

Structural Studies of *Pseudomonas aeruginosa* PilY1, a Protein Central to Infections in
Cystic Fibrosis

Jillian Orans

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in
partial fulfillment of requirements for the degree of Doctor of Philosophy in the Department
of Chemistry.

Chapel Hill
2007

Approved by:

Advisor: Matthew Redinbo, Ph.D.

Reader: Gary Pielak, Ph.D.

Reader: Matthew Wolfgang, Ph.D.

Reader: Linda Spremulli, Ph.D.

Reader: Brian Kuhlman, Ph.D.

© 2007
Jillian Orans
ALL RIGHTS RESERVED

ABSTRACT

Jillian Orans: Structural Studies of *Pseudomonas aeruginosa* PilY1, a Protein Central to Infections in Cystic Fibrosis
(Under the direction of Matthew R. Redinbo)

Pseudomonas aeruginosa is an opportunistic pathogen that is of particular concern to people afflicted with Cystic Fibrosis (CF). Patients with this disease have significantly decreased lung function that leads to airway obstruction and recurrent infections, often by *Pseudomonas*. This constant infection and its resultant lung damage are eventually fatal in over 90% of CF patients. *Pseudomonas aeruginosa* is known to adhere to and colonize human host cells using a protein called pilY1. This protein has little available functional data, but its mediation of host cell infection made it an attractive structural target. Here, we present the three dimensional crystal structure of the C-terminal domain (residues 644-1148) of pilY1. This domain is conserved (23% identity, 45% similarity) with the pilC adhesion proteins from both *Neisseria gonorrhoeae* and *Neisseria meningitidis*. The structure is a seven-bladed non-canonical β -propeller and crystallized as a dimer in the asymmetric unit. The protein contains a variety of interesting structural features, including an extensive β – strand (13 residues) that is integrated within two distinct β -blades. Structural analysis of the protein led to the hypotheses that certain regions are important in pilY1 function. Through the use of functional assays to test bacterial colonization as well as biophysical methods, we were able to pinpoint two areas of the protein that appear to be relevant to its function. One area is an extended loop region located between two β -strands on adjacent blades; the other is a calcium binding site found in a short β -turn within a propeller blade. These regions

appear to be necessary to pilY1's function in twitching motility and pilus formation. Knowledge of the nature of these areas could be critical to the design of small molecules to inhibit pilY1 activity and hence *Pseudomonas aeruginosa* bacterial infection.

DEDICATION

To my parents Hedy and Alan Orans (or Mommy and Daddy) and my little sister Sami: without your love and unfailing belief in me, this work would have never been possible.

ACKNOWLEDGEMENTS

Many people have been so helpful and supportive to me along this never ending journey; I only hope I do not forget to thank them all. First and foremost I would like to thank my advisor, Matt Redinbo. The projects he devises are some of the most interesting work I have come across, and his enthusiasm for research is infectious. My work in particular has been extremely difficult but also an absolute joy to me, and I wouldn't change anything about the experience. His support and advice have meant a lot to me during this time, and I thank him profusely.

Second, I would like to acknowledge many members of the Redinbo Laboratory. I would like to thank four previous Redinbo lab members for always being there for me along the way, both professionally and emotionally. First is Ryan Watkins, who was my first friend here in Chapel Hill and kept me sane for my first year of graduate school and more. I would like to thank Jill Chrencik, who I very quickly became friends with (and has remained one of my favorite people in the world). She took me under her wing soon after I started working here, and has remained a valuable source of research information even in the years after she departed, as well as a very close friend. Next are Chris Fleming and Denise Teotico, who started in the lab with me (and just recently left) and often helped me with my research, as well as having good times with me both inside and outside of the lab. I would also like to acknowledge Scott Lujan, who also arrived with Chris, Denise, and I, and whose crazy genius has solved more than a few complex problems for everyone mentioned. Also,

thanks to our four postdocs, Kim Lane, Laura Guogas, Mike Miley, and particularly Eric Ortlund, who were constantly relied upon (and good sports about it!) for their valuable skills and experience. Thanks to all the current lab members, specifically Yuan Cheng and especially Michael Johnson, who have provided enormous experimental assistance. I feel confident that turning this project over to you will bring the lab continued success.

Thanks to our collaborators, Professor Matthew Wolfgang and graduate student Ryan Heiniger, without whom this project would never have been possible. Also thanks to Gary Pielak, who has answered a million of my silly questions with a smile and helped me immensely in the process.

Other graduate students who have helped me along the way: Brian McNulty, Erin Milner, Dee Jacobs, Candi Cunningham, and particularly Lisa Charlton, thank you for being my very best friends here and holding off my nervous breakdown for another day. Especially Lisa (who will always be Whitlow to me), who has not only helped me immensely with both discussion and actual assistance with my research, but is also just a darn cool person to hang out with. And thanks to Dee, who's just like me, only crazier.

Thanks so much to Laurie Betts, who helped me tremendously throughout my structure determination and who gave me hope when there was none.

Last but not least, I need to thank my wonderful partner Jason. His love and support these last five years have meant the world to me, and I could not have done this without him. He is not just a partner but also my best friend, and I couldn't ask for anything more. Anyone who knows me will wholeheartedly agree with this sentiment: Jay, you really are a saint!

TABLE OF CONTENTS

	Page
LIST OF TABLES.....	xiv
LIST OF FIGURES.....	xv
LIST OF ABBREVIATIONS AND SYMBOLS.....	xix
 Chapter	
I. <i>Pseudomonas aeruginosa</i> and its role as an opportunistic pathogen.....	1
Introduction.....	1
<i>Pseudomonas aeruginosa</i>	1
<i>Pseudomonas aeruginosa</i> and antibiotic resistance.....	3
<i>Pseudomonas aeruginosa</i> and Cystic Fibrosis.....	4
Cystic Fibrosis and the CFTR.....	5
<i>Pseudomonas aeruginosa</i> infection.....	7
Type III Secretion Systems.....	7
Type IV Pili.....	8
PilY1.....	9
Background.....	9
<i>Neisseria PilC</i>	9
References.....	12
Figure Legends.....	16

II.	Cloning, Expression, and Crystallization of C-terminal pilY1	21
	Introduction	21
	Cloning and Expression of PilY1	22
	<i>Structural prediction for Pseudomonas aeruginosa PilY1</i>	22
	<i>Cloning of pilY1 gene</i>	23
	<i>Test expression of pilY1 constructs</i>	24
	<i>Expression and purification of C-terminal pilY1</i>	25
	Crystallization of pilY1 C-terminal domain	26
	<i>Initial crystallization</i>	26
	<i>Crystal screening and data collection</i>	27
	Phase Determination	27
	<i>Heavy atom soaks</i>	27
	<i>Selenomethionine derivitization</i>	28
	<i>Phasing of C-terminal PilY1 using SeMet SAD</i>	31
	References	33
	Figure Legends	35
III.	Structural Analysis of PilY1 C-terminal Domain	44
	Overall structure of PilY1 C-terminal Domain	44
	Structural Characteristics	47
	<i>β-propeller structural motifs</i>	47
	<i>WD-40 repeat proteins</i>	48
	<i>Location of methionine mutants in the structure</i>	49
	<i>PilY1 C-terminal domain is an incomplete β-propeller</i>	50

	<i>Calcium binding site</i>	51
	<i>Identification of a surface patch of positive charge</i>	53
	<i>Analysis of domain swap observed in the crystal</i>	54
	References.....	55
	Figure Legends.....	58
IV.	Functional and Biophysical Characterization of PilY1 C-terminal Domain Mutations	74
	Introduction.....	74
	Regions of the protein chosen for mutational analysis.....	75
	<i>Positively charged patch</i>	75
	<i>Calcium binding region</i>	75
	<i>Dimer and domain swap residues</i>	76
	<i>Cloning mutants and mutation of associated residues</i>	76
	<i>F670L</i>	76
	<i>P1012S</i>	77
	<i>V880A</i>	77
	<i>Mutations not expected to impact function</i>	77
	Functional and biophysical analysis of wild-type pilY1 and selected mutants.....	78
	<i>Methods</i>	78
	<i>Twitching motility assays</i>	79
	<i>Pilus preparation</i>	80
	<i>Presence of pilY1 in Pseudomonas cells</i>	80
	<i>Oligomerization state of wild-type and domain swap deletion pilY1</i>	81
	<i>Analytical ultracentrifugation</i>	81

<i>Dynamic light scattering</i>	82
<i>Fluorescence resonance energy transfer to determine calcium binding constants</i>	82
Functional assay results.....	83
<i>Positively charged patch</i>	83
<i>Calcium binding region</i>	84
<i>Dimer and domain swap residues</i>	84
<i>Cloning mutants</i>	85
<i>Cloning-associated mutants</i>	85
<i>Mutants not expected to impact function</i>	86
Biophysical characterization results.....	86
<i>AUC</i>	86
<i>Dynamic light scattering</i>	87
<i>Terbium/Tryptophan FRET of wild-type and calcium binding mutants</i>	89
Discussion.....	90
<i>P782S and Y945 do not impact pilY1 function</i>	90
<i>The calcium binding site is important to pilY1 function</i>	90
<i>Analysis of additional variants</i>	91
<i>Mutations can be functionally divided into four categories</i>	93
<i>Future experiments are required to determine exact mechanisms involved in reduced mutant functionality</i>	95
References.....	98
Figure Legends.....	99

V.	The Nuclear Xenobiotic Receptor PXR: Recent Insights and New Challenges.....	130
	Abstract.....	131
	Introduction.....	132
	PXR Function.....	132
	<i>PXR in endobiotic and xenobiotic detection.....</i>	<i>132</i>
	<i>Species-specific activation.....</i>	<i>133</i>
	<i>Cross-talk with other nuclear receptors.....</i>	<i>135</i>
	<i>Subcellular localization.....</i>	<i>136</i>
	<i>Ligand binding.....</i>	<i>137</i>
	<i>Heterodimerization with RXR.....</i>	<i>137</i>
	<i>Coregulator binding.....</i>	<i>138</i>
	Mobility in PXR structure.....	139
	Areas for future study.....	144
	References.....	146
	Figure Legends.....	155
VI.	Computational Redesign of a Nuclear Receptor/Coactivator Interface to Create Altered Binding Specificity.....	161
	Introduction.....	161
	<i>Background and significance.....</i>	<i>162</i>
	Computational redesign of PXR/SRC-1 interface using Rosetta.....	164
	<i>Methods.....</i>	<i>164</i>
	<i>Results.....</i>	<i>165</i>
	Determination of binding constants for redesigned PXR/SRC-1 interface.....	166

<i>Methods</i>	166
<i>Results</i>	168
References.....	169
Figure Legends.....	172

LIST OF TABLES

Chapter		Page
II.	Table 2.1: Summary of pilY1 constructs created for test expression.....	36
	Table 2.2: Native data reduction statistics.....	37
	Table 2.3: Heavy atom compounds used in crystal soaking experiments.....	38
	Table 2.4: Summary of leucine to methionine mutations and results.....	39
	Table 2.5: Data reduction, phasing, and refinement statistics.....	40
III.	Table 3.1: Actual and expected distances of calcium ion to ligands.....	60
V.	Table 5.1: Comparison of human and mouse EC ₅₀ values for PXR and ligands.....	160
VI.	Table 6.1: Energy values of all four wild-type and mutant complexes.....	176

LIST OF FIGURES

Chapter		Page
I.	Figure 1.1: CFTR model.....	17
	Figure 1.2: Type IV pilus structure model.....	18
	Figure 1.3: Sequence homology of pilY1 to <i>Neisseria</i> pilC proteins.....	19
	Figure 1.4: PilY1 mediates bacterial adherence to human corneal cells.....	20
II.	Figure 2.1: Fluorescence scan and scattering factor analysis of native pilY1.....	41
	Figure 2.2: ESI mass spectrometry analysis of Selenomethionine substituted pilY1.....	42
	Figure 2.3: Fluorescence scan and scattering factor analysis of triple methionine mutant pilY1.....	43
III.	Figure 3.1: PilY1 dimer in crystal structure.....	61
	Figure 3.2: Front and side view of pilY1 C-terminal monomer.....	62
	Figure 3.3: An extensive β -strand encompasses two separate propeller blades.....	63
	Figure 3.4: Topology of pilY1 C-terminal domain.....	64
	Figure 3.5: Superposition of pilY1 C-terminal Domain with <i>Comamonas Testosteroni</i> quinoxemoprotein alcohol dehydrogenase.....	65
	Figure 3.6: Standard WD-40 interaction site of WDR5 β -propeller is inaccessible in pilY1.....	66
	Figure 3.7: Methionine mutations tested in pilY1 C-terminal domain.....	67
	Figure 3.8: Superposition of C-terminal pilY1 with the receptor-binding protein P2 of bacteriophage PRD1.....	68
	Figure 3.9: Sequence alignments of WDR5 propeller blades vs. pilY1 propeller blades.....	69
	Figure 3.10: PilY1 calcium binding site.....	70
	Figure 3.11: Sequence alignment of calcium binding site with various pilY1 homologues.....	71

	Figure 3.12: Electrostatic potential of pilY1 dimer.....	72
	Figure 3.13: Analysis of potential domain swap in pilY1 C-terminal domain.....	73
IV.	Figure 4.1: Locations of cloning mutants in pilY1 C-terminal domain.....	103
	Figure 4.2: Twitching motility of positive charge patch mutations.....	104
	Figure 4.3: SDS-PAGE analysis of pilus presence in <i>Pseudomonas aeruginosa</i> with positive charge mutants.....	105
	Figure 4.4: Western blots of whole cell <i>Pseudomonas aeruginosa</i> for detection of pilY1 in positive charge patch mutants.....	106
	Figure 4.5: Twitching motility of calcium binding ligand mutations.....	107
	Figure 4.6: SDS-PAGE analysis of pilus presence in <i>Pseudomonas aeruginosa</i> with calcium binding mutants.....	108
	Figure 4.7: Western blots of whole cell <i>Pseudomonas aeruginosa</i> for detection of pilY1 in calcium binding mutants.....	109
	Figure 4.8: Twitching motility of dimer and domain swap mutants.....	110
	Figure 4.9: SDS-PAGE analysis of pilus presence in <i>Pseudomonas aeruginosa</i> with dimer and domain swap mutants.....	111
	Figure 4.10: Western blots of whole cell <i>Pseudomonas aeruginosa</i> for detection of pilY1 in dimer and domain swap mutants.....	112
	Figure 4.11: Twitching motility of cloning mutations.....	113
	Figure 4.12: SDS-PAGE analysis of pilus presence in <i>Pseudomonas aeruginosa</i> with cloning mutants.....	114
	Figure 4.13: Western blots of whole cell <i>Pseudomonas aeruginosa</i> for detection of pilY1 in cloning mutants.....	115
	Figure 4.14: Twitching motility of mutations associated with cloning mutants.....	116
	Figure 4.15: SDS-PAGE analysis of pilus presence in <i>Pseudomonas aeruginosa</i> with cloning associated mutants.....	117
	Figure 4.16: Western blots of whole cell <i>Pseudomonas aeruginosa</i> for detection of pilY1 in cloning associated mutants.....	118

	Figure 4.17: Twitching motility of mutants not expected to impact function.....	119
	Figure 4.18: SDS-PAGE analysis of pilus presence in <i>Pseudomonas aeruginosa</i> with mutants not expected to impact function.....	120
	Figure 4.19: Western blots of whole cell <i>Pseudomonas aeruginosa</i> for detection of pilY1 in mutants not expected to impact function.....	121
	Figure 4.20: Analytical Ultracentrifugation shows the pilY1 C-terminal domain is a dimer.....	122
	Figure 4.21: Dynamic light scattering analysis shows that C-terminal pilY1 is mainly found in dimeric form.....	123
	Figure 4.22: Dynamic light scattering analysis shows that C-terminal pilY1 domain swap truncation is mainly found in dimeric form.....	124
	Figure 4.23: FRET analysis of wild-type C-terminal pilY1 association with terbium.....	125
	Figure 4.24: FRET analysis of calcium binding mutant D851A and its association with terbium.....	126
	Figure 4.25: FRET analysis of calcium binding mutant N853A and its association with terbium.....	127
	Figure 4.26: FRET analysis of calcium binding mutant D855A and its association with terbium.....	128
	Figure 4.27: FRET analysis of calcium binding mutant D859A and its association with terbium.....	129
V.	Figure 5.1: Structure of the PXR LBD (blue) in complex with the small agonist SR12813 (red) and a fragment of the SRC-1 transcriptional coactivator (cyan).....	156
	Figure 5.2: Close-up of the PXR-LBD bound to the large macrolide antibiotic rifampicin (green and red).....	157
	Figure 5.3: Schematic representation of the ligand binding pocket of human PXR	158
	Figure 5.4: An LxxLL motif of the human coactivator SRC-1 bound to PXR.....	159
VI.	Figure 6.1: Structure of PXR-SR12813-SRC-1 complex.....	173

Figure 6.2: Overlay of wild-type PXR/SRC-1 complex with mutant.....	174
Figure 6.3: Response curve of wild-type PXR binding to SRC-1.....	175

LIST OF ABBREVIATIONS AND SYMBOLS

Å	angstroms
A	alanine
ABC	ATP binding cassette
ACTR	nuclear receptor coactivator
AF	activation function
AIB1	amplified in breast cancer coactivator 1
AIDS	acquired immune deficiency syndrome
ω	angular velocity
ATP	adenosine triphosphate
Au	gold
AU	asymmetric unit
AUC	analytical ultracentrifugation
b	pathlength
BM	bending magnet
°C	Celcius
c	concentration
cAMP	cyclic adnosine monophosphate
CARLA	coactivator receptor ligand assays
CARM-1	chromatin arginine methyltransferase
CBP	creb binding protein
CF	Cystic fibrosis

CFTR	cystic fibrosis transmembrane conductance regulator
C-term	C-terminal region
CYP450	cytochrome P450
D	aspartic acid
DBD	DNA-binding domain
ddH ₂ O	distilled water
°	degree
Δ	change
DLS	dynamic light scattering
DNA	deoxyribonucleic acid
DR	direct repeat
DTT	dithiothreitol
ε	extinction coefficient
E	energy
EDTA	ethylenediamine tetraacetic acid
ER	endoplasmic reticulum
ER	estrogen receptor
ER	everted repeat
F	phenylalanine
FOXO1	forkhead transcription factor
FPLC	fast protein liquid chromatography
FRET	fluorescence resonance energy transfer
FWD	forward

g	grams
g·	g-force
GC	gonococci
GR	glucocorticoid receptor
GRIP	glucocorticoid receptor interacting protein
GST	glutathione-S-transferase
H	histidine
HAT	histone acetyltransferase
Hg	mercury
His	histidine
HIV	human immunodeficiency virus
HMG-CoA	3-hydroxy-3-methylglutaryl-Coenzyme A
I	intensity
ID	insertion device
IPTG	isopropyl-beta-D-thiogalactopyranoside
k	kilo
K	lysine
K _A	association constant
K _D	dissociation constant
kDa	kilodalton
λ	wavelength
LB	luria broth
LBD	ligand binding domain

Leu	leucine
M	molarity
MC	meningococci
MDR	multi-drug resistance
Met	methionine
mg	milligram
min	minutes
mL	milliliters
mM	millimolar
mRNA	messenger ribonucleic acid
N	asparagine
n	refractive index
n	Hill coefficient
NaCl	sodium chloride
NBD	nucleotide binding domain
NCoA	nuclear receptor coactivator
NCoR	nuclear receptor corepressor
Ni	nickel
NLS	nuclear localization signal
nm	nanometer
NR	nuclear receptor
OD	optical density
P	proline

PBP	PPAR binding protein
PCN	pregnenolone 16- α -carbonitrile
PCR	polymerase chain reaction
PDB	protein data bank
PEG	polyethylene glycol
PGC-1	peroxisome proliferator activating receptor
PKA	protein kinase A
PMSF	phenylmethanesulphonylfluoride
PPAR	peroxisome-proliferator-acting receptor
PR	progesterone receptor
Pt	platinum
PVC	polyvinyl chloride
PXR	pregnane X receptor
Q	glutamine
R	arginine
r	radius
R	gas constant
RAC3	ras-related C3 botulinum toxin substrate 3
REV	reverse
RIP140	receptor interacting protein
ρ	density
RPM	rotations per minute
RXR	retinoid X receptor

S	serine
SAD	single wavelength anomalous dispersion
SBC-CAT	Structural Biology Center Collaborative Access Team
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
SeMet	selenomethionine
SER-CAT	Southeast Regional Collaborative Access Team
SHP	small heterodimer partner
σ	background level
SMRT	silencing mediator of retinoid thyroid receptor
SRC-1	steroid receptor coactivator
SUG-1	suppressor for Gal-1
T	threonine
T	temperature
T3SS	type 3 secretion system
Tb	terbium
θ	fraction of binding sites filled
TR	thyroid hormone receptor
TRAM1	thyroid hormone receptor activator molecule
Tris-HCl	tris hydrochloride
UDP	uridine diphosphate
μ L	microliter
μ M	micromolar
V	valine

VDR	vitamin D receptor
W	tryptophan
WT	wild-type
Y	tyrosine

Chapter 1:

***Pseudomonas aeruginosa* and its role as an opportunistic pathogen**

INTRODUCTION

Pseudomonas aeruginosa is a pathogenic bacterium that is of particular concern to people afflicted with cystic fibrosis. *Pseudomonas* is an extremely common hospital-acquired bacterium that can lead to urinary tract infections, pneumonia, and sepsis. In addition, chronic lung infections caused by this bacterium can result in cumulative organ damage that is eventually fatal in most patients. *Pseudomonas* attacks human cells using a pilus, a rod-like extracellular organelle responsible for attachment of the bacterial cell to the host cell. A key component of the pilus is the protein pilY1, which has been shown to be necessary for adherence to human cells and is required for pilus biogenesis. This protein is essential for *Pseudomonas* infections. PilY1 homologues are found in other infectious bacteria, including *Neisseria gonorrhoeae* and *meningitidis*. Thus, elucidating the detailed structural basis for pilY1 function would provide novel avenues for treating a wide range of bacterial infections.

PSEUDOMONAS AERUGINOSA

Pseudomonas aeruginosa is a rod-shaped, gram negative bacterium that is pathogenic to a wide variety of eukaryotic systems [1]. In humans, it is an opportunistic pathogen, only infecting tissues in people with compromised immune systems. This is particularly a problem for AIDS patients, patients undergoing chemotherapy, burn victims, and people with cystic fibrosis. *Pseudomonas aeruginosa* accounts for over 10% of all nosocomial infections that are identifiable as bacterial, ranking second only to *Staphylococcus aureus* among all gram negative pathogens, according to data from the National Nosocomial Infection Surveillance System encompassing the last decade [2, 3]. It has an incidence of 0.4%; this number is even higher in intensive care units [4-9]. It can infect a wide variety of tissues, including the eyes, lungs, urinary tract, gastrointestinal system, and skin and soft tissues. Hence, *Pseudomonas* infection is common and is also associated with higher rates of mortality from pneumonia in comparison to many other bacterial pathogens [10]. In several studies of fatal nosocomial pneumonia infection, 50% of the patients tested positive for *Pseudomonas aeruginosa*; other cases were associated with at least eight other less prominent types of bacteria, including *Staphylococcus aureus* and *Haemophilus influenzae* [11, 12].

Pseudomonas is an aerobic bacterium that is ubiquitous in the environment. It is found in soil and water, on the surface of plants, and in decaying organic matter. It is also found in many places in the hospital environment, such as sinks, respirators, humidifiers, and on the person of medical staff [13]. *P. aeruginosa* has minimal nutritional requirements, allowing it to grow in numerous and often unexpected environments including distilled water, and allowing it to tolerate a wide variety of physical conditions such as extreme temperature, high salt, and even antiseptics. It has been isolated directly from iodophor

disinfectant, phenolic germicide, and iodophor antiseptic solutions, and has also been found colonizing PVC pipes that had previously been treated with chlorine, phenolic, quaternary-ammonium, and iodophor disinfectants [14].

Pseudomonas aeruginosa and antibiotic resistance

What makes *Pseudomonas aeruginosa* such a virulent pathogen is its resistance to a broad spectrum of available antibiotics. It is intrinsically resistant to a number of antibiotic families, including the β -lactams, macrolides, tetracyclins, co-trimoxazole, and most fluoroquinolones [4]. This resistance is due to a combination of mechanisms including multidrug efflux pumps, prevention of entry due to low permeability of the outer membrane (even in comparison to other gram-negative bacteria such as *Escherichia coli*), and efficient enzymatic degradation by β -lactamases. [4]. In addition, *P. aeruginosa* is capable of developing resistance to other families of antibiotics, often benefiting from previous exposure. It can transfer antibiotic resistance (R) plasmids from one *Pseudomonas* strain to another, as well as to *E. coli* [15]. These plasmids confer resistance to various aminoglycosides, including gentamycin, kanamycin, tobramycin, and streptomycin; over 100 of these plasmids have been described [15-19]. There are a limited number of antibiotic types that are effective against this bacterium, including antipseudomonal penicillins, cephalosporins, carbapenems, and ciprofloxacin, and resistance to these drugs is emerging rapidly [20]. In one recent study of 22 patients with nosocomial multi-resistant *P. aeruginosa* infection, six people developed reversion of previous antibiotic susceptibility; in some patients this happened in as quickly as a one-day time period [21].

Pseudomonas aeruginosa and Cystic Fibrosis

Pseudomonas aeruginosa is a devastating pathogen in particular to people with cystic fibrosis. Cystic fibrosis (CF) is the most common fatal autosomal recessive genetic disorder among people with European ancestry, occurring with a frequency of 1 in every 3000 live births [22]. About 30,000 people in the United States and almost 70,000 people worldwide are affected by CF. There is currently no cure for CF; according to the Cystic Fibrosis Foundation, the predicted median survival age of patients in the US is 36.5 years. People with CF are prone to chronic lung infections, often due to *Pseudomonas aeruginosa*. Studies have shown that the bacterium colonizes the cystic fibrosis lung through the use of biofilms which form in the thickened mucus [23, 24]. These biofilms are communities of the bacteria that attach and form colonies that can grow on a wide variety of biotic and abiotic surfaces. This colonization makes clearance difficult and has also been found to contribute to antibiotic resistance [25]. A recent study determined that by the age of 3, over 95% of children with CF have shown signs of intermittent infection by the bacterium [25, 26]. *Pseudomonas* is the major pathogenic cause of persistent lung infections, which result in progressive deterioration of the lung function and eventual mortality.

CYSTIC FIBROSIS AND CFTR

Cystic fibrosis is caused by mutations in the cystic fibrosis transmembrane conductance regulator (CFTR) protein, which is necessary for chloride regulation at the epithelial cell surface, as well as regulatory interactions with other channel proteins. The CFTR gene was discovered in 1989; this discovery has been paramount in building a greater

understanding of the underlying causes of CF, the mutations associated with the disease, and the potential for therapeutics. The CFTR is a cyclic adenosine monophosphate (cAMP) regulated member of the ATP-binding cassette (ABC) transporter family of proteins which functions as a plasma membrane chloride channel under the regulation of protein kinase A (PKA) [27]. The protein is 1480 amino acids and is comprised of a two transmembrane regions containing the chloride pore, as well as two intracellular nucleotide binding domains (NBDs) and a regulatory domain. Gating of the channel is dependent on ATP binding; it is not yet known whether hydrolysis is also necessary. [28] Activation of PKA regulates opening of the pore and facilitates tissue-specific recycling [29].

The CFTR protein is found in the epithelial cells lining various organs, including the lungs, digestive tract, sweat glands, and urinary tract. Over 1,000 mutations associated with the disease have been identified to date. These genetic mutations result in a number of scenarios which leave the CFTR protein dysfunctional. They can be frameshift, deletions, insertions, or missense mutations, and the resulting alleles have been grouped into five classes [30]. Class 1 alleles contain frameshift or nonsense mutations; these result in unstable mRNA that is never translated into protein. Class 2 alleles, including the $\Delta F508$ mutation located in NBD1 and present in 70% of people with CF [31], result in a misfolded CFTR protein that is held and degraded in the ER. Class 3 alleles code for normally processed proteins that contain missense mutations that render CFTR inactive. Class 4 alleles contain missense mutations that reduce CFTR's chloride ion conductivity, attributed to a shortened time in which the channel is open or to slower ion movement [32]. In one study that characterized several of these mutations, chloride conductance was found to be ~20 fold lower than in the wild-type protein [33]. Finally, class 5 alleles encode mutations

that lead to abnormal mRNA splicing and protein maturation, resulting in fewer CFTR proteins with normal activity at the cell surface. The various types of mutations observed in the classes manifest themselves in the form of disease severity, with class 1-3 alleles associated with complete dysfunction of the CFTR and rigorous lung and pancreatic symptoms, and those of classes 4 and 5 associated with partial protein deficiency and a range of milder symptoms [34].

Mutations in the CFTR cause a host of problems for the epithelial cells and hence the organs they line. In particular, patients with CF have chronic infections in the lungs, causing progressive deterioration. Cystic fibrosis lungs are extremely susceptible to bacterial infection, especially by *Pseudomonas aeruginosa*, as a direct result of the dysfunctional CFTR protein in the epithelia. In mutated CFTR, chloride is not properly transported through the pore of the protein, leading to an ionic imbalance in the epithelial cells. This results in thick mucus which lines the lungs, causing airway obstruction and infection due to difficulty in clearance. Mucus clearance is considered an innate primary defense system for protection against bacterial infection; the ability to rid the lungs of infected mucus is necessary to prevent chronic infection and eventual respiratory failure [35]. The CF lung is incapable of effective clearance and as such is exposed to constant bacterial infection and its consequences.

***PSEUDOMONAS AERUGINOSA* INFECTION**

Type III Secretion Systems

Central to the virulence of *P. aeruginosa* is the type III secretion system (T3SS), which functions by transporting effector proteins into the eukaryotic host cell [36].

Effector proteins modify host cell processes allowing colonization of the bacteria and avoidance of the host cell immunodefense system [37]. The T3SS comprises of over 20 proteins and is considered one of the most sophisticated virulence systems in gram-negative pathogens [38]. Secretion of proteins through the bacterial envelope is moderated by what is known as a “needle complex”. This complex contains a multi-ring base, anchoring the structure to the membrane, and a needle-like projection traversed by a narrow channel through which the proteins most likely pass [38]. This needle complex has been identified in several types of gram-negative bacteria, including *P. aeruginosa* [39], and is presumed to be the main component of all type III secretion systems. Needle complexes are responsible not only for recognizing substrates so that they are appropriately targeted, but also for delivering proteins into target cells at opportune times. Translocated proteins typically contain a secretion signal that remains uncleaved. It has been proposed that the toxin proteins are delivered through the use of “translocator” proteins, which are also secreted and insert themselves into the host cell membrane, paving the way for the needle complex to dock onto the formed pore or channel to directly deliver the effector proteins [38, 40, 41].

Type IV Pili

Pseudomonas aeruginosa adheres to the eukaryotic host cell surface through the use of type IV pili. Pili are rod-like appendages anchored at the cell membrane that are used to mediate adhesion and forceful contraction [42]. The majority of the pilus is composed of a small repeating structural subunit called pilin. In *P. aeruginosa*, this protein is PilA (from this point forward, all type IV pilus proteins will be referred to using the *P. aeruginosa* nomenclature unless otherwise stated). Repeating copies of PilA make up the basic structural

unit of the pilus. Many other functional proteins are located within or associated with the pilus structure. PilB and PilT are NTPases involved in pilus assembly and retraction, respectively. PilQ is a secretin located in the outer membrane that serves as a pore through which the pilus passes during extension and retraction. Two proteins are located within the inner membrane: PilD serves as a prepilin peptidase, and PilC is thought to interact with pilB during translocation and assembly of the pilus. Associated with but not found within the extracellular pilin structure are minor proteins called pseudopilins, including PilE, PilV-X, and FimT-U, which have been found to play a part in typeIV pilus biogenesis [42]. Lastly there is PilY1, a putative adhesion protein that has been found both extracellularly in association with the pilus and in the membrane fraction of the cell [43].

PILY1

Background

The *pilY1* gene from *Pseudomonas aeruginosa* was discovered by Richard Alm and colleagues at the University of Queensland in 1996 [43]. It was identified at the same time as three other previously uncharacterized genes, *pilW*, *pilX*, and *pilY2*, all of which were initially implicated in pilus biogenesis. PilY1 is a large (127 kDa) protein with ~1163 amino acids (this value varies between different strains). As previously stated, fractionation of *Pseudomonas* cells shows that PilY1 is found mainly in the membrane fraction, with a small amount seen in the extracellular fimbrial fraction. Additional experiments indicated that the extracellular PilY1 was associated with the pilin subunit protein PilA [43].

Neisseria PilC

When the protein was compared to protein sequence databases [44], Alm noted that PilY1 was a significant match to the 110 kDa pilC protein of *Neisseria gonorrhoeae*. *N. gonorrhoeae* (gonococci or GC) is also a gram-negative human pathogen that infects mucosal surfaces, causing sexually transmitted endocervicitis in women and urethritis in men. GC is also antibiotic resistant; several families of previously effective drugs including penicillins, tetracyclines, and quinolones now have limited utility, and this resistance is increasing over time [45]. As just one example, there has been greater than a 16% increase in ciprofloxacin-resistant GC in all isolates garnered from Sydney, Australia in a span of only 12 years, from 1990-2002 [46]. To convey the gravity of this epidemic, consider recent studies showing HIV transmission rates may be up to five times greater in people with gonorrhea, attributed to an increased viral load in genital fluids [47].

The PilC protein from *N. gonorrhoeae* was identified by Jonsson and colleagues at Washington University in 1991 [48]. The protein was found to be expressed in the outer membrane as well as in purified pili preparations. Additional experiments found that PilC null mutants express dramatically reduced levels of pili, demonstrating that the protein is important in pilus biogenesis. The protein had also recently been identified as the pilus tip and the vehicle of adherence to host epithelial cells [49]. In 1995, Rudel and coworkers showed that purified PilC from *N. gonorrhoeae* effectively competes with pilus-mediated attachment of the bacteria to human ME-180 epithelial cells [49].

Both PilY1 and *N. gonorrhoeae* PilC share significant homology with the PilC protein from *Neisseria meningitidis*. *N. meningitidis* (meningococci or MC) is another gram-negative human pathogen that is responsible for the worldwide meningitis epidemic. This

bacterium is particularly dangerous because of its ability to activate genetic switches that affect expression of virulence genes, causing pathogenic and structural variability that allows MC to elude administered antibiotics [50]. MC is the leading cause of bacterial meningitis and sepsis in the US [51]. Between 1-5 people in every 100,000 become infected, and 10% of these will die, with another 10-20% affected by severe disabilities such as hearing loss and limb amputation [52].

Homology between *P. aeruginosa* PilY1 and *Neisseria* PilC proteins is restricted to the C-terminal region; Alm defined this region from residue 613 to the C-term of PilY1. In fact, this homology extends to other PilY1 proteins from various bacteria including *Xanthomonas* and *Legionella*, as well as the PilC proteins from the *Neisseria* species. It was therefore a logical conclusion that PilY1 would have similar functions to those found for PilC.

Recent cellular adherence experiments provide direct evidence identifying PilY1 as the adhesin in *Pseudomonas aeruginosa*. Unpublished data from the laboratory of Matthew Wolfgang at UNC-Chapel Hill show that PilY1 null mutants cannot adhere to human corneal cells that are easily infected by the wild-type bacteria. These results make PilY1 an attractive structural target, both as a means to identify regions of the protein that are important for adherence and pilus biogenesis, and to potentially develop a compound that could inhibit bacterial infection.

How does PilY1 function? What domains of the protein are required for pilus adherence? What are PilY1's binding partners, and which area of the protein is involved in this binding? Can we identify areas of the protein that are critical for function in either

adherence or biogenesis, and hence are potential targets for inhibition? These are several of the important questions I sought to address in the structural determination of PilY1.

REFERENCES

1. Mahajan-Miklos, S., L.G. Rahme, and F.M. Ausubel, *Elucidating the molecular mechanisms of bacterial virulence using non-mammalian hosts*. Mol Microbiol, 2000. **37**(5): p. 981-8.
2. Pellegrino, F.L., et al., *Occurrence of a multidrug-resistant Pseudomonas aeruginosa clone in different hospitals in Rio de Janeiro, Brazil*. J Clin Microbiol, 2002. **40**(7): p. 2420-4.
3. *National Nosocomial Infections Surveillance (NNIS) System report, data summary from January 1990-May 1999, issued June 1999*. Am J Infect Control, 1999. **27**(6): p. 520-32.
4. Driscoll, J.A., S.L. Brody, and M.H. Kollef, *The epidemiology, pathogenesis and treatment of Pseudomonas aeruginosa infections*. Drugs, 2007. **67**(3): p. 351-68.
5. Erbay, H., et al., *Nosocomial infections in intensive care unit in a Turkish university hospital: a 2-year survey*. Intensive Care Med, 2003. **29**(9): p. 1482-8.
6. Gaynes, R. and J.R. Edwards, *Overview of nosocomial infections caused by gram-negative bacilli*. Clin Infect Dis, 2005. **41**(6): p. 848-54.
7. Kim, J.M., et al., *Multicenter surveillance study for nosocomial infections in major hospitals in Korea. Nosocomial Infection Surveillance Committee of the Korean Society for Nosocomial Infection Control*. Am J Infect Control, 2000. **28**(6): p. 454-8.
8. Lizioli, A., et al., *Prevalence of nosocomial infections in Italy: result from the Lombardy survey in 2000*. J Hosp Infect, 2003. **54**(2): p. 141-8.
9. Pittet, D., et al., *Prevalence and risk factors for nosocomial infections in four university hospitals in Switzerland*. Infect Control Hosp Epidemiol, 1999. **20**(1): p. 37-42.
10. Osmon, S., et al., *Hospital mortality for patients with bacteremia due to Staphylococcus aureus or Pseudomonas aeruginosa*. Chest, 2004. **125**(2): p. 607-16.
11. Rello, J., et al., *Impact of previous antimicrobial therapy on the etiology and outcome of ventilator-associated pneumonia*. Chest, 1993. **104**(4): p. 1230-5.
12. Rello, J., et al., *A three-year study of severe community-acquired pneumonia with emphasis on outcome*. Chest, 1993. **103**(1): p. 232-5.
13. Pitt, T.L., *Biology of Pseudomonas aeruginosa in relation to pulmonary infection in cystic fibrosis*. J R Soc Med, 1986. **79 Suppl 12**: p. 13-8.

14. Anderson, R.L., et al., *Effect of disinfectants on pseudomonads colonized on the interior surface of PVC pipes*. Am J Public Health, 1990. **80**(1): p. 17-21.
15. Jacoby, G.A., *Properties of R plasmids determining gentamicin resistance by acetylation in Pseudomonas aeruginosa*. Antimicrob Agents Chemother, 1974. **6**(3): p. 239-52.
16. Jacoby, G.A., *Properties of an R plasmid in Pseudomonas aeruginosa producing amikacin (BB-K8), butirosin, kanamycin, tobramycin, and sisomicin resistance*. Antimicrob Agents Chemother, 1974. **6**(6): p. 807-10.
17. Holloway, B.W., V. Krishnapillai, and A.F. Morgan, *Chromosomal genetics of Pseudomonas*. Microbiol Rev, 1979. **43**(1): p. 73-102.
18. Shahid, M. and A. Malik, *Plasmid mediated amikacin resistance in clinical isolates of Pseudomonas aeruginosa*. Indian J Med Microbiol, 2004. **22**(3): p. 182-4.
19. Shahid, M. and A. Malik, *Resistance due to aminoglycoside modifying enzymes in Pseudomonas aeruginosa isolates from burns patients*. Indian J Med Res, 2005. **122**(4): p. 324-9.
20. Carmeli, Y., et al., *Emergence of antibiotic-resistant Pseudomonas aeruginosa: comparison of risks associated with different antipseudomonal agents*. Antimicrob Agents Chemother, 1999. **43**(6): p. 1379-82.
21. Harris, A., et al., *Epidemiology and clinical outcomes of patients with multiresistant Pseudomonas aeruginosa*. Clin Infect Dis, 1999. **28**(5): p. 1128-33.
22. Marcet, B. and J.M. Boeynaems, *Relationships between cystic fibrosis transmembrane conductance regulator, extracellular nucleotides and cystic fibrosis*. Pharmacol Ther, 2006. **112**(3): p. 719-32.
23. Baltimore, R.S., C.D. Christie, and G.J. Smith, *Immunohistopathologic localization of Pseudomonas aeruginosa in lungs from patients with cystic fibrosis. Implications for the pathogenesis of progressive lung deterioration*. Am Rev Respir Dis, 1989. **140**(6): p. 1650-61.
24. Singh, P.K., et al., *Quorum-sensing signals indicate that cystic fibrosis lungs are infected with bacterial biofilms*. Nature, 2000. **407**(6805): p. 762-4.
25. Murray, T.S., M. Egan, and B.I. Kazmierczak, *Pseudomonas aeruginosa chronic colonization in cystic fibrosis patients*. Curr Opin Pediatr, 2007. **19**(1): p. 83-8.
26. Burns, J.L., et al., *Longitudinal assessment of Pseudomonas aeruginosa in young children with cystic fibrosis*. J Infect Dis, 2001. **183**(3): p. 444-52.

