistance in protein purification, James Holton and the staff at ALS beamline 8.3.1 for assistance in data collection, and Paul Webb for critical review of the manuscript. This work was supported by funds from the Prostate Cancer Foundation and National Institutes of Health grant R21 CA95324 to RJF.

**Table 2-2. Summary of Structures and Crystallographic Statistics**

Structure		FxxLF	FxxFF	FxxYF	FxxLW
Space group		<i>P</i> 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	<i>P</i> 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	<i>P</i> 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	<i>P</i> 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>
Unit cell	a	55.4	54.3	55.6	55.8
dimensions (Å)	b	66.2	66.6	66.6	66.1
	c	68.8	70.2	72.5	68.4
Resolution (Å)		20–1.4	20–2.2	20–1.6	20–1.8
Unique reflections		49611	13332	36190	23771
Completeness (%) <sup>a</sup>		98.2 (88.4)	99.3 (98.6)	99.9 (99.6)	98.8 (89.6)
R <sub>sym</sub> (%) <sup>a,b</sup>		7.2 (57.1)	9.7 (54.2)	5.2 (65.0)	8.4 (59.6)
<I/I <sub>σ</sub> > <sup>a</sup>		21.3 (2.0)	20.4 (3.1)	54.1 (3.4)	19.1 (2.2)
R <sub>cryst</sub> (%) <sup>c</sup>		19.6	20.0	19.7	20.0
R <sub>free</sub> (%) <sup>c</sup>		20.5	24.5	20.8	23.4
r.m.s.d. bond lengths (Å)		0.005	0.007	0.006	0.006
r.m.s.d. bond angles (°)		1.11	1.07	1.06	1.06
Average B-factor (Å <sup>2</sup> ):	Overall	22.4	40.3	28.2	29.0
	LBD	21.2	39.7	27.1	28.3
	Peptide	30.9	56.2	35.2	37.6
Buried surface area (Å <sup>2</sup> ):	Total	1012	926	1197	937
	+1 <sup>d</sup>	305	305	311	315
	+4 <sup>d</sup>	189	204	205	187
	+5 <sup>d</sup>	276	276	283	313
PDB accession code		1T7R	1T73	1T7M	1T79

<sup>a</sup>Numbers in parenthesis denote values for the highest resolution shell.

$$^b R_{\text{sym}} = \sum |I - \langle I \rangle| / \sum (I).$$

<sup>c</sup> $R_{\text{cryst}} = \sum |F_o - F_c| / \sum |F_o|$ , where  $F_o$  and  $F_c$  are observed and calculated structure factors, respectively;  $R_{\text{free}}$  was calculated similarly with a randomly selected set of reflections consisting of 5% of total reflections that were excluded from refinement.

<sup>d</sup>Values for side chain atoms only.

**Table 2-2. Summary of Structures and Crystallographic Statistics (cont.)**

Structure		WxxLF	WxxVW	LxxLL	Unbound
Space group		<i>P</i> 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	<i>P</i> 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	<i>P</i> 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	<i>P</i> 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>
Unit cell dimensions (Å)	<i>a</i>	56.4	53.4	54.2	56.2
	<i>b</i>	66.4	66.4	66.3	66.2
	<i>c</i>	72.1	70.6	69.4	68.5
Resolution (Å)		20–2.0	20–2.1	20–1.6	20–1.7
Unique reflections		18803	15170	32737	28564
Completeness (%) <sup>a</sup>		99.8 (85.3)	99.9 (99.9)	97.2 (93.6)	99.3 (99.2)
<i>R</i> <sub>sym</sub> (%) <sup>a,b</sup>		5.9 (54.1)	6.7 (56.3)	4.5 (46.3)	6.1 (45.7)
$\langle I/I_{\sigma} \rangle^a$		28.2 (2.6)	26.2 (3.3)	31.4 (2.7)	28.5 (3.2)
<i>R</i> <sub>cryst</sub> (%) <sup>c</sup>		21.6	20.8	19.9	18.1
<i>R</i> <sub>free</sub> (%) <sup>c</sup>		23.9	23.9	21.8	20.5
r.m.s.d. bond lengths (Å)		0.007	0.006	0.006	0.006
r.m.s.d. bond angles (°)		1.02	1.08	1.07	1.07
Average B-factor (Å <sup>2</sup> ):	Overall	43.5	44.1	25.1	21.2
	LBD	42.3	43.1	23.9	20.3
	Peptide	71.0	65.9	44.1	–
Buried surface area (Å <sup>2</sup> ):	Total	970	903	984	–
	+1 <sup>d</sup>	364	351	271	–
	+4 <sup>d</sup>	194	149	193	–
	+5 <sup>d</sup>	281	287	241	–
PDB accession code		1T74	1T76	1T7F	1T7T

<sup>a</sup>Numbers in parenthesis denote values for the highest resolution shell.

<sup>b</sup> $R_{\text{sym}} = \sum |I - \langle I \rangle| / \sum (I)$ .

<sup>c</sup> $R_{\text{cryst}} = \sum |F_o - F_c| / \sum |F_o|$ , where *F*<sub>o</sub> and *F*<sub>c</sub> are observed and calculated structure factors, respectively; *R*<sub>free</sub> was calculated similarly with a randomly selected set of reflections consisting of 5% of total reflections that were excluded from refinement.

<sup>d</sup>Values for side chain atoms only.

## Chapter 3

# The AR N-terminal domain and ART-27

Collaborator on this chapter:

Peter K. Hwang

## INTRODUCTION

AR is unusual among the nuclear receptors in that ligand independent AF-1 activity is the dominant transactivation function. AF-1 has been mapped to the N-terminal domain (NTD) of the receptor. At approximately 550 amino acids, the N-terminal domain (NTD) represents more than half the protein mass of AR. It is the site of numerous post-translational modifications as well as the location of enigmatic poly-glutamine, poly-glycine, and poly-proline tracts whose role in AR function is unknown. Most significantly, it is the site of interaction for numerous coactivators, corepressors and other proteins (McEwan 2004). In the absence of these associations, the NTD is largely unstructured (Reid et al. 2002; Kumar et al. 2004). It is the synergistic folding and association of the NTD with these coregulators that presumably drive AR AF-1 activity. Due to the poor conservation of the NTD among nuclear receptors, compounds that disrupt these interactions may be an effective alternative to traditional anti-androgen compounds in the development of novel prostate cancer therapeutics. Despite the critical role AF-1 plays in AR function and prostate cancer pathology, the structural basis of AF-1 activity, and more generally of NTD function, remains unknown. As a first step toward a better understanding of NTD structure and function, we sought to isolate and characterize a complex between the NTD and the p160 coactivators SRC1, GRIP1 (SRC2), and RAC3 (SRC3), as well as the novel AR coregulator ART-27.

### High-Throughput Coexpression and Domain Mapping

In the structure determination of protein complexes a fundamental problem is that of domain boundaries. Often protein complexes consisting of full length proteins either do not express well, do not crystallize, or do not diffract to high resolution because of the



presence of non-interacting domains that contribute to disorder. It is then necessary to undertake a domain mapping experiment, where one isolates interacting fragments and then tests them for solubility, crystallizability, and diffraction. This involves the construction and screening of dozens of variants and is very time consuming.

To address this issue, we developed a high-throughput cloning and expression system, which is potentially extendable to any protein complex of interest. With the aid of robotics, protein fragments of varying length will be systematically coexpressed and screened in a combinatorial fashion for association. Like two-hybrid approaches, no prior knowledge of protein structure is necessary. However, while two-hybrid approaches may be useful to initially identify protein partnerships, fragments discovered by two-hybrid domain mapping may not necessarily express well in bacteria and yield the amounts of protein required for crystallographic analysis. Our high-throughput strategy has the advantage of screening for not only interaction, but also for soluble, highly expressed, and properly folded complexes amenable to crystallization.

This will be achieved by constructing a series of oligonucleotides that step along the NTD or coregulator gene sequence at regular intervals. These will be used to PCR fragments of varying length of both NTD and coregulator. PCR products will be recombined into a Gateway entry vector (pDONR221 or pDONR Zeo) for subsequent recombination into an expression vector. Recombination cloning eliminates the need for restriction digests, gel purification, and ligation normally required in conventional cloning strategies and makes it possible to go from PCR to expression clone within one day. *E. coli* will be systematically cotransformed with all possible combinations of generated NTD and coregulator fragments. In one scenario, 10 forward primers and 10 reverse primers will be used to generate 100

different NTD fragments; the same will be done with coregulator. NTD and coregulator fragments will then be combined to give 10,000 possible complexes.

Coexpression will initially be assayed small scale in a 96 well format using an ELISA based pulldown assay, enabling rapid readout of complex formation. In this assay, crude cell lysates will be loaded onto a Ni-NTA coated plate to immobilize the His-tagged protein. If there is association, the His-tagged partner will pulldown fragments of the GST, MBP, or S-tagged partner. The amount of associating protein will be measured using antibody directed against the partner's tag and a secondary antibody conjugated to an enzyme (ie. horseradish peroxidase) with a colorimetric or fluorescent substrate.

One limitation of this assay is that it does not directly differentiate between low expression and poor association. Since we are screening for the highest expressing, tightest binding, and most soluble fragments for crystallography, this is not of concern. However, this limitation can be addressed through the use of two control assays. The first involves the use of a protein that can be used to saturate all unbound Ni-NTA sites. The amount of reporter protein bound will be inversely proportional to the level of expression of the His-tagged partner. A large difference in ELISA signal after addition of the reporter protein is indicative of poor expression. No change in signal however, indicates poor association or low expression of the non-His-tagged partner. To differentiate between the latter two possibilities, a second control assay can be performed in parallel which will directly measure expression levels of the non-His-tagged partner. This assay can potentially involve the use of a Nano-tag, which is a peptide tag that binds to streptavidin with high affinity (Lamla and Erdmann 2004). For this assay the non-His-tagged partner must be double tagged (eg. N-terminal GST, C-terminal nano-tag or S-tag). Protein will be immobilized with one tag (eg. on a glutathione or Streptavidin coated plate) and detected using an ELISA with antibody

directed toward the second tag (ie. GST, S-tag), similar to the initial interaction assay. The most soluble, stable complexes will then be selected for large-scale expression, purification, and crystallization trials according to standard protocols.

Coexpression is a key element of our approach. It has recently been shown that the NTD and a C-terminal fragment from the p160 coactivator SRC1 possess little intrinsic structure until association (Reid et al. 2002; Kumar et al. 2004). Such synergistic folding has been observed previously in the structure of the interaction domain of CBP with ACTR (Demarest et al. 2002). Independently, each fragment was largely disordered, but upon association the fragments formed a tightly intertwined and cooperatively folded heterodimer. Expressed alone, NTD and coregulator fragments may be unstable and difficult to purify in the large quantities required for structure determination. By coexpressing individual fragments on two compatible vectors, or on a single bicistronic vector, fragments will be able to fold in the presence of their binding partner, thereby circumventing the stability and solubility problems associated with misfolded proteins. Such a coexpression strategy was critical in the structure determination of not only the CBP-ACTR complex, but also of the adaptin AP2 core and the alpha-beta subunit complex of the voltage gated calcium channel (Owen et al. 2000; Van Petegem et al. 2004).

Rapid screening of thousands of potential complexes will allow domain mapping in situations where there are multiple simultaneous protein partners. The coexpression system we have developed is not limited to two proteins and can be extended to the simultaneous expression of up to six proteins. The ability to coexpress and screen several fragments at once may prove crucial, since third partners may be necessary for NTD-coregulator interactions. Binding of SRC1 to the NTD is enhanced by the presence of a third partner, RAP74, a subunit of the general transcription factor TFIIF (Kumar et al. 2004). Recent

evidence suggests that ART-27 may be similar, as analytical ultracentrifugation of HeLa cell extracts show that native ART-27 sediments as part of a multiprotein complex (Markus et al. 2002; Gstaiger et al. 2003).

### **Construction of Gateway Coexpression Vectors**

To facilitate the rapid entry of protein fragments into expression vectors we utilized the Gateway recombination system from Invitrogen. Recombination alleviates the need for restriction digests and ligation. This system is ideal for high-throughput applications as it is rapid and robust. The Gateway system has been used successfully at several NIH Structural Genomics Centers. However, currently available Gateway vectors are not suitable for coexpression of multiple proteins. In order to coexpress proteins on different plasmids, the plasmids must have compatible origins of replication and compatible promoters. We modified the Duet series of vectors from Novagen to be Gateway compatible. The Duet vectors are a series of bicistronic vectors under the control of T7 $lac$  promoters with compatible origins of replication to be used for coexpression. Each vector contains two multiple cloning sites, one with a N-terminal His-tag and the other with a C-terminal S-tag. We have constructed a series of Gateway compatible Duet variants allowing for the coexpression of His-tagged, S-tagged, native, or GST-fusion proteins. (Table 3-1) Two compatible MBP-fusion vectors from NEB, pMAL-c2E and pMAL-p2E, were also adapted for the Gateway system. Constructed vectors are summarized in Table 3-1.

**Table 3-1. Gateway Coexpression Vectors**

Compatible vectors are shaded different colors. (Vectors of the same color are incompatible with each other.)

vector	origin	resistance	tag (N/C)	bicistronic?	base pairs
pCDFgw	CloDF13	Streptomycin	6xHis/ -	yes	5468
pCDF-GSTgw	CloDF13	Streptomycin	GST/ -	no	5917
pETgw	ColE1	Ampicillin	6xHis/S-tag	no	6938
pET•GSTgw	ColE1	Ampicillin	GST/ -	no	7556
pET•Sgw	ColE1	Ampicillin	none/S-tag	no	7119
pRSFgw	RSF1030	Kanamycin	6xHis/ -	yes	5516
pRSF•GSTgw	RSF1030	Kanamycin	GST/ -	no	5965
pMAL(c2E)gw	ColE1	Ampicillin	MBP/ -	no	8334
pMAL(p2E)gw	ColE1	Ampicillin	MBP/ -	no	8409

The first vectors to be adapted were the Duet series of vectors from Novagen, which include the pETDuet-1 (Amp<sup>R</sup>), pRSFDuet-1 (Kan<sup>R</sup>), pCDFDuet-1 (Sm<sup>R</sup>), and pACYCDuet-1 (Cm<sup>R</sup>) vectors, and the pMAL-c2E and pMAL-p2E MBP-fusion vectors from NEB. The Duet vectors are bicistronic with an N-terminal 6xHis-tag in MCS1 and an optional C-terminal S-tag in MCS2. The two different pMAL vectors vary in the presence (pMAL-p2E) or absence (pMAL-c2E) of a periplasmic signaling sequence: pMAL-p2E codes for periplasmic expression, and pMAL-c2E for cytoplasmic. The vectors enable the expression of N-terminally MBP-tagged proteins with an enterokinase protease cleavage site in the linker.

Vectors were to be converted using the Gateway conversion kit from Invitrogen. The kit includes, in all three possible reading frames, the Gateway cassette containing the *ccdB* gene and a chloramphenicol resistance marker. To convert a vector, it is restriction digested at the multiple cloning site, treated with Klenow to fill in the sticky ends and generate blunt ends, and then blunt end ligated with the proper Gateway cassette to produce the final Gateway compatible expression vector. Vectors containing the Gateway cassette must be propagated in DB3.1 cells due to the *ccdB* gene used as a selective marker in subsequent recombinations (The *ccdB* gene product interferes with *E. coli* DNA gyrase. DB3.1 cells harbor a mutation in DNA gyrase that renders them resistant to the effects of the *ccdB* gene). Initially, all four Duet vectors and the pMAL vectors were put through this protocol using the EcoR I and Hind III sites present in the multiple cloning sites of these vectors and Gateway cassettes RfC.1 for the Duet vectors and RfB for the pMAL vectors. However, we were successful in converting only the pRSFDuet-1 and pMAL vectors using this protocol: the resulting vectors were called pRSFgw, pMAL(c2E)gw, and pMAL(p2E)gw, respectively. Blunt end ligation was highly inefficient and produced colonies with the

Gateway cassette inserted backwards. Ligation of the Gateway cassette into pACYCDuet-1 appeared to be successful, however selected colonies appeared to harbor plasmids containing the product of a spurious recombination, likely due to the presence of two chloramphenicol resistance genes in one plasmid. As a result, we abandoned attempts to convert the pACYCDuet-1 vector. This was not problematic since we still had three non-chloramphenicol vectors to work with.

Since the vector conversion by blunt-end ligation of the Gateway cassette was largely unsuccessful, we took advantage of the vectors' homology to one another to construct the pCDFDuet-1 Gateway vector (pCDFgw). Utilizing the Mlu I site located 564 base pairs before MCS1, and the Afl II site located at the end of MCS1, we ligated the fragment in between these two sites from the newly constructed pRSFgw vector containing the Gateway cassette, into pCDFDuet-1 to generate pCDFgw.

To convert the pETDuet-1 vector, we took an alternative approach. The strategy was to ligate the Gateway cassette into the pETDuet-1 vector using restriction sites present in the two multiple cloning sites. The Gateway cassette was PCR amplified with primers (forward - GATE1: 5'-GT AGG AGC TCA GAT CTT CAA ACA AGT TTG TAC AAA AAA GC-3', reverse - GATE2: 5'-AGTC AAG CTT GGT ACC TCG AAC CAC TTT GTA CAAG-3') containing Sac I and Bgl II sites on the forward primer and Hind III and Kpn I sites on the reverse primer. This enabled the construction of two Gateway compatible variants of the pETDuet-1 vector. The first, utilizing the Sac I and Kpn I site, encodes an N-terminal 6xHis-tag and an optional C-terminal S-tag (pETgw). The other, utilizing the Bgl I and Kpn I sites encode a protein with no affinity tag and an optional C-terminal S-tag (pET•Sgw) allowing the production of native protein. The one disadvantage of the pETgw and pET•Sgw vectors is that by bridging the two multiple cloning sites to allow an optional

C-terminal S-tag, the ability to coexpress two proteins bicistronically is lost. However, this is not a problem because the pRSFgw and pCDFgw vectors retain their bicistronic capability. Moreover, if absolutely necessary, it is easily possible to construct a pETgw vector retaining its bicistronic expression by utilizing the Sac I and Hind III sites, instead of the Sac I and Kpn I sites.

To increase the repertoire of available tags, a series of N-terminally GST-tagged Gateway vectors (pRSF•GSTgw, pCDF•GSTgw, pET•GSTgw) was constructed. GST insert was PCR amplified off the pGEX-4T-1 vector (GE Healthcare) using a forward primer (GSTDuet\_f: 5'-GA TAT ACC ATG GGC TCC CCT ATA CTA GGT TAT TGG-3') containing a Nco I site and a reverse primer (GST\_r: 5'-TGT TTG AAG ATC TGA GCT CGC ATC CGA TTT TGG AGG ATG GTC-3') containing Sac I and Bgl II sites. The GST insert was ligated into the Duet vectors using Nco I and Sac I, swapping out the 6xHis-tag for a GST. To make these vectors Gateway compatible, the Gateway cassette was ligated into the Duet-GST using Sac I and Kpn I sites, similar to what had been done to construct the pETgw vector. All constructed Gateway vectors were successfully tested for recombination and protein expression using MBP as a positive control. Protocols are outlined below.

### Gateway Primer Design and PCR

Entry into the Gateway system requires the addition of *attB* recombination sites flanking the PCR insert. The recommended primer sequence from Invitrogen for the forward primer is 5'-GGGG ACA AGT TTG TAC AAA AAA GCA GGC TNN...-3', and for the reverse primer 5'-GGGG AC CAC TTT GTA CAA GAA AGC TGG GTN...-3' (Gateway system manual - Invitrogen). Underlined sequences represent the *attB1* and *attB2*



recombination sites in the forward and reverse primers, respectively. Since constructs would eventually be used to express protein for crystallography, we wished to add a six residue TEV protease site to enable cleavage of affinity tags. The addition of both Gateway recombination sites and TEV cleavage sites was achieved by a two-round PCR protocol whereby TEV residues would be added in the first round and *attB* recombination sites in the second round. The second round of PCR utilizes a set of “universal” Gateway primers that can be used for all proteins, while the first round primers contain template specific sequences in addition to the TEV site. To aid protein expression, codons for the TEV site were later optimized to encode the most frequently occurring codons in *E. coli*. Gateway primer sequences are summarized in Figure 3-1.

PCR for Gateway is performed using standard PCR protocols except 1  $\mu$ L of the first round PCR product is used as template for the second round of PCR. Generally, 0.3-0.4  $\mu$ M (1.5-2  $\mu$ L of 10  $\mu$ M primer) of each primer and 0.2 mM dNTP (1  $\mu$ L 10 mM dNTP) is used per 50  $\mu$ L reaction. Both first and second round PCRs involve an initial 2 minute melting step, then 25 cycles of melting, annealing, and extension followed by a 10 minute final extension - annealing and extension temperatures are dependent on the primers and polymerase used. The polymerase used with the highest success was Accuprime Pfx, a high-fidelity proofreading polymerase from Invitrogen. In cases where this polymerase failed to produce product, the less high-fidelity Platinum Taq polymerase (Invitrogen) was used.