27. Riordan, J.R., et al., *Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA*. Science, 1989. **245**(4922): p. 1066-73.
28. MacDonald, K.D., K.R. McKenzie, and P.L. Zeitlin, *Cystic fibrosis transmembrane regulator protein mutations: 'class' opportunity for novel drug innovation*. Paediatr Drugs, 2007. **9**(1): p. 1-10.
29. Okiyoneda, T. and G.L. Lukacs, *Cell surface dynamics of CFTR: the ins and outs*. Biochim Biophys Acta, 2007. **1773**(4): p. 476-9.
30. Chillon, M., et al., *Mutations in the cystic fibrosis gene in patients with congenital absence of the vas deferens*. N Engl J Med, 1995. **332**(22): p. 1475-80.
31. Moskowitz, S.M., R.L. Gibson, and E.L. Effmann, *Cystic fibrosis lung disease: genetic influences, microbial interactions, and radiological assessment*. Pediatr Radiol, 2005. **35**(8): p. 739-57.
32. Gadsby, D.C., P. Vergani, and L. Csanady, *The ABC protein turned chloride channel whose failure causes cystic fibrosis*. Nature, 2006. **440**(7083): p. 477-83.
33. Jiang, Q., et al., *Cystic fibrosis transmembrane conductance regulator-associated ATP release is controlled by a chloride sensor*. J Cell Biol, 1998. **143**(3): p. 645-57.
34. McKone, E.F., et al., *Effect of genotype on phenotype and mortality in cystic fibrosis: a retrospective cohort study*. Lancet, 2003. **361**(9370): p. 1671-6.
35. Boucher, R.C., *An overview of the pathogenesis of cystic fibrosis lung disease*. Adv Drug Deliv Rev, 2002. **54**(11): p. 1359-71.
36. Yahr, T.L. and M.C. Wolfgang, *Transcriptional regulation of the Pseudomonas aeruginosa type III secretion system*. Mol Microbiol, 2006. **62**(3): p. 631-40.
37. Hussein, M.I., F. Wartha, and M. Hensel, *Recombinant vaccines based on translocated effector proteins of Salmonella Pathogenicity Island 2*. Vaccine, 2007. **25**(1): p. 185-93.
38. Galan, J.E. and H. Wolf-Watz, *Protein delivery into eukaryotic cells by type III secretion machines*. Nature, 2006. **444**(7119): p. 567-73.
39. Quinaud, M., et al., *The PscE-PscF-PscG complex controls type III secretion needle biogenesis in Pseudomonas aeruginosa*. J Biol Chem, 2005. **280**(43): p. 36293-300.
40. Hakansson, S., et al., *The YopB protein of Yersinia pseudotuberculosis is essential for the translocation of Yop effector proteins across the target cell plasma membrane and*

- displays a contact-dependent membrane disrupting activity*. Embo J, 1996. **15**(21): p. 5812-23.
41. Tardy, F., et al., *Yersinia enterocolitica type III secretion-translocation system: channel formation by secreted Yops*. Embo J, 1999. **18**(23): p. 6793-9.
 42. Pizarro-Cerda, J. and P. Cossart, *Bacterial adhesion and entry into host cells*. Cell, 2006. **124**(4): p. 715-27.
 43. Alm, R.A., et al., *Fimbrial biogenesis genes of Pseudomonas aeruginosa: pilW and pilX increase the similarity of type 4 fimbriae to the GSP protein-secretion systems and pilY1 encodes a gonococcal PilC homologue*. Mol Microbiol, 1996. **22**(1): p. 161-73.
 44. Altschul, S.F., et al., *Basic local alignment search tool*. J Mol Biol, 1990. **215**(3): p. 403-10.
 45. Tapsall, J.W., *Antibiotic resistance in Neisseria gonorrhoeae*. Clin Infect Dis, 2005. **41 Suppl 4**: p. S263-8.
 46. Tapsall, J.W., E.A. Limnios, and T.R. Shultz, *Continuing evolution of the pattern of quinolone resistance in Neisseria gonorrhoeae isolated in Sydney, Australia*. Sex Transm Dis, 1998. **25**(8): p. 415-7.
 47. Cohen, M.S., *Sexually transmitted diseases enhance HIV transmission: no longer a hypothesis*. Lancet, 1998. **351 Suppl 3**: p. 5-7.
 48. Jonsson, A.B., G. Nyberg, and S. Normark, *Phase variation of gonococcal pili by frameshift mutation in pilC, a novel gene for pilus assembly*. Embo J, 1991. **10**(2): p. 477-88.
 49. Rudel, T., I. Scheurerpflug, and T.F. Meyer, *Neisseria PilC protein identified as type-4 pilus tip-located adhesin*. Nature, 1995. **373**(6512): p. 357-9.
 50. Stephens, D.S., B. Greenwood, and P. Brandtzaeg, *Epidemic meningitis, meningococcaemia, and Neisseria meningitidis*. Lancet, 2007. **369**(9580): p. 2196-210.
 51. Rosenstein, N.E., et al., *Meningococcal disease*. N Engl J Med, 2001. **344**(18): p. 1378-88.
 52. Healy, C.M. and C.J. Baker, *The future of meningococcal vaccines*. Pediatr Infect Dis J, 2005. **24**(2): p. 175-6.

FIGURE LEGENDS

Figure 1.1 Proposed model of the cystic fibrosis transmembrane conductance regulator (CFTR). It is a 1480 amino acid protein composed of a transmembrane region that contains the chloride pore, two nucleotide binding domains (NBDs), and a regulatory domain [28].

Figure 1.2 Proposed model for the type IV pilus in *Pseudomonas aeruginosa*. The pilin subunit, PilA, is shown in purple. The proposed adhesin PilY1 is shown in red. PilT and PilB, cytoplasmic NTPases, are shown in green and yellow triangles. Prepilin peptidase PilD is shown in olive, and secretin PilQ is shown in blue.

Figure 1.3 The PilC proteins from *Neisseria gonorrhoeae* and *Neisseria meningitidis* have sequence homology with PilY1 only within their C-terminal domains. Residue numbers indicate the sequence of PilY1; all three proteins are of the same relative size. Sequence identity is shown for both domains of the proteins with the N-terminal domain shown as transparent and the C-terminal domain as solid. Alignments were carried out using BLAST [44].

Figure 1.4 Matthew Wolfgang and colleagues (unpublished) have shown that *Pseudomonas aeruginosa* can only infect these human corneal cells if PilY1 is present. In the top panel, WT PAK strain adheres to the host cell. In the bottom panel, a PilY1 null mutant cannot bind to the same cells.

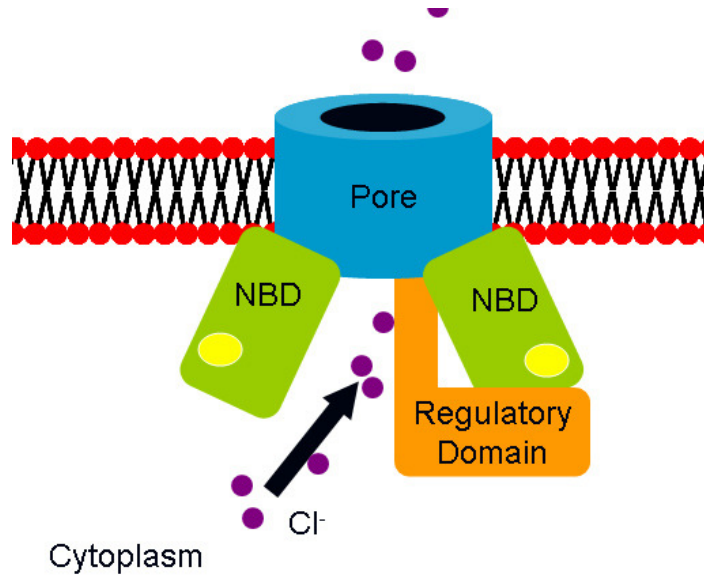


Figure 1.1 CFTR model

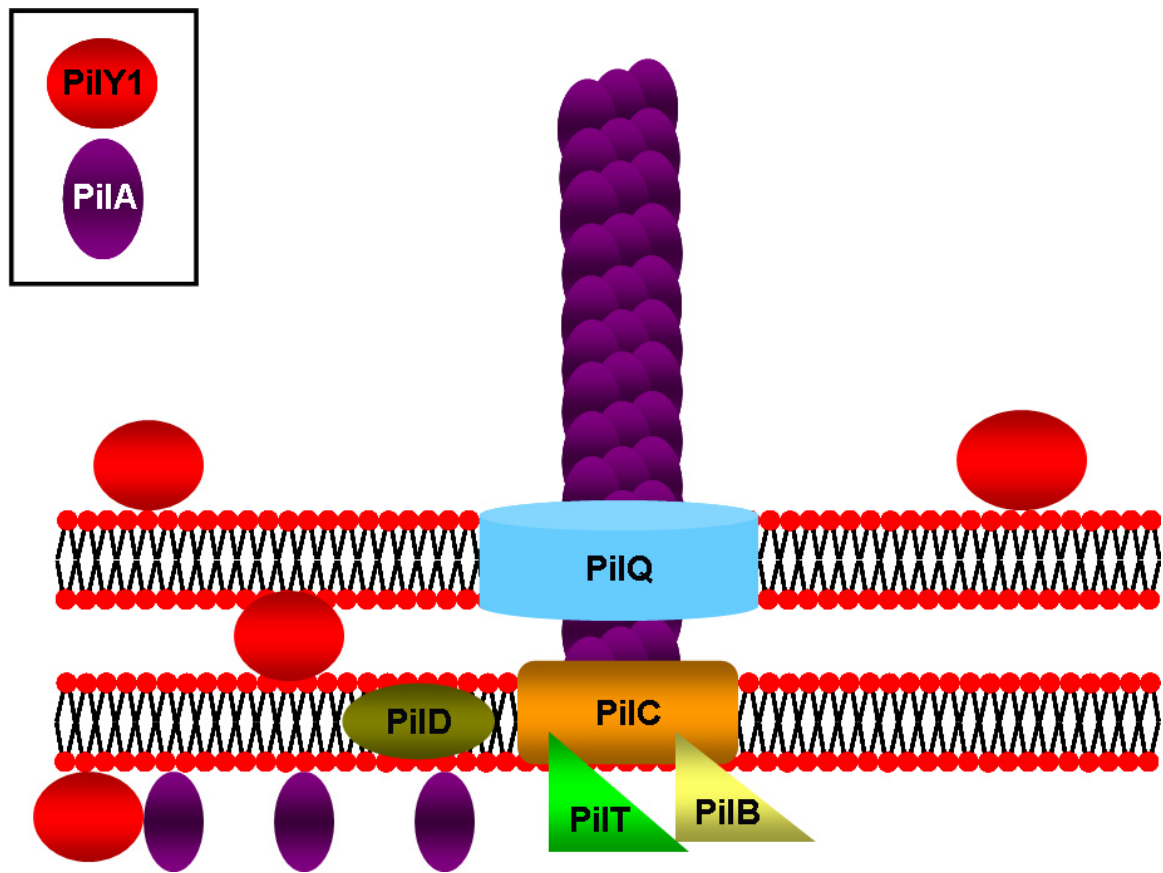


Figure 1.2 Type IV pilus structure model

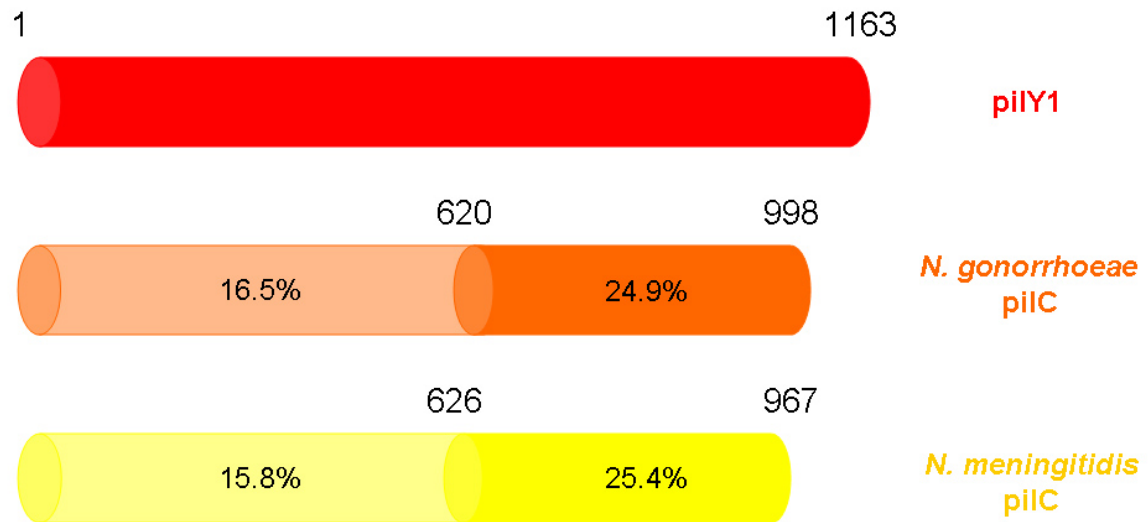


Figure 1.3 Sequence homology of PilY1 to *Neisseria* PilC proteins

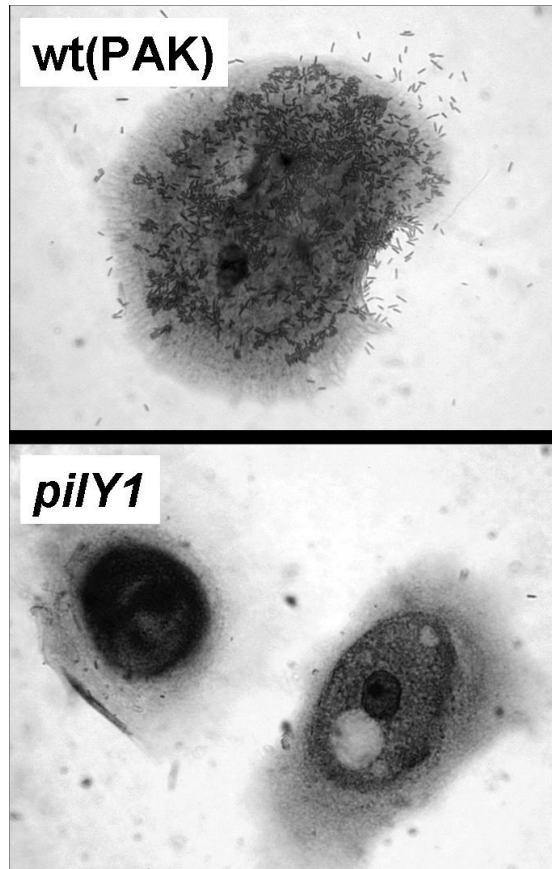


Figure 1.4 PilY1 mediates bacterial adherence to human corneal cells

Chapter 2:

Cloning, Expression, and Crystallization of C-terminal pilY1

INTRODUCTION

We acquired genomic DNA from the PAK strain of *Pseudomonas aeruginosa* for amplification of the *pilY1* gene through a collaboration with Dr. Matthew Wolfgang, a professor of Microbiology and Immunology at UNC-Chapel Hill. This chapter outlines the experimental procedures used to clone the gene, produce C-terminal PilY1 protein, crystallize the protein, and determine the three-dimensional structure. Specifically, this chapter will include construct design through the use of structural prediction servers, cloning of the gene, expression, purification and crystallization of the protein, as well as additional biophysical experiments required to characterize the PilY1 protein. It will also detail the process used to solve the crystal structure, including selenomethionine substitution of the protein, data collection and reduction, and the data processing used in determination of the final structure. Of particular interest is the successful introduction of three additional methionine residues into the C-terminal region of PilY1 such that multiwavelength anomalous dispersion methods using selenomethionine-substituted protein could be employed to determine the structure. The native PilY1 C-terminal domain contained only five methionine residues, which was not enough to obtain phases for 548 residues. A total of eight was required and achieved.

CLONING AND EXPRESSION OF PILY1

Structural prediction for Pseudomonas aeruginosa PilY1

The full length PilY1 sequence (residues 1-1163) was submitted to the program Phyre [1] for structural analysis. Phyre, a successor to 3D-PSSM [1], is a protein fold recognition server that attempts to predict a protein's three-dimensional structure and probable function based on a database of proteins with known structure. The quality of the matches was determined by the expectation value (E-value), which describes the number of alignments with the same or better E-value that would occur by chance during a database search. Therefore, a lower E-value corresponds to a more confident prediction. The sequence was also submitted to the server in two parts, the N-terminal sequence (residues 1-678) and the C-terminal sequence (residues 615-1163), chosen based on sequence alignment with the *Neisseria* PilC proteins. The output from the analysis consisted of prediction of likely disordered regions, secondary structural elements, and alignment to homologous proteins of known structure.

Analysis of the N-terminal region of pilY1 yielded no significant matches. The closest structure predicted to have homology to this region of the protein scored an E-value of 23 and a precision estimate of 5%. The precision estimate is the percentage of sequences with an E-value of equal or lesser value that were actually correct homologues; therefore, this was not considered to be significant. However, analysis of the C-terminal region of the protein provided several significant matches. This domain was predicted to be structurally homologous to several proteins with known β -propeller folds, such as those of the alcohol

dehydrogenase family. The top match was the quinohemoprotein alcohol dehydrogenase from *Comamonas testosteroni*, with an E-value of 0.014 and 95 % certainty.

The alcohol dehydrogenase (PDB ID 1KB0) identified by Phyre as a homologous structure is an 8-bladed β -propeller protein. Residues 678-1114 of the C-terminal region encompass the β -strands that comprise the predicted β -propeller. The C-terminal construct of PilY1 used in these studies includes all but one beta strand in the first expected β -propeller blade and one β -strand in the seventh expected blade. The eighth blade is not predicted in PilY1. In addition, almost all probable regions of disorder for this domain of the protein were located in between the predicted β -strands.

Cloning of pilY1 gene

The *pilY1* gene (residues 1-1163) was amplified from PAK strain genomic DNA using the following primers: FWD 5' CATGCCATGGATGAAATCGGTACTCCACCA GATCG 3'; REV 5' ATAGTTTAGCGGCCGCGTTCTTTCCTTCGATGGGGCGCCA 3' (NACF, Lineberger Comprehensive Cancer Center, UNC-CH). These primers contained NcoI and NotI enzymatic restriction sites, respectively, which were used for cloning into the pET28b vector (Novagen, San Diego, CA). The gene was amplified using a combination of two specialized PCR systems, the GC Rich PCR System (Roche, Indianapolis, IN) and the Expand Long Template PCR System (Roche, Indianapolis, IN). The PCR reaction contained 10x Buffer 1 and DNA polymerase from the Expand Long Template PCR System, as well as Resolution Solution from the GC Rich PCR System.

Various N- and C-terminal constructs of the PilY1 gene were also cloned into protein expression vectors. The gene was divided into N- and C-terminal halves using the sequence homology to *Neisseria gonorrhoeae* PilC noted by Alm *et al* upon identification of the *pilY1* gene [2]. Three new constructs were created: the so-called N- and C-terminal constructs, comprising residues 1-614 and 615-1163, respectively, as well as a modified N-terminal construct that started at residue 31 to omit a predicted signal sequence. These constructs were amplified as described above and are summarized in Table 2.1. Constructs were cloned into both a pET28 vector (N- or C-terminally His-tagged) and a pMALC2xT vector (New England Biolabs, Ipswich, MA) modified with a cleavable His-tag (gift of John Tesmer, University of Michigan).

Test Expressions of pilY1 constructs

Constructs were test-expressed in two different *E. coli* competent cell lines, BL21(DE3) Gold and Origami(DE3) cells (Novagen). The cultures were grown in Luria Broth (LB), induced with either 0.1 mM or 1 mM IPTG at an OD₆₀₀ of 0.3-0.5, and shaken at either 15 or 37°C at 210 RPM. Samples from each culture were taken pre-induction and post-induction at 2, 4, and 6-hr time points, as well as after overnight growth.

SDS-PAGE analysis of protein expression showed the two C-terminal constructs of the *pilY1* gene as the only constructs capable of expressing soluble protein. This observation applied to all test conditions. Growth at 15°C with 0.1 mM IPTG induction yielded slightly more soluble protein than other conditions. The amount of protein produced by each of the

constructs was nearly equivalent; for ease of purification, we chose to use the construct in the pET28a vector.

Expression and Purification of C-terminal PilY1

The C-terminal fragment of *pilY1* in pET28a was transformed into BL21(DE3) Gold competent cells (Novagen). A liter of LB was inoculated with 10 mL overnight culture and grown at 37°C to an OD₆₀₀ of 0.4. It was then induced with IPTG to a final concentration of 0.1 mM, the temperature was dropped to 15°C, and expression was allowed to proceed overnight. Cells were collected by gentle centrifugation at 1600g for 20 minutes at 4°C. The resulting pellet was resuspended in 20 mL of Buffer A (20 mM Tris-HCl pH 7.8, 5 mM imidazole, 50 mM NaCl, 5% glycerol), and protease inhibitors were added (0.025 mg each aprotinin and leupeptin, 0.5 mg PMSF). Cells were lysed by sonication and clarified by centrifugation at 40,000 g for 75 minutes at 4°C.

The clarified lysate was loaded onto a 10 mL Probond Ni-affinity column (Invitrogen, Carlsbad, CA). The column was washed twice with 100 mL each of Buffers B and C (same components as Buffer A but with 25 and 50 mM imidazole, respectively). The protein was eluted with Buffer D (20 mM Tris-HCl pH 7.8, 75 mM imidazole, 250 mM NaCl, 5% (v/v) glycerol). Fractions were analyzed by SDS-PAGE and concentrated to ~5 mg/mL by centrifugation at 4°C, using an Amicon Ultra-15 concentrator with 30K molecular weight cutoff (Millipore, Billerica, MA). The concentrated protein was then further purified using a Superose 12 size-exclusion column (Amersham Biosciences, Piscataway, NJ) on an AKTA FPLC system (Pharmacia, Uppsala, Sweden), which also facilitated buffer exchange into

Buffer A. The fractions were collected and analyzed by SDS-PAGE, with an estimated purity of >95%. The protein was concentrated to 10 mg/mL in the Amicon Ultra-15 and stored at 4°C, or flash-frozen for long-term storage at -80°C.

It should be noted that both SDS-PAGE analysis and the sizing column profile indicated that the protein was forming an oligomer, presumably a dimer. The addition of DTT as a reducing agent appeared to force the protein into a monomer state as assessed by SDS-PAGE, which was indicative of potential disulfide bond formation. This led to the addition of a reducing agent to the protein before crystallization experiments. After concentration, TCEP-HCl (Pierce, Rockford, IL) was added to a final concentration of 8 mM.

CRYSTALLIZATION OF PILY1 C-TERMINAL DOMAIN

Initial Crystallization

Initial crystal hits were obtained at the high-throughput crystallization laboratory at the Hauptman-Woodward Institute in Buffalo, NY. This facility utilizes microassay plates to screen over 1500 in-house and commercially available conditions using only 0.4 µL of protein (10 mg/mL) per drop; the drops are incubated at room temperature under paraffin oil and monitored at regular intervals for one month. After one day, needle crystals of C-terminal PilY1 were visible in over 60 crystallization conditions. We were able to reproduce twelve of these initial hits at UNC using the hanging drop vapor diffusion method. Subsequent improvement of crystal size and quality was achieved by lowering the incubation temperature to 15°C and by increasing the protein/crystallant ratio to 2:1. The condition that

produced the most favorable looking crystals contained 0.3 M sodium malonate (pH 6.5) and 20% PEG 3350. Use of the Additive Screen (Hampton Research, Aliso Viejo, CA) revealed that the addition of 50 mM DTT, 30-100 mM $(\text{NH}_4)_3\text{PO}_4$ and 4-8% trifluoroethanol was also beneficial to crystallization. Rectangular-shaped crystals with dimensions of 150x50x50 μm appeared overnight; size was improved significantly (up to 10 μm in each direction) when the crystals were allowed to grow for a week or more.

Crystal Screening and Data Collection

The crystals were harvested after 1-2 weeks, cryoprotected by short dips in perfluoroether oil, and screened in-house at the UNC Biomolecular X-ray Crystallography Core Facility, using an RU-type Cu-K α rotating anode generator and an R-axis IV++ image plate detector (Rigaku, Houston, TX). Crystals diffracted to ~ 3.0 Å and could be readily indexed and scaled using HKL2000 [3].

Diffraction data were collected at the Advanced Photon Source at Argonne National Laboratory on beamline 22-BM (SER-CAT). The protein crystallized in the $\text{P}2_12_12_1$ space group and diffracted to a resolution of 2.1 Å. The data were indexed, integrated and scaled using HKL2000 [3]. Data reduction statistics are summarized in Table 2.2.

PHASE DETERMINATION

Heavy Atom Soaks

As no closely related homologous structure was known, we attempted to derivatize the crystal by soaking in various heavy atom solutions. The atoms used were mainly Pt and Hg compounds, all from the Hampton Heavy Atom Screen (Hampton Research, Aliso Viejo, CA). The specific compounds, concentrations, and soaking conditions are outlined in Table 2.3. Compounds were used at 1, 5, 10, and 50 mM concentrations and soaked for 5, 30, and 60 minute intervals, as well as overnight. Crystals were backsoaked in mother liquor for 5, 30, and 60 minutes, and overnight. The heavy atoms soaks all proved unsuccessful; although several compounds provided an acceptable fluorescence signal (indicative of presence in the crystal), no heavy atom sites could be located using customary phasing methods (SOLVE [4], SHARP [5], SHELXD/E [6], and Rantan [7]).

Selenomethionine derivitization

After failing to produce a heavy atom derivative through soaking, we decided to attempt derivitization of the protein with selenomethionine. The average protein contains one methionine (Met) residue for every 59 amino acids; this is sufficient to provide a usable anomalous signal at the appropriate wavelength [8]. Inspection of the protein sequence showed that the PilY1 C-terminal construct contained five Met residues out of a total of 548, or 1 in ~110, half as many methionines as the average protein. We suspected that there were not enough Met residues in the protein to obtain successful phasing information; nevertheless, we decided to carry out the experiment.

C-terminal *pilY1* in pET28a was transformed into *E. coli* B834(DE3) competent cells (Novagen). B834 cells are derived from a methionine auxotroph strain; they are unable to

produce their own methionine and therefore cannot incorporate the residue into the nascent protein chain without supplemented methionine in the growth media. Therefore, addition of selenomethionine to the media as the only source of the Met residue results in the incorporation of SeMet instead of Met into the protein. A 100 mL overnight culture grown from a single colony in LB was centrifuged at 40,000 g for 15 minutes, and the pellet was washed twice with 50 mL of ddH₂O. This step is critical for removal of any residual methionine from the media used in the overnight culture. The washed pellet was resuspended in 1 mL ddH₂O. The cells were then ready for inoculation.

To produce the selenomethionine protein, cells were grown in SeMet Medium Base minimal media and SelenoMet Nutrient Mix (Athena Enzyme Systems, Baltimore, MD). The medium base (21.6 g) was added to 1 L of ddH₂O and autoclaved. After cooling, 5.1 g of nutrient mix was dissolved in 50 mL ddH₂O; this solution was sterilely filtered into the medium base. After adding antibiotic and supplementing the media with 50-60 mg selenomethionine, the 1 mL sample of washed and resuspended overnight culture was added. The culture was grown at 37°C to an OD₆₀₀ of 0.4-0.6, at which time it was induced with 0.1 mM IPTG and left to shake at 15°C overnight. The protein was purified and crystallized as described for the native pilY1.

Crystals of SeMet PilY1 were similar in appearance and diffraction quality to the native crystals. Although the crystals diffracted well enough to collect useful data, unfortunately the fluorescence scan did not provide a usable signal (Figure 2.1). This observation led us to believe that the protein did not contain enough methionine residues for phasing. Mass spectrometry analysis (Figure 2.2) confirmed that the protein was 100%

substituted with selenomethionine residues in place of methionine; therefore, we decided to attempt introduction of additional methionine residues into the protein.

Our initial strategy to increase the number of Met residues entailed using sequence alignments to find leucine residues in *Pseudomonas aeruginosa* PilY1 which were aligned to methionine residues in homologous proteins (i.e. PilC from *Neisseria gonorrhoeae* and *meningitidis*). Leucine is the amino acid most commonly used in methionine replacement. After comparing many closely related protein structures, it was determined that leucine was the only “safe” residue to use for methionine substitution. Safe residues were those deemed least likely to perturb local protein structure and the overall folding pathway [9].

Using the sequence alignment, we selected nine Leu residues that aligned with Met residues in the homologous *Neisseria* proteins (L690, L715, L748, L752, L773, L870, L908, L990, and L1042). These residues were mutated using the QuikChange Site-Directed Mutagenesis Kit (Stratagene, La Jolla, CA) as instructed by the manual. None of these mutant proteins were soluble when expressed in the manner of wild-type PilY1. This began an arduous process in which we selected many leucine residues to mutate, no longer based on any sequence homology, and found very few mutant proteins with soluble expression. In addition, some of the stable mutant proteins did not crystallize; of the ones that did, some would not crystallize upon selenomethionine substitution. A summary of the mutant proteins and the behavior they exhibited is shown in Table 2.4. Possible explanations as to why many of the Leu to Met mutations were so detrimental to the stability of the protein will be discussed in the following chapter.

Of the 22 mutants tested, we found a combination of 3 additional methionines (L→M 712, 812, and 823) which yielded soluble protein and crystallized as selenomethionine

substituted protein. In this manner we were able to solve the crystal structure of C-terminal PilY1.

Phasing of C-terminal PilY1 using SeMet SAD

The three dimensional structure of the C-terminal domain of PilY1 was determined through the use of selenomethionine single wavelength anomalous dispersion methods (SAD). Crystals with three additional Met residues were grown from SeMet substituted protein as described above. Crystals were similar in appearance to the native crystals; some of them were larger and displayed sharper and cleaner faces and edges.

Diffraction data were collected at the Advanced Photon Source at Argonne National Laboratory, beamline 19-ID (SBC-CAT). The protein crystallized in the space group C2 and diffracted to 2.8 Å. A fluorescence scan at the selenium peak wavelength showed excellent anomalous signal (Figure 2.3). Data were collected at one peak wavelength (0.979 Å) for 360° in 0.5° wedges, with little radiation decay observed. Data reduction, heavy atom search, and phase determination were all carried out using HKL3000 [10]. It is worthwhile to discuss HKL3000. This program is the successor to HKL2000 and not yet commercially available- by sheer luck we were granted access through a chance meeting with the authors. HKL3000 has been used to take raw diffraction images through to a 70% complete model in the span of 10 minutes (provided the resolution is 2.3 Å or better). The program combines several available programs to accomplish this feat. Data indexing, integration, and scaling is carried out with HKL2000; a heavy atom search is done with SHELXD/E; phasing and initial

phase improvement is carried out with MLPHARE [11], DM [11], FINDNCS [12], and SOLVE/RESOLVE; finally, preliminary model building is done with ARP/wARP [13]

The protein crystallized as a dimer in the asymmetric unit. HKL3000 was able to locate 14 out of 18 selenium atoms in the asymmetric unit; two of the missing atoms were within the first residue of the construct and two were in a disordered region that could not be modeled. ARP/wARP correctly built ~30% of the initial model, and further model building was done using COOT [14]. Once 80% of the residues were built, the structure and native phases were submitted to MOLREP for molecular replacement [11]. Refinement of the native structure was done using maximum likelihood restraint methods with CCP4 [11] and CNS [15]. Data reduction, phasing, and refinement statistics are summarized in Table 2.5. Analysis of the structure is discussed in greater detail in the following chapter.

REFERENCES

1. Kelley, L.A., R.M. MacCallum, and M.J. Sternberg, *Enhanced genome annotation using structural profiles in the program 3D-PSSM*. J Mol Biol, 2000. **299**(2): p. 499-520.
2. Alm, R.A., et al., *Fimbrial biogenesis genes of Pseudomonas aeruginosa: pilW and pilX increase the similarity of type 4 fimbriae to the GSP protein-secretion systems and pilY1 encodes a gonococcal PilC homologue*. Mol Microbiol, 1996. **22**(1): p. 161-73.
3. Otwinowski, Z. and W. Minor, *Processing of X-ray diffraction data collected in oscillation mode*. Methods in Enzymology. Vol. 276. 1997. 307-326.
4. Terwilliger, T.C. and J. Berendzen, *Automated MAD and MIR structure solution*. Acta Crystallogr D Biol Crystallogr, 1999. **55**(Pt 4): p. 849-61.
5. Bricogne, G., et al., *Generation, representation and flow of phase information in structure determination: recent developments in and around SHARP 2.0*. Acta Crystallogr D Biol Crystallogr, 2003. **59**(Pt 11): p. 2023-30.
6. Sheldrick, G. and T. Schneider, *SHELXL: High-resolution refinement*. Methods in Enzymology. Vol. 277. 1997. 319-343.
7. Jia-Xing, Y., *On the application of phase relationships to complex structures. XVIII. RANTAN-random MULTAN*. Acta Crystallogr A, 1981. **A37**: p. 642-644.
8. Hendrickson, W.A., J.R. Horton, and D.M. LeMaster, *Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): a vehicle for direct determination of three-dimensional structure*. Embo J, 1990. **9**(5): p. 1665-72.
9. Bordo, D. and P. Argos, *Suggestions for "safe" residue substitutions in site-directed mutagenesis*. J Mol Biol, 1991. **217**(4): p. 721-9.
10. Minor, W., et al., *HKL-3000: the integration of data reduction and structure solution-from diffraction images to an initial model in minutes*. Acta Crystallogr D Biol Crystallogr, 2006. **62**(Pt 8): p. 859-66.
11. *The CCP4 suite: programs for protein crystallography*. Acta Crystallogr D Biol Crystallogr, 1994. **50**(Pt 5): p. 760-3.
12. Lu, G., *FINDNCS: A program to detect non-crystallographic symmetries in protein crystals from heavy atoms sites*. Journal of Applied Crystallography, 1999. **32**: p. 365.

13. Lamzin, V.S. and K.S. Wilson, *Automated refinement of protein models*. Acta Crystallogr D Biol Crystallogr, 1993. **49**(Pt 1): p. 129-47.
14. Emsley, P. and K. Cowtan, *Coot: model-building tools for molecular graphics*. Acta Crystallogr D Biol Crystallogr, 2004. **60**(Pt 12 Pt 1): p. 2126-32.
15. Brunger, A.T., et al., *Crystallography & NMR system: A new software suite for macromolecular structure determination*. Acta Crystallogr D Biol Crystallogr, 1998. **54**(Pt 5): p. 905-21.

FIGURE LEGENDS

Figure 2.1 The fluorescence scan (left) shows a weak peak at the selenium absorbance wavelength, while the scattering factor analysis does not provide usable information for determination of SAD x-ray collection wavelengths.

Figure 2.2 The mass spectrometry analysis shows that selenomethionine is fully substituted for methionine in the protein used for crystallization.

Figure 2.3 This scan is from the crystals that were used to obtain phasing information. The protein used for crystallization had an additional three methionine residues in comparison to the wild-type PiY1. Note the strong peak at the characteristic selenium absorption wavelength, and the clear peaks in the scattering factor plot, which provided a wavelength for data collection with the maximum anomalous signal.

Table 2.1 Summary of pilY1 constructs created for test expression

Construct	Vector	Restriction Sites	Primers (5'-3')
1-612	pET28b	NcoI/NotI	F:CATGCCATGGATGAAATCGGTACTCCACCAGATC R:TTTTCCTTTTGCGGCCGCGCGGTCTGGCCCTTGGT
1-612	pMALC2xT	EcoRI/XbaI	F:GGAATTCATGAAATCGGTACTCCACCAGATC R:CTAGTCTAGATTATTAGCGGTCCTGGCCCTTGGT
31-612	pET28b	NcoI/NotI	F:AAGGAGATATAACCATGGGCCGCCCTGTCGGTCAGC R:GCTGACCGACAGGGCGGCCCATGGTATATCTCCTT
31-612	pMALC2xT	EcoRI/XbaI	F:GGAATTCATGGGCCGCCCTGTCGGTCAGC R:CTAGTCTAGATTATTAGCGGTCCTGGCCCTTGGT
613-1163	pET28a	NdeI/NotI	F:GGAATTCCATATGGTGGCCTTCCTGCGCGGCGA R:ATAGTTTAGCGGCCGCTTATTAGTTCTTTCCTTCGA TGGGGCG
613-1163	pMALC2xT	EcoRI/XbaI	F:CCGGAATTCGTGGCCTTCCTGCGCGGCGA R:CTAGTCTAGATTATTAGTTCTTTCCTTCGATGGGGCGC CA

Table 2.2 Native Data Reduction Statistics

Summary of Data Collection, Phasing and Refinement	
PilY1 C-terminal Domain	
Data Set	Native
Unit Cell [abc, Å; $\alpha\beta\gamma$, °]	a=64.4, b=108.3, c=159.0 $\alpha,\gamma,\beta=90$
Space Group	P2 ₁ 2 ₁ 2 ₁
Wavelength [Å]	1.0718
X-ray Source	APS SER-CAT 22-BM
Resolution [Å] (Highest shell)	50-2.10 (2.18-2.10)
No. of Reflections [unique]	65477
Completeness [%]	99.7 (98.8)
R _{sym} [%]	8.5 (43.5)
I/ σ	39.7 (6.7)
Redundancy	13.9 (13.3)

Table 2.3 Heavy Atom Compounds used in crystal soaking experiments

Compound	Crystal Reaction Notes	Positive Fluorescence Scan?
K ₂ PtCl ₄	Dissolves in > 5mM	Yes
(NH ₄) ₂ PtCl ₄		Yes
K ₂ PtCl ₆		Yes
K ₂ Pt(NO ₂) ₄	Dissolves in > 5mM	Yes
K ₂ Pt(CN) ₄ ·XH ₂ O	OK in 10 mM	Yes
PtCl ₂ (H ₂ NCH ₂ CH ₂ CH ₂ NH ₂)		Yes
K ₂ PtBr ₆		Yes
K ₂ Pt(SCN) ₆		N/A
K ₂ PtI ₆		N/A
(C ₂ H ₅ HgO)HPO ₂	Dissolves in > 5mM	No
HgCl ₂		Yes
Hg(OOCCH ₃) ₃	Dissolves in > 1mM	Yes
Ethylmercurithiosalicylic acid		N/A
C ₈ H ₈ HgO ₂		N/A
HgBr ₂		N/A
HgI ₂		N/A
Hg(CN) ₂		Yes
CH ₃ HgBr		N/A
CH ₃ HgCl	OK in 10 mM	Yes
KAu(CN) ₂	Dissolves in > 5mM	No
NaAuCl ₄	Dissolves in > 5mM	No
HAuCl ₄		No
KAuBr ₄		No
TiCl ₃	OK in 10 mM	No
Pb(CH ₃ COO) ₂	Dissolves in > 1mM	N/A
Pb(NO ₃) ₂		N/A
PbCl ₂		N/A
(C ₂ H ₅) ₃ Pb(CH ₃ COO) ₂		N/A
K ₂ IrCl ₆		Yes
IrCl ₃		No
(NH ₄) ₃ IrCl ₆		No
K ₂ OsO ₄	OK in 10 mM	Yes
(NH ₄) ₂ OsBr ₆	OK in 10 mM	No
K ₂ OsCl ₆	Dissolves in > 5mM	No
OsCl ₃	Dissolves in > 5mM	No
TaBr ₆	OK in 10 mM	Yes

Table 2.4 Summary of Leucine to Methione Mutations and Results

Leu to Met	Express solubly?	Crystallize?	Crystallize with SeMet?
690	No	N/A	N/A
715	Yes	No	N/A
748	No	N/A	N/A
752	No	N/A	N/A
773	No	N/A	N/A
870	No	N/A	N/A
908	No	N/A	N/A
990	No	N/A	N/A
1042	No	N/A	N/A
712	Yes	Yes	Yes
812	Yes	Yes	Yes
823	Yes	Yes	Yes
762	Yes	Yes	No
866	No	N/A	N/A
1076	No	N/A	N/A
657	Yes	Yes	Yes
819	No	N/A	N/A
832	Yes	Yes	Yes
1029	No	N/A	N/A
1120	No	N/A	N/A
1123	No	N/A	N/A
1142	No	N/A	N/A

Table 2.5 Data reduction, phasing and refinement statistics

PilY1 C-terminal Domain		
Data Set	Native	SeMet
Unit Cell [abc, Å; αβγ, °]	a=64.4, b=108.3, c=159.0 α,γ,β=90	a=169.8, b=64.9, c=116.6 α=γ=90, β=108.7
Space Group	P2 ₁ 2 ₁ 2 ₁	C2
Wavelength [Å]	1.0718	0.9792
X-ray Source	APS SER-CAT 22-BM	APS SBC-CAT 19-ID
Resolution [Å] (Highest shell)	50-2.10 (2.18-2.10)	50-2.80 (2.90-2.80)
No. of Reflections [unique]	65477	29869
Completeness [%]	99.7 (98.8)	99.9 (100.0)
R _{sym} ^a [%]	8.5 (43.5)	7.1 (24.4)
I/σ	39.7 (6.7)	48.6 (9.8)
Redundancy	13.9 (13.3)	7.8 (7.9)
MAD Phasing Statistics to 2.8 Å		
Se atom sites in AU (observed/total)		14/18
Mean Figure of Merit		
MLPHARE (Highest shell)		0.65 (0.60)
After DM (Highest shell)		0.82 (0.41)
Refinement Summary		
Resolution [Å] (Highest shell)	50-2.10 (2.18-2.10)	
Molecules per asymmetric unit (AU)	2	
No. of protein atoms/AU	7361	
No. of waters/AU	818	
R _{working} ^b [%]	19.6 (23.7)	
R _{free} ^c [%]	23.5 (28.1)	
Average B factor [Å ²]		
Protein	29.5	
Water	42.3	
RMS deviations		
Bonds [Å]	0.005	
Angles [°]	1.37	
Coordinate error estimates [Å]		
Luzzati	0.23	
SigmaA	0.15	
Ramachandran [%]		
Allowed, generous, disallowed	98.6, 1.0, 0.4	
PDB ID code		

^a R_{sym} = Σ|I - <I>| / ΣI, where I is the observed intensity and <I> is the average intensity of several symmetry-related observations.

^b R_{working} = Σ||F_o - F_c|| / ΣF_o, where F_o and F_c are the observed and calculated structure factors, respectively.

^c R_{free} = Σ||F_o - F_c|| / ΣF_o for 5% of the data not used at any stage of the structural refinement.

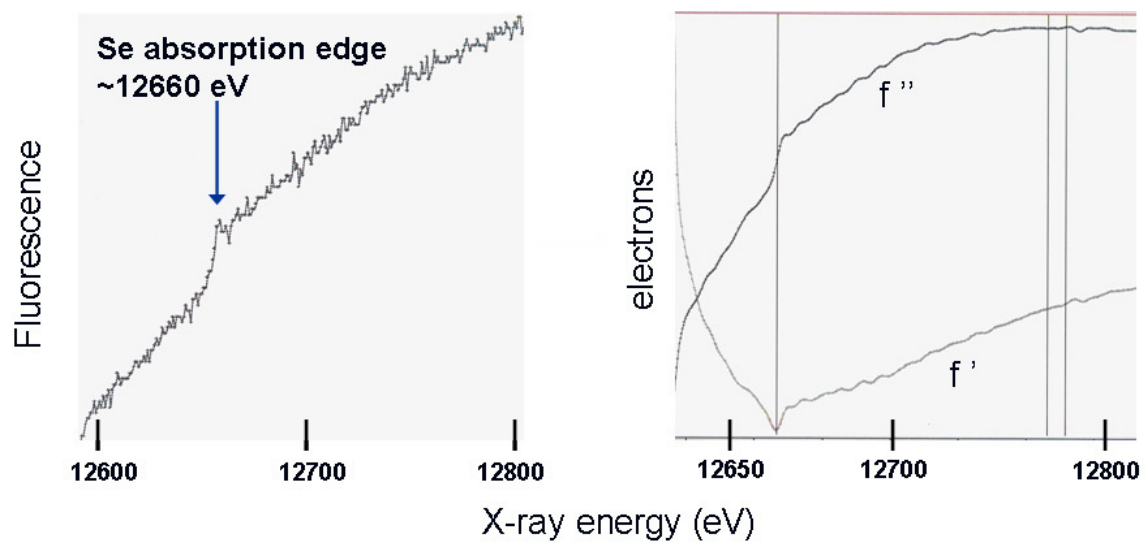


Figure 2.1 Fluorescence scan and scattering factor analysis of wild-type PilY1

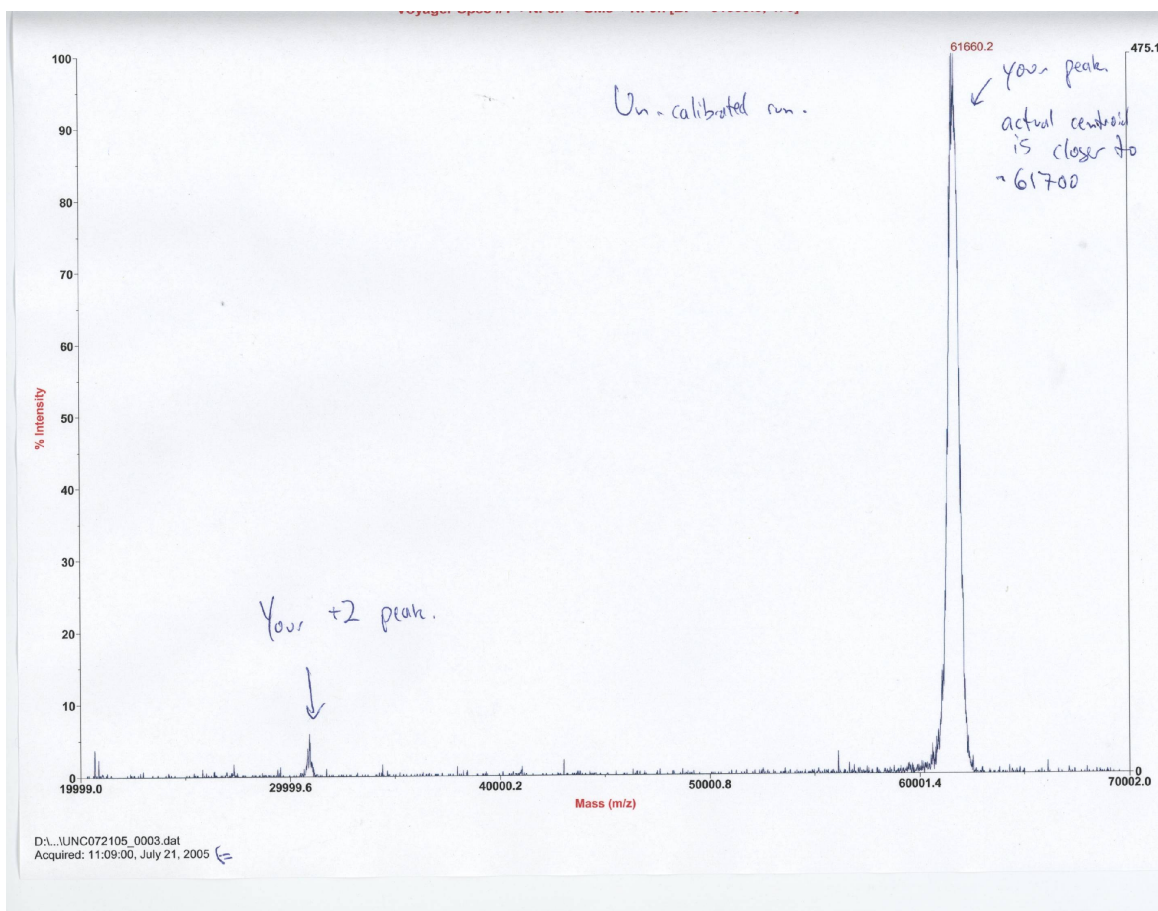


Figure 2.2 ESI Mass Spectrometry analysis of Selenomethionine substituted PilY1

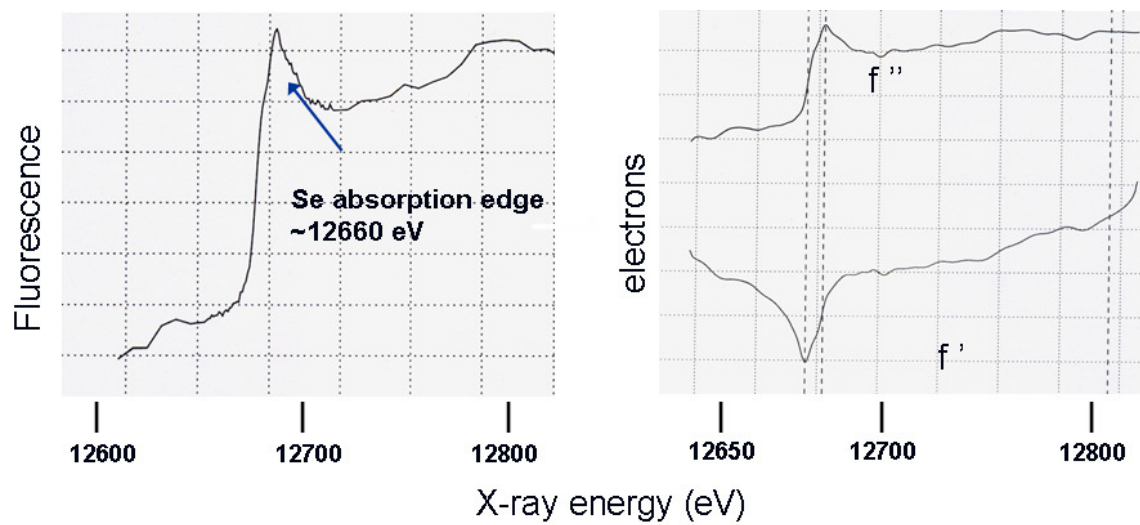


Figure 2.3 Fluorescence scan and scattering factor analysis of triple methionine mutant PilY1

Chapter 3:

Structural Analysis of PilY1 C-terminal domain

OVERALL STRUCTURE OF PILY1 C-TERMINAL DOMAIN

The overall structure of the PilY1 C-terminal domain is a modified β -propeller. Indeed, this overall fold was the structural motif predicted by the bioinformatics program Phyre [1]. The PilY1 C-terminal domain (residues 615-1163) crystallized as a dimer in the asymmetric unit, both for the selenomethionine substituted protein (space group C2) and the native protein (space group $P2_12_12_1$). This shift in space group from the native to the derivative crystal entails a reordering of cell constants (a, b, and c become c, a, and b) as well as an increase in the β angle (90° to $\sim 109^\circ$). Additionally, the lengths of the a, b and c cell constants increased, from 64.4 to 64.9, 108.3 to 116.6, and 159.0 to 169.9 Å, with respect to the native structure. We can hypothesize that the addition of selenium to the methionine residues caused a reorganization of the unit cell which required this increase in both the lengths of the cell constants and in one of the angles. Further evidence of this hypothesis is provided through generation of the symmetry mates of the two structures in their respective unit cells, which show a more tightly packed arrangement of protein molecules in the native crystal.

Both the monomer and the dimer are shown as color-coded cartoons in Figures 3.1 and 3.2. Each monomer is composed of a 7-bladed β -propeller surrounding a central pore, with 9 α -helices and 2 additional β -strands interspersed throughout the structure. The first four blades of the propeller each contain four anti-parallel β -strands, but beyond this the structure becomes more complex than the canonical β -propeller structures elucidated previously. For example, the fifth blade of the pilY1 C-terminal monomer contains five β -strands; however they are not sequentially contiguous. After the first three strands of Blade 5 comes the fifth strand, which also serves as the first strand in the sixth propeller blade. This particular β -strand is thirteen residues long and thus is able to twist in such a way as to integrate itself into two adjacent propeller blades (Figure 3.3). This strand leads into the next strand of Blade 6, followed by strand four of Blade 5, which completes this blade for a total of five strands. This is followed by the five strands of Blade 6 (for a total of six strands), and finally the three-stranded anti-parallel propeller blade 7. This complex structural arrangement is illustrated in the topology diagram in Figure 3.4.

In spite of Phyre's overall success in predicting the general fold of the pilY1 C-terminal domain, the details of this novel structure were difficult for this suite of programs to identify. For example, a structural alignment of the pilY1 C-terminal domain with the quinoxaline alcohol dehydrogenase (QAD) from *Comamonas testosteroni*, a protein predicted to be homologous by Phyre, superimposes poorly, with a C α rms deviation of 15.1 Å despite the same structural motif (Figure 3.5). This is due to the presence of β -strands of various lengths within the same blade found in the pilY1 structure. While the QAD propeller possesses a set of highly symmetrical blades with four β -strands in each, the pilY1 C-terminal domain has extremely variable blades (with 3, 4, 5, and 6 strands) and considerably

longer β -strands. This lack of symmetry is especially evident in strand 5 of the fifth blade, which as stated previously is a part of two distinct propeller blades and is located far on the outside of the canonical motif exemplified by the dehydrogenase. Hence, although the first four blades of pilY1 superimpose well with QAD, the more non-traditional blades (Blade 5 and beyond) are not capable of alignment.

Each monomer is missing two small stretches of amino acids that could not be placed into the electron density. In the N-terminal region of the model, monomer 1 (chain A) is missing residues 714-720, while monomer 2 (chain B) is missing residues 712-722. Interestingly, this area of the protein was identified as potentially unstructured and susceptible to proteolysis through the use of a chymotrypsin digest, which revealed a proteolytic site at residue 721. Briefly, the C-terminal region of pilY1 (61 kD) was subject to limited proteolysis using both trypsin and chymotrypsin from bovine pancreas (Sigma). Varying amounts (0.5-2 μ L) of protease at 40 μ M were added to 20 μ L of 16 μ M C-terminal pilY1 in 0.1 M $(\text{NH}_4)\text{HCO}_3$. The reactions were carried out at room temperature for 15 minutes, quenched with 2 μ L of 50 mM PMSF (Sigma), and analyzed by SDS-PAGE. The reaction containing the greatest amount of chymotrypsin produced a single band with a molecular weight of ~47 kD and was subsequently analyzed by electrospray ionization mass spectrometry (WM Keck Facility, Yale University). The mass spectrum showed a prominent peak at 46.5 kDa. The cleavage site was located using this mass, knowledge of standard chymotrypsin cleavage sites (Phe, Tyr, and Trp residues), and the protein analysis software PAWS (Protein Analysis WorkSheet, Genomic Solutions, AnnArbor, MI). PAWS is a program that facilitates identification of proteins and cleavage sites by highlighting potential proteolysis products using the protein sequence and results of mass spectrometry analysis. In

this manner we were able to determine the region of the protein found in the mass spectrum and hence the cleavage site. In addition, a second region of missing amino acids was located from residues 1062-1068 in both monomers. Aside from these, the structure is contiguous, missing only residues from both the N- and C-termini: residues 615-643 and residues 1149-1163 cannot be placed in the model.

STRUCTURAL CHARACTERISTICS

β -propeller structural motifs

β -propeller proteins are characterized by extreme sequence diversity and function despite similar overall three-dimensional structures [2]. The modular building block of these proteins is a four-stranded β -sheet repeated 4-8 times around a central axis [2]. Modular β -sheets from distinct structures of β -propeller proteins superimpose with rms deviations from 0.9-2.9 Å [3-8]. Due to the high sequence divergence among these proteins, alignment is only possible through superposition of the structures. In addition, β -propeller proteins have a wide range of biological functions and exist in organisms with distinct phylogenetic origins [2]. β -propeller proteins have been implicated in many diseases, such as cancer, Alzheimer's, Huntington's, and osteogenesis, and have different functions that include catalysis, scaffold formation, signaling, ligand binding and transport, and mediation of protein-protein interactions [9]. Most known β -propeller domains bind ligands or interact with other proteins or molecules through the use of the central cavity, found in the center of all of the propeller blades [4, 10]. Binding can also occur through variable loops found on

the outside of the propeller. The sequence and length diversity found in both the loops and strands of these proteins allow for a stable scaffold with loops that create versatile binding surfaces [9]. The only aspect of the β -propeller protein that is consistent among the various known structures is the presence of hydrophobic residues located in the central part of the β -strands that make up the modular sheets. These residues are important for folding and packing in propeller assembly [9].

WD-40 repeat proteins

One of the largest groups of conserved β -propeller proteins is the WD-repeat protein family. WD-repeat (or WD-40) proteins are characterized by conserved domains of ~40-60 residues that contain a glycine-histidine (GH) pair towards the N-terminus of the protein, and end with a tryptophan-aspartate (WD) dipeptide at the C-terminus. These proteins can contain from four to sixteen of these repeats, with most WD-40 repeats possessing at least 7 [11]. WD-40 proteins have no known catalytic function and are mainly involved in protein-protein interactions. They also possess a wide variety of functions, ranging from signal transduction, apoptosis, chromatin assembly, and cell cycle control [11-13]. The WD-40 proteins are conserved in many species of both prokaryotes and eukaryotes, and have been implicated in a number of human diseases, including lissencephaly, Cockayne syndrome, and Triple-A syndrome [11]. The repeating units of the WD proteins are believed to serve as a scaffold for protein-protein interactions and coordination of multiprotein complex assembly [13].

To date, there are only eight structures in the Protein Data Bank of WD-40 proteins bound to peptides of an interaction partner. These structures are of two human proteins, the Groucho/TLE corepressor bound to two repressor protein fragments, and the WDR5 protein which recognizes methylated histone proteins [14-18]. In these structures, the peptide fragments are bound within the shallow groove of the central cavity of the β -propeller, in direct contact with three or more of the propeller blades. The WD-40 protein interaction site makes the C-terminal domain of PilY1 especially interesting. Although pilY1 is not known to have catalytic activity (similar to many β -propeller proteins), it also lacks an obvious protein interaction surface. When superimposed with the WDR5/Histone 3 peptide structures, it is obvious that PilY1 contains two α -helices that cover the central cavity of the propeller and completely occlude the binding of any type of ligand at this surface. This observation is clearly illustrated in Figure 3.6. Thus, if the PilY1 C-terminal domain does mediate protein-protein interactions, it is likely that they occur at a site distinct from this more canonical surface.

Location of methionine mutants in the structure

Recall that 5 of 22 Leu-Met mutants introduced into the pilY1 C-terminal domain to provide phasing power led to unstable or uncrystallizable protein variants (see Chapter 2). It was not clear prior to structure determination why the majority of the leucine to methionine mutations tested were so detrimental to the stability of pilY1. The mutations generated are mapped out on the PilY1 model in Figure 3.7. Most of the unsuccessful mutations that resulted in expression of insoluble protein were found within β -strands of propeller blades.

As stated previously, conserved hydrophobic residues are often the only characteristic that β -propellers have in common. Because sequence diversity among these proteins is so great, recognition of conserved core hydrophobic residues can be a valuable tool when attempting to identify other proteins with this fold that lack obvious sequence repeats, such as those in the WD40 family [19]. These observations appear to explain why mutation of most of the conserved leucine residues from other bacterial PilY1 homologues was deleterious to the stability of the protein. However, two of the acceptable mutations were also found within β -strands. Of these two mutants, one is located on the outermost strand of a propeller blade (L832 of Blade 3), while the other is found at the very tip of the β -strand (L762 of Blade 2). These mutants are located at more peripheral locations, rather than at the core of the propeller. They also seem to have fewer interaction partners than some of the other deleterious mutants that are located in central strands or adjacent to outer loops that may be supported by the propeller structure. This observation could explain why the changes were not destabilizing as were many of the other intra-strand mutations.

PilY1 C-terminal domain is an incomplete β -propeller

A cursory inspection of the PilY1 model reveals that the structure does not form a complete β -propeller. There is a large opening ($\sim 24 \text{ \AA}$ at the widest point) between the first and seventh blades of the propeller; this is more than enough space to accommodate another complete blade. This does not appear to be an error of protein construct design the isolation of this domain from the full length PilY1 for structural analysis. BLAST sequence alignments with other bacterial PilY1 homologues show that all homology starts within the

C-terminal construct employed for these studies. As this is a conserved region of these proteins, one would assume that the homology would encompass the entire structural motif.

This assumption is further justified from inspection of the structure of the receptor binding protein P2 of bacteriophage PRD1 [20]. The PRD1 viron can infect a broad range of gram-negative bacteria, and P2 is thought to be responsible for recognition and attachment during infection [20]. The protein contains two domains, termed the head and the tail, which are composed of a “pseudo” β -propeller and extended loop/ β -strand region, respectively, and come together to resemble a seahorse. A superposition of the P2 protein with C-terminal PilY1 is shown in Figure 3.8. The authors refer to the head domain of the protein as a pseudo β -propeller for two reasons: first, there is little sequence identity or conservation between blades (sequence identity does occur between blades in WD-repeat proteins; see Figure 3.9), and second, the topology of the blades does not exhibit classic propeller symmetry. These features are also found in the C-terminal domain of PilY1. Another similar feature of the P2 model is the location of two short helices, which cover the central binding cavity that serves as an interaction surface in many other propellers. It is proposed that P2 uses either large outer loops or the open groove in the β -propeller as the region for protein-protein interactions. Interestingly, both the P2 and PilY1 structures contain a calcium binding site, which will be discussed in detail below. Taken together, these observations indicate that both infectious bacteria and viruses employ modified β -propeller-containing proteins to mediate host cell attachment and entry.

Calcium binding site

The C-terminal domain of pilY1 contains a single calcium binding site located within a β -hairpin turn between the first and second strands of the fourth propeller blade. The calcium binding site is shown in Figure 3.10. The calcium ion is coordinated by seven ligands: five side chain oxygens on three Asp and one Asn residue (D851, N853, D855 and D859; the interaction with D859 is bidentate), as well as the main chain carbonyl oxygen of V857 and one water molecule. Identity of the nature of the metal ion was determined through analysis of the number, nature and distance to the metal ion for all ligands involved. A summary of these distances is given in Table 3.1. All distances are well within agreement of the values found in the Cambridge Structural Database as well as the Protein Data Bank [21]. The three Asp residues are conserved among PilC from both *N. gonorrhoeae* and *meningitidis*, while the Asn residue is replaced by another Asp. This alignment is shown in Figure 3.11.

Calcium is a versatile intracellular signaling ion that is involved in a wide array of cellular processes, including fertilization, differentiation and proliferation, and apoptosis [22]. Many calcium binding proteins possess a domain known as an “EF-hand” that binds the metal ion. The domain consists of a calcium binding loop that is flanked on both sides by a two α -helices. This motif is found in a large number of Ca^{2+} binding proteins, including troponins, parvalbumin, and calmodulin [23]. The EF hand motif is often identified through sequence analysis. In this motif, the calcium ion is coordinated by side chain oxygen atoms from residues 1, 3, 5, 9, and 12 in the loop, and the carbonyl oxygen of residue 7. Recently sequence analysis of various integrin proteins has shown that the conservation of these residues is not enough to presume the type of structure in which the calcium is bound [23]. Integrins are cell adhesion receptors that are involved in cell-extracellular matrix and cell-cell

adhesion interactions. Although the integrins have always been predicted to have a calcium binding site with an EF hand motif, a more sequentially similar calcium binding site was found in a β -hairpin turn of an alkaline protease from *P. aeruginosa* [24]. This insight was confirmed through determination of a crystal structure of integrin $\alpha V\beta 3$, which contains in part a β -propeller with four calcium binding sites, all found in β -hairpin loops of four different propeller blades [25]. The positions of residues 1-7 of this motif are very similar to those in the EF hand. However, several differences are evident. First, the β -hairpin turn lacks an acidic residue at position 12 that is characteristic of the EF hand motif [26]. In addition, residue 9 is an Asp in only 34% of EF hand proteins, while an Asp is found in this position 96% of the time in integrins [26]. Both of these anti-EF hand characteristics are found in the PilY1 structure. The examination of the role this calcium binding site plays in PilY1 and *Pseudomonas* function is presented in the next chapter.

Identification of a surface patch of positive charge

Generation of an electrostatic potential of the PilY1 C-terminal domain structure revealed a large, positively charged area on the concave surface of the protein dimer as observed in the crystal. This surface is shown in Figure 3.12. The patch is located on each monomer and is composed of residues R929, R933, K934, and K970. We hypothesized that this area of the protein is important in pilus binding or other protein-protein interactions. R929 is conserved in *N. gonorrhoeae*, while a Lys replaces the Arg at position 933 in *N. meningitidis*. Additionally, Lys and Arg residues are found at the equivalent positions of 935 and 936 in both the *Neisseria* proteins. Although this sequence similarity is potentially

interesting, only the N-terminal amino acids (~60 residues) are conserved among gram-negative pilin proteins; the C-terminal region is highly variable and found on the outside of the pilus fiber, while the conserved region is packed tightly within the core [27, 28]. However, this charged region is conserved and could potentially be involved in interactions with other pilus proteins. Experimental analysis of these residues by site-directed mutagenesis and biophysical characterization is described in detail in the following chapter.

Analysis of domain swap observed in the crystal

Generation of symmetry mates for the PilY1 C-terminal domain dimer revealed a potential domain-swap at the N-terminal end of the structure. This swapped region encompasses the first β -blade of the propeller as well as the first three helices in the protein. The presence of an actual domain swap could not be determined directly due to the fact that the potential connection between the two monomers was disorder in the crystal structure (~712-722, described previously). Distance analysis of the potential connection between the two molecules was carried out and is shown in Figure 3.13. Calculation of the average length of a residue in a straight β -strand in the structure yields a distance of 2.83 Å, or 28 Å needed for the ~10 residues missing from the model. Inspection of the distances required leads to the assumption that there is not enough space to accommodate these missing residues without the presence of a domain-swap. Biophysical characterization of this region using mutation analysis is described further in the following chapter.

REFERENCES

1. Kelley, L.A., R.M. MacCallum, and M.J. Sternberg, *Enhanced genome annotation using structural profiles in the program 3D-PSSM*. J Mol Biol, 2000. **299**(2): p. 499-520.
2. Jawad, Z. and M. Paoli, *Novel sequences propel familiar folds*. Structure, 2002. **10**(4): p. 447-54.
3. Brown, K., M. Tegoni, M. Prudencio, A.S. Pereira, S. Besson, J.J. Moura, I. Moura, and C. Cambillau, *A novel type of catalytic copper cluster in nitrous oxide reductase*. Nat Struct Biol, 2000. **7**(3): p. 191-5.
4. Carr, S., C.N. Penfold, V. Bamford, R. James, and A.M. Hemmings, *The structure of TolB, an essential component of the tol-dependent translocation system, and its protein-protein interaction with the translocation domain of colicin E9*. Structure, 2000. **8**(1): p. 57-66.
5. Fulop, V., Z. Bocskei, and L. Polgar, *Prolyl oligopeptidase: an unusual beta-propeller domain regulates proteolysis*. Cell, 1998. **94**(2): p. 161-70.
6. Garcia-Higuera, I., J. Fenoglio, Y. Li, C. Lewis, M.P. Panchenko, O. Reiner, T.F. Smith, and E.J. Neer, *Folding of proteins with WD-repeats: comparison of six members of the WD-repeat superfamily to the G protein beta subunit*. Biochemistry, 1996. **35**(44): p. 13985-94.
7. Neer, E.J. and T.F. Smith, *G protein heterodimers: new structures propel new questions*. Cell, 1996. **84**(2): p. 175-8.
8. Vellieux, F.M., F. Huitema, H. Groendijk, K.H. Kalk, J.F. Jzn, J.A. Jongejan, J.A. Duine, K. Petratos, J. Drenth, and W.G. Hol, *Structure of quinoprotein methylamine dehydrogenase at 2.25 Å resolution*. Embo J, 1989. **8**(8): p. 2171-8.
9. Pons, T., R. Gomez, G. Chinea, and A. Valencia, *Beta-propellers: associated functions and their role in human diseases*. Curr Med Chem, 2003. **10**(6): p. 505-24.
10. Todd, A.E., C.A. Orengo, and J.M. Thornton, *Evolution of function in protein superfamilies, from a structural perspective*. J Mol Biol, 2001. **307**(4): p. 1113-43.
11. Li, D. and R. Roberts, *WD-repeat proteins: structure characteristics, biological function, and their involvement in human diseases*. Cell Mol Life Sci, 2001. **58**(14): p. 2085-97.
12. Neer, E.J., C.J. Schmidt, R. Nambudripad, and T.F. Smith, *The ancient regulatory-protein family of WD-repeat proteins*. Nature, 1994. **371**(6495): p. 297-300.

13. Smith, T.F., C. Gaitatzes, K. Saxena, and E.J. Neer, *The WD repeat: a common architecture for diverse functions*. Trends Biochem Sci, 1999. **24**(5): p. 181-5.
14. Dou, Y., T.A. Milne, A.J. Ruthenburg, S. Lee, J.W. Lee, G.L. Verdine, C.D. Allis, and R.G. Roeder, *Regulation of MLL1 H3K4 methyltransferase activity by its core components*. Nat Struct Mol Biol, 2006. **13**(8): p. 713-9.
15. Schuetz, A., A. Allali-Hassani, F. Martin, P. Loppnau, M. Vedadi, A. Bochkarev, A.N. Plotnikov, C.H. Arrowsmith, and J. Min, *Structural basis for molecular recognition and presentation of histone H3 by WDR5*. Embo J, 2006. **25**(18): p. 4245-52.
16. Couture, J.F., E. Collazo, and R.C. Trievel, *Molecular recognition of histone H3 by the WD40 protein WDR5*. Nat Struct Mol Biol, 2006. **13**(8): p. 698-703.
17. Jennings, B.H., L.M. Pickles, S.M. Wainwright, S.M. Roe, L.H. Pearl, and D. Ish-Horowicz, *Molecular recognition of transcriptional repressor motifs by the WD domain of the Groucho/TLE corepressor*. Mol Cell, 2006. **22**(5): p. 645-55.
18. Han, Z., L. Guo, H. Wang, Y. Shen, X.W. Deng, and J. Chai, *Structural basis for the specific recognition of methylated histone H3 lysine 4 by the WD-40 protein WDR5*. Mol Cell, 2006. **22**(1): p. 137-44.
19. Paoli, M., *Protein folds propelled by diversity*. Prog Biophys Mol Biol, 2001. **76**(1-2): p. 103-30.
20. Xu, L., S.D. Benson, S.J. Butcher, D.H. Bamford, and R.M. Burnett, *The receptor binding protein P2 of PRD1, a virus targeting antibiotic-resistant bacteria, has a novel fold suggesting multiple functions*. Structure, 2003. **11**(3): p. 309-22.
21. Harding, M.M., *Small revisions to predicted distances around metal sites in proteins*. Acta Crystallogr D Biol Crystallogr, 2006. **62**(Pt 6): p. 678-82.
22. Gifford, J.L., M.P. Walsh, and H.J. Vogel, *Structures and metal-ion-binding properties of the Ca²⁺-binding helix-loop-helix EF-hand motifs*. Biochem J, 2007. **405**(2): p. 199-221.
23. Springer, T.A., H. Jing, and J. Takagi, *A novel Ca²⁺ binding beta hairpin loop better resembles integrin sequence motifs than the EF hand*. Cell, 2000. **102**(3): p. 275-7.
24. Miyatake, H., Y. Hata, T. Fujii, K. Hamada, K. Morihara, and Y. Katsube, *Crystal structure of the unliganded alkaline protease from Pseudomonas aeruginosa IFO3080 and its conformational changes on ligand binding*. J Biochem (Tokyo), 1995. **118**(3): p. 474-9.

25. Xiong, J.P., T. Stehle, B. Diefenbach, R. Zhang, R. Dunker, D.L. Scott, A. Joachimiak, S.L. Goodman, and M.A. Arnaout, *Crystal structure of the extracellular segment of integrin alpha Vbeta3*. Science, 2001. **294**(5541): p. 339-45.
26. Oxvig, C. and T.A. Springer, *Experimental support for a beta-propeller domain in integrin alpha-subunits and a calcium binding site on its lower surface*. Proc Natl Acad Sci U S A, 1998. **95**(9): p. 4870-5.
27. Hansen, J.K., K.P. Demick, J.M. Mansfield, and K.T. Forest, *Conserved regions from Neisseria gonorrhoeae pilin are immunosilent and not immunosuppressive*. Infect Immun, 2007. **75**(8): p. 4138-47.
28. Nudleman, E. and D. Kaiser, *Pulling together with type IV pili*. J Mol Microbiol Biotechnol, 2004. **7**(1-2): p. 52-62.

FIGURE LEGENDS

Figure 3.1 This dimer comprises the asymmetric unit. The interface contains several helices and loops as well as close interaction of two beta propeller blades, shown in purple.

Figure 3.2 The seven blades of the beta propeller are color-coded. An especially interesting feature is the outer magenta beta strand, which is found in two different propeller blades. The propeller is open on the right side and is considered incomplete.

Figure 3.3 A 13-residue strand from Blades 5 and 6 (top) twists and encompass two separate β -propeller blades.

Figure 3.4 A topology diagram of the PilY1 C-terminal domain simplifies its complex tertiary structure. Each blade is color-coded in the same manner as Figures 3.1 and 3.2. An interesting feature is the sixth blade (shown in magenta), which contains a long strand that is also part of the fifth blade.

Figure 3.5 The superposition was carried out using only the β -strands in the propeller blades. Although the predicted structural motif of PilY1 was correct, the superposition is poor, with an rmsd of ~ 16 Å.

Figure 3.6 **A.** The WDR5 protein bound to a histone 3 (H3) methylated peptide at the WD40 interaction site. **B.** Similar view of the pilY1 C-terminal domain. **C, D.** Superposition of the two models shows that the potential interaction site of PilY1 is covered by two helices that occlude binding at the central cavity.

Figure 3.7 This map of the methionine mutants tested shows the soluble expressing mutants in green, and the insoluble dysfunctional mutants in red. Notice that many of the unsuccessful mutants are found within β -strands in the propeller blades.

Figure 3.8 The P2 protein contains a partial or “pseudo” β -propeller as does PilY1.

Figure 3.9 The sequence alignments of the first three propeller blades in both WDR5 and PilY1 C-terminal domain show strong sequence conservation in the WQD-40 protein, but little homology in PilY1. Identical residues are shown in red; nearly identical residues are shown in blue; and conserved residues are shown in green.

Figure 3.10 The calcium ion is coordinated by five residues and a water molecule. It makes a bidentate contact with D859. Distances shown are within good agreement with other structures from the Cambridge Database and the PDB.

Figure 3.11 Three of the Asp residues are conserved with the *Neisseria* pilC proteins, while the Asn is replaced with another D. The Gly residue at position 6 and a hydrophobic residue at position 8 are conserved as well.

Figure 3.12 The positively charged residues are shown in blue, as indicated by the key. The patch is located in the lower left hand corner of the left monomer, and the upper right hand corner on the right monomer.

Figure 3.13 The original PilY1 structure is shown in cyan, and its symmetry mate is shown in magenta. In the close up, the potential domain swap would be connected through the magenta spheres and the cyan spheres; the distances between the intramonomer residues are not long enough to accommodate the 10 amino acids missing from the structure, hence a domain swap is necessary.