Prior to the first recombination (BP recombination), PCR product is purified by using PCR cleanup kits from Qiagen or Invitrogen. The Invitrogen kit PCR cleanup kit has the advantage of being able to effectively purify small inserts (>300bp) from primers. Fragments smaller than this are gel purified using the Qiagen gel extraction kit. However, it was found that it is possible to do the BP recombination without purification of PCR

### Figure 3-1. Gateway Primers

PCR primers for entry into the Gateway system are shown. Overlapping bases in the forward primers are shown in red; those in the reverse primers are shown in blue. TEV cleavage site is indicated in green; cleavage occurs between residues in bold.

#### *Gateway 1<sup>st</sup> round:*

Forward:

5' AAC CTG TAC TTC CAG TCC + N-terminal codons

N L Y F **Q** S

Forward (codon optimized version):

5' AAT CTG TAC TTT CAG AGC + N-terminal codons

N L Y F **Q** S

Reverse:

5' GTA CAA GAA AGC TGG GTC stop + C-terminal codons

Reverse - antiparallel strand for design:

5' C-terminal codons + stop GAC CCA GCT TTC TTG TAC

#### *Gateway 2<sup>nd</sup> round:*

Forward:

*attB1*

5' GGGG ACA AGT TTG TAC AAA AAA GCA GGC TCC GAA AAC CTG TAC TTC CAG

T S L Y K K A G S E N L Y F **Q**

Forward (codon optimized version):

*attB1*

5' GGGG ACA AGT TTG TAC AAA AAA GCA GGC TCC GAA AAT CTG TAC TTT CAG

T S L Y K K A G S E N L Y F **Q**

Reverse:

*attB2*

5' GGGG AC CAC TTT GTA CAA GAA AGC TGG GTC

## Gateway BP Recombination

Construction of an expression clone by the Gateway system involves two separate recombination reactions. The first, the BP recombination, involves transfer of PCR product into an “entry vector.” These vectors are easy to manipulate as they are small, high copy, and contain M13 sites for sequencing. Once in an entry vector, it is possible to transfer insert into any number of different “destination vectors” for expression in a variety of different expression systems. We utilized the pDONR 221 (Kan<sup>R</sup>) and pDONR Zeo (Zeo<sup>R</sup>) entry vectors (Invitrogen). For the BP reaction, we largely followed the protocol in the Invitrogen manual except reaction volumes were halved to 10  $\mu$ L and incubations were carried out overnight instead of 1hr. A typical BP reaction consisted of 5  $\mu$ L PCR product, 1  $\mu$ L entry vector (from a mini-prep), 2  $\mu$ L 5x BP reaction buffer, and 2  $\mu$ L BP clonase. In most cases it is even possible to cut down the amount of BP clonase used to 1  $\mu$ L. Reactions were carried out overnight at room temperature, then treated with proteinase K for 10 minutes at 37°C. To get the most recombinants, 50  $\mu$ L of chemically competent DH5 $\alpha$  cells were transformed with 2  $\mu$ L of the BP reaction mix. After plating out on selective media (in our case kanamycin or zeocin) only cells transformed with successfully recombined vector survive, since bacteria transformed with unrecombined vector succumb to the toxic effects of the *ccdB* gene on the Gateway cassette.

BP reactions were generally high efficiency - plating out 100  $\mu$ L of a 500  $\mu$ L transformation usually yielded more than enough colonies to easily pick clones. The few cases where BP recombination failed were due to failed PCR reactions. After BP recombination, clones are picked for Qiagen minipreps. In cases where colonies were observed, BP reactions were invariably successful. Recombination can, however, be verified by restriction digest with BsrG I which cuts at the ends of the Gateway recombination sites.

Therefore if recombination was successful, one can expect to see a band corresponding to the vector and another corresponding to the size of the insert. If recombination has failed, one sees three bands: one corresponding to the vector backbone, and two characteristic bands (1454, 786) corresponding to the Gateway cassette, which has an internal BsrG I site in the *ccdB* gene in addition to ones in the recombination sites. A typical 10  $\mu$ L digest consists of 3  $\mu$ L entry clone, 1  $\mu$ L 10x NEB buffer 2, 1  $\mu$ L 10x BSA (1 mg/mL), 1  $\mu$ L BsrG I, and 4  $\mu$ L water - reactions were incubated at 37°C for at least 1 hour. A mini-prep of 4-5 mL of culture yields more than enough pDONR vector for sequencing - 6-7  $\mu$ L per sequencing reaction was generally sufficient to obtain good sequencing results

### Gateway LR Recombination

After generating an entry clone in a pDONR vector, one can then perform an LR recombination into a destination vector for expression. A typical LR recombination consisted of 3  $\mu$ L destination vector (from a miniprep), 1  $\mu$ L entry clone, 2  $\mu$ L 5x LR reaction buffer, 2  $\mu$ L LR clonase, and 2  $\mu$ L TE. Like the BP, LR recombination reactions were halved in volume to 10  $\mu$ L and carried out overnight at room temperature. Overnight incubation was far more critical for the success of the LR reaction, since in general, LR recombinations were found to be far less efficient than the BP. Halving the amount of LR clonase used to 1  $\mu$ L was attempted with varied success - generally the recommended amount of clonase (2  $\mu$ L) was used per reaction to reliably get recombinants. After transformation of 50  $\mu$ L of DH5 $\alpha$  cells with 2  $\mu$ L of the LR recombination mix, 200  $\mu$ L to all of a 500  $\mu$ L transformation was used to get a good number of colonies. Similar to BP reactions, after minipreps clones can be screened for successful recombination by digestion with BsrG I. If recombination is successful, one will see a band corresponding to the

destination vector backbone and another band corresponding to the insert. If it is not successful, one sees three bands: one corresponding to the vector backbone, and two characteristic bands (1278, 398) corresponding to the Gateway cassette (there is also a smaller 20-60 bp fragment that runs off the gel).

Unlike the BP, however, background colonies containing unrecombined vector or pDONR entry clones were sometimes present. This was puzzling as unrecombined Gateway vectors should be toxic to DH5 $\alpha$  cells and pDONR plasmids should not confer resistance to antibiotics used to select LR recombinants. The reason for this is unknown. Low recombination frequencies seemed to be dependent on insert. The LRH hinge-LBD and LBD proved to be especially difficult and many colonies were screened before a positive expression clone was found. Insert size also seemed to be a factor because full length NCoA-62 (1611 bp) also proved to be problematic. Generally, inserts less than 1 kb were successfully recombined with a high rate of success into destination vectors.

Anomalous behavior was also found using the pCDFgw vector. In one example, LR recombination of MBP-pDONR 221 with pCDFgw to generate a 6xHis-MBP pCDFgw is expected to yield a 4939 bp product yielding 3716, 1151, and 60 bp fragments upon digest with BsrG I, and 2509 and 2426 bp fragments upon digest with Hpa I and Xho I. However, upon digest with BsrG I, only one fragment of about 2500 bp was observed and digest with Hpa I and Xho I yielded fragments of about 2500 and 1100 bp. These fragments can not be explained by the presence of entry vector alone (Hpa I, Xho I digest: 3669 bp, BsrG I digest: 2510, 1151) or unrecombined pCDFgw destination vector (Hpa I, Xho I digest: 3038, 2426, BsrG I digest: 3716, 1278, 398, 60). The cause of this has yet to be identified, but likely it is due to an aberrant recombination event which results in concomitant loss of the Gateway cassette and gain of the streptomycin gene.

## ART-27

ART-27 is a recently discovered AR coactivator that is unrelated to the well characterized p160 family of coactivators. It was isolated as a NTD interacting protein in a modified yeast two-hybrid screen using a cDNA library isolated from androgen stimulated LNCaP cells (Markus et al. 2002). It is identical to a previously identified ORF of unknown function called UXT, and STAP1, a SKP2 associating member of the prefoldin family of molecular chaperones (Schroer et al. 1999; Gstaiger et al. 2003). While members of the p160 family are typically around 1400 amino acids, ART-27 is a small protein consisting of only 157 amino acids. It interacts predominantly with residues 153-336 of the AR NTD (Markus et al. 2002). AR coregulators are thought to modulate AR activity depending on biological context, and consistent with this notion, ART-27 is expressed at higher levels in differentiating prostate cells than in proliferating (Taneja et al. 2004). Expression of ART-27 in LNCaP cells inhibited cell proliferation (Taneja et al. 2004). This suggests that ART-27 may serve as a NTD adaptor protein for recruitment of coactivators involved in prostate differentiation.

### Pilot Expression Trials

All expression trials were carried out by inoculating 25 mL of LB broth 1/200 with overnight culture. If coexpression was being assessed, culture volume was doubled to 50 mL to have 25 mL of culture per affinity pulldown. Cells were grown at 37°C until an  $A_{600}$  of approximately 0.5. For initial expression trials, cells were induced with 1 mM IPTG to ensure that IPTG concentration would not be a factor if there was no expression. If induction was to be carried out at 37°C, cells were induced for 3 hours; if induction was to be carried out at 15°C, cells were moved to 15°C, allowed to equilibrate, and then induced

overnight for 16-20 hours. To assess protein expression in total cell extracts, small aliquots of culture were taken and resuspended in Bugbuster (with benzonase) (Novagen). To avoid streaky lanes on gels and to normalize culture densities, it was found that a good volume of culture to take for cell extracts could be found by dividing 250 by the  $A_{600}$  [ $V_{\text{culture}} (\mu\text{L}) = 250/A_{600}$ ]. After pelleting cells, they are resuspended in 50  $\mu\text{L}$  of Bugbuster, allowed to lyse, then boiled with 50  $\mu\text{L}$  of SDS loading buffer.

Cells were harvested by centrifugation at 5000 g for 10 minutes. Pellets were then resuspended in 1.4 mL of Bugbuster plus protease inhibitor cocktail for cell lysis (1.4 mL Bugbuster/ 25 mL culture). For affinity pulldowns, 50  $\mu\text{L}$  of glutathione, amylose, or Talon (later Ni-NTA) slurry was equilibrated in 20 mM Tris pH 7.5, 200 mM NaCl, and 0.05% Tween-20 ("TBST") or the equivalent with HEPES buffer. This gave a final bed volume of about 25  $\mu\text{L}$ . For IMAC resin, 10 mM imidazole was also added to reduce non-specific binding. Cell lysates were batch bound to affinity resin at 4°C in Eppendorf tubes for 30 minutes. After binding, beads were spun down, and washed with at least 5 tube volumes (5 x 1.5 mL) of TBST. Protein was eluted by the addition of 25  $\mu\text{L}$  of 1 M imidazole, 500 mM maltose, or 200mM glutathione to approximately 25  $\mu\text{L}$  of washed slurry.

### **Expression of ART-27**

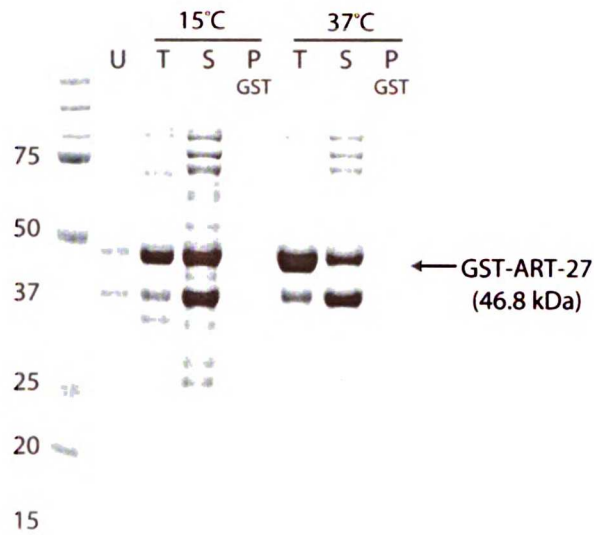
ART-27 was an ideal candidate to pilot coexpression trials with the NTD. Its small size allowed us to focus mapping efforts on the NTD, thereby reducing the number of possible fragments two-fold. Using the Gateway system, NTD residues 142-498, and full length ART-27 were amplified by PCR (Peter Hwang) and then recombined into pRSFgw (6xHis) and pDEST15 (GST) (Invitrogen), respectively. ART-27 was expressed both alone, and with NTD in BL21 Star cells. As shown in Figure 3-2, expression of GST-ART-27 was

moderately well expressed but insoluble when expressed by itself at 37°C. However when expressed at 15°C or coexpressed with NTD, expression was poor. It was then discovered that the T7 pDEST Gateway vectors from Invitrogen are incompatible for coexpression with the T7*lac* Duet vectors because of promoter differences (Novagen - Duet manual). T7 and T7*lac* promoters are incompatible with each other, likely because of differences in expression levels brought about by the tighter control of transcription with T7*lac* promoters. Therefore, ART-27 was recloned into pCDFgw and expressed either alone or with NTD-pRSFgw. However, as shown in Figure 3-3 expression was similar, with ART-27 expressing moderately well on its own and poorly when coexpressed with NTD.



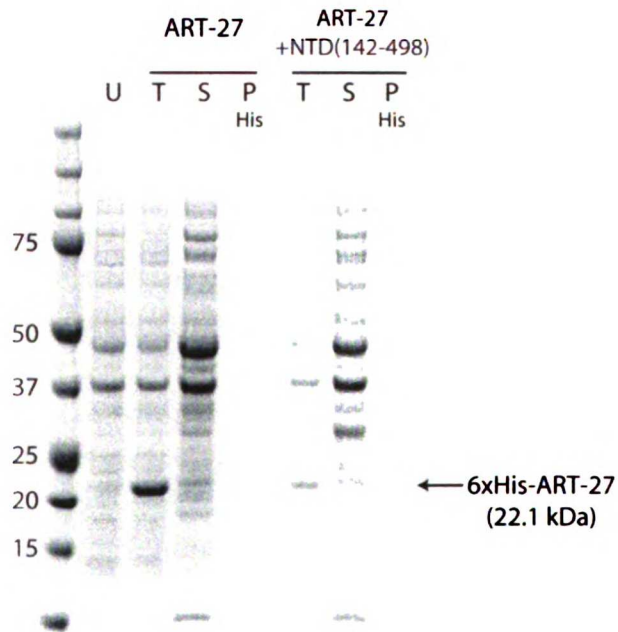
**Figure 3-2**

Expression of GST-ART-27 using pDEST15. U = uninduced cell extract, T = induced cell extract, S = soluble fraction, P = elutions from glutathione beads



**Figure 3-3**

Coexpression of 6xHis-ART-27 with 6xHis-NTD(142-498). U = uninduced cell extract, T = induced cell extract, S = soluble fraction, P = elutions from Talon beads



## ART-27 Gene Synthesis

Examination of ART-27 codon usage revealed the presence of high number of rare *E. coli* codons in the native ART-27 gene sequence that could potentially hinder expression (Figure 3-4). Because of its small size (474 bp), synthesizing in-house a version of the ART-27 gene with codons optimized for expression in *E. coli* was found to be possible without too much effort - in a previous study, it was found that 500 bp was the optimal size for gene synthesis using overlapping oligonucleotides (Kodumal et al. 2004). We used a modified version of PCR gene synthesis protocols previously published (Stemmer et al. 1995; Kodumal et al. 2004). The strategy involves two PCR reactions: the first involves assembly of the gene by combining a series of overlapping oligos that will be extended over the course of the PCR to cover the length of the entire gene; the second involves amplification of the full length gene using primers complementary to the ends of the gene.

We therefore ordered a series of 20 oligos spanning the length of ART-27 to synthesize a codon optimized version of the gene. Oligos used to construct ART-27 are listed below.

ART27-1: **aacctgtacttccagagcATGGCGACCCCGCCGAAACGTC**  
ART27-2: **TTTTCGCCGGTTCGCTTCCACCGCACGACGTTTCGGCGGGGTTCGCC**  
ART27-3: **GTGGAAGCGACCGGCGAAAAAGTGCTGCGTTATGAAACCTTTATT**  
ART27-4: **GCAGATCACGCTGCAGCACATCGCTAATAAAGGTTTCATAACGCA**  
ART27-5: **TGTGCTGCAGCGTGATCTGCGTAAAGTGCTGGATCATCGTGATAA**  
ART27-6: **ATATTTGCCAGCTGTTTCATACACTTTATCACGATGATCCAGCAC**  
ART27-7: **ATGAACAGCTGGCGAAATATCTGCAGCTGCGTAATGTGATTGAAC**  
ART27-8: **TCGCTATGTTTCGCTTCCCTGCAGACGTTCAATCACATTACGCAGC**  
ART27-9: **CAGGAAGCGAAACATAGCGAACTGTATATGCAGGTGGATCTGGGC**  
ART27-10: **CCACGGTATCCACAAA<sub>g</sub>AAATTGCAGCCCAGATCCACCTGCATAT**  
ART27-11: **TTTcTTTGTGGATACCGTGGTGCCGGATACCAGCCGTATTTATGT**  
ART27-12: **CAGAAA<sub>g</sub>AAGCCATAGCCCAGCGCCACATAAATACGGCTGGTATC**  
ART27-13: **TGGGCTATGGCTTcTTTCTGGAAGTACCCTGGCGGAAGCGCTGA**  
ART27-14: **AGGCTGCTTTTACGATCAATAAATTTTCAGCGCTTCCGCCAGGGTC**

ART27-15: **ATTGATCGTAAAAGCAGCCTGCTGACCGAACTGAGCAATAGCCTG**  
ART27-16: **CTTTAATATTCATGCTATCTTTGGTCAGGCTATTGCTCAGTTCGG**  
ART27-17: **AGATAGCATGAATATTTAAAGCGCATATTCATATGCTGCTGGAAGG**  
ART27-18: **CTGCAGGCCCTGCAGTTCACGCAGGCCTTCCAGCAGCATATGAAT**  
ART27-19: **GTGAACTGCAGGGCCTGCAGAATTTTCCGGAAAAACCGCATCATT**  
ART27-20: **gtacaagaaagctgggtcTTAATGATGCGGTTTTTCCGGA**

Codon optimization and design of oligos to construct the gene were performed with DNABuilder version 3.0 (<http://cbi.swmed.edu/computation/cbu/DNABuilder.html>). Except for the first and last one, each oligo was 45bp and overlapped with neighboring oligos by 20 bp. Each oligo was resuspended in buffer EB (10 mM Tris pH 8.5, Qiagen) to 100  $\mu$ M. A 5  $\mu$ M working stock was then made by taking 5  $\mu$ L of each oligo and then pooling them together to 100  $\mu$ L (5  $\mu$ L of each). PCR was performed using the AccuPrime Pfx polymerase (Invitrogen). The assembly PCR consisted of 5  $\mu$ L 10x AccuPrime Pfx buffer, 10  $\mu$ L ART-27 oligo mixture (5 $\mu$ M each), 0.5  $\mu$ L AccuPrime Pfx (2.5 U/ $\mu$ L), and 34.5  $\mu$ L water to bring the final volume to 50  $\mu$ L. The reaction was then cycled 1x at 95°C for 2 minutes; 25x at 95°C for 30 seconds, 60°C for 30 seconds, 68°C for 30 seconds; 1x at 68°C for 2 minutes.