Ligand	Distance in structure (Å)	Expected distance (± 0.2 Å)
D851	2.46	2.36
N853	2.25	2.36
D855	2.53	2.36
D859 (2)	2.46	2.8 (maximum)
	2.61	2.8 (maximum)
V857	2.39	2.36
H ₂ O	2.41	2.39

Table 3.1 **Actual and expected calcium ion distances to ligands**

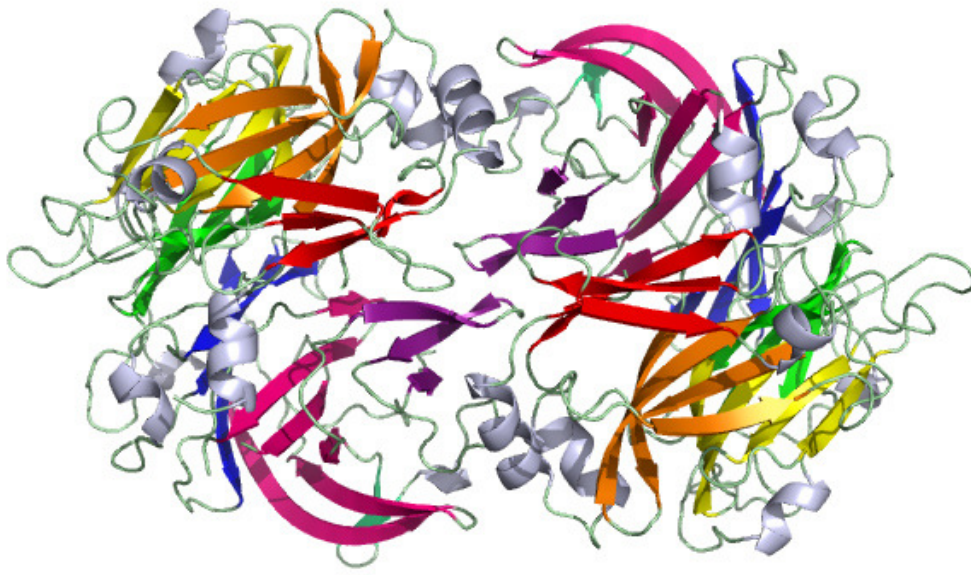


Figure 3.1 PilY1 dimer in crystal structure

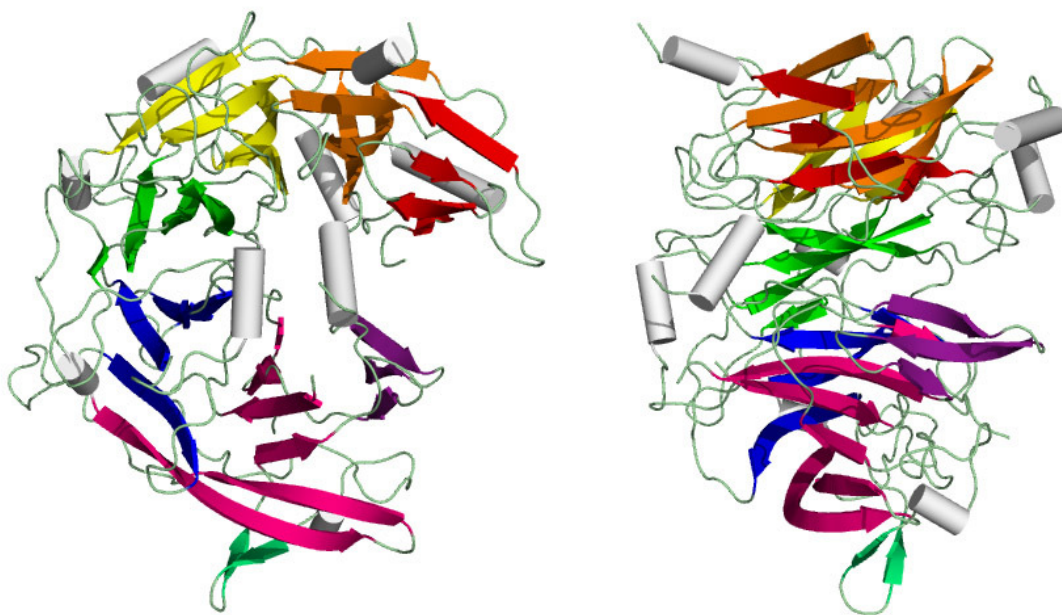


Figure 3.2 Front and side view of PilY1 C-terminal monomer

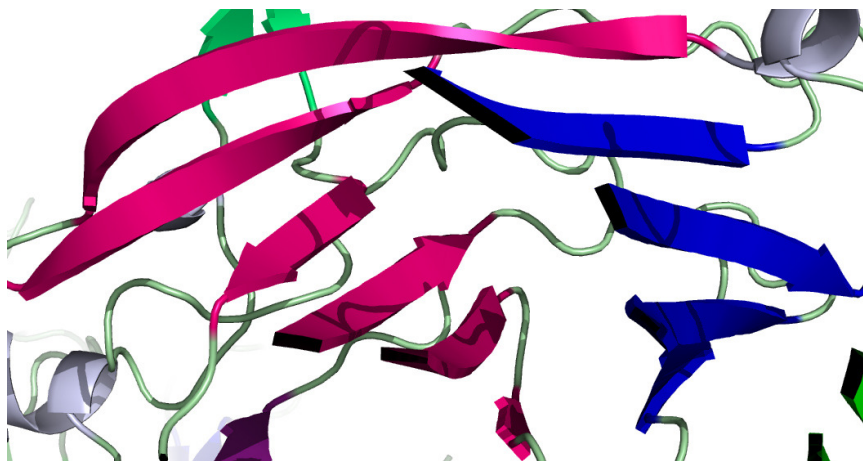


Figure 3.3 An extensive β -strand (top) encompasses two propeller blades

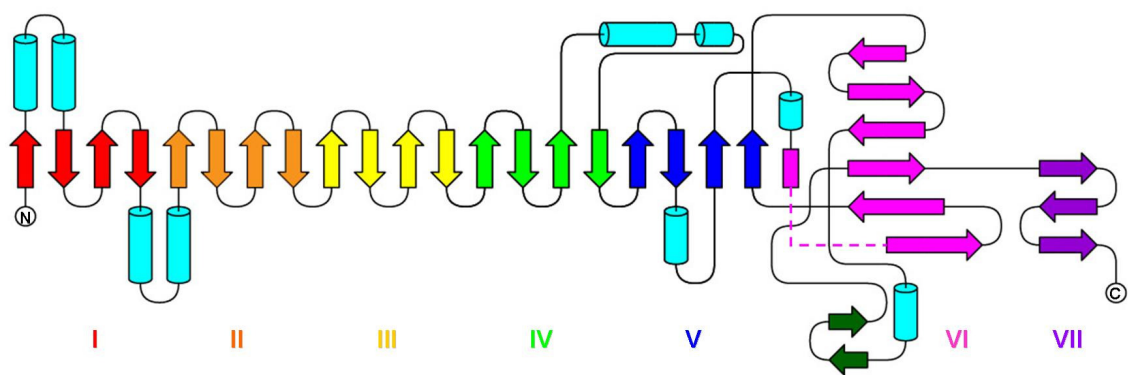


Figure 3.4 Topology of PilY1 C-terminal domain



Figure 3.5 Superposition of PilY1 C-terminal Domain with *Comamonas testosteroni* quinohemoprotein alcohol dehydrogenase

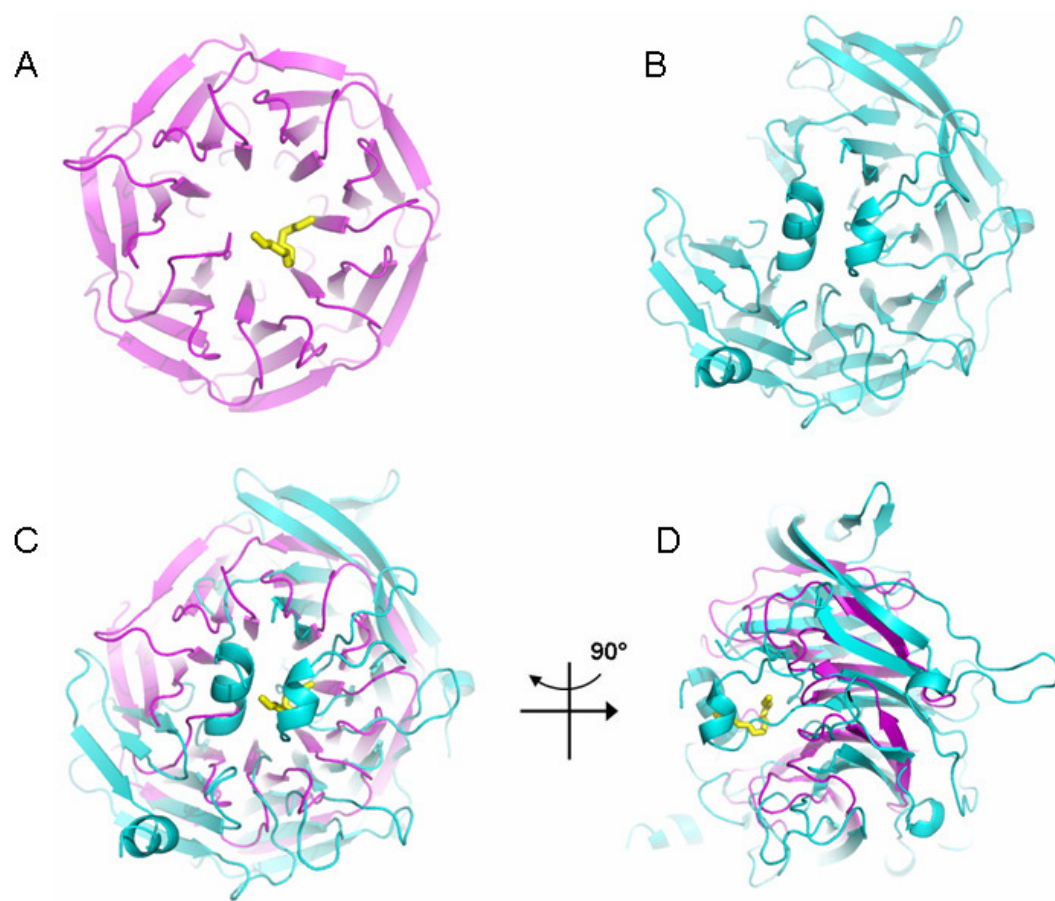


Figure 3.6 Standard WD-40 interaction site of WDR5 β -propeller is inaccessible in pilY1

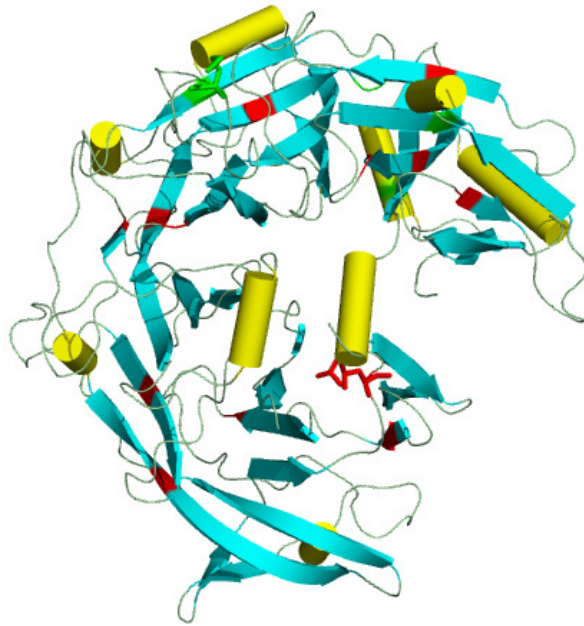


Figure 3.7 **Methionine mutations tested in PilY1 C-terminal domain**



Figure 3.8 Superposition of C-terminal PilY1 with the receptor-binding protein P2 of bacteriophage PRD1

WDR5

Blade 1	VSSVKF	WLA ^{SS} S	L ^I K ^I W ^G	FE ^K T ^I S
Blade 2	I ^S D ^V AW	LL ^V S ^A S	TL ^K I ^W D	CL ^K TL ^K
Blade 3	VFCCNF	L ^I V ^S GS	SV ^R I ^W D	CL ^K TL ^P

PilY1 C-terminal domain

Blade 1	TV ^G	RVY ^V G	LH ^G	TF ^A FI
Blade 2	VV ^A D ^A FF	AWHTV ^L I ^G S	GA ^F AL ^D	KLLWE ^I
Blade 3	TV ^A RL	GKWAV ^V T ^G	ALL ^I ID	LSSPRL ^A

Figure 3.9 Sequence alignments of WDR5 propeller blades vs. PilY1 propeller blades

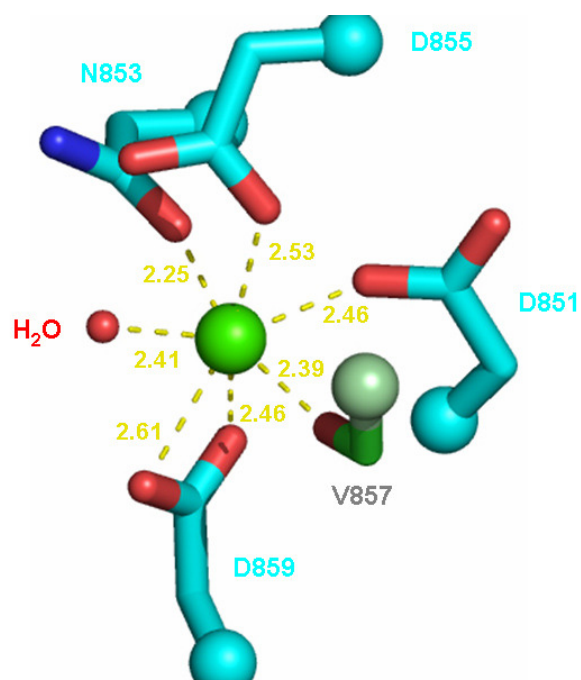


Figure 3.10 PilY1 calcium binding site

N._gonorrhoeae	DKD--LDGTV
N._meningitidis	DKD--LDGTV
L._pneumophila	QKDG LQADPVT
C._testosteroni	YGS---NGEVT
P._syringae	VN---AQNVVQ
P._aeruginosa	DNN--SDGVAD

Figure 3.11 Sequence alignment of calcium binding site with various PilY1 homologues

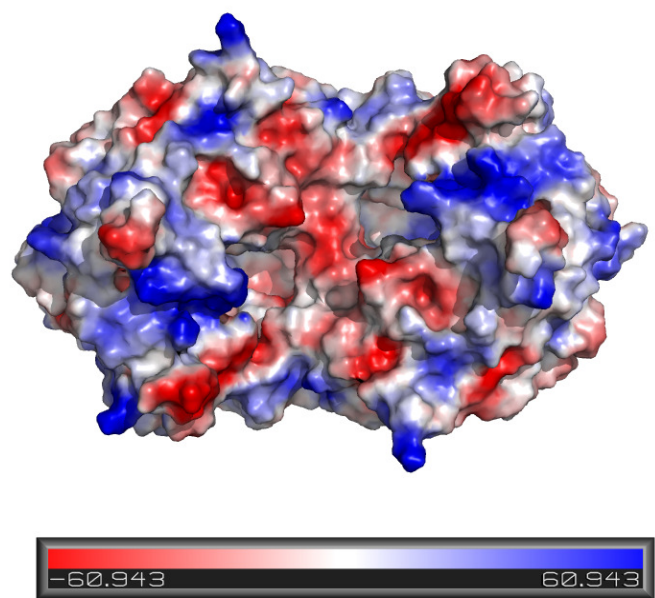


Figure 3.12 Electrostatic potential of PiliY1 dimer

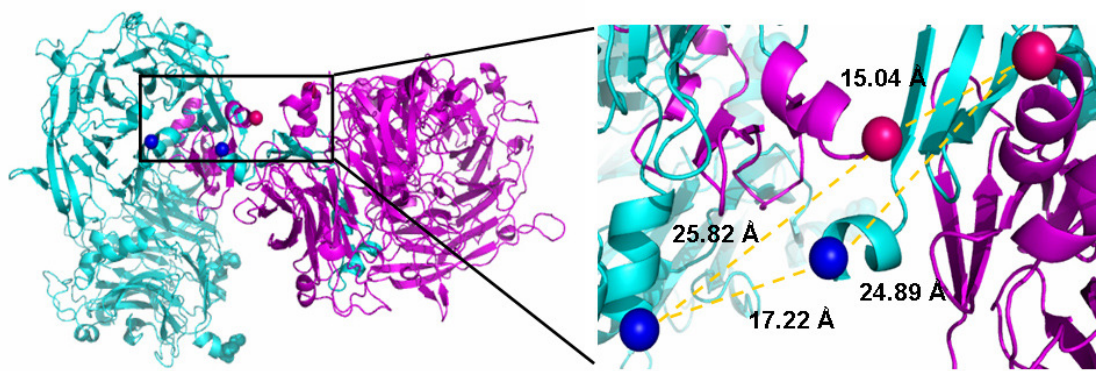


Figure 3.13 Analysis of potential domain swap in PilY1 C-terminal domain

Chapter 4:

Functional and Biophysical Characterization of PilY1 C-terminal Domain Mutations

INTRODUCTION

As stated in Chapter 1, no structural or functional data existed on *Pseudomonas aeruginosa* PilY1 when we initiated this project. Once we obtained and analyzed the three-dimensional structure of the C-terminal domain, we next wanted to determine the areas of the protein that are important for its functions, such as twitching motility, pilus biogenesis and retraction, and its interaction with other protein binding partners. In addition, we wanted to determine whether or not the dimer and the domain swap observed in the crystal structure are at all relevant to function. We answered these questions through the use of rational mutational analysis applied to specific areas of the protein that were interesting and potentially relevant to function. This chapter focuses on the rationale behind the selection of mutants, as well as descriptions and analysis of the results of the functional and biophysical methods used to characterize and evaluate these mutations. Through the use of these techniques, we discovered that the calcium binding site found in a β -turn of the protein is important for pilus formation and twitching motility. Binding affinity experiments highlighted the residues that are important for coordinating the calcium; the results of these experiments mirrored those found through the functional assays. This

discovery could be integral to understanding how PilY1 is involved in bacterial colonization and infection.

REGIONS OF THE PROTEIN CHOSEN FOR MUTATIONAL ANALYSIS

Positively charged patch

We decided to mutate the four residues identified as part of the positively charged patch on the surface of the protein. This region could potentially be involved in binding the bacterial pilus or perhaps another protein involved in the Type IV pilus system; this was described briefly as a part of the structural analysis (Chapter 3). The residues we decided to mutate were R929, R933, and K934. As stated previously in Chapter 3, R929 is conserved in *N. gonorrhoeae*, while a Lys replaces the Arg at position 933 in *N. meningitidis*. We chose to mutate all of these residues to alanine in order to eliminate the charge on the side chain.

Calcium binding region

The residues involved in the calcium binding site were described in Chapter 3. These residues are D851, N853, D855, and D859. Another residue, V857, also makes contact with the metal through its main-chain carbonyl, but this contact would be difficult to eliminate. We mutated each of the other four residues to alanine, again to eliminate interaction between the side chains and the calcium moiety. We also made an additional double mutant, D851A/D859A, to assess the impact associated with removal of more than one calcium ligand at the same time.

Dimer and domain swap residues

To assess whether the crystallographic dimer and the domain swap are of functional significance, we designed mutations to disrupt these potential interactions. To disturb the dimer, we designed two variants, each containing mutations of three to four polar or charged residues found at the dimer interface; dimer mutant 1 (T655A, K675A, R1093A, and Q1094A) and dimer mutant 2 (T655A, K675A, and S1137A).

To attempt elimination of the crystallographic domain swap, we deleted ten residues (L712-G722) that serve as the connection between the two monomers.

Cloning mutants and mutation of associated residues

During the initial cloning of the *PilY1* gene by our collaborators, three accidental mutations were discovered and found to disrupt all of the protein's measurable functionality. These mutations are not found in the construct we used for crystallization. We decided to make the individual mutations as well as mutate some of the surrounding residues in order to determine which of these residues caused the loss of functionality, as well as whether or not the particular cloning mutation regions were important to the protein's function.

F670L

This mutation is of a residue found in a ten amino acid helix located in the N-terminal region of the protein. No adjacent residue mutations were attempted.

P1012S

This mutant is found in an extended loop region (17 residues). As stated in the previous chapter, β -propellers that lack a WD40 interaction site often use large, unstructured loop regions as a binding interface for protein partners. Based on this observation we hypothesized that this could be an important part of the structure and made several changes to test this. In addition to P1012S, we tested H930A, W965A, W1015A, D1019A, and S1021A, also found in this region, to determine the impact of these residues on protein function.

V880A

This mutation is also found in an extended (18 residue) loop region towards the back side of the molecule. In addition to this mutation, we also made P855A, to see if mutating this adjacent residue could destabilize the loop and impact protein function.

Mutations not expected to impact function

As a type of negative control, we tested mutations of two residues in the protein that were found in locations not expected to be important to protein function. One of these residues, Y945, is located between the helices that cover the standard WD40 interaction site region. The other, P782, is a residue found on the back face of the protein in a region that

was not an obvious target for analysis. The Y945 residue was mutated to an alanine, while P782 was changed to a serine, to mimic the original P1012 cloning mutation.

FUNCTIONAL AND BIOPHYSICAL ANALYSIS OF WILD-TYPE PILY1 AND SELECTED VARIANTS

We used three different experiments to determine the effect each mutation had on PilY1 function. The first experiment measures twitching motility, or the ability of the bacteria to translocate and along a surface. This motility is conferred by the pilus and is necessary for the formation of biofilms [1]. As PilY1 and its *Neisseria* homologue pilC have both been implicated in pilus biogenesis (Matthew Wolfgang, UNC, unpublished) [2], this assessment is extremely important in assessing the functional effects of mutations in PilY1. The other two methods used were Western blotting for detection of PilY1 in *Pseudomonas* whole cell lysate, and SDS-PAGE analysis to detect the presence of PilA, which indicates whether or not the PilY1 mutations affect pilus biogenesis.

Methods

Full length *PilY1* was cloned from *Pseudomonas aeruginosa* genomic DNA into the pDONR221 shuttle vector (Invitrogen, Carlsbad, CA). All mutations in the *PilY1* gene were made using the Quikchange Site-Directed Mutagenesis Kit (Stratagene, La Jolla, CA). After the mutations were created and confirmed by DNA sequencing analysis, the *PilY1* gene was transferred to the pMMB67EH vector [3] using the Gateway Cloning System (Invitrogen,

Carlsbad, CA). This vector is a low copy, broad host range *E. coli* plasmid used for expression of genes in gram-negative bacteria. The pMMB67EH plasmid with the full length PilY1 gene was mated with three different *P. aeruginosa* destination strains (wild-type PAK strain, PilY1-null background) using the *E. coli* DH5 α /pRK2013 strain, which contains the helper plasmid pRK2013 used to aid in the transfer of the PilY1/pMMB67EH plasmid into the *Pseudomonas* destination strains.

Twitching motility assays

Colonies containing the mutant plasmid in *Pseudomonas aeruginosa* (PilY1 null background), as well as the PAK, PilY1 null, and PilY1 positive control strains, were streaked on LB agar plates containing 150 μ g/mL carbenecillin and 25 μ g/mL Irgacan. Plates were prepared for the twitching assay by pouring 5 mL of LB/1% agar and 50 μ M IPTG in sterile tissue culture Petri dishes (Fisher Scientific, Houston, TX). Plates were permitted to solidify for 5-10 minutes. The plated bacteria containing the mutant plasmids as well as the controls were scraped and stabbed into the middle of the LB/agar plates. The plates were then placed in a tissue culture incubator at 37°C. The radius of the distance the bacteria traveled from the center of the plate was recorded as 3-4 separate measurements at 24 and 48 hour time points. The final distance recorded was difference in the average radius measured at both time points divided by the time (radius in cm/hr). Experiments for each sample were performed in triplicate.

Pilus preparation

Patched colonies containing each mutant in the PilY1 null background (as well as the PilY1 positive control in the *Pseudomonas* PAK strain) from plates containing 150 µg/mL carbenecillin and 25 µg/mL Irgacan were swabbed and spread on LB/agar plates with 50 µM IPTG in order to obtain as much bacteria as possible. Two different clones for each mutant were selected. The plates were incubated overnight at 37°C. All of the bacteria on the plates were scraped with bent Pasteur pipets and resuspended in 1 mL of 0.15 M NaCl and 0.2% Formaldehyde. The tubes were vortexed on high for one minute and centrifuged at 16,000 G for 5 minutes. After spinning, 450 µL of the supernatant was transferred to a clean microfuge tube and 50 µL of 1 M MgCl₂ was added. The tubes were incubated either 4 hours or overnight at 4°C and then centrifuged at 16,000 G for 15 minutes. The pellet was then resuspended in SDS-PAGE loading dye and electrophoresed on an 18% acrylamide SDS gel. The amount of dye was adjusted to match the size of the pellet obtained in the first centrifugation step. The presence of pilus was determined by the amount of PilA (17 kDa) visualized on the gel.

Presence of PilY1 in Pseudomonas cells

To determine the presence of mutant PilY1 in *Pseudomonas aeruginosa* whole-cell lysate, single colonies were grown in LB overnight. The next day, 5 mL of the overnight culture was added to 5 mL LB with 50 µM IPTG and shaken at 37°C to an OD₆₀₀ of 1. The culture was then centrifuged for 10 minutes at 16,000 G and resuspended in SDS-PAGE

loading dye. The samples were electrophoresed on an 8% acrylamide SDS gel, which was then subjected to Western blotting, using a PilY1 primary antibody with an HRP secondary antibody to detect the presence of PilY1.

Oligomerization state of wild-type and domain swap deletion PilY1

We wanted to determine the oligomerization state of C-terminal PilY1 in solution for two reasons: first, to see if the dimer found in the crystal structure was present in solution; and second, to determine whether the domain swap seen in the symmetry-related molecule was actually present and not a crystallographic artifact. To accomplish this goal we used two techniques, analytical ultracentrifugation (AUC) and dynamic light scattering (DLS).

Analytical Ultracentrifugation

The AUC experiment was carried out using three different concentrations of C-terminal PilY1, ranging from 0.25-0.7 mg/mL (in 20 mM Tris-HCl, pH 7.8 and 200 mM NaCl). Sedimentation equilibrium experiments were performed at 25°C at three speeds (12, 15 and 21K rpm) for 10-12 hours each, using a Beckman XL-A analytical ultracentrifuge with scanning absorption optics (Beckman Coulter, Fullerton, CA) and an An-50 Ti rotor. Absorbance measurements were taken every 2 hours at 280 nm. Baseline absorbance offsets were determined by increasing rotor speed to 40K rpm for 4 hours. Data were analyzed with the Origin software package (Microcal, Northampton, MA).

Dynamic Light Scattering

Light scattering experiments were performed at 25°C using a Wyatt DAWN EOS multi-angle light scattering system (Wyatt Technology, Santa Barbara, CA) coupled to a 25 mL Superdex200 size-exclusion column (GE Health Sciences, Piscataway, NJ) on an AKTA FPLC (Pharmacia, Uppsala, Sweden). The two instruments were interfaced with a Wyatt Optilab refractometer and Wyatt dynamic light scattering module. The experiment was carried out using 250 μ L of C-terminal PilY1 protein at 5 mg/mL in 20 mM Tris-HCl pH 7.8, 200 mM NaCl, and 50 mM imidazole. Data were analyzed using Astra software (Wyatt Technology).

Fluorescence resonance energy transfer to determine calcium binding constants

We used fluorescence resonance energy transfer (FRET) to measure the calcium ion binding affinity of wild-type PilY1 and each of the calcium ligand mutants (D851A, N853A, D855A, and D859A). Affinity was measured using terbium in place of calcium. Terbium is often used as a calcium substitute for affinity experiments, as it has similar ionic radius and coordination properties to those of calcium [5]. In addition, terbium can be excited by the tryptophan emission wavelength, and its own emission can be measured upon binding to the protein, provided that there are one or more tryptophan residues within a usable Förster radius. Literature values of this radius for tryptophan fluorescence are widely varied (from 1 to 20 Å), thus we decided to carry out the experiment to determine if the Trp residues in the

structure were in close enough proximity to the calcium binding site to facilitate energy transfer.

C-terminal PilY1 was prepared as described in Chapter 2 and dialyzed (to remove bound calcium) against 8 L of 10 mM Tris-HCl pH 7.8, 50 mM NaCl, and 20 mM EDTA pH 8.0, over a 48 hour time period. The EDTA was then removed by dialysis in 8 L of the same buffer, without EDTA for another 48 hours. To ensure all of the calcium was removed from the protein, it was then passed over a Ni-NTA column which had previously been stripped of the nickel by using EDTA. This method has been used to effectively remove stubbornly bound metal ions from proteins by capturing the metal on the column [6].

FRET experiments were carried out at 25°C using a SPEX Fluorolog-3 Research T-format Spectrofluorometer (Horiba Jobin Yvon, Edison, NJ). Terbium chloride (TbCl_3) was diluted in protein buffer (10 mM Tris-HCl pH7.8, 50 mM NaCl) to a working concentration of 10 mM. C-terminal PilY1 (400 μL at 2 μM or ~ 0.12 mg/mL) was added to the cuvette, and terbium was added in increasing concentrations (from 0-90 μM) in 0.2 μL increments. The protein was excited at 283 nm, the tryptophan excitation wavelength, and terbium fluorescence was measured at 543 nm. Data were analyzed with SigmaPlot (Systat Software, San Jose, CA).

FUNCTIONAL ASSAY RESULTS

Positively charged patch

The R929A, R933A, and K934A were subject to the three functional assays (twitching motility, presence of PilY1, and presence of pilus) as previously described. All

three of these variants were capable of wild-type or better twitching motility (Figure 4.1). The SDS-PAGE analysis of pilus presence showed that both R929A and R933A produced a detectable amount of the pilin protein, while K934A showed no sign of PilA (Figure 4.2). Western blot analysis showed PilY1 production was that of wild-type in both R929A and R933A, but nonexistent in K934A (Figure 4.3). The implications of these results, as well as the results of these assays for all mutations, are further analyzed in the Discussion section of this chapter.

Calcium binding region

Of the four calcium binding ligand mutants in *Pseudomonas*, only bacteria harboring the N853 variant were capable of twitching like wild-type. Twitching motility was down in both D855A and D859A-containing strains (to about half that of wild-type), while strains harboring both D851A and the double mutant D851/859A were not capable of twitching (Figure 4.4). The SDS-PAGE analysis of pilus presence showed that none of these mutants produced measurable amounts of the pilin protein (Figure 4.5). Western blot analysis showed small amounts of PilY1 in the whole-cell samples from strains containing the D851A and D859A mutants, equal to about half that of wild-type expression. N853A, D855A, and D851/859A variants had very little PilY1 expression (Figure 4.6).

Dimer and domain swap residues

Both of the dimer mutants (dimer mutant 1 [T655A, K675A, R1093A, and Q1094A] and dimer mutant 2 [T655A, K675A, and S1137A]) in *Pseudomonas* were able to twitch like the wild-type bacteria; however, the domain swap deletion mutant $\Delta 712-722$ was unable to do so (Figure 4.7). SDS-PAGE analysis of the pilus showed that strains containing the dimer mutants were also producing wild-type levels of pilus, while the domain swap mutant produced none (Figure 4.8). Lastly, strains harboring the dimer mutants were able to produce wild-type levels of PilY1, while the domain swap deletion produced almost none (Figure 4.9).

Cloning mutants

Two of the cloning mutants were fully functional: both F670L and V880A are capable of twitching motility (though F670L is down to half that of wild-type) (Figure 4.10) and produce wild-type levels of both PilA and PilY1 (Figures 4.11 and 4.12). The P1012S mutation in *Pseudomonas* is incapable of twitching, and does not produce PilA (Figures 4.10 and 4.11). However, it does produce wild-type levels of PilY1 (Figure 4.12).

Cloning-associated mutants

Most of the variants chosen from the V880 and P1012 regions were able to function as wild-type, producing pilus and PilY1 and twitching normally (P885A, H930A, D1019A, and S1021A) (Figures 4.13-4.15). Neither W965A nor W1015A could make PilA or PilY1; however W965A was capable of twitching normally, while W1015A could not twitch at all.

Mutations not expected to impact function

Psuedomonas strains harboring both of these mutants (P782S and Y945A) in are able to twitch like the wild-type protein (Figure 4.15). SDS-PAGE and Western blot analysis show they are also able to make both pilus and PilY1, although pilus expression is down in Y945A, while PilY1 expression is down in P782S (Figures 4.17 and 4.18).

BIOPHYSICAL CHARACTERIZATION RESULTS

Analytical Ultracentrifugation

Sedimentation equilibrium data for the wild-type PilY1 C-terminal domain (absorbance at 280 nm versus radius in cm) were fitted using a non-linear curve fitting procedure (Origin software) and an estimated partial specific volume (0.736 cm³/g) to determine the weight-average monomer molecular weight and an association constant using the equation [7]:

$$c_r = c_{mon,r_0} e^{\frac{[(\omega^2/RT)M(1-\bar{v}\rho)(r^2-r_0^2)]}{2}} + K_a(c_{mon,r_0})^2 e^{\frac{[(\omega^2/RT)2M(1-\bar{v}\rho)(r^2-r_0^2)]}{2}} + E$$

where c_r is the concentration of the protein at the radial position r , c_{mon} is monomer concentration at the reference radius (r_0), ω is the angular velocity (radians/second), R is the gas constant ($8.314 \times 10^7 \text{ erg} \cdot \text{mol}^{-1} \cdot \text{K}^{-1}$), T is the temperature in Kelvin, M is the molecular weight of the monomer, \bar{v} is the partial specific volume (mL/g), ρ is the density of the

solvent in g/mL, K_a is the association constant, and E is the baseline offset. K_a can be converted to K_D , the dissociation constant, through the following equation:

$$K_D = \frac{2}{K_a b \epsilon}$$

where b is the path length in cm and ϵ is the molar extinction coefficient of the protein in $\text{M}^{-1}\text{cm}^{-1}$ (71,300 $\text{M}^{-1}\text{cm}^{-1}$).

When the data were fit to a single species model, the estimated molecular weight was determined as 90.3 kDa, indicating that multiple species and higher-order oligomerization were present. When the known monomeric molecular weight of 61.3 kDa was applied and fit to a monomer-dimer model, the variance value (χ^2) decreased from 1.3×10^{-4} to 1.8×10^{-5} and the residuals appeared more random, suggesting that this was the appropriate model to use. Analysis using the monomer-dimer equilibrium model yielded a dissociation constant of 27 μM . The AUC data, the fit, and the residuals are shown in Figure 4.19.

Dynamic light scattering

Light scattering was quantified as a second-order correlation function using the Astra software. This function was then subject to a non-linear least squares fitting to determine the hydrodynamic radii of the species, and hence the mass. The theory of light scattering can be condensed into the following equation:

$$\frac{K^* c}{R(\theta, c)} = \frac{1}{M_w P(\theta)} + 2A_2 c$$

where $R_{(\theta, c)}$ is the excess Rayleigh ratio of the solution as a function of the scattering angle (θ) and concentration c (directly proportional to intensity of light scattered), M_w is the weight-averaged solute molecular weight, A_2 is the second virial coefficient in the expansion of osmotic pressure, K^* is a constant ($4\pi^2 (dn/dc)^2 n_0^2 / N_a \lambda_0^4$), N_a is Avagadro's number, and $P(\theta)$ is a function which describes the angular dependence of the scattered light (related to the root mean square radius). Expansion of $P(\theta)$ gives the equation:

$$P(\theta) \approx 1 - \frac{16\pi^2 n_0^2}{3\lambda_0^2} \langle r_g^2 \rangle \sin^2 \frac{\theta}{2} + O(\sin^4 \frac{\theta}{2}) - \dots$$

where n_0 is the solvent's refraction index, λ_0 is the vacuum wavelength of the laser, and r_g is the rms radius. The mean square radius can then be calculated for the slope at $\theta=0$ of the measured ratios $1/R(\theta, c)$ with respect to $\sin^2(\theta/2)$ using the final equation:

$$\langle r_g^2 \rangle = \frac{\sum_i r_i^2 m_i}{\sum_i m_i} = \frac{1}{M} \sum_i r_i^2 m_i$$

where M is the mass of the macromolecule made up of m_i elements, and r_i is the distance between m_i and the center of the molecule's mass M .

Peak area analysis showed ~90% recovery of initial sample. Light scattering and refractive index measurements plotted against column volume for both the wild-type protein and $\Delta 712-722$ C-terminal variant are shown in Figures 4.20 and 4.21. Both samples showed three different peaks representing three species in the solution, with estimated molecular weights of 280, 120, and 60 kDa for wild-type, and 257, 124, and 64 kDa for the variant protein. When compared to the monomer weight of 61.3 kDa, these values correspond to tetramer, dimer, and monomer species, respectively. The percentage of each species was determined to be 5, 75, and 25% (for tetramer, dimer, and monomer) in the wild-type sample, and 5, 80, and 15% in the deletion mutant.

Terbium/Tryptophan FRET of wild-type and calcium binding mutants

FRET data were analyzed using SigmaPlot. Fractional change in terbium fluorescence at 543 nm was plotted against terbium chloride concentration, and data was fit to a Hill equation in order to determine the dissociation constant. In the Hill equation:

$$\theta = \frac{[L]^n}{K_D + [L]^n}$$

θ is the fraction of ligand binding sites filled, $[L]$ is the concentration of the ligand, K_D is the dissociation constant, and n is the Hill coefficient, which describes cooperativity. The Hill coefficient reflects the minimum number of binding sites in the protein; the actual number of sites could be higher than this value, but cannot be directly determined through this method.

because a lack of complete cooperativity is always present. These plots are shown in Figures 4.22-4.26. Dissociation constants were calculated for wild-type C-terminal PilY1, as well as N853A and D855A. These constants were $31 \mu\text{M} \pm 1.9$, $52 \mu\text{M} \pm 0.8$, and $62 \mu\text{M} \pm 2.4$, respectively. Terbium appeared to bind nonspecifically to both the D851A and D859A mutants, and therefore the dissociation constant could not be determined.

DISCUSSION

P782S and Y945 do not impact PilY1 function

An important set of results came from mutations selected as negative controls. Recall that these mutations were introduced into regions of the protein that did not seem to have obvious functional importance. P782 is found in a loop region on the back face of the protein, in between the second and third β -propeller blades, while the Y945 residue is located at the helices covering the WD40 repeat interaction site that is found in other β -propeller proteins. As predicted, these changes had no functional impact on the protein. Both variants were able to twitch normally, as well as express reasonable levels of pilus (PilA) and PilY1 in bacteria.

The calcium binding site is important to PilY1 function

The results of both the functional assays and the fluorescence experiments lead us to conclude that C-terminal PilY1 calcium binding is necessary to protein function. This

conclusion is drawn specifically from the results of the twitching motility assay in conjunction with the determination of calcium binding constants. Three of the mutants, N853A, D855A, and D859A, had a weak impact on twitching when harbored in bacteria. The D851A mutant in *Pseudomonas* is not able to twitch. Upon determination of dissociation constants, the wild-type PilY1 showed the lowest K_D (of 31 μM , indicating that it possesses the highest affinity for calcium), while the two other mutants capable of binding calcium had slightly increased dissociation constants (52 μM and 62 μM for N853A and D855A, respectively). Both D851A and D859A variants were not capable of specifically binding calcium. Taken together, these results show an obvious trend which seems to indicate that calcium binding affinity is directly correlated with twitching motility capability. As twitching motility is a necessary in the colonization of *Pseudomonas* biofilms on the cell surface, this result is by no means trivial.

Analysis of additional variants

Unfortunately, conclusions about many of the other mutations are not straightforward. For example, the residues we selected to mutate in the positively charged patch proved inconclusive in functional assays. Both R929A and R933A were capable of twitching like wild-type as well as making pilus and PilY1, possibly indicating that this region is not important to PilY1 function as we had originally hypothesized. However, the K934A mutant behaved unusually: it was unable to make a pilus or PilY1 but was capable of wild-type twitching motility. This is an interesting discrepancy which we will attempt to explain later on in this chapter.

Another interesting set of results are those of the dimer disruption mutants. Two of these changes were mutations in combinations of four residues found at the crystallographic dimer interface. These mutants were created with the intention of destroying this interface to see if the dimer was functionally important. We have independently confirmed the presence of the dimer in solution through the use of analytical ultracentrifugation and dynamic light scattering. However, these mutations appeared to have no effect on the functionality of the PilY1 in *Pseudomonas*. Dimer mutants 1 and 2 resulted in normal twitching and production of wild-type levels of pilus and PilY1. This result does not conclusively show that the dimer itself is irrelevant, because it these mutations may be enough to disrupt the interface. This is easily testable and thus these mutants will require further investigation, which will be outlined later on in this chapter.

The other mutant with potential to disrupt an oligomeric interface was the domain swap deletion mutant $\Delta 712-722$. This mutant lacked 11 residues that appear to connect a PilY1 C-terminal monomer to a symmetry mate in a suggested domain swap (analyzed in Chapter 3). If the protein is domain-swapped, the removal of these residues could impact the oligomerization state as well as functionality of PilY1. Interestingly, this mutant does not perform like the wild-type protein inside *Pseudomonas* cells. It could not twitch or produce pilin in bacteria, and made very little PilY1. These results seem to indicate that this region of the protein is important to PilY1 function. However, if this is true, it most likely not related to a physiological domain swap. Light scattering analysis shows that the multimeric profile of $\Delta 712-722$ PilY1 is the same as that of wild-type. It is possible that the domain swap is a crystallographic artifact; however we are fairly certain that this region is domain-swapped in the structure.

Mutations associated with the cloning mutations provided a wide and unusual array of results. Although the original cloning mutants (F670L, V880A, and P1012S) were all found to be deleterious to the function of *Pseudomonas* by our collaborators, we found that only P1012S did not behave like wild-type. Although twitching and pilin expression was slightly lower than normal in bacteria harboring both F670L and V880A, the P1012S mutant could not twitch or make pilus in the bacteria. Interestingly, this mutant makes wild-type levels of PilY1.

Analyses of the mutations associated with the cloning mutations are for the most part straightforward. Most of the mutants (P885A, H930A, D1019A, and S1021A) are consistent with normal twitching and produced wild-type levels of both PilA and PilY1 in bacteria. Neither the W965A nor W1015A changes produce PilA or PilY1; however, W965A in bacteria could twitch while W1015 could not. Again, an explanation for this discrepancy will be discussed in the following section. As at least two of these mutations (associated with the P1012S mutant) were obviously deleterious to protein function and hence, this region of the protein could be functionally relevant. As previously stated, β -propeller proteins that lack a WD40 interaction surface often bind protein partners through the use of extended loop regions, of which the P1012S and associated mutants are a part.

Mutations can be functionally divided into four categories

Analysis of the functional assays data indicate that most of the PilY1 mutations created can be categorized into four groups. The first is the “normal” group (Category I). These mutants (Y945A, H930A, R929A, R933A, P782S, D1019A, S1021A, Dimer mutant 1,

and Dimer mutant 2) have normal twitching motility profiles and produce wild-type levels of both pilus and PilY1 in bacteria. Both F670L and V880A can also be categorized in this manner, despite reduced pilin expression in both and slightly lower twitching in F670L. These mutants, and in some cases their surrounding regions, are not central to PilY1 function.

The next category of mutants is Category II, or the “dysfunctional” mutants. These mutations when harbored in bacteria produce little to no PilY1, cannot make a pilus, and are not capable of twitching motility. Mutants that fall into this category are W1015A, the double calcium binding mutant D851/859A, and the domain swap deletion $\Delta 712-722$. These variants are interesting in the sense that the residues involved are probable very important to protein function, however they are also easily understood: no pilus and no PilY1 lead to no twitching motility. In this sense these results are intuitive. Although both D851/859A and $\Delta 712-722$ produce an extremely small amount of PilY1, the mutations could be destabilizing the protein, allowing it expression of a protein lacking function.

Category III mutants are mutations that produce near wild-type levels of PilY1 but have no function: they cannot twitch and do not make a pilus. This category includes the cloning mutant P1012S and the calcium binding mutant D851A. Analysis of these mutations is problematic; although the PilY1 itself seems to be stable, the mutants do not have any associated functions. Since the Western blotting detection of PilY1 was carried out on whole cell *Pseudomonas* lysate, it is possible that the PilY1 is not being trafficked to the outside of the cell (where it would be found at the tip of the pilus) and is thus unable to perform any functional duties. This scenario is also feasible for the Category II mutants. This hypothesis is easily testable, and will be discussed in the future work.

The final category of mutants (Category IV) include those that are A) capable of twitching like wild-type and make PilY1 but produce no pilus, or B) can twitch but make no pilus and no PilY1. Category IVA mutants include W965A, N853A, and D859A, while B includes K934A and D855A. It is also interesting to note that both W965A and N853A (4A mutants) produce PilY1 that appears to be at a higher molecular weight than that of wild-type, which we are currently unable to explain. Although it is possible for *Pseudomonas* to twitch with very few pili (so few that the pilus cannot be visualized on a coomassie-stained SDS gel), we assume that these pili-depleted bacteria would not be able to twitch as well as the wild-type bacteria. This category of mutant requires further investigation in order to elucidate the mechanism behind the twitching motility.