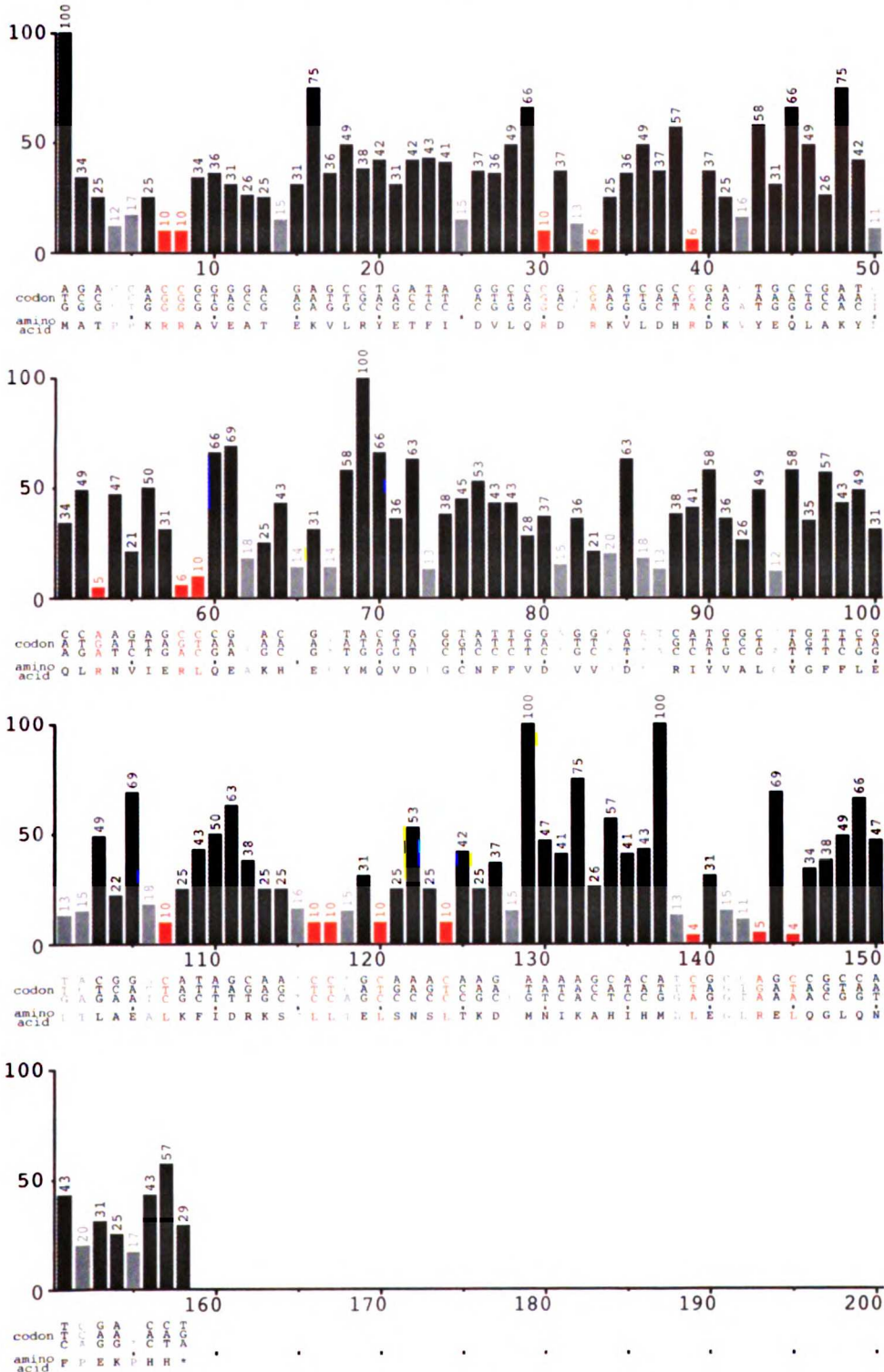
1  $\mu$ L of the first PCR was then used as template for the second PCR. This PCR consisted of 5  $\mu$ L 10x AccuPrime Pfx buffer, 2  $\mu$ L Gateway universal primer mix (10  $\mu$ M each forward and reverse), 1  $\mu$ L PCR1, 0.5  $\mu$ L AccuPrime Pfx, and 41.5  $\mu$ L water. The reaction was then cycled 1x at 95°C for 2 minutes; 25x at 95°C for 30 seconds, 55°C for 30 seconds, 68°C for 1 minute; 1x at 68°C for 2 minutes.

After the second round PCR, a band was observed corresponding to the expected product of 527 bp. This was gel-purified using the Qiagen gel extraction kit and then inserted by BP recombination into pDONR Zeo (2.5  $\mu$ L gel-purified PCR2, 1  $\mu$ L pDONR

Zeo, 2  $\mu$ L 5x BP reaction buffer, 2  $\mu$ L BP clonase, 2.5  $\mu$ L TE). After confirming recombination, two clones were sent for sequencing: one clone had a deletion, but the other had a perfect sequence - this clone was used for all subsequent ART-27 experiments.

**Figure 3-4. Codon usage of ART-27**

Numbers above the bars represent percent usage of the that codon in *E. coli*. Codons in red represent codons used 10% or less in *E. coli*; codons in grey represent 20% or less. Table generated by the Graphical Codon Usage Analyzer (<http://www.gcu.de>).



## Expression of ART-27

Codon optimized ART-27 was then cloned as a GST-fusion in the pET•GSTgw, pCDF•GSTgw, and pRSF•GSTgw vectors and expressed alone in BL21 Star cells at 15°C and 37°C (Figure 3-5). Unfortunately, expression levels were improved only marginally, if at all, versus the unoptimized version of the gene. Similarly to expression of ART-27 with native codons in the pDEST15 vector, expression was poor at 15°C and moderately well at 37°C but almost entirely insoluble.

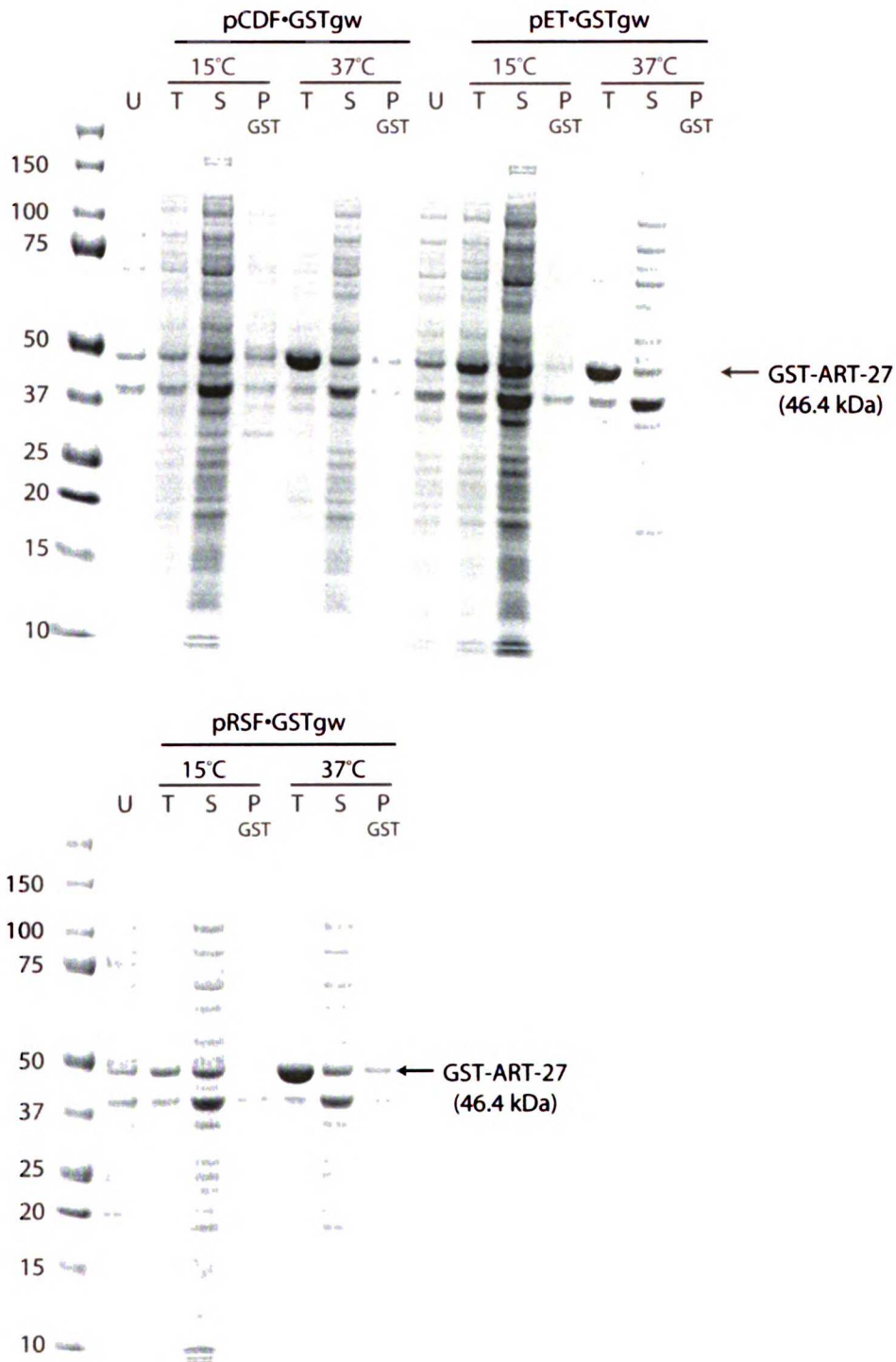
To improve solubility, ART-27 was cloned into pMAL(c2E)gw and pMAL(p2E)gw. MBP has been shown to be much more effective at improving solubility than GST when expressed as a fusion partner and this was indeed the case with ART-27 (Fox et al. 2003). When expressed alone, MBP-ART-27 expressed well and solubly at both 15°C and 37°C (Figure 3-6). Consistent with descriptions in the NEB pMAL system manual, expression was greater when expressed in the cytoplasm using pMAL(c2E)gw than when expressed in the periplasmic space using pMAL(p2E)gw.

With a soluble ART-27 construct in hand, it was hoped that coexpression with NTD would improve expression levels of NTD since one possible reason NTD expressed so poorly in previous experiments was because of the absence of a soluble, folding partner. ART-27-pMAL(c2E)gw was therefore coexpressed with NTD(142-448), which contains the entire ART-27 interacting region (Markus et al. 2002). Coexpression of MBP-ART27 with both 6xHis-NTD and GST-NTD however, failed to improve expression levels of NTD. Full length NTD(142-448) was not present - only proteolyzed fragments were recovered using glutathione or Talon beads (Figure 3-7). Moreover, MBP-ART27 failed to pulldown any of these NTD fragments. Expressing NTD and MBP-ART-27 separately, then combining the cell pellets prior to lysis produced similar results, as did coexpressing a 6xHis-tagged codon

optimized version of NTD(142-448) (discussed below) with untagged ART-27 on the same pRSFgw vector (Figure 3-11B). Similarly, 6xHis-ART-27 bicistronically coexpressed with smaller NTD fragments (142-336, 142-359, 142-383, 153-336) also produced negative results.

### Figure 3-5. Expression of codon optimized GST-ART-27

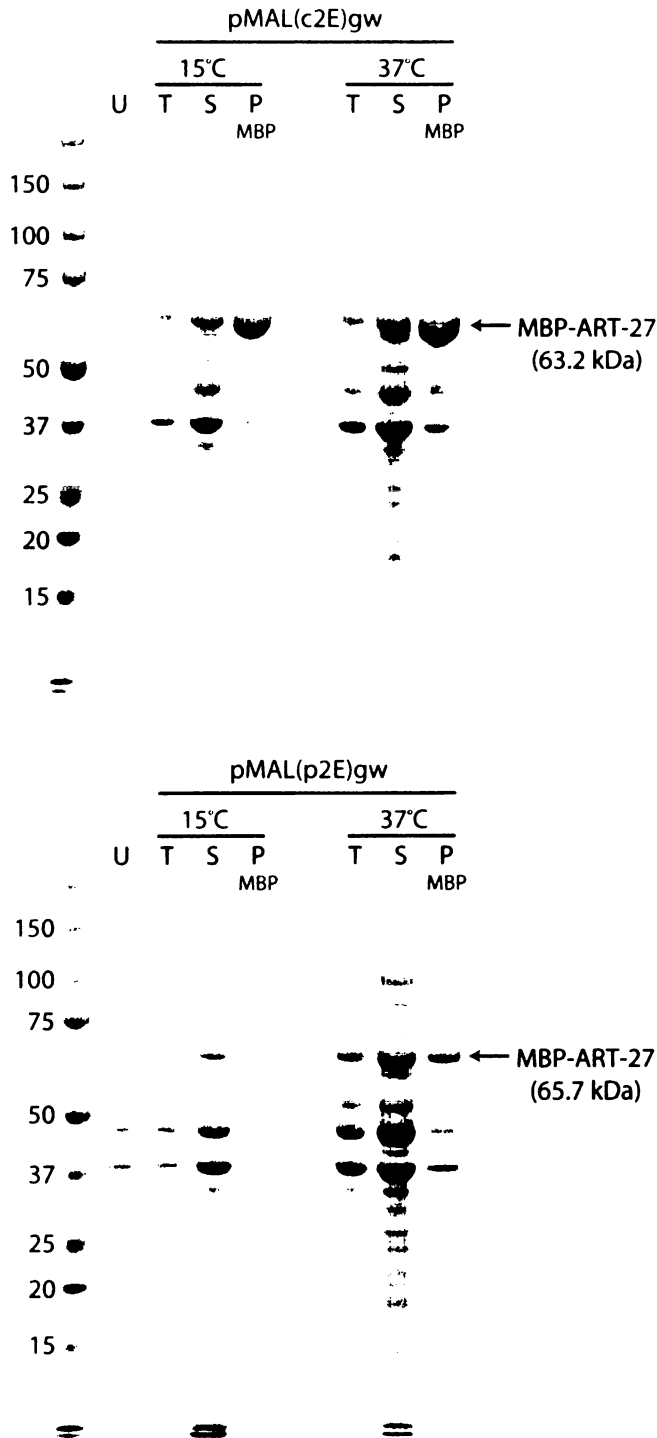
Vectors used for expression are indicated above the lanes. U = uninduced cell extract, T = induced cell extract, S = soluble fraction, P = elutions from glutathione beads.





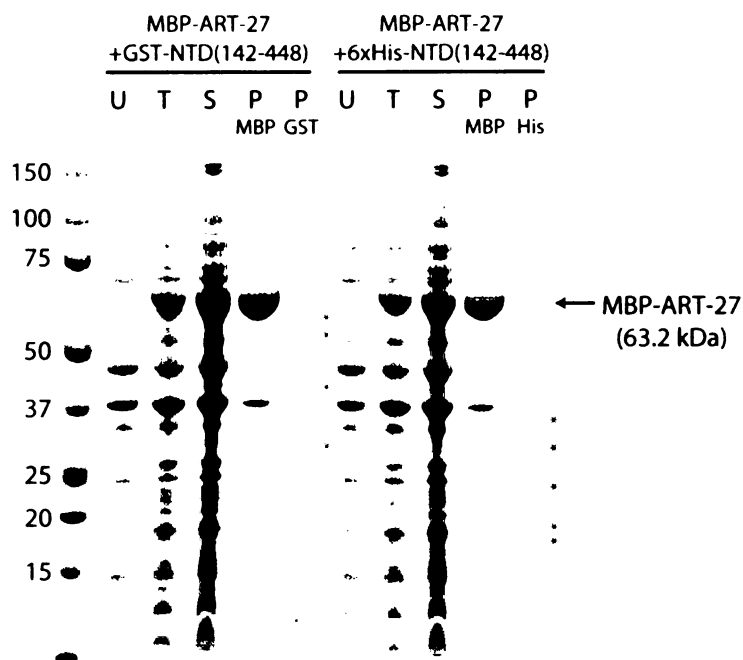
### Figure 3-6. Expression of codon optimized MBP-ART-27

Vectors used for expression are indicated above the lanes. U = uninduced cell extract, T = induced cell extract, S = soluble fraction, P = elutions from amylose beads. MBP alone is about 40 kDa.



**Figure 3-7. Coexpression of MBP-ART-27 with GST and 6xHis NTD(142-448)**

U = uninduced cell extract, T = induced cell extract, S = soluble fraction, P = elutions from amylose (MBP), Talon (His), or glutathione (GST) beads. GST-NTD(142-448) is 59.3 kDa. GST alone is 28 kDa. Bands with red asterisks (\*) represent possible GST-NTD(142-448) cleavage products. 6xHis-NTD(142-448) is 35.3 kDa. Bands with blue asterisks (\*) represent possible 6xHis-NTD(142-448) cleavage products.



From these experiments it was concluded that the interaction of ART-27 with NTD was either very weak or non-existent. However, because of low expression levels of NTD, it was difficult to conclude whether the lack of copurification was due to the true absence of an interaction or due to low concentrations of NTD which were not detectable via Coomassie stain. As a result, we set about purifying NTD and ART-27 separately to perform pulldown experiments. (Expression and purification of NTD is described below.)

### **Purification of MBP-ART-27**

MBP-ART-27 was expressed in BL21 Star cells in LB broth supplemented with 50 µg/mL carbenicillin and 0.2% glucose. Cultures were inoculated 1/200 with overnight cultures and grown at 37°C to an  $A_{600}$  of 0.5-0.6. Cells were then induced with 250 µM IPTG for 3 hours at 37°C or 15°C overnight. After harvest, cells were resuspended in lysis buffer containing 20 mM HEPES pH 7.5, 200 mM NaCl, 0.5% Tween-20, 1 mM DTT, and 1 mM EDTA. Approximately 10 mg of lysozyme was also added per liter culture. Cells were lysed by sonication on ice and clarified by centrifugation at 20,000 g for 30 minutes. Amylose resin (NEB) was incubated with lysate for 30 minutes, washed with lysis buffer, then transferred to a column for final washes with lysis buffer then with lysis buffer minus Tween-20 and EDTA (wash buffer). Protein was eluted using wash buffer plus 20 mM maltose. Final yield of protein was about 1.4 mg/L culture. To cleave off the MBP tag, TEV protease (Invitrogen) was added to MBP-ART-27 in the elution buffer at a ratio of 10 µL TEV to 500 mg protein and incubated at 4°C. Unfortunately, within 2 hours precipitate was visible (Figure 3-8). After an overnight incubation, the protein was spun down - nearly all ART-27 was present in the pellet as insoluble precipitate. Cleaved MBP remained in the supernatant.

In an attempt to crystallize ART-27 alone without cleavage of the MBP tag, MBP-ART-27 was further purified. The primary contaminant appeared to be cleaved MBP. Attempts to purify away this contaminant with a Superdex 200 size exclusion column (GE Healthsciences) were unsuccessful. Therefore, purification using a HiTrap Q anion exchange column (GE Healthcare) or a phenyl HP HiTrap hydrophobic interaction column (HIC) was attempted. For the phenyl HP HIC column, protein was dialyzed into 1 M ammonium sulfate, 20 mM HEPES pH 7.5, and 1 mM DTT. Protein was eluted with a 1-0 M ammonium sulfate gradient - MBP-ART-27 eluted at about 500 mM ammonium sulfate (Figure 3-9). For anion exchange, amylose column elutions were pooled and dialyzed into a solution consisting of 20 mM HEPES pH 7.5 and 1 mM DTT. Protein was eluted from the HiTrap Q using a 0-2 M NaCl gradient - MBP-ART-27 eluted at about 500 mM NaCl (Figure 3-9). The purity of elutions from the HiTrap Q and the phenyl HP HIC columns were similar with a slight improvement in purity versus a Superdex 200. Purified MBP-ART-27 was then concentrated to about 4 mg/mL and 12 mg/mL and screened using the Nextal prefilled MPD and PEG screens (Qiagen). Wells F5 and F6 of the Nextal PEG Screen, and G7, F8, H6, and H9 of the Nextal MPD screen produced promising leads with the more concentrated (12 mg/mL) MBP-ART-27.