Future experiments are required to determine mechanisms involved in reduced functionality

More experiments will be necessary to determine the mechanisms behind the functional phenomena we have observed in many of these PilY1 C-terminal mutations. As mentioned previously, an obvious experiment is to fractionate the *Pseudomonas* cells and separate them into membrane, periplasmic, cytoplasmic, and extracellular (pilus) fractions. This would tell us where the PilY1 in the whole cell lysate was located, and could give valuable clues in determining the potential stability of the PilY1 protein- is it being trafficked to the proper location, could it be unfolded?

Another experiment to measure stability is to determine the melting temperatures of the variants through circular dichroism spectropolarimetry and comparing them to that of the

wild-type PilY1. This would tell us which PilY1 mutants are less stable in solution and could explain some of the results (specifically, mutants that make PilY1 but cannot produce pilus or twitch normally). We plan on carrying out this experiment with all of the twitching-dysfunctional mutants to see if the stability correlates with function.

More specifically, we would like to investigate whether the observed oligomers of the protein *in vitro* are important to protein function. As was stated previously, we are not certain if we are disrupting the crystallographic dimer to the point of keeping the monomers apart. We have a new dimer disruption construct being created to interrupt all electrostatic interactions by adding repulsive charges to five of the interface residues. Analytical Ultracentrifugation or light scattering experiments will be used to see if the dimer is being destroyed by the new mutations as well as the older dimer mutant constructs. We will also perform AUC on the domain swap deletion mutant to see if the results agree with those from the light scattering experiment.

The most exciting results we have acquired are those of the calcium binding residue mutations. We can say with certainty through FRET and the twitching motility assay that there is a direct correlation between calcium binding affinity and twitching motility. We would like to pursue these mutants through two additional functional assays. One of these involves the ability of bacteria harboring the mutants to attach to lung epithelial cells. Recall from Chapter 1 (Figure 1.4) that a strains containing wild-type PilY1 adheres to these cells, while a PilY1 null mutant was unable to do so. This is the obvious experiment to determine whether or not these mutations have caused the bacteria to lose its infectivity. In addition, we would eventually like to use a mouse model to infect mice with both the wild-type and calcium binding mutant PilY1 *Pseudomonas* to see if the mutants are again unable to cause

infection in a living organism. Proving that the calcium binding site is necessary for PilY1 function and hence bacterial infectivity is an important goal and is integral to determining the mechanism and function of PilY1 in *Pseudomonas*.

REFERENCES

1. Mattick, J.S., *Type IV pili and twitching motility*. Annu Rev Microbiol, 2002. **56**: p. 289-314.
2. Pizarro-Cerda, J. and P. Cossart, *Bacterial adhesion and entry into host cells*. Cell, 2006. **124**(4): p. 715-27.
3. Morales, V.M., A. Backman, and M. Bagdasarian, *A series of wide-host-range low-copy-number vectors that allow direct screening for recombinants*. Gene, 1991. **97**(1): p. 39-47.
4. Wolfgang, M., H.S. Park, S.F. Hayes, J.P. van Putten, and M. Koomey, *Suppression of an absolute defect in type IV pilus biogenesis by loss-of-function mutations in pilT, a twitching motility gene in Neisseria gonorrhoeae*. Proc Natl Acad Sci U S A, 1998. **95**(25): p. 14973-8.
5. Wilkins, A.L., Y. Ye, W. Yang, H.W. Lee, Z.R. Liu, and J.J. Yang, *Metal-binding studies for a de novo designed calcium-binding protein*. Protein Eng, 2002. **15**(7): p. 571-4.
6. Carrer, C., M. Stolz, E. Lewitzki, C. Rittmeyer, B.O. Kolbesen, and E. Grell, *Removing coordinated metal ions from proteins: a fast and mild method in aqueous solution*. Anal Bioanal Chem, 2006. **385**(8): p. 1409-13.
7. Noble, S.M., V.E. Carnahan, L.B. Moore, T. Luntz, H. Wang, O.R. Ittoop, J.B. Stimmel, P.R. Davis-Searles, R.E. Watkins, G.B. Wisely, E. LeCluyse, A. Tripathy, D.P. McDonnell, and M.R. Redinbo, *Human PXR forms a tryptophan zipper-mediated homodimer*. Biochemistry, 2006. **45**(28): p. 8579-89.

FIGURE LEGENDS

Figure 4.1 The mutants are first shown in the context of the whole protein, with an additional close-up shot. Both the P1012 and V880 residues are located within loop regions, while F670L is in a helix.

Figure 4.2 Twitching motility of the *Pseudomonas* with mutations made in the positively charged patch region of the C-terminal domain of PilY1. When viewed in relation to the two positive controls (PAK and PilY1+) the mutants are all capable of wild-type twitching or better

Figure 4.3 This figure shows the SDS-PAGE gel bands at 17 kDA, used for detection of the PilA protein. The presence of PilA is used to indicate that the bacteria are producing a pilus. While both R929A and R933A are producing close to wild-type levels of pilin, K934A shows no sign of a pilus.

Figure 4.4 This figure shows the Western blot analysis of the positively charged patch mutants to see if they could produce PilY1. When compared to the PAK and PilY1+ controls, both R929A and R933A produced wild-type levels of PilY1, while K934A does not produce any.

Figure 4.5 The bar graph shows the twitching motility of the *Pseudomonas* with mutations made in the calcium binding region of the C-terminal domain of PilY1. When viewed in relation to the two positive controls (PAK and PilY1+), only the N853 mutant is capable of wild-type twitching. Twitching motility is down to half that of wild-type in both D855A and D859A, while D851A and the double mutant D851/859A cannot twitch.

Figure 4.6 SDS-PAGE gel bands at 17 kDA, used for detection of the PilA protein. The presence of PilA is used to indicate that the bacteria are producing a pilus. None of these five mutants are capable of producing pilin.

Figure 4.7 Western blot analysis of the calcium binding mutants to see if they could produce PilY1. When compared to the PAK and PilY1+ controls, both D851A and D859A produced about half of the wild-type levels of PilY1, while N853A, D855A, and the double mutant (D851/859A) produces almost no detectable PilY1.

Figure 4.8 Twitching motility of the *Pseudomonas* with mutations made in the dimer and domain swapped regions of the C-terminal domain of PilY1. When viewed in relation to the two positive controls (PAK and PilY1+), both of the dimer mutants are capable of wild-type twitching. Twitching motility is not seen in the domain swap deletion mutant.

Figure 4.9 SDS-PAGE gel bands at 17 kDA, used for detection of the PilA protein. The presence of PilA is used to indicate that the bacteria are producing a pilus. Both of the dimer mutants are able to produce wild-type levels of pilin, while the domain swap deletion mutant does not produce any pilin.

Figure 4.10 Western blot analysis of the dimer and domain swap deletion mutants to see if they could produce PilY1. When compared to the PAK and PilY1+ controls, both the dimer 1 and dimer 2 mutants produced wild-type levels of PilY1, while $\Delta 712-722$ produces very little PilY1.

Figure 4.11 Twitching motility of the *Pseudomonas* with the cloned mutations found in the C-terminal domain of PilY1. When viewed in relation to the two positive controls (PAK and PilY1+), both F670L and V880A are capable of near wild-type twitching. Twitching motility is not seen in P1012S.

Figure 4.12 SDS-PAGE gel bands at 17 kDA, used for detection of the PilA protein. The presence of PilA is used to indicate that the bacteria are producing a pilus. Two of the cloning mutants F670L and V880A are able to produce wild-type levels of pilin, while P1012S does not produce any pilin.

Figure 4.13 Western blot analysis of the cloning mutants to see if they could produce PilY1. When compared to the PAK and PilY1+ controls, all of the cloning mutants are capable of making wild-type levels of PilY1.

Figure 4.14 Twitching motility of the *Pseudomonas* with mutations associated with the cloned mutants found in the C-terminal domain of PilY1. When viewed in relation to the two positive controls (PAK and PilY1+), almost all of the mutants are capable of wild-type twitching. Twitching motility is not seen in W1015A.

Figure 4.15 SDS-PAGE gel bands at 17 kDA, used for detection of the PilA protein. The presence of PilA is used to indicate that the bacteria are producing a pilus. Two of the cloning associated mutants W965A and W1015 are unable to produce wild-type levels of pilin, while the rest have normal levels.

Figure 4.16 Western blot analysis of the cloning associated mutants to see if they were capable of producing PilY1. When compared to the PAK and PilY1+ controls, all of the cloning mutants but W965A and W1015A are capable of making wild-type levels of PilY1.

Figure 4.17 Twitching motility of the *Pseudomonas* with mutations not expected to impact function that are found in the C-terminal domain of PilY1. When viewed in relation to the two positive controls (PAK and PilY1+), both Y945A and P782S are capable of wild-type twitching.

Figure 4.18 SDS-PAGE gel bands at 17 kDA, used for detection of the PilA protein. The presence of PilA is used to indicate that the bacteria are producing a pilus. The two of the mutants P782S and Y945A which are not expected to impact function are able to produce near wild-type levels of pilin.

Figure 4.19 Western blot analysis of the mutants not expected to impact function, to see if they could produce PilY1. When compared to the PAK and PilY1+ controls, Y945A is capable of making near wild-type levels of PilY1, while P782S PilY1 expression is slightly lower.

Figure 4.20 The AUC data was collected at 280 nm and fit to a monomer-dimer equilibrium model with a dissociation constant of 27 μ M.

Figure 4.21 Light scattering analysis was carried out using a multi-angle light scattering system in conjunction with a size exclusion column. The peaks chosen for analysis are bracketed by black bars. Light scattering is shown in red, while the refractive index is shown in blue. The peaks were analyzed and determined to be (left to right) tetramer, dimer, and monomer species, with an estimated percentage of 5, 75, and 20%, respectively.

Figure 4.22 Light scattering analysis was carried out using a multi-angle light scattering system in conjunction with a size exclusion column. The peaks chosen for analysis are bracketed by black bars. Light scattering is shown in red, while the refractive index is shown in blue. The peaks were analyzed and determined to be (left to right) tetramer, dimer, and monomer species, with an estimated percentage of 5, 80, and 15%, respectively. This is in agreement with the wild-type C-terminal PilY1.

Figure 4.23 This graph shows the binding curve of wild-type C-terminal PilY1 and terbium chloride, a calcium substitute. The plot of fractional change in terbium fluorescence versus terbium concentration was fit to a three-parameter Hill equation, in which the calculated dissociation constant was 31 μM . This fit gives a Hill coefficient of 2.4 ± 0.3 , indicating that two binding sites are occupied, or a dimer is present.

Figure 4.24 This graph shows the binding curve of the D851A PilY1 mutant and terbium chloride, a calcium substitute. The plot of fractional change in terbium fluorescence versus terbium concentration seemed to show non-specific binding and could not be properly analyzed to provide a dissociation constant.

Figure 4.25 This graph shows the binding curve of the N853A mutant PilY1 and terbium chloride, a calcium substitute. The plot of fractional change in terbium fluorescence versus terbium concentration was fit to a five-parameter Hill equation, in which the calculated dissociation constant was 52 μM .

Figure 4.26 This graph shows the binding curve of the D855A mutant PilY1 and terbium chloride, a calcium substitute. The plot of fractional change in terbium fluorescence versus terbium concentration was fit to a four-parameter Hill equation, in which the calculated dissociation constant was 62 μM .

Figure 4.27 This graph shows the binding curve of the D859A PilY1 mutant and terbium chloride, a calcium substitute. The plot of fractional change in terbium fluorescence versus terbium concentration seemed to show non-specific binding and could not be properly analyzed to provide a dissociation constant.

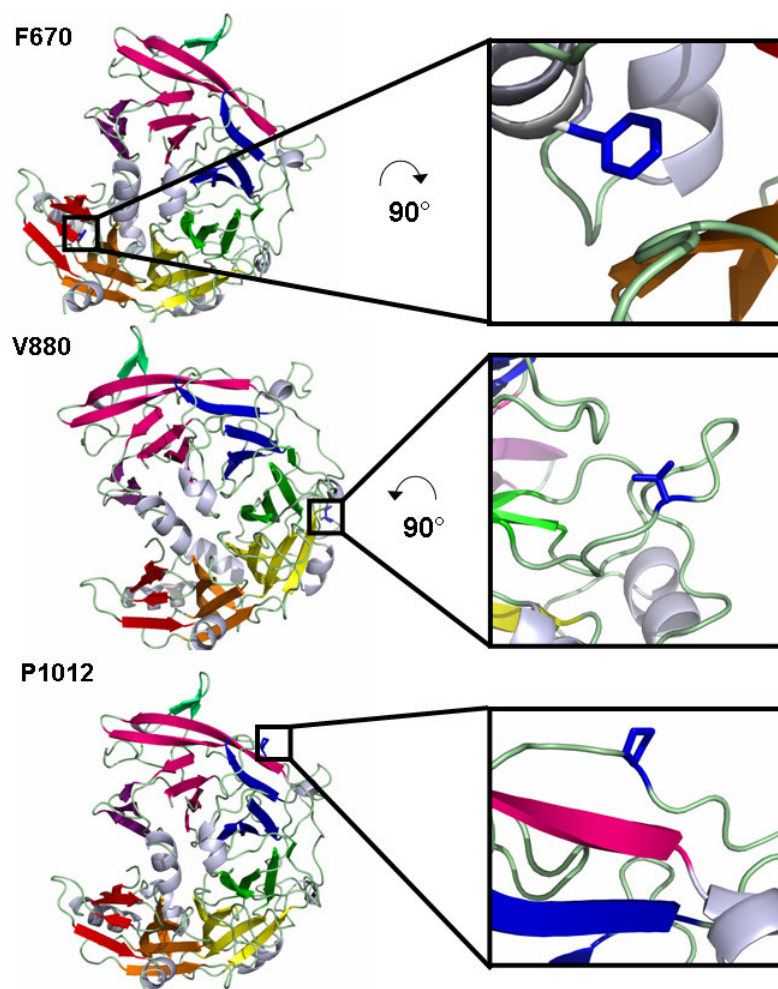


Figure 4.1 Locations of cloning mutants in PiLY1 C-terminal domain

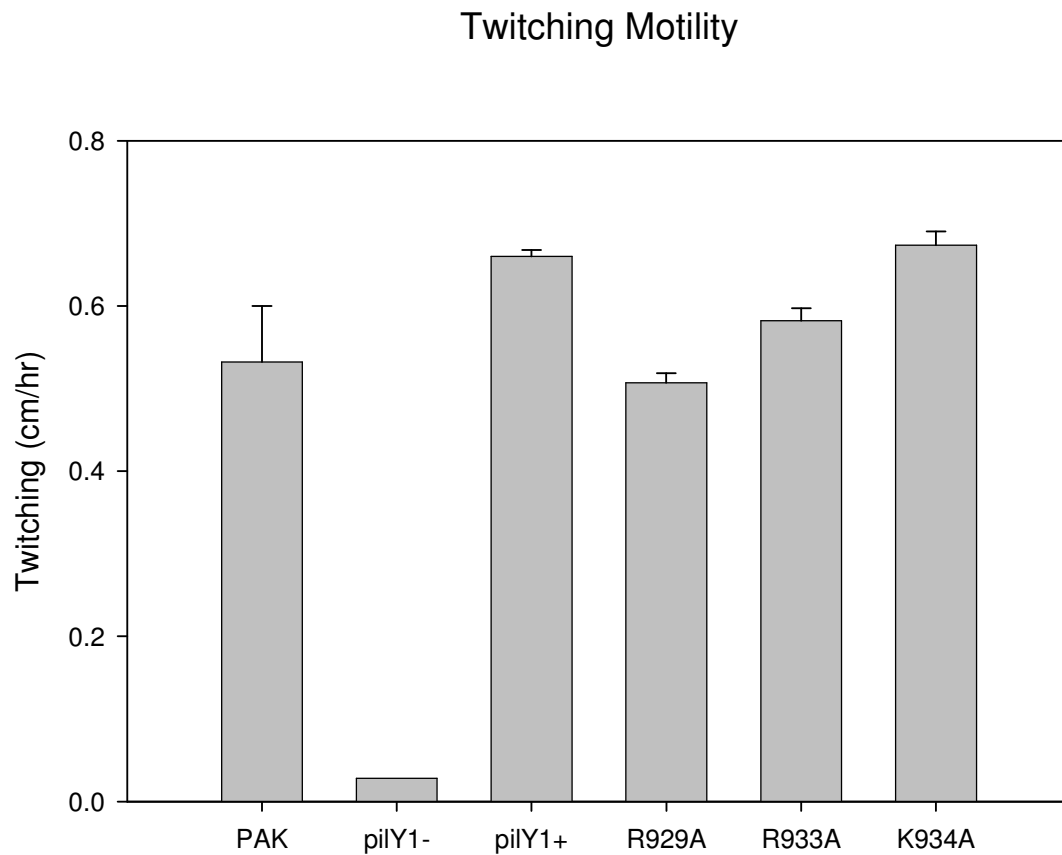


Figure 4.2 **Twitching motility of positive charge patch mutations**

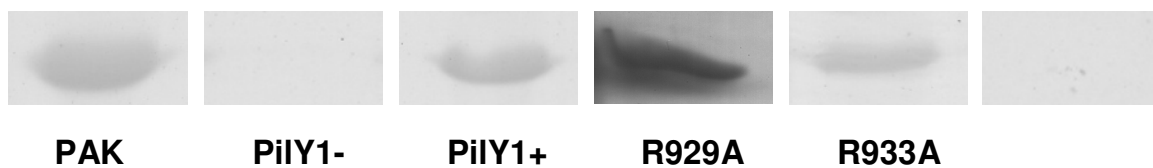


Figure 4.3 SDS-PAGE analysis of pilus presence in *Pseudomonas aeruginosa* with positive charge mutants

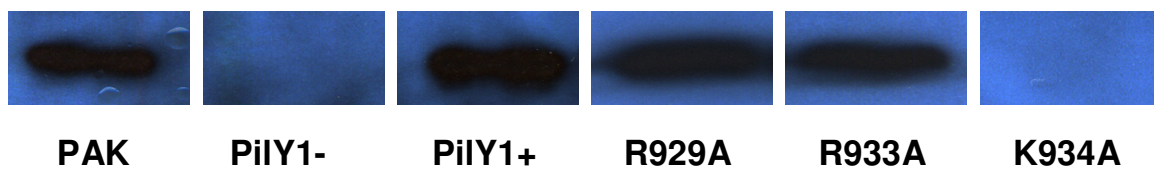


Figure 4.4 Western blots of whole cell *Pseudomonas aeruginosa* for detection of PilY1 in positive charge patch mutants

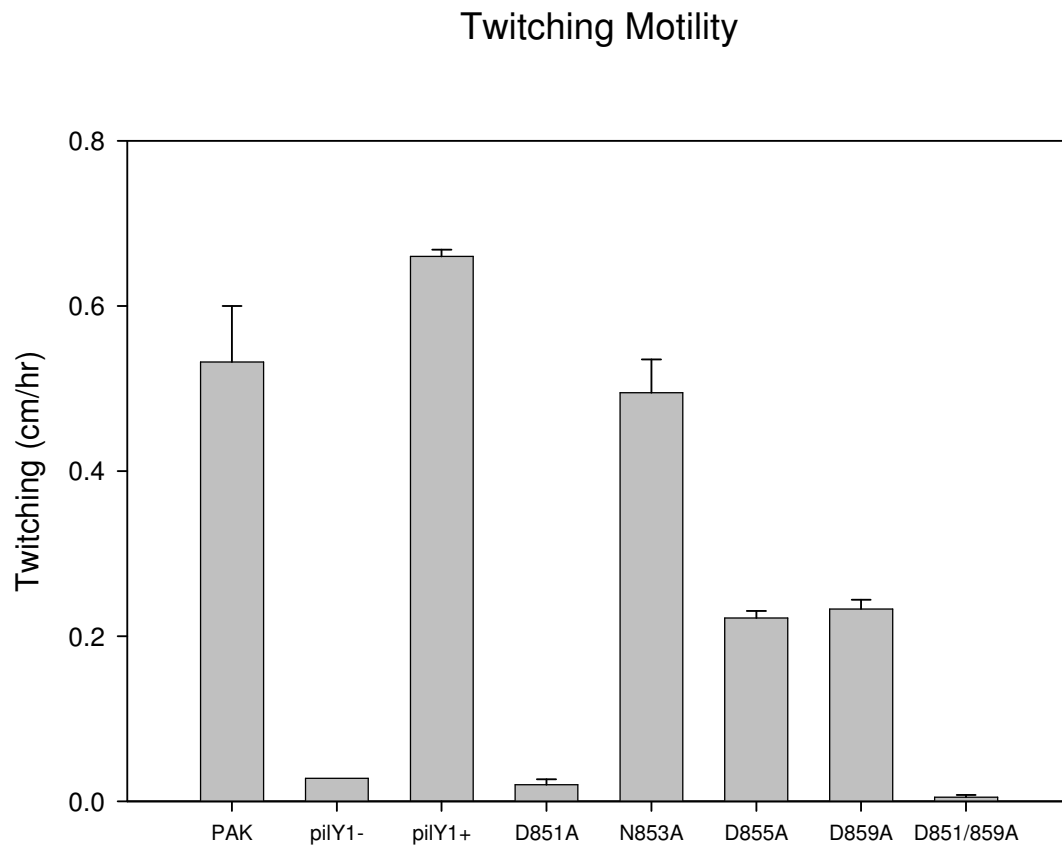


Figure 4.5 Twitching motility of calcium binding ligand mutations

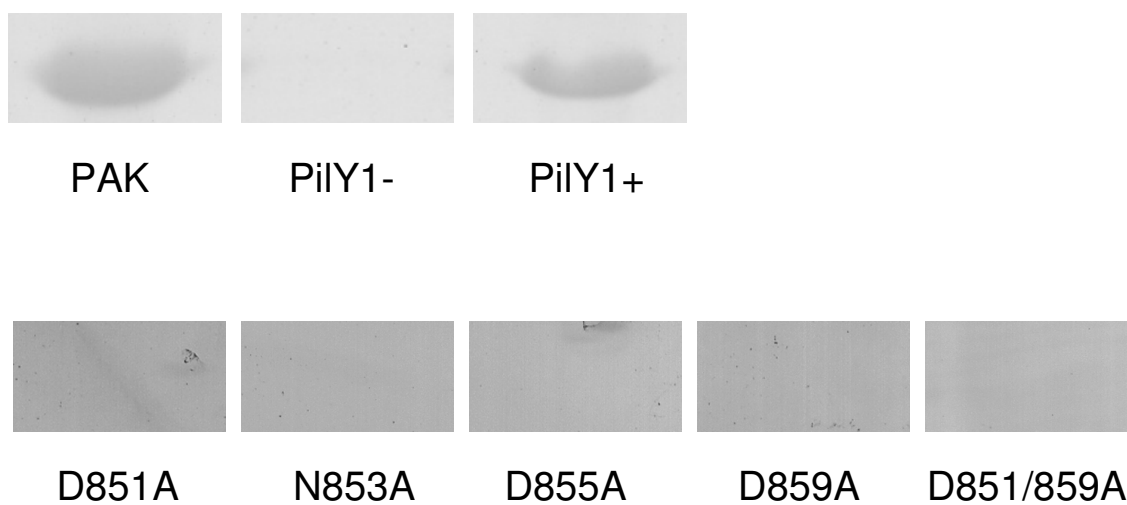


Figure 4.6 SDS-PAGE analysis of pilus presence in *Pseudomonas aeruginosa* with calcium binding mutants

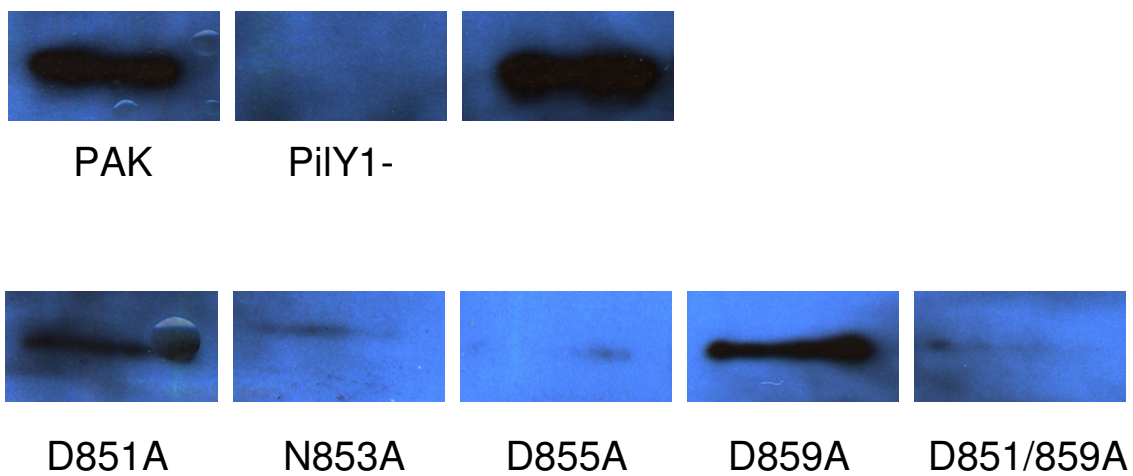


Figure 4.7 Western blots of whole cell *Pseudomonas aeruginosa* for detection of PilY1 in calcium binding mutants

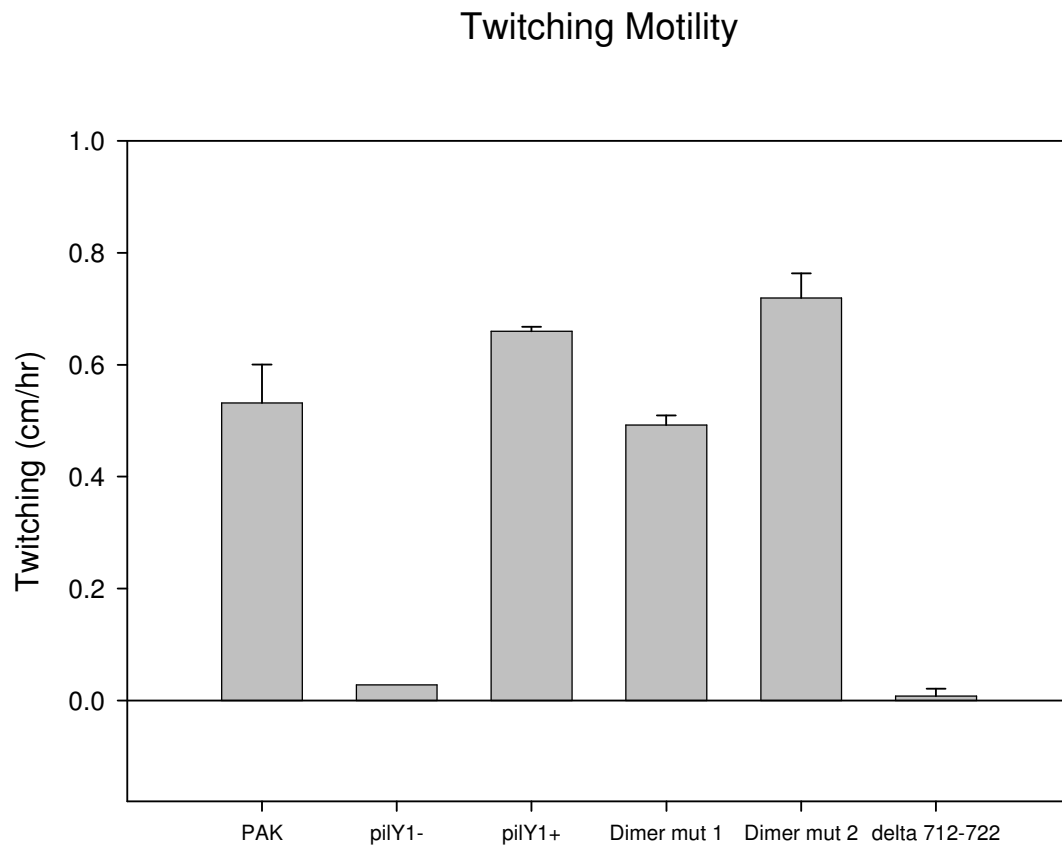


Figure 4.8 **Twitching motility of dimer and domain swap mutants**

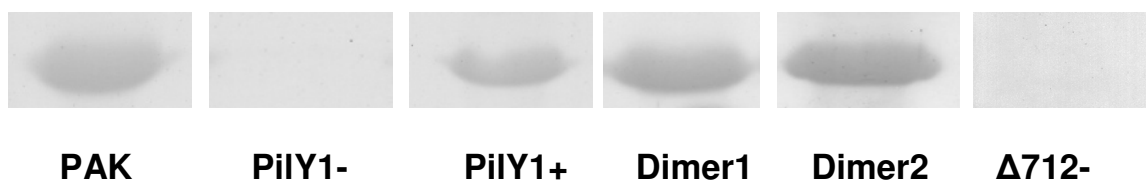


Figure 4.9 SDS-PAGE analysis of pilus presence in *Pseudomonas aeruginosa* with dimer and domain swap mutants

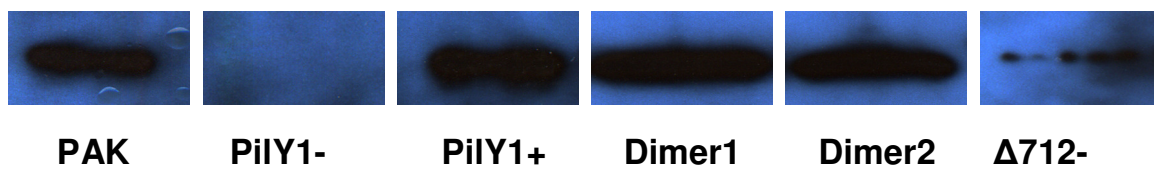


Figure 4.10 Western blots of whole cell *Pseudomonas aeruginosa* for detection of PilY1 in dimer and domain swap mutants

Twitching Motility

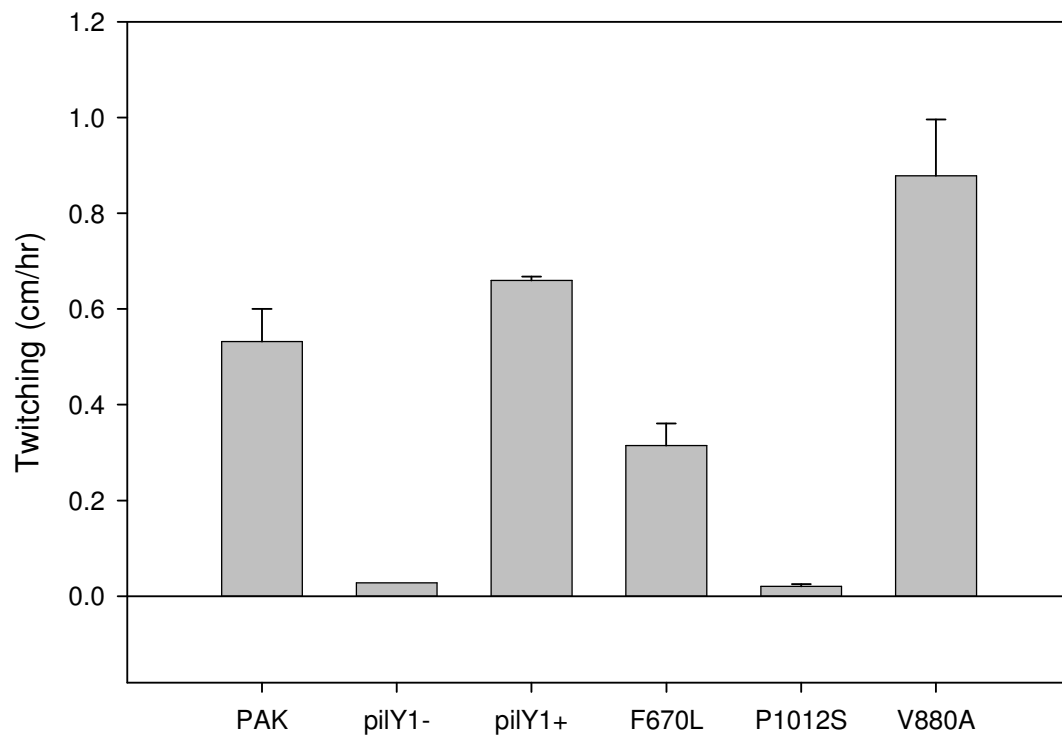


Figure 4.11 Twitching motility of cloning mutations

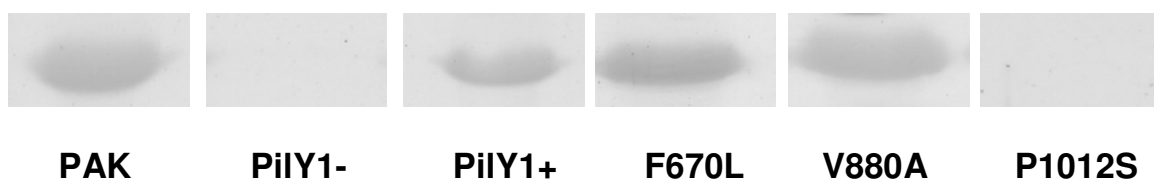


Figure 4.12 SDS-PAGE analysis of pilus presence in *Pseudomonas aeruginosa* with cloning mutants

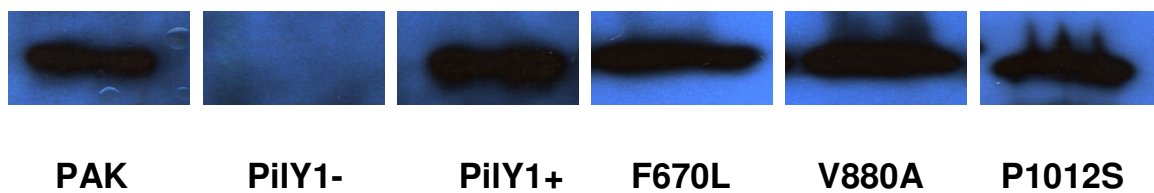


Figure 4.13 Western blots of whole cell *Pseudomonas aeruginosa* for detection of PilY1 in cloning mutants

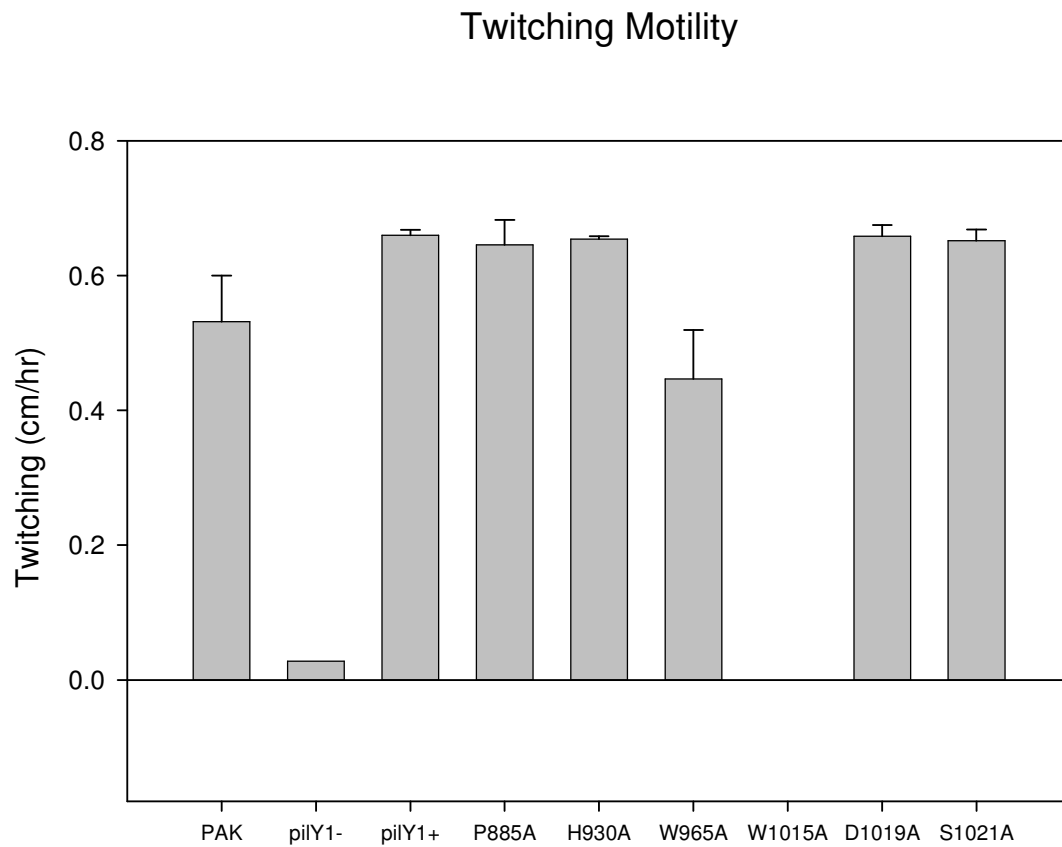


Figure 4.14 Twitching motility of mutations associated with cloning mutants

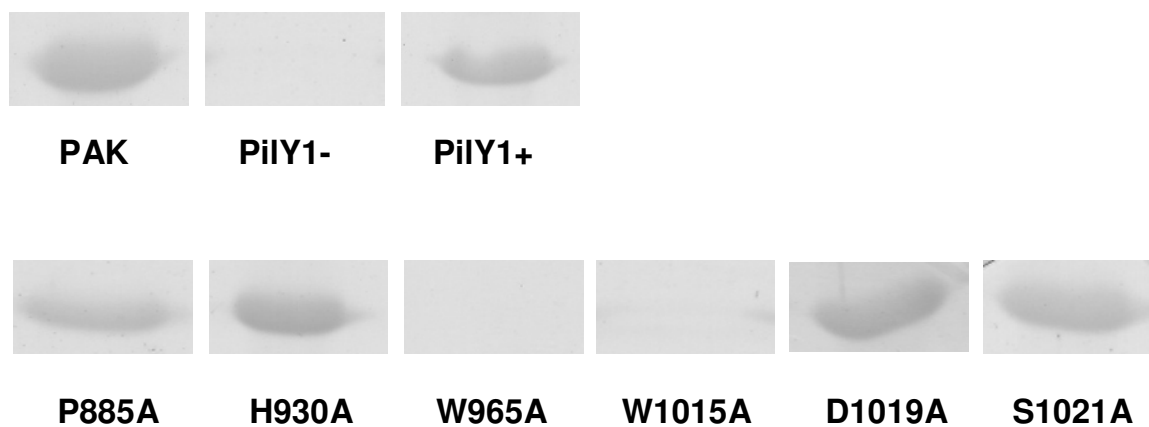


Figure 4.15 SDS-PAGE analysis of pilus presence in *Pseudomonas aeruginosa* with cloning associated mutants

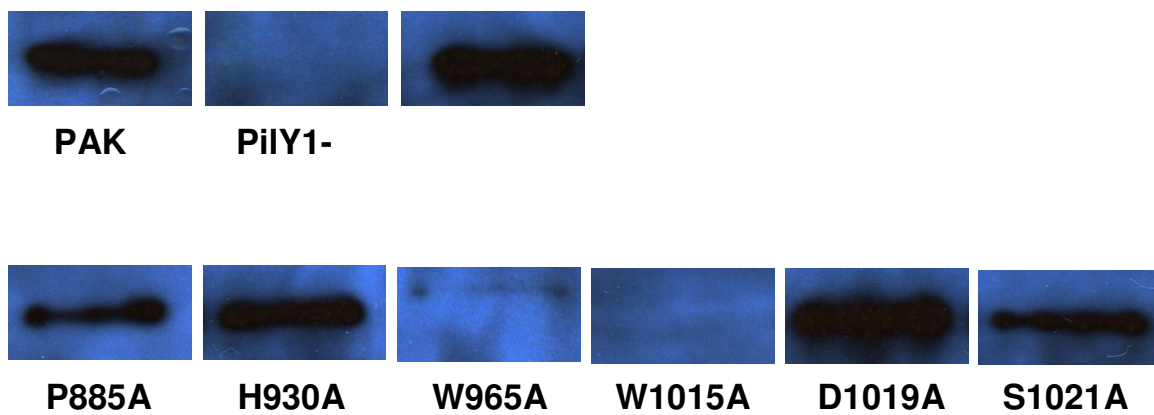


Figure 4.16 Western blots of whole cell *Pseudomonas aeruginosa* for detection of PilY1 in cloning associated mutants

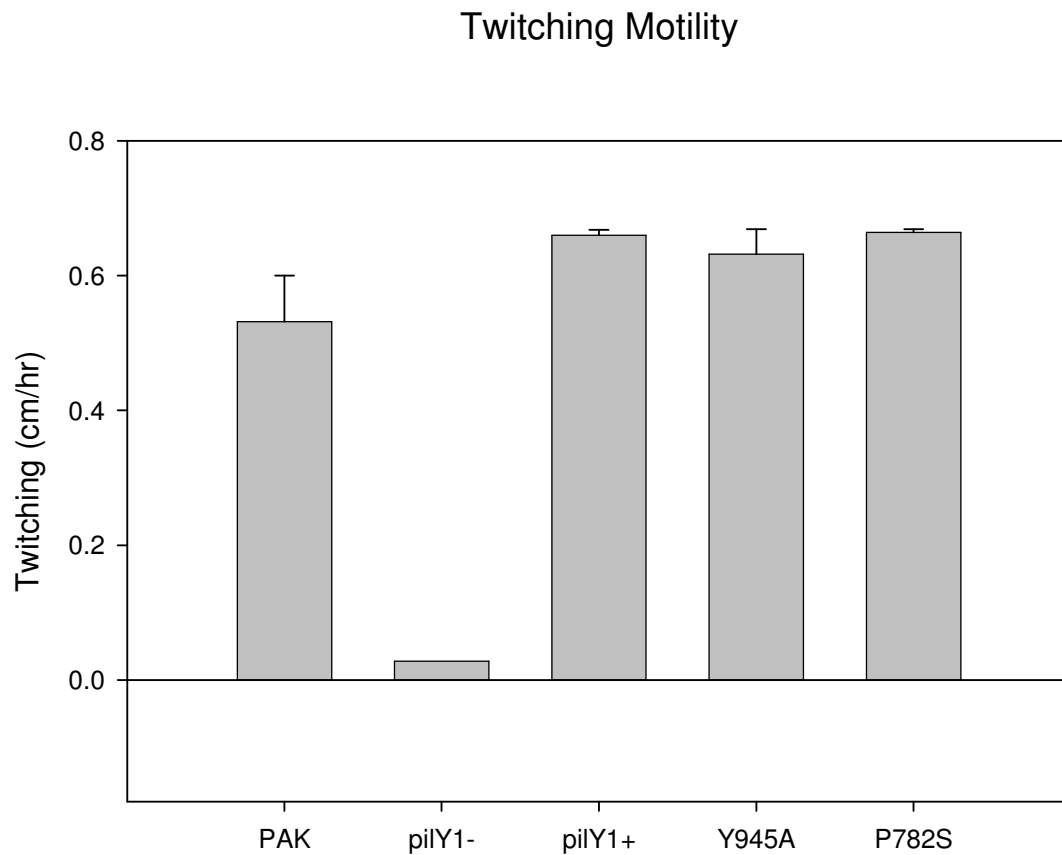


Figure 4.17 Twitching motility of mutants not expected to impact function



Figure 4.18 SDS-PAGE analysis of pilus presence in *Pseudomonas aeruginosa* with mutants not expected to impact function

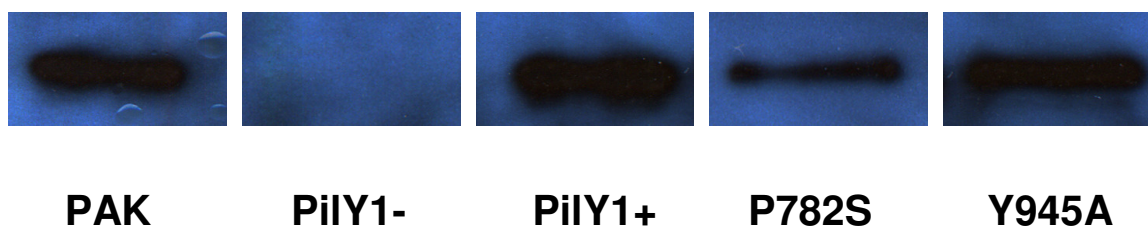


Figure 4.19 Western blots of whole cell *Pseudomonas aeruginosa* for detection of PilY1 in mutants not expected to impact function

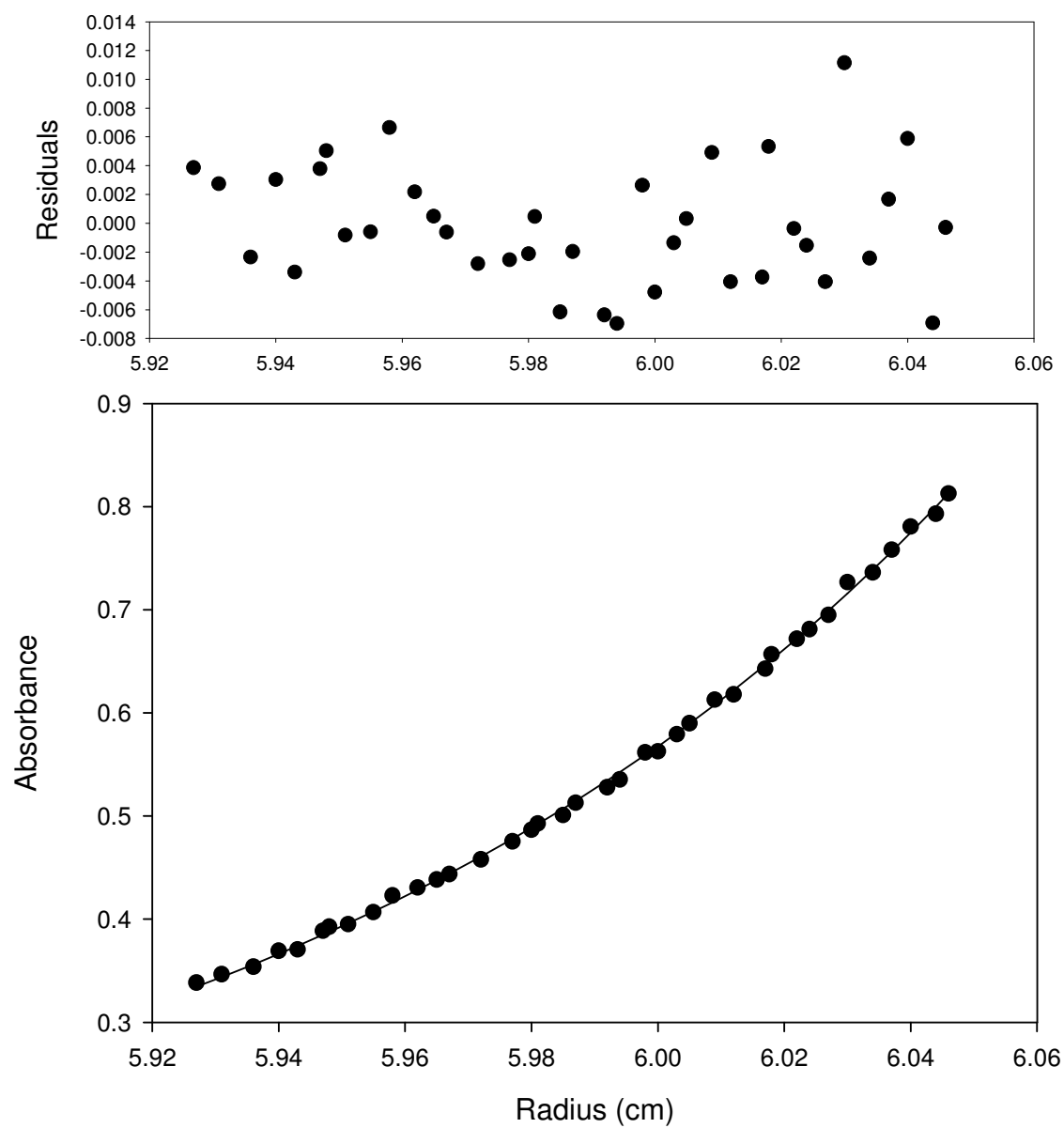


Figure 4.20 Analytical Ultracentrifugation shows the PilY1 C-terminal domain is a dimer

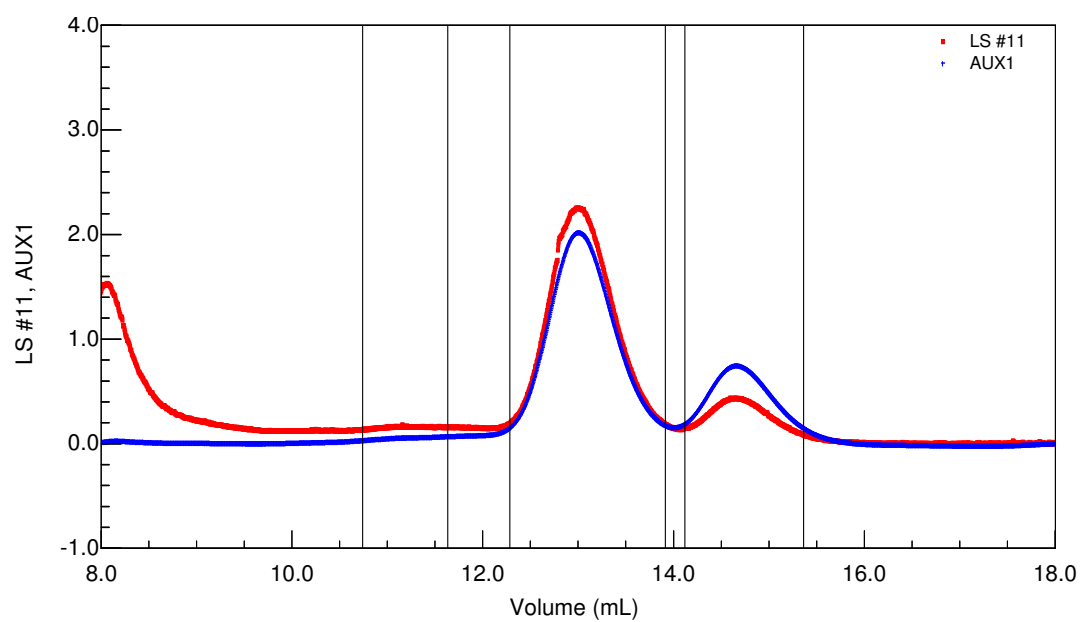


Figure 4.21 Dynamic light scattering analysis shows that C-terminal PilY1 is mainly found in dimeric form

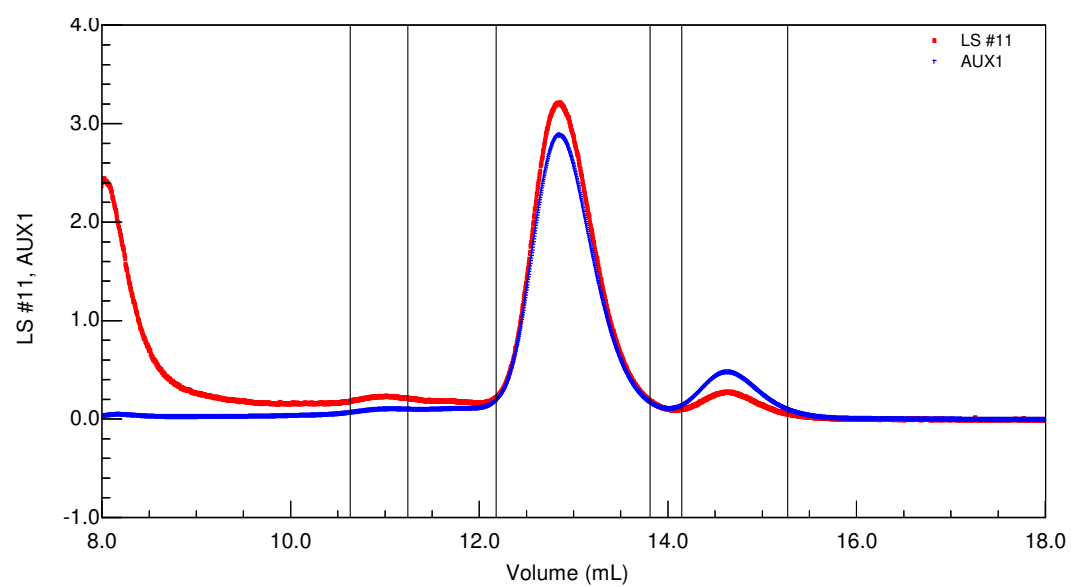


Figure 4.22 Dynamic light scattering analysis shows that C-terminal PilY1 domain swap truncation is mainly found in dimeric form

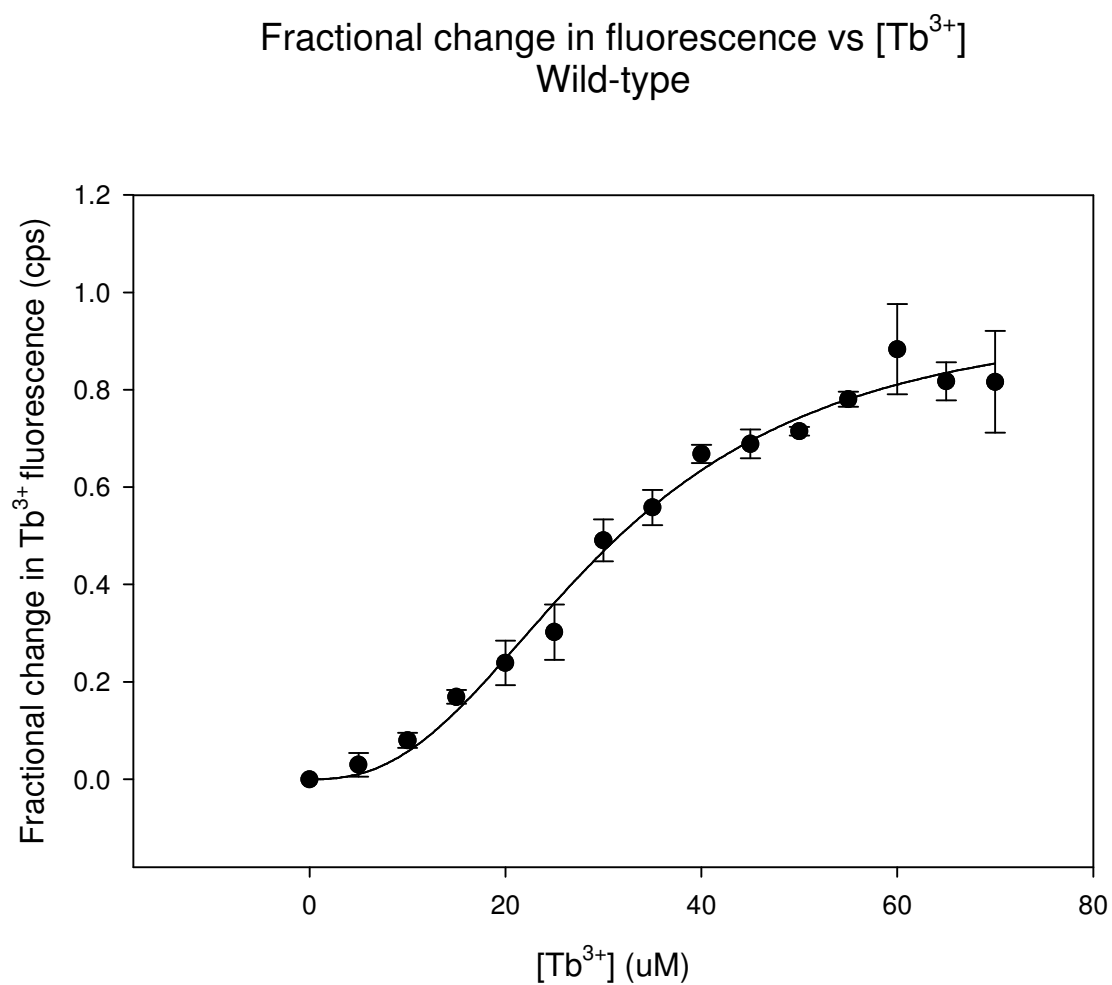


Figure 4.23 FRET analysis of wild-type C-terminal PilY1 association with terbium

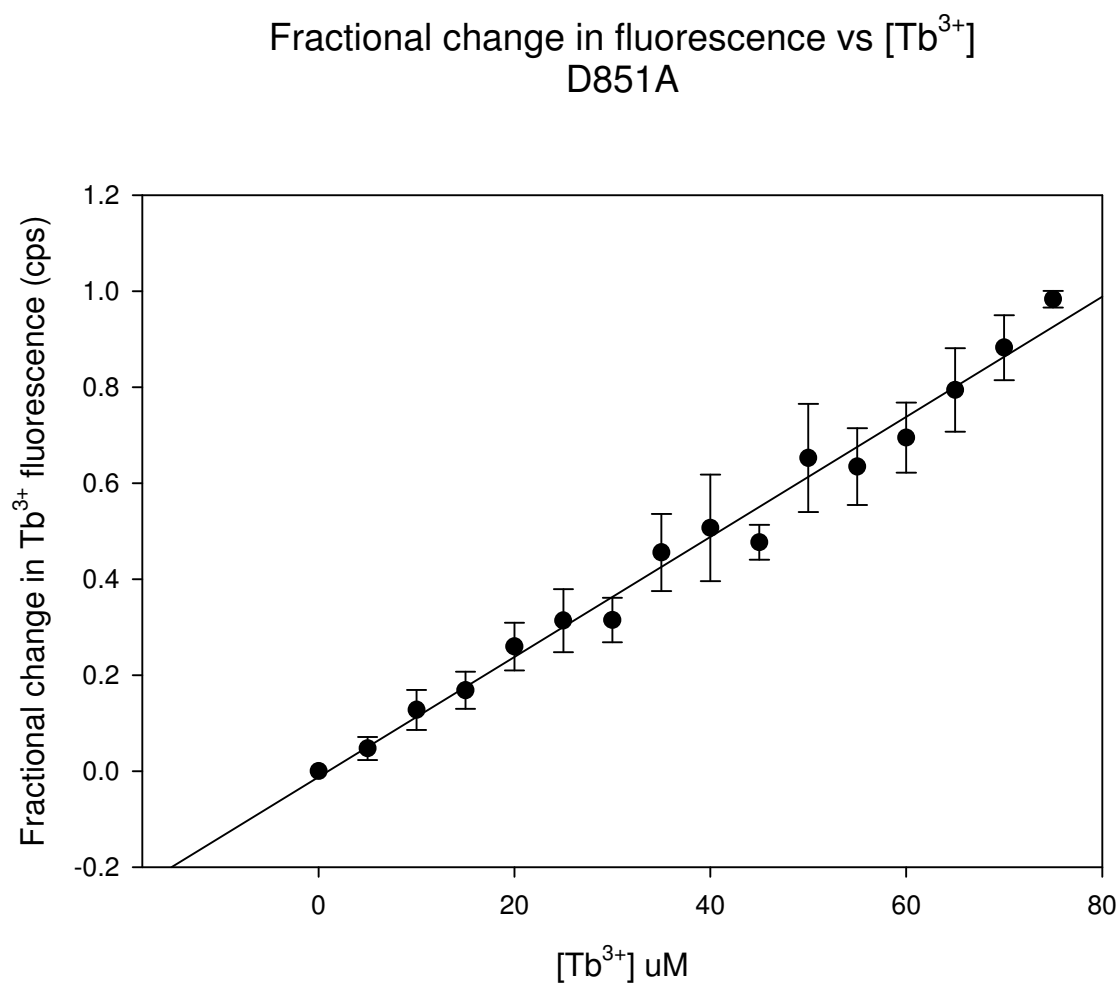


Figure 4.24 FRET analysis of calcium binding mutant D851A and its association with terbium

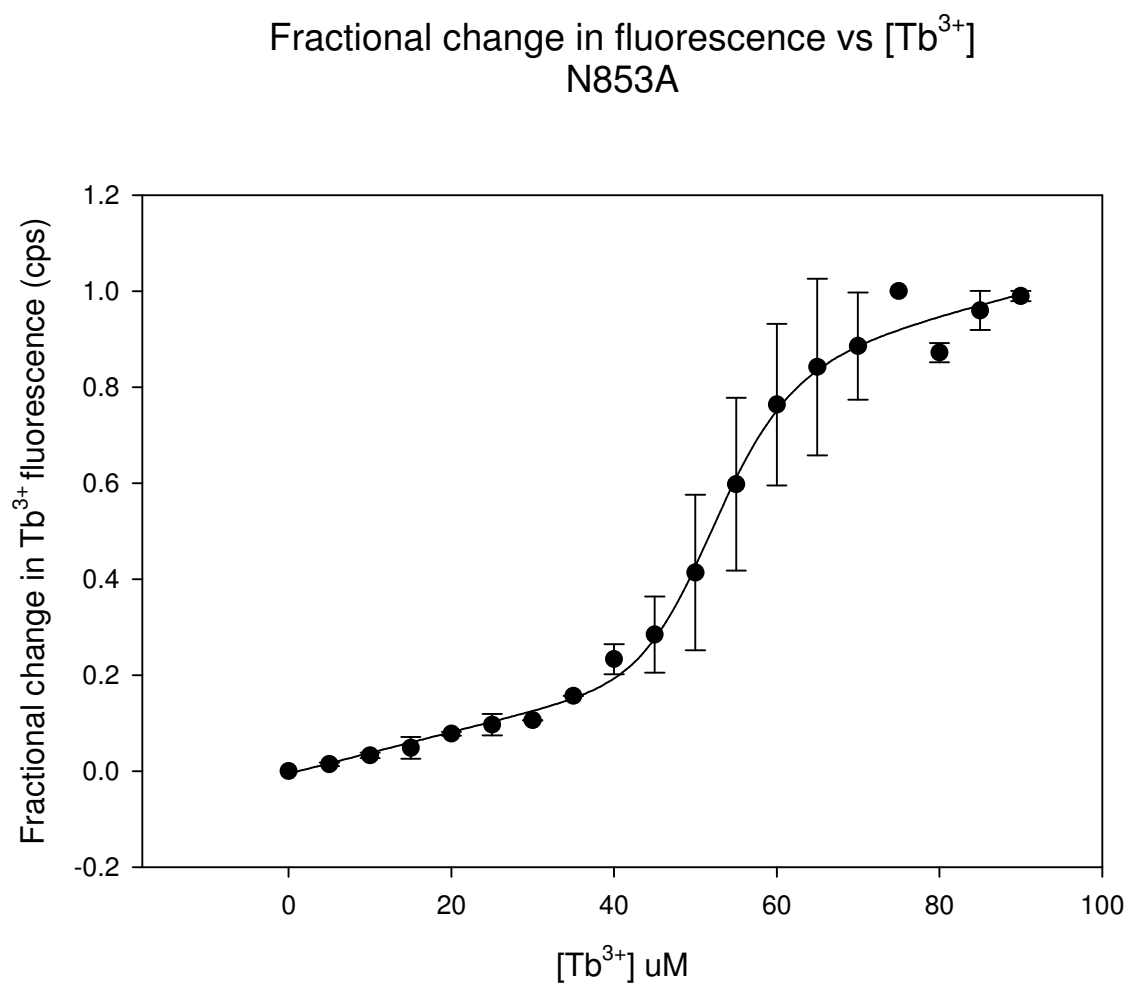


Figure 4.25 FRET analysis of binding mutant N853A and its association with terbium

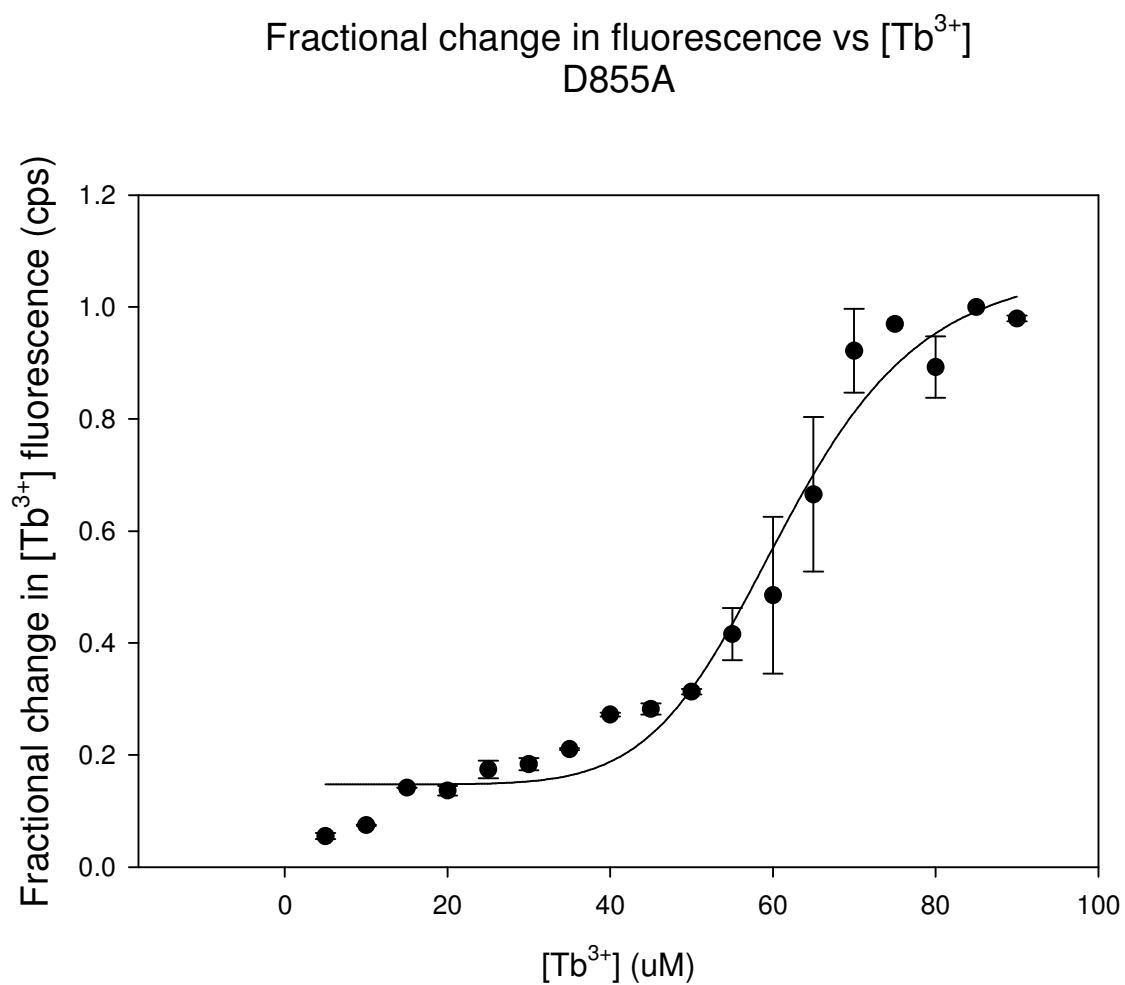


Figure 4.26 FRET analysis of binding mutant D855A and its association with terbium

Fractional change in fluorescence vs $[Tb^{3+}]$
D859A

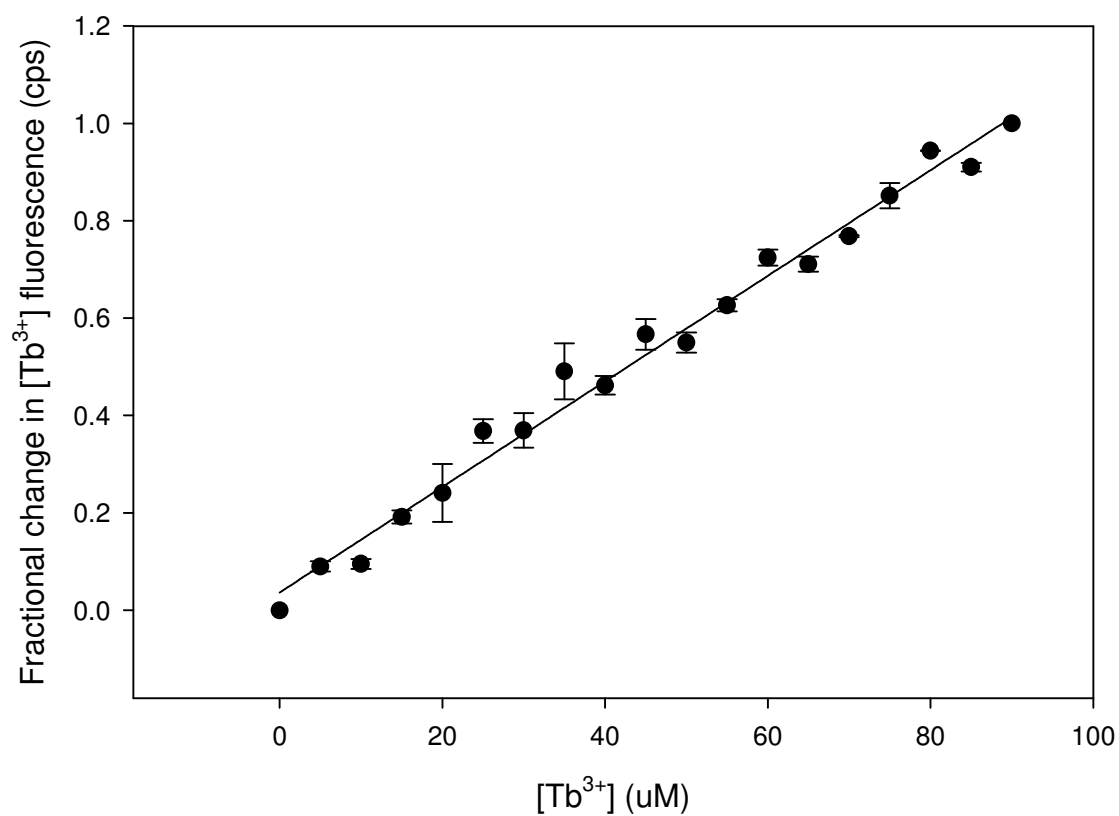


Figure 4.27 FRET analysis of binding mutant D859A and its association with terbium

Chapter 5:
The Nuclear Xenobiotic Receptor PXR: Recent Insights and New Challenges

Jillian Orans¹, Denise G. Teotico¹ and Matthew R. Redinbo^{1,2, *}

¹Department of Chemistry, University of North Carolina at Chapel Hill

²Department of Biochemistry and Biophysics, Program in Molecular Biology and
Biotechnology, and the Lineberger Comprehensive Cancer Center, University of North
Carolina at Chapel Hill

Published in Molecular Endocrinology

2005 Dec; 19(12):2891-2900

ABSTRACT

The nuclear receptor PXR plays a key but structurally enigmatic role in human biology. This ligand-regulated transcription factor responds to a wide array of chemically-distinct ligands, including many endogenous compounds and clinical drugs, and regulates the expression of a critical set of protective gene products involved in xenobiotic and endobiotic metabolism. The structural basis of this receptor's remarkable and unique promiscuity is just now coming into focus. We examine the importance of mobile regions novel to the nuclear receptor ligand binding domain fold in PXR's ability to respond to a variety of small and large agonists. We also review the functional roles played by PXR in various biological pathways, and outline emerging areas for the future examination of this key nuclear xenobiotic receptor.

INTRODUCTION

The nuclear pregnane X receptor (PXR; NR1I2, also known as SXR and PAR; [1-3] is a member of the nuclear receptor (NR) family of ligand-dependent transcriptional factors and a key regulator of genes involved in xenobiotic and endobiotic metabolism. PXR was assigned the role of detecting endogenous pregnanes by Kliewer *et al.*[4], but has subsequently been adopted as a central xenobiotic receptor that responds to many clinical drugs. PXR functions as a heterodimer with RXR α and binds to a variety of response elements (DR-3, DR-4, DR-5, ER-6, ER-8) in the promoter regions of target genes. Its moderate basal activity and up-regulation of transcriptional events are mediated by recruitment of coactivators of the p160 family (*e.g.*, SRC, GRIP); similarly, its repression of gene expression involves physical contacts with transcriptional corepressors. The subcellular localization of the receptor between the cytoplasm and nucleus, however, is emerging as an important feature in receptor, as are posttranslational modifications and tissue- and ligand-specific recruitment of transcriptional coregulators. We review recent advances in our understanding of PXR function and structure, and present some key challenges for future studies of this nuclear xenobiotic receptor.

PXR FUNCTION

PXR in Xenobiotic and Endobiotic Detection

The cytochromes P450 (CYP450s) are heme-containing mono-oxygenases involved in endobiotic and xenobiotic clearance, including the elimination of therapeutic drugs [5].

PXR is expressed predominantly in the liver and is activated by a variety of structurally-distinct ligands that are known to induce the expression of CYP450 genes central to drug metabolism. These compounds include phenobarbital, rifampicin, dexamethasone, nifedipine, taxol, and hyperforin, the active agent of the herbal remedy St. John's wort [3, 4, 6]. Phase I drug metabolism genes regulated by PXR include several CYP450s (*e.g.*, CYP3A4, CYP2B6, CYP2C8, CYP2C9, and CYP2C19), carboxylesterases, and dehydrogenases, as well as enzymes involved in heme production and the P450 reaction cycle [4, 6-13]. Indeed, PXR has been termed the master regulator of the expression of CYP3A4, which metabolizes more than 50% of human drugs. PXR also controls the expression of the Phase II drug metabolism genes encoding UDP-glucuronosyltransferases and glutathione-S-transferases (GSTs), [14-18], and Phase III drug efflux pumps like MDR1 and MRP2 [7, 19-21]. Thus, PXR is an important and efficient regulator of the expression of genes involved in all phases of drug metabolism and excretion.

PXR is also activated by a variety of endogenous ligands, including pregnanes, bile acids, hormones, and dietary vitamins [1, 4]. In response to bile acids and oxysterols, PXR regulates the expression of genes involved in bile acid metabolism and transport, including CYP7A, Oatp2, and HMG CoA synthases [22-24]. These data, and subsequent studies in animal models of cholestatic liver disease, have established that PXR plays a critical role in cholesterol homeostasis and in protecting tissues from potentially toxic endobiotics [25, 26].

Species-Specific Activation

Like most members of the nuclear receptor superfamily, PXR contains a DNA binding domain (DBD) connected by a presumably flexible hinge region to a ligand binding

domain (LBD), which contains the ligand-dependent activation function (AF-2) region. Unlike most nuclear receptors, however, the LBDs of PXR from different species exhibit significant sequence divergence. For example, mammalian isoforms of PXR share < 80% sequence identity within their LBDs compared to >90% within their DBD domains. Although each of these PXR is promiscuous in terms of ligand binding (responding to compounds of varying size, shape and chemical composition), each is also relatively specific to certain regions of chemical space. This feature of PXR activation has been termed “directed promiscuity”[27].

Kocarak *et al.* first noted striking interspecies differences in cytochrome P450 gene expression in response to known CYP3A inducers, such as the antiglucocorticoid pregnenolone 16- α -carbonitrile (PCN) and the macrolide antibiotic rifampicin [28, 29]. In-cell trans-species gene transfer studies later determined that the differential CYP3A gene expression found in rats, rabbits and humans was not derived from differences in CYP3A sequence, but rather from some other factor [30, 31]. After the initial cloning and characterization of mouse and human PXR, it was found that both forms of the receptor were not only activated by many of the compounds known to regulate CYP3A gene expression, but also that they could also bind to response elements in the promoter regions of several CYP3A genes. The mouse form of PXR was strongly activated by both PCN and dexamethasone [4], while human PXR was efficiently activated by dexamethasone, rifampicin, and RU486, another antiglucocorticoid [3, 6].

Subsequent transient-transfection experiments established that the regulation of these genes was dependent upon the activation of PXR, and that the activation profiles of the various forms of PXR were remarkably different [31]. For example, rabbit PXR could be

activated by rifampicin to induce CYP3A gene expression while rat PXR could not. Additionally, equal concentrations of PCN could effectively activate rat PXR, but not the rabbit form [31]. Jones *et al.* expanded these profiles to include human and mouse PXR through the use of a novel binding assay. Rifampicin and SR12813, a cholesterol-lowering drug, were found to effectively activate human and rabbit PXR but not rat and mouse forms. PCN was found to potently activate mouse as well as rat PXR but have little effect on the human and rabbit forms [32]. These data correlated well with the patterns of CYP3A gene expression in the liver and intestinal tissues of the various species, proving that the receptor has clearly diverged functionally through the process of evolution. Indeed, several studies have identified individual residues in the LBD that confer species-specific transcriptional activation to the PXR [27, 33, 34].

Cross-Talk with Other Nuclear Receptors

PXR overlaps functionally with CAR (constitutive androstane receptor) in terms of ligand binding and gene activation. It was noted in 1990 that the treatment of rat hepatocytes with phenobarbital caused distinct expression patterns for the cytochrome P450 isoforms CYP2B and CYP3A [29]. It was subsequently shown that both PXR and CAR could both be activated by phenobarbital, and that the effects of phenobarbital on CYP gene expression were mediated by several different nuclear receptors [6]. PXR became established as a central regulator of CYP3A genes [1, 3, 4], while CAR was found to respond to phenobarbital response elements located on CYP2B genes [35-37]. In 2000, Moore *et al.* showed that both receptors could be activated by some of the same xenobiotic and endobiotic compounds, including rifampicin and phenobarbital [38]. It was then demonstrated that CAR

and PXR cross-talk extended to DNA binding, with the finding that PXR could up-regulate CYP2B gene products using the same response element employed by CAR, and vice versa for CAR's up-regulation of CYP3A gene products [8]. Thus, PXR and CAR work in concert by binding to the same ligands and DNA response elements to control target gene expression. Recent studies have established that PXR, CAR and FOXO1, a forkhead transcription factor, function together to regulate the expression of a variety of target genes central to drug metabolism and gluconeogenesis [39].

Subcellular Localization

Studies to determine the subcellular localization of PXR have provided conflicting results. Two groups have reported that human PXR is consistently localized in the nucleus, regardless of the addition of ligand. In one such study, Kawana *et al.* used transient expression in HeLa cells to show that PXR was localized in the nucleus in the absence of ligand [40]. They also identified a nuclear localization signal (NLS) in the DNA binding domain of PXR. Removal of the DBD resulted in solely cytoplasmic localization, and mutation of the putative NLS resulted in PXR localization in both the cytoplasm and the nucleus. These results are consistent with immunostaining assays that found human PXR to be exclusively located in the nucleus both with and without ligand [41]. Other studies, however, have provided evidence to the contrary. PXR from mouse liver was found to be localized in the cytoplasm and translocated to the nucleus only upon addition of PCN or other agonists [40, 42]. The discrepancies in these findings may be linked to differences in the type of PXR employed (human vs. mouse) or in the *in vivo* vs. *in vitro* nature of the

experiments. Such results may also indicate that subcellular trafficking is an important regulatory process that tunes the function of this nuclear xenobiotic receptor.

Ligand Binding

The PXR from a variety of species are all promiscuous and can bind to a variety of chemically- and structurally-distinct xenobiotic and endobiotic compounds. As measured by scintillation binding assays [32] and coactivator receptor ligand assays (CARLA, a ligand-dependent coregulator recruitment assay) [43], PXR is activated by the direct binding of ligands within the receptor's ligand binding cavity [2]. PXR agonists include natural and synthetic steroids such as 5 β -pregnane-3,20-dione and estradiol [32], and xenobiotics like the cholesterol drugs lovastatin and SR12813 [6, 32], the anti-cancer drugs tamoxifen and taxol [7, 44], the antibiotic rifampicin [1, 6], and the active agent of St. John's wort, hyperforin [45, 46]. These ligands vary widely in shape and chemical features, and range in mass from 250 to greater than 800 Da. Clearly, PXR has a binding promiscuity unlike that of any other member of the nuclear receptor superfamily [2]. Crystallographic studies of PXR have revealed a novel insert in the ligand binding domain along with a large and conformable binding pocket [13, 27, 47]. These structures offer valuable insight into both the promiscuity and specificity of the receptor, as discussed below.

Heterodimerization with RXR

Like many other nuclear receptors, PXR controls transcriptional events as a heterodimer with RXR α . PXR is known to bind to at least four distinct DNA response

elements, including both direct and everted repeats. Several other receptors that form heterodimers with RXR α also utilize both DR and ER response elements, including CAR, the vitamin D receptor (VDR), and the thyroid hormone receptor (TR) [48, 49]. Thus, because the LBDs of RXR α and these partner receptors are expected to form only one type of heterodimer [50], the linkers connecting the DBDs and LBDs of these receptors must be flexible in order to bind to distinctly oriented DNA response elements. It is also possible that alternative DNA binding modes may influence coregulator recruitment and transcriptional activity in a ligand- and/or tissue-specific manner, providing another level of regulation of PXR action [51].