### Figure 3-8. Cleavage of MBP-ART-27 with TEV protease

Lanes are as follows:

1. Fraction 1 of amylose column elution
2. Fraction 1 of amylose column elution concentrated 5x
3. Pooled fractions of amylose column elution
4. MBP-ART-27 + TEV - 1 hour incubation
5. MBP-ART-27 + TEV (suspension) - 2 hour incubation
6. MBP-ART-27 + TEV (supernatant) - 2 hour incubation
7. MBP-ART-27 + TEV (suspension) - overnight incubation
8. MBP-ART-27 + TEV (supernatant) - overnight incubation
9. MBP-ART-27 + TEV (pellet) - overnight incubation
10. MBP-ART-27, no TEV - overnight incubation

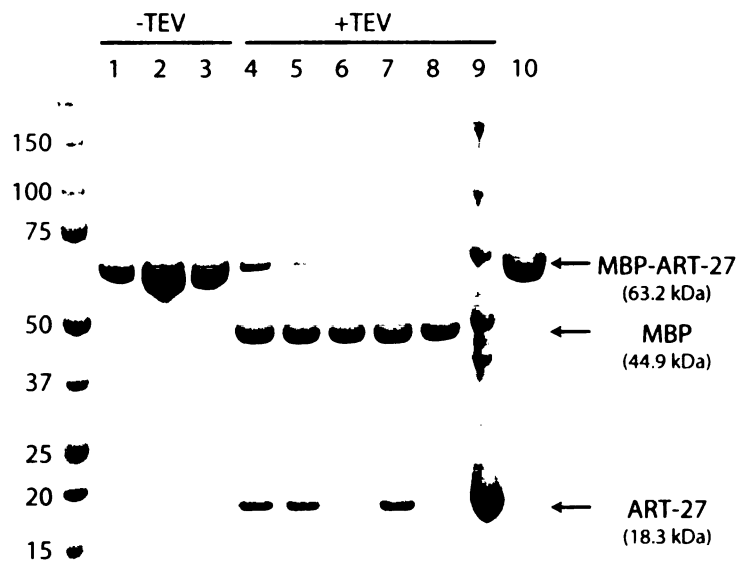
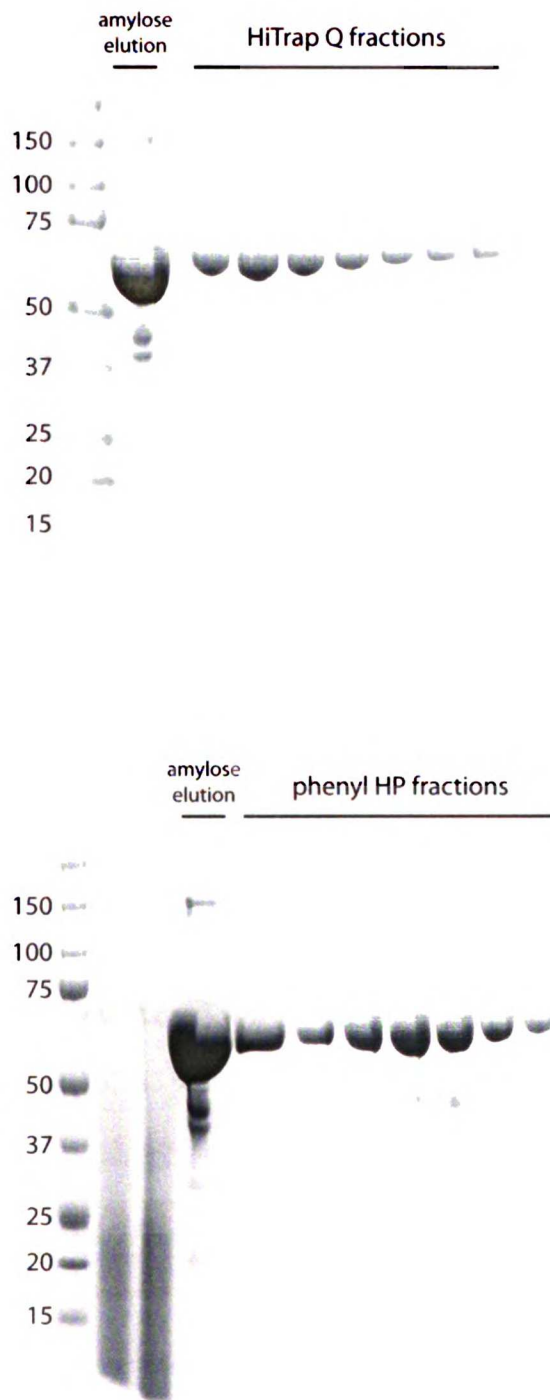


Figure 3-9. Phenyl HP and HiTrap Q purification of MBP-ART-27



### ART-27 / NTD(142-448) Pulldowns

Using purified components, pulldown experiments were performed with MBP-ART-27 and 6xHis-NTD(142-448). For pulldowns, MBP-ART-27 and 6xHis-NTD were added in an equimolar ratio to a 50  $\mu\text{L}$  slurry ( $\sim 25 \mu\text{L}$  bed volume) of Talon (BD Biosciences) and/or amylose beads [100  $\mu\text{L}$  of a 100  $\mu\text{g}/\text{mL}$  (2.9  $\mu\text{M}$ ) solution of 6xHis-NTD; 19.1  $\mu\text{L}$  of a 950  $\mu\text{g}/\text{mL}$  (15  $\mu\text{M}$ ) solution of MBP-ART-27]. NTD and ART-27 were incubated together at 4°C for 15 minutes or overnight and then incubated with Talon or amylose beads for 30 minutes. Beads were then washed several times with 20 mM HEPES pH 7.5, 200 mM NaCl, 0.5% Tween-20, plus 10 mM imidazole pH 7.5. Protein was eluted by the addition of 25  $\mu\text{L}$  of 1 M imidazole or 500 mM maltose to 25  $\mu\text{L}$  washed slurry. As shown in Figure 3-10, interaction of ART-27 with NTD was weak and improved marginally with overnight incubation. While it is possible that binding conditions were not optimal and varying the salt concentration and pH could improve interaction, it is likely that the weak interaction observed between the NTD and ART-27 is due to the absence of third partners which may be necessary to form a full binding interface. Indeed, analytical ultracentrifugation of HeLa cell extracts show that native ART-27 sediments as part of a multiprotein complex (Markus et al. 2002; Gstaiger et al. 2003). Other partner proteins in this complex may be required for full interaction of NTD with ART-27.

**Figure 3-10. MBP and 6xHis pulldowns of ART-27 and NTD(142-448)**

Pulldowns for lanes 1-4 were performed using amylose beads. Pulldowns for lanes 5-8 were performed with Talon beads. Gel was stained with Coomassie blue. 6xHis-NTD(142-448) runs larger on SDS-PAGE than its predicted molecular weight.

1/5. 6xHis-NTD(142-448) + MBP-ART-27 - 15 minute incubation

2/6. 6xHis-NTD(142-448) + MBP-ART-27 - overnight incubation

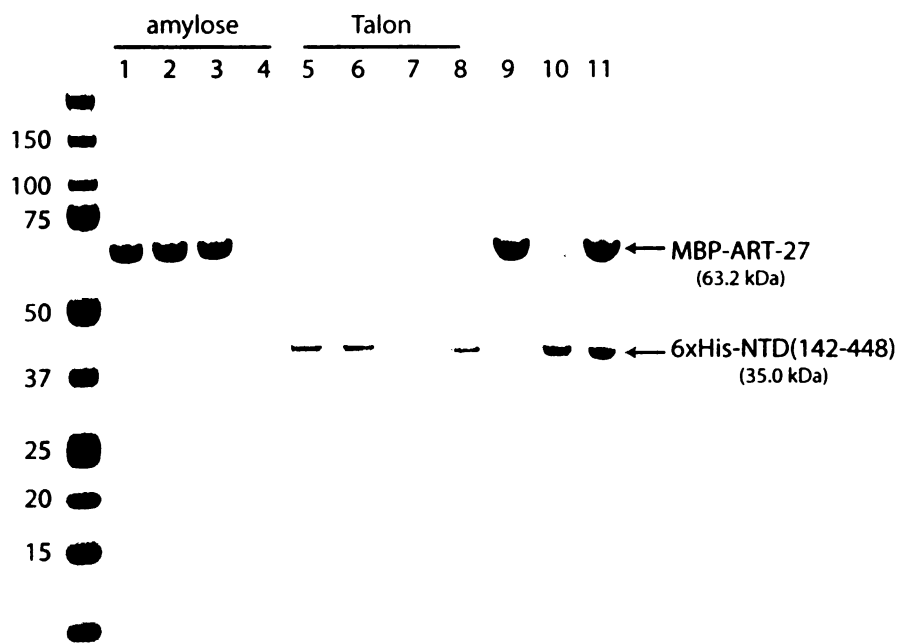
3/7. MBP-ART-27

4/8. 6xHis-NTD(142-448)

9. MBP-ART-27 - 7.5% input

10. 6xHis-NTD(142-448) - 7.5% input

11. 6xHis-NTD(142-448) + MBP-ART-27 - 7.5% input





## Expression and Purification of AR-NTD

Residues 142-498 of AR encompass the region previously mapped to contain AF-1 activity (McEwan 2004). Using pRSFgw, it was possible to express 6xHis-NTD(142-498) on its own. However, expression levels were low and full-length protein was truncated by proteolysis. Expression of a smaller fragment, NTD(142-448), which encodes AF-1 minus the poly-glycine repeat, failed to improve expression nor reduce levels of proteolysis (Figure 3-7). Coexpression with ART-27, RAP74, or with the C-terminal domains of GRIP-1, SRC-1, and RAC-3 did not alleviate these problems either. Coexpression with GRIP-1, SRC-1, and RAC-3, however, was inconclusive due to poor expression of these proteins.

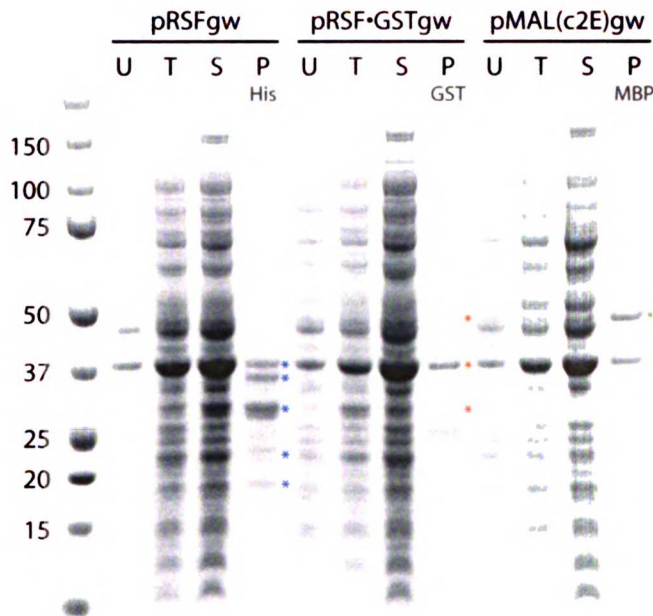
Like ART-27, examination of codon usage in the AR gene revealed the presence of many rare *E. coli* codons. We therefore expressed NTD in Rosetta 2 cells, which carry a plasmid encoding tRNA genes for the rare AGA, AGG, AUA, CUA, GGA, CCC, and CGG codons. Protein expression showed a moderate improvement over the standard BL21 Star cells (Figure 3-11). However there were a few rare codons in the NTD not present in the Rosetta 2 strain and it was hoped that elimination of these codons would further improve expression. As a result, we went ahead and had the entire AR gene synthesized with codons optimized for expression in *E. coli* (Genscript). It was hoped that a codon optimized AR gene would not only be useful for studies of the NTD, but also potentially for other parts of AR. In particular, we hoped that improved expression would facilitate structural studies of full-length AR. Unfortunately, gene synthesis of full-length AR turned out to be technically unfeasible because of poly-glutamine, poly-glycine, and poly-proline repeats. We therefore focused on the AF-1 region and synthesized in-house the AF-1 up until the poly-glycine repeat (residues 142-448). Protocols were similar to that used to construct codon-optimized ART-27. Oligonucleotides used to construct this NTD fragment are shown below.

AR448-1: **GAAaatctgtactttcagagcGGCCTGCCGCAGCAGCTGCCGG**  
 AR448-2: **GCGCTATCATCTTCATCCGGCGGGCGCCGGCAGCTGCTGCGGCAGG**  
 AR448-3: **CCGGATGAAGATGATAGCGGGCGCCGAGCACCCCTGAGCCTGCTG**  
 AR448-4: **TGCTCAGGCCCGAAAGGTCCGGGCCAGCAGGCTCAGGGTGTCTG**  
 AR448-5: **GACCTTTCCGGGCTGAGCAGCTGCAGCGGGATCTGAAAGATATtc**  
 AR448-6: **CTGCATGGTGTCTCGCTTCGCTCAGAATATCTTTCAGATCCGCGCTg**  
 AR448-7: **GCGAAGCGAGCACCATGCAGCTGCTGCAGCAACAGCAACAGGAAG**  
 AR448-8: **CCGCTGCTGCTGCCTTCGCTCACCGCTTCTGTTGCTGTTGCTGC**  
 AR448-9: **AGCGAAGGCAGCAGCAGCGGCCGTGCGCGTGAAGCGAGCGGCGCG**  
 AR448-10: **GATAATTATCTTTGCTGCTGGTCGGCGCGCCGCTCGCTTCACGCG**  
 AR448-11: **CAGCAGCAAAGATAATTATCTGGGCGGCACCAGCACCATTAGCGA**  
 AR448-12: **CGCTTTGCACAGTTCTTTCGCATTATCGCTAATGGTGTGGTGGC**  
 AR448-13: **CGAAAGAACTGTGCAAAGCGGTGAGCGTGAGCATGGGCCTGGGCG**  
 AR448-14: **GGGCTCAGATGTTCCAGCGCTTCCACGCCCAGGCCCATGCTCACG**  
 AR448-15: **GCGCTGGAACATCTGAGCCCGGGCGAACAGCTGCGTGGCGATTGC**  
 AR448-16: **GCACGCCCAGCAGCGGCGCATAACATGCAATCGCCACGCAGCTGTTc**  
 AR448-17: **TGCGCCGCTGCTGGGCGTGCCGCCGGCGGTGCGTCCGACCCCGTG**  
 AR448-18: **GCCTTTGCATTCCGCCAGCGGCGCGCACGGGGTCGGACGCACCCG**  
 AR448-19: **CGCTGGCGGAATGCAAAGGCAGCCTGCTGGATGATAGCGGGGCA**  
 AR448-20: **TATTCCGCGGTATCTTCGGTGTCTTTGCCCGCGCTATCATCCAGC**  
 AR448-21: **ACCGAAGATACCGCGGAATACAGCCCGTTTAAAGGTGGCTATACC**  
 AR448-22: **CCAGGCTTTCGCCTTCCAGGCCCTTGGTATAGCCACCTTTAAACG**  
 AR448-23: **CCTGGAAGGCGAAAGCCTGGGTTGTTCTGGTAGCGCCGCGGCAGG**  
 AR448-24: **CGGCAGTTCAGGGTGCCGCTGCTGCCTGCCGCGGCGCTACCAGA**  
 AR448-25: **GCGGCACCCTGGAAC TGCCGCTACGCTGTCTCTGTATAAAAAGCG**  
 AR448-26: **TACGCCCGGCTTCATCCAGCGCGCCGCTTTTATACAGAGACAGC**  
 AR448-27: **CTGGATGAAGCGGCGGTATCAGAGCCGTGATTATTATAATTTTc**  
 AR448-28: **GCGGTGGGCCCGCCAGCGCCAGCGGAAAATTATAATAATCACGGC**  
 AR448-29: **GGCGTGGCGGGCCACCGCCACCGCCACCGCCACCGCATCCGCA**  
 AR448-30: **CGGATTTTCCAGTTTAATACGCGCATGCGGATGCGGTGGCGGTGG**  
 AR448-31: **GTATTAAACTGGAATAATCCGCTGGATTATGGCTCTGCGTGGGCGGcag**  
 AR448-32: **CCATAACGGCACTGCGCTGCCGCTGCCGCCACGCAGAGCCATAAtc**  
 AR448-33: **GCAGCGCAGTGCCGTTATGGCGATCTGGCGAGCCTGCATGGCGCG**  
 AR448-34: **TGCCGCTACCCGGGCCCGCCGCGCCATGCAGGCTCGCCAG**  
 AR448-35: **GGCGGGCCCCGGTAGCGGCAGCCCCGAGCGGCGGCGGAGCAGCAG**  
 AR448-36: **TTCCGCGGTAAACAGGGTATGCCAGCTGCTGCTCGCCGCCGCGTc**  
 AR448-37: **ATACCCTGTTTACCGCGGAAGAAGCCAGCTGTATGGCCCCGTGCTa**  
 AR448-38: **gtacaagaagctgggtcTTAGCACGGGCCATACAGCTGG**

Optimized NTD(142-448) was expressed with pRSFgw, pRSF•GSTgw, and pMAL(c2E)gw in BL21 Star cells (Figure 3-12). Expression levels were similar to that observed in Rosetta 2 cells with significant proteolysis present. Previously, all expression had been done at 15°C to increase solubility. However, it was possible that expression overnight was contributing to high levels of proteolysis and that induction at higher temperature for a short time might increase the amounts of full-length protein. To test this idea, expression was performed at 37°C for 3 hours, with cells induced with high concentrations of IPTG (1 mM) at high cell density ( $A_{600}=1.0$ ). As shown in Figure 3-13, expression of 6xHis-NTD(142-448) at 37°C improved relative to expression at 15°C. And while proteolysis was still significant, levels of non-truncated 6xHis-NTD(142-448) now represented a significant fraction of eluted fragments. 6xHis-NTD(142-448) was found to run about 5 kDa larger than its calculated molecular weight of 35 kDa on SDS-PAGE. In an attempt to purify full-length 6xHis-NTD(142-448) from truncated fragments, Talon elutions were run through a Superdex 200 size exclusion column. Unfortunately, the resolution of the Superdex 200 was not great enough to effectively separate fragments of similar size and this method was abandoned for purification.

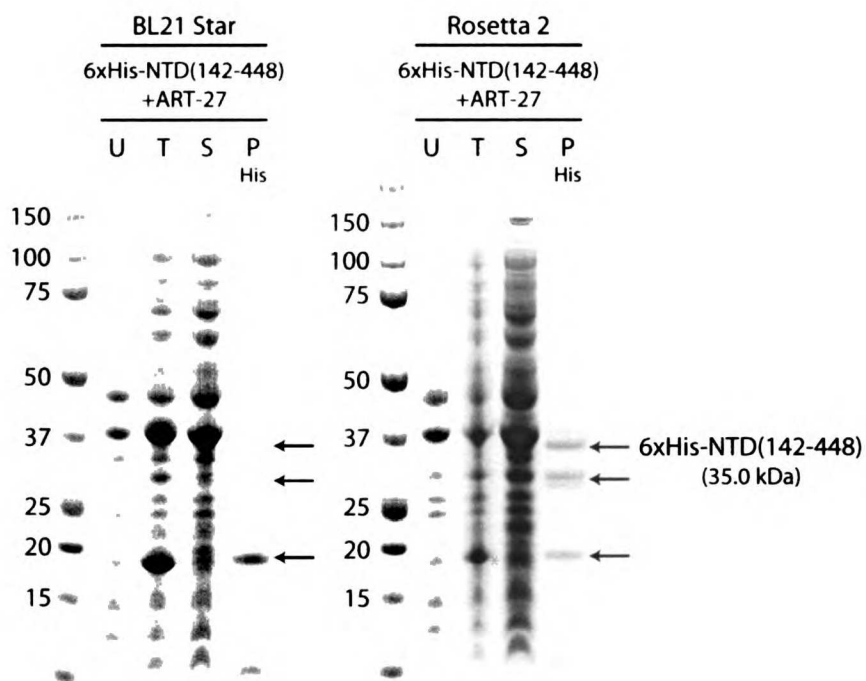
### Figure 3-11. Expression of NTD(142-448) in Rosetta 2 cells

A). Expression of NTD(142-448) in Rosetta 2 cells with an N-terminal 6xHis-tag (pRSFgw), GST-tag (pRSF•GSTgw), and MBP-tag (pMAL(c2E)gw). Eluted bands with blue, red, and green asterisks(\*) represent possible 6xHis-NTD(\*), GST-NTD(\*), and MBP-NTD( ) fragments, respectively. Full length 6xHis-NTD(142-448) is 35.0 kDa; full length GST-NTD(142-448) is 59.3 kDa; full length MBP-NTD(142-448) is 76.1 kDa. MBP and GST alone are 40 and 28 kDa, respectively.



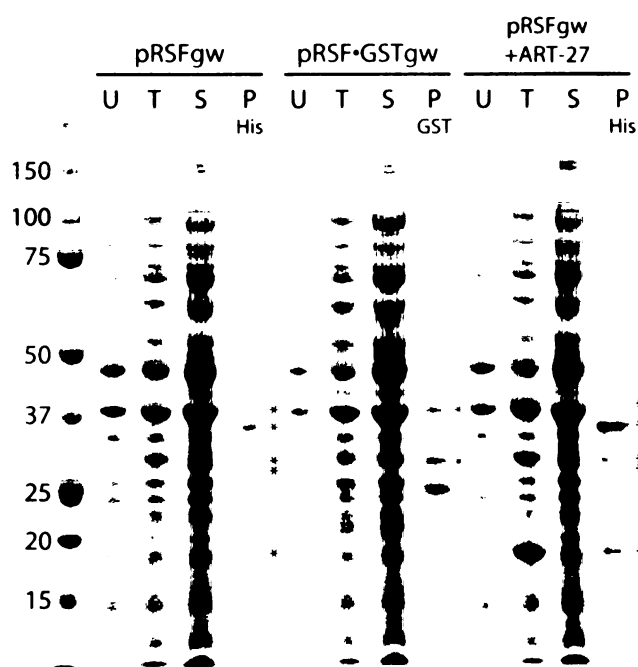
**Figure 3-11. Expression of NTD(142-448) in Rosetta 2 cells (cont.)**

B). Expression of 6xHis-NTD(142-448) in BL21 Star cells versus Rosetta 2 cells. NTD was coexpressed with ART-27 on a pRSFgw plasmid that contained untagged ART-27 in MCS2. Bands with arrows are NTD fragments. The bands at ~20 kDa and ~30 kDa were confirmed to be NTD with MALDI-TOF mass spectrometry. The overexpressed band in the cell extract with a magenta asterisk ( ) is ART-27 - this was also confirmed by MALDI-TOF mass spectrometry.