Coregulator Binding

PXR was initially found to interact with the steroid receptor coactivator 1 (SRC-1; also known as NCoA-1) [4], a member of the p160 family of coactivators that bind in a ligand-dependent manner to nuclear receptors using Leu-X-X-Leu-Leu repeats (where X is any amino acid) [52-56]. Crystal structures of several nuclear receptor LBDs in complex with coactivator fragments have revealed that binding of ligand induces a conformational change in the AF2 region at the C-terminus of the nuclear receptor LBD to create a coactivator binding cleft [47, 50, 57-59]. This interaction will be studied in further detail later in this review [60]. Other members of the p160 coactivator family known to interact with PXR include TIF-2/GRIP-1/NCoA-2 [61] and p/CIP/ACTR/AIB1/TRAM1/RAC3 [24]. Binding of a coactivator protein results in the recruitment of basal transcription factors, such as CBP (CREB-binding protein)/p300 [53, 62], as well as histone acetyltransferases (HATs) that remodel chromatin to enhance transcription [63, 64]. SRC-1 also recruits CARM-1, an

arginine methyltransferase that methylates histone H3 to loosen chromatin for transcription [65, 66]. Other coactivators known to interact functionally with PXR include the receptor interacting protein RIP140 [67, 68], Suppressor for Gal 1 SUG-1 [67], PPAR binding protein PBP [69], and the peroxisome proliferator activating receptor PGC-1 [70]. Several transcriptional corepressors that down-regulate gene expression have also been found to bind PXR. Among these are SMRT (Silencing Mediator of Retinoid Thyroid Receptor) [7], nuclear receptor corepressor NCoR [7, 69, 71], and, most recently, the small heterodimer partner SHP [72]. Unraveling the structural basis of the recruitment of coregulators to NR-DNA complexes remains a critical area for future study.

MOBILITY IN PXR STRUCTURE

Several crystal structures of the LBD of PXR have been determined in the unliganded (apo) state and in complexes activating ligands and fragments of transcriptional coregulators. Published structures include the PXR LBD bound to the cholesterol-lowering drug SR12813, both in the presence and absence of SRC-1 peptide, in complex with hyperforin, the psychoactive agent found in the herbal antidepressant St. John's Wort, and in complex with the macrolide antibiotic rifampicin [13, 27, 47, 73-75] (Table 1). The overall fold of PXR consists of a three-layered α -helical sandwich ($\alpha 1$ - $\alpha 3$ / $\alpha 4$ - $\alpha 5$ - $\alpha 8$ - $\alpha 9$ / $\alpha 7$ - $\alpha 10$) that encloses a large, conformable binding pocket (Figure 1). A five-stranded anti-parallel β -sheet ($\beta 2$, $\beta 3$, $\beta 4$, $\beta 1$ and $\beta 1'$) lies adjacent to the ligand binding pocket. This extended β -sheet is unique to PXR, as NR LBDs typically contain only two- or three-stranded β -sheets [57, 76-79]. The

PXR LBD ends with a short helix (α AF) critical to the AF-2 region of the receptor that contacts transcriptional coregulators [27].

The α -helical portion of PXR is similar in structure to the NR LBDs described previously, with root-mean-square deviations in C α positions of 1.8-2.9 Å [80-82]. Where PXR deviates most significantly in structure, and what likely contributes critically to its promiscuous ability to respond to chemically-distinct ligands, is at the bottom of the LBD as shown in Figure 1. The PXRs contain an insert of ~60 residues that is unique in the NR superfamily. This insert, amino acids 177-228 in human PXR, contains not only the β 1- β 1' regions that extend the PXR β -sheet to five strands, but also a novel α 2 that folds along the underside of the expansive PXR ligand binding pocket [73]. A portion of this sequence insert (residues 178-191) has been disordered in the PXR LBD structures examined to date [13, 27, 47, 73, 75]. Thus, the PXRs line their ligand binding pockets with novel secondary structural elements, including α 2 and β 1- β 1', many of which are structurally flexible.

The conformability of key regions of the PXR LBD is critical to the ability of the receptor to bind to compounds of varying size and shape. The recent structure of PXR in complex with the large macrolide antibiotic rifampicin has provided a direct observation of the importance of flexibility in receptor ligand binding. When the apo (unliganded) structure of the PXR LBD was first reported in 2001, it was noted that the receptor's binding pocket, while large, was not large enough to accommodate the established PXR agonist rifampicin [27]. The subsequent determination of the PXR-rifampicin complex structure reveals that three regions of the LBD become disordered to create the space necessary for this 823 Da agonist to bind. These regions are the flexible loop formed by residues 229-235, a mobile hydrophobic loop from residues 309-321, and the ~192-210 stretch that is a *bona fide* helix in

some structures, but a partially ordered pseudohelix in others (Figure 2). Each of these regions (indicated in yellow in Figure 2) is highly mobile and exhibits no clear electron density in this 2.8 Å resolution crystal structure of the PXR-LBD in complex with rifampicin. Similarly, the piperidino group on rifampicin expected to lie next to the ~192-210 loop also lacked clear electron density [73]. These observations show that PXR can bind effectively to ligands and up-regulate gene expression even when a significant portion of its ligand binding pocket is unstructured. These results also establish that the mobility of regions of PXR that are novel in the NR family is vital to the promiscuous ligand binding character of this xenobiotic receptor.

Residues 309-321 were traced as a loop in the apo and in the SR12813 complexes [27], but adopt an α -helical structure in complexes with hyperforin, rifampicin, and SR12813 with the SRC-1 coactivator peptide [13, 47, 73]. This helix, designated α_6 , is different from α_6 's found in other nuclear receptors, which are positioned at the bottom of the ligand binding cavity in the same region where the ~192-210 residues are located in PXR. Residues 229-235 support the position of the 192-210 region, and their flexibility mirrors that of the longer region nearby. The stretch from about 192 through 210 was pseudohelical in the initial apo and SR12813-bound structures, but folded into α_2 in the structures of PXR bound to hyperforin and the combination of SR12813 and the SRC-1 peptide. This region of the receptor directly contacts bound ligands and changes its position to conform to specific ligands; thus, this novel structural motif is central to the ligand binding promiscuity exhibited by the PXRs.

Even when they are not disordered, these regions of the PXR LBD are mobile and have been observed to change position to enhance contacts with distinctly shaped ligands.

Between the apo and rifampicin-bound structures, for example, the $\alpha 6$, 229-235 and 204-210 regions of the receptor exhibit main-chain shifts of 1.5 Å, 3.2 Å, and 4.5 Å, respectively, and side chain displacements of up to 7 Å (Figure 3). This conformability allows the ligand binding pocket of PXR to expand from 1280 Å³ in volume in the SR12813 complex to more than 1600 Å³ in others structures [13, 27, 47, 73].

In addition to mobility that allows the pocket to accommodate a variety of ligands, PXR's $\alpha 2$ may have another function. It is not clear how ligands enter and exit the ligand binding pocket of this promiscuous receptor. In the peroxisome proliferator-activating receptors (PPARs) [50, 57], the putative ligand entrance path occurs in a region blocked by $\alpha 6$ in PXR. The flexible and untethered $\alpha 2$ may function as a trap-door in PXR, dropping out of the way so that ligands can enter the binding pocket. In some PXR structures, a solvent accessible channel of up to 3 Å wide and 9 Å long is present in the area adjacent to $\alpha 2$ [13, 27, 47]. Thus, the sequence motif that contains $\alpha 2$ appears to play a dual role in receptor function: conforming to distinct ligand shapes to enhance promiscuity, and providing a dynamic entry and exit pathway for ligand binding and dissociation.

There are six amino acid side chains that are consistently involved in ligand binding in all PXR LBD structures determined to date: three polar residues (Ser-247, Gln-285 and His-407) and three hydrophobic residues (Met-243, Trp-299 and Phe-420) (Figure 3) [13, 27, 47, 73]. The directed promiscuity exhibited by the different species of PXR may partly be attributed to changes in these binding residues. For example, the residues Gln-285, His-407 and Met-243 are not conserved in mouse PXR. This receptor shows a lesser degree of promiscuity, and shows no or minimal activity with SR12813, hyperforin or rifampicin (Table 1). A similar trend is seen in zebra fish PXR. Alignments of mammalian PXRs reveal

that the highest degree of sequence identity occurs between the human and rhesus receptor (96%); notably, the same six binding residues are conserved, and both receptors respond largely to the same pool of compounds [83]. Differences in diet were originally thought to be the driving force for PXR's directed promiscuity. It has recently been hypothesized, however, that bile acids served as the key evolutionary ligands that drove the receptor's increasing degree of promiscuity over time [84].

NR LBDs typically contain AF-2 regions that bind to LxxLL motifs in transcriptional coactivators, and I/LxxI/VI motifs in corepressors [85]. The structure of the human PXR LBD has been determined in complex with the second LxxLL motif of the coactivator SRC-1 bound to the receptor's AF-2 region. The LxxLL motif forms an α -helix, with a second short helix kinked perpendicular to the first (**Figure 4**). The leucines in the LxxLL motif pack via hydrophobic contacts against the surface of PXR in a groove formed by α 3, α 4 and α AF. A "charge clamp" involving PXR residues Lys-259 and Glu-427 stabilizes the weak helix dipole at the C- and N-termini, respectively, of the LxxLL motif [47, 57]. Charged residues are conserved in these positions in NR LBDs, and are receptor-coactivator interactions.

The structure of the PXR LBD in complex with the SR12813 ligand alone revealed three distinct binding modes for this small agonist within the receptor's pocket. A subsequent structure with PXR in complex with both SR12813 and a fragment of SRC-1, however, revealed only a single, distinct orientation of the ligand. This observation suggests that the PXR LBD "breathes", allowing small ligands to sample multiple binding modes. In the presence of a bound coactivator fragment, however, this sampling motion is restricted, resulting in stabilization of the ligand into a single conformation [47].

Numerous single-site mutations have been introduced into the PXR LBD with varying effects on basal transcriptional activity. Some of these mutations lead to variant receptors that exhibit increased basal activation, including H407N, S247W, W299A, and R410A [73]. In the S247W mutation, the replacement of serine with a bulky tryptophan residue is expected to fill the pocket to mimic ligand binding. This may stabilize coactivator interactions and increase basal transcriptional activity. The structural basis for the effects of other mutations, however, is less clear. For example, H407N and W299A may impact receptor activity by impairing corepressor binding or by improving coactivator binding. Conversely, the mutation of charged residues (R410N, D205A, E321A, and R413A) may facilitate increased corepressor or decreased coactivator binding, causing the partial or complete loss of basal activation.

AREAS FOR FUTURE STUDY

In just a few years, PXR has moved from an orphan receptor to an adopted central xenobiotic sensor and a putative drug target. We now face new challenges to deepen our understanding of PXR's basic functions in human biology, as well as and how the receptor might be harnessed in a clinical setting. The role that distinct ligands play in PXR's regulation of tissue- and coregulator-specific transcription events is emerging as a key area of study for this xenobiotic receptor [86]. In addition, the potential impact of sites of phosphorylation on the action and stability of this and other nuclear receptors warrants detailed attention [69], as does the pursuit of structures of full-length PXR-RXR heterodimers on DNA. Finally, because PXR is upregulated in certain human cancers [87,

88], the search for selective PXR modulators (SPRMs) might provide novel therapeutic tools for the treatment of neoplastic and metabolic diseases.

REFERENCES

1. Blumberg, B., W. Sabbagh, Jr., H. Juguilon, J. Bolado, Jr., C.M. van Meter, E.S. Ong, and R.M. Evans, *SXR, a novel steroid and xenobiotic-sensing nuclear receptor*. Genes Dev, 1998. **12**(20): p. 3195-205.
2. Kliewer, S.A., B. Goodwin, and T.M. Willson, *The nuclear pregnane X receptor: a key regulator of xenobiotic metabolism*. Endocr Rev, 2002. **23**(5): p. 687-702.
3. Bertilsson, G., J. Heidrich, K. Svensson, M. Asman, L. Jendeberg, M. Sydow-Backman, R. Ohlsson, H. Postlind, P. Blomquist, and A. Berkenstam, *Identification of a human nuclear receptor defines a new signaling pathway for CYP3A induction*. Proc Natl Acad Sci U S A, 1998. **95**(21): p. 12208-13.
4. Kliewer, S.A., J.T. Moore, L. Wade, J.L. Staudinger, M.A. Watson, S.A. Jones, D.D. McKee, B.B. Oliver, T.M. Willson, R.H. Zetterstrom, T. Perlmann, and J.M. Lehmann, *An orphan nuclear receptor activated by pregnanes defines a novel steroid signaling pathway*. Cell, 1998. **92**(1): p. 73-82.
5. Maurel, P., *The CYP3A family*. In *Cytochromes P450: Metabolic and Toxicological Aspects*. 1996, Boca Raton, FL: CRC Press, Inc.
6. Lehmann, J.M., D.D. McKee, M.A. Watson, T.M. Willson, J.T. Moore, and S.A. Kliewer, *The human orphan nuclear receptor PXR is activated by compounds that regulate CYP3A4 gene expression and cause drug interactions*. J Clin Invest, 1998. **102**(5): p. 1016-23.
7. Synold, T.W., I. Dussault, and B.M. Forman, *The orphan nuclear receptor SXR coordinately regulates drug metabolism and efflux*. Nat Med, 2001. **7**(5): p. 584-90.
8. Xie, W., J.L. Barwick, C.M. Simon, A.M. Pierce, S. Safe, B. Blumberg, P.S. Guzelian, and R.M. Evans, *Reciprocal activation of xenobiotic response genes by nuclear receptors SXR/PXR and CAR*. Genes Dev, 2000. **14**(23): p. 3014-23.
9. Xie, W., J.L. Barwick, M. Downes, B. Blumberg, C.M. Simon, M.C. Nelson, B.A. Neuschwander-Tetri, E.M. Brunt, P.S. Guzelian, and R.M. Evans, *Humanized xenobiotic response in mice expressing nuclear receptor SXR*. Nature, 2000. **406**(6794): p. 435-9.
10. Goodwin, B., L.B. Moore, C.M. Stoltz, D.D. McKee, and S.A. Kliewer, *Regulation of the human CYP2B6 gene by the nuclear pregnane X receptor*. Mol Pharmacol, 2001. **60**(3): p. 427-31.
11. Gerbal-Chaloin, S., J.M. Pascussi, L. Pichard-Garcia, M. Daujat, F. Waechter, J.M. Fabre, N. Carrere, and P. Maurel, *Induction of CYP2C genes in human hepatocytes in primary culture*. Drug Metab Dispos, 2001. **29**(3): p. 242-51.

12. Gerbal-Chaloin, S., M. Daujat, J.M. Pascussi, L. Pichard-Garcia, M.J. Vilarem, and P. Maurel, *Transcriptional regulation of CYP2C9 gene. Role of glucocorticoid receptor and constitutive androstane receptor*. J Biol Chem, 2002. **277**(1): p. 209-17.
13. Watkins, R.E., J.M. Maglich, L.B. Moore, G.B. Wisely, S.M. Noble, P.R. Davis-Searles, M.H. Lambert, S.A. Kliewer, and M.R. Redinbo, *2.1 A crystal structure of human PXR in complex with the St. John's wort compound hyperforin*. Biochemistry, 2003. **42**(6): p. 1430-8.
14. Madhu, C. and C.D. Klaassen, *Protective effect of pregnenolone-16 alpha-carbonitrile on acetaminophen-induced hepatotoxicity in hamsters*. Toxicol Appl Pharmacol, 1991. **109**(2): p. 305-13.
15. Liu, L. and C.D. Klaassen, *Regulation of hepatic sulfotransferases by steroidal chemicals in rats*. Drug Metab Dispos, 1996. **24**(8): p. 854-8.
16. Dunn, R.T., 2nd, B.A. Gleason, D.P. Hartley, and C.D. Klaassen, *Postnatal ontogeny and hormonal regulation of sulfotransferase SULT1B1 in male and female rats*. J Pharmacol Exp Ther, 1999. **290**(1): p. 319-24.
17. Hosokawa, M., K. Hattori, and T. Satoh, *Differential responses of rat hepatic microsomal carboxylesterase isozymes to glucocorticoids and pregnenolone 16 alpha-carbonitrile*. Biochem Pharmacol, 1993. **45**(11): p. 2317-22.
18. Runge-Morris, M., W. Wu, and T.A. Kocarek, *Regulation of rat hepatic hydroxysteroid sulfotransferase (SULT2-40/41) gene expression by glucocorticoids: evidence for a dual mechanism of transcriptional control*. Mol Pharmacol, 1999. **56**(6): p. 1198-206.
19. Geick, A., M. Eichelbaum, and O. Burk, *Nuclear receptor response elements mediate induction of intestinal MDR1 by rifampin*. J Biol Chem, 2001. **276**(18): p. 14581-7.
20. Dussault, I., M. Lin, K. Hollister, E.H. Wang, T.W. Synold, and B.M. Forman, *Peptide mimetic HIV protease inhibitors are ligands for the orphan receptor SXR*. J Biol Chem, 2001. **276**(36): p. 33309-12.
21. Kast, H.R., B. Goodwin, P.T. Tarr, S.A. Jones, A.M. Anisfeld, C.M. Stoltz, P. Tontonoz, S. Kliewer, T.M. Willson, and P.A. Edwards, *Regulation of multidrug resistance-associated protein 2 (ABCC2) by the nuclear receptors pregnane X receptor, farnesoid X-activated receptor, and constitutive androstane receptor*. J Biol Chem, 2002. **277**(4): p. 2908-15.
22. Staudinger, J.L., B. Goodwin, S.A. Jones, D. Hawkins-Brown, K.I. MacKenzie, A. LaTour, Y. Liu, C.D. Klaassen, K.K. Brown, J. Reinhard, T.M. Willson, B.H. Koller,

- and S.A. Kliewer, *The nuclear receptor PXR is a lithocholic acid sensor that protects against liver toxicity*. Proc Natl Acad Sci U S A, 2001. **98**(6): p. 3369-74.
23. Staudinger, J., Y. Liu, A. Madan, S. Habeebu, and C.D. Klaassen, *Coordinate regulation of xenobiotic and bile acid homeostasis by pregnane X receptor*. Drug Metab Dispos, 2001. **29**(11): p. 1467-72.
 24. Xie, W., A. Radominska-Pandya, Y. Shi, C.M. Simon, M.C. Nelson, E.S. Ong, D.J. Waxman, and R.M. Evans, *An essential role for nuclear receptors SXR/PXR in detoxification of cholestatic bile acids*. Proc Natl Acad Sci U S A, 2001. **98**(6): p. 3375-80.
 25. Stedman, C.A., C. Liddle, S.A. Coulter, J. Sonoda, J.G. Alvarez, D.D. Moore, R.M. Evans, and M. Downes, *Nuclear receptors constitutive androstane receptor and pregnane X receptor ameliorate cholestatic liver injury*. Proc Natl Acad Sci U S A, 2005. **102**(6): p. 2063-8.
 26. Sonoda, J., L.W. Chong, M. Downes, G.D. Barish, S. Coulter, C. Liddle, C.H. Lee, and R.M. Evans, *Pregnane X receptor prevents hepatorenal toxicity from cholesterol metabolites*. Proc Natl Acad Sci U S A, 2005. **102**(6): p. 2198-203.
 27. Watkins, R.E., G.B. Wisely, L.B. Moore, J.L. Collins, M.H. Lambert, S.P. Williams, T.M. Willson, S.A. Kliewer, and M.R. Redinbo, *The human nuclear xenobiotic receptor PXR: structural determinants of directed promiscuity*. Science, 2001. **292**(5525): p. 2329-33.
 28. Kocarek, T.A., E.G. Schuetz, S.C. Strom, R.A. Fisher, and P.S. Guzelian, *Comparative analysis of cytochrome P4503A induction in primary cultures of rat, rabbit, and human hepatocytes*. Drug Metab Dispos, 1995. **23**(3): p. 415-21.
 29. Kocarek, T.A., E.G. Schuetz, and P.S. Guzelian, *Differentiated induction of cytochrome P450b/e and P450p mRNAs by dose of phenobarbital in primary cultures of adult rat hepatocytes*. Mol Pharmacol, 1990. **38**(4): p. 440-4.
 30. Barwick, J.L., L.C. Quattrochi, A.S. Mills, C. Potenza, R.H. Tukey, and P.S. Guzelian, *Trans-species gene transfer for analysis of glucocorticoid-inducible transcriptional activation of transiently expressed human CYP3A4 and rabbit CYP3A6 in primary cultures of adult rat and rabbit hepatocytes*. Mol Pharmacol, 1996. **50**(1): p. 10-6.
 31. Savas, U., M.R. Wester, K.J. Griffin, and E.F. Johnson, *Rabbit pregnane X receptor is activated by rifampicin*. Drug Metab Dispos, 2000. **28**(5): p. 529-37.
 32. Jones, S.A., L.B. Moore, J.L. Shenk, G.B. Wisely, G.A. Hamilton, D.D. McKee, N.C. Tomkinson, E.L. LeCluyse, M.H. Lambert, T.M. Willson, S.A. Kliewer, and J.T. Moore, *The pregnane X receptor: a promiscuous xenobiotic receptor that has diverged during evolution*. Mol Endocrinol, 2000. **14**(1): p. 27-39.

33. Ostberg, T., G. Bertilsson, L. Jendeberg, A. Berkenstam, and J. Uppenberg, *Identification of residues in the PXR ligand binding domain critical for species specific and constitutive activation*. Eur J Biochem, 2002. **269**(19): p. 4896-904.
34. Tirona, R.G., B.F. Leake, L.M. Podust, and R.B. Kim, *Identification of amino acids in rat pregnane X receptor that determine species-specific activation*. Mol Pharmacol, 2004. **65**(1): p. 36-44.
35. Trottier, E., A. Belzil, C. Stoltz, and A. Anderson, *Localization of a phenobarbital-responsive element (PBRE) in the 5'-flanking region of the rat CYP2B2 gene*. Gene, 1995. **158**(2): p. 263-8.
36. Park, Y., H. Li, and B. Kemper, *Phenobarbital induction mediated by a distal CYP2B2 sequence in rat liver transiently transfected in situ*. J Biol Chem, 1996. **271**(39): p. 23725-8.
37. Honkakoski, P. and M. Negishi, *Characterization of a phenobarbital-responsive enhancer module in mouse P450 Cyp2b10 gene*. J Biol Chem, 1997. **272**(23): p. 14943-9.
38. Moore, L.B., D.J. Parks, S.A. Jones, R.K. Bledsoe, T.G. Consler, J.B. Stimmel, B. Goodwin, C. Liddle, S.G. Blanchard, T.M. Willson, J.L. Collins, and S.A. Kliewer, *Orphan nuclear receptors constitutive androstane receptor and pregnane X receptor share xenobiotic and steroid ligands*. J Biol Chem, 2000. **275**(20): p. 15122-7.
39. Kodama, S., C. Koike, M. Negishi, and Y. Yamamoto, *Nuclear receptors CAR and PXR cross talk with FOXO1 to regulate genes that encode drug-metabolizing and gluconeogenic enzymes*. Mol Cell Biol, 2004. **24**(18): p. 7931-40.
40. Kawana, K., T. Ikuta, Y. Kobayashi, O. Gotoh, K. Takeda, and K. Kawajiri, *Molecular mechanism of nuclear translocation of an orphan nuclear receptor, SXR*. Mol Pharmacol, 2003. **63**(3): p. 524-31.
41. Koyano, S., K. Kurose, Y. Saito, S. Ozawa, R. Hasegawa, K. Komamura, K. Ueno, S. Kamakura, M. Kitakaze, T. Nakajima, K. Matsumoto, A. Akasawa, H. Saito, and J. Sawada, *Functional characterization of four naturally occurring variants of human pregnane X receptor (PXR): one variant causes dramatic loss of both DNA binding activity and the transactivation of the CYP3A4 promoter/enhancer region*. Drug Metab Dispos, 2004. **32**(1): p. 149-54.
42. Squires, E.J., T. Sueyoshi, and M. Negishi, *Cytoplasmic localization of pregnane X receptor and ligand-dependent nuclear translocation in mouse liver*. J Biol Chem, 2004. **279**(47): p. 49307-14.

43. Krey, G., O. Braissant, F. L'Horsset, E. Kalkhoven, M. Perroud, M.G. Parker, and W. Wahli, *Fatty acids, eicosanoids, and hypolipidemic agents identified as ligands of peroxisome proliferator-activated receptors by coactivator-dependent receptor ligand assay*. Mol Endocrinol, 1997. **11**(6): p. 779-91.
44. Desai, P.B., S.C. Nallani, R.S. Sane, L.B. Moore, B.J. Goodwin, D.J. Buckley, and A.R. Buckley, *Induction of cytochrome P450 3A4 in primary human hepatocytes and activation of the human pregnane X receptor by tamoxifen and 4-hydroxytamoxifen*. Drug Metab Dispos, 2002. **30**(5): p. 608-12.
45. Wentworth, J.M., M. Agostini, J. Love, J.W. Schwabe, and V.K. Chatterjee, *St John's wort, a herbal antidepressant, activates the steroid X receptor*. J Endocrinol, 2000. **166**(3): p. R11-6.
46. Moore, L.B., B. Goodwin, S.A. Jones, G.B. Wisely, C.J. Serabjit-Singh, T.M. Willson, J.L. Collins, and S.A. Kliewer, *St. John's wort induces hepatic drug metabolism through activation of the pregnane X receptor*. Proc Natl Acad Sci U S A, 2000. **97**(13): p. 7500-2.
47. Watkins, R.E., P.R. Davis-Searles, M.H. Lambert, and M.R. Redinbo, *Coactivator binding promotes the specific interaction between ligand and the pregnane X receptor*. J Mol Biol, 2003. **331**(4): p. 815-28.
48. Drocourt, L., J.C. Ourlin, J.M. Pascussi, P. Maurel, and M.J. Vilarem, *Expression of CYP3A4, CYP2B6, and CYP2C9 is regulated by the vitamin D receptor pathway in primary human hepatocytes*. J Biol Chem, 2002. **277**(28): p. 25125-32.
49. Mangelsdorf, D.J. and R.M. Evans, *The RXR heterodimers and orphan receptors*. Cell, 1995. **83**(6): p. 841-50.
50. Gampe, R.T., Jr., V.G. Montana, M.H. Lambert, A.B. Miller, R.K. Bledsoe, M.V. Milburn, S.A. Kliewer, T.M. Willson, and H.E. Xu, *Asymmetry in the PPARgamma/RXRalpha crystal structure reveals the molecular basis of heterodimerization among nuclear receptors*. Mol Cell, 2000. **5**(3): p. 545-55.
51. Auboeuf, D., A. Honig, S.M. Berget, and B.W. O'Malley, *Coordinate regulation of transcription and splicing by steroid receptor coregulators*. Science, 2002. **298**(5592): p. 416-9.
52. Onate, S.A., S.Y. Tsai, M.J. Tsai, and B.W. O'Malley, *Sequence and characterization of a coactivator for the steroid hormone receptor superfamily*. Science, 1995. **270**(5240): p. 1354-7.
53. Kamei, Y., L. Xu, T. Heinzel, J. Torchia, R. Kurokawa, B. Gloss, S.C. Lin, R.A. Heyman, D.W. Rose, C.K. Glass, and M.G. Rosenfeld, *A CBP integrator complex mediates transcriptional activation and AP-1 inhibition by nuclear receptors*. Cell, 1996. **85**(3): p. 403-14.

54. Takeshita, A., P.M. Yen, S. Misiti, G.R. Cardona, Y. Liu, and W.W. Chin, *Molecular cloning and properties of a full-length putative thyroid hormone receptor coactivator*. Endocrinology, 1996. **137**(8): p. 3594-7.
55. Heery, D.M., E. Kalkhoven, S. Hoare, and M.G. Parker, *A signature motif in transcriptional co-activators mediates binding to nuclear receptors*. Nature, 1997. **387**(6634): p. 733-6.
56. Torchia, J., D.W. Rose, J. Inostroza, Y. Kamei, S. Westin, C.K. Glass, and M.G. Rosenfeld, *The transcriptional co-activator p/CIP binds CBP and mediates nuclear-receptor function*. Nature, 1997. **387**(6634): p. 677-84.
57. Nolte, R.T., G.B. Wisely, S. Westin, J.E. Cobb, M.H. Lambert, R. Kurokawa, M.G. Rosenfeld, T.M. Willson, C.K. Glass, and M.V. Milburn, *Ligand binding and co-activator assembly of the peroxisome proliferator-activated receptor-gamma*. Nature, 1998. **395**(6698): p. 137-43.
58. Xu, L., C.K. Glass, and M.G. Rosenfeld, *Coactivator and corepressor complexes in nuclear receptor function*. Curr Opin Genet Dev, 1999. **9**(2): p. 140-7.
59. Greschik, H., R. Flaig, J.P. Renaud, and D. Moras, *Structural basis for the deactivation of the estrogen-related receptor gamma by diethylstilbestrol or 4-hydroxytamoxifen and determinants of selectivity*. J Biol Chem, 2004. **279**(32): p. 33639-46.
60. Onate, S.A., V. Boonyaratanakornkit, T.E. Spencer, S.Y. Tsai, M.J. Tsai, D.P. Edwards, and B.W. O'Malley, *The steroid receptor coactivator-1 contains multiple receptor interacting and activation domains that cooperatively enhance the activation function 1 (AF1) and AF2 domains of steroid receptors*. J Biol Chem, 1998. **273**(20): p. 12101-8.
61. Brobst, D.E., X. Ding, K.L. Creech, B. Goodwin, B. Kelley, and J.L. Staudinger, *Guggulsterone activates multiple nuclear receptors and induces CYP3A gene expression through the pregnane X receptor*. J Pharmacol Exp Ther, 2004. **310**(2): p. 528-35.
62. Yao, T.P., G. Ku, N. Zhou, R. Scully, and D.M. Livingston, *The nuclear hormone receptor coactivator SRC-1 is a specific target of p300*. Proc Natl Acad Sci U S A, 1996. **93**(20): p. 10626-31.
63. Spencer, T.E., G. Jenster, M.M. Burcin, C.D. Allis, J. Zhou, C.A. Mizzen, N.J. McKenna, S.A. Onate, S.Y. Tsai, M.J. Tsai, and B.W. O'Malley, *Steroid receptor coactivator-1 is a histone acetyltransferase*. Nature, 1997. **389**(6647): p. 194-8.
64. Pazin, M.J. and J.T. Kadonaga, *What's up and down with histone deacetylation and transcription?* Cell, 1997. **89**(3): p. 325-8.

65. Chen, D., H. Ma, H. Hong, S.S. Koh, S.M. Huang, B.T. Schurter, D.W. Aswad, and M.R. Stallcup, *Regulation of transcription by a protein methyltransferase*. Science, 1999. **284**(5423): p. 2174-7.
66. Schurter, B.T., S.S. Koh, D. Chen, G.J. Bunick, J.M. Harp, B.L. Hanson, A. Henschen-Edman, D.R. Mackay, M.R. Stallcup, and D.W. Aswad, *Methylation of histone H3 by coactivator-associated arginine methyltransferase 1*. Biochemistry, 2001. **40**(19): p. 5747-56.
67. Masuyama, H., Y. Hiramatsu, Y. Mizutani, H. Inoshita, and T. Kudo, *The expression of pregnane X receptor and its target gene, cytochrome P450 3A1, in perinatal mouse*. Mol Cell Endocrinol, 2001. **172**(1-2): p. 47-56.
68. Masuyama, H., Y. Hiramatsu, M. Kunitomi, T. Kudo, and P.N. MacDonald, *Endocrine disrupting chemicals, phthalic acid and nonylphenol, activate Pregnane X receptor-mediated transcription*. Mol Endocrinol, 2000. **14**(3): p. 421-8.
69. Ding, X. and J.L. Staudinger, *Induction of drug metabolism by forskolin: the role of the pregnane X receptor and the protein kinase a signal transduction pathway*. J Pharmacol Exp Ther, 2005. **312**(2): p. 849-56.
70. Bhalla, S., C. Ozalp, S. Fang, L. Xiang, and J.K. Kemper, *Ligand-activated pregnane X receptor interferes with HNF-4 signaling by targeting a common coactivator PGC-1alpha. Functional implications in hepatic cholesterol and glucose metabolism*. J Biol Chem, 2004. **279**(43): p. 45139-47.
71. Gonzalez, M.M. and C. Carlberg, *Cross-repression, a functional consequence of the physical interaction of non-liganded nuclear receptors and POU domain transcription factors*. J Biol Chem, 2002. **277**(21): p. 18501-9.
72. Ourlin, J.C., F. Lasserre, T. Pineau, J.M. Fabre, A. Sa-Cunha, P. Maurel, M.J. Vilarem, and J.M. Pascussi, *The small heterodimer partner interacts with the pregnane X receptor and represses its transcriptional activity*. Mol Endocrinol, 2003. **17**(9): p. 1693-703.
73. Chrencik, J.E., J. Orans, L.B. Moore, Y. Xue, L. Peng, J.L. Collins, G.B. Wisely, M.H. Lambert, S.A. Kliewer, and M.R. Redinbo, *Structural Disorder in the Complex of Human PXR and the Macrolide Antibiotic Rifampicin*. Mol Endocrinol, 2005.
74. Goodwin, B., M.R. Redinbo, and S.A. Kliewer, *Regulation of cyp3a gene transcription by the pregnane x receptor*. Annu Rev Pharmacol Toxicol, 2002. **42**: p. 1-23.
75. Watkins, R.E., S.M. Noble, and M.R. Redinbo, *Structural insights into the promiscuity and function of the human pregnane X receptor*. Curr Opin Drug Discov Devel, 2002. **5**(1): p. 150-8.

76. Lo, J.C., V.M. Schwitzgebel, J.B. Tyrrell, P.A. Fitzgerald, S.L. Kaplan, F.A. Conte, and M.M. Grumbach, *Normal female infants born of mothers with classic congenital adrenal hyperplasia due to 21-hydroxylase deficiency*. J Clin Endocrinol Metab, 1999. **84**(3): p. 930-6.
77. Bourguet, W., M. Ruff, P. Chambon, H. Gronemeyer, and D. Moras, *Crystal structure of the ligand-binding domain of the human nuclear receptor RXR-alpha*. Nature, 1995. **375**(6530): p. 377-82.
78. Renaud, J.P., N. Rochel, M. Ruff, V. Vivat, P. Chambon, H. Gronemeyer, and D. Moras, *Crystal structure of the RAR-gamma ligand-binding domain bound to all-trans retinoic acid*. Nature, 1995. **378**(6558): p. 681-9.
79. Brzozowski, A.M., A.C. Pike, Z. Dauter, R.E. Hubbard, T. Bonn, O. Engstrom, L. Ohman, G.L. Greene, J.A. Gustafsson, and M. Carlquist, *Molecular basis of agonism and antagonism in the oestrogen receptor*. Nature, 1997. **389**(6652): p. 753-8.
80. Holm, L., C. Ouzounis, C. Sander, G. Tuparev, and G. Vriend, *A database of protein structure families with common folding motifs*. Protein Sci, 1992. **1**(12): p. 1691-8.
81. Holm, L. and C. Sander, *Dali/FSSP classification of three-dimensional protein folds*. Nucleic Acids Res, 1997. **25**(1): p. 231-4.
82. Holm, L. and C. Sander, *The FSSP database of structurally aligned protein fold families*. Nucleic Acids Res, 1994. **22**(17): p. 3600-9.
83. Moore, L.B., J.M. Maglich, D.D. McKee, B. Wisely, T.M. Willson, S.A. Kliewer, M.H. Lambert, and J.T. Moore, *Pregnane X receptor (PXR), constitutive androstane receptor (CAR), and benzoate X receptor (BXR) define three pharmacologically distinct classes of nuclear receptors*. Mol Endocrinol, 2002. **16**(5): p. 977-86.
84. Krasowski, M.D., K. Yasuda, L.R. Hagey, and E.G. Schuetz, *Evolution of the pregnane X receptor: adaptation to cross-species differences in biliary bile salts*. Mol Endocrinol, 2005.
85. Hu, X. and M.A. Lazar, *The CoRNR motif controls the recruitment of corepressors by nuclear hormone receptors*. Nature, 1999. **402**(6757): p. 93-6.
86. Masuyama, H., N. Suwaki, Y. Tateishi, H. Nakatsukasa, T. Segawa, and Y. Hiramatsu, *The Pregnane X Receptor Regulates Gene Expression in a Ligand- and Promoter-selective Fashion*. Mol Endocrinol, 2005.
87. Dotzlaw, H., E. Leygue, P. Watson, and L.C. Murphy, *The human orphan receptor PXR messenger RNA is expressed in both normal and neoplastic breast tissue*. Clin Cancer Res, 1999. **5**(8): p. 2103-7.

88. Masuyama, H., Y. Hiramatsu, J. Kodama, and T. Kudo, *Expression and potential roles of pregnane X receptor in endometrial cancer*. J Clin Endocrinol Metab, 2003. **88**(9): p. 4446-54.

FIGURE LEGENDS

Figure 5.1 Structure of the PXR LBD (blue) in complex with the small agonist SR12813 (red) and a fragment of the SRC-1 transcriptional coactivator (cyan) [47]. The 60-residue sequence insert novel to the PXR and central to the receptor's promiscuity is highlighted in magenta.

Figure 5.2 Close-up of the PXR-LBD bound to the large macrolide antibiotic rifampicin (green and red) [73]. Regions of the structure disordered in this complex are highlighted in yellow. The same regions are observed to be mobile in other PXR-LBD structures, and to conform to the presence of distinct ligands and bound coactivator fragments.

Figure 5.3 Schematic representation of the ligand binding pocket of human PXR, with residues that remain static in structure in bold and those that exhibit <1 Å shifts in position in italics. The side chains of residues that move >1 Å in position are shown, along with the magnitude of maximal shifts observed upon ligand binding. Note that four of the seven residues that exhibit a high degree of ligand-induced structural flexibility are part of the sequence insert novel to the PXR.

Figure 5.4 An LxxLL motif of the human coactivator SRC-1 (residues 682-296; magenta) bound to the AF-2 region of PXR (white) via a combination of non-polar and electrostatic contacts.

Table 5.1 Comparison of human and mouse EC_{50} values and the key residues contacted in crystal structures of PXR complexed with three structurally and chemically disparate ligands: SR12813, hyperforin, and rifampicin.