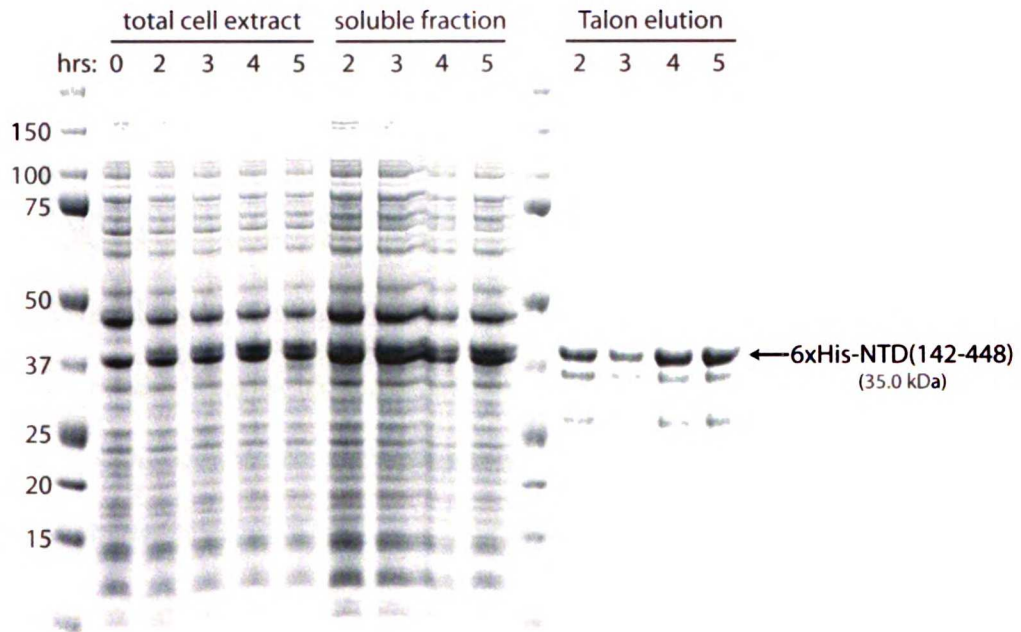
### Figure 3-12. Expression of codon optimized NTD(142-448)

Codon optimized NTD(142-448) was expressed in BL21 Star cells with a 6xHis-tag (pRSFgw) or a GST-tag (pRSF•GSTgw). 6xHis-NTD(142-448) was also coexpressed with ART-27 using a pRSFgw vector containing untagged ART-27 in MCS2. Eluted bands with blue and red asterisks represent possible 6xHis-NTD(·) and GST-NTD(·) fragments, respectively. 6xHis-NTD(142-448) runs larger than its predicted molecular weight on SDS-PAGE. The bands at ~30 kDa and ~35 kDa were confirmed to be NTD with MALDI-TOF mass spectrometry. The overexpressed band in the cell extract with a magenta asterisk ( ) is ART-27 - this was also confirmed by MALDI-TOF mass spectrometry.



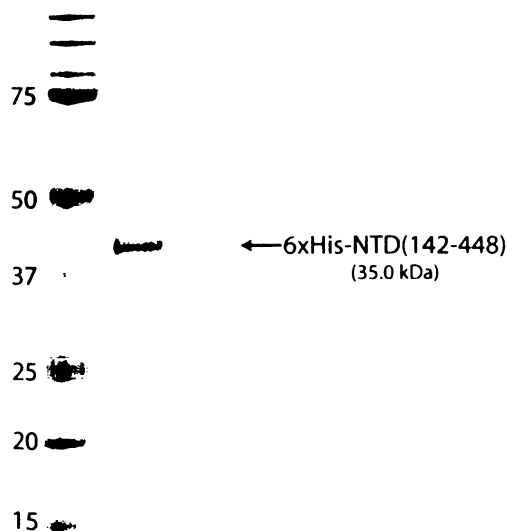
**Figure 3-13. Expression of codon optimized 6xHis-NTD(142-448) at 37°C**

Time course of expression of 6xHis-NTD(142-448) (pRSFgw) at 37°C is shown. The number of hours cells were induced is indicated. (Talon beads for the 3 hour time-point were spilled, accounting for less eluted protein.)



## Purification of NTD(142-448) Under Denaturing Conditions

To circumvent proteolysis problems, an alternative strategy was taken. Since the NTD is largely disordered in the absence of a binding partner, attempts were made to purify 6xHis-NTD(142-448) under denaturing conditions (Reid et al. 2002; Kumar et al. 2004). Protein was expressed as described above at 37°C, and lysed with 6 M urea, 50 mM HEPES pH 7.0, 200 mM NaCl, 10 mM imidazole, 5 mM  $\beta$ -mercaptoethanol. Cells were sonicated to reduce viscosity, then clarified by centrifugation at 25,000 g. Lysate was batch bound to Talon beads and incubated at 4°C for 15 minutes. Beads were washed, then transferred to a column for final washes with lysis buffer, then lysis buffer with 3 M urea, then finally lysis buffer with no urea to allow elution under native conditions. Protein was eluted with 150 mM imidazole. In another prep, protein was eluted under denaturing conditions with similar results. As shown in Figure 3-14, NTD(142-448) purified under denaturing conditions was largely free of proteolysis.



**Figure 3-14. Talon purification of 6xHis-NTD(142-448) under denaturing conditions**  
Lanes represent different fractions.



## Chapter 4

# LRH-1 and NCoA-62

Collaborators on this chapter:

Subhagya A. Wadekar, Peter K. Hwang, Holly A. Ingraham

## LRH-1 and NCoA-62

Liver receptor homologue 1 (LRH-1, NR5A2) is an orphan nuclear receptor belonging to the NR5A family of nuclear receptors whose other members include the closely related steroidogenic factor 1 (SF-1) and *Fushi tarazu* factor 1 (Ftz-F1). Members of this family are constitutively active and bind DNA as monomers. In addition to its role in regulating genes involved in bile acid and cholesterol homeostasis, LRH-1 is also critical in development and proper ovary function (Fayard et al. 2004). For many years the identity of ligands for the NR5A family were elusive. However, recent work has established that phosphatidyl inositols are the ligands for SF-1 and LRH-1 (Krylova et al. 2005). Interestingly, mouse LRH-1 has diverged from human LRH-1 in such a way that it does not bind nor require ligand for full activity, making it a true orphan (Sablin et al. 2003).

NCoA-62 (SKIP) is a highly conserved mRNA splicing factor that was independently isolated as an interacting partner for the Ski onco-protein and then later as a novel, non-p160 coactivator for the vitamin D receptor (VDR) (Baudino et al. 1998; Dahl et al. 1998). It interacts with a long list of proteins involved in mRNA processing and transcription, and is therefore thought to serve as a link between the two processes (Folk et al. 2004). NCoA-62 also participates in the TGF- $\beta$  and Notch signaling pathways implicating it as a regulator of cellular proliferation and differentiation. NCoA-62 is a 536 residue long protein and can be roughly divided into three domains: an N-terminal domain, a C-terminal domain, and a central SNW domain that contains a signature SNWK motif characterizing the members of this protein family - most protein interactions have been mapped to, roughly, residues 170-340 of this central SNW domain (Folk et al. 2004).

## Mapping LRH-1 and NCoA-62 Interactions

Recently, through work in the Ingraham lab (Subhagya Wadekar), NCoA-62 was identified as an LRH-1 coactivator through yeast 2-hybrid screens. Using full length LRH-1 as bait, a clone containing residues 3-322 of NCoA-62 was identified. Initial pulldown experiments indicated that the LRH-1 LBD was sufficient for interaction with NCoA-62 (Subhagya Wadekar). Initially, we undertook low resolution domain mapping to narrow down the LRH-1 interaction region in NCoA-62. In particular, we focused on the N-terminal and SNW domains since the clone discovered from the yeast 2-hybrid screen contained these regions. Using the Gateway system, four different NCoA-62 constructs were made: full length (1-536), N-terminal domain + SNW domain (N-SNW) (1-333), C-terminal domain + SNW domain (C-SNW)(176-536), and SNW domain alone (SNW)(176-333).

Primers used to clone these domains are listed below:

NCoA_1f:	<b>aatctgtactttcagagcATGGCGCTCACCAGCTTTTTTAC</b>
NCoA_176f:	<b>aatctgtactttcagagcTATATCCGATACACACCATCTC</b>
NCoA_333r:	<b>gtacaagaaagctgggtcTTACTCCCTGGCTTCTGGGCCATTTTC</b>
NCoA_536r:	<b>gtacaagaaagctgggtcTTATTCCTTCCTCCTCTTCTTGCCCTTC</b>

Similarly, human and mouse LRH-1 LBDs (human - 294-541) (mouse - 313-560) were cloned using the primers below:

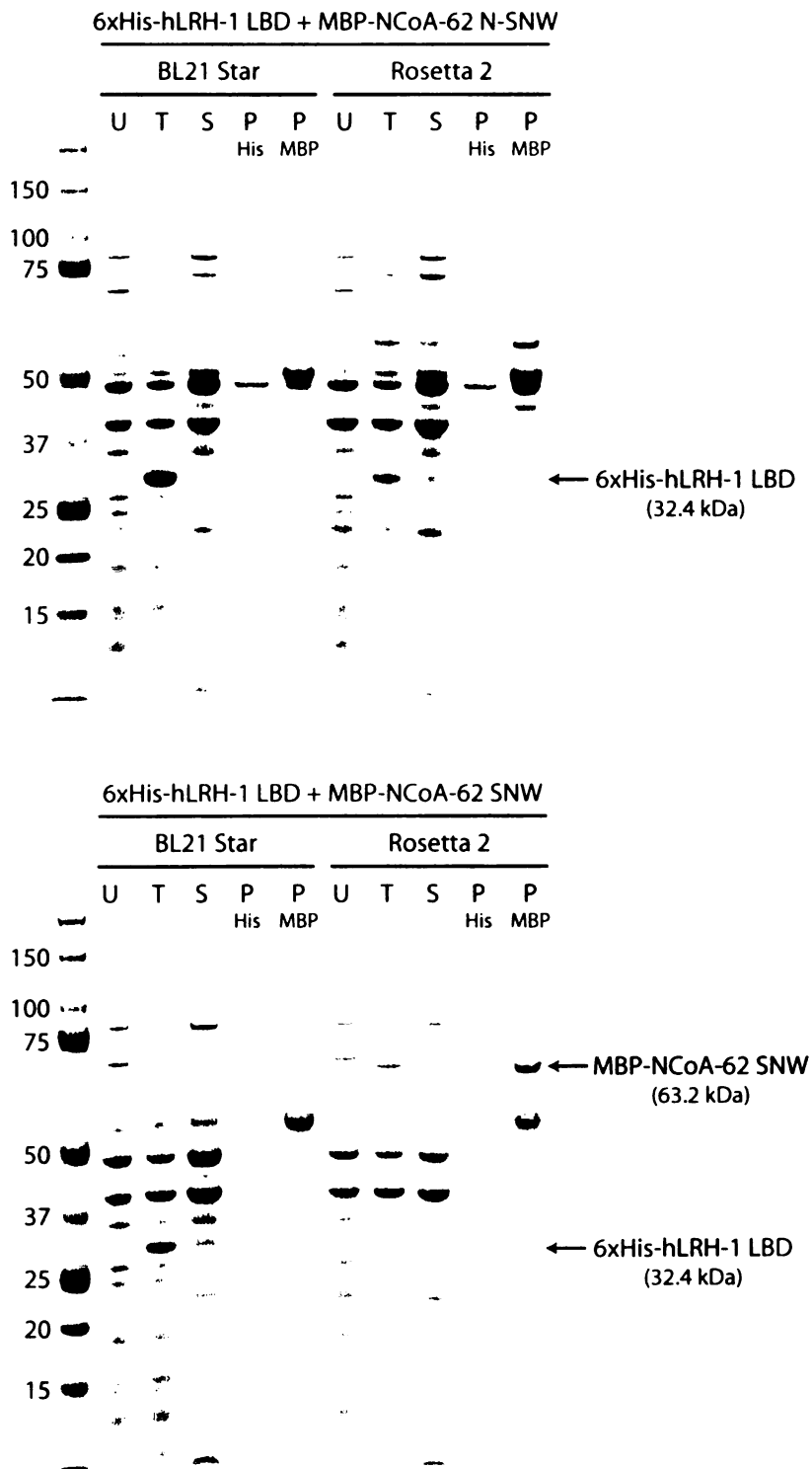
hLRH_294f:	<b>aatctgtactttcagagcCAGACGAGCTCTCCAGCAAGC</b>
hLRH_541r:	<b>gtacaagaaagctgggtcTTATGCTCTTTTGGCATGCAAC</b>
mLRH_f:	<b>aatctgtactttcagagcCAGACAAACTCCCCGGCCAGCATC</b>
mLRH_r:	<b>gtacaagaaagctgggtcTTAGGCTCTTTTGGCATGCAGCATC</b>

LRH LBD was recombined into pRSFgw and expressed with an N-terminal 6xHis-tag. LBD expressed well, but was insoluble unless expressed at 15°C. NCoA-62 domains, on the other hand, were cloned into the pCDF•GSTgw vector for expression as N-terminal GST-fusions. Expression however, of all domains was poor, both when expressed alone and

when coexpressed with LRH hinge-LBD or LBD. We therefore switched affinity tags and attempted expression of NCoA-62 domains as MBP-fusions using the pMAL(c2E)gw vector. As shown in Figure 4-1, NCoA-62 N-SNW and SNW expressed solubly as MBP-fusions. However, expressed fusion proteins were greatly truncated by proteolysis.

### Figure 4-I. Expression of MBP-NCoA-62 domains

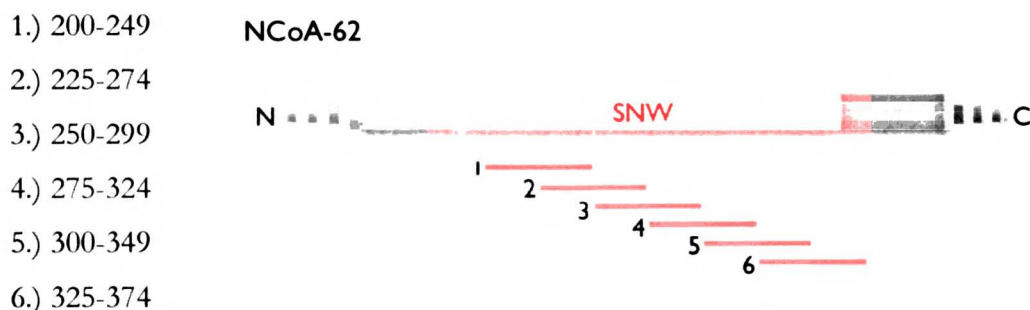
Expression of MBP-NCoA-62 N-SNW and SNW in both BL21 Star and Rosetta 2 cells are shown. U = uninduced cell extract, T = induced cell extract, S = soluble fraction, P = elutions from Talon (His) or amylose (MBP) beads.



## NCoA-62 Fragments

To circumvent the problem of poor expression and proteolysis, we reduced the number residues in our NCoA-62 constructs so that expression as a fusion partner with MBP would likely produce soluble protein that would be protected from proteolysis by MBP. We focused on a 69 residue region in the SNW domain (residues 274-342) previously mapped as a VDR interaction domain (Zhang et al. 2001). This region is consistently predicted to be helical in secondary structure prediction programs. Starting at residue 200 and ending at 374, six 50 residue NCoA-62 fragments overlapping by 25 residues were generated (see schematic below).

Fragments were as follows:



PCR and recombinations were performed according to protocols in Chapter 3.

However, because fragments were so small, second round PCR products were not purified and instead, used directly in BP recombinations (5  $\mu$ L per 10  $\mu$ L BP reaction). Primers used to construct the fragments are shown below.

NCoA_200f:	<b>aatctgtactttcagagcGTAGAAATGCAGAAAGATCCAATG</b>
NCoA_225f:	<b>aatctgtactttcagagcCCTCCTGCGCCTGTCATGCATTC</b>
NCoA_250f:	<b>aatctgtactttcagagcTGTATTTCTAACTGGAAAAATG</b>
NCoA_275f:	<b>aatctgtactttcagagcCTACAGACAGTACACATAAATG</b>
NCoA_300f:	<b>aatctgtactttcagagcGAAGCTGTGGAAATGCGTGCCCAAG</b>

NCoA\_325f:        **aatctgtacttttcagagcAGAGAAATGGCCCAGAAAGCCAG**  
 NCoA\_350f:        **aatctgtacttttcagagcCGTGAGAGGGATGAAATCCGGCATG**  
 NCoA\_249r:        **gtacaagaaagctgggtcttaAGGAGGAATCTTCCACTCTTG**  
 NCoA\_274r:        **gtacaagaaagctgggtcttaTCCTCTTCCATCAGCAGCCAG**  
 NCoA\_299r:        **gtacaagaaagctgggtcttaACGAGCCTTCCGATCAGCAATG**  
 NCoA\_324r:        **gtacaagaaagctgggtcttaAAGTTTCTCTTCATGTTTTTC**  
 NCoA\_349r:        **gtacaagaaagctgggtcttaTGCCTCCCCATCTCTTTTTTC**  
 NCoA\_374r:        **gtacaagaaagctgggtcttaAGGAGCTGCCCTGGAAAGATTC**

NCoA-62 fragments were recombined into pMAL(c2E)gw and transformed in BL21 Star for expression or coexpression with 6xHis-LRH-1 LBD. Protein expression was induced with 1 mM IPTG and carried out at 15°C for 20 hours. After harvest, cell pellets were resuspended in Ni-NTA binding buffer containing 20 mM Tris pH 8.0, 300 mM NaCl, 50 mM imidazole pH 8.0, 10% glycerol, 15 mM  $\beta$ -mercaptoethanol, and 0.1 % Triton X-100. Calbiochem protease inhibitor cocktail set V (Calbiochem) was also added. Cells were lysed by sonication and clarified by centrifugation at 20,000 g. Alternatively, Bugbuster lysis reagent (Novagen) plus protease inhibitor was used for lysis instead of sonication. Cell lysate was batch bound to amylose or Ni-NTA beads by incubation at 4°C for 30 min. Protein was washed with Ni-NTA binding buffer and eluted using Ni-NTA buffer plus 300 mM imidazole, for 6xHis-LRH LBD, or 250 mM maltose, for MBP-NCoA-62.

As shown in Figure 4-2, all fragments expressed well and solubly as MBP-fusions, with fragment 5 consistently expressing better than the rest. The reason for this is unknown, but one possibility is that this fragment possesses some additional structure that protects it from proteolysis or that it interacts particularly well with its MBP fusion partner.

Coexpression of these fragments with LRH-1 LBD show that fragments 3 (residues 250-299) and 4 (residues 275-324) copurify with LRH-1. Experiments were performed with both human and mouse LRH-1 and showed no differences between species. Since fragments 2

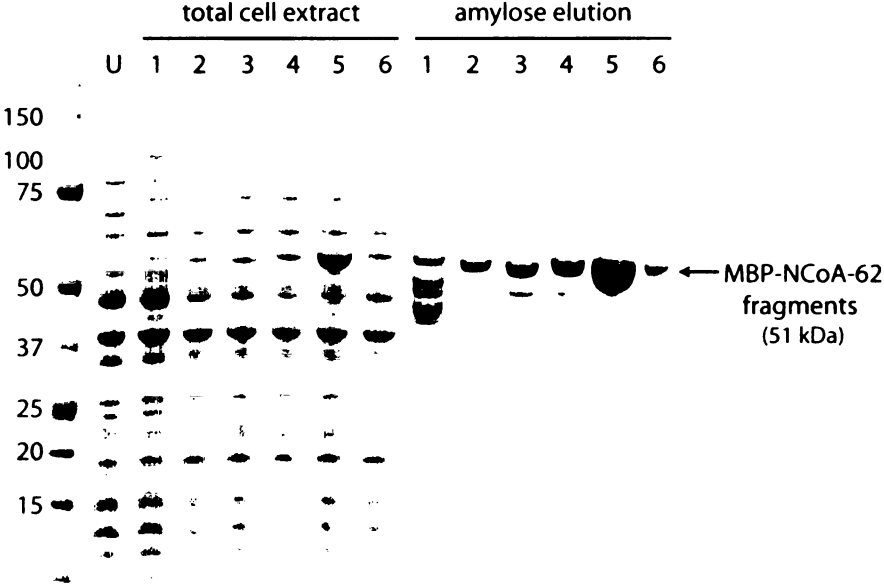
and 5 do not interact with LRH-1, these results suggest that the LRH-1 interaction site on NCoA-62 lies at the 25 residue intersection between fragments 3 and 4 (residues 275-299).

Only full length LBD interacts with these fragments. Initially there were problems with proteolysis during the purification of LRH-1. This led to a C-terminally truncated LRH-1 which coelutes with full length LRH-1 through the intact N-terminal 6xHis-tag (Figure 4-2b). This problem was eventually solved by switching to Ni-NTA resin (Qiagen) from Talon resin, and increasing the imidazole concentration during washes from 10 mM to 50 mM. This was only made possible by the stronger binding of Ni-NTA versus Talon. This not only led to a much purer elution that reduced a major contaminant at ~48 kDa, but more importantly, washed away a protease which presumably was copurifying with LRH-1 under less stringent wash conditions. Since C-terminally truncated LRH-1 does not interact with NCoA-62, this suggests that helix 12 facilitates the interaction, and that therefore, the LRH-1 AF-2 surface mediates the interaction with NCoA-62.