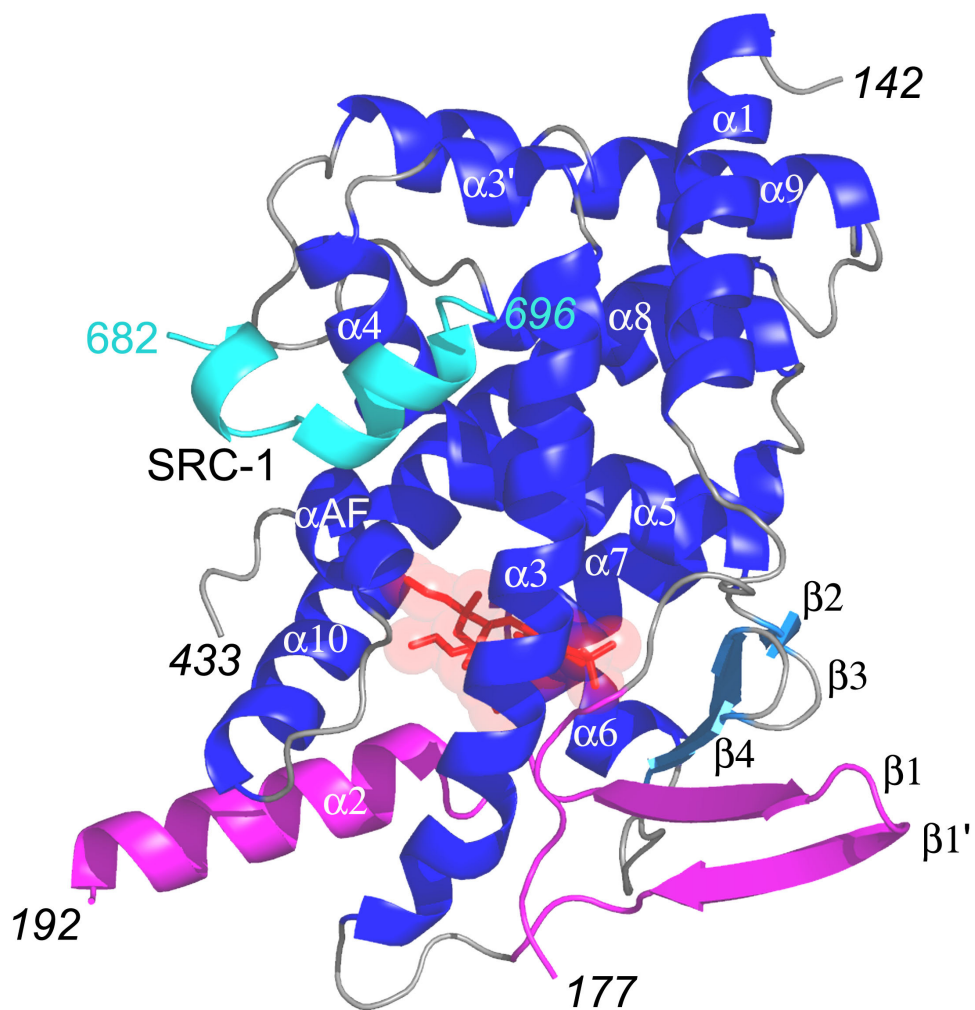


Figure 5.1 Structure of the PXR LBD (blue) in complex with the small agonist SR12813 (red) and a fragment of the SRC-1 transcriptional coactivator (cyan)

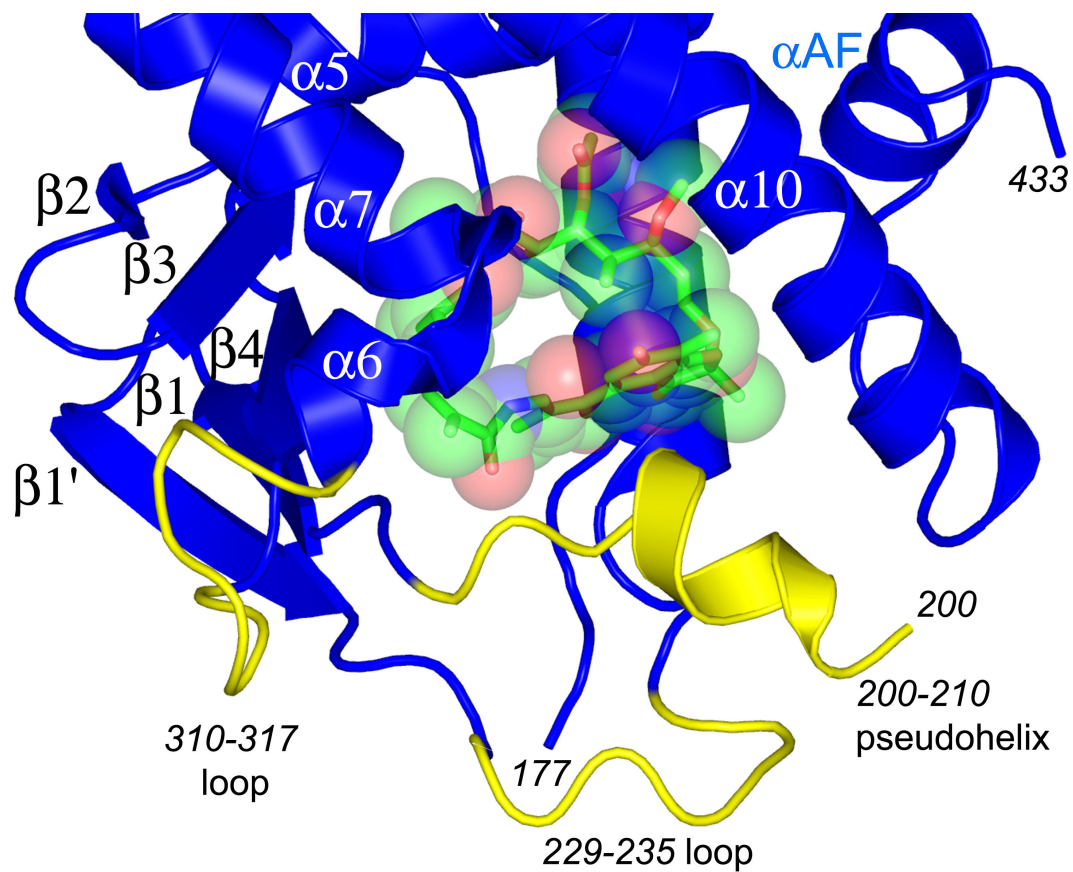


Figure 5.2 Close-up of the PXR-LBD bound to the large macrolide antibiotic rifampicin (green and red)

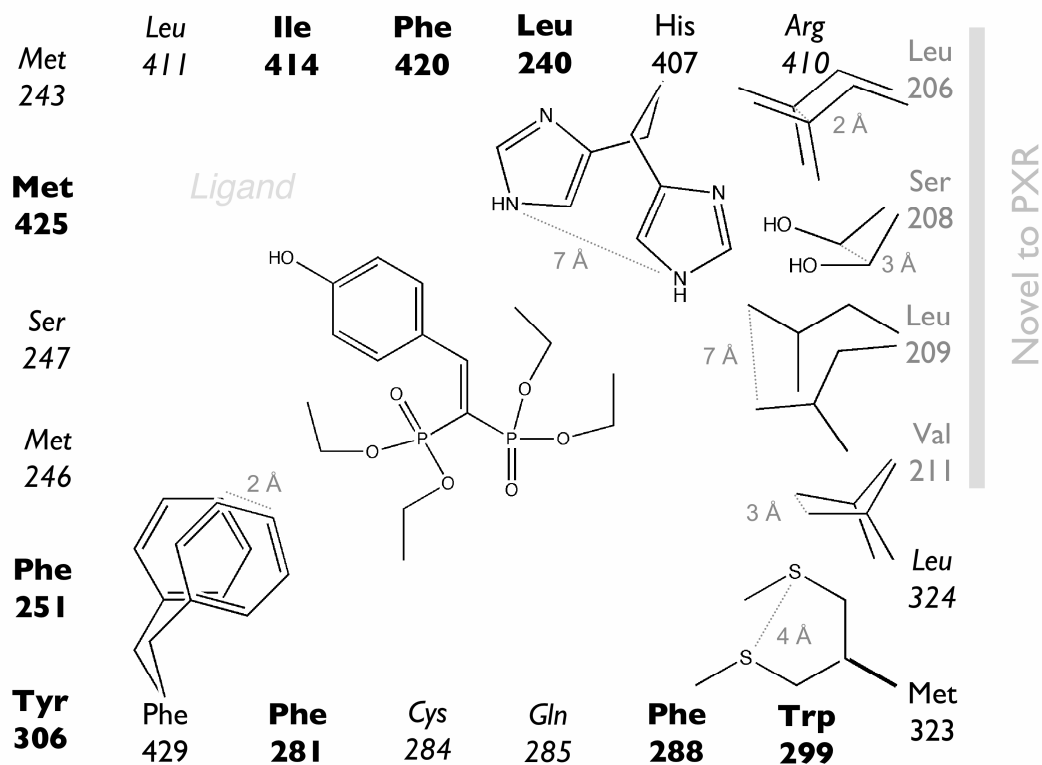


Figure 5.3 Schematic representation of the ligand binding pocket of human PXR

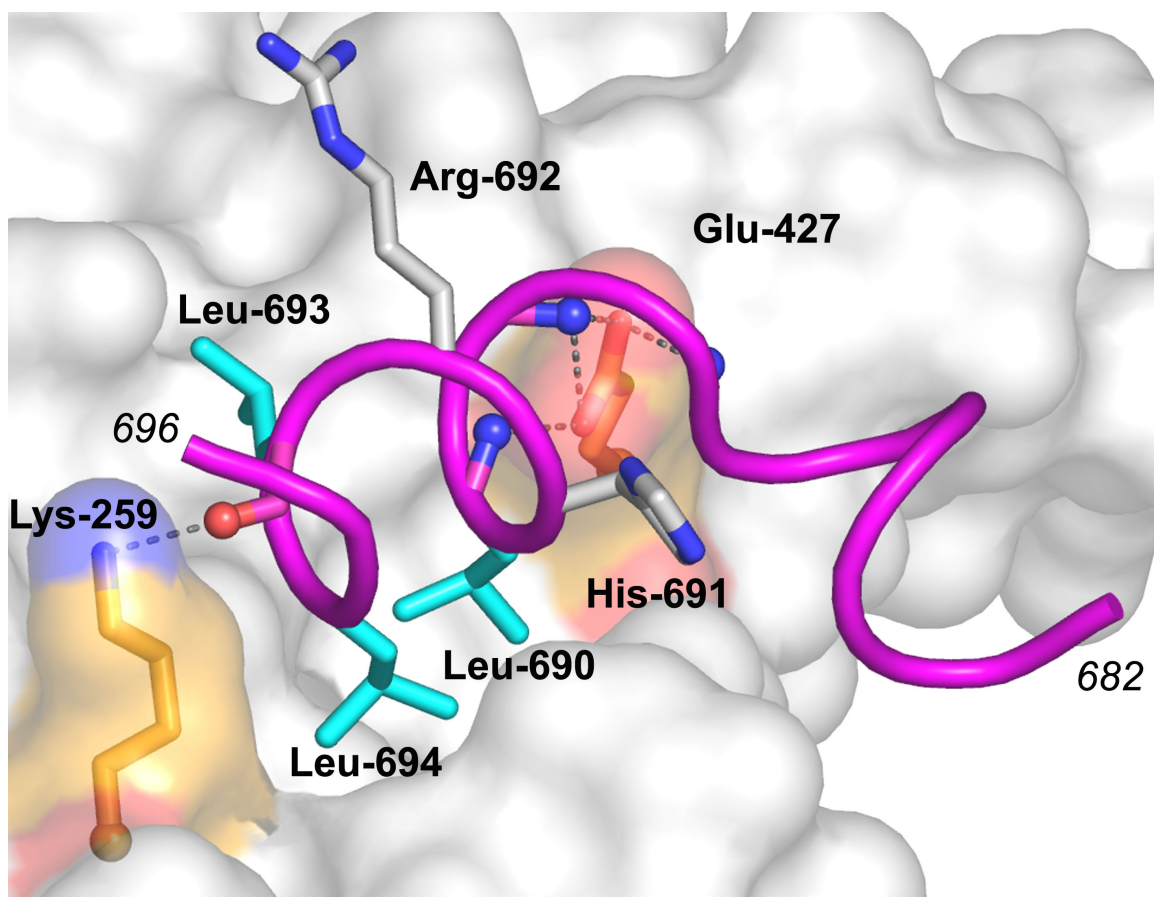
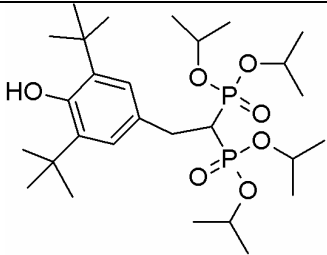
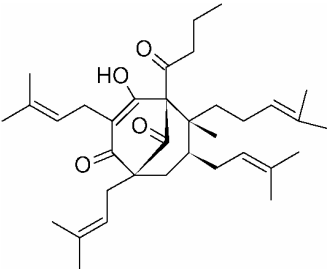
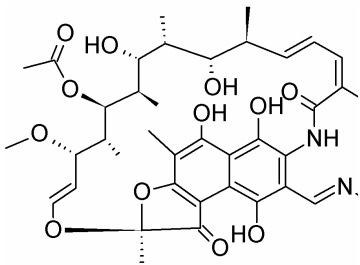


Figure 5.4 An LxxLL motif of the human coactivator SRC-1 bound to PXR

Table 5.1 Comparison of human and mouse EC₅₀ values for PXR and ligands

Ligand	Structure	MW	EC ₅₀		Key Structural Contacts
			Human (nM) ¹	Mouse (μM)	
<i>SR12813</i>		507	32.2 ± 7.7	3.4 ± 1.1 ²	Met-243, Ser247, Gln285, Trp-299 His407, Phe-420
<i>Hyperforin</i>		537	127 ± 23.4	N.D. ³	Met-243, Ser247, Gln285, Trp-299 His407, Phe-420
<i>Rifampicin</i>		823	463 ± 63.6 ¹	N.A. ²	Ser208, Met-243, Ser247, Gln285, Trp-299, His407, Phe-420, Arg410

¹ Watkins RE, Maglich JM, Moore LB, Wisely GB, Noble SM, Davis-Searles PR, Lambert MH, Kliewer SA, Redinbo MR 2003 2.1 A crystal structure of human PXR in complex with the St. John's wort compound hyperforin. *Biochemistry* 42:1430-8

² Moore LB, and Redinbo MR, unpublished; N.A. = No Activation.

³ N.D. = Not Determined

Chapter 6:
**Computational redesign of a nuclear receptor/coactivator interface to create altered
binding specificity**

INTRODUCTION

The human pregnane X nuclear receptor (PXR) regulates the transcription of a number of genes involved in xenobiotic detoxification in the liver and small intestine [1-6]. PXR is transcriptionally activated by several coactivators, including ACTR, GRIP, PBP/DRIP205, RIP140, and SRC-1 [7, 8]. Coactivators are a limiting factor in transcription regulation due to competition for binding by a variety of nuclear receptors, resulting in a “squenching” effect on transcription levels. Squenching leads to transcriptional interference between receptors and transcription factors. The goal of this project was to design mutations in both PXR and SRC-1 using a computational protein design program (Rosetta [9]) such that they will exclusively interact with one another. This would effectively eliminate the squenching phenomenon and allow for study of a more selective receptor/coactivator interaction. The hypothesis of this proposal is that Rosetta can be used to create mutations at the PXR/SRC-1 interface that will alter the binding specificity. This would generate two mutant proteins that could bind to each other with a greater affinity than to their wild-type counterparts.

Background and Significance

Nuclear receptors are ligand-regulated transcription factors that play key roles in many vital physiological processes such as development, homeostasis, cell growth, differentiation, and apoptosis. Nuclear receptors activate the transcription of target genes in a ligand-dependent manner by binding to DNA response elements, usually as homodimers or heterodimers with the retinoid X receptor (RXR). Ligands are diverse in nature and can be agonists or antagonists depending upon how binding affects transcription. They are generally small and hydrophobic, and include steroid, thyroid and retinoid hormones, vitamins, xenobiotics, and metabolic intermediates [10].

Nuclear receptors require additional factors, termed coregulators, to regulate transcription. Coactivators bind to nuclear receptors in a ligand-dependent manner and promote transcription by modifying local chromatin structure and recruiting basal transcription factors. For example, coactivators belonging to the p160/SRC (steroid receptor coactivator) family, including SRC-1, possess histone acetyltransferase (HAT) activity [11]. Hyperacetylated regions of DNA are more actively transcribed than their unacetylated counterparts [12]. The p160 family of coactivators possesses three highly conserved LXXLL motifs, termed nuclear receptor (NR) boxes, which are necessary for ligand-dependent interactions with nuclear receptors [13-18]. These motifs are found repeatedly in many coactivators; the LXXLL-containing helix interacts with hydrophobic residues in the AF-2 helix when it is in the active conformation [19].

There are a limited number of coactivators in the cell; because each coactivator can bind to more than one type of nuclear receptor, competition occurs [20]. This leads to transcriptional interference between receptors and transcription factors (squenching). Initial

evidence for squelching in cotransfection assays was observed in the late 1980's. Data suggested that the overall transcriptional activity of receptors was limited by the available cellular pool of coactivators [21, 22]. Following these findings, a yeast two-hybrid screen was used to show that SRC-1 interacted with and coactivated almost all of the steroid hormone nuclear receptors including the estrogen receptor (ER), progesterone receptor (PR), glucocorticoid receptor (GR), thyroid hormone receptor (TR) and RXR [23]. In addition, an excess of SRC-1 was found to reverse squelching of transcriptional activity observed between PR and ER. Coexpression of the two receptors reduced the activity of PR by almost 95%, but subsequent addition of excess SRC-1 reversed this repression 16 fold [23]. It was concluded that SRC-1 is a limiting factor for transactivation of steroid hormone nuclear receptors.

It would be beneficial to study the specific interaction between one receptor and coactivator by introducing mutations that promote exclusive binding between the two species. Mutation of genes encoding mutant nuclear receptors and coactivators could be used to regulate expression of enzymes involved in metabolism of steroids, carbohydrates, lipids, and xenobiotics. This could have interesting therapeutic applications *in vivo*. The goals of this project were to use the computational protein design program Rosetta to redesign the PXR/SRC-1 interface to generate an altered binding specificity, to test this redesign through the use of site-directed mutagenesis and binding characterization, and to determine the crystal structures of complexes of forms of mutant PXR and mutant SRC-1 peptides.

COMPUTATIONAL PROTEIN REDESIGN OF PXR/SRC-1 INTERFACE USING ROSETTA

The possibility of computational protein design was first considered in the early 1980's [24, 25]. Research confronted the daunting task of predicting a three dimensional protein structure based solely on amino acid sequence. Eventually it was concluded that it would be more feasible to tackle the opposite problem—designing a de novo sequence that would fold into a desired 3-D structure [26]. Protein design was concentrated on assigning energy values to the various types of naturally occurring residue interactions. The first fully automated protein was not designed until 1997. Nearly 2.0×10^{27} possible amino acid sequences from a combinatorial library were screened based on physical chemical potential functions and spatial restraints to find a sequence that would fold into the desired target motif, a zinc-finger fold [27].

Rosetta is a computational program that redesigns given protein sequences using a Monte Carlo optimization procedure. By altering both the position and character of amino acid residues at random positions using a Protein Data Bank-compiled rotamer database, a minimum energy structure can be obtained. Rosetta uses an energy function with several different parameters. Key features of the function are a Lennard-Jones potential relating to interatomic distance, a hydrogen bonding potential, an implicit solvation model which penalizes the burial of hydrophilic residues, and an amino acid rotamer database. These features are often found in protein design functions.

Methods

Rosetta was used to assign an energy value to the wild-type PXR/SRC-1 complex as well as the individual proteins. The interface was carefully examined to determine the most important interactions between amino acid residues. We chose to alter a highly conserved leucine in one of the NR box motifs in SRC-1 to an alanine, and several interacting PXR residues were selected for alteration by the program. Altered specificity can be determined as follows: Let wtPXR=A, mutant PXR=A', wtSRC-1=B, and mutant SRC-1=B';

$$E_{A'B'} - E_{A'} - E_{B'} = \Delta E_{A'B'} \quad (\text{eq. 1})$$

$$E_{A'B} - E_{A'} - E_B = \Delta E_{A'B} \quad (\text{eq. 2})$$

$$E_{AB'} - E_A - E_{B'} = \Delta E_{AB'} \quad (\text{eq. 3})$$

When $\Delta E_{A'B'}$ is greater than both $\Delta E_{A'B}$ and $\Delta E_{AB'}$ (i.e. the energy of the two-mutant species is more favorable than the energy of any wild-type-mutant combination), altered specificity has been achieved.

Results

Two residues on PXR were altered (I255L, L276A); a highly conserved leucine from an LXXLL motif on SRC-1 (L694A) was also altered. The energy values of the various complexes are shown in Table 1 and indicate that altered specificity was achieved. Figure 6.2 shows the superposition of the wild-type and mutant protein complexes. Clearly the integrity of packing is maintained in the mutant; in fact, the backbone of the proteins remains rigid throughout the simulation. Only the side chains are permitted to move or change.

In a study of nine globular proteins completely redesigned by Rosetta, denaturation experiments showed that eight of these proteins were folded, and six had stabilities equal to

or greater than their wild-type counterparts [28]. Additionally, Rosetta was recently used to design a globular protein with a novel fold, the structure of which was confirmed using x-ray crystallography [29]. The root mean square deviation between the actual and predicted structures was 1.2 Å. These results provide strong evidence that computational design methods do have the ability to identify stable amino acid sequences.

DETERMINATION OF BINDING CONSTANTS FOR REDESIGNED PXR/SRC-1 COMPLEX

The mutant PXR/mutant SRC-1 complex designed above was expected to have a smaller dissociation constant (K_D) than the wild-type complex, which is indicative of a greater affinity as predicted by Rosetta. Surface Plasmon Resonance (SPR) was used to obtain these binding constants.

Methods

Residues 130-434 of the human PXR LBD with an N-terminal polyhistidine tag were subcloned into the pRSETA expression vector (Invitrogen, Carlsbad, CA), and residues 623-710 of hSRC-1 protein were subcloned into the pACYC184 bacterial vector (American Type Culture Collection 37033). These plasmids were cotransformed into the BL21(DE3) strain of *E. coli* and expressed at 22°C without IPTG induction for 24 hours. Cotransformation of PXR and SRC-1 was necessary for the stable expression of the protein. The PXR-SRC-1 complex was purified using ProBond nickel-chelating resin (affinity) (Invitrogen, Carlsbad,

CA) and subsequently by SP Sepharose Fast Flow resin (cation exchange) (Pharmacia, Uppsala, Sweden) [30]. Protein was concentrated to 0.67 mg/mL at 4°C using an Amicon YM-10 unit (Millipore, Billerica, MA). Concentration took place in the presence of a 10-fold molar excess of the cholesterol-lowering drug SR12813, a known activator of PXR, in order to increase the stability of the protein and promote coactivator binding.

Mutant PXR plasmid DNA was obtained using the Quik-Change Site-Directed Mutagenesis Kit (Stratagene, La Jolla, CA) as instructed by the manual. Mutations were confirmed by sequencing, and expression and purification were carried out as previously described to a final concentration of 0.5 mg/mL.

Biotinylated SRC-1 peptide (residues 682-696 containing a LXXLL motif) (SynPep, Dublin, California) was immobilized on a streptavidin (SA) sensor chip (Biacore, Uppsala, Sweden). PXR in SPR buffer (0.01 M HEPES, pH 7.4) at varying concentrations (3-10 μ M) was passed over the peptide in order to obtain an association curve. Excess buffer was then used to reverse protein binding and obtain a dissociation curve. This was done for wild-type-wild-type and one wild-type-mutant combination of SRC-1 and PXR, respectively.

Since PXR must be coexpressed with SRC-1, the protein we used in the experiment was actually a complex of PXR and SRC-1. This presented a problem in that we were trying to determine a binding affinity for the complex. It has been shown previously that concentration of hPXR/SRC-1 with excess peptide displaces the larger SRC-1 fragment. This was confirmed by subsequent crystallization and structure determination [19]. Immobilization of the excess SRC-1 peptide on the sensor chip should displace the larger, expressed SRC-1 complex and allow for accurate determination of a binding constant. It was assumed that the association curve generated in the experiment would not be affected by this

displacement. To prove this was the case, fractions of protein containing only PXR without SRC-1 were identified as they eluted from the ion-exchange column by using SDS-PAGE. These fractions were then pooled and concentrated for use in SPR experiments, and gave similar results to the ones described.

Results

The binding curves for wild-type PXR and wild-type SRC-1 peptide are shown in Figure 6.3. The dissociation constant K_D was calculated using BiaEvaluation 3.0 (Biacore, Uppsala, Sweden) by fitting the curves to a 1:1 Langmuir binding model using the equation $K_D = k_d/k_a$, the acquired association and dissociation rate constants. The dissociation constant between PXR and SRC-1 peptide was determined to be 0.9 μM . This value is well within the range of other literature values of NR/coactivator binding constants [31].

Unfortunately, the binding curves for the mutant PXR and wild-type SRC-1 were not intelligible enough to produce a usable binding constant. The protein itself did not express as well as the wild-type and also behaved in a more unstable manner. The problem arose when the protein could not be flushed off of the bound SRC-1 in order to acquire a dissociation curve. This led us to believe that the protein was structurally unstable, causing non-specific binding to both the attached SRC-1 peptide and the sensor chip, and therefore we were not able to determine a binding constant.

REFERENCES

1. Bertilsson, G., J. Heidrich, K. Svensson, M. Asman, L. Jendeberg, M. Sydow-Backman, R. Ohlsson, H. Postlind, P. Blomquist, and A. Berkenstam, *Identification of a human nuclear receptor defines a new signaling pathway for CYP3A induction*. Proc Natl Acad Sci U S A, 1998. **95**(21): p. 12208-13.
2. Blumberg, B., W. Sabbagh, Jr., H. Juguilon, J. Bolado, Jr., C.M. van Meter, E.S. Ong, and R.M. Evans, *SXR, a novel steroid and xenobiotic-sensing nuclear receptor*. Genes Dev, 1998. **12**(20): p. 3195-205.
3. Kliewer, S.A., J.T. Moore, L. Wade, J.L. Staudinger, M.A. Watson, S.A. Jones, D.D. McKee, B.B. Oliver, T.M. Willson, R.H. Zetterstrom, T. Perlmann, and J.M. Lehmann, *An orphan nuclear receptor activated by pregnanes defines a novel steroid signaling pathway*. Cell, 1998. **92**(1): p. 73-82.
4. Lehmann, J.M., D.D. McKee, M.A. Watson, T.M. Willson, J.T. Moore, and S.A. Kliewer, *The human orphan nuclear receptor PXR is activated by compounds that regulate CYP3A4 gene expression and cause drug interactions*. J Clin Invest, 1998. **102**(5): p. 1016-23.
5. Maurel, P., *The CYP3A family*. In *Cytochromes P450: Metabolic and Toxicological Aspects*. 1996, Boca Raton, FL: CRC Press, Inc.
6. Xie, W., J.L. Barwick, M. Downes, B. Blumberg, C.M. Simon, M.C. Nelson, B.A. Neuschwander-Tetri, E.M. Brunt, P.S. Guzelian, and R.M. Evans, *Humanized xenobiotic response in mice expressing nuclear receptor SXR*. Nature, 2000. **406**(6794): p. 435-9.
7. Masuyama, H., Y. Hiramatsu, M. Kunitomi, T. Kudo, and P.N. MacDonald, *Endocrine disrupting chemicals, phthalic acid and nonylphenol, activate Pregnane X receptor-mediated transcription*. Mol Endocrinol, 2000. **14**(3): p. 421-8.
8. Synold, T.W., I. Dussault, and B.M. Forman, *The orphan nuclear receptor SXR coordinately regulates drug metabolism and efflux*. Nat Med, 2001. **7**(5): p. 584-90.
9. Kuhlman, B. and D. Baker, *Native protein sequences are close to optimal for their structures*. Proc Natl Acad Sci U S A, 2000. **97**(19): p. 10383-8.
10. Liao, L., S.Q. Kuang, Y. Yuan, S.M. Gonzalez, B.W. O'Malley, and J. Xu, *Molecular structure and biological function of the cancer-amplified nuclear receptor coactivator SRC-3/AIB1*. J Steroid Biochem Mol Biol, 2002. **83**(1-5): p. 3-14.

11. Spencer, T.E., G. Jenster, M.M. Burcin, C.D. Allis, J. Zhou, C.A. Mizzen, N.J. McKenna, S.A. Onate, S.Y. Tsai, M.J. Tsai, and B.W. O'Malley, *Steroid receptor coactivator-1 is a histone acetyltransferase*. *Nature*, 1997. **389**(6647): p. 194-8.
12. Pazin, M.J. and J.T. Kadonaga, *What's up and down with histone deacetylation and transcription?* *Cell*, 1997. **89**(3): p. 325-8.
13. Heery, D.M., E. Kalkhoven, S. Hoare, and M.G. Parker, *A signature motif in transcriptional co-activators mediates binding to nuclear receptors*. *Nature*, 1997. **387**(6634): p. 733-6.
14. Nolte, R.T., G.B. Wisely, S. Westin, J.E. Cobb, M.H. Lambert, R. Kurokawa, M.G. Rosenfeld, T.M. Willson, C.K. Glass, and M.V. Milburn, *Ligand binding and co-activator assembly of the peroxisome proliferator-activated receptor-gamma*. *Nature*, 1998. **395**(6698): p. 137-43.
15. Torchia, J., D.W. Rose, J. Inostroza, Y. Kamei, S. Westin, C.K. Glass, and M.G. Rosenfeld, *The transcriptional co-activator p/CIP binds CBP and mediates nuclear-receptor function*. *Nature*, 1997. **387**(6634): p. 677-84.
16. Darimont, B.D., R.L. Wagner, J.W. Apriletti, M.R. Stallcup, P.J. Kushner, J.D. Baxter, R.J. Fletterick, and K.R. Yamamoto, *Structure and specificity of nuclear receptor-coactivator interactions*. *Genes Dev*, 1998. **12**(21): p. 3343-56.
17. Feng, W., R.C. Ribeiro, R.L. Wagner, H. Nguyen, J.W. Apriletti, R.J. Fletterick, J.D. Baxter, P.J. Kushner, and B.L. West, *Hormone-dependent coactivator binding to a hydrophobic cleft on nuclear receptors*. *Science*, 1998. **280**(5370): p. 1747-9.
18. Shiau, A.K., D. Barstad, P.M. Loria, L. Cheng, P.J. Kushner, D.A. Agard, and G.L. Greene, *The structural basis of estrogen receptor/coactivator recognition and the antagonism of this interaction by tamoxifen*. *Cell*, 1998. **95**(7): p. 927-37.
19. Watkins, R.E., P.R. Davis-Searles, M.H. Lambert, and M.R. Redinbo, *Coactivator binding promotes the specific interaction between ligand and the pregnane X receptor*. *J Mol Biol*, 2003. **331**(4): p. 815-28.
20. Shemshedini, L., J.W. Ji, C. Brou, P. Chambon, and H. Gronemeyer, *In vitro activity of the transcription activation functions of the progesterone receptor. Evidence for intermediary factors*. *J Biol Chem*, 1992. **267**(3): p. 1834-9.
21. Gill, G. and M. Ptashne, *Negative effect of the transcriptional activator GAL4*. *Nature*, 1988. **334**(6184): p. 721-4.
22. Meyer, M.E., H. Gronemeyer, B. Turcotte, M.T. Bocquel, D. Tasset, and P. Chambon, *Steroid hormone receptors compete for factors that mediate their enhancer function*. *Cell*, 1989. **57**(3): p. 433-42.

23. Onate, S.A., S.Y. Tsai, M.J. Tsai, and B.W. O'Malley, *Sequence and characterization of a coactivator for the steroid hormone receptor superfamily*. Science, 1995. **270**(5240): p. 1354-7.
24. Drexler, K.E., *Molecular engineering: An approach to the development of general capabilities for molecular manipulation*. Proc Natl Acad Sci U S A, 1981. **78**(9): p. 5275-5278.
25. Pabo, C., *Molecular technology. Designing proteins and peptides*. Nature, 1983. **301**(5897): p. 200.
26. DeGrado, W.F., *Proteins from scratch*. Science, 1997. **278**(5335): p. 80-1.
27. Dahiyat, B.I. and S.L. Mayo, *De novo protein design: fully automated sequence selection*. Science, 1997. **278**(5335): p. 82-7.
28. Dantas, G., B. Kuhlman, D. Callender, M. Wong, and D. Baker, *A large scale test of computational protein design: folding and stability of nine completely redesigned globular proteins*. J Mol Biol, 2003. **332**(2): p. 449-60.
29. Kuhlman, B., G. Dantas, G.C. Ireton, G. Varani, B.L. Stoddard, and D. Baker, *Design of a novel globular protein fold with atomic-level accuracy*. Science, 2003. **302**(5649): p. 1364-8.
30. Watkins, R.E., G.B. Wisely, L.B. Moore, J.L. Collins, M.H. Lambert, S.P. Williams, T.M. Willson, S.A. Kliewer, and M.R. Redinbo, *The human nuclear xenobiotic receptor PXR: structural determinants of directed promiscuity*. Science, 2001. **292**(5525): p. 2329-33.
31. Fujino, T., Y. Sato, M. Une, T. Kanayasu-Toyoda, T. Yamaguchi, K. Shudo, K. Inoue, and T. Nishimaki-Mogami, *In vitro farnesoid X receptor ligand sensor assay using surface plasmon resonance and based on ligand-induced coactivator association*. J Steroid Biochem Mol Biol, 2003. **87**(4-5): p. 247-52.

FIGURE LEGENDS

Figure 6.1 Crystal structure of the PXR LBD in complex with the cholesterol-lowering drug SR12813 and SRC-1 peptide. The ligand is shown in blue and the coactivator is shown in cyan.

Figure 6.2 Wild-type PXR is shown in purple and wild-type SRC-1 is shown in green. The L694F SRC-1 mutation is shown in yellow, while the I255L and L276A PXR mutations are shown in magenta. Residues in blue were subject to rotamer shifts.

Figure 6.3 The plot shows the association and dissociation of varying concentrations of wild-type PXR with bound SRC-1. A 1:1 Langmuir binding model was fit to this data to give a dissociation constant of 0.9 μM .

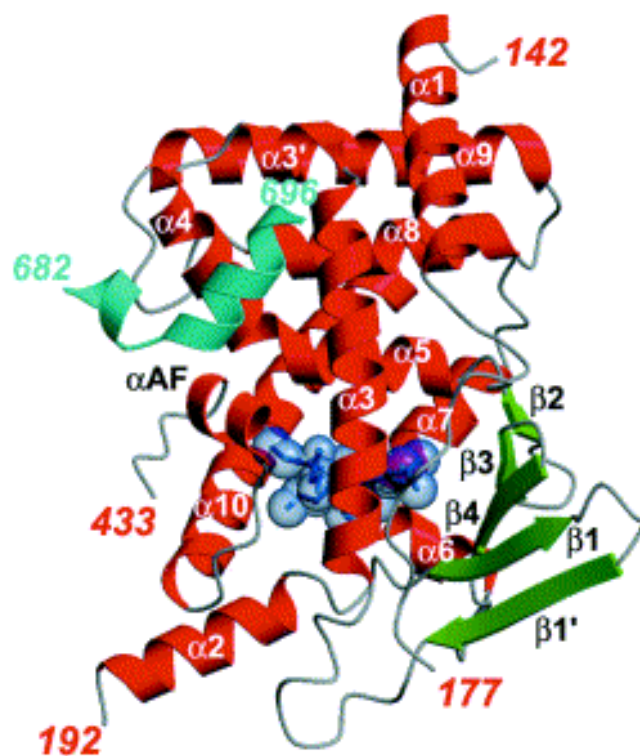


Figure 6.1 **Structure of PXR-SR12813-SRC-1 complex**

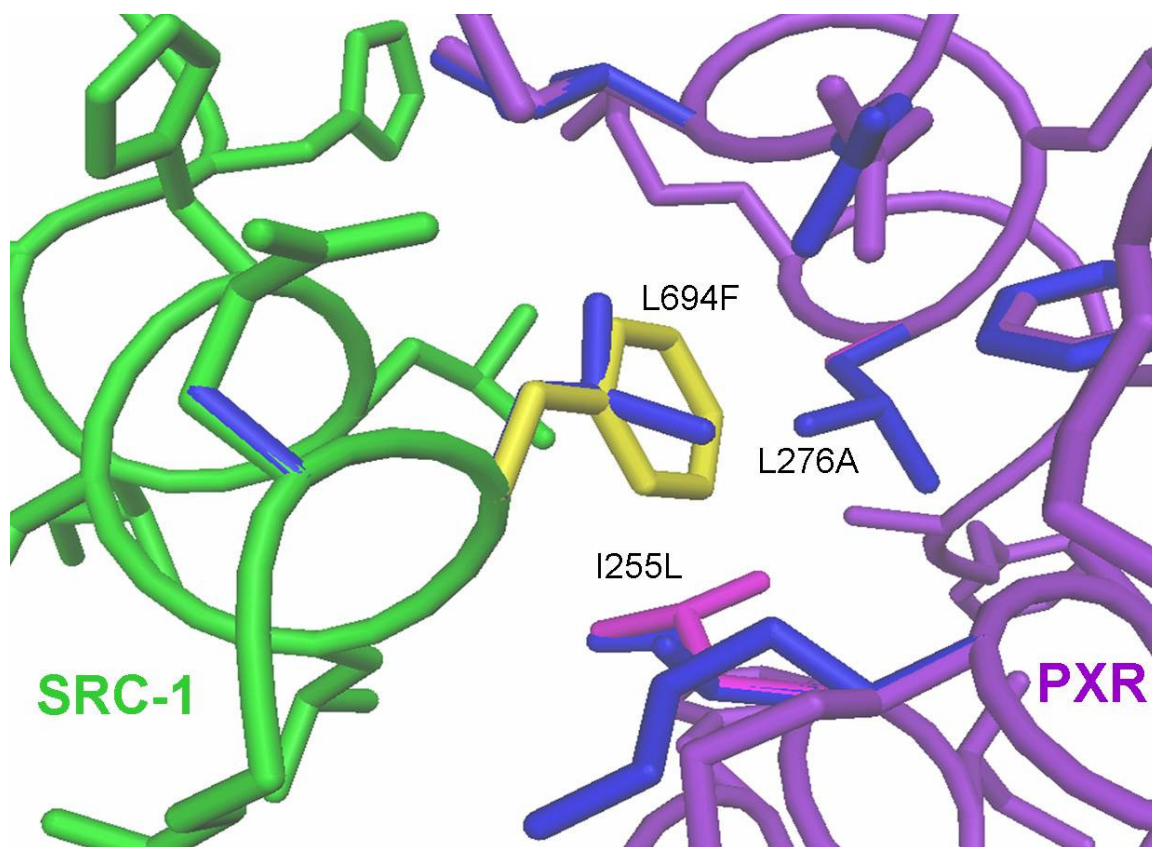


Figure 6.2 Overlay of wild-type PXR/SRC-1 complex with mutant

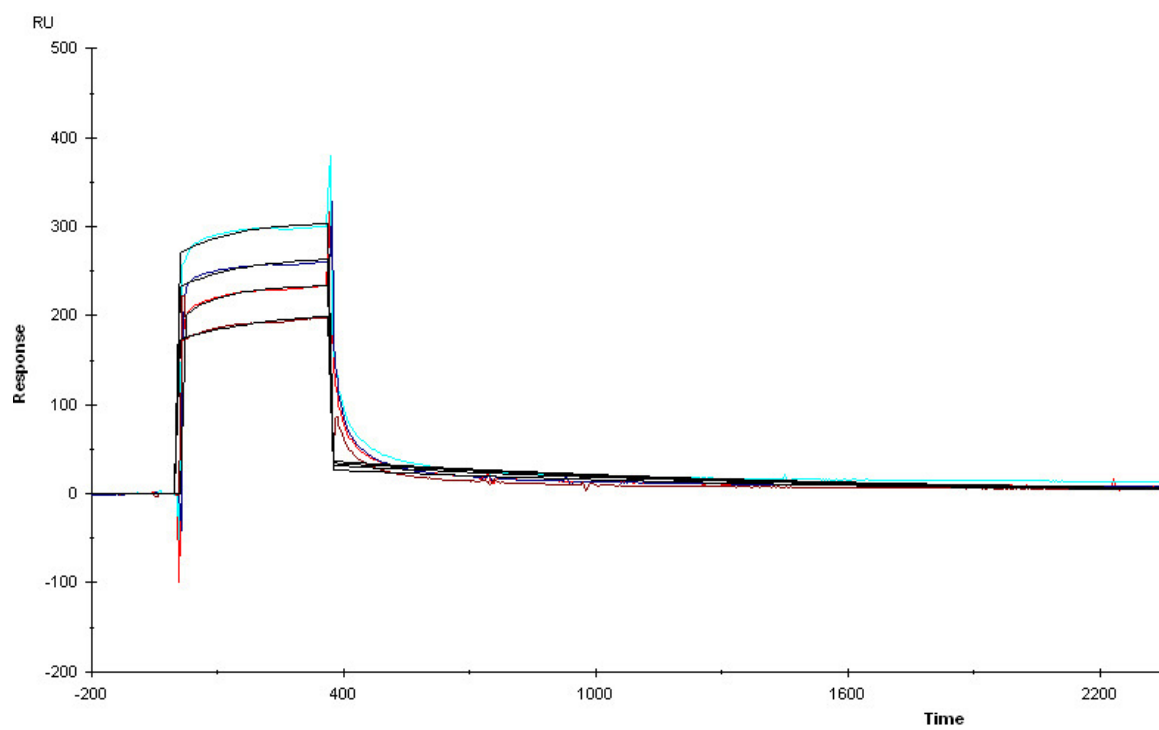


Figure 6.3 Response curve of wild-type PXR binding to SRC-1

	ΔE_{AB}	$\Delta E_{A'B'}$	$\Delta E_{A'B}$	$\Delta E_{AB'}$
Native Sequences	-14.7			
Design		-13.9	958.5	-9.2

Table 6.1 **Energy values of all four wild-type and mutant complexes.**