**Figure 4-2. Expression of MBP-NCoA-62 fragments**

Expression of MBP-NCoA-62 fragments 1-6 using the pMAL(c2E)gw vector are shown.

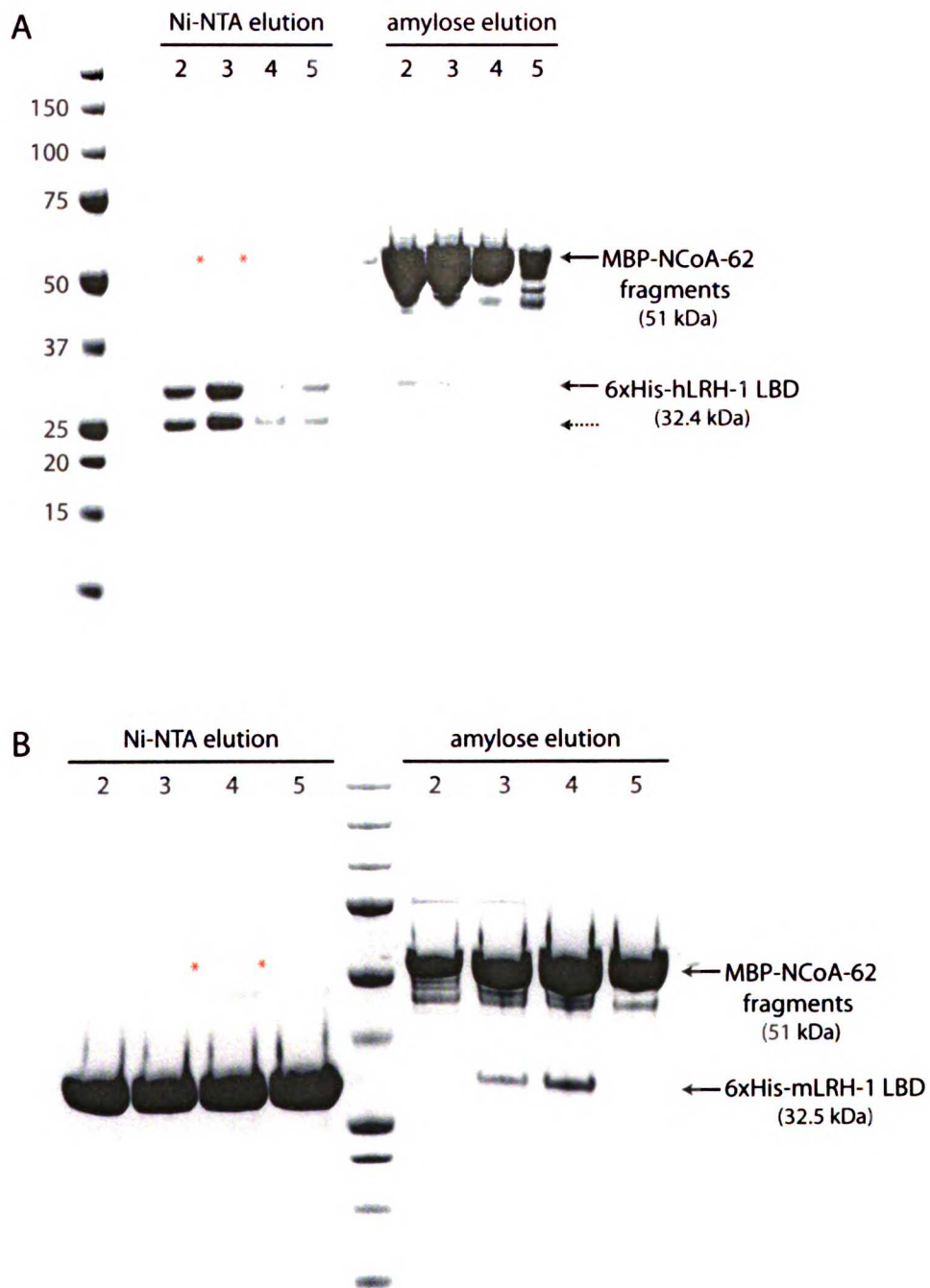


### Figure 4-3. Coexpression of LRH-1 with NCoA-62 fragments

Bands with red asterisks (\*) are NCoA-62 fragments copurified with LRH-1 LBD.

A). Coexpression of human LRH-1 LBD with NCoA-62 fragments 2-5. The dotted arrow indicates truncated hLRH-1 LBD.

B). Coexpression of mouse LRH-1 LBD with NCoA-62 fragments 2-5.

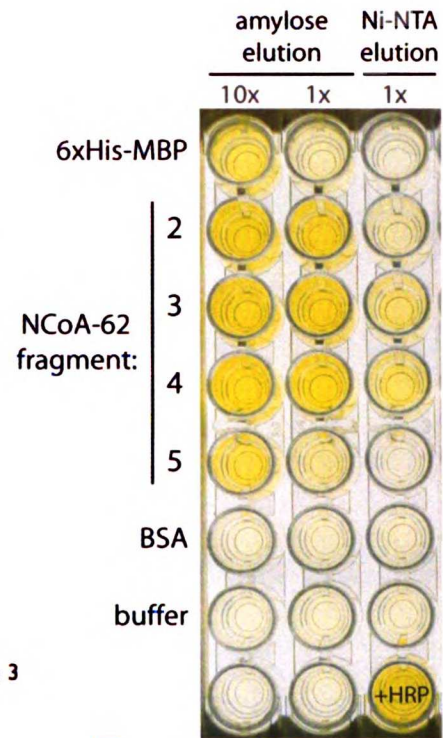


## ELISA Interaction Assay

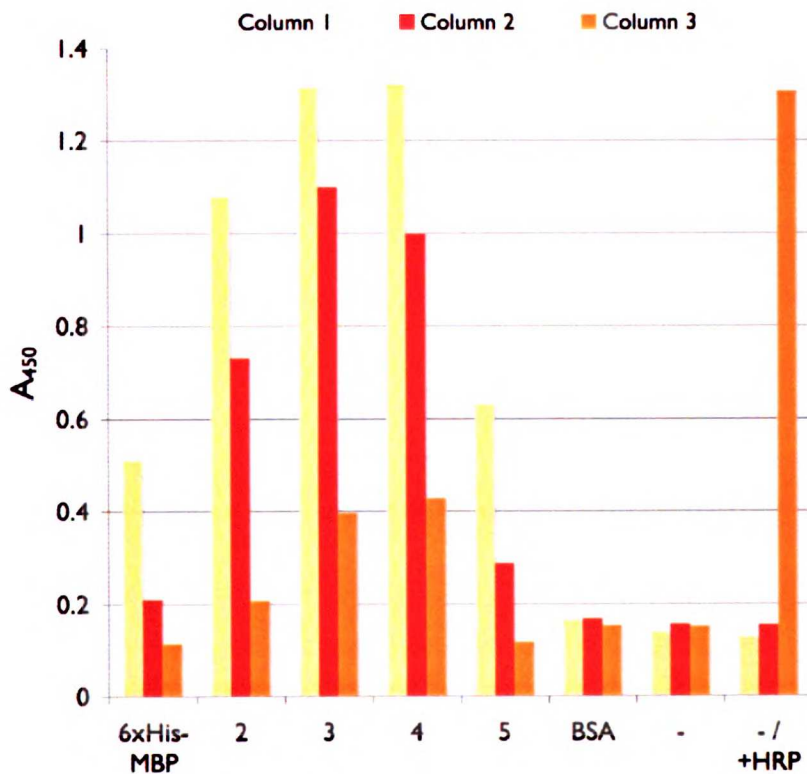
Copurified LRH-1 and NCoA-62 fragments were used to pilot the ELISA based protein interaction assay to be used for high-throughput mapping. Protein eluted from amylose and Ni-NTA pulldowns (Figure 4-2) were diluted 1/20 (total volume: 200  $\mu$ L) with Ni-NTA binding buffer and incubated in wells on a Ni-NTA coated plate (Qiagen) for 1 hour at room temperature. A 6xHis-tagged MBP was used as a positive control. Wells were washed 4x with TBST, then incubated with monoclonal anti-MBP antibody (NEB) diluted 1/10,000. After a 1 hour incubation at room temperature, wells were washed 4x with TBST and then incubated with anti-mouse HRP conjugated secondary antibody (Santa Cruz Biotechnology) diluted 1/10,000 for 30 minutes. Following incubation with secondary antibody, wells were washed 4x, then developed using the 1-Step Turbo TMB-ELISA reagent (Pierce). Color development was stopped using a 1M sulfuric acid solution and absorbances read at 450 nm. As shown in Figure 4-3, absorbances were largely consistent with pulldown assays with NCoA-62 fragments 3 and 4 giving the highest signals.

**Figure 4-4. Interaction of NCoA-62 fragments with mLRH-1 shown by ELISA**  
 LRH-1 and NCoA-62 affinity purified by amylose and Ni-NTA (Figure 4-3B) were used to pilot an ELISA based protein interaction assay.

Right). Wells in column 1 with mLRH-1 and NCoA-62 fragments have 10x more protein than similar wells in column 2. Rows with 6xHis-MBP and BSA were serially diluted 1/10 from column 1 to column 3. The bottom right well (H3) contains developer solution plus ~0.2  $\mu$ L of anti-mouse HRP conjugate.



Below). Absorbances read from the plate on the right.



## Future Directions

A peptide corresponding to residues 275-299 of NCoA-62 will be synthesized and cocrystallized with LRH-1 LBD. Because of the potential for heterogeneity caused by binding of different phospholipids to human LRH-1, mouse LRH-1 will be used for crystallization. If LRH-1 does not cocrystallize with this 25 residue fragment, fine mapping of this fragment will be performed to isolate hot spot residues and reduce the number of residues to increase the likelihood of crystallization. Concomitantly, Biacore SPR will be performed to measure the binding affinity of this interaction. In the case where LRH-1 does not crystallize, fine mapping will also be performed on a Biacore instrument, using either MBP-fused or biotinylated versions of the fragment.

ELISA interaction assays will also be further optimized. They will be piloted with crude lysate instead of affinity purified protein to facilitate future high-throughput mapping experiments. If necessary, fine mapping of NCoA-62 275-299 will also be performed with the ELISA, in parallel with SPR studies.

## Chapter 5

# The Urokinase Plasminogen Activator Receptor

Collaborator on this chapter:

Marc Shuman

## INTRODUCTION

Clinically, overexpression of the urokinase plasminogen activator (uPA) and its receptor, uPAR, are highly correlated with poor prognosis in a variety of cancer types (Schmitt et al. 1997). Their involvement in a number of cellular processes involved in tumor growth and metastasis, have made them attractive as potential therapeutic targets in cancer (Andreasen et al. 2000). Indeed, antagonizing the uPA-uPAR interaction with inactive forms of uPA have been shown to reduce the incidence of tumor metastasis (Mazar 2001).

uPA is a multi-domain protein consisting of an N-terminal epidermal growth factor like (EGF) domain followed by a kringle domain and a catalytic trypsin-like serine protease domain [The EGF domain and the kringle domain together constitute the amino-terminal fragment (ATF)]. Its receptor, uPAR, is a GPI anchored membrane protein consisting of three homologous domains (D1, D2, D3). Binding of uPA to uPAR is occurs through the N-terminal EGF domain of uPA. Together, uPA and uPAR constitute the core of the plasminogen activation system. Conversion of plasminogen to plasmin by uPA initiates a fibrinolytic pathway resulting in the degradation of extracellular matrix components in the pericellular space. In migrating cells, binding of uPA to uPAR is thought to localize uPA at the cell surface, facilitating the activation of plasminogen and amplifying the proteolytic effect at sites of extracellular matrix adhesion.

Independent of their role in plasminogen activation, it has also become apparent that uPA and uPAR are involved in signaling events affecting a number of cellular processes including growth, migration, and adhesion. Src-family tyrosine kinases, heterotrimeric G-proteins, and MAPK's have all been shown to be activated by uPA binding (Ossowski and Aguirre-Ghiso 2000; Preissner et al. 2000). The finding that uPA and uPAR can activate intracellular signaling pathways predicted the existence of transmembrane adaptor proteins

that can signal across the membrane. In the last few years, evidence has accumulated that uPAR associates with  $\beta 1$ ,  $\beta 2$ , and  $\beta 3$  integrins in a functionally significant way (Ossowski and Aguirre-Ghiso 2000; Preissner et al. 2000). This has led to the proposal that binding of uPA to uPAR triggers conformational changes in uPAR leading to its association with integrins and the induction of integrin mediated signaling pathways affecting cell migration, growth, and adhesion.

However, the observation that pertussis toxin inhibits uPA and uPAR induced chemotaxis, also suggested the involvement of G-protein coupled receptors (Fazioli et al. 1997). Recently, one such receptor has been identified. FPRL1/ Lipoxin A4R is a chemotactic receptor which binds a variety of host derived peptides as well as lipoxin A4. Lipoxins are lipid mediators, derived from lipoxygenase metabolites of arachidonic acid, that form during cell-cell interactions and down regulate the actions of the inflammatory response. uPAR, in addition to its GPI-anchored form, is also found as a soluble, cleaved form missing domain 1 which *in vivo* is thought to be generated by uPA (Hoyer-Hansen et al. 1992). Residues 88-92 of uPAR, which are exposed upon cleavage of domain 1, have been shown to be necessary and sufficient for chemotaxis (Fazioli et al. 1997). This stretch of residues was found to activate FPRL1/ Lipoxin A4R (Resnati et al. 2002; de Paulis et al. 2004). However, observations that catalytically inactive forms of uPA also induce chemotaxis (Gudewicz and Gilboa. 1987; Resnati, et al. 1996) suggest that cleavage of uPAR is not necessary for uPA induced chemotaxis and that uPA acts by inducing a conformational change in uPAR (Gudewicz and Gilboa 1987; Resnati et al. 1996; de Paulis et al. 2004).

The structure of uPA in complex with uPAR will reveal the nature of the interaction between the two proteins in atomic detail and provide a structural basis for the future design



of therapeutics which antagonize their interaction. Moreover, the structure should show how binding of uPA modifies its receptor, providing insights into the mechanism of uPA-uPAR mediated cell signaling.

## RESULTS

Previously, through work done by Jennifer Garrison, Marion Conn, Manish Butte, Cara Marks, and Marc Shuman, uPAR and full-length uPA were expressed, purified, and cocrystallized. The catalytic serine of uPA was mutated to prevent both self-cleavage and cleavage of uPAR. Nineteen data sets were collected, including many heavy atom derivatives. However, phasing the structure with phased molecular replacement using the uPA catalytic domain as a model remained elusive. When I inherited this project, the strategy was to grow more crystals and collect additional heavy atom data for MAD phasing. However, expression and purification of protein was cumbersome and unforeseen difficulties presented themselves resulting in low yields of protein. Moreover, crystals were not reproducible. As a result, I moved on to a different project and uPA-uPAR was later passed on to a post-doc in the lab, Dinlaka Sriprapundh.

However, a year ago the structure uPAR in complex with an antagonist peptide was solved and finally, earlier this year, the structure of uPAR in complex with the ATF of uPA was solved as well (Llinas et al. 2005; Huai et al. 2006). While it was disappointing to have been beaten to the structure of the uPA-uPAR complex by competing groups after so many years of effort, the newly solved structure gave us an opportunity to learn why our uPA-uPAR data sets were so unsolvable. Using the uPAR2 data set and the recently solved uPA ATF-uPAR complex (PDB ID: 2FD6) as a model, molecular replacement searches using CNS were performed (Brunger et al. 1998). Searches were straightforward. As shown below,

a cross-rotation search yielded a top solution with a correlation coefficient (RF-function) clearly above the next best solutions.

index	theta1	theta2	theta3	RF-function
1	209.563	17.372	154.291	0.0757
2	126.549	45.783	66.549	0.0563
3	184.740	28.421	166.968	0.0563
4	124.507	48.953	79.507	0.0551
5	101.990	83.678	84.217	0.0547

Similarly, translation searches also yielded solutions with clear top peaks. Previously, the space group had been identified as either  $P4_12_12$  or  $P4_32_12$ , which can not be distinguished from systematic absences (Manish Butte). Therefore translation searches were performed in both  $P4_12_12$  and  $P4_32_12$  (below). Translation searches in  $P4_32_12$  yielded the peak with the highest correlation coefficient (monitor).

**$P4_12_12$ :**

	theta1	theta2	theta3	transX	transY	transZ	monitor	packing
R# 1	205.82	16.85	158.07	1.47	38.49	-87.53	0.168	0.2902
R# 2	126.75	45.81	66.63	61.95	24.24	-88.11	0.085	0.2719
R# 3	182.69	28.60	168.40	88.93	45.88	-61.78	0.071	0.2930
R# 4	123.19	47.02	78.71	51.09	42.67	-84.21	0.078	0.3065
R# 5	100.99	83.55	83.70	84.13	10.72	-72.07	0.070	0.3142

**$P4_32_12$ :**

	theta1	theta2	theta3	transX	transY	transZ	monitor	packing
R# 1	204.71	17.09	159.25	1.75	38.77	-61.19	0.272	0.3656
R# 2	125.35	44.91	68.26	75.20	30.68	8.37	0.086	0.2924
R# 3	184.25	28.20	168.02	32.14	1.05	6.18	0.073	0.3606
R# 4	123.14	48.54	78.29	100.71	17.07	-60.17	0.079	0.3152
R# 5	100.36	83.58	84.64	43.18	27.92	-85.92	0.080	0.3122

After a round of rigid body refinement, 2.5 Å electron density maps were calculated using the top solution in both space groups. Clean, traceable density in the  $P4_32_12$  map

revealed that the space group of the uPA-uPAR complex was indeed  $P4_32_12$ . This solution was then put through one round of simulated annealing and group B-factor refinement. Coordinates refined nicely to a  $R/R_{\text{free}}$  of 31.8/37.6%. The overall structure of the complex is shown in Figure 5-1. Representative electron density is shown in Figure 5-2. Comparison of the model (2FD6) with the solution (uPAR2) reveals an RMSD of 0.7 Å (Figure 5-3). Superpositions were performed using LSQMAN by aligning uPAR from both models (Kleywegt 1996). Differences were mainly in surface loops, with loops 33-37 and 129-140 of uPAR present in our model and absent in the published structure. Interestingly, residues 88-92, present in the linker between D1 and D2 and thought to be ligands for the G-protein coupled receptor FPRL1, were disordered in both our model and the published structure (Resnati et al. 2002; de Paulis et al. 2004). Molecular replacement and one round of refinement was also carried out on the platinum soaked uPAR11 data set with similar results ( $R/R_{\text{free}} = 32.9/38.2\%$ ). RMSD of this structure with 2FD6 was 0.8 Å.

## DISCUSSION

The molecular replacement solution reveals why phased molecular replacement failed. Previously, uPAR was cocrystallized with full-length uPA containing the catalytic serine-protease domain and molecular replacement had been carried out using the catalytic domain of uPA. However, examination of electron density maps from uPAR2 shows no evidence of the catalytic domain of uPA. Since the molecular replacement model contained only the ATF of uPA, strong difference peaks would be expected if the uPA catalytic domain was present, but they are not. In addition, examination of the crystal packing reveals that there is not enough space to fully accommodate a serine-protease domain.

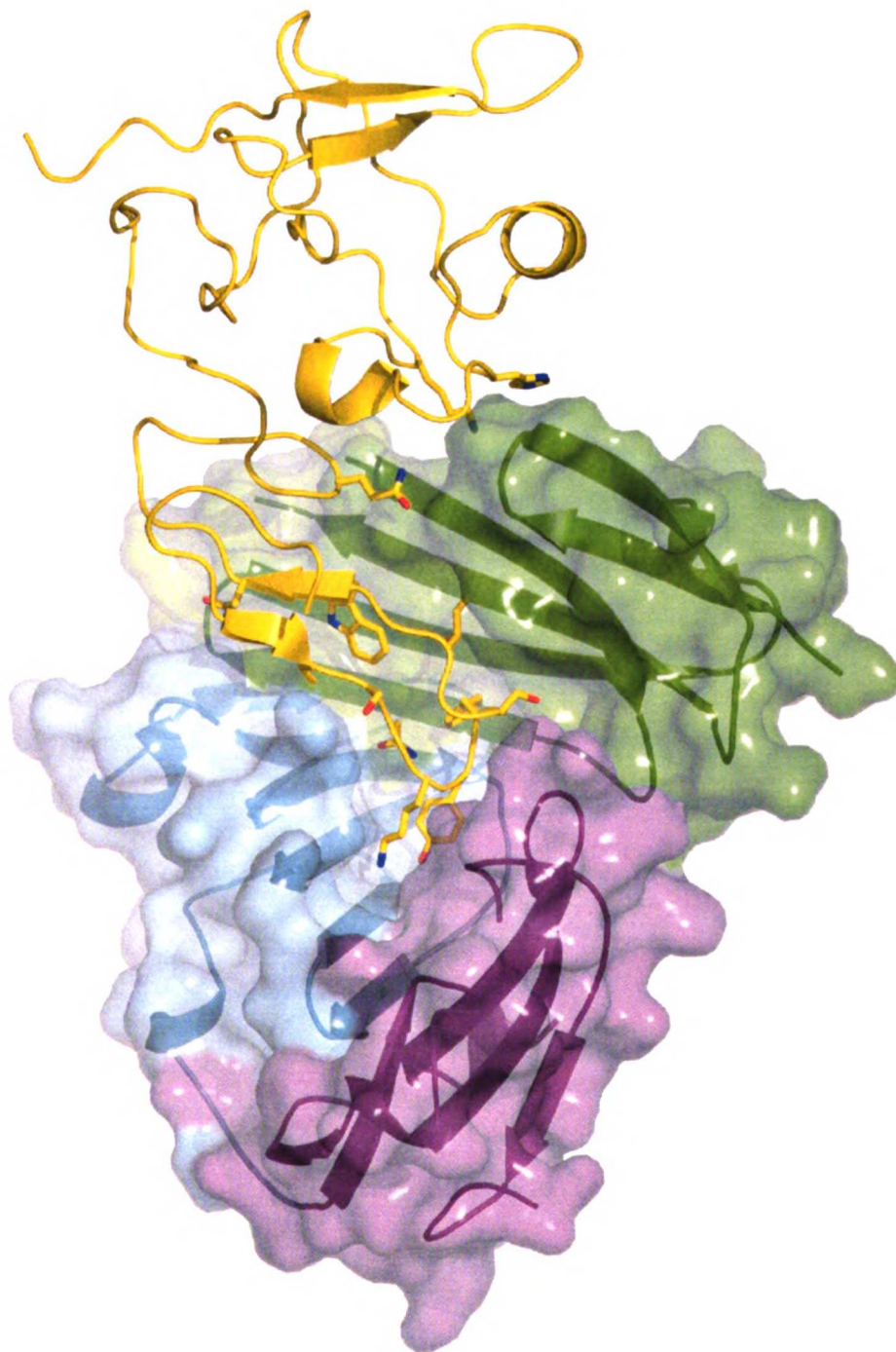
The absence of the uPA catalytic domain also helps explain why crystals were so irreproducible. Since uPAR was crystallized in the presence of full-length uPA, the presence of only the uPA ATF in the uPA-uPAR crystals suggests that cleavage of the ATF from the catalytic domain was required for crystallization. This process likely occurred slowly in the drop and was caused by either the copurification of trace amounts of protease, the slow self-cleavage of uPA caused by trace activity of the active site mutant, or oxidation of the disulfide bridge linking the ATF to the catalytic domain. Previous Western blots that showed the presence of full-length uPA and uPAR in the crystals were likely due to incomplete washing of the crystals, resulting in contamination from soluble uPA from the drop. Alternatively, if the antibody used to probe uPA had a binding epitope on the ATF, this would not distinguish between full-length uPA and ATF alone on a reducing SDS-PAGE/Western blot.

## **ACKNOWLEDGEMENTS**

I would like to thank Jennifer Garrison, Ellena Mar, and Manish Butte for their generous support and advice during the duration of this project.

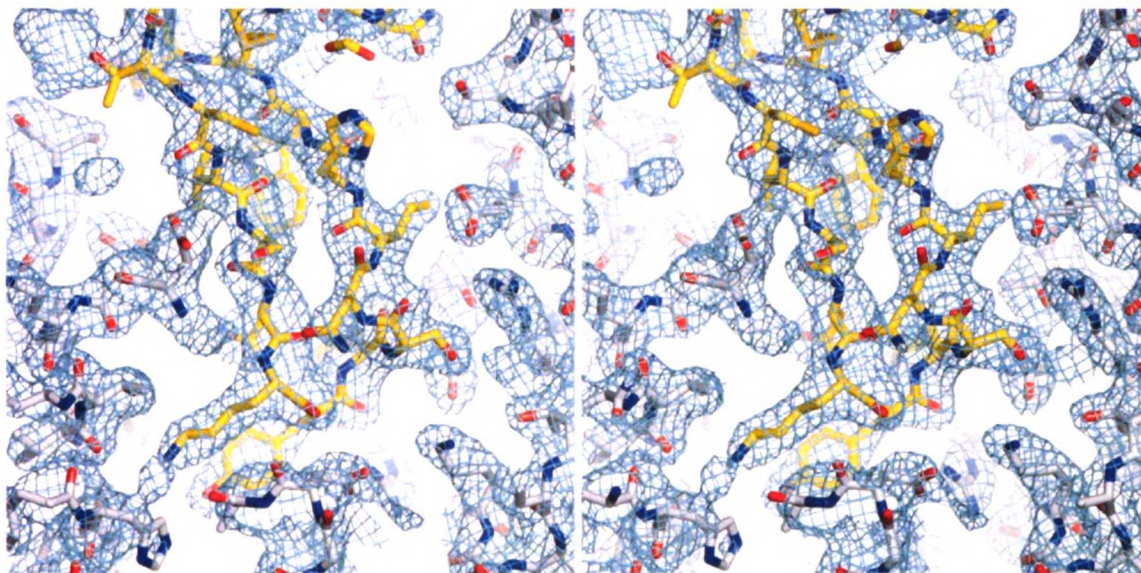
**Figure 5-1. Overall structure of the uPA-uPAR complex**

The structure of the uPA-uPAR complex from the uPAR2 dataset is shown. The ATF of uPA is colored yellow. Sidechains from the ATF making interactions with uPAR are shown as sticks. D1, D2, and D3 of uPAR are shown in green, blue, and purple, respectively.



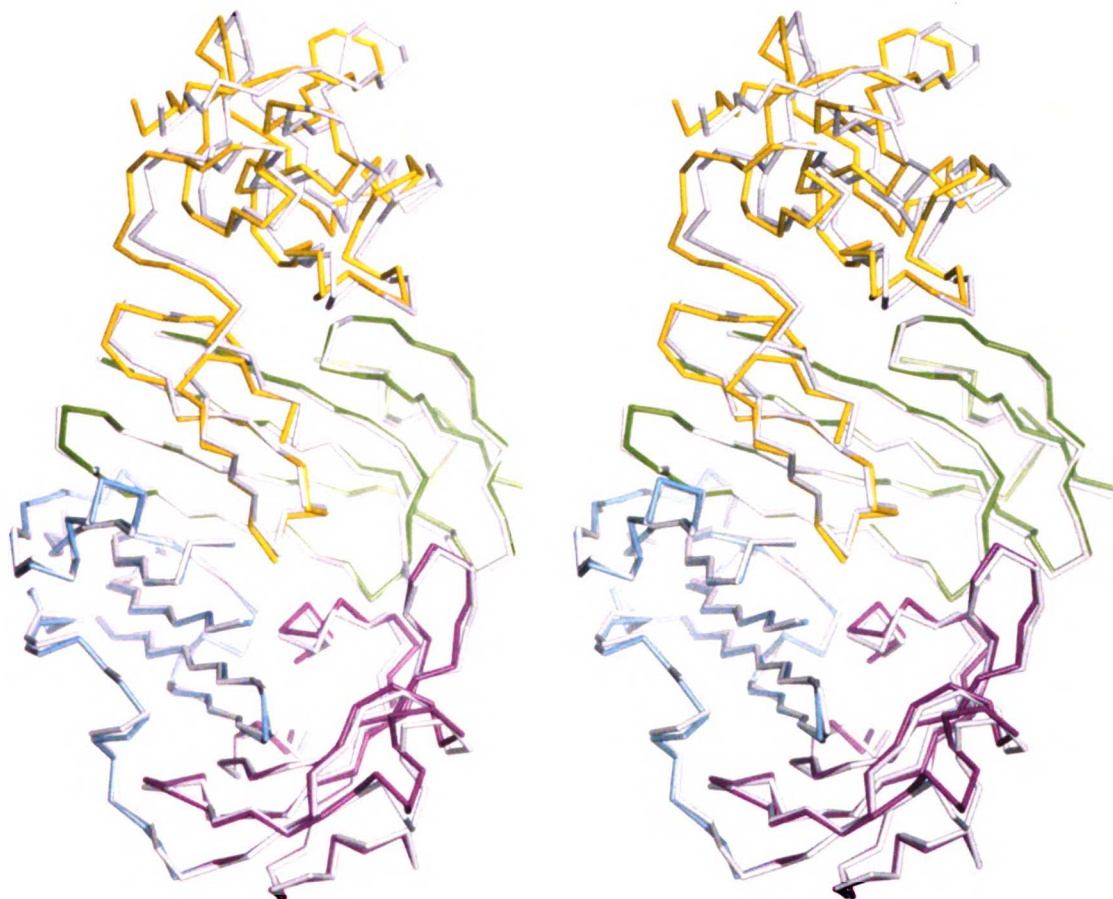
**Figure 5-2. Electron density of the uPA-uPAR interaction**

$2F_o - F_c$  electron density of the uPA-uPAR interaction is shown contoured at  $1\sigma$ . uPA and uPAR are depicted as yellow and white sticks, respectively.



### Figure 5-3. Superposition of 2FD6 with uPAR2

uPAR2 is shown superimposed with 2FD6 (Huai et al. 2006). The ATF of uPA, and D1, D2, and D3 from uPAR2 are colored yellow, green, blue, and purple, respectively. 2FD6 is colored white. RMSD between the two structures is 0.7 Å.



## REFERENCES

- Adler M, Davey DD, Phillips GB, Kim SH, Jancarik J et al. (2000) Preparation, characterization, and the crystal structure of the inhibitor ZK-807834 (CI-1031) complexed with factor Xa. *Biochemistry* 39(41): 12534-12542.
- Alen P, Claessens F, Verhoeven G, Rombauts W, Peeters B (1999) The androgen receptor amino-terminal domain plays a key role in p160 coactivator-stimulated gene transcription. *Mol Cell Biol* 19(9): 6085-6097.
- Andreasen PA, Egelund R, Petersen HH (2000) The plasminogen activation system in tumor growth, invasion, and metastasis. *Cell Mol Life Sci* 57(1): 25-40.
- Asada S, Choi Y, Uesugi M (2003) A gene-expression inhibitor that targets an alpha-helix-mediated protein interaction. *J Am Chem Soc* 125(17): 4992-4993.
- Banner DW, D'Arcy A, Chene C, Winkler FK, Guha A et al. (1996) The crystal structure of the complex of blood coagulation factor VIIa with soluble tissue factor. *Nature* 380(6569): 41-46.
- Baudino TA, Kraichely DM, Jefcoat SC, Jr., Winchester SK, Partridge NC et al. (1998) Isolation and characterization of a novel coactivator protein, NCoA-62, involved in vitamin D-mediated transcription. *J Biol Chem* 273(26): 16434-16441.
- Bianchini EP, Louvain VB, Marque PE, Juliano MA, Juliano L et al. (2002) Mapping of the catalytic groove preferences of factor xa reveals an inadequate selectivity for its macromolecule substrates. *J Biol Chem* 277(23): 20527-20534.
- Bledsoe RK, Montana VG, Stanley TB, Delves CJ, Apolito CJ et al. (2002) Crystal structure of the glucocorticoid receptor ligand binding domain reveals a novel mode of receptor dimerization and coactivator recognition. *Cell* 110(1): 93-105.
- Bloom JW, Mann KG (1978) Metal ion induced conformational transitions of prothrombin and prothrombin fragment 1. *Biochemistry* 17(21): 4430-4438.



- Borowski M, Furie BC, Bauminger S, Furie B (1986) Prothrombin requires two sequential metal-dependent conformational transitions to bind phospholipid. Conformation-specific antibodies directed against the phospholipid-binding site on prothrombin. *J Biol Chem* 261(32): 14969-14975.
- Brandstetter H, Kuhne A, Bode W, Huber R, von der Saal W et al. (1996) X-ray structure of active site-inhibited clotting factor Xa. Implications for drug design and substrate recognition. *J Biol Chem* 271(47): 29988-29992.
- Browner MF, Fauman EB, Fletterick RJ (1992) Tracking conformational states in allosteric transitions of phosphorylase. *Biochemistry* 31(46): 11297-11304.
- Broze GJ, Jr., Girard TJ, Novotny WF (1990) Regulation of coagulation by a multivalent Kunitz-type inhibitor. *Biochemistry* 29(33): 7539-7546.
- Brunger AT, Adams PD, Clore GM, DeLano WL, Gros P et al. (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr* 54 ( Pt 5): 905-921.
- Chang C, Norris JD, Gron H, Paige LA, Hamilton PT et al. (1999) Dissection of the LXXLL nuclear receptor-coactivator interaction motif using combinatorial peptide libraries: discovery of peptide antagonists of estrogen receptors alpha and beta. *Mol Cell Biol* 19(12): 8226-8239.
- Christiansen WT, Jalbert LR, Robertson RM, Jhingan A, Prorok M et al. (1995) Hydrophobic amino acid residues of human anticoagulation protein C that contribute to its functional binding to phospholipid Vesicles. *Biochemistry* 34(33): 10376-10382.
- Chuang YJ, Swanson R, Raja SM, Olson ST (2001) Heparin enhances the specificity of antithrombin for thrombin and factor Xa independent of the reactive center loop sequence. Evidence for an exosite determinant of factor Xa specificity in heparin-activated antithrombin. *J Biol Chem* 276(18): 14961-14971.

- Chung CH, Ives HE, Almeda S, Goldberg AL (1983) Purification from *Escherichia coli* of a periplasmic protein that is a potent inhibitor of pancreatic proteases. *J Biol Chem* 258(18): 11032-11038.
- Culig Z, Klocker H, Bartsch G, Hobisch A (2002) Androgen receptors in prostate cancer. *Endocr Relat Cancer* 9(3): 155-170.
- Dahl R, Wani B, Hayman MJ (1998) The Ski oncoprotein interacts with Skip, the human homolog of *Drosophila* Bx42. *Oncogene* 16(12): 1579-1586.
- Darimont BD, Wagner RL, Apriletti JW, Stallcup MR, Kushner PJ et al. (1998) Structure and specificity of nuclear receptor-coactivator interactions. *Genes Dev* 12(21): 3343-3356.
- Davie EW, Fujikawa K, Kisiel W (1991) The coagulation cascade: initiation, maintenance, and regulation. *Biochemistry* 30(43): 10363-10370.
- de Paulis A, Montuori N, Prevete N, Fiorentino I, Rossi FW et al. (2004) Urokinase induces basophil chemotaxis through a urokinase receptor epitope that is an endogenous ligand for formyl peptide receptor-like 1 and -like 2. *J Immunol* 173(9): 5739-5748.
- DeLano WL (2002) *The PyMOL Molecular Graphics System*. San Carlos, CA, USA: DeLano Scientific.
- Demarest SJ, Martinez-Yamout M, Chung J, Chen H, Xu W et al. (2002) Mutual synergistic folding in recruitment of CBP/p300 by p160 nuclear receptor coactivators. *Nature* 415(6871): 549-553.
- Dennis MS, Eigenbrot C, Skelton NJ, Ultsch MH, Santell L et al. (2000) Peptide exosite inhibitors of factor VIIa as anticoagulants. *Nature* 404(6777): 465-470.
- Ding XF, Anderson CM, Ma H, Hong H, Uht RM et al. (1998) Nuclear receptor-binding sites of coactivators glucocorticoid receptor interacting protein 1 (GRIP1) and steroid receptor coactivator 1 (SRC-1): multiple motifs with different binding specificities. *Mol Endocrinol* 12(2): 302-313.

- Falls LA, Furie BC, Jacobs M, Furie B, Rigby AC (2001) The omega-loop region of the human prothrombin gamma-carboxyglutamic acid domain penetrates anionic phospholipid membranes. *J Biol Chem* 276(26): 23895-23902.
- Fayard E, Auwerx J, Schoonjans K (2004) LRH-1: an orphan nuclear receptor involved in development, metabolism and steroidogenesis. *Trends Cell Biol* 14(5): 250-260.
- Fazioli F, Resnati M, Sidenius N, Higashimoto Y, Appella E et al. (1997) A urokinase-sensitive region of the human urokinase receptor is responsible for its chemotactic activity. *Embo J* 16(24): 7279-7286.
- Folk P, Puta F, Skruzny M (2004) Transcriptional coregulator SNW/SKIP: the concealed tie of dissimilar pathways. *Cell Mol Life Sci* 61(6): 629-640.
- Fox JD, Routzahn KM, Bucher MH, Waugh DS (2003) Maltodextrin-binding proteins from diverse bacteria and archaea are potent solubility enhancers. *FEBS Lett* 537(1-3): 53-57.
- Freedman SJ, Furie BC, Furie B, Baleja JD (1995) Structure of the calcium ion-bound gamma-carboxyglutamic acid-rich domain of factor IX. *Biochemistry* 34(38): 12126-12137.
- Freedman SJ, Blostein MD, Baleja JD, Jacobs M, Furie BC et al. (1996) Identification of the phospholipid binding site in the vitamin K-dependent blood coagulation protein factor IX. *J Biol Chem* 271(27): 16227-16236.
- Fuentes-Prior P, Iwanaga Y, Huber R, Pagila R, Rumennik G et al. (2000) Structural basis for the anticoagulant activity of the thrombin-thrombomodulin complex. *Nature* 404(6777): 518-525.
- Furie B, Furie BC (1988) The molecular basis of blood coagulation. *Cell* 53(4): 505-518.
- Geistlinger TR, Guy RK (2003) Novel selective inhibitors of the interaction of individual nuclear hormone receptors with a mutually shared steroid receptor coactivator 2. *J Am Chem Soc* 125(23): 6852-6853.

- Glass CK, Rosenfeld MG (2000) The coregulator exchange in transcriptional functions of nuclear receptors. *Genes Dev* 14(2): 121-141.
- Gottlieb B, Lehvaslaiho H, Beitel LK, Lumbroso R, Pinsky L et al. (1998) The Androgen Receptor Gene Mutations Database. *Nucleic Acids Res* 26(1): 234-238.
- Gregory CW, He B, Johnson RT, Ford OH, Mohler JL et al. (2001) A mechanism for androgen receptor-mediated prostate cancer recurrence after androgen deprivation therapy. *Cancer Res* 61(11): 4315-4319.
- Gstaiger M, Luke B, Hess D, Oakeley EJ, Wirbelauer C et al. (2003) Control of nutrient-sensitive transcription programs by the unconventional prefoldin URI. *Science* 302(5648): 1208-1212.
- Gudewicz PW, Gilboa N (1987) Human urokinase-type plasminogen activator stimulates chemotaxis of human neutrophils. *Biochem Biophys Res Commun* 147(3): 1176-1181.
- Guinto ER, Ye J, Le Bonniec BF, Esmon CT (1994) Glu192-->Gln substitution in thrombin yields an enzyme that is effectively inhibited by bovine pancreatic trypsin inhibitor and tissue factor pathway inhibitor. *J Biol Chem* 269(28): 18395-18400.
- Harris JL, Backes BJ, Leonetti F, Mahrus S, Ellman JA et al. (2000) Rapid and general profiling of protease specificity by using combinatorial fluorogenic substrate libraries. *Proc Natl Acad Sci U S A* 97(14): 7754-7759.
- He B, Wilson EM (2003) Electrostatic modulation in steroid receptor recruitment of LXXLL and FXXLF motifs. *Mol Cell Biol* 23(6): 2135-2150.
- He B, Kempainen JA, Wilson EM (2000) FXXLF and WXXLF sequences mediate the NH<sub>2</sub>-terminal interaction with the ligand binding domain of the androgen receptor. *J Biol Chem* 275(30): 22986-22994.
- He B, Lee LW, Mingos JT, Wilson EM (2002a) Dependence of selective gene activation on the androgen receptor NH<sub>2</sub>- and COOH-terminal interaction. *J Biol Chem* 277(28): 25631-25639.

- He B, Minges JT, Lee LW, Wilson EM (2002b) The FXXLF motif mediates androgen receptor-specific interactions with coregulators. *J Biol Chem* 277(12): 10226-10235.
- He B, Kemppainen JA, Voegel JJ, Gronemeyer H, Wilson EM (1999) Activation function 2 in the human androgen receptor ligand binding domain mediates interdomain communication with the NH(2)-terminal domain. *J Biol Chem* 274(52): 37219-37225.
- Hoyer-Hansen G, Ronne E, Solberg H, Behrendt N, Ploug M et al. (1992) Urokinase plasminogen activator cleaves its cell surface receptor releasing the ligand-binding domain. *J Biol Chem* 267(25): 18224-18229.
- Hsu CL, Chen YL, Yeh S, Ting HJ, Hu YC et al. (2003) The use of phage display technique for the isolation of androgen receptor interacting peptides with (F/W)XXL(F/W) and FXXLY new signature motifs. *J Biol Chem* 278(26): 23691-23698.
- Huai Q, Mazar AP, Kuo A, Parry GC, Shaw DE et al. (2006) Structure of human urokinase plasminogen activator in complex with its receptor. *Science* 311(5761): 656-659.
- Hubbard SJ, Thornton JM (1993) Naccess, Computer program. Department of Biochemistry and Molecular Biology University of College London.
- Kaiser B (2002) Factor Xa--a promising target for drug development. *Cell Mol Life Sci* 59(2): 189-192.
- Kamata K, Kawamoto H, Honma T, Iwama T, Kim SH (1998) Structural basis for chemical inhibition of human blood coagulation factor Xa. *Proc Natl Acad Sci U S A* 95(12): 6630-6635.
- Kisiel W (1979) Molecular properties of the Factor V-activating enzyme from Russell's viper venom. *J Biol Chem* 254(23): 12230-12234.
- Kleywegt GJ (1996) Use of non-crystallographic symmetry in protein structure refinement. *Acta Crystallogr D Biol Crystallogr* D52: 842-857.
- Kleywegt GJ, Brunger AT (1996) Checking your imagination: applications of the free R value. *Structure* 4(8): 897-904.

- Kodumal SJ, Patel KG, Reid R, Menzella HG, Welch M et al. (2004) Total synthesis of long DNA sequences: synthesis of a contiguous 32-kb polyketide synthase gene cluster. *Proc Natl Acad Sci U S A* 101(44): 15573-15578.
- Krem MM, Cera ED (2002) Evolution of enzyme cascades from embryonic development to blood coagulation. *Trends Biochem Sci* 27(2): 67-74.
- Krylova IN, Sablin EP, Moore J, Xu RX, Waitt GM et al. (2005) Structural analyses reveal phosphatidyl inositols as ligands for the NR5 orphan receptors SF-1 and LRH-1. *Cell* 120(3): 343-355.
- Kumar R, Betney R, Li J, Thompson EB, McEwan IJ (2004) Induced alpha-helix structure in AF1 of the androgen receptor upon binding transcription factor TFIIF. *Biochemistry* 43(11): 3008-3013.
- Lamla T, Erdmann VA (2004) The Nano-tag, a streptavidin-binding peptide for the purification and detection of recombinant proteins. *Protein Expr Purif* 33(1): 39-47.
- Langley E, Kemppainen JA, Wilson EM (1998) Intermolecular NH<sub>2</sub>-/carboxyl-terminal interactions in androgen receptor dimerization revealed by mutations that cause androgen insensitivity. *J Biol Chem* 273(1): 92-101.
- Le Bonniec BF, Esmon CT (1991) Glu-192----Gln substitution in thrombin mimics the catalytic switch induced by thrombomodulin. *Proc Natl Acad Sci U S A* 88(16): 7371-7375.
- Lee HJ, Chang C (2003) Recent advances in androgen receptor action. *Cell Mol Life Sci* 60(8): 1613-1622.
- Llinas P, Le Du MH, Gardsvoll H, Dano K, Ploug M et al. (2005) Crystal structure of the human urokinase plasminogen activator receptor bound to an antagonist peptide. *Embo J* 24(9): 1655-1663.
- Maignan S, Guilloteau JP, Pouzieux S, Choi-Sledeski YM, Becker MR et al. (2000) Crystal structures of human factor Xa complexed with potent inhibitors. *J Med Chem* 43(17): 3226-3232.

- Markus SM, Taneja SS, Logan SK, Li W, Ha S et al. (2002) Identification and characterization of ART-27, a novel coactivator for the androgen receptor N terminus. *Mol Biol Cell* 13(2): 670-682.
- Matias PM, Donner P, Coelho R, Thomaz M, Peixoto C et al. (2000) Structural evidence for ligand specificity in the binding domain of the human androgen receptor. Implications for pathogenic gene mutations. *J Biol Chem* 275(34): 26164-26171.
- Mazar AP (2001) The urokinase plasminogen activator receptor (uPAR) as a target for the diagnosis and therapy of cancer. *Anticancer Drugs* 12(5): 387-400.
- McEwan IJ (2004) Molecular mechanisms of androgen receptor-mediated gene regulation: structure-function analysis of the AF-1 domain. *Endocr Relat Cancer* 11(2): 281-293.
- McGrath ME, Erpel T, Browner MF, Fletterick RJ (1991) Expression of the protease inhibitor ecotin and its co-crystallization with trypsin. *J Mol Biol* 222(2): 139-142.
- McGrath ME, Erpel T, Bystroff C, Fletterick RJ (1994) Macromolecular chelation as an improved mechanism of protease inhibition: structure of the ecotin-trypsin complex. *Embo J* 13(7): 1502-1507.
- Merritt EA, Bacon DJ (1997) Raster3D: Photorealistic Molecular Graphics. *Methods in Enzymology* 277: 505-524.
- Mizuno H, Fujimoto Z, Atoda H, Morita T (2001) Crystal structure of an anticoagulant protein in complex with the Gla domain of factor X. *Proc Natl Acad Sci U S A* 98(13): 7230-7234.
- Nar H, Bauer M, Schmid A, Stassen JM, Wienen W et al. (2001) Structural basis for inhibition promiscuity of dual specific thrombin and factor Xa blood coagulation inhibitors. *Structure (Camb)* 9(1): 29-37.
- Needham M, Raines S, McPheat J, Stacey C, Ellston J et al. (2000) Differential interaction of steroid hormone receptors with LXXLL motifs in SRC-1a depends on residues flanking the motif. *J Steroid Biochem Mol Biol* 72(1-2): 35-46.

- Nelsestuen GL, Broderius M, Martin G (1976) Role of gamma-carboxyglutamic acid. Cation specificity of prothrombin and factor X-phospholipid binding. *J Biol Chem* 251(22): 6886-6893.
- Newmark JR, Hardy DO, Tonb DC, Carter BS, Epstein JI et al. (1992) Androgen receptor gene mutations in human prostate cancer. *Proc Natl Acad Sci U S A* 89(14): 6319-6323.
- Nolte RT, Wisely GB, Westin S, Cobb JE, Lambert MH et al. (1998) Ligand binding and co-activator assembly of the peroxisome proliferator-activated receptor-gamma. *Nature* 395(6698): 137-143.
- Norris JD, Paige LA, Christensen DJ, Chang CY, Huacani MR et al. (1999) Peptide antagonists of the human estrogen receptor. *Science* 285(5428): 744-746.
- Northrop JP, Nguyen D, Piplani S, Olivian SE, Kwan ST et al. (2000) Selection of estrogen receptor beta- and thyroid hormone receptor beta-specific coactivator-mimetic peptides using recombinant peptide libraries. *Mol Endocrinol* 14(5): 605-622.
- Ossowski L, Aguirre-Ghiso JA (2000) Urokinase receptor and integrin partnership: coordination of signaling for cell adhesion, migration and growth. *Curr Opin Cell Biol* 12(5): 613-620.
- Otwinowski Z, Minor W (1997a); Carter CW, Sweet RM, editors. New York: Academic Press. 307-326 p.
- Otwinowski Z, Minor W (1997b) Processing of X-ray Diffraction Data Collected in Oscillation Mode. In: Carter CW, Sweet RM, editors. *Methods in Enzymology: Macromolecular Crystallography, part A*. New York: Academic Press. pp. 307-326.
- Owen DJ, Vallis Y, Pearse BM, McMahon HT, Evans PR (2000) The structure and function of the beta 2-adaptin appendage domain. *Embo J* 19(16): 4216-4227.
- Padmanabhan K, Padmanabhan KP, Tulinsky A, Park CH, Bode W et al. (1993) Structure of human des(1-45) factor Xa at 2.2 Å resolution. *J Mol Biol* 232(3): 947-966.



- Paige LA, Christensen DJ, Gron H, Norris JD, Gottlin EB et al. (1999) Estrogen receptor (ER) modulators each induce distinct conformational changes in ER alpha and ER beta. *Proc Natl Acad Sci U S A* 96(7): 3999-4004.
- Perona JJ, Tsu CA, Craik CS, Fletterick RJ (1997) Crystal structure of an ecotin-collagenase complex suggests a model for recognition and cleavage of the collagen triple helix. *Biochemistry* 36(18): 5381-5392.
- Preissner KT, Kanse SM, May AE (2000) Urokinase receptor: a molecular organizer in cellular communication. *Curr Opin Cell Biol* 12(5): 621-628.
- Prendergast FG, Mann KG (1977) Differentiation of metal ion-induced transitions of prothrombin fragment 1. *J Biol Chem* 252(3): 840-850.
- Quigley CA, De Bellis A, Marschke KB, el-Awady MK, Wilson EM et al. (1995) Androgen receptor defects: historical, clinical, and molecular perspectives. *Endocr Rev* 16(3): 271-321.
- Reid J, Kelly SM, Watt K, Price NC, McEwan IJ (2002) Conformational analysis of the androgen receptor amino-terminal domain involved in transactivation. Influence of structure-stabilizing solutes and protein-protein interactions. *J Biol Chem* 277(22): 20079-20086.
- Resnati M, Guttinger M, Valcamonica S, Sidenius N, Blasi F et al. (1996) Proteolytic cleavage of the urokinase receptor substitutes for the agonist-induced chemotactic effect. *Embo J* 15(7): 1572-1582.
- Resnati M, Pallavicini I, Wang JM, Oppenheim J, Serhan CN et al. (2002) The fibrinolytic receptor for urokinase activates the G protein-coupled chemotactic receptor FPRL1/LXA4R. *Proc Natl Acad Sci U S A* 99(3): 1359-1364.
- Rezaie AR (1996) Role of residue 99 at the S2 subsite of factor Xa and activated protein C in enzyme specificity. *J Biol Chem* 271(39): 23807-23814.
- Rezaie AR (2000) Heparin-binding exosite of factor Xa. *Trends Cardiovasc Med* 10(8): 333-338.

- Rezaie AR, Esmon CT (1993) Conversion of glutamic acid 192 to glutamine in activated protein C changes the substrate specificity and increases reactivity toward macromolecular inhibitors. *J Biol Chem* 268(27): 19943-19948.
- Rezaie AR, Esmon CT (1995) Contribution of residue 192 in factor Xa to enzyme specificity and function. *J Biol Chem* 270(27): 16176-16181.
- Rezaie AR, Yang L (2001) Probing the molecular basis of factor Xa specificity by mutagenesis of the serpin, antithrombin. *Biochim Biophys Acta* 1528(2-3): 167-176.
- Sablin EP, Krylova IN, Fletterick RJ, Ingraham HA (2003) Structural basis for ligand-independent activation of the orphan nuclear receptor LRH-1. *Mol Cell* 11(6): 1575-1585.
- Sack JS, Kish KF, Wang C, Attar RM, Kiefer SE et al. (2001) Crystallographic structures of the ligand-binding domains of the androgen receptor and its T877A mutant complexed with the natural agonist dihydrotestosterone. *Proc Natl Acad Sci U S A* 98(9): 4904-4909.
- Schmitt M, Harbeck N, Thomssen C, Wilhelm O, Magdolen V et al. (1997) Clinical impact of the plasminogen activation system in tumor invasion and metastasis: prognostic relevance and target for therapy. *Thromb Haemost* 78(1): 285-296.
- Schroer A, Schneider S, Ropers H, Nothwang H (1999) Cloning and characterization of UXT, a novel gene in human Xp11, which is widely and abundantly expressed in tumor tissue. *Genomics* 56(3): 340-343.
- Seymour JL, Lindquist RN, Dennis MS, Moffat B, Yansura D et al. (1994) Ecotin is a potent anticoagulant and reversible tight-binding inhibitor of factor Xa. *Biochemistry* 33(13): 3949-3958.
- Shiau AK, Barstad D, Loria PM, Cheng L, Kushner PJ et al. (1998) The structural basis of estrogen receptor/coactivator recognition and the antagonism of this interaction by tamoxifen. *Cell* 95(7): 927-937.

- Slagsvold T, Kraus I, Bentzen T, Palvimo J, Saatcioglu F (2000) Mutational analysis of the androgen receptor AF-2 (activation function 2) core domain reveals functional and mechanistic differences of conserved residues compared with other nuclear receptors. *Mol Endocrinol* 14(10): 1603-1617.
- Soriano-Garcia M, Padmanabhan K, de Vos AM, Tulinsky A (1992) The Ca<sup>2+</sup> ion and membrane binding structure of the Gla domain of Ca-prothrombin fragment 1. *Biochemistry* 31(9): 2554-2566.
- Stemmer WP, Cramer A, Ha KD, Brennan TM, Heyneker HL (1995) Single-step assembly of a gene and entire plasmid from large numbers of oligodeoxyribonucleotides. *Gene* 164(1): 49-53.
- Stubbs MT, Bode W (1994) Coagulation factors and their inhibitors. *Curr Opin Struct Biol* 4(6): 823-832.
- Sunnerhagen M, Forsen S, Hoffren AM, Drakenberg T, Teleman O et al. (1995) Structure of the Ca(2+)-free Gla domain sheds light on membrane binding of blood coagulation proteins. *Nat Struct Biol* 2(6): 504-509.
- Taneja SS, Ha S, Swenson NK, Torra IP, Rome S et al. (2004) ART-27, an androgen receptor coactivator regulated in prostate development and cancer. *J Biol Chem* 279(14): 13944-13952.
- Tulinsky A, Park CH, Skrzypczak-Jankun E (1988) Structure of prothrombin fragment 1 refined at 2.8 Å resolution. *J Mol Biol* 202(4): 885-901.
- Ulmer JS, Lindquist RN, Dennis MS, Lazarus RA (1995) Ecotin is a potent inhibitor of the contact system proteases factor XIIa and plasma kallikrein. *FEBS Lett* 365(2-3): 159-163.
- van de Locht A, Bode W, Huber R, Le Bonniec BF, Stone SR et al. (1997) The thrombin E192Q-BPTI complex reveals gross structural rearrangements: implications for the interaction with antithrombin and thrombomodulin. *Embo J* 16(11): 2977-2984.

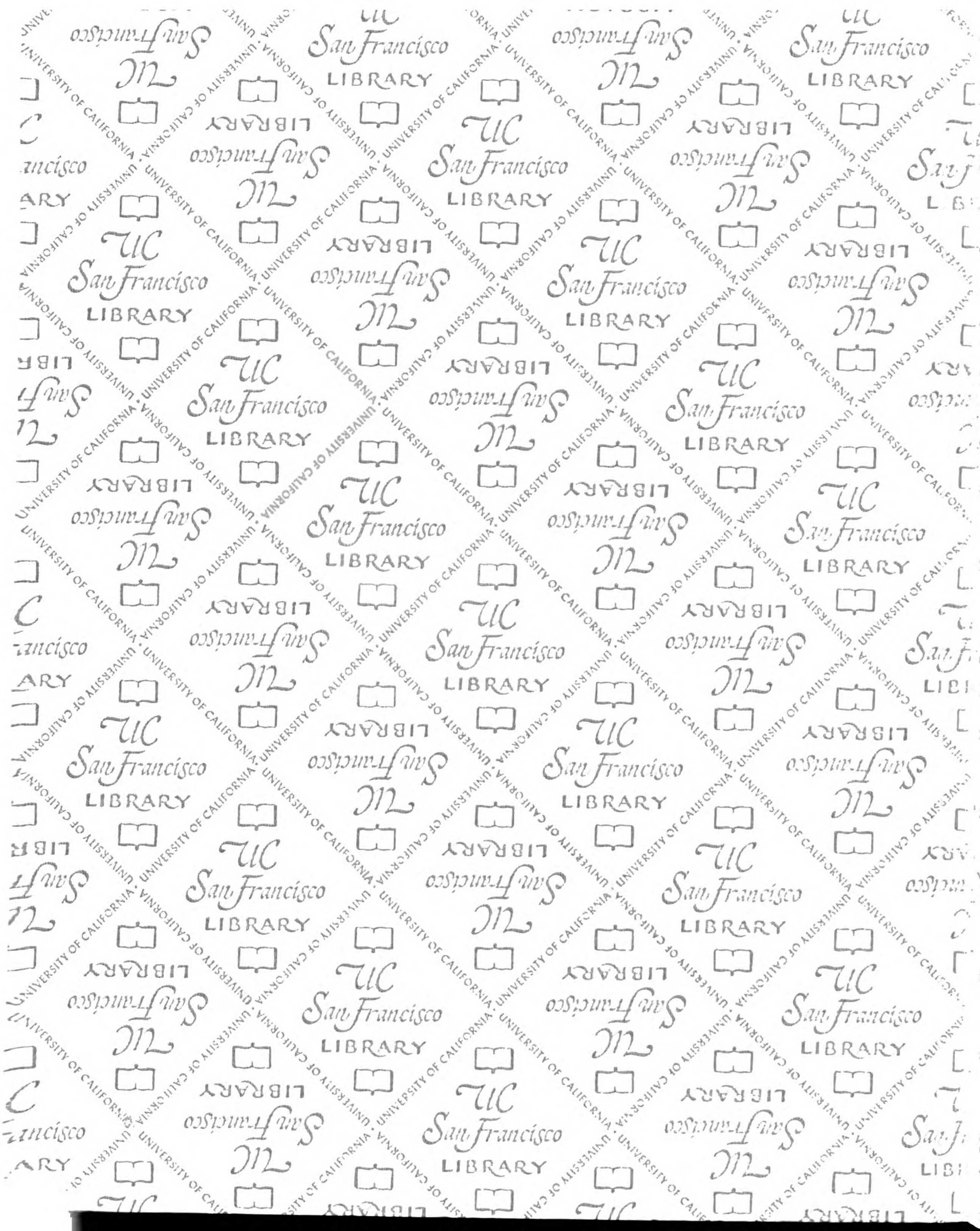
- Van Petegem F, Clark KA, Chatelain FC, Minor DL, Jr. (2004) Structure of a complex between a voltage-gated calcium channel beta-subunit and an alpha-subunit domain. *Nature* 429(6992): 671-675.
- Vassilev LT, Vu BT, Graves B, Carvajal D, Podlaski F et al. (2004) In vivo activation of the p53 pathway by small-molecule antagonists of MDM2. *Science* 303(5659): 844-848.
- Vijayalakshmi J, Padmanabhan KP, Mann KG, Tulinsky A (1994) The isomorphous structures of prethrombin2, hirugen-, and PPACK-thrombin: changes accompanying activation and exosite binding to thrombin. *Protein Sci* 3(12): 2254-2271.
- Wallace AC, Laskowski RA, Thornton JM (1995) LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein Eng* 8(2): 127-134.
- Wang Q, Lu J, Yong EL (2001a) Ligand- and coactivator-mediated transactivation function (AF2) of the androgen receptor ligand-binding domain is inhibited by the cognate hinge region. *J Biol Chem* 276(10): 7493-7499.
- Wang SX, Esmon CT, Fletterick RJ (2001b) Crystal structure of thrombin-ecotin reveals conformational changes and extended interactions. *Biochemistry* 40(34): 10038-10046.
- Warnmark A, Almlöf T, Leers J, Gustafsson JA, Treuter E (2001) Differential recruitment of the mammalian mediator subunit TRAP220 by estrogen receptors ERalpha and ERbeta. *J Biol Chem* 276(26): 23397-23404.
- Warnmark A, Treuter E, Gustafsson JA, Hubbard RE, Brzozowski AM et al. (2002) Interaction of transcriptional intermediary factor 2 nuclear receptor box peptides with the coactivator binding site of estrogen receptor alpha. *J Biol Chem* 277(24): 21862-21868.
- Waugh SM, Harris JL, Fletterick R, Craik CS (2000) The structure of the pro-apoptotic protease granzyme B reveals the molecular determinants of its specificity. *Nat Struct Biol* 7(9): 762-765.
- Wei A, Alexander RS, Duke J, Ross H, Rosenfeld SA et al. (1998) Unexpected binding mode of tick anticoagulant peptide complexed to bovine factor Xa. *J Mol Biol* 283(1): 147-154.

Yang SQ, Wang CI, Gillmor SA, Fletterick RJ, Craik CS (1998) Ecotin: a serine protease inhibitor with two distinct and interacting binding sites. *J Mol Biol* 279(4): 945-957.

Zhang C, Baudino TA, Dowd DR, Tokumaru H, Wang W et al. (2001) Ternary complexes and cooperative interplay between NCoA-62/Ski-interacting protein and steroid receptor coactivators in vitamin D receptor-mediated transcription. *J Biol Chem* 276(44): 40614-40620.

Zhang L, Castellino FJ (1994) The binding energy of human coagulation protein C to acidic phospholipid vesicles contains a major contribution from leucine 5 in the gamma-carboxyglutamic acid domain. *J Biol Chem* 269(5): 3590-3595.

UNIVERSITY OF CALIFORNIA



7537783



3 1378 00753 7783

# For reference

Not to be taken from the room.

